



Relative Variability Indexes for Nonnegative Orthant Distributions

Aboubacar Yacouba Touré

► To cite this version:

Aboubacar Yacouba Touré. Relative Variability Indexes for Nonnegative Orthant Distributions. Statistics [math.ST]. Université Bourgogne Franche Comté, 2021. English. NNT: . tel-03654853

HAL Id: tel-03654853

<https://hal.science/tel-03654853>

Submitted on 6 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE L' UNIVERSITÉ
BOURGOGNE FRANCHE-COMTÉ

ÉCOLE DOCTORALE n°553

CARNOT-PASTEUR

DOCTORAT DE MATHÉMATIQUES

SPÉCIALITÉ : STATISTIQUE

Par

M. ABOUBACAR YACOUBA TOURÉ

**RELATIVE VARIABILITY INDEXES FOR
NONNEGATIVE ORTHANT DISTRIBUTIONS
INDICES DE VARIABILITÉ RELATIVE POUR LES DISTRIBUTIONS
CONTINUES POSITIVES**

Thèse présentée et soutenue à Besançon, le 18/11/2021

Composition du Jury :

Prof. Jean-François DUPUY	Institut National des Sciences Appliquées, Rennes	Président
Prof. Cédric HEUCHENNE	University of Liège, Liège	Rapporteur
Prof. Christian H. WEISS	Helmut Schmidt University, Hamburg	Rapporteur
Prof. Amir ABOUBACAR	University of Liège, Liège	Examineur
Ass. Prof. Rahma ABID	Université Paris Dauphine de Tunis, Tunis	Examineur
Prof. Célestin C. KOKONENDJI	Université Bourgogne Franche-Comté, Besançon	Encadreur



Abstract

From the classical dispersion indexes with respect to the Poisson and binomial distributions for count data and the exponential variation index for positive continuous data, we first introduce the unified definition of the relative variability to a nonnegative natural exponential family through its variance function. Then, we show the asymptotic normality of the corresponding test statistics and give some applicable examples. Simulations have pointed out good behaviours of these asymptotic test statistics. Next, general weighted exponential distributions which are widely used with great ability in reliability are characterized. Indeed, we investigate full exponential weight functions and their extensions from any nonnegative continuous reference distribution. Several properties and their connections with the new variation phenomenon are then established. We also introduce some new indexes to measure the departure of any multivariate continuous distribution on the nonnegative orthant from a given reference distribution. The reference distribution may be an un(correlated) exponential model. The proposed multivariate variation indexes are scalar quantities, defined as ratio of two quadratic forms of the mean vector to the covariance matrix. The asymptotic behaviors and other properties are studied. Illustrative examples as well as numerical applications are analyzed under several scenarios, leading to appropriate choices of multivariate models. To finish this work, an R package has been implemented to calculate the dispersion and variation indexes in both univariate and multivariate cases. Concluding remarks, including future directions are displayed.

Keywords : Continuous distribution, count distribution, dependence, dual distribution, equi-variation, exponential dispersion models, natural exponential family, over-variation, reliability, under-variation, variance function, weighted distribution.

Mathematics Subject Classification (2020) : 60E05 ; 62E10 ; 62E15.

Résumé

À partir des indices de dispersion relatifs aux lois de Poisson et binomiale pour les données de comptage et de l'indice de variation exponentielle pour les données continues positives, nous introduisons d'abord la définition unifiée de l'indice de variabilité relative à une famille exponentielle naturelle positive à travers sa fonction variance. Ensuite, nous montrons la normalité asymptotique des statistiques de tests correspondantes et donnons des exemples applicables. Des études de simulations ont mis en évidence de bons comportements de ces statistiques de tests asymptotiques. Aussi, les lois exponentielles pondérées générales qui contiennent les lois exponentielles modifiées et qui sont largement utilisées en fiabilité ont été caractérisées. En effet, nous étudions leurs fonctions poids exponentiels et des extensions à partir d'une loi de référence continue positive. Des propriétés et leurs relations avec le nouveau phénomène de variation sont établies. Nous introduisons également de nouveaux indices pour discriminer toute distribution continue multivariée sur la demi droite positive à partir d'une distribution de référence donnée. La distribution de référence peut être un modèle exponentiel non corrélé. Des exemples illustratifs ainsi que des applications numériques sont analysés sous plusieurs scénarios, conduisant à des choix appropriés de modèles multivariés. Pour finir ce travail, un package R a été implanté pour calculer les indices de dispersion et de variation aussi bien dans le cas uni varié que multivarié. Nous terminons par des remarques finales, y compris les futures directions.

Mots clés : Dépendance, distribution de comptage, distribution continue, distribution duale, distribution pondérée, équi-variation, famille exponentielle naturelle, fiabilité, fonction variance, model exponentiel de dispersion, sous-variation, sur-variation.

Codes AMS (2020) : 60E05; 62E10; 62E15.

Remerciements

Je remercie profondément, en premier, mon directeur de thèse, Professeur **Célestin C. KOKONENDJI**, pour ce sujet épanouissant, sa rigueur professionnelle, sa disponibilité permanente et son enthousiasme débordant. Vous avez su m'aider à faire évoluer et fructifier mes recherches durant ces trois dernières années grâce à votre encadrement rigoureux. Je vous remercie pour la confiance que vous m'avez accordée, votre aide, votre soutien, vos conseils précieux et votre suivi ponctuel : *Singîla mîngi*.

J'exprime toute ma gratitude aux rapporteurs de ma thèse Prof. Cédric HEUCHENNE et Prof. Christian H. WEISS pour leurs critiques et orientations qui m'ont permis d'améliorer ce travail de recherche. Mes remerciements également à Prof. Amir ABOUBACAR et Ass. Prof. Rahma ABID pour leurs rôles respectueux et si précieux. Je remercie aussi Prof. Jean-François DUPUY pour la présidence du jury de cette thèse. Mes remerciements sincères vont également à mes collègues et amis de la faculté des sciences économiques et de gestion de Bamako (FSEG) notamment à M. Moumini Diallo et M. Ladj Kane qui m'ont beaucoup aidé dans le cadre de mes séjours en France. Une mention spéciale à M. Moumini Diallo qui m'a permis de prendre contact avec mon directeur de thèse. Je remercie également tout le personnel du LmB ainsi que tous ceux qui m'ont aidé de près ou de loin pour la réussite de ce travail passionnant. Enfin, les mots les plus simples étant les plus forts, j'adresse toute mon affection à ma famille, en particulier mes parents, mes frères et sœurs ainsi que ma femme pour leurs soutiens constants, leurs encouragements et leurs patiences tout au long de la thèse.

List of figures

2.1	A use of the exponential variation index in reliability : "Bathtub curve" of the failure rate as function of time.	45
3.1	"Bathtub curve" of the failure rate as function of time.	51
3.2	Pointwise duality of Definition 3.3.7 and Part (ii) of Remark 3.3.8 (with $\mu = 1$) in the gamma model.	56
4.1	Boxplots for the targets GVI and MVI with 100 replicates according to sample size and simulation parameters of Tables 4.5, 4.6 and 4.7.	86
4.2	Support of bivariate distributions with maximum correlations (positive in blue and negative in red) : model on \mathbb{R}^2 (left) and also finite support; model on $[0, \infty)^2$ (right), without finite support (Cuenin et al., 2016). . . .	92

List of tables

2.1	Summary of the six types of quadratic variance functions on M with support \mathcal{S} of distributions and set Λ of power of convolution (Morris, 1982); GHS stands for generalized hyperbolic secant.	31
2.2	Summary of the power variance functions $V(\mu) = \mu^p$ on M with support \mathcal{S} of distributions and $\Lambda = (0, \infty)$; see, e.g., Tweedie (1984), Jørgensen (1997, Chap. 4).	31
2.3	Some values of standard variability indexes WI (i.e., DI or VI) for the reference models.	39
2.4	Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 30$.	39
2.5	Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 50$.	40
2.6	Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 100$	40
2.7	Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 300$	41
2.8	Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 500$	41
2.9	Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 1000$	42

2.10	Asymptotic and asymmetric bootstrap confidence intervals for the Poisson dispersion (DI) and exponential variation (VI) indexes from sample size n simulated with $u = u_{0.975} = 1.96$ and $\alpha = 0.05$	44
2.11	Daily numbers of deaths of women by brain vessel disease in West Berlin, 1989 (Böhning, 1994).	46
3.1	Examples of WEDs in nonnegative continuous EDMs for using Proposition 3.4.5.	60
3.2	Summary of twelve WEDs from (3.1) and (3.2) with $X \sim \mathcal{E}(\mu)$	62
4.1	Some values of equivalences (4.10) for the univariate Weibull distribution.	75
4.2	Some scenarios from fifteen simulated bivariate datasets with $\widehat{VI}_j = \widehat{\sigma}_j^2 / \bar{y}_j^2$ and marginal variations (MV) : Over- (O), Equi- (E) and Under-variation (U).	81
4.3	Summary of real 4-variate data with $\widehat{VI}_j = \widehat{\sigma}_j^2 / \bar{y}_j^2$ and the marginal variation (MV _j) : Under-variation (U), size $n = 90$, $\det \widehat{\rho} = 0.0003$, $\widehat{GVI} = 0.1397$ and $\widehat{MVI} = 0.0771$	83
4.4	Summary of simulated 6-variate data with marginal variations (MV) : Over- (O), Equi- (E) and Under-variation (U), size $n = 560$ and $\det \widehat{\rho} = 0.2063$ such that $\widehat{GVI} = 1.0572 \approx 1$ and $\widehat{MVI} = 0.9637 \approx 1$	84
4.5	Asymptotic variances and confidence intervals ($u = u_{0.975} = 1.96$) from subsamples of simulated 6-variate data with $n = 10\,000$ having the same parameters as for Table 4.4.	84
4.6	Asymptotic variances and confidence intervals ($u = u_{0.975} = 1.96$) from subsamples of simulated over-varied 4-variate data with $n = 10\,000$	85
4.7	Asymptotic variances and confidence intervals ($u = u_{0.975} = 1.96$) from subsamples of simulated under-varied trivariate data with $n = 10\,000$	85

4.8	Asymptotic and asymmetric bootstrap confidence intervals for GVI and MVI indexes from sample size n simulated with $u = u_{0.975} = 1.96$ and $\alpha = 0.05$ using the dataset of Table 4.4.	87
4.9	Drinking water pumps trivariate data measured in the Sahel with $n = 42$	89
4.10	Empirical univariate (in diagonal), bivariate (off diagonal) and trivariate (at the corner) variation and dispersion indexes.	90

Table of contents

1	General introduction	23
2	The unified relative dispersion and relative variation indexes	29
2.1	Introduction	29
2.2	NEF and the new RWI	30
2.3	Asymptotic normality of the test statistics	33
2.4	Illustrations and simulation studies	35
2.4.1	Examples of tests and comments	36
2.4.2	Simulation studies	37
2.5	Examples of Application	44
2.6	Conclusion	46
3	General exponential weight functions and variation phenomenon	47
3.1	Introduction	47
3.2	Background in reliability	49
3.2.1	Distribution function	49
3.2.2	Reliability function and failure rate	49
3.3	Results about nonnegative continuous weighted distributions	52
3.4	WEDs in EDMs and variation phenomenon	57

3.5	Some existing WEDs in literature	60
3.5.1	Exponential weight of Gupta and Kundu (2001)	63
3.5.2	Modified exponential weight of Gupta and Kundu (2009)	64
3.6	Illustrations	64
3.6.1	Gamma and modified gamma distributions	65
3.6.2	Weibull and some of its modified distributions	66
3.6.3	Lognormal distribution as a WED	67
3.7	Conclusion	67
4	Relative variation indexes for distributions on $[0, \infty)^k$ and extensions	69
4.1	Introduction	69
4.2	Multivariate variation indexes	70
4.2.1	Basic definitions	70
4.2.2	Interpretation and properties	71
4.3	Illustrations and comments	73
4.3.1	Bivariate beta distribution of Arnold and Tony Ng (2011)	73
4.3.2	Bivariate Weibull distribution of Teimouri and Gupta (2011)	74
4.3.3	Multivariate exponential distribution of Marshall and Olkin (1967)	76
4.3.4	Multiple stable Tweedie (MST) models	76
4.4	Estimation and asymptotic properties	78
4.5	Numerical applications	80
4.5.1	Some scenarios of bivariate cases and a real 4-variate dataset	80
4.5.2	Other multivariate cases and simulation studies	83

4.5.3	Numerical application of variation and dispersion indexes in uni- variate, bivariate and multivariate setups	88
4.6	Unification of GVI and GDI	90
4.7	Conclusion	90
4.8	Complement and proofs	91
5	GWl : an R package for count and continuous generalized variability indexes	95
5.1	Introduction	95
5.2	Funtion for Poisson dispersion index	96
5.3	Function for exponential variation index	97
5.4	Funtion for binomial dispersion index	98
5.5	Funtion for negative binomial dispersion index	99
5.6	Funtion for inverse Gaussian variation index	100
5.7	Function for GDI and MDI	101
5.8	Function for GVI and MVI	102
5.9	Conclusion	103
6	Conclusions and perspectives	105
	References	109
	Appendix	119
6.1	R codes for simulation studies	119
6.1.1	Script in R for making bootstrap with bivariate GVI and MVI . .	119
6.1.2	Simulated continuous data of Table 4 from the NORTA Algorithm	121

6.1.3	Scripts in R for asymptotic variances and confidence intervals of GVI and MVI	122
6.2	Transition from discret case to continuous case	131

List of symbols and abbreviations

CI_a	Asymptotic confidence interval
RDI_B	Binomial dispersion index
CI_b	Bootstrapped confidence interval
ρ	Correlation coefficient
cov	Covariance
DFRA	Decreasing failure rate on the average
\mathbb{E}	Expectation
EDM	Exponential dispersion model
VI	Exponential variation index
GDI	Generalized dispersion index
GVI	Generalized variation index
IFRA	Increasing failure rate on the average
MDI	Marginal dispersion index
MVI	Marginal variation index
MST	Multiple stable Tweedie models
NEF	Natural exponential family
log	Natural logarithm function
NIG	Normal inverse Gaussian
NST	Normal stable Tweedie
DI	Poisson dispersion index
pdf	Probability density function
pmf	Probability mass function
RDI	Relative dispersion index
RWI	Relative variability index
RVI	Relative variation index
\mathbb{N}_0	Set of nonnegative integers
\mathbb{N}	Set of positive integers
\mathbb{R}	Set of real numbers
Var	Variance
WED	Weighted exponential distribution

Scientific productions

i. Published journal papers

- [1] **Touré, A.Y.**, Dossou-Gbété, S. and Kokonendji, C.C. (2020). Asymptotic normality of the test statistics for the unified relative dispersion and relative variation indexes, *Journal of Applied Statistics* **47**, 2479-2491.
- [2] Kokonendji, C.C., **Touré, A.Y.** and Sawadogo, A. (2021a). Relative variation indexes for multivariate continuous distributions on $[0, \infty)^k$ and extensions, *AStA Advances in Statistical Analysis* **104**, 285-307.
- [3] Kokonendji, C.C., **Touré, A.Y.** and Abid, R. (2021b). On general exponential weight functions and variation phenomenon, *Sankhya A*, in press (DOI : 10.1007/s13171-020-00226-z).
- [4] **Touré, A.Y.** and Kokonendji, C.C. (2021). GWI : Count and Continuous Generalized Variability Indexes, <https://CRAN.R-project.org/package=GWI>.
- [5] **Touré, A.Y.**, Dossou-Gbété, S. and Kokonendji, C.C. (2021). Correction to : Asymptotic normality of the test statistics for the unified relative dispersion and relative variation indexes, *Journal of Applied Statistics*, Submitted.

iii. Oral communications

- [1] "Asymptotic normality of the test statistics for the unified relative dispersion and relative variation indexes". *Séminaire de Probabilités et Statistique du LmB*, November 04 2019, Besançon, France.

- [2] "Relative variation indexes for multivariate continuous distributions on $[0, \infty)^k$ and extensions". *Séminaire de Probabilités et Statistique du LmB*, November 18 2019, Besançon, France.
- [3] "On general exponential weight functions and variation phenomenon". *Séminaire de Probabilités et Statistique du LmB*, December 14 2020, Besançon, France.
- [4] **Touré, A.Y.**, Dossou-Gbété, S. and Kokonendji, C.C. (2020). Normalité asymptotique des statistiques de tests des indices relatifs de dispersion et de variation. *Les 51èmes Journées de Statistiques (JdS 2019) de la SFdS*, 03-07 June 2019, Nancy, France.
- [5] **Touré, A.Y.**, Dossou-Gbété, S. and Kokonendji, C.C. (2020). Asymptotic normality of the test statistics for the unified relative dispersion and relative variation indexes. *1re École d'été de Statistique et Sciences des Données organisés par la SFdS et AIMS*, 01-05 July 2019, M'bour, Sénégal.
- [6] **Touré, A.Y.**, Kokonendji, C.C. and Sawadogo, A. (2020). Relative variation indexes for positive continuous distributions, *13th International Conference of the ERCIM WG on CMStatistics*, 19-21 December 2020, Virtual Conference.
- [7] Kokonendji, C.C., **Touré, A.Y.** and Abid, R. (2021). Sur les fonctions poids exponentiels et le phénomène de variation. *Les 52èmes Journées de Statistiques (JdS 2021) de la SFdS*, 07-11 June 2021, Nice, France.

Chapter 1

General introduction

Statistical modeling is an important step of data analysis in many fields of scientific research or decision making purpose. To carry out this gait, one needs to specify a probability distribution that accounts as accurately as possible the variability observed in data. The infinite number of probability distributions requires the study of different indexes for comparisons between them. Given the plethora of discrete or continuous distributions available (e.g., Johnson et al., 1994, 2005), guidelines are needed for identifying not by chance the one or two-parameter family of distributions suited for modeling data on hand. The practice in this procedure is to consider in addition certain phenomenon such as over-/under-dispersion or zero inflation/deflation for count data (e.g., Bonat et al., 2018) and, over-/under-variation or zero mass for continuous data (e.g., Abid et al., 2020). In the analysis of the count data, the presence of at least overdispersion or excess of zeros deserves special attention for the choice of the count model (e.g., Demétrio and Hinde, 1998; Akantziliotou et al., 2008; Sellers and Raim, 2016; del Castillo and Pérez-Casany, 2005). The overdispersion phenomenon which implies the fact that the variance (observed) is greater than the mean (expected variance) can also be induced by the zero-inflation one (or excess of zeros) in the sample or the heavy tail. Both measures are commonly performed with respect to the Poisson distribution and, frequently, the negative binomial and its zero-inflated version are used for modelling these kinds of count datasets. The phenomenon of overdispersion, often results from unobserved heterogeneity, i.e., the sample of responses is drawn from a population

consisting of several sub-populations. Many count models have been built through compounding and mixing Poisson or negative binomial distributions; see, e.g., Hougaard et al. (1997), Kokonendji et al. (2004; 2007), Bonat et al. (2018), Wang (2011) and Gupta et al. (2014). We also can refer to Abid and Kokonendji (2021) who invested the ratio of the maximized likelihoods within each subclass of these two models for discrimination purposes based on classic measures of dispersion, zero-inflation and heavy tail of both Poisson-Tweedie and Poisson-exponential-Tweedie models.

The Poisson dispersion phenomenon is well-known and very widely used in practice; see, e.g., Kokonendji (2014) for a review of count (or discrete integer-valued) models. There are many interpretable mechanisms leading to this phenomenon which makes it possible to classify count distributions and make inference; see, e.g., Avelino (1978), Böhning (1994) and Mizère et al. (2006) for approximative statistical tests. Introduced from Fisher (1934), the Poisson dispersion index (DI), also called the Fisher dispersion index, of a count random variable X on $\mathcal{S} = \{0, 1, 2, \dots\} =: \mathbb{N}_0$ can be defined as $DI(X) = \text{Var}X / \mathbb{E}X$, the ratio of variance to mean. In fact, the positive quantity $DI(X)$ is the ratio of two variances since $\mathbb{E}X$ is the expected variance under the Poisson distribution. Hence, one easily deduces the concept of the relative dispersion index (denoted by RDI) by choosing another reference than the Poisson distribution. Indeed, if X and Y are two count random variables on the same support $\mathcal{S} \subseteq \mathbb{N}_0$ such that $\mathbb{E}X = \mathbb{E}Y$ then

$$RDI_Y(X) := \frac{\text{Var}X}{\text{Var}Y} = \frac{DI(X)}{DI(Y)} \gtrless 1; \quad (1.1)$$

i.e., X is over-, equi- and under-dispersed compared to Y if $\text{Var}X > \text{Var}Y$, $\text{Var}X = \text{Var}Y$ and $\text{Var}X < \text{Var}Y$, respectively. For instance, one can refer to Engel and te Brake (1993) and Weiß (2018, page 15) for the binomial dispersion index defined from (1.1) as $RDI_B(X) = \text{Var}X / [(\mathbb{E}X)(1 - \mathbb{E}X/N)]$, where $N \in \mathbb{N}$ is the fixed number of trials. One also can defined from (1.1) the negative binomial dispersion index as $RDI_{NB}(X) = \text{Var}X / [(\mathbb{E}X)(1 + \mathbb{E}X/\lambda)]$, where $\lambda > 0$ is the dispersion parameter. Note that the logconcave and logconvex weighted Poisson models are Poisson under- and over-dispersed, respectively (Kokonendji et al., 2008).

The variation phenomenon was recently introduced by Abid et al. (2020) as a measure of departure from the standard exponential law of nonnegative continuous distributions. This measure is dedicated by the exponential variation index, so-called Jørgensen variation index, for nonnegative continuous random variable X on $[0, \infty)$ and defined as the ratio of variance to the squared expectation. More precisely, this positive quantity is written by

$$VI(X) := \frac{\text{Var}X}{(\mathbb{E}X)^2} \gtrless 1, \quad (1.2)$$

i.e., X is over-, equi- and under-varied compared to the exponential random variable with the same expectation $\mathbb{E}X$ if $\text{Var}X > (\mathbb{E}X)^2$, $\text{Var}X = (\mathbb{E}X)^2$ and $\text{Var}X < (\mathbb{E}X)^2$, respectively. It can be viewed as the squared of the standard coefficient of variation. Within the framework of reliability, an increasing/decreasing failure rate on the average (IFRA/DFRA) distribution, which stands also for the corresponding IFR/DFR one, implies $VI \leq (\geq) 1$; that is under- (over-) variation in the large sense (i.e., including equi-variation). See, e.g., Barlow and Proschan (1981) in the sense of the coefficient of variation. Note that IFR distributions are more frequent than DFR, but over-varied data are not the most frequent situation in reliability. Again, it is not easy to construct an exact statistical test for the index VI based on its plug-in estimator in order to assess the amount of the variability; see, e.g., Lam (1980). Following RDI of (1.1) the relative variation index (RVI) is defined, for two continuous random variables X and Y on the same support $\mathcal{S} = [0, \infty)$ with $\mathbb{E}X = \mathbb{E}Y$, by

$$RVI_Y(X) := \frac{\text{Var}X}{\text{Var}Y} = \frac{VI(X)}{VI(Y)} \gtrless 1; \quad (1.3)$$

i.e., X is over-, equi- and under-varied compared to Y if $\text{Var}X > \text{Var}Y$, $\text{Var}X = \text{Var}Y$ and $\text{Var}X < \text{Var}Y$, respectively. For instance, the inverse Gaussian variation index can be defined from (1.3) by $RVI_{IG}(X) = \lambda^2 \text{Var}X / [(\mathbb{E}X)^3]$, where λ is the dispersion parameter. It is clear that the main difference between (1.1) and (1.3) is the support \mathcal{S} of these two comparable random variables X and Y . In addition, the analogous logconcave and logconvex weighted exponential models to Kokonendji et al. (2008) can provide exponential under- and over-variation, respectively (e.g., Kokonendji et al., 2020).

The probability density function (pdf) of the standard reference which is the exponential random variable $X \sim \mathcal{E}(\mu)$ with parameter $\mu > 0$ is

$$f_X(x; \mu) = \mu \exp(-\mu x) \mathbb{1}_{[0, \infty)}(x), \quad (1.4)$$

where $\mathbb{1}_A$ denotes the indicator function of any given event A . Its mean and variance are equal to $1/\mu$ and $1/\mu^2$, respectively. However, it is always equi-varied. It may happen for the sample variance to be greater or smaller than the squared sample mean, which are referred to as *over-variation* and *under-variation*, respectively, relative to the exponential distribution. The latter is a particular case of many ones, for instance the lognormal and Weibull distributions. It has also a wide range of statistical applications in multiple fields such as the reliability; see, e.g., the monograph of Balakrishnan and Basu (1995) for a review. Both univariate lognormal $\mathcal{LN}_1(m, \sigma^2)$ and Weibull $\mathcal{W}_1(\alpha, \beta)$ models are over-, equi- and under-varied for $0 < \sigma^2 \gtrless \log 2$ with $m \in \mathbb{R}$ and, for $0 < \beta \gtrless 1$ with $\alpha > 0$, respectively (e.g., Kokonendji et al., 2020). See also Kokonendji and Sawadogo (2021) for other details on the Weibull distribution.

The choice of a multivariate model from a dataset is not an easy task (e.g., Kotz et al., 2000; Joe, 2014). Multivariate nonnegative orthant data are real vectors bounded to the left by the null vector, $\mathbf{0}$ and they can be continuous, discrete or mixed. Modeling such datasets of need nonnegative orthant distributions which are generally not easy to handle in practical data analysis. The baseline parametric distribution (e.g., Johnson et al., 2014; Kotz et al., 2000) for the analysis of nonnegative countinuous data is the exponential distribution (e.g., in Reliability). In practice, we sometimes need simple and effective indicators of multivariate distribution classes in this jungle. They must be appropriate summaries of the multivariate dataset. In the multivariate setting, there is not a unique way to define a multivariate exponential distribution; e.g., Basu (1988) and Cuenin et al. (2016).

This manuscript is organized as follows. In Chapter 2, we unify the definition of RDI (1.1) and RVI (1.3). Then, we propose the common asymptotic normality of their test statistics. Asymptotic results, illustrations and numerical studies are carried out.

Chapter 3 characterizes and provides properties for general weighted exponential distributions (WEDs) and their connections with the variation phenomenon. We investigate the two-parameters WEDs in the framework of exponential dispersion models (EDMs); e.g., Jørgensen (1997). Several results, the pointwise duality between two weighted distributions, illustrative examples for building new flexible distributions are displayed.

Chapter 4 presents notations, generalized and relative variation indexes with its interpretation and properties for practical handling. We illustrate calculations of these measures on some usual bi- and multi-variate continuous nonnegative orthant distributions such as beta, exponential and Weibull. Asymptotic properties of the corresponding estimators are displayed. We also reveal examples of application from real life and simulated continuous (nonnegative orthant) datasets under several scenarios, and produces some simulation studies.

Finally, Chapter 5 concerns an R package to computes count and continuous variability indexes in univariate and multivariate setup.

Chapter 2

The unified relative dispersion and relative variation indexes

2.1 Introduction

The Fisher dispersion index is very widely used to measure the departure of any univariate count distribution from the equi-dispersed Poisson model. It is a simple measure, the ratio of variance to mean. It is a statistic commonly used to detect departures from randomness of count data. Under the hypothesis of randomness, the true distribution of this statistic is unknown. Similarly, Abid et al. (2020) have proposed the exponential variation index (VI), so-called Jørgensen variation index to measure the departure of any univariate positive continuous distribution from the equi-varied exponential model. It is defined as the ratio of variance to the squared mean. Again, the true distribution of this statistic is unknown. However, up to date, no test of equality of VI has been developed in literature (Lam, 1980).

The two main goals of this chapter are, firstly, to unify the definition of RDI (1.1) and RVI (1.3) by distinguishing the support \mathcal{S} of their corresponding distributions and, then, to propose the common asymptotic normality of their test statistics. To get this unified treatment, we shall consider the usual framework of the natural exponential families (NEFs) with their variance functions. In general, the exact tests of indexes are unavailable in literature. Section 2.2 recalls a background about NEFs connected to RDI and RVI for producing the new common index, which shall be called the relative

variability index (RWI). The asymptotic results of the test statistics shall be shown in Section 2.3. Section 2.4 is devoted to some usefull illustrations and numerical studies for providing an idea of behaviours of the proposed tests. We provide some examples of application of dispersion and variation indexes in Section 2.5.

2.2 NEF and the new RWI

In this part, we recall some definitions, examples and properties of nonnegative NEFs which are here needed (e.g., Chapter 2 in Jørgensen, 1997) to construct the new index. For $\lambda \in \Lambda \subseteq (0, \infty)$ fixed, the density with respect to the continuous Lebesgue measure or counting measure ν of a NEF $F_\lambda := \{\mathcal{F}_\nu(\theta; \lambda); \theta \in \Theta\}$ on $\mathbb{S} \subseteq \mathbb{R}$ can be written as

$$f(x; \theta, \lambda) = a(x; \lambda) \exp\{\theta x - \lambda K(\theta)\}, \quad \theta \in \Theta \subseteq \mathbb{R}, \quad (2.1)$$

where θ is the canonical parameter and $K(\cdot)$ is the cumulant function on Θ . The fixed parameter λ is the power of convolution and, therefore, $\Lambda \supseteq \{1, 2, \dots\} =: \mathbb{N}$. Let $\lambda = 1$ and denote $F := F_1$, then the mean parameterization $\mu := K'(\theta)$ in (2.1) leads to characterize F by its variance function

$$V : M \rightarrow (0, \infty), \quad \mu \mapsto V(\mu) := K''(\psi(\mu)) = 1/\psi'(\mu), \quad (2.2)$$

where $M := K'(\Theta)$ is the means domain and $\psi(\cdot)$ is the inverse function of $K'(\cdot)$. Given $\lambda > 0$ the variance function $V_\lambda(\cdot)$ of $F_\lambda := \{\mathcal{F}_V(\mu; \lambda); \mu \in M_\lambda\}$ is obtained, via (2.2), by

$$V_\lambda(\mu) = \lambda V(\mu/\lambda) \quad \text{on } M_\lambda := \lambda M. \quad (2.3)$$

Up to an affine transformation, Morris (1982) has classified the six distributions having quadratic variance function of Table 2.1 ; see Letac and Mora (1990) for the cubic variance functions. Table 2.2 presents the class of NEFs having the power variance functions. One can refer to Abid et al. (2019, 2020, 2021) and Bonat et al. (2018) for other models. Note that, except the binomial type, all the real NEFs of Tables 2.1 and 2.2 are infinitely divisible since $\Lambda = (0, \infty)$. In order to distinguish the support \mathbb{S} of any

TABLE 2.1 – Summary of the six types of quadratic variance functions on M with support \mathcal{S} of distributions and set Λ of power of convolution (Morris, 1982); GHS stands for generalized hyperbolic secant.

Type	$V(\mu)$	M	\mathcal{S}	Λ
Gaussian	1	\mathbb{R}	\mathbb{R}	$(0, \infty)$
Poisson	μ	$(0, \infty)$	\mathbb{N}_0	$(0, \infty)$
Gamma	μ^2	$(0, \infty)$	$(0, \infty)$	$(0, \infty)$
Binomial	$\mu(1 - \mu)$	$(0, 1)$	$\{0, 1\}$	\mathbb{N}
Negative binomial	$\mu(1 + \mu)$	$(0, \infty)$	\mathbb{N}_0	$(0, \infty)$
GHS	$\mu^2 + 1$	\mathbb{R}	\mathbb{R}	$(0, \infty)$

TABLE 2.2 – Summary of the power variance functions $V(\mu) = \mu^p$ on M with support \mathcal{S} of distributions and $\Lambda = (0, \infty)$; see, e.g., Tweedie (1984), Jørgensen (1997, Chap. 4).

Type(s)	p	M	\mathcal{S}
Extreme stable	$p < 0$	$(0, \infty)$	\mathbb{R}
Normal	$p = 0$	\mathbb{R}	\mathbb{R}
[Do not exist]	$0 < p < 1$		
Poisson	$p = 1$	$(0, \infty)$	\mathbb{N}_0
Compound Poisson-gamma	$1 < p < 2$	$(0, \infty)$	$[0, \infty)$
<i>Non-central gamma</i>	$p = 3/2$	$(0, \infty)$	$[0, \infty)$
Gamma	$p = 2$	$(0, \infty)$	$(0, \infty)$
Positive stable	$p > 2$	$(0, \infty)$	$(0, \infty)$
<i>Inverse Gaussian</i>	$p = 3$	$(0, \infty)$	$(0, \infty)$

member of a nonnegative NEF from its variance function (M, V) , Jørgensen et al. (1994) are investigated the behaviour of the variance in zero. The proof of the following result can also be found in Jørgensen (1997, pages 60-63).

Proposition 2.2.1 (Jørgensen et al., 1994). *Let F be a NEF with support \mathbb{S} and variance function V on M . If $\inf \mathbb{S} = 0$ and denote by $\delta := \inf\{\mathbb{S} \setminus \{0\}\}$ the distance between 0 and the smallest positive point of the support, then V and M satisfy*

1. $\inf M = 0$;
2. $\lim_{\mu \rightarrow 0} V(\mu) = 0$;
3. $\lim_{\mu \rightarrow 0} V(\mu)/\mu = \delta$;
4. if $\mathbb{P}\{0\} > 0$ then $\lim_{\mu \rightarrow 0} V'(\mu) = \delta$;
5. if $\mathbb{P}\{0\} > 0$ then $\lim_{\mu \rightarrow 0} V'(\mu)/\mu^2 = \infty$.

Note that, in Proposition 2.2.1, $\delta = 0$ for continuous distributions (e.g., gamma) and $\delta = 1$ for count distributions (e.g., negative binomial). A distribution with $\delta > 0$ must have a positive probability in zero (e.g., non-central gamma). In general, the right derivative $V'(0^+)$ of V in zero is positive for discrete distributions and zero for continuous distributions.

Now, we propose the unified definition of the relative dispersion and the relative variation indexes with respect to a given NEF.

Définition 2.2.2. *Let X and Y be two random variables taking values on $\mathbb{S} \subseteq [0, \infty)$. Assuming $\mu = \mathbb{E}X = \mathbb{E}Y$, $\sigma^2 = \text{Var}X$ and $V_\lambda(\mu) = \text{Var}Y$, for $V_\lambda(\cdot) > 0$ known, the **relative variability index** of X with respect to Y is the positive quantity*

$$\text{RWI}_Y(X) = \text{RWI}_{V_\lambda}(X) := \frac{\sigma^2}{V_\lambda(\mu)} \gtrless 1;$$

*i.e., the **over-**, **equi-** and **under-variability** of X compared to Y , and denoted by $X > Y$, $X \asymp Y$ and $X < Y$, is realized if $\sigma^2 > V_\lambda(\mu)$, $\sigma^2 = V_\lambda(\mu)$ and $\sigma^2 < V_\lambda(\mu)$, respectively.*

In this definition, the model fitted to data comes from X and the reference model is provided by Y with $V_\lambda(\cdot) > 0$ known, meaning that Y is unobservable. More specifically,

if $\mathbb{S} \subseteq \mathbb{N}_0$ then RWI becomes RDI (1.1), and for $\mathbb{S} = [0, \infty)$ we have RVI (1.3). A trivial interpretation of RWI is the ratio of two comparable variabilities; and, the equi-variability $\text{RWI} = 1$ means no discrepancy between both variabilities.

2.3 Asymptotic normality of the test statistics

Let X_1, \dots, X_n be a sequence of independent random sample identically distributed as X on $\mathbb{S} \subseteq [0, \infty)$ with $\mathbb{E}X^4 < \infty$. Denote

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad \mu := \mathbb{E}X \quad \text{and} \quad \sigma^2 := \text{Var}X.$$

To make easier the following asymptotic results, we consider below $\lambda = 1$ in (2.3). Thus, we shall restrict to the case of nonnegative NEFs on \mathbb{S} (i.e., $\inf \mathbb{S} = 0$) and where the second derivative $V''(\cdot)$ of $V(\cdot)$ exists.

Then, the plug-in estimator $S_n^2/V(\bar{X}_n)$ of the relative variability index $\sigma^2/V(\mu)$ is asymptotically unbiased, i.e., $\lim_{n \rightarrow \infty} \mathbb{E}\{S_n^2/V(\bar{X}_n)\} = \sigma^2/V(\mu)$ according to Cramér (1974, pages 357-358). More interestingly, we establish the following central limit and strong consistency of $S_n^2/V(\bar{X}_n)$.

Theorem 2.3.1. *As $n \rightarrow \infty$ and under the null hypothesis $H_{0,V} : \sigma^2 = V(\mu)$, one has :*

$$\sqrt{n} \left(\frac{S_n^2}{V(\bar{X}_n)} - 1 \right) \rightsquigarrow \mathcal{N}(0; 2 + V''(\mu)),$$

where \rightsquigarrow stands for convergence in distribution and $\mathcal{N}(0; \tau^2)$ is the centered normal distribution with variance $\tau^2 > 0$.

From Theorem 2.3.1 which is proven below, one has the following asymptotic test statistic as $n \rightarrow \infty$

$$T_{n,V} = \sqrt{\frac{n}{2 + V''(\bar{X}_n)}} \left(\frac{S_n^2}{V(\bar{X}_n)} - 1 \right) \rightsquigarrow \mathcal{N}(0; 1). \quad (2.4)$$

A hypothesis of foremost interest to be tested in this setup is the bilateral test with the null hypothesis $H_{0,V} : \sigma^2 = V(\mu)$. It is rejected if $|T_{n,V}| > u_{1-\alpha/2}$ with $\alpha \in [0, 1]$ fixed

and u_p denotes the p th percentile of the standard normal distribution $\mathcal{N}(0; 1)$. Also, an asymptotic confidence interval for $\sigma^2/V(\mu)$ is expressed as

$$CI_a(\sigma^2/V(\mu); 1 - \alpha) := [S_n^2/V(\bar{X}_n) - u_{1-\alpha/2} \widehat{\tau}_n/\sqrt{n}, S_n^2/V(\bar{X}_n) + u_{1-\alpha/2} \widehat{\tau}_n/\sqrt{n}], \quad (2.5)$$

where $\widehat{\tau}_n^2 := 2 + V''(\bar{X}_n)$ is the corresponding empirical version of the asymptotic variance $\tau^2 := 2 + V''(\mu)$ in Theorem 2.3.1. Simulation studies shall be performed at the end of Section 2.4, in comparison to the asymmetric bootstrap version of confidence intervals. Finally, we state and prove below the following result for strong consistency.

Proposition 2.3.2. *Assume only that $\mathbb{E}X^2 < \infty$, then*

$$\frac{S_n^2}{V(\bar{X}_n)} \xrightarrow{a.s.} \frac{\sigma^2}{V(\mu)}, \text{ as } n \rightarrow \infty,$$

where $\xrightarrow{a.s.}$ stands for almost sure convergence.

Proof of Theorem 2.3.1¹. Introduce $\mathbf{U} = (X, X^2)$, $\mathbf{U}_i = (X_i, X_i^2)$ for $i \in \{1, \dots, n\}$ and the map $\Phi : (0, \infty)^2 \rightarrow (0, \infty)$ defined by $\Phi(x, y) = (y - x^2)/V(x)$ twice differentiable in (x, y) with $\Phi(\mathbb{E}\mathbf{U}) = \sigma^2/V(\mu)$ and $\mathbb{E}\mathbf{U} = (\mu, \sigma^2 + \mu^2) =: \boldsymbol{\theta}$. One has : $\Phi(n^{-1} \sum_{i=1}^n \mathbf{U}_i) = (n^{-1} \sum_{i=1}^n X_i^2 - \bar{X}_n^2)/V(\bar{X}_n) = S_n^2/V(\bar{X}_n)$. The bivariate delta method (e.g., Serfling, 1980, Theorem A of Section 3.3) allows one to deduce that, as $n \rightarrow \infty$,

$$\sqrt{n} \left\{ \Phi \left(\frac{1}{n} \sum_{i=1}^n \mathbf{U}_i \right) - \Phi(\mathbb{E}\mathbf{U}) \right\} \rightsquigarrow \mathcal{N} \left[0; (\nabla \Phi(\boldsymbol{\theta}))^\top \times \text{Cov} \mathbf{U} \times \nabla \Phi(\boldsymbol{\theta}) \right]$$

with

$$\nabla \Phi(x, y) = \left(\frac{-2xV(x) + (x^2 - y)V'(x)}{V^2(x)}, \frac{1}{V(x)} \right)^\top \text{ then } \nabla \Phi(\boldsymbol{\theta}) = \left(\frac{-2\mu V(\mu) - \sigma^2 V'(\mu)}{V^2(\mu)}, \frac{1}{V(\mu)} \right)^\top$$

$$\text{and } \text{Cov} \mathbf{U} = \begin{pmatrix} \text{Var} X & \text{cov}(X, X^2) \\ \text{Cov}(X, X^2) & \text{Var} X^2 \end{pmatrix} = \begin{pmatrix} \sigma^2 & \mu_3 - \mu\sigma^2 - \mu^3 \\ \mu_3 - \mu\sigma^2 - \mu^3 & \mu_4 - (\sigma^2 + \mu^2)^2 \end{pmatrix},$$

where μ_3 and μ_4 are the third and fourth non-centered moments, respectively.

1. We sincerely thank Prof. Christian H. Weiss (Helmut Schmidt University, Hamburg, Germany) for his remarks and suggestions for the establishment of this Theorem

From Jørgensen (1997, page 66), the 3rd and 4th cumulants are $K_X^{(3)}(0) = V(\mu)V'(\mu)$ and $K_X^{(4)}(0) = V(\mu)[V'(\mu)]^2 + [V(\mu)]^2V''(\mu)$, respectively. Since they are also defined as $K_X^{(3)}(0) := \mathbb{E}(X - \mu)^3 = \mu_3 - 3\mu\sigma^2 - \mu^3$ and $K_X^{(4)}(0) := \mathbb{E}(X - \mu)^4 - 3(\sigma^2)^2 = \mu_4 - 4\mu\mu_3 + 6\mu^2\sigma^2 - 3(\sigma^2)^2 + 3\mu^4$, it follows that

$$\mu_3 = \mu^3 + 3\mu\sigma^2 + V(\mu)V'(\mu)$$

and

$$\mu_4 = \mu^4 + 6\mu^2\sigma^2 + 3(\sigma^2)^2 + 4\mu V(\mu)V'(\mu) + V(\mu)[V'(\mu)]^2 + [V(\mu)]^2V''(\mu).$$

Therefore the asymptotic variance is deduced, in terms of μ , σ^2 , $V(\mu)$, $V'(\mu)$ and $V''(\mu)$, by

$$\begin{aligned} \tau^2 = & \left\{ \sigma^6[V'(\mu)]^2 - 2\sigma^2[V(\mu)]^2[V'(\mu)]^2 + 2\sigma^4[V(\mu)]^2 \right\} / [V(\mu)]^4 \\ & + \left\{ [V(\mu)]^3[V'(\mu)]^2 + [V(\mu)]^4V''(\mu) \right\} / [V(\mu)]^4, \end{aligned}$$

for which the desired result is finally obtained by replacing σ^2 into $V(\mu)$. ■

At this stage, let us underline that corrections have been made to the asymptotic variance τ^2 and can be found in Touré et al. (2021). Furthermore, Avelino (1978) and Mizère et al. (2006) used in their work the third and fourth centered moments for the asymptotic normality of the Poisson dispersion index and obtained $\tau^2 = 2$.

Proof of Proposition 2.3.2. According to the continuous map $\Phi : (0, \infty)^2 \rightarrow (0, \infty)$ defined through $\Phi(\mathbb{E}\mathbf{U}) = \sigma^2/V(\mu)$ and $\Phi(n^{-1} \sum_{i=1}^n \mathbf{U}_i) = S_n^2/V(\bar{X}_n)$ in the proof of Theorem 2.3.1, the desired result is easily deduced from $n^{-1} \sum_{i=1}^n \mathbf{U}_i \xrightarrow{a.s.} \mathbb{E}\mathbf{U}$. ■

2.4 Illustrations and simulation studies

In this section we first present usable examples of bilateral and unilateral tests and, then, investigate their numerical behaviours using the R software (R Core Team, 2018).

2.4.1 Examples of tests and comments

We state the following five families of the hypotheses to test with their corresponding critical regions (2.4) for $N \in \mathbb{N}$, $\lambda > 0$ and $\alpha \in (0, 1)$ fixed, and for which we find them interesting in practice :

- (i) $H_{0,G} : \sigma^2 = \mu^2/\lambda$ against $H_{1,G} : \sigma^2 \neq \mu^2/\lambda$ for the non equi-variation with respect to the gamma distribution $\mathcal{G}(\lambda; \lambda/\mu)$, where $\lambda = 1$ provides the exponential $\mathcal{E}(1/\mu)$ and $H_{0,G}$ is rejected if

$$T_{n,G} = \left| \sqrt{\frac{n}{2 + 2/\lambda}} \left(\frac{S_n^2}{\bar{X}_n^2/\lambda} - 1 \right) \right| > u_{1-\alpha/2};$$

- (ii) $H_{0,IG} : \sigma^2 \leq \mu^3/\lambda^2$ against $H_{1,IG} : \sigma^2 > \mu^3/\lambda^2$ for the over-variation with respect to the inverse Gaussian distribution $\mathcal{IG}(\lambda^{1/2}; \mu)$ and, then, $H_{0,IG}$ is rejected if

$$T_{n,IG} = \sqrt{\frac{n}{2 + 6\bar{X}_n/\lambda^2}} \left(\frac{S_n^2}{\bar{X}_n^3/\lambda^2} - 1 \right) > u_{1-\alpha};$$

- (iii) $H_{0,P} : \sigma^2 = \mu$ against $H_{1,P} : \sigma^2 \neq \mu$ for the non equi-dispersion with respect to the Poisson distribution $\mathcal{P}(\mu)$ and, therefore, $H_{0,P}$ is rejected if

$$T_{n,P} = \left| \sqrt{\frac{n}{2}} \left(\frac{S_n^2}{\bar{X}_n} - 1 \right) \right| > u_{1-\alpha/2};$$

- (iv) $H_{0,B} : \sigma^2 \geq \mu(1 - \mu/N)$ against $H_{1,B} : \sigma^2 < \mu(1 - \mu/N)$ for the under-dispersion with respect to the binomial distribution $\mathcal{B}(N; \mu/N)$ and, thus, $H_{0,B}$ is rejected if

$$T_{n,B} = \sqrt{\frac{n}{2 - 2/N}} \left(\frac{S_n^2}{\bar{X}_n - \bar{X}_n^2/N} - 1 \right) < u_{\alpha};$$

- (v) $H_{0,NB} : \sigma^2 \leq \mu(1 + \mu/\lambda)$ against $H_{1,NB} : \sigma^2 > \mu(1 + \mu/\lambda)$ for the over-dispersion with respect to the negative binomial distribution $\mathcal{NB}(\lambda; \lambda/(\lambda + \mu))$ and $H_{0,NB}$ is rejected if

$$T_{n,NB} = \sqrt{\frac{n}{2 + 2/\lambda}} \left(\frac{S_n^2}{\bar{X}_n + \bar{X}_n^2/\lambda} - 1 \right) > u_{1-\alpha}.$$

Remark that if $N \rightarrow \infty$ and $\lambda \rightarrow \infty$ both asymptotic variances of the binomial $\mathcal{B}(N; \mu/N)$ and the negative binomial $\mathcal{NB}(\lambda; \lambda/(\lambda + \mu))$ distributions are equal to the one of the Poisson distribution, that is 2; see Theorem 2.3.1 with Table 2.1 and Formula (2.3).

For Part (i) with $\lambda = 1$, the exponential variation index VI or squared coefficient of variation provides the first basical test in this area ; see Lam (1980) who tried in vain to propose an exact solution and, Balakrishnan and Basu (1995) for the importance of this reference distribution. The next situation Part (ii) is a new one in the class of inverse Gaussian over-variation such as the geometric-Mittag-Leffler models (e.g., Abid et al., 2020). The classical test Part (iii) is pointed out through the Poisson dispersion which is discussed, for instance, by Tiago de Oliveira (1965), Böhning (1994) and Mizère et al. (2006). See, e.g., Kokonendji et al. (2008) for classes of the Poisson over- and under-dispersion models. From Engel and te Brake (1993), it is also more interesting to test unilaterally the hypothesis Part (iv) of the binomial under-dispersion. Finally, since negative binomial is a very usual model of the Poisson over-dispersion phenomenon, the test Part (v) proposes to discriminate the negative binomial over-dispersion for considering for example the Poisson-exponential-Tweedie models which are so-called Poisson ultra-overdispersion models (Abid et al., 2021).

2.4.2 Simulation studies

The basic objective of these simulation studies is to investigate the proposed tests in Section 2.4.1 with regard to maintaining the pre-assigned type-I error level under the null hypothesis. Since all testing procedures based on the test statistic $T_{n,V}$ of (2.4) behave very similarly under large samples, we empirically evaluate the quality of the approximation through the standard normal distribution according to both parameters $\mu > 0$ and $\lambda \in \Lambda$. Let us recall the density expressions needed for the five reference models and their corresponding standard variability indexes WI which are equal to either VI or DI, with $\mu > 0$, $\lambda > 0$, and $N \in \mathbb{N}$, respectively.

(i) Gamma $\mathcal{G}(\lambda; \lambda/\mu)$ with $WI = 1/\lambda = VI$ and

$$f(x; \mu, \lambda) = \frac{\lambda^\lambda x^{\lambda-1} \exp(-\lambda x/\mu)}{\mu^\lambda \Gamma(\lambda)} \mathbf{1}_{(0,\infty)}(x);$$

(ii) Inverse Gaussian $\mathcal{IG}(\lambda^{1/2}; \mu)$ with $WI = \mu/\lambda = VI$ and

$$f(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3} \right)^{1/2} \exp \left\{ -\frac{(x - \mu)^2 \lambda}{2\mu^2 x} \right\} \mathbf{1}_{(0, \infty)}(x);$$

(iii) Poisson $\mathcal{P}(\mu)$ with $WI = 1 = DI$ (i.e., Poisson equi-dispersion) and

$$f(x; \mu) = \frac{\mu^x e^{-\mu}}{\Gamma(x + 1)} \mathbf{1}_{\mathbb{N}_0}(x);$$

(iv) Binomial $\mathcal{B}(N; \mu/N)$ with $WI = 1 - \mu/N = DI$ such that $\mu < N$ (i.e., Poisson under-dispersion) and

$$f(x; \mu, N) = \frac{\Gamma(N + 1)}{\Gamma(x + 1)\Gamma(N - x + 1)} \left(\frac{\mu}{N} \right)^x \left(1 - \frac{\mu}{N} \right)^{N-x} \mathbf{1}_{\{0, 1, \dots, N\}}(x);$$

(v) Negative binomial $\mathcal{NB}(\lambda; \lambda/(\lambda + \mu))$ with $WI = 1 + \mu/\lambda = DI$ (i.e., Poisson over-dispersion) and

$$f(x; \mu, \lambda) = \frac{\Gamma(\lambda + x)}{\Gamma(x + 1)\Gamma(\lambda)} \left(\frac{\lambda}{\lambda + \mu} \right)^x \left(\frac{\mu}{\lambda + \mu} \right)^\lambda \mathbf{1}_{\mathbb{N}_0}(x);$$

and, where $\Gamma(z)$ is the gamma function defined by $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.

It is noteworthy that the exponential variation index VI of gamma $\mathcal{G}(\lambda; \lambda/\mu)$ and of inverse Gaussian $\mathcal{IG}(\lambda^{1/2}; \mu)$ can be over-, equi- and under-varied for $0 < \lambda \leq 1$ and for $0 < \lambda/\mu \leq 1$, respectively.

Table 2.3 displays the theoretical values of standard variability indexes WI (i.e., DI or VI) of the corresponding reference models for small ($\mu = 0.5$ and $\mu = 1$), moderate ($\mu = 5$) and large ($\mu = 10$ and $\mu = 20$) fixed averages that we shall investigate by simulations. Indeed, the VI of gamma (i) and of inverse Gaussian (ii) models can provide all the three situations of the exponential over-, equi- and under-variations; and, we thus consider λ equal to 0.5, 1 and 3. However, the DI is always Poisson equi-, under- and over-dispersed for given Poisson (iii), binomial (iv) and negative binomial (v) model, respectively. The number of trials $N \in \mathbb{N}$ of binomial (iv) are selected to be 25, 40 and 60; and, we also take $\lambda = 0.5, 1$ and 3 for negative binomial (v) model.

These simulation studies empirically evaluate the accuracy of the extreme quantiles of the distributions of $T_{n,G}$, $T_{n,IG}$, $T_{n,P}$, $T_{n,B}$ and $T_{n,NB}$ by those of the standard normal

TABLE 2.3 – Some values of standard variability indexes WI (i.e., DI or VI) for the reference models.

Reference model	$\mu = 0.5$	$\mu = 1$	$\mu = 5$	$\mu = 10$	$\mu = 20$
(i) $\mathcal{G} : \lambda = 0.5$	2.00	2.00	2.00	2.00	2.00
(i) $\mathcal{G} : \lambda = 1$	1.00	1.00	1.00	1.00	1.00
(i) $\mathcal{G} : \lambda = 3$	0.33	0.33	0.33	0.33	0.33
(ii) $\mathcal{IG} : \lambda = 0.5$	1.00	2.00	10.00	20.00	40.00
(ii) $\mathcal{IG} : \lambda = 1$	0.50	1.00	5.00	10.00	20.00
(ii) $\mathcal{IG} : \lambda = 3$	0.17	0.33	1.67	3.33	6.67
(iii) \mathcal{P}	1.00	1.00	1.00	1.00	1.00
(iv) $\mathcal{B} : N = 25$	0.98	0.96	0.80	0.60	0.20
(iv) $\mathcal{B} : N = 40$	0.99	0.98	0.88	0.75	0.50
(iv) $\mathcal{B} : N = 60$	0.99	0.98	0.92	0.83	0.67
(v) $\mathcal{NB} : \lambda = 0.5$	2.00	3.00	11.00	21.00	41.00
(v) $\mathcal{NB} : \lambda = 1$	1.50	2.00	6.00	11.00	21.00
(v) $\mathcal{NB} : \lambda = 3$	1.17	1.33	2.67	4.33	7.67

TABLE 2.4 – Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 30$.

Hypothetical model	$\mu = 0.5$	$\mu = 1$	$\mu = 5$	$\mu = 10$	$\mu = 20$
(i) $\mathcal{G} : \lambda = 0.5$	0.022	0.023	0.023	0.021	0.025
(i) $\mathcal{G} : \lambda = 1$	0.030	0.030	0.033	0.030	0.033
(i) $\mathcal{G} : \lambda = 3$	0.042	0.043	0.039	0.038	0.040
(ii) $\mathcal{IG} : \lambda = 0.5$	0.012	0.033	0.035	0.039	0.043
(ii) $\mathcal{IG} : \lambda = 1$	0.019	0.028	0.029	0.033	0.043
(ii) $\mathcal{IG} : \lambda = 3$	0.023	0.025	0.035	0.036	0.040
(iii) \mathcal{P}	0.040	0.043	0.048	0.046	0.050
(iv) $\mathcal{B} : N = 25$	0.015	0.025	0.035	0.039	0.036
(iv) $\mathcal{B} : N = 40$	0.015	0.021	0.032	0.032	0.038
(iv) $\mathcal{B} : N = 60$	0.012	0.024	0.032	0.032	0.035
(v) $\mathcal{NB} : \lambda = 0.5$	0.033	0.032	0.035	0.036	0.039
(v) $\mathcal{NB} : \lambda = 1$	0.048	0.045	0.050	0.045	0.046
(v) $\mathcal{NB} : \lambda = 3$	0.056	0.057	0.058	0.059	0.058

TABLE 2.5 – Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 50$.

Reference model	$\mu = 0.5$	$\mu = 1$	$\mu = 5$	$\mu = 10$	$\mu = 20$
(i) $\mathcal{G} : \lambda = 0.5$	0.031	0.028	0.029	0.030	0.028
(i) $\mathcal{G} : \lambda = 1$	0.033	0.029	0.032	0.033	0.036
(i) $\mathcal{G} : \lambda = 3$	0.041	0.039	0.043	0.043	0.044
(ii) $\mathcal{IG} : \lambda = 0.5$	0.031	0.033	0.032	0.041	0.040
(ii) $\mathcal{IG} : \lambda = 1$	0.026	0.033	0.037	0.041	0.041
(ii) $\mathcal{IG} : \lambda = 3$	0.033	0.034	0.035	0.044	0.043
(iii) \mathcal{P}	0.044	0.044	0.048	0.053	0.054
(iv) $\mathcal{B} : N = 25$	0.025	0.028	0.037	0.039	0.035
(iv) $\mathcal{B} : N = 40$	0.021	0.029	0.037	0.043	0.043
(iv) $\mathcal{B} : N = 60$	0.021	0.029	0.035	0.039	0.037
(v) $\mathcal{NB} : \lambda = 0.5$	0.041	0.042	0.044	0.045	0.041
(v) $\mathcal{NB} : \lambda = 1$	0.049	0.051	0.050	0.046	0.049
(v) $\mathcal{NB} : \lambda = 3$	0.050	0.056	0.060	0.057	0.060

TABLE 2.6 – Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 100$.

Hypothetical model	$\mu = 0.5$	$\mu = 1$	$\mu = 5$	$\mu = 10$	$\mu = 20$
(i) $\mathcal{G} : \lambda = 0.5$	0.031	0.034	0.034	0.031	0.033
(i) $\mathcal{G} : \lambda = 1$	0.038	0.036	0.043	0.039	0.042
(i) $\mathcal{G} : \lambda = 3$	0.045	0.043	0.047	0.044	0.046
(ii) $\mathcal{IG} : \lambda = 0.5$	0.020	0.030	0.030	0.040	0.044
(ii) $\mathcal{IG} : \lambda = 1$	0.040	0.050	0.040	0.040	0.050
(ii) $\mathcal{IG} : \lambda = 3$	0.030	0.030	0.030	0.030	0.040
(iii) \mathcal{P}	0.047	0.049	0.052	0.054	0.049
(iv) $\mathcal{B} : N = 25$	0.029	0.038	0.044	0.043	0.041
(iv) $\mathcal{B} : N = 40$	0.026	0.039	0.044	0.039	0.040
(iv) $\mathcal{B} : N = 60$	0.032	0.040	0.041	0.043	0.040
(v) $\mathcal{NB} : \lambda = 0.5$	0.046	0.051	0.049	0.050	0.053
(v) $\mathcal{NB} : \lambda = 1$	0.051	0.055	0.053	0.052	0.056
(v) $\mathcal{NB} : \lambda = 3$	0.059	0.058	0.057	0.058	0.059

TABLE 2.7 – Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 300$.

Reference model	$\mu = 0.5$	$\mu = 1$	$\mu = 5$	$\mu = 10$	$\mu = 20$
(i) $\mathcal{G} : \lambda = 0.5$	0.043	0.042	0.043	0.039	0.044
(i) $\mathcal{G} : \lambda = 1$	0.047	0.044	0.045	0.045	0.044
(i) $\mathcal{G} : \lambda = 3$	0.049	0.049	0.050	0.050	0.050
(ii) $\mathcal{IG} : \lambda = 0.5$	0.035	0.034	0.038	0.036	0.043
(ii) $\mathcal{IG} : \lambda = 1$	0.020	0.019	0.041	0.036	0.037
(ii) $\mathcal{IG} : \lambda = 3$	0.032	0.033	0.033	0.034	0.038
(iii) \mathcal{P}	0.052	0.049	0.050	0.051	0.052
(iv) $\mathcal{B} : N = 25$	0.039	0.041	0.047	0.046	0.044
(iv) $\mathcal{B} : N = 40$	0.037	0.042	0.046	0.047	0.044
(iv) $\mathcal{B} : N = 60$	0.040	0.040	0.043	0.049	0.045
(v) $\mathcal{NB} : \lambda = 0.5$	0.054	0.055	0.055	0.059	0.053
(v) $\mathcal{NB} : \lambda = 1$	0.057	0.055	0.061	0.053	0.058
(v) $\mathcal{NB} : \lambda = 3$	0.055	0.056	0.057	0.057	0.054

TABLE 2.8 – Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 500$.

Reference model	$\mu = 0.5$	$\mu = 1$	$\mu = 5$	$\mu = 10$	$\mu = 20$
(i) $\mathcal{G} : \lambda = 0.5$	0.045	0.044	0.049	0.044	0.044
(i) $\mathcal{G} : \lambda = 1$	0.044	0.047	0.047	0.048	0.050
(i) $\mathcal{G} : \lambda = 3$	0.047	0.050	0.049	0.049	0.052
(ii) $\mathcal{IG} : \lambda = 0.5$	0.044	0.047	0.048	0.041	0.044
(ii) $\mathcal{IG} : \lambda = 1$	0.047	0.047	0.046	0.043	0.042
(ii) $\mathcal{IG} : \lambda = 3$	0.045	0.039	0.045	0.040	0.042
(iii) \mathcal{P}	0.050	0.045	0.050	0.053	0.051
(iv) $\mathcal{B} : N = 25$	0.042	0.048	0.046	0.046	0.048
(iv) $\mathcal{B} : N = 40$	0.044	0.043	0.045	0.050	0.047
(iv) $\mathcal{B} : N = 60$	0.043	0.046	0.046	0.048	0.046
(v) $\mathcal{NB} : \lambda = 0.5$	0.058	0.058	0.056	0.057	0.058
(v) $\mathcal{NB} : \lambda = 1$	0.060	0.055	0.059	0.055	0.052
(v) $\mathcal{NB} : \lambda = 3$	0.055	0.055	0.055	0.056	0.054

TABLE 2.9 – Empirical probabilities of statistical tests with respect to $\alpha = 0.05$ for $n = 1000$.

Hypothetical model	$\mu = 0.5$	$\mu = 1$	$\mu = 5$	$\mu = 10$	$\mu = 20$
(i) $\mathcal{G} : \lambda = 0.5$	0.044	0.047	0.046	0.045	0.043
(i) $\mathcal{G} : \lambda = 1$	0.048	0.050	0.049	0.047	0.052
(i) $\mathcal{G} : \lambda = 3$	0.048	0.046	0.047	0.048	0.051
(ii) $\mathcal{IG} : \lambda = 0.5$	0.042	0.043	0.040	0.040	0.040
(ii) $\mathcal{IG} : \lambda = 1$	0.049	0.048	0.049	0.043	0.043
(ii) $\mathcal{IG} : \lambda = 3$	0.043	0.045	0.045	0.046	0.040
(iii) \mathcal{P}	0.051	0.047	0.054	0.050	0.051
(iv) $\mathcal{B} : N = 25$	0.044	0.044	0.049	0.049	0.048
(iv) $\mathcal{B} : N = 40$	0.042	0.045	0.047	0.045	0.048
(iv) $\mathcal{B} : N = 60$	0.044	0.048	0.048	0.046	0.048
(v) $\mathcal{NB} : \lambda = 0.5$	0.052	0.054	0.057	0.055	0.056
(v) $\mathcal{NB} : \lambda = 1$	0.057	0.053	0.057	0.054	0.052
(v) $\mathcal{NB} : \lambda = 3$	0.052	0.054	0.050	0.052	0.054

distribution according to the sample sizes. However, we simulated 10 000 samples for each triplet (n, μ, λ) given $\mu > 0$, $\lambda > 0$ and $N \in \mathbb{N}$ with the same values as above. The sample sizes n examined are successively 30, 50, 100, 300, 500 and 1000. Tables 2.4 to 2.9 exhibit the empirical probabilities of statistical tests with respect to $\alpha = 0.05$ according to the sample size n . An empirical probability greater than $\alpha = 0.05$ shows that the accuracy of the approximation of the extreme quantiles is not sufficient. It is closed to zero for conservative tests of null hypothesis. Finally, note that for the calculation of the sample variance of the test statistics, the software R considers the empirical variance without bias, i.e, $1/(n-1) \sum_{i=1}^n (X_i - \bar{X}_n)^2$. To find the empirical variance used in page 35, we multiply by the quantity $(n-1)/n$ by the variance without bias.

For the bilateral Poisson equi-dispersion test Part (iii), the empirical probabilities revolve around the nominal value 0.05 and whatever the sample size n . This Poisson equi-dispersion test is much discussed and improved in literature; e.g., Tiago de Oliveira (1965), Böhning (1994) and Mizère et al. (2006). Concerning the unilateral binomial under-dispersion test Part (iv), the empirical probabilities are less than 0.05 for the sample sizes $n = 30$, $n = 50$, $n = 100$ and $n = 300$ and whatever the values of μ and N . For $n = 500$ and $n = 1000$ these probabilities revolve around the value 0.05. If the

mean values are small, one can consider a categorical variation index (e.g., Weiß 2019a, 2019b); and, for large mean values one can use the Poisson statistical test.

In regards to the unilateral negative binomial over-dispersion test Part (v), the empirical probabilities are on the whole less than 0.05 whatever the values of λ and μ for the sample sizes $n = 30$, $n = 50$ and $n = 100$. For the sample sizes $n = 300$, $n = 500$ and $n = 1000$ the empirical probabilities revolve around the nominal value 0.05 or not significantly different from this value. Finally, for the unilateral inverse Gaussian over-variation test Part (ii), all the empirical probabilities are less than 0.05 whatever the values of λ and μ and whatever the size n . This can be explained by the overdispersion nature of this distribution. In view of these simulation studies, one can say that the accuracy of the approximations of the extreme quantiles is sufficient.

The binomial under-dispersion Part (iv), negative binomial over-dispersion Part (v), gamma equi-variation (i) and inverse Gaussian over-variation Part (ii) tests are new in literature. It is powerful to know that our test is not robust. We have used a unification test and, therefore, we must pay attention to the case by case distributions and values of μ , λ and N whether large or small. Similar to the Poisson dispersion index test Part (iii), all these variability index tests merit to be improved for each framework. At this stage, we also omit to investigate globally the powers of these asymptotic tests for given alternative hypotheses.

In addition, Table 2.10 reveals similar behaviours of asymptotic and asymmetric confidence intervals, for instance, of standard equi-dispersion and equi-variation indexes in the situations of small and moderate sample sizes (e.g., Angelo and Brian, 2019). Indeed, we have used the empirical bootstrap method whose works well even for asymmetric distribution (e.g., Efron and Tibshirani, 1993, Chapter 13). Bootstrap methods are widely used for obtaining confidence intervals when the population distribution is not known or a non-normal. We have done 10 000 replicates given sample size n with significance level $\alpha = 0.05$ fixed. The asymmetric bootstrap confidence interval for $\sigma^2/V(\mu)$ is expressed as

$$CI_b(\sigma^2/V(\mu); 1 - \alpha) := [\widehat{b} - \widehat{\delta}_b(1 - \alpha/2) se, \widehat{b} + \widehat{\delta}_b(\alpha/2) se] = [\widehat{b} - \widehat{SE}_1, \widehat{b} + \widehat{SE}_2], \quad (2.6)$$

where $\widehat{b} = \widehat{b}(\sigma^2/V(\mu))$ is the bootstrap estimate of $\sigma^2/V(\mu)$ with standard error (se) and, $\widehat{\delta}_b(1 - \alpha/2)$ and $\widehat{\delta}_b(\alpha/2)$ are the estimated asymmetric bootstrap quantiles of the distribution $\delta = \widehat{b}(\sigma^2/V(\mu)) - \sigma^2/V(\mu)$. According to Table 2.10 with standard dispersion index DI (for Poisson) and standard variation index VI (for exponential), asymptotic and bootstrap pointwise estimates are very near for all sample size n . Afterwards, asymptotic standard error $u\widehat{\tau}_n/\sqrt{n}$ of (2.5) and bootstrap estimated standard error $\widehat{\delta}_b se$ of (2.6) decrease when sample size n increases. But, this decreasing is faster with the bootstrap method. We conclude that our procedure works well.

TABLE 2.10 – Asymptotic and asymmetric bootstrap confidence intervals for the Poisson dispersion (DI) and exponential variation (VI) indexes from sample size n simulated with $u = u_{0.975} = 1.96$ and $\alpha = 0.05$.

n	$\widehat{DI}_n \pm u\widehat{\tau}_n/\sqrt{n}$	$\widehat{DI}_b(-\widehat{SE}_1; +\widehat{SE}_2)$	$\widehat{VI}_n \pm u\widehat{\tau}_n/\sqrt{n}$	$\widehat{VI}_b(-\widehat{SE}_1; +\widehat{SE}_2)$
30	1.2520 ± 0.5061	$1.0296(-0.0757; +0.0777)$	1.1676 ± 0.7157	$1.2448(-0.1862; +0.2555)$
50	1.3748 ± 0.3920	$1.1449(-0.0441; +0.0449)$	0.9892 ± 0.5544	$1.1740(-0.1262; +0.1335)$
100	0.7970 ± 0.2779	$0.9567(-0.0206; +0.0212)$	0.8355 ± 0.3920	$0.9782(-0.0590; +0.0614)$
300	0.9866 ± 0.1600	$1.0302(-0.0125; +0.0120)$	0.8811 ± 0.2263	$1.0738(-0.0194; +0.0191)$
500	1.0678 ± 0.1240	$1.0310(-0.0086; +0.0079)$	0.9457 ± 0.1753	$1.1021(-0.0185; +0.0149)$
1000	0.9555 ± 0.0877	$1.0125(-0.0049; +0.0045)$	0.9867 ± 0.1240	$0.9846(-0.0053; +0.0051)$

Note that this bootstrap representation was motivated by the fact of comparison with the asymptotic confidence interval. In addition, there is another representation of confidence interval which consists to give the lower and upper bounds of the confidence intervals with the width of the intervals (Altunkaynak and Gamgam, 2018). For the sake of comparison, we opted for the first method.

2.5 Examples of Application

In this section, we shall give examples of dispersion and variation indexes.

The first is an example of application of the variation index in reliability. Figure 2.1 represents the "bathtub curve" of failure rate of an electronic component for example. In the case $VI < 1$, the curve decreases and we have the under-variation. For $VI = 1$, the curve is constant and thus we have the exponential equi-variation. Finally, if $VI > 1$ the curve is increasing, this corresponds to the over-variation.

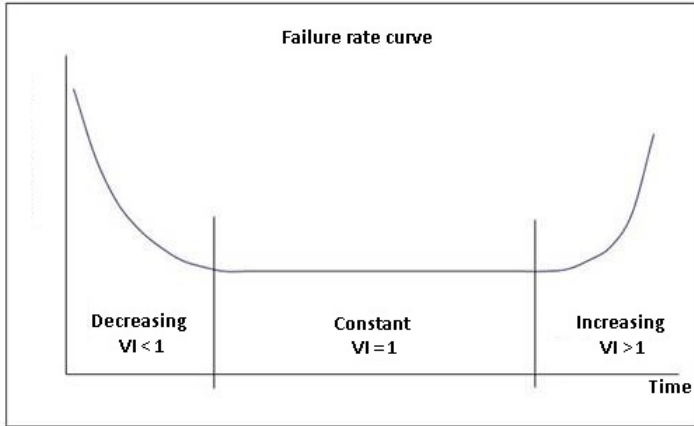


FIGURE 2.1 – A use of the exponential variation index in reliability : "Bathtub curve" of the failure rate as function of time.

The second example is the one of dispersion index and defined as follows. Böhning (1994) has considered the count dataset of Table 2.11 for producing different conclusions between the test statistic $T_T = \sqrt{n}(S_n^2 - \bar{X}_n)/(1 - 2\bar{X}_n^{1/2} + 3\bar{X}_n)^{1/2}$ of Tiago de Oliveira (1965) and the corrected version of his own given by $T_B = \sqrt{(n-1)/2}(S_n^2 - \bar{X}_n)/\bar{X}_n$. Indeed, the numerical values $n = 366$, $\bar{x} = 6.3634$ and $s^2 = 6.8238$ pointwisely shown a slight Poisson over-dispersion through $s^2 - \bar{x} = 0.4604 > 0$ or $s^2/\bar{x} = 1.0724 > 1$. However, the two statistical test values $t_T = 2.2706$ and $t_B = 0.9774$ lead to opposite conclusions via their p-values $\hat{\alpha}_T = 0.0116$ and $\hat{\alpha}_B = 0.1642$, respectively. From our Poisson dispersion index test Part (iii), the statistical test $t_{n,p} = 0.9787$ provides the p-value $\hat{\alpha} = 0.3277$ (for a bilateral test) which clearly rejects the Poisson over-dispersion much better than the Böhning procedure.

Several other data in (the statistical) literature are applicable to decide which models

TABLE 2.11 – Daily numbers of deaths of women by brain vessel disease in West Berlin, 1989 (Böhning, 1994).

Deaths per day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Frequency	1	4	15	31	39	55	54	49	47	31	16	9	8	4	3

envisaged. Now, we introduce original data of the failure times (in months) of drinking water pumps installed in the Sahel. They are 23, 261, 87, 10, 120, 14, 62, 15, 47, 225, 71, 20, 246, 21, 19, 42, 20, 5, 12, 120, 17, 11, 3, 14, 71, 11, 5, 14, 11, 16, 90, 1, 16, 52, 95, 10, 1, 14, 4, 7, 14, 20 and have been provided by Sobom M. Somé of an statistical study office of Burkina-Faso. Given that $n = 42$, $\bar{x} = 46.1191$ and $s^2 = 4131.6200$, one has pointwisely the exponential over-variation through $VI = s^2/\bar{x}^2 = 1.9425 > 1$. In addition, the bilateral statistical test Part (i) with $\lambda = 1$ and $t_{n,E} = 3.0540$ provides the p-value $\widehat{\alpha}_E = 0.0023$ (for a bilateral test) which rejects the exponential equi-variation of these data with respect to $\alpha = 0.05$. More generally, from the maximum likelihood estimate $\widehat{\lambda} = 0.7664$ of λ in the gamma model $\mathcal{G}(\lambda; \lambda/\mu)$, one gets $RVI_G = 1.4887 > 1$ and $t_{n,G} = 1.4752$ for the gamma equi-variation test Part (i) leading to the p-value $\widehat{\alpha}_G = 0.1402$. Hence, we suggest to use the flexible two-parameter gamma model for analyzing these over-varied data. Remember that the exponential model is suited for equi-varied data.

2.6 Conclusion

In this chapter, we have proposed a new index RWI that takes into account both RDI for count data and RVI for positive continuous data. We also proposed asymptotic tests for RWI which provided good results even for small sample sizes. However, these tests deserve to be improved by a more robust test. Also in our work we did not study the power of these tests which can provide better results. We were able to show on an example that our dispersion index is better than those of Böhning (1994) and Tiago de Oliveira (1965) insofar as our p-value gives a good result.

Chapter 3

General exponential weight functions and variation phenomenon

3.1 Introduction

We define a general weighted exponential random variable, denoted by X^w , from the reference $X \sim \mathcal{E}(\mu)$ of (1.4) and a measurable nonnegative weight function $w : [0, \infty) \rightarrow [0, \infty)$ such that

$$f_{X^w}(\cdot; \mu) := w(\cdot) f_X(\cdot; \mu) \quad (3.1)$$

is a pdf, a so-called weighted exponential distribution (WED). More practically, if the exponential weight function $w_0(\cdot)$ is such that $1 \neq \mathbb{E}w_0(X) < \infty$, then $w(\cdot)$ introduced in (3.1) becomes

$$w(\cdot) = \frac{w_0(\cdot)}{\mathbb{E}w_0(X)}, \quad (3.2)$$

which is finally self-normalized. Notice that, from (3.2), the (non-normalized) exponential weight function $w_0(\cdot) \equiv w_0(\cdot; \boldsymbol{\alpha})$ can depend on a nonnegative number k of parameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k) \in \Theta_k \subseteq \mathbb{R}^k$. Therefore, the (normalized) exponential weight function $w(\cdot) \equiv w(\cdot; \mu, \boldsymbol{\alpha})$ of (3.1) is in terms of $k + 1$ parameters. Introducing an extra parameter to an existing distribution (for instance via weight function) often brings more flexibility to a class of distributions, and can be very useful for data analysis purposes. The WEDs can be considered as modified exponential distributions that provide a unified approach to handle both over-variation and under-variation.

Several WEDs with great ability in statistical applications, more particularly in reliability and life data analysis, include the exponentiated exponential distribution (Gupta and Kundu, 2001), the weighted exponential distribution of Gupta and Kundu (2009), the gamma-exponentiated exponential distribution (Ristić and Balakrishnan, 2012), the weighted generalized exponential distribution (Mahdavi, 2015), the length biased weighted exponential distribution of Das and Kundu (2016), the extended weighted exponential distribution (Mahdavi and Jabbari, 2017) and the logarithmic weighted exponential distribution (Bakouch et al., 2021). See also Table 3.2 in Section 3.5 for a detailed summary of WEDs. A brief review on weighted distributions can be found in Saghir et al. (2017). The statistical literature of this millennium is already wealthy and growing fast on proposing new WEDs, generally grounded on Azzalini's (1985) method. It is noteworthy that the parameter estimations can be performed by using maximum likelihood, moments, weighted least square and maximum product of spacings methods; see, e.g., Alqallaf et al. (2015), Dey et al. (2015), Kharazmi et al. (2015), Oguntunde (2015), Famoye (2019) and Bakouch et al. (2021).

In this chapter, we provide a few properties for general WEDs and their connections with the variation phenomenon. In particular, we investigate the two-parameters WEDs in the framework of exponential dispersion models (EDMs); e.g., Jørgensen (1997). In section 3.2 we will recall some important notions in reliability which motivates this part of the thesis. In Section 3.3, we establish several results related to the representation of any nonnegative continuous distribution as a weighted version of another one, in particular of the very useful exponential distribution. We also introduce the pointwise duality between two weighted distributions. Section 3.4 presents an effective connection between WEDs in EDMs and the variation phenomenon. In Section 3.5 we provide details on some examples of WEDs of Table 3.2. Section 3.6 is devoted to illustrative examples for building new flexible distributions.

3.2 Background in reliability

According to Aissani (1992), Reliability is the ability of a system to perform a given function during a given period and under specified operating conditions. The conditions are all external constraints, be they human, climatic, physical, etc. Reliability is a hallmark of device expressed by the probability that it performs a function under given conditions for a given time. It is therefore characterized by four concepts : probability, performance of a function, given conditions, time.

3.2.1 Distribution function

We denote by $F(x)$ the distribution function associated with the lifetimes X . It can be interpreted as the probability that the lifetime of the component is less than t .

$$F(x) = \mathbb{P}[\text{component lifetime (system)} < x] = \mathbb{P}(X < x) = \int_0^x f(t) dt.$$

$F(x) dx$ is the probability that the lifetime of a system is between x and $x + dx$, or the probability that it fails between x and $x + dx$:

$$F(x) dx = \mathbb{P}(x < X < x + dx) = \int_x^{x+dx} f(t) dt.$$

3.2.2 Reliability function and failure rate

Reliability function $R(x)$ represents the probability of functioning correctly for a given period $[0, x]$, in other words the probability that no failure will occur :

$$R(x) = \mathbb{P}(X \geq x) = 1 - F(x) = \int_x^{\infty} f(t) dt.$$

Equipment failure can be characterized by a rate called the failure rate. This rate is also called the failure rate, chance rate or death rate. It is defined as the conditional probability that the equipment will fail between time $[x; x + dx)$ knowing that it has survived until time x . It is defined by

$$\Lambda(x) = \frac{F'(x)}{R(x)} = -\frac{R'(x)}{R(x)} = \frac{f(x)}{R(x)}.$$

By integrating the differential equation, the reliability function will be written :

$$R(x) = \exp\left(-\int_0^x \Lambda(t) dt\right).$$

We can notice that all the functions mentioned are naturally linked to each other : the knowledge of $R(x)$ implies that of $f(x)$ and therefore that of $\Lambda(x)$. Thus, it suffices to specify the distribution of the lifetimes (random variable X) to determine all the quantities of interest relating to reliability.

Example 3.2.1. — For $Y \sim \mathcal{E}(\mu)$, the reference exponential distribution of pdf (1.4), the reliability and failure rate functions are $R(x) = \exp(-\mu x)$ and $\Lambda(x) = f_Y(x)/R(x) = \mu$, respectively.

— For $Y \sim \mathcal{G}(\mu, \alpha)$, the two-parameter gamma random variable with $\mu, \alpha > 0$ such that

$$f_Y(x) = \frac{\mu^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\mu x) \mathbb{1}_{(0, \infty)}(x).$$

One major disadvantage of the gamma distribution is that the distribution function or survival function can not be expressed in a closed form if the shape parameter is not an integer. The survival function and the failure rate of this model are not explicit, but they are defined as follows :

$$R(x) = 1 - F_Y(x) = 1 - \int_0^x f_Y(t) dt,$$

and

$$\Lambda(x) = f_Y(x)/R(x).$$

— Consider $Y \sim \mathcal{W}_\gamma(\mu, \alpha)$, the three-parameter Weibull (1933) random variable with $\mu, \alpha > 0$ and $\gamma \geq 0$ such that

$$f_Y(x) = \alpha \mu [\mu(x - \gamma)]^{\alpha-1} \exp\{-[\mu(x - \gamma)]^\alpha\} \mathbb{1}_{[\gamma, \infty)}(x).$$

For $\gamma = 0$ we have the 2-parameter Weibull random variable which the reliability and rate functions are respectively

$$R(x) = \exp\left(-\left(\frac{x}{\mu}\right)^\alpha\right)$$

and

$$\Lambda(x) = \frac{\alpha}{\mu} \left(\frac{x}{\mu} \right)^{\alpha-1}.$$

These three examples clearly show that the failure rate function is easy to use for the exponential distribution and very complicated for the gamma distribution. Hence the advantage of the exponential distribution in reliability.

The failure rate is interpreted as a power of time, which makes it possible to model many situations given the value of the shape parameter α :

1. If $\alpha > 1$, $\Lambda(x)$ is decreasing and therefore the system improves.
2. If $\alpha < 1$, $\Lambda(x)$ is increasing then the system wears out.
3. If $\alpha = 1$, $\Lambda(x)$ is constant and we have the exponential distribution as a special case of Weibull ($\gamma = 0$) and gamma distributions.

Weibull distribution seems to be more able to model a lifetime showing a period known as “youth”, “useful life” and “wear-out” materialized below in Figure 3.1 by the so-called “bathtub” curve.

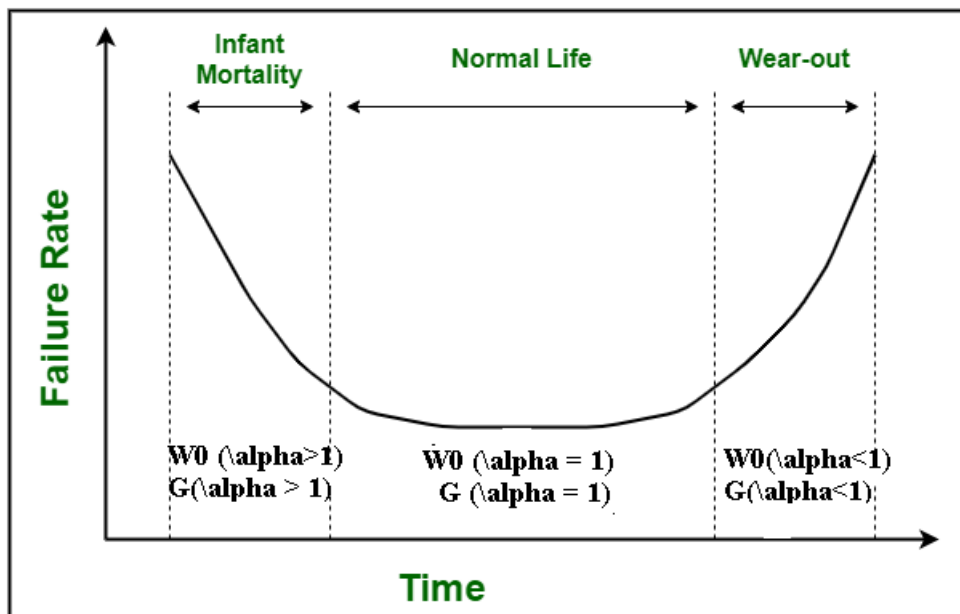


FIGURE 3.1 – "Bathtub curve" of the failure rate as function of time.

Figure 3.1 shows the following results. Let $Y \sim \mathcal{W}(\gamma; \mu, \alpha) = \mathcal{W}_\gamma(\mu, \alpha)$ be the 3-parameter Weibull (1933) random variable with

$$f_Y(x) = \alpha \mu [\mu(x - \gamma)]^{\alpha-1} \exp\{-[\mu(x - \gamma)]^\alpha\} \mathbb{1}_{[\gamma, \infty)}(x), \quad \mu > 0, \alpha > 0, \gamma \geq 0,$$

and $Y \sim \mathcal{G}(\mu, \alpha)$ be the 2-parameter gamma random variable with

$$f_Y(x) = [\mu^\alpha / \Gamma(\alpha)] x^{\alpha-1} \exp(-\mu x) \mathbb{1}_{(0, \infty)}(x), \quad \mu > 0, \alpha > 0,$$

then

$$\boxed{\mathcal{G}(\mu, 1) = \mathcal{E}(\mu) = \mathcal{W}_0(\mu, 1)}$$

3.3 Results about nonnegative continuous weighted distributions

In this section, we explore several properties related to the representation of any nonnegative continuous distribution as a weighted distribution with respect to a given reference one. The first results make this investigation possible through the following decomposition with respect to the reference exponential distribution before generalizing. We shall proceed that way because the exponential distribution is the most used in practice, and the corresponding exponential weight function provides some properties or characteristics on the original nonnegative continuous random variables.

Theorem 3.3.1. *Let Y be a nonnegative continuous random variable with pdf $f_Y(\cdot)$ and support $\mathbb{S}_Y \subseteq [0, \infty)$ and let $X \sim \mathcal{E}(\mu)$ of (1.4) with $\mu > 0$. Then*

$$Y \stackrel{d}{=} X^w, \tag{3.3}$$

where $\stackrel{d}{=}$ stands for equality in distribution and the exponential weight function $w(\cdot)$ is given by

$$w(x) = \frac{\exp(\mu x)}{\mu} f_Y(x), \quad \forall x \in \mathbb{S}_Y. \tag{3.4}$$

Proof. Since the Radon-Nikodym theorem states that given a measurable space $(\mathbb{T}, \mathcal{A})$, and given two σ -finite nonnegative measures λ_1 and λ_2 on it; if λ_2 is absolutely

continuous with respect to λ_1 , then there exists a unique measurable nonnegative real function $f(\cdot)$ defined on \mathbb{T} , such that for any measurable set A ,

$$\lambda_2(A) = \int_A f(x) d\lambda_1(x).$$

In particular, this Radon-Nikodym theorem obviously holds for probability measures. Hence, by taking $(\mathbb{T}, \mathcal{A}) := ([0, \infty), \mathcal{B})$ with \mathcal{B} the Borel σ -algebra on $[0, \infty)$ and λ_1 which is equal to the exponential measure of (1.4), we obtain the desired result (3.3). Thus, the weight function (3.4) is easily deduced by rewriting (3.3) from (3.1) as $f_Y(\cdot) = f_Y(\cdot)f_X(\cdot; \mu)/f_X(\cdot; \mu) = f_Y(\cdot) * \mu \exp(-\mu x)/\mu \exp(-\mu x) = w(\cdot)f_X(\cdot; \mu)$. ■

Example 3.3.2. Let $Y \sim \mathcal{G}(\mu, \alpha)$ be the gamma distribution with parameters $\mu, \alpha > 0$ such that

$$f_Y(x) = \frac{\mu^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\mu x) \mathbb{1}_{(0, \infty)}(x),$$

where $\Gamma(z)$ is the gamma function defined by $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.

The gamma distribution is a natural WED : $\mathcal{G}(\mu, 1) = \mathcal{E}(\mu)$ with exponential weight function $w(x) = [\mu^{\alpha-1}/\Gamma(\alpha)]x^{\alpha-1}$.

Theorem 3.3.1 allows even for the beta distribution $\mathcal{B}_e(\alpha, \beta)$ with support $[0, 1]$ to have a WED representation from (3.4). See also No. 12 of Table 3.2 which are wrapped WEDs of Roy and Adnan (2012).

The next proposition reveals the connection between two representations of a non-negative continuous distribution as WEDs. It can be also used to demonstrate the uniqueness of (3.3).

Proposition 3.3.3. Let X_1 and X_2 be two exponential random variables with parameters μ_1 and μ_2 respectively. If a nonnegative continuous random variable Y is such that $Y \stackrel{d}{=} X_1^{w_1}$ with $w_1(\cdot)$ defined as in (3.4), then $Y \stackrel{d}{=} X_2^{w_2}$ with

$$w_2(x) = \frac{\mu_1}{\mu_2} \exp[(\mu_2 - \mu_1)x] w_1(x) \mathbb{1}_{\mathbb{S}_Y}(x). \quad (3.5)$$

Proof. It is immediate. ■

The following result shows that any nonnegative continuous random variable has its own weighted exponential representation of (3.3), that we called the self-decomposition in a WED. It will be used to design a semiparametric approach for estimating the pdf of an unknown WED.

Corollaire 3.3.4. *Let Y be a nonnegative continuous random variable with $\mathbb{S}_Y \subseteq [0, \infty)$ and such that $\mathbb{E}Y < \infty$. Then, Y follows a WED with respect to the reference exponential distribution of parameter $1/\mathbb{E}Y$.*

Proof. It is easily deduced from Theorem 3.3.1 with $\mu = 1/\mathbb{E}Y > 0$. ■

The previous representation of Corollary 3.3.4 allows to interpret a given nonnegative continuous random variable as a weighted version of the exponential one with the same mean. As examples, we shall state the gamma, the lognormal and the three-parameter Weibull distributions in the last section with some of their modifications.

Now, we exhibit the commutativity of the weightening operation under a more general reference than the standard exponential random variable.

Theorem 3.3.5. *Let Y be a nonnegative continuous random variable and let $w_1(\cdot)$ and $w_2(\cdot)$ be two positive weight functions such that $0 < \mathbb{E}w_j(Y) < \infty$ for $j = 1, 2$ and $0 < \mathbb{E}[w_1(Y)w_2(Y)] < \infty$. Then :*

$$(Y^{w_1})^{w_2} \stackrel{d}{=} Y^{w_1 w_2} \stackrel{d}{=} (Y^{w_2})^{w_1}.$$

Proof. Given that

$$\mathbb{E}w_2(Y^{w_1}) = \frac{1}{\mathbb{E}w_1(Y)} \int_{\mathbb{S}_Y} w_2(y)w_1(y)f_Y(y)dy = \frac{\mathbb{E}[w_1(Y)w_2(Y)]}{\mathbb{E}w_1(Y)},$$

the pdf of $(Y^{w_1})^{w_2}$ is successively expressed by

$$\begin{aligned} f_{(Y^{w_1})^{w_2}}(y) &= \frac{1}{\mathbb{E}w_2(Y^{w_1})} w_2(y)f_{Y^{w_1}}(y) \\ &= \frac{1}{\mathbb{E}[w_1(Y)w_2(Y)]} w_2(y)w_1(y)f_Y(y), \end{aligned}$$

for all $y \in \mathbb{S}_Y$, which is the same pdf of $Y^{w_1 w_2}$ and $(Y^{w_2})^{w_1}$ by interchanging $w_1(\cdot)$ and $w_2(\cdot)$. ■

The following remark brings together some other properties, weighting operations and advanced extensions of the previous results on WEDs like Theorem 3.3.1 as well as Proposition 3.3.3. We therefore omit their proofs.

Remarque 3.3.6. (i) *The positivity of $w_j(\cdot)$ (i.e., $w_j(\cdot) > 0$) implies that the supports of the corresponding positive continuous weighted distributions are $(0, \infty)$.*

(ii) *It is noteworthy that $0 < \mathbb{E}w_j(Y) < \infty$ for $j = 1, 2$ do not necessarily imply $0 < \mathbb{E}[w_1(Y)w_2(Y)] < \infty$.*

(iii) *If $Y \stackrel{d}{=} Z^{w_1}$ and $Z \stackrel{d}{=} U^{w_2}$, then one has $Y = U^{w_1 w_2}$ from Theorem 3.3.5, especially for $U \sim \mathcal{E}(\mu)$ relative to WEDs.*

(iv) *If $w := w_1 \circ w_2$ is a composed of two weight functions, then $Y^w = Y^{w_1(w_2)}$.*

(v) *Theorem 3.3.1a : one generally has $Y \stackrel{d}{=} Z^w$ with $w(\cdot) = f_Y(\cdot)/f_Z(\cdot)$ and $f_Z(\cdot) > 0$ for $\mathbb{S}_Y \subseteq \mathbb{S}_Z \subseteq [0, \infty)$.*

(vi) *Proposition 3.3.3a : if $Y \stackrel{d}{=} Z_1^{w_1}$ and $Y \stackrel{d}{=} Z_2^{w_2}$ with $\mathbb{S}_Y \subseteq \mathbb{S}_{Z_1} = \mathbb{S}_{Z_2} \subseteq [0, \infty)$, then*

$$w_2(\cdot) = w_1(\cdot) \frac{f_{Z_1}(\cdot)}{f_{Z_2}(\cdot)} \frac{\mathbb{E}w_2(Z_2)}{\mathbb{E}w_1(Z_1)}.$$

The concept of duality introduced by Kokonendji et al. (2008, Section 3) for weighted Poisson distributions is here extended to any reference nonnegative continuous distribution, including the standard exponential distribution.

Définition 3.3.7. *Let Y be a positive continuous random variable on $\mathbb{S}_Y \subseteq [0, \infty)$, and let $w_0(\cdot)$ and $w_0^*(\cdot)$ be two positive weight functions. The two corresponding weighted versions Y^{w_0} and $Y^{w_0^*}$ are said to be a **dual pair** with respect to Y if and only if*

$$w_0(y) w_0^*(y) = 1, \quad \forall y \in \mathbb{S}_Y.$$

A practical meaning of duality for WEDs is that this distribution provides the opposite variation (i.e., over-variation for under-variation and conversely). The standard exponential distribution $\mathcal{E}(\mu)$ is self-dual because its weight function is $w(x) = 1, \forall x \in [0, \infty)$. As an example of (closed) dual pair of WEDs, consider the gamma distribution

with weight function $w_0(x) = x^{\alpha-1}$ and $\alpha \in (1, \infty)$ (see No. 1 of Table 3.2). There exists another gamma distribution with $\alpha = 2 - \alpha^* \in (1, \infty)$ or $\alpha^* \in (0, 1)$ which is its dual. Notice that $\alpha_1 = 1$ produces the self-dual of the exponential distribution.

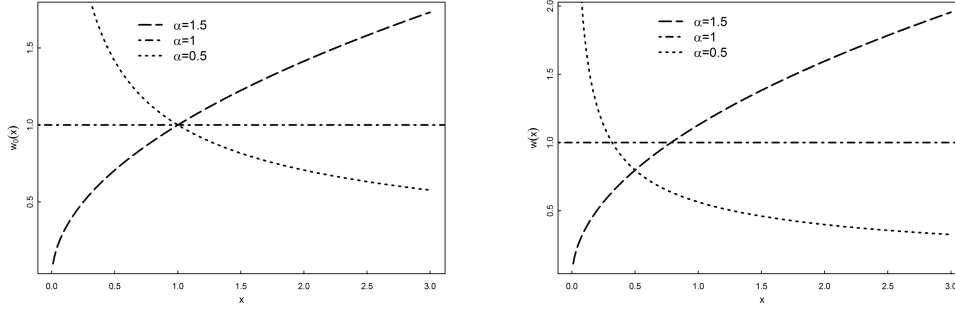


FIGURE 3.2 – Pointwise duality of Definition 3.3.7 and Part (ii) of Remark 3.3.8 (with $\mu = 1$) in the gamma model.

Figure 3.2 displays some plots of (closed) dual pairs of the gamma distribution as a WED which is further discussed in Section 3.6.1. We can note the interchange in the dispersion of the roles of the weight function $w_0(x)$ for x below 1 (which is the same to the equi-varied exponential distribution : $w_0(1) = 1$) and x above 1 ($w_0(x) \geq 1$ for over- or under-variation). The previous situations are translated in the second graphics of $w(\cdot)$.

Remarque 3.3.8. (i) *The dual of Y^{w_0} exists if and only if $\mathbb{E}[1/w_0(Y)] < \infty$.*

(ii) *If $w_0(\cdot)$ is such that $1 \neq \mathbb{E}w_0(Y) < \infty$, then Definition 3.3.7 is to be understand as*

$$w_0(y)w_0^*(y) = \mathbb{E}w_0(Y)\mathbb{E}w_0^*(Y) = cst, \quad \forall y \in \mathbb{S}_Y.$$

In the case $\mathbb{E}w_0(Y)\mathbb{E}w_0^(Y) = 1$, we can talk about the mean duality.*

Finally, if $Y \stackrel{d}{=} Z^w$ with positive $w(\cdot)$ and $\mathbb{E}Y = \mathbb{E}Z$, then from (1.2) and some of previous results, the relative variation index $\text{RVI}_Z(Y)$ of Y with respect to Z satisfies the following intuitive equivalence :

$$\text{RVI}_Z(Y) := \frac{\text{VI}(Y)}{\text{VI}(Z)} = \frac{\text{Var}Y}{\text{Var}Z} \geq 1 \iff w(\cdot) \geq 1. \quad (3.6)$$

That is, Y being over-, equi- and under-varied compared to Z , is interpreted through the corresponding weight (exponential) function $w(\cdot) \gtrless 1$. Especially for standard WEDs and basic VIs, we deduce similar connections (3.6) between an exponential weight function $w(\cdot)$ and $VI(Y) = RVI_Z(Y)$ with $Z \sim \mathcal{E}(\mu)$; e.g., $\mu = 1/\mathbb{E}Y$ from Corollary 3.3.4.

3.4 WEDs in EDMs and variation phenomenon

In this section, we establish an effective connection (3.6) between VIs and WEDs as members of nonnegative continuous EDMs like gamma, inverse Gaussian and nonnegative (geometric) Tweedie distributions; see, for example, Jørgensen (1997), Jørgensen and Kokonendji (2011), Abid et al. (2019, 2020), Kokonendji et al. (2020) and Bar-Lev (2020).

Let us first recall that the pdf of nonnegative continuous EDMs has the following form

$$f(x; \theta, \phi) = a(x; \phi) \exp[\theta x - K(\theta; \phi)] \mathbb{1}_{[0, \infty)}(x), \quad (3.7)$$

with $\phi > 0$ the dispersion parameter, $\theta \in \Theta \cap (-\infty, 0) \neq \emptyset$ the canonical parameter, $a(x; \phi)$ the normalizing function and $K(\theta; \phi)$ the cumulant function. When ϕ is fixed (e.g., $\phi = 1$), one gets the nonnegative natural exponential families.

Example 3.4.1. — *The reference exponential distribution $X \sim \mathcal{E}(\theta)$ of (1.4) is a NEF ($\phi = 1$ in EDM) with pdf*

$$f_X(x; \theta) = \exp[\theta x + \log(-\theta)] \mathbb{1}_{[0, \infty)}(x), \quad (3.8)$$

and $\mathbb{E}X = 1/(-\theta) =: \mu, \theta < 0$.

— *The gamma distribution $X \sim \mathcal{G}(\theta, \phi)$ is also an EDM with pdf*

$$f_X(x; \theta, \phi) = \frac{1}{\Gamma(\phi)} x^{\phi-1} \exp[x\theta + \phi \log(-\theta)] \mathbb{1}_{[0, \infty)}(x),$$

with $\mathbb{E}X = \phi/(-\theta) =: \mu, \theta < 0$ and $\alpha := \phi > 0$.

The following first result stands for a consequence of Theorem 3.3.1 in nonnegative continuous EDMs setup.

Proposition 3.4.2. *Any nonnegative continuous EDM with pdf given in (3.7) is a WED with*

$$w(\cdot; \theta, \phi) = \frac{a(\cdot; \phi)}{\mathbb{E}_\theta[a(X; \phi)]} = a(\cdot; \phi) \exp[-K(\theta; \phi)]/(-\theta), \quad \phi > 0, \theta < 0,$$

its corresponding exponential weight function and $X \sim \mathcal{E}(\theta)$ of (3.8).

Proof. From both pdfs (3.7) and (3.8), one first has

$$\begin{aligned} \mathbb{E}_\theta[a(X; \phi)] &= \int_0^\infty a(x; \phi) \exp\{\theta x + \log(-\theta)\} dx \\ &= \int_0^\infty \frac{f(x; \theta, \phi)}{\exp[\theta x - K(\theta; \phi)]} \exp[\theta x + \log(-\theta)] dx \\ &= \exp[\log(-\theta) + K(\theta; \phi)]. \end{aligned}$$

Then, one can always rewrite $f(x; \theta, \phi)$ of (3.7) as follows :

$$\begin{aligned} f(x; \theta, \phi) &= a(x; \phi) \exp[-K(\theta; \phi) - \log(-\theta)] \exp[\theta x + \log(-\theta)] \mathbb{1}_{[0, \infty)}(x) \\ &= \frac{a(x; \phi)}{\exp[\log(-\theta) + K(\theta; \phi)]} \times \exp[\theta x + \log(-\theta)] \mathbb{1}_{[0, \infty)}(x) \\ &= \frac{a(x; \phi)}{\mathbb{E}_\theta[a(X; \phi)]} \times f_X(x; \theta) \\ &= f_{X^w}(x; \theta, \phi). \end{aligned}$$

This completes the proof. ■

Next, we aim to prove an effective connection between variation phenomenon and WEDs in EDMs. To establish this result with respect to the parameter $\mu = 1/(-\theta) = \mu(\theta)$, we first need the following lemma.

Lemme 3.4.3. *If X^w follows a WED in EDMs of Proposition 3.4.2. Then, for $\mu = 1/(-\theta)$,*

$$\begin{aligned} \text{Var}_\mu(X^w) &= (\mathbb{E}_\mu X^w)^2 + \mu^4 \frac{d^2}{d\mu^2} \log \mathbb{E}_\mu[a(X; \phi)] - 2\mu^3 \frac{d}{d\mu} \log \mathbb{E}_\mu[a(X; \phi)] \\ &\quad - \mu^4 \left\{ \frac{d}{d\mu} \log \mathbb{E}_\mu[a(X; \phi)] \right\}^2. \end{aligned}$$

Proof. From Proposition 3.4.2, the pdf (3.7) of X^w can be expressed as

$$\begin{aligned} f_{X^w}(x; \theta, \phi) &= \frac{a(x; \phi)}{\mathbb{E}_\theta[a(X; \phi)]} \exp[\theta x + \log(-\theta)] \mathbb{1}_{[0, \infty)}(x) \\ &= a(x; \phi) \exp\left\{\theta x + \log(-\theta) - \log \mathbb{E}_\theta[a(X; \phi)]\right\} \mathbb{1}_{[0, \infty)}(x). \end{aligned}$$

For fixed ϕ , the above pdf $f_{X^w}(x; \theta, \phi)$ is therefore an element of the natural exponential family on $[0, \infty)$ with cumulant function

$$\theta \mapsto K(\theta; \phi) = -\log(-\theta) + \log \mathbb{E}_\theta[a(X; \phi)];$$

see, e.g., Jørgensen (1997). Given that $\theta = -1/\mu$ with $d\theta = d\mu/\mu^2$, the mean $\mathbb{E}_{\mu(\theta)}(X^w) = dK(\theta; \phi)/d\theta$ of X^w is

$$\mathbb{E}_\mu(X^w) = \frac{1}{-\theta} + \frac{d}{d\theta} \log \mathbb{E}_\theta[a(X; \phi)] = \mu + \mu^2 \frac{d}{d\mu} \log \mathbb{E}_\mu[a(X; \phi)]. \quad (3.9)$$

Next, the characteristic variance $\text{Var}_{\mu(\theta)}(X^w) = d^2K(\theta; \phi)/d\theta^2$ (function of the mean) of X^w is derived successively from (3.9) as

$$\begin{aligned} \text{Var}_\mu(X^w) &= \frac{1}{\theta^2} + \frac{d}{d\theta} \left(\frac{d}{d\theta} \log \mathbb{E}_\theta[a(X; \phi)] \right) \\ &= (\mathbb{E}_\theta X^w)^2 - \left(\frac{d}{d\theta} \log \mathbb{E}_\theta[a(X; \phi)] \right)^2 - \frac{2}{(-\theta)} \frac{d}{d\theta} \log \mathbb{E}_\theta[a(X; \phi)] \\ &\quad + \frac{d^2}{d\theta^2} \log \mathbb{E}_\theta[a(X; \phi)]. \end{aligned}$$

Using again $\theta = -1/\mu$ with $d\theta = d\mu/\mu^2$, the lemma is proved. ■

Theorem 3.4.4. *If X^w follows a WED in EDMs of Proposition 3.4.2 with*

$$\Delta_\mu := \mu \frac{d^2}{d\mu^2} \log \mathbb{E}_\mu[a(X; \phi)] - 2 \frac{d}{d\mu} \log \mathbb{E}_\mu[a(X; \phi)] - \mu \left(\frac{d}{d\mu} \log \mathbb{E}_\mu[a(X; \phi)] \right)^2.$$

Then

$$\Delta_\mu \gtrless 0 \iff X^w \gtrless X \sim \mathcal{E}(\mu);$$

i.e., Δ_μ is positive, null and negative if and only if X^w is over-, equi- and under-varied, respectively, with respect to $X \sim \mathcal{E}(\mu)$.

Proof. Using Lemma 3.4.3, the exponential variation index (1.2) of X^w is easily given by $VI_\mu(X^w) = \text{Var}_\mu(X^w)/(\mathbb{E}_\mu X^w)^2 = 1 + \mu^3 \Delta_\mu$ with $\mu > 0$. ■

Theorem 3.4.4 will be useful in situations where Δ_μ is available in an explicit form; see, for example, Section 3.6.1 for an illustration of the gamma model. Furthermore, we can point out that Theorem 3.4.4 is very different from that of weighted Poisson distributions of Kokonendji et al. (2008, Theorem 3) for which the corresponding Δ_μ is simply equal to the single term $d^2 \log \mathbb{E}_\mu[a(X; \phi)]/d\mu^2$.

As a matter of fact, the very practical use of the variation phenomenon (1.2) for any WED in EDMs shall be performed through the following proposition with direct applications of Table 3.1, which may be considered within the framework of generalized linear models. This proposition can be viewed as an analogue to the result of Corollary 3.3.4.

Proposition 3.4.5. *If X^w follows a WED in EDMs of Proposition 3.4.2 with positive $m = \mathbb{E}X^w$ and $\phi V(m/\phi) = \text{Var}_m(X^w)$, where $V(\cdot)$ is its unit variance function, then*

$$VI_m(X^w) := \phi V(m/\phi)/m^2 \gtrless 1 \iff X^w \gtrless X \sim \mathcal{E}(1/m).$$

Proof. It comes from $VI_m(X^w) = \text{Var}_m(X^w)/(\mathbb{E}_m X^w)^2$ and (1.2). ■

TABLE 3.1 – Examples of WEDs in nonnegative continuous EDMs for using Proposition 3.4.5.

Type(s) of EDMs	$V(m)$	$\phi V(m/\phi)/m^2$	Author(s)
Gamma ($p = 2$)	m^2	$1/\phi$	Morris (1982)
Inverse Gaussian ($p = 3$)	m^3	m/ϕ^2	Letac and Mora (1990)
Ressel-Kendall* ($q = 3$)	$m^2 + m^3$	$1/\phi + m/\phi^2$	Letac and Mora (1990)
Tweedie ($p > 1$)	m^p	$m^{p-2}\phi^{1-p}$	Jørgensen (1997)
Geometric Tweedie ($q > 1$)	$m^2 + m^q$	$1/\phi + m^{q-2}\phi^{1-q}$	Abid et al. (2020)

* Ressel-Kendall \equiv Geometric Inverse Gaussian.

3.5 Some existing WEDs in literature

Here, we present some WEDs available in literature. Table 3.2 depicts twelve WEDs of the statistical literature which are generally recent. The presented $w_0(\cdot)$ is normalized

(3.1) or non-normalized (3.2) when one has $\mathbb{E}w_0(X) = 1$ or $\mathbb{E}w_0(X) \neq 1$, respectively. These different weighted exponential distributions have two or three parameters. In fact, the three non-normalized $w_0(\cdot)$ of Table 3.2 depend on one parameter $\alpha > 0$ for Nos. 1 and 5 and two parameters $\alpha, \mu > 0$ for No. 6, where μ is coming from the reference exponential model $\mathcal{E}(\mu)$. The other normalized $w_0(\cdot)$ are parameterized by either $\mu, \alpha > 0$ or $\mu, \alpha, \beta > 0$. Among these twelve WEDs of Table 3.2, there are certain connections between them. For instance, No. 4 is a particular case of No. 8 with $\beta = 2$.

TABLE 3.2 – Summary of twelve WEDs from (3.1) and (3.2) with $X \sim \mathcal{E}(\mu)$.

No. $w_0(x)$	$\mathbb{E}w_0(X)$	Distributions	Author(s)
1. $x^{\alpha-1}$	$\Gamma(\alpha)\mu^{1-\alpha}$	Gamma(μ, α)	—
2. $\alpha(1 - e^{-\mu x})^{\alpha-1}$	1	EED(μ, α)	Gupta and Kundu (2001)
3. $\frac{\alpha+1}{\alpha} (1 - e^{-\alpha\mu x})$	1	WED(μ, α)	Gupta and Kundu (2009)
4. $\frac{[(\alpha+1)\mu]^2}{\alpha(\alpha+2)\mu} (1 - e^{-\alpha\mu x})x$	1	LBWED(μ, α)	Das and Kundu (2016)
5. $1 + \alpha x$	$\mu/(\alpha + \mu)$	ExtED(μ, α)	Gómez et al. (2014)
6. $\log(\alpha + e^{\mu x})$	$(\alpha + 1)\log(\alpha + 1)/\alpha$	LWED(μ, α)	Bakouch et al. (2021)
7. $\frac{\beta\Gamma(\alpha + \beta + 1)}{(1 - \beta)\Gamma(\alpha + 1)\Gamma(\beta)} \frac{(1 - e^{-\mu\alpha x})(1 - e^{-\mu x})^{\beta-1}}{\mu^{\beta-1}}$	1	WGEED(μ, α, β)	Mahdavi (2015)
8. $\frac{[1 - (1 + \alpha)^{-\beta}]\Gamma(\beta)}{(1 + \alpha)(1 + \beta)(1 + \alpha + \beta)} \frac{(1 - e^{-\mu\alpha x})x^{\beta-1}}{\alpha\beta(2 + \alpha + \beta)}$	1	WGED(μ, α, β)	Mahdavi (2015)
9. $\frac{\alpha\beta(2 + \alpha + \beta)}{(1 + \alpha)^2} \frac{(1 - e^{-\alpha\mu x})(1 - e^{-\beta\mu x})}{[\mu + \beta - (\mu + \beta + \alpha\beta\mu x)]e^{-\alpha\mu x}}$	1	TWED(μ, α, β)	Shakhatreh (2012)
10. $\frac{\alpha[\mu(1 + \alpha) + \alpha\beta]}{\alpha^{\beta}} \frac{[\mu + \beta - (\mu + \beta + \alpha\beta\mu x)]e^{-\alpha\mu x}}{(1 - e^{-\mu x})^{\alpha-1}[-\log(1 - e^{-\mu x})]^{\beta-1}}$	1	EWED(μ, α, β)	Mahdavi and Jabbari (2017)
11. $\frac{\alpha^{\beta}}{\Gamma(\beta)} (1 - e^{-\mu x})^{\alpha-1} [-\log(1 - e^{-\mu x})]^{\beta-1}$	1	GEED(μ, α, β)	Ristić and Balakrishnan (2012)
12. $\frac{\alpha + 1}{\alpha} \mathbb{1}_{[0, 2\pi]}(x) \sum_{m=0}^{\infty} \exp[-2m\pi\mu(1 - e^{-\alpha\mu(x+2m\pi)})]$	1	WWED(μ, α)	Roy and Adnan (2012)

From Theorem 3.3.5, No. 9 is regarded as a product of two weight functions of No. 3 that can be expressed in terms of : $w_{9;\mu,\alpha,\beta}(\cdot) = w_{3;\mu,\alpha}(\cdot) \times w_{3;\mu,\beta}(\cdot)$. No. 7 appears to be a product of weight functions of No. 3 and No. 2 : $w_{7;\mu,\alpha,\beta}(\cdot) = w_{3;\mu,\alpha}(\cdot) \times w_{2;\mu,\beta}(\cdot)$. No. 8 is linked to No. 3 and No. 1 by : $w_{8;\mu,\alpha,\beta}(\cdot) = w_{3;\mu,\alpha}(\cdot) \times w_{1;\beta}(\cdot)$. Finally, we get : $w_{2;\mu,2}(\cdot) = w_{3;\mu,1}(\cdot)$.

3.5.1 Exponential weight of Gupta and Kundu (2001)

In this part, we consider a new two-parameter exponentiated exponential distribution (EED) introduced by Gupta and Kundu (2001). The author has introduced it to make comments both positive and negative of this family with respect to a Weibull family and a gamma family and give the practitioner one more option, with a hope that it may have a "better fit" compared to a Weibull family or a gamma family in certain situations. The two parameters of this distribution represent the shape and the scale parameter like as for the gamma and Weibull distributions. It is observed that many properties of this new family are quite similar to those of a Weibull or a gamma family, therefore this distribution can be used as a possible alternative to a Weibull or a gamma distribution. The pdf of the weighted exponential distribution EED (μ, α) is defined by

$$f_2(x; \mu, \alpha) = \alpha\mu(1 - \exp(-\mu x))^{\alpha-1} \exp(-\mu x) \mathbb{1}_{(0,\infty)}(x),$$

with $\mu > 0, \alpha > 0$.

Notice that, for the shape parameter $\alpha = 1$, we have the pdf of the exponential distribution. Therefore, the three families gamma, Weibull and EED are a generalization of the exponential family, but in different ways. According to the author, the pdf of the EED (μ, α) distribution has also a good interpretation in physical domain. Assuming there are n -components in a parallel system and the distribution of the life of each component is independent and identically distributed. If the distribution of each component follows an EED distribution, then the lifetime distribution of the entire system is also an EED distribution. It is observed that it has lots of properties which are quite

similar to those of a gamma distribution but it has an explicit expression of the distribution function or the survival function like a Weibull distribution. It has also likelihood ratio ordering with respect to the shape parameter, when the scale parameter is kept constant.

3.5.2 Modified exponential weight of Gupta and Kundu (2009)

Azzalini (1985) proposed a new distribution to introduce an additional parameter on the normal distribution. This distribution is well known in the statistical literature as skew-normal and it has received considerable attention over the past two decades. Kundu and Gupta (2009) introduced a new class of weighted exponential distribution (WED) using the idea of Azzalini (1985). This distribution is similar to the EED distribution except for a small change in the weight function. A random variable X follows a WED, with parameters $\mu, \alpha > 0$, if the pdf of X is written by

$$f_2(x; \mu, \alpha) = \frac{\alpha + 1}{\alpha} \mu \exp(-\mu x) (1 - \exp(-\mu \alpha x)) \mathbb{1}_{(0, \infty)}(x).$$

This model can be obtained from two independent and identically distributed random variables which is exactly the same way that Azzalini obtained the skewed normal distribution from two iid variables with normal distributions. Assuming that X_1 and X_2 are two iid random variables with density $f_Y(\cdot)$ and the distribution function $F_Y(\cdot)$, then for all $\alpha > 0$, consider a new random variable $X = X_1$ such that $\alpha X_1 > X_2$. Note that the density of the new random variable X is

$$f(x) = \frac{1}{\mathbb{P}(\alpha X_1 > X_2)} f_Y(x) F_Y(\alpha x), \quad x > 0.$$

Then, the density $f_2(x; \mu, \alpha)$ of WED (μ, α) is obtained from the above equation by replacing $f_Y(x) = \exp(-\mu x)$ and $F_Y(x) = 1 - \exp(-\mu x)$.

3.6 Illustrations

In this section, we shall exhibit some examples where results displayed in the preceding sections prove to be very useful. These examples of WEDs with possible

extensions cover three families of weight functions from the usual gamma, Weibull and lognormal distributions.

3.6.1 Gamma and modified gamma distributions

The gamma distribution (No. 1 of Table 3.2) is a natural WED in the sense that a particular value ($\alpha = 1$) of its parameters gives the reference exponential distribution. Its has been already discussed for illustrations of the duality concept and as an EDM membership through Table 3.1. Consider the gamma random variable $Y \sim \mathcal{G}(\mu, \alpha)$ for $\mu, \alpha > 0$ such that $\mathbb{E}Y = \alpha/\mu$ with the normalized exponential weight function $w_{1;\mu,\alpha}(x) = (x\mu)^{\alpha-1}/\Gamma(\alpha)$. Its self-decomposition in WED from Corollary 3.3.4 provides the weight function

$$\widetilde{w}_{1;\mu,\alpha}(x) = \frac{\alpha(\mu x)^{\alpha-1}}{\Gamma(\alpha)} \exp\left\{\left(\frac{1}{\alpha} - 1\right)\mu x\right\} = \alpha \exp\left\{\left(\frac{1}{\alpha} - 1\right)\mu x\right\} w_{1;\mu,\alpha}(x),$$

which also illustrates Proposition 3.3.3.

An application of Theorem 3.4.4 can be done here for gamma $\mathcal{G}(\mu, \alpha)$ as an EDM with its normalizing function $a(x; \alpha) = x^{\alpha-1}/\Gamma(\alpha)$ and $\alpha = \phi > 0$. In fact, one has successively $\mathbb{E}_\mu[a(X; \alpha)] = \mu^{\alpha-1}$, $d \log \mathbb{E}_\mu[a(X; \alpha)]/d\mu = (\alpha - 1)/\mu$, $d^2 \log \mathbb{E}_\mu[a(X; \alpha)]/d\mu^2 = (1 - \alpha)/\mu^2$ and, therefore, $\Delta_\mu = (1 - \alpha)(2 + \alpha)/\mu$ which is positive, null and negative if and only if $\alpha < 1$, $\alpha = 1$ and $\alpha > 1$, respectively. As matter of fact, one obtains the same conditions for gamma to be exponentially over-, equi- and under-varied, respectively.

In order to illustrate the composition of two weight functions (i.e., Part (iv) of Remark 3.3.6) with respect to gamma, we utilize once again Table 3.2 as follows : $w_{2;\mu,\alpha}(\cdot) = w_{3;\mu,1}(\cdot) \circ w_{1;\alpha}(\cdot)$ and $w_{7;\mu,\alpha,\beta}(\cdot) = w_{3;\mu,\alpha}(\cdot) \times \{w_{2;\mu,2}(\cdot) \circ w_{1;\beta}(\cdot)\}$. In other words, by changing the standard reference exponential to the gamma one (i.e., Part (iii) of Remark 3.3.6), the weighted gamma-exponential distribution (WGED) of Mahdavi (2015) in Table 3.2 (No. 8) can be represented through the (normalized) gamma weight function as follows : $w_{8;\mu,\alpha,\beta}(x)/w_{1;\mu,\beta}(x) = [1 - (1 + \alpha)^{-\beta}]^{-1}(1 - e^{-\mu\alpha x})\mathbb{1}_{(0,\infty)}(x)$.

3.6.2 Weibull and some of its modified distributions

Let $Y \sim \mathcal{W}(\mu, \alpha, \gamma)$ be the three-parameter Weibull random variable with pdf

$$f_Y(y) = \alpha \mu [\mu(y - \gamma)]^{\alpha-1} \exp\{-\mu(y - \gamma)^\alpha\} \mathbb{1}_{[\gamma, \infty)}(y),$$

for $\mu, \alpha > 0$ and $\gamma \geq 0$ such that $\mathbb{E}Y = \gamma + \mu\Gamma(1/\alpha)/\alpha$. It is a natural WED for $\alpha = 1$ and $\gamma = 0$. Its exponential weight function is provided by $w_{\gamma; \mu, \alpha}(x) = \alpha [\mu(x - \gamma)]^{\alpha-1} \exp\{-\mu(x - \gamma)^\alpha\}$, where $w_{0; \mu, \alpha}(\cdot)$ is the one of the (two-parameter) Weibull distribution with $\gamma = 0$. From Corollary 3.3.4, we obtain the weight function of self-decomposition as WED by :

$$\widetilde{w}_{\gamma; \mu, \alpha}(x) = \mu[\alpha\gamma + \mu\Gamma(1/\alpha)][\mu(x - \gamma)]^{\alpha-1} \exp\left\{\frac{\alpha x}{\alpha\gamma + \mu\Gamma(1/\alpha)} - \mu(x - \gamma)^\alpha\right\}.$$

Using Proposition 3.3.3, it is easy to connect $\widetilde{w}_{\gamma; \mu, \alpha}(\cdot)$ with $w_{\gamma; \mu, \alpha}(\cdot)$.

Ramadan (2013) has proposed a class of weighted Weibull distributions having the following normalized weight function :

$$w_{0; \mu, \alpha, \beta}(x) = (1 + \beta^\alpha)[1 - \exp\{-\mu(x\beta)^\alpha\}]/\beta^\alpha.$$

Previous investigation could be carried out to get its weight function of self-decomposition $\widetilde{w}_{0; \mu, \alpha, \beta}(\cdot)$. Given $\alpha = 1$, one gets the weighted exponential distribution (No. 3 of Table 3.2) of Gupta and Kundu (2009) : $w_{0; \mu, 1, \beta}(\cdot) = w_{3; \mu, \beta}(\cdot)$. Other particular cases can be obtained for weighted extreme value distribution and also for weighted Rayleigh distribution.

Another weighted Weibull distribution has been introduced by Shahbaz et al. (2010) with normalized weight function

$$w_{0^*; \mu, \alpha, \beta}(x) = (\alpha + 1)[1 - \exp(-\mu\beta x^\alpha)]/\alpha.$$

In addition, for $\alpha = 1$ one has : $w_{0^*; \mu, 1, \beta}(\cdot) = w_{3; \mu, \beta}(\cdot)$ of Table 3.2. Similar to exponential and gamma models, several weighting operations could be conducted to build new weighted Weibull distributions. See, for example, Domma et al. (2017) and Elsherpieny et al. (2017) for some recent weighted Weibull distributions. Finally, one can refer to Al-Mutairia et al. (2018) for weighted Weibull distributions in bivariate and multivariate cases.

3.6.3 Lognormal distribution as a WED

Considering $Y \sim \mathcal{LN}(m, \sigma^2)$ for $m \in \mathbb{R}, \sigma > 0$ the lognormal random variable with pdf

$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log y - m)^2}{2\sigma^2}\right\} \mathbb{1}_{(0,\infty)}(y),$$

such that $\mathbb{E}Y = \exp(m + \sigma^2/2)$. It is clear to see that the lognormal distribution is not a natural WED. Afterwards, Corollary 3.3.4 provides its self-decomposition as a WED with normalized weight function :

$$\tilde{w}_{m,\sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{m + \frac{\sigma^2}{2} - \frac{(\log x - m)^2}{2\sigma^2} + x \exp\left(m + \frac{\sigma^2}{2}\right)\right\}.$$

To our knowledge, there is not yet a weighted lognormal distribution. However, the previous weighting operations could be considered for constructing new distributions grounded on the lognormal one if they are needed in applications.

3.7 Conclusion

In this chapter, we have shown that any positive continuous random variable can be a weighted exponential distribution. We have also proposed several weighting properties including a given reference distribution. Finally, we made the connection with the recent variation index. We have seen that many weighted exponential distributions are related to each other by properties. This shows the importance of weighted exponential distributions in statistical modeling such as reliability.

Chapter 4

Relative variation indexes for distributions on $[0, \infty)^k$ and extensions

4.1 Introduction

Behind the Gaussian distribution and similar to the Poisson distribution for count models (e.g., Kokonendji, 2014), we probably have the exponential distribution on the positive half real line which is the most common probability distributions for this support. Since both univariate concepts of VI and of the well-known Fisher (1934) dispersion index with respect to the equi-dispersed Poisson model are similar, we here suggest first a useful and appropriate definition of multivariate over-, equi- and under-variation following the multivariate dispersion indexes of Kokonendji and Puig (2018). Then, we mainly propose an extension for unifying multivariate dispersion and multivariate variation indexes in the framework of natural exponential families.

This chapter is structured as follows. Section 4.2 presents notations, generalized and relative variation indexes with its interpretation and properties for practical handling. Section 4.3 illustrates calculations of these measures on some usual bi- and multivariate continuous non-negative orthant distributions such as beta, exponential and Weibull. Section 4.4 provides asymptotic properties of the corresponding estimators. Section 4.5 reveals example applications from real life and simulated continuous (non-negative orthant) datasets under several scenarios, and produces some simulation studies. Section 4.6 is reserved for the unification of the new generalized variation

index with the existing generalized dispersion index of Kokonendji and puig (2018).

Finally, Section 4.8 gathers some demonstrations and complements for this chapter.

4.2 Multivariate variation indexes

Let $\mathbf{Y} = (Y_1, \dots, Y_k)^\top$ be a non-negative continuous k -variate random vector on $[0, \infty)^k$, $k \geq 1$. We consider the following notations : $\sqrt{\text{var}\mathbf{Y}} = (\sqrt{\text{var}Y_1}, \dots, \sqrt{\text{var}Y_k})^\top$ is the elementwise square root of the variance vector of \mathbf{Y} ; $\text{diag } \sqrt{\text{var}\mathbf{Y}} = \text{diag}_k(\sqrt{\text{var}Y_j})$ is the $k \times k$ diagonal matrix with diagonal entries $\sqrt{\text{var}Y_j}$ and 0 elsewhere; and, $\text{cov}\mathbf{Y} = (\text{cov}(Y_i, Y_j))_{i,j \in \{1, \dots, k\}}$ denotes the covariance matrix of \mathbf{Y} which is a $k \times k$ symmetric matrix with entries $\text{cov}(Y_i, Y_j)$ such that $\text{cov}(Y_i, Y_i) = \text{var}Y_i$ is the variance of Y_i . Then

$$\text{cov}\mathbf{Y} = (\text{diag } \sqrt{\text{var}\mathbf{Y}})(\rho_{\mathbf{Y}})(\text{diag } \sqrt{\text{var}\mathbf{Y}}), \quad (4.1)$$

where $\rho_{\mathbf{Y}} = \rho(\mathbf{Y})$ is the correlation matrix of \mathbf{Y} ; see, e.g., Johnson and Wichern (2007, Eq. 2-36). Note that there are infinitely many multivariate distributions with exponential margins. We denote a generic k -variate exponential distribution by $\mathcal{E}_k(\boldsymbol{\mu}, \boldsymbol{\rho})$, given specific positive mean vector $\boldsymbol{\mu}^{-1} := (\mu_1^{-1}, \dots, \mu_k^{-1})^\top$ and correlation matrix $\boldsymbol{\rho} = (\rho_{ij})_{i,j \in \{1, \dots, k\}}$; see, e.g., Appendix A for a broader one. The uncorrelated or independent k -variate exponential will be written as $\mathcal{E}_k(\boldsymbol{\mu})$, for $\boldsymbol{\rho} = \mathbf{I}_k$ the $k \times k$ unit matrix.

4.2.1 Basic definitions

Proceeding along similar lines as Albert and Zhang (2010) and also as Kokonendji and Puig (2018), we define the *generalized variation index* of \mathbf{Y} by

$$\text{GVI}(\mathbf{Y}) = \frac{\mathbb{E}\mathbf{Y}^\top (\text{cov}\mathbf{Y}) \mathbb{E}\mathbf{Y}}{(\mathbb{E}\mathbf{Y}^\top \mathbb{E}\mathbf{Y})^2}; \quad (4.2)$$

see Section 4.8 for its construction and below for its interpretation and more properties with illustrations. Remark that when $k = 1$, GVI is the univariate variation index VI (Abid et al., 2020). The *relative (generalized) variation index* is defined, for two continuous

random vectors \mathbf{X} and \mathbf{Y} on the same support $\mathcal{S} = [0, \infty)^k$ with $\mathbb{E}\mathbf{X} = \mathbb{E}\mathbf{Y}$ and $\text{GVI}(\mathbf{X}) > 0$, by

$$\text{RVI}_{\mathbf{X}}(\mathbf{Y}) := \frac{\text{GVI}(\mathbf{Y})}{\text{GVI}(\mathbf{X})} = \frac{\mathbb{E}\mathbf{Y}^\top (\text{cov}\mathbf{Y}) \mathbb{E}\mathbf{Y}}{\mathbb{E}\mathbf{X}^\top (\text{cov}\mathbf{X}) \mathbb{E}\mathbf{X}} \gtrless 1; \quad (4.3)$$

i.e., the **over-** (**equi-** and **under-variation**) of \mathbf{Y} compared to \mathbf{X} , and denoted by $\mathbf{Y} > \mathbf{X}$ ($\mathbf{Y} \asymp \mathbf{X}$ and $\mathbf{Y} < \mathbf{X}$), is realized if $\text{GVI}(\mathbf{Y}) > \text{GVI}(\mathbf{X})$ ($\text{GVI}(\mathbf{Y}) = \text{GVI}(\mathbf{X})$ and $\text{GVI}(\mathbf{Y}) < \text{GVI}(\mathbf{X})$, respectively). For instance, it is possible to derive asymptotic properties of the empirical estimator of RVI as well as GVI in Section 4.4. Afterwards, this would be of interest for conclusions like $\mathbf{Y} > \mathbf{X}$ significantly.

In the framework of the natural exponential family F_Y on $[0, \infty)^k$ (e.g., Chapter 54 in Kotz et al., 2000), generated by the distribution of \mathbf{Y} and characterized by its variance function $\mathbf{m} \mapsto \mathbf{V}_{F_Y}(\mathbf{m})$, the index GVI can also be rewritten via \mathbf{m} and $\mathbf{V}_{F_Y}(\mathbf{m})$. As for (4.2), the “generalized variation function” defined on the mean domain $\mathbf{M}_{F_Y} \subseteq (0, \infty)^k$ to $(0, \infty)$ by

$$\mathbf{m} \mapsto \text{GVI}_{F_Y}(\mathbf{m}) = \frac{\mathbf{m}^\top \{\mathbf{V}_{F_Y}(\mathbf{m})\} \mathbf{m}}{(\mathbf{m}^\top \mathbf{m})^2} \quad (4.4)$$

appears to be very useful through this parameterization.

4.2.2 Interpretation and properties

Concerning an interpretation of GVI, we first express the denominator of (4.2) as $\mathbb{E}\mathbf{Y}^\top \mathbb{E}\mathbf{Y} = \sqrt{\mathbb{E}\mathbf{Y}^\top} (\text{diag} \mathbb{E}\mathbf{Y}) \sqrt{\mathbb{E}\mathbf{Y}}$, using then (4.1) to rewrite $\text{cov}\mathbf{Y}$, obtaining

$$\text{GVI}(\mathbf{Y}) = \frac{\{(\text{diag} \sqrt{\text{var}\mathbf{Y}}) \mathbb{E}\mathbf{Y}\}^\top (\rho_Y) \{(\text{diag} \sqrt{\text{var}\mathbf{Y}}) \mathbb{E}\mathbf{Y}\}}{\left\{[(\text{diag} \sqrt{\mathbb{E}\mathbf{Y}}) \sqrt{\mathbb{E}\mathbf{Y}}]^\top (\mathbf{I}_k) [(\text{diag} \sqrt{\mathbb{E}\mathbf{Y}}) \sqrt{\mathbb{E}\mathbf{Y}}]\right\}^2}. \quad (4.5)$$

From (4.5), it is clear that $\text{GVI}(\mathbf{Y})$ makes it possible to compare the full variability of \mathbf{Y} (in the numerator) with respect to its expected uncorrelated exponential variability (in the denominator) which depends only on $\mathbb{E}\mathbf{Y}$.

Next, the GVI index can be considered in itself as a notion of k -variate over-, equi- and under-variation.

Proposition 4.2.1. *For all positive continuous random vector \mathbf{Y} on $(0, \infty)^k$, $k \geq 1$, and $\mathbf{X} \sim \mathcal{E}_k(\boldsymbol{\mu})$ then $\text{RVI}_{\mathbf{X}}(\mathbf{Y}) = \text{GVI}(\mathbf{Y})$. Furthermore, one has*

$$\mathbf{Y} \sim \mathcal{E}_k(\boldsymbol{\mu}, \boldsymbol{\rho}) \quad \Rightarrow \quad \text{GVI}(\mathbf{Y}) = 1 + \frac{\boldsymbol{\mu}^{-\top}(\boldsymbol{\rho} - \mathbf{I}_k)\boldsymbol{\mu}^{-1}}{(\boldsymbol{\mu}^{-\top}\boldsymbol{\mu}^{-1})^2} \gtrless 1, \quad (4.6)$$

with $\boldsymbol{\mu}^{-\top} := (\boldsymbol{\mu}^{-1})^\top = (\mu_1^{-1}, \dots, \mu_k^{-1})$.

Proof. It is trivial from (4.5), with $\text{GVI}(\mathbf{Y}) = 1$ if and only if $\boldsymbol{\rho} = \mathbf{I}_k$, i.e., $\mathbf{Y} \sim \mathcal{E}_k(\boldsymbol{\mu})$. ■

From Proposition 4.2.1, the multivariate exponential model $\mathbf{Y} \sim \mathcal{E}_k(\boldsymbol{\mu}, \boldsymbol{\rho})$ can be over-, equi- or under-varied (with respect to the uncorrelated exponential) according to its correlation structure. For instance, if $k = 2$ then (4.6) clearly gives the one-to-one relationship, viz.

$$\text{GVI}(\mathbf{Y}) = 1 + \frac{2\rho\mu_1^{-2}\mu_2^{-2}}{(\mu_1^{-2} + \mu_2^{-2})^2} \gtrless 1 \quad \Leftrightarrow \quad \rho \gtrless 0.$$

Finally, if we only want to take into account the variation information coming from the margins, we can modify GVI by replacing $\text{cov}\mathbf{Y}$ in (4.2) with $\text{diag}\text{var}\mathbf{Y}$, that is $\boldsymbol{\rho} = \mathbf{I}_k$ in (4.1), obtaining the “multiple marginal variation index”, viz.

$$\text{MVI}(\mathbf{Y}) = \frac{\mathbb{E}\mathbf{Y}^\top(\text{diag}\text{var}\mathbf{Y})\mathbb{E}\mathbf{Y}}{(\mathbb{E}\mathbf{Y}^\top\mathbb{E}\mathbf{Y})^2} = \sum_{j=1}^k \frac{(\mathbb{E}Y_j)^4}{(\mathbb{E}\mathbf{Y}^\top\mathbb{E}\mathbf{Y})^2} \text{VI}(Y_j). \quad (4.7)$$

The expression on the right-hand side of (4.7) provides a representation of MVI as a weighted average of the univariate variation indexes VI of the components. MVI could be used for exploring profile distributions in multiple positive response regression models (Bonat and Jørgensen, 2016) or in multivariate continuous time series. Similarly to (4.4), the corresponding “multiple marginal variation function” is defined on the mean domain $\mathbf{M}_{F_Y} \subseteq (0, \infty)^k$ to $(0, \infty)$ by

$$\mathbf{m} \mapsto \text{MVI}_{F_Y}(\mathbf{m}) = \frac{\mathbf{m}^\top \{\text{diag}\mathbf{V}_{F_Y}(\mathbf{m})\}\mathbf{m}}{(\mathbf{m}^\top \mathbf{m})^2}. \quad (4.8)$$

In the same way as (4.3), the relative versions of MVI can be introduced.

4.3 Illustrations and comments

The use of our variation indexes shall be illustrated by two bivariate models and two families of general k -variate ones. It will then be seen how the marginal VIs interplay with the correlation structure in the multivariate variation measures discussed previously. Considering $\mathbf{Y} = (Y_1, Y_2)^\top$ and using (4.7), we explicitly write

$$\text{GVI}(Y_1, Y_2) = \text{MVI}(Y_1, Y_2) + \rho \frac{2(\mathbb{E}Y_1)^2(\mathbb{E}Y_2)^2 \sqrt{\text{VI}(Y_1)\text{VI}(Y_2)}}{\{(\mathbb{E}Y_1)^2 + (\mathbb{E}Y_2)^2\}^2}, \quad (4.9)$$

which highlights that GVI is not a weighted average of VIs, as MVI is. It is noteworthy that $\text{GVI} \geq \text{MVI}$ accordingly to $\rho \geq 0$, with $\text{GVI} = \text{MVI}$ for $\rho = 0$. Similar remarks hold for the k -variate cases, where the correlation matrix ρ is reduced to \mathbf{I}_k .

4.3.1 Bivariate beta distribution of Arnold and Tony Ng (2011)

The flexible bivariate beta $\mathbf{Y} = (Y_1, Y_2)^\top \sim \mathcal{B}_2(\boldsymbol{\alpha})$ of Arnold and Tony Ng (2011) which exhibits both positive and negative correlation between random variables can be defined as follows. Suppose that U_1, U_2, V_1, V_2 and W are independent gamma random variables with common unit scale parameter, i.e., $U_j \sim \mathcal{G}_1(\alpha_j, 1)$, $V_j \sim \mathcal{G}_1(\alpha'_j, 1)$, $j = 1, 2$ and $W \sim \mathcal{G}_1(\alpha_0, 1)$ with $\alpha_j > 0$, $\alpha'_j > 0$ and $\alpha_0 > 0$. Then, for $j = 1, 2$ and $\boldsymbol{\alpha} := (\alpha_0, \alpha_1, \alpha_2, \alpha'_1, \alpha'_2)$, one has

$$Y_j := \frac{U_j + V_j}{U_j + V_1 + V_2 + W} \sim \mathcal{B}_1(\alpha_j + \alpha'_j, \alpha_0 + \alpha'_1 + \alpha'_2 - \alpha'_j).$$

Since the covariance of Y_1 and Y_2 cannot be expressed in closed form, it has been numerically shown that the correlation $\rho = \rho(Y_1, Y_2) = \rho(\boldsymbol{\alpha})$ belongs into $[-1, 1]$. In fact, from the proposed construction, the positive correlations are obtained when $\alpha'_1 = \alpha'_2 = 0$. For negative correlations, one can consider $\alpha_0 = 0$ with α_1 and α_2 fixed, and it will be get closer to -1 as α'_1 and α'_2 get larger. See Arnold and Tony Ng (2011) for more details and connected references.

For given $\rho(\boldsymbol{\alpha}) \in [-1, 1]$ of $\mathbf{Y} = (Y_1, Y_2)^\top \sim \mathcal{B}_2(\boldsymbol{\alpha})$, the direct calculations of $\text{GVI}(\mathbf{Y})$ through (4.9) and $\text{MVI}(\mathbf{Y})$ via (4.7) are obtained from the following first moments and

VIs of the univariate beta random variables $Y_j, j = 1, 2$:

$$\mathbb{E}Y_j = \frac{\alpha_j + \alpha'_j}{\alpha_0 + \alpha'_1 + \alpha'_2 + \alpha_j} > 0$$

and

$$\begin{aligned} \text{VI}(Y_j) &= \left(\frac{1 + \alpha_j + \alpha'_j}{1 + \alpha_0 + \alpha'_1 + \alpha'_2 + \alpha'_j} - \mathbb{E}Y_j \right) / \mathbb{E}Y_j \\ &= \frac{\alpha_0 + \alpha'_1 + \alpha'_2 - \alpha'_j}{(1 + \alpha_0 + \alpha'_1 + \alpha'_2 + \alpha'_j)(\alpha_j + \alpha'_j)} > 0. \end{aligned}$$

Knowing that Y_1 and Y_2 are over-, equi- and under-varied, then $\mathbf{Y} = (Y_1, Y_2)^\top \sim \mathcal{B}_2(\boldsymbol{\alpha})$ can also be in the bivariate sense and according to the values of $\boldsymbol{\alpha}$. A k -variate extension of this bivariate beta distribution is available for tedious calculations of their multivariate variation indexes.

4.3.2 Bivariate Weibull distribution of Teimouri and Gupta (2011)

Given the bivariate Weibull $\mathbf{Y} = (Y_1, Y_2)^\top \sim \mathcal{W}_2(\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma, \delta)$ of Teimouri and Gupta (2011) built by a copula with density

$$\begin{aligned} f_Y(y_1, y_2) &= \frac{\beta_1 \beta_2}{\alpha_1 \alpha_2} \left(\frac{y_1}{\alpha_1} \right)^{\beta_1-1} \left(\frac{y_2}{\alpha_2} \right)^{\beta_2-1} \exp \left\{ - \left(\frac{y_1}{\alpha_1} \right)^{\beta_1} - \left(\frac{y_2}{\alpha_2} \right)^{\beta_2} \right\} \\ &\quad \times (1 + \delta \exp \left[-(\gamma - 1) \left\{ \left(\frac{y_1}{\alpha_1} \right)^{\beta_1} + \left(\frac{y_2}{\alpha_2} \right)^{\beta_2} \right\} \right] \\ &\quad \times \left[\exp \left\{ - \left(\frac{y_1}{\alpha_1} \right)^{\beta_1} \right\} - \gamma \right] \left[\exp \left\{ - \left(\frac{y_2}{\alpha_2} \right)^{\beta_2} \right\} - \gamma \right] \right), \end{aligned}$$

such that $\alpha_j > 0$ and $\beta_j > 0, j = 1, 2, \gamma > 1$ and $\delta \in [0, 1]$. For $\delta = 0$ one gets the uncorrelated bivariate Weibull distribution which depends on both scale parameters α_j and shape parameters $\beta_j, j = 1, 2$.

Since we explicitly have the first, second and product moments of Y_1 and Y_2 (Teimouri and Gupta, 2011), the calculations of $\text{GVI}(\mathbf{Y})$ using (4.9) and $\text{MVI}(\mathbf{Y})$ through (4.7) are derived from the following first moments, VIs and correlation of $Y_j, j = 1, 2$:

$$\mathbb{E}Y_j = \alpha_j \Gamma(1 + 1/\beta_j),$$

$$VI(Y_j) = \frac{\Gamma(1 + 2/\beta_j)}{\{\Gamma(1 + 1/\beta_j)\}^2} - 1$$

depending only on shape parameter β_j , and

$$\begin{aligned} \rho = \rho(Y_1, Y_2) &= \delta \Gamma(1 + 1/\beta_1) \Gamma(1 + 1/\beta_2) \\ &\times [\gamma^{-2-1/\beta_1-1/\beta_2} - (\gamma + 1) \gamma^{-1-1/\beta_1} \{\gamma^{-1-1/\beta_2} - (\gamma + 1)^{-1-1/\beta_2}\} \\ &- (\gamma + 1) \gamma^{-1-1/\beta_2} \{\gamma^{-1-1/\beta_1} - (\gamma + 1)^{-1-1/\beta_1}\} \\ &+ (\gamma + 1)^2 \{\gamma^{-1-1/\beta_1} - (\gamma + 1)^{-1-1/\beta_1}\} \{\gamma^{-1-1/\beta_2} - (\gamma + 1)^{-1-1/\beta_2}\}] \\ &\times \left[\{\Gamma(1 + 2/\beta_1) - \Gamma^2(1 + 1/\beta_1)\} \{\Gamma(1 + 2/\beta_2) - \Gamma^2(1 + 1/\beta_2)\} \right]^{-1/2}, \end{aligned}$$

where $\Gamma(\cdot)$ is the classical gamma function. The above univariate variation indexes $VI(Y_j)$, $j = 1, 2$, satisfy the following equivalences :

$$0 < VI(Y_j) \lessgtr 1 \stackrel{(i)}{\iff} 0 < \beta_j \Gamma(2/\beta_j) \{\Gamma(1/\beta_j)\}^{-2} \lessgtr 1 \stackrel{(ii)}{\iff} 0 < \beta_j \lessgtr 1, \quad (4.10)$$

for all $\alpha_j > 0$ fixed. Indeed, the first equivalence (i) of (4.10) is derived from the gamma duplication formula or Legendre's doubling formula - revisited (e.g., Abramowitz and Stegun, 1972 ; Chapter 6), and the second one (ii) stems from the function study (see, e.g., Table 4.1) which will still needs formal proof. Note that, in reliability, conditions (4.10) on the shape parameter $\beta \lessgtr 1$ of the univariate Weibull distribution $\mathcal{W}_1(\alpha, \beta)$ are known in the opposite sense as $\beta \lessgtr 1$ for the classical failure rate (or bathtub) curve from $\lambda_{\mathcal{W}_1}(t) = (\beta/\alpha)(t/\alpha)^{\beta-1} \mathbb{1}_{(0, \infty)}(t)$; i.e., $\lambda_{\mathcal{W}_1}(\cdot)$ is decreasing/constant/increasing if and only if $\beta \lessgtr 1$, respectively. One can refer to Barlow and Proschan (1981) using the standard coefficient of variation. Hence, according to its parameters this bivariate Weibull distribution can be over-, equi- and under-varied with respect to the uncorrelated bivariate exponential distribution.

TABLE 4.1 – Some values of equivalences (4.10) for the univariate Weibull distribution.

β	0.1	0.3	0.5	0.8	1	2	4	10	100
$\beta \Gamma(2/\beta) \{\Gamma(1/\beta)\}^{-2}$	92378	15.12	3	1.29	1	0.64	0.54	0.5072	0.5001
$VI(\beta)$	184755	29.24	5	1.59	1	0.27	0.08	0.0145	0.0002

4.3.3 Multivariate exponential distribution of Marshall and Olkin (1967)

The k -variate exponential $\mathbf{Y} = (Y_1, \dots, Y_k)^\top \sim \mathcal{E}_k(\mu_1, \dots, \mu_k, \mu_0)$ of Marshall and Olkin (1967) is constructed as follows. Let X_1, \dots, X_k and Z be univariate exponential random variables with parameters $\mu_1 > 0, \dots, \mu_k > 0$ and $\mu_0 \geq 0$, respectively. Then, by setting $Y_j := X_j + Z$ for $j = 1, \dots, k$, one easily has $\mathbb{E}Y_j = 1/(\mu_j + \mu_0) = \sqrt{\text{var}Y_j}$ and $\text{cov}(Y_j, Y_\ell) = \mu_0/[(\mu_j + \mu_0)(\mu_\ell + \mu_0)(\mu_j + \mu_\ell + \mu_0)]$ for all $j \neq \ell$. Point out that each correlation $\rho(Y_j, Y_\ell) = \mu_0/(\mu_j + \mu_\ell + \mu_0)$ lies in $[0, 1]$ if and only if $\mu_0 \geq 0$.

Using (4.6) appropriately, we obtain

$$\text{GVI}(\mathbf{Y}) = 1 + \frac{\mu_0 \sum_{j=1}^k (\mu_j + \mu_0)^{-1} \{ \sum_{\ell \neq j} (\mu_j + \mu_\ell + \mu_0)^{-1} (\mu_\ell + \mu_0)^{-1} \}}{\{(\mu_1 + \mu_0)^{-2} + \dots + (\mu_k + \mu_0)^{-2}\}^2} \geq 1 \quad (\Leftrightarrow \mu_0 \geq 0).$$

Through (4.7), we easily have

$$\text{MVI}(\mathbf{Y}) = \frac{\sum_{j=1}^k (\mu_j + \mu_0)^{-4}}{\sum_{j=1}^k (\mu_j + \mu_0)^{-4} + 2 \sum_{1 \leq j < \ell \leq k} (\mu_j + \mu_0)^{-2} (\mu_\ell + \mu_0)^{-2}} < 1.$$

Hence, this multivariate exponential model is always under-varied with respect to the MVI and over- or equi-varied with respect to GVI. If $\mu_0 = 0$ then this k -variate exponential distribution is reduced to $\mathcal{E}_k(\boldsymbol{\mu})$ with $\text{GVI}(\mathbf{Y}) = 1$. However, the assumption of non-negative correlations between components is sometimes insufficient for some analyzes. We can refer to Appendix A for a more extensive exponential model which is derived as a particular case of a full multivariate Tweedie (1984) models with a flexible dependence structure (Cuenin et al., 2016).

4.3.4 Multiple stable Tweedie (MST) models

Consider the huge k -variate MST class of families $F_{\mathbf{p}} = F_{\mathbf{p}}(m_1, \dots, m_k, \lambda)$ of models which have been introduced in Boubacar Maïnassara and Kokonendji (2014) to extend the so-called normal stable Tweedie (NST) with $\mathbf{p} = (p_1, 0, \dots, 0)$ for $p_1 \geq 1$. The normal inverse Gaussian (NIG) model is a common particular case of NST with

$p_1 = 3$ (Barndorff-Nielsen, 1997). This MST class contains infinite subclasses of multivariate distributions among others the gamma-MST with $\mathbf{p} = (2, p_2, \dots, p_k)$ and inverse Gaussian-MST with $\mathbf{p} = (3, p_2, \dots, p_k)$. Also, we have some particular models as the multiple gamma with $\mathbf{p} = (2, \dots, 2)$, multiple inverse Gaussian with $\mathbf{p} = (3, \dots, 3)$ and the gamma-Gaussian with $\mathbf{p} = (2, 0, \dots, 0)$ of Casalis (1996). In fact, the MST models are composed by a fixed univariate stable Tweedie variable having a positive mean domain and random variables that, given the fixed one, are real independent stable Tweedie variables, possibly different, with the same dispersion parameter equal to the fixed component.

Precisely and for short, within the framework of natural exponential families Kokonendji and Moypemna Sembona (2018) have completely characterized the MST models through their variance functions as follows. Let $\mathbf{p} = (p_1, \dots, p_k)$ with $p_1 \geq 1$ and $p_j \in \{0\} \cup [1, \infty)$ for $j = 2, \dots, k$. Then, the variance function of $F_{\mathbf{p}}$ is given by $\mathbf{V}_{F_{\mathbf{p}}}(\mathbf{m}) = \lambda^{1-p_1} m_1^{p_1-2} \mathbf{m} \mathbf{m}^\top + \text{diag}(0, m_1^{1-p_2} m_2^{p_2}, \dots, m_1^{1-p_k} m_k^{p_k})$ for all $\lambda > 0$ and $\mathbf{m} = (m_1, \dots, m_k)^\top \in \mathbf{M}_{F_{\mathbf{p}}} = (0, \infty) \times M_{F_{p_2}} \times \dots \times M_{F_{p_k}}$. Therefore, from (4.4) and (4.8), one has

$$\text{GVI}_{F_{\mathbf{p}}}(\mathbf{m}) = \lambda^{1-p_1} m_1^{p_1-2} + \frac{\sum_{j=2}^k m_1^{1-p_j} m_j^{p_j+2}}{(\mathbf{m}^\top \mathbf{m})^2} > 0$$

and

$$\text{MVI}_{F_{\mathbf{p}}}(\mathbf{m}) = \frac{\lambda^{1-p_1} m_1^{p_1-2} \sum_{j=1}^k m_j^4 + \sum_{j=2}^k m_1^{1-p_j} m_j^{p_j+2}}{\sum_{j=1}^k m_j^4 + \sum_{1 \leq j < \ell \leq k} m_j^2 m_\ell^2} > 0.$$

According to different classifications of Kokonendji and Moypemna Sembona (2018), several scenarios occur for k -variate over-, equi- and under-variation with respect to GVI and MVI. For instance, let $\lambda = 1$ and $p_1 = 2$ for the exponential-MST subclass then one deduces $\text{GVI}_{F_{\mathbf{p}}}(\mathbf{m}) > 1$ for all $p_j > 1$ with $j = 2, \dots, k$. In order to investigate both indexes GVI and MVI for k -variate (semi-)continuous models on $[0, \infty)^k$, we finally exclude cases $p_1 = 1$ for the Poisson-MST and also for all $p_j = 0$ and $p_j = 1$, $j = 2, \dots, k$, related to the normal and Poisson components, respectively. Hence, the NST class is removed from this study.

4.4 Estimation and asymptotic properties

Let Y_1, \dots, Y_n be a random sample from Y with support on $(0, \infty)^k$, where for each $i \in \{1, \dots, n\}$, $Y_i = (Y_{i1}, \dots, Y_{ik})^\top$. It is common to consider the statistically consistent empirical versions

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i = (\bar{Y}_1, \dots, \bar{Y}_k)^\top \quad \text{and} \quad \widehat{\text{cov}Y} = \frac{1}{n-1} \sum_{i=1}^n Y_i Y_i^\top - \bar{Y}_n \bar{Y}_n^\top \quad (4.11)$$

of the mean vector and covariance matrix of Y , respectively. An estimator of $\text{GVI}(Y)$ directly derived from (4.11) is determined by

$$\widehat{\text{GVI}}_n(Y) = \frac{\bar{Y}_n^\top \widehat{\text{cov}Y} \bar{Y}_n}{(\bar{Y}_n^\top \bar{Y}_n)^2}. \quad (4.12)$$

Since all the univariate positive continuous variables take positive values, we deduce from Cramér (1974, pp. 357-358) that $\widehat{\text{GVI}}_n(Y)$ is an asymptotically unbiased estimator, i.e., $\mathbb{E}\{\widehat{\text{GVI}}_n(Y)\} \approx \text{GVI}(Y)$. Concerning the theoretical variance of $\widehat{\text{GVI}}_n$, we would need at least the moments of fourth order of the components of Y .

More interestingly, we establish the following central limit and strong consistency results of $\widehat{\text{GVI}}_n$ and $\widehat{\text{MVI}}_n$. The proofs are given in Appendix C.

Proposition 4.4.1. *Let $Y = (Y_1, \dots, Y_k)^\top$ be a positive continuous k -variate random vector on $(0, \infty)^k$, $k \geq 1$, such that $\mathbb{E}(Y_{\ell_1} Y_{\ell_2} Y_{\ell_3} Y_{\ell_4}) < \infty$. Let also Y_1, \dots, Y_n be a random sample from Y .*

(i) *As $n \rightarrow \infty$,*

$$\sqrt{n} \{\widehat{\text{GVI}}_n(Y) - \text{GVI}(Y)\} \rightsquigarrow \mathcal{N}(0, \sigma_{gvi}^2),$$

where \rightsquigarrow stands for convergence in distribution and $\mathcal{N}(0, \sigma_{gvi}^2)$ is the centered normal distribution with variance $\sigma_{gvi}^2 = \Delta^\top \Gamma \Delta$. The $\{k + k(k+1)/2\} \times 1$ vector $\Delta = (\dots, \Delta_j, \dots; \dots, \Delta_{j\ell}, \dots)_{j \in \{1, \dots, k\}; \ell \in \{j, \dots, k\}}^\top$ and the $\{k + k(k+1)/2\} \times \{k + k(k+1)/2\}$ four-block symmetric matrix

$$\Gamma = \begin{bmatrix} \Sigma & \Gamma_3^\top \\ \Gamma_3 & \Gamma_4 \end{bmatrix}$$

are such that, for all $j, j', j'' \in \{1, \dots, k\}$,

$$\Delta_j = \left\{ 2 \sum_{j'=1}^k \mathbb{E} Y_{j'} \text{cov}(Y_j, Y_{j'}) - 4 \mathbb{E} Y_j \left(\sum_{j'=1}^k (\mathbb{E} Y_{j'})^2 \right) \text{GVI}(Y) \right\} / (\mathbb{E} Y^\top \mathbb{E} Y)^2,$$

$\Delta_{jj} = (\mathbb{E}Y_j)^2 / (\mathbb{E}Y^\top \mathbb{E}Y)^2$, $\Delta_{j\ell} = 2\mathbb{E}Y_j \mathbb{E}Y_\ell / (\mathbb{E}Y^\top \mathbb{E}Y)^2$ for $\ell \in \{j+1, \dots, k\}$, $\Sigma(j; j') = \text{cov}(Y_j, Y_{j'})$, $\Gamma_3^\top(j; j', \ell') = \text{cov}(Y_j, Y_{j'} Y_{\ell'})$ for $\ell' \in \{j', \dots, k\}$ and $\Gamma_4(j', \ell'; j'', \ell'') = \text{cov}(Y_{j'} Y_{\ell'}, Y_{j''} Y_{\ell''})$ for $\ell'' \in \{j'', \dots, k\}$;

(ii) As $n \rightarrow \infty$,

$$\sqrt{n} \{\widehat{\text{MVI}}_n(\mathbf{Y}) - \text{MVI}(\mathbf{Y})\} \rightsquigarrow \mathcal{N}(0, \sigma_{mvi}^2),$$

with $\sigma_{mvi}^2 = \mathbf{\Lambda}^\top \mathbf{\Pi} \mathbf{\Lambda}$. The $2k \times 1$ vector $\mathbf{\Lambda} = (\Lambda_1, \dots, \Lambda_k; \Lambda_{11}, \dots, \Lambda_{kk})^\top$ and the $2k \times 2k$ four-block symmetric matrix

$$\mathbf{\Pi} = \begin{bmatrix} \Sigma & \Pi_3^\top \\ \Pi_3 & \Pi_4 \end{bmatrix}$$

are such that, for $j, j' \in \{1, \dots, k\}$,

$$\Lambda_j = \left\{ 2\mathbb{E}Y_j \text{var}Y_j - 4\mathbb{E}Y_j \left(\sum_{j'=1}^k (\mathbb{E}Y_{j'})^2 \right) \text{MVI}(\mathbf{Y}) \right\} / (\mathbb{E}Y^\top \mathbb{E}Y)^2,$$

$\Lambda_{jj} = (\mathbb{E}Y_j)^2 / (\mathbb{E}Y^\top \mathbb{E}Y)^2$, $\Sigma(j; j') = \text{cov}(Y_j, Y_{j'})$, $\Pi_3^\top(j; j') = \text{cov}(Y_j, Y_{j'}^2)$ and $\Pi_4(j; j') = \text{cov}(Y_j^2, Y_{j'}^2)$.

Note that Parts (i) and (ii) of Proposition 4.4.1 provide the same result for $k = 1$ with

$$\begin{aligned} \sigma_{gvi}^2 = \sigma_{mvi}^2 &= \frac{1}{(\mathbb{E}Y)^8} [3(\mathbb{E}Y)^8 + (\mathbb{E}Y)^2 \text{var}Y \{10(\mathbb{E}Y)^4 - 4\mathbb{E}Y \mathbb{E}Y^3 + 4(\text{var}Y)^2\} \\ &\quad - 4(\mathbb{E}Y)^5 \mathbb{E}Y^3 + (\mathbb{E}Y)^4 \{11(\text{var}Y)^2 + \mathbb{E}Y^4\}]; \end{aligned}$$

see Touré et al. (2021). Therefore we find the same asymptotic variance for the exponential distribution as Theorem 2.3.1 of the univariate case for $k = 1$. Also, an asymptotic confidence interval for $\text{GVI}(\mathbf{Y})$ is expressed as

$$CI_a(\text{GVI}; 1 - \alpha) := [\widehat{\text{GVI}}_n - u_{1-\alpha/2} \widehat{\sigma}_{gvi} / \sqrt{n}, \widehat{\text{GVI}}_n + u_{1-\alpha/2} \widehat{\sigma}_{gvi} / \sqrt{n}], \quad (4.13)$$

where $\alpha \in [0, 1]$ is fixed, u_p is the p th percentile of the standard normal distribution $\mathcal{N}(0, 1)$ and $\widehat{\sigma}_{gvi}^2 = \widehat{\Delta}_n^\top \widehat{\Gamma}_n \widehat{\Delta}_n$ is the corresponding empirical version of σ_{gvi}^2 (Part (i) of Proposition 4.4.1). A similar result also holds for the intuitive index MVI. Finally, we state the following results for strong consistency.

Proposition 4.4.2. *Let $\mathbf{Y} = (Y_1, \dots, Y_k)^\top$ be a positive continuous k -variate random vector on $(0, \infty)^k$, $k \geq 1$, such that $\mathbb{E}(Y_{\ell_1} Y_{\ell_2}) < \infty$. If $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ be a random sample from \mathbf{Y} , then*

$$\widehat{\text{GVI}}_n(\mathbf{Y}) \xrightarrow{a.s.} \text{GVI}(\mathbf{Y}) \text{ and } \widehat{\text{MVI}}_n(\mathbf{Y}) \xrightarrow{a.s.} \text{MVI}(\mathbf{Y}), \text{ as } n \rightarrow \infty,$$

where $\xrightarrow{a.s.}$ stands for almost sure convergence.

In finite samples, we shall suggest the use of a bootstrapped approach (e.g., Efron and Tibshirani, 1993) for reducing the estimated variances and their corresponding lengths of confidence intervals; see the end of the next section.

4.5 Numerical applications

All computations have been done with the Python (Python Software Foundation, 2019) and R software (R Core Team, 2018). To generate a k -variate continuous positive orthant distribution given k (over-, equi- and under-varied) marginals and correlation matrix ρ , we have used the NORmal To Anything (NORTA) method (e.g., Su, 2015).

In practical way, we shall consider the exponential, lognormal and Weibull distributions (e.g., Dey and Kundu, 2009) as marginals. They have been used quite effectively in analyzing positively skewed data, which play important roles in the reliability analysis. Recall here that the univariate exponential distribution $\mathcal{E}_1(\mu)$ is always equi-varied for all $\mu > 0$ and, both univariate lognormal $\mathcal{LN}_1(m, \sigma^2)$ and Weibull $\mathcal{W}_1(\alpha, \beta)$ models are over- (equi- and under-) varied for $0 < \sigma^2 \leq \log 2$ with $m \in \mathbb{R}$ and, from (4.10) for $0 < \beta \leq 1$ with $\alpha > 0$, respectively. All these theoretical behaviors work well on simulated datasets of univariate exponential, lognormal and Weibull distributions that we omit presenting here.

4.5.1 Some scenarios of bivariate cases and a real 4-variate dataset

Firstly, we consider Table 4.2 consisting of fifteen simulated bivariate datasets from exponential, lognormal and Weibull distributions, presenting several scenarios of correlation (positive or negative) and marginal over-, equi- or under-variation. Table 4.2

unveils a summary of these datasets, along with the sample values of the indexes GVI and MVI.

TABLE 4.2 – Some scenarios from fifteen simulated bivariate datasets with $\widehat{VI}_j = \widehat{\sigma}_j^2 / \bar{y}_j^2$ and marginal variations (MV) : Over- (O), Equi- (E) and Under-variation (U).

Dataset	n	\bar{y}_1	\bar{y}_2	$\widehat{\sigma}_1^2$	$\widehat{\sigma}_2^2$	\widehat{VI}_1	\widehat{VI}_2	$\widehat{\rho}_{12}$	MV	\widehat{MVI}	\widehat{GVI}
No 1	50	2.58	3.80	13.36	31.08	2.02	2.15	-0.23	O/O	2.11	1.00
No 2	80	1.34	1.31	3.98	4.05	2.21	2.36	-0.33	O/O	2.14	0.76
No 3	100	1.94	1.96	3.74	3.76	0.99	0.98	0.20	E/E	0.99	1.98
No 4	120	5.76	5.04	34.97	29.47	1.02	1.01	-0.40	E/E	1.00	0.34
No 5	150	0.27	0.26	0.02	0.04	0.24	0.51	0.56	U/U	0.39	2.28
No 6	300	8.77	8.55	22.87	20.82	0.30	0.28	0.03	U/U	0.25	0.15
No 7	500	5.60	2.10	299.03	20.32	9.53	4.63	-0.06	O/O	7.40	7.32
No 8	800	10.59	10.39	110.44	111.03	0.98	1.02	0.05	E/E	0.98	1.01
No 9	1000	2.66	1.76	2.35	0.41	0.33	0.13	0.78	U/U	0.17	1.02
No 10	1000	6.10	1.74	323.00	0.52	8.68	0.17	0.49	O/U	7.42	7.50
No 11	1500	1.77	3.41	0.44	17.88	0.14	1.55	-0.57	U/O	0.96	0.87
No 12	3000	0.75	0.62	1.62	0.11	2.86	0.28	-0.29	O/U	1.06	0.99
No 13	3000	1.00	1.09	1.00	3.95	1.00	3.33	-0.29	E/O	1.19	0.82
No 14	5000	0.68	1.00	0.42	1.02	0.89	1.02	0.80	U/E	0.98	2.23
No 15	8000	1.98	3.37	3.83	17.97	0.98	1.58	-0.38	E/O	1.33	0.99

In order to measure the departure from the bivariate uncorrelated exponential of the considered datasets, our estimated index \widehat{GVI} provides a very good summary of the bivariate variation by taking into account both marginal variation and the non-null correlation value $\widehat{\rho}_{12}$. As a matter of fact, the bivariate equi-variation ($\widehat{GVI} = 1$) is significantly obtained here for both over-varied marginals with negative correlation (No 1), for both under-varied marginals with large positive correlation (No 9), for both equi-varied marginals with very weak correlation (No 8), and either for one marginal over-varied and the other under-varied (No 12) or equi-varied (No 15) with negative correlations. The bivariate over-variation ($\widehat{GVI} > 1$) is pointed out for both over-varied marginals with weak negative correlation (No 7), for both equi-varied marginals with positive correlation (No 3), for both under-varied marginals with positive correlation (No 5), and either for one marginal under-varied and the other over-varied (No 10) or equi-varied (No 14) with positive correlation. Concerning the bivariate under-variation

($\widehat{GVI} \in (0, 1)$), this is highlighted for both over-varied marginals with negative correlation (No 2), for both equi-varied marginals with negative correlation (No 4), for both under-varied marginals with weak positive correlation (No 6), and either for one marginal over-varied and the other under-varied (No 11) or equi-varied (No 13) with negative correlation. In the common sense, we always have the bivariate over-/under-variation for both over-/under-varied marginals with positive/negative correlation. The values of \widehat{GVI} provide the corresponding degree of (over-/under-) variation with respect to the reference value 1 of the bivariate equi-variation. For instance, we detect a higher degree of over-variation in No 7 and No 10 than in No 3, No 5 or No 14; similarly, we discern a weaker degree of bivariate under-variation (close to 1) in No 11 and No 13 than in Nos. 4 or 6.

In the same way, the marginal index \widehat{MVI} also works very well, summarizing both marginal variations (without correlation). The two indexes \widehat{MVI} and \widehat{GVI} are close when the correlation is quasi-null (Nos. 7 or 8). For the sake of brevity, we omit here an analysis of the standard errors of the estimated indexes of these datasets; a complete analysis shall be done in the next section for 6-variate datasets.

In summary, multivariate variation indexes MVI and GVI are meaningful because they summarize the variation behavior from each individual variable. In addition, GVI also contains information about their correlation. They can be used for descriptive analysis, for clustering, for comparing different datasets and for testing departures from known multivariate distributions like Touré et al. (2020) are done for univariate cases.

Secondly, we explore the real 4-variate dataset which refers to the annual observations from 1900 to 1989 of the United States. It is reported by Hayashi (2000) : the first variable (m) is the natural log of the money M1, the second (p) is the natural log of the net national product price deflator, the third (y) is the natural log of the net national product, and the fourth (r) is the commercial paper rate in percent at an annual rate.

To evaluate the departure from the 4-variate uncorrelated exponential distribution

of the considered dataset, our estimated indexes still provide very good summaries through $\widehat{\text{GVI}} = 0.1397$ and $\widehat{\text{MVI}} = 0.0771$. Indeed, both indexes strongly show a 4-variate under-variation with $0 < \widehat{\text{MVI}} < \widehat{\text{GVI}} < 1$. Since $\widehat{\text{MVI}}$ is very close to 0 than $\widehat{\text{GVI}}$, each of the four marginal distributions must be univariate under-varied with the correlation matrix having only positive coefficients. Table 4.3 corroborates this analysis only from results of MVI and GVI. At this stage, one can select an appropriate theoretical 4-variate distribution for modelling this dataset and their (interest) parameters adjusted directly by estimation.

TABLE 4.3 – Summary of real 4-variate data with $\widehat{\text{VI}}_j = \widehat{\sigma}_j^2 / \bar{y}_j^2$ and the marginal variation (MV_j): Under-variation (U), size $n = 90$, $\det \widehat{\rho} = 0.0003$, $\widehat{\text{GVI}} = 0.1397$ and $\widehat{\text{MVI}} = 0.0771$.

j	\bar{y}_j	$\widehat{\sigma}_j^2$	$\widehat{\text{VI}}_j$ (MV_j)	$\widehat{\rho}_{j1}$	$\widehat{\rho}_{j2}$	$\widehat{\rho}_{j3}$	$\widehat{\rho}_{j4}$
1	4.1476	1.9630	0.1141 (U)	1.0000	0.9579	0.9905	0.3926
2	3.1709	0.6049	0.0602 (U)	0.9579	1.0000	0.9552	0.6002
3	2.2610	0.6330	0.1238 (U)	0.9905	0.9552	1.0000	0.4331
4	4.5547	8.4074	0.4053 (U)	0.3926	0.6002	0.4331	1.0000

4.5.2 Other multivariate cases and simulation studies

In this section, we first study a 6-variate simulated dataset. We then analyze the behavior of the asymptotic variances and confidence intervals by simulation. Finally, we compare the asymptotic standard errors of GVI and MVI to those obtained from a bootstrap method.

The 6-variate dataset of size $n = 560$ is simulated following the following scenario. We have considered two over-, two equi- and two under-variations as univariate marginals with the theoretical correlation matrix such that

$$\det \rho := \det \begin{pmatrix} 1 & -0.03 & 0.57 & 0.12 & 0.24 & -0.57 \\ -0.03 & 1 & -0.05 & -0.09 & 0.36 & 0.04 \\ 0.57 & -0.05 & 1 & 0.35 & 0.27 & -0.58 \\ 0.12 & -0.09 & 0.35 & 1 & 0.18 & 0.04 \\ 0.24 & 0.36 & 0.27 & 0.18 & 1 & 0.06 \\ -0.57 & 0.04 & -0.58 & 0.04 & 0.06 & 1 \end{pmatrix} = 0.2051.$$

The summary needed to compute the variation indexes $\widehat{\text{GVI}}$ and $\widehat{\text{MVI}}$ is reported in Table 4.4. As commented before for Table 4.2, we also observe a different behavior of the two variation indexes in this 6-variate example. We here obtain $\widehat{\text{GVI}} = 1.0572 \approx 1$ and $\widehat{\text{MVI}} = 0.9637 \approx 1$, both indicating a 6-variate phenomenon of quasi-equi-variation.

TABLE 4.4 – Summary of simulated 6-variate data with marginal variations (MV) : Over- (O), Equi- (E) and Under-variation (U), size $n = 560$ and $\det \widehat{\rho} = 0.2063$ such that $\widehat{\text{GVI}} = 1.0572 \approx 1$ and $\widehat{\text{MVI}} = 0.9637 \approx 1$.

j	\bar{y}_j	$\widehat{\sigma}_j^2$	MV _{j}	$\widehat{\rho}_{j1}$	$\widehat{\rho}_{j2}$	$\widehat{\rho}_{j3}$	$\widehat{\rho}_{j4}$	$\widehat{\rho}_{j5}$	$\widehat{\rho}_{j6}$
1	1.2245	3.2031	O	1.0000	-0.0197	0.5572	0.1074	0.2939	-0.5586
2	0.4929	0.2324	E	-0.0197	1.0000	-0.0683	-0.1078	0.3293	0.0102
3	1.1548	0.4834	U	0.5572	-0.0683	1.0000	0.3136	0.3116	-0.5946
4	0.9507	0.9236	E	0.1074	-0.1078	0.3136	1.0000	0.1451	0.0264
5	4.3871	28.6346	O	0.2939	0.3293	0.3116	0.1451	1.0000	0.0310
6	0.9093	0.1039	U	-0.5586	0.0102	-0.5946	0.0264	0.0310	1.0000

TABLE 4.5 – Asymptotic variances and confidence intervals ($u = u_{0.975} = 1.96$) from subsamples of simulated 6-variate data with $n = 10\,000$ having the same parameters as for Table 4.4.

n	$\det \widehat{\rho}$	$\widehat{\sigma}_{gvi}^2$	$\widehat{\sigma}_{mvi}^2$	$\widehat{\text{GVI}}_n \pm u \widehat{\sigma}_{gvi} / \sqrt{n}$	$\widehat{\text{MVI}}_n \pm u \widehat{\sigma}_{mvi} / \sqrt{n}$
50	0.2030	27.1916	13.2311	1.0933 ± 1.3712	1.0322 ± 0.9234
100	0.1571	37.1112	14.1102	1.1511 ± 1.2111	1.0133 ± 0.7748
300	0.2092	34.1805	11.2333	1.0442 ± 1.0743	1.0211 ± 0.3233
500	0.2050	30.0203	10.3422	1.0487 ± 0.9385	1.0713 ± 0.1876
1 000	0.1958	30.2236	10.8976	1.0197 ± 0.4863	1.0342 ± 7.4500
3 000	0.2017	29.7601	9.1233	1.0622 ± 0.2986	1.0211 ± 0.1074
5 000	0.2067	29.0823	9.0303	1.0123 ± 0.1099	1.1111 ± 0.0738
10 000	0.2067	29.0876	9.1111	1.0511 ± 0.0971	1.0455 ± 0.0322

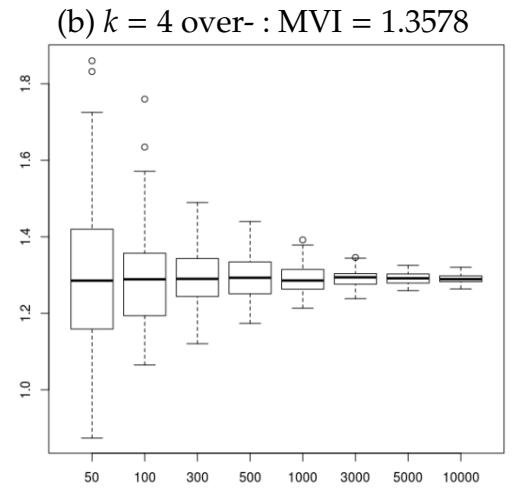
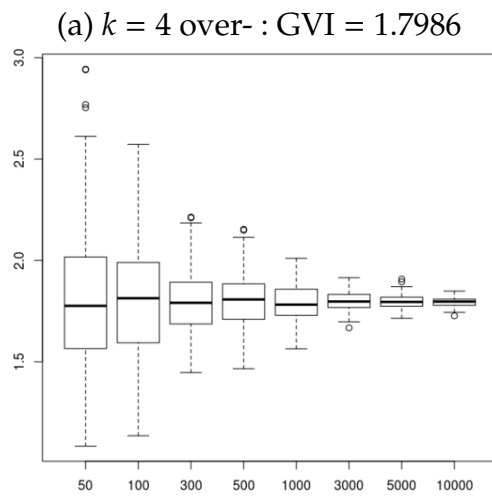
Table 4.5 depicts an evolution of the asymptotic variances and confidence intervals of GVI and MVI from subsamples of a simulated 6-variate dataset with a maximum size $n = 10\,000$, having the same parameters as those of Table 4.4. We observe that both estimated standard errors $u \widehat{\sigma} / \sqrt{n}$ decrease when sample size n increases ; this fact looks

TABLE 4.6 – Asymptotic variances and confidence intervals ($u = u_{0.975} = 1.96$) from subsamples of simulated over-varied 4-variate data with $n = 10\,000$.

n	$\det \widehat{\rho}$	$\widehat{\sigma}_{gvi}^2$	$\widehat{\sigma}_{mvi}^2$	$\widehat{GVI}_n \pm u \widehat{\sigma}_{gvi} / \sqrt{n}$	$\widehat{MVI}_n \pm u \widehat{\sigma}_{mvi} / \sqrt{n}$
50	0.3209	154.1411	21.3344	3.1123 ± 2.3107	1.3311 ± 1.3822
100	0.3452	145.6324	19.1212	3.0103 ± 1.6798	1.2721 ± 1.1811
300	0.6915	138.8999	18.7866	2.2155 ± 1.2132	1.2722 ± 0.8465
500	0.7071	132.4122	17.3333	2.0215 ± 0.9278	1.2530 ± 0.5440
1 000	0.6490	123.2211	16.4411	1.7019 ± 0.7555	1.2633 ± 0.3097
3 000	0.6582	118.1155	14.8965	1.4899 ± 0.4001	1.3811 ± 0.1233
5 000	0.5998	115.0967	13.0988	1.6667 ± 0.2064	1.3111 ± 0.0711
10 000	0.6069	110.2222	12.0015	1.9244 ± 0.1234	1.2344 ± 0.0322

TABLE 4.7 – Asymptotic variances and confidence intervals ($u = u_{0.975} = 1.96$) from subsamples of simulated under-varied trivariate data with $n = 10\,000$.

n	$\det \widehat{\rho}$	$\widehat{\sigma}_{gvi}^2$	$\widehat{\sigma}_{mvi}^2$	$\widehat{GVI}_n \pm u \widehat{\sigma}_{gvi} / \sqrt{n}$	$\widehat{MVI}_n \pm u \widehat{\sigma}_{mvi} / \sqrt{n}$
50	0.9174	17.3211	12.6643	0.8213 ± 0.6123	0.6534 ± 0.4922
100	0.9634	17.2233	12.7755	0.8345 ± 0.4305	0.6755 ± 0.2674
300	0.9551	16.5678	11.4455	0.8743 ± 0.2617	0.7867 ± 0.1944
500	0.9446	16.0145	10.9066	0.8123 ± 0.2012	0.7133 ± 0.1169
1 000	0.9281	17.0123	10.0342	0.8368 ± 0.1388	0.7411 ± 0.0823
3 000	0.9262	15.3344	11.7711	0.7614 ± 0.0922	0.6977 ± 0.0411
5 000	0.9221	14.8877	10.1188	0.8544 ± 0.0514	0.7123 ± 0.0233
10 000	0.9195	14.9988	9.6511	0.7970 ± 0.0317	0.6955 ± 0.0124



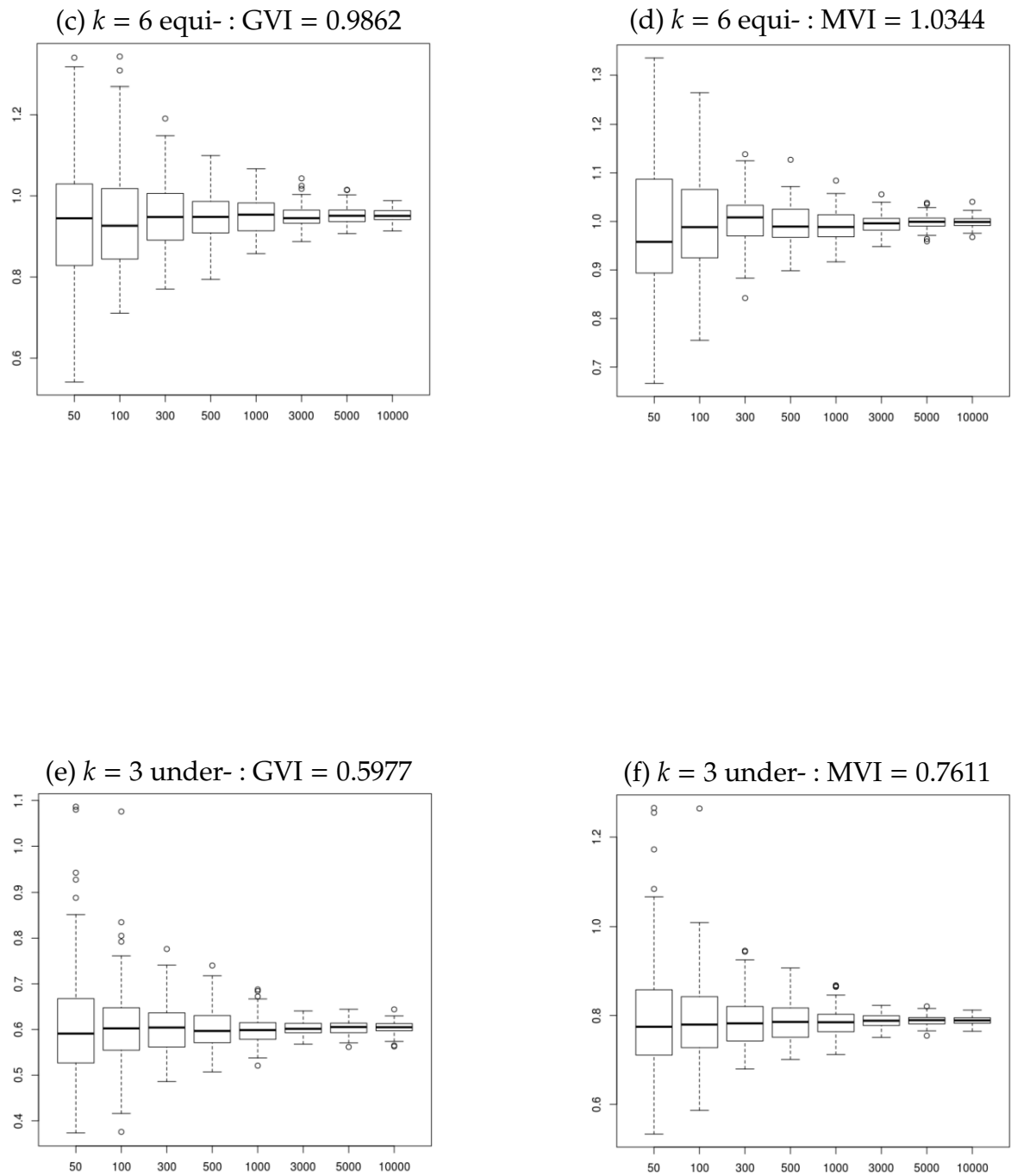


FIGURE 4.1 – Boxplots for the targets GVI and MVI with 100 replicates according to sample size and simulation parameters of Tables 4.5, 4.6 and 4.7.

like to the context of 6-variate quasi-equi-variation for GVI and also for MVI. The stable behavior of variances agrees with Proposition 4.4.1.

Similar studies have also been performed by simulating a 4-variate over-varied distribution (Table 4.6) and a trivariate under-varied distribution (Table 4.7). More precisely, Table 4.6 have been obtained by simulating four marginal (over-, equi-, and under-varied) Weibull distributions, with the cross correlation matrix such that

$$\det \rho := \det \begin{pmatrix} 1 & -0.03 & 0.27 & 0.35 \\ -0.03 & 1 & -0.05 & 0.04 \\ 0.27 & -0.05 & 1 & 0.48 \\ 0.35 & 0.04 & 0.48 & 1 \end{pmatrix} = 0.6588.$$

As for Table 4.7, we have simulated one marginal Weibull distribution, one marginal exponential distribution and one marginal lognormal distribution, with the correlation matrix such that

$$\det \rho := \det \begin{pmatrix} 1 & -0.03 & -0.27 \\ -0.03 & 1 & -0.05 \\ -0.27 & -0.05 & 1 \end{pmatrix} = 0.9229.$$

Also, we notice that all estimated standard errors $u\widehat{\sigma}/\sqrt{n}$ decrease when sample size n increases, but more slowly for GVI than for MVI in Table 4.6 of the 4-variate phenomenon of over-variation. Figure 4.1 clearly presents typical behaviors of boxplots related to Tables 4.5, 4.6 and 4.7. However, the estimated variances in Table 4.6 of the 4-variate over-variation are much larger than those in Table 4.7 of the 3-variate under-variation.

TABLE 4.8 – Asymptotic and asymmetric bootstrap confidence intervals for GVI and MVI indexes from sample size n simulated with $u = u_{0.975} = 1.96$ and $\alpha = 0.05$ using the dataset of Table 4.4.

n	$\widehat{\text{GVI}}_n \pm u\widehat{\sigma}_{gvi}/\sqrt{n}$	$\widehat{\text{GVI}}_b(-\widehat{SE}_1; +\widehat{SE}_2)$	$\widehat{\text{MVI}}_n \pm u\widehat{\sigma}_{mvi}/\sqrt{n}$	$\widehat{\text{MVI}}_b(-\widehat{SE}_1; +\widehat{SE}_2)$
30	1.2166 \pm 1.3344	0.9816(−0.1722;+0.1868)	0.9798 \pm 37.9062	0.9723(−0.1506;+0.1602)
50	1.0322 \pm 1.1200	0.9699(−0.1140;+0.1180)	1.0407 \pm 36.0743	0.9776(−0.0659;+0.0467)
100	1.0255 \pm 1.0876	0.9628(−0.0342;+0.0334)	0.9950 \pm 26.0580	0.9654(−0.0434;+0.0162)
300	0.9944 \pm 0.9466	1.0114(−0.0260;+0.0257)	1.0703 \pm 23.4215	0.9896(−0.0275;+0.0130)
500	1.0699 \pm 0.7935	1.0259(−0.0251;+0.0244)	1.1648 \pm 15.5693	0.9827(−0.0101;+0.0009)
1000	1.2345 \pm 0.6392	1.0310(−0.0044;+0.0144)	1.1952 \pm 08.9738	0.9864(−0.0010;+0.0002)

Table 4.8 exhibits behaviors of both asymptotic and asymmetric bootstrap confidence intervals for GVI and MVI in the situations of small and moderate sample sizes (e.g., Angelo and Brian, 2019). As a matter of fact, we have used the empirical bootstrap method whose works well even for asymmetric distribution (e.g., Efron and Tibshirani, 1993, Chapter 13). We have done 10 000 replicates for given sample size n with significance level $\alpha = 0.05$ fixed. The asymmetric bootstrap confidence interval for GVI is expressed as

$$CI_b(\text{GVI}; 1 - \alpha) := [\widehat{\text{GVI}}_b - \widehat{\delta}_b(1 - \alpha/2) se, \widehat{\text{GVI}}_b + \widehat{\delta}_b(\alpha/2) se] = [\widehat{\text{GVI}}_b - \widehat{SE}_1; \widehat{\text{GVI}}_b + \widehat{SE}_2], \quad (4.14)$$

where $\widehat{\text{GVI}}_b$ is the bootstrap estimate of GVI with standard error (se) and, $\widehat{\delta}_b(1 - \alpha/2)$ and $\widehat{\delta}_b(\alpha/2)$ are the estimated asymmetric bootstrap quantiles of the distribution $\delta = \widehat{\text{GVI}}_b - \text{GVI}$. A similar asymmetric bootstrap confidence interval holds for the intuitive index MVI. As shown Table 4.8, asymptotic and bootstrap pointwise estimates for GVI and MVI are very close for all sample size n . For these 6-variate equi-varied datasets, asymptotic estimated standard error $u\widehat{\sigma}_{gvi}/\sqrt{n}$ of GVI from (4.13) and bootstrap estimated standard error $\widehat{\delta}_b se$ of GVI from (4.14) decrease when sample size n increases. But this decrease is faster with bootstrap method. Similar results hold for the MVI index. Accordingly, in order to reduce the estimated variances of our indexes for small and moderate sample sizes, we attempt to recommend a bootstrapped approach such as $\widehat{\sigma}_{b,gvi}^2 = n(se)^2$ from (4.14).

4.5.3 Numerical application of variation and dispersion indexes in univariate, bivariate and multivariate setups

In this part, we calculate the GDI and GVI indexes in the univariate and multivariate case on real data. For that, we consider the following dataset of Table 4.9. It concerns three measurements (with $n = 42$) of drinking water pumps installed in the Sahel. See, e.g., Kokonendji and Somé (2021) who used these data in semi-parametric modeling.

The first variable X_1 represents the *failure times* (in months) and it is recently used by Touré et al. (2020, 2021). The second variable X_2 refers to the *distance* (in kilometers) between each water pump and the repair center in the Sahel, while the third one X_3 stands for *average volume* (in m^3) of water per day.

TABLE 4.9 – Drinking water pumps trivariate data measured in the Sahel with $n = 42$.

X_1 :	23	261	87	10	120	14	62	15	47	225	71	20	246	21
X_2 :	97	93	94	100	98	84	96	110	121	73	90	93	103	116
X_3 :	26	52	22	39	23	26	32	17	10	39	31	42	52	26
X_1 :	19	42	20	5	12	120	17	11	3	14	71	11	5	14
X_2 :	114	82	96	94	77	91	117	103	99	113	79	109	84	118
X_3 :	26	36	43	36	6	27	15	36	9	52	11	20	25	37
X_1 :	11	16	90	1	16	52	95	10	1	14	4	7	14	20
X_2 :	98	93	94	103	109	110	89	108	101	93	102	138	103	96
X_3 :	25	18	43	43	24	38	6	40	21	34	15	23	68	37

Table 4.10 displays all empirical univariate, bivariate and trivariate variation (4.2) and dispersion (see, Kokonendji and Puig, 2018) indexes from Table 4.9. Hence, each X_j , (X_j, X_k) and (X_1, X_2, X_3) is over-dispersed compared to the corresponding uncorrelated Poisson distribution. But, only (X_1, X_3) (resp. X_1) can be considered as a bivariate equi-variation (resp. univariate over-variation) with respect to the corresponding uncorrelated exponential distribution; and, other X_j , (X_j, X_k) and (X_1, X_2, X_3) are under-varied. It is noteworthy that, we just compute dispersion indexes for curiosity since all values in Table 4.9 are positive integers; and, we herenow omit the counting point of view in the remainder of the analysis.

TABLE 4.10 – Empirical univariate (in diagonal), bivariate (off diagonal) and trivariate (at the corner) variation and dispersion indexes.

$\widehat{\text{GVI}}_3 = 0.0533$	X_1	X_2	X_3	$\widehat{\text{GDI}}_3 = 15.1230$	X_1	X_2	X_3
X_1	1.9425	0.0558	1.0549	X_1	89.5860	14.3223	70.7096
X_2	0.0558	0.0167	0.0157	X_2	14.3223	1.6623	2.0884
X_3	1.0549	0.0157	0.2122	X_3	70.7096	2.0884	6.3192

4.6 Unification of GVI and GDI

Concerning a generalization of the basal GVI of (4.2) which is also viewed as a particular RVI with respect to the uncorrelated exponential model, the recent univariate unification of dispersion and variation indexes by Touré et al. (2020, 2021) is used in the multivariate framework of natural exponential families as follows. Let X and Y be two random vectors on the same support $\mathcal{S} \subseteq \mathbb{R}^k$ and assume $\mathbf{m} := \mathbb{E}X = \mathbb{E}Y$, $\Sigma_Y := \text{cov}Y$ and $\mathbf{V}_{F_X}(\mathbf{m}) := \text{cov}(X)$ fixed, then the *relative variability index* of Y with respect to X can be defined as

$$\text{RWI}_X(Y) := \text{tr}[\Sigma_Y \mathbf{W}_{F_X}^+(\mathbf{m})] \gtrless 1,$$

where $\mathbf{W}_{F_X}^+(\mathbf{m})$ is the unique Moore-Penrose inverse of the associated matrix $\mathbf{W}_{F_X}(\mathbf{m}) := [\mathbf{V}_{F_X}(\mathbf{m})]^{1/2} [\mathbf{V}_{F_X}(\mathbf{m})]^\top$ to $\mathbf{V}_{F_X}(\mathbf{m})$; see Section 4.8 (part B) for GVI. Thus, we unify the construction of GDI and GVI by choosing $\mathbf{W}_{F_X}(\mathbf{m}) = \sqrt{\mathbf{m}} \sqrt{\mathbf{m}}^\top$ and $\mathbf{W}_{F_X}(\mathbf{m}) = \mathbf{m} \mathbf{m}^\top$, respectively. It is noteworthy that one can consider $\mathbf{V}_{F_X}(\mathbf{m})$ as a particular case of the MST variance function of Section 4.3.4; but, it will be equivalent to the proposed GVI via RVI for supports $\mathcal{S} = [0, \infty)^k$ of distributions.

4.7 Conclusion

In this chapter, we have defined the index GVI for multivariate positive continuous data. The index MVI is used regardless of the correlation structure. We also have

illustrated these indexes on bi and multivariate distributions. Simulation studies were also carried out on these GVI and MVI indices. Finally, we made the unification of GVI with GDI of Kokonendji and Puig (2018).

4.8 Complement and proofs

A On a broader multivariate exponential distribution

According to Cuenin et al. (2016), taking $p = 2$ in their multivariate Tweedie (1984) models of flexible dependence structure, another way to define a k -variate exponential distribution is given by $\mathcal{E}_k(\Lambda)$. The $k \times k$ symmetric variation matrix $\Lambda = (\lambda_{ij})_{i,j \in \{1, \dots, k\}}$ is such that $\lambda_{ij} = \lambda_{ji} \geq 0$, the mean of the marginal exponential is $\lambda_{ii} > 0$, and the non-negative correlation terms satisfy

$$\rho_{ij} = \frac{\lambda_{ij}}{\sqrt{\lambda_{ii}\lambda_{jj}}} \in [0, \min\{R(i, j), R(j, i)\}], \quad (4.15)$$

with $R(i, j) = \sqrt{\lambda_{ii}/\lambda_{jj}}(1 - \lambda_{ii}^{-1} \sum_{\ell \neq i, j} \lambda_{i\ell}) \in (0, 1)$. The construction of Cuenin et al. (2016) is perfectly defined having $k(k+1)/2$ parameters as in $\mathcal{E}_k(\mu, \rho)$. Furthermore, we attain the exact bounds of the correlation terms in (4.15). The main fact is that Cuenin et al. (2016) pointed out the construction and simulation of the negative correlation structure from the positive one of (4.15) by using the inversion method.

The negativity of a correlation component is important for the rare phenomenon of under-variation in a bivariate/multivariate positive continuous model. Figure 4.2 (right) plots a limit shape of any bivariate positive continuous distribution with very strong negative correlation (in red), which is not the diagonal line of the upper bound (+1) of positive correlation (in blue); see, e.g., Cuenin et al. (2016) for bivariate count model. Contrarily, Figure 4.2 (left) represents the classic lower (−1) and upper (+1) bounds of correlations on \mathbb{R}^2 or finite support.

B Construction of GVI

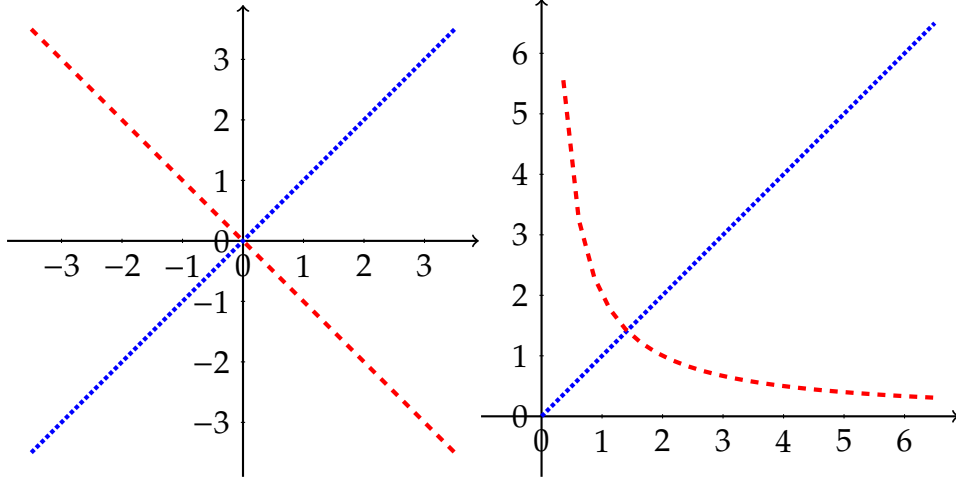


FIGURE 4.2 – Support of bivariate distributions with maximum correlations (positive in blue and negative in red) : model on \mathbb{R}^2 (left) and also finite support; model on $[0, \infty)^2$ (right), without finite support (Cuenin et al., 2016).

In order to extend appropriately the univariate $VI = \sigma^2 m^{-2}$ to the k -dimensional one for any positive continuous random vector \mathbf{Y} on $(0, \infty)^k$ having positive (elementwise) mean vector $\mathbf{m} = (m_1, \dots, m_k)^\top$ and covariance matrix Σ , we consider the product of two matrices, namely $\Sigma \mathbf{M}^{-1}$, where $\mathbf{M} = \mathbf{m} \mathbf{m}^\top$ is the $k \times k$ matrix outer product of \mathbf{m} and which is well-defined. According to the singularity of $\mathbf{M} = \mathbf{m} \mathbf{m}^\top$, the unique Moore-Penrose inverse \mathbf{M}^+ of \mathbf{M} is therefore

$$\mathbf{M}^+ = \frac{\mathbf{M}}{(\mathbf{m}^\top \mathbf{m})^2}.$$

Then, we have $\Sigma \mathbf{M}^+ = (\mathbf{M}^+ \Sigma)^\top$. Since the rank of \mathbf{M} is equal to 1, then $\mathbf{M}^+ \Sigma$ is also of rank 1 and has only one positive eigenvalue :

$$\lambda = \text{tr}(\mathbf{M}^+ \Sigma) = \text{tr}(\Sigma \mathbf{M}^+) = \frac{\mathbf{m}^\top \Sigma \mathbf{m}}{(\mathbf{m}^\top \mathbf{m})^2} =: \text{GVI},$$

where “ $\text{tr}(\cdot)$ ” stands for the trace operator.

This quantity λ does not depend on the number k of variables and it is numerically comparable to the univariate $VI = \sigma^2 m^{-2}$. Also, it characterizes uniquely the $\mathbf{M}^+ \Sigma$ matrix, leading to the following definition of GVI. Note finally that if $\Sigma = \mathbf{0}$ then we easily deduce $\lambda = 0$, and conversely. We thus have the natural ordering of the half non-negative real line for $\lambda \geq 0$.

C Proofs of the asymptotic results

C1 Proof of Proposition 4.4.1.

Part (i) : Let $\mathbf{Z} = (\dots, Y_j, \dots; \dots, Y_j Y_\ell, \dots)_{j \in \{1, \dots, k\}; \ell \in \{j, \dots, k\}}^\top$
 $\mathbf{Z}_i = (\dots, Y_{ij}, \dots; \dots, Y_{ij} Y_{i\ell}, \dots)_{j \in \{1, \dots, k\}; \ell \in \{j, \dots, k\}}^\top$ for $i \in \{1, \dots, n\}$, and the map $\Phi : (0, \infty)^k \times \mathbb{R}^{k(k+1)/2} \rightarrow (0, \infty)$ given through $\Phi(\mathbb{E}\mathbf{Z}) = \text{GVI}(\mathbf{Y})$ and $\Phi(n^{-1} \sum_{i=1}^n \mathbf{Z}_i) = \widehat{\text{GVI}}_n(\mathbf{Y})$; i.e., for $\boldsymbol{\theta} = (\dots, m_j, \dots; \dots, \sigma_{j\ell}, \dots)_{j \in \{1, \dots, k\}; \ell \in \{j, \dots, k\}}^\top$, $\Phi(\boldsymbol{\theta}) = (\mathbf{m}^\top \boldsymbol{\Sigma} \mathbf{m}) / (\mathbf{m}^\top \mathbf{m})^2$, where $\mathbf{m} = (m_1, \dots, m_k)^\top$ is the mean vector of \mathbf{Y} and $\boldsymbol{\Sigma} = (\sigma_{j\ell})_{j, \ell \in \{1, \dots, k\}}$ is the covariance matrix of \mathbf{Y} with $\sigma_{j\ell} = \sigma_{\ell j}$. Since Φ is differentiable at $\boldsymbol{\theta}$, the multivariate delta method (e.g., Serfling, 1980, Theorem A of Section 3.3) allows one to deduce that, as $n \rightarrow \infty$,

$$\sqrt{n} \left\{ \Phi \left(\frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \right) - \Phi(\mathbb{E}\mathbf{Z}) \right\} \rightsquigarrow \mathcal{N} \left[0, \left(\frac{\partial \Phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \times \text{cov} \mathbf{Z} \times \frac{\partial \Phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right].$$

To check that $\text{cov} \mathbf{Z} = \boldsymbol{\Gamma}$ of the proposition under the assumption on the fourth order moments of Y_j , one can rewrite \mathbf{Z} in the following order : $\mathbf{Z} = (\mathbf{Y}, \widetilde{\mathbf{Y}})$ with $\mathbf{Y} = (Y_1, \dots, Y_k)^\top$ and $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \dots, \widetilde{Y}_{\widetilde{k}})^\top$ such that $\widetilde{k} = k + (k-1) + \dots + 1 = k(k+1)/2$ and

$$\begin{aligned} \widetilde{\mathbf{Y}} &= (Y_1 Y_1 = \widetilde{Y}_1, Y_1 Y_2 = \widetilde{Y}_2, \dots, Y_1 Y_k = \widetilde{Y}_k, Y_2 Y_2 = \widetilde{Y}_{k+1}, \dots, Y_2 Y_k = \widetilde{Y}_{2k-1}, \\ &\quad Y_3 Y_3 = \widetilde{Y}_{2k}, \dots, Y_k Y_k = \widetilde{Y}_{k(k+1)/2})^\top. \end{aligned}$$

Then, the three main block matrices of $\boldsymbol{\Gamma}$ are successively found to be

$$\begin{aligned} \boldsymbol{\Sigma} &= \text{cov} \mathbf{Y} = (\text{cov}(Y_j, Y_{j'}))_{j, j' \in \{1, \dots, k\}}, \\ \boldsymbol{\Gamma}_4 &= \text{cov} \widetilde{\mathbf{Y}} = (\text{cov}(\widetilde{Y}_j, \widetilde{Y}_{j'}))_{j, j' \in \{1, \dots, k(k+1)/2\}} = (\text{cov}(Y_{j_1} Y_{j_2}, Y_{j'_1} Y_{j'_2}))_{j_1, j'_1 \in \{1, \dots, k\}; j_2 \in \{j_1, \dots, k\}; j'_2 \in \{j'_1, \dots, k\}}, \\ \boldsymbol{\Gamma}_3^\top &= (\text{cov}(Y_j, \widetilde{Y}_{j'}))_{j \in \{1, \dots, k\}; j' \in \{1, \dots, k(k+1)/2\}} = (\text{cov}(Y_j, Y_{j'_1} Y_{j'_2}))_{j, j'_1 \in \{1, \dots, k\}; j'_2 \in \{j'_1, \dots, k\}}. \end{aligned}$$

To see that $\partial \Phi(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} = \boldsymbol{\Delta}$, we first expand Φ as follows :

$$\Phi(\boldsymbol{\theta}) = \frac{\sum_{j=1}^k \sum_{\ell=1}^k \sigma_{j\ell} m_j m_\ell}{(\sum_{j=1}^k m_j^2)^2} = \frac{m_j^2 \sigma_{jj} + m_j \sum_{j' \neq j} \sigma_{jj'} m_{j'} + \sum_{\ell \neq j} m_\ell (\sigma_{\ell j} m_j + \sum_{j' \neq j} \sigma_{\ell j'} m_{j'})}{(m_j^2 + \sum_{j' \neq j} m_{j'}^2)^2}.$$

Then, direct calculations provide all components of Δ : for $j \in \{1, \dots, k\}$, one has

$$\begin{aligned}\Delta_j = \frac{\partial}{\partial m_j} \Phi(\theta) &= \left\{ 2m_j \sigma_{jj} + \sum_{j' \neq j} m_{j'} \sigma_{jj'} + \sum_{\ell \neq j} m_\ell \sigma_{\ell j} - 4m_j \left(\sum_{j'=1}^k m_{j'}^2 \right) \Phi(\theta) \right\} (\mathbf{m}^\top \mathbf{m})^{-2} \\ &= \left\{ 2 \sum_{j'=1}^k m_{j'} \sigma_{jj'} - 4m_j \left(\sum_{j'=1}^k m_{j'}^2 \right) \Phi(\theta) \right\} (\mathbf{m}^\top \mathbf{m})^{-2},\end{aligned}$$

and $\Delta_{jj} = \partial \Phi(\theta) / \partial \sigma_{jj} = m_j^2 / (\mathbf{m}^\top \mathbf{m})^2$ while for $\ell \in \{j+1, \dots, k\}$, $\Delta_{j\ell} = \partial \Phi(\theta) / \partial \sigma_{j\ell} = 2m_j m_\ell / (\mathbf{m}^\top \mathbf{m})^2$. This ends the proof of Part (i).

Part (ii) : Introduce $\mathbf{W} = (Y_1, \dots, Y_k; Y_1^2, \dots, Y_k^2)^\top$, $\mathbf{W}_i = (Y_{i1}, \dots, Y_{ik}; Y_{i1}^2, \dots, Y_{ik}^2)^\top$ for $i \in \{1, \dots, n\}$ and the map $\Psi : (0, \infty)^{2k} \rightarrow (0, \infty)$ defined by

$\Psi(\theta) = (\sum_{j=1}^k m_j^2 \sigma_{jj}) / (\sum_{j=1}^k m_j^2)^2$ with $\theta = (m_1, \dots, m_k; \sigma_{11}, \dots, \sigma_{kk})^\top$. Then, one has $\Psi(\mathbb{E}\mathbf{W}) = \text{MVI}(\mathbf{Y})$ and $\Psi(n^{-1} \sum_{i=1}^n \mathbf{W}_i) = \widehat{\text{MVI}}_n(\mathbf{Y})$. The function Ψ is differentiable at the point θ and, therefore, a straightforward application of the multivariate delta method leads to the conclusion that, as $n \rightarrow \infty$,

$$\sqrt{n} \left\{ \Psi \left(\frac{1}{n} \sum_{i=1}^n \mathbf{W}_i \right) - \Psi(\mathbb{E}\mathbf{W}) \right\} \rightsquigarrow \mathcal{N} \left[0, \left(\frac{\partial \Psi(\theta)}{\partial \theta} \right)^\top \times \text{cov} \mathbf{W} \times \frac{\partial \Psi(\theta)}{\partial \theta} \right].$$

Here, it is now trivial that $\text{cov} \mathbf{W} = \mathbf{\Pi}$ of the theorem under the assumption of the finite moments on Y_j and also that $\partial \Psi(\theta) / \partial \theta = \mathbf{\Lambda}$ with $\Lambda_j = \partial \Psi(\theta) / \partial m_j = \{2m_j \sigma_{jj} - 4m_j (\sum_{j'=1}^k m_{j'}^2) \Psi(\theta)\} (\mathbf{m}^\top \mathbf{m})^{-2}$ and $\Lambda_{jj} = \partial \Psi(\theta) / \partial \sigma_{jj} = m_j^2 (\mathbf{m}^\top \mathbf{m})^{-2}$ for all $j \in \{1, \dots, k\}$. This concludes the proof. ■

C2 Proof of Proposition 4.4.2. According to the both continuous maps $\Phi : (0, \infty)^k \times \mathbb{R}^{k(k+1)/2} \rightarrow (0, \infty)$ defined through $\Phi(\mathbb{E}\mathbf{Z}) = \text{GVI}(\mathbf{Y})$ and $\Phi(n^{-1} \sum_{i=1}^n \mathbf{Z}_i) = \widehat{\text{GVI}}_n(\mathbf{Y})$ and $\Psi : (0, \infty)^{2k} \rightarrow (0, \infty)$ such that $\Psi(\mathbb{E}\mathbf{W}) = \text{MVI}(\mathbf{Y})$ and $\Psi(n^{-1} \sum_{i=1}^n \mathbf{W}_i) = \widehat{\text{MVI}}_n(\mathbf{Y})$ in the proof of Proposition 4.4.1, the desired result is easily deduced from $n^{-1} \sum_{i=1}^n \mathbf{Z}_i \xrightarrow{a.s.} \mathbb{E}\mathbf{Z}$ and $n^{-1} \sum_{i=1}^n \mathbf{W}_i \xrightarrow{a.s.} \mathbb{E}\mathbf{W}$, respectively. ■

Chapter 5

GWl : an R package for count and continuous generalized variability indexes

5.1 Introduction

Firstly, both functions of the univariate Poisson dispersion index (DI) for count data and the univariate exponential variation index (VI) for nonnegative continuous data are performed. Next, other functions of univariate indexes such the binomial dispersion index (DIb), the negative binomial dispersion index (DI_{nb}) and the inverse Gaussian variation index (VI_{IG}) are given. Finally, we are computed some multivariate versions of these functions such that the generalized dispersion index (GDI) with its marginal one (MDI) and the generalized variation index (GVI) with its marginal one (MVI) too. This package is available at <https://cran.r-project.org/web/packages/GWI/index.html>.

This Chapter is organized as follows. In Section 5.2, the function of univariate Poisson or Fisher dispersion index (Fisher, 1934) for count data function is displayed. The function of univariate exponential or Jørgensen variation index (Abid et al., 2020) for positive continuous data is performed in Section 5.3. The functions of binomial dispersion index and negative binomial dispersion index are computed in Section 5.4 and Section 5.5, respectively. The function of inverse Gaussian variation index is displayed in Section 5.6. In Section 5.7, the generalized dispersion index (GDI) and its

marginal one (MDI) of Kokonendji and Puig (2018) functions are carried out. Finally, the generalized variation index (GVI) and its marginal ones (MVI) functions are computed in Section 5.8.

5.2 Funtion for Poisson dispersion index

Description

The function computes the univariate Poisson dispersion index (DI) for a count random variable.

Usage

```
di.fun (X)
```

Arguments

X a count random variable

Details

di.fun provides the univariate Poisson dispersion index.

Value

Returns

di The Poisson dispersion index

Examples

```
X<- c(6,7,8,9,8,4,7,6,12,8,0)
```

```
di.fun(X)
```

```
T<- c(61,72,83,94,85,46,77,68,129,80,10,12,12,3,4,5)
```

`di.fun(T)`

5.3 Function for exponential variation index

Description

The function calculates the univariate exponential variation index (VI) for a positive continuous random variable.

Usage

`vi.fun (X)`

Arguments

`X` a positive continuous random variable

Details

`vi.fun` computes the univariate exponential variation index.

Value

Returns

`vi` The exponential variation index

Examples

```
X<-c(6.23,7.02,8.94,9.56,8.01,4.34,7.44,6.66,12.72,8.34,0)
```

```
vi.fun(X)
```

```
T<-c(6.2,7.1,8.9,9.7,8.1,4.4,7.5,6.3,12.2,8.3)
```

```
vi.fun(T)
```

5.4 Funtion for binomial dispersion index

Description

The function computes the binomial dispersion index for a given number of trials $N \in \{1, 2, \dots\}$.

Usage

```
dib.fun (X, N)
```

Arguments

X A count random variable

N The number of trials of binomial distribution

Details

`dib.fun` provides the Poisson dispersion index with respect to the binomial distribution.

Value

Returns

`dib` The binomial dispersion index

Examples

```
X<- c(6,7,8,9,8,4,7,6,12,8,0)
```

```
dib.fun(X)
```

```
T<- c(61,72,83,94,85,46,77,68,129,80,10,12,12,3,4,5)
```

```
dib.fun(T)
```

5.5 Funtion for negative binomial dispersion index

Description

The function computes the negative binomial dispersion index for a given dispersion parameter $l \in (0, \infty)$.

Usage

```
dinb.fun (X, l)
```

Arguments

X A count random variable

l The dispersion parameter of negative binomial distribution

Details

dinb.fun computes the dispersion index with respect to negative binomial distribution. See Touré et al. (2020) and Abid et al. (2021) for more details.

Value

Returns

dinb The negative binomial dispersion index

Examples

```
X<- c(6,7,8,9,8,4,7,6,12,8,0)
```

```
dinb.fun(X)
```

```
T<- c(61,72,83,94,85,46,77,68,129,80,10,12,12,3,4,5)
```

```
dinb.fun(T)
```

5.6 Funtion for inverse Gaussian variation index

Description

The function computes the inverse Gaussian variation index with shape parameter $l \in (0, \infty)$.

Usage

```
viiG.fun (X,l)
```

Arguments

X A positive continuous random variable

l The shape parameter of the inverse Gaussian distribution

Details

viiG.fun provides the univariate Poisson dispersion index.

Value

Returns

viiG The inverse Gaussian variation index

Examples

```
X<- c(6,7,8,9,8,4,7,6,12,8,0)
```

```
viiG.fun(X)
```

```
T<- c(61,72,83,94,85,46,77,68,129,80,10,12,12,3,4,5)
```

```
viiG.fun(T)
```

5.7 Function for GDI and MDI

Description

The function computes GDI and MDI indexes introduced by Kokonendji and Puig (2018).

Usage

```
gmdi.fun (Y)
```

Arguments

Y a matrix of count random variables

Details

gmi.fun computes GDI and MDI introduced by Kokonendji and Puig (2018).

Indeed, Kokonendji and Puig (2018, Section 3) defined the *generalized dispersion index* of Y by

$$\text{GDI}(Y) = \frac{\sqrt{EY^\top} (\text{cov} Y) \sqrt{EY}}{EY^\top EY}.$$

Note that when $k = 1$, $\text{GDI}(Y)$ is just the classical Fisher dispersion index DI . $\text{GDI}(Y)$ makes it possible to compare the full variability of Y (in the numerator) with respect to its expected uncorrelated Poissonian variability (in the denominator) which depends only on EY .

If we only want to take into account the dispersion information coming from the margins, the authors defined the *multiple marginal dispersion index* by

$$\text{MDI}(Y) = \frac{\sqrt{EY^\top} (\text{diag var} Y) \sqrt{EY}}{EY^\top EY} = \sum_{j=1}^k \frac{\{E(Y_j)\}^2}{EY^\top EY} \text{DI}(Y_j).$$

Value

Returns a vector containing :

GDI The generalized dispersion index

MDI The marginal dispersion index

Examples

```
Y<-cbind(c(1,2,3,4,5,6,7,8),c(1,2,3,4,5,6,7,8))
```

```
gmdi.fun(Y)
```

```
Z<-cbind(c(1,2,3,4,5),c(1,2,3,4,5),c(1,2,3,4,5))
```

```
gmdi.fun(Z)
```

5.8 Function for GVI and MVI

Description

The function computes GVI and MVI indexes.

Usage

```
gmvi.fun (Y)
```

Arguments

Y a matrix of positive continuous random variables

Details

gmvi.fun computes GVI of (4.2) and MVI of (4.7).

Value

Returns a vector containing :

GVI The generalized variation index

MVI The marginal variation index

Examples

```
Y<-cbind(c(1.32,2.55,3.06,4.12,5.37),c(1.11,2.09,3.24,4.33,5.11))
```

```
gmvi.fun(Y)
```

```
Z<-cbind(c(1.34,2.89,3.01),c(1.66,2.88,3.23),c(1.88,2.22,3.23))
```

```
gmvi.fun(Z)
```

5.9 Conclusion

In this last chapter, we implemented an R package which allows to calculate for a given dataset the dispersion and variation indexes in the univariate as well as in the multivariate case. However, for a very large dataset, we can improve this package in order to have the correlation matrices of the smaller sub-matrices of the main matrix. We can also consider the construction of confidence intervals for these different indexes. These last works constitute a future field of research.

Chapter 6

Conclusions and perspectives

In this thesis, we have first introduced the new relative variability index RWI which brings together the classical RDI for count models and the more recent RVI for non-negative continuous models. We have subsequently proposed the unified asymptotic tests having very good behaviours. From the above examples of application, the very common case of the Poisson dispersion index test Part (iii) has been well improved and the new exponential variation index test Part (i) works wonderfully. In finite samples, we can suggest a bootstrapped approach for reducing the estimated variances. At this stage of analysis, other variability index tests will need some modifications of their statistical tests. Otherwise, the RWI may be extended to comparable variables of the same mean and support. We have a simple situation where the support $\mathcal{S} = \mathbb{R}$ and we compare, for example, the centered Gaussian with respect to the Student distribution, which is not an NEF. As for the multivariate setup, we already have the multivariate version of RDI in Kokonendji and Puig (2018); and, Kokonendji et al. (2020) are recently proposed the multivariate RVI and the corresponding unified version RWI. Eventually, work is in progress for (robust) test statistics of the multivariate RWI.

In the third chapter, we have first exhibited an efficient representation of any nonnegative continuous distribution as a WED in different points of view or another weighted version of any previously fixed nonnegative continuous distribution. We then established several characterizations and weighing operations with possible extensions. As a matter of fact, a nonnegative continuous weight function involves some informa-

tion for comparing or investigating the original distribution. For instance, the duality (if there exists) and variation phenomenon have been investigated through the nature of their weight functions. Given the parametric representation of a WED in reliability, we can also derive all the functions needed for life data analysis such as survival, cumulative density, hazard rate, moment generating and residuals life functions. According to various degrees of difficulty depending on the complexity of the pdf, the maximum product of spacings method considered as a powerful alternative to maximum likelihood estimation can be applied as follows. Let x_1, x_2, \dots, x_n be an n -ordered sample from the WED of parameters θ . Let $F(x; \theta)$ be the cumulative density function and $D_i(\theta) := F(x_i; \theta) - F(x_{i-1}; \theta)$ for $i = 1, 2, \dots, n+1$, be the uniform spacings of a random sample from the WED, where $F(x_0; \theta) = 0$ and $F(x_{n+1}; \theta) = 1$. The maximum product spacings estimators $\hat{\theta}$ are obtained by maximizing the geometric mean of the spacings given by

$$\left[\prod_{i=1}^{n+1} D_i(\theta) \right]^{\frac{1}{n+1}},$$

with respect to θ . In general, the WEDs used for reliability and life data analysis are usually limited to three parameters, known as scale, shape and location parameters. These WEDs complete the panoply of parametric families of life distributions (e.g., Marshall and Olkin, 2007). Finally, from the self-decomposition result of Corollary 3.3.4 and in the direction that we have unified WEDs, we can also continue in the nonparametric and semiparametric approaches of Marshall and Olkin (2007). In fact, in order to estimate any nonnegative continuous pdf $f : [0, \infty) \rightarrow \mathbb{R}$ under nonparametric assumption, we can consider a semiparametric approach by decomposing

$$f(\cdot) = w(\cdot)f_{\mathcal{E}}(\cdot; \theta) =: f_w(\cdot; \theta),$$

with $f_w(\cdot; \theta)$ being the unknown but purely parametric part, well specified and depending on θ , namely exponential or Weibull, as well as $w(\cdot)$ being the unknown continuous weight function or nonparametric part on $[0, \infty)$ for fixed θ . See, e.g., Kokonendji and Somé (2021) for more details on semiparametric approach in multivariate setup. According to appropriate smoothers on $[0, \infty)$ (e.g., Kokonendji and Somé, 2018), further

work in this direction is underway. We can refer to Kokonendji et al. (2009, 2017) for count distributions.

In the fourth chapter, from the univariate variation index (Abid et al., 2020) and the multivariate dispersion indexes for count models (Kokonendji and Puig, 2018), we have first introduced multivariate variation indexes GVI, MVI and RVI for continuous distributions on non-negative orthant. All the proposed indexes are easy to handle from a theoretical and practical point of view. Unlike the intuitive marginal variation index MVI, the index GVI takes into account the correlations between variables. The ratio of two GVI provides the index RVI for changing the reference distribution of the measure of over-, equi- and under-variation in the multivariate framework. The interpretation and some properties of GVI and MVI are provided. Also, the asymptotic variances of GVI and MVI obtained from Proposition 4.4.1 seem to provide large standard errors for small and moderate sample sizes; they can be improved, for instance, through a bootstrap method. An example of real data analysis is presented, helping to select an appropriate multivariate model.

Then, from $RVI_X(Y)$ given in (4.3) one exactly obtains its equivalent (i.e., relative dispersion index) $RDI_X(Y)$ for count models by changing the support $\mathcal{S} = \mathbb{N}^k$ of X and Y ; see Formula (9) of Kokonendji and Puig (2018). Tests of hypothesis relying on the corresponding estimators as test statistics with their asymptotic normality distributions should be deduced. Finally, let us note the following problems which are in advanced discussion. For the unification of GVI with GDI of Kokonendji and Puig (2018), we need the multivariate equivalent to Proposition 2.2.1. Then, a global test would deserve to be seen as in the univariate case after unification. It would be reasonable to first characterize the bivariate weighted exponential distributions to control the correlations between the variables. Therefore, how to investigate the multivariate connections to over-, equi- and under-variation through $m \mapsto GVI_F(m)$ or $MVI_F(m)$? How, for instance, to discriminate some closed distributions from these indexes? See, e.g., Dey and Kundu (2009) for a univariate case. Statistical tests of these multivariate variation indexes can

be produced in the direction of Aerts and Haesbroeck (2017); see also Feltz and Miller (1996). This is the reason why the work must continue in multivariate setup in order to be able to do the similar in univariate case, namely statistical tests. Therefore, the work done in Section 4.4 deserves to be continued. How to define the multivariate equivalent of the categorial indexes of Weiß (2019a, 2019b)? All these extensions will form the basis of future work.

References

- [1] Abid, R., Kokonendji, C.C. and Masmoudi, A. (2019). Geometric dispersion models with real quadratic v-functions, *Statistics and Probability Letters* **145**, 197–204.
- [2] Abid, R., Kokonendji, C.C. and Masmoudi, A. (2020). Geometric Tweedie regression models for continuous and semicontinuous data with variation phenomenon, *ASTA Advances in Statistical Analysis* **104**, 33–58.
- [3] Abid, R., Kokonendji, C.C. and Masmoudi, A. (2021). On Poisson-exponential-Tweedie models for ultra-overdispersed count data, *ASTA Advances in Statistical Analysis* **105**, 1-23.
- [4] Abid, R. and Kokonendji, C.C. (2021). Choice between and within the classes of Poisson-Tweedie and Poisson-exponential-Tweedie count models, *Communications in Statistics - Simulation and Computation*, in press (DOI : 10.1080/03610918.2021.1898635), pp.1-15.
- [5] Abramowitz, M., Stegun, I.A. (1972). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, *Dover Publications*, New York.
- [6] Aerts, S., Haesbroeck, G. (2017). Robust asymptotic tests for the equality of multivariate coefficients of variation, *TEST* **26**, 163–187.
- [7] Akantziliotou, C., Rigby, R.A. and Stasinopoulos, D.M. (2008). A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution, *Computational Statistics and Data Analysis* **53**, 381-393.

- [8] Albert, A., Zhang, L. (2010). A novel definition of the multivariate coefficient of variation, *Biometrical Journal* **52**, 667–675.
- [9] Altunkaynak, B. and Gamgam, H. (2018). Bootstrap confidence intervals for the coefficient of quartile variation, *Communications in Statistics-Simulation and Computation* **48**, 2138–2146.
- [10] Alqallaf, F., Ghitany, M.E. and Agostinelli, C. (2015). Weighted exponential distribution : different methods of estimations, *Applied Mathematical and Information Science* **9**, 1167–1173.
- [11] Al-Mutairia, D.K., Ghitany, M.E. and Kundu, D. (2018). Weighted Weibull distribution : bivariate and multivariate cases, *Brazilian Journal of Probability and Statistics* **32**, 20–43.
- [12] Angelo, C., Brian, R. (2019). Package boot, <https://cran.r-project.org/web/packages/boot/>. Accessed 20 Dec 2019.
- [13] Arnold, B.C. and Tony Ng, H.K. (2011). Flexible bivariate beta distributions, *Journal of Multivariate Analysis* **102**, 1194–1202.
- [14] Aissani, A. (1992). Modèles Stochastique de la Théorie de Fiabilité, *Springer*, New York.
- [15] Avelino, E.G. (1978). The Index of Dispersion, Ph.D. thesis, *University of British Columbia*, Vancouver.
- [16] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics* **12**, 171–178.
- [17] Bakouch, H., Chesneau, C. and Khan, M.N. (2021). On a weighted exponential distribution with a logarithmic weight : Theory and applications, *Afrika Matematika* **32**, DOI :10.1007/s13370-020-00861-7.
- [18] Balakrishnan, N. and Basu, A.P. (1995). The Exponential Distribution : Theory, Models and Applications, *Gordon and Breach*, Amsterdam.

-
- [19] Bar-Lev, S.K. (2020). Independent, tough identical results : The class of Tweedie on power variance functions and the class of Bar-Lev and Enis on reproducible natural exponential families, *International Journal of Statistics and Probability* **9**, 30–35.
- [20] Barlow, R.A. and Proschan, F. (1981). Statistical Theory of Reliability and Life Testing : Probability Models, *Silver Springs, Maryland*.
- [21] Barndorff-Nielsen, O.E. (1997). Normal inverse Gaussian distribution and stochastic volatility modelling, *Scandinavian Journal of Statistics* **24**, 1–13.
- [22] Basu, A.P. (1988). Multivariate exponential distributions and their applications in reliability. In : Handbook of Statistics, **vol. 7**, Quality Control and Reliability, P.R. Krishnaiah and C.R. Rao (eds), *Elsevier*, Amsterdam, 467–477.
- [23] Bonat, W.H. and Jørgensen, B. (2016). Multivariate covariance generalized linear models, *Journal of the Royal Statistical Society* **65**, 649–675.
- [24] Bonat, W.H., Jørgensen, B., Kokonendji, C.C., Hinde, J. and Demétrio, C.G.B. (2018). Extended Poisson-Tweedie : Properties and regression models for count data, *Statistical Modelling* **18**, 24–49.
- [25] Böhning, D. (1994). A note on a test for Poisson overdispersion, *Biometrika* **81**, 418–419.
- [26] Boubacar Maïnassara, Y. and Kokonendji, C.C. (2014). On normal stable Tweedie models and power-generalized variance functions of only one component, *TEST* **23**, 585–606.
- [27] Casalis, M. (1996). The $2d + 4$ simple quadratic natural exponential families on \mathbb{R}^d , *Annal of Statistics* **24**, 1828–1854.
- [28] Cramér, H. (1974). Mathematical Methods of Statistics, *Princeton University Press*, Princeton.
- [29] Cuenin, J., Jørgensen, B., and Kokonendji, C.C. (2016). Simulations of full multivariate Tweedie with flexible dependence structure, *Computational Statistics* **31**, 1477–1492.

- [30] Das, S. and Kundu, D. (2016). On weighted exponential distribution and its length biased version, *Journal of the Indian Society for Probability and Statistics* **17**, 57–77.
- [31] del Castillo, J. and Pérez-Casany, M. (2005). Overdispersed and underdispersed Poisson generalizations, *Journal of Statistical Planning and Inference* **134**, 486-500.
- [32] Dey, S., Ali, S. and Park, C. (2015). Weighted exponential distribution : properties and different methods of estimation, *Journal of Statistical Computation and Simulation* **85**, 3641–3661.
- [33] Dey, A.K. and Kundu, D. (2009). Discriminating among the log-normal, Weibull, and generalized exponential distributions, *IEEE Transactions on Reliability* **58**, 416–424.
- [34] Domma, F., Condino, F. and Popović, B.V. (2017). A new generalized weighted Weibull distribution with decreasing, increasing, upside-down bathtub, N-shape and M-shape hazard rate, *Journal of Applied Statistics* **44**, 2978–2993.
- [35] Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap, *Chapman & Hall*, London.
- [36] Elsherpieny, E-S.A., Abdelall, Y.Y. and Mohamed, A.A. (2017). On the exponentiated new weighted Weibull distribution, *International Journal of Engineering and Applied Sciences* **4**, 41–50.
- [37] Engel, B. and te Brake, J. (1993). Analysis of embryonic development with a model for under- or overdispersion relative to binomial variation, *Biometrics* **49**, 269-279.
- [38] Famoye, F. (2019). Exponentiated Weibull-geometric distribution and its application to count data, *Journal of Data Science* **17**, 712-725.
- [39] Feltz, C.J. and Miller, G.E. (1996). An asymptotic test for the equality of coefficients of variation from k populations, *Statistics in Medicine* **15**, 647–658.
- [40] Fisher, R.A. (1934). The effects of methods of ascertainment upon the estimation of frequencies, *Annals of Eugenics* **6**, 13-25.

-
- [41] Gómez, Y. M., Bolfarine, H. and Gómez, H. W. (2014). A new extension of the exponential distribution, *Revista Colombiana de Estadística* **37**, 25–34.
- [42] Gupta, R.D. and Kundu, D. (2001). Exponentiated exponential family : An alternative to gamma and Weibull, *Biometrical Journal* **33**, 117–130.
- [43] Gupta, R.D. and Kundu, D. (2009). New class of weighted exponential distribution, *Statistics* **43**, 621–634.
- [44] Gupta, R.C., Sim, S.Z. and Ong, S.H. (2014). Analysis of discrete data by Conway-Maxwell Poisson distribution, *AStA Advances in Statistical Analysis* **4**, 327–343.
- [45] Hayashi, F. (2000). Econometrics, Chapter 10, *Princeton University Press*, URL : http://fhayashi.fc2web.com/hayashi_econometrics.htm, 665–667.
- [46] Hinde, J. and Demétrio, C.G.B. (1998). Overdispersion : Models and Estimation, *Associacao Brasileira de Estatística*, Sao Paulo.
- [47] Hougaard, P., Lee, M-L.T. and Whitmore, G.A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes, *Biometrics* **53**, 1225–1238.
- [48] Joe, H. (2014). Dependence Modeling with Copulas, *Monographs on Statistics and Applied Probability* **134**. Chapman & Hall - CRC Press, London.
- [49] Johnson, N.L., Kemp, A.W. and Kotz, S. (2005). Univariate Discrete Distributions, 3rd Edition, *Wiley*, New York.
- [50] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). Continuous Univariate Distributions, 2nd Edition, *Wiley*, New York.
- [51] Johnson, R.A. and Wichern, D.W. (2007). Applied Multivariate Statistical Analysis, 6th Edition, *Pearson Prentice Hall*, New Jersey.
- [52] Jørgensen, B. (1997). The Theory of Dispersion Models, *Chapman & Hall*, London.
- [53] Jørgensen, B. and Kokonendji, C.C. (2011). Dispersion models for geometric sums, *Brazilian Journal of Probability and Statistics* **25**, 263–293.

- [54] Jørgensen, B., Martínez, J.R. and Tsao, M. (1994). Asymptotic behaviour of the variance function, *Scandinavian Journal of Statistics* **21**, 223-243.
- [55] Kharazmi, O., Mahdavi, A. and Fathizadeh, M. (2015). Generalized weighted exponential distribution, *Communications in Statistics-Simulation and Computation* **44**, 1557–1569.
- [56] Kokonendji, C.C. (2014). Over- and underdispersion models. In : N. Balakrishnan (Ed.), *The Wiley Encyclopedia of Clinical Trials- Methods and Applications of Statistics in Clinical Trials*, **Vol.2** (Chap.30), pp. 506-526, Wiley, New York.
- [57] Kokonendji, C.C., Bonat, W.H. and Abid, R. (2021). Tweedie regression models and its geometric sums for (semi-)continuous data, *WIREs Computational Statistics* **13**, e1496, 17 pp.
- [58] Kokonendji, C.C., Demétrio, C.G.B. and Zocchi, S.S. (2007). On Hinde-Demétrio regression models for overdispersed count data, *Statistical Methodology*. **4**, 271-291.
- [59] Kokonendji, C.C., Dossou-Gbete, S. and Demétrio, C.G.B. (2004). Some discrete exponential dispersion models : Poisson-Tweedie and Hinde-Demetrio classes, *Statistics and Operations Research Transactions* **28**, 201-214.
- [60] Kokonendji, C.C., Mizère, D. and Balakrishnan, N. (2008). Connections of the Poisson weight function to overdispersion and underdispersion, *Journal of Statistical Planning and Inference* **138**, 1287-1296.
- [61] Kokonendji, C.C. and Moypemna Sembona, C.C. (2018). Characterization and classification of multiple stable Tweedie models, *Lithuanian Mathematical Journal* **58**, 441-456.
- [62] Kokonendji, C.C. and Puig, P. (2018). Fisher dispersion index for multivariate count distributions : A review and a new proposal, *Journal of Multivariate Analysis* **165**, 180-193.
- [63] Kokonendji, C.C., Senga Kiessé, T. and Balakrishnan, N. (2009). Semiparametric estimation for count data through weighted distributions, *Journal of Statistical Planning and Inference* **139** , 3625–3638.

-
- [64] Kokonendji, C.C. and Somé, S.M. (2018). On multivariate associated kernels to estimate general density functions, *Journal of the Korean Statistical Society* **47**, 112–126.
- [65] Kokonendji, C.C. and Somé, S.M. (2021). Bayesian bandwidths in semiparametric modelling for nonnegative orthant data with diagnostics, *Stats* **4**, 162-183.
- [66] Kokonendji, C.C. and Sawadogo, A. (2021). The exponential variation property for the Weibull distribution, *International Journal of Applied Mathematics and Statistics* **60**, 90-97.
- [67] Kokonendji, C.C., Touré, A.Y. and Sawadogo, A. (2021a). Relative variation indexes for multivariate continuous distributions on $[0, \infty)^k$ and extensions, *AStA Advances in Statistical Analysis* **104**, 285-307.
- [68] Kokonendji, C.C., Zougab, N. and Senga Kiessé, T. (2017). Poisson-weighted estimation by discrete kernel with application to radiation biodosimetry. In *Biomedical Big Data and Statistics for Low Dose Radiation Research - Extended Abstracts Fall 2015*, vol. **VII**, Part II, Chap.19, pp. 115-120 (Editors : Ainsbury, E.A., Calle, M.L., Cardis, E., Einbeck, J., Gómez, G., Puig, P.), Springer Birkhauser, Basel.
- [69] Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). Continuous Multivariate Distributions - Models and Applications, Second Edition, *Wiley*, New York.
- [70] Lam, H-K. (1980). Remarks on the distribution of the sample variance in exponential sampling, *Communications in Statistics - Simulation and Computation* **9**, 639-647.
- [71] Letac, G., Mora, M. (1990). Natural real exponential families with cubic variance functions, *The Annals of Statistics* **18**, 1-37.
- [72] Mahdavi, A. (2015). Two weighted distributions generated by exponential distribution, *Journal of Mathematical Extension* **9**, 1–12.
- [73] Mahdavi, A. and Jabbari, L. (2017). An extended weighted exponential distribution, *Journal of Modern Applied Statistical Methods* **16**, 296–307.
- [74] Marshall, A.W. and Olkin, I. (1967). A multivariate exponential distribution, *Journal of American Statistical Association* **62**, 30–44.

- [75] Marshall, A.W. and Olkin, I. (2007). Life Distributions : Structure of Nonparametric, Semiparametric, and Parametric Families, *Springer*, New York.
- [76] Mizère, D., Kokonendji, C.C. and Dossou-Gbété, S. (2006). Quelques tests de la loi de Poisson contre des alternatives générales basées sur l'indice de dispersion de Fisher, *Revue de Statistique Appliquée* **54**, 61-84.
- [77] Morris, C.N. (1982). Natural exponential families with quadratic variance functions, *The Annals of Statistics* **10**, 65-80.
- [78] Oguntunde, P.E. (2015). On the exponentiated weighted exponential distribution and its basic statistical properties, *Applied Science Reports* **10**, 160–167.
- [79] Python Software Foundation, (2020). Python Language Reference, Version 3.7.3, Available at <http://www.python.org> . Accessed 24 Feb 2020.
- [80] R Core Team, R, (2018). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
- [81] Ramadan, M.M. (2013). A class of weighted Weibull distributions and its properties, *Studies in Mathematical Science* **6**, 35–45.
- [82] Ristić, M.M. and Balakrishnan, N. (2012). The gamma-exponentiated exponential distribution, *Journal of Statistical Computation and Simulation* **82**, 1191–1206.
- [83] Roy, S. and Adnan, M.A.S. (2012). Wrapped weighted exponential distributions, *Statistics and Probability Letters* **82**, 77–83.
- [84] Saghir, A., Hamedani, G.G., Tazeem, S. and Khadim, A. (2017). Weighted distributions : A brief review, perspective and characterizations, *International Journal of Statistics and Probability* **6**, 109–131.
- [85] Sellers, K.F. and Raim, A. (2016). A flexible zero-inflated model to address data dispersion, *Computational Statistics and Data Analysis* **99** , 68-80.
- [86] Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics, *Wiley*, New York.

-
- [87] Shahbaz, S., Shahbaz, M.Q. and Butt, N.S. (2010). A class of weighted Weibull distribution, *Pakistan Journal of Statistics and Operation Research* **6**, 53–59.
- [88] Shakhathreh, M.K. (2012). A two-parameter of weighted exponential distribution, *Statistics and Probability Letters* **82**, 252–261.
- [89] Somé, S.M. and Kokonendji, C.C. (2021). Bayesian selector of adaptive bandwidth for multivariate gamma kernel estimator on $[0, \infty)^d$, *Journal of Applied Statistics*, in press (DOI : 10.1080/02664763.2021.1881456), pp.1-22.
- [90] Su, P. (2020). Generation of Multivariate Data with Arbitrary Marginals : Package, <https://cran.r-project.org/web/packages/NORTARA/>. Accessed 19 Feb 2020.
- [91] Teimouri, M. and Gupta, A.K. (2011). On a bivariate Weibull distribution, *Advances and Applications in Statistics* **22**, 77–106.
- [92] Tiago de Oliveira, J. (1965). Some elementary tests of mixtures of discrete distributions. In : G.P. Patil (Ed.), *Classical and Contagious Discrete Distributions*, pp. 379-384, Pergamon, New York.
- [93] Touré, A.Y., Dossou-Gbété, S. and Kokonendji, C.C. (2020). Asymptotic normality of the test statistics for relative dispersion and relative variation indexes, *Journal of Applied Statistics* **47**, 2479-2491.
- [94] Touré, A.Y., Dossou-Gbété, S. and Kokonendji, C.C. (2021). Correction to : Asymptotic normality of the test statistics for the unified relative dispersion and relative variation indexes, *Journal of Applied Statistics*, Submitted.
- [95] Tweedie, M.C.K. (1984). An index which distinguishes between some important exponential families. In : Ghosh, J.K., Roy, J. (eds.) *Statistics : Applications and New Directions, Proceedings of the Indian Statistical Golden Jubilee International Conference*, Indian Statistical Institute, Calcutta, pp. 579–604.
- [96] Wang, Z. (2011). One mixed negative binomial distribution with application, *Journal of Statistical Planning and Inference* **141**, 1153-1160.
- [97] Weiß, C.H. (2018). *An Introduction to Discrete-Valued Times Series*, Wiley, Hoboken NJ.

- [98] Weiß, C.H. (2019a). On the sample coefficient of nominal variation. In : A. Steland et al. (eds), *Stochastic Models, Statistics and Their Applications, Springer Proceedings in Mathematics & Statistics*, **vol. 294** (Chap.18), pp. 239-250, Dresden, Germany.
- [99] Weiß, C.H. (2019b). On some measures of ordinal variation, *Journal of Applied Statistics* **46**, 2905-2926.

Appendix

6.1 R codes for simulation studies

6.1.1 Script in R for making bootstrap with bivariate GVI and MVI

```
# Simulation data No.1
# Simulation of bivariate continuous data using NORTARA package
# Loading extra packages
require(NORTARA)
# Marginal distributions:
#  $X_1 \sim W(\beta = 0.6, \alpha = 2)$ 
#  $X_2 \sim W(\beta = 0.4, \alpha = 1)$ 
# Correlation matrix
COR <- matrix(c(1,-0.23,-0.23,1),2,2)
invcdfnames <- c("qweibull","qweibull")
paramslists <- list(
  m1 = list(shape=0.6, scale=2),
  m2 = list(shape=0.4, scale=1)
)
data <- NORTARA::genNORTARA(n = 50, cor_matrix = COR,
                           invcdfnames = invcdfnames,
                           paramslists = paramslists)

data=cbind(y,x)
```



```
m=c(mean(data[,1]),mean(data[,2]))

# GVI
(t(m)%*%var(data)%*%(m))/(t(m)%*%m)^2

# MVI
(t(m)%*%diag(diag(var(data)))%*%(m))/(t(m)%*%m)^2

# bootstrap gvi and mvi (\textbf{10000 replicates})
n=length(y)
nb=10000; z=seq(1,n);gvi=numeric(nb);mvi=numeric(nb)
for(i in 1:nb){
  zb=sample(z,n,replace=T)
  datab=cbind(data[zb,1],data[zb,2])
  m=c(mean(datab[,1]),mean(datab[,2]))
  gvi[i]=(t(m)%*%var(data)%*%(m))/(t(m)%*%m)^2
  mvi[i]=(t(m)%*%diag(diag(var(datab)))%*%(m))/(t(m)%*%m)^2
}
hist(gvi)
hist(mvi)

# bootstrap mean and sd
c(mean(gvi),sd(gvi))
c(mean(mvi),sd(mvi))

# 95% CI approx.
quantile(gvi,c(0.025,0.975))
quantile(mvi,c(0.025,0.975))
```

6.1.2 Simulated continuous data of Table 4 from the NORTA Algorithm

```
# Simulation of multivariate count data using NORTARA package
# Loading extra packages
require(NORTARA)
# Marginal distributions:
# X1 ~ W(\beta = 0.7, alpha = 1)
# X2 ~ E(\mu = 2)
# X3 ~ LN(\mu = 0, sigma = 0.6)
# X4 ~ W(\beta = 1, alpha = 1)
# X5 ~ LN(\mu = 1, sigma = 1)
# X6 ~ W(\beta = 3, alpha = 1)
# Correlation matrix
COR<- matrix(c(1,-0.03,0.57,0.12,0.24,-0.57,-0.03,1,-0.05,-0.09,
               0.36,0.04,0.57,-0.05,1,0.35,0.27,-0.58,0.12,
               -0.09,0.35,1,0.18,0.04,0.24,0.36,0.27,0.18,
               1,0.06,-0.57,0.04,-0.58,0.04,0.06,1),6,6)

invcdfnames <- c("qweibull","qexp","qlnorm","qweibull","qlnorm","qweibull")
paramslists <- list(
  m1 = list(shape = 0.7, scale = 1),
  m2 = list(rate = 2),
  m3 = list(meanlog = 0, sdlog = 0.6),
  m4 = list(shape = 1, scale = 1),
  m5 = list(meanlog = 1, sdlog = 1),
  m6 = list(shape = 3, scale = 1)
)
data <- NORTARA::genNORTARA(n = 560, cor_matrix = COR,
```

```
invcdfnames = invcdfnames,  
paramslists = paramslists)
```

```
apply(data, 2, mean)
```

```
apply(data, 2, var)
```

6.1.3 Scripts in R for asymptotic variances and confidence intervals of GVI and MVI

```
# Reading the table of data
```

```
# Function that returns the vector composed by the averages of each variable
```

```
aux_esp<-function(data){  
  res<-c()  
  n=dim(data)  
  n=n[2]  
  for (i in 1:n){  
    res[i]=mean(data[,i])  
  }  
  return(res)  
}
```

```
# Function that returns the vector composed by the variances of each variable
```

```
aux_var<-function(data){  
  res<-c()  
  n=dim(data)  
  n=n[2]  
  for (i in 1:n){  
    res[i]=var(data[,i])  
  }  
  return(res)
```

```
}  
  
# Function that returns the matrix composed by the variances on its diagonal  
# and 0 else  
aux_diag_var<-function(data){  
  res<-c()  
  n=dim(data)  
  n=n[2]  
  for (i in 1:n){  
    res[i]=var(data[,i])  
  }  
  return(diag(res))  
}  
  
# Function that calculates the MVI  
MVI<-function(data){  
  tmp_E=aux_esp(data)  
  tmp_D=aux_diag_var(data)  
  res=(tmp_E**tmp_D**tmp_E)/(tmp_E**(tmp_E))^2  
  return(res)  
}  
  
# Function that calculates the GVI  
GVI<-function(data){  
  tmp_E=aux_esp(data)  
  tmp_C=cov(data)  
  res=(tmp_E**tmp_C**tmp_E)/(tmp_E**(tmp_E))^2  
  return(res)  
}  
  
# Fonction that returns the delta vector  
vec_delta<-function(data){
```

```
k=dim(data)
k=k[2]
res<-c()
tmp_E=aux_esp(data)
for (i in 1:k){
  res[i]=0
  for (j in 1:k){
    "res[i]=res[i]-2*(tmp_E[i])^3"
  }
  " res[i]=(res[i]+(2*tmp_E[j]*cov(data[,i],data[,j]))
-4*tmp_E[i]*(tmp_E[j])^2*GVI(data))/(tmp_E%*(tmp_E))^2"

  }
  for (i in 1:k){
    for(j in i:k){
      if (i==j){
        tmp=(tmp_E[i])^2/(tmp_E%*(tmp_E))^2
      }
      else{
        tmp=2*(tmp_E[i]*tmp_E[j])/(tmp_E%*(tmp_E))^2
      }
      res<-c(res,tmp)
    }
  }
  return(res)
}

# Function that returns the gamma3 matrix
mat_gamma3<-function(data){
```

```
k=dim(data)
k=k[2]
tmp_E=aux_esp(data)
res<-c()
for (j in 1:k){
  tmp_vec<-c()
  for (jj in 1:k){
    for (ll in jj:k){
      tmp=cov(data[,j],data[,jj]*data[,ll])
      tmp_vec=c(tmp_vec,tmp)
    }
  }
  res=rbind(res,tmp_vec)
}
rownames(res)=c()
return(res)
}

# Function that returns the gamma4 matrix
mat_gamma4<-function(data){
  k=dim(data)
  k=k[2]
  tmp_E=aux_esp(data)
  res<-c()
  for (j in 1:k){
    for (l in j:k){
      tmp_vec<-c()
      for (jj in 1:k){
        for (ll in jj:k){
```

```
tmp=cov(data[,j]*data[,1],data[,jj]*data[,11])
tmp_vec=c(tmp_vec,tmp)
}
}
res=rbind(res,tmp_vec)
}
}
rownames(res)=c()
return(res)
}

# Function that returns the gamma matrix
mat_gamma<-function(data){
tmp=cbind(cov(data),mat_gamma3(data))
tmp2=cbind(t(mat_gamma3(data)),mat_gamma4(data))
res=rbind(tmp,tmp2)
return(res)
}

# Function that calculates the variance of GVI
sigma_GVI<-function(data){
delta=vec_delta(data)
gamma=mat_gamma(data)
res=delta%%gamma%%delta
return(res)
}

vec_lambda<-function(data){ # Function that returns the lambda vector
k=dim(data)
k=k[2]
res<-c()
```

```

tmp_E=aux_esp(data)
tmp_V=aux_var(data)
for (i in 1:k){
  res[i]=0
  for (j in 1:k){

    "res[i]=res[i]-2*(tmp_E[i])^3"
  }
  "res[i]=(res[i]+(2*tmp_E[i]*tmp_V[i])
-4*tmp_E[i]*(tmp_E[j])^2*MVI(data))/(tmp_E%*(tmp_E))^2"
}

for (i in 1:k){
  res[i+k]=(tmp_E[i])^2/(tmp_E%*tmp_E)^2
}
return(res)
}

mat_pi3<-function(data){ # Function that returns the pi3 matrix
k=dim(data)
k=k[2]
tmp_E=aux_esp(data)
res<-c()
for (j in 1:k){
  tmp_vec<-c()
  for (jj in 1:k){
    tmp=cov(data[,j],data[,jj]*data[,jj])
    tmp_vec=c(tmp_vec,tmp)

```



```
    }
    res=rbind(res,tmp_vec)
  }
  rownames(res)=c()
  return(res)
}

mat_pi4<-function(data){ # Function that returns the pi4 matrix
  k=dim(data)
  k=k[2]
  tmp_E=aux_esp(data)
  res<-c()
  for (j in 1:k){
    tmp_vec<-c()
    for (jj in 1:k){
      tmp=cov(data[,j]*data[,j],data[,jj]*data[,jj])
      tmp_vec=c(tmp_vec,tmp)
    }
    res=rbind(res,tmp_vec)
  }
  rownames(res)=c()
  return(res)
}

mat_pi<-function(data){ # Function that returns the pi matrix
  tmp=cbind(cov(data),mat_pi3(data))
  tmp2=cbind(t(mat_pi3(data)),mat_pi4(data))
  res=rbind(tmp,tmp2)
  return(res)
}
```

```

sigma_MVI<-function(data){ # Function that calculates the variance of MVI
  lambda=vec_lambda(data)
  pi=mat_pi(data)
  res=lambda%%pi%%lambda
  return(res)
}

ind_tab<-function(data,n){ # Function that returns a table with
  #all the information
  k=length(n)
  res<-c()
  for (i in 1:k){
    tmp_vec<-c()
    tdata=data[1:n[i],]
    tmp_vec=c(det(cor(tdata)),sigma_GVI(tdata),sigma_MVI(tdata),GVI(tdata)
, qnorm(0.975)*sqrt(sigma_GVI(tdata)/n[i]),MVI(tdata),
  qnorm(0.975)*sqrt(sigma_MVI(tdata)/n[i]))
    res=rbind(res,tmp_vec)
  }
  rownames(res)=n
  colnames(res)=c("det(rho)","sigma2GVI","sigma2MVI","GVI +-", "E.t(int)",
  "MVI+-", "E.t(int)")
  return(res)
}

# For this function ind_tab, data is the data set and n a vector
# with values (sample size) to calculate the different indices
n=c(50,100,300,500,1000,3000,5000,10000)
ind_tab(data,n)

invcdfnames <- c("qweibull","qexp","qlnorm","qweibull","qlnorm","qweibull")

```

```
paramslists <- list(
  m1 = list(shape=0.7,scale=1),
  m2 = list(rate=2),
  m3 = list(meanlog=0,sdlog=0.6),
  m4 = list(shape=1,scale=1),
  m5 = list(meanlog=1,sdlog=1),
  m6 = list(shape=3,scale=1)
)
# Matrix of specified input correlation
COR<-matrix(c(1,-0.03,0.57,0.12,0.24,-0.57,-0.03,1,-0.05,-0.09,
              0.36,0.04,0.57,-0.05,1,0.35,0.27,-0.58,0.12,-0.09,
              0.35,1,0.18,0.04,0.24,0.36,0.27,0.18,1,0.06,-0.57,
              0.04,-0.58,0.04,0.06,1),ncol=6,nrow=6)
# Function which calculates the correlation matrix
signal<- BoundingRA(cor_matrix = COR, invcdfnames, paramslists)
data<-NORTARA::genNORTARA(n=10000, cor_matrix=COR, invcdfnames=invcdfnames,
paramslists = paramslists)
apply(data, 2, mean) # Function that applies the average of each column
apply(data, 2, var) # Function that applies the variance of each column
# Function that returns two matrix containing the GVI and MVI in n
#simulated replicates
# n is a vector with the same purpose as previously
n=c(50,100,300,500,1000,3000,5000,10000)
ind_tab(data,n)
#Function that returns two matrices gvi and mvi for n replicates
aux_box<-function(n,nrep){
mat_gvi=c()
mat_mvi=c()
```

```

nsimu=n[length(n)]
for (i in 1:nrep){
data <- NORTARA::genNORTARA(nsimu, cor_matrix = COR,
                             invcdfnames = invcdfnames,
                             paramslists = paramslists)

tmp=ind_tab(data,n)
mat_gvi=rbind(mat_gvi,tmp[,4])
mat_mvi=rbind(mat_mvi,tmp[,6])
}
return(list(mat_gvi,mat_mvi))
}

n=c(50,100,300,500,1000,3000,5000,10000)
data_box=aux_box(n,100) #Function that returns boxplots for 100 replicates
gvi=data_box[[1]]
mvi=data_box[[2]]
boxplot(gvi)
boxplot(mvi)

```

6.2 Transition from discret case to continuous case

- The transition from discrete to continuous is not simply. For instance, to obtain the reference exponential distribution $X \sim \mathcal{E}(\mu)$ as a weighted version of a positive continuous distribution Y with support $[0, \infty)$ one has : $Y^{1/w} \stackrel{d}{=} X$ where w is the positive exponential weight function.
- In the discrete case with the criterion of d'Alembert, if f is a probability mass function we can calculate the sum $1/f$.
- But, in the continuous case the density $1/f$ can not be integrated. Indeed, assuming that for $n \in \mathbb{N}_0$, $u_n = \int_n^{n+1} f(x)dx$, $u'_n = \int_n^{n+1} \sqrt{f(x)}dx$, $v_n = \int_n^{n+1} (1/f(x))dx$,

$v'_n = \int_n^{n+1} 1/(\sqrt{f(x)})dx$. From Schwarz one has $(u'_n)^2 \leq u_n < \infty$, $(v'_n)^2 \leq v_n < \infty$.

Again from Schwarz one can write : $1 = \left(\int_n^{n+1} \sqrt{f(x)}/(\sqrt{f(x)})dx \right)^2 \leq u_n v_n \implies$

$$1/u_n \leq v_n.$$

Since $\sum_n u_n < \infty$ and $\sum_n v_n < \infty$ then u_n and v_n tend towards zero and contradict

$$1/u_n \leq v_n.$$