Tail Probability Estimation of Factor Models with Regularly-Varying Tails: Asymptotics and Efficient Estimation

Farzad Pourbabaee [†] and Omid Shams Solari [‡]

April 2022

Abstract

We study the tail probability of linear factor models generated from non-identically distributed components with *regularly-varying* tails, a large subclass of heavy-tailed distributions. An efficient sampling method for tail probability estimation for this class is introduced and theoretically shown to exponentially outperform the crude Monte-Carlo estimator, in terms of the coverage probability and the confidence interval's length. The theoretical results are empirically validated through stochastic simulations on independent non-identically Pareto distributed factors. The proposed estimator is available as part of a more comprehensive TPE package.

1 Introduction

Tail probability (TP) estimation is a well-studied problem in various branches of research; from finance and economics, to particle physics and weather forecasting. Researchers are often interested in the probability of occurrence of catastrophes, namely the major over-shoots or under-shoots of an outcome comprised of a few input resources. A prominent example is the TP estimation of the factor models. A well-studied case is estimating the TP of sums of i.i.d. random variables. Namely, estimating $\mathsf{P}\left(X_1 + \cdots + X_N > x\right)$ for finite N, when x is very large. The main contribution of this paper is to develop an estimator for

Authors contributed equally to this manuscript. We gratefully acknowledge Robert M. Anderson for the support and guidance over the course of this paper, and thank Lisa Goldberg, Markus Pelger, Alex Shkolnik and Johan Walden for the valuable discussions.

[†]California Institute of Technology. Email: far@caltech.edu.

[‡]University of California, Berkeley. Email: solari@berkeley.edu.

the TP of the sum of independent (but not necessarily identical) – henceforth referred to by '*ind*' – random variables that have **regularly-varying tails**.

Such TP estimation is well-explored when the distribution of factors are thin-tailed and/or their class of distribution functions is stable under addition, e.g., Gaussian or Gamma factors. The statistical analysis is particularly straightforward in these cases, because of the available closed-form expressions for the right or left tail probability. However, the majority of cases do not fall in this line, as in many cases this stability does not hold, and we cannot appeal to analytical expressions for the tail probability. An important example is the class of heavy-tailed distributions. Loosely speaking, for this class of random variables the rare events occur more frequently than in a light-tailed distribution such as Gaussian. Gabaix (2016) enumerated many examples in which Power law distribution (a prominent member of heavy-tailed random variables) emerges, such as firm and city size, income and wealth distribution, and CEO compensations. Asmussen et al. (2000) proposed the first efficient algorithm for TP estimation of linear factor models with heavy-tailed i.i.d. components. They introduce a *Conditional Monte-Carlo* algorithm which benefits from conditioning on the order statistics. Chan and Kroese (2011) utilize the same estimator of Asmussen et al. (2000) in specific setting of ind random variables, where the factors' distribution are restricted to be either Weibull or Pareto. Our paper establishes the bounded relative error for the larger class of ind random variables with regularly-varying (RV) tails, thus encompassing these previous works. Additionally, we express a finite sample result, showing that this estimator concentrates around to the true tail probability exponentially faster than the crude Monte-Carlo.

Conditioning is quite appealing since the classical Monte-Carlo methods for estimating TP fall short, precisely because a large number of samples need to be drawn to get non-zero realizations of the sampling event. However, in some settings there are more efficient tools to address this problem, such as *importance sampling*. The idea is essentially to sample from another probability measure that assigns more weight to the regions where the sampling function takes larger values, and then correct for the transformation of the sampling measure. Ackerberg (2009) showed that importance sampling can reduce the computational burden for smoothing the simulated moments, as first suggested by McFadden (1989).

However, for the case of heavy-tailed distributions, the common measure transformation methods such as importance sampling are not quite applicable. This is mainly because the systematic approach in importance sampling relies on the existence of the moment generating function, and leveraging it as the Radon-Nikodym derivative to find a candidate distribution that assigns more weight to improbable events. However, for many heavy-tailed distributions (e.g., Pareto) the moment generating function do not exist entirely on the real line. Secondly,

due to the degeneracy of the likelihood ratio in high-dimensional models, these methods are less useful (Rubinstein and Kroese (2016)). In this paper, a novel technique (based on the conditional Monte-Carlo sampling) is introduced to address the problem of TP estimation for the sum of ind random variables belonging to a large subclass of heavy-tailed distributions, namely regularly-varying (RV) random variables.

In the context of insurance risk, Goovaerts et al. (2005) studies the tail asymptotics of randomly weighted sum of i.i.d. Pareto factors. Further, Foss and Richards (2010) find asymptotic results for the sum of conditionally independent factors under rather stringent conditions on the structure of factors' dependencies. Albrecher et al. (2006) and Kortschak and Albrecher (2009) use Copulas to capture the dependence structure of the factors, and derived similar asymptotic results for the tail probability. The main contributions of our paper are to provide asymptotics for the tail probability of the sum and maximum of independent, $\mathbb{R}$-valued, RV random variables; and to propose an improved Monte-Carlo method for estimating these probabilities.

Probability estimation of such extreme events arises in many places: notably the extreme losses or profits of a portfolio exposed to multiple independent risk factors. Another example studied in Acemoglu et al. (2017) is the frequency of large economic downturns, and significant GDP departures from equilibrium trend, caused by the heavy tail nature of micro shocks, wherein independent factors with Pareto tails add up and create large swings. The challenge is that for all these cases, the extreme tail probabilities are excessively small.[1] Therefore, finding non-trivial confidence intervals for them is not just a matter of their size, but more importantly how big or small are they relative to the sought probability. Namely, it is the relative error, the length of the confidence interval divided by the point estimate, that matters for reporting the estimation precision. The proposed estimation method in this paper (based on the conditional Monte-Carlo) achieves a *bounded relative error* (contrary to the crude Monte-Carlo estimator) as the probability of the tail event shrinks.

The paper is organized as follows. In section 2, we provide motivations for the estimation of tail probability and briefly review a systematic application of importance sampling for Gaussian factors. We further argue why such a method falls short for heavy-tailed distributions, thereby setting the stage for our proposed estimator. In section 3, we establish some results on the asymptotics of the tail probability of RV sums, and present our proposed conditional Monte-Carlo algorithm, along with its concentration analysis and comparisons with the crude Monte-Carlo. Section 4 studies the implications for portfolios of many assets

[1] A non-exhaustive list of other economic research relying on the sums of many factors includes the Pareto distribution of productivity process in Lucas Jr. and Moll (2014), high dimensional decision making under many additive factors Pourbabaee (2021) and the relation between interest rates, economic growth and the tail of income distribution by Piketty and Saez (2014).

with heavy-tailed components. In section 5, the efficiency of our proposed estimator relative to the crude Monte-Carlo is demonstrated through the simulations. The proofs as well as simulation details are presented in the appendix.

2 Gaussian Factor Model

In this section, we present a brief overview of the application of importance sampling as a method of variance reduction in the estimation of tail probability under Gaussian factors. We further study the conditions that prevent using importance sampling and thus setting the stage for Conditional Monte-Carlo in the upcoming sections.

Assume there are M assets available in the market whose returns are driven by k latent factors $\phi = (\phi_1, \ldots, \phi_k)$. The return to the i-th security η_i is captured as a linear combination of the latent factors and the idiosyncratic risk ε_i, which is assumed uncorrelated with ϕ:

$$\eta_i = \langle \beta_i, \phi \rangle + \varepsilon_i.$$

Asset returns are all evaluated over the time interval $[t, t+\tau]$, where τ is the investment horizon. Observations of high-frequency data confirm that the distribution of returns deviates more intensely from Gaussianity as the investment horizon becomes shorter. For the moment, suppose that τ is long enough that we can assume Normal distributions both for the factors and asset specific risks, in particular assume $\phi \sim \mathcal{N}(0, I_k)$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ are mutually independent. Let ξ represent the return to the market index, which is typically calculated as the market-cap weighted sum of security returns, but here for simplicity is taken as the uniform average of M returns:

$$\xi = \frac{1}{M} \sum_{i=1}^{M} \eta_i = \bar{\eta} = \langle \bar{\beta}, \phi \rangle + \bar{\varepsilon}, \tag{2.1}$$

which has the Normal distribution $\mathcal{N}\left(0, \|\bar{\beta}_2\|^2 + \sum_{i=1}^{M} \sigma_i^2 / M^2\right)$.

One can think of periods of market turmoil as the times when the market index reflects large downswings and upswings, namely $|\xi| > \lambda$, and one might want to estimate the probability of these large fluctuations, e.g., $\mathsf{P}(\xi > \lambda)$ for large λ. Since no closed form expression for this integral exists, we have to resort to simulation methods. However, crude Monte-Carlo sampling from the distribution of ξ requires drawing a large number of samples to find some that surpass the threshold λ. Importance sampling can help to reduce the required number of sample points, or alternatively reduce the variance of the point estimator. Given that

the cumulative generating function $\psi(\theta)$ exists for Gaussian distribution for all $\theta \in \mathbb{R}$, one possible choice to get an appropriate importance sampling distribution is the exponential measure change through

$$\psi(\theta) = \log \mathsf{E}\left[e^{\theta\xi}\right] = \frac{\theta^2}{2}\left(\|\bar{\beta}\|_2^2 + \frac{1}{M^2}\sum_{i=1}^{M}\sigma_i^2\right).$$

Specifically, If P denotes the actual probability measure for ξ, the exponentially twisted measure P_θ is obtained by

$$\frac{\mathrm{d}\mathsf{P}_\theta}{\mathrm{d}\mathsf{P}} = e^{\theta\xi - \psi(\theta)}. \tag{2.2}$$

Now we can generate n samples from P_θ, and form the following sample average, that represents the unbiased estimator under the new measure P_θ:

$$\frac{1}{n}\sum_{i=1}^{n} 1_{\{\xi_i > \lambda\}} \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{P}_\theta}(\xi_i).$$

Denote the per-sample estimator by $Z(\lambda) = 1_{\{\xi>\lambda\}} \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{P}_\theta}(\xi)$. The next definition spells out two notions of relative error.

Definition 2.1. The estimator $Z(\lambda)$ has *bounded relative error* if

$$\limsup_{\lambda\to\infty} \frac{\mathsf{Var}(Z(\lambda))}{\mathsf{E}\left[Z(\lambda)\right]^2} < \infty,$$

and is *logarithmically efficient* (a weaker notion) if for some $\varepsilon > 0$

$$\limsup_{\lambda\to\infty} \frac{\mathsf{Var}(Z(\lambda))}{\mathsf{E}\left[Z(\lambda)\right]^{2-\varepsilon}} = 0.$$

The following result, which is proved in Asmussen (2008), sheds light on the efficiency of exponential twisting for a certain value of θ.

Theorem 2.2. *The exponential change of measure in* (2.2) *is logarithmically efficient for the unique parameter* θ *that solves* $\lambda = \psi'(\theta)$.

As a result of this theorem, the optimal parameter for the measure change is

$$\theta^* = \frac{\lambda}{\|\bar{\beta}\|_2^2 + \sum_{i=1}^{M}\sigma_i^2/M^2}. \tag{2.3}$$

Having stated this theorem, the following lines summarize the simulation steps for estimating the probability of large fluctuations of market index in the Gaussian case:

(i) Find θ^* from (2.3).

(ii) Draw random samples ξ_i, $i = 1, \ldots, M$ from P_{θ^*}.

(iii) Calculate $\frac{1}{n}\sum_{i=1}^{n} 1_{\{\xi_i > \lambda\}}\, e^{\psi(\theta^*)-\theta^*\xi_i}$ as an estimator of $\mathsf{P}(\xi > \lambda)$.

As a result of twisting the sampling distribution, the relative error now scales as $\mathsf{P}(\xi > \lambda)^{-\varepsilon/2}$, compared to $\mathsf{P}(\xi > \lambda)^{-1/2}$ for the classical Monte-Carlo. Equivalently, this boost shows us how to achieve a certain level of relative error with fewer sample points. However, this machinery can not always be employed, because the moment generating function need not always exist. Therefore, to find the optimal measure change we have to appeal to heuristic methods, or use other Monte-Carlo methods as explained further in the proceeding sections.

3 Regularly-Varying Factors

In this section we study the consequences of dealing with independent factors that are assumed to have regularly-varying tails, for example the Pareto distribution. This class of distribution functions is contained in the larger family of sub-exponential distributions as defined below.

Definition 3.1. The distribution F of a non-negative random variable X is called sub-exponential, if

$$\lim_{x\to\infty} \frac{\mathsf{P}(X_1 + \cdots + X_N > x)}{\mathsf{P}(X_1 > x)} = N \text{ for all } N \geq 1, \tag{3.1}$$

where X_i's are i.i.d. copies drawn from F [2].

This definition extends to probability distributions on the entire real line by restriction to the positive and negative halves. Then, the random variable $X \sim F$, taking values in $\mathbb{R}$, is called sub-exponential if $X_+ = (X \vee 0)$ and $X_- = -(X \wedge 0)$ are both sub-exponentials. Equation (3.1) says that the probability that the sum of N i.i.d. sub-exponential random variables exceeds a certain threshold is roughly N times the probability that one of them exceeds that level. The question is thus what happens if the random variables are independent and individually sub-exponential but not necessarily identically distributed? Is the tail probability for the sum related to the sum of tail probabilities of the summands, and if so, under what conditions? As pointed out in the introduction, variations of these questions are studied under different conditions for the factors.

[2] For more information, check definition 1.3.3 in Embrechts et al. (2013).

In the remainder of this paper, we restrict ourselves to the case of sum of ind random variables that satisfy two additional conditions: (i) *h-condition* and (ii) regularly-varying tails.

Notation 3.2 (Asymptotic equivalence)**.** $f(x) \sim g(x)$ if $f(x)/g(x) \to 1$, as $x \to \infty$.

Condition 3.3 (h-condition)**.** Given the distribution F, there exists an eventually increasing function $h(x)$ such that $\lim_{x\to\infty} h(x) = \infty$ and

$$\lim_{x\to\infty} \frac{\bar{F}(x + h(x))}{\bar{F}(x)} = 1, \text{ where } \bar{F}(x) := 1 - F(x).$$

Definition 3.4 (Regularly-Varying (RV) distribution)**.** A distribution function F has a *regularly-varying* tail, if $\bar{F}(x) \sim L(x)/x^{\alpha}$ as $x \to \infty$, where $\alpha > 0$ and $L(\cdot)$ varies *slowly* at infinity, namely

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1 \text{ for all } t > 0.$$

For more details about the notion of RV distribution check Feller (2008).

Example 3.5. Suppose X has Power law distribution with coefficient μ, i.e., $\mathsf{P}(X > x) \propto x^{-\mu}$. Then, one can check that $h(x) = x^{\delta}$, for any $0 < \delta < 1$, satisfies the h-condition.

Functions such as $\log(x)$, $\log(\log(x))$ and any convergent function to a bounded level are examples of slow-variation. The RV property depends only on the behavior of the distribution at infinity, so it does not matter how it behaves at intermediate points. One stylized observation about this family of distributions is that they have finite moments of order less than α, but not higher. This prevents us from using moment generating function (and thus toward a methodological search for an exponential measure change – such as the one in previous section for Gaussian factors).

Claim 3.6. For all distribution functions of regular variation, we can take $h(x) = x^{\delta}$ with any $0 < \delta < 1$, and the h-condition will hold. The corollary of theorem 1 in section 8.8 of Feller (2008) paves the way to prove this claim, which allows us to represent the slowly varying function $L(\cdot)$ as

$$L(x) = a(x) \exp\left(\int_1^x \frac{\varepsilon(y)}{y} dy\right),$$

where $\varepsilon(x) \to 0$ and $a(x) \to c$ as $x \to \infty$. Therefore,

$$\lim_{x\to\infty} \frac{L(x + x^{\delta})}{L(x)} = \lim_{x\to\infty} \frac{a(x + x^{\delta})}{a(x)} \lim_{x\to\infty} \exp\left(\int_x^{x+x^{\delta}} \frac{\varepsilon(y)}{y} dy\right),$$

where the first term converges to 1, and the second term's exponent is approaching zero, because

$$\left| \int_x^{x+x^\delta} \frac{\varepsilon(y)}{y} dy \right| \leq \frac{\sup_{y \in (x, x+x^\delta)} |\varepsilon(y)|}{x^{1-\delta}} \to 0, \quad \text{as } x \to \infty.$$

Remark 3.7. The next two results on the tail probability of the sum and the maximum of a sequence of ind random variables are built on the well-known properties of i.i.d. $\mathbb{R}^+$-valued RV random variables in Embrechts et al. (2013). Precisely, for non-negative, sub-exponential and i.i.d. random variables $(X_i)_{i=1,\cdots,N}$, $\mathsf{P}(X_1 + \cdots + X_N > x) \sim \mathsf{P}(\max_{1 \leq i \leq N} X_i > x) \sim N\mathsf{P}(X_i > x)$ when $x \to \infty$.

The following theorem assumes condition 3.3, and expresses an asymptotic equivalence result for the tail probability of an independent RV sum. Its proof is relegated to the appendix.

Theorem 3.8. *Suppose $X_1, \ldots, X_N$ are independent random variables in $\mathbb{R}$, such that:*

(i) An RV distribution F exists, where $\bar{F}_i(x) \sim c_i \bar{F}(x)$ for all i's and at least one $c_i \neq 0$;

(ii) A function $h(\cdot)$ exists that satisfies the h-condition in 3.3 for F,

then the following asymptotic result holds:

$$\mathsf{P}(X_1 + \cdots + X_N > x) \sim \sum_{i=1}^{N} \mathsf{P}(X_i > x) \sim \left(\sum_{i=1}^{N} c_i \right) \bar{F}(x). \tag{3.2}$$

Another interesting feature of sub-exponential distributions (and consequently RV distribution) is the so-called *catastrophe principle*, that roughly states that the i.i.d. sum of non-negative sub-exponential random variables is large if and only if one of them is large. To put it in a more precise way, here is the formal definition of this property:

Definition 3.9. The distribution function F supported on $[0, \infty)$ is said to satisfy the catastrophe principle, if

$$\mathsf{P}\left(\max_{1 \leq i \leq N} X_i > x \right) \sim \mathsf{P}(X_1 + \cdots + X_N > x), \text{ as } x \to \infty,$$

where $X_1, \ldots, X_N$ are i.i.d. draws from F.

However, we want to know what happens to the maximum factor under the more general conditions of theorem 3.8, namely when the random variables are independently drawn from non-identical distributions, and can take negative as well as positive values. The next

theorem examines the behavior of the maximum term up to a certain constant, and its proof is presented in the appendix.

Theorem 3.10. *Suppose $X_1, \ldots, X_N$ are independently drawn from $F_1, \ldots, F_N$, and take values in $\mathbb{R}$. Then, under the same conditions (i) and (ii) of theorem 3.8, the following asymptotic result holds:*

$$\sum_{i=1}^{N} \mathsf{P}\left(X_i > x\right) + o(\bar{F}(x)) \leq \mathsf{P}\left(\max_{1\leq i\leq N} X_i > x\right) \leq (1-e^{-1})^{-1} \sum_{i=1}^{N} \mathsf{P}\left(X_i > x\right) + o(\bar{F}(x)).$$

Remark 3.11. We argue in the appendix that under a bit more stringent conditions, the exact statement of the catastrophe principle can be obtained, namely $\mathsf{P}\left(\max_{1\leq i\leq N} X_i > x\right) \sim \sum_{i=1}^{N} \mathsf{P}\left(X_i > x\right)$ in this case.

An important take-away from this result is that even under the extended case (ind $\mathbb{R}$-valued random variables), the catastrophe principle asymptotically holds up to a constant. More precisely, the probability that the sum exceeds a large value is of the *same order* of the maximum summand exceeding the same threshold. This can also be interpreted in another sense: aggregate fluctuations do not become extremely large by accumulating small variations; rather, there has to be a single factor with large deviation to support such an extreme event.

3.1 Conditional Monte-Carlo Algorithm

The asymptotic result in theorem 3.8 regarding the tail probability of the sum can be used to take $(\sum_{i=1}^{N} c_i)\bar{F}(x)$ as an estimator for $\mathsf{P}\left(X_1 + \cdots + X_N > x\right)$. However, this estimation performs weakly in many cases, and simulation based on that will be inaccurate. A conditional Monte-Carlo algorithm is developed in Asmussen and Kroese (2006) and Chan and Kroese (2011) to cope with the TP estimation of i.i.d. heavy tails. That idea is incorporated here to obtain an estimator for the sum of ind factors. The algorithm goes as follows:

(i) Sample X_i from its corresponding distribution F_i for $i = 1, \ldots, N$.

(ii) Let $M_N = \max\{X_i : i \in [N]\}$.

(iii) Compute $Z(x) = \sum_{i=1}^{N} \mathsf{P}\left(S_N > x, M_N = X_i | X_{-i}\right)$

The proposed $Z(x)$ is an unbiased estimator of $\mathsf{P}(S_N > x)$. The notation X_{-i} is used to denote all random variables excluding X_i, and S_N represents the sum of generated random variables from independent distributions, i.e., $X_1 + \cdots + X_N$. It is shown in Asmussen and Kroese (2006) that the estimator in step 3 of the algorithm 3.1 has *bounded relative error* for non-negative i.i.d. case, when the common distribution F has RV form.

The following theorem establishes the same result of Asmussen and Kroese (2006), but for the extended case of ind $\mathbb{R}$-valued factors.

Theorem 3.12. *Suppose F satisfies the two conditions of theorem 3.8. Then, the estimator $Z(x)$ in algorithm 3.1 has bounded relative error, namely*

$$\limsup_{x\to\infty} \frac{\mathsf{Var}(Z(x))}{\mathsf{E}[Z(x)]^2} < \infty.$$

Proof. Since F satisfies RV, then there exist a slowly varying function L and $\alpha > 0$ such that $\bar{F}(x) = L(x)/x^\alpha$. Denote $M_{N,-i} = \max\{X_{-i}\}$, $S_{N,-i} = \sum_{j\neq i} X_j$, and let $\widetilde{X_i}$ be an independent copy of X_i. Note that $Z(x)$ is implicitly a statistic generated from $X_1, \ldots, X_N$, thereby a random variable.

$$\begin{aligned} Z(x) &= \sum_{i=1}^{N} \mathsf{P}(S_N > x, M_N = X_i | X_{-i}) = \sum_{i=1}^{N} \mathsf{P}\left(\widetilde{X_i} > (x - S_{N,-i}) \vee M_{N,-i} | X_{-i}\right) \\ &= \sum_{i=1}^{N} \bar{F}_i\left((x - S_{N,-i}) \vee M_{N,-i}\right) \sim \sum_{i=1}^{N} c_i \bar{F}\left((x - S_{N,-i}) \vee M_{N,-i}\right) \end{aligned}$$

One can check that if $M_{N,-i} \leq x/N$ then $x - S_{N,-i} \geq x/N$, and hence $M_{N,-i} \vee (x - S_{N,-i}) \geq x/N$ always. Consequently, $Z(x)$ is asymptotically upper bounded by $\left(\sum_{i=1}^{N} c_i\right) \bar{F}(x/N)$, which yields to

$$\limsup_{x\to\infty} \frac{\mathsf{E}[Z(x)^2]}{\mathsf{E}[Z(x)]^2} \leq \lim_{x\to\infty} \frac{\left(\sum_{i=1}^{N} c_i\right)^2 \bar{F}(x/N)^2}{\left(\sum_{i=1}^{N} c_i\right)^2 \bar{F}(x)^2} = \frac{L^2(x/N)/(x/N)^{2\alpha}}{L(x)^2/x^{2\alpha}} = N^{2\alpha}.$$

□

3.2 Concentration and Efficiency Analysis

The proposed algorithm can be repeated n times with the outcome of the i-th step being referred as $Z_i(x)$, and the sample average is denoted by $\bar{Z}_n(x)$. Let $\mu(x) := \mathsf{P}(X_1 + \cdots + X_N > x)$,

and $\sigma(x)^2 := \mathrm{Var}(Z(x))$. Then, a simple application of central limit theorem yields

$$\frac{\bar{Z}_n(x) - \mu(x)}{\sigma(x)/\sqrt{n}} \xrightarrow{d} Z \stackrel{d}{=} \mathcal{N}(0,1).$$

Therefore, one can get the following asymptotic confidence interval for the tail probability of $\bar{Z}_n(x)$:

$$\begin{aligned}
\mathsf{P}\left(\left|\bar{Z}_n(x) - \mu(x)\right| > \kappa\mu(x)\right) &= 1 - \mathsf{P}\left(\left|\bar{Z}_n(x) - \mu(x)\right| \leq \kappa\mu(x)\right) \\
&= 1 - \mathsf{P}\left(|Z| \leq \frac{\kappa\mu(x)}{\sigma(x)/\sqrt{n}}\right) + o_n(1) \\
&\leq 1 - \mathsf{P}\left(|Z| \leq \frac{\kappa\sqrt{n}}{N^\alpha}\right) + o_x(1) + o_n(1) \\
&= 2\left(1 - \Phi\left(\frac{\kappa\sqrt{n}}{N^\alpha}\right)\right) + o_x(1) + o_n(1),
\end{aligned} \tag{3.3}$$

where $\Phi(\cdot)$ is the standard Gaussian CDF. Thus, for large enough n and x we have

$$\left|\bar{Z}_n(x) - \mu(x)\right| \leq \kappa\mu(x)\,,$$

with probability of at least $(2\Phi(\kappa\sqrt{n}N^{-\alpha}) - 1)$. Another way to find the concentration bound on $\bar{Z}_n(x)$ is to use Markov's inequality:

$$\mathsf{P}\left(\left|\bar{Z}_n(x) - \mu(x)\right| > \kappa\mu(x)\right) \leq \frac{\mathsf{E}\left[\left(\bar{Z}_n(x) - \mu(x)\right)^2\right]}{\kappa^2\mu(x)^2} \leq \frac{N^{2\alpha}}{\kappa^2 n} + o_x(1), \tag{3.4}$$

where the last inequality uses the final bound in theorem 3.12 and holds for large enough x. Finally, we express a stronger approach to get a concentration bound based on the notion of sub-Gaussian random variables.

Definition 3.13 (van Der Vaart and Wellner (1996))**.** A random variable X with mean $\mu = \mathsf{E}X$ is called *sub-Gaussian* with parameter $\sigma > 0$, if

$$\mathsf{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}}, \text{ for all } \lambda \in \mathbb{R}.$$

Remark 3.14. Suppose that the random variable X with mean μ is sub-Gaussian with parameter σ, then the following *Chernoff* deviation bound would immediately fall out:

$$\mathsf{P}\left(|X - \mu| > t\right) \leq 2e^{-t^2/2\sigma^2}.$$

One can show that if X takes value in $[a, b]$, then its sub-Gaussianity parameter is $(b - a)/2$. Since the estimator $Z(x)$ has bounded support, then it is sub-Gaussian and we can apply the preceding Chernoff bound. Inspecting the computations in theorem 3.12, we can confirm that $Z(x) \in [0, \sum_{i=1}^{N} \bar{F}_i(x/N)]$, therefore $Z(x)$ is sub-Gaussian with parameter $\sum_{i=1}^{N} \bar{F}_i(x/N)/2$, and the following deviation bound results from remark 3.14:

$$\mathsf{P}\Big(|Z(x) - \mu(x)| > \kappa\mu(x)\Big) \leq 2\exp\left\{ \frac{-2\kappa^2\mu(x)^2}{\Big(\sum_{i=1}^{N} \bar{F}_i(x/N)\Big)^2} \right\} = e^{-2\kappa^2/N^{2\alpha}} + o_x(1).$$

In the last step we use the tail approximation for both $\mu(x)$ and the sum in the exponent's denominator. This is a one-shot bound, namely just for one trial of our algorithm, whereas if we repeat this process n times, and take the sample average, then we get a much sharper precision:

$$\mathsf{P}\left(\left|\bar{Z}_n(x) - \mu(x)\right| > \kappa\mu(x)\right) \leq 2\exp\left\{ \frac{-2n\kappa^2\mu(x)^2}{\Big(\sum_{i=1}^{N} \bar{F}_i(x/N)\Big)^2} \right\} = e^{-2n\kappa^2/N^{2\alpha}} + o_x(1). \tag{3.5}$$

As can be viewed in all three bounds (3.3), (3.4) and (3.5) the ratio $N^{2\alpha}/n$ turns out to be the key parameter controlling the decay rate of error probability. For instance, if $N = 10$, and $\alpha = 2$, we need to repeat our algorithm 10^4 times to get small error probability. An important observation here is that n scales proportional to $N^{2\alpha}$, thus for fixed error rate smaller values of α lead to faster convergence rate, which makes more sense once we recall that the smaller levels of α correspond to the fatter tails. Therefore, the tail asymptotic equivalence relation will be achieved at smaller x's, equivalently, the error in tail probability estimation would be smaller for fixed x.

The proposed algorithm asymptotically outperforms the crude Monte-Carlo sampling in the sense of estimator's efficiency, namely for certain level κ, the deviation probability of our estimator is smaller than the crude Monte-Carlo estimator denoted by

$$\hat{\mu}_n(x) := \frac{1}{n}\sum_{k=1}^{n} 1_{\left\{X_1^{(k)} + \cdots + X_N^{(k)} > x\right\}},$$

where $X_i^{(k)}$ is the k-th independent draw from F_i. The main theoretical result of the paper is presented next, in that we establish the exponential boost obtained via the proposed estimator relative to the crude Monte-Carlo counterpart.

Theorem 3.15. *For any level $0 < \kappa < 1$, the proposed estimator $\bar{Z}_n(x)$ is exponentially more efficient than $\hat{\mu}_n(x)$. Namely, for any $0 < r < 2\kappa^2 N^{-2\alpha}$,*

$$\frac{\mathsf{P}\big(\,|\bar{Z}_n(x) - \mu(x)| > \kappa\mu(x)\big)}{\mathsf{P}\big(\,|\hat{\mu}_n(x) - \mu(x)| > \kappa\mu(x)\big)} \leq (n+1)\exp\Big\{ -n\kappa^2 \left(\frac{2}{N^{2\alpha}} - \mu(x) + o(\mu(x))\right) + 2\kappa + o_x(1) + \frac{1}{n\mu(x)}\big(1 + o_x(1)\big)\Big\},$$

and therefore,

$$\limsup_{x\to\infty} \lim_{n\to\infty} \left\{ rn + \log\left(\frac{\mathsf{P}\big(\,|\bar{Z}_n(x) - \mu(x)| > \kappa\mu(x)\big)}{\mathsf{P}\big(\,|\hat{\mu}_n(x) - \mu(x)| > \kappa\mu(x)\big)}\right)\right\} \leq 0. \tag{3.6}$$

In particular, the above result justifies that our proposed estimator $\bar{Z}_n(x)$ concentrates around the true value $\mu(x)$ exponentially faster than the speed of crude Monte-Carlo. Equivalently, to achieve a certain level of coverage probability of the true mean ($\mu(x)$), our estimator requires only a fraction of samples used by the crude Monte-Carlo estimator, where higher r in equation (3.6) corresponds to smaller fractions, and thus fewer samples.

4 Tail Probability of Market Portfolio

One of the main motivations of studying RV distributions in this paper was to capture the large deviations of asset returns, as initially laid out for the Gaussian case. Now, consider the scenario in which the factor returns have Power law tails, i.e., $\mathsf{P}(\phi_i > x) \propto x^{-\tau_i}$, that happens to be the case in many empirical stock return observations, see for example Cont (2001) and Gopikrishnan et al. (1998). The demeaned market index return can be modeled as the sum of independent zero mean factors combined with an independent noise, as seen before in (2.1):

$$\xi = \sum_{i=1}^{k} \bar{\beta}_i \phi_i + \bar{\varepsilon}.$$

Since ϕ_i is assumed to have Power law tail, so does $\bar{\beta}_i\phi_i$ with the same tail coefficient. Therefore, letting $\tau = \min\{\tau_i : i \in [k]\}$ and $\gamma \in \{i : \tau_i = \tau\}$, the supporting distribution F in the sense of theorem 3.8 would be a Power law with coefficient τ (more precisely $F \overset{d}{=} \beta_\gamma\phi_\gamma$), and

$$c_i = \begin{cases} \lim\limits_{x\to\infty} \frac{\mathsf{P}\left(\bar{\beta}_i\phi_i > x\right)}{\mathsf{P}\left(\bar{\beta}_\alpha\phi_\alpha > x\right)} & \tau_i = \tau, \\ 0 & \tau_i > \tau. \end{cases}$$

Moreover, one can impose Gaussian structure on the idiosyncratic risk terms, and treat them as independent factors that have vanishing tail probabilities relative to the heaviest tail component, $\beta_\gamma \phi_\gamma$, namely

$$\lim_{x\to\infty} \frac{\mathsf{P}\left(\bar{\varepsilon} > x\right)}{\mathsf{P}\left(\bar{\beta}_\gamma \phi_\gamma > x\right)} = 0.$$

Then, the result of theorem 3.8 implies that, for large λ:

$$\mathsf{P}\left(\xi > \lambda\right) \sim \left(\sum_{i=1}^{k} c_i\right) \mathsf{P}\left(\bar{\beta}_\gamma \phi_\gamma > \lambda\right) \propto x^{-\tau_\gamma}.$$

As a result of this asymptotic tail equivalence, we can contemplate that only the factors with the heaviest tails contribute to the extreme events, and the market large fluctuations are mainly driven by them. Particularly, in terms of hedging against extreme events, the risk managers shall not worry about the factors with fat body distribution but light tails, even if they add a sizable portion of the portfolio variance, rather they should mainly concern about highly skewed ones.

Next, let us investigate the case, where the market portfolio is generated by aggregating a *large* number of individual stocks, uniformly weighted without loss of generality in this context. It is often observed that after factor extraction the remaining idiosyncratic parts reflect heavy-tailed dispersions and treating them as Gaussians is quite unrealistic. Therefore, their variations could possibly affect the aggregate index fluctuations. However, we show this is not true in the sense that each one can individually affect the fluctuations of its corresponding security, but once added together and averaged out, the aggregate noise tail probability would have negligible effect compared to the contribution of factors with heavier tails. More precisely, as described above let $\eta_i = \langle \beta_i, \phi \rangle + \varepsilon_i$ be the return to the i-th security, while ε_i is no longer required to be Gaussian, but can take any RV form with associated parameters L and $\alpha > 0$. The following proposition asserts this claim in a more definitive form.

Proposition 4.1. *Let $\eta_i = \langle \beta_i, \phi \rangle + \varepsilon_i$ be the return to the i-th security, such that idiosyncratic residuals likewise the factor returns are independent and have RV tails. Then, given the existence of a supporting distribution F with parameters L and $\alpha > 1$, satisfying the h-condition 3.3, and uniformly bounded proportionality coefficients $\{c_i\}$ of individual noise*

distributions with respect to F (i.e., $\max_{1\le i\le M} c_i < c$), we get

$$\lim_{M\to\infty} \mathsf{P}\left(\frac{1}{M}\sum_{i=1}^{M}\varepsilon_i > x\right) = 0,$$

for fixed large x.

The important result of this proposition is that under some regularity conditions on the residual security risks, the aggregate effect of these factors to the frequency of market index fluctuations will vanish for portfolios of many assets. Therefore, the tail probability of such portfolios is mainly controlled by the common factors, which appear in all individual asset returns. One can think of this result as a version of the central limit theorem type argument across independent residuals, but in the case of independent and non-identical variables with heavy tails. The market index tail probability can then be approximated as:

$$\begin{aligned}\mathsf{P}\left(\langle\bar{\beta},\phi\rangle + \frac{1}{M}\sum_{i=1}^{M}\varepsilon_i > x\right) &\sim \mathsf{P}\left(\langle\bar{\beta},\phi\rangle > x\right) + \mathsf{P}\left(\frac{1}{M}\sum_{i=1}^{M}\varepsilon_i > x\right)\\ &\overset{(*)}{\sim} \mathsf{P}\left(\langle\beta,\phi\rangle > x\right).\end{aligned}$$

The first asymptotic equivalence simply follows from theorem 3.8 as x gets large, and the second equivalence $(*)$ falls out by sending $M\to\infty$, in addition to the assumption that the average factor loading vector converges as $M\to\infty$, namely,

$$M^{-1}\sum_{i=1}^{M}\beta_i \to \beta.$$

Our proposed conditional Monte-Carlo estimator, that was introduced in the previous section can now be employed to estimate the tail probability of market index return.

5 Simulations

In this section we numerically validate the result presented in theorem 3.15. We implement simulations, which numerically demonstrate that our proposed estimator is exponentially more efficient than the crude Monte-Carlo estimator. More formally, we show that for any precision level $0<\kappa<1$ and finite number N of independent Pareto factors

$$\Lambda_n \equiv \log\left(\frac{\mathsf{P}\big(\,|\bar{Z}_n(x)-\mu(x)| > \kappa\mu(x)\big)}{\mathsf{P}\big(\,|\hat{\mu}_n(x)-\mu(x)| > \kappa\mu(x)\big)}\right) \le -rn, \tag{5.1}$$

for large n as $x \to \infty$, where $0 < r < 2\kappa^2 N^{-2\alpha}$ and $\alpha = \min_{1\le i\le N} \alpha_i$ is the shape parameter corresponding to the factor with the heaviest tail. The procedure for estimating the maximum rate r is explained in appendix B.

To find the ratio of deviation probabilities in equation (5.1), we appeal to the natural Monte-Carlo method, that simply counts the number of times each estimator ($\bar{Z}$ and $\hat{\mu}$) deviates from the true mean μ by more than $\kappa\mu$. Each iteration involves finding a fresh valuation for our estimator $\bar{Z}_n(x)$ (based on algorithm 3.1) and the crude Monte-Carlo counterpart $\hat{\mu}_n(x)$. Also, to see whether the deviation from the true $\mu(x)$ exceeds the $\kappa\mu(x)$ threshold, we need to know the true mean $\mu(x)$. After several experiments with our TP estimator, we observed that as the sample size increases, estimation variance decreases and the mean estimate for $\mu(x)$ stays very close to those of crude Monte-Carlo. Therefore, in order to estimate μ (to be inserted in equation (5.1)), we used our TP estimator but with a very large sample size, $n = 10^7$.

We set the number of factors in each model $N = 10$ and the precision parameter $\kappa = 0.005$. The chosen level for κ should be small enough such that the difference between the estimators becomes more clear, and large enough so that simulations are numerically stable.

5.1 Variable Thresholds for Tail Probability

Here we examine the relationship between Λ_n and the sample size n as we move the threshold x further to the right (see figure 1). In this simulation, we estimate Λ_n in equation (5.1) when $\mu(x) = \mathsf{P}\big(\sum_{j=1}^N X_j > x\big)$, and the random variables $\{X_j\}$ are independent Pareto factors. The threshold x changes from 100 to 1000 with steps of length 100. The number of factors is set to $N = 10$, with the shape parameters $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{10}) \in \mathbb{R}^{10}$ that are uniformly scattered between 1 and 3.

As it appears from figure 1, Λ_n decays in n (almost linearly), thus verifying the exponential efficiency concept in equation (5.1). In addition the estimated slope $r(x)$ – measured by the slope of each curve in figure 1 – is higher for larger values of threshold x. This means as x moves along the right tail, the relative efficiency of our TP estimator over the crude Monte-Carlo increases.

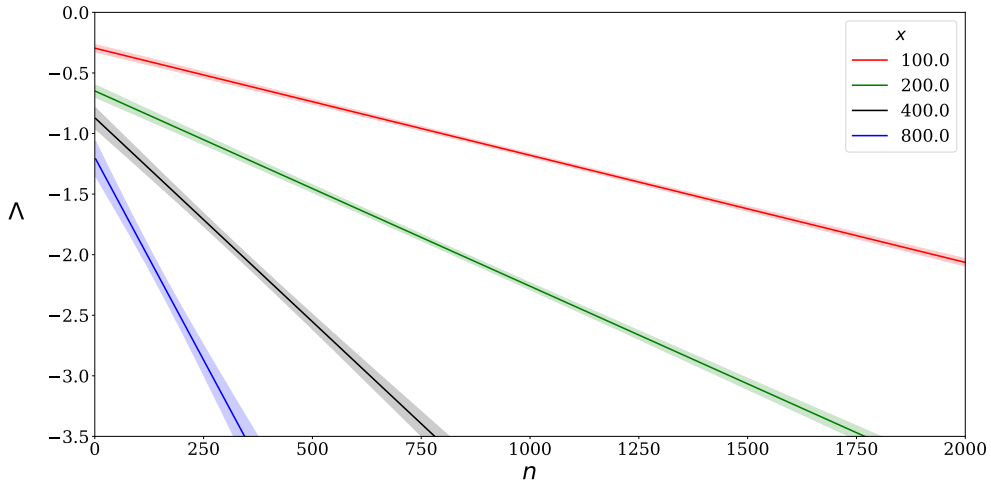

Figure 1: Regression lines fitted to Λ_n vs. sample size n for increasing values of the threshold x along with the 95% confidence intervals.

Table 1 further highlights the result of figure 1. As x moves along the right tail, the estimated $\mu(x)$ naturally falls, and more importantly the estimated efficiency rate $r(x)$ increases.

Table 1: $\{x, \mu(x), r(x)\}$ corresponding to simulation 5.1.

$x \times 100$	1	2	3	4	5	6	7	8	9	10
$\mu(x) \times 10^{-3}$	19.21	8.22	5.14	3.71	2.89	2.36	1.99	1.72	1.51	1.35
$r(x) \times 10^{-3}$	0.87	1.45	1.99	2.30	3.03	3.64	4.25	5.15	5.74	5.69

5.2 Examining the Catastrophe Principle

In this subsection two simulations are performed which aim to validate the Catastrophe Principle. To this end, we simulate M different factor models, where each model contains N Pareto factors with shape parameters $(\alpha_{i1}, \ldots, \alpha_{iN})$, $i = 1, \ldots, M$. We consider two cases:

(i) Groups of factors that share the same $\alpha_{\min}$ (corresponding to the maximum tail thickness), but the average shape parameter $\bar{\alpha}_i$ is different between models (figure 2).

(ii) $\alpha_{\min}$ is different between groups but each group shares the same average shape parameter $\bar{\alpha}$ with a group in the first case (figure 3).

In both cases we repeat the simulation 50 times for each sample size n and compute Λ_n. In figure 2, each curve shows the estimated Λ_n with respect to the sample size n for a certain

level of $\bar{\alpha}$. All of them share the same $\alpha_{\min}$. Evidently, as $\bar{\alpha}$ decreases, meaning as the tails become heavier, the efficiency rate r falls. That is the exponential improvement of our TP estimator relative to the crude Monte-Carlo in the sense of equation (5.1) is maintained across all these curves, but is higher when the average shape parameter $\bar{\alpha}$ is larger.

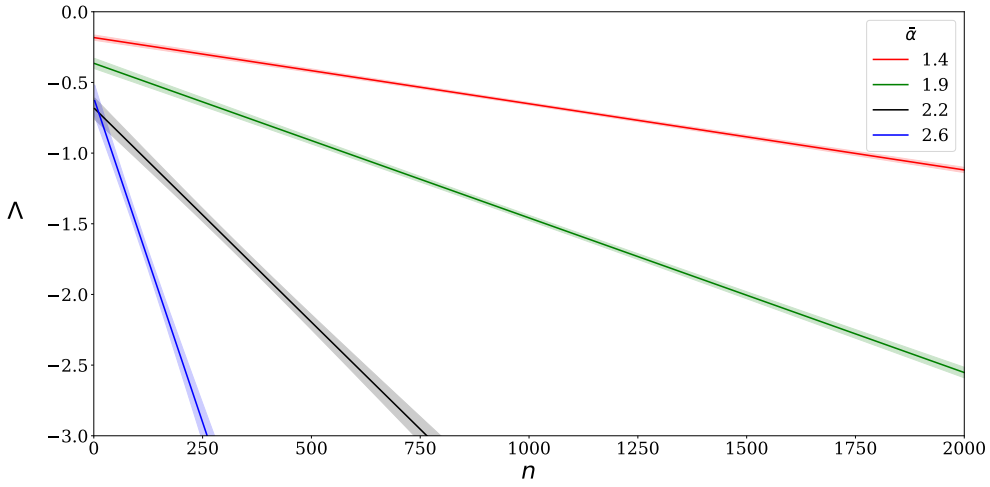

Figure 2: Λ_n vs. sample size n for various $\bar{\alpha}$ and commonly shared $\alpha_{\min}$.

Table 2 further illuminates the pattern of figure 2. According to this table, $\mu(x)$ is not so much sensitive to $\bar{\alpha}$ so long as $\alpha_{\min}$ remains the same, thereby confirming the Catastrophe Principle. However, the estimated slope r increases orders of magnitude which points to the high variability inherent in the crude Monte-Carlo method.

Table 2: $\{\bar{\alpha}, \mu(x), r(x)\}$ while $\alpha_{\min}$ remains constant.

$\bar{\alpha}$	1.45	1.85	2.25	2.65	3.05	3.45	3.85	4.25	4.65	5.05	5.45
$\mu(x) \times 10^{-2}$	3.367	1.389	1.109	1.05	1.042	1.034	1.029	1.025	1.022	1.020	1.018
$r(x) \times 10^{-3}$	0.47	1.1	3.2	9.8	18.9	44.3	53.6	269.1	301.8	417.8	446.2

In figure 3, we investigate the sensitivity of the estimated efficiency rate r (i.e., the slope of Λ_n w.r.t n) to the maximum tail thickness, that corresponds to $\alpha_{\min}$. In contrast with the simulation of figure 2, this time $\alpha_{\min}$ is not constant across different M curves. In fact, the shape parameters are chosen in a way that for each model in the previous simulation, there exists a model in this simulation with equal $\bar{\alpha}$. In essence, mean tail thicknesses are similar in the two simulations. As before, our TP estimator maintains its dominance as $\alpha_{\min}$ varies across the curves; however, upon consulting table 3, we observe that, contrary to the

previous simulation, while $\mu(x)$ decreases by orders of magnitude between factor models as $\alpha_{\min}$ increases, the estimated efficiency rate r does not change drastically.

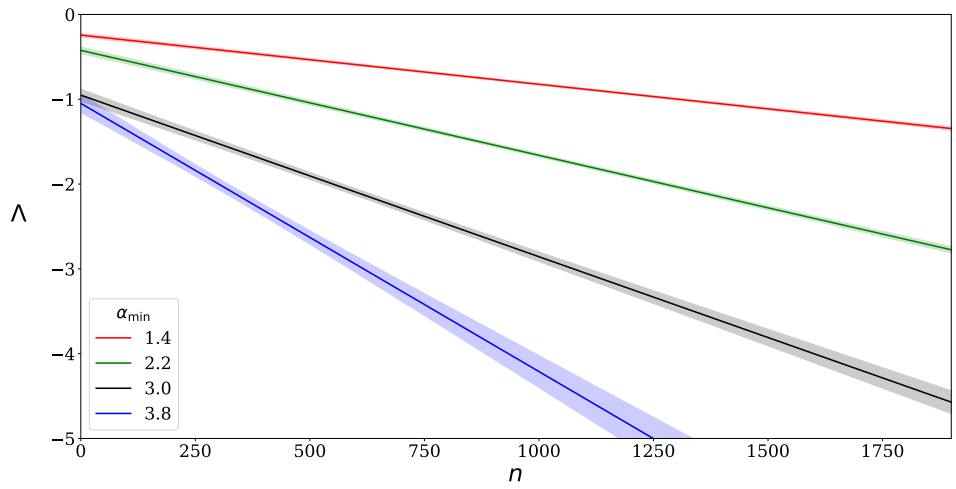

Figure 3: Same simulation as in figure 2, but with varying levels of $\alpha_{\min}$.

Table 3: $\{\alpha_{\min}, \mu(x), r(x)\}$ with $\bar{\alpha}$ kept equal between each column of this table and table 2.

$\alpha_{\min}$	1.0	1.4	1.8	2.2	2.6	3.0	3.4	3.8	4.2	4.6	5.0
$\mu(x) \times 10^{-2}$	3.36	5.14e-01	7.90e-02	1.22e-02	1.92e-03	3.02e-04	4.77e-05	7.54e-06	1.19e-06	1.88e-07	2.98e-8
$r(x) \times 10^{-4}$	4.6	5.9	8.7	12.1	15.1	18.4	25.3	31.2	37.0	60.6	62.6

6 Conclusion

This paper covers a comprehensive asymptotic characterization of tail probability estimation in the case of linear factor models with *independent but not necessarily identical regularly-varying* factors. Exploiting this characterization, a Conditional Monte-Carlo estimator is proposed, which is exponentially more efficient than the crude Monte-Carlo in terms of the concentration around the true tail probability (theorem (3.6)). The simulations are carried out for different levels of primitives, all confirming the dominance of our estimator compared to the crude Monte-Carlo. As a future direction, we hope to generalize the results of this article to study the tail probability of factor models with dependent factors and regularly-varying tails.

A Proofs

A.1 Proof of Theorem 3.8

First, the following lemma is proven, then the theorem's proof follows.

Lemma A.1. *Let F have regularly varying tail, namely $\bar{F}(x) \sim L(x)x^{-\alpha}$ for some $\alpha > 0$. Then, there exists $0 < \delta < 1$, such that for $h(x) = x^{\delta}$,*

$$\bar{F}(h(x))^2 = o(\bar{F}(x)).$$

Proof. Lemma 2 of chapter 8 in Feller (2008) ensures that for every $\varepsilon > 0$, there exists x_0, such that for all $x > x_0$: $x^{-\varepsilon} < L(x) < x^{\varepsilon}$. Now one can check that by taking $\varepsilon < \alpha/5$ and $1 > \delta > 3/4$ the desired result follows:

$$\frac{\bar{F}(h(x))^2}{\bar{F}(x)} \leq x^{\varepsilon(2\delta+1)-\alpha(2\delta-1)} \to 0, \text{ as } x \to \infty. \ \|$$

We justify equation (3.2) for the case of two random variables, X_1 and X_2, then the general case will follow by a straight induction. The argument goes through a similar line of proof as in Foss and Richards (2010), but we leverage the independence to relax some of its necessary conditions. The idea is to upper and lower bound $\mathsf{P}(X_1 + X_2 > x)$ by $\mathsf{P}(X_1 > x) + \mathsf{P}(X_2 > x)$ with some vanishing approximation errors (that are approaching 0 as $x \to \infty$, faster than $\bar{F}(x)$, henceforth denoted by $o(\bar{F}(x))$). First, the upper-bound is verified:

$$\begin{aligned} \mathsf{P}(X_1 + X_2 > x) \leq \mathsf{P}\big(X_1 > x - h(x)\big) + \mathsf{P}\big(X_2 > x - h(x)\big) \\ + \mathsf{P}\big(h(x) < X_1 \leq x - h(x), X_2 > x - X_1\big). \end{aligned} \tag{A.1}$$

The first two terms can be approximated by leveraging assumptions (i) and (ii) of the theorem. For example:

$$\mathsf{P}\big(X_1 > x - h(x)\big) \sim c_1\bar{F}(x - h(x)) \sim c_1\bar{F}(x) \sim \mathsf{P}(X_1 > x),$$

where the first and last approximations hold because of the assumption (i) in the theorem, and the middle one is guaranteed by assumption (ii) and the h-condition 3.3. Furthermore,

the last probability will be of order $\bar{F}(x)$ as $x \to \infty$:

$$\begin{aligned}\mathsf{P}\big(h(x) < X_1 \leq x - h(x), X_2 > x - X_1\big) &\leq \mathsf{P}\big(h(x) < X_1 \leq x - h(x)\big)\mathsf{P}\big(X_2 > h(x)\big) \\ &\leq \mathsf{P}\big(X_1 > h(x)\big)\mathsf{P}\big(X_2 > h(x)\big) \\ &\sim c_1 c_2 \bar{F}(h(x))^2,\end{aligned}$$

where the last term is of order $o(\bar{F}(x))$ because of lemma A.1, thus concluding the upper bound.

Next, the lower bounding follows as:

$$\begin{aligned}\mathsf{P}\big(X_1 + X_2 > x\big) \geq \mathsf{P}\big(X_1 > x + h(x), X_2 > -h(x)\big) + \mathsf{P}\big(X_2 > x + h(x), X_1 > -h(x)\big) \\ \mathsf{P}\big(X_1 > x + h(x), X_2 > x + h(x)\big),\end{aligned} \tag{A.2}$$

where each of the first two terms decouples, and again because of assumptions (i) and (ii) of the theorem, the first one for instance can be approximated as

$$\begin{aligned}\mathsf{P}\big(X_1 > x + h(x), X_2 > -h(x)\big) &= \mathsf{P}\big(X_1 > x + h(x)\big)\mathsf{P}\big(X_2 > -h(x)\big) \\ &\sim c_1 \bar{F}(x + h(x)) \sim c_1 \bar{F}(x) \sim \mathsf{P}\left(X_1 > x\right).\end{aligned}$$

Similar reasoning implies that the third term in (A.2) is of order $o(\bar{F}(x))$, and is therefore vanishing compared to the first two terms in (A.2). The lower bound is now justified, hence the first approximation in equation (3.2) is concluded. Finally, approximation of the sum of tail probabilities with $\bar{F}$ follows immediately as a result of the first assumption of the theorem. □

A.2 Proof of Theorem 3.10

First, the lower bound is shown:

$$\begin{aligned}\mathsf{P}\left(\max_{1\leq i\leq N} X_i > x\right) &= 1 - \mathsf{P}\left(\max_{1\leq i\leq N} X_i \leq x\right) = 1 - \prod_{i=1}^{N}\big(1 - \mathsf{P}(X_i > x)\big) \\ &\geq 1 - \prod_{i=1}^{N} \exp\big(-\mathsf{P}\left(X_i > x\right)\big) = 1 - \exp\left(-\sum_{i=1}^{N}\mathsf{P}(X_i > x)\right) \\ &= \sum_{i=1}^{N}\mathsf{P}(X_i > x) + o(\bar{F}(x)),\end{aligned}$$

where the last equality is an immediate application of the Taylor's lemma. Showing the upper bound mainly falls in the same steps, but requires invoking the inequality $e^{-x} \leq 1-(1-e^{-1})x$, that holds for $x \in [0,1]$:

$$\begin{aligned} 1-\prod_{i=1}^{N}\big(1-\mathsf{P}(X_i>x)\big) &\leq 1-\prod_{i=1}^{N}\exp\big(-(1-e^{-1})^{-1}\mathsf{P}(X_i>x)\big) \\ &= 1-\exp\left(-(1-e^{-1})^{-1}\sum_{i=1}^{N}\mathsf{P}\left[X_i>x\right]\right) \\ &= (1-e^{-1})^{-1}\sum_{i=1}^{N}\mathsf{P}(X_i>x)+o(\bar{F}(x)). \end{aligned} \tag{A.3}$$

Through a geometric argument it becomes clear that $e^{-x} \leq 1-ax$ for $a<1$, and small enough x. Therefore, it is possible to send $a \uparrow 1$ and control for the size of all $\bar{F}_i(x)$, $i=1,\ldots,N$. Under the case where the convergence of $\bar{F}_i(x)/c_i\bar{F}(x)$ (in condition (i) of theorem 3.8) is uniform over all $i=1,\ldots,N$, one can send a to 1 from below slower than the speed of $\bar{F}(x)\to 0$, thereby a tighter upper bound will be obtained in (A.3) with the pre-factor of 1 rather than $(1-e^{-1})^{-1}$.

A.3 Proof of Theorem 3.15

To prove the proposition we need the following lemma, that paves the way for the main verification.

Lemma A.2. *Let $S_n := \sum_{k=1}^{n}\xi_k$, where ξ_k's are i.i.d. Bernoulli random variables with success probability of α, then*

$$\mathsf{P}(S_n \leq n\delta) \geq \frac{1}{n+1}e^{-nD\left(\frac{\lfloor n\delta\rfloor}{n}||\alpha\right)},$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler divergence, that is

$$D(\delta||\alpha) = \delta\log\left(\frac{\delta}{\alpha}\right)+(1-\delta)\log\left(\frac{1-\delta}{1-\alpha}\right).$$

Proof. For the notational simplicity let $m=\lfloor n\delta\rfloor$, and $\tilde{\delta}=m/n$, then

$$\begin{aligned} \mathsf{P}(S_n\leq n\delta) &= \sum_{k=0}^{m}\binom{n}{k}\alpha^k(1-\alpha)^{n-k} \\ &\geq \binom{n}{m}\alpha^m(1-\alpha)^{n-m} = \binom{n}{m}e^{n\left(\tilde{\delta}\log\alpha+(1-\tilde{\delta})\log(1-\alpha)\right)}. \end{aligned} \tag{A.4}$$

Take the auxiliary Binomial random variable $Y \sim \text{Bin}(n, \tilde{\delta})$, then $\mathsf{P}(Y = \ell)$ is maximized when $\ell = m = \lfloor n\delta \rfloor$. The following loose bound results for $\binom{n}{m}$:

$$\begin{aligned} 1 &= \sum_{\ell=0}^{n} \binom{n}{\ell} \tilde{\delta}^{\ell}(1-\tilde{\delta})^{n-\ell} \leq (n+1)\binom{n}{m}\tilde{\delta}^{m}(1-\tilde{\delta})^{n-m} \\ &= (n+1)\binom{n}{m} e^{n\left(\tilde{\delta}\log\tilde{\delta}+(1-\tilde{\delta})\log(1-\tilde{\delta})\right)}, \end{aligned}$$

implying that $\binom{n}{m} \geq (n+1)^{-1} e^{-n\left(\tilde{\delta}\log\tilde{\delta}+(1-\tilde{\delta})\log(1-\tilde{\delta})\right)}$. Then, the proposed bound in the lemma drops out once this lower bound for $\binom{n}{m}$ is substituted in (A.4).‖

Now we can return to the proof of the theorem, first by finding the lower bound for the deviation probability of $\hat{\mu}$:

$$\mathsf{P}\big(|\hat{\mu}_n(x) - \mu(x)| > \kappa\mu(x)\big) = \mathsf{P}\big(\hat{\mu}_n(x) < (1-\kappa)\mu(x)\big) + \mathsf{P}\big(\hat{\mu}_n(x) > (1+\kappa)\mu(x)\big). \quad \text{(A.5)}$$

The first term is lower bounded using the result of lemma A.2 as:

$$\begin{aligned} \mathsf{P}\big(\hat{\mu}_n(x) < (1-\kappa)\mu(x)\big) &\geq \frac{1}{n+1}\exp\left\{-nD\left(\frac{\lfloor n(1-\kappa)\mu(x)\rfloor}{n} \;||\; \mu(x)\right)\right\} \\ &\geq \frac{1}{n+1}\exp\left\{-nD\big((1-\kappa)\mu(x) - 1/n \;||\; \mu(x)\big)\right\}. \end{aligned} \quad \text{(A.6)}$$

In a same manner the second term in (A.5) is lower bounded, with this in mind that $1-\hat{\mu}_n(x)$ is the Binomial sample mean in its own right, but with the different success probability of $1-\mu(x)$:

$$\begin{aligned} \mathsf{P}\big(\hat{\mu}_n(x) > (1+\kappa)\mu(x)\big) &= \mathsf{P}\big(1-\hat{\mu}_n(x) < 1-(1+\kappa)\mu(x)\big) \\ &\geq \frac{1}{n+1}\exp\left\{-nD\left(\frac{\lfloor n(1-(1+\kappa)\mu(x))\rfloor}{n} \;||\; 1-\mu(x)\right)\right\} \\ &\geq \frac{1}{n+1}\exp\left\{-nD\big(1-(1+\kappa)\mu(x) - 1/n \;||\; 1-\mu(x)\big)\right\}. \end{aligned} \quad \text{(A.7)}$$

Denote the KL-divergences in the exponents of (A.6) and (A.7) with D_1 and D_2, respectively. Then, the convexity of $x \mapsto e^{-nx}$ implies:

$$\begin{aligned} \mathsf{P}\big(|\hat{\mu}_n(x) - \mu(x)| > \kappa\mu(x)\big) &\geq \frac{1}{n+1}\left(e^{-nD_1} + e^{-nD_2}\right) \\ &\geq \frac{2}{n+1}e^{-n(D_1+D_2)/2}. \end{aligned} \quad \text{(A.8)}$$

Consequently, it is left to simplify and find an upper bound for $D_1 + D_2$, which is mainly

carried out by leveraging the inequality $x \geq \log(1+x)$ for $x \in (-1,1)$:

$$\begin{aligned} D_1 &= \big((1-\kappa)\mu(x) - 1/n\big) \log\left(\frac{(1-\kappa)\mu(x) - 1/n}{\mu(x)}\right) \\ &\quad + \big(1 - (1-\kappa)\mu(x) + 1/n\big) \log\left(\frac{1-(1-\kappa)\mu(x)+1/n}{1-\mu(x)}\right) \\ &\leq \big((1-\kappa)\mu(x) - 1/n\big)\left(-\kappa - \frac{1}{n\mu(x)}\right) \\ &\quad + \big(1 - (1-\kappa)\mu(x) + 1/n\big)\left(\frac{\kappa\mu(x)+1/n}{1-\mu(x)}\right), \end{aligned} \tag{A.9}$$

and

$$\begin{aligned} D_2 &= \big(1 - (1+\kappa)\mu(x) - 1/n\big) \log\left(\frac{1-(1+\kappa)\mu(x)-1/n}{1-\mu(x)}\right) \\ &\quad + \big((1+\kappa)\mu(x) + 1/n\big) \log\left(\frac{(1+\kappa)\mu(x)+1/n}{\mu(x)}\right) \\ &\leq \big(1 - (1+\kappa)\mu(x) - 1/n\big)\left(-\frac{\kappa\mu(x)+1/n}{1-\mu(x)}\right) \\ &\quad + \big((1+\kappa)\mu(x) + 1/n\big)\left(\kappa + \frac{1}{n\mu(x)}\right). \end{aligned} \tag{A.10}$$

Therefore, the following upper bound on $D_1 + D_2$ falls out by adding (A.9) and (A.10):

$$D_1 + D_2 \leq 2\frac{\big(\kappa\mu(x) + 1/n\big)^2}{\mu(x)\big(1-\mu(x)\big)}.$$

After substitution of this bound in (A.8), it follows

$$\mathsf{P}\big(|\hat{\mu}_n(x) - \mu(x)| > \kappa\mu(x)\big) \geq \frac{2}{n+1} \exp\left\{\frac{-n\big(\kappa\mu(x)+1/n\big)^2}{\mu(x)\big(1-\mu(x)\big)}\right\}.$$

Using the upper bound on the deviation probability of $\bar{Z}_n(x)$ in (3.5), we can see that

$$\begin{aligned}\log\left(\frac{\mathsf{P}\big(\left|\bar{Z}_n(x)-\mu(x)\right|>\kappa\mu(x)\big)}{\mathsf{P}\big(|\hat{\mu}_n(x)-\mu(x)|>\kappa\mu(x)\big)}\right) &\leq (n+1)\exp\Bigg\{\frac{-2n\kappa^2\mu(x)^2}{\Big(\sum_{i=1}^{N}\bar{F}_i(x/N)\Big)^2}+\\ &\qquad\frac{n\kappa^2\mu(x)}{1-\mu(x)}+\frac{1}{n\mu(x)(1-\mu(x))}+\frac{2\kappa}{1-\mu(x)}\Bigg\}\\ &\leq (n+1)\exp\Big\{-n\kappa^2\left(\frac{2}{N^{2\alpha}}-\mu(x)+o(\mu(x))\right)+\\ &\qquad 2\kappa+o_x(1)+\frac{1}{n\mu(x)}\big(1+o_x(1)\big)\Big\},\end{aligned}$$

where in the last equation, we used the large x asymptotics. Now, for any rate r smaller than $2\kappa^2N^{-2\alpha}$, x can be taken large enough, so that the ratio of deviation probabilities decays faster than e^{-rn}. Consequently, the ratio of our estimator's deviation probability over its crude Monte-Carlo counterpart decays exponentially in n, thus establishing the claim of the theorem 3.15. □

A.4 Proof of Proposition 4.1

Theorem 3.8 can be employed again to asymptotically approximate the deviation sum with sum of deviations:

$$\mathsf{P}\left(\sum_{i=1}^{M}\varepsilon_i>Mx\right)\sim\sum_{i=1}^{M}\mathsf{P}\left(\varepsilon_i>Mx\right)\sim\left(\sum_{i=1}^{M}c_i\right)\frac{L(Mx)}{(Mx)^{\alpha}}. \tag{A.11}$$

Proportionality coefficients c_i are defined in the usual fashion: $c_i=\lim_{x\to\infty}\mathsf{P}\left(\varepsilon_i>x\right)/\bar{F}(x)$, and as stated in the theorem are uniformly bounded by a constant say c. Therefore, relation (A.11) can be upper bounded as:

$$\left(\sum_{i=1}^{M}c_i\right)\frac{L(Mx)}{(Mx)^{\alpha}}\leq\frac{cL(Mx)}{M^{\alpha-1}x^{\alpha}}\longrightarrow 0,\quad\text{as } M\to\infty.$$

The last conclusion holds because $L(\cdot)$ is slowly varying by definition and grows at a slower rate than any polynomial growth of M (assuming that $\alpha>1$). □

B Simulations

Estimating r: In order to empirically study equation (3.6), it is sufficient to show that the convex hull of $\{n_i, \Lambda_{n_i}\}$ is bounded above by the affine function $f(n_i) = -rn_i$ where $r > 0$. For each n_i, Λ is evaluated 50 times, for all of which we use a single estimate for $\mu(x)$.

Code Availability: Accompanying code for all the simulations are available as a python package on https://github.com/osolari/tpe.

References

[1] Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2017). "Microeconomic Origins of Macroeconomic Tail Risks," *American Economic Review*, 107(1): 54–108.

[2] Daniel A Ackerberg (2009). "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation," *Quantitative Marketing and Economics*, 7(4): 343–376.

[3] Hansjörg Albrecher, Søren Asmussen, and Dominik Kortschak (2006). "Tail Asymptotics for the Sum of Two Heavy-Tailed Dependent Risks," *Extremes*, 9(2): 107–130.

[4] Søren Asmussen (2008). *Applied Probability and Queues*, volume 51, Springer Science & Business Media.

[5] Søren Asmussen, Klemens Binswanger, and Bjarne Højgaard (2000). "Rare Events Simulation for Heavy-Tailed Distributions," *Bernoulli*, 6(2): 303–322.

[6] Søren Asmussen and Dirk P Kroese (2006). "Improved Algorithms for Rare Event Simulation with Heavy Tails," *Advances in Applied Probability*, 38(2): 545–558.

[7] Joshua C.C. Chan and Dirk P. Kroese (2011). "Rare-Event Probability Estimation with Conditional Monte Carlo," *Annals of Operations Research*, 189(1): 43–61.

[8] Rama Cont (2001). "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues," *Quantitative Finance*, 1(2): 223–236.

[9] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch (2013). *Modelling Extremal Events: for Insurance and Finance*, volume 33, Springer Science & Business Media.

[10] Willliam Feller (2008). *An Introduction to Probability Theory and its Applications*, volume 2, John Wiley & Sons.

[11] Serguei Foss and Andrew Richards (2010). "On Sums of Conditionally Independent Subexponential Random Variables," *Mathematics of Operations Research*, 35(1): 102–119.

[12] Xavier Gabaix (2016). "Power Laws in Economics: An Introduction," *The Journal of Economic Perspectives*, 30(1): 185–205.

[13] Marc J. Goovaerts, Rob Kaas, Roger J.A. Laeven, Qihe Tang, and Raluca Vernic (2005). "The Tail Probability of Discounted Sums of Pareto-Like Losses in Insurance," *Scandinavian Actuarial Journal*, 2005(6): 446–461.

[14] Parameswaran Gopikrishnan, Martin Meyer, LA Nunes Amaral, and H Eugene Stanley (1998). "Inverse Cubic Law for the Distribution of Stock Price Variations," *The European Physical Journal B-Condensed Matter and Complex Systems*, 3(2): 139–140.

[15] Dominik Kortschak and Hansjörg Albrecher (2009). "Asymptotic Results for the Sum of Dependent Non-Identically Distributed Random Variables," *Methodology and Computing in Applied Probability*, 11(3): 279–306.

[16] Robert E. Lucas Jr. and Benjamin Moll (2014). "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122(1): 1–51.

[17] Daniel McFadden (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica: Journal of the Econometric Society*: 995–1026.

[18] Thomas Piketty and Emmanuel Saez (2014). "Inequality in the Long Run," *Science*, 344(6186): 838–843.

[19] Farzad Pourbabaee (2021). "High Dimensional Decision Making, Upper and Lower Bounds," *Economics Letters*, 204: 109894.

[20] Reuven Y. Rubinstein and Dirk P. Kroese (2016). *Simulation and the Monte Carlo Method*, volume 10, John Wiley & Sons.

[21] Aad W. van Der Vaart and Jon A. Wellner (1996). *Weak Convergence and Empirical Processes*, Springer.