A precise bare simulation approach to the minimization of some distances. II. Further Foundations

Michel Broniatowski, Wolfgang Stummer

A precise bare simulation approach to the minimization of some distances. II. Further Foundations

Michel Broniatowski and Wolfgang Stummer

Abstract

The constrained minimization (respectively maximization) of directed distances and of related generalized entropies is a fundamental task in information theory as well as in the adjacent fields of statistics, machine learning, artificial intelligence, signal processing and pattern recognition. In our previous paper [1], we obtained such kind of constrained optima by a new dimension-free precise bare (pure) simulation method, provided basically that (i) the underlying directed distance is of $f-$divergence type, and that (ii) this can be connected to a light-tailed probability distribution in a certain manner. In the present paper, we extend this approach such that constrained optimization problems of a very huge amount of directed distances and generalized entropies — and beyond — can be tackled by a newly developed dimension-free *extended bare simulation* method, for obtaining both optima as well as optimizers. Almost no assumptions (like convexity) on the set of constraints are needed, within our discrete setup of arbitrary dimension, and our method is precise (i.e., converges in the limit). For instance, we cover constrained optimizations of *arbitrary* $f-$divergences, Bregman distances, scaled Bregman distances and weighted ℓ_r-distances. The potential for wide-spread applicability is indicated, too; in particular, we deliver many recent references for uses of the involved distances/divergences in various different research fields (which may also serve as an interdisciplinary interface).

Index Terms

f-divergences (of Csiszar-Ali-Silvey-Morimoto type), Bregman distances, scaled Bregman distances, Kullback-Leibler information distance, relative entropy, (density) power divergences, Tsallis (cross) entropies, Cressie-Read measures, Burbea-Rao divergences, weighted ℓ_r-distances, $\varphi-$entropies, minimum-distance estimators, generalized maximum entropy method, importance sampling.

1st February 2024

I. Introduction

In our previous paper [1], we developed a new random-simulation-based approach — called the *(narrow-sense) bare simulation method* — to obtain the deterministic constrained minimum $\inf\{\Phi_{\mathbf{P}}(\mathbf{Q}), \mathbf{Q} \in \mathbf{\Omega}\}$ of (generalized versions of) $\varphi-$*divergences* $\Phi_{\mathbf{P}}(\mathbf{Q}) = D_\varphi(\mathbf{Q}, \mathbf{P})$ (cf. Csiszar [2], Ali & Silvey [3], Morimoto [4]), where $\mathbf{P} \in \mathbb{R}^K$, $\mathbf{\Omega} \subset \mathbb{R}^K$ and $\varphi : \mathbb{R} \mapsto \mathbb{R}$ is a pregiven convex function with certain additional properties (namely, φ is appropriately linked to an instrumental probability distribution, under which the crucial auxiliary simulations are performed). In particular, amongst others, in [1] we covered (generalized versions of) the Kullback-Leibler information distance (relative entropy), the (squared) Hellinger distance, the Pearson chi-square divergence and the more general power divergences (also called Cressie-Read measures [5] [6], Tsallis cross-entropies [7], Amari's alpha-divergences [8]), and the Jensen-Shannon divergence/distance; some other important quantities which are closely connected to those $\varphi-$divergences — such as e.g. the Bhattacharyya coefficient (cf. [9],[10],[11]), the Bhattacharyya distance [9] and the more general Renyi divergences [12] (see also the comprehensive exposition in van Erven & Harremoës [13]), as well as the maximization/minimization of some $\varphi-$entropies (cf. Burbea & Rao [14]) such as e.g. the Shannon entropy [15], the Renyi entropies [12], the Havrda-Charvat entropies [16] (also called Tsallis entropies [17]) and of some Euclidean norms — were treated in [1], too.

Let us briefly recall the *core steps* of our bare-simulation minimization method in [1]. The *first step* is to normalize the vector [1] $\mathbf{P}$ into a probability vector $\mathbf{P}$ [2] (e.g. the $\varphi-$entropy triggering case $\mathbf{P} = (1, \ldots, 1)$ is converted into the uniform-probability vector $\mathbf{P} = (1/K, ..., 1/K)$). The *second step* follows from expressing the function φ in form of the Fenchel-Legendre transform of the cumulant (i.e., log moment) generating function of some random variable W; a probabilistic construction based on i.i.d. copies W_i of W allows to interpret $\inf\{D_\varphi(\mathbf{Q}, \mathbf{P}), \mathbf{Q} \in \mathbf{\Omega}\}$ as an asymptotic characteristic for some explicitly constructable scheme involving both $\mathbf{P}$ and the W_i's. The *third and final step* consists in the construction of this probabilistic scheme, and it differs for the specific problem context.

M. Broniatowski is with the LPSM, Sorbonne Université, 4 place Jussieu, 75252 Paris, France. ORCID 0000-0001-6301-5531.

W. Stummer is with the Department of Mathematics, Friedrich-Alexander-Universität Erlangen–Nürnberg (FAU), Cauerstrasse 11, 91058 Erlangen, Germany; e-mail: stummer@math.fau.de . ORCID 0000-0002-7831-4558. Corresponding author.

[1] in this paper, vectors are taken to be row vectors

[2] with a slight abuse of notation; see the main text for a more comprehensive notation

More detailed, for a deterministic setup where the (transformed) probability vector $\mathbf{P} = (p_1, \ldots, p_K)$ is completely known and $\mathbf{\Omega}$ has non-void interior, we construct in [1] the integer part $n_i := \lfloor np_i \rfloor$, partition the index set $\{1, \ldots, n\}$ into K sets of size $n_1, \ldots, n_K$ and build a K-component vector; *each component* of this vector is an ad hoc weighted empirical mean of the W_i 's ; up to standard transformations the empirical count of the visits of this vector in $\mathbf{\Omega}$ approximates the solution of the optimization problem $\inf \{D_\varphi(\mathbf{Q}, \mathbf{P}), \mathbf{Q} \in \mathbf{\Omega}\}$. Therefore, the resulting approximation can be performed straightforwardly: the (typically) very complicated minimization task is replaced by a much more comfortable — nevertheless convergent — random count procedure which can be based on a fast and accurate — pseudo, true, natural, quantum — random number generator. In case of a statistical/risk optimizing context, one has instead of a *known* probability vector $\mathbf{P}$ a data-describing *sample* $X_1, \ldots, X_n$ of n i.i.d. (and even more general) copies of a discrete random variable X with *unknown* distribution (described by an unknown probability vector) $\mathbf{P}$, and $\mathbf{\Omega}$ is now a subset of the probability simplex $\mathbb{S}^K$ in $\mathbb{R}^K$. For such contexts, in [1] we appropriately adapted the above-described method, by — amongst other things — basically using the corresponding (vectorized) sample-based histogram as $\mathbf{P}$.

Although [1] covers a considerable amount of important $\varphi-$divergences, as shown by numerous solved cases, the question arises whether the above-described bare-simulation method can be extended to achieve the minimum $\inf \{\Phi_{\mathbf{P}}(\mathbf{Q}), \mathbf{Q} \in \mathbf{\Omega}\}$ of other important classes $\Phi_{\mathbf{P}}(\mathbf{Q}) = D(\mathbf{Q}, \mathbf{P})$ of directed (i.e. non-symmetric) distances, divergences and (dis)similarity measures. To give the corresponding positive answer, is the main goal of this paper. Moreover, as opposed to [1], not only the minimum itself but also the corresponding minimizer(s) will be treated as well. Furthermore, we also tackle the maximum $\sup \{\Phi_{\mathbf{P}}(\mathbf{Q}), \mathbf{Q} \in \mathbf{\Omega}\}$ as well as the corresponding maximizer(s).

Indeed, in this paper we particularly investigate the following important classes of directed distances (divergences, dissimilarity measures) and connected entropies:

- general $\varphi-$divergences $\Phi_{\mathbf{P}}(\mathbf{Q}) := D_\varphi(\mathbf{Q}, \mathbf{P})$ where φ *does not satisfy* the above-mentioned Fenchel-Legendre-transform respresentability. For instance, this covers the omnipresent ℓ_1-distance (also known as *total variation distance*) between $\mathbf{Q}$ and $\mathbf{P}$. For some comprehensive overviews on $\varphi-$divergences, the reader is referred to the insightful books of e.g. Liese & Vajda [18], Read & Cressie [6], Vajda [19], Csiszar & Shields [20], Stummer [21], Pardo [22], Liese & Miescke [23], Basu et al. [24], the survey articles of e.g. Liese & Vajda [25], Vajda & van der Meulen [26], Reid & Williamson [27], Basseville [28], and the references therein. Moreover, we are interested in the very important special case $\mathbf{P} = (1, \ldots, 1) =: \mathbf{1}$ which leads to the corresponding general *non-probability-vector extension* of the $\varphi-$*entropies* in the sense of Burbea & Rao [14] (see also Csiszar [29], Ben-Bassat [30], Ben-Tal & Teboulle [31], Kesavan & Kapur [32], Dacunha-Castelle & Gamboa [33], Teboulle & Vajda [34], Gamboa & Gassiat [35], Vajda & Zvarova [36]). Those general $\varphi-$entropies can also be interpreted as an "index/degree of (in)equality of the set $\mathbf{\Omega}$", respectively as an "index/degree of diversity of the set $\mathbf{\Omega}$". We can also deal with the corresponding general *non-probability-vector extension* of the even wider class of $(h, \varphi)-$entropies in the sense of Salicru et al. [37] (see also e.g. Pardo [22], Vajda & Vasek [38]).

- *separable (ordinary) Bregman distances* (cf. [39]) $\Phi_{\mathbf{P}}(\mathbf{Q}) := D_\varphi^{OBD}(\mathbf{Q}, \mathbf{P})$ with strictly convex function $\varphi : \mathbb{R} \mapsto \mathbb{R}$. Aside from the vast applications in engineering, some general connections to probability and statistics are given e.g. in Csiszár [40], [41], [42], Pardo & Vajda [43], [44], Stummer & Vajda [45], and Broniatowski & Stummer [46],[47]. Important special cases include (i) the omnipresent *squared* ℓ_2-*distance*, (ii) the more general *density power divergences* of Basu et al. [48] (see also e.g. Stummer & Vajda [45] for a different parameter scaling which covers the prominent *Itakura-Saito distance* (in the version of e.g. Banerjee et al. [49])), as well as (iii) their rescaled versions called *beta-divergences* (cf. Eguchi & Kano [50], Mihoko & Eguchi [51]). Another recently launched special case is the *Bregman exponential divergence* of Mukherjee et al. [52] (see also Basak & Basu [53]).

- *scaled Bregman distances* (with scaling vector $\mathbf{M}$) $\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\varphi,\mathbf{M}}^{SBD}(\mathbf{Q}, \mathbf{P})$ of Stummer [54] and Stummer & Vajda [45], which cover both $\varphi-$divergences as well as separable (ordinary) Bregman distances. Further investigations (e.g. on robustness issues) can be found in Kißlinger & Stummer [55], [56], [57], [58]; Broniatowski & Stummer [46] flexibilized/widened the involved domains, and Broniatowski & Stummer [47] give a comprehensive survey on various different kinds of applications to statistics, and to the adjacent fields of machine learning and artificial intelligence. Moreover, Stummer & Kißlinger [59] give some structural flexibilizations/generalizations of scaled Bregman distances[3] which cover as *special cases* (the separable versions of) (i) the total Bregman divergences of Liu et al. [60],[61], Vemuri et al. [62], (ii) its variant given in Nock et al. [63], as well as (iii) the conformal divergences and the scaled conformal divergences of Nock et al. [64]. Furthermore, we can also deal with the even wider class of the *distances of Broniatowski & Stummer [46]* (see also Broniatowski & Stummer [47]), which cover all the above-mentioned (generalizations of) scaled Bregman distances as special cases.

[3](even for non-convex φ)

- the very prominent *non-separable (ordinary) Bregman distances* (cf. [39]) $\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\boldsymbol{\varphi}}^{gnOBD}(\mathbf{Q},\mathbf{P})$ with strictly convex *multivariate* function $\boldsymbol{\varphi} : \mathbb{R}^K \mapsto \mathbb{R}$ (as usual, if $\boldsymbol{\varphi}$ is of the separable form $\boldsymbol{\varphi}(\mathbf{Q}) := \sum_{k=1}^{K} \varphi(q_k)$, then $D_{\boldsymbol{\varphi}}^{gnOBD}(\mathbf{Q},\mathbf{P})$ collapses to the separable (ordinary) Bregman distance $D_{\varphi}^{OBD}(\mathbf{Q},\mathbf{P})$). A very important special case is the omnipresent (squared) *Mahalanobis distance* [65].

- the omnipresent *weighted ℓ_r-distances* $\Phi_{\mathbf{P}}(\mathbf{Q}) := ||\mathbf{Q}-\mathbf{P}||_{r,\mathbf{w}}$ where $r \in]0,\infty[$ and $\mathbf{w}$ is a vector of weights.

- *Burbea-Rao divergences* [14] (see also e.g. Pardo & Vajda [43],[44], as well as Stummer & Kißlinger [59] for an imbedding into their scaled-Bregman-distance-flexibilizations).

In the light of the above explanations, the goals of this paper are:

(G1) to *extend* the results of [1] on — narrow sense — bare-simulation minimization $\inf\{D_{\varphi}(\mathbf{Q},\mathbf{P}), \mathbf{Q} \in \boldsymbol{\Omega}\}$ of $\varphi-$divergences with instrumentally linked divergence generator φ, *to* the — narrow sense — bare-simulation minimization of the wider class of scaled Bregman distances $D_{\varphi,\mathbf{M}}^{SBD}(\mathbf{Q},\mathbf{P})$ (including separable Bregman distances $D_{\varphi}^{OBD}(\mathbf{Q},\mathbf{P})$) with the same type of divergence generator φ;
(G2) to solve — by means of a newly developed *extension* of our narrow-sense bare-simulation method — both constrained *minimization and maximization problems* for the above-mentioned huge range of directed distances, divergences, (dis-)similarity measures, entropies (and beyond); our new method is precise (i.e., converges in the limit) and needs almost no assumptions (like convexity) on the set $\boldsymbol{\Omega}$ of constraints of arbitrary finite dimension.
(G3) to deliver *sharp* estimates of the desired *minima and maxima*, and to derive the left-open estimates of the *minimizers* for the context of [1], as well as the of the minimizers and maximizers for the contexts (G1) and (G2).

This agenda is achieved in the following way. In the next Section II, we briefly introduce the principal idea of our new *extended* bare-simulation optimization paradigm, in addition to the — now called *narrow-sense* — one of [1]. In order to lay a solid explanatory basis for our new developments, we first recall in Section III our main results of [1] on narrow-sense bare-simulation minimization of $\varphi-$divergences with instrumentally linked divergence generator φ, for "general" constraints sets $\boldsymbol{\Omega}$ with non-void interior. For the same type of φ and $\boldsymbol{\Omega}$, in Section IV we carry out the above-mentioned goal (G1) on narrow-sense bare-simulation minimization of separable Bregman distances and even more general scaled Bregman distances. Based on the results of the previous two chapters, we then achieve in Section V the goal (G2) and derive four fundamental (non-narrow-sense) bare-simulation minimization and maximization results on all the above-mentioned general directed distances $D(\mathbf{Q},\mathbf{P})$ and friends, for the context of $\boldsymbol{\Omega}$ with non-void interior. In the next three Sections VI, VII, VIII we carry out the same program as in the Sections III, IV, V, but for the case that $\boldsymbol{\Omega}$ contains the side constraint that for each member $\mathbf{Q}$ the sum of the components equals the same fixed constant $A > 0$ (implying that $\boldsymbol{\Omega}$ has void interior). Furthermore, in the Sections IX, X, XI we carry out the same program as in the Sections III, IV, V, but for the risk-carrying case that $\mathbf{P}$ — respectively some involved parameter — is unknown (and $A = 1$). In the Sections XII, XIII and XIV we provide corresponding estimators for the minima, minimizers, maxima and maximizers of the above-mentioned sections (cf. Goal (G3)). Finally, all the proofs are given Appendix A.

II. A NEW MINIMIZATION PARADIGM

We concern with minimization and maximization problems of the following type, where $\mathcal{M}$ is a topological space and $\mathcal{T}$ is the Borel $\sigma-$field over a given base on $\mathcal{M}$; e.g. take $\mathcal{M} = \mathbb{R}^K$ to be the $K-$dimensional Euclidean space equipped with the Borel $\sigma-$ field $\mathcal{T}$.

Definition 1: A measurable function $\Phi : \mathcal{M} \mapsto \mathbb{R} \cup \{-\infty,\infty\}$ and measurable set $\Omega \subset \mathcal{M}$ [4] are called "bare-simulation minimizable" (BS-minimizable) respectively "bare-simulation maximizable" (BS-maximizable) if for

$$\Phi(\Omega) := \inf_{Q\in\Omega}\{\Phi(Q)\} \in]-\infty,\infty[\qquad \text{respectively} \qquad \Phi(\Omega) := \sup_{Q\in\Omega}\{\Phi(Q)\} \in]-\infty,\infty[\tag{1}$$

there exists a measurable function $G : [0,\infty[\mapsto \mathbb{R}$, a sequence $(f_n)_{n\in\mathbb{N}}$ of measurable functions $f_n : \mathcal{M} \mapsto [0,\infty[$ as well as a sequence $((\mathfrak{X}_n,\mathcal{A}_n,\mathbb{\Pi}_n))_{n\in\mathbb{N}}$ of probability spaces and on them a sequence $(\xi_n)_{n\in\mathbb{N}}$ [5] of $\mathcal{M}-$valued random variables such that

$$G\Big(-\lim_{n\to\infty}\frac{1}{n}\log \mathbb{E}_{\mathbb{\Pi}_n}\Big[f_n(\xi_n)\cdot 1_{\Omega}(\xi_n)\Big]\Big) = \inf_{Q\in\Omega}\Phi(Q) = \Phi(\Omega) \tag{2}$$

$$\text{respectively} \quad G\Big(-\lim_{n\to\infty}\frac{1}{n}\log \mathbb{E}_{\mathbb{\Pi}_n}\Big[f_n(\xi_n)\cdot 1_{\Omega}(\xi_n)\Big]\Big) = \sup_{Q\in\Omega}\Phi(Q) = \Phi(\Omega), \tag{3}$$

[4] i.e. $\Omega \in \mathcal{T}$

[5] in order to emphasize the dependence on Φ, one should use the notations $(\xi_{\Phi,n})_{n\in\mathbb{N}}$, $\mathbb{\Pi}_{\Phi,n}$, etc.; this is avoided for the sake of a better readability.

where $\mathbb{E}_{\mathbb{\Pi}_n}[\,\cdot\,]$ denotes the expectation with respect to $\mathbb{\Pi}_n$ and $1_B(\cdot)$ denotes the indicator function on the set B; in situations where Φ is fixed and different Ω's are considered, we say that "Φ is bare-simulation minimizable (BS-minimizable) on Ω" respectively "Φ is bare-simulation maximizable (BS-maximizable) on Ω". In case that one can even choose $f_n(\cdot) \equiv 1$ — and hence $\mathbb{E}_{\mathbb{\Pi}_n}\big[\, f_n(\xi_n) \cdot 1_\Omega(\xi_n)\,\big] = \mathbb{\Pi}_n\big[\,\xi_n \in \Omega\,\big]$ — then we speak of "bare-simulation minimizable/maximizable *in the narrow sense*".

Remark 2: (a) The above-mentioned Definition 1 extends the Definition 1 of Browniatowski & Stummer [1] who deal with the narrow-sense-case $f_n(\cdot) \equiv 1$.
(b) We could even extend the above-mentioned Definition 1 to allow for more general $f_n : \mathcal{M} \mapsto \mathbb{R}$ such that $\mathbb{E}_{\mathbb{\Pi}_n}\Big[\, f_n(\xi_n) \cdot 1_\Omega(\xi_n)\,\Big] \geq 0$ for all large enough $n \in \mathbb{N}$.
(c) As usual, we call Ω the *constraint set* (alternatively used names are e.g. *choice set* or *search space*).

The basic idea/incentive of this new approach is: if a minimization problem (1) has no explicit solution and is computationally intractable (or unfeasible) but can be shown to be BS-minimizable with concretely constructable G, $(f_n)_{n\in\mathbb{N}}$, $(\xi_n)_{n\in\mathbb{N}}$ and $(\mathbb{\Pi}_n)_{n\in\mathbb{N}}$, then one can basically simulate the log-expectations $-\frac{1}{n}\log \mathbb{E}_{\mathbb{\Pi}_n}\big[\, f_n(\xi_n) \cdot 1_\Omega(\xi_n)\,\big]$ for large enough integer $n \in \mathbb{N}$ to obtain an approximation of the minimum/maximum (1) without having to evaluate the corresponding (not necessarily unique) minimizer/maximizer. As explained in [1], this is for instance important for fast and efficient model search. However, in contrast to [1], we show in this paper also how one can "nearly synchronously" achieve an approximation of the corresponding minimizer(s)/maximizer(s).

For reasons of transparency, we *start* to demonstrate this approach for the following important/prominent class of *deterministic* constrained minimization problems with the following components:

(i) $\mathcal{M}$ is the $K-$dimensional Euclidean space $\mathbb{R}^K$, i.e. Ω is a set of vectors Q with a number of K components (where K may be huge, as it is e.g. the case in big data contexts);
(ii) $\Phi(\cdot) := \Phi_P(\cdot)$ depends on some known vector P in $\mathbb{R}^K$ with K nonnegative components;
(iii) $\Phi_P(\cdot)$ is a "directed distance" (divergence) from P into Ω in the sense of $\Omega \ni Q \mapsto \Phi_P(Q) := D(Q,P)$, where $D(\cdot,\cdot)$ has the the two properties "$D(Q,P) \geq 0$" and "$D(Q,P) = 0$ if and only if $Q = P$". In particular, $D(\cdot,\cdot)$ needs neither satisfy the symmetry $D(Q,P) = D(P,Q)$ nor the triangular inequality.

In other words, the left-hand part of (1) together with (i)-(iii) constitutes a *deterministic* constrained distance/divergence-minimization problem; we design a "universal" method to solve such problems by constructing appropriate (cf.(2)) G, $(f_n)_{n\in\mathbb{N}}$ and sequences $(\xi_n)_{n\in\mathbb{N}}$ of $\mathbb{R}^K-$valued random variables. The latter will be first constructed — for a first insight — with the help of *narrow-sense methods* developed in [1] for directed distances $D(\cdot,\cdot)$ from a large subclass of the important omnipresent Csiszar-Ali-Silvey-Morimoto $\varphi-$divergences (also called $f-$divergences) given in Definition 3 below.

III. Deterministic Narrow-Sense Bare-Simulation-Optimization of $\varphi-$divergences

To begin with, concerning the above-mentioned point (i) we take the $K-$ dimensional Euclidean space $\mathcal{M} = \mathbb{R}^K$, denote from now on — as usual — its elements (i.e. vectors) in boldface letters, and also employ the subsets

$$\mathbb{R}^K_{\neq 0} := \{\mathbf{Q} := (q_1,\ldots,q_K) \in \mathbb{R}^K : q_i \neq 0 \text{ for all } i = 1,\ldots,K\},$$
$$\mathbb{R}^K_{>0} := \{\mathbf{Q} := (q_1,\ldots,q_K) \in \mathbb{R}^K : q_i > 0 \text{ for all } i = 1,\ldots,K\},$$
$$\mathbb{R}^K_{\geq 0} := \{\mathbf{Q} := (q_1,\ldots,q_K) \in \mathbb{R}^K : q_i \geq 0 \text{ for all } i = 1,\ldots,K\},$$
$$\mathbb{R}^K_{\gneq 0} := \mathbb{R}^K_{\geq 0}\backslash\{\mathbf{0}\} := \{\mathbf{Q} := (q_1,\ldots,q_K) \in \mathbb{R}^K_{\geq 0} : q_i \neq 0 \text{ for some } i = 1,\ldots,K\},$$
$$\mathbb{S}^K := \{\mathbf{Q} := (q_1,\ldots,q_K) \in \mathbb{R}^K_{\geq 0} : \textstyle\sum_{i=1}^K q_i = 1\} \quad \text{(simplex of probability vectors, probability simplex)},$$
$$\mathbb{S}^K_{>0} := \{\mathbf{Q} := (q_1,\ldots,q_K) \in \mathbb{R}^K_{>0} : \textstyle\sum_{i=1}^K q_i = 1\}.$$

Concerning the directed distances $D(\cdot,\cdot)$ in (ii) and (iii), as a basis we *first* deal with the following

Definition 3: (a) Let the "divergence-generator" be a lower semicontinuous convex function $\varphi :]-\infty,\infty[\to [0,\infty]$ satisfying $\varphi(1) = 0$. Furthermore, for the effective domain $dom(\varphi) := \{t \in \mathbb{R} : \varphi(t) < \infty\}$ we assume that its interior $int(dom(\varphi))$ is non-empty which implies that $int(dom(\varphi)) =]a,b[$ for some $-\infty \leq a < 1 < b \leq \infty$. Moreover, we suppose that φ is strictly convex at the point $t = 1$ [6] (very often in practice, φ is strictly convex even in a non-empty neighborhood $]t^{sc}_-, t^{sc}_+[\subseteq]a,b[$ of one ($t^{sc}_- < 1 < t^{sc}_+$)). Also, we set $\varphi(a) := \lim_{t\downarrow a}\varphi(t)$ and $\varphi(b) := \lim_{t\uparrow b}\varphi(t)$ (these limits always exist). The class of all such functions φ will be denoted by $\widetilde{\Upsilon}(]a,b[)$. A frequent choice is e.g. $]a,b[=]0,\infty[$ or $]a,b[=]-\infty,\infty[$.

[6] in line with e.g. Liese & Miescke [23], here a convex function φ is called strictly convex at the point $t = 1$ if the function φ is not linear in the open interval $]1-\varepsilon, 1+\varepsilon[$ for any $\varepsilon > 0$

(b) For $\varphi \in \widetilde{\Upsilon}(]a,b[)$, $\mathbf{P} := (p_1,\ldots,p_K) \in \mathbb{R}^K_{\geq 0}$ and $\mathbf{Q} := (q_1,\ldots,q_K) \in \mathbf{\Omega} \subset \mathbb{R}^K$, we define the Csiszar-Ali-Silvey-Morimoto (CASM) $\varphi-$divergence

$$\Phi_{\mathbf{P}}(\mathbf{Q}) := D_\varphi(\mathbf{Q},\mathbf{P}) := \sum_{k=1}^K p_k \cdot \varphi\left(\frac{q_k}{p_k}\right) \geq 0. \tag{4}$$

As usual, in (4) we employ the three conventions that $p \cdot \varphi\left(\frac{0}{p}\right) = p \cdot \varphi(0) > 0$ for all $p > 0$, and $0 \cdot \varphi\left(\frac{q}{0}\right) = q \cdot \lim_{x\to\infty} \frac{\varphi(x\cdot \mathrm{sgn}(q))}{x\cdot \mathrm{sgn}(q)} > 0$ for $q \neq 0$ (employing the sign of q), and $0 \cdot \varphi\left(\frac{0}{0}\right) := 0$. Throughout the paper, we only consider constellations $(\varphi, \mathbf{P}, \mathbf{\Omega})$ for which the very mild condition $\Phi_{\mathbf{P}}(\Omega) := \inf_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q},\mathbf{P}) \neq \infty$ holds.

For probability vectors $\mathbb{P}$ and $\mathbb{Q}$ in $\mathbb{S}^K$, the $\varphi-$divergences $D_\varphi(\mathbb{Q},\mathbb{P})$ were introduced by Csiszar [2], Ali & Silvey [3] and Morimoto [4] (where the first two references even deal with more general probability distributions); for some comprehensive overviews — including statistical applications to goodness-of-fit testing and minimum distance estimation — the reader is referred to the insightful books the reader is referred to the insightful books of e.g. Liese & Vajda [18], Read & Cressie [6], Vajda [19], Csiszar & Shields [20], Stummer [21], Pardo [22], Liese & Miescke [23], Basu et al. [24], the survey articles of e.g. Liese & Vajda [25], Vajda & van der Meulen [26], Reid & Williamson [27], Basseville [28], and the references therein; For the setup of $D_\varphi(\mathbf{Q},\mathbf{P})$ for vectors $\mathbf{P}$, $\mathbf{Q}$ with non-negative components the reader is referred to e.g. Stummer & Vajda [66] (who deal with even more general nonnegative measures and give some statistical as well as information-theoretic applications) and Gietl & Reffel [67]. The case of $\varphi-$divergences for vectors with arbitrary components can be extracted from e.g. Broniatowski & Keziou [68] who actually deal with finite *signed* measures. For a comprehensive technical treatment, see also Browniatowski & Stummer [46].

As an important special case, we get for the choice $\mathbf{P} := (1,\ldots,1) := \mathbf{1}$ the quantity

$$\Phi_{\mathbf{1}}(\mathbf{Q}) := D_\varphi(\mathbf{Q},\mathbf{1}) = \sum_{k=1}^K \varphi(q_k) \tag{5}$$

with $\varphi \in \widetilde{\Upsilon}(]a,b[)$. As is well known, there are numerous applications of $\sum_{k=1}^K \varphi(q_k)$ where φ is e.g. interpreted as cost function respectively energy function respectively purpose function. Furthermore, $\sum_{k=1}^K \varphi(q_k)$ can be interpreted as (non-probability extension of a) $\varphi-$entropy in the sense of Burbea & Rao [14] (see also Csiszar [29], Ben-Bassat [30], Ben-Tal & Teboulle [31], Kesavan & Kapur [32], Dacunha-Castelle & Gamboa [33], Teboulle & Vajda [34], Gamboa & Gassiat [35], Vajda & Zvarova [36]). Moreover, since $\mathbf{1}$ can be seen as a reference vector with (normalized) equal components, $D_\varphi(\mathbf{Q},\mathbf{1})$ in (5) can be interpreted as an "index/degree of (in)equality of the set $\mathbf{\Omega}$", respectively as an "index/degree of diversity of the set $\mathbf{\Omega}$". A comprehensive BS-concerning discussion with references on the theory and applications of the quantities in (5), is given e.g. in Broniatowski & Stummer [1].

Returning to the general case, from (4) it is obvious that in general $D_\varphi(\mathbf{Q},\mathbf{P}) \neq D_\varphi(\mathbf{P},\mathbf{Q})$ (non-symmetry). Moreover, it is straightforward to deduce that $D_\varphi(\mathbf{Q},\mathbf{P}) = 0$ if and only if $\mathbf{Q} = \mathbf{P}$ (reflexivity). By appropriate choice of φ, one can get as special cases many very prominent divergences which are frequently used in information theory and its applications to e.g. statistics, artificial intelligence, and machine learning.

For reasons of a more compact representation, we shall henceforth assume that $\mathbf{P} := (p_1,\ldots,p_K) \in \mathbb{R}^K_{>0}$, unless stated otherwise.

As a fundamental tool for later purposes, let us now briefly explain how the *BS method in the narrow sense* of Broniatowski & Stummer [1] can be used to tackle the following deterministic optimization problems:

Problem 4: For pregiven $\varphi \in \widetilde{\Upsilon}(]a,b[)$, positive-components vector $\mathbf{P} := (p_1,..,p_K) \in \mathbb{R}^K_{>0}$ (or from some subset thereof), and subset $\mathbf{\Omega} \subset \mathbb{R}^K$ (also denoted in boldface letters, with a slight abuse of notation) with regularity properties

$$cl(\mathbf{\Omega}) = cl\left(int\left(\mathbf{\Omega}\right)\right), \qquad int\left(\mathbf{\Omega}\right) \neq \emptyset, \tag{6}$$

find

$$\Phi_{\mathbf{P}}(\mathbf{\Omega}) := \inf_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q},\mathbf{P}), \tag{7}$$

provided that

$$\inf_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q},\mathbf{P}) < \infty \tag{8}$$

and that divergence generator φ additionally satisfies the following Condition 5.

Condition 5: Let $\varphi \in \widetilde{\Upsilon}(]a,b[)$ and $M_{\mathbf{P}} = \sum_{i=1}^K p_i > 0$. Then the multiple $\widetilde{\varphi} := M_{\mathbf{P}} \cdot \varphi$ should satisfy the representation

$$\widetilde{\varphi}(t) = \sup_{z\in\mathbb{R}} \Big(z \cdot t - \log \int\limits_{\mathbb{R}} e^{zy} d\widetilde{\zeta}(y) \Big), \qquad t \in \mathbb{R}, \tag{9}$$

for some probability distribution $\widetilde{\zeta}$ on the real line such that the function $z \mapsto MGF_{\widetilde{\zeta}}(z) := \int_{\mathbb{R}} e^{zy} d\widetilde{\zeta}(y)$ is finite on some open interval containing zero.

Remark 6: The change from φ to $\widetilde{\varphi} := M_{\mathbf{P}} \cdot \varphi$ in Condition 5 stems from the fact that one can equivalently rewrite (7) such that the vector $\mathbf{P}$ "turns into" a probability vector $\widetilde{\mathbb{P}}$; the latter will be essentially needed for our BS method (cf. Broniatowski & Stummer [1]). Indeed, one can construct the probability vector $\widetilde{\mathbb{P}} := \mathbf{P}/M_{\mathbf{P}}$ and analogously $\widetilde{\mathbf{Q}} := \mathbf{Q}/M_{\mathbf{P}}$ (notice that $\widetilde{\mathbf{Q}}$ may be a non-probability vector). With this, one can equivalently rewrite

$$D_{\varphi}(\mathbf{Q}, \mathbf{P}) = \sum_{k=1}^{K} p_k \cdot \varphi\Big(\frac{q_k}{p_k}\Big) = \sum_{k=1}^{K} M_{\mathbf{P}} \cdot \widetilde{p_k} \cdot \varphi\Big(\frac{M_{\mathbf{P}} \cdot \widetilde{q_k}}{M_{\mathbf{P}} \cdot \widetilde{p_k}}\Big) = D_{\widetilde{\varphi}}(\widetilde{\mathbf{Q}}, \widetilde{\mathbb{P}}). \tag{10}$$

and thus the solution of (7) coincides with the one of the problem of finding

$$\widetilde{\Phi}_{\widetilde{\mathbb{P}}}(\widetilde{\mathbf{\Omega}}) := \inf_{\widetilde{\mathbf{Q}} \in \widetilde{\mathbf{\Omega}}} D_{\widetilde{\varphi}}(\widetilde{\mathbf{Q}}, \widetilde{\mathbb{P}}), \qquad \text{with } \widetilde{\mathbf{\Omega}} := \mathbf{\Omega}/M_{\mathbf{P}}. \tag{11}$$

Remark 7: A comprehensive study on Condition 5 is given in Section XI of Broniatowski & Stummer [1] as well as in Broniatowski & Stummer [69]; numerous explicitly solved cases can be found in Section XII of [1]. In particular, Condition 5 implies in particular that
(i) φ is strictly convex in a non-empty neighborhood $]t_-^{sc}, t_+^{sc}[\subseteq]a, b[$ of one ($t_-^{sc} < 1 < t_+^{sc}$) and affine linear on the rest $]-\infty, \infty[\setminus]t_-^{sc}, t_+^{sc}[$ in case that this rest is non-empty,
(ii) $\varphi(t) > 0$ for all $t \in]a, b[\setminus \{1\}$,
(iii) φ is continuously differentiable on $]a, b[$; accordingly, we denote the corresponding derivative by φ',
(iv) $\varphi'(1) = 0$.

Remark 8: (a) The purpose of assumption (6) is to get rid of the $\limsup$ type and $\liminf$ type results in our below-mentioned "bare-simulation" approach and to obtain *limit*-statements which motivate our construction. In practice, it is enough to verify $\mathbf{\Omega} \subseteq cl\,(int\,(\mathbf{\Omega}))$, which is equivalent to the left-hand part of (6). Clearly, any open set $\mathbf{\Omega} \subset \mathbb{R}^K$ satisfies the left-hand part of (6). In the subsetup where $\mathbf{\Omega}$ is a closed convex set and $int(\mathbf{\Omega}) \neq \emptyset$, (6) is satisfied and the minimizer $\mathbf{Q}_{min} \in \mathbf{\Omega}$ in (7) is attained and even unique. When $\mathbf{\Omega}$ is open and satisfies (6), then the infimum in (7) exists but is reached at some generalized projection of $\mathbf{P}$ on $\mathbf{\Omega}$ (see Csiszar [70] for the Kullback-Leibler-information-distance case of probability measures, which extends to any $\varphi-$divergence in our framework).
(b) Without further mentioning, the regularity assumption (6) is supposed to hold in the *full* topology. Of course, $int\left(\mathbb{S}^K\right) = \emptyset$ and thus, for the important probability-vector setup $\mathbf{\Omega} \subset \mathbb{S}^K$ the assumption (6) is violated which requires extra refinements (cf. Section VI below). The same is needed for $\mathbf{\Omega} \subset A \cdot \mathbb{S}^K$ for some $A \neq 1$, since obviously $int\left(A \cdot \mathbb{S}^K\right) = \emptyset$; such a context appears naturally e.g. in connection with mass transportation problems and with distributed energy management (see e.g. Chapter IX of Broniatowski & Stummer [1] and the references therein).
(c) Our approach is predestined for *non- or semiparametric* models, see Broniatowski & Stummer [1] for a detailed discussion.
(d) The Condition 5 implies in particular that φ satisfies $\int_{\mathbb{R}} y d\widetilde{\zeta}(y) = 1$ and that $\widetilde{\zeta}$ has light tails; moreover, $\widetilde{\zeta}$ may depend on $M_{\mathbf{P}}$ in a highly non-trivial way. For details — including also methods for finding $\widetilde{\zeta}$ as well as numerous examples — the reader is referred to Broniatowski & Stummer [1].

Returning to the distance-minimizing Problem 4, we proceed by constructing an appropriate sequence $(\boldsymbol{\xi}_n)_{n \in \mathbb{N}}$ of $\mathbb{R}^K-$valued random variables (cf. (2) in Definition 1 and the special case of Remark 2(a)) as follows: we first transform $\widetilde{\mathbb{P}} := \mathbf{P}/M_{\mathbf{P}}$ having components $(\widetilde{p}_1, \ldots, \widetilde{p}_K)$. Moreover, for any $n \in \mathbb{N}$ and any $k \in \{1, \ldots, K-1\}$, let $n_k := \lfloor n \cdot \widetilde{p}_k \rfloor$ (where $\lfloor x \rfloor$ denotes the integer part of x) and $n_K := n - \sum_{k=1}^{K-1} n_k$; for this, we assume that $n \in \mathbb{N}$ is large enough, namely $n \geq \max_{k \in \{1,\ldots,K\}} \frac{1}{\widetilde{p}_k}$, such that all the integers n_k ($k = 1, \ldots, K$) are non-zero. Since we assume $\mathbf{P} \in \mathbb{R}^K_{>0}$ and thus none of the $\widetilde{p}_k$'s is zero, one has

$$\lim_{n \to \infty} \frac{n_k}{n} = \widetilde{p}_k, \qquad k = 1, \ldots, K. \tag{12}$$

With this at hand, we decompose the set $\{1, \ldots, n\}$ of all integers from 1 to n into the following disjoint blocks: $I_1^{(n)} := \{1, \ldots, n_1\}$, $I_2^{(n)} := \{n_1 + 1, \ldots, n_1 + n_2\}$, and so on until the last block $I_K^{(n)} := \{\sum_{k=1}^{K-1} n_k + 1, \ldots, n\}$ which therefore contains all integers from $n_1 + \ldots + n_{K-1} + 1$ to n. Due to our construction, $I_k^{(n)}$ has $n_k \geq 1$ elements (i.e. $card(I_k^{(n)}) = n_k$ where $card(B)$ denotes the number of elements in a set B) for all $k \in \{1, \ldots, K\}$ [7]. Furthermore, consider a vector $\widetilde{\mathbf{W}} := \left(\widetilde{W}_1, \ldots, \widetilde{W}_n\right)$ where the $\widetilde{W}_i$'s are i.i.d. copies of the random variable $\widetilde{W}$ whose distribution is associated with the divergence-generator $\widetilde{\varphi} := M_{\mathbf{P}} \cdot \varphi$ through (9), in the sense that $\mathbb{\Pi}[\widetilde{W} \in \cdot\,] = \widetilde{\zeta}[\,\cdot\,]$ on some probability space $(\mathfrak{X}, \mathcal{A}, \mathbb{\Pi})$. We

[7] if all $\widetilde{p}_k$ ($k = 1, \ldots, K$) are rational numbers in $]0, 1[$ with $\sum_{k=1}^{K} \widetilde{p}_k = 1$ and N is the (always existing) smallest integer such that all $N \cdot \widetilde{p}_k$ ($k = 1, \ldots, K$) are integers (i.e. $\in \mathbb{N}$), then for any multiple $n = m \cdot N$ ($m \in \mathbb{N}$) one gets that all $n \cdot \widetilde{p}_k$ are integers and hence $n_k = \lfloor n \cdot \widetilde{p}_k \rfloor = n \cdot \widetilde{p}_k$ ($k = 1, \ldots, K$)

group the $\widetilde{W}_i$'s according to the above-mentioned blocks and sum them up blockwise, in order to build the following $K-$component random vector

$$\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}} := \Big(\frac{1}{n}\sum_{i\in I_1^{(n)}} \widetilde{W}_i, \ldots, \frac{1}{n}\sum_{i\in I_K^{(n)}} \widetilde{W}_i\Big); \tag{13}$$

notice that the signs of its components may be negative, depending on the nature of the $\widetilde{W}_i$'s; moreover, the expectation of its $k-$th component converges to $\widetilde{p}_k$ as n tends to infinity (since the expectation of $\widetilde{W}_1$ is 1), whereas the $n-$fold of the corresponding variance converges to $\widetilde{p}_k$ times the variance of $\widetilde{W}_1$.

For such a context, Broniatowski & Stummer [1] obtain the following assertion[8] :

Theorem 9: Let $\mathbf{P} \in \mathbb{R}^K_{>0}$, $M_{\mathbf{P}} := \sum_{i=1}^K p_i > 0$, and suppose that the divergence generator φ satisfies Condition 5 above, with $\widetilde{\zeta}$ (cf. (9)). Additionally, let $\widetilde{W} := (\widetilde{W}_i)_{i\in\mathbb{N}}$ be a sequence of random variables where the $\widetilde{W}_i$'s are i.i.d. copies of the random variable $\widetilde{W}$ whose distribution is $\mathbb{\Pi}[\widetilde{W} \in \cdot\,] = \widetilde{\zeta}[\,\cdot\,]$ [9]. Then, in terms of the random vectors $\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}$ (cf. (13)) there holds

$$\inf_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q},\mathbf{P}) \;=\; -\lim_{n\to\infty}\frac{1}{n}\log \mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}} \in \mathbf{\Omega}/M_{\mathbf{P}}\Big] \tag{14}$$

for any $\mathbf{\Omega} \subset \mathbb{R}^K$ with regularity properties (6) and finiteness property (8). In particular, for each $\mathbf{P} \in \mathbb{R}^K_{>0}$ the function $\Phi_{\mathbf{P}}(\cdot) := D_\varphi(\cdot,\mathbf{P})$ (cf. (4)) is bare-simulation minimizable (BS-minimizable) in the narrow sense (cf. (2) in Definition 1 and the special case of Remark 2(a)) on any such $\mathbf{\Omega} \subset \mathbb{R}^K$.

Remark 10: (i) For some contexts, we can *explicitly* give the distribution of each of the independent (non-deterministic parts of the) components $\Big(\sum_{i\in I_k^{(n)}} \widetilde{W}_i\Big)_{k=1,\ldots,K}$ of the vector $\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}$; this will ease the corresponding concrete simulations. For corresponding examples, see [1].
(ii) Let us emphasize that we have assumed $\mathbf{P} \in \mathbb{R}^K_{>0}$ in Theorem 9 which excludes $\mathbf{P}$ from having zero components. However, in cases where $\lim_{x\to\infty}\left|\frac{\varphi(x\cdot sgn(q))}{x\cdot sgn(q)}\right| = +\infty$ for $q \neq 0$, then if $p_{k_0} = 0$ for some k_0 it follows that $q_{k_0} = 0$, which proves that $\mathbf{P} \in \mathbb{R}^K_{>0}$ imposes no restriction in Theorem 9, since the projection of $\mathbf{P}$ in $\mathbf{\Omega}$ then belongs to the subspace of $\mathbb{R}^K$ generated by the non-null components of $\mathbf{P}$; such a situation appears e.g. for power divergence generators φ_γ with $\gamma > 2$. So there is no loss of generality assuming $\mathbf{P} \in \mathbb{R}^K_{>0}$ in this case.
(iii) Notice that (14) even holds when $\mathbf{P} \in \mathbf{\Omega}$ for which the left-hand side becomes zero.

Returning to the general context, the limit statement (14) provides the principle for the approximation of the solution of Problem (7). Indeed, by replacing the right-hand side in (14) by its finite counterpart, we deduce for given large n

$$-\frac{1}{n}\log \mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}} \in \mathbf{\Omega}/M_{\mathbf{P}}\Big] \approx \inf_{Q\in\mathbf{\Omega}} D_\varphi(\mathbf{Q},\mathbf{P}); \tag{15}$$

it remains to estimate the left-hand side of (15). The latter can be performed either by a *naive estimator* of the frequency of those replications of $\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}$ which hit $\mathbf{\Omega}/M_{\mathbf{P}}$, or more efficiently by some improved estimator; for details, the reader is referred to Section X of Broniatowski & Stummer [1], see also the corresponding extensions given in Section XII below, where the latter also provides e.g. estimates of the *minimizers*).

Remark 11: According to (14) of Theorem 9 as well as (15), we can principally tackle the (approximative) computation of the minimum value $\inf_{Q\in\mathbf{\Omega}} D_\varphi(\mathbf{Q},\mathbf{P}) = \inf_{Q\in\mathbf{\Omega}} \sum_{k=1}^K p_k \cdot \varphi\left(\frac{q_k}{p_k}\right)$ and in particular of $\inf_{\mathbf{Q}\in\mathbf{\Omega}} \sum_{k=1}^K \varphi(q_k) = \inf_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q},\mathbf{1})$ (cf. (5)) by basically *only employing a fast and accurate — pseudo, true, natural, quantum — random number generator*[10], provided that the constraint set $\mathbf{\Omega}$ satisfies the mild assumptions (6) and (8). Notice that (6) also covers constraint sets $\mathbf{\Omega}$ (of arbitrary dimension K) which are *non-convex* and even *highly disconnected*, and for which other minimization methods (e.g. pure enumeration, gradient or steepest descent methods, etc.) may be problematic or intractable.

Returning to the general context, notice that Theorem 9 does not cover cases where $\mathbf{\Omega}$ consists of $\mathbf{Q}$ satisfying the additional constraint $\sum_{i=1}^K q_i = A$ for some fixed $A > 0$ (and thus $int\,(\mathbf{\Omega}) = \emptyset$ which violates (6)). However, such situations can be still handled with an adaption of the above-described narrow-sense BS method, see Broniatowski & Stummer [1] and also the beginning of Section VIII below.

[8] with $(\mathfrak{X}_n, \mathcal{A}_n, \mathbb{\Pi}_n) = (\mathfrak{X}, \mathcal{A}, \mathbb{\Pi})$ for all $n \in \mathbb{N}$
[9] and thus, $E_{\mathbb{\Pi}}[\widetilde{W}_i] = 1$ and $\widetilde{W}_i$ has light tails
[10] for exemplary recent references on this very active research field on fast and accurate random number generation see e.g. [71]–[163]

To end up this section, let us present some examples of $\varphi-$divergence generators which satisfy Condition 5 and for which thus the Theorem 9 can be applied; the corresponding essential distributions $\widetilde{\mathfrak{l}}$ can be obtained from $\mathfrak{l}$ in the last column of Table 1 below, by employing $M_{\mathbf{P}} \cdot \varphi$ instead of φ. For details, construction methods and further examples, the reader is referred to Broniatowski & Stummer [1] (see also Broniatowski & Stummer [164] for applications to fuzzy sets and basic belief assignments).

Example 12: Let us take the important case of power-divergence generators $\varphi_\gamma : \mathbb{R} \mapsto [0,\infty]$ defined by

$$\varphi_\gamma(t) := \begin{cases} \frac{t^\gamma - \gamma\cdot t + \gamma - 1}{\gamma\cdot(\gamma-1)}, & \text{if } \gamma \in]-\infty,0[\text{ and } t \in]0,\infty[, \\ -\log t + t - 1, & \text{if } \gamma = 0 \text{ and } t \in]0,\infty[, \\ \frac{t^\gamma - \gamma\cdot t + \gamma - 1}{\gamma\cdot(\gamma-1)}, & \text{if } \gamma \in]0,1[\text{ and } t \in [0,\infty[, \\ t\cdot\log t + 1 - t, & \text{if } \gamma = 1 \text{ and } t \in [0,\infty[, \\ \frac{t^\gamma - \gamma\cdot t + \gamma - 1}{\gamma\cdot(\gamma-1)} \cdot 1_{]0,\infty[}(t) + (\frac{1}{\gamma} - \frac{t}{\gamma-1}) \cdot 1_{]-\infty,0]}(t), & \text{if } \gamma \in]1,2[\text{ and } t \in]-\infty,\infty[, \\ \frac{(t-1)^2}{2}, & \text{if } \gamma = 2 \text{ and } t \in]-\infty,\infty[, \\ \frac{t^\gamma - \gamma\cdot t + \gamma - 1}{\gamma\cdot(\gamma-1)} \cdot 1_{]0,\infty[}(t) + (\frac{1}{\gamma} - \frac{t}{\gamma-1}) \cdot 1_{]-\infty,0]}(t), & \text{if } \gamma \in]2,\infty[\text{ and } t \in]-\infty,\infty[, \\ \infty, & \text{else,} \end{cases} \tag{16}$$

which — by (4) — for arbitrary multiplier $\widetilde{c} \in]0,\infty[$ generate (the vector-valued form of) the *generalized power divergences* given by

$$D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P}) := \begin{cases} \widetilde{c}\cdot\Big\{ \frac{\sum_{k=1}^K (q_k)^\gamma \cdot (p_k)^{1-\gamma}}{\gamma\cdot(\gamma-1)} - \frac{1}{\gamma-1}\cdot\sum_{k=1}^K q_k + \frac{1}{\gamma}\cdot\sum_{k=1}^K p_k \Big\}, & \text{if } \gamma \in]-\infty,0[,\ \mathbf{P} \in \mathbb{R}^K_{\gneqq 0} \text{ and } \mathbf{Q} \in \mathbb{R}^K_{>0}, \\ \widetilde{c}\cdot\Big\{ \sum_{k=1}^K p_k \cdot \log\Big(\frac{p_k}{q_k}\Big) + \sum_{k=1}^K q_k - \sum_{k=1}^K p_k \Big\}, & \text{if } \gamma = 0,\ \mathbf{P} \in \mathbb{R}^K_{\gneqq 0} \text{ and } \mathbf{Q} \in \mathbb{R}^K_{>0}, \\ \widetilde{c}\cdot\Big\{ \frac{\sum_{k=1}^K (q_k)^\gamma \cdot (p_k)^{1-\gamma}}{\gamma\cdot(\gamma-1)} - \frac{1}{\gamma-1}\cdot\sum_{k=1}^K q_k + \frac{1}{\gamma}\cdot\sum_{k=1}^K p_k \Big\}, & \text{if } \gamma \in]0,1[,\ \mathbf{P} \in \mathbb{R}^K_{\gneqq 0} \text{ and } \mathbf{Q} \in \mathbb{R}^K_{\geq 0}, \\ \widetilde{c}\cdot\Big\{ \sum_{k=1}^K q_k \cdot \log\Big(\frac{q_k}{p_k}\Big) - \sum_{k=1}^K q_k + \sum_{k=1}^K p_k \Big\}, & \text{if } \gamma = 1,\ \mathbf{P} \in \mathbb{R}^K_{>0} \text{ and } \mathbf{Q} \in \mathbb{R}^K_{\geq 0}, \\ \widetilde{c}\cdot\Big\{ \sum_{k=1}^K \frac{(q_k)^\gamma \cdot (p_k)^{1-\gamma}}{\gamma\cdot(\gamma-1)} \cdot 1_{[0,\infty[}(q_k) - \frac{1}{\gamma-1}\cdot\sum_{k=1}^K q_k + \frac{1}{\gamma}\cdot\sum_{k=1}^K p_k \Big\}, & \text{if } \gamma \in]1,2[,\ \mathbf{P} \in \mathbb{R}^K_{>0} \text{ and } \mathbf{Q} \in \mathbb{R}^K, \\ \widetilde{c}\cdot\sum_{k=1}^K \frac{(q_k - p_k)^2}{2\cdot p_k}, & \text{if } \gamma = 2,\ \mathbf{P} \in \mathbb{R}^K_{>0} \text{ and } \mathbf{Q} \in \mathbb{R}^K, \\ \widetilde{c}\cdot\Big\{ \sum_{k=1}^K \frac{(q_k)^\gamma \cdot (p_k)^{1-\gamma}}{\gamma\cdot(\gamma-1)} \cdot 1_{[0,\infty[}(q_k) - \frac{1}{\gamma-1}\cdot\sum_{k=1}^K q_k + \frac{1}{\gamma}\cdot\sum_{k=1}^K p_k \Big\}, & \text{if } \gamma \in]2,\infty[,\ \mathbf{P} \in \mathbb{R}^K_{>0} \text{ and } \mathbf{Q} \in \mathbb{R}^K, \\ \infty, & \text{else;} \end{cases} \tag{17}$$

notice that one has the straightforward relationship $D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\cdot) = \widetilde{c}\cdot D_{\varphi_\gamma}(\cdot,\cdot)$.

For $\widetilde{c} = 1$ and probability vectors $\mathbb{Q}$, $\mathbb{P}$ in $\mathbb{S}^K$ respectively in $\mathbb{S}^K_{>0}$, the divergences (17) simplify considerably, namely to the well-known *power divergences* $D_{\varphi_\gamma}(\mathbb{Q},\mathbb{P})$ in the scaling of e.g. Liese & Vajda [18] (in other scalings they are also called *Rathie & Kannapan's non-additive directed divergences of order* γ [165], *Cressie-Read divergences* [5] [6], *relative Tsallis entropies or Tsallis cross-entropies* [7] (see also Shiino [166]), *Amari's alpha-divergences* [8]); for some comprehensive overviews on power divergences $D_{\varphi_\gamma}(\mathbb{Q},\mathbb{P})$ — including statistical applications to goodness-of-fit testing and minimum distance estimation — the reader is referred to the above-mentioned insightful books [18]–[23], the survey articles [25],[26], and the references therein. Prominent and widely used special cases of $D_{\varphi_\gamma}(\mathbb{Q},\mathbb{P})$ are the omnipresent *Kullback-Leibler information divergence (relative entropy)* where $\gamma = 1$, the equally important *reverse Kullback-Leibler information divergence (reverse relative entropy)* where $\gamma = 0$, the *Pearson chi-square divergence* ($\gamma = 2$), the *Neyman chi-square divergence* ($\gamma = -1$), the *Hellinger divergence* ($\gamma = \frac{1}{2}$, also called squared Hellinger distance, squared Matusita distance [167] or squared Hellinger-Kakutani metric, see e.g. Deza & Deza [168]). Some exemplary (relatively) recent studies and applications of power divergences $D_{\varphi_\gamma}(\mathbb{Q},\mathbb{P})$ — aside from the vast statistical literature (including in particular maximum likelihood estimation and Pearson's chi-square test) — are cited e.g. in Broniatowski & Stummer [1]. For $\widetilde{c} = 1$ and nonnegative-component vectors $\mathbf{Q}$, $\mathbf{P}$ in $\mathbb{R}^K_{\geq 0}$ respectively $\mathbb{R}^K_{>0}$ respectively $\mathbb{R}^K_{\gneqq 0}$, the generalized power divergences $D_{\varphi_\gamma}(\mathbf{Q},\mathbf{P})$ of (17) were treated by Stummer & Vajda [66] (for even more general probability measures, deriving e.g. also generalized Pinsker's inequalities); for a more general comprehensive technical treatment see also e.g. Broniatowski & Stummer [46].

For any fixed $M_{\mathbf{P}} \in]0,\infty[$, Condition 5 is satisfied for $\varphi := \widetilde{c}\cdot\varphi_\gamma$ — and thus the *narrow-sense* BS-minimizability concerning Theorem 9 can be applied — for all $\widetilde{c} \in]0,\infty[$ and all $\gamma \in \mathbb{R}\backslash]1,2[$ (cf. [1]). (As far as we know at the moment) For the

case $\gamma \in]1,2[$ one can not verify Condition 5, but BS-minimizability (in the sense of (2) with $f_n(\cdot) \not\equiv 1$) will be shown in Example 28 below.
For the latter, we shall employ the next example, which is of interest on its own due to its finite-derivative-behaviour at $t = \pm\infty$ (and which will be also helpful for e.g. the study of the total variation distance in Example 26 below).

Example 13: For any parameter-triple $\alpha, \beta, \widetilde{c} \in]0,\infty[$ we choose $]a,b[:=]-\infty,\infty[$ and

$$\varphi_{\alpha,\beta,\widetilde{c}}(t) := \widetilde{c}\cdot\alpha\cdot\Big\{\sqrt{1+\beta^2\cdot\Big(\frac{1-t}{\alpha}\Big)^2}\;-\;1+\log\frac{2\cdot\Big(\sqrt{1+\beta^2\cdot\Big(\frac{1-t}{\alpha}\Big)^2}\;-\;1\Big)}{\beta^2\cdot\Big(\frac{1-t}{\alpha}\Big)^2}\Big\}\;\in[0,\infty[,\quad t\in]-\infty,\infty[\,. \tag{18}$$

Notice that $\varphi_{\alpha,\beta,\widetilde{c}}(1)=0$, $\varphi'_{\alpha,\beta,\widetilde{c}}(1)=0$, $\varphi_{\alpha,\beta,\widetilde{c}}(-\infty)=\infty$ and $\varphi_{\alpha,\beta,\widetilde{c}}(\infty)=\infty$. Moreover, $\varphi'_{\alpha,\beta,\widetilde{c}}(-\infty)=\varphi'_{\alpha,\beta,\widetilde{c}}(a)=-\widetilde{c}\cdot\beta$ and $\varphi'_{\alpha,\beta,\widetilde{c}}(\infty)=\varphi'_{\alpha,\beta,\widetilde{c}}(b)=\widetilde{c}\cdot\beta$. Furthermore, $\varphi_{\alpha,\beta,\widetilde{c}}(\cdot)$ is strictly convex and smooth (i.e. of $C^\infty-$type). From (18), we construct the corresponding divergence (cf. (4))

$$\begin{aligned}
&D_{\varphi_{\alpha,\beta,\widetilde{c}}}(\mathbf{Q},\mathbf{P}) = \sum_{k=1}^K p_k\cdot\varphi_{\alpha,\beta,\widetilde{c}}\Big(\frac{q_k}{p_k}\Big)\\
&=\widetilde{c}\cdot\alpha\cdot\sum_{k=1}^K p_k\cdot\Big\{\sqrt{1+\beta^2\cdot\Big(\frac{1-\frac{q_k}{p_k}}{\alpha}\Big)^2}\;-\;1+\log\frac{2\cdot\Big(\sqrt{1+\beta^2\cdot\Big(\frac{1-\frac{q_k}{p_k}}{\alpha}\Big)^2}\;-\;1\Big)}{\beta^2\cdot\Big(\frac{1-\frac{q_k}{p_k}}{\alpha}\Big)^2}\Big\},\qquad \mathbf{P}\in\mathbb{R}^K_{>0},\mathbf{Q}\in\mathbb{R}^K.
\end{aligned}\tag{19}$$

For any fixed $M_{\mathbf{P}} \in]0,\infty[$, Condition 5 is satisfied for $\varphi := \varphi_{\alpha,\beta,\widetilde{c}}$ — and thus the *narrow-sense* BS-minimizability concerning Theorem 9 can be applied — for all parameter-triples $\alpha,\beta,\widetilde{c}\in]0,\infty[$ (cf. [1]). For the important special case $\alpha=\beta$ the formula (18) collapses to

$$\varphi_{\beta,\beta,\widetilde{c}}(t) = \widetilde{c}\cdot\beta\cdot\Big\{\sqrt{1+(1-t)^2}\;-\;1+\log\frac{2\cdot\Big(\sqrt{1+(1-t)^2}\;-\;1\Big)}{(1-t)^2}\Big\}\;\in[0,\infty[,\quad t\in]-\infty,\infty[\,, \tag{20}$$

and (19) collapses to

$$\begin{aligned}
&D_{\varphi_{\beta,\beta,\widetilde{c}}}(\mathbf{Q},\mathbf{P}) = \sum_{k=1}^K p_k\cdot\varphi_{\beta,\beta,\widetilde{c}}\Big(\frac{q_k}{p_k}\Big)\\
&=\widetilde{c}\cdot\beta\cdot\sum_{k=1}^K p_k\cdot\Big\{\sqrt{1+\Big(1-\frac{q_k}{p_k}\Big)^2}\;-\;1+\log\frac{2\cdot\Big(\sqrt{1+\Big(1-\frac{q_k}{p_k}\Big)^2}\;-\;1\Big)}{\Big(1-\frac{q_k}{p_k}\Big)^2}\Big\},\qquad \mathbf{P}\in\mathbb{R}^K_{>0},\mathbf{Q}\in\mathbb{R}^K.
\end{aligned}$$

To end this section, let us present Table 1 which gives a compact summary of several important solved cases; for details, see Section XII of Broniatowski & Stummer [1]. Notice that — as already explained above — in the current setup one has to take in Table 1 the divergence generator $M_{\mathbf{P}}\cdot\varphi$ instead of φ (cf. (9) instead of (69)). Here and henceforth, we employ the following notations:

- $GAM(\alpha,\beta)$ denotes the Gamma distribution with rate parameter (inverse scale parameter) $\alpha\in]0,\infty[$ and shape parameter $\beta\,]0,\infty[$ having (Lebesgue-)density $f(y) := \frac{\alpha^\beta\cdot y^{\beta-1}\cdot e^{-\alpha\cdot y}}{\Gamma(\beta)}\cdot 1_{]0,\infty[}(y),\;\; y\in\mathbb{R}$;
- $POI(\kappa)$ denotes the Poisson distribution with parameter $\kappa\in]0,\infty[$;
- Compound $POI(\kappa)$-$GAM(\alpha,\beta)$ denotes the corresponding "compound Poisson-Gamma distribution";
- $NOR(\mu,\sigma^2)$ denotes the Normal (i.e. Gaussian) distribution with mean $\mu\in\mathbb{R}$ and variance $\sigma^2\in]0,\infty[$;
- $NB(\tau,p)$ denotes the Negative-Binomial distribution with parameters $\tau\in]0,\infty[$ and $p\in]0,1[$.

$]a,b[$	$\varphi(t)$ for $t\in]a,b[$	$\mathbf{P}\in$	$\mathbf{Q}\in$	$D_\varphi(\mathbf{Q},\mathbf{P})$, $P=(p_k)_{k=1,\dots K}$ $\mathbf{Q}=(q_k)_{k=1,\dots K}$	$]t_-^{sc},t_+^{sc}[$	$\varphi(a)$	$\varphi(b)$	$\varphi'(a)$	$\varphi'(b)$	support of ζ	$\zeta=\mathbb{\Pi}[W\in\cdot\,]$ cf. (69)
$]0,\infty[$	$\widetilde{c}\cdot\varphi_\gamma(t):=\widetilde{c}\cdot\frac{t^\gamma-\gamma\cdot t+\gamma-1}{\gamma\cdot(\gamma-1)}$ for $\gamma\in\,]-\infty,0[,\,\widetilde{c}>0$	$\mathbb{R}^K_{\gneq 0}$	$\mathbb{R}^K_{>0}$	$\widetilde{c}\cdot\Big\{\frac{\sum_{k=1}^K (q_k)^\gamma\cdot(p_k)^{1-\gamma}}{\gamma\cdot(\gamma-1)}$ $-\frac{1}{\gamma-1}\cdot\sum_{k=1}^K q_k+\frac{1}{\gamma}\cdot\sum_{k=1}^K p_k\Big\}$	$]0,\infty[$	∞	∞	$-\infty$	$\frac{\widetilde{c}}{1-\gamma}$	$]0,\infty[$	dampened stable distribution on $[0,\infty[$
$]0,\infty[$	$\widetilde{c}\cdot\varphi_\gamma(t)$ for $\gamma\in]0,1[,\,\widetilde{c}>0$	$\mathbb{R}^K_{\gneq 0}$	$\mathbb{R}^K_{\geq 0}$	as above	$]0,\infty[$	$\frac{\widetilde{c}}{\gamma}$	∞	$-\infty$	$\frac{\widetilde{c}}{1-\gamma}$	$[0,\infty[$	Compound $POI(\frac{\widetilde{c}}{\gamma})$ $-\,GAM(\frac{\widetilde{c}}{1-\gamma},\frac{\gamma}{1-\gamma})$
$]-\infty,\infty[$	$\widetilde{c}\cdot\{\varphi_\gamma(t)\cdot\mathbf{1}_{]0,\infty[}(t)$ $+(\frac{1}{\gamma}-\frac{t}{\gamma-1})\cdot\mathbf{1}_{]-\infty,0]}(t)\}$ for $\gamma\in]2,\infty[,\,\widetilde{c}>0$	$\mathbb{R}^K_{\gneq 0}$	$\mathbb{R}^K$	$\widetilde{c}\cdot\Big\{\sum_{k=1}^K\frac{(q_k)^\gamma\cdot(p_k)^{1-\gamma}}{\gamma\cdot(\gamma-1)}\cdot\mathbf{1}_{[0,\infty[}(q_k)$ $-\frac{1}{\gamma-1}\cdot\sum_{k=1}^K q_k+\frac{1}{\gamma}\cdot\sum_{k=1}^K p_k\Big\}$	$]0,\infty[$	∞	∞	$\frac{\widetilde{c}}{1-\gamma}$	∞	$]-\infty,\infty[$	distorted stable distri- bution on $]-\infty,\infty[$
$]-\infty,\infty[$	$\widetilde{c}\cdot\varphi_2(t)=\widetilde{c}\cdot\frac{(t-1)^2}{2}$ for $\widetilde{c}>0$	$\mathbb{R}^K_{\gneq 0}$	$\mathbb{R}^K$	$\widetilde{c}\cdot\sum_{k=1}^K\frac{(q_k-p_k)^2}{2\cdot p_k}$	$]-\infty,\infty[$	∞	∞	$-\infty$	∞	$]-\infty,\infty[$	$NOR(1,\frac{1}{\widetilde{c}})$
$]0,\infty[$	$\widetilde{c}\cdot\varphi_0(t):=\widetilde{c}\cdot\{-\log t$ $+t-1\}$ for $\widetilde{c}>0$	$\mathbb{R}^K_{\gneq 0}$	$\mathbb{R}^K_{>0}$	$\widetilde{c}\cdot\Big\{\sum_{k=1}^K p_k\cdot\log\big(\frac{p_k}{q_k}\big)$ $+\sum_{k=1}^K q_k-\sum_{k=1}^K p_k\Big\}$	$]0,\infty[$	∞	∞	$-\infty$	1	$]0,\infty[$	$GAM(\widetilde{c},\widetilde{c})$
$]0,\infty[$	$\widetilde{c}\cdot\varphi_1(t):=\widetilde{c}\cdot\{t\cdot\log t$ $+1-t\}$ for $\widetilde{c}>0$	$\mathbb{R}^K_{\gneq 0}$	$\mathbb{R}^K_{\geq 0}$	$\widetilde{c}\cdot\Big\{\sum_{k=1}^K q_k\cdot\log\big(\frac{q_k}{p_k}\big)$ $-\sum_{k=1}^K q_k+\sum_{k=1}^K p_k\Big\}$	$]0,\infty[$	1	∞	$-\infty$	∞	$\{0,\frac{1}{c},\dots\}$ $=\frac{1}{c}\cdot\mathbb{N}_0$	$\frac{1}{c}-$fold of $POI(\widetilde{c})$
$]0,\infty[$	$\widetilde{c}\cdot\{t\cdot\log t$ $+(t+1)\cdot\log(\frac{2}{t+1})\},\,\widetilde{c}>0$	$\mathbb{R}^K_{\gneq 0}$	$\mathbb{R}^K_{\gneq 0}$ with $\mathbf{P}+\mathbf{Q}\in\mathbb{R}^K_{>0}$	$\widetilde{c}\cdot\Big\{\sum_{k=1}^K q_k\cdot\log\big(\frac{2q_k}{q_k+p_k}\big)$ $+\sum_{k=1}^K p_k\cdot\log\big(\frac{2p_k}{q_k+p_k}\big)\Big\}$	$]0,\infty[$	$\widetilde{c}\log 2$	∞	$-\infty$	$\widetilde{c}\log 2$	$\{0,\frac{1}{c},\dots\}$ $=\frac{1}{c}\cdot\mathbb{N}_0$	$\frac{1}{c}-$fold of $NB(\widetilde{c},\frac{1}{2})$
$]\frac{\beta-1}{\beta},\infty[$ $\beta\in]0,1]$	$\widetilde{c}\cdot\frac{(t-1)^2}{2(\beta\cdot t+1-\beta)}$ for $\widetilde{c}>0$	$\mathbb{R}^K_{\gneq 0}$	$]\frac{\beta-1}{\beta}\mathbf{P},\infty[$ component-wise	$\frac{\widetilde{c}}{2}\cdot\sum_{k=1}^K\frac{(q_k-p_k)^2}{\beta\cdot q_k+(1-\beta)\cdot p_k}$	$]\frac{\beta-1}{\beta},\infty[$	∞	∞	$-\infty$	$\frac{\widetilde{c}}{2\beta}$	$]\frac{\beta-1}{\beta},\infty[$	modified dampened stable distribution
$]z_1,z_2[$ $z_1<1$ $z_2>1$	$\frac{(t-z_1)}{z_2-z_1}\cdot\log\big(\frac{(t-z_1)\cdot p}{(z_2-t)\cdot(1-p)}\big)$ $-\log\big(\frac{(z_2-z_1)\cdot p}{(z_2-t)}\big)$ for $p:=\frac{z_2-1}{z_2-z_1}\in]0,1[$	$\mathbb{R}^K_{\gneq 0}$	$[z_1\mathbf{P},z_2\mathbf{P}]$ comp.-wise	$\sum_{k=1}^K\frac{q_k-z_1\cdot p_k}{z_2-z_1}\cdot\log\Big(\frac{p\cdot(q_k-z_1\cdot p_k)}{(1-p)\cdot(z_2\cdot p_k-q_k)}\Big)$ $-\sum_{k=1}^K p_k\cdot\log\Big(\frac{p\cdot(z_2-z_1)\cdot p_k}{z_2\cdot p_k-q_k}\Big)$	$]z_1,z_2[$	$\log\frac{1}{p}$	$\log\frac{1}{1-p}$	$-\infty$	∞	$\{z_1,z_2\}$	$\zeta[\{z_1\}]=p,$ $\zeta[\{z_2\}]=1-p$
$]-\infty,\infty[$	$\widetilde{c}\cdot\varphi_{\alpha,\beta_1,\beta_2}(t):=\widetilde{c}\cdot\alpha\cdot\{$ $\frac{\sqrt{4+(\beta_1+\beta_2)^2g^2}-(\beta_1-\beta_2)g-2}{2}$ $+\log\frac{\sqrt{4+(\beta_1+\beta_2)^2g^2}-2}{\beta_1\beta_2g^2}\}$ for $\alpha,\widetilde{c},\beta_1,\beta_2>0$, $g:=g(t):=\frac{1-t}{\alpha}+\frac{1}{\beta_2}-\frac{1}{\beta_1}$,	$\mathbb{R}^K_{\gneq 0}$	$\mathbb{R}^K$	$\widetilde{c}\cdot\sum_{k=1}^K p_k\cdot\varphi_{\alpha,\beta_1,\beta_2}\big(\frac{q_k}{p_k}\big)$	$]-\infty,\infty[$	∞	∞	$-\widetilde{c}\beta_2$	$\widetilde{c}\beta_1$	$]-\infty,\infty[$	law of $\breve{\theta}+Z_1-Z_2$ $\breve{\theta}:=1+\alpha\cdot\big(\frac{1}{\beta_2}-\frac{1}{\beta_1}\big)$ Z_1,Z_2 independent, $Z_1\sim GAM(\widetilde{c}\beta_1,\widetilde{c}\alpha)$, $Z_2\sim GAM(\widetilde{c}\beta_2,\widetilde{c}\alpha)$ (generalized asymmetric Laplace distribution)
$]-\infty,\infty[$	$\widetilde{c}\cdot\varphi_{\alpha,\beta}(t):=\widetilde{c}\cdot\alpha\cdot\{\sqrt{1+\beta^2g^2}$ $-1+\log\frac{2\cdot(\sqrt{1+\beta^2g^2}-1)}{\beta^2g^2}\}$ for $\beta,\alpha,\widetilde{c}>0$ and $g:=g(t):=\frac{1-t}{\alpha}$	$\mathbb{R}^K_{\gneq 0}$	$\mathbb{R}^K$	$\widetilde{c}\cdot\sum_{k=1}^K p_k\cdot\varphi_{\alpha,\beta}\big(\frac{q_k}{p_k}\big)$	$]-\infty,\infty[$	∞	∞	$-\widetilde{c}\beta$	$\widetilde{c}\beta$	$]-\infty,\infty[$	as above, but for $\breve{\theta}=1$, $Z_1\sim GAM(\widetilde{c}\beta,\widetilde{c}\alpha)$, $Z_2\sim GAM(\widetilde{c}\beta,\widetilde{c}\alpha)$

Table 1. Selection of concrete examples including some of their important features.

IV. Deterministic Narrow-Sense Bare-Simulation-Optimization of Bregman divergences

In the previous section, we have recalled/summarized recently achieved (cf. Broniatowski & Stummer [1]) *narrow-sense* bare-simulation minimization results for a special subclass of — discrete smooth — CASM $\varphi-$divergences. Let us now present a *first generalization* thereof, namely *narrow-sense* bare-simulation minimization results for a special subclass of — discrete smooth — *Scaled Bregman Distances* of Stummer [54] and Stummer & Vajda [45]; this will particularly cover narrow-sense bare-simulation minimization results for (a special subclass of) *ordinary/classical Bregman Distances*.

To start with, we fix a (scaling) vector $\mathbf{P}$ with strictly positive components $p_k > 0$ as well as a divergence generator $\varphi \in \widetilde{\Upsilon}(]a,b[)$ having representability (9) (i.e. φ satisfies Condition 5); recall from Remark 7 that $]t_-^{sc}, t_+^{sc}[$ (covering $t=1$) denotes the interior of its (maximum) domain of strict convexity, that φ is finite and differentiable on $]a,b[$, that $\varphi(t)=0$ if and only if $t=1$, and that $\varphi'(t)=0$ if and only if $t=1$. Moreover, we fix a second vector $\mathbf{Q}^{**} \in \mathbb{R}^K$ such that

$$t_k^{**} := \frac{q_k^{**}}{p_k} \in]t_-^{sc}, t_+^{sc}[\quad \text{for all } k=1,\ldots,K. \tag{21}$$

Notice that this implies $\varphi(t_k^{**}) < \infty$, and $\varphi'(t_k^{**}) \in]-\infty,\infty[$ $(k=1,\ldots,K)$. From this, for each $k=1,\ldots,K$ we construct the function

$$\varphi_k(t) \ := \ \varphi(t) - \varphi(t_k^{**}) - \varphi'(t_k^{**}) \cdot (t - t_k^{**}), \qquad t \in \mathbb{R}, \tag{22}$$

whose effective domain is $dom(\varphi_k) = dom(\varphi)$ and thus $int(dom(\varphi_k)) =]a,b[$. Notice that $\varphi_k(t) \geq 0$ for all $t \in \mathbb{R}$, that $\varphi_k(t)=0$ if and only if $t = t_k^{**}$, and that $\varphi_k'(t)=0$ if and only if $t=t_k^{**}$. Moreover, $]t_-^{sc}, t_+^{sc}[$ is also the interior of the (maximum) domain of strict convexity of φ_k, and consequently, φ_k is affine-linear on $]a, t_-^{sc}[$ (provided that this interval is non-empty) and on $]t_+^{sc}, b[$ (provided that this interval is non-empty). Clearly, we interpret $\varphi_k(a) := \lim_{t \downarrow a}(\varphi(t) - \varphi(t_k^{**}) - \varphi'(t_k^{**}) \cdot (t - t_k^{**})) = \varphi(a) - \varphi(t_k^{**}) - \varphi'(t_k^{**}) \cdot (a - t_k^{**})$ and $\varphi_k(b) := \lim_{t \uparrow b}(\varphi(t) - \varphi(t_k^{**}) - \varphi'(t_k^{**}) \cdot (t - t_k^{**})) = \varphi(b) - \varphi(t_k^{**}) - \varphi'(t_k^{**}) \cdot (b - t_k^{**})$ (where the limits always exist but may be infinite). By means of these φ_k's, under the assumption (21) we construct the — discrete smooth special cases of — *scaled Bregman distances* (between $\mathbf{Q}$ and $\mathbf{Q}^{**}$ [11]) of Stummer [54] and Stummer & Vajda [45] as

$$D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) \ := \ \sum_{k=1}^K p_k \cdot \varphi_k\left(\frac{q_k}{p_k}\right) \ = \ \sum_{k=1}^K p_k \cdot \left[\varphi\left(\frac{q_k}{p_k}\right) - \varphi\left(\frac{q_k^{**}}{p_k}\right) - \varphi'\left(\frac{q_k^{**}}{p_k}\right) \cdot \left(\frac{q_k}{p_k} - \frac{q_k^{**}}{p_k}\right)\right]. \tag{23}$$

Notice that $D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) < \infty$ if and only if $\frac{q_k}{p_k} \in dom(\varphi)$ for all $k \in \{1,\ldots,K\}$ (mostly, we deal with cases where $\frac{q_k}{p_k} \in]t_-^{sc}, t_+^{sc}[$ with eventual involvement of the boundary points). Moreover, there hold the above-mentioned divergence properties (iii), i.e. "$D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) \geq 0$" and "$D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) = 0$ if and only if $\mathbf{Q} = \mathbf{Q}^{**}$". Furthermore, the corresponding CASM $\varphi-$divergence can be recovered as special case $D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{P}) = D_\varphi(\mathbf{Q},\mathbf{P})$. Moreover, the particular choice $\mathbf{P} := (1,\ldots,1) := \mathbf{1}$ leads to — discrete smooth special cases of — the omnipresent important class of *separable ordinary/classical (i.e. unscaled) Bregman distances* $D_\varphi^{OBD}(\mathbf{Q},\mathbf{Q}^{**}) := D_{\varphi,\mathbf{1}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**})$ [12]. Detailed references on the theory and applications of ordinary Bregman distances respectively on scaled Bregman distances will be given below, after (30) as well as in (D2) respectively (D3) of the next Section V. Notice also that scaled Bregman distances of the form (23) also appear *in a natural way* in a new context of speed-up simulation, see Subsection XII-G below.

Let us also mention that in (22) we could also *equivalently* replace the generator $\varphi \in \widetilde{\Upsilon}(]a,b[)$ (with (9)) by $\breve{\varphi}$ defined through $\breve{\varphi}(t) := \varphi(t) + c_1 + c_2 \cdot t$ $(t \in dom(\varphi))$ with arbitrary $c_1, c_2 \in \mathbb{R}$; in particular, one has $\breve{\varphi}(1) = c_1 + c_2$ instead of $\varphi(1)=0$, and $\breve{\varphi}'(1) = c_2$ instead of $\varphi(1)' = 0$. This replacement can be done since for both $\breve{\varphi}$ and φ the respective two intervals $]a,b[$ and $]t_-^{sc}, t_+^{sc}[$ coincide, and since $D_{\breve{\varphi},\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) = D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**})$ by straightforward calculations. For the sake of brevity, for the rest of this section we stick to φ rather than $\breve{\varphi}$.

Remark 14: In the analogous spirit of Remark 6 — in terms of $\widetilde{\varphi} := M_{\mathbf{P}} \cdot \varphi$, the probability vector $\widetilde{\mathbb{P}} := \mathbf{P}/M_{\mathbf{P}}$, as well as the vectors $\widetilde{\mathbf{Q}} := \mathbf{Q}/M_{\mathbf{P}}$ and $\widetilde{\mathbf{Q}}^{**} := \mathbf{Q}^{**}/M_{\mathbf{P}}$ — one can equivalently rewrite (23) as

$$D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) = D_{\widetilde{\varphi},\widetilde{\mathbb{P}}}^{SBD}(\widetilde{\mathbf{Q}},\widetilde{\mathbf{Q}}^{**}) \tag{24}$$

and thus the two following minimization problems coincide:

$$\inf_{\mathbf{Q}\in\mathbf{\Omega}} D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) \qquad \text{respectively} \qquad \inf_{\widetilde{\mathbf{Q}}\in\widetilde{\mathbf{\Omega}}} D_{\widetilde{\varphi},\widetilde{\mathbb{P}}}^{SBD}(\widetilde{\mathbf{Q}},\widetilde{\mathbf{Q}}^{**}) \quad \text{with } \widetilde{\mathbf{\Omega}} := \mathbf{\Omega}/M_{\mathbf{P}}.$$

Notice that the roles of $\mathbf{P}$ respectively $\widetilde{\mathbb{P}}$ are now that of a scaling vector, whereas in (7) respectively (11) of the previous Section III they act as points which are projected on the constraint set $\mathbf{\Omega}$ respectively $\widetilde{\mathbf{\Omega}}$.

[11] note that for the finiteness of the divergence we allow $\mathbf{Q}$ to be in a domain which is larger or equal to the domain of $\mathbf{Q}^{**}$

[12] where φ may even have an affine-linear part, but all the components of $\mathbf{Q}^{**}$ are in its strictly-convex part

Below we shall show that the above-mentioned scaled Bregman distances can be BS-minimized *in the narrow sense* in their first component. For this, recall that we have fixed $\mathbf{P} \in \mathbb{R}^K_{>0}$ and $\mathbf{Q}^{**} \in \mathbb{R}^K$ such that (21) holds. By means of them, and by employing the above-mentioned notations $n_k := \lfloor n \cdot \widetilde{p}_k \rfloor$ ($k \in \{1,\ldots,K-1\}$), $n_K := n - \sum_{k=1}^{K-1} n_k$ (recall (12)) as well as

$$I_1^{(n)} := \{1,\ldots,n_1\}\,, \quad I_2^{(n)} := \{n_1+1,\ldots,n_1+n_2\}\,, \quad \ldots, \quad I_K^{(n)} := \{\sum_{k=1}^{K-1} n_k + 1,\ldots,n\}, \quad card(I_k^{(n)}) = n_k,$$

we construct the $n-$dimensional vector of random variables (with a slight abuse of notation)

$$\widetilde{\mathbf{V}} := \widetilde{\mathbf{V}}_n := \left(\widetilde{V}_1,\ldots,\widetilde{V}_n\right) = \left(\widetilde{V}_1,\ldots,\widetilde{V}_{n_1},\widetilde{V}_{n_1+1},\ldots,\widetilde{V}_{n_1+n_2},\ldots,\widetilde{V}_{\sum_{k=1}^{K-1} n_k+1},\ldots,\widetilde{V}_n\right), \tag{25}$$

as follows: independently for each $k \in \{1,\ldots,K\}$ we employ n_k i.i.d. random variables $\widetilde{V}_i$, $i \in I_k^{(n)}$, with common distribution $\mathbb{\Pi}[\widetilde{V}_i \in \cdot\,] = \widetilde{U}_k[\,\cdot\,]$ ($i \in I_k^{(n)}$) given by (the $\widetilde{\zeta}-$distortion type)

$$d\widetilde{U}_k(v) \;:=\; \frac{\exp\left(\tau_k \cdot v\right)}{MGF_{\widetilde{\zeta}}(\tau_k)}\, d\widetilde{\zeta}(v), \tag{26}$$

where $\tau_k := \; M_{\mathbf{P}} \cdot \varphi'\Big(\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\Big) = \; M_{\mathbf{P}} \cdot \varphi'\Big(\frac{q_k^{**}}{p_k}\Big)$. From this, we construct

$$\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}} := \Big(\frac{1}{n}\sum_{i\in I_1^{(n)}} \widetilde{V}_i, \ldots, \frac{1}{n}\sum_{i\in I_K^{(n)}} \widetilde{V}_i\Big) \tag{27}$$

for which (as a consequence of (9)) one can prove $\mathbb{E}_{\mathbb{\Pi}}\Big[\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big] = \Big(\frac{\lfloor n\cdot\widetilde{p}_1\rfloor}{n\cdot\widetilde{p}_1}\cdot\widetilde{q}_1^{**},\ldots,\frac{\lfloor n\cdot\widetilde{p}_{K-1}\rfloor}{n\cdot\widetilde{p}_{K-1}}\cdot\widetilde{q}_{K-1}^{**},\frac{n-\sum_{i=1}^{K-1}\lfloor n\cdot\widetilde{p}_i\rfloor}{n\cdot\widetilde{p}_K}\cdot\widetilde{q}_K^{**}\Big)$ (cf. Broniatowski & Stummer [1]) and thus — due to (12) — $\lim_{n\to\infty}\mathbb{E}_{\mathbb{\Pi}}\Big[\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big] = \widetilde{\mathbf{Q}}^{**}$ or equivalently $\lim_{n\to\infty}\mathbb{E}_{\mathbb{\Pi}}\Big[M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big] = \mathbf{Q}^{**}$.

Remark 15: Notice all the differences to the construction in the previous Section III. There, instead of the $\widetilde{\mathbf{V}}$ we have employed $\widetilde{\mathbf{W}} := \left(\widetilde{W}_1,\ldots,\widetilde{W}_n\right)$ where the $\widetilde{W}_i$'s are i.i.d. copies of the random variable $\widetilde{W}$ whose (block-neutral) distribution is $\mathbb{\Pi}[\widetilde{W} \in \cdot\,] = \widetilde{\zeta}[\,\cdot\,]$. We have partitioned these $\widetilde{\mathbf{W}}$ into the same blocks as the $\widetilde{\mathbf{V}}$, and transformed the former into the random vectors $\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}$ (cf. (13)) which are formally the same as $\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ (cf. (27)) where $\widetilde{\mathbf{W}}$ is just replaced by $\widetilde{\mathbf{V}}$. It is straightforward to see that for the special case $\mathbf{P} = \mathbf{Q}^{**}$ the two constructions coincide (since then $\tau_k = M_{\mathbf{P}}\cdot\varphi'(1) = 0$); this is consistent with the above-mentioned collapse-property $D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{P}) = D_\varphi(\mathbf{Q},\mathbf{P})$.

We are now in the position to formulate the following new assertion on the BS-minimizability of scaled Bregman distances:

Theorem 16: Let $\mathbf{P} \in \mathbb{R}^K_{>0}$, $M_{\mathbf{P}} := \sum_{i=1}^K p_i > 0$, and suppose that the divergence generator φ satisfies the above Condition 5, with $\widetilde{\zeta}$ (cf. (9)). Additionally, let $\mathbf{Q}^{**} \in \mathbb{R}^K$ such that (21) holds. Moreover, we assume that $\boldsymbol{\Omega}$ satisfies the regularity properties (6) as well as the finiteness property

$$\inf_{\mathbf{Q}\in\boldsymbol{\Omega}} D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) < \infty.$$

Furthermore, let $\widetilde{V} := (\widetilde{\mathbf{V}}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors constructed via (25) and (26). Then, in terms of the random vectors $\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ (cf. (27)) there holds

$$\inf_{\mathbf{Q}\in\boldsymbol{\Omega}} D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) \;=\; -\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}} \in \boldsymbol{\Omega}/M_{\mathbf{P}}\Big]\,. \tag{28}$$

In particular, for each $\mathbf{P} \in \mathbb{R}^K_{>0}$ and each $\mathbf{Q}^{**} \in \mathbb{R}^K$ with (21) the function $\Phi_{\mathbf{P},\mathbf{Q}^{**}}(\cdot) := D^{SBD}_{\varphi,\mathbf{P}}(\cdot,\mathbf{Q}^{**})$ (cf. (23)) is bare-simulation minimizable (BS-minimizable) *in the narrow sense* (cf. (2) in Definition 1 and the special case of Remark 2(a)) on any such $\boldsymbol{\Omega} \subset \mathbb{R}^K$.

The proof of Theorem 16 will be given in Appendix A.

Remark 17: Analogously to Remark 10(iii), the limit relation (28) even holds when $\mathbf{Q}^{**} \in \boldsymbol{\Omega}$ for which the left-hand side becomes zero.

Remark 18: For some contexts, we can *explicitly* give the distribution of each of the independent (non-deterministic parts of the) components $\Big(\sum_{i\in I_k^{(n)}}\widetilde{V}_i\Big)_{k=1,\ldots,K}$ of the vector $\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ (cf. [1] for other purposes than here); this will ease the corresponding concrete simulations.

Analogously to (15), the limit statement (28) provides the principle for the approximation of the solution of the minimization problem $\Phi_{\mathbf{P}}(\mathbf{\Omega}) := \inf_{\mathbf{Q}\in\mathbf{\Omega}} D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$. Namely, by replacing the right-hand side in (28) by its finite counterpart, we derive for given large n

$$-\frac{1}{n}\log \mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}} \in \mathbf{\Omega}/M_{\mathbf{P}}\Big] \approx \inf_{Q\in\mathbf{\Omega}} D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}); \tag{29}$$

it remains to estimate the left-hand side of (29) (see Section XII below, where the latter also provides estimates of the *minimizers*).

Example 19: Let us take the important case of the generators $\varphi_\gamma : \mathbb{R} \mapsto [0,\infty]$ defined by (16) in Example 12, which — by (23) — for arbitrary multiplier $\widetilde{c} \in]0,\infty[$, and arbitrary $\mathbf{P} \in \mathbb{R}^K_{>0}$ generate (the vector-valued form of) the *generalized scaled Bregman power distances* given by

$$D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) := \begin{cases} \widetilde{c}\cdot\Big\{\frac{\sum_{k=1}^K (q_k)^\gamma\cdot(p_k)^{1-\gamma}}{\gamma\cdot(\gamma-1)} + \frac{\sum_{k=1}^K (q_k^{**})^\gamma\cdot(p_k)^{1-\gamma}}{\gamma} - \frac{\sum_{k=1}^K q_k\cdot(q_k^{**})^{\gamma-1}\cdot(p_k)^{1-\gamma}}{\gamma-1}\Big\}, \\ \qquad\qquad \text{if } \gamma\in]-\infty,0[,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0} \text{ and } \mathbf{Q}\in\mathbb{R}^K_{>0}, \\ \widetilde{c}\cdot\Big\{-\sum_{k=1}^K p_k\cdot\log\Big(\frac{q_k}{q_k^{**}}\Big) + \sum_{k=1}^K \frac{q_k\cdot p_k}{q_k^{**}} - \sum_{k=1}^K p_k\Big\}, \quad \text{if } \gamma=0,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0} \text{ and } \mathbf{Q}\in\mathbb{R}^K_{>0}, \\ \widetilde{c}\cdot\Big\{\frac{\sum_{k=1}^K (q_k)^\gamma\cdot(p_k)^{1-\gamma}}{\gamma\cdot(\gamma-1)} + \frac{\sum_{k=1}^K (q_k^{**})^\gamma\cdot(p_k)^{1-\gamma}}{\gamma} - \frac{\sum_{k=1}^K q_k\cdot(q_k^{**})^{\gamma-1}\cdot(p_k)^{1-\gamma}}{\gamma-1}\Big\}, \\ \qquad\qquad \text{if } \gamma\in]0,1[,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0} \text{ and } \mathbf{Q}\in\mathbb{R}^K_{\geq 0}, \\ \widetilde{c}\cdot\Big\{\sum_{k=1}^K q_k\cdot\log\Big(\frac{q_k}{q_k^{**}}\Big) - \sum_{k=1}^K q_k + \sum_{k=1}^K q_k^{**}\Big\}, \quad \text{if } \gamma=1,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0} \text{ and } \mathbf{Q}\in\mathbb{R}^K_{\geq 0}, \\ \widetilde{c}\cdot\Big\{\frac{\sum_{k=1}^K (q_k)^\gamma\cdot(p_k)^{1-\gamma}\cdot 1_{[0,\infty[}(q_k)}{\gamma\cdot(\gamma-1)} + \frac{\sum_{k=1}^K (q_k^{**})^\gamma\cdot(p_k)^{1-\gamma}}{\gamma} - \frac{\sum_{k=1}^K q_k\cdot(q_k^{**})^{\gamma-1}\cdot(p_k)^{1-\gamma}}{\gamma-1}\Big\}, \\ \qquad\qquad \text{if } \gamma\in]1,2[,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0} \text{ and } \mathbf{Q}\in\mathbb{R}^K, \\ \widetilde{c}\cdot\sum_{k=1}^K \frac{(q_k-q_k^{**})^2}{2\cdot p_k}, \quad \text{if } \gamma=2,\ \mathbf{Q}^{**}\in\mathbb{R}^K \text{ and } \mathbf{Q}\in\mathbb{R}^K, \\ \widetilde{c}\cdot\Big\{\frac{\sum_{k=1}^K (q_k)^\gamma\cdot(p_k)^{1-\gamma}\cdot 1_{[0,\infty[}(q_k)}{\gamma\cdot(\gamma-1)} + \frac{\sum_{k=1}^K (q_k^{**})^\gamma\cdot(p_k)^{1-\gamma}}{\gamma} - \frac{\sum_{k=1}^K q_k\cdot(q_k^{**})^{\gamma-1}\cdot(p_k)^{1-\gamma}}{\gamma-1}\Big\}, \\ \qquad\qquad \text{if } \gamma\in]2,\infty[,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0} \text{ and } \mathbf{Q}\in\mathbb{R}^K, \\ \infty, \qquad\qquad \text{else;} \end{cases} \tag{30}$$

notice that one has the straightforward relationship $D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot,\cdot) = \widetilde{c}\cdot D^{SBD}_{\varphi_\gamma,\mathbf{P}}(\cdot,\cdot)$. The special case of (30) for probability vectors $\mathbb{Q}\in\mathbb{S}$, $\mathbb{Q}^{**}\in\mathbb{S}$ has been introduced in Stummer & Vajda [45] and e.g. applied in Kißlinger & Stummer [55],[56],[57],[58] as well as in Roensch & Stummer [169],[170],[171]. The case of (30) for "general" vectors (and even "general" measures) has been first indicated in Kißlinger & Stummer [57] and fully worked out in Broniatowski & Stummer [46]; for corresponding applications see e.g. Krömer & Stummer [172] (to mortality data analytics) and Broniatowski & Stummer [47] (to statistics and the adjacent fields of machine learning and artificial intelligence).

In line with the above-mentioned general imbedding, the particular choice $\mathbf{P} := (1,\ldots,1) := \mathbf{1}$ leads to — discrete smooth special cases of — the omnipresent important class of *separable ordinary/classical (i.e. unscaled) Bregman power distances* $D^{OBD}_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{Q}^{**}) := D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{1}}(\mathbf{Q},\mathbf{Q}^{**})$ which can be explicitly deduced from (30) by setting $p_k := 1$ for all $k = 1,\ldots,K$. For instance, for the case $\gamma = 2$ the omnipresent *squared weighted ℓ_2-distance* $D^{SBD}_{\widetilde{c}\cdot\varphi_2,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$ reduces to the omnipresent (multiple of) *squared ℓ_2-distance*

$$0 \leq D^{OBD}_{\widetilde{c}\cdot\varphi_2}(\mathbf{Q},\mathbf{Q}^{**}) \ = \ \frac{\widetilde{c}}{2}\cdot\sum_{k=1}^K (\,q_k - q_k^{**}\,)^2 \quad \mathbf{Q}\in\mathbb{R}^K,\ \mathbf{Q}^{**}\in\mathbb{R}^K. \tag{31}$$

In the special case of probability vectors $\mathbb{Q}^{**}\in\mathbb{S}^K_{>0}$ and $\mathbb{Q}\in\mathbb{S}^K_{>0}$ together with the special choice $\gamma > 0$, the divergences $D^{OBD}_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{Q}^{**})$ reduce to the prominent (finite-discrete-special-cases of) "order$-\gamma$" density power divergences of Basu et al. [48] (also known as Basu-Harris-Hjort-Jones distances) and their rescaled versions called *beta-divergences* (cf. Eguchi & Kano [50], Mihoko & Eguchi [51]). For general $\gamma\in\mathbb{R}$ and general probability measures see e.g. Stummer & Vajda [45]. The general case of non-probability vectors and measures is treated in Broniatowski & Stummer [46],[47] (see also Hennequin et al. [173] for finite discrete beta-divergences).

As far as further important particular parameter-cases (other than $\gamma = 2$) is concerned, $D^{OBD}_{\widetilde{c}\cdot\varphi_1}(\mathbf{Q},\mathbf{Q}^{**})$ (with $\gamma = 1$) amounts to the *discrete generalized relative entropy*, whereas $D^{OBD}_{\widetilde{c}\cdot\varphi_0}(\mathbf{Q},\mathbf{Q}^{**})$ (with $\gamma = 0$) leads to the *discrete (e.g. sampled) Itakura-Saito distance* [174].

Some exemplary (relatively) recent studies and applications of (generalized) density power divergences $D_{\widetilde{c}\cdot\varphi_\gamma}^{OBD}(\mathbf{Q},\mathbf{Q}^{**})$ (including beta-divergences, Itakura-Saito-divergences, continuous versions) — aside from the vast literature on the omnipresent cases $\gamma=1$ and $\gamma=2$ — appear e.g. in Basu et al. [175], Basu et al. [176], Ghosh & Basu [177],[178], Basu et al. [179], Martin et al. [180], Basu et al. [181], Balakrishnan et al. [182], Balakrishnan et al. [183], Ghosh & Majumdar [184], Leplat et al. [185], Vandecappelle et al. [186], Balakrishnan et al. [187], Basak et al. [188], Basu et al. [189], Calvino et al. [190], Castilla et al. [191], Castilla et al. [192], Legros et al. [193], Pu et al. [194], Ramirez et al. [195], Castilla & Chocano [196], Marmin et al. [197], Saraceno et al. [198], Sharma & Pradhan [199].

Returning to the general setup of generalized scaled Bregman power distances (30), let us point out that for any fixed $M_{\mathbf{P}}\in\,]0,\infty[$ the Condition 5 is satisfied for $\varphi:=\widetilde{c}\cdot\varphi_\gamma$ — and thus the *narrow-sense* BS-minimizability concerning Theorem 16 can be applied — for all $\widetilde{c}\in]0,\infty[$ and all $\gamma\in\mathbb{R}\backslash]1,2[$. (As far as we know at the moment) For the case $\gamma\in]1,2[$ one can not verify Condition 5, but BS-minimizability (in the sense of (2) with $f_n(\cdot)\not\equiv 1$) will be shown in Example 28 below.

As indicated in the rows 1 to 6 in the above Table 1 (with $M_{\mathbf{P}}\cdot\varphi$ instead of φ), the representability (9) of the power divergence generators $\varphi:=\widetilde{c}\cdot\varphi_\gamma$ has been completely worked out in Section XII of Broniatowski & Stummer [1]; there (for other purposes than here) we have also *explicitly* derived the corresponding crucial block-wise sampling distributions (cf. (26)) $\widetilde{U}_k[\,\cdot\,]=\mathbb{\Pi}[\,\widetilde{V}_i\in\cdot\,]$ ($i\in I_k^{(n)}$) and even more comfortably the block-sum distributions $\widetilde{U}_k^{*n_k}[\,\cdot\,]=\mathbb{\Pi}[\sum_{i\in I_k^{(n)}}\widetilde{V}_i\in\cdot\,]$, as follows:

- $\gamma\in]-\infty,0[$: $\widetilde{U}_k^{*n_k}$ has the (Lebesgue-)density
$$f_{\widetilde{U}_k^{*n_k}}(x)\;:=\;\frac{\exp((\tau_k-\frac{\widetilde{c}\cdot M_{\mathbf{P}}}{1-\gamma})\cdot x)}{\exp\left(n_k\cdot\frac{\widetilde{c}\cdot M_{\mathbf{P}}}{\gamma}\cdot(1+\frac{\gamma-1}{\widetilde{c}\cdot M_{\mathbf{P}}}\cdot\tau_k)^{\gamma/(\gamma-1)}\right)}\cdot f_{\breve{\breve{Z}}}(x)\cdot 1_{]0,\infty[}(x),\qquad x\in\mathbb{R},$$
where $\tau_k=\widetilde{c}\cdot M_{\mathbf{P}}\cdot\frac{1-\left(\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\right)^{\gamma-1}}{1-\gamma}$ for $\widetilde{q}_k^{**}>0$, and $\breve{\breve{Z}}$ is a random variable with density $f_{\breve{\breve{Z}}}$ of a stable law with parameter-quadruple $(\frac{-\gamma}{1-\gamma},1,0,-n_k\cdot\frac{(\widetilde{c}\cdot M_{\mathbf{P}})^{1/(1-\gamma)}\cdot(1-\gamma)^{-\gamma/(1-\gamma)}}{\gamma})$ (in terms of the "form-B notation" on p.12 in Zolotarev [200]);
- $\gamma=0$: $\widetilde{U}_k^{*n_k}=GAM\,(\widetilde{c}\cdot M_{\mathbf{P}}-\tau_k,n_k\cdot\widetilde{c}\cdot M_{\mathbf{P}})$, with $\tau_k=\widetilde{c}\cdot M_{\mathbf{P}}\cdot(1-\frac{\widetilde{p}_k}{\widetilde{q}_k^{**}})$ for $\widetilde{q}_k^{**}>0$;
- $\gamma\in]0,1[$: $\widetilde{U}_k^{*n_k}$ is Compound $POI(n_k\cdot\breve{\theta})-GAM\big(\frac{\widetilde{c}\cdot M_{\mathbf{P}}}{1-\gamma}-\tau_k,\frac{\gamma}{1-\gamma}\big)\big)$ with $\breve{\theta}:=\frac{\widetilde{c}\cdot M_{\mathbf{P}}}{\gamma}\cdot\big(\frac{(\gamma-1)\cdot\tau_k}{\widetilde{c}\cdot M_{\mathbf{P}}}+1\big)^{\gamma/(\gamma-1)}$ and $\tau_k=\widetilde{c}\cdot M_{\mathbf{P}}\cdot\frac{1-\left(\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\right)^{\gamma-1}}{1-\gamma}$ for $\widetilde{q}_k^{**}>0$;
- $\gamma=1$: $\widetilde{U}_k^{*n_k}$ is the probability distribution $\frac{1}{\widetilde{c}\cdot M_{\mathbf{P}}}\cdot POI\left(n_k\cdot\widetilde{c}\cdot M_{\mathbf{P}}\cdot\exp(\frac{\tau_k}{\widetilde{c}\cdot M_{\mathbf{P}}})\right)$ with support on the lattice $\left\{\frac{j}{\widetilde{c}\cdot M_{\mathbf{P}}},\,j\in\mathbb{N}_0\right\}$, where $\tau_k=\widetilde{c}\cdot\log\left(\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\right)$ for $\widetilde{q}_k^{**}>0$;
- $\gamma=2$: $\widetilde{U}_k^{*n_k}=NOR(n_k\cdot(1+\frac{\tau_k}{\widetilde{c}\cdot M_{\mathbf{P}}}),\frac{n_k}{\widetilde{c}\cdot M_{\mathbf{P}}})$ with $\tau_k=\widetilde{c}\cdot M_{\mathbf{P}}\cdot(\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}-1)$ for $\widetilde{q}_k^{**}\in\mathbb{R}$;
- $\gamma\in]2,\infty[$: $\widetilde{U}_k^{*n_k}$ has the (Lebesgue-)density
$$f_{\widetilde{U}_k^{*n_k}}(x)\;:=\;\frac{\exp((\tau_k+\frac{\widetilde{c}\cdot M_{\mathbf{P}}}{\gamma-1})\cdot x)}{\exp\left(n_k\cdot\frac{\widetilde{c}\cdot M_{\mathbf{P}}}{\gamma}\cdot(1+\frac{\gamma-1}{\widetilde{c}\cdot M_{\mathbf{P}}}\cdot\tau_k)^{\gamma/(\gamma-1)}\right)}\cdot f_{\breve{\breve{Z}}}(-x),\qquad x\in\mathbb{R},$$
where $\tau_k=-\frac{\widetilde{c}\cdot M_{\mathbf{P}}}{\gamma-1}\cdot\left(1-\left(\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\right)^{\gamma-1}\right)$ for $\widetilde{q}_k^{**}>0$, and $\breve{\breve{Z}}$ is a random variable with density $f_{\breve{\breve{Z}}}$ of a stable law with parameter-quadruple $(\frac{\gamma}{\gamma-1},1,0,n_k\cdot\frac{(\widetilde{c}\cdot M_{\mathbf{P}})^{1/(1-\gamma)}\cdot(\gamma-1)^{\gamma/(\gamma-1)}}{\gamma})$.

The explicit block-sum distributions $\widetilde{U}_k^{*n_k}$ for the other $\varphi-$generator examples in Table 1 (rows 7 to 11) can also be found in Section XII of Broniatowski & Stummer [1]. Let us present a new interesting completely worked out case:

Example 20: For $\beta\in\mathbb{R}\backslash\{0\}$ and $\widetilde{c}\in]0,\infty[$ and we employ the strictly increasing, smooth function
$$F_{\beta,\widetilde{c}}(t)\;:=\;\frac{2\widetilde{c}}{\beta}\cdot\left(e^{\beta\cdot t}-e^{\beta}\right),\qquad\qquad t\in]-\infty,\infty[,$$
which satisfies $F_{\beta,\widetilde{c}}(1)=0$ and which has strictly increasing, smooth inverse
$$F_{\beta,\widetilde{c}}^{\leftarrow}(x)\;:=\;\frac{1}{\beta}\cdot\log\left(\frac{\beta}{2\widetilde{c}}\cdot x+e^{\beta}\right),\qquad\qquad\text{if }\beta\cdot x>-2\widetilde{c}\cdot e^{\beta}.$$

By applying Theorem 22 and the corresponding Remark 23(b) of Broniatowski & Stummer [1], we obtain the divergence generator

$$\begin{aligned}\varphi_{\beta,\widetilde{c}}(t) &:= t\cdot F_{\beta,\widetilde{c}}(t) - \int_0^{F_{\beta,\widetilde{c}}(t)} F^{-1}_{\beta,\widetilde{c}}(u)\,du \;=\; t\cdot\frac{2\widetilde{c}}{\beta}\cdot\left(e^{\beta\cdot t}-e^{\beta}\right) - \frac{2\widetilde{c}}{\beta^2}\cdot\Big\{[\beta\cdot t\cdot e^{\beta\cdot t}-e^{\beta\cdot t}]-[\beta\cdot e^{\beta}-e^{\beta}]\Big\}\\ &= \frac{2\widetilde{c}}{\beta^2}\cdot\Big\{e^{\beta\cdot t}-t\cdot\beta\cdot e^{\beta}+(\beta-1)\cdot e^{\beta}\Big\},\qquad t\in\,]-\infty,\infty[;\end{aligned} \tag{32}$$

as well as the cumulant-generating-function-*candidate*

$$\Lambda_{\beta,\widetilde{c}}(z) := \int_0^z F^{\leftarrow}_{\beta,\widetilde{c}}(u)\,du \;=\; \frac{2\widetilde{c}}{\beta^2}\cdot\Big\{\Big[(\frac{\beta}{2\widetilde{c}}\cdot z+e^{\beta})\cdot\log(\frac{\beta}{2\widetilde{c}}\cdot z+e^{\beta})-(\frac{\beta}{2\widetilde{c}}\cdot z+e^{\beta})\Big]-\Big[\beta\cdot e^{\beta}-e^{\beta}\Big]\Big\},\quad\text{if } \beta\cdot z>-2\widetilde{c}\cdot e^{\beta}, \tag{33}$$

which satisfy the representability (69) (which is (9) with $M_{\mathbf{P}}=1$) below.
The function $\varphi_{\beta,\widetilde{c}}$ of (32) (and thus, equivalently, its linear-part-cleaned sibling $\breve{\varphi}_{\beta,\widetilde{c}}(t) := \frac{2\widetilde{c}}{\beta^2}\cdot e^{\beta\cdot t}$) — having $]t^{sc}_-,t^{sc}_+[=\,]-\infty,\infty[$ — generates the new scaled Bregman divergence (cf. (23))

$$D^{SBD}_{\varphi_{\beta,\widetilde{c}},\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) \;=\; D^{SBD}_{\breve{\varphi}_{\beta,\widetilde{c}},\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) = \frac{2\widetilde{c}}{\beta^2}\cdot\sum_{k=1}^{K}\Big\{p_k\cdot\exp\Big(\beta\cdot\frac{q_k}{p_k}\Big)-\Big(p_k+\beta\cdot(q_k-q_k^{**})\Big)\cdot\exp\Big(\beta\cdot\frac{q_k^{**}}{p_k}\Big)\Big\},$$
$$\text{for } \mathbf{P}\in\mathbb{R}^K_{>0},\mathbf{Q}\in\mathbb{R}^K,\mathbf{Q}^{**}\in\mathbb{R}^K. \tag{34}$$

Taking in (34) the special case $\mathbf{P}:=\mathbf{1}=(1,\ldots,1)$ and probability vectors $\mathbf{Q}:=\mathbb{Q}\in\mathbb{S}^K$, $\mathbf{Q}^{**}:=\mathbb{Q}^{**}\in\mathbb{S}^K$, one ends up at the discrete version of the *Bregman exponential divergence* of Mukherjee et al. [52]; the latter has been generalized by Basak & Basu [53] (in a continuous setting) to the *extended Bregman exponential divergence* by employing $\mathbf{Q}:=(\mathbb{Q})^{\alpha_1}$ and $\mathbf{Q}^{**}:=(\mathbb{Q}^{**})^{\alpha_2}$ (i.e. componentwise α_i-th powers), which also arises as a special case of (34) with $\mathbf{P}:=\mathbf{1}$. Another interesting special case of (34) is to choose $\mathbf{Q}:=\mathbb{Q}\in\mathbb{S}^K$ and $\mathbf{Q}^{**}:=\mathbb{Q}^{**}\in\mathbb{S}^K_{>0}$ together with $\mathbf{P}:=\mathbb{P}:=\mathbb{Q}^{**}$; the outcoming $\phi-$divergence $D^{SBD}_{\varphi_{\beta,\widetilde{c}},\mathbb{P}}(\mathbb{Q},\mathbb{Q}^{**})=D_{\varphi_{\beta,\widetilde{c}}}(\mathbb{Q},\mathbb{Q}^{**})=D_{\varphi_{\beta,\widetilde{c}}}(\mathbb{Q},\mathbb{P})$ is a multiple of the $\beta-$order *generalized negative exponential disparity/divergence GNED* of Jeong & Sarkar [201] and Bhandari et al. [202] (see also Basu et al. [24]), which — already in the special case $\beta:=-1$ called *negative exponential disparity/divergence NED* (cf. Lindsay [203] and Basu & Sarkar [204]) — performs good estimation-robustness against both outliers and inliers. For a more general, comprehensive, systematic treatment of the robustness-design of scaled Bregman divergences (including ordinary separable Bregman distances and $\varphi-$divergences), the reader is e.g. referred to Kißlinger & Stummer [57].

As far as (33) is concerned, one can show that the involved $\Lambda_{\beta,\widetilde{c}}$ is the cumulant generating function of a "distorted stable distribution" $\mathfrak{Q}[\,\cdot\,]=\mathbb{\Pi}[W\in\cdot\,]$ of a random variable W, which can be constructed as follows: let Z be an auxiliary random variable (having density f_Z and support $supp(Z)=]-\infty,\infty[$) of a stable law with parameter-quadruple $(1,1,1,\frac{2\widetilde{c}}{\beta^2})$ in terms of the above-mentioned "form-B notation"; by applying a general Laplace-transform result on p. 112 of [200] we can derive

$$M_Z(z):=E_{\mathbb{\Pi}}[\exp(z\cdot Z)]=\int_{-\infty}^{\infty}\exp(z\cdot y)\cdot f_Z(y)\,dy \;=\; \begin{cases}\exp\Big(\frac{2\widetilde{c}}{\beta^2}\cdot[(-z)\cdot\log(-z)-(-z)]\Big), & \text{if } z\in\,]-\infty,0],\\ \infty, & \text{if } z\in\,]0,\infty[.\end{cases}$$

Since $0\notin int(dom(M_Z))$ (and thus, Z does not have light-tails) we have to distort the involved density in order to extend the effective domain. Accordingly, let W be a random variable having density

$$f_W(y) \;:=\; \frac{2\widetilde{c}}{|\beta|}\cdot\frac{\exp\{\frac{2\widetilde{c}\cdot e^{\beta}}{\beta}\cdot y\}}{\exp\{\frac{2\widetilde{c}\cdot(\beta-1)\cdot e^{\beta}}{\beta^2}\}}\cdot f_Z\Big(-\frac{2\widetilde{c}}{\beta}\cdot y\Big),\qquad y\in\,]-\infty,\infty[. \tag{35}$$

Then one can straightforwardly deduce from (35) that $\int_{-\infty}^{\infty}f_W(y)\,dy=1$ and that

$$\begin{aligned}M_W(z) &:= E_{\mathbb{\Pi}}[\exp(z\cdot W)]=\int_{-\infty}^{\infty}\exp(z\cdot y)\cdot f_W(y)\,dy\\ &= \begin{cases}\exp\Big(\frac{2\widetilde{c}}{\beta^2}\cdot\Big\{(\frac{\beta}{2\widetilde{c}}\cdot z+e^{\beta})\cdot\log(\frac{\beta}{2\widetilde{c}}\cdot z+e^{\beta})-(\frac{\beta}{2\widetilde{c}}\cdot z+e^{\beta})-(\beta-1)\cdot e^{\beta}\Big\}\Big), & \text{if } z\cdot\beta\geq-2\widetilde{c}\cdot e^{\beta},\\ \infty, & \text{if } z\cdot\beta<-2\widetilde{c}\cdot e^{\beta},\end{cases}\end{aligned} \tag{36}$$

where the first line in (36) coincides with the exponentiated right-hand side (33). Notice that $\mathfrak{Q}$ is an infinitely divisible (cf. Proposition 27 in [1]) continuous distribution with density f_W, and that $\mathfrak{Q}[\,]0,\infty[\,]=\mathbb{\Pi}[W>0]=\int_0^{\infty}f_W(u)\,du\in\,]0,1[$, $\mathfrak{Q}[\,\{0\}\,]=\mathbb{\Pi}[W=0]=0$. Having derived the distribution $\mathfrak{Q}[\,\cdot\,]=\mathbb{\Pi}[W\in\cdot\,]$ of W, by replacing $\widetilde{c}$ with $\widetilde{c}\cdot M_{\mathbf{P}}$ in all the above

construction, we end up with the divergence generator $\widetilde{\varphi}_{\beta,\widetilde{c}} := M_{\mathbf{P}} \cdot \varphi_{\beta,\widetilde{c}} = \varphi_{\beta,\widetilde{c}\cdot M_{\mathbf{P}}}$ and the distribution $\widetilde{\mathfrak{l}}[\,\cdot\,] = \mathbb{\Pi}[\widetilde{W} \in \cdot\,]$ of a random variable $\widetilde{W}$, such that the representability (9) is satisfied. With those ingredients, we can apply Theorem 9 for the narrow-sense BS-minimization $\inf_{\mathbf{Q}\in\mathbf{\Omega}} D^{SBD}_{\varphi_{\beta,\widetilde{c}},\mathbf{P}}(\mathbf{Q},\mathbf{P}) = \inf_{\mathbf{Q}\in\mathbf{\Omega}} D_{\varphi_{\beta,\widetilde{c}}}(\mathbf{Q},\mathbf{P})$ of the non-probability version of the above-described multiple $\beta-$order GNED, for any $\mathbf{\Omega} \subset \mathbb{R}^K$ with regularity properties (6) and finiteness property (8). For the latter-type constraint sets, in order to perform the BS-minimization $\inf_{\mathbf{Q}\in\mathbf{\Omega}} D^{SBD}_{\varphi_{\beta,\widetilde{c}},\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$ for the more general scaled Bregman divergences of (34), according to Theorem 16 we have to further derive the corresponding distributions $\widetilde{U}_k[\,\cdot\,] =: \mathbb{\Pi}[\widetilde{V}_i \in \cdot\,]$ (cf. (26)) of the $\widetilde{V}_i$ ($i \in I_k^{(n)}$), or — for a better simulation comfort (cf. Remark 18) — even the block-sum distributions $\widetilde{U}_k^{*n_k}[\,\cdot\,] = \mathbb{\Pi}[\sum_{i\in I_k^{(n)}} \widetilde{V}_i \in \cdot\,]$. We achieve this by employing (26) to compute the moment generating function of $\sum_{i\in I_k^{(n)}} \widetilde{V}_i$ as the (due to the assumed independence) n_k-th power of the moment generating function of $\widetilde{V}_i$. From this, and with the help of $\tau_k := M_{\mathbf{P}} \cdot \varphi'_{\beta,\widetilde{c}}\Big(\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\Big) = \frac{2\widetilde{c}\cdot M_{\mathbf{P}}}{\beta} \cdot \Big\{\exp\Big(\beta \cdot \frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\Big) - \exp(\beta)\Big\}$ for arbitrary $\widetilde{q}_k^{**} \in \mathbb{R}$, we identify that $\widetilde{U}_k^{*n_k}$ has the (Lebesgue-)density

$$f_{\widetilde{U}_k^{*n_k}}(y) \;:=\; \frac{2\widetilde{c}\cdot M_{\mathbf{P}}}{|\beta|} \cdot \frac{\exp\Big\{\frac{2\widetilde{c}\cdot M_{\mathbf{P}}}{\beta} \cdot \exp\Big(\beta \cdot \frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\Big) \cdot y\Big\}}{\exp\Big\{\frac{2\widetilde{c}\cdot M_{\mathbf{P}}\cdot n_k}{\beta^2} \cdot \exp\Big(\beta \cdot \frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\Big) \cdot \Big(\beta \cdot \frac{\widetilde{q}_k^{**}}{\widetilde{p}_k} - 1\Big)\Big\}} \cdot f_{\breve{\breve{Z}}}\Big(-\,\frac{2\widetilde{c}\cdot M_{\mathbf{P}}}{\beta} \cdot y\Big), \qquad y \in \,]-\infty,\infty[,$$

where $\breve{\breve{Z}}$ is a random variable with density $f_{\breve{\breve{Z}}}$ of a stable law with parameter-quadruple $(1,1,1,\frac{2\widetilde{c}\cdot M_{\mathbf{P}}\cdot n_k}{\beta^2})$.

V. Bare-Simulation-Method for General Deterministic Divergence-Optimization-Problems

A. Further Divergences and Friends

Although the BS-*minimization* results of

- the above Section III on $\varphi-$divergences $D_\varphi(\mathbf{Q},\mathbf{P})$ with divergence generator φ satisfying Condition 5 — and thus particularly φ is continuously differentiable (cf. Remark 7) — as well as
- the above Section IV on scaled (including ordinary) Bregman divergences $D^{SBD}_{\varphi,\mathbf{M}}(\mathbf{Q},\mathbf{P})$ (with slight re-notation) with the same divergence generator φ (with possibly affine-linear addition term)

cover a substantial amount of concrete divergences and related generalized entropies, (in addition to the desir of *maximization* of the latter two) we would like to go one step beyond and e.g. also minimize and maximize the *following* directed distances $\mathbf{Q} \mapsto \Phi_{\mathbf{P}}(\mathbf{Q}) := D(\mathbf{Q},\mathbf{P})$ and connected functions of great importance in information theory as well as in the adjacent fields of statistics, machine learning, artificial intelligence, signal processing and pattern recognition:

(D1) CASM $\varphi-$divergences $\Phi_{\mathbf{P}}(\mathbf{Q}) := D_\varphi(\mathbf{Q},\mathbf{P})$ which *can not be covered* by the narrow-sense BS minimizability results of Section III, i.e. $\varphi \in \widetilde{\Upsilon}(]a,b[)$ *does not satisfy Condition 5*; for instance, φ may be *non*-differentiable which is e.g. the case with the choice $\varphi_{TV}(t) := |t-1|$ (with $t \in]a,b[=]-\infty,\infty[$) leading to the very important (discrete special case of the) *total variation distance*

$$\Phi_{\mathbf{P}}(\mathbf{Q}) \;:=\; D_{\varphi_{TV}}(\mathbf{Q},\mathbf{P}) := \sum_{k=1}^K p_k \cdot \varphi_{TV}\Big(\frac{q_k}{p_k}\Big) \;=\; \sum_{k=1}^K |\,q_k - p_k\,| \;\geq 0, \quad \mathbf{Q} \in \mathbb{R}^K,\ \mathbf{P} \in \mathbb{R}^K_{\geq 0}, \tag{37}$$

which is equal to the ℓ_1-distance $\Phi^{\ell_1}_{\mathbf{P}}(\mathbf{Q}) := ||\mathbf{Q}-\mathbf{P}||$ between $\mathbf{Q}$ and $\mathbf{P}$ (notice that $\Phi^{\ell_1}_{\mathbf{P}}(\mathbf{Q})$ can be trivially extended to $\mathbf{P} \in \mathbb{R}^K$ which (in case of some strictly negative components $p_k < 0$) is not a CASM divergence anymore). As far as literature on $\varphi-$divergences is concerned, e.g. recall the references given in the paragraph after (4); some further exemplary (relatively) recent studies and applications of $D_\varphi(\mathbf{Q},\mathbf{P})$ (partially with (37) as well as more general setups) — in addition to the vast statistical literature (including in particular maximum likelihood estimation and Pearson's chi-square test) — appear e.g. in Berend et al. [205], Jiao et al. [206], Como & Fagnani [207], Han et al. [208], Sason [209], Batsidis et al. [210], Böcherer & Geiger [211], Das & Kashyap [212], Alonso-Revenga et al. [213], Keziou & Regnault [214], Liu et al. [215], Tzortzis et al. [216], Castilla et al. [217], Csiszár & Breuer [218], El Gheche et al. [219], Felipe et al. [220], Markatou & Chen [221], Sun et al. [222], Asadi et al. [223], Broniatowski et al. [224], Collet [225], Sason [226], Yagli & Cuff [227], Zhao et al. [228], De Ponti [229], Kammerer & Stummer [230], Nishiyama & Sason [231], Nomura [232], Rassouli & Gündüz [233], Esposito et al. [234], Markatou et al. [235], Salehkalaibar et al. [236], Stummer [237], Tzortzis et al. [238], Birrell et al. [239], Castilla & Chocano [240], Dixit et al. [241], Hyun et al. [242], Melbourne et al. [243], Peng et al. [244], Tan & Zhang [245], Zhang et al. [246], Alba-Fernández & Jiménez-Gamero [247], Baudry et al. [248], Boukeloua & Keziou [249], Cressie et al. [250], Kateri [251], Manole & Ramdas [252], Markatou & Liu [253], Masiha et al. [254], Miranda et al. [255], Nielsen & Okamura [256], Perrone [257], Velasco-Forero [258] and Nielsen & Okamura [259].

(D2) (discrete special case of) *separable "ordinary/classical" Bregman distances* (cf. [39])

$$\Phi_{\mathbf{P}}(\mathbf{Q}) := D^{OBD}_\varphi(\mathbf{Q},\mathbf{P}) \;:=\; \sum_{k=1}^K \Big[\varphi(q_k) - \varphi(p_k) - \varphi'(p_k)\cdot(q_k - p_k)\Big] \;\geq 0\,, \tag{38}$$

where (opposed to Section IV) the divergence generator φ is *not necessarily in* $\widetilde{\Upsilon}(]a,b[)$ *satisfying additionally the Condition 5*, but is a lower semicontinuous convex function $\varphi :]-\infty,\infty[\rightarrow]-\infty,\infty]$ which is (continuously) differentiable on $int(dom(\varphi)) =]a,b[$ and strictly convex on (say, only) some interval $]t_-^{sc},t_+^{sc}[\subset]a,b[$ [13] such that $dom(\varphi)$ covers all the involved components q_k of all $\mathbf{Q} \in \mathbf{\Omega}$ as well as $]t_-^{sc},t_+^{sc}[$ covers all p_k ($k = 1,\ldots,K$); accordingly, $D_\varphi^{OBD}(\mathbf{Q},\mathbf{P})$ in (38) is always finite and the overall sum can be split into finite "partial" sums. For instance, $]a,b[=]0,\infty[$ or $]a,b[=]-\infty,\infty[$.

For the special case of probability vectors $\mathbb{P}$ and $\mathbb{Q} \in \mathbb{S}^K$, $D_\varphi^{OBD}(\mathbb{Q},\mathbb{P})$ are studied e.g. in Csiszár [40], [41], [42], Pardo & Vajda [43], [44], and Stummer & Vajda [45]). The general case of non-probability vectors (and even measures) is treated e.g. in Broniatowski & Stummer [46],[47]. Some exemplary (relatively) recent applications on separable ordinary Bregman distances (including continuous versions) — aside from the literature on the special case of separable ordinary Bregman *power* distances already cited after (31) — appear e.g. in Jiao et al. [206], Csiszár & Breuer [260], Jana & Basu [261], Painsky & Wornell [262], Vial et al. [263], Tan & Zhang [245].

(D3) (discrete special case of) *scaled Bregman distances* of Stummer [54] and Stummer & Vajda [45]

$$\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\varphi,\mathbf{M}}^{SBD}(\mathbf{Q},\mathbf{P}) \ := \ \sum_{k=1}^{K} \left[\varphi\left(\frac{q_k}{m_k}\right) - \varphi\left(\frac{p_k}{m_k}\right) - \varphi'_+\left(\frac{p_k}{m_k}\right) \cdot \left(\frac{q_k}{m_k} - \frac{p_k}{m_k}\right) \right] \cdot m_k \ \geq \ 0\,, \qquad (39)$$

where $\mathbf{M} \in \mathbb{R}_{>0}^K$ is a scaling vector with strictly positive components $m_k > 0$ and (opposed to Section IV) the divergence generator φ is *not necessarily in* $\widetilde{\Upsilon}(]a,b[)$ *satisfying additionally the Condition 5*, but is a lower semicontinuous convex function $\varphi :]-\infty,\infty[\rightarrow]-\infty,\infty]$ which is strictly convex (with right-hand derivative φ'_+) on (say, only) some interval $]t_-^{sc},t_+^{sc}[\subset\ int(dom(\varphi)) =]a,b[$ such that $dom(\varphi)$ covers all the involved components $\frac{q_k}{m_k}$ for all $\mathbf{Q} \in \mathbf{\Omega}$ as well as $]t_-^{sc},t_+^{sc}[$ covers all $\frac{p_k}{m_k}$ ($k = 1,\ldots,K$); accordingly, $D_{\varphi,\mathbf{M}}^{SBD}(\mathbf{Q},\mathbf{P})$ in (39) is always finite and the overall sum can be split into finite "partial" sums. For instance, $]a,b[=]0,\infty[$ or $]a,b[=]-\infty,\infty[$. Notice that *scaled Bregman divergences* have been first defined in Stummer [54], Stummer & Vajda [45] for the context of *probability measures and probability vectors*, see also Kißlinger & Stummer [55], [56], [57] for the "purely adaptive" case (i.e. $m_k = w\big(q_k,p_k\big)$ for some so-called *scale-connector function* $w(\cdot)$ respectively its continuous version) and for indications on non-probability measures and non-probability vectors. Later on, Broniatowski & Stummer [46] flexibilized/widened this to *scaled Bregman distances of arbitrary functions and vectors*, see also Broniatowski & Stummer [47] for various different kinds of applications to statistics, and to the adjacent fields of machine learning and artificial intelligence.

(D4) (discrete special case of) the *distances of Broniatowski & Stummer [46]* (see also Broniatowski & Stummer [47])

$$\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\varphi,\mathbf{M_1},\mathbf{M_2},\mathbf{M_3}}^{BSD,c}(\mathbf{Q},\mathbf{P}) := \sum_{k=1}^{K} \left[\varphi\left(\frac{q_k}{m_{1,k}}\right) - \varphi\left(\frac{p_k}{m_{2,k}}\right) - \varphi'_{+,c}\left(\frac{p_k}{m_{2,k}}\right) \cdot \left(\frac{q_k}{m_{1,k}} - \frac{p_k}{m_{2,k}}\right) \right] \cdot m_{3,k} \geq 0, \qquad (40)$$

where $\mathbf{M}_1, \mathbf{M}_2 \in \mathbb{R}_{>0}^K$ are scaling vectors and $\mathbf{M}_3 \in \mathbb{R}_{>0}^K$ is an aggregation (weighting) vector with strictly positive components $m_{i,k} > 0$ ($k = 1,\ldots,K,\ i = 1,2,3$). Moreover, the divergence generator φ is a lower semicontinuous convex function $\varphi :]-\infty,\infty[\rightarrow]-\infty,\infty]$ which is strictly convex (with right-hand derivative φ'_+, left-hand derivative φ'_- and intermediate values $\varphi'_{+,c}(t) := c \cdot \varphi'_+(t) + (1-c) \cdot \varphi'_-(t)$ ($c \in [0,1]$)) on (say, only) some interval $]t_-^{sc},t_+^{sc}[\subset int(dom(\varphi)) =]a,b[$ such that $dom(\varphi)$ covers all the involved components $\frac{q_k}{m_{1,k}}$ for all $\mathbf{Q} \in \mathbf{\Omega}$ and as well as $]t_-^{sc},t_+^{sc}[$ covers all $\frac{p_k}{m_{2,k}}$ ($k = 1,\ldots,K$). Accordingly, $D_{\varphi,\mathbf{M_1},\mathbf{M_2},\mathbf{M_3}}^{BSD,c}(\mathbf{Q},\mathbf{P})$ in (40) is always finite and the overall sum can be split into finite "partial" sums. Under some mild assumptions, [46] verify the axioms of divergence (i.e. non-negativity and reflexivity). Within this general framework, all the above distances/divergences (D1), (D2), (D3) appear as special cases. If φ is also differentiable (and thus, $\varphi'_{+,c}(\cdot)$ becomes the derivative $\varphi'(\cdot)$ and we omit the obsolete index c), other important special cases are (the separable versions of)

(i) the *total Bregman distances* of Liu et al. [60],[61], Vemuri et al. [62] (see also e.g. the recent application in Lohit & Kumar [264]) which are representable as $\Phi_{\mathbf{P}}(\mathbf{Q}) := D_\varphi^{TBD}(\mathbf{Q},\mathbf{P}) := D_{\varphi,\mathbf{1},\mathbf{1},\mathbf{M}^{to}(\mathbf{P})}^{BSD}(\mathbf{Q},\mathbf{P})$ with constant $m_{1,k} = m_{2,k} = 1$ and $m_{3,k} = m_3 := \frac{1}{\sqrt{1+\sum_{i=1}^K (\varphi'(p_i))^2}}$ for all $k = 1,\ldots,K$ and $\mathbf{M}^{to}(\mathbf{P})$ is the vector where each of the K components is equal to m_3,

(ii) the variants of (i) with constant $m_{1,k} = m_1 = H(\mathbf{Q})$, $m_{2,k} = m_2 = H(\mathbf{P})$ (for some differentiable positive real-valued function $H(\cdot)$) and $m_{3,k} = 1$ (cf. Nock et al. [63]),

(iii) the *conformal divergences* of Nock et al. [64] which can be represented as $\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\varphi,\mathbf{1},\mathbf{1},\mathbf{M}^{conf}(\mathbf{P})}^{BSD}(\mathbf{Q},\mathbf{P})$ where $\mathbf{M}^{conf}(\mathbf{P})$ is the vector where each of the K components is equal to $H(\mathbf{P})$ for some technically adequate positive real-valued function $H(\cdot)$,

(iv) the *scaled conformal divergences* of Nock et al. [64] which can be represented as $\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\varphi,m\cdot\mathbf{1},m\cdot\mathbf{1},m\cdot\mathbf{M}^{conf}(\mathbf{P})}^{BSD}(\mathbf{Q},\mathbf{P})$ for some strictly positive constant $m > 0$.

[13] recall that the divergence generator φ may contain some line parts, see e.g. $\varphi := \varphi_\gamma$ of (16) with $\gamma \in]1,2[$, where $]t_-^{sc},t_+^{sc}[=]0,\infty[$ and $]a,b[=]-\infty,\infty[$

A more detailed discussion on (i)–(iv) (and their continuous versions and further generalizations) can be found in Stummer & Kißlinger [59].

(D5) (discrete special case of the *non-probability extension* of) the $\varphi-$*entropies* in the sense of Burbea & Rao [14] (see also Csiszar [29], Ben-Bassat [30], Ben-Tal & Teboulle [31], Kesavan & Kapur [32], Dacunha-Castelle & Gamboa [33], Teboulle & Vajda [34], Gamboa & Gassiat [35], Vajda & Zvarova [36])

$$\Phi_{\mathbf{1}}(\mathbf{Q}) \ := \ \mathcal{E}_{\varphi}(\mathbf{Q}) := \sum_{k=1}^{K} \varphi(q_k) \tag{41}$$

for (say) $\varphi \in \widetilde{\Upsilon}(]a,b[)$ *not necessarily satisfying additionally the Condition 5*. Clearly, the constrained maximization (minimization) of $\mathcal{E}_{\varphi}(\cdot)$ corresponds to a *generalized entropy maximization (minimization)* task.
As explained right after (5), there are also numerous other applications of $\sum_{k=1}^{K} \varphi(q_k)$ where φ is e.g. a cost function respectively energy function respectively purpose function, and $\sum_{k=1}^{K} \varphi(q_k)$ can also be interpreted as an "index/degree of (in)equality of the set $\mathbf{\Omega}$", respectively as an "index/degree of diversity of the set $\mathbf{\Omega}$". A corresponding comprehensive BS-concerning discussion is given e.g. in Broniatowski & Stummer [1].
A flexibilization of (41) is given by

$$\Phi_{\mathbf{1}}(\mathbf{Q}) \ := \ \mathcal{E}_{\varphi,h}(\mathbf{Q}) := h\Big(\sum_{k=1}^{K} \varphi(q_k)\Big) \tag{42}$$

where $h : \mathcal{H} \mapsto \mathbb{R}$ is a continuous strictly increasing (respectively strictly decreasing) function with $\mathcal{H} \subseteq [0,\infty[$. The quantity $h\Big(\sum_{k=1}^{K} \varphi(q_k)\Big)$ in (42) can be seen as (non-probability extension of an) $(h,\varphi)-$entropy in the sense of Salicru et al. [37] (see also e.g. Pardo [22], Vajda & Vasek [38], as well as e.g. Chen et al. [265], Girardin & Lhote [266], Ren et al. [267], Girardin et al. [268] for exemplary applications).

(D6) Burbea-Rao divergences [14] (see also e.g. Pardo & Vajda [43],[44])

$$\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\varphi}^{BR}(\mathbf{Q},\mathbf{P}) \ := \ \sum_{k=1}^{K} \Big[\frac{\varphi(q_k)+\varphi(p_k)}{2} \ - \ \varphi\Big(\frac{q_k+p_k}{2}\Big)\Big]$$

with some lower semicontinuous convex function $\varphi :]-\infty,\infty[\to]-\infty,\infty]$ which is strictly convex on $int(dom(\varphi)) =]a,b[$ such that $dom(\varphi)$ covers all the involved components q_k of all $\mathbf{Q} \in \mathbf{\Omega}$ as well as all p_k $(k = 1,\ldots,K)$. We can also handle their straightforward generalizations (called (separable form of) *skew Burbea-Rao divergences* in e.g. Nielsen & Boltz [269] respectively *skew Jensen divergences* in e.g. Nielsen & Nock [270])

$$\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\varphi,\beta}^{sBR}(\mathbf{Q},\mathbf{P}) \ := \ \sum_{k=1}^{K} \Big[\beta\cdot\varphi(q_k) + (1-\beta)\cdot\varphi(p_k) \ - \ \varphi\Big(\beta\cdot q_k + (1-\beta)\cdot p_k\Big)\Big]$$

with some $\beta \in]0,1[$ (and we can even deal with corresponding non-separable generalizations of the latter); see also Stummer & Kißlinger [59] for obtaining separable skew Burbea-Rao divergences as a special case of their scaled-Bregman-distance-flexibilizations.

(D7) (discrete special case of) *generally non-separable "ordinary/classical" Bregman distances* (cf. [39])

$$\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\psi}^{gnOBD}(\mathbf{Q},\mathbf{P}) \ := \ \psi(\mathbf{Q}) - \psi(\mathbf{P}) - \nabla\psi(\mathbf{P})\cdot(\mathbf{Q}-\mathbf{P}) \ \geq 0\,, \tag{43}$$

where $\psi : \mathbf{B} \mapsto \mathbb{R}$ is strictly convex and continuously differentiable (with gradient $\nabla\psi$) on an open domain $\mathbf{B} \subset \mathbb{R}^K$ such that its closure $cl(\mathbf{B})$ covers all the involved $\mathbf{Q} \in \mathbf{\Omega}$ as well as all $\mathbf{P}$. For the separable choice $\psi(\mathbf{Q}) := \psi_{\varphi}(\mathbf{Q}) := \sum_{k=1}^{K} \varphi(q_k)$, the distance (43) turns into the OBD (38). Moreover, for the choice $\psi(\mathbf{X}) := \psi_{MAH,\underline{\underline{A}}}(\mathbf{x}) := \mathbf{X}^{tr}\underline{\underline{A}}\mathbf{X}$ with some positive definite $K \times K-$matrix $\underline{\underline{A}} := (a_{i,j})_{i,j=1,\ldots,K}$, the distance (43) leads to the prominently used *squared Mahalanobis distance* [65]

$$\Phi_{\mathbf{P}}(\mathbf{Q}) := D_{\psi_{MAH,\underline{\underline{A}}}}^{gnOBD}(\mathbf{Q},\mathbf{P}) \ = \ \sum_{k=1}^{K}\sum_{j=1}^{K} a_{i,j}\cdot(\,q_k - p_k\,)\cdot(\,q_j - p_j\,)\,; \tag{44}$$

clearly, in the special case where $\underline{\underline{A}} := \frac{1}{2}\cdot\underline{\underline{I}}$ is half of the unit matrix $\underline{\underline{I}}$, the squared Mahalanobis distance (44) simplifies to the squared ℓ_2-distance (31).

Some exemplary (relatively) recent applications of (squared) Mahalanobis distance can be found e.g. in Xu et al. [271], Kim et al. [272], Mei et al. [273], Zhang et al. [274], Li et al. [275], Xu et al. [276], Mahony & Cannon [277], Xu et al. [278], Etherington [279], Fitzpatrick & Dunn [280], Kakavand et al. [281], Sun et al. [282], Bai et al. [283], Li et al. [284], Naveed & ur Rehman [285], Wang et al. [286], Winter et al. [287], Bartlett et al. [288], Butterfield et al. [289], Chamberland et al. [290], Kang et al. [291], Sato et al. [292], Zheng et al. [293], A et al. [294], Chakraborty et al. [295],

Chen et al. [296], dos Santos et al. [297], Guerra et al. [298], Huang et al. [299], Nomoto et al. [300], Reichen et al. [301], Sun et al. [302], Timmermann et al. [303], Wauchope et al. [304], Weinberger [305], Wen et al. [306], Yang et al. [307], Zhang et al. [308], Burssens et al. [309], Choi et al. [310], Choi et al. [311], Dahlin et al. [312], Ebrahimi et al. [313], Jeong et al. [314], Kim et al. [315], Nowakowski et al. [316], Qu et al. [317], Rabby et al. [318], Sarno et al. [319], Tang et al. [320], Tsvieli & Weinberger [321], Zhang et al. [322], Zhou et al. [323].

Some exemplary (relatively) recent applications of generally non-separable (ordinary/classical) Bregman distances appear e.g. in Jiao et al. [206], Varshney & Varshney [324], Hu et al. [325], Nock et al. [326], Raskutti & Mukherjee [327], Wang et al. [328], He et al. [329], Li et al. [330], Harremoës [331], Xu et al. [332], Halder [333], Zhang et al. [334], Shao et al. [335], Tembine [336], Brécheteau et al. [337], Lin et al. [338], Yuan et al. [339], Azizan et al. [340], Dytso et al. [341], Gruzdeva & Ushakov [342], Song et al. [343], Tan & Zhang [245], Yu et al. [344], Capó et al. [345], Chen et al. [346], Fernández-Rodriguez [347], Hayashi [348], Li & Ralescu [349], Xiong et al. [350], Liu et al. [351].

(D8) *weighted ℓ_r-distances* ($r \in]0,\infty[$)

$$\Phi_{\mathbf{P}}(\mathbf{Q}) \,:=\, D_{\ell_r,\mathbf{M}}(\mathbf{Q},\mathbf{P}) \,:=\, \left(\sum_{k=1}^{K} \frac{1}{m_k} \cdot |\, q_k - p_k \,|^r\right)^{1/r} \geq 0 \tag{45}$$

where $\mathbf{M}$ is a weighting (scaling) vector with strictly positive components $m_k > 0$.

Remark 21: (a) The above-mentioned contexts of Section III, Section IV as well as (D1) to (D8) also cover corresponding divergences $D(X,Y)$ (including entropies) between (possible complex-valued) $M\times N$ matrices X and Y, by taking $\breve{D}(X,Y) := D(f(X),f(Y))$ where f is an appropriate mapping from (a subset of) the space of all (possible complex-valued) $M \times N$ matrices to (a subset of) $\mathbb{R}^K$. For instance, for real-valued $M \times N$ matrices one can employ their *vectorization* by taking (say) $f(X) := \mathbf{Q} := (q_1,\ldots,q_{M\cdot N})$ such that $q_{(i-1)\cdot N+j} := x_{ij}$ ($i = 1,\ldots,M$, $j = 1,\ldots,N$) and hence $K := M \cdot N$; more flexible versions where $i \in \{1,\ldots,M\}$, $j \in J_i$ for some $J_i \subseteq \{1,\ldots,N\}$ as well as multidimensional-array/tensor versions can be transformed in a similar book-keeping manner, too.
(b) Another important special case of (a) is to take for $K \times K$ Hermitian (e.g. real symmetric) matrices (say) X the function $f(X) = (\lambda_1,\ldots,\lambda_K)$ to be the $K-$dimensional real-valued vector of its eigenvalues λ_k in e.g. decreasing order. Accordingly, $\breve{D}(X,Y) := D(f(X),f(Y))$ measures the dissimilarity between the vectors of ordered eigenvalues of the two Hermitian matrices X and Y. Depending on the nature of the underlying vector-divergence $D(\cdot,\cdot)$ one may need further restrictions on the involved matrices X,Y in order to achieve the finiteness $\breve{D}(X,Y) < \infty$. For instance, if $D(\cdot,\cdot) := D_\varphi(\cdot,\cdot)$ is of CASM $\varphi-$divergence type (D1), then "basically" all the ratios of the involved eigenvalues should lie in $dom(\varphi)$; for $dom(\varphi) = [0,\infty[$ (respectively $dom(\varphi) =]0,\infty[$) this is the case for positive semi-definite (respectively positive definite) Hermitian matrices X,Y. As another example, one can take a vector-divergence $D(\cdot,\cdot)$ from the class (D2) of separable ordinary/classical Bregman distances (D2) or from the class (D7) of generally non-separable ordinary/classical Bregman distances (see e.g. Bauschke & Borwein [352], Dhillon & Tropp [353] and Kulis et al. [354]) which covers in particular the *von Neumann divergence*, the *log-determinant divergence*, the squared Frobenius distance and the more general *spectral Bregman matrix divergences*.
(c) According to (a) and (b), all the above-mentioned and below-mentioned BS-optimization results on vector-divergences $D(\cdot,\cdot)$ carry over to corresponding BS-optimization results on matrix-divergences $\breve{D}(\cdot,\cdot)$.

B. Minimization via Base-Divergence-Method 1

All the above-mentioned contexts (D1) to (D8) share basically the same property that the involved function (to be constrained-optimized) $\mathbf{Q} \mapsto \Phi_{\mathbf{P}}(\mathbf{Q})$ is continuous. For such a context we obtain the following new fundamental

Theorem 22: Let us arbitrarily fix some $\mathbf{P} \in \mathbb{R}^K_{>0}$, $M_{\mathbf{P}} > 0$, φ, $\widetilde{\zeta}$, $\widetilde{W} := (\widetilde{W}_i)_{i\in\mathbb{N}}$ and $\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}$ (cf. (13)) as in Theorem 9; we call the corresponding $D_\varphi(\cdot,\mathbf{P})$ the *base-divergence* (function).
(a) Furthermore, suppose that $\boldsymbol{\Omega} \subset \mathbb{R}^K$ is compact and satisfies the regularity properties (6), and that $\Phi : \boldsymbol{\Omega} \mapsto \mathbb{R}$ is a continuous function on $\boldsymbol{\Omega}$. Then, there holds

$$\inf_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q}) \;=\; -\lim_{n\to\infty} \frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \Big(D_\varphi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}, \mathbf{P}\big) - \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big)\Big) \cdot 1\!\!1_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big) \Big]\Big) \tag{46}$$

and the infimum is attained at some (not necessarily unique) point in $\boldsymbol{\Omega}$. In particular, the function $\Phi(\cdot)$ is bare-simulation minimizable (BS-minimizable) on $\boldsymbol{\Omega}$ (cf. (2) in Definition 1).
(b) If $\boldsymbol{\Omega} \subset \mathbb{R}^K$ is not necessarily compact but satisfies the regularity properties (6) and the finiteness property (8), and $\Phi : \boldsymbol{\Omega} \mapsto \mathbb{R}$ is a continuous function which satisfies the lower-bound condition

$$\text{there exists a constant } c_1 \in \mathbb{R} \text{ such that for all } \mathbf{Q} \in \boldsymbol{\Omega} \text{ there holds} \quad \Phi(\mathbf{Q}) \geq D_\varphi(\mathbf{Q},\mathbf{P}) - c_1\,, \tag{47}$$

then the representation/convergence (46) — and hence the corresponding BS-minimizability — still holds, but the infimum may not necessarily be attained/reached at some point in $\boldsymbol{\Omega}$.

The proof of Theorem 22 will be given in Appendix A below.

Remark 23: (i) In Theorem 22(b), one gets (46) with exponent $D_{\varphi}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\leq c_1$, which turns into an *exponential dampening* in case of $c_1<0$. Examples for the applicability of (47) and Theorem 22(b) will be given right below.
(ii) In Theorem 22 we have allowed for the special case that $\mathbf{P}$ can be in $\boldsymbol{\Omega}$ (and thus, $\inf_{\mathbf{Q}\in\boldsymbol{\Omega}}D_{\varphi}(\mathbf{Q},\mathbf{P})=0$ of Remark 10(iii) applies); however, in such a situation one gets $\inf_{\mathbf{Q}\in\boldsymbol{\Omega}}\Phi(\mathbf{Q})\neq 0$ in general.
(iii) As indicated above, in Theorem 22 the function $\Phi(\cdot)$ can be e.g. of the form $\Phi(\cdot):=\Phi_{\breve{\mathbf{P}}}(\cdot):=D_{\breve{\varphi}}(\cdot,\breve{\mathbf{P}})$ where the pregiven $\breve{\mathbf{P}}$ and $\breve{\varphi}$ do not necessarily coincide with the $\mathbf{P}$ and φ of the base-divergence $D_{\varphi}(\cdot,\mathbf{P})$ employed in (46).

Analogously to (15), the limit statement (46) provides the principle for the approximation of the solution of the minimization problem $\Phi(\boldsymbol{\Omega}):=\inf_{\mathbf{Q}\in\boldsymbol{\Omega}}\Phi(\mathbf{Q})$. This can be achieved by replacing the right-hand side in (46) by its finite counterpart, from which we obtain for given large n

$$-\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_{\varphi}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big)\Big)\cdot 1_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\ \Big]\Big)\ \approx\ \inf_{Q\in\boldsymbol{\Omega}}\Phi(\mathbf{Q});\tag{48}$$

it remains to estimate the left-hand side of (48) (see Section XII below, where the latter also provides estimates of the *minimizers*).

Clearly, Theorem 22(a) can be applied to obtain $\inf_{\mathbf{Q}\in\boldsymbol{\Omega}}\Phi(\mathbf{Q})$ for all the directed distances/divergences $\Phi(\cdot):=\Phi_{\mathbf{P}}(\cdot)$ and friends given in (D1) to (D8), which are therefore all BS-minimizable on compact $\boldsymbol{\Omega}$ with regularity (6). As far as the application of Theorem 22(b) to the case (D1) of CASM $\varphi-$divergences is concerned, we derive the following

Corollary 24: Let us arbitrarily fix some $\mathbf{P}\in\mathbb{R}^K_{>0}$, $M_{\mathbf{P}}>0$, φ, $\widetilde{\zeta}$, $\widetilde{W}:=(\widetilde{W}_i)_{i\in\mathbb{N}}$ and $\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}$ (cf. (13)) as in Theorem 9. If $\breve{\varphi}\in\widetilde{\Upsilon}(]a,b[)$ satisfies the lower-bound condition

$$\text{there exists a constant } \breve{c}\in]0,\infty[\text{ such that for all } t\in]a,b[\text{ there holds } \ \breve{c}\cdot\breve{\varphi}(t)\geq\varphi(t) \ \text{ with equality if and only if } t=1\,,\tag{49}$$

then there holds

$$\inf_{\mathbf{Q}\in\boldsymbol{\Omega}}D_{\breve{c}\cdot\breve{\varphi}}(\mathbf{Q},\mathbf{P})\ =\ -\lim_{n\to\infty}\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_{\varphi}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)-D_{\breve{c}\cdot\breve{\varphi}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)\Big)\Big)\cdot 1_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\ \Big]\Big)\tag{50}$$

for all $\boldsymbol{\Omega}\subset\mathbb{R}^K$ satisfying the regularity properties (6) and the finiteness property (8). In particular, the CASM divergence $\Phi_{\mathbf{P}}(\cdot)\ :=\ D_{\breve{c}\cdot\breve{\varphi}}(\cdot,\mathbf{P})$ is BS-minimizable (on any such $\boldsymbol{\Omega}$). The exponent in (50) satisfies $D_{\varphi}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)-D_{\breve{c}\cdot\breve{\varphi}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)\leq 0$ [14] with equality if and only if $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}=\mathbf{P}$ (where typically the latter happens at most very occasionally).

The assertion of Corollary 24 follows immediately from Theorem 22(b) and the fact that (49) implies

$$D_{\breve{c}\cdot\breve{\varphi}}(\mathbf{Q},\mathbf{P})\geq D_{\varphi}(\mathbf{Q},\mathbf{P})\quad\text{for all }\mathbf{Q}\in\boldsymbol{\Omega}$$

(with equality if and only if $\mathbf{Q}=\mathbf{P}$).

Remark 25: In Corollary 24, the assumed *narrow-sense* BS-minimizability of $D_{\varphi}(\cdot,\mathbf{P})$ is transformed into the (non-narrow-sense) BS-minimizability of $D_{\breve{c}\cdot\breve{\varphi}}(\cdot,\mathbf{P})$.

Example 26: Let us show how Corollary 24 can be used to tackle the BS-minimizability of the very important total variation distance (ℓ_1-distance)

$$\Phi_{\mathbf{P}}(\mathbf{Q})\ :=\ D_{\breve{\varphi}}(\mathbf{Q},\mathbf{P}):=D_{\varphi_{TV}}(\mathbf{Q},\mathbf{P}):=\sum_{k=1}^{K}p_k\cdot\varphi_{TV}\left(\frac{q_k}{p_k}\right)\ =\ \sum_{k=1}^{K}|\,q_k-p_k\,|\ \geq 0\qquad\text{(cf. (37))}$$

where $\breve{\varphi}(t):=\varphi_{TV}(t):=|t-1|$ for all $t\in]a,b[=]-\infty,\infty[$. For this, as generator of the base-divergence we take $\varphi:=\varphi_{\alpha,\beta,\widetilde{c}}$ of Example 13, with (for simplicity) parameters $\widetilde{c}:=1$ and $\alpha=\beta\in]0,1[$. From $\varphi_{\beta,\beta,1}(1)=0=\varphi_{TV}(1)$, $\varphi'_{\beta,\beta,1}(1)=0$, $\varphi'_{\beta,\beta,1}(-\infty)=-\beta>\varphi'_{TV,-}(1)=-1$, $\varphi'_{\beta,\beta,1}(\infty)=\beta<\varphi'_{TV,+}(1)=1$ and the strict convexity of $\varphi_{\beta,\beta,1}(\cdot)$ (and thus the strict increasingness of $\varphi'_{\beta,\beta,1}$), one can easily see that $\varphi_{\beta,\beta,1}(t)\leq|t-1|$ for all $t\in\mathbb{R}$, with equality if and only if $t=1$. Hence, (49) holds with $\breve{c}:=1$. To proceed, let us arbitrarily fix some $\mathbf{P}\in\mathbb{R}^K_{>0}$. The transformed generator $\widetilde{\varphi_{\beta,\beta,1}}:=M_{\mathbf{P}}\cdot\varphi_{\beta,\beta,1}=\varphi_{\beta,\beta,M_{\mathbf{P}}}$ satisfies Condition 5 (i.e. (9)) with $\widetilde{\zeta}$ such that the $\widetilde{W}_i$'s are i.i.d. copies of the random variable $\widetilde{W}$ of the form $\widetilde{W}:=1+Z_1-Z_2$, where Z_1 and Z_2 are auxiliary random variables which are independent and

[14] in other words, the exponential-dampening concerning Remark 23 is applicable with $c_1=0$.

$GAM(M_{\mathbf{P}} \cdot \beta, M_{\mathbf{P}} \cdot \beta)-$distributed (cf. Broniatowski & Stummer [1], see also Table 1). With these choices, (50) specializes to

$$\inf_{\mathbf{Q}\in\mathbf{\Omega}} D_{\varphi_{TV}}(\mathbf{Q},\mathbf{P}) \;=\; -\lim_{n\to\infty}\frac{1}{n}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_{\varphi_{\beta,\beta,1}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)-D_{\varphi_{TV}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)\Big)\Big)\cdot 1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\;\Big]\Big) \tag{51}$$

for all $\mathbf{\Omega}\subset\mathbb{R}^K$ satisfying the regularity properties (6) and the finiteness property (8). In particular, the total variation distance (ℓ_1-distance) $\Phi_{\mathbf{P}}(\cdot) := D_{\varphi_{TV}}(\cdot,\mathbf{P})$ is BS-minimizable (on any such $\mathbf{\Omega}$). As an exemplary application, let us take $\mathbf{P}:=\mathbf{1}$ and

$$\mathbf{\Omega}\subsetneqq\mathbf{\Omega}_0:=\Big\{\mathbf{Q}\in\mathbb{R}^K:\sum_{i=1}^{d}\Big(y_i+\sum_{k=1}^{K}x_{i,k}-\sum_{k=1}^{K}x_{i,k}\cdot q_k\Big)^2\le\varepsilon\Big\} \tag{52}$$

with pregiven $y_i\in\mathbb{R}$ and $x_{i,k}\in\mathbb{R}$ ($i=1,\ldots,d$, $k=1,\ldots,K$) — where the integer d is (say) much smaller than K — and pregiven $\varepsilon\in]0,\infty[$. Non-degeneration assumptions on the entries $(x_{i,k})_{i=1,\ldots,d;k=1,\ldots,K}$ have to be considered so that $\mathbf{\Omega}$ (e.g. $\mathbf{\Omega}_0$) satisfies the regularity properties (6) and the corresponding finiteness property (8).
Accordingly, with our BS method (51) we can approximate/tackle the corresponding minimimum $\inf_{\mathbf{Q}\in\mathbf{\Omega}} D_{\varphi_{TV}}(\mathbf{Q},\mathbf{1})$ which with $\widetilde{q}_k := q_k-1$ transforms equivalently into the constrained ℓ_1-norm minimum

$$\inf_{\widetilde{\mathbf{Q}}}\;||\widetilde{\mathbf{Q}}||_1 \tag{53}$$

such that

$$\sum_{i=1}^{d}\Big(y_i-\sum_{k=1}^{K}x_{i,k}\cdot\widetilde{q}_k\Big)^2\le\varepsilon, \tag{54}$$

eventually (cf. $\subsetneqq$ in (52)) with further (i.e. side) constraints on $\mathbf{Q}$. (55)

In a multiple-linear-regression context where ε is a pregiven small error and each data observation y_i is of the form $y_i=\widetilde{q}_1\cdot x_{i,1}+\cdots+\widetilde{q}_K\cdot x_{i,K}+\eta_i$ with deterministic explanatory variables $x_{i,k}$, parameters $\widetilde{q}_k$ and i.i.d. homoscedastic/standard Gaussian noise η_i, the problem (53) to (55) corresponds to the (minimum value of a formulation of the) well-known *basis pursuit denoising problem* (cf. Donoho et al. [355], see also e.g. Candès et al. [356], Lustig et al. [357], Candès [358], Candès et al. [359], Goldstein & Osher [360], Zhang et al. [361], Edgar et al. [362]) — with eventual further constraints. As a side remark, let us mention the *direct* application of Theorem 22(a) to the continuous weighted ℓ_r-distance ($r\in]0,\infty[$)

$$\Phi_{\mathbf{P}}(\mathbf{Q}) \;:=\; D_{\ell_r,\mathbf{M}}(\mathbf{Q},\mathbf{P}) \;:=\; \left(\sum_{k=1}^{K}\frac{1}{m_k}\cdot|\,q_k-p_k\,|^r\right)^{1/r}\ge 0\,,\qquad \mathbf{M}\in\mathbb{R}^K_{>0},\qquad\qquad \text{(cf. (45))}$$

with the special constellation $\varphi=\varphi_{\beta,\beta,1}$, $\mathbf{P}:=\mathbf{1}$ and compact $\mathbf{\Omega}_0$; by proceeding analogously as above, we can tackle by our BS-method the constrained weighted ℓ_r-norm minimum

$$\inf_{\widetilde{\mathbf{Q}}}\;\left(\sum_{k=1}^{K}\frac{1}{m_k}\cdot|\,\widetilde{q}_k\,|^r\right)^{1/r} \tag{56}$$

such that the constraint (54) is satisfied. (57)

For the subcase $m_k=1$ for all $k=1,\ldots,K$, the problem (56),(57) has been tackled by other methods e.g. in Foucart & Lai [363] and Liu et al. [364], whereas e.g. Bruckstein et al. [365] deal with general $m_k>0$ together with $r=1$; notice that for $m_k=1$ the minimization problem (56),(57) is a *relaxation* of the "pure" sparseness optimization (dimension reduction) problem

$$\inf_{\widetilde{\mathbf{Q}}}\;||\widetilde{\mathbf{Q}}||_0$$

such that the constraint (54) is satisfied,

since

$$\lim_{r\to 0}\left(\sum_{k=1}^{K}|\,\widetilde{q}_k\,|^r\right)^{1/r} \;=\; ||\widetilde{\mathbf{Q}}||_0 \;=:\; card(\{k\in\{1,\ldots,K\}:\widetilde{q}_k\neq 0\}).$$

Remark 27: As a side effect of the above considerations (in slightly more general form), for any parameter-triple $\alpha,\beta,\widetilde{c}\in]0,\infty[$ with $\widetilde{c}\cdot\beta<1$ we can show the new divergence inequality

$$\text{for all } \mathbf{Q}\in\mathbb{R}^K \text{ and } \mathbf{P}\in\mathbb{R}^K_{>0} \text{ there holds } \;D_{\varphi_{\alpha,\beta,\widetilde{c}}}(\mathbf{Q},\mathbf{P})\le D_{\varphi_{TV}}(\mathbf{Q},\mathbf{P}) \quad\text{with equality if and only if } \mathbf{Q}=\mathbf{P}\,,$$

since $\varphi_{\alpha,\beta,\widetilde{c}}(t)\le|t-1|$ for all $t\in\mathbb{R}$, with equality if and only if $t=1$.

Example 28: Let us now show how Corollary 24 can be used to tackle the BS-minimizability of the generalized power divergences $\Phi(\cdot) := \Phi_{\mathbf{P}}(\cdot) := D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbf{P})$ (cf. (17)) for the missing case $\gamma \in]1,2[$ (and $\widetilde{c} \in]0,\infty[$) in Example 12. For this, as generator of the base-divergence we take $\varphi := \varphi_{\alpha,\beta,\frac{\widetilde{c}}{\gamma}}$ of Example 13, with (for simplicity) parameters $\alpha = \beta \in]0,1[$. In Appendix A we shall prove the bound

$$\text{for all } \gamma \in]1,\infty[,\ \beta \in]0,1[,\ \widetilde{c} \in]0,\infty[\text{ and } t \in \mathbb{R} \text{ there holds } \varphi_{\beta,\beta,\frac{\widetilde{c}}{\gamma}}(t) \leq \widetilde{c}\cdot\varphi_\gamma(t) \quad \text{with equality if and only if } t = 1\,. \tag{58}$$

Thus, by confining ourselves to $\gamma \in]1,2[$ we can proceed analogously to Example 26, with the difference that $\widetilde{\varphi_{\alpha,\beta,\frac{\widetilde{c}}{\gamma}}} := M_{\mathbf{P}}\cdot\varphi_{\alpha,\beta,\frac{\widetilde{c}}{\gamma}} = \varphi_{\beta,\beta,\frac{M_{\mathbf{P}}\cdot\widetilde{c}}{\gamma}}$ and hence the $\widetilde{W}_i$'s are i.i.d. copies of the random variable $\widetilde{W}$ of the form $\widetilde{W} := 1 + Z_1 - Z_2$, where Z_1 and Z_2 are auxiliary random variables which are independent and $GAM(\frac{M_{\mathbf{P}}\cdot\widetilde{c}}{\gamma}\cdot\beta, \frac{M_{\mathbf{P}}\cdot\widetilde{c}}{\gamma}\cdot\beta)-$distributed (cf. Section XII of Broniatowski & Stummer [1], see also Table 1). With these choices, (50) specializes to

$$\inf_{\mathbf{Q}\in\mathbf{\Omega}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P}) \;=\; -\lim_{n\to\infty}\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_{\varphi_{\beta,\beta,\frac{\widetilde{c}}{\gamma}}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big) - D_{\widetilde{c}\cdot\varphi_\gamma}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)\Big)\Big)\cdot 1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\ \Big]\Big)$$

for all $\mathbf{\Omega} \subset \mathbb{R}^K$ satisfying the regularity properties (6) and the finiteness property (8). In particular, for $\gamma \in]1,2[$ and $\widetilde{c} \in]0,\infty[$ the generalized power divergence $\Phi_{\mathbf{P}}(\cdot) := D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbf{P})$ is BS-minimizable (on any such $\mathbf{\Omega}$). This contrasts to Example 12, where $D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbf{P})$ is BS-minimizable *in the narrow sense* for all the other cases $\gamma \in \mathbb{R}\backslash]1,2[$ and $\widetilde{c} \in]0,\infty[$ (cf. Broniatowski & Stummer [1]).

Remark 29: (a) As a side effect of the above considerations (in slightly more general form), for all parameters $\widetilde{c} \in]0,\infty[$, $\beta \in]0,1[$ and $\gamma \in]1,\infty[$ we have shown the new divergence inequality

$$\text{for all } \mathbf{Q} \in \mathbb{R}^K \text{ and } \mathbf{P} \in \mathbb{R}^K_{>0} \text{ there holds } \ D_{\varphi_{\beta,\beta,\frac{\widetilde{c}}{\gamma}}}(\mathbf{Q},\mathbf{P}) \leq D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P}) \quad \text{with equality if and only if } \mathbf{Q} = \mathbf{P}\,.$$

(b) For the restricted cases where $\mathbf{\Omega} \subset \mathbb{R}^K_{>0}$ and $\gamma \in]1,2[$, one can e.g. alternatively use as the base-divergence-generator $\varphi := \frac{\widetilde{c}}{\gamma}\cdot\varphi_1$ and employ the well-known bound (cf. e.g. Liese & Vajda [18])

$$\text{for all } \gamma \in]1,\infty[,\ \widetilde{c} \in]0,\infty[\text{ and } t \in\ [0,\infty[\text{ there holds } \frac{\widetilde{c}}{\gamma}\cdot\varphi_1(t) \leq \widetilde{c}\cdot\varphi_\gamma(t) \quad \text{with equality if and only if } t = 1\,, \tag{59}$$

and hence

$$\text{for all } \mathbf{Q} \in \mathbb{R}^K_{\geq 0} \text{ and } \mathbf{P} \in \mathbb{R}^K_{>0} \text{ there holds } \ D_{\frac{\widetilde{c}}{\gamma}\cdot\varphi_1}(\mathbf{Q},\mathbf{P}) \leq D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P}) \quad \text{with equality if and only if } \mathbf{Q} = \mathbf{P}\,.$$

Thus, we can proceed analogously to Example 26, with the difference that $\widetilde{\frac{\widetilde{c}}{\gamma}\cdot\varphi_1} := M_{\mathbf{P}}\cdot\frac{\widetilde{c}}{\gamma}\cdot\varphi_1$ and hence the $\widetilde{W}_i$'s are i.i.d. copies of the random variable $\widetilde{W}$ of the form $\widetilde{W} = \frac{\gamma}{\widetilde{c}\cdot M_{\mathbf{P}}}\cdot Z$ for a Poisson $POI(\frac{\widetilde{c}\cdot M_{\mathbf{P}}}{\gamma})-$distributed random variable Z (cf. Section XII of Broniatowski & Stummer [1], see also Table 1). With these choices, (50) specializes accordingly.
(c) As a side effect, we can even show the new divergence bounds

$$\text{for all } \gamma \in]1,\infty[,\ \beta \in]0,\frac{8}{5}],\ \widetilde{c} \in]0,\infty[\text{ and } t \in\ [0,\infty[\text{ there holds } \varphi_{\beta,\beta,\frac{\widetilde{c}}{\gamma}}(t) \leq \frac{\widetilde{c}}{\gamma}\cdot\varphi_1(t) \quad \text{with equality if and only if } t = 1, \tag{60}$$

and hence

$$\text{for all } \mathbf{Q} \in \mathbb{R}^K_{\geq 0} \text{ and } \mathbf{P} \in \mathbb{R}^K_{>0} \text{ there holds } \ D_{\varphi_{\beta,\beta,\frac{\widetilde{c}}{\gamma}}}(\mathbf{Q},\mathbf{P}) \leq D_{\frac{\widetilde{c}}{\gamma}\cdot\varphi_1}(\mathbf{Q},\mathbf{P}) \quad \text{with equality if and only if } \mathbf{Q} = \mathbf{P}\,;$$

see Appendix A for a proof.

C. Minimization via Base-Divergence-Method 2

Due to our investigations in Section IV, as an alternative to the new Theorem 22 we can also derive the following new assertions, by using a different base-divergence (function):

Theorem 30: Let us arbitrarily fix some $\mathbf{P} \in \mathbb{R}^K_{>0}$, $M_{\mathbf{P}} > 0$, $\mathbf{Q}^{**}$, φ, $\widetilde{\zeta}$, $\mathbf{\Omega}$, $\widetilde{V}$ and $\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ (cf. (27)) as in Theorem 16; we call the corresponding $D^{SBD}_{\varphi,\mathbf{P}}(\cdot,\mathbf{Q}^{**})$ (cf. (23)) the *base-SBD-divergence* (function).
(a) Furthermore, suppose that $\mathbf{\Omega} \subset \mathbb{R}^K$ is also compact and that $\Phi : \mathbf{\Omega} \mapsto \mathbb{R}$ is a continuous function on $\mathbf{\Omega}$. Then, there holds

$$\inf_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \;=\; -\lim_{n\to\infty}\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}},\mathbf{Q}^{**}\Big) - \Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot 1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\ \Big]\Big) \tag{61}$$

and the infimum is attained at some (not necessarily unique) point in $\mathbf{\Omega}$. In particular, the function $\Phi(\cdot)$ is bare-simulation minimizable (BS-minimizable) on $\mathbf{\Omega}$ (cf. (2) in Definition 1), *in terms of the SBD method*.
(b) If $\mathbf{\Omega} \subset \mathbb{R}^K$ is not necessarily compact and $\Phi : \mathbf{\Omega} \mapsto \mathbb{R}$ is a continuous function which satisfies the lower-bound condition

$$\text{there exists a constant } c_1 \in \mathbb{R} \text{ such that for all } \mathbf{Q} \in \mathbf{\Omega} \text{ there holds} \quad \Phi(\mathbf{Q}) \geq D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q}, \mathbf{Q}^{**}) - c_1\,,$$

then the representation/convergence (61) — and hence the corresponding BS-minimizability — still holds, but the infimum may not necessarily be attained/reached at some point in $\mathbf{\Omega}$.

The proof of Theorem 30 is given in Appendix A below.

Remark 31: (a) In Theorem 30 we have allowed for the special case that $\mathbf{Q}^{**}$ can be in $\mathbf{\Omega}$ (and thus, $\inf_{\mathbf{Q}\in\mathbf{\Omega}} D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q}, \mathbf{Q}^{**}) = 0$ of Remark 17 applies); however, in such a situation one gets $\inf_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \neq 0$ in general.
(b) For the special case $\Phi(\cdot) := D^{SBD}_{\varphi,\mathbf{P}}(\cdot, \mathbf{Q}^{**})$, the exponent in (61) becomes zero and then Theorem 30(b) collapses to Theorem 16.

Alternatively to (48), the limit statement (61) provides another principle for the approximation of the solution of the minimization problem $\Phi(\mathbf{\Omega}) := \inf_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$. Indeed, by replacing the right-hand side in (61) by its finite counterpart, we get for given large n

$$-\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\tilde{\mathbf{V}}}_n, \mathbf{Q}^{**}\Big) - \Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\tilde{\mathbf{V}}}_n\big)\Big)\Big)\cdot 1\!\!1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\tilde{\mathbf{V}}}_n\big)\ \Big]\Big) \ \approx\ \inf_{Q\in\mathbf{\Omega}} \Phi(\mathbf{Q}); \tag{62}$$

it remains to estimate the left-hand side of (62) (see Section XII below, where the latter also provides estimates of the *minimizers*).

D. Maximization via Base-Divergence-Method 1

In the previous two subsections, for the above-mentioned contexts (D1) to (D8) — and beyond — we have only dealt with BS-minimizability so far. For their *BS-maximizability* we obtain the following new fundamental

Theorem 32: Let us arbitrarily fix some $\mathbf{P} \in \mathbb{R}^K_{>0}$, $M_{\mathbf{P}} > 0$, φ, $\widetilde{\zeta}$, $\widetilde{W} := (\widetilde{W}_i)_{i\in\mathbb{N}}$ and $\boldsymbol{\xi}^{\widetilde{\mathbf{W}}}_n$ (cf. (13)) as in Theorem 9; recall that we have named $D_\varphi(\cdot, \mathbf{P})$ as the corresponding *base-divergence* (function).
(a) Furthermore, suppose that $\mathbf{\Omega} \subset \mathbb{R}^K$ is compact and satisfies the regularity properties (6), and that $\Phi : \mathbf{\Omega} \mapsto \mathbb{R}$ is a continuous function on $\mathbf{\Omega}$. Then, there holds

$$\sup_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \ =\ \lim_{n\to\infty} \frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{W}}}_n, \mathbf{P}\big) + \Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{W}}}_n\big)\Big)\Big)\cdot 1\!\!1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{W}}}_n\big)\ \Big]\Big) \tag{63}$$

and the supremum is attained at some (not necessarily unique) point in $\mathbf{\Omega}$. In particular, the function $\Phi(\cdot)$ is bare-simulation maximizable (BS-maximizable) on $\mathbf{\Omega}$ (cf. (3) in Definition 1).
(b) If $\mathbf{\Omega} \subset \mathbb{R}^K$ is not necessarily compact but satisfies only the regularity properties (6), and $\Phi : \mathbf{\Omega} \mapsto \mathbb{R}$ is a continuous function which satisfies the upper-bound condition

$$\text{there exists a constant } c_1 \in \mathbb{R} \text{ such that for all } \mathbf{Q} \in \mathbf{\Omega} \text{ there holds} \quad \Phi(\mathbf{Q}) \leq c_1 - D_\varphi(\mathbf{Q}, \mathbf{P})\,,$$

then the representation/convergence (63) — and hence the corresponding BS-maximizability — still holds, but the supremum may not necessarily be attained/reached at some point in $\mathbf{\Omega}$.

The proof of Theorem 32 will be given in Appendix A below.

Remark 33: (i) In Theorem 32(b), one gets (63) with exponent $D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{W}}}_n, \mathbf{P}\big) + \Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{W}}}_n\big) \leq c_1$, which turns into an *exponential dampening* in case of $c_1 \leq 0$.
(ii) In Theorem 32 we have allowed for the special case that $\mathbf{P}$ can be in $\mathbf{\Omega}$.

Analogously to (48), the limit statement (63) provides the principle for the approximation of the solution of the maximization problem $\Phi(\mathbf{\Omega}) := \sup_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$. This can be achieved by replacing the right-hand side in (63) by its finite counterpart, from which we obtain for given large n

$$\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{W}}}_n, \mathbf{P}\big) + \Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{W}}}_n\big)\Big)\Big)\cdot 1\!\!1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{W}}}_n\big)\ \Big]\Big) \ \approx\ \sup_{Q\in\mathbf{\Omega}} \Phi(\mathbf{Q}); \tag{64}$$

it remains to estimate the left-hand side of (64) (see Section XII below, where the latter also provides estimates of the *maximizers*).

E. Maximization via Base-Divergence-Method 2

Due to our investigations in Section IV, as an alternative to the new Theorem 32 we can also derive the following new assertions, by employing a different base-divergence (function):

Theorem 34: Let us arbitrarily fix some $\mathbf{P} \in \mathbb{R}^K_{>0}$, $M_{\mathbf{P}} > 0$, $\mathbf{Q}^{**}$, φ, $\widetilde{\mathfrak{l}}$, $\mathbf{\Omega}$, $\widetilde{V}$ and $\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ (cf. (27)) as in Theorem 16; recall that we have named $D^{SBD}_{\varphi,\mathbf{P}}(\cdot,\mathbf{Q}^{**})$ as the corresponding *base-SBD-divergence* (function).
(a) Furthermore, suppose that $\mathbf{\Omega} \subset \mathbb{R}^K$ is also compact and that $\Phi : \mathbf{\Omega} \mapsto \mathbb{R}$ is a continuous function on $\mathbf{\Omega}$. Then, there holds

$$\sup_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \;=\; \lim_{n\to\infty} \frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}, \mathbf{Q}^{**}\Big) + \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big) \cdot \mathfrak{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big) \Big]\Big) \tag{65}$$

and the supremum is attained at some (not necessarily unique) point in $\mathbf{\Omega}$. In particular, the function $\Phi(\cdot)$ is bare-simulation maximizable (BS-maximizable) on $\mathbf{\Omega}$ (cf. (3) in Definition 1), *in terms of the SBD method*.
(b) If $\mathbf{\Omega} \subset \mathbb{R}^K$ is not necessarily compact and $\Phi : \mathbf{\Omega} \mapsto \mathbb{R}$ is a continuous function which satisfies the upper-bound condition

$$\text{there exists a constant } c_1 \in \mathbb{R} \text{ such that for all } \mathbf{Q} \in \mathbf{\Omega} \text{ there holds} \quad \Phi(\mathbf{Q}) \leq c_1 - D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}),$$

then the representation/convergence (65) — and hence the corresponding BS-maximizability — still holds, but the supremum may not necessarily be attained/reached at some point in $\mathbf{\Omega}$.

The proof of Theorem 34 is given in Appendix A below.

Remark 35: In Theorem 34 we have allowed for the special case that $\mathbf{Q}^{**}$ can be in $\mathbf{\Omega}$ (and thus, $\inf_{\mathbf{Q}\in\mathbf{\Omega}} D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) = 0$ of Remark 17 applies).

Analogously to (64), the limit statement (65) provides the principle for the approximation of the solution of the maximization problem $\Phi(\mathbf{\Omega}) := \sup_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$. Indeed, by replacing the right-hand side in (65) by its finite counterpart, we derive for given large n

$$\frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}, \mathbf{Q}^{**}\Big) + \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big) \cdot \mathfrak{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big) \Big]\Big) \;\approx\; \sup_{Q\in\mathbf{\Omega}} \Phi(\mathbf{Q}); \tag{66}$$

it remains to estimate the left-hand side of (66) (see Section XII below, where the latter also provides estimates of the *maximizers*).

VI. Deterministic Narrow-Sense Bare-Simulation-Optimization of $\varphi-$divergences with Constant-Component-Sum Side Constraint

Recall that we have denoted by $\mathbb{S}^K := \{\mathbf{Q} := (q_1,\ldots,q_K) \in \mathbb{R}^K_{\geq 0} : \sum_{i=1}^K q_i = 1\}$ the simplex of probability vectors (probability simplex) and its interior by $\mathbb{S}^K_{>0} := \{\mathbf{Q} := (q_1,\ldots,q_K) \in \mathbb{R}^K_{>0} : \sum_{i=1}^K q_i = 1\}$. For better emphasis, for elements of these two sets we use the symbols $\mathbb{Q}$,$\mathbb{P}$ instead of $\mathbf{Q}$,$\mathbf{P}$, etc., but for their components we still use our notation q_k,p_k. Moreover, subsets of $\mathbb{S}^K$ or $\mathbb{S}^K_{>0}$ will be denoted by $\mathbb{\Omega}$ instead of $\mathbf{\Omega}$ etc. In this section, we work with constraint sets of the form $A \cdot \mathbb{\Omega}$ for some arbitrary $A \in]0,\infty[$ (sometimes even $A \in \mathbb{R}\backslash\{0\}$) which satisfy $int\,(A \cdot \mathbb{\Omega}) = \emptyset$ (cf. Remark 8(b)) and thus need extra refinements. In more detail, we deal with

Problem 36: For pregiven $\varphi \in \widetilde{\Upsilon}(]a,b[)$, positive-components probability vector $\mathbb{P} := (p_1,..,p_K) \in \mathbb{S}^K_{>0}$, and subset $A \cdot \mathbb{\Omega} \subset A \cdot \mathbb{S}^K$ with regularity properties — in the relative topology (!!) —

$$cl(A \cdot \mathbb{\Omega}) = cl\,(int\,(A \cdot \mathbb{\Omega}))\,, \qquad int\,(A \cdot \mathbb{\Omega}) \neq \emptyset, \tag{67}$$

find

$$\Phi_{\mathbb{P}}(A \cdot \mathbb{\Omega}) := \inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} D_\varphi(\mathbf{Q},\mathbb{P}),$$

provided that

$$\inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} D_\varphi(\mathbf{Q},\mathbb{P}) < \infty \tag{68}$$

and that divergence generator φ additionally satisfies the following Condition 37.

Condition 37: Let $\varphi \in \widetilde{\Upsilon}(]a,b[)$ and satisfy the representation

$$\varphi(t) = \sup_{z\in\mathbb{R}} \Big(z \cdot t - \log \int_{\mathbb{R}} e^{zy} d\mathfrak{l}(y) \Big), \qquad t \in \mathbb{R}, \tag{69}$$

for some probability distribution $\mathfrak{l}$ on the real line such that the function $z \mapsto MGF_{\mathfrak{l}}(z) := \int_{\mathbb{R}} e^{zy} d\mathfrak{l}(y)$ is finite on some open interval containing zero.

Remark 38: Since here $M_{\mathbb{P}} = \sum_{i=1}^{K} p_i = 1$ and hence (in the notation of the previous Sections III and V) $\widetilde{\mathbb{P}} := \mathbb{P}/M_{\mathbb{P}} = \mathbb{P}$, the Condition 5 collapses to Condition 37.

For the directed-distance-minimization Problem 36, we proceed (mostly analogously to Section III above) by constructing an appropriate sequence $(\boldsymbol{\xi}_n)_{n\in\mathbb{N}}$ of $\mathbb{R}^K-$valued random variables (cf. (2) in Definition 1 and the special case of Remark 2(a)) as follows: for any $n \in \mathbb{N}$ and any $k \in \{1,\ldots,K-1\}$, let $n_k := \lfloor n \cdot p_k \rfloor$ (where $\lfloor x \rfloor$ denotes the integer part of x) and $n_K := n - \sum_{k=1}^{K-1} n_k$; for this, we assume that $n \in \mathbb{N}$ is large enough, namely $n \geq \max_{k\in\{1,\ldots,K\}} \frac{1}{p_k}$, such that all the integers n_k $(k = 1,\ldots,K)$ are non-zero. Since we assume $\mathbf{P} \in \mathbb{R}^K_{>0}$ and thus none of the p_k's is zero, one has

$$\lim_{n\to\infty} \frac{n_k}{n} = p_k, \qquad k = 1,\ldots,K.$$

With this at hand, we decompose the set $\{1,\ldots,n\}$ of all integers from 1 to n into the following disjoint blocks: $I_1^{(n)} := \{1,\ldots,n_1\}$, $I_2^{(n)} := \{n_1+1,\ldots,n_1+n_2\}$, and so on until the last block $I_K^{(n)} := \{\sum_{k=1}^{K-1} n_k + 1,\ldots,n\}$ which therefore contains all integers from $n_1+\ldots+n_{K-1}+1$ to n. Due to our construction, $I_k^{(n)}$ has $n_k \geq 1$ elements (i.e. $card(I_k^{(n)}) = n_k$) for all $k \in \{1,\ldots,K\}$ [15]. Furthermore, consider a vector $\mathbf{W} := (W_1,\ldots,W_n)$ where the W_i's are i.i.d. copies of the random variable W whose distribution is associated with the divergence-generator φ through (69), in the sense that $\mathbb{\Pi}[W \in \cdot\,] = \mathbb{\wr}[\,\cdot\,]$. We group the W_i's according to the above-mentioned blocks and sum them up blockwise, in order to build the following $K-$component random vector (instead of $\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}$ in (13))

$$\boldsymbol{\xi}_n^{w\mathbf{W}} \quad := \quad \begin{cases} \left(\frac{\sum_{i\in I_1^{(n)}} W_i}{\sum_{k=1}^{K}\sum_{i\in I_k^{(n)}} W_i}, \ldots, \frac{\sum_{i\in I_K^{(n)}} W_i}{\sum_{k=1}^{K}\sum_{i\in I_k^{(n)}} W_i} \right), & \text{if } \sum_{j=1}^{n} W_j \neq 0, \\ (\infty,\ldots,\infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^{n} W_j = 0. \end{cases} \tag{70}$$

Remark 39: (i) (Concerning e.g. computer-program command availability) In case of $\sum_{j=1}^{n} W_j = 0$, in (70) we may equivalently assign to $\boldsymbol{\xi}_n^{w\mathbf{W}}$ any vector outside of $\boldsymbol{\mathfrak{Q}}$ instead of $\boldsymbol{\infty}$.
(ii) By construction, in case of $\sum_{j=1}^{n} W_j \neq 0$, the sum of the random K vector components of (70) is now automatically equal to 1, but — as (depending on φ) the W_i's may take both positive and negative values — these random components may be negative with probability strictly greater than zero (respectively nonnegative with probability strictly less than 1). However, $\mathbb{\Pi}[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{W}}} \in \mathbb{S}^K_{>0}] > 0$ since all the (identically distributed) random variables W_i have expectation 1 (as a consequence of the assumed representability (69); in case of $\mathbb{\Pi}[W_1 > 0] = 1$ one has even $\mathbb{\Pi}[\boldsymbol{\xi}_n^{w\mathbf{W}} \in \mathbb{S}^K_{>0}] = 1$. Summing up things, the probability $\mathbb{\Pi}[\boldsymbol{\xi}_n^{w\mathbf{W}} \in \boldsymbol{\mathfrak{Q}}] = \mathbb{\Pi}[A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}} \in A \cdot \boldsymbol{\mathfrak{Q}}]$ is strictly positive and finite at least for large n, whenever $\Phi_{\mathbb{P}}(\boldsymbol{\mathfrak{Q}}) = \inf_{\mathbb{Q}\in\boldsymbol{\mathfrak{Q}}} D_\varphi(\mathbb{Q},\mathbb{P})$ is finite.

With the above-mentioned ingredients, we have proven in Theorem 12 of Broniatowski & Stummer [1] (an even more general, conditional-expectations-involving version of) the following assertion:

Theorem 40: Let $\mathbb{P} \in \mathbb{S}^K_{>0}$, and suppose that the divergence generator φ satisfies Condition 37. Additionally, let $(W_i)_{i\in\mathbb{N}}$ be a family of independent and identically distributed $\mathbb{R}-$valued random variables with probability distribution $\mathbb{\wr}[\,\cdot\,] := \mathbb{\Pi}[W_1 \in \cdot\,]$ being connected with the divergence generator $\varphi \in \Upsilon(]a,b[)$ via the representability (69). Then there holds (in terms of (70))

$$\inf_{\mathbb{Q}\in\boldsymbol{\mathfrak{Q}}} \inf_{m\neq 0} D_\varphi(m\cdot\mathbb{Q},\mathbb{P}) = \inf_{m\neq 0} \inf_{\mathbb{Q}\in\boldsymbol{\mathfrak{Q}}} D_\varphi(m\cdot\mathbb{Q},\mathbb{P}) = -\lim_{n\to\infty} \frac{1}{n} \log \mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\mathbf{W}} \in \boldsymbol{\mathfrak{Q}}\Big] \tag{71}$$

for all sets $\boldsymbol{\mathfrak{Q}}$ satisfying the regularity properties (67) *in the relative topology* and the finiteness property (68) (the latter two with $A = 1$).

From this and with the help of (10), one gets immediately

Corollary 41: Let $A \in \mathbb{R}\backslash\{0\}$, $\mathbf{P} \in \mathbb{R}^K_{>0}$ with $M_{\mathbf{P}} := \sum_{i=1}^{K} p_i > 0$, and suppose that the divergence generator φ satisfies Condition 5. Additionally, let $(\widetilde{W}_i)_{i\in\mathbb{N}}$ be a family of independent and identically distributed $\mathbb{R}-$valued random variables with probability distribution $\widetilde{\mathbb{\wr}}[\cdot\,] := \mathbb{\Pi}[\widetilde{W}_1 \in \cdot\,]$ being connected with the divergence generator $\varphi \in \Upsilon(]a,b[)$ via the representability (9). Then there holds (in terms of (70) with $\widetilde{W}$ instead of W)

$$\inf_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}} \inf_{m\neq 0} D_\varphi(m\cdot\mathbf{Q},\mathbf{P}) = \inf_{m\neq 0} \inf_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}} D_\varphi(m\cdot\mathbf{Q},\mathbf{P}) = -\lim_{n\to\infty} \frac{1}{n} \log \mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{W}}} \in \boldsymbol{\mathfrak{Q}}\Big] \tag{72}$$

for all sets $\boldsymbol{\mathfrak{Q}}$ satisfying the regularity properties (67) *in the relative topology* and the (68).

[15] if all p_k $(k = 1,\ldots,K)$ are rational numbers in $]0,1[$ with $\sum_{k=1}^{K} p_k = 1$ and N is the (always existing) smallest integer such that all $N \cdot p_k$ $(k = 1,\ldots,K)$ are integers (i.e. $\in \mathbb{N}$), then for any multiple $n = m \cdot N$ $(m \in \mathbb{N})$ one gets that all $n \cdot p$ are integers and hence $n_k = \lfloor n \cdot p_k \rfloor = n \cdot p_k$ $(k = 1,\ldots,K)$

Remark 42: (a) As shown in [1], the involved "inner" $m-$minimizations on the left-hand side of (71) (and analogously, of (72)) can be solved explicitly in the important special case of the power divergences (17).
(b) Even more, we have worked out in [1] that the outcomes of (a) can be rewritten in terms of invertible functions $F(\cdot)$ of the divergences $D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P})$ (and analogously of $D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P})$).

Let us illuminate the details of the previous Remark 42. The required representability (69) is satisfied for all (multiple of) the generators $\varphi(\cdot):=\widetilde{c}\cdot\varphi_\gamma(\cdot)$ of (16) with $\widetilde{c}\in]0,\infty[$ and $\gamma\in\mathbb{R}\backslash]1,2[$ (cf. Broniatowski & Stummer [1]); the corresponding crucial probability laws $\mathbb{Q}$ can be found in Table 1. The corresponding generalized power divergences $D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P})$ given by (17) will be used as tools to derive our base-divergences for our new fundamental Theorem 52 below. In order to obtain this, for fixed $A\in]0,\infty[$ and $\mathbb{P}\in\mathbb{S}^K_{>0}$ we employ the following auxiliary notations:

(S1) $\widetilde{\mathcal{M}}_\gamma:=A\cdot\mathbb{S}^K_{>0}$ for $\gamma\in]-\infty,0]$, respectively, $\widetilde{\mathcal{M}}_\gamma:=A\cdot\mathbb{S}^K$ for $\gamma\in]0,1]\cup[2,\infty[$.
(S2) the *modified $\gamma-$order Hellinger integral of $\mathbf{Q}$ and $\mathbb{P}$* given by

$$0<H_\gamma(\mathbf{Q},\mathbb{P}):=\sum_{k=1}^K(q_k)^\gamma\cdot(p_k)^{1-\gamma}\;=\;1+\gamma\cdot(A-1)+\gamma\cdot(\gamma-1)\cdot D_{\varphi_\gamma}(\mathbf{Q},\mathbb{P}),\quad \gamma\in]-\infty,0[\cup]0,1[\cup[2,\infty[,\;\mathbf{Q}\in\widetilde{\mathcal{M}}_\gamma;$$

(S3) for $\gamma\in]-\infty,0[\cup]0,1[\cup[2,\infty[$ and $\widetilde{c}\in]0,\infty[$ we define the function $F_{\gamma,\widetilde{c},A}:]-\infty,\infty[\mapsto]-\infty,\infty]$ by

$$F_{\gamma,\widetilde{c},A}(x)\;:=\;\begin{cases}\frac{\widetilde{c}}{\gamma}\cdot\left\{1-A^{\gamma/(\gamma-1)}\cdot\left[1+\gamma\cdot(A-1)+\frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x\right]^{-1/(\gamma-1)}\right\},\\ \qquad\text{if } \gamma\in]-\infty,0[\cup]0,1[\text{ and } 1+\gamma\cdot(A-1)+\frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x\geq 0,\\ \qquad\text{or if } \gamma\in[2,\infty[\text{ and } 1+\gamma\cdot(A-1)+\frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x>0,\\ \infty,\qquad\text{if } \gamma\in[2,\infty[\text{ and } 1+\gamma\cdot(A-1)+\frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x\leq 0,\\ \infty,\qquad\text{if } \gamma\in]-\infty,0[\cup]0,1[\text{ and } 1+\gamma\cdot(A-1)+\frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x<0.\end{cases}\tag{73}$$

Notice that in the constellations of (S2) only the first line of (73) is relevant, and we can even show $F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\Big)\in[0,\infty[$ for all $\mathbf{Q}\in\widetilde{\mathcal{M}}_\gamma$ (cf. Broniatowski & Stummer [1]). Moreover, the function $x\mapsto F_{\gamma,\widetilde{c},A}(x)$ is on its effective domain $dom(F_{\gamma,\widetilde{c},A}):=\{x:F_{\gamma,\widetilde{c},A}(x)\in]-\infty,\infty[\}$ strictly increasing and has the strictly increasing inverse function

$$F^{\leftarrow}_{\gamma,\widetilde{c},A}(z):=\frac{\widetilde{c}}{\gamma\cdot(\gamma-1)}\cdot\left\{A^\gamma\cdot\left[1-\frac{\gamma}{\widetilde{c}}\cdot z\right]^{-(\gamma-1)}-1-\gamma\cdot(A-1)\right\},\qquad\text{for all } z\in\mathbb{R}\text{ such that }\gamma\cdot z\leq\widetilde{c}.$$

(S4) for $\gamma=1$ we employ for $\widetilde{c}\in]0,\infty[$ the strictly increasing function $F_{1,\widetilde{c},A}:]-\infty,\infty[\mapsto]-\infty,\infty[$ defined by

$$F_{1,\widetilde{c},A}(x):=\widetilde{c}\cdot\left\{1-A\cdot\exp\left(\frac{1}{A}-1-\frac{x}{A\cdot\widetilde{c}}\right)\right\}.\tag{74}$$

Notice that $F_{1,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_1}(\mathbf{Q},\mathbb{P})\Big)\in[0,\infty[$ for all $\mathbf{Q}\in\widetilde{\mathcal{M}}_1$ (cf. Broniatowski & Stummer [1]). The corresponding inverse is given by

$$F^{\leftarrow}_{1,\widetilde{c},A}(z):=\widetilde{c}\cdot\left\{1-A-A\cdot\left[\log\left(1-\frac{z}{\widetilde{c}}\right)-\log A\right]\right\},\qquad\text{for all } z\in]-\infty,\widetilde{c}[.$$

(S5) for $\gamma=0$ we employ for $\widetilde{c}\in]0,\infty[$ the strictly increasing function $F_{0,\widetilde{c},A}:]-\infty,\infty[\mapsto]-\infty,\infty[$ defined by

$$F_{0,\widetilde{c},A}(x):=\widetilde{c}\cdot(1-A+\log A)+x.$$

Notice that $F_{0,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_1}(\mathbf{Q},\mathbb{P})\Big)\in[0,\infty[$ for all $\mathbf{Q}\in\widetilde{\mathcal{M}}_0$ (cf. Broniatowski & Stummer [1]). The corresponding inverse is given by

$$F^{\leftarrow}_{0,\widetilde{c},A}(z):=z-\widetilde{c}\cdot(1-A+\log A),\qquad\text{for all } z\in\mathbb{R}.$$

Remark 43: For the special case $A=1$, $D(\cdot,\cdot):=F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\cdot)\Big)$ is a divergence on $\mathfrak{M}$ (i.e. for all $\mathbb{Q},\mathbb{P}\in\mathfrak{M}$ there holds $F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P})\Big)\geq 0$ with equality if and only if $\mathbb{Q}=\mathbb{P}$). Moreover, due to the above-mentioned strict increasingness of $F_{\gamma,\widetilde{c},1}$, the (not necessarily unique) minimizer respectively maximizer of $D(\cdot,\mathbb{P})$ and $D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbb{P})$ on (say) compact $\mathfrak{M}$ coincide.

For such a context, Broniatowski & Stummer [1] (cf. Theorem 12 [16], Formula (39) and Lemma 14 therein) obtain the following assertion on BS-minimizability in the narrow sense:

[16] which coincides with the above Theorem 40

Theorem 44: Let $\mathbb{P} \in \mathbb{S}_{>0}^K$, $\widetilde{c} \in]0,\infty[$, $\gamma \in \mathbb{R}\setminus]1,2[$ and $A \in]0,\infty[$ be arbitrary but fixed. Moreover, let $(W_i)_{i\in\mathbb{N}}$ be a family of independent and identically distributed $\mathbb{R}-$valued random variables with probability distribution $\mathfrak{l}[\cdot\,] := \mathbb{\Pi}[W_1 \in \cdot\,]$ being connected with the divergence generator $\varphi := \widetilde{c}\cdot\varphi_\gamma(\cdot) \in \widetilde{\Upsilon}(]a,b[)$ via the representability (69).
(a) Then there holds

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{W}} F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\Big) = -\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\mathbf{W}}\in\mathfrak{W}\Big] \tag{75}$$
$$= \inf_{\mathbb{Q}\in\mathfrak{W}}\inf_{m\neq 0} D_{\widetilde{c}\cdot\varphi_\gamma}(m\cdot\mathbb{Q},\mathbb{P}) = \inf_{\mathbf{Q}\in A\cdot\mathfrak{W}}\inf_{\widetilde{m}\neq 0} D_{\widetilde{c}\cdot\varphi_\gamma}(\widetilde{m}\cdot\mathbf{Q},\mathbb{P}) = -\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}\in A\cdot\mathfrak{W}\Big]$$

for all sets $A\cdot\mathfrak{W}\subset\widetilde{\mathcal{M}}_\gamma$ satisfying the regularity properties (67) *in the relative topology*. In particular, for each $\mathbb{P}\in\mathbb{S}_{>0}^K$ the function $\Phi_{\mathbb{P}}(\cdot) := F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbb{P})\Big)$ is bare-simulation minimizable (BS-minimizable) in the narrow sense (cf. (2) in Definition 1 and the special case of Remark 2(a)) on all sets $A\cdot\mathfrak{W}\subset\widetilde{\mathcal{M}}_\gamma$ satisfying (67) *in the relative topology*.
(b) Moreover, there holds

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{W}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P}) = F^{\leftarrow}_{\gamma,\widetilde{c},A}\Big(-\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\mathbf{W}}\in\mathfrak{W}\Big]\Big) \tag{76}$$

for all sets $A\cdot\mathfrak{W}\subset\widetilde{\mathcal{M}}_\gamma$ satisfying the regularity properties (67) *in the relative topology*. In particular, for each $\mathbb{P}\in\mathbb{S}_{>0}^K$ the function $\Phi_{\mathbb{P}}(\cdot) := D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbb{P})$ is bare-simulation minimizable (BS-minimizable) in the narrow sense on all sets $A\cdot\mathfrak{W}\subset\widetilde{\mathcal{M}}_\gamma$ satisfying (67) *in the relative topology*.

Remark 45: (a) By straightforward rescaling (cf. Corollary 41 above), for $\mathbf{P}\in\mathbb{R}_{>0}^K$ with $M_{\mathbf{P}} := \sum_{i=1}^K p_i$, $\widetilde{c}\in]0,\infty[$, $\gamma\in\mathbb{R}\setminus]1,2[$ and $A\in]0,\infty[$ one can generalize (76) to

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{W}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P}) = F^{\leftarrow}_{\gamma,\breve{c},\breve{A}}\Big(-\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{W}}}\in\mathfrak{W}\Big]\Big) \tag{77}$$

for all sets $A\cdot\mathfrak{W}\subset\widetilde{\mathcal{M}}_\gamma$ satisfying the regularity properties (67) *in the relative topology*; here, $\breve{c} := M_{\mathbf{P}}\cdot\widetilde{c}$, $\breve{A} := \frac{A}{M_{\mathbf{P}}}$, and $(\widetilde{W}_i)_{i\in\mathbb{N}}$ is a family of independent and identically distributed $\mathbb{R}-$valued random variables with probability distribution $\widetilde{\mathfrak{l}}[\cdot\,] := \mathbb{\Pi}[W_1\in\cdot\,]$ being connected with the divergence generator $\varphi := \widetilde{c}\cdot\varphi_\gamma(\cdot)$ via the representability (9).
(b) For the case $\gamma = 2$ in Theorem 44 as well as in (a) of this remark, one can even take $A\in\mathbb{R}\setminus\{0\}$ instead of $A\in]0,\infty[$.
(c) For applications of Theorem 44 to fuzzy sets and basic belief assignments, see Broniatowski & Stummer [164].

Analogously to (14), the limit statement (77) provides the principle for the approximation of the solution of the minimization problem $\Phi_{\mathbf{P}}(\mathfrak{Q}) := \inf_{\mathbf{Q}\in A\cdot\mathfrak{W}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P})$. Indeed, by replacing the right-hand side in (77) by its finite counterpart, we deduce for given large n

$$F^{\leftarrow}_{\gamma,\breve{c},\breve{A}}\Big(-\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{W}}}\in\mathfrak{W}\Big]\Big) \;\approx\; \inf_{\mathbf{Q}\in A\cdot\mathfrak{W}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P}); \tag{78}$$

it remains to estimate the left-hand side of (78) (see Section XIII below, where the latter also provides estimates of the *minimizers*).

VII. Deterministic Narrow-Sense Bare-Simulation-Optimization of Bregman Distances with Constant-Component-Sum Side Constraint

In the previous Section VI, we have recalled/summarized recently achieved (cf. Broniatowski & Stummer [1]) *narrow-sense* bare-simulation minimization results on CASM $\varphi-$divergences $\mathbf{Q}\mapsto D_\varphi(\mathbf{Q},\mathbf{P})$ under the *additional constraints* that all the components q_k are nonnegative/strictly positive and that $\sum_{k=1}^K q_k = A$ for some (say) $A>0$, where we have mostly concentrated on the explicitly solvable subcase $\varphi = \widetilde{c}\cdot\varphi_\gamma$ ($\widetilde{c}\in]0,\infty[$, $\gamma\in\mathbb{R}\setminus]1,2[$). On the other hand, independently of any constraints, we have seen in Section IV that the $\varphi-$divergences generalize to the scaled Bregman distances (cf. (23)) $\mathbf{Q}\mapsto D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$ (recall that $D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{P}) = D_\varphi(\mathbf{Q},\mathbf{P})$). In the following, we work out how the results of Section VI generalize to the bare-simulation minimization of these scaled Bregman distances under the above-mentioned additional constraints. We first achieve

Theorem 46: Let $\mathbb{P}\in\mathbb{S}_{>0}^K$, and suppose that the divergence generator φ satisfies Condition 37. Additionally, let $\mathbf{Q}^{**}\in\mathbb{R}^K$ such that (21) holds. Moreover, we assume that $\mathfrak{W}\subset\mathbb{S}^K$ satisfies the regularity properties (67) in the relative topology as well as the finiteness property

$$\inf_{\mathbb{Q}\in\mathfrak{W}} D^{SBD}_{\varphi,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**}) < \infty. \tag{79}$$

Additionally, let $V := (\mathbf{V}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors constructed via (25) and (26) (where we write V instead of $\widetilde{V}$, since $M_{\mathbb{P}} = 1$ and thus $\widetilde{p}_k = p_k/M_{\mathbb{P}} = p_k$ as well as $\widetilde{\mathfrak{l}} = \mathfrak{l}$). Then, in terms of the random vectors $\boldsymbol{\xi}_n^{w\mathbf{V}}$ given by

$$\boldsymbol{\xi}_n^{w\mathbf{V}} := \begin{cases} \left(\frac{\sum_{i\in I_1^{(n)}} V_i}{\sum_{k=1}^K \sum_{i\in I_k^{(n)}} V_i}, \ldots, \frac{\sum_{i\in I_K^{(n)}} V_i}{\sum_{k=1}^K \sum_{i\in I_k^{(n)}} V_i}\right), & \text{if } \sum_{j=1}^n V_j \neq 0, \\ (\infty,\ldots,\infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^n V_j = 0, \end{cases} \tag{80}$$

there holds

$$\inf_{\mathbb{Q}\in\boldsymbol{\mathfrak{Q}}} \inf_{m\neq 0} D_{\varphi,\mathbb{P}}^{SBD}(m\cdot\mathbb{Q},\mathbf{Q}^{**}) = \inf_{m\neq 0} \inf_{\mathbb{Q}\in\boldsymbol{\mathfrak{Q}}} D_{\varphi,\mathbb{P}}^{SBD}(m\cdot\mathbb{Q},\mathbf{Q}^{**}) = -\lim_{n\to\infty}\frac{1}{n}\log \mathbb{\Pi}\left[\boldsymbol{\xi}_n^{w\mathbf{V}} \in \boldsymbol{\mathfrak{Q}}\right]. \tag{81}$$

The proof of Theorem 46 will be given in Appendix A below.

From Theorem 46 and with the help of (24), one gets immediately

Corollary 47: Let $A \in \mathbb{R}$, $\mathbf{P} \in \mathbb{R}_{>0}^K$ with $M_{\mathbf{P}} := \sum_{i=1}^K p_i > 0$, and suppose that the divergence generator φ satisfies Condition 5. Additionally, let $\mathbf{Q}^{**} \in \mathbb{R}^K$ such that (21) holds. Moreover, we assume that $\boldsymbol{\mathfrak{Q}} \subset \mathbb{S}^K$ satisfies the regularity properties (67) in the relative topology as well as the finiteness property

$$\inf_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}} D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) < \infty.$$

Additionally, let $\widetilde{V} := (\widetilde{\mathbf{V}}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors constructed via (25) and (26). Then there holds (in terms of (80) with $\widetilde{V}$ instead of V)

$$\inf_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}} \inf_{m\neq 0} D_{\varphi,\mathbf{P}}^{SBD}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) = \inf_{m\neq 0} \inf_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}} D_{\varphi,\mathbf{P}}^{SBD}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) = -\lim_{n\to\infty}\frac{1}{n}\log \mathbb{\Pi}\left[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}} \in \boldsymbol{\mathfrak{Q}}\right]. \tag{82}$$

Remark 48: (a) *Analogously* to Remark 42(a), the involved "inner" $m-$minimizations on the left-hand side of (81) (and analogously, of (82)) can be solved explicitly in the important special case of the scaled Bregman power distances (30).
(b) *Contrary* to Remark 42(b), the outcomes of (a) can generally *NOT* be rewritten in terms of invertible functions $F(\cdot)$ of the divergences $D_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}^{SBD}(\mathbb{Q},\mathbf{Q}^{**})$ (and analogously of $D_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**})$). This means that we can generally *NOT* use our *narrow-sense* bare-simulation procedure for the minimization of $\mathbb{Q} \mapsto D_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}^{SBD}(\mathbb{Q},\mathbf{Q}^{**}) =: \Phi_{\mathbb{P}}(\mathbb{Q})$ (respectively $\mathbf{Q} \mapsto D_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) =: \Phi_{\mathbf{P}}(\mathbf{Q})$). However, we are able to employ our bare-simulation procedure in the (non-narrow-)sense of (2) for these problems (by applying the below-mentioned Theorem 52 or alternatively, Theorem 55 and Remark 57.
(c) For the special case $\mathbb{P} = \mathbb{Q}^{**}$ (respectively $\mathbf{P} = \mathbf{Q}^{**}$), the Theorem 46 collapses to Theorem 40 (respectively, Corollary 47 collapses to Corollary 41).
(d) Notice that the quantity $\breve{D}_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) := \inf_{m\neq 0} D_{\varphi,\mathbf{P}}^{SBD}(m\cdot\mathbf{Q},\mathbf{Q}^{**})$ satisfies the axioms of a divergence, that is, $\breve{D}_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) \geq 0$, as well as $\breve{D}_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**}) = 0$ if and only if $\mathbb{Q} = \mathbf{Q}^{**}$ (reflexivity). Hence, (82) reflects a corresponding narrow-sense bare-simulation for this.

In the following, we illuminate the details of the previous Remark 48(a),(b). The required representability (69) respectively (9) is satisfied for all the generators $\varphi(\cdot) := \widetilde{c}\cdot\varphi_\gamma(\cdot)$ of (16) with $\widetilde{c} \in]0,\infty[$ and $\gamma \in \mathbb{R}\setminus]1,2[$ (cf. Broniatowski & Stummer [1]); the corresponding crucial probability laws $\mathfrak{l}$ (respectively $\widetilde{\mathfrak{l}}$ by taking $\widetilde{c}\cdot M_{\mathbf{P}}$ instead of $\widetilde{c}$) can be found in Table 1. Moreover, we fix $A \in]0,\infty[$, $\mathbf{P} \in \mathbb{R}_{>0}^K$ and employ the following auxiliary notations:

(T1) recall from (S1) that $\widetilde{\mathcal{M}}_\gamma := A\cdot\mathbb{S}_{>0}^K$ for $\gamma \in]-\infty,0]$, respectively, $\widetilde{\mathcal{M}}_\gamma := A\cdot\mathbb{S}^K$ for $\gamma \in]0,1]\cup[2,\infty[$ and also set $\widetilde{\mathcal{M}}_\gamma := A\cdot\mathbb{S}^K$ for $\gamma \in]1,2[$ (the latter will be mainly employed for the purposes of the next theorem only), $\widetilde{\mathcal{N}}_\gamma := \mathbb{R}_{>0}^K$ for $\gamma \in \mathbb{R}^K\setminus\{2\}$, respectively, $\widetilde{\mathcal{N}}_2 := \mathbb{R}$ (for consistency of $\mathbf{Q}^{**}$ in (30) and $\mathbf{Q}^{**}$ in (85)).
(T2) the *modified (discrete) $\gamma-$order Hellinger integral of* $\mathbf{Q}$ *and* $\mathbf{P}$ given by

$$0 < H_\gamma(\mathbf{Q},\mathbf{P}) := \sum_{k=1}^K (q_k)^\gamma\cdot(p_k)^{1-\gamma}, \qquad \gamma\in\mathbb{R}\setminus\{0,1\},\ \mathbf{Q}\in\widetilde{\mathcal{M}}_\gamma; \tag{83}$$

(T3) the *modified (discrete) $\gamma-$order Hellinger integral of* $\mathbf{Q}^{**}$ *and* $\mathbf{P}$ given by

$$H_\gamma(\mathbf{Q}^{**},\mathbf{P}) = \sum_{k=1}^K (q_k^{**})^\gamma\cdot(p_k)^{1-\gamma}, \qquad \gamma\in\mathbb{R}\setminus\{0,1\},\ \mathbf{Q}^{**}\in\widetilde{\mathcal{N}}_\gamma; \tag{84}$$

(T4) the *$\gamma-$order triple power sum of* $\mathbf{Q}$, $\mathbf{Q}^{**}$ *and* $\mathbf{P}$ given by

$$T_\gamma(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P}) := \sum_{k=1}^K q_k\cdot(q_k^{**})^{\gamma-1}\cdot(p_k)^{1-\gamma}, \qquad \gamma\in\mathbb{R}\setminus\{1\},\ \mathbf{Q}\in\mathbb{R}^K,\ \mathbf{Q}^{**}\in\widetilde{\mathcal{N}}_\gamma; \tag{85}$$

(T5) the *modified Kullback-Leibler information distance (modified relative entropy)* given by

$$I(\mathbf{Q},\mathbf{Q}^{**}) := \sum_{k=1}^{K} q_k \cdot \log\left(\frac{q_k}{q_k^{**}}\right), \qquad \mathbf{Q} \in \widetilde{\mathcal{M}}_1 = A \cdot \mathbb{S}^K,\ \mathbf{Q}^{**} \in \widetilde{\mathcal{N}}_1 = \mathbb{R}^K_{>0}; \tag{86}$$

(T6) the *logarithmic* $0-$*order triple power sum of* $\mathbf{Q}$, $\mathbf{Q}^{**}$ *and* $\mathbf{P}$ given by

$$\breve{T}_0(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P}) := -\sum_{k=1}^{K} p_k \cdot \log\left(\frac{q_k}{q_k^{**}}\right), \qquad \mathbf{Q} \in \mathbb{R}^K_{>0},\ \mathbf{Q}^{**} \in \widetilde{\mathcal{N}}_0 = \mathbb{R}^K_{>0}. \tag{87}$$

In terms of (T1) to (T6), we obtain the following assertions:

Theorem 49: Let $A \in]0,\infty[$ and $\mathbf{P} \in \mathbb{R}^K_{>0}$ be arbitrarily fixed.
(a) Let $\widetilde{c} \in]0,\infty[$ be arbitrary, $\gamma \in]-\infty,0[\cup]0,1[\cup]1,\infty[$, $\mathbf{Q} \in \widetilde{\mathcal{M}}_\gamma$ and $\mathbf{Q}^{**} \in \mathbb{R}^K_{>0}$. Then one has

$$\begin{aligned}\inf_{m\neq 0} D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) &= \frac{\widetilde{c}}{\gamma}\cdot\Big[H_\gamma(\mathbf{Q}^{**},\mathbf{P}) - T_\gamma(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P})^{\gamma/(\gamma-1)}\cdot H_\gamma(\mathbf{Q},\mathbf{P})^{-1/(\gamma-1)}\Big]\\ &=: \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}),\end{aligned} \tag{88}$$

and consequently (cf. (82)) for any $\gamma \in]-\infty,0[\cup]0,1[\cup[2,\infty[$ and any subset $A\cdot\mathbb{\Omega} \subset \widetilde{\mathcal{M}}_\gamma$ with (67) one gets

$$\begin{aligned}&\inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} \frac{\widetilde{c}}{\gamma}\cdot\Big[H_\gamma(\mathbf{Q}^{**},\mathbf{P}) - T_\gamma(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P})^{\gamma/(\gamma-1)}\cdot H_\gamma(\mathbf{Q},\mathbf{P})^{-1/(\gamma-1)}\Big]\\ &= \inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) = -\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\in\mathbb{\Omega}\Big].\end{aligned} \tag{89}$$

For the case $\gamma = 2$ one can even allow for $A < 0$ and $\mathbf{Q}^{**} \in \mathbb{R}^K$.
(b) For any $\widetilde{c} > 0$, $\mathbf{Q} \in \widetilde{\mathcal{M}}_1 = A\cdot\mathbb{S}^K$ and $\mathbf{Q}^{**} \in \mathbb{R}^K_{>0}$ one gets

$$\inf_{m\neq 0} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) = \widetilde{c}\cdot\Big[M_{\mathbf{Q}^{**}} - A\cdot\exp\Big(-\frac{1}{A}\cdot I(\mathbf{Q},\mathbf{Q}^{**})\Big)\Big] =: \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) \tag{90}$$

(which exceptionally does not depend on $\mathbf{P}$) and consequently (cf. (82)) for any subset $A\cdot\mathbb{\Omega} \subset \widetilde{\mathcal{M}}_1 = A\cdot\mathbb{S}^K$ with (67) one has

$$\begin{aligned}&\inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} \widetilde{c}\cdot\Big[M_{\mathbf{Q}^{**}} - A\cdot\exp\Big(-\frac{1}{A}\cdot I(\mathbf{Q},\mathbf{Q}^{**})\Big)\Big]\\ &= \inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) = -\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\in\mathbb{\Omega}\Big].\end{aligned} \tag{91}$$

(c) For any $\widetilde{c} > 0$, $\mathbf{Q} \in \widetilde{\mathcal{M}}_0 = A\cdot\mathbb{S}^K_{>0}$ and $\mathbf{Q}^{**} \in \mathbb{R}^K_{>0}$ we obtain

$$\inf_{m\neq 0} D^{SBD}_{\widetilde{c}\cdot\varphi_0,\mathbf{P}}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) = \widetilde{c}\cdot\Big[M_{\mathbf{P}}\cdot\log\Big(T_0(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P})\Big) + \breve{T}_0(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P}) - M_{\mathbf{P}}\cdot\log(M_{\mathbf{P}})\Big] =: \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_0,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) \tag{92}$$

and consequently (cf. (82)) for any set subset $A\cdot\mathbb{\Omega} \subset \widetilde{\mathcal{M}}_0 = A\cdot\mathbb{S}^K_{>0}$ with (67) one gets

$$\begin{aligned}&\inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} \widetilde{c}\cdot\Big[M_{\mathbf{P}}\cdot\log\Big(T_0(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P})\Big) + \breve{T}_0(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P}) - M_{\mathbf{P}}\cdot\log(M_{\mathbf{P}})\Big]\\ &= \inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_0,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) = -\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\in\mathbb{\Omega}\Big].\end{aligned} \tag{93}$$

Remark 50: In accordance with Remark 48(d), the quantity $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$ in Theorem 49 can be regarded as (new class of) divergences between $\mathbf{Q}$ and $\mathbf{Q}^{**}$. Accordingly, we call $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$ the *inner-minimization-scaled-Bregman-distance* (in short, innmin-SBD), which is BS-minimized — in a narrow sense — by (89) respectively (91) respectively (93). The BS-minimization — in a wide sense — of the "original" scaled Bregman divergence $\mathbf{Q} \mapsto D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$ will be treated in Remark 57 below.

The proof of Theorem 49 will be given in Appendix A.

Analogously to (78), the limit statements (89), (91) and (93), provide the principle for the approximation of the solution of the divergence minimization problem $\Phi_{\mathbf{P}}(\mathbb{\Omega}) := \inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$. Indeed, by replacing the right-hand side in those by their finite counterparts, we deduce for given large n

$$-\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\in\mathbb{\Omega}\Big] \approx \inf_{\mathbf{Q}\in A\cdot\mathbb{\Omega}} \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}); \tag{94}$$

it remains to estimate the left-hand side of (78) (see Section XIII below, where the latter also provides estimates of the *minimizers*).

Theorem 49 establishes the part (a) of our above-mentioned Remark 48. As far as the corresponding part (b) is concerned, it is clear that the involved functions $\mathbf{Q} \mapsto \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$ can generally *NOT* be rewritten in terms of $\mathbf{Q}-$independent invertible functions $\breve{F}(\cdot)$ of the divergences $D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$. This means that we can generally *NOT* employ our *narrow-sense* bare-simulation procedure for the minimization of $\mathbf{Q} \mapsto D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$. However, there are a few important exceptions given in the following

Theorem 51: Let $\mathbf{P} \in \mathbb{R}^K_{>0}$, $\widetilde{c} \in]0,\infty[$, $\gamma \in \mathbb{R}$ and $A \in]0,\infty[$ be arbitrary but fixed. Moreover, let $(W_i)_{i\in\mathbb{N}}$ be a family of independent and identically distributed $\mathbb{R}-$valued random variables with probability distribution $\mathbb{Q}[\cdot\,] := \mathbb{P}[W_1 \in \cdot\,]$ being connected with the divergence generator $\varphi := \widetilde{c}\cdot\varphi_\gamma(\cdot) \in \widetilde{\Upsilon}(]a,b[)$ via the representability (69).
(a) If $\gamma \in \mathbb{R}\backslash\{1\}$, then for all $C \in]0,\infty[$, $\widetilde{c} \in]0,\infty[$ and $\mathbf{Q} \in \widetilde{\mathcal{M}}_\gamma$ there holds in the special subsetup $\mathbf{Q}^{**} := C\cdot\mathbf{P}$ the representation

$$\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},C\cdot\mathbf{P}) = \breve{F}_{\gamma,\widetilde{c},A,M_\mathbf{P},C}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},C\cdot\mathbf{P})\Big) \tag{95}$$

with function $\breve{F}_{\gamma,\widetilde{c},A,M_\mathbf{P},C} :]-\infty,\infty[\mapsto]-\infty,\infty]$ given by

$$\breve{F}_{\gamma,\widetilde{c},A,M_\mathbf{P},C}(x) \;:=\; \begin{cases} \frac{\widetilde{c}\cdot C^\gamma}{\gamma}\cdot\big\{M_\mathbf{P} - A^{\gamma/(\gamma-1)}\cdot\big[C^\gamma\cdot M_\mathbf{P} + \gamma\cdot C^{\gamma-1}\cdot(A - C\cdot M_\mathbf{P}) + \frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x\big]^{-1/(\gamma-1)}\big\}, \\ \qquad \text{if } \gamma\in]-\infty,0[\cup]0,1[\text{ and } C^\gamma\cdot M_\mathbf{P} + \gamma\cdot C^{\gamma-1}\cdot(A - C\cdot M_\mathbf{P}) + \frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x \geq 0, \\ \qquad \text{or if } \gamma\in]1,2[\cup]2,\infty[\text{ and } C^\gamma\cdot M_\mathbf{P} + \gamma\cdot C^{\gamma-1}\cdot(A - C\cdot M_\mathbf{P}) + \frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x > 0, \\ \frac{\widetilde{c}\cdot C^2}{2}\cdot\big\{M_\mathbf{P} - A^2\cdot\big[C^2\cdot M_\mathbf{P} + 2\cdot C\cdot(A - C\cdot M_\mathbf{P}) + \frac{2}{\widetilde{c}}\cdot x\big]^{-1}\big\}, \\ \qquad\qquad \text{if } \gamma = 2 \text{ and } x\in]\frac{\widetilde{c}\cdot C}{2}\cdot(C\cdot M_\mathbf{P} - 2A),\infty[, \\ \widetilde{c}\cdot\Big(M_\mathbf{P} - \frac{A}{C} + M_\mathbf{P}\cdot\log\Big(\frac{A}{C}\Big) - M_\mathbf{P}\cdot\log(M_\mathbf{P})\Big) + x, \qquad \text{if } \gamma = 0 \text{ and } x\in]-\infty,\infty[, \\ \infty, \quad \text{if } \gamma\in]-\infty,0[\cup]0,1[\text{ and } C^\gamma\cdot M_\mathbf{P} + \gamma\cdot C^{\gamma-1}\cdot(A - C\cdot M_\mathbf{P}) + \frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x < 0, \\ \qquad \text{or if } \gamma\in]1,2[\cup]2,\infty[\text{ and } C^\gamma\cdot M_\mathbf{P} + \gamma\cdot C^{\gamma-1}\cdot(A - C\cdot M_\mathbf{P}) + \frac{\gamma\cdot(\gamma-1)}{\widetilde{c}}\cdot x \leq 0, \\ \qquad \text{or if } \gamma = 2 \text{ and } x\in]-\infty,\frac{\widetilde{c}\cdot C}{2}\cdot(C\cdot M_\mathbf{P} - 2A)], \end{cases} \tag{96}$$

which on its effective domain $dom(\breve{F}_{\gamma,\widetilde{c},A,M_\mathbf{P},C})$ is strictly increasing with strictly increasing inverse function

$$\breve{F}^{\leftarrow}_{\gamma,\widetilde{c},A,M_\mathbf{P},C}(z) \;:=\; \begin{cases} \frac{\widetilde{c}}{\gamma\cdot(\gamma-1)}\cdot\Big\{A^\gamma\cdot\big[M_\mathbf{P} - \frac{\gamma}{\widetilde{c}\cdot C^\gamma}\cdot z\big]^{-(\gamma-1)} - C^\gamma\cdot M_\mathbf{P} - \gamma\cdot C^{\gamma-1}\cdot(A - C\cdot M_\mathbf{P})\Big\}, \\ \qquad\qquad \text{if } \gamma\in]-\infty,0[\cup]0,1[\text{ and } \gamma\cdot z \leq \widetilde{c}\cdot C^\gamma\cdot M_\mathbf{P}, \\ \qquad\qquad \text{or if } \gamma\in]1,2[\cup]2,\infty[\text{ and } \gamma\cdot z < \widetilde{c}\cdot C^\gamma\cdot M_\mathbf{P}, \\ \frac{\widetilde{c}}{2}\cdot\Big\{A^2\cdot\big[M_\mathbf{P} - \frac{2}{\widetilde{c}\cdot C^2}\cdot z\big]^{-1} - C^2\cdot M_\mathbf{P} - 2\cdot C\cdot(A - C\cdot M_\mathbf{P})\Big\}, \\ \qquad\qquad \text{if } \gamma = 2 \text{ and } 2\cdot z < \widetilde{c}\cdot C^2\cdot M_\mathbf{P}, \\ z - \widetilde{c}\cdot\Big(M_\mathbf{P} - \frac{A}{C} + M_\mathbf{P}\cdot\log\Big(\frac{A}{C}\Big) - M_\mathbf{P}\cdot\log(M_\mathbf{P})\Big), \qquad \text{if } \gamma = 0 \text{ and } z\in]-\infty,\infty[. \end{cases} \tag{97}$$

Consequently (cf. (89), (93)) for any $\gamma \in \mathbb{R}\backslash[1,2[$ and any subset $A\cdot\mathfrak{Q} \subset \widetilde{\mathcal{M}}_\gamma$ with (67) one gets

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} \breve{F}_{\gamma,\widetilde{c},A,M_\mathbf{P},C}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},C\cdot\mathbf{P})\Big) = -\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}} \in \mathfrak{Q}\Big]$$

and

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},C\cdot\mathbf{P}) = \breve{F}^{\leftarrow}_{\gamma,\widetilde{c},A,M_\mathbf{P},C}\Big(-\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}} \in \mathfrak{Q}\Big]\Big). \tag{98}$$

In particular, the functions $\Phi_\mathbf{P}(\cdot) := \breve{F}_{\gamma,\widetilde{c},A,M_\mathbf{P},C}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot,C\cdot\mathbf{P})\Big)$ and $\Phi_\mathbf{P}(\cdot) := D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot,C\cdot\mathbf{P})$ are bare-simulation minimizable (BS-minimizable) in the narrow sense (cf. (2) in Definition 1 and the special case of Remark 2(a)) on all sets $A\cdot\mathfrak{Q} \subset \widetilde{\mathcal{M}}_\gamma$ satisfying (67) *in the relative topology*.
(b) If $\gamma = 1$, then for all $\widetilde{c} \in]0,\infty[$, $\mathbf{Q}^{**} \in \mathbb{R}^K_{>0}$ and $\mathbf{Q} \in \widetilde{\mathcal{M}}_1 = A\cdot\mathbb{S}^K$ there holds the representation

$$\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) = \breve{F}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})\Big) \tag{99}$$

with strictly increasing function $\breve{F}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}} :]-\infty,\infty[\mapsto]-\infty,\infty[$ given by

$$\breve{F}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}(x) \;:=\; \widetilde{c}\cdot\Big\{M_{\mathbf{Q}^{**}} - A\cdot\exp\Big(\frac{M_{\mathbf{Q}^{**}}}{A} - 1 - \frac{x}{A\cdot\widetilde{c}}\Big)\Big\}, \qquad x\in]-\infty,\infty[, \tag{100}$$

having strictly increasing inverse function

$$\breve{F}^{\leftarrow}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}(z) \;:=\; \widetilde{c}\cdot\Big\{M_{\mathbf{Q}^{**}} - A - A\cdot\Big[\log\Big(M_{\mathbf{Q}^{**}}-\frac{z}{\widetilde{c}}\Big)-\log A\Big]\Big\}, \qquad \text{for all } z\in\,]-\infty,\widetilde{c}\cdot M_{\mathbf{Q}^{**}}[. \tag{101}$$

Consequently (cf. (91)) for any set subset $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_1$ with (67) one gets

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\breve{F}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})\Big) = -\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\in\mathfrak{Q}\Big]$$

and

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) = \breve{F}^{\leftarrow}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}\Big(-\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\in\mathfrak{Q}\Big]\Big). \tag{102}$$

In particular, the functions $\Phi_{\mathbf{P}}(\cdot) := \breve{F}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\cdot,\mathbf{Q}^{**})\Big)$ and $\Phi_{\mathbf{P}}(\cdot) := D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\cdot,\mathbf{Q}^{**})$ are bare-simulation minimizable (BS-minimizable) in the narrow sense (cf. (2) in Definition 1 and the special case of Remark 2(a)) on all sets $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_1$ satisfying (67) *in the relative topology*.

The assertions of Theorem 51 follow by straightforward caluclations from Theorem 49. For the special case $C=1$, Theorem 51 collapses to Theorem 44.

Analogously to (78), the limit statements (98) and (102) provide the principle for the approximation of the solution of the minimization problems $\Phi_{\mathbf{P}}(\mathfrak{Q}) := \inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},C\cdot\mathbf{P})$ (for $\gamma\in\mathbb{R}\backslash[1,2[$ and $\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$ (for $\gamma=1$). Indeed, we replace the right-hand side in those by their finite counterparts, and accordingly obtain for given large n

$$\breve{F}^{\leftarrow}_{\gamma,\widetilde{c},A,M_{\mathbf{P}},C}\Big(-\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\in\mathfrak{Q}\Big]\Big) \;\approx\; \inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},C\cdot\mathbf{P}), \qquad \text{for } \gamma\in\mathbb{R}\backslash[1,2[, \tag{103}$$

$$\breve{F}^{\leftarrow}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}\Big(-\frac{1}{n}\log\mathbb{\Pi}\Big[\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\in\mathfrak{Q}\Big]\Big) \;\approx\; \inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}), \qquad \text{for } \gamma=1; \tag{104}$$

it remains to estimate the left-hand sides of (103) and (104) (see Section XIII below, where the latter also provides estimates of the *minimizers*).

VIII. Bare-Simulation-Method for General Deterministic Divergence-Optimization-Problems with Constant-Component-Sum Side Constraint

A. Minimization via Base-Divergence-Method 1

Recall that we are interested in the constrained optimization of the *continuous* functions $\mathbf{Q}\mapsto\Phi_{\mathbf{P}}(\mathbf{Q})$ in the above-mentioned cases (D1) to (D8) of Subsection V-A, and beyond. Notice that — on constraint sets of the form $A\cdot\mathfrak{Q}$ with (67) (to be treated in this section) — the class (D1) of CASM $\varphi-$divergences $\Phi_{\mathbf{P}}(\mathbf{Q}) := D_\varphi(\mathbf{Q},\mathbf{P})$ which *can not be covered* by the *narrow-sense* BS minimizability results of Section VI (namely seemingly all $\varphi-$divergences which are not generalized power divergences $D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{P})$, cf. (17)) is now *much larger* than the class (D1) of CASM $\varphi-$divergences which *can not be covered* by the narrow-sense BS minimizability results of Section III dealing with constraint sets of the form $\mathfrak{Q}$ with (6). One corresponding example is the prominent *Jensen-Shannon divergence* (being also called symmetrized and normalized Kullback-Leibler information distance, symmetrized and normalized relative entropy, capacitory discrimination), see Broniatowski & Stummer [1] for details.

However, for such cases — and beyond — we can apply the following new fundamental *non-narrow-sense* BS-minimizability:

Theorem 52: Let us arbitrarily fix some $\mathbb{P}\in\mathbb{S}^K_{>0}$, $\widetilde{c}\in\,]0,\infty[$, $A\in\,]0,\infty[$, φ_γ with $\gamma\in\mathbb{R}\backslash\,]1,2[$, $\mathbb{\Pi}$, $W:=(W_i)_{i\in\mathbb{N}}$ and $\boldsymbol{\xi}_n^{w\mathbf{W}}$ (cf. (70)) as in Theorem 44.
(a) Furthermore, suppose that $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_\gamma$ is compact and satisfies the regularity properties (67) in the relative topology, and that $\Phi: A\cdot\mathfrak{Q}\mapsto\mathbb{R}$ is a continuous function on $A\cdot\mathfrak{Q}$. Then, there holds

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\Phi(\mathbf{Q}) \;=\; -\lim_{n\to\infty}\frac{1}{n}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}},\mathbb{P})\Big)-\Phi\big(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\Big)\Big)\cdot\mathbb{1}_{\mathfrak{Q}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\;\Big]\Big) \tag{105}$$

and the infimum is attained at some (not necessarily unique) point in $A\cdot\mathfrak{Q}$. In particular, the function $\Phi(\cdot)$ is bare-simulation minimizable (BS-minimizable) on $A\cdot\mathfrak{Q}$ (cf. (2) in Definition 1).
(b) If $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_\gamma$ is not necessarily compact but satisfies the regularity properties (67) in the relative topology and the finiteness property (68) with $\varphi:=\widetilde{c}\cdot\varphi_\gamma$, and $\Phi: A\cdot\mathfrak{Q}\mapsto\mathbb{R}$ is a continuous function which satisfies the lower-bound condition

$$\text{there exists a constant } c_1\in\mathbb{R} \text{ such that for all } \mathbf{Q}\in A\cdot\mathfrak{Q} \text{ there holds} \quad \Phi(\mathbf{Q})\geq F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\Big)-c_1\,, \tag{106}$$

then the representation/convergence (105) — and hence the corresponding BS-minimizability — still holds, but the infimum may not necessarily be attained/reached at some point in $A \cdot \mathfrak{Q}$.

The proof of Theorem 52 will be given in Appendix A below. Clearly, Theorem 52(a) can be applied to obtain $\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} \Phi(\mathbf{Q})$ for all the directed distances/divergences $\Phi(\cdot) := \Phi_{\breve{\mathbf{P}}}(\cdot)$ (where $\breve{\mathbf{P}}$ needs not coincide with $\mathbb{P}$) and friends given in (D1) to (D8) of Subsection V-A, which are therefore all BS-minimizable on compact $A \cdot \mathfrak{Q} \subset \widetilde{\mathcal{M}}_\gamma$ with regularity (67) in the relative topology.

Analogously to (78), the limit statement (105) provides the principle for the approximation of the solution of the minimization problem $\Phi(\mathfrak{Q}) := \inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} \Phi(\mathbf{Q})$. This can be achieved by replacing the right-hand side in (105) by its finite counterpart, from which we obtain for given large n

$$-\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}},\mathbb{P})\Big)-\Phi\big(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\Big)\Big)\cdot 1\!\!1_{\mathfrak{Q}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\ \Big]\Big)\ \approx\ \inf_{Q\in\mathfrak{Q}}\Phi(\mathbf{Q}); \tag{107}$$

it remains to estimate the left-hand side of (107) (see Section XIII below, where the latter also provides estimates of the *minimizers*).

Example 53: As a continuation of Example 28 in connection with Remark 29(b), let us how Theorem 52 can be used to tackle the BS-minimizability — on $A \cdot \mathfrak{Q}$ — of the generalized power divergences $\Phi(\cdot) := \Phi_{\mathbb{P}}(\cdot) := D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbb{P})$ (cf. (17)) for the missing case $\gamma \in]1,2[$ (and $\widetilde{c}\in]0,\infty[$). Indeed, by employing (74) it is easy to see that

$$\text{for all } x\in[0,\infty[,\ A\in]0,\infty[,\ \widetilde{c}\in]0,\infty[,\ \gamma\in]1,2[\text{ there holds } x\ \geq\ F_{1,\frac{\widetilde{c}}{\gamma},A}(x)=\frac{\widetilde{c}}{\gamma}\cdot\Big\{1-A\cdot\exp\Big(\frac{1}{A}-1-\frac{x}{A\cdot\frac{\widetilde{c}}{\gamma}}\Big)\Big\} \tag{108}$$

$$\text{with equality if and only if } x=\frac{\widetilde{c}}{\gamma}\cdot(1-A)\geq 0$$

(and even for a wider range of γ). By using (59) and (108) we obtain

$$\text{for all } \mathbf{Q}\in A\cdot\mathfrak{Q} \text{ and all } \gamma\in]1,2[\text{ there holds}\quad \Phi_{\mathbb{P}}(\mathbf{Q}):=D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\ \geq\ D_{\frac{\widetilde{c}}{\gamma}\cdot\varphi_1}(\mathbf{Q},\mathbb{P})\ \geq\ F_{1,\frac{\widetilde{c}}{\gamma},A}\Big(D_{\frac{\widetilde{c}}{\gamma}\cdot\varphi_1}(\mathbf{Q},\mathbb{P})\Big); \tag{109}$$

notice that the inequalities in (109) turn into equalities if $q_k = p_k$ for all $k\in\{1,\dots,K\}$ (and hence, $A=1$) but also e.g. if $q_K = 0$ and $q_k = p_k$ for all $k\in\{1,\dots,K-1\}$ (and hence, $A<1$). According to (109), the bound (106) is satisfied with $c_1 = 0$. Thus, we can proceed analogously to Remark 29(b), and choose the W_i's to be i.i.d. copies of the random variable W of the form $W = \frac{\gamma}{\widetilde{c}}\cdot Z$ for a Poisson $POI(\frac{\widetilde{c}}{\gamma})-$distributed random variable Z (cf. Broniatowski & Stummer [1], see also Table 1). With these choices, (105) specializes to

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\ =\ -\lim_{n\to\infty}\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(F_{1,\frac{\widetilde{c}}{\gamma},A}\Big(D_{\frac{\widetilde{c}}{\gamma}\cdot\varphi_1}(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}},\mathbb{P})\Big)-D_{\widetilde{c}\cdot\varphi_\gamma}(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}},\mathbb{P})\Big)\Big)\cdot 1\!\!1_{\mathfrak{Q}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\ \Big]\Big)$$

for all $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_\gamma$ satisfying the regularity properties (67) in the relative topology and the finiteness property (68) with $\varphi := \frac{\widetilde{c}}{\gamma}\cdot\varphi_1$.

B. Maximization via Base-Divergence-Method 1

For the *non-narrow-sense BS-maximizability* we obtain the following new fundamental

Theorem 54: Let us arbitrarily fix some $\mathbb{P}\in\mathbb{S}_{>0}^K$, $\widetilde{c}\in]0,\infty[$, $A\in]0,\infty[$, φ_γ with $\gamma\in\mathbb{R}\setminus]1,2[$, (, $W := (W_i)_{i\in\mathbb{N}}$ and $\boldsymbol{\xi}_n^{w\mathbf{W}}$ (cf. (70)) as in Theorem 44.
(a) Furthermore, suppose that $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_\gamma$ is compact and satisfies the regularity properties (67) in the relative topology, and that $\Phi : A\cdot\mathfrak{Q}\mapsto\mathbb{R}$ is a continuous function on $A\cdot\mathfrak{Q}$. Then, there holds

$$\sup_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\Phi(\mathbf{Q})\ =\ \lim_{n\to\infty}\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}},\mathbb{P})\Big)+\Phi\big(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\Big)\Big)\cdot 1\!\!1_{\mathfrak{Q}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\ \Big]\Big) \tag{110}$$

and the supremum is attained at some (not necessarily unique) point in $A\cdot\mathfrak{Q}$. In particular, the function $\Phi(\cdot)$ is bare-simulation maximizable (BS-maximizable) on $A\cdot\mathfrak{Q}$ (cf. (3) in Definition 1).
(b) If $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_\gamma$ is not necessarily compact but satisfies the regularity properties (67) in the relative topology and $\Phi : A\cdot\mathfrak{Q}\mapsto\mathbb{R}$ is a continuous function which satisfies the upper-bound condition

$$\text{there exists a constant } c_1\in\mathbb{R} \text{ such that for all } \mathbf{Q}\in A\cdot\mathfrak{Q} \text{ there holds}\quad \Phi(\mathbf{Q})\leq c_1 - F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\Big),$$

then the representation/convergence (110) — and hence the corresponding BS-maximizability — still holds, but the supremum may not necessarily be attained/reached at some point in $A\cdot\mathfrak{Q}$.

The proof of Theorem 54 will be given in Appendix A below.

Analogously to (107), the limit statement (110) provides the principle for the approximation of the solution of the maximization problem $\Phi(\mathfrak{Q}) := \sup_{\mathbf{Q}\in A\cdot\mathfrak{Q}} \Phi(\mathbf{Q})$. This can be achieved by replacing the right-hand side in (110) by its finite counterpart, from which we obtain for given large n

$$\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}},\mathbb{P})\Big)+\Phi\big(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\Big)\Big)\cdot 1_{\mathfrak{Q}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\ \Big]\Big)\ \approx\ \sup_{Q\in\mathfrak{Q}}\Phi(\mathbf{Q}); \tag{111}$$

it remains to estimate the left-hand side of (111) (see Section XIII below, where the latter also provides estimates of the *maximizers*).

C. Minimization via Base-Divergence-Method 2

Due to our investigations in Section VII, as an alternative to the new Theorem 52 we can also derive the following new assertions, by switching the involved base-divergences:

Theorem 55: Let us arbitrarily fix some $\mathbf{P}\in\mathbb{R}^K_{>0}$ (recall $M_{\mathbf{P}} := \sum_{i=1}^K p_i > 0$), $\mathbf{Q}^{**}$, $\widetilde{c}\in]0,\infty[$, $A\in]0,\infty[$, φ_γ with $\gamma\in\mathbb{R}\setminus]1,2[$, $\widetilde{V}$ and $\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}$ (cf. (80) with $\widetilde{V}$ instead of V) as in Theorem 49; we call the corresponding $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot,\mathbf{Q}^{**})$ (cf. (88),(90),(92)) the *base-innmin-SBD-divergence* (function).
(a) Furthermore, suppose that $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_\gamma$ is compact and satisfies the regularity properties (67) in the relative topology, and that $\Phi: A\cdot\mathfrak{Q}\mapsto\mathbb{R}$ is a continuous function on $A\cdot\mathfrak{Q}$. Then, there holds

$$\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\Phi(\mathbf{Q})\ =\ -\lim_{n\to\infty}\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}},\mathbf{Q}^{**}\Big)-\Phi\big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot 1_{\mathfrak{Q}}\big(\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\big)\ \Big]\Big) \tag{112}$$

and the infimum is attained at some (not necessarily unique) point in $A\cdot\mathfrak{Q}$. In particular, the function $\Phi(\cdot)$ is bare-simulation minimizable (BS-minimizable) on $A\cdot\mathfrak{Q}$ (cf. (2) in Definition 1), *in terms of the SBD method*.
(b) If $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_\gamma$ is not necessarily compact but satisfies the regularity properties (67) in the relative topology with $\varphi := \widetilde{c}\cdot\varphi_\gamma$, and $\Phi: A\cdot\mathfrak{Q}\mapsto\mathbb{R}$ is a continuous function which satisfies the lower-bound condition

$$\text{there exists a constant } c_1\in\mathbb{R} \text{ such that for all } \mathbf{Q}\in A\cdot\mathfrak{Q} \text{ there holds}\quad \Phi(\mathbf{Q})\geq\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})-c_1\,,$$

then the representation/convergence (112) — and hence the corresponding BS-minimizability — still holds, but the infimum may not necessarily be attained/reached at some point in $A\cdot\mathfrak{Q}$.

Remark 56: (a) For the case $\gamma\in\mathbb{R}\setminus\{1\}$ and the special subsetup $\mathbf{Q}^{**} := C\cdot\mathbf{P}$ (cf. Theorem 51(a)), in Theorem 55 one can equivalently replace $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot,C\cdot\mathbf{P})$ by $\breve{F}_{\gamma,\widetilde{c},A,M_{\mathbf{P}},C}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot,C\cdot\mathbf{P})\Big)$ (cf. (95),(96)). By additionally taking $C=1$ and $\mathbf{P}:=\mathbb{P}$ to be a probability vector, one collapsedly ends up with the corresponding assertions of Theorem 52.
(b) For the case $\gamma=1$, in Theorem 55 one can equivalently replace $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\cdot,\mathbf{Q}^{**})$ by $\breve{F}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\cdot,\mathbf{Q}^{**})\Big)$ (cf. (99),(100)).

Remark 57: The BS-minimization — in a wide sense — $\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\Phi(\mathbf{Q})$ of the scaled Bregman divergence $\Phi(\mathbf{Q}) := D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$ follows immediately as special case of Theorem 55. Notice that in (112) the corresponding exponential part $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}},\mathbf{Q}^{**}\Big)-\Phi\big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\big)$ is generally non-zero.

The proof of Theorem 55 will be given in Appendix A below. Clearly, Theorem 55(a) can be applied to obtain $\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\Phi(\mathbf{Q})$ for all the directed distances/divergences $\Phi(\cdot) := \Phi_{\mathbf{P}}(\cdot)$ and friends given in (D1) to (D8) of Subsection V-A, which are therefore all BS-minimizable on compact $A\cdot\mathfrak{Q}\subset\widetilde{\mathcal{M}}_\gamma$ with regularity (67) in the relative topology.

Analogously to (107), the limit statement (112) provides the principle for the approximation of the solution of the minimization problem $\Phi(\mathfrak{Q}) := \inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\Phi(\mathbf{Q})$. This can be achieved by replacing the right-hand side in (112) by its finite counterpart, from which we obtain for given large n

$$-\frac{1}{n}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}},\mathbf{Q}^{**}\Big)-\Phi\big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot 1_{\mathfrak{Q}}\big(\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\big)\ \Big]\Big)\ \approx\ \inf_{Q\in\mathfrak{Q}}\Phi(\mathbf{Q}); \tag{113}$$

it remains to estimate the left-hand side of (113) (see Section XIII below, where the latter also provides estimates of the *minimizers*).

D. Maximization via Base-Divergence-Method 2

For the *non-narrow-sense BS-maximizability* via innmin-SBDs as base divergences, we obtain the following new fundamental

Theorem 58: Let us arbitrarily fix some $\mathbf{P} \in \mathbb{R}^K_{>0}$ (recall $M_{\mathbf{P}} := \sum_{i=1}^K p_i > 0$), $\mathbf{Q}^{**}$, $\widetilde{c} \in]0,\infty[$, $A \in]0,\infty[$, φ_γ with $\gamma \in \mathbb{R}\backslash\,]1,2[$, $\widetilde{V}$ and $\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}$ (cf. (80) with $\widetilde{V}$ instead of V) as in Theorem 49; recall that we have named the corresponding $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot,\mathbf{Q}^{**})$ (cf. (88),(90),(92)) the *base-innmin-SBD-divergence* (function).
(a) Furthermore, suppose that $A\cdot\mathfrak{\mathbf{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ is compact and satisfies the regularity properties (67) in the relative topology, and that $\Phi : A\cdot\mathfrak{\mathbf{Q}} \mapsto \mathbb{R}$ is a continuous function on $A\cdot\mathfrak{\mathbf{Q}}$. Then, there holds

$$\sup_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}} \Phi(\mathbf{Q}) \;=\; \lim_{n\to\infty} \frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot \Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}, \mathbf{Q}^{**}\Big) + \Phi\big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{Q}}}\big(\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\big)\Big]\Big) \tag{114}$$

and the supremum is attained at some (not necessarily unique) point in $A\cdot\mathfrak{\mathbf{Q}}$. In particular, the function $\Phi(\cdot)$ is bare-simulation maximizable (BS-maximizable) on $A\cdot\mathfrak{\mathbf{Q}}$ (cf. (3) in Definition 1), *in terms of the SBD method*.
(b) If $A\cdot\mathfrak{\mathbf{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ is not necessarily compact but satisfies the regularity properties (67) in the relative topology, and $\Phi : A\cdot\mathfrak{\mathbf{Q}} \mapsto \mathbb{R}$ is a continuous function which satisfies the lower-bound condition

$$\text{there exists a constant } c_1 \in \mathbb{R} \text{ such that for all } \mathbf{Q} \in A\cdot\mathfrak{\mathbf{Q}} \text{ there holds} \quad \Phi(\mathbf{Q}) \le c_1 - \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})\ ,$$

then the representation/convergence (114) — and hence the corresponding BS-minimizability — still holds, but the supremum may not necessarily be attained/reached at some point in $A\cdot\mathfrak{\mathbf{Q}}$.

Remark 59: (a) For the case $\gamma \in \mathbb{R}\backslash\{1\}$ and the special subsetup $\mathbf{Q}^{**} := C\cdot\mathbf{P}$ (cf. Theorem 51(a)), in Theorem 58 one can equivalently replace $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot, C\cdot\mathbf{P})$ by $\breve{F}_{\gamma,\widetilde{c},A,M_{\mathbf{P}},C}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot,C\cdot\mathbf{P})\Big)$ (cf. (95),(96)). By additionally taking $C=1$ and $\mathbf{P} := \mathbb{P}$ to be a probability vector, one collapsedly ends up with the corresponding assertions of Theorem 54.
(b) For the case $\gamma = 1$, in Theorem 58 one can equivalently replace $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\cdot,\mathbf{Q}^{**})$ by $\breve{F}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(\cdot,\mathbf{Q}^{**})\Big)$ (cf. (99),(100)).

The proof of Theorem 58 will be given in Appendix A below. Certainly, Theorem 58(a) can be applied to obtain $\sup_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q})$ for all the directed distances/divergences $\Phi(\cdot) := \Phi_{\breve{\mathbf{P}}}(\cdot)$ (where $\breve{\mathbf{P}}$ needs not coincide with $\mathbf{P}$) and friends given in (D1) to (D8) of Subsection V-A, which are therefore all BS-maximizable on compact $A\cdot\mathfrak{\mathbf{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ with regularity properties (67) in the relative topology.

Analogously to (111), the limit statement (114) provides the principle for the approximation of the solution of the maximization problem $\Phi(\mathfrak{\mathbf{Q}}) := \sup_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q})$. This can be achieved by replacing the right-hand side in (114) by its finite counterpart, from which we obtain for given large n

$$\frac{1}{n}\log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot \Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}, \mathbf{Q}^{**}\Big) + \Phi\big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{Q}}}\big(\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\big)\Big]\Big) \;\approx\; \sup_{Q\in\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q}); \tag{115}$$

it remains to estimate the left-hand side of (115) (see Section XIII below, where the latter also provides estimates of the *maximizers*).

IX. Narrow-Sense Bare-Simulation-Minimization of $\varphi-$divergences under risk

A. The statistical view

In contrast to the previous Sections VI,VII,VIII, we now work out our BS method for the important setup where basically $\mathbb{P}$ is a *random* (unknown) element of the (open) simplex $\mathbb{S}^K_{>0}$ of zeros-free $K-$component probability (frequency) vectors and $\mathfrak{\mathbf{Q}} \subset \mathbb{S}^K$ (i.e. $A=1$). Its importance stems from the fact that in the statistics of discrete data — and in the adjacent research fields of information theory, artificial intelligence and machine learning — one often encounters the following *minimum distance estimation (MDE) problem* which is often also named as *estimation of the asymptotic risk*:

(MDE1) for index $i \in \mathbb{N}$, let the generation of the $i-$th (uncertainty-prone) data point be represented by the random variable X_i which takes values in the discrete set $\mathcal{Y} := \{d_1,\cdots,d_K\}$ of K distinct values "of any kind". It is assumed that there exists a — true unknown — probability distribution $\mathbb{P}[\cdot]$ on $\mathcal{Y}$ which is the a.s. limit (as n tends to infinity) of the empirical distributions $\mathbb{P}_n^{emp}(\mathbf{X}_1^n)$ defined by the collection $\mathbf{X}_1^n := (X_1,\ldots,X_n)$, i.e.

$$\lim_{n\to\infty}\mathbb{P}_n^{emp}(\mathbf{X}_1^n) := \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \delta_{X_i} = \mathbb{P} \qquad \text{a.s.} \tag{116}$$

where δ_y denotes the one-point distribution (Dirac mass) at point y [17]. We assume that none of the entries of $\mathbb{P}$ bears zero mass so that $\mathbb{P}$ is identified with a point in the interior of $\mathbb{S}^K$ (see below). The underlying probability space (say, $(\mathfrak{X}, \mathcal{A}, \mathbb{\Pi})$) where the above a.s. convergence holds, pertains to the random generation of the sequence $(X_i)_{i\in\mathbb{N}}$, of which we do not need to know but for (116). Examples include the i.i.d. case (where the X_i's are independent and have common distribution $\mathbb{P}$), ergodic Markov chains on $\mathcal{Y}$ with stationary distribution $\mathbb{P}$, more globally autoregressive chains with stationary distribution $\mathbb{P}$, etc.

Remark 60: (a) Let us briefly discuss our assumption (116) (resp. its vector form (119) below) on the limit behavior of the empirical distribution of the sample $\mathbf{X}_1^n = (X_1, \ldots, X_n)$ as n tends to infinity. In the "basic" statistical context, the sample $\mathbf{X}_1^n$ consists of i.i.d. replications of a generic random variable X with probability distribution $\mathbb{IP}$. However, our approach captures many other sampling schemes, where the distribution $\mathbb{IP}$ is defined implicitly through (116) for which we aim at some estimate of $\Phi_{\mathbb{IP}}(\mathbb{\Omega})$ of a family $\mathbb{\Omega}$ of probability distributions on $\mathcal{Y}$. Sometimes the sequence of samples may stem from a "triangular" array $(X_{1,n}, .., X_{k_n,n})_{n\in\mathbb{N}}$ with $k_n \to \infty$ (as n tends to infinity) and (116) is substituted by

$$\lim_{n\to\infty} \frac{1}{k_n} \sum_{j=1}^{k_n} \delta_{X_{j,n}} = \mathbb{IP} \quad \text{a.s.}$$

which does not alter the results of this paper by any means.
(b) As another alternative to (116) (resp. its vector form (119) below), in the following we could also (verbatim) employ any other sequence $\mathbb{P}_n(\mathbf{X}_1^n)$ of probability-distribution-valued function(al)s for which $lim_{n\to\infty}\mathbb{P}_n(\mathbf{X}_1^n) = \mathbb{P}$ a.s.

(MDE2) given a *model* $\mathbb{\Omega}$, i.e. a family $\mathbb{\Omega}$ of probability distributions $\mathbb{Q}$ on $\mathcal{Y}$ each of which serves as a potential description of the underlying (unknown) data-generating mechanism $\mathbb{P}$, one would like to find the *minimum directed distance/divergence* (cf. (iii) in Section II)

$$\Phi_{\mathbb{P}}(\mathbb{\Omega}) := \inf_{\mathbb{Q}\in\mathbb{\Omega}} D(\mathbb{Q}, \mathbb{P}) \tag{117}$$

which quantifies the *adequacy* [18] of the model $\mathbb{\Omega}$ for modelling the true unknown data-generating mechanism $\mathbb{P}$, *via* the minimal distance/dissimilarity of $\mathbb{\Omega}$ to $\mathbb{P}$; a lower $\Phi_{\mathbb{P}}-$value means a better adequacy (in the sense of a lower departure between the model and the truth, cf. Lindsay [366], Lindsay et al. [367], Markatou & Chen [221], Markatou & Chen [368]). Hence, especially in the context of *model selection* within complicated big-data contexts, for the *search of appropriate models* $\mathbb{\Omega}$ and model elements/members therein, the (fast and efficient) computation of $\Phi_{\mathbb{P}}(\mathbb{\Omega})$ constitutes a decisive first step, since if the latter is "too large" (respectively "much larger than" $\Phi_{\mathbb{P}}(\overline{\mathbb{\Omega}})$ for some competing model $\overline{\mathbb{\Omega}}$), then the model $\mathbb{\Omega}$ is "not adequate enough" (respectively "much less adequate than" $\overline{\mathbb{\Omega}}$); in such a situation, the effort of computing the (not necessarily unique) best model element/member $\arg\inf_{\mathbb{Q}\in\mathbb{\Omega}} D(\mathbb{Q}, \mathbb{P})$ within the model $\mathbb{\Omega}$ is "not very useful" and is thus a "waste of computational time". Because of such considerations, we concentrate *first* on finding the infimum (117); finding the corresponding *minimizer(s)* $\operatorname{arginf}_{\mathbb{Q}\in\mathbb{\Omega}} D(\mathbb{Q}, \mathbb{P})$ — which can be interpreted as the (not necessarily unique) model member $\mathbb{Q} \in \mathbb{\Omega}$ which *most adequately* describes the true unknown data-generating mechanism $\mathbb{P}$ — will be treated *later* in Section XIV.

Since $int(\mathbb{\Omega})$ is required to be a non-empty set (in the relative topology) in the space of probability distributions on $\mathcal{Y}$, the present procedure is fitted for semi-parametric models $\mathbb{\Omega}$, e.g. defined through moment conditions (as extensions of the Empirical Likelihood paradigm, see e.g. Broniatowski & Keziou [369]), or through L-moment conditions (i.e. moment conditions pertaining to quantile measures, see Broniatowski & Decurninge [370]), or even more involved non-parametric models where the geometry of $\mathbb{\Omega}$ does not allow for ad-hoc procedures. In such setups, there is typically no closed form of the divergence with respect to any probability distribution available.

The measurement or the estimation of $\Phi_{\mathbb{P}}(\mathbb{\Omega})$ is a tool for the choice of pertinent putative models $\mathbb{\Omega}$ among a class of specifications. The case when $\Phi_{\mathbb{P}}(\mathbb{\Omega}) > 0$ is interesting in its own, since it is quite common in engineering modelling to argue in favor of misspecified models (or (non-void) neighborhoods of such models for sake of robustness issues), due to quest for conservatism; the choice between them is a widely open field e.g. in the practice of reliability.

An estimate of $\Phi_{\mathbb{P}}(\mathbb{\Omega})$ can be used as a statistics for some test of fit, and indeed the likelihood ratio test adapted to some semi-parametric models has been generalized to the divergence setting (see Broniatowski & Keziou [369]). The statement of the limit distributions of our estimate, under the model and under misspecification, is postponed to future work. In the following, we compute/approximate (117) — and beyond — by our *bare simulation (BS)* method, by appropriately adapting our methods of the previous Sections VI,VII,VIII. To achieve this, let us first remark that, as usual, each probability distribution (probability measure) $\mathbb{P}$ on $\mathcal{Y} = \{d_1, \ldots, d_K\}$ can be uniquely identified with the (row) vector $\mathbb{IP} := (p_1, \ldots, p_K) \in \mathbb{S}^K$ of the corresponding

[17] notice that $\mathbb{P}_n^{emp}(\mathbf{X}_1^n)$ a probability distribution (on the data space $\mathcal{Y}$), which is random due to its dependence on the X_i's
[18] also called *misspecification error*

probability masses (frequencies) $p_k = \mathbb{P}[\{d_k\}]$ via $\mathbb{P}[B] = \sum_{k=1}^{K} p_k \cdot 1_B(d_k)$ for each $B \subset \mathcal{Y}$, where $1_B(\cdot)$ denotes the indicator function on the set B. In particular, the probability distribution $\mathbb{P}$ in (MDE1) can be identified with $(p_1, \ldots, p_K)$ in terms of $p_k = \mathbb{P}[\{d_k\}]$ (which in the i.i.d. case turns into $p_k = \mathbb{\Pi}[X_1 = d_k]$). Along this line, the family $\mathbb{\Omega}$ of probability distributions in (MDE2) can be identified with a subset $\boldsymbol{\mathfrak{Q}} \subset \mathbb{S}^K$ of probability vectors (viz. of vectors of probability masses). Analogously, each finite nonnegative measure Q on $\mathcal{Y}$ can be uniquely identified with a vector $\mathbf{Q} := (q_1, \ldots, q_K) \in \mathbb{R}^K_{\geq 0}$, and each finite signed measure Q with a vector $\mathbf{Q} := (q_1, \ldots, q_K) \in \mathbb{R}^K$. The corresponding divergences between distributions/measures are then, as usual, defined through the divergences between their respective masses/frequencies:

$$D(Q, \mathbb{P}) := D(\mathbf{Q}, \mathbb{P}).$$

In particular, $\mathbb{P}_n^{emp}(\mathbf{X}_1^n)$ can be identified with the vector $\mathbb{P}_n^{emp}(\mathbf{X}_1^n) := (p_{n,1}^{emp}(\mathbf{X}_1^n), \ldots, p_{n,K}^{emp}(\mathbf{X}_1^n))$ where

$$p_{n,k}^{emp}(\mathbf{X}_1^n) \;:=\; \frac{1}{n} \cdot n_k(\mathbf{X}_1^n) \;:=\; \frac{1}{n} \cdot card(\{i \in \{1, \ldots, n\} :\; X_i = d_k\}) \;=:\; \frac{1}{n} \cdot card(I_k^{(n)}(\mathbf{X}_1^n)), \quad k \in \{1, \ldots, K\}, \tag{118}$$

and accordingly the required limit behaviour (116) is equivalent to the vector-convergence

$$\lim_{n \to \infty} \Big(\frac{n_1(\mathbf{X}_1^n)}{n}, \ldots, \frac{n_K(\mathbf{X}_1^n)}{n} \Big) = (p_1, \ldots, p_K) \qquad \text{a.s.} \tag{119}$$

As usual, $\mathbb{P}_n^{emp}(\mathbf{X}_1^n)$ can be interpreted as the *vectorized histogram of the sample* $(X_1, \ldots, X_n)$.
Notice that, in contrast to the above Section VIII, the sets $I_k^{(n)}(\mathbf{X}_1^n)$ of indexes introduced in (118) and their numbers $n_k(\mathbf{X}_1^n) = card(I_k^{(n)})$ of elements are now *random* (due to their dependence on the X_i's). Of course, $M_{\mathbb{P}_n^{emp}(\mathbf{X}_1^n)} = 1$. In a *batch procedure*, when inference is done once/after the sample $\mathbf{X}_1^n = (X_1, \ldots, X_n)$ is observed, we may reorder this sample by putting the n_1 sample points X_i which are equal to d_1 in the first places, and so on; accordingly one ends up with index sets $I_k^{(n)}(\mathbf{X}_1^n)$ which are completely analogue to the ones used in the Sections VI,VII,VIII. When the *online acquisition* of the data X_i's is required, then we usually do not reorder the sample, and the $I_k^{(n)}(\mathbf{X}_1^n)$'s do generally not consist in consecutive indexes, which does not make any change with respect to the resulting construction nor to the estimator. Let us first concentrate on the well-known CASM $\varphi-$divergences $D(\cdot, \cdot) := D_\varphi(\cdot, \cdot)$ — in the probability-vector form context — and solve the following

Problem 61: For pregiven $\varphi \in \widetilde{\Upsilon}(]a, b[)$, unknown zeros-free probability vector $\mathbb{P} := (p_1, .., p_K) \in \mathbb{S}^K_{>0}$ satisfying (119), and subset $\boldsymbol{\mathfrak{Q}} \subset \mathbb{S}^K$ with regularity properties — in the relative topology (!!) —

$$cl(\boldsymbol{\mathfrak{Q}}) = cl\,(int\,(\boldsymbol{\mathfrak{Q}}))\,, \qquad int\,(\boldsymbol{\mathfrak{Q}}) \neq \emptyset \qquad \text{(cf. (67))},$$

find

$$\Phi_{\mathbb{P}}(\boldsymbol{\mathfrak{Q}}) := \inf_{\mathbb{Q} \in \boldsymbol{\mathfrak{Q}}} D_\varphi(\mathbb{Q}, \mathbb{P}),$$

provided that

$$\inf_{\mathbb{Q} \in \boldsymbol{\mathfrak{Q}}} D_\varphi(\mathbb{Q}, \mathbb{P}) < \infty$$

and that the divergence generator φ additionally satisfies the Condition 37 (i.e. the representability (69) holds with probability distribution $\mathbb{\Pi}$).

To tackle Problem 61, we proceed analogously to Section VI and employ a family of random variables $(W_i)_{i \in \mathbb{N}}$ of independent and identically distributed $\mathbb{R}-$valued random variables with probability distribution $\mathbb{\Pi}[\cdot] := \mathbb{\Pi}[W_1 \in \cdot\,]$ — being connected with the divergence generator $\varphi \in \widetilde{\Upsilon}(]a, b[)$ via the representability (69) — such that $(W_i)_{i \in \mathbb{N}}$ is independent of $(X_i)_{i \in \mathbb{N}}$ [19]. Moreover, we use

$$\boldsymbol{\xi}_{n,\mathbf{X}}^{w\mathbf{W}} \;:=\; \begin{cases} \left(\frac{\sum_{i \in I_1^{(n)}(\mathbf{X}_1^n)} W_i}{\sum_{k=1}^K \sum_{i \in I_k^{(n)}(\mathbf{X}_1^n)} W_i}, \ldots, \frac{\sum_{i \in I_K^{(n)}(\mathbf{X}_1^n)} W_i}{\sum_{k=1}^K \sum_{i \in I_k^{(n)}(\mathbf{X}_1^n)} W_i} \right), & \text{if } \sum_{j=1}^n W_j \neq 0, \\ (\infty, \ldots, \infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^n W_j = 0. \end{cases} \tag{120}$$

Notice that the right-hand side of (120) structurally coincides with the right-hand side of (70), however — as explained above — the construction of the involved partition $I_k^{(n)}(\mathbf{X}_1^n)$ differs from the non-random $I_k^{(n)}$ which we also indicate with (respective) different indexing on the corresponding left-hand sides.

With the above-mentioned ingredients, we have proven (cf. Theorem 12 of [1]) a more general, conditional-expectations-involving version of Theorem 40; from this, we have proceeded analogously to the derivation of the Theorem 44. Indeed, as mentioned above, the required representability (69) is satisfied for all (multiple of) the generators $\widetilde{c} \cdot \varphi_\gamma(\cdot)$ of (16) with $\widetilde{c} \in]0, \infty[$ and $\gamma \in \mathbb{R} \setminus]1, 2[$. In terms of the auxiliary notations in (S1) to (S5) with $A := 1$ and the *conditional* distributions

[19] notice that all the W_i and X_i live on the *same* underlying probability space $(\mathfrak{X}, \mathcal{A}, \mathbb{\Pi})$

$\mathbb{\Pi}_n[\,\cdot\,] := \mathbb{\Pi}_{\mathbf{X}_1^n}[\,\cdot\,] := \mathbb{\Pi}[\,\cdot\,|\,X_1,\ldots,X_n]$ [20], we have obtained in in Broniatowski & Stummer [1] the following assertion on BS-minimizability in the narrow sense:

Theorem 62: Suppose that $(X_i)_{i\in\mathbb{N}}$ is a sequence of random variables with values in $\mathcal{Y} := \{d_1,\cdots,d_K\}$ such that (119) holds for some probability vector $\mathbb{P} \in \mathbb{S}_{>0}^K$. Furthermore, we arbitrarily fix $\widetilde{c} \in]0,\infty[$ and $\gamma \in \mathbb{R}\setminus]1,2[$. Moreover, let $(W_i)_{i\in\mathbb{N}}$ be a family of independent and identically distributed $\mathbb{R}-$valued random variables with probability distribution $\mathfrak{l}[\,\cdot\,] := \mathbb{\Pi}[W_1 \in \cdot\,]$ being connected with the divergence generator $\widetilde{c}\cdot\varphi_\gamma \in \Upsilon(]a,b[)$ via the representability (69), such that $(W_i)_{i\in\mathbb{N}}$ is independent of $(X_i)_{i\in\mathbb{N}}$.
(a) Then there holds

$$\inf_{\mathbb{Q}\in\mathfrak{\mathbb{Q}}} F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P})\Big) = -\lim_{n\to\infty}\frac{1}{n}\log\,\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}_{n,\mathbf{X}}^{w\mathbf{W}} \in \mathfrak{\mathbb{Q}}\Big]$$

for all sets $\mathfrak{\mathbb{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ satisfying the regularity properties (67) *in the relative topology*. In particular, for each such $\mathbb{P} \in \mathbb{S}_{>0}^K$ with (119) the function $\Phi_{\mathbb{P}}(\cdot) := F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbb{P})\Big)$ is bare-simulation minimizable (BS-minimizable) in the narrow sense (cf. (2) in Definition 1 and the special case of Remark 2(a)) on all sets $\mathfrak{\mathbb{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ satisfying (67) *in the relative topology*.
(b) Moreover, there holds

$$\inf_{\mathbb{Q}\in\mathfrak{\mathbb{Q}}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P}) = F^{\leftarrow}_{\gamma,\widetilde{c},1}\Big(-\lim_{n\to\infty}\frac{1}{n}\log\,\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}_{n,\mathbf{X}}^{w\mathbf{W}} \in \mathfrak{\mathbb{Q}}\Big]\Big) = \lim_{n\to\infty} F^{\leftarrow}_{\gamma,\widetilde{c},1}\Big(-\frac{1}{n}\log\,\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}_{n,\mathbf{X}}^{w\mathbf{W}} \in \mathfrak{\mathbb{Q}}\Big]\Big) \tag{121}$$

for all sets $\mathfrak{\mathbb{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ satisfying the regularity properties (67) *in the relative topology*. In particular, for each such $\mathbb{P} \in \mathbb{S}_{>0}^K$ with (119) the function $\Phi_{\mathbb{P}}(\cdot) := D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbb{P})$ is bare-simulation minimizable (BS-minimizable) in the narrow sense on all sets $\mathfrak{\mathbb{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ satisfying (67) *in the relative topology*.

Remark 63: For an equivalent version of Theorem 62 in terms of probability *distributions* $\mathbb{Q} \in \mathfrak{Q}$ rather than probability *vectors* $\mathbb{Q} \in \mathfrak{\mathbb{Q}}$, the reader is referred to Broniatowski & Stummer [1] (cf. Theorem 12, Formula (39) and Lemma 14 therein).

The limit statement (121) provides the principle for the *estimation* — based on the random sample $\mathbf{X}_1^n = (X_1,\ldots,X_n)$ — of the solution of the minimization problem

$$\Phi_{\mathbb{P}}(\mathfrak{\mathbb{Q}}) := \inf_{\mathbb{Q}\in\mathfrak{\mathbb{Q}}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P})$$

which quantifies the adequacy (misspecification error) of the model $\mathfrak{\mathbb{Q}}$ for modelling the true unknown data-generating mechanism $\mathbb{P}$. Indeed, by replacing the right-hand side in (121) by its finite counterpart, we deduce for given large n (with a slight abuse of notation)

$$\underline{M}_n^{BS,PD}\Big(X_1,\ldots,X_n\Big) := \underline{M}_n^{BS,PD}\Big(\mathbb{P}_n^{emp}(\mathbf{X}_1^n)\Big) := F^{\leftarrow}_{\gamma,\widetilde{c},1}\Big(-\frac{1}{n}\log\,\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}_{n,\mathbf{X}}^{w\mathbf{W}} \in \mathfrak{\mathbb{Q}}\Big]\Big) \;\approx\; \inf_{\mathbb{Q}\in\mathfrak{\mathbb{Q}}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P}), \tag{122}$$

where PD is the abbreviation for *power divergences*; it remains to estimate the involved conditional probability $\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}_{n,\mathbf{X}}^{w\mathbf{W}} \in \mathfrak{\mathbb{Q}}\Big]$ in (122). The latter can be performed either by a *naive estimator* of the frequency of those replications of $\boldsymbol{\xi}_{n,\mathbf{X}}^{\mathbf{W}}$ which hit $\mathfrak{\mathbb{Q}}$, or more efficiently by some improved estimator; for details, the reader is referred to Section X of Broniatowski & Stummer [1], see also the corresponding extensions given in Section XIV below, where the latter also provides e.g. estimates of the *minimizers*).

B. The pure data-analytic view on the risk case

Let us begin with the remark that — by construction[21] — the above-mentioned sample-based function (statistical functional) $\underline{M}_n^{BS,PD}\Big(X_1,\ldots,X_n\Big)$ *does not* coincide with the sample-based function (with slight abuse of notation)

$$\underline{FM}_n^{BS,PD}\Big(X_1,\ldots,X_n\Big) := \underline{FM}_n^{BS,PD}\Big(\mathbb{P}_n^{emp}(\mathbf{X}_1^n)\Big) := \inf_{\mathbb{Q}\in\mathfrak{\mathbb{Q}}} D_{\widetilde{c}\cdot\varphi_\gamma}\Big(\mathbb{Q},\mathbb{P}_n^{emp}(\mathbf{X}_1^n)\Big)$$

which quantifies the adequacy (misspecification error) of the model $\mathfrak{\mathbb{Q}}$ for describing — the vectorized histogram of — the *sample* $X_1,\ldots,X_n$; the corresponding minimizer corresponds to the prominent (non-parametric) *sample-based minimum (power) divergence estimator*. For more details on the latter, see Section XIV (and in particular, Remark 102) below.

Now suppose that we have observed some *concrete data* $\mathbf{x}_1^n := (x_1,\ldots,x_n)$ (which are, as usual, interpreted as realizations of the random variables $\mathbf{X}_1^n = (X_1,\ldots,X_n)$). Moreover, let us summarize the data by the corresponding vectorized histogram

[20] recall from basic probability theory that for any event $B \in \mathcal{A}$ the conditional probability $\mathbb{\Pi}_{\mathbf{X}_1^n}[\,B\,] := \mathbb{\Pi}[\,B\,|\,X_1,\ldots,X_n]$ can in particular be rewritten as $h(X_1,\ldots,X_n)$ for some ($B-$dependent) measurable function $h : \mathcal{Y}^n \mapsto \mathbb{R}$; accordingly, for observed *concrete data* $\mathbf{x}_1^n := (x_1,\ldots,x_n)$ (which are, as usual, interpreted as realizations of the random variables $\mathbf{X}_1^n = (X_1,\ldots,X_n)$) one takes $\mathbb{\Pi}_{\mathbf{x}_1^n}[\,B\,] := \mathbb{\Pi}[\,B\,|\,X_1 = x_1,\ldots,X_n = x_n] := h(x_1,\ldots,x_n)$. The corresponding expectations will be denoted by $\mathbb{E}_{\mathbb{\Pi}_{\mathbf{X}_1^n}}[\,\cdot\,] := \mathbb{E}_{\mathbb{\Pi}}[\,\cdot\,|\,X_1,\ldots,X_n]$ respectively $\mathbb{E}_{\mathbb{\Pi}_{\mathbf{x}_1^n}}[\,\cdot\,] := \mathbb{E}_{\mathbb{\Pi}}[\,\cdot\,|\,x_1,\ldots,x_n]$.

[21] $\mathbb{P}_n^{emp}(\mathbf{X}_1^n)$ only enters in the construction of the partition $I_k^{(n)}(\mathbf{X}_1^n)$ $(k = 1,\ldots,K)$

$\mathbb{P}_n^{emp}(\mathbf{x}_1^n) \in \mathbb{S}_{>0}^K$ (or, more generally, any data-dependent vector $\mathbb{P}_n(\mathbf{x}_1^n) \in \mathbb{S}_{>0}^K$) so that (for large enough n) each data value d_k ($k = 1, \ldots, K$) is observed at least once. Independently of the nature of the true unknown original data-generating meachanism $\mathbb{P}$ (and thus, independently on the validity of (119)), let us find — as usual — the adequacy (misspecification error) of the model $\mathfrak{Q}$ for describing — the vectorized histogram of — the *concrete data* $x_1, \ldots, x_n$; in other words, in the following let us solve the *deterministic* minimization problem (also called estimation of the *empirical* risk)

$$\underline{FM}_n^{BS,PD}\Big(x_1, \ldots, x_n\Big) = \underline{FM}_n^{BS,PD}\Big(\mathbb{P}_n^{emp}(\mathbf{x}_1^n)\Big) = \inf_{\mathbb{Q} \in \mathfrak{Q}} D_{\widetilde{c} \cdot \varphi_\gamma}\Big(\mathbb{Q}, \mathbb{P}_n^{emp}(\mathbf{x}_1^n)\Big); \tag{123}$$

the corresponding minimizer, i.e. the *data-based* minimum (power) divergence estimator, will be given in Section XIV below.

The deterministic minimization problem (123) can be treated via our BS-method of the above Section VI, by taking $\mathbf{P} := \mathbb{P}_n^{emp}(\mathbf{x}_1^n)$ and thus $A = M_{\mathbb{P}_n^{emp}(\mathbf{x}_1^n)} = 1$ (and accordingly, we omit the tildes for the involved variables). Since the index n now plays a completely different role as in Section VI, we introduce a *new additional index* $m \in \mathbb{N}$ which will be the analogue of the approximation-step-indicating index n used in Section VI. Accordingly, for any $m \in \mathbb{N}$ and any $k \in \{1, \ldots, K-1\}$, let $m_k(\mathbf{x}_1^n) := \lfloor m \cdot p_{n,k}^{emp}(\mathbf{x}_1^n) \rfloor$ and $m_K(\mathbf{x}_1^n) := m - \sum_{k=1}^{K-1} m_k(\mathbf{x}_1^n)$; for this, we assume that $m \in \mathbb{N}$ is large enough, namely $m \geq \max_{k \in \{1,\ldots,K\}} \frac{1}{p_{n,k}^{emp}(\mathbf{x}_1^n)}$, such that all the integers $m_k(\mathbf{x}_1^n)$ ($k = 1, \ldots, K$) are non-zero (recall that we have $p_{n,k}^{emp}(\mathbf{x}_1^n) > 0$ for large enough $n \in \mathbb{N}$). Clearly, one gets

$$\lim_{m \to \infty} \frac{m_k(\mathbf{x}_1^n)}{m} = p_{n,k}^{emp}(\mathbf{x}_1^n), \qquad k = 1, \ldots, K. \tag{124}$$

With this at hand, we decompose the set $\{1, \ldots, m\}$ of all integers from 1 to m into the disjoint blocks $I_1^{(m)}(\mathbf{x}_1^n) := \{1, \ldots, m_1(\mathbf{x}_1^n)\}$, $I_2^{(m)}(\mathbf{x}_1^n) := \{m_1(\mathbf{x}_1^n) + 1, \ldots, m_1(\mathbf{x}_1^n) + m_2(\mathbf{x}_1^n)\}$, and so on until the last block $I_K^{(m)}(\mathbf{x}_1^n) := \{\sum_{k=1}^{K-1} m_k(\mathbf{x}_1^n) + 1, \ldots, m\}$. Thus, $I_k^{(m)}(\mathbf{x}_1^n)$ has $m_k(\mathbf{x}_1^n) \geq 1$ elements. From this, we construct

$$\boldsymbol{\xi}_{n,m,\mathbf{x}}^{w\mathbf{W}} := \begin{cases} \left(\frac{\sum_{i \in I_1^{(m)}(\mathbf{x}_1^n)} W_i}{\sum_{k=1}^K \sum_{i \in I_k^{(m)}(\mathbf{x}_1^n)} W_i}, \ldots, \frac{\sum_{i \in I_K^{(m)}(\mathbf{x}_1^n)} W_i}{\sum_{k=1}^K \sum_{i \in I_k^{(m)}(\mathbf{x}_1^n)} W_i} \right), & \text{if } \sum_{j=1}^m W_j \neq 0, \\ (\infty, \ldots, \infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^m W_j = 0. \end{cases} \tag{125}$$

Remark 64: Alternatively to the above procedure, we could also introduce a *new additional index* (with a slight abuse of notation) $m \in \mathbb{N}$ such that $m \cdot n$ (rather than m itself) will be the analogue of the approximation-step-indicating index n used in Section VI. Indeed, by construction, all the $p_{n,k}^{emp}(\mathbf{x}_1^n)$ are rational numbers and n is the smallest integer such that all $n \cdot p_{n,k}^{emp}(\mathbf{x}_1^n) = n_k(\mathbf{x}_1^n)$ ($k = 1, \ldots, K$) are integers (which for large enough data size n are non-zero); consequently, also all $m \cdot n \cdot p_{n,k}^{emp}(\mathbf{x}_1^n) = m \cdot n_k(\mathbf{x}_1^n)$ are (non-zero) integers. With this, we replace the partitions $I_k^{(n)}$ (having n_k elements, cf. (118)) by the new "$m-$fold blown-up" partitions $A_k^{(m \cdot n)}(\mathbf{x}_1^n)$ of size $m_k(\mathbf{x}_1^n) := m \cdot n_k(\mathbf{x}_1^n)$ and thus, $\frac{m_k(\mathbf{x}_1^n)}{m \cdot n} = p_{n,k}^{emp}(\mathbf{x}_1^n)$, $k \in \{1, \ldots, K\}$ (which means that here the appropriately adapted condition (124) holds even pointwise rather than only in the limit). More detailed, in the above-mentioned *batch procedure* we may reorder such that $A_1^{(m \cdot n)}(\mathbf{x}_1^n) := \{1, \ldots, m \cdot n_1(\mathbf{x}_1^n)\}$, $A_2^{(m \cdot n)}(\mathbf{x}_1^n) := \{m \cdot n_1(\mathbf{x}_1^n) + 1, \ldots, m \cdot (n_1(\mathbf{x}_1^n) + n_2(\mathbf{x}_1^n))\}$, …, $A_K^{(m \cdot n)}(\mathbf{x}_1^n) := \{m \cdot \sum_{i=1}^{K-1} n_i(\mathbf{x}_1^n) + 1, \ldots, m \cdot \sum_{i=1}^{K} n_i(\mathbf{x}_1^n)\}$, such that $\bigcup_{k=1}^K A_k^{(m \cdot n)}(\mathbf{x}_1^n) = \{1, \ldots, m \cdot n\}$. When the above-mentioned *online acquisition* of the data x_i's is required, then we usually do not reorder but after the appearance of a data point (say) $x_{n+1} = d_k$ in the $(n+1)-$th concrete data acquisition we immediately add m *new* (e.g. the next available) indices to the old index set $A_k^{(m \cdot n)}(\mathbf{x}_1^n)$ to end up with the new index subset $A_k^{(m \cdot (n+1))}(\mathbf{x}_1^{n+1})$ (and keep all the other index subsets the same, i.e. $A_j^{(m \cdot (n+1))}(\mathbf{x}_1^{n+1}) := A_j^{(m \cdot n)}(\mathbf{x}_1^n)$ for $j \neq k$). Accordingly, the $A_k^{(m \cdot n)}(\mathbf{x}_1^n)$'s do generally not consist in consecutive indexes, which does not make any change with respect to the resulting construction. With all this in hand, in the following investigations we can alternatively employ

$$\boldsymbol{\xi}_{n,m,\mathbf{x}}^{w\mathbf{W}} := \begin{cases} \left(\frac{\sum_{i \in A_1^{(m \cdot n)}(\mathbf{x}_1^n)} W_i}{\sum_{k=1}^K \sum_{i \in A_k^{(m \cdot n)}(\mathbf{x}_1^n)} W_i}, \ldots, \frac{\sum_{i \in A_K^{(m \cdot n)}(\mathbf{x}_1^n)} W_i}{\sum_{k=1}^K \sum_{i \in A_k^{(m \cdot n)}(\mathbf{x}_1^n)} W_i} \right), & \text{if } \sum_{j=1}^{m \cdot n} W_j \neq 0, \\ (\infty, \ldots, \infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^{m \cdot n} W_j = 0, \end{cases} \tag{126}$$

instead of (125) together with appropriately using $m \cdot n$ instead of m for the corresponding rates and limit-takings; such a construction has been employed in Broniatowski & Stummer [1].

In terms of (125) (with fixed n), we can apply Theorem 44(b) to end up with

$$\underline{FM}_n^{BS,PD}\Big(x_1, \ldots, x_n\Big) = \underline{FM}_n^{BS,PD}\Big(\mathbb{P}_n^{emp}(\mathbf{x}_1^n)\Big) = \inf_{\mathbf{Q} \in \mathfrak{Q}} D_{\widetilde{c} \cdot \varphi_\gamma}(\mathbf{Q}, \mathbb{P}_n^{emp}(\mathbf{x}_1^n)) = F_{\gamma, \widetilde{c}, 1}^{\leftarrow}\Big(- \lim_{m \to \infty} \frac{1}{m} \log \mathbb{\Pi}_{\mathbf{x}_1^n}\Big[\boldsymbol{\xi}_{n,m,\mathbf{x}}^{w\mathbf{W}} \in \mathfrak{Q}\Big] \Big) \tag{127}$$

for all sets $\mathfrak{Q} \subset \widetilde{\mathcal{M}}_\gamma$ (with $A = 1$) satisfying the regularity properties (67) *in the relative topology*; here, in (127) we have used the abbreviation $\mathbb{\Pi}_{\mathbf{x}_1^n}[\,\cdot\,] := \mathbb{\Pi}[\,\cdot \,|\, X_1 = x_1, \ldots, X_n = x_n]$ (recall footnote 20).

C. The combined view

Of course, one may also want to combine the statistical view (cf. Subsection IX-A) with the pure data-analytic view on the risk case (cf. Subsection IX-B). This amounts to replace the concrete (deterministic) data $\mathbf{x}_1^n := (x_1, \ldots, x_n)$ with the (random) sample $\mathbf{X}_1^n := (X_1, \ldots, X_n)$ assuming (116). Then the following coherence/consistency result

$$\Phi_{\mathbb{P}}(\mathfrak{Q}) = \inf_{\mathbb{Q}\in\mathfrak{Q}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P}) = \lim_{n\to\infty} \underline{FM}_n^{BS,PD}\Big(X_1,\ldots,X_n\Big) = \lim_{n\to\infty}\lim_{m\to\infty} F^{\leftarrow}_{\gamma,\widetilde{c},1}\Big(-\frac{1}{m}\log \mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{W}}_{n,m,\mathbf{X}} \in \mathfrak{Q}\Big]\Big) \qquad a.s. \tag{128}$$

should hold — under appropriate analytic conditions — for all compact sets $\mathfrak{Q} \subset \widetilde{\mathcal{M}}_\gamma$ (with $A = 1$) satisfying (67) and (68). Indeed, by denoting

$$\mathbf{R}_n := \mathbf{R}_n(\mathbf{X}_1^n) := \mathbb{P}_n^{emp}(\mathbf{X}_1^n), \quad \Phi_{\mathbf{R}_n}(\mathbb{Q}) := D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{R}_n), \quad \mathbf{R} := \mathbb{P}, \quad \Phi_{\mathbf{R}}(\mathbb{Q}) := D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbf{R}),$$

the desired result (128) follows from (127) together with

$$\Phi_{\mathbf{R}_n}(\cdot) \text{ converges (as } n \text{ tends to } \infty) \text{ a.s. to } \Phi_{\mathbf{R}}(\cdot) \text{ uniformly on } \mathfrak{Q} \text{ whenever } \mathbf{R}_n \text{ converges (as } n \text{ tends to } \infty) \text{ a.s. to } \mathbf{R}. \tag{129}$$

As far as the latter (which will be also used in other contexts below) is concerned, in the case where the minimizer-set $\mathcal{A} := \operatorname{argmin}_{\mathbb{Q}\in\mathfrak{Q}} \Phi_{\mathbf{R}}(\mathbb{Q})$ basically consists of isolated points, the a.s. convergence of the family of minimizers of $\Phi_{\mathbf{R}_n}$ towards $\mathcal{A}$ can be proved under adequate analytical conditions (see e.g. Theorem 5.7 of Van der Vaart [371]). The latter can, for instance, be tackled by appropriately carrying over the parametric-case-concerning *grand consistency theorem* (i.e. Theorem 2.1) of Kuchibhotla & Basu [372] to our non-parametric set-up with compact $\mathfrak{Q}$ (e.g. for the divergence generator $\varphi_{\alpha,\beta,\widetilde{c}}$ of Example 13, which satisfies $\varphi_{\alpha,\beta,\widetilde{c}}(0) + \varphi'_{\alpha,\beta,\widetilde{c}}(\infty) < \infty$ and thus their Assumption (C2)).

X. Narrow-Sense Bare-Simulation-Minimization of innmin-Bregman divergences under risk

For scaled Bregman distances we first derive the following fundamental risk-case-extension of Theorem 46:

Theorem 65: Suppose that $(X_i)_{i\in\mathbb{N}}$ is a sequence of random variables with values in $\mathcal{Y} := \{d_1, \cdots, d_K\}$ such that (119) holds for some probability vector $\mathbb{P} \in \mathbb{S}_{>0}^K$. Moreover, assume that the divergence generator φ satisfies Condition 37 and let $\mathbf{Q}^{**} \in \mathbb{R}^K$ be such that (21) holds. Additionally, we suppose that $\mathfrak{Q} \subset \mathbb{S}^K$ satisfies the regularity properties (67) in the relative topology as well as the finiteness property $\inf_{\mathbb{Q}\in\mathfrak{Q}} D^{SBD}_{\varphi,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**}) < \infty$ (cf. (79)). Moreover, let $V := (\mathbf{V}_i)_{i\in\mathbb{N}}$ be a sequence of random vectors which are independent of $(X_i)_{i\in\mathbb{N}}$ [22] and which are constructed by (25) and (26) (where we write V instead of $\widetilde{V}$, since $M_{\mathbb{P}} = 1$ and thus $\widetilde{p}_k = p_k/M_{\mathbb{P}} = p_k$ as well as $\widetilde{\zeta} = \zeta$). Moreover, let $I_k^{(n)}(\mathbf{X}_1^n) := \big\{i \in \{1,\ldots,n\} :\ X_i = d_k\big\}$ $(k = 1,\ldots,K)$ be the partitions constructed in (118). Then, in terms of the random vectors $\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}$ given by

$$\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}} \quad := \quad \begin{cases} \left(\frac{\sum_{i\in I_1^{(n)}(\mathbf{x}_1^n)} V_i}{\sum_{k=1}^K \sum_{i\in I_k^{(n)}(\mathbf{x}_1^n)} V_i}, \ldots, \frac{\sum_{i\in I_K^{(n)}(\mathbf{x}_1^n)} V_i}{\sum_{k=1}^K \sum_{i\in I_k^{(n)}(\mathbf{x}_1^n)} V_i}\right), & \text{if } \sum_{j=1}^n V_j \neq 0, \\ (\infty,\ldots,\infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^n V_j = 0, \end{cases} \tag{130}$$

there holds

$$\inf_{\mathbb{Q}\in\mathfrak{Q}} \inf_{m\neq 0} D^{SBD}_{\varphi,\mathbb{P}}(m\cdot\mathbb{Q},\mathbf{Q}^{**}) = \inf_{m\neq 0} \inf_{\mathbb{Q}\in\mathfrak{Q}} D^{SBD}_{\varphi,\mathbb{P}}(m\cdot\mathbb{Q},\mathbf{Q}^{**}) = -\lim_{n\to\infty} \frac{1}{n}\log \mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}} \in \mathfrak{Q}\Big]. \tag{131}$$

The proof of Theorem 65 will be given in Appendix A below.

By combining Theorem 65 with (88), (90) and (92) we straightforwardly deduce the following

Theorem 66: Let $\mathbb{P} \in \mathbb{S}_{>0}^K$, $(X_i)_{i\in\mathbb{N}}$, $V := (\mathbf{V}_i)_{i\in\mathbb{N}}$, $\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}$ as in Theorem 65, for the special choice $\varphi := \widetilde{c}\cdot\varphi_\gamma$ (cf. (16)) with some arbitrarily fixed $\widetilde{c} \in]0,\infty[$, $\gamma \in \mathbb{R}\setminus]1,2[$ together with some arbitrarily fixed $\mathbf{Q}^{**} \in \widetilde{\mathcal{N}}_\gamma$ (cf. (T1)) implying (21). Then there holds

$$\inf_{\mathbb{Q}\in\mathfrak{Q}} \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**}) = -\lim_{n\to\infty} \frac{1}{n}\log \mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}} \in \mathfrak{Q}\Big] \tag{132}$$

for all sets $\mathfrak{Q} \subset \widetilde{\mathcal{M}}_\gamma$ (cf. (T1)) satisfying the regularity properties (67) *in the relative topology*. In particular, for each such $\mathbb{P} \in \mathbb{S}_{>0}^K$ with (119) the function $\Phi_{\mathbb{P}}(\cdot) := \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\cdot,\mathbf{Q}^{**})$ is bare-simulation minimizable (BS-minimizable) in the narrow sense (cf. (2) in Definition 1 and the special case of Remark 2(a)) on all sets $\mathfrak{Q} \subset \widetilde{\mathcal{M}}_\gamma$ satisfying (67) *in the relative topology*.

[22] notice that all the V_i and X_i live on the *same* underlying probability space $(\mathfrak{X},\mathcal{A},\mathbb{\Pi})$

Recall from Remark 50 that $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**})$ is a divergence (which we have called innmin-SBD). Accordingly, based on the limit statement (132) (rather than (121)) we can follow the spirit of the previous Section IX to estimate — based on random scaling-samples $\mathbf{X}_1^n=(X_1,\ldots,X_n)$ satisfying (116) — the directed distance (divergence) from $\mathbf{Q}^{**}$ to the model $\mathfrak{M}$ under a true unknown "scaling-data-generating" mechanism $\mathbb{P}$, *given by*

$$\Phi_{\mathbb{P}}(\mathfrak{M}):=\inf_{\mathbb{Q}\in\mathfrak{M}}\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**}),$$

via (with a slight abuse of notation)

$$\underline{M}_n^{BS,SBD}\Big(X_1,\ldots,X_n\Big):=\underline{M}_n^{BS,SBD}\Big(\mathbb{P}_n^{emp}(\mathbf{X}_1^n)\Big):=-\frac{1}{n}\log\,\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}\in\mathfrak{M}\Big]\;\approx\;\inf_{\mathbb{Q}\in\mathfrak{M}}\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**})\tag{133}$$

together with an estimate the involved conditional probability $\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}\in\mathfrak{M}\Big]$ in (133) (for given large n). Moreover, in the spirit of Subsection IX-B we can approximate the above-mentioned randomly-scaled directed distance — in terms of the vectorized histogram of the *concrete scaling-data* $x_1,\ldots,x_n$ — by

$$\underline{FM}_n^{BS,SBD}\Big(x_1,\ldots,x_n\Big)=\underline{FM}_n^{BS,SBD}\Big(\mathbb{P}_n^{emp}(\mathbf{x}_1^n)\Big):=\inf_{\mathbb{Q}\in\mathfrak{M}}\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}_n^{emp}(\mathbf{x}_1^n)}(\mathbb{Q},\mathbf{Q}^{**}),\tag{134}$$

via an appropriate application of Theorem 49 — by substituting $\mathbf{P}$ by $\mathbb{P}_n^{emp}(\mathbf{x}_1^n)$ and $\mathbb{\Pi}[\,\cdot\,]$ by $\mathbb{\Pi}_{\mathbf{x}_1^n}[\,\cdot\,]$ — and end up with

$$\underline{FM}_n^{BS,SBD}\Big(x_1,\ldots,x_n\Big)=-\lim_{m\to\infty}\frac{1}{m}\log\,\mathbb{\Pi}_{\mathbf{x}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,m,\mathbf{x}}\in\mathfrak{M}\Big]\tag{135}$$

for all sets $\mathfrak{M}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying the regularity properties (67) *in the relative topology*. Here, in (135) we have employed

$$\boldsymbol{\xi}^{w\mathbf{V}}_{n,m,\mathbf{x}}\;:=\;\begin{cases}\left(\frac{\sum_{i\in I_1^{(m)}(\mathbf{x}_1^n)}V_i}{\sum_{k=1}^K\sum_{i\in I_k^{(m)}(\mathbf{x}_1^n)}V_i},\ldots,\frac{\sum_{i\in I_K^{(m)}(\mathbf{x}_1^n)}V_i}{\sum_{k=1}^K\sum_{i\in I_k^{(m)}(\mathbf{x}_1^n)}V_i}\right), & \text{if }\sum_{j=1}^m V_j\neq 0,\\ (\infty,\ldots,\infty)=:\boldsymbol{\infty}, & \text{if }\sum_{j=1}^m V_j=0.\end{cases}\tag{136}$$

where the partitions $I_k^{(m)}(\mathbf{x}_1^n)$ are the same as in (125).

Remark 67: In analogy with Remark 64, alternatively to (136) we can use

$$\boldsymbol{\xi}^{w\mathbf{V}}_{n,m,\mathbf{x}}\;:=\;\begin{cases}\left(\frac{\sum_{i\in A_1^{(m\cdot n)}(\mathbf{x}_1^n)}V_i}{\sum_{k=1}^K\sum_{i\in A_k^{(m\cdot n)}(\mathbf{x}_1^n)}V_i},\ldots,\frac{\sum_{i\in A_K^{(m\cdot n)}(\mathbf{x}_1^n)}V_i}{\sum_{k=1}^K\sum_{i\in A_k^{(m\cdot n)}(\mathbf{x}_1^n)}V_i}\right), & \text{if }\sum_{j=1}^{m\cdot n}V_j\neq 0,\\ (\infty,\ldots,\infty)=:\boldsymbol{\infty}, & \text{if }\sum_{j=1}^{m\cdot n}V_j=0,\end{cases}\tag{137}$$

(where the partitions $A_k^{(m\cdot n)}(\mathbf{x}_1^n)$ are the same as in (126)) to derive

$$\underline{FM}_n^{BS,SBD}\Big(x_1,\ldots,x_n\Big)=-\lim_{m\to\infty}\frac{1}{m\cdot n}\log\,\mathbb{\Pi}_{\mathbf{x}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,m,\mathbf{x}}\in\mathfrak{M}\Big].$$

In the spirit of Subsection IX-C, by combining (134), (135) and (129) with

$$\mathbf{R}_n:=\mathbf{R}_n(\mathbf{X}_1^n):=\mathbb{P}_n^{emp}(\mathbf{X}_1^n),\quad \Phi_{\mathbf{R}_n}(\mathbb{Q}):=\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{R}_n}(\mathbb{Q},\mathbf{Q}^{**}),\quad \mathbf{R}:=\mathbb{P},\quad \Phi_{\mathbf{R}}(\mathbb{Q}):=\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{R}}(\mathbb{Q},\mathbf{Q}^{**}),$$

it is possible — under appropriate analytic conditions — to obtain the following coherence/consistency result

$$\Phi_{\mathbb{P}}(\mathfrak{M})=\inf_{\mathbb{Q}\in\mathfrak{M}}\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**})=\lim_{n\to\infty}\underline{FM}_n^{BS,SBD}\Big(X_1,\ldots,X_n\Big)=-\lim_{n\to\infty}\lim_{m\to\infty}\frac{1}{m}\log\,\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,m,\mathbf{X}}\in\mathfrak{M}\Big]\qquad a.s.$$

for all sets $\mathfrak{M}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying the regularity properties (67) *in the relative topology*.

Due to the lack of representability (cf. the explanations right after (94)) we *generally* can not give analogous results for the *original* scaled Bregman distance $D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**})$. The latter can be achieved, however, in the subsetup $\gamma=1$ as well as in the subsetup where $\mathbf{Q}^{**}:=C\cdot\mathbf{P}$ and $\gamma\in\mathbb{R}\backslash[1,2[$. To start with, by combining Theorem 66 with (95), (97), (99) and (101) for the special choice $A=1$, we can straightforwardly derive the following

Theorem 68: Let $\mathbb{P} \in \mathbb{S}^K_{>0}$, $(X_i)_{i\in\mathbb{N}}$, $V := (\mathbf{V}_i)_{i\in\mathbb{N}}$, $\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}$ as in Theorem 65, for the special choice $\varphi := \widetilde{c}\cdot\varphi_\gamma$ (cf. (16)) with some arbitrarily fixed $\widetilde{c}\in]0,\infty[$, $\gamma \in \mathbb{R}\setminus]1,\underline{2}[$. Then the following holds for any $C>0$:
(a) In case of $\gamma \in \mathbb{R}\setminus[1,2[$, for any subset $\mathfrak{W} \subset \mathcal{M}_\gamma$ (with $A=1$) with (67) one gets

$$\inf_{\mathbb{Q}\in\mathfrak{W}} \breve{F}_{\gamma,\widetilde{c},1,1,C}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},C\cdot\mathbb{P})\Big) = -\lim_{n\to\infty}\frac{1}{n}\log \mathbb{\Pi}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}\in\mathfrak{W}\Big]$$

and (equivalently)

$$\inf_{\mathbb{Q}\in\mathfrak{W}} D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},C\cdot\mathbb{P}) = \breve{F}^{\leftarrow}_{\gamma,\widetilde{c},1,1,C}\Big(-\lim_{n\to\infty}\frac{1}{n}\log \mathbb{\Pi}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n}\in\mathfrak{W}\Big]\Big). \tag{138}$$

In particular, the functions $\Phi_{\mathbf{P}}(\cdot) := \breve{F}_{\gamma,\widetilde{c},1,1,C}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\cdot,C\cdot\mathbb{P})\Big)$ and $\Phi_{\mathbf{P}}(\cdot) := D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\cdot,C\cdot\mathbb{P})$ are bare-simulation minimizable (BS-minimizable) in the narrow sense (cf. (2) in Definition 1 and the special case of Remark 2(a)) on all sets $\mathfrak{W}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying (67) *in the relative topology*.
(b) If $\gamma=1$, then for any $\mathbf{Q}^{**}\in\mathbb{R}^K_{>0}$ and any subset $\mathfrak{W}\subset\mathbb{S}^K$ with (67) one gets

$$\inf_{\mathbb{Q}\in\mathfrak{W}} \breve{F}_{1,\widetilde{c},1,M_{\mathbf{Q}^{**}}}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**})\Big) = -\lim_{n\to\infty}\frac{1}{n}\log \mathbb{\Pi}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}\in\mathfrak{W}\Big]$$

and (equivalently)

$$\inf_{\mathbb{Q}\in\mathfrak{W}} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**}) = \breve{F}^{\leftarrow}_{1,\widetilde{c},1,M_{\mathbf{Q}^{**}}}\Big(-\lim_{n\to\infty}\frac{1}{n}\log \mathbb{\Pi}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}\in\mathfrak{W}\Big]\Big). \tag{139}$$

In particular, the functions $\Phi_{\mathbb{P}}(\cdot) := \breve{F}_{1,\widetilde{c},A,M_{\mathbf{Q}^{**}}}\Big(D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbb{P}}(\cdot,\mathbf{Q}^{**})\Big)$ and $\Phi_{\mathbb{P}}(\cdot) := D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbb{P}}(\cdot,\mathbf{Q}^{**})$ are bare-simulation minimizable (BS-minimizable) in the narrow sense (cf. (2) in Definition 1 and the special case of Remark 2(a)) on all sets $\mathfrak{W}\subset\mathbb{S}^K$ satisfying (67) *in the relative topology*.

For the special case $C=1$ — respectively $\mathbf{Q}^{**}=\mathbb{P}$ if $\gamma=1$ — Theorem 68 collapses to Theorem 44.

Remark 69: Analogously as in Theorem 12 of Broniatowski & Stummer [1], we can also give equivalent versions of the three Theorems 65, 66 and 68 (and the proofs thereof) in terms of probability *distributions* $\mathbb{Q}\in\mathfrak{Q}$ and $\mathbb{P}$ as well as measure Q^{**}, rather than probability *vectors* $\mathbb{Q}\in\mathfrak{W}$ and $\mathbb{P}$ as well as vector $\mathbf{Q}^{**}$; for the sake of brevity, we omit the corresponding explicitness.

Now, analogously to the above investigations, we employ the limit statements (138) and (139) (rather than (132)) and estimate — based on random scaling-samples $\mathbf{X}^n_1=(X_1,\ldots,X_n)$ satisfying (116) — the directed distance (divergence) from $\mathbf{Q}^{**}$ to the model $\mathfrak{W}$ under a true unknown "scaling-data-generating" mechanism $\mathbb{P}$, *given by*

$$\Phi_{\mathbb{P}}(\mathfrak{W}) := \inf_{\mathbb{Q}\in\mathfrak{W}} D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},C\cdot\mathbb{P}), \quad \text{for } \gamma\in\mathbb{R}\setminus[1,2[,$$
$$\Phi_{\mathbb{P}}(\mathfrak{W}) := \inf_{\mathbb{Q}\in\mathfrak{W}} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**}), \quad \text{for } \gamma=1,$$

via (with a slight abuse of notation)

$$\underline{M}^{BS,SBD}_n\Big(X_1,\ldots,X_n\Big) := \underline{M}^{BS,SBD}_n\Big(\mathbb{P}^{emp}_n(\mathbf{X}^n_1)\Big) := \breve{F}^{\leftarrow}_{\gamma,\widetilde{c},1,1,C}\Big(-\frac{1}{n}\log \mathbb{\Pi}_{\mathbf{X}^n_1}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}\in\mathfrak{W}\Big]\Big) \approx \inf_{\mathbb{Q}\in\mathfrak{W}} D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q},C\cdot\mathbb{P}),$$
$$\text{for } \gamma\in\mathbb{R}\setminus[1,2[, \tag{140}$$

$$\underline{M}^{BS,SBD}_n\Big(X_1,\ldots,X_n\Big) := \underline{M}^{BS,SBD}_n\Big(\mathbb{P}^{emp}_n(\mathbf{X}^n_1)\Big) := \breve{F}^{\leftarrow}_{1,\widetilde{c},1,M_{\mathbf{Q}^{**}}}\Big(-\frac{1}{n}\log \mathbb{\Pi}_{\mathbf{X}^n_1}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}\in\mathfrak{W}\Big]\Big) \approx \inf_{\mathbb{Q}\in\mathfrak{W}} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**}),$$
$$\text{for } \gamma=1, \tag{141}$$

together with an estimate of the involved conditional probability $\mathbb{\Pi}_{\mathbf{X}^n_1}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}\in\mathfrak{W}\Big]$ in (140) and (141) (for given large n). Moreover, in the spirit of Subsection IX-B we can approximate the above-mentioned randomly-scaled directed distance — in terms of the vectorized histogram of the *concrete scaling-data* $x_1,\ldots,x_n$ — by

$$\underline{FM}^{BS,SBD}_n\Big(x_1,\ldots,x_n\Big) = \underline{FM}^{BS,SBD}_n\Big(\mathbb{P}^{emp}_n(\mathbf{x}^n_1)\Big) := \inf_{\mathbb{Q}\in\mathfrak{W}} D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}^{emp}_n(\mathbf{x}^n_1)}(\mathbb{Q},C\cdot\mathbb{P}^{emp}_n(\mathbf{x}^n_1)), \quad \text{for } \gamma\in\mathbb{R}\setminus[1,2[,$$
$$\underline{FM}^{BS,SBD}_n\Big(x_1,\ldots,x_n\Big) = \underline{FM}^{BS,SBD}_n\Big(\mathbb{P}^{emp}_n(\mathbf{x}^n_1)\Big) := \inf_{\mathbb{Q}\in\mathfrak{W}} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbb{P}^{emp}_n(\mathbf{x}^n_1)}(\mathbb{Q},\mathbf{Q}^{**}), \quad \text{for } \gamma=1,$$

via an appropriate application of Theorem 51, and end up with

$$\underline{FM}^{BS,SBD}_n\Big(x_1,\ldots,x_n\Big) = \breve{F}^{\leftarrow}_{\gamma,\widetilde{c},1,1,C}\Big(-\lim_{m\to\infty}\frac{1}{m}\log \mathbb{\Pi}_{\mathbf{x}^n_1}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,m,\mathbf{x}}\in\mathfrak{W}\Big]\Big), \quad \text{for } \gamma\in\mathbb{R}\setminus[1,2[,$$
$$\underline{FM}^{BS,SBD}_n\Big(x_1,\ldots,x_n\Big) = \breve{F}^{\leftarrow}_{1,\widetilde{c},1,M_{\mathbf{Q}^{**}}}\Big(-\lim_{m\to\infty}\frac{1}{m}\log \mathbb{\Pi}_{\mathbf{x}^n_1}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,m,\mathbf{x}}\in\mathfrak{W}\Big]\Big), \quad \text{for } \gamma=1,$$

for all sets $\mathfrak{M} \subset \widetilde{\mathcal{M}}_\gamma$ (with $A = 1$) satisfying the regularity properties (67) *in the relative topology*.

Finally, following the spirit of Subsection IX-C in an analogous fashion as above, it is possible — under appropriate analytic conditions — to obtain the following coherence/consistency result

$$\Phi_{\mathbb{P}}(\mathfrak{M}) = \inf_{\mathbb{Q}\in\mathfrak{M}} \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}}(\mathbb{Q}, C\cdot\mathbb{P}) = \lim_{n\to\infty} \underline{FM}^{BS,SBD}_n\Big(X_1,\ldots,X_n\Big) = \breve{F}^{\leftarrow}_{\gamma,\widetilde{c},1,1,C}\Big(-\lim_{n\to\infty}\lim_{m\to\infty}\frac{1}{m}\log \mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,m,\mathbf{X}} \in \mathfrak{M}\Big]\Big) \; a.s. \quad \text{for } \gamma \in \mathbb{R}\backslash[1,2[,$$

$$\Phi_{\mathbb{P}}(\mathfrak{M}) = \inf_{\mathbb{Q}\in\mathfrak{M}} \breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbb{P}}(\mathbb{Q}, \mathbf{Q}^{**}) = \lim_{n\to\infty} \underline{FM}^{BS,SBD}_n\Big(X_1,\ldots,X_n\Big) = \breve{F}^{\leftarrow}_{1,\widetilde{c},1,M_{\mathbf{Q}^{**}}}\Big(-\lim_{n\to\infty}\lim_{m\to\infty}\frac{1}{m}\log \mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,m,\mathbf{X}} \in \mathfrak{M}\Big]\Big) \; a.s. \quad \text{for } \gamma = 1,$$

for all sets $\mathfrak{M} \subset \widetilde{\mathcal{M}}_\gamma$ (with $A = 1$) satisfying the regularity properties (67) *in the relative topology*.

XI. Bare-Simulation-Method For General Divergence-Optimization-Problems Under Risk

A. Divergences under Risk and Friends

In the two previous Sections IX and X we have dealt — under risk (i.e. with unknown $\mathbb{P}$) — with *narrow-sense* bare-simulation *minimization* of $\varphi-$divergences $\Phi_{\mathbb{P}}(\mathbb{Q}) := D_\varphi(\mathbb{Q},\mathbb{P})$, innmin scaled Bregman distances $\Phi_{\mathbb{P}}(\mathbb{Q}) := \breve{D}^{SBD}_{\varphi,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**})$ and scaled Bregman distances $\Phi_{\mathbb{P}}(\mathbb{Q}) := D^{SBD}_{\varphi,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**})$ where (except for Theorem 65) as divergence generator φ we have taken the power functions $\varphi = \widetilde{c}\cdot\varphi_\gamma$ (cf. (16) with $\gamma \in \mathbb{R}\backslash[1,2[$ (and for Bregman distances restrictively $\mathbf{Q}^{**} = C\cdot\mathbb{P}$ in case of $\gamma \neq 1$); recall the different role of $\mathbb{P}$ ("point to be projected" vs. scaling) in these two contexts.

Also recall that we are interested in the (generally, non-narrow sense) bare-simulation constrained *minimization and maximization* of the *continuous* functions $\mathbf{Q} \mapsto \Phi_{\mathbf{P}}(\mathbf{Q})$ — given in Subsection V-A, and beyond. In fact, the there-involved functions (**supposedly on a compact $\mathfrak{M} \subset \mathbb{S}^K$ in this section**) $\mathbb{Q} \mapsto \Phi_{\mathbf{R}}(\mathbb{Q})$ partially carry even another (respectively, more general) vector-valued "parameter" $\mathbf{R}$ which in this section we *consider as unknown*. For instance, in (D1) and (D2) we have the only possibility $\Phi_{\mathbf{R}}(\mathbb{Q}) := D_\varphi(\mathbb{Q},\mathbf{R})$ and $\Phi_{\mathbf{R}}(\mathbb{Q}) := D^{OBD}_\varphi(\mathbb{Q},\mathbf{R})$ with $\mathbf{R} := \mathbf{P}$, whereas in (D3) we have three possibilities of an unknown "parameter" $\mathbf{R}$: $\Phi_{\mathbf{R}}(\mathbb{Q}) := D^{SBD}_{\varphi,\mathbf{M}}(\mathbb{Q},\mathbf{R})$ with $\mathbf{R} := \mathbf{P}$, $\Phi_{\mathbf{R}}(\mathbb{Q}) := D^{SBD}_{\varphi,\mathbf{R}}(\mathbb{Q},\mathbf{P})$ with $\mathbf{R} := \mathbf{M}$[23], and $\Phi_{\mathbf{R}}(\mathbb{Q}) := D^{SBD}_{\varphi,\mathbf{R}_1}(\mathbb{Q},\mathbf{R}_2)$ with 2K-dimensional concatenation $\mathbf{R} = (\mathbf{R}_1,\mathbf{R}_2) = (\mathbf{M},\mathbf{P})$. In (D4) we have even more possibilities, unless we deal e.g. with the interesting special case of the total Bregman distance $D^{TBD}_\varphi(\mathbb{Q},\mathbf{R})$ with $\mathbf{R} := \mathbf{P}$. The cases (D6), (D7) and (D8) can take unknown parameters $\mathbf{R}$ in an analogous way, whereas (D5) has no parameter.

Returning to the general context, in case that $\mathbf{R}$ is *known*, we can simply apply the two Theorems 52 and 54 (or alternatively, 55 and 58) with the special subsetup $A = 1$ to $\Phi(\cdot) := \Phi_{\mathbf{R}}(\cdot)$. In case that $\mathbf{R}$ is *unknown* (i.e. under risk), we have to adapt the above-mentioned procedures of the Sections IX respectively X. For this, we assume that for index $i \in \mathbb{N}$ the generation of the $i-$th (uncertainty-prone) data point is represented by the random variable Y_i [24] which takes values in the discrete set $\widetilde{\mathcal{Y}} := \Big\{\widetilde{d}_1,\cdots,\widetilde{d}_K\Big\}$ of K distinct values "of any kind"`. It is assumed that — with the notation $\mathbf{Y}_1^n := (Y_1,\ldots,Y_n)$ — there is a sequence $(\mathbf{R}_n(\mathbf{Y}_1^n))_{n\in\mathbb{N}}$ of (say) vector-valued function(al)s $\mathbf{R}_n(\mathbf{Y}_1^n)$ for which

$$\lim_{n\to\infty} \mathbf{R}_n(\mathbf{Y}_1^n) = \mathbf{R} \quad \text{a.s.} \tag{142}$$

B. Minimization via Base-Divergence-Method 1

We first choose the "transformed" power divergences $F_{\gamma,\widetilde{c},1}(D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\cdot))$ as the involved base-divergences. In analogy with the considerations of the Subsection IX-B, let us examine the pure data-analytic view on the risk case. Accordingly, we can approximate — based on concrete data $\mathbf{y}_1^n = (y_1,\ldots,y_n)$ from the samples $\mathbf{Y}_1^n = (Y_1,\ldots,Y_n)$ satisfying (142) — the *unknown-parameter-carrying function*

$$\Phi_{\mathbf{R}}(\mathfrak{M}) := \inf_{\mathbb{Q}\in\mathfrak{M}} \Phi_{\mathbf{R}}(\mathbb{Q})$$

by the corresponding empirical (i.e. concrete-data-dependent) minimum

$$\underline{FM}^{BS}_n\Big(y_1,\ldots,y_n\Big) = \underline{FM}^{BS}_n\Big(\mathbf{R}_n(\mathbf{y}_1^n)\Big) := \inf_{\mathbb{Q}\in\mathfrak{M}} \Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\mathbb{Q}), \tag{143}$$

via an appropriate application of Theorem 52, and end up with

[23] e.g. take $\varphi(t) := 2\cdot\varphi_2(t) = (t-1)^2$ (cf. (16)) to end up with the squared weighted ℓ_2-distance $G := D^{SBD}_{2\varphi_2,\mathbf{R}}(\mathbb{Q},\mathbf{P}) = \sum_{k=1}^K \frac{1}{r_k}\cdot|\,q_k - p_k\,|^2 = \big(D_{\ell_2,\mathbf{R}}(\mathbb{Q},\mathbf{P})\big)^2$ (cf. (45)), and choose $q_k > 0$, $p_k > 0$ to be two alternative fixed (i.e. non-random) cash-flows at (say) some future time $t_k > 0$ as well as r_k to be a random (e.g. random-term-structure-of-interest-rates connected) discount factor from t_k to $t_0 := 0$ (e.g. $t_0 \mathrel{\widehat{=}} 12.00$ hours tomorrow); accordingly, G is connected with the aggregated dissimilarity of the corresponding (interest-rate-prone) present-values at t_0.

[24] on the underlying probability space $(\mathfrak{X},\mathcal{A},\mathbb{\Pi})$

$$\underline{FM}_n^{BS}\Big(y_1,\ldots,y_n\Big) = -\lim_{m\to\infty}\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)-\Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{W}})\Big)\Big)\cdot 1\!\!1_{\mathfrak{W}}(\boldsymbol{\xi}_m^{w\mathbf{W}})\ \Big]\Big) \tag{144}$$

for all sets $\mathfrak{W}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying the regularity properties (67) *in the relative topology*. Here, in (144) we have employed the random vectors $\boldsymbol{\xi}_m^{w\mathbf{W}}$ constructed from an auxiliary deterministic probability vector $\mathbb{P}^{aux}\in\mathbb{S}_{>0}^K$ by — cf. (70) with m instead of n and blocks $I_k^{(m)}$ of sizes $m_k:=\lfloor m\cdot p_k^{aux}\rfloor$ $(k=1,\ldots,K-1)$ and $m_K:=m-\sum_{k=1}^{K-1}m_k$ instead of n_k, n_K —

$$\boldsymbol{\xi}_m^{w\mathbf{W}} \ := \ \begin{cases}\left(\frac{\sum_{i\in I_1^{(m)}}W_i}{\sum_{k=1}^K\sum_{i\in I_k^{(m)}}W_i},\ldots,\frac{\sum_{i\in I_K^{(m)}}W_i}{\sum_{k=1}^K\sum_{i\in I_k^{(m)}}W_i}\right), & \text{if } \sum_{j=1}^m W_j\neq 0,\\ (\infty,\ldots,\infty)=:\boldsymbol{\infty}, & \text{if } \sum_{j=1}^m W_j=0;\end{cases} \tag{145}$$

moreover, $(W_i)_{i\in\mathbb{N}}$ and $(Y_i)_{i\in\mathbb{N}}$ are supposed to be independent[25]. In (144) we have also used the *conditional* expectations $\mathbb{E}_{\mathbb{\Pi}_{\mathbf{y}_1^n}}[\,\cdot\,]:=\mathbb{E}_{\mathbb{\Pi}}[\,\cdot\,|\,Y_1=y_1,\ldots,Y_n=y_n]$ (cf. footnote 20).

Remark 70: The involved $\mathbb{P}^{aux}$ may also depend on the deterministic data sequence $\mathbf{y}_1^n$; for the special case $\mathbb{P}^{aux}:=\mathbb{P}_n^{emp}(\mathbf{y}_1^n)$, we obtain $\boldsymbol{\xi}_m^{w\mathbf{W}}=\boldsymbol{\xi}_{n,m,\mathbf{y}}^{w\mathbf{W}}$ (cf. (125) respectively (126)).

Based on (144), one can (for fixed n) approximate the corresponding empirical minimum $\inf_{\mathbb{Q}\in\mathfrak{W}}\Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\mathbb{Q})$ by

$$-\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)-\Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{W}})\Big)\Big)\cdot 1\!\!1_{\mathfrak{W}}(\boldsymbol{\xi}_m^{w\mathbf{W}})\ \Big]\Big)\ \approx\ \inf_{\mathbb{Q}\in\mathfrak{W}}\Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\mathbb{Q})$$

for large m.

Clearly, from (144) we can immediately derive for the random sample $\mathbf{Y}_1^n=(Y_1,\ldots,Y_n)$

$$\underline{FM}_n^{BS}\Big(Y_1,\ldots,Y_n\Big) = -\lim_{m\to\infty}\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)-\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{W}})\Big)\Big)\cdot 1\!\!1_{\mathfrak{W}}(\boldsymbol{\xi}_m^{w\mathbf{W}})\ \Big]\Big) \tag{146}$$

for all sets $\mathfrak{W}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying the regularity properties (67) *in the relative topology*; also the Remark 70 carries over correspondingly. Here, in (146) we have employed the *conditional* expectations $\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}[\,\cdot\,]:=\mathbb{E}_{\mathbb{\Pi}}[\,\cdot\,|\,Y_1,\ldots,Y_n]$ (see also footnote 20).

Based on (146), one can (for fixed n) approximate the corresponding sample-dependent minimum $\inf_{\mathbb{Q}\in\mathfrak{W}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$ by

$$-\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)-\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{W}})\Big)\Big)\cdot 1\!\!1_{\mathfrak{W}}(\boldsymbol{\xi}_m^{w\mathbf{W}})\ \Big]\Big)\ \approx\ \inf_{\mathbb{Q}\in\mathfrak{W}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \tag{147}$$

for large m.

In the spirit of Subsection IX-C, by combining (143), (146), (142) and (129) with $\mathbf{R}_n:=\mathbf{R}_n(\mathbf{Y}_1^n)$, it is possible — under appropriate analytic conditions — to obtain the following coherence/consistency result

$$\begin{aligned}\Phi_{\mathbf{R}}(\mathfrak{W}) &:= \inf_{\mathbb{Q}\in\mathfrak{W}}\Phi_{\mathbf{R}}(\mathbb{Q}) = \lim_{n\to\infty}\underline{FM}_n^{BS}\Big(Y_1,\ldots,Y_n\Big)\\ &= -\lim_{n\to\infty}\lim_{m\to\infty}\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)-\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{W}})\Big)\Big)\cdot 1\!\!1_{\mathfrak{W}}(\boldsymbol{\xi}_m^{w\mathbf{W}})\ \Big]\Big)\quad a.s.\end{aligned}$$

for all compact sets $\mathfrak{W}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying the regularity properties (67) *in the relative topology*.

Clearly, all the above considerations can be applied to tackle $\inf_{\mathbb{Q}\in\mathfrak{W}}\Phi(\mathbf{Q})$ for all the directed distances/divergences $\Phi(\cdot):=\Phi_{\mathbf{R}}(\cdot)$ and friends discussed in the above Subsection XI-A.

C. Maximization via Base-Divergence-Method 1

For the *non-narrow-sense* BS-maximizability under risk, one can proceed analogously to the previous Subsection XI-B, and approximate — based on concrete data $\mathbf{y}_1^n=(y_1,\ldots,y_n)$ from the samples $\mathbf{Y}_1^n=(Y_1,\ldots,Y_n)$ satisfying (142) — the *unknown-parameter-carrying function*

$$\Phi_{\mathbf{R}}(\mathfrak{W}) := \sup_{\mathbb{Q}\in\mathfrak{W}}\Phi_{\mathbf{R}}(\mathbb{Q}) \tag{148}$$

[25]notice that all the Y_i and W_i live on the *same* underlying probability space $(\mathfrak{X},\mathcal{A},\mathbb{\Pi})$

by the empirical (i.e. concrete-data-dependent) maximum given by

$$\overline{FM}_n^{BS}\Big(y_1,\ldots,y_n\Big) = \overline{FM}_n^{BS}\Big(\mathbf{R}_n(\mathbf{y}_1^n)\Big) := \sup_{\mathbb{Q}\in\mathfrak{\mathbf{\Omega}}} \Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\mathbb{Q}), \tag{149}$$

via an appropriate application of Theorem 54, and end up with

$$\overline{FM}_n^{BS}\Big(y_1,\ldots,y_n\Big) = \lim_{m\to\infty}\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)+\Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{\Omega}}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\ \Big]\Big) \tag{150}$$

for all sets $\mathfrak{\mathbf{\Omega}}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying the regularity properties (67) *in the relative topology*; in accordance with Remark 70, the involved $\mathbb{P}^{aux}$ may also depend on the deterministic data sequence $\mathbf{y}_1^n$; for the special case $\mathbb{P}^{aux}:=\mathbb{P}_n^{emp}(\mathbf{y}_1^n)$, we obtain $\boldsymbol{\xi}_m^{w\mathbf{W}}=\boldsymbol{\xi}_{n,m,\mathbf{y}}^{w\mathbf{W}}$ (cf. (125) respectively (126)). Based on (150), one can approximate (for fixed n) the corresponding empirical maximum by

$$\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)+\Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{\Omega}}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\ \Big]\Big)\ \approx\ \sup_{\mathbb{Q}\in\mathfrak{\mathbf{\Omega}}}\Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\mathbb{Q})$$

for large m. Clearly, from (150) we can immediately derive for the random sample $\mathbf{Y}_1^n=(Y_1,\ldots,Y_n)$

$$\overline{FM}_n^{BS}\Big(Y_1,\ldots,Y_n\Big) = \lim_{m\to\infty}\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)+\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{\Omega}}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\ \Big]\Big) \tag{151}$$

for all sets $\mathfrak{\mathbf{\Omega}}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying the regularity properties (67) *in the relative topology*; also the Remark 70 carries over correspondingly.

Based on (151), one can approximate (for fixed n) the corresponding sample-dependent maximum $\sup_{\mathbb{Q}\in\mathfrak{\mathbf{\Omega}}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$ by

$$\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)+\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{\Omega}}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\ \Big]\Big)\ \approx\ \sup_{\mathbb{Q}\in\mathfrak{\mathbf{\Omega}}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \tag{152}$$

for large m.

In the spirit of Subsection IX-C, by combining (149), (151), (142) and (129) with $\mathbf{R}_n:=\mathbf{R}_n(\mathbf{Y}_1^n)$, it is possible — under appropriate analytic conditions — to obtain the following coherence/consistency result

$$\begin{aligned}\Phi_{\mathbf{R}}(\mathfrak{\mathbf{\Omega}}) &:= \sup_{\mathbb{Q}\in\mathfrak{\mathbf{\Omega}}}\Phi_{\mathbf{R}}(\mathbb{Q}) = \lim_{n\to\infty}\overline{FM}_n^{BS}\Big(Y_1,\ldots,Y_n\Big)\\ &= \lim_{n\to\infty}\lim_{m\to\infty}\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)+\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{\Omega}}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\ \Big]\Big)\quad a.s.\end{aligned}$$

for all compact sets $\mathfrak{\mathbf{\Omega}}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying the regularity properties (67) *in the relative topology*.

Clearly, all the above considerations can be applied to tackle $\sup_{\mathbb{Q}\in\mathfrak{\mathbf{\Omega}}}\Phi(\mathbf{Q})$ for all the directed distances/divergences $\Phi(\cdot):=\Phi_{\mathbf{R}}(\cdot)$ and friends discussed in the above Subsection XI-A.

D. Minimization via Base-Divergence-Method 2

For the *non-narrow-sense* BS-minimizability under risk, one proceed analogously to the Subsection XI-B, but now alternatively choose the innmin scaled Bregman power divergences $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_{\gamma,\cdot}}(\cdot,\mathbf{Q}^{**})$ as involved base-divergences. Firstly, we approximate — based on the *concrete data* $\mathbf{y}_1^n:=(y_1,\ldots,y_n)$ and $\mathbf{R}_n(\mathbf{y}_1^n)$ — the corresponding empirical (i.e. concrete-data-dependent) minimum given by

$$\underline{FM}_n^{BS}\Big(y_1,\ldots,y_n\Big) = \underline{FM}_n^{BS}\Big(\mathbf{R}_n(\mathbf{y}_1^n)\Big) := \inf_{\mathbb{Q}\in\mathfrak{\mathbf{\Omega}}}\Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\mathbb{Q})\qquad\text{(cf. (143))}$$

via an appropriate application of Theorem 55, and end up with

$$\underline{FM}_n^{BS}\Big(y_1,\ldots,y_n\Big) = -\lim_{m\to\infty}\frac{1}{m}\log\Big(\ \mathbb{E}_{\mathbb{\Pi}_{\mathbf{y}_1^n}}\Big[\exp\Big(m\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}},\mathbf{Q}^{**})-\Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{V}}\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{\Omega}}}\big(\boldsymbol{\xi}_m^{w\mathbf{V}}\big)\ \Big]\Big) \tag{153}$$

for all sets $\mathfrak{\mathbf{\Omega}}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$) satisfying the regularity properties (67) *in the relative topology*. Here, in (153) we have employed the random vectors $\boldsymbol{\xi}_m^{w\mathbf{V}}$ constructed from an auxiliary deterministic probability vector $\mathbb{P}^{aux}\in\mathbb{S}^K_{>0}$ by — cf. (80)

with m instead of n and blocks $I_k^{(m)}$ of sizes $m_k := \lfloor m \cdot p_k^{aux} \rfloor$ $(k = 1, \ldots, K-1)$ and $m_K := m - \sum_{k=1}^{K-1} m_k$ instead of n_k, n_K —

$$\boldsymbol{\xi}_m^{w\mathbf{V}} \quad := \quad \begin{cases} \left(\frac{\sum_{i \in I_1^{(m)}} V_i}{\sum_{k=1}^K \sum_{i \in I_k^{(m)}} V_i}, \ldots, \frac{\sum_{i \in I_K^{(m)}} V_i}{\sum_{k=1}^K \sum_{i \in I_k^{(m)}} V_i} \right), & \text{if } \sum_{j=1}^m V_j \neq 0, \\ (\infty, \ldots, \infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^m V_j = 0; \end{cases} \tag{154}$$

moreover, $(V_i)_{i \in \mathbb{N}}$ and $(Y_i)_{i \in \mathbb{N}}$ are supposed to be independent[26]. In accordance with Remark 70, the involved $\mathbb{P}^{aux}$ may also depend on the deterministic data sequence $\mathbf{y}_1^n$; for the special case $\mathbb{P}^{aux} := \mathbb{P}_n^{emp}(\mathbf{y}_1^n)$, we obtain $\boldsymbol{\xi}_m^{w\mathbf{V}} = \boldsymbol{\xi}_{n,m,\mathbf{y}}^{w\mathbf{V}}$ (cf. (136) respectively (137)). Based on (153), one can (for fixed n) approximate the corresponding empirical minimum by

$$-\frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{y}_1^n}} \Big[\exp\Big(m \cdot \Big(\breve{D}^{SBD}_{\widetilde{c} \cdot \varphi_\gamma, \mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}}, \mathbf{Q}^{**}) - \Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big) \Big) \cdot 1\!\!1_{\boldsymbol{\mathfrak{Q}}}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big] \Big) \;\approx\; \inf_{\mathbb{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\mathbb{Q})$$

for large m. Certainly, from (153) we can straightforwardly deduce for the random sample $\mathbf{Y}_1^n = (Y_1, \ldots, Y_n)$

$$\underline{FM}_n^{BS}\Big(Y_1, \ldots, Y_n\Big) = -\lim_{m \to \infty} \frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}} \Big[\exp\Big(m \cdot \Big(\breve{D}^{SBD}_{\widetilde{c} \cdot \varphi_\gamma, \mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}}, \mathbf{Q}^{**}) - \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big) \Big) \cdot 1\!\!1_{\boldsymbol{\mathfrak{Q}}}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big] \Big) \tag{155}$$

for all sets $\boldsymbol{\mathfrak{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ (with $A = 1$) satisfying the regularity properties (67) *in the relative topology*; also the Remark 70 carries over correspondingly.

Based on (155), one can (for fixed n) approximate the corresponding sample-dependent minimum by

$$-\frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}} \Big[\exp\Big(m \cdot \Big(\breve{D}^{SBD}_{\widetilde{c} \cdot \varphi_\gamma, \mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}}, \mathbf{Q}^{**}) - \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big) \Big) \cdot 1\!\!1_{\boldsymbol{\mathfrak{Q}}}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big] \Big) \;\approx\; \inf_{\mathbb{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \tag{156}$$

for large m.

Finally, in the spirit of Subsection IX-C, by combining (143), (155), (142) and (129) with $\mathbf{R}_n := \mathbf{R}_n(\mathbf{Y}_1^n)$, it is possible — under appropriate analytic conditions — to obtain the following coherence/consistency result

$$\begin{aligned} \Phi_{\mathbf{R}}(\boldsymbol{\mathfrak{Q}}) &:= \inf_{\mathbb{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi_{\mathbf{R}}(\mathbb{Q}) = \lim_{n \to \infty} \underline{FM}_n^{BS}\Big(Y_1, \ldots, Y_n\Big) \\ &= -\lim_{n \to \infty} \lim_{m \to \infty} \frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}} \Big[\exp\Big(m \cdot \Big(\breve{D}^{SBD}_{\widetilde{c} \cdot \varphi_\gamma, \mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}}, \mathbf{Q}^{**}) - \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big) \Big) \cdot 1\!\!1_{\boldsymbol{\mathfrak{Q}}}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big] \Big) \quad a.s. \end{aligned}$$

for all sets $\boldsymbol{\mathfrak{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ (with $A = 1$) satisfying the regularity properties (67) *in the relative topology*.

Clearly, all the above considerations can be applied to tackle $\inf_{\mathbb{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q})$ for all the directed distances/divergences $\Phi(\cdot) := \Phi_{\mathbf{R}}(\cdot)$ and friends discussed in the above Subsection XI-A.

E. Maximization via Base-Divergence-Method 2

For the *non-narrow-sense* BS-maximizability under risk, one can proceed analogously to the previous Subsection XI-D, and approximate — based on concrete data $\mathbf{y}_1^n = (y_1, \ldots, y_n)$ from the samples $\mathbf{Y}_1^n = (Y_1, \ldots, Y_n)$ satisfying (142) — the *unknown-parameter-carrying function*

$$\Phi_{\mathbf{R}}(\boldsymbol{\mathfrak{Q}}) := \sup_{\mathbb{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi_{\mathbf{R}}(\mathbb{Q}) \qquad \text{(cf. (148))}$$

by the empirical (i.e. concrete-data-dependent) maximum

$$\overline{FM}_n^{BS}\Big(y_1, \ldots, y_n\Big) = \overline{FM}_n^{BS}\Big(\mathbf{R}_n(\mathbf{y}_1^n)\Big) := \sup_{\mathbb{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\mathbb{Q}) \qquad \text{(cf. (149))}$$

via an appropriate application of Theorem 58, and end up with

$$\overline{FM}_n^{BS}\Big(y_1, \ldots, y_n\Big) = \lim_{m \to \infty} \frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{y}_1^n}} \Big[\exp\Big(m \cdot \Big(\breve{D}^{SBD}_{\widetilde{c} \cdot \varphi_\gamma, \mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}}, \mathbf{Q}^{**}) + \Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big) \Big) \cdot 1\!\!1_{\boldsymbol{\mathfrak{Q}}}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big] \Big) \tag{157}$$

for all sets $\boldsymbol{\mathfrak{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ (with $A = 1$) satisfying the regularity properties (67) *in the relative topology*; in accordance with Remark 70, the involved $\mathbb{P}^{aux}$ may also depend on the deterministic data sequence $\mathbf{y}_1^n$; for the special case $\mathbb{P}^{aux} := \mathbb{P}_n^{emp}(\mathbf{y}_1^n)$, we obtain $\boldsymbol{\xi}_m^{w\mathbf{V}} = \boldsymbol{\xi}_{n,m,\mathbf{y}}^{w\mathbf{V}}$ (cf. (136) respectively (137)). Based on (157), one can (for fixed n) approximate the corresponding empirical maximum by

$$\frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{y}_1^n}} \Big[\exp\Big(m \cdot \Big(\breve{D}^{SBD}_{\widetilde{c} \cdot \varphi_\gamma, \mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}}, \mathbf{Q}^{**}) + \Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big) \Big) \cdot 1\!\!1_{\boldsymbol{\mathfrak{Q}}}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big] \Big) \;\approx\; \sup_{\mathbb{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi_{\mathbf{R}_n(\mathbf{y}_1^n)}(\mathbb{Q})$$

[26]notice that all the Y_i and V_i and live on the *same* underlying probability space $(\mathfrak{X}, \mathcal{A}, \mathbb{\Pi})$

for large m. Clearly, from (157) we can immediately derive for the random sample $\mathbf{Y}_1^n = (Y_1, \ldots, Y_n)$

$$\overline{FM}_n^{BS}\Big(Y_1, \ldots, Y_n\Big) = \lim_{m\to\infty} \frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m \cdot \Big(\breve{D}^{SBD}_{\breve{c}\cdot\varphi_\gamma, \mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}}, \mathbf{Q}^{**}) + \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{V}})\Big)\Big) \cdot \mathrm{l}_{\mathfrak{Q}}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big]\Big) \quad (158)$$

for all sets $\mathfrak{Q} \subset \widetilde{\mathcal{M}}_\gamma$ (with $A = 1$) satisfying the regularity properties (67) *in the relative topology*; also the Remark 70 carries over correspondingly.

Based on (158), one can (for fixed n) approximate the corresponding sample-dependent maximum by

$$\frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m \cdot \Big(\breve{D}^{SBD}_{\breve{c}\cdot\varphi_\gamma, \mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}}, \mathbf{Q}^{**}) + \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{V}})\Big)\Big) \cdot \mathrm{l}_{\mathfrak{Q}}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big]\Big) \;\approx\; \sup_{\mathbb{Q}\in\mathfrak{Q}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \quad (159)$$

for large m.

In the spirit of Subsection IX-C, by combining (149), (158), (142) and (129) with $\mathbf{R}_n := \mathbf{R}_n(\mathbf{Y}_1^n)$, it is possible — under appropriate analytic conditions — to obtain the following coherence/consistency result

$$\begin{aligned}\Phi_{\mathbf{R}}(\mathfrak{Q}) &:= \sup_{\mathbb{Q}\in\mathfrak{Q}} \Phi_{\mathbf{R}}(\mathbb{Q}) = \lim_{n\to\infty} \overline{FM}_n^{BS}\Big(Y_1, \ldots, Y_n\Big) \\ &= \lim_{n\to\infty} \lim_{m\to\infty} \frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m \cdot \Big(\breve{D}^{SBD}_{\breve{c}\cdot\varphi_\gamma, \mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}}, \mathbf{Q}^{**}) + \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{V}})\Big)\Big) \cdot \mathrm{l}_{\mathfrak{Q}}(\boldsymbol{\xi}_m^{w\mathbf{V}}) \Big]\Big) \quad a.s.\end{aligned}$$

for all sets $\mathfrak{Q} \subset \widetilde{\mathcal{M}}_\gamma$ (with $A = 1$) satisfying the regularity properties (67).

Clearly, all the above considerations can be applied to tackle $\sup_{\mathbb{Q}\in\mathfrak{Q}} \Phi(\mathbf{Q})$ for all the directed distances/divergences $\Phi(\cdot) := \Phi_{\mathbf{R}}(\cdot)$ and friends discussed in the above Subsection XI-A.

XII. Bare-Simulation Estimators For General Deterministic Divergence-Optimization-Problems

Recall that we are interested in the constrained optimization of the *continuous* distance-connected functions $\mathbf{\Omega} \ni \mathbf{Q} \mapsto \Phi(\mathbf{Q})$ in the above-mentioned cases (D1) to (D8) of Subsection V-A, and beyond (e.g. $\Phi(\cdot) := D_{\breve{\varphi}}(\cdot, \breve{\mathbf{P}})$ may be a $\breve{\varphi}-$divergence with pregiven $\breve{\mathbf{P}}$ and $\breve{\varphi}$, cf. Remark 23(iii)). In this section, we consider constraint sets $\mathbf{\Omega} \in \mathbb{R}^K$ with regularity properties (6) such that the function $\Phi(\cdot)$ possesses a (not necessarily unique) minimizer; for this, we construct *naive estimators* as well as *speed-up estimators* (approximations) of the *minimum value* $\inf_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) = \min_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$ and of the corresponding (set of) *minimizers* $\arg\inf_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) = \arg\min_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$. The case of maximum values and maximizers will be treated analogously. We mainly focus on compact sets $\mathbf{\Omega}$ but also discuss some relaxations thereof.

A. Naive estimators of min and argmin — Base-Divergence-Method 1, compact case

Recall that from Theorem 22(a) we obtain for any continuous function $\Phi : \mathbf{\Omega} \mapsto \mathbb{R}$ on a compact set $\mathbf{\Omega} \subset \mathbb{R}^K$ with (6) the assertion

$$\min_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \;=\; -\lim_{n\to\infty} \frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \Big(D_\varphi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}, \mathbf{P}\big) - \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big)\Big) \cdot \mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big) \Big]\Big)\,, \quad (160)$$

where $M_{\mathbf{P}} = \sum_{i=1}^K p_i > 0$ and

$$\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}} := \Big(\frac{1}{n} \sum_{i\in I_1^{(n)}} \widetilde{W}_i, \ldots, \frac{1}{n} \sum_{i\in I_K^{(n)}} \widetilde{W}_i\Big) \qquad \text{(cf. (13))}$$

is constructed from a sequence $\widetilde{W} := (\widetilde{W}_i)_{i\in\mathbb{N}}$ of random variables, where the $\widetilde{W}_i$'s are i.i.d. copies of the random variable $\widetilde{W}$ whose distribution is $\mathbb{\Pi}[\widetilde{W} \in \cdot\,] = \widetilde{\zeta}[\,\cdot\,]$ being attributed to the divergence generator $\widetilde{\varphi} := M_{\mathbf{P}} \cdot \varphi \in \widetilde{\Upsilon}(]a, b[)$ by the representability (9). Within such a set-up, we thus obtain for large $n \in \mathbb{N}$ (cf. (160)) the approximation

$$\Phi(\mathbf{\Omega}) := \min_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \;\approx\; -\frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \Big(D_\varphi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}, \mathbf{P}\big) - \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big)\Big) \cdot \mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big) \Big]\Big)\,, \quad (161)$$

and hence for getting an estimator of the minimum value $\Phi(\mathbf{\Omega})$ one can estimate the right-hand side of (161). To achieve this, for the rest of this section we assume that n is chosen such that all $n \cdot \frac{p_k}{M_{\mathbf{P}}}$ are integers (and hence, $n = \sum_{k=1}^K n_k$ with $n_k = n \cdot \frac{p_k}{M_{\mathbf{P}}}$) — the remaining case works analogously. Extending the lines of Broniatowski & Stummer [1] (who deal with the narrow-sense BS minimizability), a corresponding *naive (crude) estimator* can be constructed by

$$\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,1} \;:=\; -\frac{1}{n} \log \frac{1}{L} \sum_{\ell=1}^L \exp\Big(n \cdot \Big(D_\varphi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}, \mathbf{P}\big) - \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}\big)\Big)\Big) \cdot \mathbf{1}_{\mathbf{\Omega}}\Big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}\Big)\,, \quad (162)$$

where we simulate independently L copies $\widetilde{\mathbf{W}}^{(1)}, \ldots, \widetilde{\mathbf{W}}^{(L)}$ of the vector $\widetilde{\mathbf{W}} := \left(\widetilde{W}_1, \ldots, \widetilde{W}_n\right)$ with independent entries under $\widetilde{\zeta}$, and compute each of $\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(1)}}, \ldots, \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(L)}}$ according to (13). Clearly, with the help of the strong law of large numbers we get with $\widehat{\Phi(\mathbf{\Omega})}_{n,\infty}^{naive,1} := \lim_{L\to\infty} \widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,1}$ the following assertion:

Proposition 71:

$$\lim_{n\to\infty} \widehat{\Phi(\mathbf{\Omega})}_{n,\infty}^{naive,1} \;=\; \lim_{n\to\infty} \lim_{L\to\infty} \widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,1} \;=\; \Phi(\mathbf{\Omega}) \qquad \text{a.s.} \tag{163}$$

As the corresponding *very natural naive (crude) estimator* of the minimizer-set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$, we take

$$\widehat{\operatorname*{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})}_{n,L}^{naive,1} := \operatorname*{argmin}_{\boldsymbol{\nu}\in\mathcal{W}_{n,L}} \Phi(\boldsymbol{\nu}) \tag{164}$$

where $\mathcal{W}_{n,L} := \{M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}} : \ell \in \{1, \ldots, L\}\} \cap \mathbf{\Omega}$. In other words, as a corresponding *naive (crude) estimator* of the (not necessarily unique) element $\mathbf{Q}^*$ of the minimizer-set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$, we take the (not necessarily unique) element $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}$ of the set $\{M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}} : \ell \in \{1, \ldots, L\}\}$ such that $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}} \in \mathbf{\Omega}$ and

$$\Phi(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}) \;\le \Phi(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}) \qquad \text{for all } \ell = 1, \ldots, L \text{ for which } M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}} \text{ belongs to } \mathbf{\Omega}.$$

In short, $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}$ minimizes $\Phi(\cdot)$ amongst all values $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}$ at hand which fall into $\mathbf{\Omega}$. For large enough $n \in \mathbb{N}$ and $L \in \mathbb{N}$, such $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}$ exists since $\mathbf{\Omega}$ has non-void interior, by assumption (6). We prove that if L and n tend to infinity, then $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}$ concentrates to the above-mentioned set $\mathcal{Q}^*$ of minimizers of $\Phi(\cdot)$ on $\mathbf{\Omega}$. As usual in similar procedures, L is assumed to be large enough in order to justify some approximation for fixed n, typically the substitution of empirical means by expectations, since L is at disposal.
Next we show that $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}$ is a proxy minimizer of $\Phi(\cdot)$ on $\mathbf{\Omega}$, by proving the following

Proposition 72: There holds

$$\min_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \;\le\; \Phi\left(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}\right) \;\le\; \widehat{\Phi(\mathbf{\Omega})}_{n,\infty}^{naive,1} \;+\; o_{\mathbb{\Pi}}(1) \tag{165}$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L \to \infty$ and $n \to \infty$ under the distribution $\mathbb{\Pi}$ (recall that $\mathbb{\Pi}[(\widetilde{W}_1, \ldots, \widetilde{W}_n) \in \cdot\,] = \widetilde{\zeta}^{\otimes n}[\,\cdot\,]$).

The proof of Proposition 72 is given in Appendix A below.

Remark 73: In the current set-up of compact $\mathbf{\Omega}$ with (6), by taking the special case $\Phi(\mathbf{Q}) := D_\varphi(\mathbf{Q}, \mathbf{P})$ we obtain the naive BS-estimator $\widehat{D_\varphi(\mathbf{\Omega}, \mathbf{P})}_{n,L}^{naive,1}$ of the minimum value $\min_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q}, \mathbf{P})$ given in Broniatowski & Stummer [1]. In the latter paper, we had left the *open gap* of constructing a naive BS-estimator of the corresponding minimizer $\operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q}, \mathbf{P})$, which we have now *filled/resolved* by taking the corresponding special case $\widehat{\operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q}, \mathbf{P})}_{n,L}^{naive,1}$ of (164) (for a speed-up version thereof see $\widehat{\operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q}, \mathbf{P})}_{n,L}^{speedup,1}$ of (177) below).

Proposition 74: In the above set-up, one has

$$\lim_{n\to\infty} \lim_{L\to\infty} \Phi\Big(\widehat{\operatorname*{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})}_{n,L}^{naive,1}\Big) \;=\; \Phi(\mathbf{\Omega}) \qquad \text{a.s.},$$

and thus the quantity $\Phi\Big(\widehat{\operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})}_{n,L}^{naive,1}\Big)$ is a natural alternative to the estimate $\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,1}$ given in (162).

B. Naive estimators of max and argmax — Base-Divergence-Method 1, compact case

In the set-up of compact $\mathbf{\Omega}$ with (6), we can treat the maximizing problem completely analogously to the method in the previous Subsection XII-A. Indeed, by employing Theorem 32(a) instead of Theorem 22(a) we construct the naive BS-estimator

$$\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,1} \;:=\; \frac{1}{n} \log \frac{1}{L} \sum_{\ell=1}^{L} \exp\Big(n \cdot \Big(D_\varphi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}, \mathbf{P}\big) + \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}\big)\Big)\Big) \cdot \mathbf{1}_{\mathbf{\Omega}}\Big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}\Big), \tag{166}$$

of the maximum value $\Phi(\mathbf{\Omega}) := \max_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$; for this, we get

$$\lim_{n\to\infty} \widehat{\Phi(\mathbf{\Omega})}_{n,\infty}^{naive,1} \;=\; \lim_{n\to\infty} \lim_{L\to\infty} \widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,1} \;=\; \Phi(\mathbf{\Omega}) \qquad \text{a.s.}$$

instead of (163) (notice the different definition of the involved quantities). As the corresponding very natural naive BS-estimator of the maximizer $\operatorname{argmax}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$ we take

$$\widehat{\operatorname*{argmax}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi}(\mathbf{Q})^{naive,1}_{n,L} := \operatorname*{argmax}_{\boldsymbol{\nu}\in\mathcal{W}_{n,L}} \Phi(\boldsymbol{\nu})$$

instead of (164); in short, we take any $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}$ which maximizes $\Phi(\cdot)$ amongst all values $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}$ at hand which fall into $\mathbf{\Omega}$. For this, we obtain — instead of (165) — the assertion

$$\max_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \;\geq\; \Phi\left(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}\right) \;\geq\; \widehat{\Phi(\mathbf{\Omega})}^{naive,1}_{n,\infty} \;-\; o_{\mathbb{\Pi}}(1)$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L \to \infty$ and $n \to \infty$ under the distribution $\mathbb{\Pi}$. Hence, as L and n tend to infinity, $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}$ concentrates to the set of maximizers of $\Phi(\cdot)$ on $\mathbf{\Omega}$.

Proposition 75: In the above set-up, one has

$$\lim_{n\to\infty} \lim_{L\to\infty} \Phi\Big(\widehat{\operatorname*{argmax}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi}(\mathbf{Q})^{naive,1}_{n,L}\Big) \;=\; \Phi(\mathbf{\Omega}) \qquad \text{a.s.},$$

and thus the quantity $\Phi\Big(\widehat{\operatorname{argmax}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi}(\mathbf{Q})^{naive,1}_{n,L}\Big)$ is a natural alternative to the estimate $\widehat{\Phi(\mathbf{\Omega})}^{naive,1}_{n,L}$ given in (166).

C. Naive estimators — Base-Divergence-Method 1, non-compact case

Suppose that we are in the set-up of Theorem 22(b), which particularly means that $\mathbf{\Omega} \subset \mathbb{R}^K$ is not necessarily compact but satisfies (6) and (8), and that $\Phi : \mathbf{\Omega} \mapsto \mathbb{R}$ is a continuous function which satisfies the lower-bound condition (47) (notice that (47) trivially holds for the special case $\Phi(\mathbf{Q}) := D_\varphi(\mathbf{Q}, \mathbf{P})$, or if $\mathbf{\Omega}$ is bounded but not necessarily closed). Additionally, let us now assume that the minimum value is achieved, i.e. $\inf_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) = \min_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) = \Phi(\mathbf{Q}_{min})$ for some (not necessarily unique) point $\mathbf{Q}_{min} \in \mathbf{\Omega}$, and that the corresponding set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$ of minimizers is covered by a compact set $\mathbf{B} \in \mathbb{R}^K$ (e.g. we take $\mathbf{B} := cl(\mathbf{\Omega})$ in case that $\mathbf{\Omega}$ is bounded but not necessarily closed). In such a context, we can proceed as in Subsection XII-A by replacing $\mathbf{\Omega}$ with $\mathbf{\Omega} \cap \mathbf{B}$. The analogous procedure applies to the maximization problem, by proceeding as in Subsection XII-B by replacing $\mathbf{\Omega}$ with $\mathbf{\Omega} \cap \mathbf{B}$.

D. Naive estimators of min and argmin — Base-Divergence-Method 2, compact case

In the above Subsections XII-A, XII-B, XII-C, in order to optimize $\Phi(\cdot)$ we have chosen an appropriate $\varphi-$divergence $D_\varphi(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}, \mathbf{P})$ as the so-called *base divergence* (cf. Theorem 22 and Theorem 32); we have referred to this choice as *Base-Divergence-Method 1*. However, as can be seen from the alternative Theorem 30 and Theorem 34, for the optimization of the same $\Phi(\cdot)$ we can also choose — as Base-Divergence-Method 2 — as *base divergence* an appropriate scaled Bregman distance $D^{SBD}_{\varphi,\mathbf{P}}(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}, \mathbf{Q}^{**})$ (cf. (23)), where $\mathbf{Q}^{**} \in \mathbb{R}^K$ *NEED NOT* be in $\mathbf{\Omega}$; for instance, if the goal is to optimize $\mathbf{Q} \mapsto \Phi(\mathbf{Q}) := D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q}, \mathbf{Q}^{**})$ for application-context-specific $\mathbf{Q}^{**} \notin \mathbf{\Omega}$, the use of the base-divergence $D^{SBD}_{\varphi,\mathbf{P}}(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}, \mathbf{Q}^{**})$ appears naturally (as a side remark, recall that separable ordinary Bregman distances are subsumed as special cases $D^{OBD}_{\varphi}(\mathbf{Q}, \mathbf{Q}^{**}) := D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q}, \mathbf{Q}^{**})$ with $\mathbf{P} := \mathbf{1} = (1, \ldots, 1)$).

Remark 76: As will be seen in Subsection XII-G below, if we can freely choose $\mathbf{Q}^{**} \in int(\mathbf{\Omega})$ in an appropriate way, then we shall end up with *improved/speed-up estimator versions* of the above-mentioned *naive extimators* of Base-Divergence-Method 1.

However, let us start with the general case of any fixed $\mathbf{Q}^{**} \in \mathbb{R}^K$ and any fixed $\mathbf{P} \in \mathbb{R}^K_{>0}$ (with $M_{\mathbf{P}} = \sum_{i=1}^K p_i > 0$) satisfying

$$t^{**}_k := \frac{q^{**}_k}{p_k} \in]t^{sc}_-, t^{sc}_+[\quad \text{for all } k = 1, \ldots, K \qquad \text{(cf. (21))}.$$

In such a setup, recall from Theorem 30(a) that we have obtained for any continuous function $\Phi : \mathbf{\Omega} \mapsto \mathbb{R}$ on a compact set $\mathbf{\Omega} \subset \mathbb{R}^K$ with (6) the assertion

$$\min_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \;=\; -\lim_{n\to\infty} \frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}, \mathbf{Q}^{**}\Big) - \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big) \cdot \mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big) \Big]\Big) \tag{167}$$

by the following construction: in terms of the notations $n_k := \lfloor n \cdot \widetilde{p}_k \rfloor$ ($k \in \{1, \ldots, K-1\}$), $n_K := n - \sum_{k=1}^{K-1} n_k$ (recall (12)),

$$I^{(n)}_1 := \{1, \ldots, n_1\}, \quad I^{(n)}_2 := \{n_1 + 1, \ldots, n_1 + n_2\}, \quad \ldots, \quad I^{(n)}_K := \{\sum_{k=1}^{K-1} n_k + 1, \ldots, n\}, \quad card(I^{(n)}_k) = n_k,$$

$$\widetilde{\mathbf{V}} := \widetilde{\mathbf{V}}_n := \Big(\widetilde{V}_1,\ldots,\widetilde{V}_n\Big) = \Big(\widetilde{V}_1,\ldots,\widetilde{V}_{n_1},\widetilde{V}_{n_1+1},\ldots,\widetilde{V}_{n_1+n_2},\ldots,\widetilde{V}_{\sum_{k=1}^{K-1}n_k+1},\ldots,\widetilde{V}_n\Big) \qquad \text{(cf. (25))},$$

$$\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}} := \Big(\frac{1}{n}\sum_{i\in I_1^{(n)}}\widetilde{V}_i,\ldots,\frac{1}{n}\sum_{i\in I_K^{(n)}}\widetilde{V}_i\Big) \qquad \text{(cf. (27))},$$

we have employed — independently for each $k\in\{1,\ldots,K\}$ — n_k i.i.d. random variables $\widetilde{V}_i$, $i\in I_k^{(n)}$, with common distribution $\mathbb{\Pi}[\widetilde{V}_i\in\cdot\,] = \widetilde{U}_k[\,\cdot\,]$ ($i\in I_k^{(n)}$) given by

$$d\widetilde{U}_k(v) \;:=\; \frac{\exp\big(\tau_k\cdot v\big)}{MGF_{\widetilde{\zeta}}(\tau_k)}\,d\widetilde{\zeta}(v), \qquad \text{(cf. (26))}$$

with $\tau_k := M_{\mathbf{P}}\cdot\varphi'\Big(\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\Big) = M_{\mathbf{P}}\cdot\varphi'\Big(\frac{q_k^{**}}{p_k}\Big)$. Recall (cf. the lines right below (27)) that $\mathbb{E}_{\mathbb{\Pi}}\Big[\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big] = \Big(\frac{\lfloor n\cdot\widetilde{p}_1\rfloor}{n\cdot\widetilde{p}_1}\cdot\widetilde{q}_1^{**},\ldots,\frac{\lfloor n\cdot\widetilde{p}_{K-1}\rfloor}{n\cdot\widetilde{p}_{K-1}}\cdot\widetilde{q}_{K-1}^{**},\frac{n-\sum_{i=1}^{K-1}\lfloor n\cdot\widetilde{p}_i\rfloor}{n\cdot\widetilde{p}_K}\cdot\widetilde{q}_K^{**}\Big)$ and thus $\lim_{n\to\infty}\mathbb{E}_{\mathbb{\Pi}}\Big[\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big] = \widetilde{\mathbf{Q}}^{**}$ or equivalently $\lim_{n\to\infty}\mathbb{E}_{\mathbb{\Pi}}\Big[M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big] = \mathbf{Q}^{**}$. Within such a set-up, we thus obtain for large $n\in\mathbb{N}$ (cf. (167)) the approximation

$$\Phi(\boldsymbol{\Omega}) := \min_{\mathbf{Q}\in\boldsymbol{\Omega}}\Phi(\mathbf{Q}) \;\approx\; -\frac{1}{n}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_{\varphi,\mathbf{P}}^{SBD}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}},\mathbf{Q}^{**}\Big) - \Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot 1_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\;\Big]\Big) \tag{168}$$

and hence for getting an estimator of the minimum value $\Phi(\boldsymbol{\Omega})$ one can estimate the right-hand side of (168). To achieve this, for the rest of this section we assume that n is chosen such that all $n\cdot\frac{p_k}{M_{\mathbf{P}}}$ are integers (and hence, $n=\sum_{k=1}^K n_k$ with $n_k = n\cdot\frac{p_k}{M_{\mathbf{P}}}$) — the remaining case works analogously. With this, we construct the corresponding *naive estimator* for the minimum value $\min_{\mathbf{Q}\in\boldsymbol{\Omega}}\Phi(\mathbf{Q})$ as

$$\widehat{\Phi(\boldsymbol{\Omega})}_{n,L}^{\,naive,2} \;:=\; -\frac{1}{n}\log\frac{1}{L}\sum_{\ell=1}^{L}\exp\Big(n\cdot\Big(D_{\varphi,\mathbf{P}}^{SBD}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}},\mathbf{Q}^{**}\Big) - \Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}\big)\Big)\Big)\cdot 1_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}\big)\,, \tag{169}$$

where we simulate independently L copies $\widetilde{\mathbf{V}}^{(1)},\ldots,\widetilde{\mathbf{V}}^{(L)}$ of the random vector $\widetilde{\mathbf{V}} := \Big(\widetilde{V}_1,\ldots,\widetilde{V}_n\Big)$ constructed by the above-mentioned method, and compute each of $\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(1)}},\ldots,\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(L)}}$ according to (27). Clearly, with the help of the strong law of large numbers we get with $\widehat{\Phi(\boldsymbol{\Omega})}_{n,\infty}^{\,naive,2} := \lim_{L\to\infty}\widehat{\Phi(\boldsymbol{\Omega})}_{n,L}^{\,naive,2}$ the following assertion:

Proposition 77: Let the assumptions of Theorem 30(a) be satisfied. Then one has

$$\lim_{n\to\infty}\widehat{\Phi(\boldsymbol{\Omega})}_{n,\infty}^{\,naive,2} \;=\; \lim_{n\to\infty}\lim_{L\to\infty}\widehat{\Phi(\boldsymbol{\Omega})}_{n,L}^{\,naive,2} \;=\; \Phi(\boldsymbol{\Omega}) \qquad \text{a.s.} \tag{170}$$

As the corresponding *very natural naive BS-estimator of the minimizer-set* $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q}\in\boldsymbol{\Omega}}\Phi(\mathbf{Q})$, we take

$$\widehat{\operatorname*{argmin}_{\mathbf{Q}\in\boldsymbol{\Omega}}\Phi(\mathbf{Q})}_{\;n,L}^{\;naive,2} := \operatorname*{argmin}_{\boldsymbol{\nu}\in\mathcal{V}_{n,L}}\Phi(\boldsymbol{\nu}) \tag{171}$$

where $\mathcal{V}_{n,L} := \{M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}} : \ell\in\{1,\ldots,L\}\,\}\cap\boldsymbol{\Omega}$. In short, we take as minimizer-estimate the (not necessarily unique) element $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}$ which minimizes $\Phi(\cdot)$ amongst all values $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}$ at hand which fall into $\boldsymbol{\Omega}$. We prove that if L and n tend to infinity, then $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}$ concentrates to the above-mentioned set $\mathcal{Q}^*$ of minimizers of $\Phi(\cdot)$ on $\boldsymbol{\Omega}$. Indeed, we show that $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}$ is a proxy minimizer of $\Phi(\cdot)$ on $\boldsymbol{\Omega}$, by proving the following

Proposition 78: There holds

$$\min_{\mathbf{Q}\in\boldsymbol{\Omega}}\Phi(\mathbf{Q}) \;\le\; \Phi\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}\Big) \;\le\; \widehat{\Phi(\boldsymbol{\Omega})}_{n,\infty}^{\,naive,2} \;+\; o_{\mathbb{\Pi}}(1) \tag{172}$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L\to\infty$ and $n\to\infty$ under the distribution $\mathbb{\Pi}$ (recall that $\mathbb{\Pi}[(\widetilde{V}_1,\ldots,\widetilde{V}_n)\in\cdot\,] = \bigotimes_{k=1}^K\widetilde{U}_k^{\otimes n_k}[\,\cdot\,]$)

The proof of Proposition 78 works completely analogous to that of Proposition 72.

Remark 79: In the current set-up of compact $\boldsymbol{\Omega}$ with (6), by taking the special case $\Phi(\mathbf{Q}) := D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**})$ we obtain the naive BS-estimator $\widehat{D_{\varphi,\mathbf{P}}^{SBD}(\boldsymbol{\Omega},\mathbf{Q}^{**})}_{\;n,L}^{\;naive,2}$ of the minimum value $\min_{\mathbf{Q}\in\boldsymbol{\Omega}}D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**})$, as well as the corresponding naive BS-estimator $\widehat{\operatorname{argmin}_{\mathbf{Q}\in\boldsymbol{\Omega}}D_{\varphi,\mathbf{P}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**})}_{\;n,L}^{\;naive,2}$ of the minimizer(s); recall that the fixed pregiven $\mathbf{Q}^{**}$ NEED NOT be in $\boldsymbol{\Omega}$ (e.g. in some application-specific contexts). For the special case $\mathbf{P} := \mathbf{1} = (1,\ldots,1)$, this leads to the corresponding estimators for the minimum and the minimizer(s) of the separable ordinary Bregman distance $\mathbf{Q}\mapsto D_{\varphi}^{OBD}(\mathbf{Q},\mathbf{Q}^{**}) = D_{\varphi,\mathbf{1}}^{SBD}(\mathbf{Q},\mathbf{Q}^{**})$.

Proposition 80: In the above set-up, one has

$$\lim_{n\to\infty}\lim_{L\to\infty}\Phi\Big(\widehat{\operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})}_{n,L}^{naive,2}\Big) \;=\; \Phi(\mathbf{\Omega}) \qquad \text{a.s.},$$

and thus the quantity $\Phi\Big(\widehat{\operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})}_{n,L}^{naive,2}\Big)$ is a natural alternative to the estimate $\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,2}$ given in (169).

E. Naive estimators of max and argmax — Base-Divergence-Method 2, compact case

In the set-up of compact $\mathbf{\Omega}$ with (6), we can treat the maximizing problem completely analogously to the method in the previous Subsection XII-D. Indeed, by employing Theorem 34(a) instead of Theorem 30(a) we construct the naive BS-estimator

$$\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,2} \;:=\; \frac{1}{n}\log\frac{1}{L}\sum_{\ell=1}^{L}\exp\Big(n\cdot\Big(D_{\varphi,\mathbf{P}}^{SBD}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}},\mathbf{Q}^{**}\Big)+\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}\big)\Big)\Big)\cdot\mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}\big)\,, \tag{173}$$

of the maximum value $\Phi(\mathbf{\Omega}):=\max_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})$; for this, we get

$$\lim_{n\to\infty}\widehat{\Phi(\mathbf{\Omega})}_{n,\infty}^{naive,2} \;=\; \lim_{n\to\infty}\lim_{L\to\infty}\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,2} \;=\; \Phi(\mathbf{\Omega}) \qquad \text{a.s.}$$

instead of (170) (notice the different definition of the involved quantities). As the corresponding very natural naive BS-estimator of the maximizer $\operatorname{argmax}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})$ we take

$$\widehat{\operatorname*{argmax}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})}_{n,L}^{naive,2} := \operatorname*{argmax}_{\boldsymbol{\nu}\in\mathcal{V}_{n,L}}\Phi(\boldsymbol{\nu})$$

instead of (171); in short, we take any $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}$ which maximizes $\Phi(\cdot)$ amongst all values $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}$ at hand which fall into $\mathbf{\Omega}$. For this, we obtain — instead of (172) — the assertion

$$\max_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q}) \;\geq\; \Phi\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}\Big) \;\geq\; \widehat{\Phi(\mathbf{\Omega})}_{n,\infty}^{naive,2} \;-\; o_{\mathbb{\Pi}}(1)$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L\to\infty$ and $n\to\infty$ under the distribution $\mathbb{\Pi}$. Hence, as L and n tend to infinity, $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}$ concentrates to the set of maximizers of $\Phi(\cdot)$ on $\mathbf{\Omega}$.

Proposition 81: In the above set-up, one has

$$\lim_{n\to\infty}\lim_{L\to\infty}\Phi\Big(\widehat{\operatorname*{argmax}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})}_{n,L}^{naive,2}\Big) \;=\; \Phi(\mathbf{\Omega}) \qquad \text{a.s.},$$

and thus the quantity $\Phi\Big(\widehat{\operatorname{argmax}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})}_{n,L}^{naive,2}\Big)$ is a natural alternative to the estimate $\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,2}$ given in (173).

F. Naive estimators — Base-Divergence-Method 2, non-compact case

We can proceed analogously to the Base-Divergence-Method-1-treating Subsection XII-C (e.g. by employing Theorem 30(b)).

G. Improved/Speed-up estimators of min and argmin — compact case

G.1) The underlying method

For the sake of brevity, we deal here only with improving/speeding-up the naive estimators given in Subsection XII-A which concerns about estimating the min and argmin of $\Phi(\cdot)$ on a compact constraint set $\mathbf{\Omega}$ with (6), via Base-Divergence-Method 1, by choosing the $\varphi-$divergence $D_\varphi(\cdot,\mathbf{P})$ as the so-called *base divergence*. To start with, recall that

$$\min_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q}) \;=\; -\lim_{n\to\infty}\frac{1}{n}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big)\Big)\cdot\mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big]\Big) \qquad \text{(cf. (160))},$$

which for large $n\in\mathbb{N}$ leads to the approximation

$$\Phi(\mathbf{\Omega}):=\min_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q}) \;\approx\; -\frac{1}{n}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big)\Big)\cdot\mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big]\Big) \qquad \text{(cf. (161))}.$$

Notice that by construction of $\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}$ (cf. (13)) one has

$$\mathbb{E}_{\mathbb{\Pi}}\Big[M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\Big]=\left(M_{\mathbf{P}}\cdot\frac{\lfloor n\cdot\widetilde{p}_1\rfloor}{n},\ldots,M_{\mathbf{P}}\cdot\frac{\lfloor n\cdot\widetilde{p}_{K-1}\rfloor}{n},M_{\mathbf{P}}\cdot\frac{n-\sum_{i=1}^{K-1}\lfloor n\cdot\widetilde{p}_i\rfloor}{n}\right)$$

and $\lim_{n\to\infty}\mathbb{E}_{\mathbb{\Pi}}\Big[M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\Big]=\mathbf{P}$. Since typically $\mathbf{P}\notin\mathbf{\Omega}$, the method of Subsection XII-A can be time costly (relative to the employed computer-power), since the *runtime* which is required in order to obtain some accuracy while estimating — via $\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,1}$ given in (162) — the expectation (cf. (161))

$$\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_{\varphi}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big)\Big)\cdot 1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big]$$

depends on the hit rate of the set $\mathbf{\Omega}$ by the replicates/copies $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}$ of the random vector $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}$. We thus may substitute $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}$ by the replicates/copies $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}$ of some new random vector $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ whose visits in $\mathbf{\Omega}$ will be much more frequent. For instance, we may have the goal that the new replication mean $\mathbb{E}_{\mathbb{\Pi}}\Big[M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}\Big]$ should be equal to some point $\mathbf{Q}^*$ which lies in the interior $int(\mathbf{\Omega})$ of $\mathbf{\Omega}$; for pregiven $\mathbf{P}\in\mathbb{R}^K_{>0}$, such a vector $\mathbf{Q}^*$ may be either pregiven (e.g. by the nature of the application context) or it may be simulatively achieved by e.g. proxy method 1 or proxy method 2 of Subsection X-A of Broniatowski & Stummer [1].

One *comfortable* way to achieve from $\mathbf{Q}^*\in int(\mathbf{\Omega})$ the above-mentioned desired construction $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}$, is to employ (25), (26) and (27) with the special choice $\mathbf{Q}^{**}:=\mathbf{Q}^*$ (under the assumption (21)), and accordingly obtain from Theorem 30(a) the limit assertion

$$\min_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})\;=\;-\lim_{n\to\infty}\frac{1}{n}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}},\mathbf{Q}^*\Big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot 1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big]\Big)\,. \tag{174}$$

Notice that (cf. the lines right below (27) with $\mathbf{Q}^{**}:=\mathbf{Q}^*$)

$$\begin{aligned}\mathbb{E}_{\mathbb{\Pi}}\Big[M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big] &= \left(M_{\mathbf{P}}\cdot\frac{\lfloor n\cdot\widetilde{p}_1\rfloor}{n\cdot\widetilde{p}_1}\cdot\widetilde{q}_1^*,\ldots,M_{\mathbf{P}}\cdot\frac{\lfloor n\cdot\widetilde{p}_{K-1}\rfloor}{n\cdot\widetilde{p}_{K-1}}\cdot\widetilde{q}_{K-1}^*,M_{\mathbf{P}}\cdot\frac{n-\sum_{i=1}^{K-1}\lfloor n\cdot\widetilde{p}_i\rfloor}{n\cdot\widetilde{p}_K}\cdot\widetilde{q}_K^*\right)\\ &= \left(\frac{\lfloor n\cdot\widetilde{p}_1\rfloor}{n\cdot\widetilde{p}_1}\cdot q_1^*,\ldots,\frac{\lfloor n\cdot\widetilde{p}_{K-1}\rfloor}{n\cdot\widetilde{p}_{K-1}}\cdot q_{K-1}^*,\frac{n-\sum_{i=1}^{K-1}\lfloor n\cdot\widetilde{p}_i\rfloor}{n\cdot\widetilde{p}_K}\cdot q_K^*\right)\end{aligned}$$

which lies in $int(\mathbf{\Omega})$ for large enough n; moreover, $\lim_{n\to\infty}\mathbb{E}_{\mathbb{\Pi}}\Big[M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big]=\mathbf{Q}^*$. Within such a set-up, we thus obtain for large $n\in\mathbb{N}$ the approximation

$$\Phi(\mathbf{\Omega}):=\min_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})\;\approx\;-\frac{1}{n}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}},\mathbf{Q}^*\Big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot 1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big]\Big) \tag{175}$$

and hence for getting an estimator of the minimum value $\Phi(\mathbf{\Omega})$ one can estimate the involved expectation

$$\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}},\mathbf{Q}^*\Big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot 1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big]$$

with higher hit rate than above. Indeed, we proceed analogously to Subsection XII-D (with $\mathbf{Q}^{**}:=\mathbf{Q}^*$), assume that n is chosen such that all $n\cdot\frac{p_k}{M_{\mathbf{P}}}$ are integers (and hence, $n=\sum_{k=1}^K n_k$ with $n_k=n\cdot\frac{p_k}{M_{\mathbf{P}}}$; the remaining case works analogously) and construct the corresponding *speed-up BS-estimator* for the minimum value $\min_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})$ as

$$\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{speedup,1}\;:=\;-\frac{1}{n}\log\frac{1}{L}\sum_{\ell=1}^{L}\exp\Big(n\cdot\Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}},\mathbf{Q}^*\Big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}\big)\Big)\Big)\cdot 1_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}\big)\,; \tag{176}$$

notice the difference to the *naive estimator* $\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{naive,2}$ for the minimum value $\min_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})$ given in (169): due to the different nature of $\mathbf{Q}^*$ and $\mathbf{Q}^{**}$, the construction of $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}$ in (176) typically hits $\mathbf{\Omega}$ *more often* than the construction of $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}$ in (169) does. From the corresponding special case of Subsection XII-D we obtain with $\widehat{\Phi(\mathbf{\Omega})}_{n,\infty}^{speedup,1}:=\lim_{L\to\infty}\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{speedup,1}$ the following assertion:

Proposition 82: Let the assumptions of Theorem 30(a) be satisfied (with $\mathbf{Q}^*$ instead of $\mathbf{Q}^{**}$). Then one has

$$\lim_{n\to\infty}\widehat{\Phi(\mathbf{\Omega})}_{n,\infty}^{speedup,1}\;=\;\lim_{n\to\infty}\lim_{L\to\infty}\widehat{\Phi(\mathbf{\Omega})}_{n,L}^{speedup,1}\;=\;\Phi(\mathbf{\Omega})\qquad\text{a.s.}$$

As the corresponding *very natural speed-up BS-estimator of the minimizer-set* $\mathcal{Q}^*:=\operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})$, we take

$$\widehat{\operatorname*{argmin}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})}_{n,L}^{speedup,1}:=\operatorname*{argmin}_{\boldsymbol{\nu}\in\mathcal{V}_{n,L}}\Phi(\boldsymbol{\nu}) \tag{177}$$

where $\mathcal{V}_{n,L}:=\{M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}:\ell\in\{1,\ldots,L\}\}\cap\mathbf{\Omega}$. In short, we take as minimizer-estimate the (not necessarily unique) element $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}$ which minimizes $\Phi(\cdot)$ amongst all values $M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{(\ell)}}$ at hand which fall into $\mathbf{\Omega}$. From the corresponding special

case of Subsection XII-D we derive that — as L and n tend to infinity — $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}$ concentrates to the above-mentioned set $\mathcal{Q}^*$ of minimizers of $\Phi(\cdot)$ on $\boldsymbol{\Omega}$; indeed, $M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}$ is a proxy minimizer of $\Phi(\cdot)$ on $\boldsymbol{\Omega}$, due to the following

Proposition 83: There holds

$$\min_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q}) \;\leq\; \Phi\Big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}^{L,*}}\Big) \;\leq\; \widehat{\Phi(\boldsymbol{\Omega})}_{n,\infty}^{speedup,1} \;+\; o_{\mathbb{\Pi}}(1)$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L \to \infty$ and $n \to \infty$ under the distribution $\mathbb{\Pi}$ (recall that $\mathbb{\Pi}[(\widetilde{V}_1, \ldots, \widetilde{V}_n) \in \cdot\,] = \bigotimes_{k=1}^{K} \widetilde{U}_k^{\otimes n_k}[\,\cdot\,]$).

Proposition 84: In the above set-up, one has

$$\lim_{n\to\infty} \lim_{L\to\infty} \Phi\Big(\widehat{\operatorname{argmin}_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q})}_{n,L}^{speedup,1}\Big) \;=\; \Phi(\boldsymbol{\Omega}) \qquad \text{a.s.},$$

and thus the quantity $\Phi\Big(\widehat{\operatorname{argmin}_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q})}_{n,L}^{speedup,1}\Big)$ is a natural alternative to the estimate $\widehat{\Phi(\boldsymbol{\Omega})}_{n,L}^{speedup,1}$ given in (176).

Analogously to the above investigations, by applying the results of Subsection XII-E to the special choice $\mathbf{Q}^{**} := \mathbf{Q}^* \in int(\boldsymbol{\Omega})$ we can also derive improved/speed-up estimators $\widehat{\Phi(\boldsymbol{\Omega})}_{n,L}^{speedup,1}$ and $\Phi\Big(\widehat{\operatorname{argmax}_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q})}_{n,L}^{speedup,1}\Big)$ for $\Phi(\boldsymbol{\Omega}) := \max_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q})$, as well as $\widehat{\operatorname{argmax}_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q})}_{n,L}^{speedup,1}$ for $\operatorname{argmax}_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q})$. For the sake of brevity, the details are omitted.

G.2) Connections to importance sampling

Usually, the importance-sampling paradigm basically aims for the more accurate simulative estimation of an (say) integral $\int_{\breve{\boldsymbol{\Omega}}} \breve{g}(\breve{\mathbf{y}})\, \mathrm{d}\breve{\mathfrak{l}}(\breve{\mathbf{y}})$ represented as

$$\int\limits_{\breve{\boldsymbol{\Omega}}} \breve{g}(\breve{\mathbf{y}})\, \mathrm{d}\breve{\mathfrak{l}}(\breve{\mathbf{y}}) = \int\limits_{\mathbb{R}^K} \mathbb{1}_{\breve{\boldsymbol{\Omega}}}(\breve{\mathbf{y}}) \cdot \breve{g}(\breve{\mathbf{y}})\, \mathrm{d}\breve{\mathfrak{l}}(\breve{\mathbf{y}}) = \mathbb{E}_{\mathbb{\Pi}}\Big[\, \breve{g}(\breve{\mathbf{Y}}) \cdot \mathbb{1}_{\breve{\boldsymbol{\Omega}}}(\breve{\mathbf{Y}}) \,\Big]$$

(with distribution $\mathbb{\Pi}[\breve{\mathbf{Y}} \in \cdot\,] = \breve{\mathfrak{l}}[\,\cdot\,]$), by simulatively estimating the rewritten integral

$$\int\limits_{\breve{\boldsymbol{\Omega}}} \breve{g}(\breve{\mathbf{y}})\, \mathrm{d}\breve{\mathfrak{l}}(\breve{\mathbf{y}}) = \int\limits_{\mathbb{R}^K} \mathbb{1}_{\breve{\boldsymbol{\Omega}}}(\breve{\mathbf{z}}) \cdot \breve{g}(\breve{\mathbf{z}}) \cdot \frac{\mathrm{d}\breve{\mathfrak{l}}}{\mathrm{d}\breve{\mathbb{S}}}(\breve{\mathbf{z}})\, \mathrm{d}\breve{\mathbb{S}}(\breve{\mathbf{z}}) = \mathbb{E}_{\mathbb{\Pi}}\Big[\, \frac{\mathrm{d}\breve{\mathfrak{l}}}{\mathrm{d}\breve{\mathbb{S}}}(\breve{\mathbf{Z}}) \cdot\; \breve{g}(\breve{\mathbf{Z}}) \cdot \mathbb{1}_{\breve{\boldsymbol{\Omega}}}(\breve{\mathbf{Z}}) \,\Big]$$

(with distribution $\mathbb{\Pi}[\breve{\mathbf{Z}} \in \cdot\,] = \breve{\mathbb{S}}[\,\cdot\,]$), where $\breve{\mathbb{S}}$ and thus $\breve{\mathbf{Z}}$ are constructed in a (goal-depending) appropriate manner. To establish some connection with our *BS-minimization* context, recall that in the above Subsection XII-A we have employed $\mathbb{\Pi}[\widetilde{\mathbf{W}} \in \cdot\,] = \mathbb{\Pi}[(\widetilde{W}_1, \ldots, \widetilde{W}_n) \in \cdot\,] = \widetilde{\mathfrak{l}}^{\otimes n}[\,\cdot\,]$ and obtained for large $n \in \mathbb{N}$ the approximation

$$\begin{aligned}
\Phi(\boldsymbol{\Omega}) := \min_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q}) \;&\approx\; -\frac{1}{n} \log\Big(\; \mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \Big(D_\varphi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}, \mathbf{P}\big) - \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big)\Big) \cdot \mathbb{1}_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big) \,\Big]\Big) \qquad \text{(cf. 161)} \\
&=\; -\frac{1}{n} \log\left(\int\limits_{\mathbb{R}^K} \mathbb{1}_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\mathbf{y}}\big) \cdot \exp\Big(n \cdot \Big(D_\varphi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\mathbf{y}}, \mathbf{P}\big) - \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\mathbf{y}}\big)\Big)\Big)\, \mathrm{d}\widetilde{\mathfrak{l}}^{\otimes n}(\mathbf{y}) \right) \qquad (178)
\end{aligned}$$

where $\boldsymbol{\xi}_n^{\mathbf{y}} := \Big(\frac{1}{n}\sum_{i\in I_1^{(n)}} y_i, \ldots, \frac{1}{n}\sum_{i\in I_K^{(n)}} y_i\Big)$ (analogously to (13)). We can apply the above-mentioned general importance-sampling procedure to the *very special case* (amongst many other alternative possibilities)

$$\breve{\boldsymbol{\Omega}} := \boldsymbol{\Omega}, \quad \breve{\mathbf{Y}} := M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}, \quad \breve{g}(\breve{\mathbf{Y}}) := \exp\Big(n \cdot \Big(D_\varphi\big(\breve{\mathbf{Y}}, \mathbf{P}\big) - \Phi\big(\breve{\mathbf{Y}}\big)\Big)\Big), \quad \breve{\mathfrak{l}}[\,\cdot\,] := \mathbb{\Pi}[\breve{\mathbf{Y}} \in \cdot\,], \quad \breve{\mathbf{Z}} := M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}, \quad \breve{\mathbb{S}}[\,\cdot\,] := \mathbb{\Pi}[\breve{\mathbf{Z}} \in \cdot\,],$$

where we employ — with the same choice $\mathbf{Q}^{**} := \mathbf{Q}^* \in int(\boldsymbol{\Omega})$ as in the previous Subsection XII-G.1) — the random vector $\widetilde{\mathbf{V}}$ of (25) having distribution $\mathbb{\Pi}[\widetilde{\mathbf{V}} \in \cdot\,] = \mathbb{\Pi}[(\widetilde{V}_1, \ldots, \widetilde{V}_n) \in \cdot\,] = \bigotimes_{k=1}^{K} \widetilde{U}_k^{\otimes n_k}[\,\cdot\,]$, and $\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ of (27); accordingly, by rewriting (161) and (178) we end up with the approximation

$$\begin{aligned}
&\Phi(\boldsymbol{\Omega}) := \min_{\mathbf{Q}\in\boldsymbol{\Omega}} \Phi(\mathbf{Q}) \\
&\approx\; -\frac{1}{n} \log\left(\mathbb{E}_{\mathbb{\Pi}}\Big[\mathbb{1}_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big) \cdot \exp\Big(n \cdot \Big(D_\varphi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}, \mathbf{P}\big) - \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big) \cdot \frac{\mathrm{d}\breve{\mathfrak{l}}}{\mathrm{d}\breve{\mathbb{S}}}\Big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big) \,\Big] \right) \qquad (179) \\
&=\; -\frac{1}{n} \log\left(\int\limits_{\mathbb{R}^K} \mathbb{1}_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\mathbf{z}}\big) \cdot \exp\Big(n \cdot \Big(D_\varphi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\mathbf{z}}, \mathbf{P}\big) - \Phi\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\mathbf{z}}\big)\Big)\Big) \cdot \frac{\mathrm{d}\breve{\mathfrak{l}}}{\mathrm{d}\breve{\mathbb{S}}}\Big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\mathbf{z}}\Big)\; \mathrm{d}\bigotimes_{k=1}^{K} \widetilde{U}_k^{\otimes n_k}(z) \right).
\end{aligned}$$

Notice that the above-mentioned underlying importance-sampling goal is in our current context the improved hit rate of the set $\mathbf{\Omega}$ (and this is achieved by construction). At this point, one may wonder about the connection between the two alternative approximations (179) and

$$\Phi(\mathbf{\Omega}) := \min_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q}) \;\approx\; -\frac{1}{n}\log\Big(\mathbb{E}_{\mathbb{I}\!\mathbb{P}}\Big[\exp\Big(n\cdot\Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n,\mathbf{Q}^*\Big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n\big)\Big)\Big)\cdot \mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n\big)\,\Big]\Big) \text{ (cf. (175))}.$$

The corresponding answer is that they *essentially coincide*:

Proposition 85: Under the assumptions of the Subsections XII-A and XII-G.1), if $n\in\mathbb{N}$ is fixed and $\frac{n_k}{n}=\widetilde{p}_k$ for all $k=1,\ldots,K$, then the term in formula (179) is equal to the term in formula (175).

Remark 86: Concerning the above Proposition 85, recall that $\lim_{n\to\infty}\frac{n_k}{n}=\widetilde{p}_k$ (cf. (12)); moreover, if all $\widetilde{p}_k$ ($k=1,\ldots,K$) are rational numbers in $]0,1[$ with $\sum_{k=1}^K\widetilde{p}_k=1$ and N is the (always existing) smallest integer such that all $N\cdot\widetilde{p}_k$ ($k=1,\ldots,K$) are integers (i.e. $\in\mathbb{N}$), then for any multiple $n=\ell\cdot N$ ($\ell\in\mathbb{N}$) one gets that all $n_k=n\cdot\widetilde{p}_k$ are integers and that $card(I^{(n)}_K)=n_K$.

The proof of Proposition 85 will be given in Appendix A.

For the corresponding *importance-sampling estimate* of the approximation (179), we take

$$\begin{aligned}\widehat{\Phi(\mathbf{\Omega})}^{IS,1}_{n,L} &:= -\frac{1}{n}\log\frac{1}{L}\sum_{\ell=1}^{L}\mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(\ell)}}_n\big)\cdot\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(\ell)}}_n,\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(\ell)}}_n\big)\Big)\Big)\cdot\frac{\mathrm{d}\breve{\mathfrak{C}}}{\mathrm{d}\breve{\mathbb{S}}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(\ell)}}_n\Big)\\ &= -\frac{1}{n}\log\frac{1}{L}\sum_{\ell=1}^{L}\mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(\ell)}}_n\big)\cdot\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(\ell)}}_n,\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(\ell)}}_n\big)\Big)\Big)\cdot\prod_{k=1}^{K}IS_k(\widetilde{\mathbf{V}}^{(\ell)}_k) \qquad (180)\end{aligned}$$

where we simulate independently L copies $\widetilde{\mathbf{V}}^{(1)},\ldots,\widetilde{\mathbf{V}}^{(L)}$ of the random vector $\widetilde{\mathbf{V}}:=\big(\widetilde{V}_1,\ldots,\widetilde{V}_n\big)$ constructed by the above-mentioned method, and compute each of $\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(1)}}_n,\ldots,\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(L)}}_n$ according to (27); notice that in (180) we have employed (the $k-$block importance-sampling factor) $\widetilde{IS}_k(z_1,\ldots,z_{n_k}) := \exp\Big(n_k\cdot\Lambda_{\widetilde{\zeta}}(\tau_k)-\tau_k\cdot\sum_{i=1}^{n_k}z_i\Big)$ (cf. Subsection X-A of Broniatowski & Stummer [1]).

As can be seen straightforwardly from the above calculations, due to (12) one gets even the limit relation

$$\begin{aligned}\Phi(\mathbf{\Omega}) &:= \min_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})\\ &= -\lim_{n\to\infty}\frac{1}{n}\log\Bigg(\mathbb{E}_{\mathbb{I}\!\mathbb{P}}\Big[\mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n\big)\cdot\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n,\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n\big)\Big)\Big)\cdot\frac{\mathrm{d}\breve{\mathfrak{C}}}{\mathrm{d}\breve{\mathbb{S}}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n\Big)\,\Big]\Bigg) \qquad (181)\\ &= -\lim_{n\to\infty}\frac{1}{n}\log\Big(\mathbb{E}_{\mathbb{I}\!\mathbb{P}}\Big[\mathrm{l}_{\mathbf{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n\big)\cdot\exp\Big(n\cdot\Big(D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n,\mathbf{Q}^*\Big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}}_n\big)\Big)\Big)\Big]\Big)\,. \qquad \text{(cf. (174))}\end{aligned}$$

Summing up things, for an *improved* estimation of the *minimum value* $\min_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})$ the two above-described approaches are *asymptotically* equal. However, notice that for the — quite comfortable — essential convergence proofs of the Propositions 82 and 83 (including the treatment of estimates of *minimizers*) we rely heavily on the use of our newly developed Theorems 16 and 30, which put us in position to further proceed completely analogously to the case of naive estimators.

To end up this section, let us mention that as the corresponding *very natural importance-sampling BS-estimator of the minimizer-set* $\mathcal{Q}^*:=\mathrm{argmin}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})$, one can take

$$\widehat{\operatorname*{argmin}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})}^{IS,1}_{n,L} := \operatorname*{argmin}_{\boldsymbol{\nu}\in\mathcal{V}_{n,L}}\Phi(\boldsymbol{\nu})$$

where $\mathcal{V}_{n,L}:=\{M_{\mathbf{P}}\cdot\boldsymbol{\xi}^{\widetilde{\mathbf{V}}^{(\ell)}}_n:\ell\in\{1,\ldots,L\}\}\cap\mathbf{\Omega}$, for which we can prove — due to the equality of (181) and (174) — that the limit behaviour of $\widehat{\mathrm{argmin}_{\mathbf{Q}\in\mathbf{\Omega}}\Phi(\mathbf{Q})}^{speedup}_{n,L}$ carries over. Hence, also the above-mentioned Remark 73 carries over.

XIII. Bare-Simulation Estimators For General Deterministic Divergence-Optimization-Problems with Constant-Component-Sum Side Constraint

Recall that we are interested in the constrained optimization of the *continuous* distance-connected functions $\mathbf{\Omega} \ni \mathbf{Q} \mapsto \Phi(\mathbf{Q})$ in the above-mentioned cases (D1) to (D8) of Subsection V-A, and beyond (e.g. $\Phi(\cdot) := D_{\breve{\varphi}}(\cdot, \breve{\mathbf{P}})$ may be a $\breve{\varphi}-$divergence with pregiven $\breve{\mathbf{P}}$ and $\breve{\varphi}$, cf. Remark 23(iii)). Contrary to the previous Section XII, we now involve constraint sets $A \cdot \mathbf{\Omega}$ where $A \in]0, \infty[$ is a pregiven constant and $\mathbf{\Omega} \in \mathbb{S}^K$ (respectively $\mathbb{S}^K_{>0}$); this means that we employ the side constraints $q_k \geq 0$ (respectively, $q_k > 0$) for all $k = 1, \ldots, K$ as well as $\sum_{k=1}^K q_k = A$. Clearly, one has $int\,(A \cdot \mathbf{\Omega}) = \emptyset$ in the *full* topology (cf. Remark 8(b)), and thus we can not apply the estimator-results of Section XII but we need some extra refinements. Those will be worked out in the following, for the context that $A \cdot \mathbf{\Omega}$ satisfies the regularity properties (67) — in the *relative* topology (!!) — such that the function $\Phi(\cdot)$ possesses a (not necessarily unique) minimizer; for this, we construct *naive estimators* as well as *speed-up estimators* (approximations) of the *minimum value* $\inf_{\mathbf{Q} \in A \cdot \mathbf{\Omega}} \Phi(\mathbf{Q}) = \min_{\mathbf{Q} \in A \cdot \mathbf{\Omega}} \Phi(\mathbf{Q})$ and of the corresponding (set of) *minimizers* $\arg\inf_{\mathbf{Q} \in A \cdot \mathbf{\Omega}} \Phi(\mathbf{Q}) = \arg\min_{\mathbf{Q} \in A \cdot \mathbf{\Omega}} \Phi(\mathbf{Q})$. The case of maximum values and maximizers will be treated analogously. We mainly focus on compact sets $A \cdot \mathbf{\Omega}$ (in the relative toplogy) but also discuss some relaxations thereof.

A. Naive estimators of min and argmin — Base-Divergence-Method 1, compact case

Recall that from Theorem 52(a) we obtain for any continuous function $\Phi : A \cdot \mathbf{\Omega} \mapsto \mathbb{R}$ on a compact set $A \cdot \mathbf{\Omega} \subset \widetilde{\mathcal{M}}_\gamma$ with (67) the assertion

$$\min_{\mathbf{Q} \in A \cdot \mathbf{\Omega}} \Phi(\mathbf{Q}) \;=\; -\lim_{n \to \infty} \frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \Big(F_{\gamma, \widetilde{c}, A}\Big(D_{\widetilde{c} \cdot \varphi_\gamma}(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}}, \mathbb{P})\Big) - \Phi\big(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}}\big)\Big)\Big) \cdot 1_{\mathbf{\Omega}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}}\big) \Big]\Big) \;, \tag{182}$$

where $M_{\mathbf{P}} = \sum_{i=1}^K p_i > 0$, $\gamma \in \mathbb{R} \setminus]1, 2[$, $\widetilde{\mathcal{M}}_\gamma$ as defined in (S1), and

$$\boldsymbol{\xi}_n^{w\mathbf{W}} \;:=\; \begin{cases} \left(\frac{\sum_{i \in I_1^{(n)}} W_i}{\sum_{k=1}^K \sum_{i \in I_k^{(n)}} W_i}, \ldots, \frac{\sum_{i \in I_K^{(n)}} W_i}{\sum_{k=1}^K \sum_{i \in I_k^{(n)}} W_i} \right), & \text{if } \sum_{j=1}^n W_j \neq 0, \\ (\infty, \ldots, \infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^n W_j = 0, \end{cases} \qquad \text{(cf. (70))}$$

is constructed from a sequence $W := (W_i)_{i \in \mathbb{N}}$ of random variables, where the W_i's are i.i.d. copies of the random variable W whose distribution is $\mathbb{\Pi}[W \in \cdot\,] = \mathbb{\zeta}[\,\cdot\,]$ being attributed to the power divergence generator $\varphi := \widetilde{c} \cdot \varphi_\gamma$ by the representability (69). Within such a set-up, we thus obtain for large $n \in \mathbb{N}$ (cf. (182)) the approximation

$$\Phi(A \cdot \mathbf{\Omega}) := \min_{\mathbf{Q} \in A \cdot \mathbf{\Omega}} \Phi(\mathbf{Q}) \;\approx\; -\frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \Big(F_{\gamma, \widetilde{c}, A}\Big(D_{\widetilde{c} \cdot \varphi_\gamma}(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}}, \mathbb{P})\Big) - \Phi\big(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}}\big)\Big)\Big) \cdot 1_{\mathbf{\Omega}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}}\big) \Big]\Big) \;, \tag{183}$$

and hence for getting an estimator of the minimum value $\Phi(A \cdot \mathbf{\Omega})$ one can estimate the right-hand side of (183). To achieve this, for the rest of this section we assume that n is chosen such that all $n \cdot \frac{p_k}{M_{\mathbf{P}}}$ are integers (and hence, $n = \sum_{k=1}^K n_k$ with $n_k = n \cdot \frac{p_k}{M_{\mathbf{P}}}$) — the remaining case works analogously. As above, a corresponding *naive (crude) estimator* can be constructed by

$$\widehat{\Phi(A \cdot \mathbf{\Omega})}_{n,L}^{naive,1} \;:=\; -\frac{1}{n} \log \frac{1}{L} \sum_{\ell=1}^L \exp\Big(n \cdot \Big(F_{\gamma, \widetilde{c}, A}\Big(D_{\widetilde{c} \cdot \varphi_\gamma}(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}, \mathbb{P})\Big) - \Phi\big(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}\big)\Big)\Big) \cdot 1_{\mathbf{\Omega}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}\big) \,, \tag{184}$$

where we simulate independently L copies $\mathbf{W}^{(1)}, \ldots, \mathbf{W}^{(L)}$ of the vector $\mathbf{W} := (W_1, \ldots, W_n)$ with independent entries under $\mathbb{\zeta}$, and compute each of $\boldsymbol{\xi}_n^{w\mathbf{W}^{(1)}}, \ldots, \boldsymbol{\xi}_n^{w\mathbf{W}^{(L)}}$ according to (70). Clearly, with the help of the strong law of large numbers we get with $\widehat{\Phi(A \cdot \mathbf{\Omega})}_{n,\infty}^{naive,1} := \lim_{L \to \infty} \widehat{\Phi(A \cdot \mathbf{\Omega})}_{n,L}^{naive,1}$ the following assertion:

Proposition 87:

$$\lim_{n \to \infty} \widehat{\Phi(A \cdot \mathbf{\Omega})}_{n,\infty}^{naive,1} \;=\; \lim_{n \to \infty} \lim_{L \to \infty} \widehat{\Phi(A \cdot \mathbf{\Omega})}_{n,L}^{naive,1} \;=\; \Phi(A \cdot \mathbf{\Omega}) \qquad \text{a.s.} \tag{185}$$

As the corresponding *very natural naive (crude) estimator* of the minimizer-set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q} \in A \cdot \mathbf{\Omega}} \Phi(\mathbf{Q})$, we take

$$\widehat{\operatorname*{argmin}_{\mathbf{Q} \in A \cdot \mathbf{\Omega}} \Phi(\mathbf{Q})}_{n,L}^{naive,1} := \operatorname*{argmin}_{\boldsymbol{\nu} \in \mathcal{W}_{n,L}} \Phi(\boldsymbol{\nu}) \tag{186}$$

where $\mathcal{W}_{n,L} := \{A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}} : \ell \in \{1, \ldots, L\}\} \cap A \cdot \mathbf{\Omega}$. In other words, as a corresponding *naive (crude) estimator* of the (not necessarily unique) element $\mathbf{Q}^*$ of the minimizer-set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q} \in A \cdot \mathbf{\Omega}} \Phi(\mathbf{Q})$, we take the (not necessarily unique) element $A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}}$ of the set $\{A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}} : \ell \in \{1, \ldots, L\}\}$ such that $A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}} \in A \cdot \mathbf{\Omega}$ and

$$\Phi(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}}) \;\leq \Phi(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}) \qquad \text{for all } \ell = 1, \ldots, L \text{ for which } A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}} \text{ belongs to } A \cdot \mathbf{\Omega}.$$

In short, $A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}}$ minimizes $\Phi(\cdot)$ amongst all values $A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}$ at hand which fall into $A \cdot \boldsymbol{\mathfrak{Q}}$. For large enough $n \in \mathbb{N}$ and $L \in \mathbb{N}$, such $A \cdot \boldsymbol{\xi}_n^{w\widetilde{\mathbf{W}}^{L,*}}$ exists since $A \cdot \boldsymbol{\mathfrak{Q}}$ has non-void interior in the relative topology, by assumption (67). We prove that if L and n tend to infinity, then $A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}}$ concentrates to the above-mentioned set $\mathcal{Q}^*$ of minimizers of $\Phi(\cdot)$ on $A \cdot \boldsymbol{\mathfrak{Q}}$. As usual in similar procedures, L is assumed to be large enough in order to justify some approximation for fixed n, typically the substitution of empirical means by expectations, since L is at disposal.
Next we derive that $A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}}$ is a proxy minimizer of $\Phi(\cdot)$ on $A \cdot \boldsymbol{\mathfrak{Q}}$, manifested by the following

Proposition 88: There holds

$$\min_{\mathbf{Q} \in A \cdot \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q}) \;\leq\; \Phi\left(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}}\right) \;\leq\; \widehat{\Phi(A \cdot \boldsymbol{\mathfrak{Q}})}_{n,\infty}^{naive,1} \;+\; o_{\mathbb{\Pi}}(1) \tag{187}$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L \to \infty$ and $n \to \infty$ under the distribution $\mathbb{\Pi}$ (recall that $\mathbb{\Pi}[(W_1, \ldots, W_n) \in \cdot\,] = \mathfrak{Q}^{\otimes n}[\,\cdot\,]$).

The proof of Proposition 88 works analogously to the proof of Proposition 72, by replacing Theorem 22(a) with Theorem 52(a).

Remark 89: In the current set-up of compact $A \cdot \boldsymbol{\mathfrak{Q}}$ with (67), by taking the special case $\Phi(\mathbf{Q}) := D_\varphi(\mathbf{Q}, \mathbb{P})$ we obtain the naive BS-estimator $\widehat{D_\varphi(A \cdot \boldsymbol{\mathfrak{Q}}, \mathbb{P})}_{n,L}^{naive,1}$ of the minimum value $\min_{\mathbf{Q} \in A \cdot \boldsymbol{\mathfrak{Q}}} D_\varphi(\mathbf{Q}, \mathbb{P})$.

Proposition 90: In the above set-up, one has

$$\lim_{n\to\infty} \lim_{L\to\infty} \Phi\Big(\widehat{\operatorname*{argmin}_{\mathbf{Q} \in A \cdot \boldsymbol{\mathfrak{Q}}} \Phi}(\mathbf{Q})_{n,L}^{naive,1}\Big) \;=\; \Phi(A \cdot \boldsymbol{\mathfrak{Q}}) \qquad \text{a.s.},$$

and thus the quantity $\Phi\Big(\widehat{\operatorname{argmin}_{\mathbf{Q} \in A \cdot \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q})}_{n,L}^{naive,1}\Big)$ is a natural alternative to the estimate $\widehat{\Phi(A \cdot \boldsymbol{\mathfrak{Q}})}_{n,L}^{naive,1}$ given in (184).

Remark 91: By applying the above-mentioned results (with $A = 1$) to the special *divergence* (cf. Remark 43) $\Phi(\mathbf{Q}) := F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q}, \mathbb{P})\Big)$, we obtain for $\Phi(\boldsymbol{\mathfrak{Q}}) := \min_{\mathbf{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q})$ the estimator $\widehat{\Phi(\boldsymbol{\mathfrak{Q}})}_{n,L}^{naive,1} := -\frac{1}{n} \log \frac{1}{L} \sum_{\ell=1}^{L} 1_{\boldsymbol{\mathfrak{Q}}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}\big)$ as well as the alternative estimator $\Phi\Big(\widehat{\operatorname{argmin}_{\mathbf{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q})}_{n,L}^{naive,1}\Big)$. The involved estimator $\widehat{\operatorname{argmin}_{\mathbf{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q})}_{n,L}^{naive,1}$ of the minimizer set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q})$ is — at the same time (due to the strict increasingness of $F_{\gamma,\widetilde{c},1}$) — also the estimator of the minimizer set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q} \in \boldsymbol{\mathfrak{Q}}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q}, \mathbb{P})$; giving the latter had been left as an *open gap* in Broniatowski & Stummer [1], which we have now *filled/resolved*. Moreover, $D_{\widetilde{c}\cdot\varphi_\gamma}(\widehat{\operatorname{argmin}_{\mathbf{Q} \in \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q})}_{n,L}^{naive,1}, \mathbb{P})$ is an estimator of $\min_{\mathbf{Q} \in \boldsymbol{\mathfrak{Q}}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q}, \mathbb{P})$ which serves an alternative to the one given in Broniatowski & Stummer [1].

B. Naive estimators of max and argmax — Base-Divergence-Method 1, compact case

In the set-up of compact $A \cdot \boldsymbol{\mathfrak{Q}}$ with (67), we can handle the maximizing problem completely analogously to the method in the previous Subsection XIII-A. Indeed, by applying Theorem 54(a) instead of Theorem 52(a) we construct the naive BS-estimator

$$\widehat{\Phi(A \cdot \boldsymbol{\mathfrak{Q}})}_{n,L}^{naive,1} \;:=\; \frac{1}{n} \log \frac{1}{L} \sum_{\ell=1}^{L} \exp\Big(n \cdot \Big(F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}, \mathbb{P})\Big) + \Phi\big(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}\big)\Big)\Big) \cdot 1_{\boldsymbol{\mathfrak{Q}}}\big(\boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}\big)\,, \tag{188}$$

of the maximum value $\Phi(A \cdot \boldsymbol{\mathfrak{Q}}) := \max_{\mathbf{Q} \in A \cdot \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q})$; for this, we get

$$\lim_{n\to\infty} \widehat{\Phi(A \cdot \boldsymbol{\mathfrak{Q}})}_{n,\infty}^{naive,1} \;=\; \lim_{n\to\infty} \lim_{L\to\infty} \widehat{\Phi(A \cdot \boldsymbol{\mathfrak{Q}})}_{n,L}^{naive,1} \;=\; \Phi(\boldsymbol{\Omega}) \qquad \text{a.s.}$$

instead of (185) (notice the different definition of the involved quantities). As the corresponding very natural naive BS-estimator of the maximizer $\operatorname{argmax}_{\mathbf{Q} \in A \cdot \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q})$ we take

$$\widehat{\operatorname*{argmax}_{\mathbf{Q} \in A \cdot \boldsymbol{\mathfrak{Q}}} \Phi}(\mathbf{Q})_{n,L}^{naive,1} := \operatorname*{argmax}_{\boldsymbol{\nu} \in \mathcal{W}_{n,L}} \Phi(\boldsymbol{\nu})$$

instead of (186); in short, we take any $A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}}$ which maximizes $\Phi(\cdot)$ amongst all values $A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{(\ell)}}$ at hand which fall into $A \cdot \boldsymbol{\mathfrak{Q}}$. For this, we obtain — instead of (187) — the assertion

$$\max_{\mathbf{Q} \in A \cdot \boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q}) \;\geq\; \Phi\left(A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}}\right) \;\geq\; \widehat{\Phi(A \cdot \boldsymbol{\mathfrak{Q}})}_{n,\infty}^{naive,1} \;-\; o_{\mathbb{\Pi}}(1)$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L \to \infty$ and $n \to \infty$ under the distribution $\mathbb{\Pi}$. Hence, as L and n tend to infinity, $A \cdot \boldsymbol{\xi}_n^{w\mathbf{W}^{L,*}}$ concentrates to the set of maximizers of $\Phi(\cdot)$ on $A \cdot \boldsymbol{\mathfrak{Q}}$.

Proposition 92: In the above set-up, one has

$$\lim_{n\to\infty}\lim_{L\to\infty}\Phi\Big(\widehat{\operatorname{argmax}_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi}(\mathbf{Q})^{naive,1}_{n,L}\Big) \;=\; \Phi(A\cdot\mathfrak{\mathbf{Q}}) \qquad \text{a.s.},$$

and thus the quantity $\Phi\Big(\widehat{\operatorname{argmax}_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q})}^{naive,1}_{n,L}\Big)$ is a natural alternative to the estimate $\widehat{\Phi(A\cdot\mathfrak{\mathbf{Q}})}^{naive,1}_{n,L}$ given in (188).

C. Naive estimators — Base-Divergence-Method 1, non-compact case

Suppose that we are in the set-up of Theorem 52(b), which particularly means that $A\cdot\mathfrak{\mathbf{Q}}$ is not necessarily compact but satisfies (67) and (68), and that $\Phi : A\cdot\mathfrak{\mathbf{Q}}\mapsto\mathbb{R}$ is a continuous function which satisfies the lower-bound condition (106) (notice that (106) trivially holds if $A\cdot\mathfrak{\mathbf{Q}}$ is bounded but not necessarily closed). Additionally, let us now assume that the minimum value is achieved, i.e. $\inf_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q}) = \min_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q}) = \Phi(\mathbf{Q}_{min})$ for some (not necessarily unique) point $\mathbf{Q}_{min}\in A\cdot\mathfrak{\mathbf{Q}}$, and that the corresponding set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q})$ of minimizers is covered by a compact set $\mathbf{B}\in A\cdot\mathbb{S}^K$ (e.g. we take $\mathbf{B} := cl(A\cdot\mathfrak{\mathbf{Q}})$ in case that $A\cdot\mathfrak{\mathbf{Q}}$ is bounded but not necessarily closed). In such a context, we can proceed as in Subsection XIII-A by replacing $A\cdot\mathfrak{\mathbf{Q}}$ with $A\cdot\mathfrak{\mathbf{Q}}\cap\mathbf{B}$. The analogous procedure applies to the maximization problem, by proceeding as in Subsection XIII-B by replacing $A\cdot\mathfrak{\mathbf{Q}}$ with $A\cdot\mathfrak{\mathbf{Q}}\cap\mathbf{B}$.

D. Naive estimators of min and argmin — Base-Divergence-Method 2, compact case

In the above Subsections XIII-A, XIII-B, XIII-C we have chosen the divergence $F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbb{P})\Big)$ as the *base divergence* (cf. Theorem 52 and Theorem 54); we have referred to this choice as *Base-Divergence-Method 1*. However, as can be seen from the alternative Theorem 55 and Theorem 58, we can also choose — as Base-Divergence-Method 2 — as *base divergence* an appropriate "innmin scaled Bregman distance" $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\cdot,\mathbf{Q}^{**})$ (cf. (88),(90),(92)), where $\mathbf{Q}^{**}\in\mathbb{R}^K_{>0}$ *NEED NOT* be in $A\cdot\mathfrak{\mathbf{Q}}$.

Let us start with any fixed $\mathbf{Q}^{**}\in\mathbb{R}^K_{>0}$ and any fixed $\mathbf{P}\in\mathbb{R}^K_{>0}$ (with $M_{\mathbf{P}}=\sum_{i=1}^K p_i>0$) satisfying

$$t^{**}_k := \frac{q^{**}_k}{p_k}\in\,]t^{sc}_-,t^{sc}_+[\quad\text{for all } k=1,\ldots,K \qquad \text{(cf. (21))}.$$

In such a setup, recall from Theorem 55(a) that we have obtained for any continuous function $\Phi:\mathfrak{\mathbf{\Omega}}\mapsto\mathbb{R}$ on a compact set $A\cdot\mathfrak{\mathbf{Q}}\subset\widetilde{\mathcal{M}}_\gamma$ ($\gamma\in\mathbb{R}\setminus]1,2[$) with (67) the assertion

$$\min_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q}) \;=\; -\lim_{n\to\infty}\frac{1}{n}\log\Big(\,\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}}_n,\mathbf{Q}^{**}\Big)-\Phi\big(A\cdot\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}}_n\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{Q}}}\big(\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}}_n\big)\,\Big]\Big)\,, \tag{189}$$

where $\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}}_n$ is as in (80) with $\widetilde{V}$ instead of V (where $\widetilde{V}$ is constructed via (25) and (26) with the special choice $\widetilde{\varphi}(t) := M_{\mathbf{P}}\cdot\widetilde{c}\cdot\varphi_\gamma(t)$). Within such a set-up, we thus obtain for large $n\in\mathbb{N}$ (cf. (189)) the approximation

$$\Phi(A\cdot\mathfrak{\mathbf{Q}}) := \min_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q}) \;\approx\; -\frac{1}{n}\log\Big(\,\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}}_n,\mathbf{Q}^{**}\Big)-\Phi\big(A\cdot\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}}_n\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{Q}}}\big(\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}}_n\big)\,\Big]\Big) \tag{190}$$

and hence for getting an estimator of the minimum value $\Phi(A\cdot\mathfrak{\mathbf{Q}})$ one can estimate the right-hand side of (190). To achieve this, for the rest of this section we assume that n is chosen such that all $n\cdot\frac{p_k}{M_{\mathbf{P}}}$ are integers (and hence, $n=\sum_{k=1}^K n_k$ with $n_k = n\cdot\frac{p_k}{M_{\mathbf{P}}}$) — the remaining case works analogously. With this, we construct the corresponding *naive estimator* for the minimum value $\min_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q})$ as

$$\widehat{\Phi(A\cdot\mathfrak{\mathbf{Q}})}^{naive,2}_{n,L} \;:=\; -\frac{1}{n}\log\frac{1}{L}\sum_{\ell=1}^{L}\exp\Big(n\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}^{(\ell)}}_n,\mathbf{Q}^{**}\Big)-\Phi\big(A\cdot\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}^{(\ell)}}_n\big)\Big)\Big)\cdot 1_{\mathfrak{\mathbf{Q}}}\big(\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}^{(\ell)}}_n\big)\,, \tag{191}$$

where we simulate independently L copies $\widetilde{\mathbf{V}}^{(1)},\ldots,\widetilde{\mathbf{V}}^{(L)}$ of the random vector $\widetilde{\mathbf{V}} := \big(\widetilde{V}_1,\ldots,\widetilde{V}_n\big)$, and compute each of $\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}^{(1)}}_n,\ldots,\boldsymbol{\xi}^{w\widetilde{\mathbf{V}}^{(L)}}_n$ according to (80) (with $\widetilde{V}$ instead of V). Clearly, with the help of the strong law of large numbers we get with $\widehat{\Phi(A\cdot\mathfrak{\mathbf{Q}})}^{naive,2}_{n,\infty} := \lim_{L\to\infty}\widehat{\Phi(A\cdot\mathfrak{\mathbf{Q}})}^{naive,2}_{n,L}$ the following assertion:

Proposition 93: Let the assumptions of Theorem 55(a) be satisfied. Then one has

$$\lim_{n\to\infty}\widehat{\Phi(A\cdot\mathfrak{\mathbf{Q}})}^{naive,2}_{n,\infty} \;=\; \lim_{n\to\infty}\lim_{L\to\infty}\widehat{\Phi(A\cdot\mathfrak{\mathbf{Q}})}^{naive,2}_{n,L} \;=\; \Phi(A\cdot\mathfrak{\mathbf{Q}}) \qquad \text{a.s.} \tag{192}$$

As the corresponding *very natural naive BS-estimator of the minimizer-set* $\mathcal{Q}^* := \operatorname{argmin}_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi(\mathbf{Q})$, we take

$$\widehat{\operatorname*{argmin}_{\mathbf{Q}\in A\cdot\mathfrak{\mathbf{Q}}}\Phi}(\mathbf{Q})^{naive,2}_{n,L} := \operatorname*{argmin}_{\boldsymbol{\nu}\in\mathcal{V}_{n,L}}\Phi(\boldsymbol{\nu}) \tag{193}$$

where $\mathcal{V}_{n,L} := \{A \cdot \boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{(\ell)}} : \ell \in \{1,\dots,L\}\} \cap A \cdot \boldsymbol{\mathfrak{Q}}$. In short, we take as minimizer-estimate the (not necessarily unique) element $A \cdot \boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{L,*}}$ which minimizes $\Phi(\cdot)$ amongst all values $A \cdot \boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{(\ell)}}$ at hand which fall into $A \cdot \boldsymbol{\mathfrak{Q}}$. We prove that if L and n tend to infinity, then $A \cdot \boldsymbol{\xi}_n^{\mathbf{w}\widetilde{\mathbf{V}}^{L,*}}$ concentrates to the above-mentioned set $\mathcal{Q}^*$ of minimizers of $\Phi(\cdot)$ on $A \cdot \boldsymbol{\mathfrak{Q}}$. Indeed, we show that $A \cdot \boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{L,*}}$ is a proxy minimizer of $\Phi(\cdot)$ on $A \cdot \boldsymbol{\mathfrak{Q}}$, by proving the following

Proposition 94: There holds

$$\min_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}} \Phi(\mathbf{Q}) \;\le\; \Phi\left(A \cdot \boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{L,*}}\right) \;\le\; \widehat{\Phi(A\cdot\boldsymbol{\mathfrak{Q}})}_{n,\infty}^{naive,2} \;+\; o_{\mathbb{\Pi}}(1) \tag{194}$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L \to \infty$ and $n \to \infty$ under the distribution $\mathbb{\Pi}$ (recall that $\mathbb{\Pi}[(\widetilde{V}_1,\dots,\widetilde{V}_n) \in \cdot\,] = \bigotimes_{k=1}^{K} \widetilde{U}_k^{\otimes n_k}[\,\cdot\,]$)

The proof of Proposition 94 works analogously to the proof of Proposition 72, by replacing Theorem 22(a) with Theorem 55(a).

Proposition 95: In the above set-up, one has

$$\lim_{n\to\infty}\lim_{L\to\infty} \Phi\Big(\widehat{\operatorname{argmin}\,\Phi}(\mathbf{Q})_{n,L}^{naive,2}\Big) \;=\; \Phi(A\cdot\boldsymbol{\mathfrak{Q}}) \qquad \text{a.s.},$$

and thus the quantity $\Phi\Big(\widehat{\operatorname{argmin}_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}}\,\Phi(\mathbf{Q})}_{n,L}^{naive,2}\Big)$ is a natural alternative to the estimate $\widehat{\Phi(A\cdot\boldsymbol{\mathfrak{Q}})}_{n,L}^{naive,2}$ given in (191).

E. Naive estimators of max and argmax — Base-Divergence-Method 2, compact case

In the set-up of compact constraint set $A\cdot\boldsymbol{\mathfrak{Q}} \subset \widetilde{\mathcal{M}}_\gamma$ ($\gamma \in \mathbb{R}\setminus]1,2[$) with (67), we can treat the maximizing problem completely analogously to the method in the previous Subsection XIII-D. Indeed, by employing Theorem 58(a) instead of Theorem 55(a) we construct the naive BS-estimator

$$\widehat{\Phi(A\cdot\boldsymbol{\mathfrak{Q}})}_{n,L}^{naive,2} \;:=\; \frac{1}{n}\log\frac{1}{L}\sum_{\ell=1}^{L}\exp\Big(n\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{(\ell)}},\mathbf{Q}^{**}\Big)+\Phi\big(A\cdot\boldsymbol{\xi}_n^{\mathbf{w}\widetilde{\mathbf{V}}^{(\ell)}}\big)\Big)\Big)\cdot 1_{\boldsymbol{\mathfrak{Q}}}\big(\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{(\ell)}}\big)\,, \tag{195}$$

of the maximum value $\Phi(A\cdot\boldsymbol{\mathfrak{Q}}) := \max_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}}\Phi(\mathbf{Q})$; for this, we get

$$\lim_{n\to\infty}\widehat{\Phi(A\cdot\boldsymbol{\mathfrak{Q}})}_{n,\infty}^{naive,2} \;=\; \lim_{n\to\infty}\lim_{L\to\infty}\widehat{\Phi(A\cdot\boldsymbol{\mathfrak{Q}})}_{n,L}^{naive,2} \;=\; \Phi(A\cdot\boldsymbol{\mathfrak{Q}}) \qquad \text{a.s.}$$

instead of (192) (notice the different definition of the involved quantities). As the corresponding very natural naive BS-estimator of the maximizer $\operatorname{argmax}_{\mathbf{Q}\in A\cdot\boldsymbol{\Omega}}\Phi(\mathbf{Q})$ we take

$$\widehat{\operatorname*{argmax}_{\mathbf{Q}\in A\cdot\boldsymbol{\Omega}}\Phi}(\mathbf{Q})_{n,L}^{naive,2} := \operatorname*{argmax}_{\boldsymbol{\nu}\in\mathcal{V}_{n,L}}\Phi(\boldsymbol{\nu})$$

instead of (193); in short, we take any $A\cdot\boldsymbol{\xi}_n^{\mathbf{w}\widetilde{\mathbf{V}}^{L,*}}$ which maximizes $\Phi(\cdot)$ amongst all values $A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{(\ell)}}$ at hand which fall into $A\cdot\boldsymbol{\mathfrak{Q}}$. For this, we obtain — instead of (194) — the assertion

$$\max_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}}\Phi(\mathbf{Q}) \;\ge\; \Phi\left(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{L,*}}\right) \;\ge\; \widehat{\Phi(A\cdot\boldsymbol{\mathfrak{Q}})}_{n,\infty}^{naive,2} \;-\; o_{\mathbb{\Pi}}(1)$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L\to\infty$ and $n\to\infty$ under the distribution $\mathbb{\Pi}$. Hence, as L and n tend to infinity, $A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}^{L,*}}$ concentrates to the set of maximizers of $\Phi(\cdot)$ on $A\cdot\boldsymbol{\mathfrak{Q}}$.

Proposition 96: In the above set-up, one has

$$\lim_{n\to\infty}\lim_{L\to\infty}\Phi\Big(\widehat{\operatorname{argmax}\,\Phi}(\mathbf{Q})_{n,L}^{naive,2}\Big) \;=\; \Phi(A\cdot\boldsymbol{\mathfrak{Q}}) \qquad \text{a.s.},$$

and thus the quantity $\Phi\Big(\widehat{\operatorname{argmax}_{\mathbf{Q}\in A\cdot\boldsymbol{\mathfrak{Q}}}\,\Phi(\mathbf{Q})}_{n,L}^{naive,2}\Big)$ is a natural alternative to the estimate $\widehat{\Phi(A\cdot\boldsymbol{\mathfrak{Q}})}_{n,L}^{naive,2}$ given in (195).

F. Naive estimators — Base-Divergence-Method 2, non-compact case

We can proceed analogously to the above Base-Divergence-Method-1-treating Subsection XIII-C (e.g. by employing Theorem 55(b)).

G. Improved/Speed-up estimators of min, argmin, max and argmax — compact case

Recall from Subsection XII-G.1) that — for the case of compact constraint set $\mathbf{\Omega}$ with (6) — we have constructed improved/speed-up estimators of $\Phi(\mathbf{\Omega}) := \min_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$ and the corresponding minimizers $\operatorname{argmin}_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$ by switching from the base-divergence $D_{\varphi}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}, \mathbf{P}\big)$ to the different base-divergence $D^{SBD}_{\varphi,\mathbf{P}}\Big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}, \mathbf{Q}^*\Big)$ where $\mathbf{Q}^*$ is chosen to be in $int(\mathbf{\Omega})$ such that by construction the involved random vectors $\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ satisfy $\mathbb{E}_{\mathbb{\Pi}}\Big[M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big] \in int(\mathbf{\Omega})$ for all large enough n and even $\lim_{n\to\infty} \mathbb{E}_{\mathbb{\Pi}}\Big[M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big] = \mathbf{Q}^*$. For the current set-up of compact constraint sets $A \cdot \mathbf{\Omega}$ — with fixed constant-component-sum $A \in]0,\infty[$ and $\mathbf{\Omega} \in \mathbb{S}^K$ (respectively $\mathbb{S}^K_{>0}$) — with regularity assumptions (67), we can proceed similarly. Indeed, for achieving improved/speed-up estimators of the *minimum value* $\inf_{\mathbf{Q}\in A\cdot\mathbf{\Omega}} \Phi(\mathbf{Q}) = \min_{\mathbf{Q}\in A\cdot\mathbf{\Omega}} \Phi(\mathbf{Q})$ and of the corresponding (set of) *minimizers* $\arg\inf_{\mathbf{Q}\in A\cdot\mathbf{\Omega}} \Phi(\mathbf{Q}) = \arg\min_{\mathbf{Q}\in A\cdot\mathbf{\Omega}} \Phi(\mathbf{Q})$, we switch (say) from the base-divergence $F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}, \mathbb{P})\Big)$ to the different base-divergence $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A \cdot \boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}, \mathbf{Q}^{**}\Big)$ where $\mathbf{Q}^{**}$ should be chosen in an appropriate way, e.g. as follows: suppose that we have at hand some point $\mathbf{Q}^* \in int(A \cdot \mathbf{\Omega})$; for pregiven $\mathbf{P} \in \mathbb{R}^K_{>0}$, such a vector $\mathbf{Q}^*$ may be either pregiven (e.g. by the nature of the application context) or it may be simulatively achieved by e.g. proxy method 1 or proxy method 2 of Subsection X-A of Broniatowski & Stummer [1]. From this, we construct $\mathbf{Q}^{**} := \mathbf{Q}^*$. Also recall (24) with the setting $\widetilde{\varphi} := M_{\mathbf{P}} \cdot \varphi$, $\widetilde{\mathbb{P}} := \mathbf{P}/M_{\mathbf{P}}$, $\widetilde{\mathbf{Q}} := \mathbf{Q}/M_{\mathbf{P}}$ and $\widetilde{\mathbf{Q}}^{**} := \mathbf{Q}^{**}/M_{\mathbf{P}}$. Clearly, $\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k} = \frac{q_k^{**}}{p_k} = \frac{q_k^*}{p_k} \in]t_-^{sc}, t_+^{sc}[$ for all $k = 1,\ldots,K$ (recall from the first six rows in Table 1 that $]t_-^{sc}, t_+^{sc}[=]0,\infty[$ for $\gamma \in \mathbb{R}\backslash\{2\}$ and $]t_-^{sc}, t_+^{sc}[=]-\infty,\infty[$ for $\gamma = 2$) and thus (21) is satisfied. Accordingly, by the law of large numbers, the unweighted random vector $\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ (cf. (27)) performs the a.s. convergence

$$\begin{aligned}
\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}} \quad &:= \quad \Big(\frac{1}{n}\cdot\sum_{i\in I_1^{(n)}} \widetilde{V}_i,\, \ldots,\, \frac{1}{n}\cdot\sum_{i\in I_{K-1}^{(n)}} \widetilde{V}_i,\, \frac{1}{n}\cdot\sum_{i\in I_K^{(n)}} \widetilde{V}_i\Big) \\
&= \quad \Big(\frac{\lfloor n\cdot\widetilde{p}_1\rfloor}{n}\cdot\frac{1}{n_1}\cdot\sum_{i\in I_1^{(n)}} \widetilde{V}_i,\, \ldots,\, \frac{\lfloor n\cdot\widetilde{p}_{K-1}\rfloor}{n}\cdot\frac{1}{n_{K-1}}\cdot\sum_{i\in I_{K-1}^{(n)}} \widetilde{V}_i,\, \frac{n-\sum_{k=1}^{K-1}\lfloor n\cdot\widetilde{p}_k\rfloor}{n}\frac{1}{n_K}\cdot\sum_{i\in I_K^{(n)}} \widetilde{V}_i\Big) \\
&\xrightarrow[n\to\infty]{a.s.} \quad \Big(\widetilde{q}_1^{**},\, \ldots,\, \widetilde{q}_{K-1}^{**},\, \widetilde{q}_K^{**}\Big) = \widetilde{\mathbf{Q}}^{**},
\end{aligned}$$

where we have particularly used the fact that within each block $I_k^{(n)}$ all the $\widetilde{V}_i$ are independent and identically distributed with mean $E_{\mathbb{\Pi}}[\widetilde{V}_i] = \frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}$ and finite variance. Hence, the corresponding total sum of components $M_{\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}} = \frac{1}{n}\sum_{i=1}^n \widetilde{V}_i$ converges a.s. (as $n\to\infty$) to $M_{\widetilde{\mathbf{Q}}^{**}} = \sum_{k=1}^K \widetilde{q}_k^{**} = \frac{A}{M_{\mathbf{P}}}$. Moreover, — since all $\widetilde{q}_k^{**}$ are strictly positive — the indicator $1_{\mathbb{R}\backslash\{0\}}(\sum_{i=1}^n \widetilde{V}_i)$ converges a.s. (as $n\to\infty$) to 1 and, thus, $1_{\{0\}}(\sum_{i=1}^n \widetilde{V}_i)$ converges a.s. (as $n\to\infty$) to zero. Consequently, $\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}} = \frac{\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}}{\frac{1}{n}\sum_{i=1}^n \widetilde{V}_i} \cdot 1_{\mathbb{R}\backslash\{0\}}(\sum_{i=1}^n \widetilde{V}_i) + \boldsymbol{\infty} \cdot 1_{\{0\}}(\sum_{i=1}^n \widetilde{V}_i)$ converges a.s. (as $n\to\infty$) to $\frac{M_{\mathbf{P}}}{A}\cdot\widetilde{\mathbf{Q}}^{**} = \frac{1}{A}\cdot\mathbf{Q}^{**}$. Hence, we have derived the following

Proposition 97:

$$\lim_{n\to\infty} \mathbb{\Pi}[A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}} \in A\cdot\mathbf{\Omega}] \;=\; 1.$$

In other words, with the specific deliberate choice $\mathbf{Q}^{**} := \mathbf{Q}^* \in int(A\cdot\mathbf{\Omega})$, we obtain — for large enough approximation step n — the desired good hit-rate needed for the construction of the improved/speed-up estimators. For the latter we can now proceed, indeed, analogously to Subsection XII-G.1), by appropriately applying the results of the Subsections XIII-D and XIII-E (instead of Subsections XII-D and XII-E). For the sake of brevity, the details are omitted.

XIV. Bare-Simulation Estimators For General Divergence-Optimization-Problems Under Risk

Recall from Subsection XI-A that we are interested in the constrained optimization of the *continuous* distance-connected functions $\mathbf{\Omega} \ni \mathbb{Q} \mapsto \Phi_{\mathbf{R}}(\mathbb{Q})$ where $\mathbf{R}$ is an *unknown* (say) vector-valued "parameter" (i.e. under risk). In continuation of Section XI we now give *estimators* of both the minimum value $\min_{\mathbb{Q}\in\mathbf{\Omega}} \Phi_{\mathbf{R}}(\mathbf{Q})$ and the corresponding minimizer(s) $\operatorname{argmin}_{\mathbb{Q}\in\mathbf{\Omega}} \Phi_{\mathbf{R}}(\mathbf{Q})$. The case of maximum values and maximizers will be treated analogously. For the sake of brevity, we only focus on compact constraint sets $\mathbf{\Omega}$.

A. Naive estimators of min and argmin — Base-Divergence-Method 1, compact case

Let the constraint set $\mathbf{\Omega}$ be compact — with (67) — in the relative topology, and thus the minimum value of $\Phi_{\mathbf{R}}(\cdot)$ is achieved at some (not necessarily unique) point in $\mathbf{\Omega}$. Applied to such a situation, (143) and (146) give — for the random sample

$\mathbf{Y}_1^n = (Y_1, \ldots, Y_n)$ and the corresponding sequence $(\mathbf{R}_n(\mathbf{Y}_1^n))_{n\in\mathbb{N}}$ of (say) vector-valued function(al)s $\mathbf{R}_n(\mathbf{Y}_1^n)$ for which $\lim_{n\to\infty} \mathbf{R}_n(\mathbf{Y}_1^n) = \mathbf{R}$ a.s. (cf. (142)) the *sample-dependent minimum*

$$\begin{aligned}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{W}) &:= \min_{\mathbb{Q}\in\mathfrak{W}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \\ &= -\lim_{m\to\infty} \frac{1}{m} \log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}}, \mathbb{P}^{aux})\Big) - \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big)\Big)\cdot 1_{\mathfrak{W}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big]\Big), \end{aligned} \tag{196}$$

where

$$\boldsymbol{\xi}_m^{w\mathbf{W}} := \begin{cases} \left(\frac{\sum_{i\in I_1^{(m)}} W_i}{\sum_{k=1}^K \sum_{i\in I_k^{(m)}} W_i}, \ldots, \frac{\sum_{i\in I_K^{(m)}} W_i}{\sum_{k=1}^K \sum_{i\in I_k^{(m)}} W_i}\right), & \text{if } \sum_{j=1}^m W_j \neq 0, \\ (\infty,\ldots,\infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^m W_j = 0, \end{cases} \quad \text{(cf. (145))}$$

is the random vector constructed from (i) an auxiliary deterministic probability vector $\mathbb{P}^{aux} \in \mathbb{S}_{>0}^K$ inducing index blocks $I_k^{(m)}$ of sizes $m_k := \lfloor m\cdot p_k^{aux}\rfloor$ $(k = 1,\ldots,K-1)$ and $m_K := m - \sum_{k=1}^{K-1} m_k$, and (ii) from W_i's which are i.i.d. copies of the random variable W whose distribution is $\mathbb{\Pi}[W\in\cdot\,] = \mathfrak{l}[\,\cdot\,]$ being attributed to the power divergence generator $\varphi := \widetilde{c}\cdot\varphi_\gamma$ $(\gamma\in\mathbb{R}\setminus]1,2[)$ by the representability (69); moreover, $(W_i)_{i\in\mathbb{N}}$ and $(Y_i)_{i\in\mathbb{N}}$ are supposed to be independent. Recall from Remark 70 that the involved $\mathbb{P}^{aux}$ may also depend on the sample $\mathbf{Y}_1^n$; for the special case $\mathbb{P}^{aux} := \mathbb{P}_n^{emp}(\mathbf{Y}_1^n)$, we obtain $\boldsymbol{\xi}_m^{w\mathbf{W}} = \boldsymbol{\xi}_{n,m,\mathbf{Y}}^{w\mathbf{W}}$ (cf. (125) respectively (126)).

From (196), we obtain for large m the minimal-empirical-risk approximation (cf. (147))

$$\begin{aligned}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{W}) &= \min_{\mathbb{Q}\in\mathfrak{W}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \\ &\approx -\frac{1}{m}\log\Big(\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}}, \mathbb{P}^{aux})\Big) - \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big)\Big)\cdot 1_{\mathfrak{W}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big]\Big). \end{aligned} \tag{197}$$

Hence for getting an estimator of the minimal empirical risk $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{W}) = \min_{\mathbb{Q}\in\mathfrak{W}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$ one can estimate the right-hand side of (197).

To achieve this, for the rest of this section we assume that m is chosen such that all $m\cdot p_k^{aux}$ are integers (and hence, $m = \sum_{k=1}^K m_k$ with $m_k = m\cdot p_k^{aux}$) — the remaining case works analogously. Following the lines of Subsection XIII-A (with $A = 1$), a corresponding *naive (crude) estimator* can be constructed by

$$\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{W})}_{m,L}^{naive,1} := -\frac{1}{m}\log\frac{1}{L}\sum_{\ell=1}^L \exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}}, \mathbb{P}^{aux})\Big) - \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}}\big)\Big)\Big)\cdot 1_{\mathfrak{W}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}}\big), \tag{198}$$

where we (independently of $\mathbf{Y}_1^n$) simulate independently L copies $\mathbf{W}^{(1)},\ldots,\mathbf{W}^{(L)}$ of the vector $\mathbf{W} := (W_1,\ldots,W_m)$ with independent entries under $\mathfrak{l}$, and compute each of $\boldsymbol{\xi}_m^{w\mathbf{W}^{(1)}},\ldots,\boldsymbol{\xi}_m^{w\mathbf{W}^{(L)}}$ according to (145). Clearly, with the help the strong law of large numbers we get with $\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{W})}_{m,\infty}^{naive,1} := \lim_{L\to\infty} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{W})}_{m,L}^{naive,1}$ the following assertion:

Proposition 98:

$$\lim_{m\to\infty} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{W})}_{m,\infty}^{naive,1} = \lim_{m\to\infty}\lim_{L\to\infty} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{W})}_{m,L}^{naive,1} = \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{W}) \quad \text{a.s.} \tag{199}$$

As the corresponding *very natural naive (crude) estimator* of the minimizer-set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbb{Q}\in\mathfrak{W}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$, we take

$$\operatorname*{argmin}_{\mathbb{Q}\in\mathfrak{W}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})}_{m,L}^{naive,1} := \operatorname*{argmin}_{\mathbb{\nu}\in\mathcal{W}_{m,L}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{\nu}) \tag{200}$$

where $\mathcal{W}_{m,L} := \{\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}} : \ell\in\{1,\ldots,L\}\}\cap\mathfrak{W}$. In other words, as a corresponding *naive (crude) estimator* of the (not necessarily unique) element $\mathbb{Q}^*$ of the minimizer-set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbb{Q}\in\mathfrak{W}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$, we take the (not necessarily unique) element $\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}$ of the set $\{\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}} : \ell\in\{1,\ldots,L\}\}$ such that $\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}\in\mathfrak{W}$ and

$$\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}) \leq \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}}) \quad \text{for all } \ell = 1,\ldots,L \text{ for which } \boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}} \text{ belongs to } \mathfrak{W}.$$

In short, $\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}$ minimizes $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\cdot)$ amongst all values $\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}}$ at hand which fall into $\mathfrak{W}$. For large enough $m\in\mathbb{N}$ and $L\in\mathbb{N}$, such $\boldsymbol{\xi}_m^{w\widetilde{\mathbf{W}}^{L,*}}$ exists since $\mathfrak{W}$ has non-void interior in the relative topology, by assumption (67). We prove that if L and m tend to infinity, then $\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}$ concentrates to the above-mentioned set $\mathcal{Q}^*$ of minimizers of $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\cdot)$ on $\mathfrak{W}$. As usual in similar procedures, L is assumed to be large enough in order to justify some approximation for fixed m, typically the substitution of empirical means by expectations, since L is at disposal.

Next we derive that $\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}$ is a proxy minimizer of $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\cdot)$ on $\mathfrak{Q}$, manifested by the following

Proposition 99: There holds

$$\min_{\mathbb{Q}\in\mathfrak{Q}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \;\le\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\left(\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}\right) \;\le\; \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,\infty}^{naive,1} \;+\; o_{\mathbb{\Pi}}(1) \tag{201}$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L\to\infty$ and $m\to\infty$ under the distribution $\mathbb{\Pi}$ (recall that $\mathbb{\Pi}[(W_1,\ldots,W_m)\in\cdot\,] = \zeta^{\otimes m}[\,\cdot\,]$).

The proof of Proposition 99 works analogously to the proof of Proposition 72, by replacing Theorem 22(a) with (196).

Remark 100: In the current set-up of compact $\mathfrak{Q}$ with (67), by taking the special case $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) := D_\varphi(\mathbb{Q},\mathbf{R}_n(\mathbf{Y}_1^n))$ with $\mathbf{R}_n(\mathbf{Y}_1^n) := \mathbb{P}_n^{emp}(\mathbf{Y}_1^n)$ we obtain the naive BS-estimator $D_\varphi(\mathfrak{Q},\widehat{\mathbb{P}_n^{emp}}(\mathbf{Y}_1^n))_{n,L}^{naive,1}$ of the minimum value $\min_{\mathbb{Q}\in\mathfrak{Q}} D_\varphi(\mathbb{Q},\mathbb{P}_n^{emp}(\mathbf{Y}_1^n))$.

Proposition 101: In the above set-up, one has

$$\lim_{m\to\infty}\lim_{L\to\infty} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\operatorname*{argmin}_{\mathbb{Q}\in\mathfrak{Q}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,1}\Big) \;=\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q}) \qquad \text{a.s.},$$

and thus the quantity $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\operatorname{argmin}_{\mathbb{Q}\in\mathfrak{Q}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,1}\Big)$ is a natural alternative to the estimate $\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,L}^{naive,1}$ given in (198).

Remark 102: (a) By applying the above-mentioned results to the special *divergence* (cf. Remark 43) $\Phi_{\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)}(\mathbb{Q}) := F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P}_n^{emp}(\mathbf{Y}_1^n))\Big)$, we obtain for $\Phi_{\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)}(\mathfrak{Q}) := \min_{\mathbb{Q}\in\mathfrak{Q}}\Phi_{\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)}(\mathbb{Q})$ the estimator $\widehat{\Phi_{\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,L}^{naive,1} := -\frac{1}{m}\log\frac{1}{L}\sum_{\ell=1}^L 1\!\!1_{\mathfrak{Q}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}}\big)$ as well as the alternative estimator $\Phi_{\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)}\Big(\operatorname{argmin}_{\mathbb{Q}\in\mathfrak{Q}} \widehat{\Phi_{\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,1}\Big)$. The involved estimator $\operatorname{argmin}_{\mathbb{Q}\in\mathfrak{Q}} \widehat{\Phi_{\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,1}$ of the minimizer set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbb{Q}\in\mathfrak{Q}}\Phi_{\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)}(\mathbb{Q})$ is — at the same time (due to the strict increasingness of $F_{\gamma,\widetilde{c},1}$) — also the estimator of the minimizer set $\mathcal{Q}^* := \operatorname{argmin}_{\mathbb{Q}\in\mathfrak{Q}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P}_n^{emp}(\mathbf{Y}_1^n))$; giving the latter had been left as an *open gap* in Broniatowski & Stummer [1], which we have now *filled/resolved*. By the way, these newly developed minimizer(s) are the *non-parametric* analogues of the very prominent *parametric* minimum-$\varphi-$divergence estimator(s) (e.g. the omnipresent *maximum-likelihood estimator(s)* corresponds to the minimum-Kullback-Leibler-information-distance estimator, i.e. $\varphi := \varphi_1$ given in (16)).
(b) By construction, $D_{\widetilde{c}\cdot\varphi_\gamma}(\operatorname{argmin}_{\mathbb{Q}\in\mathfrak{Q}} \widehat{\Phi_{\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,1},\mathbb{P})$ is an estimator of $\min_{\mathbb{Q}\in\mathfrak{Q}} D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbb{Q},\mathbb{P}_n^{emp}(\mathbf{Y}_1^n))$ which serves an alternative to the one given in Broniatowski & Stummer [1].

B. Naive estimators of max and argmax — Base-Divergence-Method 1, compact case

In the set-up of compact $\mathfrak{Q}$ — with (67) — in the relative topology, we can handle the maximizing problem completely analogously to the method in the previous Subsection XIV-A. In fact, instead of (143) and (146) (leading to (147)) we employ (149) and (151) (leading to (152)) — for the random sample $\mathbf{Y}_1^n = (Y_1,\ldots,Y_n)$ — to obtain (for large m) the following approximation of the sample-dependent maximum

$$\begin{aligned}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q}) &:= \max_{\mathbb{Q}\in\mathfrak{Q}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})\\ &\approx\; \frac{1}{m}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}},\mathbb{P}^{aux})\Big)+\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\Big)\Big)\cdot 1\!\!1_{\mathfrak{Q}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}}\big)\;\Big]\Big),\end{aligned}$$

from which we construct the corresponding *naive BS-estimator* as

$$\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,L}^{naive,1} \;:=\; \frac{1}{m}\log\frac{1}{L}\sum_{\ell=1}^L \exp\Big(m\cdot\Big(F_{\gamma,\widetilde{c},1}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}},\mathbb{P}^{aux})\Big)+\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}}\big)\Big)\Big)\cdot 1\!\!1_{\mathfrak{Q}}\big(\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}}\big)\,; \tag{202}$$

for this, we get

$$\lim_{m\to\infty}\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,\infty}^{naive,1} \;=\; \lim_{m\to\infty}\lim_{L\to\infty}\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,L}^{naive,1} \;=\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q}) \qquad \text{a.s.}$$

instead of (199) (notice the different definition of the involved quantities). As the corresponding very natural naive BS-estimator of the maximizer(-set) $\operatorname{argmax}_{\mathbb{Q}\in\mathfrak{Q}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$ we take

$$\operatorname*{argmax}_{\mathbb{Q}\in\mathfrak{Q}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,1} := \operatorname*{argmax}_{\mathbb{\nu}\in\mathcal{W}_{m,L}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{\nu})$$

instead of (200); in short, we take any $\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}$ which maximizes $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\cdot)$ amongst all values $\boldsymbol{\xi}_m^{w\mathbf{W}^{(\ell)}}$ at hand which fall into $\mathfrak{Q}$. For this, we obtain — instead of (201) — the assertion

$$\max_{\mathbf{Q}\in\mathfrak{Q}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbf{Q}) \;\geq\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}\Big) \;\geq\; \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,\infty}^{naive,1} \;-\; o_{\mathbb{\Pi}}(1)$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L\to\infty$ and $m\to\infty$ under the distribution $\mathbb{\Pi}$. Hence, as L and m tend to infinity, $\boldsymbol{\xi}_m^{w\mathbf{W}^{L,*}}$ concentrates to the set of maximizers of $\Phi(\cdot)$ on $\mathfrak{Q}$.

Proposition 103: In the above set-up, one has

$$\lim_{m\to\infty}\lim_{L\to\infty} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\operatorname*{argmax}_{\mathbb{Q}\in\mathfrak{Q}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})}_{m,L}^{naive,1}\Big) \;=\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q}) \qquad \text{a.s.},$$

and thus the quantity $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\operatorname{argmax}_{\mathbb{Q}\in\mathfrak{Q}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbf{Q})}_{m,L}^{naive,1}\Big)$ is a natural alternative to the estimate $\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,L}^{naive,1}$ given in (202).

C. Naive estimators of min and argmin — Base-Divergence-Method 2, compact case

Relying on Subsection XI-B, in the above Subsections XIV-A and XIV-B we have chosen the divergence $F_{\gamma,\widetilde{c},1}(D_{\widetilde{c}\cdot\varphi_\gamma}(\cdot,\mathbb{P}^{aux}))$ as the *base divergence*; we have referred to this choice as *Base-Divergence-Method 1*. However, relying on the alternative Subsection XI-D, we can also choose — as Base-Divergence-Method 2 — as *base divergence* the "innmin scaled Bregman distance" $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}^{aux}}(\cdot,\mathbf{Q}^{**})$ (cf. (88),(90),(92)), where $\mathbf{Q}^{**}\in\mathbb{R}^K_{>0}$ *NEED NOT* be in $\mathfrak{Q}$.

Let us start with any fixed $\mathbf{Q}^{**}\in\mathbb{R}^K_{>0}$ and any fixed $\mathbb{P}^{aux}\in\mathbb{S}^K_{>0}$ satisfying

$$t_k^{**} := \frac{q_k^{**}}{p_k^{aux}} \in]t_-^{sc}, t_+^{sc}[\quad \text{for all } k=1,\ldots,K \qquad \text{(cf. (21))}.$$

In such a setup, recall from (156) that for all compact sets $\mathfrak{Q}\subset\widetilde{\mathcal{M}}_\gamma$ (with $A=1$ and $\gamma\in\mathbb{R}\setminus]1,2[$) satisfying the regularity properties (67) *in the relative topology* one gets the approximation (for large m)

$$\begin{aligned}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q}) &:= \min_{\mathbb{Q}\in\mathfrak{Q}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})\\ &\approx\; -\frac{1}{m}\log\Big(\,\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}},\mathbf{Q}^{**}) - \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{V}}\big)\Big)\Big)\cdot \mathbb{1}_{\mathfrak{Q}}\big(\boldsymbol{\xi}_m^{w\mathbf{V}}\big)\Big]\Big), \end{aligned} \tag{203}$$

where $\boldsymbol{\xi}_m^{w\mathbf{V}}$ is as in (154). Also recall that — in accordance with Remark 70 — the involved $\mathbb{P}^{aux}$ may also depend on the sample $\mathbf{Y}_1^n$; for the special case $\mathbb{P}^{aux}:=\mathbb{P}_n^{emp}(\mathbf{Y}_1^n)$ we obtain $\boldsymbol{\xi}_m^{w\mathbf{V}}=\boldsymbol{\xi}_{n,m,\mathbf{Y}}^{w\mathbf{V}}$ (cf. (136) respectively (137), with $\mathbf{Y}$ instead of $\mathbf{y}$).

Accordingly, for getting an estimator of the sample-dependent minimum $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q}) = \min_{\mathbb{Q}\in\mathfrak{Q}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$ one can estimate the right-hand side of (203).

To achieve this, for the rest of this section we assume that m is chosen such that all $m\cdot p_k^{aux}$ are integers (and hence, $m=\sum_{k=1}^K m_k$ with $m_k = m\cdot p_k^{aux}$) — the remaining case works analogously. Following the lines of Subsection XIII-D (with $A=1$), a corresponding *naive (crude) estimator* can be constructed by

$$\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,L}^{naive,2} \;:=\; -\frac{1}{m}\log\frac{1}{L}\sum_{\ell=1}^{L}\exp\Big(m\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}^{aux}}\Big(\boldsymbol{\xi}_m^{w\mathbf{V}^{(\ell)}},\mathbf{Q}^{**}\Big)\Big) - \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{V}^{(\ell)}}\big)\Big)\Big)\cdot\mathbb{1}_{\mathfrak{Q}}\big(\boldsymbol{\xi}_m^{w\mathbf{V}^{(\ell)}}\big), \tag{204}$$

where we simulate independently L copies $\mathbf{V}^{(1)},\ldots,\mathbf{V}^{(L)}$ of the random vector $\mathbf{V}:=(V_1,\ldots,V_m)$, and compute each of $\boldsymbol{\xi}_m^{w\mathbf{V}^{(1)}},\ldots,\boldsymbol{\xi}_m^{w\mathbf{V}^{(L)}}$ according to (80) (with m instead of n). Clearly, with the help of the strong law of large numbers we get with $\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,\infty}^{naive,2} := \lim_{L\to\infty}\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,L}^{naive,2}$ the following assertion:

Proposition 104:

$$\lim_{m\to\infty}\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,\infty}^{naive,2} \;=\; \lim_{m\to\infty}\lim_{L\to\infty}\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q})}_{m,L}^{naive,2} \;=\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathfrak{Q}) \qquad \text{a.s.} \tag{205}$$

As the corresponding *very natural naive (crude) estimator* of the minimizer-set $\mathcal{Q}^*:=\operatorname{argmin}_{\mathbb{Q}\in\mathfrak{Q}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$, we take

$$\operatorname*{argmin}_{\mathbb{Q}\in\mathfrak{Q}}\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})}_{m,L}^{naive,2} := \operatorname*{argmin}_{\mathbb{v}\in\mathcal{V}_{m,L}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{v}) \tag{206}$$

where $\mathcal{V}_{m,L} := \{\boldsymbol{\xi}_m^{w\mathbf{V}^{(\ell)}} : \ell \in \{1,\dots,L\}\} \cap \boldsymbol{\Omega}$. In short, we take as minimizer-estimate the (not necessarily unique) element $\boldsymbol{\xi}_m^{w\mathbf{V}^{L,*}}$ which minimizes $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\cdot)$ amongst all values $\boldsymbol{\xi}_m^{w\mathbf{V}^{(\ell)}}$ at hand which fall into $\boldsymbol{\Omega}$. We prove that if L and m tend to infinity, then $\boldsymbol{\xi}_m^{w\mathbf{V}^{L,*}}$ concentrates to the above-mentioned set $\mathcal{Q}^*$ of minimizers of $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\cdot)$ on $\boldsymbol{\Omega}$. As usual in similar procedures, L is assumed to be large enough in order to justify some approximation for fixed m, typically the substitution of empirical means by expectations, since L is at disposal.

Next we derive that $\boldsymbol{\xi}_m^{w\mathbf{V}^{L,*}}$ is a proxy minimizer of $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\cdot)$ on $\boldsymbol{\Omega}$, manifested by the following

Proposition 105: There holds

$$\min_{\mathbb{Q}\in\boldsymbol{\Omega}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \;\le\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\boldsymbol{\xi}_m^{w\mathbf{V}^{L,*}}\Big) \;\le\; \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega})}_{m,\infty}^{naive,2} \;+\; o_{\mathbb{\Pi}}(1) \tag{207}$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L\to\infty$ and $m\to\infty$ under the distribution $\mathbb{\Pi}$ (recall that $\mathbb{\Pi}[(V_1,\dots,V_m)\in\cdot\,] = \bigotimes_{k=1}^K U_k^{\otimes m_k}[\,\cdot\,]$).

The proof of Proposition 105 works analogously to the proof of Proposition 72, by replacing Theorem 22(a) with (155).

Proposition 106: In the above set-up, one has

$$\lim_{m\to\infty}\lim_{L\to\infty} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\operatorname*{argmin}_{\mathbb{Q}\in\boldsymbol{\Omega}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,2}\Big) \;=\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega}) \qquad \text{a.s.},$$

and thus the quantity $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\operatorname{argmin}_{\mathbb{Q}\in\boldsymbol{\Omega}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,2}\Big)$ is a natural alternative to the estimate $\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega})}_{m,L}^{naive,2}$ given in (204).

D. Naive estimators of max and argmax — Base-Divergence-Method 2, compact case

In the set-up of compact $\boldsymbol{\Omega}$ — with (67) — in the relative topology, we can handle the maximizing problem completely analogously to the method in the previous Subsection XIV-C. In fact, instead of (203) we employ (159) — for the random sample $\mathbf{Y}_1^n = (Y_1,\dots,Y_n)$ — to obtain (for large m) the approximation of the sample-dependent maximum

$$\begin{aligned}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega}) &:= \max_{\mathbb{Q}\in\boldsymbol{\Omega}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \\ &\approx\; \frac{1}{m}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}_{\mathbf{Y}_1^n}}\Big[\exp\Big(m\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}},\mathbf{Q}^{**}) + \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{V}}\big)\Big)\Big)\cdot 1\!\!1_{\boldsymbol{\Omega}}\big(\boldsymbol{\xi}_m^{w\mathbf{V}}\big)\;\Big]\Big)\end{aligned}$$

(where where $\mathbf{Q}^{**}\in\mathbb{R}^K_{>0}$ *NEED NOT* be in $\boldsymbol{\Omega}$), from which we construct the corresponding *naive BS-estimator* as

$$\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega})}_{m,L}^{naive,2} \;:=\; \frac{1}{m}\log\frac{1}{L}\sum_{\ell=1}^{L}\exp\Big(m\cdot\Big(\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}^{aux}}\Big(\boldsymbol{\xi}_m^{w\mathbf{V}^{(\ell)}},\mathbf{Q}^{**}\Big) + \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\big(\boldsymbol{\xi}_m^{w\mathbf{V}^{(\ell)}}\big)\Big)\Big)\cdot 1\!\!1_{\boldsymbol{\Omega}}\big(\boldsymbol{\xi}_m^{w\mathbf{V}^{(\ell)}}\big)\,; \tag{208}$$

for this, we get

$$\lim_{m\to\infty}\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega})}_{m,\infty}^{naive,2} \;=\; \lim_{m\to\infty}\lim_{L\to\infty}\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega})}_{m,L}^{naive,2} \;=\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega}) \qquad \text{a.s.}$$

instead of (205) (notice the different definition of the involved quantities). As the corresponding very natural naive BS-estimator of the maximizer(-set) $\operatorname{argmax}_{\mathbb{Q}\in\boldsymbol{\Omega}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$ we take

$$\operatorname*{argmax}_{\mathbb{Q}\in\boldsymbol{\Omega}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,2} := \operatorname*{argmax}_{\mathbb{\nu}\in\mathcal{V}_{m,L}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{\nu})$$

instead of (206); in short, we take any $\boldsymbol{\xi}_m^{w\mathbf{V}^{L,*}}$ which maximizes $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\cdot)$ amongst all values $\boldsymbol{\xi}_m^{w\mathbf{V}^{(\ell)}}$ at hand which fall into $\boldsymbol{\Omega}$. For this, we obtain — instead of (207) — the assertion

$$\max_{\mathbb{Q}\in\boldsymbol{\Omega}} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) \;\ge\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\boldsymbol{\xi}_m^{w\mathbf{V}^{L,*}}\Big) \;\ge\; \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega})}_{m,\infty}^{naive,2} \;-\; o_{\mathbb{\Pi}}(1)$$

where $o_{\mathbb{\Pi}}(1)$ goes to 0 as $L\to\infty$ and $m\to\infty$ under the distribution $\mathbb{\Pi}$. Hence, as L and m tend to infinity, $\boldsymbol{\xi}_m^{w\mathbf{V}^{L,*}}$ concentrates to the set of maximizers of $\Phi(\cdot)$ on $\boldsymbol{\Omega}$.

Proposition 107: In the above set-up, one has

$$\lim_{m\to\infty}\lim_{L\to\infty} \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\operatorname*{argmax}_{\mathbb{Q}\in\boldsymbol{\Omega}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,2}\Big) \;=\; \Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega}) \qquad \text{a.s.},$$

and thus the quantity $\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}\Big(\operatorname{argmax}_{\mathbb{Q}\in\boldsymbol{\Omega}} \widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}}(\mathbb{Q})_{m,L}^{naive,2}\Big)$ is a natural alternative to the estimate $\widehat{\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\boldsymbol{\Omega})}_{m,L}^{naive,2}$ given in (208).

E. Improved/Speed-up estimators of min, argmin, max and argmax — compact case

Recall from Subsection XIII-G that — for the case of compact constraint sets $A \cdot \mathbf{\Omega}$ with (67) — we have constructed improved/speed-up estimators of $\inf_{\mathbf{Q}\in A\cdot\mathbf{\Omega}} \Phi(\mathbf{Q}) = \min_{\mathbf{Q}\in A\cdot\mathbf{\Omega}} \Phi(\mathbf{Q})$ and of the corresponding (set of) *minimizers* $\arg\inf_{\mathbf{Q}\in A\cdot\mathbf{\Omega}} \Phi(\mathbf{Q}) = \arg\min_{\mathbf{Q}\in A\cdot\mathbf{\Omega}} \Phi(\mathbf{Q})$, by switching from the base-divergence $F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}},\mathbb{P})\Big)$ to the different base-divergence $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}\Big(A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}},\mathbf{Q}^*\Big)$ where $\mathbf{Q}^*$ is chosen to be in $int(A\cdot\mathbf{\Omega})$ such that by construction the involved random vector $A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}$ converges a.s. (as $n\to\infty$) to $\mathbf{Q}^*$ and consequently $\lim_{n\to\infty}\mathbb{\Pi}[A\cdot\boldsymbol{\xi}_n^{w\widetilde{\mathbf{V}}}\in A\cdot\mathbf{\Omega}] = 1$. Accordingly, we obtained — for large enough approximation step n — the desired good hit-rate needed for the construction of the improved/speed-up estimators.

For the current *risk-case* set-up on compact constraint sets $\mathbf{\Omega}\in\mathbb{S}^K$ with (67), we can act similarly. Indeed, for achieving improved/speed-up estimators of of the *minimum value* $\inf_{\mathbb{Q}\in\mathbf{\Omega}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) = \min_{\mathbb{Q}\in\mathbf{\Omega}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$ and of the corresponding (set of) *minimizers* $\arg\inf_{\mathbb{Q}\in\mathbf{\Omega}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q}) = \arg\min_{\mathbb{Q}\in\mathbf{\Omega}}\Phi_{\mathbf{R}_n(\mathbf{Y}_1^n)}(\mathbb{Q})$ we employ the base-divergence $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbb{P}^{aux}}(\boldsymbol{\xi}_m^{w\mathbf{V}},\mathbb{Q}^*)$ where $\mathbb{Q}^*$ is chosen to be in $int(\mathbf{\Omega})$ such that by construction the involved random vector $\boldsymbol{\xi}_m^{w\mathbf{V}}$ converges — with respect to the conditional distribution $\mathbb{\Pi}_{\mathbf{Y}_1^n}$ — a.s. (as $m\to\infty$) to $\mathbb{Q}^*$ and consequently $\lim_{m\to\infty}\mathbb{\Pi}_{\mathbf{Y}_1^n}[\boldsymbol{\xi}_m^{w\mathbf{V}}\in\mathbf{\Omega}] = 1$. Accordingly, we obtain — for large enough approximation step m — the desired good hit-rate needed for the construction of the improved/speed-up estimators, and proceed analogously to Subsection XII-G.1), by appropriately applying the results of the Subsections XIV-C and XIV-D (instead of Subsections XII-D and XII-E). For the sake of brevity, the details are omitted.

APPENDIX A
PROOFS

Proof of Theorem 16. We first prove that for all $k=1,\ldots,K$ the function $\widetilde{\varphi}_k(\cdot) := M_{\mathbf{P}}\cdot\varphi_k(\cdot)$ (cf. (22)) satisfies the representability

$$\widetilde{\varphi}_k(t) \;=\; \sup_{z\in\mathbb{R}}\Big(z\cdot t-\log\int_{\mathbb{R}} e^{z\cdot y}d\widetilde{U}_k(y)\Big),\qquad t\in\mathbb{R}. \tag{209}$$

In other words, we have to show that for all $k=1,\ldots,K$ the function $t\mapsto\widetilde{\varphi}_k(t)$ is the Fenchel-Legendre transform of the cumulant-generating function (log-moment-generating function) $\Lambda_{\widetilde{U}_k}(z) := \log\int_{\mathbb{R}} e^{z\cdot y}d\widetilde{U}_k(y)$ of the distribution $\widetilde{U}_k$. To start with, let us fix a $k\in\{1,\ldots,K\}$ and recall that we have constructed

$$d\widetilde{U}_k(v) \;:=\; \frac{\exp(\tau_k\cdot v)}{MGF_{\widetilde{\zeta}}(\tau_k)}\,d\widetilde{\zeta}(v) = \frac{\exp(\tau_k\cdot v)}{\int_{\mathbb{R}}e^{\tau_k\cdot y}d\widetilde{\zeta}(y)}\,d\widetilde{\zeta}(v)\qquad\text{(cf. (26))}$$

where $\tau_k := M_{\mathbf{P}}\cdot\varphi'\Big(\frac{q_k^{**}}{p_k}\Big) = \widetilde{\varphi}'\Big(\frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\Big) = \widetilde{\varphi}'(t_k^{**}) < \infty$ with $t_k^{**} = \frac{\widetilde{q}_k^{**}}{\widetilde{p}_k}\in\,]t_-^{sc},t_+^{sc}[$ (cf. (21)). From this, we get for all $t\in\mathbb{R}$

$$\sup_{z\in\mathbb{R}}\Big(z\cdot t-\log\int_{\mathbb{R}}e^{z\cdot y}d\widetilde{U}_k(y)\Big) \;=\; \sup_{z\in\mathbb{R}}\Big(z\cdot t-\Big(\Lambda_{\widetilde{\zeta}}(z+\tau_k)-\Lambda_{\widetilde{\zeta}}(\tau_k)\Big)\Big) \;=\; \sup_{z\in\mathbb{R}}\Big(z\cdot t-\Lambda_{\widetilde{\zeta}}(z)\Big)+\Lambda_{\widetilde{\zeta}}(\tau_k)-t\cdot\tau_k \tag{210}$$

$$=\; \widetilde{\varphi}(t)+\Lambda_{\widetilde{\zeta}}(\tau_k)-t\cdot\tau_k \;=\; \widetilde{\varphi}(t)+\Lambda_{\widetilde{\zeta}}(\widetilde{\varphi}'(t_k^{**}))-t\cdot\widetilde{\varphi}'(t_k^{**}) \;=\; \widetilde{\varphi}(t)+t_k^{**}\cdot\widetilde{\varphi}'(t_k^{**})-\widetilde{\varphi}(t_k^{**})-t\cdot\widetilde{\varphi}'(t_k^{**}) = \widetilde{\varphi}_k(t), \tag{211}$$

where in the first equality of (211) we have employed our basic representability assumption

$$\widetilde{\varphi}(t) \;=\; \sup_{z\in\mathbb{R}}\Big(z\cdot t-\Lambda_{\widetilde{\zeta}}(z)\Big),\qquad t\in\mathbb{R},\qquad\text{(cf. (9))}$$

and in the third equality of (211) that the corresponding explicit solution of (9) for $t=t_k^{**}\in\,]t_-^{sc},t_+^{sc}[$ is given by [27]

$$\widetilde{\varphi}(t_k^{**}) \;=\; t_k^{**}\cdot\Lambda'^{\leftarrow}_{\widetilde{\zeta}}(t_k^{**})-\Lambda_{\widetilde{\zeta}}\Big(\Lambda'^{\leftarrow}_{\widetilde{\zeta}}(t_k^{**})\Big) \;=\; t_k^{**}\cdot\widetilde{\varphi}'(t_k^{**})-\Lambda_{\widetilde{\zeta}}(\widetilde{\varphi}'(t_k^{**}))\,.$$

A comprehensive study of the properties of $\Lambda_{\widetilde{\zeta}}$ is given in Section XI of Broniatowski & Stummer [1] as well as in Broniatowski & Stummer [69].

Summing up, so far we have proved the representability (209). Moreover, recall from Condition 5 that the moment-generating function $z\mapsto MGF_{\widetilde{\zeta}}(z)=\int_{\mathbb{R}}e^{z\cdot y}d\widetilde{\zeta}(y)$ (and equivalently, $z\mapsto\Lambda_{\widetilde{\zeta}}(z)$) is supposed to be finite on some open interval containing zero. This implies — due to $\Lambda_{\widetilde{U}_k}(z)=\Lambda_{\widetilde{\zeta}}(z+\tau_k)-\Lambda_{\widetilde{\zeta}}(\tau_k)$ (cf. the first equality in (210)) and $\tau_k\in int(dom(\Lambda_{\widetilde{\zeta}}))$ — that also $z\mapsto\Lambda_{\widetilde{U}_k}(z)$ is finite on some open interval containing zero ($k=1,\ldots,K$). Hence, one can prove for $D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**}) = \sum_{k=1}^K p_k\cdot\varphi_k\Big(\frac{q_k}{p_k}\Big)$ (cf. (23)) the narrow-sense BS-minimizability (28) *analogously* to the way in which the narrow-sense

[27] recall that $h^{\leftarrow}$ denotes the inverse of a function h

BS-minimizability (14) was proved for $D_\varphi(\mathbf{Q},\mathbf{P}) = \sum_{k=1}^K p_k \cdot \varphi\left(\frac{q_k}{p_k}\right)$ (cf. (4)) in Appendix A of Broniatowski & Stummer [1]. For the sake of brevity, we omit the details. ■

Proof of Theorem 22. It suffices to prove part (b); from there, the part (a) follows immediately as a special case. Let us define the auxiliary function $\widetilde{F}(\mathbf{Q}) := D_\varphi(\mathbf{Q},\mathbf{P}) - \Phi(\mathbf{Q})$ which by our assumption (47) satisfies $\widetilde{F}(\mathbf{Q}) \leq c_1$ for all $\mathbf{Q} \in \mathbf{\Omega}$. Moreover, by rewriting (14) of Theorem 9 we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{\Pi}\Big[M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}} \in \mathbf{\Omega}\Big] = - \inf_{\mathbf{Q}\in\mathbf{\Omega}} D_\varphi(\mathbf{Q},\mathbf{P})\,.$$

By applying the Varadhan Lemma (see e.g. the lecture notes of Franco [373] and Swart [374], and the references therein) to the sequence $(\mu_n)_{n\in\mathbb{N}}$ of probability distributions on $\mathbb{R}^K$ (equipped with the corresponding Borel $\sigma-$field) given by $\mu_n[\,\cdot\,] := \mathbb{\Pi}\big[M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}} \in \cdot\,\big]$ we obtain (with a slight abuse of notation)

$$\lim_{n\to\infty} \frac{1}{n} \log \left(\int\limits_{\mathbb{R}^K} \exp(n \cdot \widetilde{F}(\mathbf{Q})) \cdot 1_{\mathbf{\Omega}}(\mathbf{Q})\, \mathrm{d}\mu_n(\mathbf{Q}) \right) = \sup_{\mathbf{Q}\in\mathbf{\Omega}} \Big(\widetilde{F}(\mathbf{Q}) \,-\, D_\varphi(\mathbf{Q},\mathbf{P}) \Big)\,. \tag{212}$$

As usual, the left-hand side of (212) can be equivalently rewritten as

$$\lim_{n\to\infty} \frac{1}{n} \log\Big(\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n \cdot \widetilde{F}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big)\Big) \cdot 1_{\mathbf{\Omega}}\big(M_{\mathbf{P}} \cdot \boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\big) \Big]\Big)$$

whereas the right-hand side of (212) is nothing else but $\sup_{\mathbf{Q}\in\mathbf{\Omega}} \big(-\Phi(\mathbf{Q})\big) = -\inf_{\mathbf{Q}\in\mathbf{\Omega}} \Phi(\mathbf{Q})$. ■

Proof of bound (60)**.** Clearly, due to (16) and (20) it suffices to consider $\frac{\widetilde{c}}{\gamma} = 1$. The case $t = 0$ is clear:

$$\varphi_{\beta,\beta,1}(0) \;=\; \beta \cdot \Big\{ \sqrt{2} \,-\, 1 + \log\Big(2 \cdot \Big(\sqrt{2} \,-\, 1\Big)\Big)\Big\} \;<\; 1 \;=\; \varphi_1(0)\,.$$

So let $t \in]0,\infty[$. For each $\beta \in]0, \frac{8}{5}]$ we show the derivatives relations

$$0 \;>\; \varphi'_{\beta,\beta,1}(t) \;>\; \varphi'_1(t) \qquad \text{for all } t \in]0,1[, \tag{213}$$
$$0 \;=\; \varphi'_{\beta,\beta,1}(t) \;=\; \varphi'_1(t) \qquad t = 1, \tag{214}$$
$$0 \;<\; \varphi'_{\beta,\beta,1}(t) \;<\; \varphi'_1(t) \qquad \text{for all } t \in]1,\infty[, \tag{215}$$

Clearly, it suffices to consider $\beta = \frac{8}{5}$. By straightforward calculations one gets from (20) for all $t \in \mathbb{R}\backslash\{1\}$

$$\varphi'_{\beta,\beta,1}(t) = \frac{\beta}{1-t} \cdot \Big(1 - \sqrt{1 + (1-t)^2}\Big)$$

(with $\lim_{t\to 1} \varphi'_{\beta,\beta,1}(t) = 0$) and from (16) for all $t \in]0,\infty[$

$$\varphi'_1(t) = \log(t) = \log(1 + (t-1)).$$

Moreover, for fixed $t \in]0,1[$ its transform $x := t - 1 \in]-1,0[$ satisfies

$$\log(1+x) \;<\; \frac{2\cdot x}{2+x} \;<\; -\frac{8}{5\cdot x} \cdot \Big(1 - \sqrt{1+x^2}\Big);$$

indeed, the left-hand inequality is well-known (see e.g. Topsoe [375]) whereas the right-hand inequality follows from

$$\sqrt{1+x^2} \;<\; 1 + \frac{5\cdot x^2}{4\cdot(2+x)} \iff x^2\cdot(2+x)^2 \;<\; \frac{5\cdot x^2\cdot(2+x)}{2} + \frac{25\cdot x^4}{16} \iff 0 \;<\; \Big(\frac{3\cdot x}{4} - 1\Big)^2.$$

Analogously, one can show that for fixed $t \in]1,\infty[$ its transform $x := t - 1 \in]0,\infty[$ satisfies

$$\log(1+x) \;>\; \frac{2\cdot x}{2+x} \;\geq\; -\frac{8}{5\cdot x} \cdot \Big(1 - \sqrt{1+x^2}\Big)$$

where the right-hand inequality turns into an equality if and only of $x = 4/3$. Hence, we have shown (213),(214),(215) from which (60) follows immediately by deducing the monotonicity properties of the function $t \mapsto g(t) := \varphi_1(t) - \varphi_{\beta,\beta,1}(t)$ which satisfies $g(1) = 0$. ■

Proof of bound (58)**.** Let us arbitrarily fix $\gamma \in]1,\infty[$, $\beta \in]0,1[$ and $\widetilde{c} \in]0,\infty[$. For $t \in]0,\infty[$ the bound (58) follows immediately from (60) and (59). Furthermore, for each $\gamma \in]1,\infty[$ one gets from (16) and (20)

$$\varphi_{\beta,\beta,\frac{\widetilde{c}}{\gamma}}(0) \;=\; \frac{\widetilde{c}}{\gamma} \cdot \beta \cdot \Big\{ \sqrt{2} \,-\, 1 + \log\Big(2 \cdot \Big(\sqrt{2} \,-\, 1\Big)\Big)\Big\} \;<\; \frac{\widetilde{c}}{\gamma} \;=\; \widetilde{c} \cdot \varphi_\gamma(0)\,.$$

Moreover, by means of Example 26 one can even show for all $t \in]-\infty, 0]$

$$\varphi_{\beta,\beta,\frac{\widetilde{c}}{\gamma}}(t) \;<\; \frac{\widetilde{c}}{\gamma}\cdot(1-t) \;\leq\; \frac{\widetilde{c}}{\gamma}\cdot(\gamma\cdot\varphi_\gamma(t))$$

where the right-hand inequality follows immediately from (16). ■

Proof of Theorem 30. We can proceed analogously as in the above proof of Theorem 22, by replacing the use of Theorem 9 with the use of the analogous Theorem 16. For the sake of brevity, we omit the details. ■

Proof of Theorem 32. The assertions follow analogously to those of Theorem 22, by instead taking the auxiliary function $\widetilde{F}(\mathbf{Q}) := D_\varphi(\mathbf{Q},\mathbf{P}) + \Phi(\mathbf{Q})$. ■

Proof of Theorem 34. We can proceed analogously as in the proof of Theorem 32, by replacing the use of Theorem 9 with the use of the analogous Theorem 16. For the sake of brevity, we omit the details. ■

Proof of Theorem 49. From (30) we obtain straightforwardly for arbitrary $A > 0$, $\widetilde{c} > 0$ and $\mathbf{P} \in \mathbb{R}^K_{>0}$

$$D^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) := \begin{cases} \widetilde{c}\cdot\left\{\frac{m^\gamma\cdot H^{(1)}_\gamma}{\gamma\cdot(\gamma-1)} + \frac{H^{(2)}_\gamma}{\gamma} - \frac{m\cdot T_\gamma}{\gamma-1}\right\}, & \text{if } \gamma\in]-\infty,0[,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0},\ \mathbf{Q}\in A\cdot\mathbb{S}^K_{>0} \text{ and } m>0,\\ \widetilde{c}\cdot\left\{\breve{T}_0 - M_{\mathbf{P}}\cdot\log m + m\cdot T_0 - M_{\mathbf{P}}\right\}, & \text{if } \gamma=0,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0},\ \mathbf{Q}\in A\cdot\mathbb{S}^K_{>0} \text{ and } m>0,\\ \widetilde{c}\cdot\left\{\frac{m^\gamma\cdot H^{(1)}_\gamma}{\gamma\cdot(\gamma-1)} + \frac{H^{(2)}_\gamma}{\gamma} - \frac{m\cdot T_\gamma}{\gamma-1}\right\}, & \text{if } \gamma\in]0,1[,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0},\ \mathbf{Q}\in A\cdot\mathbb{S}^K \text{ and } m\geq 0,\\ \widetilde{c}\cdot\left\{A\cdot m\cdot\log m + m\cdot(I-A) + M_{\mathbf{Q}^{**}}\right\}, & \text{if } \gamma=1,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0},\ \mathbf{Q}\in A\cdot\mathbb{S}^K \text{ and } m\geq 0,\\ \widetilde{c}\cdot\left\{\frac{m^\gamma\cdot H^{(1)}_\gamma\cdot\mathbb{1}_{[0,\infty[}(m)}{\gamma\cdot(\gamma-1)} + \frac{H^{(2)}_\gamma}{\gamma} - \frac{m\cdot T_\gamma}{\gamma-1}\right\}, & \\ & \text{if } \gamma\in]1,2[,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0},\ \mathbf{Q}\in A\cdot\mathbb{S}^K \text{ and } m\in]-\infty,\infty[,\\ \widetilde{c}\cdot\left\{\frac{m^2\cdot H^{(1)}_2}{2} + \frac{H^{(2)}_2}{2} - m\cdot T_2\right\}, & \text{if } \gamma=2,\ \mathbf{Q}^{**}\in\mathbb{R}^K,\ \mathbf{Q}\in A\cdot\mathbb{S}^K \text{ and } m\in]-\infty,\infty[,\\ \widetilde{c}\cdot\left\{\frac{m^\gamma\cdot H^{(1)}_\gamma\cdot\mathbb{1}_{[0,\infty[}(m)}{\gamma\cdot(\gamma-1)} + \frac{H^{(2)}_\gamma}{\gamma} - \frac{m\cdot T_\gamma}{\gamma-1}\right\}, & \\ & \text{if } \gamma\in]2,\infty[,\ \mathbf{Q}^{**}\in\mathbb{R}^K_{>0},\ \mathbf{Q}\in A\cdot\mathbb{S}^K \text{ and } m\in]-\infty,\infty[,\\ \infty, & \text{else,} \end{cases} \tag{216}$$

where we have employed the five $m-$independent abbreviations

$$H^{(1)}_\gamma := H_\gamma(\mathbf{Q},\mathbf{P}) = \sum_{k=1}^K (q_k)^\gamma\cdot(p_k)^{1-\gamma}, \qquad H^{(2)}_\gamma := H_\gamma(\mathbf{Q}^{**},\mathbf{P}), \qquad \text{(cf. (83),(84))}$$

$$T_\gamma := T_\gamma(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P}) := \sum_{k=1}^K q_k\cdot(q^{**}_k)^{\gamma-1}\cdot(p_k)^{1-\gamma}, \qquad \text{(cf. (85))}$$

$$I := I(\mathbf{Q},\mathbf{Q}^{**}) := \sum_{k=1}^K q_k\cdot\log\left(\frac{q_k}{q^{**}_k}\right), \qquad \text{(cf. (86))}$$

$$\breve{T}_0 := \breve{T}_0(\mathbf{Q},\mathbf{Q}^{**},\mathbf{P}) := -\sum_{k=1}^K p_k\cdot\log\left(\frac{q_k}{q^{**}_k}\right). \qquad \text{(cf. (87))}$$

To proceed, let us fix an arbitrary constant $\widetilde{c} > 0$.
(i) Case $\gamma\cdot(1-\gamma)\neq 0$.
(ia) Let us start with the subcase $\gamma\in]-\infty,0[$. From the first and the last line of (216), it is clear that the corresponding $m-$infimum can not be achieved for $m\leq 0$; since $H^{(1)}_\gamma > 0$ and $T_\gamma > 0$, one gets the unique minimizer $m_{min} = \left(\frac{H^{(1)}_\gamma}{T_\gamma}\right)^{1/(1-\gamma)} > 0$ and the minimum

$$\inf_{m\neq 0} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) = D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(m_{min}\cdot\mathbf{Q},\mathbf{Q}^{**}) = \frac{\widetilde{c}}{\gamma}\cdot\left[H^{(2)}_\gamma - (T_\gamma)^{\gamma/(\gamma-1)}\cdot(H^{(1)}_\gamma)^{-1/(\gamma-1)}\right].$$

Thus, (88) is established and from there, (89) follows by means of (82).
(ib) The subcase $\gamma\in]0,1[$ (cf. the third line of (216)) works analogously to the subcase (ia) since $H^{(1)}_\gamma > 0$ and $T_\gamma > 0$ due to the fact that $\mathbf{Q}\in A\cdot\mathbb{S}^K$ implies that all the components q_k are nonnegative and that at least one component q_k must be strictly positive.

(ic) In the subcase $\gamma \in]1,2[\cup]2,\infty[$ (cf. the fifth and seventh line of (216)) it is straightforward to see that the desired infimum can not be achieved for $m<0$. Hence, one can proceed analogously to subcase (ia).
(id) Let $\gamma=2$, $\mathbf{Q}^{**}\in\mathbb{R}^K$ and $\mathbf{Q}\in A\cdot\mathbb{S}^K$ for some $A\in\mathbb{R}\backslash\{0\}$. The latter implies that either "all components q_k are nonnegative and at least one component is strictly positive" (if $A>0$) or "all components q_k are nonpositive and at least one component is strictly negative" (if $A<0$). There holds $H_2^{(1)}>0$, $H_2^{(2)}\geq 0$ and $T_2\in]-\infty,\infty[$. For $T_\gamma\neq 0$, we obtain the unique minimizer $m_{min}=\frac{T_2}{H_2^{(1)}}\neq 0$ and the minimum

$$\inf_{m\neq 0} D^{SBD}_{\widetilde{c}\cdot\varphi_2,\mathbf{P}}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) = D^{SBD}_{\widetilde{c}\cdot\varphi_2,\mathbf{P}}(m_{min}\cdot\mathbf{Q},\mathbf{Q}^{**}) = \frac{\widetilde{c}}{2}\cdot\Big[H_2^{(2)}-(T_2)^2\cdot(H_2^{(1)})^{-1}\Big]. \tag{217}$$

In the subcase $T_2=0$, one arrives immediately at

$$\inf_{m\neq 0} D^{SBD}_{\widetilde{c}\cdot\varphi_2,\mathbf{P}}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) = \inf_{m\neq 0}\frac{\widetilde{c}}{2}\cdot\Big[m^2\cdot H_2^{(1)}+H_2^{(2)}\Big] = \frac{\widetilde{c}}{2}\cdot H_2^{(2)}$$

which is the corresponding special case of (217).
(ii) Case $\gamma=1$. From the fourth and the last line of (216), it is clear that the corresponding $m-$infimum can not be achieved for $m\leq 0$. We obtain the unique minimizer $m_{min}=\exp\{-I/A\}>0$ and the minimum

$$\inf_{m\neq 0} D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) = D^{SBD}_{\widetilde{c}\cdot\varphi_1,\mathbf{P}}(m_{min}\cdot\mathbf{Q},\mathbf{Q}^{**}) = \widetilde{c}\cdot\Big[M_{\mathbf{Q}^{**}}-A\cdot\exp\Big(-\frac{I}{A}\Big)\Big].$$

Thus, (90) is established and from there, (91) follows by means of (82).
(iii) Case $\gamma=0$. From the second and the last line of (216), it is clear that the corresponding $m-$infimum can not be achieved for $m\leq 0$. By straightforward calculations, we deduce the unique minimizer $m_{min}=\frac{M_{\mathbf{P}}}{T_0}>0$ and the corresponding minimum

$$\inf_{m\neq 0} D^{SBD}_{\widetilde{c}\cdot\varphi_0,\mathbf{P}}(m\cdot\mathbf{Q},\mathbf{Q}^{**}) = D^{SBD}_{\widetilde{c}\cdot\varphi_0,\mathbf{P}}(m_{min}\cdot\mathbf{Q},\mathbf{Q}^{**}) = \widetilde{c}\cdot\Big[M_{\mathbf{P}}\cdot\log(T_0)+\breve{T}_0-M_{\mathbf{P}}\cdot\log(M_{\mathbf{P}})\Big].$$

Thus, (92) is proved and from there, (93) can be derived by means of (82). ■

Proof of Theorem 52. We follow the lines of the above proof of Theorem 22. It suffices to prove part (b). Let us define the auxiliary function $\widetilde{F}(\mathbf{Q}):=F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\Big)-\Phi(\mathbf{Q})$ which by (106) satisfies $\widetilde{F}(\mathbf{Q})\leq c_1$ for all $\mathbf{Q}\in A\cdot\mathfrak{Q}$. Moreover, by rewriting (75) of Theorem 44 we have

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{\Pi}\Big[A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}\in A\cdot\mathfrak{Q}\Big] = -\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}} F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\Big).$$

By applying the Varadhan Lemma to the sequence $(\mu_n)_{n\in\mathbb{N}}$ of probability distributions on $\mathbb{R}^K$ (equipped with the corresponding Borel $\sigma-$field) given by $\mu_n[\,\cdot\,]:=\mathbb{\Pi}\big[A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}\in\,\cdot\,\big]$ we obtain (with a slight abuse of notation)

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\int_{\mathbb{R}^K}\exp(n\cdot\widetilde{F}(\mathbf{Q}))\cdot 1_{A\cdot\mathfrak{Q}}(\mathbf{Q})\,\mathrm{d}\mu_n(\mathbf{Q})\right) = \sup_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\Big(\widetilde{F}(\mathbf{Q})\;-\;F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\Big)\Big). \tag{218}$$

As usual, the left-hand side of (218) can be equivalently rewritten as

$$\lim_{n\to\infty}\frac{1}{n}\log\Big(\;\mathbb{E}_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\widetilde{F}\big(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\Big)\cdot 1_{A\cdot\mathfrak{Q}}\big(A\cdot\boldsymbol{\xi}_n^{w\mathbf{W}}\big)\;\Big]\Big)$$

whereas the right-hand side of (218) is nothing else but $\sup_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\big(-\Phi(\mathbf{Q})\big)=-\inf_{\mathbf{Q}\in A\cdot\mathfrak{Q}}\Phi(\mathbf{Q})$. ■

Proof of Theorem 54. The assertions follow analogously to those of Theorem 52, by instead taking the auxiliary function $\widetilde{F}(\mathbf{Q}):=F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\Big)+\Phi(\mathbf{Q})$. ■

Proof of Theorem 55. We can proceed analogously as in the proof of Theorem 52, by replacing the use of (75) of Theorem 44 with the use of the analogous (89) (respectively (91) respectively (93)) of Theorem 49. In particular, $F_{\gamma,\widetilde{c},A}\Big(D_{\widetilde{c}\cdot\varphi_\gamma}(\mathbf{Q},\mathbb{P})\Big)$ is replaced by $\breve{D}^{SBD}_{\widetilde{c}\cdot\varphi_\gamma,\mathbf{P}}(\mathbf{Q},\mathbf{Q}^{**})$. For the sake of brevity, we omit the details. ■

Proof of Theorem 58. The assertions follow analogously to those of Theorem 54, by employing the changes indicated in the above proof of Theorem 55. For the sake of brevity, we omit the details. ■

Proof of Theorem 65. Recall from (22) and (23) that we can represent the involved scaled Bregman distances as $D^{SBD}_{\varphi,\mathbb{P}}(\mathbb{Q},\mathbf{Q}^{**}) = \sum_{k=1}^K p_k \cdot \varphi_k\Big(\frac{q_k}{p_k}\Big)$, where for each $k = 1,\ldots,K$ one has

$$\varphi_k(t) \;=\; \sup_{z\in\mathbb{R}}\Big(z\cdot t - \log\int\limits_{\mathbb{R}} e^{z\cdot y}dU_k(y)\Big), \qquad t\in\mathbb{R},$$

cf. (209) with omittance of the tildes since $\widetilde{\varphi}_k(\cdot) := M_{\mathbb{P}}\cdot\varphi_k(\cdot) = \varphi_k(\cdot)$ and $\widetilde{\mathfrak{l}} = \mathfrak{l}$; moreover, let $V := (\mathbf{V}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors constructed (with the help of $\mathfrak{l}$) via (25) and (26) (without tildes). Moreover, recall the vector

$$\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}} \;:=\; \begin{cases} \Big(\frac{\sum_{i\in I_1^{(n)}(\mathbf{X}_1^n)} V_i}{\sum_{k=1}^K\sum_{i\in I_k^{(n)}(\mathbf{X}_1^n)} V_i},\ldots,\frac{\sum_{i\in I_K^{(n)}(\mathbf{X}_1^n)} V_i}{\sum_{k=1}^K\sum_{i\in I_k^{(n)}(\mathbf{X}_1^n)} V_i(\mathbf{X}_1^n)}\Big), & \text{if } \sum_{j=1}^n V_j\neq 0,\\ (\infty,\ldots,\infty) =: \boldsymbol{\infty}, & \text{if } \sum_{j=1}^n V_j = 0,\end{cases} \qquad \text{(cf. (130))}$$

and define

$$\boldsymbol{\xi}^{\mathbf{V}}_{n,\mathbf{X}} := \Big(\frac{1}{n}\sum_{i\in I_1^{(n)}(\mathbf{X}_1^n)} V_i,\ldots,\frac{1}{n}\sum_{i\in I_K^{(n)}(\mathbf{X}_1^n)} V_i\Big)$$

where the right-hand side structurally coincides with (27) (without tildes) but however — as explained above — the construction of the involved $I_k^{(n)}$ differs which is reflected in different indexing on the corresponding left-hand sides. Recall that inside the block $I_k^{(n)}$, the V_i's are independent with the same distribution U_k (cf. (26)) (depending on the block). With the help of these ingredients, to prove (131) we can follow nearly verbatim our proof of Theorem 12 in Broniatowski & Stummer [1]; in the following, for the sake of brevity we explain only major steps. Firstly, we show for $m\neq 0$ that

$$\Big\{\boldsymbol{\xi}^{\mathbf{V}}_{n,\mathbf{X}}\in m\cdot\boldsymbol{\mathfrak{M}}\Big\}\subset\Big\{\frac{1}{n}\sum_{i=1}^n V_i = m\Big\}. \tag{219}$$

To see this, $\boldsymbol{\xi}^{\mathbf{V}}_{n,\mathbf{X}}\in m\cdot\boldsymbol{\mathfrak{M}}$ means that there exists a probability vector $\mathbb{Q} := \{q_1,\ldots,q_K\}$ such that $\boldsymbol{\xi}^{\mathbf{V}}_{n,\mathbf{X}} = m\cdot\mathbb{Q}$ and thus the sum of the components of $\boldsymbol{\xi}^{\mathbf{V}}_{n,\mathbf{X}}$ (which is equal to $\frac{1}{n}\sum_{i=1}^n V_i$) must be m. To proceed, (219) implies

$$\frac{1}{n}\log\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}\in\boldsymbol{\mathfrak{M}}\Big] = \frac{1}{n}\log\mathbb{\Pi}_{\mathbf{X}_1^n}\Big[\boldsymbol{\xi}^{\mathbf{V}}_{n,\mathbf{X}}\in\bigcup_{m\neq 0} m\cdot\boldsymbol{\mathfrak{M}}\Big]$$

(analogously to the proof of equality (172) in [1]). Secondly, in the present context the Lemma 42 of [1] still holds, and the same is true for inequality (175) of [1] with $\Phi_{\mathbb{P}}(\cdot) := D^{SBD}_{\varphi,\mathbb{P}}(\cdot,\mathbf{Q}^{**})$ as well as for inequality (173) of [1] with $\mathbf{V}$ instead of $\mathbf{W}$ (where we employ the corresponding part of the proof of Theorem 16 (of the current paper) instead of Proposition 39 of [1]). From this, (131) (of the current paper) follows by considerations which are analogous to those in the last paragraph of the proof of Theorem 12 in [1]. ■

Proof of Theorem 46. The corresponding assertions follow analogously to the above proof of Theorem 65, by taking $\boldsymbol{\xi}^{w\mathbf{V}}_n$ instead of $\boldsymbol{\xi}^{w\mathbf{V}}_{n,\mathbf{X}}$, and $\mathbb{\Pi}$ instead of $\mathbb{\Pi}_{\mathbf{X}_1^n}$. ■

Proof of Proposition 72. We get, by the definition of $\Phi\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}\Big)$ and (162) that

$$\begin{aligned}\widehat{\Phi(\boldsymbol{\Omega})}^{naive,1}_{n,L} \;&=\; -\frac{1}{n}\log\frac{1}{L}\sum_{\ell=1}^L\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}},\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}\big)\Big)\Big)\cdot\mathbf{1}_{\boldsymbol{\Omega}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}\Big)\\ &\geq\; \Phi\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{L,*}}\Big) \;-\; \frac{1}{n}\log\frac{1}{L}\sum_{\ell=1}^L\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}},\mathbf{P}\big)\Big)\Big)\cdot\mathbf{1}_{\boldsymbol{\Omega}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}\Big)\;.\end{aligned}$$

For fixed n, by the strong law of large numbers there holds a.s.

$$\begin{aligned}&\lim_{L\to\infty}\frac{1}{L}\sum_{\ell=1}^L\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}},\mathbf{P}\big)\Big)\Big)\cdot\mathbf{1}_{\boldsymbol{\Omega}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}^{(\ell)}}\Big)\\ &=\; E_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)\Big)\Big)\cdot\mathbf{1}_{\boldsymbol{\Omega}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\Big)\Big].\end{aligned}$$

Since we have assumed that $\boldsymbol{\Omega}$ is compact, we get from the application of Theorem 22(a) to the constant null function $\breve{\phi}(\cdot)\equiv 0$ on $\boldsymbol{\Omega}$ that

$$\lim_{n\to\infty}\frac{1}{n}\log E_{\mathbb{\Pi}}\Big[\exp\Big(n\cdot\Big(D_\varphi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}},\mathbf{P}\big)\Big)\Big)\cdot\mathbf{1}_{\boldsymbol{\Omega}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{W}}}\Big)\Big] \;=\; 0,$$

which completes the proof of (165). ■

Proof of Proposition 85. First, we apply the following representation

$$\frac{\mathrm{d}\breve{\zeta}}{\mathrm{d}\breve{\mathbb{S}}}\Big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\Big) \;=\; \exp\Big(\sum_{k=1}^{K}\Big(n_k\cdot\Lambda_{\widetilde{\zeta}}(\tau_k)-\tau_k\cdot\sum_{i\in I_k^{(n)}}\widetilde{\mathbf{V}}_i\Big)\Big) \tag{220}$$

(cf. (114) in Subsection X-A of Broniatowski & Stummer [1]). By plugging (220) into (179), the involved expectation becomes

$$\mathbb{E}_{\mathbb{\Pi}}\Big[\,\mathbb{1}_{\boldsymbol{\Omega}}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\cdot\exp\Big(n\cdot\Big(D_{\varphi}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}},\mathbf{P}\big)-\Phi\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}\big)\Big)\Big)\cdot\exp\Big(\sum_{k=1}^{K}\Big(n_k\cdot\Lambda_{\widetilde{\zeta}}(\tau_k)-\tau_k\cdot\sum_{i\in I_k^{(n)}}\widetilde{\mathbf{V}}_i\Big)\Big)\,\Big].$$

Moreover,

$$\begin{aligned}
&D_{\varphi}\big(M_{\mathbf{P}}\cdot\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}},\mathbf{P}\big)+\frac{1}{n}\cdot\sum_{k=1}^{K}\Big(n_k\cdot\Lambda_{\widetilde{\zeta}}(\tau_k)-\tau_k\cdot\sum_{i\in I_k^{(n)}}\widetilde{\mathbf{V}}_i\Big)=D_{\widetilde{\varphi}}(\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}},\widetilde{\mathbb{P}})+\sum_{k=1}^{K}\frac{n_k}{n}\cdot\Lambda_{\widetilde{\zeta}}(\tau_k)-\sum_{k=1}^{K}\tau_k\cdot\frac{1}{n}\sum_{i\in I_k^{(n)}}\widetilde{\mathbf{V}}_i\\
&\quad=\; D_{\widetilde{\varphi}}(\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}},\widetilde{\mathbb{P}})+\sum_{k=1}^{K}\widetilde{p}_k\cdot\Lambda_{\widetilde{\zeta}}(\tau_k)-\sum_{k=1}^{K}\widetilde{p}_k\cdot\tau_k\cdot\frac{1}{\widetilde{p}_k}\cdot\frac{1}{n}\sum_{i\in I_k^{(n)}}\widetilde{\mathbf{V}}_i \qquad (221)\\
&\quad=\;\sum_{k=1}^{K}\widetilde{p}_k\cdot\Big[\,\widetilde{\varphi}\Big(\frac{\widetilde{x}_k}{\widetilde{p}_k}\Big)+\Lambda_{\widetilde{\zeta}}(\tau_k)-\tau_k\cdot\frac{\widetilde{x}_k}{\widetilde{p}_k}\,\Big]\;=\;\sum_{k=1}^{K}\widetilde{p}_k\cdot\widetilde{\varphi}_k\Big(\frac{\widetilde{x}_k}{\widetilde{p}_k}\Big)\;=\;D^{SBD}_{\widetilde{\varphi},\widetilde{\mathbb{P}}}(\widetilde{\mathbf{x}},\widetilde{\mathbf{Q}}^*)\;=\;D^{SBD}_{\varphi,\mathbf{P}}(\mathbf{x},\mathbf{Q}^*) \qquad (222)
\end{aligned}$$

which finishes the proof of Proposition 85. In the above display, in (221) we have used the assumption $\frac{n_k}{n}=\widetilde{p}_k$ and in the first equality of (222) we have employed the divergence definition (4) as well as the abbreviation $\widetilde{\mathbf{x}}:=\boldsymbol{\xi}_n^{\widetilde{\mathbf{V}}}$ which by construction leads to the corresponding components $\widetilde{x}_k=\frac{1}{n}\sum_{i\in I_k^{(n)}}\widetilde{\mathbf{V}}_i$; moreover, the second equality in (222) follows — in terms of $\widetilde{\varphi}_k:=M_{\mathbf{P}}\cdot\varphi_k$ — from (211) in the above proof of Theorem 16, the third equality comes from (the tilted version of) the divergence definition (23), and the last equality is nothing but the divergence rewritability (24). ■

ACKNOWLEDGMENT

W. Stummer is grateful to the Sorbonne Université Paris for its multiple partial financial support and especially to the LPSM for its multiple great hospitality. M. Broniatowski thanks very much the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) for its partial financial support and hospitality.

REFERENCES

[1] M. Broniatowski and W. Stummer, "A precise bare simulation approach to the minimization of some distances. I. Foundations," *IEEE Trans. Inf. Theory*, Vol. 69, no. 5, pp. 3062–3120, 2023.
[2] I. Csiszár, "Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten," *Publ. Math. Inst. Hungar. Acad. Sci.*, Vol. 8, pp. 85–108, 1963.
[3] M.S. Ali and D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Roy. Statist. Soc. Series B (Methodological)*, Vol. 28, no. 1, pp. 131–140, 1966.
[4] T. Morimoto, "Markov processes and the H-theorem," *J. Phys. Soc. Jpn.*, Vol. 18, no. 3, pp. 328–331, 1963.
[5] N. Cressie and T.R.C. Read, "Multinomial goodness-of-fit tests," *J. R. Statist. Soc. B*, Vol. 46, no. 3, pp. 440–464, 1984.
[6] T.R.C. Read and N.A.C. Cressie, *Goodness-of-Fit Statistics for Discrete Multivariate Data*. New York, USA: Springer, 1988.
[7] C. Tsallis, "Generalized entropy-based criterion for consistent testing," *Phys. Rev. E*, Vol. 58, No. 2, pp. 1442–1445, 1998.
[8] S.-I. Amari, *Differential-Geometrical Methods in Statistics*. Berlin, Germany: Springer, 1985.
[9] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, Vol. 35, pp. 99–109, 1943.
[10] A. Bhattacharyya, "On a measure of divergence between two multinomial populations," *Sankhya*, Vol. 7, no. 4, pp. 401–406, 1946.
[11] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation (contd.)," *Sankhya*, Vol. 8, no. 3, pp. 201–218, 1947.
[12] A. Renyi, "On measures of entropy and information," in: J. Neyman (ed.), *Proc. 4th Berkeley Symp. Math. Stat. Probab. Vol.1*, pp. 547–561. Berkeley, CA, USA: Univ. of California Press, 1961.
[13] T. van Erven and P. Harremoës, "Renyi divergence and Kullback-Leibler divergence," *IEEE Trans. Inf. Theory*, Vol. 60, no. 7, pp. 3797–3820, 2014.
[14] C. Burbea and C.R. Rao, "On the convexity of some divergence measures based on entropy functions," *IEEE Trans. Inf. Theory*, Vol. 28, no. 3, pp. 489–495, 1982.
[15] C.E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27, no. 3, pp. 379–423, 1948.
[16] J. Havrda and F. Charvat, "Quantification method of classification process," *Kybernetika*, Vol. 3, pp. 30–34, 1967.
[17] C. Tsallis, "Possible generalization of Boltzmann-Gibbs Statistics," *Journal of Statistical Physics*, Vol. 52, no. 1/2, pp. 479–487, 1988.
[18] F. Liese and I. Vajda, *Convex Statistical Distances*. Leipzig, Germany: Teubner, 1987.
[19] I. Vajda, *Theory of Statistical Inference and Information*. Dordrecht, NL: Kluwer, 1989.
[20] I. Csiszár and P.C. Shields, *Information Theory and Statistics: a Tutorial*. Hanover, MA, USA: now Publishers, 2004.
[21] W. Stummer, *Exponentials, Diffusions, Finance, Entropy and Information*. Aachen, Germany: Shaker, 2004.
[22] L. Pardo, *Statistical Inference Based on Divergence Measures*. Boca Raton, USA: Chapman & Hall/CRC, 2006.
[23] F. Liese and K.J. Miescke, *Statistical Decision Theory: Estimation, Testing, and Selection*. New York, USA: Springer, 2008.
[24] A. Basu, H. Shioya and C. Park, *Statistical Inference: The Minimum Distance Approach*. Boca Raton, USA: CRC Press, 2011.

[25] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Trans. Inf. Theory*, Vol. 52, no. 10, pp. 4394–4412, 2006.

[26] I. Vajda and E.C. van der Meulen, "Goodness-of-fit criteria based on observations quantized by hypothetical and empirical percentiles," in: Z.A. Karian ZA and E.J. Dudewicz (eds.), *Handbook of Fitting Statistical Distributions with R*, pp. 917 – 994. Heidelberg, Germany: CRC, 2010.

[27] M.D. Reid and R.C. Williamson, "Information, divergence and risk for binary experiments," *J. Machine Learn. Res.*, Vol. 12, pp. 731–817, 2011.

[28] M. Basseville, "Divergence measures for statistical data processing - an annotated bibliography," *Signal Process.*, Vol. 93, pp. 621–633, 2013.

[29] I. Csiszár, "A class of measures of informativity of observation channels," *Periodica Mathem. Hungar.*, Vol. 2, no. 1-4, pp. 191–213, 1972.

[30] M. Ben-Bassat, "f-entropies, probability of error, and feature selection," *Information and Control*, Vol. 39, pp. 227–242, 1978.

[31] A. Ben-Tal and M. Teboulle, "Rate-distortion theory with generalized information measures via convex programming duality," *IEEE Trans. Inf. Theory*, Vol. 32, no. 5, pp. 630–641, 1986.

[32] H.K. Kesavan and J.N. Kapur, "The generalized maximum entropy principle," *IEEE Trans. Syst. Man Cyb.*, Vol. 19, no. 5, pp. 1042–1052, 1989.

[33] D. Dacunha-Castelle and F. Gamboa, "Maximum d'entropie et probleme des moments," *Ann. Inst. Henri Poincare*, Vol. 26, no. 4, pp. 567–596, 1990.

[34] M. Teboulle and I. Vajda, "Convergence of best $\phi-$entropy estimates," *IEEE Trans. Inf. Theory*, Vol. 39, no. 1, pp. 297–301, 1993.

[35] F. Gamboa and E. Gassiat, "Asymptotic distribution of $(h,\phi)-$entropies," *Ann. Stat.*, Vol. 25, no. 1, pp. 328–350, 1997.

[36] I. Vajda and J. Zvarova, "On generalized entropies, Bayesian decisions and statistical diversity," *Kybernetika*, Vol. 43, no. 5, pp. 675–696, 2007.

[37] M. Salicru, M.L. Menendez, D. Morales and L. Pardo, "Asymptotic distribution of $(h,\phi)-$entropies," *Commun. Statist. - Theory Meth.*, Vol. 22, no. 7, pp. 2015–2031, 1993.

[38] I. Vajda and K. Vasek, "Majorizations, concave entropies, and comparison of experiments," *Problems of Control and Information Theory*, Vol. 14, no. 2, pp. 105–115, 1985.

[39] L.M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Comput. Math. Math. Phys.*, Vol. 7, no. 3, pp. 200–217, 167.

[40] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Ann. Statist.*, Vol. 19, no. 4, pp. 2032–2066, 1991.

[41] I. Csiszár, "Maximum entropy and related methods," in: *Trans. 12th Prague Conf. Information Theory, Statistical Decision Functions and Random Processes*, pp. 58–62. Prague, Czech Republic: Czech Acad. Sci., 1994.

[42] I. Csiszár, "Generalized projections for non-negative functions," *Acta Math. Hung.*, Vol. 68, pp. 161–186, 1995.

[43] M.C. Pardo and I. Vajda, "About distances of discrete distributions satisfying the data processing theorem of information theory," *IEEE Trans. Inf. Theory*, Vol. 43, no. 4, pp. 1288–1293, 1997.

[44] M.C. Pardo and I. Vajda, "On asymptotic properties of information-theoretic divergences," *IEEE Trans. Inf. Theory*, Vol. 49, no. 7, pp. 1860–1868, 2003.

[45] W. Stummer and I. Vajda, "On Bregman distances and divergences of probability measures," *IEEE Trans. Inf. Theory*, Vol. 58, no. 3, pp. 1277–1288, 2012.

[46] M. Broniatowski and W. Stummer, "Some universal insights on divergences for statistics, machine learning and artificial intelligence," in: F. Nielsen (ed.), *Geometric Structures of Information*, Ser. Signals and Communications Technology, pp. 149–211. Cham, Switzerland: Springer Nature Switzerland, 2019.

[47] M. Broniatowski and W. Stummer, "A unifying framework for some directed distances in statistics," in: F. Nielsen, A.S.R.S. Rao, C.R. Rao (eds.), *Geometry and Statistics*, Handbook of Statistics, Vol. 46, pp. 145–223. Cambrigde MA, USA: Academic Press, 2022.

[48] A. Basu, I.R. Harris, N.L. Hjort and M.C. Jones, "Robust and efficient estimation by minimizing a density power divergence," *Biometrika*, Vol. 85, no. 3, pp. 549–559, 1998.

[49] A. Banerjee, X. Guo and H. Wang, "On the Optimality of Conditional Expectation as a Bregman Predictor," *IEEE Trans. Inf. Theory*, Vol. 51, no. 7, pp. 2664–2669, 2005.

[50] S. Eguchi and Y. Kano, "Robustifing Maximum Likelihood Estimation by Psi-divergence," *Research Memorandum 802*, 2001. Tokyo: Institute of Statistical Mathematics.

[51] M. Mihoko and S. Eguchi, "Robust blind source separation by beta divergence," *Neural Comput.*, Vol. 14, pp. 1859–1886, 2002.

[52] T. Mukherjee, A. Mandal and A. Basu, "The B-exponential divergence and its generalizations with applications to parametric estimation," *Stat. Methods Appl.*, Vol. 28, pp. 241—257, 2019.

[53] S. Basak and A. Basu, "The extended Bregman divergence and parametric estimation," *Statistics*, Vol. 56, no. 3, pp. 699—718, 2022.

[54] W. Stummer, "Some Bregman distances between financial diffusion processes," *Proc. Appl. Math. Mech.*, Vol. 7, no. 1, pp. 1050503–1050504, 2007.

[55] A.-L. Kißlinger and W. Stummer, "Some decision procedures based on scaled Bregman distance surfaces," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2013*, Lecture Notes in Computer Science, vol. 8085, pp. 479–486. Berlin, Germany: Springer, 2013.

[56] A.-L. Kißlinger and W. Stummer, "New model search for nonlinear recursive models, regressions and autoregressions," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2015*, Lecture Notes in Computer Science, vol. 9389, pp. 693–701. Berlin, Germany: Springer, 2015.

[57] A.-L. Kißlinger and W. Stummer, "Robust statistical engineering by means of scaled Bregman distances," in: C. Agostinelli, A. Basu, P. Filzmoser and D. Mukherjee (eds.), *Recent Advances in Robust Statistics – Theory and Applications*, pp. 81–113. New Delhi, India: Springer, 2016.

[58] A.-L. Kißlinger and W. Stummer, "A new toolkit for robust distributional change detection," *Appl. Stochastic Models Bus. Ind.*, Vol. 34, pp. 682–699, 2018.

[59] W. Stummer and A.-L. Kißlinger, "Some new flexibilizations of Bregman divergences and their asymptotics," In: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2017*, Lecture Notes in Computer Science, vol. 10589, pp. 514–522. Cham, Switzerland: Springer International Publishing, 2017.

[60] M. Liu, B.C. Vemuri, S.-I. Amari and F. Nielsen, "Total Bregman divergence and its applications to shape retrieval," in: *Proc. 23rd IEEE CVPR*, pp. 3463–3468 , 2010.

[61] M. Liu, B.C. Vemuri, S.-I. Amari and F. Nielsen, "Shape retrieval using hierarchical total Bregman soft clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 34, no. 12, pp. 2407–2419, 2012.

[62] B.C. Vemuri, M. Liu, S.-I. Amari and F. Nielsen, "Total Bregman divergence and its applications to DTI analysis," *IEEE Trans. Med. Imag.*, Vol. 30, no. 2, pp. 475–483, 2011.

[63] R. Nock, A.K. Menon and C.S. Ong, "A scaled Bregman theorem with applications," in: *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, 2016, 9 pages.

[64] R. Nock, F. Nielsen and S.-I. Amari, "On conformal divergences and their population minimizers," *IEEE Trans. Inf. Theory*, Vol. 62, no. 1, pp. 527–538, 2016.

[65] P.C. Mahalanobis, "On the generalized distance in statistics," *Proc. Nat. Inst. Sci. Indida*, Vol. 2, no. 1, pp. 49–55, 1936.

[66] W. Stummer and I. Vajda, "On divergences of finite measures and their applicability in statistics and information theory," *Statistics*, Vol. 44, no. 2, pp. 169–187, 2010.

[67] C. Gietl and and F.P. Reffel, "Continuity of $f-$projections and applications to the iterative proportional fitting procedure," *Statistics*, Vol. 51, no. 3, pp. 668–684, 2017.

[68] M. Broniatowski and A. Keziou, "Minimization of ϕ-divergences on sets of signed measures," *Stud. Scient. Math. Hungar.*, Vol. 43, pp. 403–442, 2006.

[69] M. Broniatowski and W. Stummer, "On a cornerstone of bare-simulation distance/divergence optimization," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2023, Part I*, Lecture Notes in Computer Science, vol. 14071, pp. 105–116. Cham, Switzerland: Springer Nature Switzerland, 2023.

[70] I. Csiszár, "Sanov property, generalized I-projection and a conditional limit theorem," *Ann. Probab.*, Vol. 12, no. 3, pp. 768–793, 1984.
[71] R.R. Tucci, "Method for sampling probability distributions using a quantum computer," *United States Patent*, Patent No. US 8543627 B1, 24th Sep. 2013.
[72] J.S. Teh, A. Samsudin, M. Al-Mazrooie and A. Akhavan, "GPUs and chaos: a new true random number generator," *Nonlinear Dyn.*, Vol. 82, pp. 1913–1922, 2015.
[73] C. Aghamohammadi and J.P. Crutchfield, "Thermodynamics of random number generation," *Phys. Rev. E*, Vol. 95, pp. 062139-1–062139-11, 2017.
[74] M. Herrero-Collantes and J.C. Garcia-Escartin, "Quantum random number generators," *Rev. Mod. Phys.*, Vol. 89, no. 1, pp. 015004-1 – 015004-48, 2017.
[75] K.A. Balygin, V.I. Zaitsev, A.N. Klimov, S.P. Kulik and S.N. Molotkov, "A quantum random number generator based on the 100-Mbit/s Poisson photocount statistics," *J. Exp. Theor. Phys.*, Vol. 126, no. 6, pp. 728–740, 2018.
[76] B. Dang, J. Sun, T. Zhang, S. Wang, M. Zhao, K. Liu, L. Xu, J. Zhu , C. Cheng, L. Bao, Y. Yang, H. Wang, Y. Hao and R. Huang, "Physically transient true random number generators based on paired threshold switches enabling Monte Carlo method applications," *IEEE Electron. Device Lett.*, Vol. 40, no. 7, pp. 1096–1099, 2019.
[77] L. Gong, J. Zhang, H. Liu, L. Sang and Y. Wang, "True random number generators using electrical noise," *IEEE Access*, Vol. 7, pp. 125796–125805, 2019.
[78] S.T. Chandrasekaran, V.E.G. Karnam and A. Sanya, "0.36-mW, 52-Mbps true random number generator based on a stochastic delta-sigma modulator," *IEEE Solid-State Lett.*, Vol. 3, pp. 190–193, 2020.
[79] D. Drahi, N. Walk, M.J. Hoban, A.K. Fedorov, R. Shakhovoy, A. Feimov, Y. Kurochkin, W.S. Kolthammer, J. Nunn, J. Barrett and I.A. Walmsley, "Certified quantum random numbers from untrusted light," *Phys. Rev. X*, Vol. 10, pp. 041048-1 – 041048-32, 2020.
[80] C. Kollmitzer, S. Schauer, S. Rass and B. Rainer (eds.), *Quantum Random Number Generation*. Cham, Switzerland: Springer Nature, 2020.
[81] Y. Liu, C. Chen, D.D. Yang, Q. Li and X. Li, "Fast true random number generator based on chaotic oscillation in self-feedback weakly coupled superlattices," *IEEE Access*, Vol. 8, pp. 182693–182703, 2020.
[82] T. Arciuolo and K.M. Elleithy, "Parallel, true random number generator (P-TRNG): using parallelism for fast true random number generation in hardware," in: *Proc. 2021 IEEE 11th Ann. Comput. and Commun. Workshop Conf. (CCWC)*, 0987–0992, 2021; doi:10.1109/CCWC51732.2021.9375939.
[83] A.M. Awaludin, D. Pratama and H. Kim, "AnyTRNG: generic, high-throughput, low-area true random number generator based on synchronous edge sampling," in: H. Kim (ed.), *Information Security Applications WISA 2021*, Lecture Notes in Computer Science, vol. 13009, pp. 157—168. Cham, Switzerland: Springer Nature Switzerland, 2021.
[84] B. Bai, J. Huang, G.-R. Qiao, Y.-Q. Nie, W. Tang, T. Chu, J. Zhang and J.-W. Pan, "18.8 Gbps real-time quantum random number generator with a photonic integrated chip," *Appl. Phys. Lett.*, Vol. 118, No. 264001, 2021; doi:10.1063/5.0056027.
[85] G. Cao, L. Zhang, X. Huang, W. Hu and X. Yang, "16.8 Tb/s True Random Number Generator Based on Amplified Spontaneous Emission," *IEEE Photonics Technol. Lett.*, Vol. 33, no. 14, pp. 699–702, 2021.
[86] S.T. Chandrasekaran, A. Jayaraj, N. Ramesh and A. Sanyal, "33-200Mbps, 3pJ/bit true random number generator based on CT Delta-Sigma modulator," in: *Proc. 2021 IEEE Intern. Symp. Circuits Systems (ISCAS)*, 2021; doi:10.1109/ISCAS51556.2021.9401507.
[87] A. Degada and H. Thapliyal, "An integrated TRNG-PUF architecture based on photovoltaic solar cells," *IEEE Consumer Electron. Magaz.*, Vol. 10, no. 4, pp. 99–105, 2021.
[88] I. Fischer and D.J. Gauthier, "High-speed harvesting of random numbers," *Science*, Vol. 371, 26 February 2021, pp. 889–890, 2021.
[89] Z. Fu, Y. Tang, X. Zhao, K. Lu, Y. Dong, A. Shukla, Z. Zhu and Y. Yang, "An overview of spintronic true random number generator," *Front. Phys.*, Vol. 9, No. 638207, 2021; doi:10.3389/fphy.2021.638207.
[90] T. Gehring, C. Lupo, A. Kordts, D.S. Nikolic, N. Jain, T. Rydberg, T.B. Pedersen, S. Pirandola and U.L. Andersen, "Homodyne-based quantum random number generator at 2.9 Gbps secure against quantum side-information," *Nature Comm.*, Vol. 12, No. 605, 2021; doi:10.1038/s41467-020-20813-w.
[91] G. Gras, A. Martin, J.W. Choi and F. Bussières, "Quantum entropy model of an integrated quantum-random-number-generator chip," *Phys. Rev. Applied*, Vol. 15, No. 054048, 2021; doi:10.1103/PhysRevApplied.15.054048.
[92] Y. Guo, Q. Cai, P. Li, Z. Jia, B. Xu, Q. Zhang, Y. Zhang, R. Zhang, Z. Gao, K.A. Shore and Y. Wang, "40 Gb/s quantum random number generation based on optically sampled amplified spontaneous emission," *APL Photonics*, Vol. 6, No. 066105, 2021; doi:10.1063/5.0040250.
[93] M. Hofert, "Random number generators produce collisions: why, how many and more," *Americ. Statistician*, Vol. 75, no. 4, pp. 394–402, 2021.
[94] M.M. Jacak, P. Jóźwiak, J. Niemczuk and J.E. Jacak, "Quantum generators of random numbers," *Sci. Rep.*, Vol. 11, No. 16108, 2021; doi:10.1038/s41598-021-95388-7.
[95] K. Kim, S. Bittner, Y. Zeng, S. Guazzotti, O. Hess, Q.J. Wang and Hui Cao, "Massively parallel ultrafast random bit generation with a chip-scale laser," *Science*, Vol. 371, 26 February 2021, pp. 948–952, 2021.
[96] S. Kim, M.-S. KIM, Y. Lee, H.-D. KIM and S.-J. Choi, "Low-power true random number generator based on randomly distributed carbon nanotube networks," *IEEE Access*, Vol. 9, pp. 91341–91346, 2021.
[97] Y. Li, Y. Fei, W. Wang, X. Meng, H. Wang, Q. Duan and Z. Ma, "Quantum random number generator using a cloud superconducting quantum computer based on source-independent protocol," *Sci. Rep.*, Vol. 11, 23873, 2021; doi:10.1038/s41598-021-03286-9.
[98] Y. Li, Y. Fei, W. Wang, X. Meng, H. Wang, Q. Duan and Z. Ma, "Experimental study on the security of superluminescentvLED-based quantum random generator," *Opt. Eng.*, Vol. 60, No. 11, 116106-01–116106-22, 2021; doi:10.1117/1.OE.60.11.116106.
[99] Z. Lu, S. Yang, J. Liu, X. Wang and Y. Li "Efficient FPGA implementation of high-speed true random number generator," *Rev. Sci. Instrum.*, Vol. 92, No. 024706, 2021; doi:10.1063/5.0035519.
[100] W. Luo, N. Takeuchi, O. Chen and N. Yoshikawa, "Low-autocorrelation random number generator based on adiabatic quantum-flux-parametron logic," *IEEE Trans. Appl. Supercond.*, Vol. 31, no. 5, pp. 1302305, 2021.
[101] Y. Luo, S. Han, S. Zhang, Y. Wang and J. Liu, "High speed true random number generator controlled by logistic map," in: *Proc. 2021 IEEE Int. Conf. High Perform. Comp. & Commun.; 7th Int. Conf. Data Science & Systems; 19th Int. Conf. Smart City; 7th Int. Conf. Depend. Sensor, Cloud & Big Data Systems & Application*, pp. 57–62, 2021; doi:10.1109/HPCC-DSS-SMARTCITY-DEPENDSYS53884.2021.00035.
[102] K. Malik, J. Pulikkotil and A. Sharma, "Comparison of pseudorandom number generators and their application for uncertainty estimation using Monte Carlo imulation," *MAPAN-J. Metrology Soc. India*, Vol. 36, no. 3, pp. 481–496, 2021.
[103] F. Monet, J.-S. Boisvert and R. Kashyap, "A simple high-speed random number generator with minimal post-processing using a random Raman fiber laser," *Sci. Rep.*, Vol. 11, 13182, 2021; doi:10.1038/s41598-021-92668-0.
[104] R. Patgiri, "Rando: a general-purpose true random number generator for conventional computers," in: *Proc. 2021 IEEE 20th Intern. Conf. Trust, Security Privacy Comput. Commun.(TrustCom)*, pp. 107–113, 2021; doi: 10.1109/TRUSTCOM53373.2021.00032.
[105] R. Serrano, C. Duran, T.-T. Hoang, M. Sarmiento, K.-D. Nguyen, K. Suzaki and C.-K. Pham, "A fully digital true random number generator with entropy source based in frequency collapse," *IEEE Access*, Vol. 9, pp. 105748–105755, 2021.
[106] S. Stoller and K.A. Campbell, "Demonstration of three true random number generator circuits using memristor created entropy and commercial off-the-shelf components," *Entropy*, Vol. 23, No. 371, 2021; doi:10.3390/e23030371.
[107] S. Taneja and M. Alioto, "Fully synthesizable unified true random number generator and cryptographic core," *IEEE J. Solid-State Circ.*, Vol. 56, no. 10, pp. 3049–3061, 2021.
[108] P.-H. Tseng, M.-H. Lee, Y.-H. Lin, H.-L. Lung, K.-C. Wang and C.-Y. Lu, "ReRAM-based pseudo-true random number generator With high throughput and unpredictability characteristics," *IEEE Trans. Electron. Devices*, Vol. 68, no. 4, pp. 1593–1597, 2021.
[109] X. Wang, H. Liang, Y. Wang, L. Yao, Y. Guo, M. Yi, Z. Huang, H. Qi and Y. Lu, "High-throughput portable true random number generator based on jitter-latch structure," *IEEE Trans. Circ. Syst.-I: Reg. Papers*, Vol. 68, no. 2, pp. 741–750, 2021.

[110] N. Abraham, K. Watanabe, T. Taniguchi and K. Majumdar, "A high-quality entropy source using van der Waals heterojunction for true random number generation," *ACS Nano*, Vol. 16, pp. 5898–5908, 2022.
[111] N. Akashi, K. Nakajima, M. Shibayama and Y. Kuniyoshi, "A mechanical true random number generator," *New J. Phys.*, Vol. 24, No. 013019, 2022; doi:10.1088/1367-2630/ac45ca.
[112] Y. Alloun, M.S. Azzaz, A. Kifouche and R. Kaibou, "Pseudo random number generator based on chaos theory and artificial neural networks," in: *Proc. 2022 2nd Int. Conf. on Advanced Electr. Eng. (ICAEE)*, 2022, 6 pages.
[113] M.J. Bae, "Quantum walk random number generation: memory-based models," in: *Proc. 2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*, 2022, 12 pages.
[114] H. Bhatia, E. Tretschk, C. Theobalt and V. Golyanik, "Generation of truly random numbers on a quantum annealer," *IEEE Access*, Vol. 10, pp. 112832–112844, 2022.
[115] J. Cui, M. Yi, D. Cao, L. Yao, X. Wang, H. Liang, Z. Huang, H. Qi, T. Ni and Y. Lu, "Design of True Random Number Generator Based on Multi-Stage Feedback Ring Oscillator," *IEEE Trans. Circ. Syst.-II: Expr. Briefs*, Vol. 69, no. 3, pp. 1752–1756, 2022.
[116] P. Debashis, H. Li, D. Nikonov and I. Young, "Spin electronics Gaussian random number generator with reconfigurable mean and variance using stochastic magnetic tunnel junctions," *IEEE Magnetics Lett.*, Vol. 13, No. 4502905, 2022; doi:10.1109/LMAG.2022.3152991.
[117] K. Demir and S. Ergün, "A comparative analysis on chaos-based random number generation methods," *Eur. Phys. J. Plus*, Vol. 137, No. 591, 2022; doi:10.1140/epjp/s13360-022-02793-6.
[118] S. Dutta, A. Arunachalam and S. Misailovic, "To seed or not to seed? An empirical analysis of usage of seeds for testing in machine learning projects," in: *Proc. 2022 IEEE Conf. Softw. Test. Verif. Valid. (ICST)*, pp. 151–161, 2022; doi:10.1109/ICST53961.2022.00026.
[119] A.M. Garipcan and E. Erdem, "A gigabit TRNG with novel lightweight post-processing method for cryptographic applications," *Eur. Phys. J. Plus*, Vol. 137, No. 493, 2022; doi:10.1140/epjp/s13360-022-02679-7.
[120] M. Grujić and I. Verbauwhede, "TROT: a three-edge ring oscillator based true random number generator with time-to-digital vonversion," *IEEE Trans. Circ. Syst.-I: Reg. Papers*, Vol. 69, no. 6, pp. 2435–2448, 2022.
[121] K. Kim, S. Bittner, Y. Zeng, S. Guazzotti, O. Hess, Q.J. Wang and H. Cao, "Massively parallel generation of random numbers using a semiconductor laser," in: *Proc. 2022 Conf. Lasers Electro-Optics (CLEO)*, AW4M.1, 2022.
[122] J. Kim and H. Chae, "A 10-Gbps, 0.121-pJ/bit, All-Digital True Random-Number Generator using Middle Square Method," in: *Proc. 2022 IEEE Asian Solid-State Circ. Conf. (A-SSCC)*, Paper 9.4, 2022; doi:10.1109/A-SSCC56115.2022.9980831.
[123] T.L. Liao, P.Y. Wan and J.-J. Yan, "Design and synchronization of chaos-based true random number generators and its FPGA implementation," *IEEE Access*, Vol. 10, pp. 8279–8286, 2022.
[124] Y. Lv, B.R. Zink and J.-P. Wang, "Bipolar random spike and bipolar random number generation by two magnetic tunnel junctions," *IEEE Trans. Electron. Devices*, Vol. 69, no. 3, pp. 1582–1587, 2022.
[125] R.B. Naik and U. Singh, "A review on applications of chaotic maps in pseudo-random number generators and encryption," *Ann. Data Science*, 18-01-2022, 26 pages; doi:10.1007/s40745-021-00364-7.
[126] J. Park, B. Kim and J.-Y. Sim, "A PVT-tolerant oscillation-collapse-based true random number generator with an odd number of inverter stages," *IEEE Trans. Circ. Syst.-II: Expr. Briefs*, Vol. 69, no. 10, pp. 4058–4062, 2022.
[127] R. Della Sala, D. Bellizia and G. Scotti, "High-throughput FPGA-compatible TRNG architecture exploiting multistimuli metastable cells," *IEEE Trans. Circ. Syst.-I: Reg. Papers*, Vol. 69, no. 12, pp. 4886–4897, 2022.
[128] R. Zhang, X. Wang, K. Liu and H. Shinohara, "A 0.186-pJ per Bit Latch-Based True Random Number Generator Featuring Mismatch Compensation and Random Noise Enhancement," *IEEE J. Solid-State Circ.*, Vol. 57, no. 8, pp. 2498–2508, 2022.
[129] X. Zhou, Z. Hu, Z. Chai, W. Zhang, S. Clima, R. Degraeve, J.F. Zhang, A. Fantini, D. Garbin, R. Delhougne, L. Goux and G. S. Kar, "Impact of relaxation on the performance of GeSe true random number generator based on ovonic threshold switching," *IEEE Electron Device Lett.*, Vol. 43, no. 7, pp. 1061–1064, 2022.
[130] Y. Zhu, Y. Bian, J. Yang, Y. Zhang and S. Yu, "21 Gbps source-independent quantum random number generator based on vacuum fluctuations," in: *Proc. 2022 Asia Commun. Photonics Conf. (ACP) and Intern. Conf. Information Photonics Optical Commun. (IPOC)*, 2140–2142, 2022.
[131] J. Argillander, A. Alarcon, C. Bao C. Kuang, G. Lima, F. Gao and G.B. Xavier, "Quantum random number generation based on a perovskite light emitting diode," *Commun. Phys.*, Vol. 6, No. 157, 2023; doi:10.1038/s42005-023-01280-3.
[132] C. Bruynsteen, T. Gehring, C. Lupo, J. Bauwelinck and X. Yin, "100-Gbit/s Integrated Quantum Random Number Generator Based on Vacuum Fluctuations ," *PRX QUANTUM*, Vol. 4, No. 010330, 2023; doi: 10.1103/PRXQuantum.4.010330.
[133] Q. Cai, P. Li, Y. Shi, Z. Jia, L. Ma, B. Xu, X. Chen, K.A. Shore and Y. Wang, "Tbps parallel random number generation based on a single quarter-wavelength-shifted DFB laser," *Optics & Laser Technol.*, Vol. 162, No. 109273, 2023; doi:10.1016/j.optlastec.2023.109273.
[134] S. Chatterjee, N. Rangarajan, S. Patniak, D. Rajasekharan, O. Sinanoglu and Y.S. Chauhan, "FerroCoin: ferroelectric tunnel junction-based true random number generator," *IEEE Trans. Emerg. Topics Comput.*, Vol. 11, no. 2, pp. 541–547, 2023.
[135] C.-B. Chen, T. Chen, Y.-H. Huang and Y.-H. Huang, "35.56-Gbits/sec chaos random number generator supporting 1.2-GSamples/sec noise generation," *IEEE Trans. Circ. Syst.-II: Expr. Briefs*, Vol. 70, no. 10, pp. 3802–3806, 2023.
[136] X. Cheng, Y. Zhang, H. Zhuand Y. Zhou, "A true random number generator with high bit rate and low energy efficiency," *Int. J.Circ. Theor. Appl.*, Vol. 51, pp. 3415–3431, 2023.
[137] M. Eaton, A. Hossameldin, R.J. Birrittella, P.M. Alsing, C.C. Gerry, H. Dong, C. Cuevas and O. Pfister, "Resolution of 100 photons and quantum generation of unbiased random numbers," *Nature Photonics*, Vol. 17, pp. 106–111, 2023.
[138] E. Elmitwalli and S. Köse, "Bistable Josephson Junction-based true random number generator without inductors," *IEEE Trans. Circ. Syst.-II: Expr. Briefs*, Vol. 70, no. 4, pp. 1615–1619, 2023.
[139] M.S. Equbal, T. Ketkar and S. Sahay, "Hybrid CMOS-RRAM true random number generator exploiting coupled entropy sources," *IEEE Trans. Electron. Devices*, Vol. 70, no. 3, pp. 1061–1066, 2023.
[140] F. Frustaci, F. Spagnolo, S. Perri and P. Corsonello, "A high-speed FPGA-based true random number generator using metastability with clock managers," *IEEE Trans. Circ. Syst.-II: Expr. Briefs*, Vol. 70, no. 2, pp. 756–760, 2023.
[141] Y. Fu, J. Wen, L. Wang, L. Yang, Q. Zhu, W. Zuo, P. Zhang, Y. Li, H. Tong, G. Ma, H. Wang and X. Miao, "A 2.22 Mb/s true random number generator based on a GeTex ovonic threshold switching memristor," *IEEE Electron Device Lett.*, Vol. 44, no. 5, pp. 853–856, 2023.
[142] O. Guillan-Lorenzo, M. Troncoso-Costas, D. Alvarez-Outarelo, F.J. Diaz-Otero and J.C. Garcia-Escartin, "Optical quantum random number generators: a comparative study," *Opt. Quant. Electron.*, Vol. 55, No. 185, 2023; doi:10.1007/s11082-022-04396-y.
[143] J. Hanlon and S. Felix, "A fast hardware pseudorandom number generator based on xoroshiro128," *IEEE Trans. Computers*, Vol. 72, no. 5, pp. 1518–1524, 2023.
[144] P. Keshavarzian, K. Ramu, D. Tang, C. Weill, F. Gramuglia, S.S. Tan, M. Tng, L. Lim, E. Quek, D. Mandich, M. Stipcević and E. Charbon, "A 3.3-Gb/s SPAD-Based Quantum Random Number Generator," *IEEE J. Solid-State Circ.*, Vol. 58, no. 9, pp. 2632–2646, 2023.
[145] F. Khodayari, A. Amirany, M.H. Moaiyeri and K. Jafari, "A variation-aware ternary true random number generator using magnetic tunnel junction at subcritical current regime," *IEEE Trans. Magnetics*, Vol. 59, no. 3, 3400208, 2023.
[146] D. Kumar, L.L. Mankali, P.K. Misra and M. Goswami, "A 0.7 pJ/bit, 1.5 Gbps energy-efficient image-based true random number generator," *IETE J. Research*, Vol. 69, no. 3, pp. 1260–1270, 2023.
[147] S. Lee, "Strategies for ultra-fast bit generation of two-terminal threshold switch-based true random number generator using drift-free Ge-doped SiO2 threshold switch device," *Solid-State Electron.*, Vol. 201, No. 108609, 2023; doi:10.1016/j.sse.2023.108609.

[148] J. Li, Y. Li, Y. Dong, Y. Yuang, J. Xiao and Y. Huang, "400 Gb/s physical random number generation based on deformed square self-chaotic lasers," *Chin. Optics Lett.*, Vol. 21, No. 6, 061901, 2023; doi:10.3788/COL202321.061901.
[149] J. Li, J. Liu, D. Liu, W. Tian, S. Jin, S. Hu, J. Guo and Y. Jin, "Ultrafast random number generation based on random laser ," *J. Lightwave Technol.*, Vol. 41, no. 16, pp. 5233–5243, 2023.
[150] S. Li, Y. Liu, F. Ren and Z. Yang, "Design of a high throughput pseudorandom number generator based on discrete hyper-chaotic system," *IEEE Trans. Circ. Syst.-II: Expr. Briefs*, Vol. 70, no. 2, pp. 806–810, 2023.
[151] X. Li, Y. Wang, Y. Yang, S. Lv, Q. Luo, X. Wang, X. Xu, D. Lei and F. Zhang, "A 144-fJ/bit reliable and compact TRNG based on the diffusive resistance of 3-D resistive random access memory ," *IEEE Trans. Electron. Devices*, Vol. 70, no. 8, pp. 4139–4144, 2023.
[152] Y. Luo, J. Zhang, J. Hao and X. Zhao, "A 2.5 pJ/bit PVT-tolerant true random number generator based on native-NMOS-regulated ring oscillator," *IEEE Trans. Circ. Syst.-II: Expr. Briefs*, Vol. 70, no. 10, pp. 3927–3931, 2023.
[153] S. Maity, A. Prosad, H. Natarajan and V. Raghunathan "Comparison of high speed quantum random number generators based on ASE-ASE and ASE-LASER beating," in: P.R. Hemmer and A.L. Migdall (eds.), *Quantum Computing, Communication, and Simulation III*, Proceedings of SPIE, vol. 12446, No. 1244610, 2023; doi:10.1117/12.2649916.
[154] H. Ming, H. Hu, F. Lv and R. Yu, "A high-performance hybrid random number generator based on a nondegenerate coupled chaos and its practical implementation," *Nonlinear Dyn.*, Vol. 111, pp. 847–869 2023.
[155] F. Monet and R. Kashyap, "On multiplexing in physical random number generation, and conserved total entropy content," *Sci. Rep.*, Vol. 13, 7892, 2023; doi:10.1038/s41598-023-35130-7.
[156] P. Nahar, P. Khandekar, M. Deshmukh, H.S. Jatana and U. Khambete, "Survey of Stochastic Number Generators and Optimizing Techniques," in: A.J. Kulkarni et al. (eds.), *Intelligent Systems and Applications*, Lecture Notes in Electrical Engineering, vol. 959, pp. 405—415. Singapore: Springer Nature Singapore Pte Ltd., 2023.
[157] B. Peng, Q. Wu, Z. Wang and J. Yang, "A RRAM-based true random number generator with 2T1R architecture for hardware security applications," *Micromachines*, Vol. 14, No. 1213, 2023; doi:10.3390/mi14061213.
[158] K. Tanizawa, K. Kato and F. Futami, "Real-time 50-Gbit/s spatially multiplexed quantum random number generator based on vacuum fluctuation," in: *Proc. 2023 Opt. Fiber Commun. Conf. Exh. (OCE)*, Th4A.8, 2023.
[159] X. Wang, R. Zhang, K. Liu and H. Shinohara, "A 0.116 pJ/bit latch-based true random number generator featuring static inverter selection and noise enhancement," *IEEE Trans. Very Large Scale Integr. (VLSI) Systems*, to appear, early access at doi:10.1109/TVLSI.2023.3328602, 2023.
[160] Z. Yang, Y. Liu, Y. Wu, Y. Qi, F. Ren and S. Li, "A high speed pseudo-random bit generator driven by 2D-discrete hyperchaos," *Chaos Solit. Fract.*, Vol. 167, No. 113039, 2023; doi:10.1016/j.chaos.2022.113039.
[161] F. Yuan, S. Li, Y. Deng, Y. Li and G. Chen, "Cu-doped TiO2-x nanoscale memristive applications in chaotic circuit and true random number generator," *IEEE Trans. Ind. Electron.*, Vol. 70, no. 4, pp. 4120–4127, 2023.
[162] W. Zhang, "Analysis and Construction of Nonlinear Correctors Used in True Random Number Generators," *IEEE Trans. Inf. Theory*, Vol. 69, no. 10, pp. 6671–6681, 2023.
[163] J. Zheng and J. Li, "An ultrafast cryptographically secure pseudorandom number generator," in: S.-H. Seo and H. Seo (eds.), *Int. Conf. Inf. Security Crypt. ICISC 2022*, Lecture Notes in Computer Science, vol. 13849, pp. 267–291. Cham, Switzerland: Springer Nature Switzerland, 2023.
[164] M. Broniatowski and W. Stummer, "Bare-Simulation optimization of some distances between fuzzy sets respectively basic belief assignments," Preprint 2023.
[165] P.N. Rathie and P. Kannappan, "A directed-divergence function of type β," *Information and Control*, Vol. 20, pp. 38–45, 1972.
[166] M. Shiino, "H-Theorem with generalized relative entropies and the Tsallis statistics," *J. Phys. Soc. Japan*, Vol. 67, no. 11, pp. 3658–3660, 1998.
[167] K. Matusita, "On the theory of statistical functions," *Ann. Inst. Statist. Math. Tokyo*, Vol. 3, pp. 17–35, 1951; corrections in *Ann. Inst. Statist. Math. Tokyo*, Vol. 4, pp. 51–53, 1952.
[168] M.M. Deza and E. Deza, *Encyclopedia of Distances*, 4th ed. Berlin, Germany: Springer, 2016.
[169] B. Roensch and W. Stummer, "3D insights to some divergences for robust statistics and machine learning," In: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2017*, Lecture Notes in Computer Science, vol. 10589, pp. 460–469. Cham, Switzerland: Springer International Publishing, 2017.
[170] B. Roensch and W. Stummer, "Robust estimation by means of scaled Bregman power distances; part I; non-homogeneous data," In: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2019*, Lecture Notes in Computer Science, vol. 11712, pp. 319–330. Cham, Switzerland: Springer Nature, 2019.
[171] B. Roensch and W. Stummer, "Robust estimation by means of scaled Bregman power distances; part II; extreme values," In: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2019*, Lecture Notes in Computer Science, vol. 11712, pp. 331–340. Cham, Switzerland: Springer Nature, 2019.
[172] S. Krömer and W. Stummer, "A new toolkit for mortality data analytics," in: A. Steland, E. Rafajlowicz, and O. Okhrin (eds.), *Stochastic Models, Statistics and Their Applications*, pp. 393–407. Cham, Switzerland: Springer Nature Switzerland, 2019.
[173] R. Hennequin, B. David and R. Badeau, "Beta-Divergence as a Subclass of Bregman Divergence," *IEEE Signal Proces. Lett.*, Vol. 18, no. 2, pp. 83–86, 2011.
[174] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in: *Proc. 6th Int. Congr. Acoust.*, 1968, pp. C-17—C-20. Los Alamitos, CA, USA.
[175] A. Basu, A. Mandal, N. Martin and L. Pardo, "Robust tests for the equality of two normal means based on the density power divergence," *Metrika*, Vol. 78, pp. 611–634, 2015.
[176] A. Basu, A. Mandal, N. Martin and L. Pardo, "Generalized Wald-type tests on minimum power divergence estimators," *Statistics*, Vol. 50, no. 1, pp. 1–26, 2016.
[177] A. Ghosh and A. Basu, "Robust Bayes estimation using the density power divergence," *Ann. Inst. Statist. Math.*, Vol. 68, pp. 413–437, 2016.
[178] A. Ghosh and A. Basu, "Robust estimation in generalized linear models: the density power divergence approach," *TEST*, Vol. 25, pp. 269–290, 2016.
[179] A. Basu, A. Ghosh, A. Mandal, N. Martin and L. Pardo, "A Wald-type test statistic for testing linear hypothesis in logistic regression models based on minimum density power divergence estimator," *Electr. J. Statist.*, Vol. 11, pp. 2741–2772, 2017.
[180] F. Martin, J. Carballeira, L. Moreno, S. Garrido and P. González, "Using the Jensen-Shannon, density power, and Itakura-Saito divergences to implement an evolutionary-based global localization filter for mobile robots," *IEEE Access*, Vol. 5, pp. 13922–13940, 2017.
[181] A. Basu, A. Mandal, N. Martin and L. Pardo, "Testing composite hypothesis based on the density power divergence," *Sankhya*, Vol. 80-B, Part 2, pp. 222–262, 2017.
[182] N. Balakrishnan, E. Castilla, N. Martin and L. Pardo, "Robust estimators and test statistics for one-shot device testing under the exponential distribution," *IEEE Trans. Inf. Theory*, Vol. 65, no. 5, pp. 3080–3096, 2019.
[183] N. Balakrishnan, E. Castilla, N. Martin and L. Pardo, "Robust inference for one-shot device testing data under Weibull lifetime model," *IEEE Trans. Reliab.*, Vol. 69, no. 3, pp. 937–953, 2020.
[184] A. Ghosh and S. Majumdar, "Ultrahigh-dimensional robust and efficient sparse regression using non-concave penalized density power divergence," *IEEE Trans. Inf. Theory*, Vol. 66, no. 12, pp. 7812–7827, 2020.
[185] V. Leplat, N. Gillis and A.M.S. Ang, "Blind audio source separation with minimum-volume beta-divergence NMF," *IEEE Trans. Signal Proces.*, Vol. 68, pp. 3400–3410, 2020.

[186] M. Vandecappelle, N. Vervliet and L. De Lathauwer, "A second-order method for fitting the canonical polyadic decomposition with non-least-squares cost," *IEEE Trans. Signal Proces.*, Vol. 68, pp. 4454–4465, 2020.
[187] N. Balakrishnan, E. Castilla, N. Martin and L. Pardo, "Divergence-based robust inference under proportional hazards model for one-shot device life-test," *IEEE Trans. Reliab.*, Vol. 70, no. 4, pp. 1355–1367, 2021.
[188] S. Basak, A. Basu and M.C. Jones, "On the 'optimal' density power divergence tuning parameter," *J. Appl. Statist.*, Vol. 48, no. 3, pp.536–556, 2021.
[189] A. Basu, A. Ghosh, A. Mandal, N. Martin and L. Pardo, "Robust Wald-type tests in GLM with random design based on minimum density power divergence estimators," *Statist. Meth. Appl.*, Vol. 30, pp.973–1005, 2021.
[190] A. Calvino, N. Martin and L. Pardo, "Robustness of minimum density power divergence estimators and Wald-type test statistics in loglinear models with multinomial sampling," *J. Computat. Appl. Math.*, Vol. 386, No. 113214, 2021; doi:10.1016/j.cam.2020.113214.
[191] E. Castilla, A. Ghosh, N. Martin and L. Pardo, "Robust semiparametric inference for polytomous logistic regression with complex survey design," *Adv. Data Anal. Classif.*, Vol. 15, pp. 701–734, 2021.
[192] E. Castilla, N. Martin, L. Pardo and K. Zografos, "Composite likelihood methods: Rao-type tests based on composite minimum density power divergence estimator," *Statist. Papers*, Vol. 62, pp. 1003–1041, 2021.
[193] Q. Legros, J. Tachella, R. Tobin, A. McCarthy, S. Meignen, G.S. Buller, Y. Altmann, S. McLaughlinand M.E. Davies, "Robust 3D reconstruction of dynamic scenes from single-photon Lidar using beta-divergences," *IEEE Trans. Image Process.*, Vol. 30, pp. 1716–1727, 2021.
[194] W. Pu, S. Ibrahim, X. Fu and M. Hong, "Stochastic mirror descent for low-rank tensor decomposition under non-Euclidean losses," *IEEE Trans. Signal Proces.*, Vol. 70, pp. 1803–1818, 2022.
[195] M.A. Ramirez, W. Beccaro, D.Z. Rodriguez and R.L. Rosa, "Differentiable measures for speech modeling" *IEEE Access*, Vol. 10, pp. 17609–17618, 2022.
[196] E. Castilla and P.J. Chocano, "On the choice of the optimal tuning parameter in robust one-shot device testing analysis," in: N. Balakrishnan et al. (eds.), *Trends in Mathematical, Information and Data Sciences*, Studies in Systems, Decision and Control, vol. 445, pp. 169–180. Cham, Switzerland: Springer Nature Switzerland, 2023.
[197] A. Marmin, J.H. de Morais Goulart and C. Févotte, "Majorization-minimization for sparse nonnegative matrix factorization with the $\beta-$divergence," *IEEE Trans. Signal Proces.*, Vol. 71, pp. 1435–1447, 2023.
[198] G. Saraceno, A. Ghosh, A. Basu and C. Agostinelli, "Robust estimation of fixed effect parameters and variances of linear mixed models: the minimum density power divergence approach," *AStA Adv. Stat. Anal.*, 2023; doi:10.1007/s10182-023-00473-z.
[199] P. Sharma and P.M. Pradhan, "A novel family of robust incremental adaptive algorithms for distributed estimation based on Bregman divergence," *IEEE Sensor Lett.*, Vol. 7, No. 8, 7004104, 2023; doi:10.1109/LSENS.2023.3296396.
[200] V.M. Zolotarev, *One-dimensional Stable Distributions*. Providence, USA: American Mathematical Society, 1986.
[201] D.-B. Jeong and S. Sarkar, "Negative exponential disparity based family of goodness-of-fit tests for multinomial models," *J. Statist. Comput. Simul.*, Vol. 65, pp. 43–61, 2000.
[202] S.K. Bhandari, A. Basu and S. Sarkar, "Robust inference in parametric models using the family of generalized negative exponential disparities," *Aust. N. Z. J. Stat.*, Vol. 48, no. 1, pp. 95–114, 2006.
[203] B.G. Lindsay, "Efficiency versus robustness: the case for Hellinger distance and related methods," *Ann. Stat.*, Vol. 22, no. 2, pp. 1081–1114, 1994.
[204] A. Basu and S. Sarkar, "The trade-off between robustness and efficiency and the effect of model smoothing in minimum disparity inference," *J. Statist. Comput. Simul.*, Vol. 50, pp. 173–185, 1994.
[205] D. Berend, P. Harremoës and A. Kontorovich, "Minimum KL-divergence on complements of L_1 balls," *IEEE Trans. Inf. Theory*, Vol. 60, no. 6, pp. 3172–3177, 2014.
[206] J. Jiao, T.A. Courtade, A. No, K. Venkat and T. Weissman, "The curious case of the binary alphabet," *IEEE Trans. Inf. Theory*, Vol. 60, no. 12, pp. 7616–7626, 2014.
[207] G. Como and F. Fagnani, "Robustness of large-scale stochastic matrices to localized perturbations," *IEEE Trans. Netw. Sci. Eng.*, Vol. 2, no. 2, pp. 53–64, 2015.
[208] Y. Han, J. Jiao and T. Weissman, "Minimax Estimation of Discrete Distributions Under ℓ_1 loss," *IEEE Trans. Inf. Theory*, Vol. 61, no. 11, pp. 6343–6354, 2015.
[209] I. Sason, "Tight bounds for symmetric divergence measures and a refined bound for lossless source coding," *IEEE Trans. Inf. Theory*, Vol. 61, no. 2, pp. 701–707, 2015.
[210] A. Batsidis, N. Martin, L. Pardo and K. Zografos, "$\phi-$Divergence based procedure for parametric change-point problems," *Methodol. Comput. Appl. Probab.*, Vol. 18, pp. 21–35, 2016.
[211] G. Böcherer and B.C. Geiger, "Optimal quantization for distribution synthesis," *IEEE Trans. Inf. Theory*, Vol. 62, no. 11, pp. 6162–6172, 2016.
[212] N.V.A. Das and N. Kashyap, "MCMC methods for drawing random samples from the discrete-grains model of a Magnetic Medium," *IEEE J. Sel. Areas Commun.*, Vol. 34, no. 9, pp. 2430–2438, 2016.
[213] J.M. Alonso-Revenga, N. Martin and L. Pardo, "New improved estimators for overdispersion in models with clustered multinomial data and unequal cluster sizes," *Stat. Comput.*, Vol. 27, pp. 193–217, 2017.
[214] A. Keziou and P. Regnault, "Semiparametric estimation of mutual information and related criteria: optimal test of independence," *IEEE Trans. Inf. Theory*, Vol. 63, no. 1, pp. 57–71, 2017.
[215] J. Liu, P. Cuff and S. Verdú, "$E_\gamma-$resolvability," *IEEE Trans. Inf. Theory*, Vol. 63, no. 5, pp. 2629–2658, 2017.
[216] I. Tzortzis, C.D. Charalambous, T. Charalambous, C.N. Hadjicostis and M. Johansson, "Approximation of Markov processes by lower dimensional processes via total variation metrics," *IEEE Trans. Autom. Control*, Vol. 62, no. 3, pp. 1030–1045, 2017.
[217] E. Castilla, N. Martin and L. Pardo, "Minimum phi-divergence estimators for multinomial logistic regression with complex sample design," *AStA Adv. Stat. Anal.*, Vol. 102, pp. 381–411, 2018.
[218] I. Csiszár and T. Breuer, "Expected value minimization in information theoretic multiple priors models," *IEEE Trans. Inf. Theory*, Vol. 64, no. 6, pp. 3957–3974, 2018.
[219] M. El Gheche, G. Chierchia and J.-C. Pesquet, "Proximity operators of discrete information divergences," *IEEE Trans. Inf. Theory*, Vol. 64, no. 2, pp. 1092–1104, 2018.
[220] A. Felipe, N. Martin, P. Miranda and L. Pardo, "Statistical inference in constrained latent class models for multinomial data based on $\phi-$divergence measures," *Adv. Data Anal. Classif.*, Vol. 12, pp. 605–636, 2018.
[221] M. Markatou and Y. Chen, "Non-quadratic distances in model assessment," *Entropy*, Vol. 20, No. 464, 2018; doi:10.3390/e20060464.
[222] Y. Sun, S. Schaefer and W. Wang, "Image structure retrieval via L_0 minimization," *IEEE Trans. Visual. Comput. Graph.*, Vol. 24, no. 7, pp. 2129–2139, 2018.
[223] M. Asadi, N. Ebrahimi, O. Kharazmi and E.S. Soofi, "Mixture models, Bayes Fisher information, and divergence measures," *IEEE Trans. Inf. Theory*, Vol. 65, no. 4, pp. 2316–2321, 2019.
[224] M. Broniatowski, E. Miranda and W. Stummer, "Testing the number and the nature of the components in a mixture distribution," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2019*, Lecture Notes in Computer Science, vol. 11712, pp. 309–318. Cham, Switzerland: Springer Nature Switzerland, 2019.
[225] J.-F. Collet, "An exact expression for the gap in the data processing inequality for f-divergences," *IEEE Trans. Inf. Theory*, Vol. 65, no. 7, pp. 4387–4391, 2019.

[226] I. Sason, "On data-processing and majorization inequalities for f-Divergences with applications," *Entropy*, Vol. 21, No. 1022, 2019; doi:10.3390/e21101022.
[227] S. Yagli and P. Cuff, "Exact exponent for soft covering," *IEEE Trans. Inf. Theory*, Vol. 65, no. 10, pp. 6234–6262, 2019.
[228] K. Zhao, Y. Liu, H. Du and B. Yuan, "Scaling analysis of multilevel polar coded modulation" *IEEE Access*, Vol. 7, pp. 65129–65138, 2019.
[229] N. De Ponti, "Metric properties of homogeneous and spatially inhomogeneous F-Divergences," *IEEE Trans. Inf. Theory*, Vol. 66, no. 5, pp. 2872–2890, 2020.
[230] N.B. Kammerer and W. Stummer, "Some dissimilarity measures of branching processes and optimal decision making in the presence of potential pandemics," *Entropy*, Vol. 22(8), No. 874, 2020 (123 pages); doi:10.3390/e22080874.
[231] T. Nishiyama and I. Sason, "On relations between the relative entropy and χ^2-divergence, generalizations and applications," *Entropy*, Vol. 22, No. 563, 2020; doi:10.3390/e22050563.
[232] R. Nomura, "Source resolvability and intrinsic randomness: two random number generation problems With respect to a subclass of f-divergences," *IEEE Trans. Inf. Theory*, Vol. 66, no. 12, pp. 7588–7601, 2020.
[233] B. Rassouli and D. Gündüz, "Optimal utility-privacy trade-off with total variation distance as a privacy measure," *IEEE Trans. Inf. Forensics Sec.*, Vol. 15, pp. 594–603, 2020.
[234] A.R. Esposito, M. Gastpar and I. Issa, "Generalization error bounds via Rényi-, f-divergences and maximal leakage," *IEEE Trans. Inf. Theory*, Vol. 67, no. 8, pp. 4986–5004, 2021.
[235] M. Markatou, D. Karlis and Y. Ding, "Distance-based statistical inference," *Annu. Rev. Stat. Appl.*, Vol. 8, pp. 301–327, 2021.
[236] S. Salehkalaibar, M.H. Yassaee, V.Y.F. Tan and M. Ahmadipour, "State masking over a two-state compound channel," *IEEE Trans. Inf. Theory*, Vol. 67, no. 9, pp. 5651–5673, 2021.
[237] W. Stummer, "Optimal transport with some directed distances," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2021*, Lecture Notes in Computer Science, vol. 12829, pp. 829–840. Cham, Switzerland: Springer Nature Switzerland, 2021.
[238] I. Tzortzis, C.D. Charalambous and C.N. Hadjicostis, "Jump LQR systems with unknown transition probabilities," *IEEE Trans. Autom. Control*, Vol. 66, no. 6, pp. 2693–2708, 2021.
[239] J. Birrell, M.A. Katsoulakis and Y. Pantazis, "Optimizing variational representations of divergences and accelerating their statistical estimation," *IEEE Trans. Inf. Theory*, Vol. 68, no. 7, pp. 4553–4572, 2022.
[240] E. Castilla and P.J. Chocano, "A new robust approach for multinomial logistic regression With complex design model," *IEEE Trans. Inf. Theory*, Vol. 68, no. 11, pp. 7379–7395, 2022.
[241] A. Dixit, M. Ahmadi and J.W. Burdick, "Distributionally robust model predictive control with total variation distance," *IEEE Control Syst. Lett.*, Vol. 6, pp. 3325–3330, 2022.
[242] D. Hyun, G.B. Kim, N. Bottenus and J.J. Dahl, "Ultrasound lesion detectability as a distance between probability measures," *IEEE Trans. Ultrason. Ferr. Frequ. Control*, Vol. 69, no. 2, pp. 732–743, 2022.
[243] J. Melbourne, S. Talukdar, S. Bhaban, M. Madiman and M.V. Salapaka, "The differential entropy of mixtures: new bounds and applications," *IEEE Trans. Inf. Theory*, Vol. 68, no. 4, pp. 2123–2146, 2022.
[244] J. Peng, G. Serrano, I.M. Traniello, M.E. Calleja-Cervantes, U.V. Chembazhi, S. Bangru, T. Ezponda, J.R. Rodriguez-Madoz, A. Kalsotra, F. Prosper, I. Ochoa and M. Hernaez, "SimiC enables the inference of complex gene regulatory dynamics across cell phenotypes," *Commun. Biol.*, Vol. 5, No. 531, 2022; doi:10.1038/s42003-022-03319-7.
[245] Z. Tan and X. Zhang, "On loss functions and regret bounds for multi-category classification," *IEEE Trans. Inf. Theory*, Vol. 68, no. 8, pp. 5295–5313, 2022.
[246] F. Zhang, J. Li, and H.K.T. Ng, "Minimum f-divergence estimation with applications to degradation data analysis," *IEEE Trans. Inf. Theory*, Vol. 68, no. 10, pp. 6774–6789, 2022.
[247] M.V. Alba-Fernández and M.D. Jiménez-Gamero, "Equivalence tests for multinomial data based on $\phi-$divergences," in: N. Balakrishnan et al. (eds.), *Trends in Mathematical, Information and Data Sciences*, Studies in Systems, Decision and Control, vol. 445, pp. 121–129. Cham, Switzerland: Springer Nature Switzerland, 2023.
[248] J.-P. Baudry, M. Broniatowski and C. Thommeret, "Aggregated tests based on supremal divergence estimators for non-regular statistical models," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2023, Part I*, Lecture Notes in Computer Science, vol. 14071, pp. 136–144. Cham, Switzerland: Springer Nature Switzerland, 2023.
[249] M. Boukeloua and A. Keziou, "Empirical likelihood with censored data?," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2023, Part I*, Lecture Notes in Computer Science, vol. 14071, pp. 125–135. Cham, Switzerland: Springer Nature Switzerland, 2023.
[250] N. Cressie, A.R. Pearse and D. Gunawan, "Optimal spatial prediction for non-negative spatial processes using a phi-divergence loss function," in: N. Balakrishnan et al. (eds.), *Trends in Mathematical, Information and Data Sciences*, Studies in Systems, Decision and Control, vol. 445, pp. 181–197. Cham, Switzerland: Springer Nature Switzerland, 2023.
[251] M. Kateri, "Generalized Models for Binary and Ordinal Responses," in: N. Balakrishnan et al. (eds.), *Trends in Mathematical, Information and Data Sciences*, Studies in Systems, Decision and Control, vol. 445, pp. 63–71. Cham, Switzerland: Springer Nature Switzerland, 2023.
[252] T. Manole and A. Ramdas, "Martingale methods for sequential estimation of convex functionals and divergences," *IEEE Trans. Inf. Theory*, Vol. 69, no. 7, pp. 4641–4658, 2023.
[253] M. Markatou and A. Liu, "Statistical distances in goodness-of-fit," in: N. Balakrishnan et al. (eds.), *Trends in Mathematical, Information and Data Sciences*, Studies in Systems, Decision and Control, vol. 445, pp. 213—222. Cham, Switzerland: Springer Nature Switzerland, 2023.
[254] S. Masiha, A. Gohari and M.H. Yassaee, "f-Divergences and their applications in lossy compression and bounding generalization error," *IEEE Trans. Inf. Theory*, Vol. 69, no. 12, pp. 7538–7564, 2023.
[255] P. Miranda, A. Felipe and N. Martin, "Phi-divergence test statistics applied to latent class models for binary data," in: N. Balakrishnan et al. (eds.), *Trends in Mathematical, Information and Data Sciences*, Studies in Systems, Decision and Control, vol. 445, pp. 223—231. Cham, Switzerland: Springer Nature Switzerland, 2023.
[256] F. Nielsen and K. Okamura, "On the f-divergences between hyperboloid and Poincaré distributions," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2023, Part I*, Lecture Notes in Computer Science, vol. 14071, pp. 176–185. Cham, Switzerland: Springer Nature Switzerland, 2023.
[257] P. Perrone, "Categorical information geometry," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2023, Part I*, Lecture Notes in Computer Science, vol. 14071, pp. 268–277. Cham, Switzerland: Springer Nature Switzerland, 2023.
[258] S. Velasco-Forero, "Can Generalised Divergences Help for Invariant Neural Networks?," in: F. Nielsen and F. Barbaresco (eds.), *Geometric Science of Information GSI 2023, Part I*, Lecture Notes in Computer Science, vol. 14071, pp. 82–90. Cham, Switzerland: Springer Nature Switzerland, 2023.
[259] F. Nielsen and K. Okamura, "On the f-divergences between densities of a multivariate location or scale family," *Statist. Comput.*, Vol. 34, No. 60, 2024; doi:10.1007/s11222-023-10373-6.
[260] I. Csiszár and T. Breuer, "Expected value minimization in information theoretic multiple priors models," *IEEE Trans. Inf. Theory*, Vol. 64, no. 6, pp. 3957–3974, 2018.
[261] S. Jana and A. Basu, "A characterization of all single-integral, non-kernel divergence estimators," *IEEE Trans. Inf. Theory*, Vol. 65, no. 12, pp. 7976–7984, 2019.
[262] A. Painsky and G.W. Wornell, "Bregman divergence bounds and universality properties of the logarithmic loss," *IEEE Trans. Inf. Theory*, Vol. 66, no. 3, pp. 1658–1673, 2020.

[263] P.-H. Vial, P. Magron, T. Oberlin and C. Févotte, "Phase retrieval with Bregman divergences and application to audio signal recovery," *IEEE J. Sel. Topics Signal Proces.*, Vol. 15, no. 1, pp. 51–64, 2021.
[264] H. Lohit and D. Kumar, "Modified total Bregman divergence driven picture fuzzy clustering with local information for brain MRI image segmentation," *Appl. Soft Comput.*, Vol. 144, No. 110460, 2023; doi:10.1016/j.asoc.2023.110460.
[265] B. Chen, J. Hu, L. Pu, and Z. Sun, "Stochastic gradient algorithm under $(h, \phi)-$entropy," *Circuits Syst. Signal Process.*, Vol. 26, pp. 941–960, 2007.
[266] V. Girardin and L. Lhote, "Rescaling entropy and divergence Rates," *IEEE Trans. Inf. Theory*, Vol. 61, no. 11, pp. 5868–5882, 2015.
[267] M. Ren, J. Zhang, M. Jiang, M. Yu, and J. Xu, "Minimum $(h, \phi)-$entropy control for non-Gaussian stochastic networked control systems and its application to a networked DC motor control system," *IEEE Trans. Control Syst. Technol.*, Vol. 23, no. 1, pp. 406–411, 2015.
[268] V. Girardin, L. Lhote and P. Regnault, "Different closed-form expressions for generalized entropy rates of Markov chains," *Methodol. Comput. Appl. Probab.*, Vol. 21, pp. 1431–1452, 2019.
[269] F. Nielsen and S. Boltz, "The Burbea-Rao and Bhattacharyya centroids," *IEEE Trans. Inf. Theory*, Vol. 57, no. 8, pp. 5455–5466, 2011.
[270] F. Nielsen and R. Nock, "Generalizing skew Jensen divergences and Bregman divergences with comparative convexity," *IEEE Signal Proces. Lett.*, Vol. 24, no. 8, pp. 1123–5466, 2017.
[271] H. Xu, C. Caramanis and S. Mannor, "Outlier-robust PCA: the high-dimensional case," *IEEE Trans. Inf. Theory*, Vol. 59, no. 1, pp. 546–572, 2013.
[272] S. Kim, B. Ham, B. Kim and K. Sohn, "Mahalanobis distance cross-correlation for illumination-invariant stereo matching," *IEEE Trans. Circ. Syst. Video Techn.*, Vol. 24, no. 11, pp. 1844–1859, 2014.
[273] J. Mei, M. Liu, Y.-F. Wang and H. Gao, "Learning a Mahalanobis distance-based dynamic time warping measure for multivariate time series classification," *IEEE Trans. Cybernetics*, Vol. 46, no. 6, pp. 1363–1374, 2016.
[274] Y. Zhang, B. Du, L. Zhang and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, Vol. 54, no. 3, pp. 1376–1389, 2016.
[275] Z. Li, D.P. Filev, I. Kolmanovsky, E. Atkins and J. Lu, "A new clustering algorithm for processing GPS-based road anomaly reports with a Mahalanobis distance," *IEEE Trans. Intell. Transp. Syst.*, Vol. 18, no. 7, pp. 1980–1988, 2017.
[276] Y. Xu, H. Min, Q. Wu, H. Song and B. Ye, "Multi-instance metric transfer learning for genome-wide protein function prediction," *Sci. Rep.*, Vol. 7, 41831, 2017; doi:10.1038/srep41831.
[277] C.R. Mahony1 and A.J. Cannon, "Wetter summers can intensify departures from natural variability in a warming climate," *Nature Commun.*, Vol. 9, 783, 2018; doi:10.1038/s41467-018-03132-z.
[278] Y. Xu, Z. Wu, J. Chanussot and Z. Wei, "Joint reconstruction and anomaly detection from compressive hyperspectral images using Mahalanobis distance-regularized tensor RPCA," *IEEE Trans. Geosci. Remote Sens.*, Vol. 56, no. 5, pp. 2919–2930, 2018.
[279] T.R. Etherington, "Mahalanobis distances and ecological niche modelling: orrecting a chi-squared probability error," *PeerJ*, Vol. 7, e6678, 2019; doi:10.7717/peerj.6678.
[280] M.C. Fitzpatrick and R.R. Dunn, "Contemporary climatic analogs for 540 North American urban areas in the late 21st century," *Nature Commun.*, Vol. 10, 614, 2019; doi:10.1038/s41467-019-08540-3.
[281] M. Kakavand, A. Mustapha, Z. Tan, S.F. Yazdani and L. Arulsamy, "Using Mahalanobis distance map for web service attacks classification" *IEEE Access*, Vol. 7, pp. 167141–167156, 2019.
[282] S. Sun, H. Wang, Z. Chang, B. Mao and Y. Liu, "On the Mahalanobis distance classification criterion for a ventricular septal defect diagnosis system," *IEEE Sensors J.*, Vol. 19, no. 7, pp. 2665–2674, 2019.
[283] Z. Bai, X.-L. Zhang and J. Chen, "Speaker verification by partial AUC optimization with Mahalanobis distance metric learning," *IEEE/ACM Trans. Audio Speech Lang. Proces.*, Vol. 28, pp. 1533–1548, 2020.
[284] Q. Li, Z. Wu, L. Lin, J. Zeng, J. Zhang, H. Yan and S. Min, "High-level fusion coupled with Mahalanobis distance weighted (MDW) method for multivariate calibration," *Sci. Rep.*, Vol. 10, 5478, 2020; doi:10.1038/s41598-020-62396-y.
[285] K. Naveed and N. ur Rehman, "Wavelet based multivariate signal denoising using Mahalanobis distance and EDF statistics," *IEEE Trans. Signal Proces.*, Vol. 68, pp. 5997–6010, 2020.
[286] R. Wang, J.-W. Chen, Y. Wang, L. Jiao and M. Wang, "SAR image change detection via spatial metric learning with an improved Mahalanobis distance," *IEEE Geosci. Remote Sens. Lett.*, Vol. 17, no. 1, pp. 77–81, 2020.
[287] A.J. Winter, J.M.D. Kruijssen, S.N. Longmore and M. Chevance, "Stellar clustering shapes the architecture of planetary systems," *Nature*, Vol. 586, pp. 528–532, 2020, and Supplement on Methods.
[288] T.E. Bartlett, P. Jia, S. Chandna and S. Roy, "Inference of tissue relative proportions of the breast epithelial cell types luminal progenitor, basal, and luminal mature," *Sci. Rep.*, Vol. 11, 23702, 2021; doi:10.1038/s41598-021-03161-7.
[289] N.C. Butterfield, K.F. Curry, J. Steinberg , H. Dewhurst, D. Komla-Ebri, N.S. Mannan, A.-T. Adoum, V.D. Leitch, J.G. Logan, J.A. Waung, E. Ghirardello, L. Southam, S.E. Youlten, J.M. Wilkinson, E.A. McAninch, V.E. Vancollie, F. Kussy, J.K. White, C.J. Lelliott, D.J. Adams, R. Jacques, A.C. Bianco, A. Boyde, E. Zeggini, P.I. Croucher, G.R. Williams and J.H.D. Bassett, "Accelerating functional gene discovery in osteoarthritis," *Nature Commun.*, Vol. 12, 467, 2021; doi:0.1038/s41467-020-20761-5.
[290] M. Chamberland, S. Genc, C.M.W. Tax, D. Shastin, K. Koller, E.P. Raven, A. Cunningham, J. Dohert, M.B.M. van den Bree, G.D. Parker, K. Hamandi, W.P. Gray and D.K. Jones, "Detecting microstructural deviations in individuals with deep diffusion MRI tractometry," *Nature Commput. Sci.*, Vol. 1, 598–606, 2021;
[291] J.B. Kang , A. Nathan, K. Weinand, F. Zhang, N. Millard, L. Rumker, D.B. Moody, I. Korsunsky and S. Raychaudhuri, "Efficient and precise single-cell reference atlas mapping with Symphony," *Nature Commun.*, Vol. 12, 5890, 2021; doi:10.1038/s41467-021-25957-x.
[292] H. Sato, H. Ogihara, K. Takahashi, Y. Kawata, Y. Kojima, K. Tominaga, J. Yokoyama, Y. Hamamoto and S. Terai, "New cine magnetic resonance imaging parameters for the differential diagnosis of chronic intestinal pseudo-obstruction," *Sci. Rep.*, Vol. 11, 22974, 2021; doi:10.1038/s41598-021-02268-1.
[293] J. Zheng, F. Shen and L. Ye, "Improved Mahalanobis distance based JITL-LSTM soft sensor for multiphase batch processes" *IEEE Access*, Vol. 9, pp. 72172–72182, 2021.
[294] R. A, X. Mu and J. He, "Enhance tensor RPCA-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, Vol. 19, No. 6008305, 2022.
[295] A. Chakraborty, J.G. Wang and F. Ay, "dcHiC detects differential compartments across multiple Hi-C datasets," *Nature Commun.*, Vol. 13, 6827, 2022; doi:10.1038/10.1038/s41467-022-34626-6.
[296] Y. Chen, C. Han, J. He and G. Wang, "A framework of Mahalanobis-distance metric with supervised learning for clustering multipath components in MIMO channel analysis," *IEEE Trans. Antenn. Propag.*, Vol. 70, no. 6, pp. 4069–4081, 2022.
[297] R.F. dos Santos, M. Paraskevaidi, D.M.A. Mann, D. Allsop, M.C.D. Santos, C.L.M. Morais and K.M.G. Lima, "Alzheimer's disease diagnosis by blood plasma molecular fluorescence spectroscopy (EEM)," *Sci. Rep.*, Vol. 12, 16199, 2022; doi:10.1038/10.1038/s41598-022-20611-y.
[298] C.A. Guerra, M. Berdugo, D.J. Eldridge, N. Eisenhauer, B.K. Singh, H. Cui, S. Abades, F.D. Alfaro, A.R. Bamigboye, F. Bastida, J.L. Blanco-Pastor, A. de los Rios15, J. Durán, T. Grebenc, J.G. Illán, Y.-R. Liu, T.P. Makhalanyane, S. Mamet, M.A. Molina-Montenegro, J.L. Moreno, A. Mukherjee, T.U. Nahberger, G.F. Penaloza-Bojacá, César Plaza, Sergio Picó, J.P. Verma, A. Rey, A. Rodriguez, L. Tedersoo, A.L. Teixido, C. Torres-Diaz, P. Trivedi, J. Wang, L. Wang, J. Wang, E. Zaady, X. Zhou, X.-Q. Zhou and M. Delgado-Baquerizo, "Global hotspots for soil nature conservation," *Nature*, Vol. 610, pp. 693–698, 2022, and Supplement on Methods.
[299] H. Huang, J. Huang, X. Li, W. Zhuo, Y. Wu, Q. Niu, W. Su and W. Yuan, "A dataset of winter wheat aboveground biomass in China during 2007-2015 based on data assimilation," *Sci. Data*, Vol. 9, 200, 2022; doi:10.1038/s41597-022-01305-6.

[300] M. Nomoto, E. Murayama, S. Ohno, R. Okubo-Suzuki, S.-i. Muramatsu and K. Inokuchi, "Hippocampus as a sorter and reverberatory integrator of sensory inputs," *Nature Commun.*, Vol. 13, 7413, 2022; doi:10.1038/s41467-022-35119-2.
[301] L. Reichen, A.-M. Burgdorf, S. Brönnimann, J. Franke, R. Hand, V. Valler, E. Samakinwa, Y. Brugnara and T. Rutishauser, "A decade of cold Eurasian winters reconstructed for the early 19th century," *Nature Commun.*, Vol. 13, 2116, 2022; doi:10.1038/s41467-022-29677-8.
[302] S. Sun, T. Huang, B. Zhang, P. He, L. Yan, D. Fan, J. Zhang and J. Chen, "A novel intelligent system based on adjustable classifier models for diagnosing heart sounds," *Sci. Rep.*, Vol. 12, 1283, 2022; doi:10.1038/s41598-021-04136-4.
[303] A. Timmermann, K.-S. Yu, P. Raia, J. Ruan, A. Mondanaro, E. Zeller, C. Zollikofer, M.P. de León, D. Lemmon, M. Willeit and A. Ganopolski, "Climate effects on archaic human habitats and species successions," *Nature*, Vol. 604, pp. 495–501, 2022, and Supplement on Methods.
[304] H.S. Wauchope, J.P.G. Jones, J. Geldmann, B.I. Simmons, T. Amano, D.E. Blanco, R.A. Fuller, A. Johnston, T. Langendoen, T. Mundkur, S. Nagy and W.J. Sutherland, "Protected areas have a mixed impact on waterbirds, but management helps," *Nature*, Vol. 605, pp. 103–107, 2022, and Supplement on Methods.
[305] N. Weinberger, "Generalization bounds and algorithms for learning to communicate over additive noise channels," *IEEE Trans. Inf. Theory*, Vol. 68, no. 3, pp. 1886–1921, 2022.
[306] S. Wen, W. Guo, Y. Liu and R. Wu, "Rotated object detection via scale-invariant Mahalanobis distance in aerial images," *IEEE Geosci. Remote Sens. Lett.*, Vol. 19, No. 6514505, 2022.
[307] B. Yang, T. Karigo and D.J. Anderson, "Transformations of neural representations in a social behaviour network," *Nature*, Vol. 608, pp. 741–749, 2022, and Supplement on Methods.
[308] M. Zhang, Y. Zhang and G. Shen, "PPDDS: a privacy-preserving disease diagnosis scheme based on the secure Mahalanobis distance evaluation model," *IEEE Systems J.*, Vol. 16, no. 3, pp. 4552–4562, 2022.
[309] S. Burssens, D.M. Bowman, M. Michielsen, S. Simón-Diaz, C. Aerts, V. Vanlaer, G. Banyard, N. Nardetto, R.H.D. Townsend, G. Handler, J.S.G. Mombarg, R. Vanderspek and G. Ricker, "A calibration point for stellar evolution from massive star asteroseismology" *Nature Astronom.*, Vol. 7, pp. 913—930, 2023; doi:10.1038/s41550-023-01978-y.
[310] K. Choi, S. Woo and J. Lee, "Motor-effector dependent modulation of sensory-motor processes identified by the multivariate pattern analysis of EEG activity," *Sci. Rep.*, Vol. 13, 3161, 2023; doi:10.1038/s41598-023-30324-5.
[311] H.-K. Choi, P. Cong, C. Ge, A. Natarajan, B. Liu, Y. Zhang, K. Li, M.N. Rushdi, W. Chen, J. Lou, M. Krogsgaard and C. Zhu, "Catch bond models may explain how force amplifies TCR signaling and antigen discrimination," *Nature Commun.*, Vol. 14, 2616, 2023; doi:10.1038/s41467-023-38267-1.
[312] J.L. Dahlin, B.K. Hua, B.E. Zucconi, S.D. Nelson Jr, S. Singh, A.E. Carpenter, J.H. Shrimp, E. Lima-Fernandes, M.J. Wawer, L.P.W. Chung, A. Agrawal, M. OŔeilly, D. Barsyte-Lovejoy, M. Szewczyk, F. Li, P. Lak, M. Cuellar, P.A. Cole, J.L. Meier, T. Thomas, J.B. Baell, P.J. Brown, M.A. Walters, P.A. Clemons, S.L. Schreiber and B.K. Wagner, "Reference compounds for characterizing cellular injury in high-content cellular morphology assays," *Nature Commun.*, Vol. 14, 1364, 2023; doi:10.1038/s41467-023-36829-x.
[313] A.S. Ebrahimi, P. Orlowska-Feuer, Q. Huang, A.G. Zippo, F.P. Martial, R.S. Petersen and R. Storchi, "Three-dimensional unsupervised probabilistic pose reconstruction (3D-UPPER) for freely moving animals," *Sci. Rep.*, Vol. 13, 155, 2023; doi:10.1038/s41598-022-25087-4.
[314] W. Jeong, S. Kim, J. Park and J. Lee, "Multivariate EEG activity reflects the Bayesian integration and the integrated Galilean relative velocity of sensory motion during sensorimotor behavior," *Commun. Biol.*, Vol. 6, 113, 2023; doi:10.1038/s42003-023-04481-2.
[315] M. Kim, K.-R. Moon and B.-D. Lee, "Motor-effector dependent modulation of sensory-motor processes identified by the multivariate pattern analysis of EEG activity," *Sci. Rep.*, Vol. 13, 3415, 2023; doi:10.1038/s41598-023-30589-w.
[316] A.J. Nowakowski, S.W.J. Canty, N.J. Bennett, C.E. Cox, A. Valdivia, J.L. Deichmann, T.S. Akre, S.E. Bonilla-Anariba, S. Costedoat and M. McField, "Co-benefits of marine protected areas for nature and people," *Nature Sustainab.*, Vol. 6, No. 10, pp. 1210–1218, 2023; doi:110.1038/s41893-023-01150-4.
[317] C. Qu, Y. Zhang, K. Huang, S. Wang and Y. Yang, "Point clouds outlier removal method based on improved Mahalanobis and completion," *IEEE Robot. Automat. Lett.*, Vol. 8, no. 1, pp. 17–24, 2023.
[318] Y.W. Rabby, Y. Li and H. Hilafu, "An objective absence data sampling method for landslide susceptibility mapping," *Sci. Rep.*, Vol. 13, 1740, 2023; doi:10.1038/10.1038/s41598-023-28991-5.
[319] J. Sarno, P. Domizi, Y. Liu, M. Merchant, C.B. Pedersen, D. Jedoui, A. Jager, G.P. Nolan, G. Gaipa, S.C. Bendall, F.-A. Bavaand K.L. Davis, "Dasatinib overcomes glucocorticoid resistance in B-cell acute lymphoblastic leukemia," *Nature Commun.*, Vol. 14, 2935, 2023; doi:10.1038/s41467-023-38456-y.
[320] Y. Tang, Y. Zhou, X. Ren, Y. Sun, Y. Huang and D. Zhou, "A new basic probability assignment generation and combination method for conflict data fusion in the evidence theory," *Sci. Rep.*, Vol. 13, 8443, 2023; doi:0.1038/s41598-023-35195-4.
[321] A. Tsvieli and N. Weinberger, "Learning maximum margin channel decoders," *IEEE Trans. Inf. Theory*, Vol. 69, no. 6, pp. 3597–3626, 2023.
[322] J. Zhang, Y. Feng, F.T. Maestre M. Berdugo, J. Wang, C. Coleine, T. Sáez-Sandino, L. Garcia-Velázquez, B.K. Singh and M. Delgado-Baquerizo, "Water availability creates global thresholds in multidimensional soil biodiversity and functions," *Nature Ecol. Evol.*, Vol. 7, pp. 1002–1011, 2023, and Supplement.
[323] M. Zhou, C. Shang, G. Li, L. Shen, N. Naik, S. Jin, J. Peng and Q. Shen, "Transformation-based fuzzy rule interpolation with Mahalanobis distance measures supported by Choquet integral," *IEEE Trans. Fuzzy Syst.*, Vol. 31, no. 4, pp. 1083–1097, 2023.
[324] K.R. Varshney and L.R. Varshney, "Optimal grouping for group minimax hypothesis testing," *IEEE Trans. Inf. Theory*, Vol. 60, no. 10, pp. 6511–6521, 2014.
[325] Z. Hu, Y. Zhu, J. Xu and Y. Yang, "Bregman-based inexact excessive gap method for multiservice resource allocation," *IEEE Trans. Wireless Commun.*, Vol. 14, no. 2, pp. 1115–1130, 2015.
[326] R. Nock, W. Bel Haj Ali, R. D'Ambrosio, F. Nielsen and M. Barlaud, "Gentle nearest neighbors boosting over proper scoring rules," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 37, no. 1, pp. 80–93, 2015.
[327] G. Raskutti and S. Mukherjee, "The information geometry of mirror descent," *IEEE Trans. Inf. Theory*, Vol. 61, no. 3, pp. 1451–1457, 2015.
[328] H. Wang, X. Liao, T. Huang and C. Li, "Cooperative distributed optimization in multiagent networks with delays," *IEEE Trans. Syst. Man Cyb.*, Vol. 45, no. 2, pp. 363–369, 2015.
[329] W. He, J.T.-Y. Kwok, J. Zhu and Y. Liu, "A note on the unification of adaptive online learning," *IEEE Trans. Neural Netw. Learn. Syst.*, Vol. 28, no. 5, pp. 1178–1191, 2017.
[330] X. Li, T. Pi, Z. Zhang, X. Zhao, M. Wang, X. Li and P.S. Yu, "Learning Bregman distance functions for structural learning to rank," *IEEE Trans. Knowledge Data Engin.*, Vol. 29, no. 9, pp. 1916–1927, 2017.
[331] P. Harremoës, "Entropy on spin factors," in: N. Ay et al. (eds.), *Information Geometry and Its Applications*, Springer Proceedings in Mathematics & Statistics, vol. 252, pp. 247—278. Cham, Switzerland: Springer Nature Switzerland, 2018.
[332] J. Xu, S. Zhu, Y. C. Soh and L. Xie, "A Bregman splitting scheme for distributed optimization over networks," *IEEE Trans. Autom. Control*, Vol. 63, no. 11, pp. 3809–3824, 2018.
[333] A. Halder, "DeGroot–Friedkin map in opinion dynamics is mirror descent," *IEEE Control Syst. Lett.*, Vol. 3, no. 2, pp. 463–468, 2019.
[334] Q. Zhang, Y. Zhang, Y. Huang and Y. Zhang, "Azimuth superresolution of forward-looking radar imaging which relies on linearized Bregman," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, Vol. 12, no. 7, pp. 2032–2043, 2019.
[335] X. Shao, X. Chen, D.W.K. Ng, C. Zhong and Z. Zhang, "Cooperative activity detection: sourced and unsourced massive random access paradigms," *IEEE Trans. Signal Proces.*, Vol. 68, pp. 6578–6593, 2020.
[336] H. Tembine, "Deep learning meets game theory: Bregman-based algorithms for interactive deep generative adversarial networks," *IEEE Trans. Cybernetics*, Vol. 50, no. 3, pp. 1132–1145, 2020.
[337] C. Brécheteau, A. Fischer and C. Levrard, "Robust Bregman clustering," *Ann. Statist.*, Vol. 49, no. 3, pp. 1679–1701, 2021.

[338] C.-W. Lin, S.-H. Liao, H.-S. Huang, L.-M. Wang, J.-H. Chen, C.-H. Su and K.-L. Chen, "Improvement of multisource localization of magnetic particles in an animal," *Sci. Rep.*, Vol. 11, 9628, 2021; doi:110.1038/s41598-021-88847-8.
[339] D. Yuan, Y. Hong, D.W.C. Ho and S. Xu, "Convergence of best $\phi-$entropy estimates," *IEEE Trans. Autom. Control*, Vol. 66, no. 2, pp. 714–729, 2021.
[340] N. Azizan, S. Lale and B. Hassibi, "Stochastic mirror descent on overparameterized nonlinear models," *IEEE Trans. Neural Netw. Learn. Syst.*, Vol. 33, no. 12, pp. 7717–7727, 2022.
[341] A. Dytso, M. Fauß and H.V. Poor, "Bayesian risk with Bregman loss: A Cramér–Rao type bound and linear estimation," *IEEE Trans. Inf. Theory*, Vol. 68, no. 3, pp. 1985–2000, 2022.
[342] T.V. Gruzdeva and A.V. Ushakov, "On a nonconvex distance-based clustering problem," in: P. Pardalos et al. (eds.), *Mathematical Optimization Theory and Operations Research MOTOR 2022*, Lecture Notes in Computer Science, vol. 13367, pp. 139–152. Cham, Switzerland: Springer Nature Switzerland, 2022.
[343] Y. Song, Y. Gu, R. Zhang and G. Yu, "BrePartition: optimized high-dimensional kNN search with Bregman distances," *IEEE Trans. Knowledge Data Engin.*, Vol. 34, no. 3, pp. 1053–1065, 2022.
[344] Z. Yu, D.W.C. Ho and D. Yuan, "Distributed randomized gradient-free mirror descent algorithm for constrained optimization," *IEEE Trans. Autom. Control*, Vol. 67, no. 2, pp. 957–964, 2022.
[345] M. Capó, A. Pérez and J.A. Lozano, "Fast computation of cluster validity measures for Bregman divergences and benefits," *Pattern Recogn. Lett.*, Vol. 170, pp. 100–105, 2023.
[346] G. Chen, G. Xu, W. Li and Y. Hong, "Distributed mirror descent algorithm With Bregman damping for nonsmooth constrained optimization," *IEEE Trans. Autom. Control*, Vol. 68, no. 11, pp. 6921–6928, 2023.
[347] J.D. Fernández-Rodriguez, E.J. Palomo, J. Benito-Picazo, E. Dominguez, E. Lóopez-Rubio and F. Ortega-Zamorano, "A convolutional autoencoder and a neural gas model based on Bregman divergences for hierarchical color quantization," *Neurocomputing*, Vol. 544, No. 126288, 2023; doi:10.1016/j.neucom.2023.126288.
[348] M. Hayashi, "Bregman divergence based Em algorithm and its application to classical and quantum rate distortion theory," *IEEE Trans. Inf. Theory*, Vol. 69, no. 6, pp. 3460–3492, 2023.
[349] Z. Li and A. Ralescu, "Generalized self-supervised contrastive learning with Bregman divergence for image recognition," *Pattern Recogn. Lett.*, Vol. 171, pp. 155–161, 2023.
[350] M. Xiong, B. Zhang, D.W.C. Ho, D. Yuan and S. Xu, "Event-triggered distributed stochastic mirror descent for convex optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, Vol. 34, no. 9, pp. 6480–6491, 2023.
[351] J. Liu, Z. Yu and D.W.C. Ho, "Distributed constrained optimization with delayed subgradient information over time-varying network under adaptive quantization," *IEEE Trans. Neural Netw. Learn. Syst.*, Vol. 35, no. 1, pp. 143–156, 2024.
[352] H.H. Bauschke and J.M. Borwein, "Legendre functions and the method of random Bregman projections," *J. Convex Anal.*, Vol. 4, no. 1, pp. 27–67, 1997.
[353] I.S. Dhillon and J.A. Tropp, "Matrix nearness problems with Bregman divergences," *SIAM J. Matrix Anal. Appl.*, Vol. 29, no. 4, pp. 1120–1146, 2007.
[354] B. Kulis, M.A. Sustik and I.S. Dhillon, "Low-rank kernel learning with Bregman matrix divergences," *J. Machine Learn. Res.*, Vol. 10, pp. 341–376, 2009.
[355] D.L. Donoho, M. Elad and V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, Vol. 52, no. 1, pp. 6–18, 2006.
[356] E.J.Candès, J.K. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, Vol. LIX, pp. 1207–1223, 2006.
[357] M. Lustig, D. Donoho and J.M. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Reson. Med.*, Vol. 58, pp. 1182–1195, 2007.
[358] E.J.Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, Vol. 346, pp. 589–592, 2008.
[359] E.J.Candès, M.B. Wakin and S.P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, Vol. 14, pp. 877–905, 2008.
[360] T. Goldstein and S. Osher, "The split Bregman method for $L1-$regularized problems," *SIAM J. Imaging Sci.*, Vol. 2, no. 2, pp. 323–343, 2009.
[361] Y. Zhang, B.S. Peterson, G. Ji and Z. Dong, "Energy preserved sampling for compressed sensing MRI," *Comput. Mathem. Meth. Med.*, Vol. 2014, No. 546814, 2014; doi:10.1155/2014/546814.
[362] M.P. Edgar, G.M. Gibson and M.J. Padgett, "Principles and prospects for single-pixel imaging," *Nature Photonics*, Vol. 13, pp. 13–20, 2019.
[363] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q--minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, Vol. 26, pp. 395–407, 2009.
[364] J. Liu, J. Jin and Y. Gu, "Robustness of sparse recovery via F-minimization: a topological viewpoint," *IEEE Trans. Inf. Theory*, Vol. 61, no. 7, pp. 3996–4014, 2015.
[365] A.M. Bruckstein, D.L. Donoho and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, Vol. 51, no. 1, pp. 34–81, 2009.
[366] B.G. Lindsay, "Statistical distances as loss functions in assessing model adequacy," in: M.P. Taper and S.R. Lele (eds.), *The Nature of Scientific Evidence*, pp. 439–487. Chicago, IL, USA: The University of Chicago Press, 2004. This includes comments by D.R. Cox and S.P. Ellner as well as a rejoinder by the author.
[367] B.G. Lindsay, M. Markatou, S. Ray, K. Yang, and S.-C. Chen, "Quadratic distances on probabilities: a unified foundation," *Ann. Statist.*, Vol. 36, no. 2, pp. 983–1006, 2008.
[368] M. Markatou and E.M. Sofikitou, "Statistical distances and the construction of evidence functions for model adequacy," *Front. Ecol. Evol.*, Vol. 7, No. 447, 2019; doi: 10.3389/fevo.2019.00447.
[369] M. Broniatowski and A. Keziou, "Divergences and duality for estimation and test under moment condition models," *J. Statistical Planning and Inference*, Vol. 142, pp. 2554–2573, 2012.
[370] M. Broniatowski and A. Decurninge, "Estimation for models defined by conditions on their L-moments," *IEEE Trans. Inf. Theory*, Vol. 62, no. 9, pp. 5181–5198, 2016.
[371] A.W. Van der Vaart, *Asymptotic Statistics, 8th Printing*. Cambridge, UK: Cambridge University Press, 2007.
[372] A.K. Kuchibhotla and A. Basu, "On the asymptotics of minimum disparity estimation," *TEST*, Vol. 26, pp. 481–502, 2017.
[373] T. Franco, "A mini-course in large deviations," *https://w3.impa.br/∼tertu/archives/LDP_Notes_2015.pdf*, 49 pages. Salvador, Brazil: Universidade Federal da Bahia (UFBA), March 2015.
[374] J. Swart, "Large deviation theory, 6th edition," *http://staff.utia.cas.cz/swart/lecture_notes/LDP23_04_15.pdf*, 179 pages. Prague, Czech Republic: Institute of Information Theory and Automation (UTIA), The Czech Academy of Sciences, April 2023.
[375] F. Topsoe, "Some bounds for the logarithmic function," in: Y.J. Cho, J.K. Kim and S.S. Dragomir (eds.), *Inequality Theory and Applications 4*, pp. 137—151. New York, USA: Nova Sci. Publ., 2007.