

# Transform MCMC schemes for sampling intractable factor copula models <sup>\*</sup>

Cyril Bénézet<sup>†</sup>

Emmanuel Gobet <sup>‡</sup>

Rodrigo Targino<sup>§</sup>

September 3, 2021

## Abstract

In financial risk management, modelling dependency within a random vector  $\mathcal{X}$  is crucial, a standard approach is the use of a copula model. Say the copula model can be sampled through realizations of  $\mathcal{Y}$  having copula function  $C$ : had the marginals of  $\mathcal{Y}$  been known, sampling  $\mathcal{X}^{(i)}$ , the  $i$ -th component of  $\mathcal{X}$ , would directly follow by composing  $\mathcal{Y}^{(i)}$  with its cumulative distribution function (c.d.f.) and the inverse c.d.f. of  $\mathcal{X}^{(i)}$ . In this work, the marginals of  $\mathcal{Y}$  are not explicit, as in a factor copula model. We design an algorithm which samples  $\mathcal{X}$  through an empirical approximation of the c.d.f. of the  $\mathcal{Y}$ -marginals. To be able to handle complex distributions for  $\mathcal{Y}$  or rare-event computations, we allow Markov Chain Monte Carlo (MCMC) samplers. We establish convergence results whose rates depend on the tails of  $\mathcal{X}$ ,  $\mathcal{Y}$  and the Lyapunov function of the MCMC sampler. We present numerical experiments confirming the convergence rates and also revisit a real data analysis from financial risk management.

MSC 2010: 62H05, 60J22, 91G60

KEYWORDS: copula model, Markov chain Monte Carlo, sampling algorithm

## 1 Introduction

### 1.1 Statement of the problem

Let  $\mathcal{X}$  be a  $\mathbb{R}^d$ -valued random variable, we aim at computing statistics of the form

$$\mathbb{E}(g(\mathcal{X})).$$

---

<sup>\*</sup>Acknowledgement: this research has benefited from the support of the Chaire Stress Test, RISK Management and Financial Steering of the Ecole Polytechnique Foundation.

<sup>†</sup>Université Paris-Saclay, CNRS, Univ Évre and ENSIIE, Laboratoire de Mathématiques et Modélisation d'Évre, 91037, Évre, France. Email: cyril.benezet@ensiie.fr

<sup>‡</sup>Center for Applied Mathematics (CMAP), CNRS, Ecole polytechnique, Institut Polytechnique de Paris, France. Email: emmanuel.gobet@polytechnique.edu

<sup>§</sup>School of Applied Mathematics (EMAp), Getulio Vargas Foundation (FGV), Brazil. Email: rodrigo.targino@fgv.br

We focus on the case where sampling directly  $\mathcal{X}$  is not feasible. Namely, the marginal cumulative distribution functions (c.d.f.'s)  $F^{(i)}$  of  $\mathcal{X}^{(i)}$  are known, but the copula  $C$ , although existing owing to Sklar's Theorem (see [Joe14]), is not explicit. In summary, we assume the probabilistic model for  $\mathcal{X}$  is given by

$$\mathcal{X} = (\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(d)}) \sim C(F^{(1)}(\cdot), \dots, F^{(d)}(\cdot)). \quad (1.1)$$

We are investigating a framework where the copula  $C$  coincides with that of a vector  $\mathcal{Y} = \Phi(\mathcal{Z})$  with a measurable function  $\Phi : \mathbb{R}^D \rightarrow \mathbb{R}^d$  and a  $\mathbb{R}^D$ -valued random variable  $\mathcal{Z}$  which can be sampled, by Monte Carlo (MC) or Markov Chain Monte Carlo (MCMC) schemes. The marginals  $G^{(i)}$  of  $\mathcal{Y}^{(i)}$  may be not explicit or tractable: hence the following usual simulation scheme can not be applied (because at line 3,  $G^{(i)}$  is unknown).

---

**Algorithm 1: Usual sampling of  $\mathcal{X}$  through sampling of  $\mathcal{Z}$**  (see [CLV04, Th. 4.8 p.143])

---

**Input:**  $(F^{(i)})^{-1}$  the quantile of  $\mathcal{X}^{(i)}$ ,  $G^{(i)}$  the c.d.f. of  $\mathcal{Y}^{(i)}$

**Output:**  $\mathcal{X} = (\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(d)})$  sampled with the distribution (1.1)

- 1 Sample  $\mathcal{Z}$
  - 2 Compute  $\mathcal{Y} = \Phi(\mathcal{Z})$
  - 3 Get  $U = (U_1, \dots, U_d) = (G^{(1)}(\mathcal{Y}^{(1)}), \dots, G^{(d)}(\mathcal{Y}^{(d)}))$
  - 4 Set  $\mathcal{X}^{(i)} = (F^{(i)})^{-1}(U_i)$
- 

In the following we assume that the marginal c.d.f.'s  $G^{(i)}$  and  $F^{(i)}$  are continuous.

To overcome the unfeasibility of Algorithm 1, our strategy is to build empirically, using  $n$  samples of  $\mathcal{Y}$ , the c.d.f.  $G^{(i)}$  of  $\mathcal{Y}^{(i)}$ : this approximation will induce some error on  $G^{(i)}$  which will have an impact of the resulting generated sample  $(\hat{\mathcal{X}}_k : 1 \leq k \leq n)$ . The goal of this paper is to analyse the overall error as a function of  $n$ , by considering two criterions:

- the strong error in  $L_p$  norm (using  $n$  samples)  $\left| g(\mathcal{X}_n) - g(\hat{\mathcal{X}}_n) \right|_p$  (see Theorem 2.7),
- the statistical fluctuation  $\left| \frac{1}{N} \sum_{n=1}^N g(\hat{\mathcal{X}}_n) - \mathbb{E}(g(\mathcal{X})) \right|_p$  (see Corollary 2.9).

To achieve this research program, we have to cope with two major difficulties, which makes our contributions non trivial, interesting and new:

- In many applications, the samples  $(\mathcal{Y}_n, n \geq 1)$  may not be independent and may not have the target distribution of  $\mathcal{Y}$ . Indeed, sampling  $\mathcal{Y}$  may be possible only through a MCMC scheme, see Section 3 below.

Along the mathematical analysis of the approximation of  $G^{(i)}$ , the c.d.f. of  $\mathcal{Y}^{(i)}$ , we will have to measure how fast the sampled  $\mathcal{Y}_n$  will (weakly) converge to the target distribution and how well the statistical fluctuation of ergodic averages are controlled: we will be working under standard assumptions of geometric ergodicity (via a Lyapunov condition on the Markov kernel). We will provide simplified Lyapunov functions leading to geometric ergodicity, see Section A.2, some of them being seemingly new to the best of our knowledge. Fortunately (see Section A.2.2) these assumptions are also compatible with standard i.i.d. sampler.

- The quantile functions  $(F^{(i)})^{-1}$  are far from being Lipschitz bounded (e.g. for a Pareto distribution, it behaves as  $(1 - u)^{-1/\eta}$  as the quantile  $u \rightarrow 1$ , see Example 2.2). Hence the error on  $G^{(i)}$  will propagate highly non-linearly on the sample of  $\mathcal{X}$ : it may affect the global rate of convergence, which instead of being  $N^{-1/2}$  for usual MC procedure may be  $N^{-\zeta}$  with an order  $\zeta < \frac{1}{2}$  depending on the trade-off between the singularity/growth of the quantiles  $(F^{(i)})^{-1}$  and the Lyapunov function. Our numerical experiments confirm this trade-off (see Section 3).

To better understand the set of assumptions that will be used throughout this work, let us now give some examples. The first example shows that a wide family of interesting copulas are obtained as described above, but with unknown marginals. The second example exhibits some of the statistics we are interested in this work.

## 1.2 Examples

**Example 1.1** (Factor copula). *We consider the factor copulas, investigated by [OP17], which are the copulas for  $\mathcal{Y}$  obtained as follows. Take a vector of factors  $\mathcal{M} = (\mathcal{M}^{(1)}, \dots, \mathcal{M}^{(J)})$  and idiosyncratic errors  $\epsilon = (\epsilon^{(1)}, \dots, \epsilon^{(d)})$ ; the components of  $(\mathcal{M}^{(1)}, \dots, \mathcal{M}^{(J)}, \epsilon^{(1)}, \dots, \epsilon^{(d)})$  are independent and those of  $(\epsilon^{(i)})_{i=1}^d$  are identically distributed (thus i.i.d.).*

*Then, for each  $1 \leq i \leq d$ , set  $\mathcal{Y}^{(i)} := h^{(i)}(\mathcal{M}, \epsilon)$  for some measurable function  $h^{(i)} : \mathbb{R}^{J+d} \rightarrow \mathbb{R}$ . In the notations of our work, one can set  $D = J + d$ ,  $\mathcal{Z} := (\mathcal{M}^{(1)}, \dots, \mathcal{M}^{(J)}, \epsilon^{(1)}, \dots, \epsilon^{(d)})$  and  $\Phi^{(i)} := h^{(i)}$  ( $1 \leq i \leq d$ ).*

*More generally, one can also consider  $f : \mathbb{R}^D \ni \mathcal{Z} \mapsto (\mathcal{M}^{(1)}, \dots, \mathcal{M}^{(J)}, \epsilon^{(1)}, \dots, \epsilon^{(d)}) \in \mathbb{R}^{J+d}$  and  $\Phi^{(i)} := h^{(i)} \circ f$  ( $1 \leq i \leq d$ ). This more general formulation will be used in our second numerical example, see Section 3.2.*

*Several well known copulas can be retrieved through judicious choices of  $\mathcal{M}, \epsilon$  and  $\Phi$  (see Table 1). When  $\Phi^{(i)}$  is linear we have a particular sub-class of copulas, the linear factor copulas. Even in the linear case, depending on the distributions of  $\mathcal{M}$  and  $\epsilon$ , the marginals of  $\mathcal{Y}$  may not be explicitly known.*

*To the best of our knowledge, the idea of using a factor model to extract the dependence structure was first proposed in [MDCL13], but only explored for Gaussian copulas. An application can be found in [CMS15], where the authors use a linear, one-factor copula in the context of mortality modeling of multiple populations. The models tested in [CMS15] are a combination of Gaussian, student t and skew-t (see Appendix A.3) distributions for the factor and idiosyncratic shocks. A similar copula model is used in [BZ16], for modeling stock returns. In Section 3.2 we follow the example from [OP17, Section 4.2] and use a one-factor, skew-t-t copula.*

**Example 1.2** (Multivariate extreme risks). *We are concerned with statistics of  $\mathcal{X}$  when some components are extreme: it takes usually the generic form*

$$\mathbb{E} \left( g(\mathcal{X}) \mid \mathcal{X}^{(i)} \geq x^{(i)}, i \in \mathcal{I} \right), \quad \mathbb{E} \left( g(\mathcal{X}) \mathbf{1}_{\mathcal{X}^{(i)} \geq x^{(i)}, i \in \mathcal{I}} \right) \quad (1.2)$$

*for some subset  $\mathcal{I} \subset \{1, \dots, d\}$  and  $x^{(i)}$  possibly large.*

*Various quantities of interest take this form:*

Copula	$h^{(i)}(\mathcal{M}, \epsilon)$	$F_{\mathcal{M}}$	$F_{\epsilon}$
Normal	$\mathcal{M} + \epsilon^{(i)}$	$\mathcal{N}(0, \sigma_{\mathcal{M}}^2)$	$\mathcal{N}(0, \sigma_{\epsilon}^2)$
Student's $t$	$\mathcal{M}^{1/2} \epsilon^{(i)}$	$InvGa(\nu/2, \nu/2)$	$\mathcal{N}(0, \sigma_{\epsilon}^2)$
Skew $t$	$\lambda \mathcal{M} + \mathcal{M}^{1/2} \epsilon^{(i)}$	$InvGa(\nu/2, \nu/2)$	$\mathcal{N}(0, \sigma_{\epsilon}^2)$
Clayton	$(1 + \epsilon^{(i)}/\mathcal{M})^{-\alpha}$	$\Gamma(\alpha, 1)$	$Exp(1)$
Gumbel	$-(\log \mathcal{M}/\epsilon^{(i)})^{\alpha}$	$Stable(1/\alpha, 1, 1, 0)$	$Exp(1)$

Table 1: Special cases of known copulas as one factor copulas (adapted from [OP17]).

1. (Stable tail dependence function (s.t.d.f.)). *Assessing multivariate extreme risks is closely related to estimate the extremal dependence structure, which is described by the s.t.d.f.  $L$  given by*

$$\lim_{t \rightarrow +\infty} t \mathbb{P} \left[ 1 - F^{(1)}(\mathcal{X}^{(1)}) \leq x_1/t \text{ or } \dots \text{ or } 1 - F^{(1)}(\mathcal{X}^{(1)}) \leq x_d/t \right] = L(x_1, \dots, x_d),$$

$$\forall x_1 > 0, \dots, x_d > 0. \quad (1.3)$$

The existence of the above limit is discussed in [DF07, Chapter 7]. See [Hua92] for pioneering works about s.t.d.f., [dR93] for statistical estimation, [Fal19] for a recent account on multivariate extreme value theory.

Suppose we are able to generate a sample of  $\mathcal{X}$ , as in the goal of this work. Then an empirical version of (1.3) can be designed by taking  $t$  large; bias corrections (accounting for finite  $t$ ) are discussed in [BEGG16]. The resulting approximate s.t.d.f.  $L$  is then a linear combination of quantities like the RHS of (1.2).

2. (Tail dependence coefficients). When  $x^{(i)}$  are taken as  $a$ -quantiles, denoted in Risk Management context by  $x^{(i)} = VaR_a(\mathcal{X}^{(i)}) = (F^{(i)})^{-1}(a)$ , we obtain several measures of interest. For instance,

(a) Coefficients of upper and lower tail dependence ([MFE10, Section 7.2.4]):

$$\lambda_{ij}^u = \lim_{a \rightarrow 1^-} \mathbb{P} \left[ \mathcal{X}^{(i)} > VaR_a(\mathcal{X}^{(i)}) \mid \mathcal{X}^{(j)} > VaR_a(\mathcal{X}^{(j)}) \right] = \lim_{a \rightarrow 1^-} \frac{C_{ij}(1-a, 1-a)}{1-a},$$

$$\lambda_{ij}^l = \lim_{a \rightarrow 0^+} \mathbb{P} \left[ \mathcal{X}^{(i)} < VaR_a(\mathcal{X}^{(i)}) \mid \mathcal{X}^{(j)} < VaR_a(\mathcal{X}^{(j)}) \right] = \lim_{a \rightarrow 0^+} \frac{C_{ij}(a, a)}{a},$$

where  $C_{ij}$  is the copula between  $\mathcal{X}^{(i)}$  and  $\mathcal{X}^{(j)}$ . In Finance and Insurance one is particularly interested in  $\lambda_{ij}^u$ , which is understood as the probability of asset  $i$  having a very large loss given that asset  $j$  has had one itself.

(b) Coefficients of multivariate upper and lower tail dependence:

$$\lambda_{1,\dots,d}^u = \lim_{a \rightarrow 1^-} \mathbb{P} \left[ \mathcal{X}^{(1)} > VaR_a(\mathcal{X}^{(1)}) \mid \mathcal{X}^{(2)} > VaR_a(\mathcal{X}^{(2)}), \dots, \mathcal{X}^{(d)} > VaR_a(\mathcal{X}^{(d)}) \right],$$

$$\lambda_{1,\dots,d}^l = \lim_{a \rightarrow 0^+} \mathbb{P} \left[ \mathcal{X}^{(1)} < VaR_a(\mathcal{X}^{(1)}) \mid \mathcal{X}^{(2)} < VaR_a(\mathcal{X}^{(2)}), \dots, \mathcal{X}^{(d)} < VaR_a(\mathcal{X}^{(d)}) \right],$$

see [DR12] for explicit computations with an Archimedean copula.

3. (Semi-correlation). The upper and lower semi-correlations were introduced in [AC02] and [Gab05], respectively, as

$$\begin{aligned}\rho_{ij}^+ &= \text{Corr}(\mathcal{X}^{(i)}, \mathcal{X}^{(j)} \mid \mathcal{X}^{(i)} > 0, \mathcal{X}^{(j)} > 0), \\ \rho_{ij}^- &= \text{Corr}(\mathcal{X}^{(i)}, \mathcal{X}^{(j)} \mid \mathcal{X}^{(i)} < 0, \mathcal{X}^{(j)} < 0).\end{aligned}$$

Here, the above quantities are of the form (1.2) but with non-large  $x^{(i)}$ .

Explicit computations in the Gaussian case can be found in [Joe14, Section 2.17].

4. ( $k$ -Expected Shortfall). If  $\mathcal{X}$  is a vector of (negative) returns, a Risk Manager can be interested in computing the expected return on its  $i$ -th position conditional on  $k+1$  stocks in the market having negative returns above some threshold  $C$ , i.e.

$$(k - ES)^{(i)} = \mathbb{E} \left( \mathcal{X}^{(i)} \mid \left( \sum_{j=1}^d \mathbf{1}_{\{\mathcal{X}^{(j)} \geq C\}} \right) > k \right).$$

This measure has been studied in [OP17] and is also discussed in Section 3.2, below.

### 1.3 Organization of the paper

In Section 2, we define the model and related notations, provide the sampling algorithm, and state main convergence results under some assumptions. Extensions are also discussed. Section 3 is dedicated to numerical experiments, to illustrate the algorithm and the convergence rates derived in the previous section. Section 4 gathers the proofs of convergence results. We conclude in Section 5 and collect some technical results in Appendice A.

Throughout the paper, the c.d.f. of the standard Gaussian random variable  $N \sim \mathcal{N}(0, 1)$  is denoted by  $F_N(x) = \mathbb{P}[N \leq x]$ . We recall Mill's inequalities: for any  $x > 0$ ,

$$\frac{x}{x^2 + 1} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \leq F_N(-x) \leq \frac{1}{x} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}. \quad (1.4)$$

## 2 Main results

### 2.1 Setting

On a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , let  $\mathcal{Z}$  be a  $D$ -dimensional random vector ( $D \geq 1$ ) with distribution  $\pi$ , taking values in  $\mathcal{A}^{\mathcal{Z}} \subset \mathbb{R}^D$ .

For a measurable map  $\Phi = (\Phi^{(i)})_{i=1}^d : \mathcal{A}^{\mathcal{Z}} \rightarrow \mathbb{R}^d$ , let  $\mathcal{Y} := \Phi(\mathcal{Z})$  be a  $d$ -dimensional random vector. For each  $1 \leq i \leq d$ , we denote by  $G^{(i)}$  the  $i$ -th marginal c.d.f. of  $\mathcal{Y}$ , which is not

assumed to be known.

We set  $U := (U^{(i)})_{i=1}^d := (G^{(i)}(\mathcal{Y}^{(i)}))_{i=1}^d$  and  $\mathcal{X} := (\mathcal{X}^{(i)})_{i=1}^d := \left( (F^{(i)})^{-1}(U^{(i)}) \right)_{i=1}^d$  where, for each  $1 \leq i \leq d$ ,  $F^{(i)}$  is a fixed cumulative distribution function.

The relationships are summarized in the following diagram:

$$\mathcal{Z} = \begin{bmatrix} \mathcal{Z}^{(1)} \\ \vdots \\ \mathcal{Z}^{(D)} \end{bmatrix} \xrightarrow{\Phi} \mathcal{Y} = \begin{bmatrix} \mathcal{Y}^{(1)} = \Phi^{(1)}(\mathcal{Z}) \\ \vdots \\ \mathcal{Y}^{(d)} = \Phi^{(d)}(\mathcal{Z}) \end{bmatrix} \rightarrow U = \begin{bmatrix} U^{(1)} = G^{(1)}(\mathcal{Y}^{(1)}) \\ \vdots \\ U^{(d)} = G^{(d)}(\mathcal{Y}^{(d)}) \end{bmatrix} \rightarrow \begin{bmatrix} \mathcal{X}^{(1)} = (F^{(1)})^{-1}(U^{(1)}) \\ \vdots \\ \mathcal{X}^{(d)} = (F^{(d)})^{-1}(U^{(d)}) \end{bmatrix} = \mathcal{X}.$$

To compute statistics on  $\mathcal{X}$ , one needs to sample from  $\mathcal{Z}$  and perform the transforms above. One could, of course, use a crude Monte Carlo independent sample of  $\mathcal{Z}$  (and empirical c.d.f. approximations for  $G^{(i)}$ ,  $1 \leq i \leq d$ ) to obtain an (approximately) independent sample of  $\mathcal{X}$ . However, as shown in Example 1.2, we might be interested in computing conditional expectations with respect to a (rare) event  $A$ . A naive simulation scheme with acceptance-rejection in  $A$  is inefficient, especially if the event  $A$  has a small probability, see [AG07, Chapter 6]. In case of a rare-event regime, one is better off using dedicated numerical schemes, one of them being a MCMC sampler to approximate  $\mathcal{X}$  conditionally to  $A$ ; see [AB01], [GL15] for instance. More generally, direct independent MC sampling of  $\mathcal{Z}$  may not be feasible (because its distribution is too complex) and the only alternative may be to sample a Markov chain which on the long run has the same distribution of  $\mathcal{Z}$ . These situations motivate our general set-up where  $\mathcal{Z}$  is sampled (possibly approximately) using a MCMC scheme.

Let us mention already that the set of assumptions under which we work will allow to consider the independent sampler as well, see also Section A.2.2.

We now introduce the algorithm, making use of a MCMC sampler which invariant distribution is the distribution of  $\mathcal{Z}$ .

## 2.2 The algorithm

As explained above, we introduce a  $D$ -dimensional ergodic Markov chain  $\left( \mathcal{Z}_k = (\mathcal{Z}_k^{(i)}, 1 \leq i \leq D) \right)_{k \geq 0}$ ,  $\mathcal{A}^{\mathcal{Z}}$ -valued, with Markov kernel  $\mathcal{P}$  and invariant probability measure  $\pi$ .

As the marginal c.d.f.'s  $G^{(i)}$  of  $\mathcal{Y}^{(i)}$  are not explicitly known, we cannot directly use Algorithm 1. The aim of this work is to study the following algorithm, analogous to Algorithm 1, where

the  $G^{(i)}$ 's are computed along the sampling scheme.

---

**Algorithm 2: sampling of  $\mathcal{X}$  via sampling of  $\mathcal{Z}$  and approximation of  $G^{(i)}$**

---

**Input:**  $(F^{(i)})^{-1}$  the quantile of  $\mathcal{X}^{(i)}$ ,  $\mathcal{Z}_0 \in \mathcal{A}^{\mathcal{Z}}$  (deterministic)

**Output:**  $\mathcal{X}_k = (\mathcal{X}_k^{(1)}, \dots, \mathcal{X}_k^{(d)})$  for  $1 \leq k \leq n$ .

**for**  $k \leftarrow 1$  **to**  $n$  **do**

1     Sample  $\mathcal{Z}_k$  from  $\mathcal{P}(\mathcal{Z}_{k-1}, \cdot)$ .

2     Compute  $\mathcal{Y}_k = \Phi(\mathcal{Z}_k)$ .

3     Approximate and mollify  $G^{(i)}$  by

$$\tilde{G}_k^{(i)}(y) := \frac{1}{2\sqrt{k}} + \left(1 - \frac{1}{\sqrt{k}}\right) \left(\frac{1}{k} \sum_{\ell=1}^k \mathbf{1}_{\mathcal{Y}_\ell^{(i)} \leq y}\right).$$

4     Set  $V_k^{(i)} := \tilde{G}_k^{(i)}(\mathcal{Y}_k^{(i)})$  and  $V_k := (V_k^{(i)})_{i=1}^d$ .

5     Set  $\mathcal{X}_k^{(i)} := (F^{(i)})^{-1}(V_k^{(i)})$  and  $\mathcal{X}_k := (\mathcal{X}_k^{(i)})_{i=1}^d$ .

---

For all  $1 \leq k \leq n$  and  $1 \leq i \leq d$ , we also define  $W_k^{(i)} := G^{(i)}(\mathcal{Y}_k^{(i)})$  for  $W_k = (W_k^{(i)})_{i=1}^d$ .

Heuristically,  $\tilde{G}_k^{(i)}(y)$  converges to  $G^{(i)}(y)$  in  $L_p$ -sense as  $k \rightarrow +\infty$ , for any  $y$  (thanks to ergodic properties of the MCMC sampler) and in fact, the convergence is uniform in  $y$  (see Theorem 2.5). Consequently,  $V_k$  and  $W_k$  will be close to each other (see Theorem 2.7), and the latter is close in distribution to  $(G^{(i)}(\mathcal{Y}^{(i)}))_{i=1}^d$  as requested. The role of terms  $1/\sqrt{k}$  in  $\tilde{G}^{(i)}$  is to slightly mollify the empirical c.d.f. to discard values 0 and 1 before generating  $\mathcal{X}_k^{(i)}$ : this is somehow important for the algorithm stability (see Proposition 2.6) since the  $\mathcal{X}^{(i)}$  may be unbounded (resulting in singular  $(F^{(i)})^{-1}$  at 0 or 1).

Let  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  be a measurable map. Our interest lies in the approximation of  $\mathbb{E}(g(\mathcal{X}))$ . Using this algorithm, we consider the two approximations

$$\mathbb{E}(g(\mathcal{X}_n)) \quad \text{and} \quad \frac{1}{n} \sum_{k=1}^n g(\mathcal{X}_k).$$

Under the hypotheses that we introduce in the next subsection, we quantify the convergence rates of these two approximations.

We first obtain a weak convergence result, showing that the first approximation converges to  $\mathbb{E}(g(\mathcal{X}))$ , see Corollary 2.8. We moreover prove that a strong convergence theorem holds for the Monte Carlo averages, see Corollary 2.9.

## 2.3 Assumptions

(A-i) The transition kernel  $\mathcal{P}$  of the Markov chain  $(\mathcal{Z}_k : k \geq 0)$  satisfies the following:

- $\mathcal{P}$  is  $\Psi$ -irreducible aperiodic transition kernel on some measurable subset  $\mathcal{A}^{\mathcal{Z}} \subset \mathbb{R}^D$ ;

- there exist some constants  $0 \leq \rho^{(\mathcal{P})} < 1$ ,  $b < \infty$ ,  $\nu > \frac{b}{1-\rho^{(\mathcal{P})}}$  and a measurable Lyapunov function  $\mathcal{L} : \mathcal{A}^{\mathcal{Z}} \mapsto [1, \infty)$  such that  $\{\mathcal{L} \leq \nu\}$  is 1-small and satisfying to

$$\mathcal{P}\mathcal{L} \leq \rho^{(\mathcal{P})}\mathcal{L} + b. \quad (2.1)$$

(A-ii) The initial point of  $\mathcal{Z}_0 \in \mathcal{A}^{\mathcal{Z}}$  is deterministic.

(A-iii) There exists  $q_{\max} \in [-1, 0)$  such that for all  $q > q_{\max}$  and all  $i \in \{1, \dots, d\}$ , the map  $(G^{(i)} \circ \Phi^{(i)})^q + (1 - G^{(i)} \circ \Phi^{(i)})^q$  from  $\mathcal{A}^{\mathcal{Z}}$  to  $\mathbb{R}$  is bounded in  $\mathcal{L}$ -norm, i.e.

$$\sup_{\mathcal{A}^{\mathcal{Z}}} \frac{(G^{(i)} \circ \Phi^{(i)})^q + (1 - G^{(i)} \circ \Phi^{(i)})^q}{\mathcal{L}} < +\infty.$$

(A-iv) For each  $1 \leq i \leq d$ , the marginal c.d.f.  $G^{(i)}$  of  $\mathcal{Y}^{(i)}$  in the stationary regime is continuous.

(A-v) Let  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  such that

$$\varphi := g \circ \left( (F^{(1)})^{-1}, \dots, (F^{(i)})^{-1}, \dots, (F^{(d)})^{-1} \right) \quad (2.2)$$

is measurable on  $[0, 1]^d$ , continuous on  $(0, 1)^d$  and that there exists a slowly varying<sup>1</sup> function  $\ell : (0, 1] \rightarrow (0, \infty)$  at 0, which is bounded on any closed interval of  $(0, 1]$ , and a parameter  $0 \leq \gamma < -q_{\max}$  such that, for all  $(u, v) \in (0, 1)^d \times (0, 1)^d$ ,

$$|\varphi(u) - \varphi(v)| \leq \sum_{i=1}^d \frac{\ell(u_i \wedge v_i)|u_i - v_i|}{(u_i \wedge v_i)^{\gamma+1}} + \sum_{i=1}^d \frac{\ell(1 - u_i \vee v_i)|u_i - v_i|}{(1 - u_i \vee v_i)^{\gamma+1}}, \quad (2.3)$$

$$|\varphi(u)| \leq \sum_{i=1}^d \frac{\ell(u_i)}{u_i^\gamma} + \sum_{i=1}^d \frac{\ell(1 - u_i)}{(1 - u_i)^\gamma}. \quad (2.4)$$

**Remark 2.1.** (a) Assumption (A-i) is natural in our setting in order to use geometric convergence theorems which we recall in the Appendix, see Theorem A.1. Moreover, this assumption is satisfied for the Gaussian sampler (see Section A.2.1), which allows us to construct numerical examples in the spirit of Example 1.1. In addition, the independent sampler (sampling i.i.d.  $\mathcal{Z}$ ) also satisfies to this assumption. See Section A.2.2 for getting the Lyapunov function which is involved in the next assumptions.

(b) Assumption (A-iii) is a technical assumption which allows to show that, for each  $1 \leq i \leq d$  and  $k \geq 1$ , the random variable  $V_k^{(i)}$  have negative moments up to  $q_{\max}$ , see Proposition 2.6.

Note that for non-negative  $q$ , the inequality obviously holds since  $G^{(i)} \in [0, 1]$  and  $\mathcal{L}$  is bounded from below by 1. Interesting values of  $q$  are those in the interval  $(q_{\max}, 0)$ .

---

<sup>1</sup>According to [BGT87, Section 1.2],  $\ell$  is slowly varying in the Karamata sense if  $\frac{\ell(\lambda u)}{\ell(u)} \rightarrow 1$  as  $u \rightarrow 0$  for any  $\lambda > 0$ .



- (c) Assumption (A-iv) ensures that  $U^{(i)}$  is uniformly distributed for all  $1 \leq i \leq d$ .
- (d) Assumption (A-v) is a quantitative control for the behaviour of  $\varphi$  near  $u_i = 0$  and  $u_i = 1$ . In the case where  $g$  is Lipschitz-continuous, this is directly related to the heavyness of the tails of  $\mathcal{X}^{(i)}$ . This hypothesis is minimal in the sense that we include the marginals with the heaviest tails which are approximated using a MCMC sampler satisfying the previous assumptions. We have a particular view towards applications where  $\mathcal{X}^{(i)}$  is heavy tailed, see Example 3.2.

We now compute, for some given functions  $g$  and  $F^{(i)}$ , the number  $\gamma$  from Assumption (A-v). The particular example of a marginal Pareto distribution for  $\mathcal{X}$  is explored in details in Section 3.1.

**Example 2.2** (Pareto distribution with index  $\eta_i > 0$  for  $\mathcal{X}^{(i)}$ ). Take  $F^{(i)}(x) := 1 - \left(\frac{x}{\underline{x}^i}\right)^{-\eta_i}$  for  $x \geq \underline{x}^i > 0$ . Then  $(F^{(i)})^{-1}(u_i) = \underline{x}^i(1 - u_i)^{-1/\eta_i}$  for  $u_i \in (0, 1)$ . For smooth  $g$  (with bounded first derivative) and  $u = (u_1, \dots, u_d)$ , we have

$$|\partial_{u_i} \varphi(u)| = \left| ((F^{(i)})^{-1})'(u_i) \right| \left| \partial_i g \left( (F^{(1)})^{-1}(u_1), \dots, (F^{(d)})^{-1}(u_d) \right) \right| \leq \frac{|g'|_\infty}{\eta_i} \underline{x}^i (1 - u_i)^{-1/\eta_i - 1}.$$

This gives  $\gamma = 1/\min_{1 \leq i \leq d} \eta_i$ .

Had  $g$  been bounded with compact support, then we would assume

$$|\varphi(u) - \varphi(v)| \leq C \sum_{i=1}^d |u_i - v_i|, \quad |\varphi(u)| \leq C,$$

meaning that  $\gamma = -1$ . This simpler case, not covered by Assumption (A-v), is discussed in Section 2.5.

**Example 2.3** (Exponential distribution for  $\mathcal{X}^{(i)}$ ). Take  $F^{(i)}(x) = 1 - \exp(-x)$  for  $x \geq 0$ . Then  $(F^{(i)})^{-1}(u) = -\ln(1 - u)$  for  $u \in (0, 1)$ . For Lipschitz  $g$ , it corresponds to  $\gamma = 0$  and the slow varying function is like a log.

**Example 2.4** (Gaussian distribution for  $\mathcal{X}^{(i)}$ ). Take  $F^{(i)}(x) = \int_{-\infty}^x \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$ . Then  $(F^{(i)})^{-1}(u) \sim -\sqrt{2\ln(1/u)}$  for  $u \rightarrow 0$  (see [Dom03, Proposition 21]). For Lipschitz  $g$ , it corresponds to  $\gamma = 0$  and the slow varying function is like the square root of a log.

To summarize, as a rule of thumb, for Lipschitz  $g$ , the lighter the tails of  $X$ , the smaller the parameter  $\gamma$ , the better the convergence orders in next results.

## 2.4 Main results

From now on, we assume that Assumptions (A-i)–(A-v) from Section 2.3 are in force. We state the convergence of statistics from Algorithm 2. Owing to (2.2), note that

$$\varphi(U) = g(\mathcal{X}) \quad \text{and} \quad \varphi(V_k) = g(\mathcal{X}_k),$$

thus stating convergence on  $\varphi$  or  $g$  is the same up to changing the function variables, we prefer to consider  $\varphi$ .

**Theorem 2.5** (Uniform convergence of the c.d.f. of  $\mathcal{Y}$  in  $L_p$ -norm). *For any  $p \geq 1$ ,  $n \geq 1$  and  $i \in \{1, \dots, d\}$ , we have*

$$\left| \sup_{y \in \mathbb{R}} |\tilde{G}_n^{(i)}(y) - G^{(i)}(y)| \right|_p \leq C_{p, (2.5)} n^{-\frac{p}{2(p+1)}}, \quad (2.5)$$

for some finite constant  $C_{p, (2.5)}$ .

**Proposition 2.6** (Stability). *Let  $q > q_{\max}$  be fixed. For any  $i \in \{1, \dots, d\}$ , we have*

$$\sup_{n \geq 1} \mathbb{E} \left( [W_n^{(i)}]^q \right) < +\infty, \quad \sup_{n \geq 1} \mathbb{E} \left( [1 - W_n^{(i)}]^q \right) < +\infty,$$

and

$$\sup_{n \geq 1} \mathbb{E} \left( [V_n^{(i)}]^q \right) < +\infty, \quad \sup_{n \geq 1} \mathbb{E} \left( [1 - V_n^{(i)}]^q \right) < +\infty.$$

For  $n \geq 1$  and  $0 \leq k \leq d$ , define  $\chi_n^k := \left( \left( V_n^{(i)} \right)_{i=1}^k, \left( W_n^{(i)} \right)_{i=k+1}^d \right)$ . Let  $1 \leq p < \frac{-q_{\max}}{\gamma}$  (with the convention  $\frac{-q_{\max}}{0} = +\infty$  if  $\gamma = 0$ ). Then

$$\sup_{n \geq 1, 0 \leq k \leq d} \mathbb{E} \left( |\varphi(\chi_n^k)|^p \right) < +\infty.$$

**Theorem 2.7** (Strong approximation). *For all  $\iota > 0$  and any  $p \in [1, \frac{-q_{\max}}{\gamma})$ , there exists a constant  $C_{\iota, p} > 0$  such that, for any  $n \geq 1$ ,*

$$|\varphi(V_n) - \varphi(W_n)|_p \leq C_{p, \iota} n^{-\frac{1}{2p} + \frac{\gamma}{2|q_{\max}|} + \iota}.$$

**Corollary 2.8** (Weak convergence). *For all  $\iota > 0$ , there exists a constant  $C_\iota > 0$  such that, for any  $n \geq 1$ ,*

$$|\mathbb{E}(\varphi(V_n)) - \mathbb{E}(\varphi(U))| \leq C_\iota n^{-\frac{1}{2} + \frac{\gamma}{2|q_{\max}|} + \iota}.$$

**Corollary 2.9** (Convergence of Monte Carlo averages). *For all  $\iota > 0$  and for any  $p \geq 1$  satisfying  $p \vee 2 < \frac{|q_{\max}|}{\gamma}$ , there exists a positive constant  $C_{p, \iota}$  such that for any  $n \geq 1$ ,*

$$\left| \frac{1}{n} \sum_{k=1}^n \varphi(V_k) - \mathbb{E}(\varphi(U)) \right|_p \leq C_{p, \iota} n^{-\frac{1}{2p} + \frac{\gamma}{2|q_{\max}|} + \iota}.$$

The exponent  $\frac{\gamma}{2|q_{\max}|}$  in the convergence rate in Theorem 2.7, Corollary 2.8 and Corollary 2.9 is linked to the unboundedness of  $\varphi$  near the boundary of  $[0, 1]^d$  and to the minimal negative integrability of the random variables  $V_k^{(i)}$  ( $k \geq 1$  and  $1 \leq i \leq d$ ), see Proposition 2.6. When  $\gamma$  tends to  $|q_{\max}|$ , the growth rate for  $\varphi$  increases to the maximal allowed growth and this results in the convergence order going to 0: one can observe that if  $\gamma = |q_{\max}|$ , convergence cannot occur as  $g(\mathcal{X}) = \varphi(U)$  is not integrable in general.

In Theorem 2.7 and Corollary 2.9, we observe that for  $1 < p < \frac{|q_{\max}|}{\gamma}$ , the exponent  $\frac{1}{2p}$  in the convergence rate is also smaller than the expected  $\frac{1}{2}$ . This is also connected to the unboundedness of  $\varphi$ . From (2.4), if  $\gamma$  were to be negative, it is clear that  $\varphi$  would be bounded on  $[0, 1]^d$ , while its variations would still grow to infinity near the boundary of  $[0, 1]^d$ , see (2.3): this situation is somehow simpler and presented in extensions below.

## 2.5 Extensions

### 2.5.1 Negative $\gamma$

We introduce here a modification of the previous Assumption (A-v) to the case where  $\gamma$  is negative.

(v') There exists a slowly varying function  $\ell : (0, 1] \rightarrow (0, \infty)$  at 0, bounded on any closed interval of  $(0, 1]$ , and a parameter  $-1 \leq \gamma < 0$  such that, for all  $(u, v) \in (0, 1)^d \times (0, 1)^d$ ,

$$|\varphi(u) - \varphi(v)| \leq \sum_{i=1}^d \frac{\ell(u_i \wedge v_i)|u_i - v_i|}{(u_i \wedge v_i)^{\gamma+1}} + \sum_{i=1}^d \frac{\ell(1 - u_i \vee v_i)|u_i - v_i|}{(1 - u_i \vee v_i)^{\gamma+1}}, \quad (2.6)$$

$$|\varphi(u)| \leq \sum_{i=1}^d \ell(u_i) + \sum_{i=1}^d \ell(1 - u_i).$$

In this situation, the counterpart to Theorem 2.7, Corollary 2.8 and Corollary 2.9 is the following:

**Theorem 2.10.** (i) For all  $\iota > 0$  and any  $p \geq 1$ , there exists a constant  $C_{\iota, p} > 0$  such that, for any  $n \geq 1$ ,

$$|\varphi(V_n) - \varphi(W_n)|_p \leq C_{p, \iota} n^{-\frac{1}{2(p(\gamma+1)-\gamma)} + \iota}.$$

(ii) For all  $\iota > 0$ , there exists a constant  $C_\iota > 0$  such that, for any  $n \geq 1$ ,

$$|\mathbb{E}(\varphi(V_n)) - \mathbb{E}(\varphi(U))| \leq C_\iota n^{-\frac{1}{2} + \iota}.$$

(iii) For all  $\iota > 0$  and for any  $p \geq 1$ , there exists a positive constant  $C_{p, \iota}$  such that for any  $n \geq 1$ ,

$$\mathbb{E} \left( \left| \frac{1}{n} \sum_{k=1}^n \varphi(V_k) - \mathbb{E}(\varphi(U)) \right|^p \right)^{1/p} \leq C_{p, \iota} n^{-\frac{1}{2(p(\gamma+1)-\gamma)} + \iota}.$$

We indeed observe an improvement in the convergence rate. For  $\gamma = -1$ , we recover an expected  $\frac{1}{2}$  convergence rate in  $p$ -norm for all  $p \geq 1$ . Notice also that when  $\gamma \rightarrow 0^-$  in the previous Theorem, we recover the rate of  $\frac{1}{2p}$  from Theorem 2.7 and Corollary 2.9. The proof of the first statement of this theorem is quite similar to the proof of Theorem 2.7, and the convergence results (ii)–(iii) are deduced from (i) the same way Corollaries 2.8–2.9 are from Theorem 2.7. Thus we only provide the main differences for proving Theorem 2.10 (i) after the proof of Theorem 2.7, in Remark 4.2.

### 2.5.2 Conditional expectations

We now extend the previous results to the case where we want to approximate a conditional expectation of the form  $\mathbb{E}(g(\mathcal{X}) \mid A)$ , where the event  $A$  takes the form

$$A := \{\mathcal{Y} \in \mathcal{A}^{\mathcal{Y}}\} = \{\mathcal{Z} \in \mathcal{A}^{\mathcal{Z}}\}$$

with known sets  $\mathcal{A}^{\mathcal{Y}}$  and  $\mathcal{A}^{\mathcal{Z}}$  (recall that  $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$  are related to each other via the relation  $\mathcal{X}^{(i)} = (F^{(i)})^{-1}(G^{(i)}(\mathcal{Y}_i))$  and  $\mathcal{Y} = \Phi(\mathcal{Z})$ ). Observe that whenever a conditional distribution of  $\mathcal{X}$  is targeted, one needs the unconditional marginal c.d.f.'s of  $\mathcal{Y}$  to obtain a  $\mathcal{X}$ -sample. Let us design an algorithm coherent with these conditional expectations. The Bayes formula yields

$$G^{(i)}(y_i) = \mathbb{P}[\mathcal{Y}^{(i)} \leq y_i] = \mathbb{P}[\mathcal{Y}^{(i)} \leq y_i \mid A] \mathbb{P}[A] + \mathbb{P}[\mathcal{Y}^{(i)} \leq y_i \mid A^c] \mathbb{P}[A^c]; \quad (2.7)$$

therefore, to extend Algorithm 2, one needs to compute two conditional c.d.f.'s and the probability  $\mathbb{P}[A]$ . For these three quantities, combining three different sampling schemes is a priori necessary. For the ease of presentation, we may skip the simplest quantity, i.e.  $\mathbb{P}[A]$ : if  $A$  is a rare (or a very likely) event, one should use dedicated methods, like splitting methods combined with interacting particle system or MCMC methods, see [GL15] and references therein. For other types of  $A$ , usual MC methods are sufficient.

We now assume that  $\mathbb{P}[A]$  is accurately estimated during a first stage and we focus our discussion on the two conditional c.d.f.'s.. Assume that one can use a MCMC sampler for  $\mathcal{Z}$  with transition kernel  $\mathcal{P}$ , that the related Markov Chain is symmetric with respect to the distribution  $\pi$  of  $\mathcal{Z}$ : then, adding an acceptance-rejection step in the set  $\mathcal{A}^{\mathcal{Z}}$  allows to generate a Markov chain with stationary distribution given by the restriction of  $\pi$  to  $\mathcal{A}^{\mathcal{Z}}$  (see [CDFG12, Proposition 1]). In the case of the preconditioned Crank–Nicolson sampler as described in Section A.2.1, assuming (A.4), the resulting Markov chain satisfies again to Assumption (A-i) with some quadratic exponential Lyapunov functions, see Proposition A.2. Now, to get the two conditional c.d.f.'s in (2.7), one can run synchronously two Markov chains, one with

acceptance-rejection in  $\mathcal{A}^{\mathcal{Z}}$ , the other in  $(\mathcal{A}^{\mathcal{Z}})^c$ . Algorithm 2 becomes as follows.

---

**Algorithm 3: sampling of  $\mathcal{X} \mid A$  via sampling of  $\mathcal{Z}$  conditionally to  $\mathcal{Z} \in \mathcal{A}^{\mathcal{Z}}$  and  $\mathcal{Z} \in (\mathcal{A}^{\mathcal{Z}})^c$**

---

**Input:**  $(F^{(i)})^{-1}$  the quantile of  $\mathcal{X}^{(i)}$ ,  $\mathcal{Z}_{0,A} \in \mathcal{A}^{\mathcal{Z}}$ ,  $\mathcal{Z}_{0,A^c} \in (\mathcal{A}^{\mathcal{Z}})^c$

**Output:**  $\mathcal{X}_k = (\mathcal{X}_k^{(1)}, \dots, \mathcal{X}_k^{(d)})$  for  $1 \leq k \leq n$ .

**for**  $k \leftarrow 1$  **to**  $n$  **do**

1     Sample  $\mathcal{Z}_{k,A}$  from  $\mathcal{P}(\mathcal{Z}_{k-1,A}, \cdot)$  and accept the proposal if it is in  $\mathcal{A}^{\mathcal{Z}}$ .

2     Compute  $\mathcal{Y}_{k,A} = \Phi(\mathcal{Z}_{k,A})$ .

3     Sample  $\mathcal{Z}_{k,A^c}$  from  $\mathcal{P}(\mathcal{Z}_{k-1,A^c}, \cdot)$  and accept the proposal if it is in  $(\mathcal{A}^{\mathcal{Z}})^c$ .

4     Compute  $\mathcal{Y}_{k,A^c} = \Phi(\mathcal{Z}_{k,A^c})$ .

5     Approximate and mollify  $G^{(i)}$  by

$$\tilde{G}_k^{(i)}(y) := \frac{1}{2\sqrt{k}} + \left(1 - \frac{1}{\sqrt{k}}\right) \left( \left( \frac{1}{k} \sum_{\ell=1}^k \mathbf{1}_{\mathcal{Y}_{\ell,A}^{(i)} \leq y} \right) \mathbb{P}[A] + \left( \frac{1}{k} \sum_{\ell=1}^k \mathbf{1}_{\mathcal{Y}_{\ell,A^c}^{(i)} \leq y} \right) \mathbb{P}[A^c] \right).$$

6     Set  $V_k^{(i)} := \tilde{G}_k^{(i)}(\mathcal{Y}_{k,A}^{(i)})$  and  $V_k := (V_k^{(i)})_{i=1}^d$ .

7     Set  $\mathcal{X}_k^{(i)} := (F^{(i)})^{-1}(V_k^{(i)})$  and  $\mathcal{X}_k := (\mathcal{X}_k^{(i)})_{i=1}^d$ .

---

The output of the algorithm enables to approximate  $\mathbb{E}(g(X) \mid A)$  by  $\frac{1}{n} \sum_{k=1}^n g(\mathcal{X}_k)$ . Let us conclude with few comments.

- If  $A$  and  $A^c$  are not rare events, one could directly compute the c.d.f. of  $\mathcal{Y}^{(i)}$  as in Algorithm 2, and keep only the samples  $\mathcal{X}_k$  in  $A$  to compute the final empirical mean. But in case of a rare  $A$ , this would be very inefficient, since only a few samples  $\mathcal{X}_k$  would be relevant. Also, the estimation of  $\tilde{G}^{(i)}$  would be presumably very inaccurate in the region of interest, whence two sampling schemes as in Algorithm 3.
- In the above algorithm, there is no reason for which Markov kernels for proposing  $\mathcal{Z}_{k,A}$  and  $\mathcal{Z}_{k,A^c}$  should be the same, it was done just for the purpose of concise explanation.
- The convergence analysis of Algorithm 3 works really the same as for Algorithm 2, and yields similar convergence rates, details are left to the reader.

### 3 Numerical experiments

#### 3.1 Pareto independent marginals with pCN sampler

As a toy example, we want to compute statistics of a random vector  $\mathcal{X}$  with independent marginals with c.d.f.  $F^{(i)}$ ,  $1 \leq i \leq d$ . Using our machinery and especially Corollary 2.8, for a measurable function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ , we approximate  $\mathbb{E}(g(\mathcal{X}))$  with

$$\mathbb{E} \left( \varphi \left( \tilde{G}_n^{(1)}(\mathcal{Y}_n^{(1)}), \dots, \tilde{G}_n^{(d)}(\mathcal{Y}_n^{(d)}) \right) \right), \quad (3.1)$$

where  $n \geq 1$  is fixed. In this example,  $D = d$  and  $\Phi$  is the identity function on  $\mathbb{R}^d = \mathbb{R}^D$ , and  $(\mathcal{Z}_k, k \geq 0) = (\mathcal{Y}_k, k \geq 0)$  is a Markov chain which stationary distribution has independent copula.

In practice, we approximate the expectation (3.1) with  $M \geq 1$  independent macro-runs. If  $\{(\mathcal{Y}_k^m)_{k=0,\dots,n}, 1 \leq m \leq M\}$  is a  $M$ -sample drawn with the Markov chain  $(\mathcal{Y}_k)_{k \geq 0}$ , we compute

$$e_n^M := \frac{1}{M} \sum_{m=1}^M \varphi \left( \tilde{G}_n^{(1),m}(\mathcal{Y}_n^{(1),m}), \dots, \tilde{G}_n^{(d),m}(\mathcal{Y}_n^{(d),m}) \right). \quad (3.2)$$

In view of Corollary 2.9, we also approximate  $\left| \mathbb{E}(g(\mathcal{X})) - \frac{1}{n} \sum_{k=1}^n \varphi \left( G^{(1)}(\mathcal{Y}_k^{(1)}), \dots, G^{(d)}(\mathcal{Y}_k^{(d)}) \right) \right|_p$  with

$$E_n^M := \left( \frac{1}{M} \sum_{m=1}^M \left| \mathbb{E}(g(\mathcal{X})) - \frac{1}{n} \sum_{k=1}^n \varphi \left( \tilde{G}_k^{(1),m}(\mathcal{Y}_k^{(1),m}), \dots, \tilde{G}_k^{(d),m}(\mathcal{Y}_k^{(d),m}) \right) \right|^p \right)^{\frac{1}{p}}. \quad (3.3)$$

We consider here the case where  $\mathcal{X}$  has Pareto marginals on  $[1, \infty)$  with index  $\eta_i > 0$ , so that  $F^{(i)}(x) = 1 - x^{-\eta_i}$  for all  $1 \leq i \leq d$  (see Example 2.2), and we consider the function  $g$  defined by

$$g(x_1, \dots, x_d) := \sum_{i=1}^d x_i.$$

Since  $g$  is smooth with bounded first derivative, we obtain (see Example 2.2) that  $\varphi$  satisfies to Assumption (A-v) with  $\gamma = 1/\min_{1 \leq i \leq d} \eta_i$ . Moreover, we compute easily  $\mathbb{E}(g(\mathcal{X})) = \sum_{i=1}^d \frac{\eta_i}{\eta_i - 1}$  for the reference value.

The Markov chain  $(\mathcal{Y}_n)_{n \geq 0}$  we consider is the preconditioned Crank-Nicolson (pCN) sampler [CRSW13], defined by

$$\begin{aligned} \mathcal{Y}_k &= \kappa \mathcal{Y}_{k-1} + \varepsilon_k, \quad k \geq 1, \\ \mathcal{Y}_0 &\text{ deterministic,} \end{aligned}$$

where  $\kappa \in (-1, 1)$  and  $(\varepsilon_k)_{k \geq 1}$  is an i.i.d. sequence such that  $\varepsilon_k \stackrel{d}{=} \mathcal{N}(0, (1 - \kappa^2)I_d)$ .  $\mathcal{Y}_0$  is deterministic so Assumption (A-ii) is satisfied, and its stationary distribution is  $\mathcal{N}(0, I_d)$ , with independent and continuous marginals, so that (A-iv) is also verified in this context. It is shown in Subsection A.2.1 that its associated transition kernel satisfies to Assumption (A-i) with the Lyapunov functions from Proposition A.2, and we now show that Assumption (A-iii) is satisfied with every  $q_{\max} > -1$ , hence one can take the limit  $q_{\max} \rightarrow -1$  in the convergence rates we obtain in Corollary 2.8 and Corollary 2.9.

**Proposition 3.1.** *For all  $q_{\max} \in (-1, 0)$ , there is a  $s \in (0, \frac{1}{2})$  such that Assumptions (A-i)-(A-iii) are satisfied for the preconditioned Crank-Nicolson sampler with  $\mathcal{L}(y) = e^{sy^2}$ .*

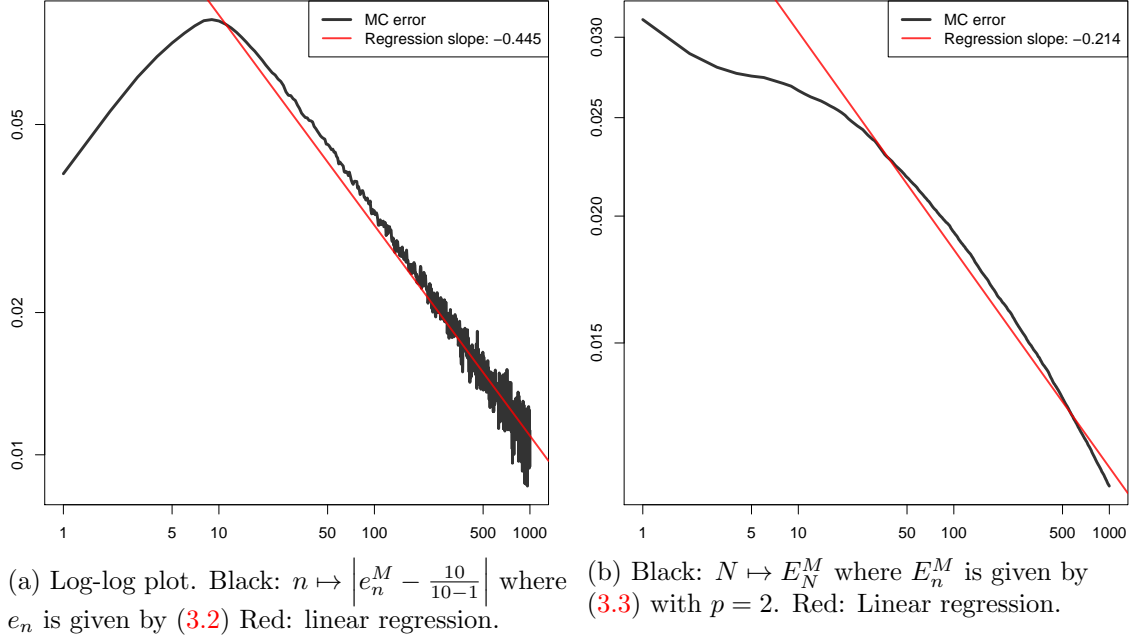


Figure 1: Log-log plots of empirical errors.

*Proof.* Let  $0 > q > q_{\max} > -1$  and  $1 \leq i \leq d$ . Use the lower bound of Mill's inequalities (1.4) for  $G^{(i)}(x) = F_N(x)$ : for  $-\frac{q}{2} < s < \frac{1}{2}$ , and using  $G^{(i)}(x) = 1 - G^{(i)}(-x)$ , we get

$$\begin{aligned} \sup_{y \in \mathbb{R}^d} \frac{G^{(i)}(y_i)^q + (1 - G^{(i)}(y_i))^q}{\mathcal{L}(y)} &= \sup_{x \geq 0} \frac{G^{(i)}(x)^q + (1 - G^{(i)}(x))^q}{e^{sx^2}} \\ &\leq \sup_{x \geq 0} \left( 2^{-q} e^{-sx^2} + (2\pi)^{-\frac{q}{2}} \left( \frac{x}{x^2 + 1} \right)^q e^{-sx^2 - qx^2/2} \right) < +\infty. \end{aligned}$$

□

We consider the case  $d = 1$  with  $\eta_1 = 10$  so that  $\gamma = 0.1$ . The Gaussian sampler is considered with  $\kappa = 0.5$  and  $\mathcal{Y}_0^{(1)} = 1000$ .

We approximate  $\mathbb{E}(g(\mathcal{X}))$  with (3.2) with  $1 \leq n \leq 1000$  and  $M = 10000$ .

Corollary 2.8 gives that the theoretical order of convergence for the weak error is  $\frac{1}{2} - \frac{\gamma}{2|q_{\max}|} = 0.45$ : this is confirmed in practice, as is shown in Figure 1a.

Corollary 2.9 states that the theoretical order of convergence for the time-average in  $L_p$ -norm is  $\frac{1}{2p} - \frac{\gamma}{2|q_{\max}|}$ : for  $p = 2$ , it equals 0.2 (in the limit  $q_{\max} \rightarrow -1^-$ ), very close to the empirical one (see Figure 1b).

### 3.2 kES for factor skew t-t copula with t marginals

We now follow the example in [OP17, Section 4.2] and compute the kES for the constituents of the S&P 100 index using the same range of dates: from April 2008 to December 2010.

Table 2: Marginal parameters for the losses (negative of the returns).

<b>Ticker</b>	<b>m</b> ( $\times 100$ )	<b>s</b> ( $\times 100$ )	<b>d.f.</b>	<b>SIC</b>	<b>Ticker</b>	<b>m</b> ( $\times 100$ )	<b>s</b> ( $\times 100$ )	<b>d.f.</b>	<b>SIC</b>	<b>Ticker</b>	<b>m</b> ( $\times 100$ )	<b>s</b> ( $\times 100$ )	<b>d.f.</b>	<b>SIC</b>
AA	-0.01	2.67	2.95	3	F	-0.09	2.71	2.72	3	NOV	-0.21	2.52	2.62	3
AAPL	-0.18	1.75	3.46	3	FCX	-0.21	3.00	3.28	1	NSC	-0.09	1.90	3.65	6
ABT	-0.01	1.05	3.67	2	FDX	-0.02	2.09	4.43	4	NYX	0.03	2.15	2.21	6
AEP	-0.02	1.14	2.98	4	GD	0.01	1.48	3.65	3	ORCL	-0.03	1.49	3.16	7
ALL	-0.02	1.52	1.77	6	GE	-0.02	1.68	2.22	3	OXY	-0.14	1.90	2.62	1
AMGN	0.02	1.16	2.49	2	GILD	0.07	1.40	3.23	2	PEP	0.00	0.95	2.99	2
AMZN	-0.01	2.06	3.10	7	GOOG	-0.05	1.71	6.13	7	PFE	0.02	1.38	3.56	2
AVP	-0.01	1.67	2.88	2	GS	-0.00	1.85	2.16	6	PG	-0.03	0.88	2.43	2
AXP	-0.04	2.38	2.60	6	HAL	-0.17	2.39	3.24	1	PM	-0.02	1.23	3.27	2
BA	-0.02	1.85	3.87	3	HD	0.01	1.64	3.30	5	QCOM	0.01	1.55	2.92	3
BAC	0.14	2.60	1.82	6	HNZ	-0.05	0.86	2.52	2	RF	0.14	3.52	2.30	6
BAX	-0.06	1.14	3.63	3	HON	-0.05	1.70	3.52	3	RTN	0.03	1.32	3.91	3
BK	0.11	1.88	1.85	6	HPQ	-0.06	1.37	2.73	3	S	0.25	3.10	2.65	4
BMJ	-0.01	1.19	2.96	2	IBM	-0.06	1.11	2.86	3	SLB	-0.05	2.20	3.31	1
C	0.12	2.78	1.91	6	INTC	-0.02	1.73	3.92	3	SO	-0.03	0.82	2.74	4
CAT	-0.07	2.03	3.39	3	JNJ	-0.03	0.70	2.28	2	T	0.02	1.09	2.57	4
CL	-0.03	0.93	2.62	2	JPM	0.09	2.26	2.18	6	TGT	0.01	1.56	2.49	5
CMCSA	0.01	1.77	3.00	4	KO	-0.04	0.95	2.61	2	TWX	0.00	1.72	2.91	7
COF	0.06	2.81	2.23	6	LMT	0.00	1.44	5.00	3	TXN	-0.11	1.83	4.64	3
COP	-0.07	1.49	2.38	2	LOW	0.03	1.94	3.96	5	UNH	-0.01	1.90	2.83	6
COST	-0.01	1.23	3.24	5	MA	-0.06	1.97	3.09	6	UPS	0.00	1.29	2.66	4
CPB	-0.06	0.93	3.16	2	MCD	-0.06	1.09	3.68	5	USB	-0.00	2.14	2.26	6
CSCO	-0.03	1.52	2.80	3	MDT	-0.02	1.23	2.89	3	UTX	-0.03	1.38	3.23	3
CVS	-0.02	1.47	3.42	5	MET	0.01	2.19	1.80	6	VZ	-0.03	1.10	2.57	4
CVX	-0.07	1.30	2.56	2	MMM	-0.06	1.27	3.18	3	WFC	0.14	2.31	1.93	6
DD	-0.03	2.24	2.92	2	MO	-0.12	0.97	2.91	2	WMB	-0.13	2.10	2.71	4
DIS	0.02	1.52	2.73	7	MRK	-0.04	1.43	2.98	2	WMT	-0.03	0.95	3.01	5
DVN	-0.05	2.08	3.16	1	MS	0.01	2.37	2.00	6	WY	-0.05	2.27	3.48	2
ETR	0.04	1.24	3.30	4	MSFT	-0.03	1.38	2.60	7	XOM	0.02	1.13	2.38	2
EXC	0.05	1.31	2.70	4	NKE	-0.07	1.49	2.90	3	XRX	0.03	2.04	3.18	3



### 3.2.1 The model

Let us denote the **negative** daily returns (i.e., the losses) of the  $i$ -th stock by  $\mathcal{X}^{(i)}$ . We assume that

$$\mathcal{X}^{(i)} \sim t_{\nu_i}(m_i, s_i),$$

i.e., a  $t$  distribution with  $\nu_i$  degrees of freedom, location  $m_i$  and scale parameter  $s_i$ . The stocks tickers and their marginal parameters can be found in Table 2.<sup>2</sup>

We assume the dependence structure to be given by the block equidependence copula of [OP17]:

$$\begin{aligned} \mathcal{Y}^{(i)} &= \beta_{S(i)}\mathcal{M}^{(0)} + \gamma_{S(i)}\mathcal{M}^{(S(i))} + \epsilon^{(i)}, \quad i = 1, \dots, d, \quad S(i) \in \{1, \dots, J-1\}, \\ \mathcal{M}^{(0)} &\sim \text{skew } t(\nu, \lambda), \\ \mathcal{M}^{(S)} &\stackrel{iid}{\sim} t(\nu), \quad S = 1, \dots, J-1, \quad \text{with } \mathcal{M}^{(S)} \perp\!\!\!\perp \mathcal{M}^{(0)}, \\ \epsilon^{(i)} &\stackrel{iid}{\sim} t(\nu), \quad i = 1, \dots, d, \quad \epsilon^{(i)} \perp\!\!\!\perp \mathcal{M}^{(j)}, \end{aligned}$$

where  $S(i)$  denotes the industry group, extracted from the first digit of the SIC (Standard Industrial Classification) for the  $i$ -th stock;  $\mathcal{M}^{(0)}$  is a common (market-wide) factor and  $\mathcal{M}^{(S)}$  the sector specific factor. Using the notation from Example 1.1 the dimensions are  $d = 90$ ,  $J = 7 + 1$  and  $D = 98$ . The copula parameters are given in Table 3.

Parameter	Value	Parameter	Value	Parameter	Value
		$\beta_1$	1.25	$\gamma_1$	1.09
		$\beta_2$	0.88	$\gamma_2$	0.22
		$\beta_3$	1.03	$\gamma_3$	0.17
		$\beta_4$	0.91	$\gamma_4$	0.27
		$\beta_5$	0.94	$\gamma_5$	0.55
		$\beta_6$	1.07	$\gamma_6$	0.57
		$\beta_7$	1.12	$\gamma_7$	0.39
$\nu$	10.08				
$\lambda$	-0.22				

Table 3: Copula parameters

To illustrate the flexibility of the factor copula model, we present the coefficients of lower (Table 4) and upper (Table 5) tail dependence, computed using [OP17, Proposition 2]. From these statistics one can see stocks in different sectors can have substantially different upper and lower tail dependence, a feature that would be impossible to observe in any elliptically symmetric copula (such as the Gaussian and the  $t$  copulas).

In Figure 2 we show the effect of this flexibility in a graphical way. Each plot is related to either two assets (Freeport-McMoRan, Inc. – FCX – and Halliburton Company – HAL) in

<sup>2</sup>The daily returns were collected from Yahoo Finance using the R library `quantmod` [RUT<sup>+</sup>20]. Even though 10 stocks from the index couldn't be retrieved (BHI, DELL, DOW, EMC, KFT, MON, NWSA, SLE, WAG) by `quantmod`, we proceed to model the remaining 90 stocks, with no material loss for the numerical exercise below.

	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6	Sector 7
Sector 1	0.99	0.70	0.92	0.75	0.81	0.94	0.96
Sector 2		0.70	0.70	0.70	0.70	0.70	0.70
Sector 3			0.92	0.75	0.81	0.92	0.92
Sector 4				0.75	0.75	0.75	0.75
Sector 5					0.81	0.81	0.81
Sector 6						0.94	0.94
Sector 7							0.96

Table 4: Lower tail dependence for the factor copula of Section 3.2

	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6	Sector 7
Sector 1	0.33	0.02	0.07	0.02	0.03	0.09	0.14
Sector 2		0.02	0.02	0.02	0.02	0.02	0.02
Sector 3			0.07	0.02	0.03	0.07	0.07
Sector 4				0.02	0.02	0.02	0.02
Sector 5					0.03	0.03	0.03
Sector 6						0.09	0.09
Sector 7							0.14

Table 5: Upper tail dependence for the factor copula of Section 3.2

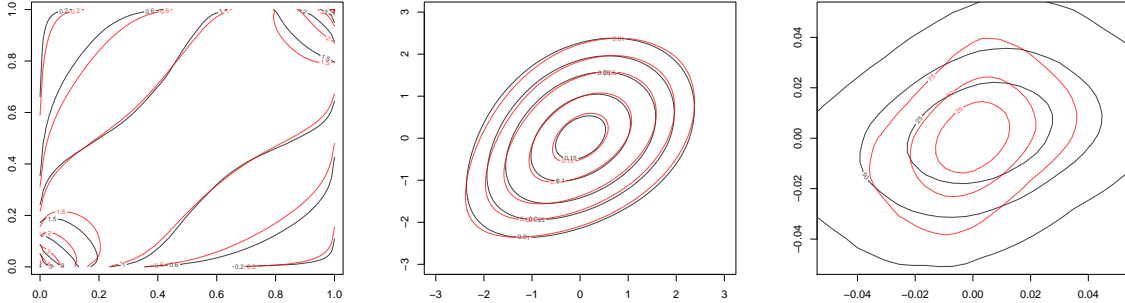


Figure 2: First column: Copula density for (black) two assets in sector 1 and (red) one asset from sector 7 and one from sector 2; Second column: Joint density for (black) two assets in Sector 1 (red) one asset from sector 7 and one from sector 2, assuming standard Gaussian marginals; Third column: Joint density of (black) two assets in sector 1 (FCX and HAL) and (red) one asset from sector 7 (DIS) and one from sector 2 (AVP).

sector 1 (black curves) or one asset (Walt Disney Co – DIS) in sector 7 and the other one (Avcorp Industries Inc. – AVP) in sector 2 (red curves), which are the two combinations with

most distinctive tail coefficients. On the left-most image we show the contour plots of the copula densities, where the differences in the tails can be readily seen. In the middle figure we show the contour plots of joint distributions with the same copulas as in the previous plot, but transformed to have standard Gaussian marginals. Although visually less pronounced, one can still see some difference between the lower joint tails (when losses are very negative, i.e., both stocks present high positive returns) and the upper joint tails of these two distributions. Finally, the right-most plot shows the empirical joint density for the two pairs of stocks. Throughout this example we denote by  $G_{\nu,\lambda}(x) := \mathbb{P}[\mathcal{M}^{(0)} \leq x]$  and  $G_\nu(x) := \mathbb{P}[\mathcal{M}^{(1)} \leq x]$ , respectively, the c.d.f.'s for the skew- $t(\nu, \lambda)$  and the  $t(\nu)$  distributions. Our interest is to compute, for some threshold  $C$ , the kES (see Example 1.2 item 4) when all losses are larger than  $C$ :

$$\begin{aligned} (k - \text{ES})^{(i)} &= \mathbb{E} \left( \mathcal{X}^{(i)} \left| \left( \sum_{j=1}^d \mathbf{1}_{\{\mathcal{X}^{(j)} \geq C\}} \right) > d-1 \right. \right) \\ &= \mathbb{E} \left( \mathcal{X}^{(i)} \left| \mathcal{X}^{(1)} > C, \dots, \mathcal{X}^{(d)} > C \right. \right). \end{aligned}$$

We follow the strategy outlined in Section 2.5.2, which is now discussed in details.

- *Interpretation of  $A$  in terms of  $\mathcal{Z}$ .*

Note that, since each  $F^{(i)}$  is continuous,  $C$  is the quantile at level  $a^{(i)} := F^{(i)}(C)$  for the distribution of  $\mathcal{X}^{(i)}$ . Since  $G^{(i)}$  is also continuous and  $F^{(i)}$  and  $G^{(i)}$  are non-decreasing, we obtain, for  $y^{(i)}$  the quantile at level  $a^{(i)}$  for the distribution of  $\mathcal{Y}^{(i)}$ ,

$$A = \bigcap_{1 \leq i \leq d} \{\mathcal{X}^{(i)} \geq C\} = \bigcap_{1 \leq i \leq d} \{\mathcal{Y}^{(i)} \geq y^{(i)}\} = \bigcap_{1 \leq i \leq d} \{\Phi^{(i)}(\mathcal{Z}) \geq y^{(i)}\},$$

where the last equality follows from the relationship between  $\mathcal{Z}$  and  $\mathcal{Y}$ .

In the sequel we use the following definitions:

$$\{\mathcal{Y} \in \mathcal{A}^{\mathcal{Y}}\} := \bigcap_{1 \leq i \leq d} \{\mathcal{Y}^{(i)} \geq y^{(i)}\} \text{ and } \{\mathcal{Z} \in \mathcal{A}^{\mathcal{Z}}\} := \bigcap_{1 \leq i \leq d} \{\Phi^{(i)}(\mathcal{Z}) \geq y^{(i)}\}.$$

- *Approximating  $\mathbb{P}[A]$ .*

As discussed in Section 2.5.2, we assume that  $\mathbb{P}[A]$  is estimated in a preliminary step. Here we use a crude MC procedure for  $\mathcal{Y}$  to estimate  $\mathbb{P}[A]$  (see details below).

- *Approximating  $\mathbb{P}[\mathcal{Y}^{(i)} \leq \cdot \mid A^c]$ .*

As we are interested in studying the proposed methodology for events  $A$  with low probability, we don't create a Markov Chain to sample from  $\mathcal{Y} \mid \mathcal{Y} \in (\mathcal{A}^{\mathcal{Y}})^c$  (or, equivalently, from  $\mathcal{Z} \mid \mathcal{Z} \in (\mathcal{A}^{\mathcal{Z}})^c$ ). Instead, we generate i.i.d. samples from  $\mathcal{Y}$  and reject those not in  $(\mathcal{A}^{\mathcal{Y}})^c$ . As we only have one Markov Chain, we denote  $(\mathcal{Z}_{k,A})_{k \geq 0}$  from Algorithm 3 as simply  $(\mathcal{Z}_k)_{k \geq 0}$ .

### 3.2.2 The sampler

We sample  $\mathcal{Z} = (\mathcal{Z}^{(1)}, \dots, \mathcal{Z}^{(D)})$  using a Markov Chain whose stationary distribution  $\pi_{\mathcal{A}^{\mathcal{Z}}}(z)dz$  is Gaussian restricted to  $\mathcal{A}^{\mathcal{Z}}$ , i.e.,

$$\pi_{\mathcal{A}^{\mathcal{Z}}}(z) := \frac{\mathbf{1}_{\mathcal{A}^{\mathcal{Z}}}(z)\pi(z)}{\int_{\mathcal{A}^{\mathcal{Z}}} \pi(t)dt}, \quad \text{with} \quad \pi(z) := \frac{e^{-\frac{|z|^2}{2}}}{(2\pi)^{D/2}}. \quad (3.4)$$

We use the preconditioned Crank-Nicolson sampler.

Moreover, we define  $\mathcal{M}^{(0)} := G_{\nu, \lambda}^{-1} \circ F_{\mathcal{N}}(\mathcal{Z}^{(1)})$ ,  $\mathcal{M}^{(i)} := G_{\nu}^{-1} \circ F_{\mathcal{N}}(\mathcal{Z}^{(i+1)})$  for  $i = 1, \dots, J-1$  and  $\epsilon^{(i)} := G_{\nu}^{-1} \circ F_{\mathcal{N}}(\mathcal{Z}^{(i+J)})$  for  $i = 1, \dots, d$ . In the notations of the paper, we then obtain  $\mathcal{Y} = \Phi(\mathcal{Z}) = (\Phi^{(i)}(\mathcal{Z}))_{i=1}^d$  with, for  $1 \leq i \leq d$ ,

$$\begin{aligned} \Phi^{(i)} : \mathbb{R}^D &\rightarrow \mathbb{R}, \\ z &\mapsto \beta_{S(i)} G_{\nu, \lambda}^{-1} \circ F_{\mathcal{N}}(z^{(1)}) + \gamma_{S(i)} G_{\nu}^{-1} \circ F_{\mathcal{N}}(z^{(S(i)+1)}) + G_{\nu}^{-1} \circ F_{\mathcal{N}}(z^{(i+J)}). \end{aligned}$$

This sampler is studied in Appendix A.2.1, and we assume that the geometric Assumption (A.4) is satisfied for the set  $\mathcal{A}^{\mathcal{Z}}$  relevant to the example, namely defined by

$$\mathcal{A}^{\mathcal{Z}} = \left\{ z \in \mathbb{R}^D \mid \Phi^{(i)}(z) \geq y^{(i)}, 1 \leq i \leq d \right\}.$$

Indeed, the proof can be made in the case where  $\mathcal{A}^{\mathcal{Z}}$  is an polyhedron, while the set  $\mathcal{A}^{\mathcal{Z}}$  in this context looks like a polyhedron at infinity. A rigorous verification of (A.4) is left for further research.

We now check that Assumption (A-iii) holds in this context, namely for the pCN sampler in  $\mathbb{R}^D$  with invariant distribution  $\mathcal{N}(0_D, I_D)$  restricted to  $\mathcal{A}^{\mathcal{Z}}$ .

**Proposition 3.2.** *Assume that  $\beta_S, \gamma_S \geq 0$  for all  $1 \leq S \leq J-1$ . Then for all  $q_{\max} > -1$ , there exists a quadratic Lyapunov function  $\mathcal{L}$  for which Assumption (A-iii) is satisfied.*

*Proof.* For all  $1 \leq i \leq d$  and  $z \in \mathbb{R}^D$ , we have (accounting for the signs of  $\beta_S, \gamma_S$ )

$$\begin{aligned} G^{(i)} \circ \Phi^{(i)}(z) &= \mathbb{P} \left[ \Phi^{(i)}(\mathcal{Z}) \leq \Phi^{(i)}(z) \right] \\ &\geq \mathbb{P} \left[ \mathcal{Z}^{(1)} \leq z^{(1)} \right] \mathbb{P} \left[ \mathcal{Z}^{(S(i)+1)} \leq z^{(S(i)+1)} \right] \mathbb{P} \left[ \mathcal{Z}^{(i+J)} \leq z^{(i+J)} \right]. \end{aligned}$$

Similarly

$$\begin{aligned} 1 - G^{(i)} \circ \Phi^{(i)}(z) &= \mathbb{P} \left[ \Phi^{(i)}(\mathcal{Z}) \geq \Phi^{(i)}(z) \right] \\ &\geq \mathbb{P} \left[ \mathcal{Z}^{(1)} \geq z^{(1)} \right] \mathbb{P} \left[ \mathcal{Z}^{(S(i)+1)} \geq z^{(S(i)+1)} \right] \mathbb{P} \left[ \mathcal{Z}^{(i+J)} \geq z^{(i+J)} \right]. \end{aligned}$$

Now, let  $q_{\max} > -1$  be fixed. Let  $-\frac{q}{2} < -\frac{q_{\max}}{2} < s < \frac{1}{2}$  and consider  $\mathcal{L}(z) := e^{s \sum_{i=1}^D |z^{(i)}|^2}$  for  $z \in \mathbb{R}^D$ . From Proposition A.2, since (A.4) is satisfied, this is a Lyapunov function for this

sampler. Moreover, since  $\mathcal{L}(z) \geq e^{s((z^{(1)})^2 + (z^{(S(i)+1)})^2 + (z^{(i+J)})^2)}$ , we obtain,

$$\begin{aligned} & \sup_{\mathcal{A}^{\mathcal{Z}}} \frac{(G^{(i)} \circ \Phi^{(i)})^q + (1 - G^{(i)} \circ \Phi^{(i)})^q}{\mathcal{L}} \\ & \leq \sup_{(u,v,w) \in \mathbb{R}^3} \frac{F_{\mathcal{N}}(u)^q F_{\mathcal{N}}(v)^q F_{\mathcal{N}}(w)^q + F_{\mathcal{N}}(-u)^q F_{\mathcal{N}}(-v)^q F_{\mathcal{N}}(-w)^q}{e^{s(u^2+v^2+w^2)}} \\ & \leq \sup_{(u,v,w) \in \mathbb{R}_+^3} 2^{-3q} e^{-s(u^2+v^2+w^2)} + (2\pi)^{-\frac{3q}{2}} \frac{uvw}{(u^2+1)(v^2+1)(w^2+1)} e^{-(s+\frac{q}{2})(u^2+v^2+w^2)} < +\infty, \end{aligned}$$

where we have used the lower Mill inequality (1.4) and  $s + \frac{q}{2} > s + \frac{q_{\max}}{2} > 0$ .  $\square$

### 3.2.3 Implementation details

For the results shown below, we use  $C = 0.01$ , meaning that  $(k - \text{ES})^{(i)}$  computes the expected loss in the  $i$ -th asset when all assets experience losses larger than 1%. For completeness, Figure 3 presents the distribution of  $a^{(j)}$ , the  $C = 0.01$  quantiles of  $\mathcal{X}^{(j)}$  (which can be computed, as  $F^{(i)}$  is perfectly known).

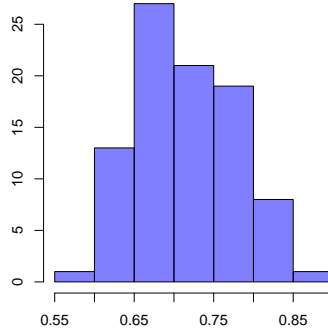


Figure 3: Histogram of  $a^{(j)} = \mathbb{P}[\mathcal{X}^{(j)} \leq C]$  for  $C = 0.01$ .

We estimate  $\mathbb{P}[A] \approx 1.42 \times 10^{-4}$  using a crude MC procedure for  $\mathcal{Y}$ , with sample size  $10^6$ . For computational speed, we use a subsample of size  $10^4$  only, to compute the empirical c.d.f. conditional to  $A^c$ . Note that in our proposed approach we completely avoid sampling from  $\mathcal{X}$ , i.e., we don't use Algorithm 1. Nevertheless, Algorithm 1, with a mollified c.d.f.  $G^{(i)}$ , is used to compute the “ground truth”, with which the proposed estimator will be compared to. Similarly to Section 3.1, we use  $M = 100$  independent macro-runs of Algorithm 3, modified to only consider one Markov Chain with values in  $\mathcal{A}^{\mathcal{Z}}$ , as discussed at the end of Section 3.2.1. The pCN sampler is used with parameters  $\kappa = 0.9$  and  $\mathcal{Z}_0 = 0_D$ . All the chains are iterated for  $n = 20\,000$  steps.

As previously discussed, the quantity  $\mathbb{P}[\mathcal{A}^{\mathcal{Z}}]$  involved in Algorithm 3 and the ground truth  $(k - \text{ES})^{(i)}$  can't be computed exactly and can only be estimated in this example. Therefore,

the study of the convergence rates in this example is deemed unnecessary, due to the extra source of noise added by these estimations.

Notwithstanding the fact that we don't present a formal convergence analysis we are still able to provide advices based on empirical observations. We point out that, even if the theoretical properties of the algorithms are independent of parameters such as  $\mathcal{Z}_0$  and  $\kappa$ , we observe in practice a substantial impact in the chains and the empirical convergence rates.

In addition, the mollification terms of the form  $\frac{1}{\sqrt{k}}$  can be replaced with  $\frac{1}{\sqrt{c_m k}}$  (for some constant  $c_m > 0$ ) without theoretical implications, but we observe in practice an improvement by good choice of  $c_m$ , and the numerical results presented below were generated with  $c_m = 100$ . Although the theoretical convergence rates are not impacted by these modifications, the multiplying constants are and we suspect that it may cause some numerical instability with respect to the parameters.

### 3.2.4 Results

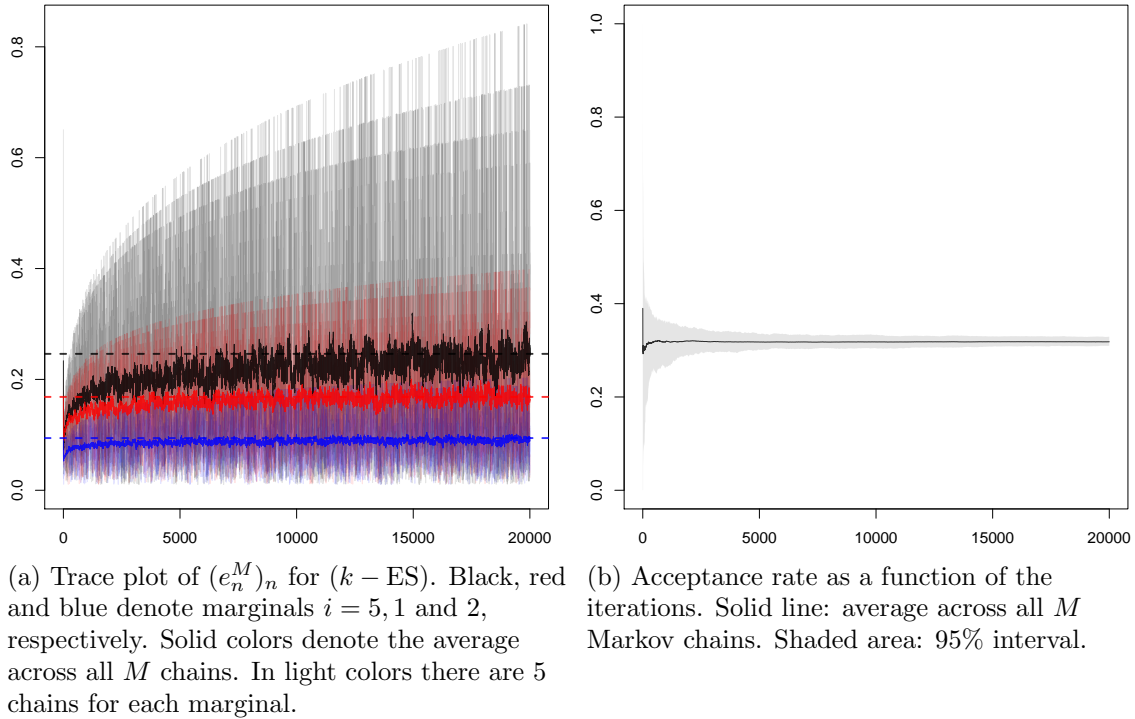


Figure 4: Qualitative analysis of the convergence.

In the sequel we discuss the results for three marginals:  $i = 5, 1$  and  $2$ . The 5th component, i.e., the stock of The Allstate Corporation (ALL), was included due to its marginal degree of freedom, which was estimated as 1.77. This is the smallest value in our sample and implies that these losses have infinite variance (which happens whenever the d.f. is less than 2), but

the mean is finite. This is a challenging case. The other two stocks have lighter tails and are included for comparison.

As in the Pareto example of Section 3.1, we estimate  $(k - \text{ES})^{(i)}$  through Equation (3.2). Figure 4a presents plot of  $(e_n^M)_n$  for each one of the marginals (solid colors). The dashed lines denote the estimate obtained by crude Monte Carlo using  $10^6$  samples. We can see that after 10 000 iterations the estimates from the Markov chain are close to their true value (computed through the Monte Carlo simulation with  $10^6$  samples). As previously mentioned, even though we don't present a formal convergence analysis here, we point out here that the chain for marginal  $i = 5$  takes longer to attain its correct value (illustrating once again that heavy tails slow down convergence, coherently with our theoretical results – the case of large  $\alpha$ ). Also in Figure 4a we present, in lighter colors, the estimates base on a single Markov chain (instead of the average across  $M$  chains). Due to the mollification, we see the maximum value of the estimates (with one chain) is an increasing function of the iteration, as expected. Note that for each marginal there are 5 chains plotted in light colors.

For each one of the  $M$  Markov chains, we compute the acceptance rate at each up to the  $n$ -th iteration. In Figure 4b we present three statistics for the acceptance rate across the  $M$  chains: the mean (solid line) and its 2.5% and 97.5% quantiles (shaded area). The reader is reminded the probability of the event  $A$  is of order  $10^{-4}$ , which would imply that about 0.01% of the samples would be accepted in a simple rejection Monte Carlo, in comparison with about 30% accepted samples for the proposed MCMC scheme. We remind that targeting 20-50% acceptance rate is a common practice which is supported by theoretical results, see e.g. [Ros11].

## 4 Proofs of the main results

### 4.1 Proof of Theorem 2.5

From Theorem A.1, we know that for any function  $\varphi : \mathcal{Z} \rightarrow \mathbb{R}$  bounded by 1,

$$\left| \frac{1}{n} \sum_{k=1}^n \varphi(\mathcal{Z}_k) - \mathbb{E}(\varphi(\mathcal{Z})) \right|_p \leq \left| \frac{1}{n} \sum_{k=1}^n \varphi(\mathcal{Z}_k) - \mathbb{E}(\varphi(\mathcal{Z})) \right|_{p \vee 2} \leq C_{p \vee 2, \mathcal{P}} \frac{\mathcal{L}^{1/(p \vee 2)}(\mathcal{Z}_0)}{\sqrt{n}}$$

for a constant depending only on the Markov kernel  $\mathcal{P}$  and the norm exponent  $p$ . This inequality applied to  $\varphi(z) = \mathbf{1}_{\Phi^{(i)}(z) \leq y}$  and  $\varphi(z) = \mathbf{1}_{\Phi^{(i)}(z) < y}$  ( $y \in \mathbb{R}$  fixed) gives

$$\left| \hat{G}_n^{(i)}(y) - G^{(i)}(y) \right|_p \vee \left| \hat{G}_n^{(i)}(y^-) - G^{(i)}(y^-) \right|_p \leq C_{p \vee 2, \mathcal{P}} \frac{\mathcal{L}^{1/p \vee 2}(\mathcal{Z}_0)}{\sqrt{n}}. \quad (4.1)$$

Let  $\epsilon > 0$ ; now design a finite decreasing sequence  $y_1, \dots, y_{k^*}$  of points which constitutes a  $\epsilon/2$ -covering of the c.d.f.  $G^{(i)}(\cdot)$ , we proceed by induction. Let  $y_1 = \inf \{y \in \mathbb{R} : G^{(i)}(y) \geq 1 - \frac{\epsilon}{2}\}$ , and assume that  $y_1, \dots, y_{k-1}$  are already constructed. If  $G^{(i)}(y_{k-1}^-) \leq \frac{\epsilon}{2}$ , then stop. Otherwise, we set  $y_k := \inf \{y \in \mathbb{R} : G^{(i)}(y) \geq G^{(i)}(y_{k-1}^-) - \frac{\epsilon}{2}\}$ . Notice that for all  $k$ , we have  $G^{(i)}(y_k^-) \leq 1 - k \frac{\epsilon}{2}$  so that  $k^* \leq \frac{2}{\epsilon}$ .

Now we prove that

$$\sup_{y \in \mathbb{R}} |G^{(i)}(y) - \hat{G}_n^{(i)}(y)| \leq \max \left\{ \left| G^{(i)}(y_k) - \hat{G}_n^{(i)}(y_k) \right| \vee \left| G^{(i)}(y_k^-) - \hat{G}_n^{(i)}(y_k^-) \right| : 1 \leq k \leq k^* \right\} + \frac{\epsilon}{2}. \quad (4.2)$$

For this, take  $y \in \mathbb{R}$ .

- If  $y \in [y_{k+1}, y_k)$  for some  $1 \leq k < k^*$ , we have, since  $G^{(i)}(y) - G^{(i)}(y_{k+1}) \leq G^{(i)}(y_k^-) - G^{(i)}(y_k^-) + \frac{\epsilon}{2} = \frac{\epsilon}{2}$ ,

$$\begin{aligned} G^{(i)}(y) - \hat{G}_n^{(i)}(y) &\leq G^{(i)}(y) - \hat{G}_n^{(i)}(y_{k+1}) \\ &= G^{(i)}(y) - G^{(i)}(y_{k+1}) + G^{(i)}(y_{k+1}) - \hat{G}_n^{(i)}(y_{k+1}) \\ &\leq G^{(i)}(y_{k+1}) - \hat{G}_n^{(i)}(y_{k+1}) + \frac{\epsilon}{2}, \end{aligned}$$

and also, since  $G^{(i)}(y) - G^{(i)}(y_k^-) \geq G^{(i)}(y_k^-) - \frac{\epsilon}{2} - G^{(i)}(y_k^-) = -\frac{\epsilon}{2}$ ,

$$\begin{aligned} G^{(i)}(y) - \hat{G}_n^{(i)}(y) &\geq G^{(i)}(y) - \hat{G}_n^{(i)}(y_k^-) \\ &= G^{(i)}(y) - G^{(i)}(y_k^-) + G^{(i)}(y_k^-) - \hat{G}_n^{(i)}(y_k^-) \\ &\geq G^{(i)}(y_k^-) - \hat{G}_n^{(i)}(y_k^-) - \frac{\epsilon}{2}. \end{aligned}$$

Thus  $|G^{(i)}(y) - \hat{G}_n^{(i)}(y)| \leq \left( \left| G^{(i)}(y_k^-) - \hat{G}_n^{(i)}(y_k^-) \right| \vee \left| G^{(i)}(y_{k+1}) - \hat{G}_n^{(i)}(y_{k+1}) \right| \right) + \frac{\epsilon}{2}$ .

- If  $y \in [y_1, +\infty)$ , we have, using  $G^{(i)}(y) \leq 1$  and  $G^{(i)}(y_1) \geq 1 - \frac{\epsilon}{2}$ ,

$$G^{(i)}(y) - \hat{G}_n^{(i)}(y) \leq G^{(i)}(y) - G^{(i)}(y_1) + G^{(i)}(y_1) - \hat{G}_n^{(i)}(y_1) \leq G^{(i)}(y_1) - \hat{G}_n^{(i)}(y_1) + \frac{\epsilon}{2},$$

and

$$G^{(i)}(y) - \hat{G}_n^{(i)}(y) \geq 1 - \frac{\epsilon}{2} - 1 = -\frac{\epsilon}{2},$$

thus we deduce  $|G^{(i)}(y) - \hat{G}_n^{(i)}(y)| \leq |G^{(i)}(y_1) - \hat{G}_n^{(i)}(y_1)| + \frac{\epsilon}{2}$ .

- Similarly, if  $y \in (-\infty, y_{k^*})$ ,

$$G^{(i)}(y) - \hat{G}_n^{(i)}(y) \leq G^{(i)}(y) \leq \frac{\epsilon}{2},$$

and

$$G^{(i)}(y) - \hat{G}_n^{(i)}(y) \geq G^{(i)}(y) - G^{(i)}(y_{k^*}^-) + G^{(i)}(y_{k^*}^-) - \hat{G}_n^{(i)}(y_{k^*}^-) \geq G^{(i)}(y_{k^*}^-) - \hat{G}_n^{(i)}(y_{k^*}^-) - \frac{\epsilon}{2},$$

so  $|G^{(i)}(y) - \hat{G}_n^{(i)}(y)| \leq |G^{(i)}(y_{k^*}^-) - \hat{G}_n^{(i)}(y_{k^*}^-)| + \frac{\epsilon}{2}$ .



All in all, we have justified (4.2). Furthermore, we deduce that

$$\left\{ \sup_{y \in \mathbb{R}} \left| G^{(i)}(y) - \hat{G}_n^{(i)}(y) \right| > \epsilon \right\} \subset \bigcup_{k=1}^{k^*} \left( \left\{ \left| G^{(i)}(y_k) - \hat{G}_n^{(i)}(y_k) \right| > \frac{\epsilon}{2} \right\} \cup \left\{ \left| G^{(i)}(y_k^-) - \hat{G}_n^{(i)}(y_k^-) \right| > \frac{\epsilon}{2} \right\} \right).$$

We then have, using Markov inequality, the bound (4.1) and the fact that  $k^* \leq \frac{2}{\epsilon}$ ,

$$\begin{aligned} \mathbb{P} \left[ \sup_{y \in \mathbb{R}} \left| G^{(i)}(y) - \hat{G}_n^{(i)}(y) \right| > \epsilon \right] &\leq \sum_{k=1}^{k^*} \left( \mathbb{P} \left[ \left| G^{(i)}(y_k) - \hat{G}_n^{(i)}(y_k) \right| > \frac{\epsilon}{2} \right] + \mathbb{P} \left[ \left| G^{(i)}(y_k^-) - \hat{G}_n^{(i)}(y_k^-) \right| > \frac{\epsilon}{2} \right] \right) \\ &\leq \frac{2^p}{\epsilon^p} \sum_{k=1}^{k^*} \left( \mathbb{E} \left( \left| G^{(i)}(y_k) - \hat{G}_n^{(i)}(y_k) \right|^p \right) + \mathbb{E} \left( \left| G^{(i)}(y_k^-) - \hat{G}_n^{(i)}(y_k^-) \right|^p \right) \right) \\ &\leq \frac{4}{\epsilon} C_{p \vee 2, \mathcal{P}}^p \mathcal{L}^{p/p \vee 2}(\mathcal{Z}_0) \frac{2^p}{\epsilon^p} \frac{1}{n^{\frac{p}{2}}} =: \frac{C_{p, \mathcal{P}, \mathcal{Z}_0}}{\epsilon^{p+1} n^{\frac{p}{2}}}. \end{aligned}$$

Last,

$$\begin{aligned} \left| \sup_{y \in \mathbb{R}} \left| G^{(i)}(y) - \hat{G}_n^{(i)}(y) \right| \right|_p^p &= \int_0^\infty \mathbb{P} \left[ \sup_{y \in \mathbb{R}} \left| G^{(i)}(y) - \hat{G}_n^{(i)}(y) \right| > \epsilon^{\frac{1}{p}} \right] d\epsilon \\ &\leq \epsilon_0 + \frac{C_{p, \mathcal{P}, \mathcal{Z}_0}}{n^{\frac{p}{2}}} \int_{\epsilon_0}^\infty \frac{1}{\epsilon^{\frac{p+1}{p}}} d\epsilon \\ &= \epsilon_0 + p \frac{C_{p, \mathcal{P}, \mathcal{Z}_0}}{n^{\frac{p}{2}}} \epsilon_0^{-\frac{1}{p}}. \end{aligned}$$

Thus, taking  $p$ -roots and setting  $\epsilon_0 = n^{-\frac{p^2}{2(p+1)}}$ , we obtain

$$\left| \sup_{y \in \mathbb{R}} \left| \hat{G}_n^{(i)}(y) - G^{(i)}(y) \right| \right|_p \leq (1 + p C_{p, \mathcal{P}, \mathcal{Z}_0})^{\frac{1}{p}} n^{-\frac{p}{2(p+1)}}.$$

To conclude, notice that

$$\begin{aligned} \left| \sup_{y \in \mathbb{R}} \left| \tilde{G}_n^{(i)}(y) - G^{(i)}(y) \right| \right|_p &\leq \left| \sup_{y \in \mathbb{R}} \left| \tilde{G}_n^{(i)}(y) - \hat{G}_n^{(i)}(y) \right| \right|_p + \left| \sup_{y \in \mathbb{R}} \left| \hat{G}_n^{(i)}(y) - G^{(i)}(y) \right| \right|_p \\ &\leq \frac{1}{2\sqrt{n}} + (1 + p C_{p, \mathcal{P}, \mathcal{Z}_0})^{\frac{1}{p}} n^{-\frac{p}{2(p+1)}} \leq \left( \frac{1}{2} + (1 + p C_{p, \mathcal{P}, \mathcal{Z}_0})^{\frac{1}{p}} \right) n^{-\frac{p}{2(p+1)}}, \end{aligned}$$

as we observe that

$$\left| \tilde{G}_n^{(i)}(y) - \hat{G}_n^{(i)}(y) \right| = \left| \frac{1}{2\sqrt{n}} - \frac{1}{\sqrt{n}} \hat{G}_n^{(i)}(y) \right| \leq \frac{1}{2\sqrt{n}} \leq \frac{1}{2} n^{-\frac{p}{2(p+1)}}, \text{ a.s. .}$$

We are done.  $\square$

## 4.2 Proof of Proposition 2.6

We start first a simple lemma.

**Lemma 4.1.** *Let  $(I_n)_{n \geq 1}$  be a sequence of measurable sets with  $I_n \subset [0, 1]$  and such that  $\text{Leb.}(I_n) \leq cn^{-\gamma}$  for some  $c, \gamma > 0$  independent of  $n$ . We then have, for a constant  $C > 0$  independent of  $n$ ,*

$$\sup_{i \in \{1, \dots, d\}} \mathbb{P} \left[ W_n^{(i)} \in I_n \right] \leq Cn^{-\gamma}.$$

*Proof.* Let  $n \geq 1$  and  $1 \leq i \leq d$ . Note that  $W_n^{(i)} = G^{(i)}(\mathcal{Y}_n^{(i)}) = G^{(i)}(\Phi^{(i)}(\mathcal{Z}_n))$ . Then, apply Theorem A.1, inequality (A.2), to the function  $\Phi_n(\cdot) = \mathbf{1}_{G^{(i)} \circ \Phi^{(i)}(\cdot) \in I_n}$ : since  $G^{(i)}(\Phi^{(i)}(\mathcal{Z}))$  is uniformly distributed on  $[0, 1]$  (Assumption (A-iv)), we get

$$\mathbb{P} \left[ W_n^{(i)} \in I_n \right] \leq \text{Leb.}(I_n) + C_{(\text{A.2})} \mathcal{L}(\mathcal{Z}_0) \rho^n.$$

Using that  $\text{Leb.}(I_n) \leq cn^{-\gamma}$  and exploiting that the  $\rho^n$ -term is negligible compared to the other one, we get the result.  $\square$

Now let us go back to the proof of Proposition 2.6. Let  $n \geq 1$  and  $q \in (q_{\max}, 0)$  be given (the case  $q \geq 0$  is obvious).

- We prove that  $\mathbb{E} \left( \left[ W_n^{(i)} \right]^q \right)$  and  $\mathbb{E} \left( \left[ 1 - W_n^{(i)} \right]^q \right)$  are bounded uniformly in  $n$ : we only justify the first statement, one can deal with the second quantity similarly. First, we have

$$\mathbb{E} \left( \left[ W_n^{(i)} \right]^q \right) = \mathbb{E} \left( \left[ G^{(i)}(\mathcal{Y}_n^{(i)}) \right]^q \right) = \mathbb{E} \left( \left[ G^{(i)}(\Phi^{(i)}(\mathcal{Z}_n)) \right]^q \right).$$

Thus, from Theorem A.1, we have, for some  $C_{(\text{A.2})} \geq 0$  and  $\rho \in (0, 1)$ , since  $G^{(i)}(\Phi^{(i)}(\mathcal{Z})) = G^{(i)}(\mathcal{Y}^{(i)})$  is uniformly distributed on  $[0, 1]$ ,

$$\mathbb{E} \left( \left[ W_n^{(i)} \right]^q \right) \leq \int_0^1 x^q dx + C_{(\text{A.2})} \left( \sup_{\mathcal{A}^{\mathcal{Z}}} \frac{(G^{(i)} \circ \Phi^{(i)})^q}{\mathcal{L}} \right) \mathcal{L}(\mathcal{Z}_0) \rho^n.$$

Since  $\sup_{\mathcal{A}^{\mathcal{Z}}} \frac{(G^{(i)} \circ \Phi^{(i)})^q}{\mathcal{L}}$  is finite by Assumption (A-iii), the above quantity is bounded uniformly in  $n$ .

- Let us now handle the proof of the uniform bound on  $\mathbb{E} \left( \left[ V_n^{(i)} \right]^q \right)$ , the one on  $\mathbb{E} \left( \left[ 1 - V_n^{(i)} \right]^q \right)$  is quite similar and left to the reader. Since

$$0 \leq \mathbb{E} \left( \left[ V_n^{(i)} \right]^q \right) \leq \mathbb{E} \left( \left| \left[ V_n^{(i)} \right]^q - \left[ W_n^{(i)} \right]^q \right| \right) + \mathbb{E} \left( \left[ W_n^{(i)} \right]^q \right),$$

it is enough to prove  $\mathbb{E} \left( \left| \left[ V_n^{(i)} \right]^q - \left[ W_n^{(i)} \right]^q \right| \right)$  is upper bounded, uniformly in  $n$ .

Let us fix  $J = \left\lceil n^{\frac{1-q}{4}} \right\rceil$  and  $h := \frac{1}{J}$ , so that  $cn^{\frac{q-1}{4}} \leq h \leq n^{\frac{q-1}{4}}$  for some  $0 < c < 1$  independent of  $n$ . We estimate the quantity of interest according to the partitioning event  $\{jh < W_n^{(i)} \leq (j+1)h\}$  for  $j = 0, \dots, J-1$ .

We start with the term  $j = 0$ . We have, since  $V_n^{(i)} \in \left[\frac{1}{2\sqrt{n}}, 1 - \frac{1}{2\sqrt{n}}\right]$  and using Lemma 4.1, for some  $C > 0$ ,

$$\begin{aligned} & \mathbb{E} \left( \left| \left[ V_n^{(i)} \right]^q - \left[ W_n^{(i)} \right]^q \right| \mathbf{1}_{W_n^{(i)} \leq h} \right) \\ & \leq \mathbb{E} \left( \left[ V_n^{(i)} \right]^q \mathbf{1}_{W_n^{(i)} \leq h} \right) + \mathbb{E} \left( \left[ W_n^{(i)} \right]^q \mathbf{1}_{W_n^{(i)} \leq h} \right) \\ & \leq \left( \frac{1}{2\sqrt{n}} \right)^q \mathbb{P} \left[ W_n^{(i)} \leq h \right] + \mathbb{E} \left( \left[ W_n^{(i)} \right]^q \mathbf{1}_{W_n^{(i)} \leq h} \right) \\ & \leq Cn^{-\frac{q}{2}}h + \mathbb{E} \left( \left[ W_n^{(i)} \right]^q \mathbf{1}_{W_n^{(i)} \leq h} \right) \\ & \leq Cn^{-\frac{q+1}{4}} + \int_0^h x^q dx + C_{(\text{A.2})} \left( \sup_{\mathcal{A}^{\mathbb{Z}}} \frac{(G^{(i)} \circ \Phi^{(i)})^q}{\mathcal{L}} \right) \mathcal{L}(\mathcal{Z}_0) \rho^n, \end{aligned}$$

which is bounded uniformly in  $n$ .

We consider the case  $0 < j < J$ . Using the Hölder inequality for  $p > 1$  (to be chosen later), Lemma 4.1 and Theorem 2.5 as  $\left| V_n^{(i)} - W_n^{(i)} \right| \leq \sup_{\mathbb{R}} \left| \tilde{G}_n^{(i)} - G^{(i)} \right|$ , we get, for some  $C > 0$ ,

$$\begin{aligned} & \mathbb{E} \left( \left| \left[ V_n^{(i)} \right]^q - \left[ W_n^{(i)} \right]^q \right| \mathbf{1}_{jh < W_n^{(i)} \leq (j+1)h} \mathbf{1}_{\left| V_n^{(i)} - W_n^{(i)} \right| \leq \frac{h}{2}} \right) \\ & \leq \mathbb{E} \left( \left| V_n^{(i)} - W_n^{(i)} \right| \left( W_n^{(i)} \wedge V_n^{(i)} \right)^{q-1} \mathbf{1}_{jh < W_n^{(i)} \leq (j+1)h} \mathbf{1}_{\left| V_n^{(i)} - W_n^{(i)} \right| \leq \frac{h}{2}} \right) \\ & \leq \left( j - \frac{1}{2} \right)^{q-1} h^{q-1} \mathbb{E} \left( \left| V_n^{(i)} - W_n^{(i)} \right|^p \right)^{\frac{1}{p}} \mathbb{P} \left[ jh < W_n^{(i)} \leq (j+1)h \right]^{\frac{p-1}{p}} \\ & \leq C \left( j - \frac{1}{2} \right)^{q-1} h^{q-1} n^{-\frac{p}{2(p+1)}} h^{\frac{p-1}{p}}. \end{aligned}$$

In addition, using the Markov inequality for  $r \geq 1$  (to be chosen later),

$$\begin{aligned} \mathbb{E} \left( \left| \left[ V_n^{(i)} \right]^q - \left[ W_n^{(i)} \right]^q \right| \mathbf{1}_{jh < W_n^{(i)} \leq (j+1)h} \mathbf{1}_{\left| V_n^{(i)} - W_n^{(i)} \right| > \frac{h}{2}} \right) & \leq 2^r \left( \left( \frac{1}{2\sqrt{n}} \right)^q + (jh)^q \right) \frac{\mathbb{E} \left( \left| V_n^{(i)} - W_n^{(i)} \right|^r \right)}{h^r} \\ & \leq 2^r C_{r,(\text{2.5})}^r \left( \left( \frac{1}{2\sqrt{n}} \right)^q + (jh)^q \right) n^{-\frac{r^2}{2(r+1)}} h^{-r}. \end{aligned}$$

Summing over  $0 < j < J$ , by our choice of  $p, r$  and since  $\sum_{j \geq 1} \left( j - \frac{1}{2} \right)^{q-1} < +\infty$  and  $\sum_{j=1}^{J-1} j^q = O(h^{-q-1})$  as  $-1 < q < 0$ , one obtains for some positive constant  $C$  (changing from

term to term),

$$\mathbb{E} \left( \left| \left[ V_n^{(i)} \right]^q - \left[ W_n^{(i)} \right]^q \right| \mathbf{1}_{W_n^{(i)} > h} \right) \leq C h^{q-1+\frac{p-1}{p}} n^{-\frac{p}{2(p+1)}} + C n^{-\frac{q}{2}-\frac{r^2}{2(r+1)}} h^{-r-1} + C n^{-\frac{r^2}{2(r+1)}} h^{-r-1}.$$

Because  $q < 0$ , the third term can be neglected compared to the second term.

After replacing  $h$  by its lower bound, the first term can be bounded by  $C n^{[\dots]}$  with an exponent equal to

$$\frac{q-1}{4} \left( q - \frac{1}{p} \right) - \frac{p}{2(p+1)} :$$

as  $p \rightarrow +\infty$ , it converges to  $\frac{q(q-1)}{4} - \frac{1}{2} < 0$  (here we use  $q \in (-1, 0)$ ). Therefore, for  $p$  large enough, the first term is uniformly bounded in  $n$ .

Similarly, the second term is bounded by  $C n^{[\dots]}$  with an exponent equal to  $-\frac{q}{2} - \frac{r^2}{2(r+1)} - \frac{q-1}{4}(r+1)$ : when  $r \rightarrow +\infty$ , it is equivalent to  $-(\frac{1}{2} + \frac{q-1}{4})r \rightarrow -\infty$ . Therefore, for  $r$  large enough, the second term is uniformly bounded in  $n$ . We are done with the bounds on  $\mathbb{E} \left( [V_n^{(i)}]^q \right)$  and  $\mathbb{E} \left( [1 - V_n^{(i)}]^q \right)$ .

• We now turn to the proof for the last statement. Let  $n \geq 1$ ,  $0 \leq k \leq d$  and  $1 \leq p < \frac{-q_{\max}}{\alpha}$  (with the convention  $\frac{-q_{\max}}{0} = +\infty$  if  $\alpha = 0$ ). Then there exists  $\epsilon > 0$  such that  $1 \leq p < \frac{-q_{\max}}{\alpha+\epsilon}$ . As the function  $\ell$  in the definition of  $\varphi$  (see Assumption (A-v)) is slowly varying at 0 and bounded on any interval excluding  $u = 0$ , it follows from [BGT87, Proposition 1.3.6 (v)] that

$$|\ell(u)| \leq C_{\ell, \epsilon} u^{-\epsilon}, \quad \forall u \in (0, 1]. \quad (4.3)$$

Furthermore, by Jensen's inequality  $(\sum_{i=1}^d x_i + \sum_{i=1}^d x'_i)^p \leq (2d)^{p-1} \left( \sum_{i=1}^d x_i^p + \sum_{i=1}^d (x'_i)^p \right)$  for non-negative  $x_i, x'_i$ . Putting these results together, we obtain

$$\begin{aligned} \mathbb{E} \left( \left| \varphi(\chi_n^k) \right|^p \right) &\leq (2d)^{p-1} C_{\ell, \epsilon}^p \left[ \sum_{i=1}^k \left( \mathbb{E} \left( \left( V_n^{(i)} \right)^{-p(\alpha+\epsilon)} \right) + \mathbb{E} \left( \left( 1 - V_n^{(i)} \right)^{-p(\alpha+\epsilon)} \right) \right) \right. \\ &\quad \left. + \sum_{i=k+1}^d \left( \mathbb{E} \left( \left( W_n^{(i)} \right)^{-p(\alpha+\epsilon)} \right) + \mathbb{E} \left( \left( 1 - W_n^{(i)} \right)^{-p(\alpha+\epsilon)} \right) \right) \right]. \end{aligned}$$

Since  $-p(\alpha+\epsilon) =: q > q_{\max}$ , the first part of the proposition implies that the above quantity is bounded independently of  $k, n$ .  $\square$

### 4.3 Proof of Theorem 2.7

We fix  $1 \leq p < \frac{-q_{\max}}{\alpha}$  (recall that  $\frac{-q_{\max}}{0} = +\infty$ ). For each  $0 \leq k \leq d$ , we define  $\chi_n^k = \left( \left( V_n^{(i)} \right)_{i=1}^k, \left( W_n^{(i)} \right)_{i=k+1}^d \right)$ . In particular,  $\chi_n^0 = W_n$  and  $\chi_n^d = V_n$ . By the Minkowski inequality, we have

$$|\varphi(V_n) - \varphi(W_n)|_p \leq \sum_{k=1}^d \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right|_p.$$

We now fix  $1 \leq k \leq d$  and we estimate the corresponding term above.

Let  $b \in (0, \frac{1}{2})$  close enough to  $\frac{1}{2}$  so that  $b(1 + \frac{p\alpha}{q_{\max}}) \geq \frac{1}{2} + \frac{p\alpha}{2q_{\max}} - \frac{p\ell}{2}$  and  $\frac{b}{p} \geq \frac{1}{2p} - \frac{\ell}{2}$ . We fix  $J = \lceil n^b \rceil$  and  $h = \frac{1}{J}$  so that  $n^{-b} \geq h \geq \frac{1}{2}n^{-b}$  for any  $n \geq 1$ .

Using the Hölder inequality for some  $1 < p_1 < \frac{-q_{\max}}{p\alpha}$  (we recall the convention  $\frac{-q_{\max}}{0} = +\infty$ , and that  $1 \leq p < \frac{-q_{\max}}{\alpha}$  so there exists such  $p_1$ ),

$$\begin{aligned} & \left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \left( \mathbf{1}_{W_n^{(k)} \leq h} + \mathbf{1}_{W_n^{(k)} > 1-h} \right) \right|_p \\ & \leq \left| \varphi(\chi_n^k) \left( \mathbf{1}_{W_n^{(k)} \leq h} + \mathbf{1}_{W_n^{(k)} > 1-h} \right) \right|_p + \left| \varphi(\chi_n^{k-1}) \left( \mathbf{1}_{W_n^{(k)} \leq h} + \mathbf{1}_{W_n^{(k)} > 1-h} \right) \right|_p \\ & \leq \left[ \left| \varphi(\chi_n^k) \right|_{p_1 p} + \left| \varphi(\chi_n^{k-1}) \right|_{p_1 p} \right] \left| \mathbf{1}_{W_n^{(k)} \leq h} + \mathbf{1}_{W_n^{(k)} > 1-h} \right|_{p \frac{p_1}{p_1-1}}, \end{aligned} \quad (4.4)$$

and the term in brackets is bounded by a non-negative constant  $C_{p_1}$  independent of  $n$  by Proposition 2.6. Using Lemma 4.1 with the set  $I_n = [0, h] \cup (1-h, 1]$ , we obtain

$$\left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \left( \mathbf{1}_{W_n^{(k)} \leq h} + \mathbf{1}_{W_n^{(k)} > 1-h} \right) \right|_p \leq C_{p_1} (2h)^{\frac{p_1-1}{p_1 p}} \leq C_{p_1} n^{-b \frac{(p_1-1)}{p_1 p}},$$

where we have used  $\frac{1}{2}n^{-b} \leq h \leq n^{-b}$ . We have  $\frac{p_1-1}{p_1} = 1 - \frac{1}{p_1} < 1 + \frac{p\alpha}{q_{\max}}$  and  $\lim_{p_1 \rightarrow \frac{-q_{\max}}{p\alpha}} \frac{p_1-1}{p_1} = 1 + \frac{p\alpha}{q_{\max}}$ , thus taking  $p_1$  close enough to  $\frac{-q_{\max}}{p\alpha}$  so that  $\frac{p_1-1}{p_1} \geq 1 + \frac{p\alpha}{q_{\max}} - \frac{p\ell}{2b}$  gives the existence of  $C_1(\ell) > 0$  such that

$$\left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \left( \mathbf{1}_{W_n^{(k)} \leq h} + \mathbf{1}_{W_n^{(k)} > 1-h} \right) \right|_p \leq C_1(\ell) n^{-b(\frac{1}{p} + \frac{\alpha}{q_{\max}} - \frac{\ell}{2b})}.$$

By our choice of  $b$ , we obtain

$$\left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \left( \mathbf{1}_{W_n^{(k)} \leq h} + \mathbf{1}_{W_n^{(k)} > 1-h} \right) \right|_p \leq C_1(\ell) n^{-\frac{1}{2p} + \frac{\alpha}{2|q_{\max}|} + \ell}. \quad (4.5)$$

We now fix  $0 < j < J-1$ . Owing to the bounds (2.3) and (4.3) on  $\varphi$  and  $\ell$ , we have, for any  $\epsilon \in (0, -q_{\max} - \alpha)$ ,

$$\begin{aligned} & \left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \mathbf{1}_{jh < W_n^{(k)} \leq (j+1)h} \mathbf{1}_{|V_n^{(k)} - W_n^{(k)}| \leq \frac{h}{2}} \right|_p \\ & \leq C_{\ell, \epsilon} \left| \left( \left( V_n^{(k)} \wedge W_n^{(k)} \right)^{-\alpha-\epsilon-1} + \left( 1 - V_n^{(k)} \vee W_n^{(k)} \right)^{-\alpha-\epsilon-1} \right) \left| V_n^{(k)} - W_n^{(k)} \right| \mathbf{1}_{jh < W_n^{(k)} \leq (j+1)h} \mathbf{1}_{|V_n^{(k)} - W_n^{(k)}| \leq \frac{h}{2}} \right|_p \\ & \leq C_{\ell, \epsilon} h^{-\alpha-\epsilon-1} \left( \left( j - \frac{1}{2} \right)^{-\alpha-\epsilon-1} + \left( J - j - \frac{3}{2} \right)^{-\alpha-\epsilon-1} \right) \left| \left| V_n^{(k)} - W_n^{(k)} \right| \mathbf{1}_{jh < W_n^{(k)} \leq (j+1)h} \right|_p, \end{aligned}$$

where we used that  $V_n^{(k)}, W_n^{(k)} \in [jh - \frac{h}{2}, (j+1)h + \frac{h}{2}]$  on the event  $\{jh < W_n^{(k)} \leq (j+1)h\} \cap \{|V_n^{(k)} - W_n^{(k)}| \leq \frac{h}{2}\}$ . Now apply Hölder inequality with some  $p_2 > 1$ , use Theorem 2.5 and

Lemma 4.1 with  $I_n = (jh, (j+1)h]$ : it readily follows (with a constant  $C_{p_2}$  changing from line to line)

$$\begin{aligned}
& \left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \mathbf{1}_{jh < W_n^{(k)} \leq (j+1)h} \mathbf{1}_{|V_n^{(k)} - W_n^{(k)}| \leq \frac{h}{2}} \right|_p \\
& \leq C_{p_2} h^{-\alpha-\epsilon-1} \left( \left( j - \frac{1}{2} \right)^{-\alpha-\epsilon-1} + \left( J - j - \frac{3}{2} \right)^{-\alpha-\epsilon-1} \right) n^{-\frac{p_2 p}{2(p_2 p + 1)}} h^{\frac{p_2 - 1}{p_2 p}} \\
& \leq C_{p_2} n^{b(\alpha+\epsilon+1)} \left[ \left( j - \frac{1}{2} \right)^{-\alpha-\epsilon-1} + \left( J - j - \frac{3}{2} \right)^{-\alpha-\epsilon-1} \right] n^{-\frac{p_2 p}{2(p_2 p + 1)} - b \frac{p_2 - 1}{p_2 p}}. \quad (4.6)
\end{aligned}$$

Now sum over  $0 < j < J - 1$ : the sum of terms in [...] is bounded by  $2 \sum_{j \geq 1} (j - \frac{1}{2})^{-\alpha-\epsilon-1}$  which is finite since  $\alpha + \epsilon > 0$ . This gives (for a new constant)

$$\left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \mathbf{1}_{h < W_n^{(k)} \leq 1-h} \mathbf{1}_{|V_n^{(k)} - W_n^{(k)}| \leq \frac{h}{2}} \right|_p \leq C_{p_2} n^{b(\alpha+\epsilon+1-\frac{1}{p}+\frac{1}{p_2 p})-\frac{p_2 p}{2(p_2 p + 1)}}. \quad (4.7)$$

As  $p_2 \rightarrow +\infty$  and  $\epsilon \rightarrow 0$ , the above exponent of  $n$  converges to  $b(\alpha+1-\frac{1}{p})-\frac{1}{2} = b\alpha - \frac{b}{p} + b - \frac{1}{2}$ , and therefore by our choice of  $b$  and for some appropriate choices of  $p_2$  and  $\epsilon$  (depending on  $b$  and  $\iota$ ), we get

$$\left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \mathbf{1}_{h < W_n^{(k)} \leq 1-h} \mathbf{1}_{|V_n^{(k)} - W_n^{(k)}| \leq \frac{h}{2}} \right|_p \leq C_{p_2} n^{b\alpha - \frac{b}{p} + b - \frac{1}{2} + \frac{\iota}{2}} \leq C_{p_2} n^{-\frac{1}{2p} + \frac{\alpha}{2|q_{\max}|} + \iota} \quad (4.8)$$

where we have used the crude estimates  $b < \frac{1}{2} \leq \frac{1}{2|q_{\max}|}$  and  $b - \frac{1}{2} < 0$  at the last inequality. Last, using Hölder inequality with some  $p_3 \in (1, \frac{-q_{\max}}{p\alpha})$ , Proposition 2.6, Markov inequality and Theorem 2.5, for all  $r \geq 1$  we get (with a constant  $C_{p_3}$  changing from line to line)

$$\begin{aligned}
\left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \mathbf{1}_{|V_n^{(k)} - W_n^{(k)}| > \frac{h}{2}} \right|_p & \leq C_{p_3} \left( \mathbb{P} \left[ |V_n^{(k)} - W_n^{(k)}| > \frac{h}{2} \right] \right)^{\frac{p_3 - 1}{p_3 p}} \\
& \leq C_{p_3} \left( n^{br} \mathbb{E} \left( |V_n^{(k)} - W_n^{(k)}|^r \right) \right)^{\frac{p_3 - 1}{p_3 p}} \\
& \leq C_{p_3} \left( n^{br - \frac{r^2}{2(r+1)}} \right)^{\frac{p_3 - 1}{p_3 p}}.
\end{aligned}$$

Since  $b < \frac{1}{2}$  the exponent of  $n$  goes to  $-\infty$  as  $r$  goes to  $+\infty$  (whatever the value of  $p_3 > 1$  is); therefore, obviously for  $r$  large, we obtain, for some  $C_3(\iota) > 0$ ,

$$\left| \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right| \mathbf{1}_{|V_n^{(k)} - W_n^{(k)}| > \frac{h}{2}} \right|_p \leq C_3(\iota) n^{-\frac{1}{2p} + \frac{\alpha}{2|q_{\max}|} + \iota}. \quad (4.9)$$

Combining (4.5)-(4.8)-(4.9), we obtain

$$\sup_{k=1 \dots d} \left| \varphi(\chi_n^k) - \varphi(\chi_n^{k-1}) \right|_p \leq C_{p,\iota} n^{-\frac{1}{2p} + \frac{\alpha}{2|q_{\max}|} + \iota}$$

for some constant  $C_{p,\iota}$ . □

**Remark 4.2.** 1. We notice that the terms contributing the more to the overall error are the border terms (associated to  $W_n^{(k)} \leq h$  or  $W_n^{(k)} > 1 - h$ , see (4.4)). One could argue that instead of using the bound assumption (2.4) on  $\varphi$ , one would have better to leverage the regularity assumption (2.3): in fact, for border terms, the regularity assumption does not bring any valuable information. Indeed the superiority of the regularity assumption would write as  $\frac{|v-w|}{(v \wedge w)^{\alpha+1}} \ll \frac{1}{(v \wedge w)^\alpha}$ , which is equivalent to  $|v-w| \ll v \wedge w$ : the latter inequality likely does not hold for  $v = V_n^{(k)}$  and  $w = W_n^{(k)}$ , since to the best of our knowledge, their difference is approximately of order  $n^{-1/2}$  and  $V_n^{(k)} \geq \frac{1}{2}n^{-1/2}$ .

2. In the case where  $-1 \leq \alpha < 0$ , see Theorem 2.10 and the assumption above it, one can indeed make use of the regularity assumption (2.6) to improve the convergence rate, leading to the result in Theorem 2.10 (i). Since the above proof is quite similar to the proof of Theorem 2.7, we only provide a sketch of proof and leave details to the reader.

Let  $1 \leq p < \infty$ , and we are interested as above in  $|\varphi(\chi_n^k) - \varphi(\chi_n^{k-1})|_p$ . We fix  $h = n^{-b}$  with  $b \in (0, \frac{1}{2})$ .

- Following the arguments above, the term  $\left| |\varphi(\chi_n^k) - \varphi(\chi_n^{k-1})| \mathbf{1}_{|V_n^{(k)} - W_n^{(k)}| > \frac{h}{2}} \right|_p$  goes to 0 faster than any negative power of  $n$ .
- For  $0 < j < J - 1$ , we similarly obtain (4.6). When summing over  $j$ , since  $\alpha + \epsilon < 0$  for small  $\epsilon$ , the sum  $\sum_{j=1}^{J-1} (j - \frac{1}{2})^{-\alpha-\epsilon-1}$  is upper bounded by  $O(J^{-\alpha-\epsilon}) = O(h^{\alpha+\epsilon})$ , so the upper bound in (4.7) becomes  $O(n^{b(1-\frac{1}{p}+\frac{1}{p_2p})-\frac{p_2p}{2(p_2p+1)}})$ , and the exponent goes to  $b(1-\frac{1}{p}) - \frac{1}{2}$  as  $p_2 \rightarrow \infty$ .
- We introduce  $0 < h' < h$  so that  $n^{-b'} \geq h' \geq \frac{1}{2}n^{-b'}$  with  $0 < b < \frac{1}{2} \leq b'$ , and we study the two terms  $\left| |\varphi(\chi_n^k) - \varphi(\chi_n^{k-1})| \left( \mathbf{1}_{W_n^{(k)} \leq h'} + \mathbf{1}_{W_n^{(k)} > 1-h'} \right) \right|_p$  and  $\left| |\varphi(\chi_n^k) - \varphi(\chi_n^{k-1})| \left( \mathbf{1}_{h \geq W_n^{(k)} > h'} + \mathbf{1}_{1-h < W_n^{(k)} \leq 1-h'} \right) \right|_p$ . As in the proof of the theorem, the first term is dealt with using the bound assumption (2.4), leading to an upper bound of the form  $O(n^{-b' \frac{(p_1-1)}{p_1p}})$ . Here  $p_1 \in (1, \infty)$  owing to the boundedness of  $\varphi$ , so the exponent goes to  $\frac{b'}{p}$  as  $p_1 \rightarrow \infty$ . The second term is dealt with using the regularity assumption (2.3) and then following the same steps as above, and we obtain the upper bound  $O(n^{b'(\alpha+1+\epsilon)-\frac{p_1p}{2(p_1p+1)}-b\frac{p_1-1}{p_1p}})$  for all  $1 < p_1 < \infty$ . The exponent goes to  $b'(\alpha+1+\epsilon) - \frac{1}{2} - \frac{b}{p}$  as  $p_1 \rightarrow \infty$ .
- We then choose (with  $\epsilon \rightarrow 0$ )  $b' = \frac{\frac{1}{2}+\frac{b}{p}}{\alpha+1+\frac{1}{p}}$  and  $b = \frac{p(\alpha+1)}{2(p(\alpha+1)-\alpha)} < \frac{1}{2}$ , so that all the previous exponents are equal. Then we have  $b' = \frac{p}{2(p(\alpha+1)-\alpha)} \geq \frac{1}{2}$ , and we obtain a convergence order  $\frac{1}{2(p(\alpha+1)-\alpha)} - \iota$  for all  $\iota > 0$ . We are done with Theorem 2.10 (i).

#### 4.4 Proof of Corollary 2.8

Let  $n \geq 1$  and let  $\iota > 0$ . We have

$$|\mathbb{E}(\varphi(V_n)) - \mathbb{E}(\varphi(U))| \leq |\mathbb{E}(\varphi(V_n)) - \mathbb{E}(\varphi(W_n))| + |\mathbb{E}(\varphi(W_n)) - \mathbb{E}(\varphi(U))| \quad (4.10)$$

and we estimate the two terms separately.

- Firstly, we have, by Theorem A.1, with  $\Phi = \varphi(G^{(1)} \circ \Phi^{(1)}, \dots, G^{(d)} \circ \Phi^{(d)})$ ,

$$|\mathbb{E}(\varphi(W_n)) - \mathbb{E}(\varphi(U))| = |\mathbb{E}(\Phi(\mathcal{Z}_n)) - \mathbb{E}(\Phi(\mathcal{Z}))| \leq C_{(A.2)} \mathcal{L}(\mathcal{Z}_0) \left( \sup_{\mathcal{A}^{\mathcal{Z}}} \frac{|\Phi|}{\mathcal{L}} \right) \rho^n. \quad (4.11)$$

Let us check that  $\Phi$  has a finite  $\mathcal{L}$ -norm. Since  $\ell$  is slowly varying, we can use the bound (4.3) available for any  $\epsilon > 0$ . Let  $\epsilon \in (0, -\alpha - q_{\max})$  (we recall that  $\alpha < -q_{\max}$  is assumed): then, the  $\varphi$ -assumption bound (2.4) yields

$$\sup_{\mathcal{A}^{\mathcal{Z}}} \frac{|\Phi|}{\mathcal{L}} \leq C_{\ell, \epsilon} \sum_{i=1}^d \sup_{\mathcal{A}^{\mathcal{Z}}} \frac{(G^{(i)} \circ \Phi^{(i)})^{-\alpha-\epsilon} + (1 - G^{(i)} \circ \Phi^{(i)})^{-\alpha-\epsilon}}{\mathcal{L}}.$$

The above bound is finite by assumption (A-iii) about the sampler owing to  $-\alpha - \epsilon > q_{\max}$ .

- Secondly, invoking Theorem 2.7 with  $p = 1$ , we have

$$\begin{aligned} |\mathbb{E}(\varphi(V_n)) - \mathbb{E}(\varphi(W_n))| &\leq \mathbb{E}(|\varphi(V_n) - \varphi(W_n)|) \\ &= |\varphi(V_n) - \varphi(W_n)|_1 \\ &\leq C_{1, \iota} n^{-\frac{1}{2} + \frac{\alpha}{2|q_{\max}|} + \iota}, \end{aligned} \quad (4.12)$$

for some positive constant  $C_{1, \iota}$ .

- To conclude, using estimates (4.11)-(4.12) in (4.10) and exploiting that the  $\rho^n$ -term is negligible compared to the other one, we obtain the announced result.  $\square$

#### 4.5 Proof of Corollary 2.9

We have, by the Minkowski inequality,

$$\left| \frac{1}{n} \sum_{k=1}^n \varphi(V_k) - \mathbb{E}(\varphi(U)) \right|_p \leq \frac{1}{n} \sum_{k=1}^n |\varphi(V_k) - \varphi(W_k)|_p + \left| \frac{1}{n} \sum_{k=1}^n \varphi(W_k) - \mathbb{E}(\varphi(U)) \right|_p. \quad (4.13)$$

By Theorem 2.7 we have, for each  $k \geq 1$ ,

$$|\varphi(V_k) - \varphi(W_k)|_p \leq C_{p, \iota} k^{-\frac{1}{2p} + \frac{\alpha}{2|q_{\max}|} + \iota}, \quad (4.14)$$

thus,

$$\frac{1}{n} \sum_{k=1}^n |\varphi(V_k) - \varphi(W_k)|_p \leq \frac{C_{p, \iota}}{n} \sum_{k=1}^n k^{-\frac{1}{2p} + \frac{\alpha}{2|q_{\max}|} + \iota}$$



$$\leq C_{p,\iota} n^{-\frac{1}{2p} + \frac{\alpha}{2|q_{\max}|} + \iota}, \quad (4.15)$$

since the exponent of  $k$  in (4.14) is strictly larger than  $-1$ . Here the constant  $C_{p,\iota}$  has changed. We now deal with the second term. Using (A.1) from Theorem A.1 and the fact that, if  $p < 2$ , the  $L^p$ -norm is controlled by the  $L^2$ -norm, we obtain

$$\begin{aligned} \left| \frac{1}{n} \sum_{k=1}^n \varphi(W_k) - \mathbb{E}(\varphi(U)) \right|_p &\leq \left| \frac{1}{n} \sum_{k=1}^n \varphi \circ (G^{(1)}, \dots, G^{(d)}) \circ \Phi(Z_k) - \mathbb{E}(\varphi \circ (G^{(1)}, \dots, G^{(d)}) \circ (\mathcal{Z})) \right|_{p \vee 2} \\ &\leq C_{p \vee 2} n^{-\frac{1}{2}}, \end{aligned} \quad (4.16)$$

provided that  $\varphi \circ (G^{(1)}, \dots, G^{(d)}) \circ \Phi$  is of finite  $\mathcal{L}^{\frac{1}{p \vee 2}}$ -norm. This is indeed the case as we have, owing to the bound (2.4) on  $\varphi$  and to the bound (4.3) on  $\ell$ ,

$$\begin{aligned} \sup_{\mathcal{A}^{\mathcal{Z}}} \frac{\varphi \circ (G^{(1)}, \dots, G^{(d)}) \circ \Phi}{\mathcal{L}^{\frac{1}{p \vee 2}}} &\leq 2 \sum_{i=1}^d \sup_{\mathcal{A}^{\mathcal{Z}}} \left( \frac{\ell(G^{(i)} \circ \Phi^{(i)})}{(G^{(i)} \circ \Phi^{(i)})^\alpha \mathcal{L}^{\frac{1}{p \vee 2}}} \vee \frac{\ell(1 - G^{(i)} \circ \Phi^{(i)})}{(1 - G^{(i)} \circ \Phi^{(i)})^\alpha \mathcal{L}^{\frac{1}{p \vee 2}}} \right) \\ &\leq 2C_{\ell, \epsilon} \sum_{i=1}^d \sup_{\mathcal{A}^{\mathcal{Z}}} \left( \frac{(G^{(i)} \circ \Phi^{(i)})^{-(p \vee 2)(\alpha + \epsilon)} \vee (1 - G^{(i)} \circ \Phi^{(i)})^{-(p \vee 2)(\alpha + \epsilon)}}{\mathcal{L}} \right)^{\frac{1}{p \vee 2}} \\ &< +\infty, \end{aligned}$$

since  $-(p \vee 2)(\alpha + \epsilon) > q_{\max}$  for  $\epsilon$  small enough (remind that  $p \vee 2 > -\frac{q_{\max}}{\alpha}$ ).

The theorem is proved using the estimates (4.15) and (4.16) in (4.13), and observing that the bound (4.15) dominates that of (4.16).  $\square$

## 5 Conclusion

In this work, we studied the theoretical and numerical properties of a transform MCMC scheme developed to efficiently compute expectations, conditional to rare events, in which the unconditional distribution is given by an intractable copula. The intractability of the copula arises from the fact that it is defined as a copula of a different random vector, for example obtained with a factor model, which leads to the so-called factor copulas. Under mild and natural hypotheses, we are able to derive the convergence rates for our proposed estimators, which are reproduced in a simulation example. We also revisit the computation of a challenging statistic originated in the financial risk management literature.

## A Appendix

### A.1 Geometric convergence rate for Markov chains

We recall here some useful results.

**Theorem A.1** ([MT09, Theorem 15.0.1], [FM03, Proposition 2]). *Assume that the Markov kernel  $\mathcal{P}$  satisfies Assumptions (A-i) and (A-ii). Then, for any  $p \geq 2$ , for any function  $\Phi : \mathcal{A}^{\mathcal{Z}} \subset \mathbb{R}^D \rightarrow \mathbb{R}$  bounded in  $\mathcal{L}^{1/p}$ -norm,*

$$\left| \frac{1}{n} \sum_{k=1}^n \Phi(\mathcal{Z}_k) - \mathbb{E}(\Phi(\mathcal{Z})) \right|_p \leq C_{p,\mathcal{P}} \left( \sup_{\mathcal{A}^{\mathcal{Z}}} \frac{|\Phi|}{\mathcal{L}^{\frac{1}{p}}} \right) \frac{\mathcal{L}^{1/p}(\mathcal{Z}_0)}{\sqrt{n}} \quad (\text{A.1})$$

for a constant depending only on the Markov kernel  $\mathcal{P}$  and the norm exponent  $p$ .

In addition, there exists  $\rho \in (0, 1)$  and a finite constant  $C_{(\text{A.2})} > 0$  such that for any function measurable function  $\Phi : \mathcal{A}^{\mathcal{Z}} \rightarrow \mathbb{R}$  bounded in  $\mathcal{L}$ -norm,

$$|\mathbb{E}(\Phi(\mathcal{Z}_n)) - \mathbb{E}(\Phi(\mathcal{Z}))| \leq C_{(\text{A.2})} \left( \sup_{\mathcal{A}^{\mathcal{Z}}} \frac{|\Phi|}{\mathcal{L}} \right) \mathcal{L}(\mathcal{Z}_0) \rho^n. \quad (\text{A.2})$$

For an up-to-date overview of the literature, see the recent book [DMPS18], in particular Chapter 15.

## A.2 Lyapunov functions for some usual schemes

### A.2.1 The preconditioned Crank-Nicolson sampler

In our applications,  $\mathcal{Z}$  is often Gaussian distributed (possibly with restriction on a set  $\mathcal{A}^{\mathcal{Z}}$  in the case of conditional computations). We collect here related MCMC results available in this Gaussian context. Define

$$\begin{aligned} \mathcal{Z}_k &= (\kappa \mathcal{Z}_{k-1} + \varepsilon_k) \mathbf{1}_{\mathcal{A}^{\mathcal{Z}}}(\kappa \mathcal{Z}_{k-1} + \varepsilon_k) + \mathcal{Z}_{k-1} \mathbf{1}_{(\mathcal{A}^{\mathcal{Z}})^c}(\kappa \mathcal{Z}_{k-1} + \varepsilon_k), \quad k \geq 1, \\ \mathcal{Z}_0 &\in \mathcal{A}^{\mathcal{Z}} \quad \text{deterministic,} \end{aligned} \quad (\text{A.3})$$

where  $\kappa \in (0, 1)$  and  $\varepsilon_k \stackrel{d}{=} \mathcal{N}(0, (1 - \kappa^2)I_D)$  independently of everything else.

We assume that  $\mathcal{A}^{\mathcal{Z}}$  is a measurable subset of  $\mathbb{R}^D$  with nonempty interior.

Its Markov transition kernel is defined, for  $y \in \mathcal{A}^{\mathcal{Z}}$  and  $B \subset \mathcal{A}^{\mathcal{Z}}$  measurable, by

$$\mathcal{P}(y, B) = \int_{\mathcal{A}^{\mathcal{Z}}} \mathbf{1}_B(z) p(y, z) dz + \delta_y(B) \int_{(\mathcal{A}^{\mathcal{Z}})^c} p(y, z) dz,$$

where  $p$  is the (unconditional) transition density (w.r.t. the Lebesgue measure), given by, for  $y, z \in \mathbb{R}^D$ ,

$$p(y, z) = \exp \left( - \sum_{j=1}^D \frac{|z_j - \kappa y_j|^2}{2(1 - \kappa^2)} \right) (2\pi(1 - \kappa^2))^{-D/2}.$$

Since  $\pi_{\mathcal{A}^{\mathcal{Z}}}(z)p(y, z) = \pi_{\mathcal{A}^{\mathcal{Z}}}(y)p(z, y)$  for all  $y, z \in \mathcal{A}^{\mathcal{Z}}$ , where  $\pi_{\mathcal{A}^{\mathcal{Z}}}$  (defined in (3.4)) is the density with respect to the Lebesgue measure of the standard Gaussian distribution  $\mathcal{N}(0, I_D)$  restricted to  $\mathcal{A}^{\mathcal{Z}}$ , this process is an autoregressive process with acceptance-rejection and which stationary distribution has density  $\pi_{\mathcal{A}^{\mathcal{Z}}}$ .

In the MCMC literature, the sampler (A.3) is called preconditioned Crank-Nicolson (pCN). It can be defined in infinite dimension too. This is not the only MCMC scheme available: let us mention the random walk Metropolis (RWM), the Metropolis Adjusted Langevin Algorithm (MALA), Hybrid Monte-Carlo (HMC) among some well known schemes, see [CRSW13] for a broad overview. However, pCN is among the simplest ones and it is more robust to increasing dimension and generally improves upon naive RWM, see [CRSW13] and [HSV14]. We now assume, in addition, that there exists  $\delta \in (0, 1)$  such that

$$\sup_{y \in \mathcal{A}^{\mathcal{Z}}} \int_{(\mathcal{A}^{\mathcal{Z}})^c} p(y, z) dz \leq \delta. \quad (\text{A.4})$$

The above condition, which reads as a lower bound on the acceptance rate, is standard in the MCMC literature, see [FGM17, Corollary 3]. The condition below, although not surprising, is seemingly less usual to the best of our knowledge: we could not find a reference.

**Proposition A.2** (The quadratic exponential Lyapunov function). *The Markov kernel  $\mathcal{P}$  is a  $\Psi$ -irreducible aperiodic transition kernel on  $\mathcal{A}^{\mathcal{Z}}$  for which any compact subset  $\mathcal{C}$  of  $\mathcal{A}^{\mathcal{Z}}$  with nonzero Lebesgue measure is an accessible petite set.*

Let  $s \in (0, \frac{1}{2})$  and set

$$\mathcal{L}(z) = e^{s \sum_{i=1}^D |z_i|^2}.$$

*This is a Lyapunov function satisfying the drift condition (2.1) and results of Theorem A.1 holds with this  $\mathcal{L}$ .*

A similar statement with  $\mathcal{L}(z) = e^{s \sum_{i=1}^D |z_i|}$  would be presumably much more standard (see e.g. [DMPS18, p.318–319]), but having a quadratic exponential growth instead of linear one plays an important role for enlarging the scope of applicability of our method (in view of assumption (A-iii)).

*Proof.* The proof that  $\mathcal{P}$  is  $\Psi$ -irreducible, aperiodic, and that any compact subset with positive Lebesgue measure is an accessible petite set is standard and thus omitted.

It remains to prove that (2.1) holds and that  $\{\mathcal{L} \leq \nu\}$  is 1-small for some  $\nu > \frac{b}{1-\rho(\mathcal{P})}$ .

We first prove (2.1). Starting from the quadratic exponential form for  $\mathcal{L}$  and using (A.4), we get

$$\begin{aligned} \mathcal{P}\mathcal{L}(y) &= \int_{\mathcal{A}^{\mathcal{Z}}} \mathcal{L}(z) p(y, z) dz + \mathcal{L}(y) \int_{(\mathcal{A}^{\mathcal{Z}})^c} p(y, z) dz \\ &\leq \mathcal{L}(y) \left( \int_{\mathbb{R}^D} p(y, z) \frac{\mathcal{L}(z)}{\mathcal{L}(y)} dz + \delta \right) \\ &= \mathcal{L}(y) \left( \prod_{j=1}^D \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(1-\kappa^2)}} \exp\left(-\frac{(z_j - \kappa y_j)^2}{2(1-\kappa^2)}\right) e^{s|z_j|^2 - s|y_j|^2} dz_j + \delta \right). \end{aligned} \quad (\text{A.5})$$

A direct computation shows

$$-\frac{(z_j - \kappa y_j)^2}{2(1-\kappa^2)} + s|z_j|^2 - s|y_j|^2$$

$$\begin{aligned}
&= -\left(\frac{1}{2(1-\kappa^2)} - s\right)z_j^2 + \frac{\kappa y_j z_j}{(1-\kappa^2)} - \frac{\kappa^2 y_j^2}{2(1-\kappa^2)} - s|y_j|^2 \\
&= -\left(\frac{1}{2(1-\kappa^2)} - s\right) \left(z_j - \frac{\kappa y_j}{2(1-\kappa^2)} \frac{1}{\left(\frac{1}{2(1-\kappa^2)} - s\right)}\right)^2 - \frac{\kappa^2 y_j^2}{2(1-\kappa^2)} - s|y_j|^2 \\
&\quad + \left(\frac{1}{2(1-\kappa^2)} - s\right) \left(\frac{\kappa y_j}{2(1-\kappa^2)} \frac{1}{\left(\frac{1}{2(1-\kappa^2)} - s\right)}\right)^2 \\
&= -\left(\frac{1}{2(1-\kappa^2)} - s\right) (z_j - [\dots]y_j)^2 - \frac{\kappa^2 y_j^2}{2(1-\kappa^2)} - s|y_j|^2 + \frac{\kappa^2 y_j^2}{2(1-\kappa^2)(1-2s(1-\kappa^2))}, \\
-\theta &:= -\frac{\kappa^2}{2(1-\kappa^2)} - s + \frac{\kappa^2}{2(1-\kappa^2)(1-2s(1-\kappa^2))} \\
&= \frac{-\kappa^2(1-2s(1-\kappa^2)) - 2s(1-\kappa^2)(1-2s(1-\kappa^2)) + \kappa^2}{2(1-\kappa^2)(1-2s(1-\kappa^2))} \\
&= \frac{2\kappa^2 s(1-\kappa^2) - 2s(1-\kappa^2)(1-2s(1-\kappa^2))}{2(1-\kappa^2)(1-2s(1-\kappa^2))} \\
&= \frac{\kappa^2 s - s(1-2s(1-\kappa^2))}{(1-2s(1-\kappa^2))} \\
&= -s \frac{(1-\kappa^2) - 2s(1-\kappa^2)}{(1-2s(1-\kappa^2))} = -\frac{(1-\kappa^2)s(1-2s)}{(1-2s(1-\kappa^2))},
\end{aligned}$$

hence  $\theta > 0$  since  $1 - 2s(1 - \kappa^2) > 1 - 2s > 0$  and  $1 - \kappa^2 > 0$ . Plugging this in the previous equalities and integrating in  $z_j$  the new Gaussian density in (A.5), we get

$$\mathcal{P}\mathcal{L}(y) \leq \mathcal{L}(y) \left( \frac{1}{(1-2s(1-\kappa^2))^{D/2}} e^{-\theta|y|^2} + \delta \right).$$

Let  $\rho^{(\mathcal{P})} \in (0, 1)$  be given arbitrarily. Because  $\theta > 0$  and  $0 < \delta < 1$ , for any  $|y| > R$  with  $R$  large enough, we get

$$\mathcal{P}\mathcal{L}(y) \leq \rho^{(\mathcal{P})} \mathcal{L}(y), \quad \forall |y| > R.$$

On the other hand,  $\mathcal{P}\mathcal{L}(y)$  is readily uniformly bounded on the compact set  $\{y : |y| \leq R\}$ . Therefore, the inequality (2.1) is proved.

We now prove the small set assumption. Let  $\nu > \frac{b}{1-\rho^{(\mathcal{P})}}$ , and notice that  $\mathcal{C} := \{z \in \mathcal{A}^{\mathcal{Z}} : \mathcal{L}(z) \leq \nu\} = \{z : |z| \leq s^{-\frac{1}{2}} \sqrt{\ln(\nu)}\} \cap \mathcal{A}^{\mathcal{Z}}$ , so that  $\int_{\mathcal{C}} \pi(z) dz > 0$  for  $\nu$  large enough, where  $\pi$  is the standard Gaussian density (3.4). We set

$$\pi_{\mathcal{C}}(dz) := \frac{\mathbf{1}_{\mathcal{C}} \pi(z) dz}{\int_{\mathcal{C}} \pi(z) dz},$$

and we compute, for  $y \in \mathcal{C}$  and any measurable subset  $B$  of  $\mathcal{A}^{\mathcal{Z}}$ ,

$$\mathcal{P}(y, B) \geq \int_{\mathcal{A}^{\mathcal{Z}}} \mathbf{1}_B(z) p(y, z) dz$$

$$\begin{aligned}
&= \int_{\mathcal{A}^{\mathcal{Z}}} \mathbf{1}_B(z) \frac{p(y, z)}{\pi(z)} \pi(z) dz \\
&\geq \inf_{(y, z) \in \mathcal{C}^2} \left( \frac{p(y, z)}{\pi(z)} \right) \int_{\mathcal{C}} \mathbf{1}_B(z) \pi(z) dz = \inf_{(y, z) \in \mathcal{C}^2} \left( \frac{p(y, z)}{\pi(z)} \right) \pi(\mathcal{C}) \pi_{\mathcal{C}}(B).
\end{aligned}$$

Since  $\mathcal{C}$  is bounded, it is clear that the above infimum is nonzero, and the proof is complete.  $\square$

### A.2.2 The independent sampler

Here the sequence of  $(\mathcal{Z}_k)_{k \geq 1}$  is i.i.d. with distribution  $\pi$ :

$$\mathcal{Z}_k \stackrel{d}{=} \pi(.), \quad \mathcal{P}(z, dz') = \pi(dz').$$

We quickly justify the following.

**Proposition A.3.** *The resulting “Markov” kernel  $\mathcal{P}$  satisfies the assumption (A-i) with some Lyapunov function  $\mathcal{L}$  given in the proof.*

*Proof.* First, obviously there is a closed ball  $\mathcal{C}_0 = B(0, R_0)$  with radius  $R_0 > 0$  such that  $\pi(\mathcal{C}_0) > 0$ , provided that  $R_0$  is large enough.

- $\mathcal{P}$  is  $\pi$ -irreducible [MT09, Proposition 4.2.1 (ii)], since

$$A \in \mathcal{B}(\mathbb{R}^D), \quad z \in \mathbb{R}^D : \quad \pi(A) > 0 \implies \mathcal{P}(z, A) > 0$$

(owing to  $\mathcal{P}(z, A) = \pi(A)$ ).

- The set  $\mathcal{C}_0$  is a  $\pi_1$ -small set [MT09, Section 5.2], hence a petite set [MT09, Proposition 5.5.3] since obviously,

$$A \in \mathcal{B}(\mathbb{R}^D), \quad z \in \mathcal{C}_0 : \quad \mathcal{P}(z, A) \geq \pi(A).$$

- $\mathcal{P}$  is strongly aperiodic (and thus aperiodic), as a consequence of the existence of the  $\pi_1$ -small set  $\mathcal{C}_0$  with  $\pi_1(\mathcal{C}_0) > 0$  and irreducibility, see [MT09, p. 114].
- The set  $\mathcal{C}_0$  is clearly accessible [MT09, p. 86] since from any starting point  $z \in \mathbb{R}^D$ , the probability of hitting  $\mathcal{C}$  in finite time is strictly positive  $\mathbb{P}[\tau_{\mathcal{C}} < +\infty] \geq \pi(\mathcal{C}_0) > 0$ .

We now exhibit a candidate for the Lyapunov function  $\mathcal{L}$ , which goes from  $\mathbb{R}^D$  to  $[1, \infty)$ , is radial and non-decreasing w.r.t. the radius, goes to infinity at infinity, and such that  $\mathbb{E}(\mathcal{L}(\mathcal{Z})) \leq 2$ . We distinguish two cases.

1.  $\mathcal{Z}$  is a bounded random variable, hence integrable. Define  $\mathcal{L}(z) = 1 + \frac{|z|}{1 + \mathbb{E}(|\mathcal{Z}|)}$ : this function  $\mathcal{L}$  clearly meets our requirements.

2.  $\mathcal{Z}$  is unbounded. The construction of  $\mathcal{L}$  requires more work. For any  $t \geq 0$ , there is a  $l(t) \geq 0$  such that

$$\mathbb{P}[|\mathcal{Z}| \leq l(t)] \geq 1 - \frac{1}{(1+t)^2}. \quad (\text{A.6})$$

We take the smallest of such  $l(t)$ : this is the quantile of  $|\mathcal{Z}|$  at level  $1 - \frac{1}{(1+t)^2} \in [0, 1]$ . The function  $l$  is càglàd non-decreasing, from  $\mathbb{R}^+$  to  $\mathbb{R}^+$  [EH13, Proposition 2.3, item (2)]. Since  $\mathcal{Z}$  is unbounded,  $l(t) \rightarrow +\infty$  as  $t \rightarrow +\infty$ .

Set  $\mathcal{L}(z) := l^{-1}(|z|) + 1$  where  $l^{-1}$  is the generalized inverse of  $l$  (see [EH13]):  $\mathcal{L}$  is a measurable function from  $\mathbb{R}^D$  to  $[1, \infty)$ , going to infinity at infinity. Then,

$$\begin{aligned} \mathbb{E}(\mathcal{L}(\mathcal{Z}) - 1) &= \int_0^\infty \mathbb{P}[\mathcal{L}(\mathcal{Z}) - 1 > t] dt \\ &= \int_0^\infty \mathbb{P}[l^{-1}(|\mathcal{Z}|) > t] dt \leq \int_0^\infty \mathbb{P}[|\mathcal{Z}| > l(t)] dt \leq \int_0^\infty \frac{1}{(1+t)^2} dt = 1, \end{aligned}$$

where we have used at the first inequality [EH13, Proposition 2.3, item (5)] and (A.6) at the second one. Therefore,  $\mathbb{E}(\mathcal{L}(\mathcal{Z})) \leq 2$ .

Now, taking advantage of  $\mathbb{E}(\mathcal{L}(\mathcal{Z})) \leq 2$ , we verify the inequality  $\mathcal{P}\mathcal{L} \leq \rho^{(\mathcal{P})}\mathcal{L} + b\mathbf{1}_{\mathcal{C}}$  for some  $\rho^{(\mathcal{P})}, b, \mathcal{C}$ . Actually we can take any  $\rho^{(\mathcal{P})} \in (0, 1)$ ,  $b = 2$  and  $\mathcal{C} = \mathcal{C}_0$  with  $R_0$  large enough to have  $\mathcal{L}(z) \geq 2/\rho^{(\mathcal{P})}$  for any  $z$  such that  $|z| = R_0$ : indeed,

- on the one hand for  $z \in \mathcal{C}_0$ ,

$$\mathcal{P}\mathcal{L}(z) = \mathbb{E}(\mathcal{L}(\mathcal{Z})) \leq 2 = b\mathbf{1}_{\mathcal{C}_0}(z) \leq \rho^{(\mathcal{P})}\mathcal{L}(z) + b\mathbf{1}_{\mathcal{C}_0}(z);$$

- on the other hand, for  $z \notin \mathcal{C}_0$  (i.e.  $|z| > R_0$ ), we have

$$\mathcal{P}\mathcal{L}(z) = \mathbb{E}(\mathcal{L}(\mathcal{Z})) \leq 2 \leq \rho^{(\mathcal{P})}\mathcal{L}\left(\frac{z}{|z|}R_0\right) \leq \rho^{(\mathcal{P})}\mathcal{L}(z) = \rho^{(\mathcal{P})}\mathcal{L}(z) + b\mathbf{1}_{\mathcal{C}_0}(z).$$

We are done. □

### A.3 The skew-t distribution

We follow [Han94] and for  $\nu \in (2, \infty)$  and  $\lambda \in (-1, 1)$  we say that  $\mathcal{Z} \sim st(\nu, \lambda)$  has skew- $t$  distribution with  $\nu$  degrees of freedom and parameter of asymmetry  $\lambda$  if its density is given by

$$f_{\mathcal{Z}}(z; \nu, \lambda) = \begin{cases} bc \left(1 + \frac{1}{\nu-2} \left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2}, & z < -a/b \\ bc \left(1 + \frac{1}{\nu-2} \left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-(\nu+1)/2}, & z \geq -a/b \end{cases},$$

where  $a = 4\lambda c \left(\frac{\nu-2}{\nu-1}\right)$ ,  $b = \sqrt{1 + 3\lambda^2 - a^2}$  and  $c = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi(\nu-2)}}$ .

## References

- [AB01] S.-K. Au and J.L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic engineering mechanics*, 16(4):263–277, 2001.
- [AC02] A. Ang and J. Chen. Asymmetric correlations of equity portfolios. *Journal of financial Economics*, 63(3):443–494, 2002.
- [AG07] S. Asmussen and P.W. Glynn. *Stochastic simulation: Algorithms and analysis*. Stochastic Modelling and Applied Probability 57. New York, NY: Springer, 2007.
- [BEGG16] J. Beirlant, M. Escobar-Bach, Y. Goegebeur, and A. Guillou. Bias-corrected estimation of stable tail dependence function. *Journal of Multivariate Analysis*, 143:453 – 466, 2016.
- [BGT87] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular variation*, volume 27. Cambridge University Press, 1987.
- [BZ16] M. Bartels and F.A. Ziegelmann. Market risk forecasting for high dimensional portfolios via factor copulas with gas dynamics. *Insurance: Mathematics and Economics*, 70:66–79, 2016.
- [CDFG12] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte-Carlo for rare event estimation. *Stat. Comput.*, 22(3):795–808, 2012.
- [CLV04] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula methods in finance*. John Wiley & Sons, 2004.
- [CMS15] H. Chen, R. MacMinn, and T. Sun. Multi-population mortality models: A factor copula approach. *Insurance: Mathematics and Economics*, 63:135–146, 2015.
- [CRSW13] S.L. Cotter, G.O. Roberts, A.M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 2013.
- [DF07] L. De Haan and A. Ferreira. *Extreme value theory: an introduction*. Springer Science & Business Media, 2007.
- [DMPS18] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains*. Springer, 2018.
- [Dom03] D.E. Dominici. The inverse of the cumulative standard normal probability function. *Integral Transforms and Special Functions*, 14(4):281–292, 2003.
- [dR93] L. de Haan and S.I. Resnick. Estimating the limit distribution of multivariate extremes. *Communications in Statistics. Stochastic Models*, 9(2):275–309, 1993.

- [DR12] G. De Luca and G. Riveccio. Multivariate tail dependence coefficients for Archimedean copulae. In *Advanced Statistical Methods for the Analysis of Large Data-Sets*, pages 287–296. Springer, 2012.
- [EH13] P. Embrechts and M. Hofert. A note on generalized inverses. *Math. Methods Oper. Res.*, 77(3):423–432, 2013.
- [Fal19] M. Falk. *Multivariate extreme value theory and D-norms*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2019.
- [FGM17] G. Fort, E. Gobet, and E. Moulines. MCMC design-based non-parametric regression for rare-event. Application to nested risk computations. *Monte Carlo Methods and Applications*, 23(1):21–42, 2017.
- [FM03] G. Fort and E. Moulines. Convergence of the Monte Carlo expectation maximization for curved exponential families. *Annals of Statistics*, 31(4):1220–1259, 2003.
- [Gab05] G. Gabbi. Semi-correlations as a tool for geographical and sector asset allocation. *European Journal of Finance*, 11(3):271–281, 2005.
- [GL15] E. Gobet and G. Liu. Rare event simulation using reversible shaking transformations. *SIAM Scientific Computing*, 37(5):A2295–A2316, 2015.
- [Han94] B.E. Hansen. Autoregressive conditional density estimation. *International Economic Review*, 35(3):705–730, 1994.
- [HSV14] M. Hairer, A.M. Stuart, and S.J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014.
- [Hua92] X. Huang. *Statistics of bivariate extreme values*. Thesis Publishers Amsterdam, 1992.
- [Joe14] H. Joe. *Dependence modeling with copulas*. CRC press, 2014.
- [MDCL13] J.S. Murray, D.B. Dunson, L. Carin, and J.E. Lucas. Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association*, 108(502):656–665, 2013.
- [MFE10] A. J McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, 2010.
- [MT09] S. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009.
- [OP17] D.H. Oh and A.J. Patton. Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, 35(1):139–154, 2017.



- [Ros11] J.S. Rosenthal. Optimal proposal distributions and adaptive MCMC. In *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC Handb. Mod. Stat. Methods, pages 93–111. CRC Press, Boca Raton, FL, 2011.
- [RUT<sup>+</sup>20] J.A. Ryan, J.M. Ulrich, W. Thielen, P. Teetor, and S. Bronder. *quantmod: Quantitative Financial Modelling Framework*, 2020. R package version 0.4.17.