

SAMPLE PATH LARGE DEVIATIONS FOR ORDER STATISTICS

KEN R. DUFFY,* *National University of Ireland, Maynooth*

CLAUDIO MACCI,** *Università di Roma ‘Tor Vergata’*

GIOVANNI LUCA TORRISI,*** *Consiglio Nazionale delle Ricerche*

Abstract

We consider the sample paths of the order statistics of independent and identically distributed random variables with common distribution function F . If F is strictly increasing but possibly having discontinuities, we prove that the sample paths of the order statistics satisfy the large deviation principle in the Skorokhod M_1 topology. Sanov’s theorem is deduced in the Skorokhod M_1' topology as a corollary to this result. A number of illustrative examples are presented, including applications to the sample paths of trimmed means and Hill plots.

Keywords: Large deviation; order statistic; empirical law; Skorokhod topology; weak convergence

2010 Mathematics Subject Classification: Primary 60F10; 62G30

1. Introduction

Let $\{X_n : n \geq 1\}$ be a sequence of independent and identically distributed (i.i.d.) real-valued random variables with distribution function $F(x) = P(X_1 \leq x)$ that is assumed to be strictly increasing, but possibly having discontinuities. Define $a := \inf\{x : F(x) > 0\} \in [-\infty, \infty)$ and $b = \inf\{x : F(x) = 1\} \in (-\infty, \infty]$. For each $n \geq 1$, let $X_{1,n}, \dots, X_{n,n}$ denote the ascending order statistics of X_1, \dots, X_n , so that $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$, and define $X_{0,n} := a$ and $X_{n+1,n} := b$. For each $n \geq 1$, define the sample path of the order statistics by

$$X_n(t) := X_{[(n+1)t],n} \quad \text{for all } t \in [0, 1], \quad (1)$$

where $[x]$ is the greatest integer that is less than x .

The purpose of the present article is to prove the functional large deviation principle (LDP) for order statistics in the sense of Varadhan [33]. We consider $X_n(\cdot)$ as a random element of the space of nondecreasing càdlàg functions (i.e. those which are right continuous and have left-hand limits) ϕ such that $\phi(0) \geq a$ and $\phi(1) = b$.

We equip this space with the Skorokhod M_1 topology [30] everywhere, apart from for Sanov’s theorem where we use the Skorokhod M_1' topology [24], [36]. We prove that the random paths

Received 6 July 2009; revision received 25 November 2010.

* Postal address: Hamilton Institute, National University of Ireland Maynooth, Co. Kildare, Ireland.

Email address: ken.duffy@nuim.ie

** Postal address: Dipartimento di Matematica, Università di Roma ‘Tor Vergata’, Via della Ricerca Scientifica, I-00133 Rome, Italy. Email address: macci@mat.uniroma2.it

*** Postal address: Istituto per le Applicazioni del Calcolo ‘Mauro Picone’, Consiglio Nazionale delle Ricerche (CNR), Via dei Taurini 19, 00185, Rome, Italy. Email address: torrisci@iac.rm.cnr.it

$\{X_n(\cdot)\}$ satisfy the LDP. That is, for all Borel sets B ,

$$-\inf_{\phi \in B^\circ} J^F(\phi) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(X_n(\cdot) \in B) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(X_n(\cdot) \in B) \leq -\inf_{\phi \in \bar{B}} J^F(\phi), \tag{2}$$

where B° denotes the interior of B and \bar{B} denotes its closure. The rate function J^F takes values in $[0, \infty]$, is lower semicontinuous, and has compact level sets (i.e. it is a good rate function—see, e.g. [7]).

The sequence of order statistics sample paths $\{X_n(\cdot)\}$ defined in (1) is closely related to the sequence of empirical distribution functions $\{F_n\}$ defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_n \leq x\}}$$

for all $n \geq 1$. Indeed, the right-continuous generalized inverse of $X_n(\cdot)$ is approximately $F_n(\cdot)$ (in a sense that is made precise in the proof of Corollary 2). This relationship suggests that one way to prove that order statistics sample paths satisfy the LDP is to begin with Sanov’s theorem [27], the LDP for empirical measures in the space of probability measures, and to deduce the LDP for $\{X_n(\cdot)\}$ from it. In a recent article, this is the approach taken by Boistard [4] in order to prove large deviation results for L -statistics such as the trimmed mean and Gini’s mean difference. For distribution functions with lighter than exponential tails, she strengthened the topology in Sanov’s theorem from the topology of weak convergence to the topology generated by the L_2 -Wasserstein metric. This enabled her to deduce the LDP for L -statistics of not necessarily bounded random variables by use of the contraction principle. The case of L -statistics for exponentially distributed random variables falls outside the conditions of her general approach, but is treated using alternate arguments.

The work in this article differs from [4] in two significant ways: (i) the method of proof and (ii) the topology in which the result holds. We take a completely different approach to prove the LDP for $\{X_n(\cdot)\}$. We begin by proving that the sample paths of the order statistics for i.i.d. uniformly distributed random variables satisfy the LDP. This is achieved by using an alternate characterization of the distribution of the sample paths of the order statistics of uniformly distributed random variables in terms of self-normalized sums of i.i.d. exponentially distributed random variables. By recalling a version of Mogul’skii’s theorem [19] due to Puhalskii [22] and then applying Puhalskii’s extension of the contraction principle [22], [23] with a function that embodies this representation, we obtain the LDP for the sample paths of the order statistics of the i.i.d. uniformly distributed random variables. An additional application of the contraction principle recovers the result for more general distributions than the uniform distribution. When the underlying distribution function is strictly increasing (although possibly discontinuous), this approach leads to the functional LDP holding in the Skorokhod M_1 topology. From this result we deduce the LDP for trimmed means for any strictly increasing distribution function.

We comment that the topology of uniform convergence would be too strong for these results as even in the limit it is possible to have discontinuous sample paths. This is embodied by the resultant rate functions being finite at discontinuous paths.

An expression for the rate function, J^F in (2), is given in (6). If $F(x) = \int_a^x f(y) dy$, where $f(y) > 0$ almost everywhere (a.e.), this reduces to the formula in (9). This functional LDP enables not only the calculation of the exponential decay in the probability of seeing unlikely sample paths of order statistics, but also the identification of the most likely paths of the order statistics given that a rare event occurred.

We illustrate the merits of this LDP by deducing the sample path LDP for the trimmed means of any strictly increasing distribution function, even those with infinite mean. We also establish the LDP for Hill plots, which enables estimates on the likelihood that Hill’s [13] widely used methodology misclassifies a non-Pareto law as being a Pareto law.

This article is organized as follows. In Section 2 we introduce the basic setup and notation. The functional LDP for order statistics is presented in Section 3. Applications of the results are presented in Section 4.

2. Notation and terminology

We equip the real line \mathbb{R} and its subsets with the Euclidean metric $\rho_1(x, y) = |x - y|$, but we equip the extensions of the real line ($\mathbb{R} \cup \{+\infty\}$, $\mathbb{R} \cup \{-\infty\}$, and $\mathbb{R} \cup \{-\infty, +\infty\}$) with an alternate metric, $\rho_2(x, y) = |\arctan(x) - \arctan(y)|$, to ensure that they are Polish spaces [9]. The metrics ρ_1 and ρ_2 are topologically equivalent when restricted to $[0, 1]$. The use of ρ_1 or ρ_2 is solely a technicality with the usage being dependent on whether we are working with real- or extended real-valued functions.

Let $D[0, 1]$ denote the space of real-valued (or extended real-valued) càdlàg functions on the closed interval $[0, 1]$ equipped with the Skorokhod M_1 topology [30], [36] induced by the metric

$$d_{M_1}(\phi_1, \phi_2) = \inf_{(u_j, r_j) \in \Pi(\phi_j), j=1,2} \max\{\|u_1 - u_2\|_\infty, \|r_1 - r_2\|_\infty\},$$

where $\|u\|_\infty = \sup_{s \in [0, 1]} |u(s)|$ and $\Pi(\phi)$ is the set of all parametric representations (u, r) of ϕ . A parametric representation (u, r) is a continuous nondecreasing function of the interval $[0, 1]$ onto the completed graph Γ_ϕ of ϕ , where the function u gives the spatial component, while the function r gives the time component. In this context a completed graph of ϕ means that

$$\Gamma_\phi = \{(u, t) \in \mathbb{R} \times (0, \infty) : u \in [\min\{\phi(t-), \phi(t)\}, \max\{\phi(t-), \phi(t)\}]\} \cup \{(\phi(0), 0)\},$$

where $\phi(t-)$ denotes the left limit of ϕ at t and we define an order on Γ_ϕ by saying that $(u_1, t_1) \leq (u_2, t_2)$ if either (i) $t_1 < t_2$ or (ii) $t_1 = t_2$ and $|\phi(t_1-) - u_1| \leq |\phi(t_2-) - u_2|$. In Corollary 2 below we shall also consider $D[0, 1]$ equipped with the weaker M'_1 topology [24], [36] which is defined in the same way as M_1 except that we change Γ_ϕ to

$$\Gamma'_\phi = \{(u, t) \in \mathbb{R} \times (0, \infty) : u \in [\min\{\phi(t-), \phi(t)\}, \max\{\phi(t-), \phi(t)\}]\},$$

where $\phi(0-) = 0$.

For each $-\infty \leq a < b \leq \infty$, let $\mathcal{V}_{a,b}^+ \subset D[0, 1]$ denote the closed set of nondecreasing functions ϕ such that $\phi(t) \geq a$ for all $t \in [0, 1]$, and $\phi(1) = b$. We will equip each $\mathcal{V}_{a,b}^+$ with the Skorokhod M_1 topology, apart from for Sanov’s theorem, where we will use the M'_1 variant. For both topologies, the space is metrizable as a separable metric space.

For each function $\phi \in D[0, 1]$, we use the following notation for its Lebesgue decomposition with respect to Lebesgue measure:

$$\phi(t) = \phi^{(a)}(t) + \phi^{(s)}(t) = \int_0^t \dot{\phi}^{(a)}(s) ds + \phi^{(s)}(t).$$

Here $\phi^{(a)}$ is its absolutely continuous component with $\phi^{(a)}(0) := 0$ and $\phi^{(s)}$ is its singular component.

The quantile function $F^{-1} : [0, 1] \mapsto [a, b]$, defined by

$$F^{-1}(u) := \inf\{x : F(x) > u\} \quad \text{if } u \in [0, 1) \quad \text{and} \quad F^{-1}(1) := b, \tag{3}$$

is the right-continuous generalized inverse of F .

3. Functional LDP for order statistics

Theorem 3 below is the cornerstone result. It proves the functional LDP in the Skorokhod M_1 topology for the sample paths of order statistics where the distribution function F is strictly increasing, but possibly discontinuous. In order to establish the result, we shall appeal to the following version of Mogul’skiĭ’s theorem. The version of Mogul’skiĭ’s theorem reported in Theorem 5.1.2 of [7] is insufficient for our needs, as it does not encompass the case of exponentially distributed random variables. See also [17], [20], [21], and [25].

Theorem 1. ([22].) *If $\{Y_i\}$ is i.i.d. with $P(Y_i > 0) = 1$ and $E(\exp(\theta Y_1)) < \infty$ for some $\theta > 0$, then the sample paths*

$$S_n(t) := \frac{1}{n} \sum_{i=1}^{[(n+1)t]} Y_i \tag{4}$$

satisfy the LDP in $D[0, 1]$, equipped with the Skorokhod M_1 topology, with a rate function that is finite only for functions, ϕ , that are nondecreasing and of finite variation. For such a ϕ , the rate function is

$$I(\phi) = \int_0^1 I_t(\dot{\phi}^{(a)}(t)) dt + \phi^{(s)}(1), \tag{5}$$

where $I_t(x)$ is the rate function for the partial sums $\{n^{-1} \sum_{i=1}^n Y_i\}$.

This is a deduction from Lemma 3.2 of [22]. Under the conditions of Theorem 1, it proves that the LDP holds for $\{S_n\}$ in the space of divergent càdlàg functions on the interval $[0, \infty)$ equipped with the topology of weak convergence. We first restrict the argument to $[0, 1]$ by the contraction principle, which gives the rate function in (5). To obtain the final result, we note equivalence between the topology of weak convergence and the Skorokhod M_1 topology for monotone functions (see, for example, Corollary 12.5.1 of [36]).

We shall also make extensive use of Puhalskii’s extension of the contraction principle [22]. In particular, we have the following result.

Theorem 2. ([23, Corollary 3.1.15].) *Assume that $\{X_n\}$ satisfies the LDP in a Hausdorff topological space E with rate function I_E . If $f : E \mapsto E'$, where E' is a Tychonoff space, is continuous at all x such that $I_E(x) < \infty$, then $\{f(X_n)\}$ satisfies the LDP in E' with rate function $I_{E'}(y) = \inf\{I_E(x) : f(x) = y\}$.*

All of the spaces we consider are metric spaces, so the topological conditions of this theorem are met.

Armed with Theorems 1 and 2, we now prove our cornerstone result.

Theorem 3. (LDP for order statistics.) *The sample paths $\{X_n(\cdot)\}$ satisfy the LDP in $\mathcal{V}_{a,b}^+$, equipped with the Skorokhod M_1 topology with rate function*

$$J^F(\chi) = \inf_{\phi \in \mathcal{V}_{0,1}^+} \left\{ - \int_0^1 \log(\dot{\phi}^{(a)}(t)) dt : F^{-1}(\phi(t)) = \chi(t) \text{ for all } t \in [0, 1] \right\}. \tag{6}$$

Note that $J^F(\chi) = 0$ if $\chi(t) = F^{-1}(t)$.

Proof. Begin by considering $\{U_n : n \geq 1\}$, a sequence of i.i.d. random variables that are uniformly distributed on $[0, 1]$. For each $n \geq 1$, let $U_{1,n}, \dots, U_{n,n}$ be the order statistics of U_1, \dots, U_n , with $U_{0,n} := 0$ and $U_{n+1,n} := 1$. For each $n \geq 1$, define the sample path of the order statistics by $U_n(t) := U_{[(n+1)t],n}$ for all $t \in [0, 1]$.

The distribution of the sample path $U_n(\cdot)$ is equal to a distribution that can be constructed from a sequence of i.i.d. exponentially distributed random variables. Let the $\{Y_n\}$ be i.i.d. exponentially distributed random variables with mean 1. Define the self-normalized random functions $\{N_n\}$ by

$$N_n(t) := \left(\sum_{i=1}^{[(n+1)t]} Y_i \right) / \left(\sum_{j=1}^{n+1} Y_j \right) \quad \text{if } \sum_{j=1}^{n+1} Y_j > 0. \tag{7}$$

As a consequence of Proposition 8.2.1 of [29], $N_n(\cdot)$ is equal in distribution to $U_n(\cdot)$. As an application of Theorem 1, the sample paths

$$S_n(t) := \frac{1}{n} \sum_{i=1}^{[(n+1)t]} Y_i$$

satisfy the LDP in $D[0, 1]$ with rate function that is finite only for functions ϕ that are nondecreasing and of finite variation, $I(\phi) = \int_0^1 I_l(\dot{\phi}^{(a)}(t)) dt + \phi^{(s)}(1)$, where $I_l(x) = x - \log(x) - 1$ is the rate function for the partial sums of i.i.d. exponentially distributed random variables $\{n^{-1} \sum_{i=1}^n Y_i\}$.

Define $g: \{\phi \in D[0, 1]: \phi(1) \neq 0\} \mapsto D[0, 1]$ by $g(\phi)(t) = \phi(t)/\phi(1)$. Note that if $S_n(1) \neq 0$ then $g(S_n) = N_n$, where $S_n(\cdot)$ is defined in (4) and $N_n(\cdot)$ is defined in (7). The map g is continuous at all ϕ such that $\phi(1) > 0$, as if $\phi_n \rightarrow \phi$ in $D[0, 1]$ equipped with the Skorokhod M_1 topology then $\phi_n(1) \rightarrow \phi(1)$. As $I_l(0) = -\log(0) - 1 = \infty$, $I(\phi) < \infty$ only if $\phi(1) > 0$. Puhalskii’s extension of the contraction principle, Theorem 2, requires only that g be continuous at all limit points where the rate function is finite in order for the usual contraction principle result to hold (see, e.g. Theorem 4.2.1 of [7]). Thus, as g is continuous at all ϕ such that $I(\phi) < \infty$, we deduce that $\{N_n(\cdot)\}$ satisfies the LDP in $\mathcal{V}_{0,1}^+$ with the following rate function:

$$\begin{aligned} J^U(\psi) &= \inf\{I(\phi) : g(\phi) = \psi\} \\ &= \inf\left\{I(\phi) : \frac{\phi(t)}{\phi(1)} = \psi(t) \text{ for all } t \in [0, 1]\right\} \\ &= \inf_{\phi(1)>0} I(\phi(1)\psi) \\ &= \inf_{z>0} I(z\psi). \end{aligned}$$

For fixed $z > 0$ and $\psi \in \mathcal{V}_{0,1}^+$, we have

$$\begin{aligned} I(z\psi) &= \int_0^1 I_l(z\dot{\psi}^{(a)}(t)) dt + z\psi^{(s)}(1) \\ &= \int_0^1 (z\dot{\psi}^{(a)}(t) - \log(z) - \log(\dot{\psi}^{(a)}(t)) - 1) dt + z\psi^{(s)}(1) \\ &= z(1 - \psi^{(s)}(1)) - \log(z) - \int_0^1 \log(\dot{\psi}^{(a)}(t)) dt - 1 + z\psi^{(s)}(1) \\ &= z - \log(z) - 1 - \int_0^1 \log(\dot{\psi}^{(a)}(t)) dt, \end{aligned}$$

where we have used the fact that $\psi(1) = 1$ to deduce that $\int_0^1 \dot{\psi}^{(a)}(t) dt = 1 - \psi^{(s)}(1)$. However, $\inf_{z>0} (z - \log(z) - 1) = 0$ and is attained at $z = 1$; thus,

$$J^U(\psi) = \inf_{z>0} I(z\psi) = - \int_0^1 \log(\dot{\psi}^{(a)}(t)) dt. \tag{8}$$

As the order statistics sample path $U_n(\cdot)$ has the same distribution as the self-normalized sample path $N_n(\cdot)$, the sample paths of the order statistics of uniformly distributed random variables $\{U_n(\cdot)\}$ satisfy the LDP in $\mathcal{V}_{0,1}^+$ with the rate function given in (8). Since $\log(0) = -\infty$, $J^U(\psi) = \infty$ unless the absolutely continuous component of ψ 's Lebesgue decomposition is strictly increasing a.e. with respect to the Lebesgue measure.

Consider a sequence of i.i.d. random variables $\{X_n\}$ with common distribution function $F(\cdot)$. As is well known (see, e.g. Theorem 14.1 of [3]), with the quantile function $F^{-1}(u)$ defined in (3) and the $\{U_n\}$ being i.i.d. random variables distributed uniformly on $[0, 1]$, then $\{F^{-1}(U_n)\}$ is an i.i.d. sequence of random variables with distribution function $F(\cdot)$. With a slight abuse of notation, define the map $F^{-1}: \mathcal{V}_{0,1}^+ \mapsto \mathcal{V}_{a,b}^+$ by $F^{-1}(\phi)(t) = F^{-1}(\phi(t))$ for all $t \in [0, 1]$. As $F^{-1}(u)$ is a nondecreasing function of u , $F^{-1}(U_n(\cdot))$ is exactly the sample path of the order statistics of $F^{-1}(U_1), \dots, F^{-1}(U_n)$. As we have proved that $\{U_n(\cdot)\}$ satisfies the LDP, to deduce the LDP for the sample paths $\{X_n(\cdot)\}$ of the order statistics of $\{X_n\}$, it suffices to show that the map $F^{-1}: \mathcal{V}_{0,1}^+ \mapsto \mathcal{V}_{a,b}^+$ is sufficiently well behaved that the contraction principle (see, e.g. Theorem 4.2.1 of [7]) can be applied.

As F is assumed to be strictly increasing (although it can have discontinuities), F^{-1} is continuous on $[0, 1]$ (see, e.g. Lemma 13.6.4 of [36]). Note that $F^{-1}(\phi) := F^{-1} \circ \phi$ and Theorem 13.2.3 of [36] proves that composition on $D[0, 1] \times D[0, 1]$ is continuous at all (F^{-1}, ϕ) such that F^{-1} is continuous and ϕ is nondecreasing. Thus, $F^{-1}: \mathcal{V}_{0,1}^+ \mapsto \mathcal{V}_{a,b}^+$ is continuous and Theorem 3 follows from an application of the contraction principle (see, e.g. Theorem 4.2.1 of [7]).

We now state a corollary of Theorem 3 that follows from the chain rule [34].

Corollary 1. (Distribution functions with positive densities a.e.) *If $F(x) = \int_a^x f(y) dy$ and f is continuous and positive a.e., so that $F(x)$ is strictly increasing and continuous, then $J^F(\chi) = \infty$ unless $\chi \in \mathcal{V}_{a,b}^+$ (or, equivalently, $F \circ \chi \in \mathcal{V}_{0,1}^+$) is strictly increasing, in which case*

$$J^F(\chi) = - \int_0^1 (\log(f(\chi(t))) + \log(\dot{\chi}^{(a)}(t))) dt. \tag{9}$$

In the next subsection we present some illustrative examples based on Theorem 3 and Corollary 1, which, *inter alia*, demonstrate that the rate functions defined in (6) and (9) are not convex in general.

3.1. Examples

Example 1 below demonstrates why $J^F(\cdot)$ is finite at paths with discontinuities: they correspond to ranges where no sample has been observed. It also illustrates how the functional LDP enables the deduction of conditional laws of large numbers. We say that the order statistics of random variables with distribution function F can or cannot emulate the order statistics of random variables with distribution function G if $J^F(G^{-1}) < \infty$ or $J^F(G^{-1}) = \infty$, respectively. Examples 2 to 5 below concern order statistics of given laws that can or cannot emulate the order statistics of other distributions. In particular, Example 4 shows that the order statistics of Pareto distributions, even those with finite mean, can emulate those with infinite

mean. Example 5 shows that the order statistics of Pareto distributions can emulate those of any exponential distribution, but the order statistics of exponential distributions cannot emulate the order statistics of Pareto distributions with infinite mean. We return to this final point in Example 6 in Section 4.2 when we consider trimmed means.

Example 1. (*Discontinuous paths.*) If X_1 is uniformly distributed on $[0, 1]$, denoted $F = U$, then $F(x) = \int_0^x dx$. Thus, $F^{-1}(u) = u$, so that

$$J^U(\chi) = - \int_0^1 \log(\dot{\chi}^{(a)}(t)) dt$$

for any $\chi \in \mathcal{V}_{0,1}^+$. As $x \mapsto -\log(x)$ is a strictly convex function, note that J^U is a strictly convex rate function. Define the set $A := \{\phi : \phi(t) \leq \frac{1}{3} \text{ for all } t \in [0, 1]\}$. Note that $X_n(\cdot) \in A$ if and only if $X_{n,n} \in [0, \frac{1}{3}]$, i.e. $X_i \in [0, \frac{1}{3}]$ for all $i \in \{1, \dots, n\}$, and, therefore, $P(X_n(\cdot) \in A) = (\frac{1}{3})^n$. The exponent in the decay of this probability is $-\log(\frac{1}{3})$. This can also be calculated from the LDP by considering the sample path large deviations for $P(X_n(\cdot) \in A)$ and, in particular, by determining $\inf\{J^U(\chi) : \chi \in A\}$. As $-\log(u)$ is a convex function for $u > 0$, we can use Jensen’s inequality to show that the infimum $\inf\{J^U(\chi) : \chi \in A\}$ is attained at $\hat{\chi}(t) = t/3$ for $t \in [0, 1)$ and $\hat{\chi}(1) = 1$. For this path, $J^U(\hat{\chi}) = -\log(\frac{1}{3})$ and, therefore, $\lim n^{-1} \log P(X_n(\cdot) \in A) = -\log(\frac{1}{3})$.

The sample path LDP gives more information than the direct calculation. It shows that the most likely path to this event is that $X_{1,n}, \dots, X_{n,n}$ be spread uniformly over $[0, \frac{1}{3}]$, in the following sense. By, for example, Theorem 3.1(b) of [16], for any $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(X_n(\cdot) \notin B_\varepsilon(\hat{\chi}) \mid X_n(\cdot) \in A) = 0,$$

where $B_\varepsilon(\hat{\chi})$ is the open ball of radius ε around $\hat{\chi} \in \mathcal{V}_{0,1}^+$. That is, conditioned on $X_{n,n} \leq \frac{1}{3}$, the sample paths of the order statistics $\{X_n(\cdot)\}$ satisfy a weak law of large numbers at $\hat{\chi}$, the path where the samples are uniformly distributed in $[0, \frac{1}{3}]$.

Example 2. (*Rate function for the beta distribution.*) Assume that X_1 is distributed as a Beta(α, β) distribution, so that X_1 takes values in $[0, 1]$ with a strictly increasing continuous distribution function $F(x; \alpha, \beta)$ with density

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1},$$

where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ and $\alpha, \beta > 0$. By Corollary 1, $\{X_n(\cdot)\}$ satisfies the LDP in $\mathcal{V}_{0,1}^+$ with rate function

$$J^{\text{Beta}(\alpha,\beta)}(\chi) = -\log\left(\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\right) - (\alpha - 1) \int_0^1 \log(\chi(t)) dt - (\beta - 1) \int_0^1 \log(1 - \chi(t)) dt - \int_0^1 \log(\dot{\chi}^{(a)}(t)) dt$$

for any $\chi \in \mathcal{V}_{0,1}^+$. Considering $J^{\text{Beta}(\alpha,\beta)}(\hat{\chi})$, where $\hat{\chi}(t) := t$ for $t \in [0, 1]$, we are evaluating the large deviations rate of seeing the quantile function of a uniform law given that the underlying distribution is actually a Beta(α, β) distribution. We obtain

$$J^{\text{Beta}(\alpha,\beta)}(\hat{\chi}) = -\log\left(\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\right) + \alpha + \beta - 2.$$

This function has its minimum, $J^{\text{Beta}(\alpha,\beta)}(\hat{\chi}) = 0$, when the random variables $\{X_n\}$ have a uniform distribution, $\alpha = \beta = 1$. Note that if $\alpha = 1$ then, as $\Gamma(1 + \beta) = \beta\Gamma(\beta)$, $J^{\text{Beta}(\alpha,\beta)}(\hat{\chi}) = \beta - \log(\beta) - 1$. This is the rate function evaluated at β for the partial sums $\{n^{-1} \sum_{i=1}^n Y_i\}$ of i.i.d. exponentially distributed random variables $\{Y_i\}$ with mean 1. By symmetry, the same result holds if $\beta = 1$ and α is varied.

Example 3. (Rate function for the exponential distribution.) If $F(x) = 1 - \exp(-\lambda x)$ for all $x \geq 0$ so that $a = 0$ and $b = +\infty$, then $F^{-1}(u) = -\log(1 - u)/\lambda$. Thus, by Corollary 1,

$$J^{\text{Exp}(\lambda)}(\chi) = -\log(\lambda) + \lambda \int_0^1 \chi(t) dt - \int_0^1 \log(\dot{\chi}^{(a)}(t)) dt, \tag{10}$$

which can be readily seen to be strictly convex. For example, if $\hat{\chi}(t) = F^{-1}(t) = -\log(1 - t)/\lambda$ then $J^{\text{Exp}(\lambda)}(\hat{\chi}) = 0$. That is, if the sample path is the quantile function of an exponential distribution with rate λ then the rate function is 0. If, for some $K > 0$, $\hat{\chi}_K(t) = Kt$ for $t \in [0, 1)$ and $\hat{\chi}_K(1) = \infty$, then $J^{\text{Exp}(\lambda)}(\hat{\chi}_K) = -\log(\lambda) + \lambda K/2 - \log(K)$. Thus, the most likely λ to give rise to the quantile function of a uniform law on $[0, K)$ is when $\lambda_K = 2/K$ and the mean of the exponential distribution corresponds to the mean of the corresponding uniform distribution. For $\lambda_K = 2/K$, $J^{\text{Exp}(\lambda_K)}(\hat{\chi}_K) = -\log(2) + 1 \approx 0.307$, irrespective of K .

Example 4. (Rate function for the Pareto distribution.) If $F(x) = 1 - x^{-\alpha}$ for $\alpha > 0$ so that $a = 1$ and $b = +\infty$, then $F^{-1}(u) = (1 - u)^{-1/\alpha}$. Thus, by Corollary 1,

$$J^{\text{Pareto}(\alpha)}(\chi) = -\log(\alpha) + (\alpha + 1) \int_0^1 \log(\chi(t)) dt - \int_0^1 \log(\dot{\chi}^{(a)}(t)) dt,$$

which is an example of a nonconvex rate function. To see this, consider, for $k = 1, 2$ and $\rho \in (0, 1)$, the functions

$$\chi_{k,\varepsilon}(t) = \begin{cases} t + k & \text{for } t \in [0, 1 - \varepsilon), \\ \exp((1 - t)^{-\rho}) & \text{for } t \in [1 - \varepsilon, 1). \end{cases}$$

Then, for any $\gamma \in (0, 1)$,

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} J^{\text{Pareto}(\alpha)}(\gamma \chi_{1,\varepsilon} + (1 - \gamma) \chi_{2,\varepsilon}) \\ &= -\log(\alpha) + (\alpha + 1) \int_0^1 \log(\gamma(t + 1) + (1 - \gamma)(t + 2)) dt \\ &> -\log(\alpha) + (\alpha + 1) \int_0^1 (\gamma \log(t + 1) + (1 - \gamma) \log(t + 2)) dt, \\ &= \lim_{\varepsilon \rightarrow 0} (\gamma J^{\text{Pareto}(\alpha)}(\chi_{1,\varepsilon}) + (1 - \gamma) J^{\text{Pareto}(\alpha)}(\chi_{2,\varepsilon})), \end{aligned}$$

by the strict concavity of $x \mapsto \log(x)$. Thus, for any sufficiently small ε , the lack of convexity of $J^{\text{Pareto}(\alpha)}(\cdot)$ is demonstrated.

If $\hat{\chi}$ corresponds to the quantile function of the Pareto(α), $\hat{\chi}(t) = (1 - t)^{-1/\alpha}$, then $J^{\text{Pareto}(\alpha)}(\hat{\chi}) = 0$. If $K > 0$ and $\hat{\chi}(t) = 1 + Kt$ for $t < 1$ and $\hat{\chi}(1) = \infty$, corresponding to the quantile function of a uniform distribution on $[1, 1 + K)$, then $J^{\text{Pareto}(\alpha)}(\hat{\chi}) = -\log(\alpha K) + (\alpha + 1)((K + 1)K^{-1} \log(K + 1) - 1)$. The minimum over α is attained at $\alpha_K = K/((K + 1) \log(K + 1) - K)$ for which

$$J^{\text{Pareto}(\alpha_K)}(\hat{\chi}) = -2 \log(K) + \log((K + 1) \log(K + 1) - K) + (K + 1)K^{-1} \log(K + 1).$$

This has its infimum as $K \rightarrow 0$, so that $\alpha_K \rightarrow \infty$ and $J^{\text{Pareto}(\alpha_K)}(\hat{\chi})$ tends to $-\log(2) + 1 \approx 0.307$. If $\hat{\chi}(t) = (1 - t)^{-1/\beta}$ for any $\beta > 0$ then

$$J^{\text{Pareto}(\alpha)}(\hat{\chi}) = \frac{\alpha}{\beta} - \log\left(\frac{\alpha}{\beta}\right) - 1.$$

The order statistics path $\hat{\phi}$ of the uniformly distributed random variables on $[0, 1]$ that attains this is $\hat{\phi}(t) = F \circ \hat{\chi}(t) = 1 - (1 - t)^{\alpha/\beta}$. Thus, it is possible on the scale of large deviations for the sample path of the order statistics of any Pareto law to emulate that of any other.

Example 5. (*Exponential and Pareto distributions.*) If $\hat{\chi}(t) = -\log(1 - t)/\lambda + 1$, corresponding to a quantile function of an exponential law on $[1, \infty)$, then

$$J^{\text{Pareto}(\alpha)}(\hat{\chi}) = \log\left(\frac{\lambda}{\alpha}\right) - 1 - \exp(\lambda)(\alpha + 1) \text{Ei}(-\lambda),$$

where $\text{Ei}(-\lambda) = -\int_{\lambda}^{\infty} (\exp(-t)/t) dt$. Thus, in the large deviations limit with a finite rate, the order statistics of any i.i.d. Pareto-distributed random variables can emulate the quantile function of any i.i.d. exponentially distributed random variables. On the other hand, if $\hat{\chi}(t) = (1 - t)^{-1/\alpha} - 1$, corresponding to the quantile function of a Pareto distribution on $[0, \infty)$, then $J^{\text{Exp}(\lambda)}(\hat{\chi}) = +\infty$ if $\alpha \leq 1$, and $J^{\text{Exp}(\lambda)}(\hat{\chi}) = \log(\alpha/\lambda) + (\lambda\alpha + 1 - \alpha^2)/(\alpha(\alpha - 1)) < \infty$ if $\alpha > 1$. That is, in the large deviations limit, the order statistics of exponentially distributed random variables *cannot* emulate the order statistics of Pareto distributed random variables with infinite mean. We return to this point in Section 4.2, Example 6.

3.2. Comment on Sanov’s theorem

Sanov’s theorem (see, e.g. [7, Section 6.2]) considers the empirical laws of a process of i.i.d. random variables. With the laws considered as random elements of the space of probability measures equipped with either the topology of weak convergence or the τ topology, Sanov’s theorem proves that the empirical laws satisfy the LDP with relative entropy as the rate function. For some modern developments, see, for example, [15] and the references therein.

As stated in the introduction, the empirical laws and the sample paths of order statistics are closely related. The following corollary shows that a version of Sanov’s theorem for the empirical distribution functions can be recovered from the sample path LDP for the order statistics. For the sequence $\{X_n\}$, the empirical distribution functions $F_n \in D[0, 1]$ are defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}$$

for all $n \geq 1$.

For $-\infty \leq a < b \leq \infty$ and each $\chi \in \mathcal{V}_{a,b}^+$, define the right inverse χ^{-1} by $\chi^{-1}(t) := \inf\{s: \chi(s) > t\}$ for all $t \in [a, b)$, $\chi^{-1}(1) := b$. Thus, χ^{-1} is an element of $\mathcal{V}_{0,1}^+[a, b]$, the set of nondecreasing elements in the space of càdlàg functions on $[a, b]$ with $\chi^{-1}(a) \geq 0$ and $\chi^{-1}(b) = 1$. For our purposes, it suffices to equip $\mathcal{V}_{0,1}^+[a, b]$ with the Skorokhod M_1' topology, which is finer than the M_1 topology and, thus, than the topology of weak convergence.

Corollary 2. (*Sanov’s theorem.*) Assume that $F(x) = \int_a^x f(y) dy$, where $f(y) > 0$ a.e. and $-\infty \leq a < b \leq \infty$. Then $\{F_n\}$ satisfies the LDP in $\mathcal{V}_{0,1}^+[a, b]$ equipped with the Skorokhod

M'_1 topology with rate function $H^F(\eta) = \infty$ unless η is absolutely continuous, in which case

$$H^F(\eta) = \int_a^b \dot{\eta}(s) \log\left(\frac{\dot{\eta}(s)}{f(s)}\right) ds.$$

Proof. For any $x \in (a, b)$,

$$\begin{aligned} F_n(x) &= \frac{1}{n}(\inf\{m \in \{0, 1, \dots, n + 1\}: X_{m,n} > x\} - 1) \\ &= \frac{1}{n}\left((n + 1) \inf\left\{r \in \left\{0, \frac{1}{n + 1}, \dots, \frac{n + 1}{n + 1}\right\}: X_{(n+1)r,n} > x\right\} - 1\right) \\ &= \left(1 + \frac{1}{n}\right) \inf\{s \in [0, 1]: X_{[(n+1)s],n} > x\} - \frac{1}{n} \\ &= \left(1 + \frac{1}{n}\right) X_n^{-1}(x) - \frac{1}{n}, \end{aligned}$$

where we have used the definition of pseudo-inverse above in our identification of $X_n^{-1}(x)$. Hence, $\sup_{x \in [a,b]} |F_n(x) - X_n^{-1}(x)| \leq 2/n$, and the sequences $\{F_n\}$ and $\{X_n^{-1}(\cdot)\}$ are exponentially equivalent (see, e.g. Definition 4.2.10 of [7]) in the uniform topology. Thus, to prove that $\{F_n\}$ satisfies the LDP, it suffices to show that $\{X_n^{-1}(\cdot)\}$ does. By Theorem 13.6.2 of [36] (with straightforward modifications if $b < \infty$), the function $\mathcal{V}_{a,b}^+ \mapsto \mathcal{V}_{0,1}^+[a, b]$ such that $\phi \mapsto \phi^{-1}$ can be seen to be continuous from the M_1 to M'_1 topologies at all strictly increasing ϕ . It is necessary to move to the M'_1 topology as, in general, inversion is not continuous in the M_1 topology. As $J^F(\phi) = \infty$ unless ϕ is strictly increasing, the LDP for $\{X_n^{-1}(\cdot)\}$ follows from an application of Theorem 2. Take $\eta \in \mathcal{V}_{0,1}^+[a, b]$ so that η^{-1} is strictly increasing (i.e. η is absolutely continuous), and let $\ell = \ell_1 + \ell_2$ be the Lebesgue decomposition of the Lebesgue measure ℓ with respect to the measure $\ell \circ \eta$, i.e. the image measure of ℓ under η^{-1} . By Lemma 3.6 of [22] we have

$$(\eta^{-1})^{(a)}(t) = \frac{d\ell_1}{d\eta}(\eta^{-1}(t)) \quad \ell\text{-a.e.}$$

So, by Corollary 1 we deduce that

$$\begin{aligned} H^F(\eta) &= J^F(\eta^{-1}) \\ &= - \int_0^1 (\log(f(\eta^{-1}(t))) + \log(\eta^{-1})^{(a)}(t)) dt \\ &= - \int_0^1 \log\left(f(\eta^{-1}(t)) \frac{d\ell_1}{d\eta}(\eta^{-1}(t))\right) dt \\ &= - \int_{\eta^{-1}(0)}^b \log\left(f(s) \frac{d\ell_1}{d\eta}(s)\right) \dot{\eta}(s) ds \\ &= - \int_{\eta^{-1}(0)}^b \log\left(f(s) \frac{d\ell}{d\eta}(s)\right) \dot{\eta}(s) ds \tag{11} \\ &= \int_a^b \dot{\eta}(s) \log\left(\frac{\dot{\eta}(s)}{f(s)}\right) ds; \tag{12} \end{aligned}$$

indeed, equality (11) follows by the absolute continuity of η (which implies that $\ell = \ell_1$), and (12) follows from the facts that $\dot{\eta}(s) = 0$ for $s \in [a, \eta^{-1}(0))$ and $0 \log(0) := 0$.

4. Applications

4.1. Sample path large deviations for L -statistics

Let $K : [0, 1] \mapsto \mathbb{R}$, and consider the sequence of random variables called L -statistics:

$$T_n := \frac{1}{n + 1} \sum_{i=1}^n K\left(\frac{i}{n + 1}\right) X_{i,n} \quad \text{for } n \geq 1.$$

Much is known regarding the large- n behaviour of the sequence of random variables $\{T_n\}$ and, in particular, its central limit behaviour; see, e.g. [2], [5], [31], and [32]. Large deviation results for L -statistics can be found in [4], [10], and [11].

Consider the measure on $[0, 1]$ defined by

$$\mu_n(ds) := \frac{1}{n + 1} \sum_{i=1}^n \delta_{i/(n+1)}(ds),$$

where δ_x is the Dirac delta measure at x , and set

$$V_n(t) := \int_0^t K(s) X_{[(n+1)s],n} \mu_n(ds), \quad t \in [0, 1].$$

Clearly, $T_n = V_n(1)$. For large values of n , approximating the empirical measure with the Lebesgue measure on $[0, 1]$, we are led to consider the following sample paths:

$$T_n(t) := \int_0^t K(s) X_{[(n+1)s],n} ds, \quad t \in [0, 1].$$

The sample paths $\{V_n(\cdot)\}$ and $\{T_n(\cdot)\}$ record the shape of the L -statistics for the complete range of quantile values, while T_n records it only over the whole range.

4.2. Trimmed means

The function K can be chosen to remove outliers. When

$$K(u) = \begin{cases} 1/(1 - 2\gamma) & \text{if } u \in [\gamma, 1 - \gamma], \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

for $\gamma \in (0, \frac{1}{2})$, T_n is called the trimmed mean and is a robust statistic. It provides an average value of the observations after the exclusion of large and small observations. It is used, for example, in scoring at the Olympic Games. In the case where $\gamma = \frac{1}{4}$, it is called the interquartile mean.

In this case the sample paths $T_n(\cdot)$ are given by

$$T_n(t) = (1 - 2\gamma)^{-1} \int_{\gamma}^t X_{[(n+1)s],n} ds \quad \text{for } t \in [\gamma, 1 - \gamma].$$

Theorem 4. (Trimmed means.) *Assume that F is strictly increasing. If K is of the form in (13) with $\gamma \in (0, \frac{1}{2})$ then $\{T_n(\cdot)\}$ satisfies the LDP in $C[\gamma, 1 - \gamma]$, the space of continuous functions on $[\gamma, 1 - \gamma]$ with $\phi(\gamma) = 0$, equipped with the topology induced by the uniform norm $\|\phi\| := \sup_{t \in [\gamma, 1 - \gamma]} |\phi(t)|$ and with rate function*

$$H^\gamma(\chi) = \inf_{\phi \in \mathcal{V}_{a,b}^+} \left\{ J^F(\phi) : (1 - 2\gamma)^{-1} \int_{\gamma}^t \phi(s) ds = \chi(t) \text{ for all } t \in [\gamma, 1 - \gamma] \right\},$$

where $J^F(\cdot)$ is defined in (6). Note that $H^\gamma(\chi) = 0$ if $\chi(t) = (1 - 2\gamma)^{-1} \int_\gamma^t F^{-1}(u) du$ for all $t \in [\gamma, 1 - \gamma]$.

Proof. Consider the map $g_\gamma : D[0, 1] \mapsto C[\gamma, 1 - \gamma]$ defined by

$$g_\gamma(\phi)(t) = \frac{1}{1 - 2\gamma} \int_\gamma^t \phi(s) ds \quad \text{for } t \in [\gamma, 1 - \gamma],$$

and note that $g_\gamma(X_{[(n+1)\cdot], n}) = T_n(\cdot)$. As Theorem 3 proves that $\{X_n(\cdot)\}$ satisfies the LDP in $\mathcal{V}_{a,b}^+$, in order to deduce the LDP by invoking the extended contraction principle, Theorem 2, we must check that g_γ is continuous at all ϕ such that $J^F(\phi) < \infty$. When $C[\gamma, 1 - \gamma]$ is equipped with the topology of uniform convergence, g_γ is continuous (see, e.g. Theorem 11.5.1 of [36]) at all ϕ taking values in \mathbb{R} . Thus, concern only arises if $a = -\infty$ or $b = +\infty$. This causes no difficulty as $J^F(\phi) < \infty$ only if $a < \phi(\gamma) < \phi(1 - \gamma) < b$. To see this, consider (for example) $\phi(1 - \gamma) = b$. Then, using (6),

$$\begin{aligned} J^F(\phi) &\geq \inf_{\psi \in \mathcal{V}_{0,1}^+} \left\{ - \int_0^1 \log(\dot{\psi}^{(a)}(t)) dt : F^{-1}(\psi(1 - \gamma)) = \phi(1 - \gamma) \right\} \\ &\geq \inf_{\psi \in \mathcal{V}_{0,1}^+} \left\{ - \int_0^1 \log(\dot{\psi}^{(a)}(t)) dt : \psi(1 - \gamma) = 1 \right\}. \end{aligned}$$

As $\psi(1 - \gamma) = 1$, $\psi(1) = 1$ and ψ is nondecreasing, $\dot{\psi}^{(a)}(t) = 0$ for $t \in (1 - \gamma, 1]$, $-\log(0) = \infty$, and, thus, $J^F(\phi) = \infty$. Hence, the LDP follows from Theorem 3 and an application of Theorem 2.

In contrast to Example 5, the following example shows that, for trimmed means, exponential laws can emulate Pareto laws with finite rate, even Pareto laws with infinite mean.

Example 6. (*Exponential emulating Pareto.*) If $F(x) = 1 - \exp(-\lambda x)$ for all $x \geq 0$ so that $a = 0$ and $b = +\infty$, then $F^{-1}(u) = -\log(1 - u)/\lambda$ and $J^{\text{Exp}(\lambda)}(\chi)$ is given in (10). Consider the test function $\chi(t) = (1 - 2\gamma)^{-1} \int_\gamma^t ((1 - s)^{-1/\alpha} - 1) ds$ for $t \in [\gamma, 1 - \gamma]$, corresponding to the trimmed mean of a Pareto distribution on $[0, \infty)$ with parameter α (see Example 5). Using the expression in Theorem 3,

$$\begin{aligned} H^\gamma(\chi) &= \inf_{\phi \in \mathcal{V}_{0,\infty}^+} \{ J^{\text{Exp}(\lambda)}(\phi) : \phi(t) = (1 - t)^{-1/\alpha} - 1 \text{ for all } t \in [\gamma, 1 - \gamma] \} \\ &= \inf_{\psi \in \mathcal{V}_{0,1}^+} \{ J^U(\psi) : \psi(t) = 1 - e^{-\lambda((1-t)^{-1/\alpha} - 1)} \text{ for all } t \in [\gamma, 1 - \gamma] \} \\ &= \inf_{\psi \in \mathcal{V}_{0,1}^+} \left\{ - \int_0^1 \log(\dot{\psi}^{(a)}(s)) ds : \psi(\gamma) = 1 - e^{-\lambda((1-\gamma)^{-1/\alpha} - 1)}, \right. \\ &\quad \left. \psi(1 - \gamma) = 1 - e^{-\lambda(\gamma^{-1/\alpha} - 1)}, \right. \\ &\quad \left. \dot{\psi}(t) = \frac{\lambda}{\alpha} (1 - t)^{-(1+\alpha)/\alpha} e^{-\lambda((1-t)^{-1/\alpha} - 1)} \text{ for all } t \in [\gamma, 1 - \gamma] \right\}. \end{aligned}$$

Our aim is to show that the right-hand side of this expression is finite for all $\alpha > 0$, demonstrating that the trimmed means of exponentially distributed random variables can emulate those of

Pareto distributed random variables with infinite mean. Consider the following function that is defined by its derivative:

$$\dot{\psi}(t) = \begin{cases} \frac{1 - \exp(-\lambda((1 - \gamma)^{-1/\alpha} - 1))}{\gamma} & \text{if } t \in [0, \gamma), \\ \frac{\lambda}{\alpha}(1 - t)^{-(1+\alpha)/\alpha} \exp(-\lambda((1 - t)^{-1/\alpha} - 1)) & \text{if } t \in [\gamma, 1 - \gamma), \\ \frac{\exp(-\lambda(\gamma^{-1/\alpha} - 1))}{\gamma} & \text{if } t \in [1 - \gamma, 1]. \end{cases}$$

As ψ meets the constraints in the infimum, we can upper bound $H^\gamma(\chi)$ by evaluating $J^U(\psi)$. If $\alpha \neq 1$,

$$\begin{aligned} J^U(\psi) &= - \int_0^1 \log(\dot{\psi}(t)) dt \\ &= 2\gamma \log \gamma - \gamma \log(1 - \exp(-\lambda((1 - \gamma)^{-1/\alpha} - 1))) + \lambda\gamma(\gamma^{-1/\alpha} - 1) \\ &\quad - (1 - 2\gamma) \log\left(\frac{\lambda}{\alpha}\right) + \frac{\alpha + 1}{\alpha}(2\gamma - 1 + (1 - \gamma) \log(1 - \gamma) - \gamma \log \gamma) \\ &\quad - \lambda(1 - 2\gamma) + \frac{\lambda\alpha}{\alpha - 1}((1 - \gamma)^{(\alpha-1)/\alpha} - \gamma^{(\alpha-1)/\alpha}). \end{aligned}$$

If $\alpha = 1$, the last term is replaced with $\lambda(\log(1 - \gamma) - \log(\gamma))$. For any $\lambda > 0$ and any $\gamma \in (0, \frac{1}{2})$, this expression is finite for $\alpha \in (0, 1)$, although growing quickly as $\alpha \rightarrow 0$. Thus, the paths of trimmed means of exponential distributions can mimic those of a Pareto distribution with infinite mean, even though Example 5 shows that this is *not* possible for untrimmed means.

4.3. Hill plots

Consider a sequence of i.i.d. random variables $\{X_n\}$ with common distribution function F supported on $[1, \infty)$. Given a sample set of order statistics $X_{1,n}, \dots, X_{n,n}$, we wish to determine if the original distribution is ultimately a Pareto(α) law on $[1, \infty)$, i.e. if the distribution function is $F(x) = 1 - x^{-\alpha}$ for sufficiently large x . Hill’s [13] widely used methodology to answer this question employs the following empirical quantities: for each $1 \leq k \leq n$, define

$$H_{k,n} := \frac{1}{k + 1} \sum_{i=n-k+1}^n \log(X_{i,n}) - \frac{k}{k + 1} \log(X_{n-k,n}).$$

The approach is based on the following observation: if it was indeed the case that X_i ultimately behaves as a Pareto(α) law then $\log(X_i)$ is ultimately an exponentially distributed random variable with mean $1/\alpha$. Thus, determining that the tail of F is a Pareto distribution function is equivalent to determining that the tail of the distribution function of $\log(X_i)$ is exponential. In Hill’s methodology, we create a Hill plot: for a contiguous range of small k , e.g. $k \in \{[n/100], [n/100] + 1, \dots, [n/10]\}$, we plot $H_{k,n}$ versus k . If the resulting plot is almost a straight line, we deduce that the tail of the distribution function of $\log(X_1)$ is exponential and, thus, F ultimately coincides with a Pareto law with a parameter given by 1 over the height of the line.

Owing to its practical importance, much is known about the properties of Hill’s estimator (see, e.g. [6], [8], [12], [26], [28], and the references therein). Here, as an application of

Theorem 3, by considering the sample paths of Hill’s estimator, we prove the LDP for Hill plots and use it to estimate the likelihood that a non-Pareto distribution is misidentified as being a Pareto distribution.

Consider the sample paths of Hill’s estimator defined by $H_n(0) := 0$ and, for $t \in (0, 1]$,

$$\begin{aligned} H_n(t) &:= \frac{1}{t} \int_{1-t}^1 \log(X_n(s)) \, ds - \log(X_n(1-t)) \\ &= \frac{1}{(n+1)t} \sum_{i=[(n+1)(1-t)]+1}^n \log(X_{i,n}) \\ &\quad - \frac{1}{(n+1)t} (n - [(n+1)(1-t)]) \log(X_{[(n+1)(1-t),n]).} \end{aligned}$$

For $k \in \{1, \dots, n\}$, we have

$$H_n\left(\frac{k+1}{n+1}\right) = \frac{1}{k+1} \sum_{i=n-k+1}^n \log(X_{i,n}) - \frac{k}{k+1} \log(X_{n-k,n}) = H_{k,n},$$

so that $H_n(\cdot)$ is, indeed, the sample path of Hill’s estimator with sample size n . That is, $H_n(\cdot)$ is the Hill plot with a sample of size n .

The following theorem proves that Hill plots satisfy the LDP for i.i.d. random variables with a continuous increasing distribution function F that have bounded support or satisfy tail conditions. After the theorem, we will show that these conditions are verified, for example, for any Weibull law, including those with heavier than exponential tails.

Theorem 5. (LDP for Hill plots.) *Assume that the same hypotheses as in Theorem 3 with $a \geq 1$ hold. In addition, suppose that $b < \infty$, or, alternatively, $F(x)$ is differentiable for all sufficiently large x , there exists $\beta \in (0, 1)$ such that*

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log(1 - F(\exp(\varepsilon^{-\beta}))) = -\infty \tag{14}$$

and, defining the function $x_F(t) := F(\exp((1-t)^{-\beta} + 1))$, for all t sufficiently close to 1, $x_F(t) > t$ and

$$\frac{d}{dt} x_F(t) \geq \frac{x_F(t)(1 - x_F(t))}{x_F(t) - t} \log\left(\frac{(1-t)x_F(t)}{t(1 - x_F(t))}\right). \tag{15}$$

Then $\{H_n(\cdot)\}$ satisfies the LDP in $D[0, 1]$, equipped with the Skorokhod M_1 topology with rate function

$$L^F(\chi) = \inf_{\phi \in \mathcal{V}_{a,b}^+} \left\{ J^F(\phi) : \frac{1}{t} \int_{1-t}^1 \log(\phi(s)) \, ds - \log(\phi(1-t)) = \chi(t) \text{ for all } t \in (0, 1] \right\}.$$

Note that $L^F(\chi) = 0$ if $\chi(t) = t^{-1} \int_{1-t}^1 \log(F^{-1}(s)) \, ds - \log(F^{-1}(1-t))$.

Proof. Define the function $h: \mathcal{V}_{a,b}^+ \mapsto D[0, 1]$ by $h(\chi)(0) := 0$ and

$$h(\chi)(t) = \frac{1}{t} \int_{1-t}^1 \log(\chi(s)) \, ds - \log(\chi(1-t)).$$

To prove the LDP for $\{h(X_n)(\cdot)\}$, we apply extensions of the contraction principle with h after noting the following. The function h can be written as $h = h_4 \circ h_3 \circ h_2 \circ h_1$, where

$$h_1(\chi)(t) = (\chi(t), \chi(1 - t)), \quad h_2(\chi, \psi)(t) = (\log \chi(t), \log \psi(t)),$$

$$h_3(\chi, \psi)(t) = \left(\int_{1-t}^1 \chi(s) ds, \psi(t) \right), \quad \text{and} \quad h_4(\chi, \psi)(t) = \frac{1}{t} \chi(t) - \psi(t).$$

The function h_1 is continuous by arguments analogous to those in Theorem 8.1 of [35], while h_2 , using the continuity of $\log(\cdot)$, is continuous by Theorem 13.2.3 of [36]. For continuous χ , the function h_4 is continuous by Corollary 12.7.1 of [36]. If $b < \infty$ then we can appeal to Theorem 11.5.1 of [36] to deduce the continuity of h_3 and the result follows from an application of the contraction principle. However, if $b = +\infty$, the function h_3 is not continuous. In this case, if the second set of additional conditions in the statement of the theorem holds, then we will show that an approximate version of the contraction principle can be employed.

Consider the first component of the function $h_3 \circ h_2 \circ h_1$. That is, g defined by

$$g(\chi)(t) = \int_{1-t}^1 \log \chi(s) ds.$$

If $\chi(1) = +\infty$ then we cannot appeal to Theorem 11.5.1 of [36] to deduce the continuity of g , as this theorem holds only if χ is real valued. Instead, we consider the family of functions $\{g_\varepsilon : \varepsilon > 0\}$ defined by

$$g_\varepsilon(\chi)(t) = \int_{1-t}^{1-\varepsilon} \log \chi(s) ds,$$

which approximate the behaviour of $g(\chi)$. By similar logic to that in Theorem 4, for any $\varepsilon > 0$, the function g_ε is continuous at all χ such that $J^F(\chi) < \infty$, so that Theorem 2 can be applied, obtaining the LDP for $\{g_\varepsilon(X_n)\}$.

We will thus show that $\{g(X_n)\}$ satisfies the LDP by applying the approximate contraction principle, Theorem 4.2.23 of [7]. This approach requires that the sequences $\{g_\varepsilon(X_n)\}$ are exponentially good approximations of $\{g(X_n)\}$, i.e.

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P \left(\int_{1-\varepsilon}^1 \log(X_n(s)) ds \geq \delta \right) = -\infty, \tag{16}$$

as well as the verification of Equation (4.2.24) of [7] for which it suffices to prove that

$$\limsup_{\varepsilon \rightarrow 0} \sup_{\{\chi \in \mathcal{V}_{a,b}^+ : J^F(\chi) \leq \alpha\}} \int_{1-\varepsilon}^1 \log(\chi(s)) ds = 0 \quad \text{for every } \alpha \in (0, \infty). \tag{17}$$

Given $\delta > 0$, recalling that $\beta \in (0, 1)$, choose $\varepsilon_\delta > 0$ such that $\varepsilon^{1-\beta}/(1-\beta) + \varepsilon < \delta$ for all $0 < \varepsilon \leq \varepsilon_\delta$. Then, with $\chi \in \mathcal{V}_{a,b}^+$,

$$\begin{aligned} \left\{ \chi : \int_{1-\varepsilon}^1 \log(\chi(s)) ds \geq \delta \right\} &\subset \left\{ \chi : \int_{1-\varepsilon}^1 \log(\chi(s)) ds > \frac{\varepsilon^{1-\beta}}{1-\beta} + \varepsilon \right\} \\ &= \left\{ \chi : \int_{1-\varepsilon}^1 \log(\chi(s)) ds > \int_{1-\varepsilon}^1 (\log(\exp((1-s)^{-\beta})) + 1) ds \right\} \\ &\subset \left\{ \chi : \int_{1-\varepsilon}^1 \log \left(\frac{\chi(s)}{\exp((1-s)^{-\beta}) + 1} \right) ds > 0 \right\} \end{aligned}$$

$$\begin{aligned} & \subset \left\{ \chi : \sup_{t \in [1-\varepsilon, 1]} (\chi(t) - \exp((1-t)^{-\beta})) \geq 1 \right\} \\ & =: A_\varepsilon. \end{aligned}$$

We can apply the large deviations upper bound on the closure of $A_\varepsilon, \bar{A}_\varepsilon$, to obtain, for any $\rho > \varepsilon$,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left(\int_{1-\varepsilon}^1 \log(X_n(s)) \, ds \geq \delta \right) \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(X_n \in A_\varepsilon) \\ & \leq - \inf_{\chi \in \mathcal{V}_{a,b}^+} \{J^F(\chi) : \chi \in \bar{A}_\varepsilon\} \\ & \leq - \inf_{t \in [1-\rho, 1]} \inf_{\chi \in \mathcal{V}_{a,b}^+} \{J^F(\chi) : \chi(t) \geq \exp((1-t)^{-\beta}) + 1\} \\ & = - \inf_{t \in [1-\rho, 1]} \inf_{\phi \in \mathcal{V}_{0,1}^+} \left\{ - \int_0^1 \log(\dot{\phi}^{(a)}(s)) \, ds : \phi(t) \geq x_F(t) \right\}. \end{aligned}$$

Using Jensen’s inequality, for $x \geq t$, we have

$$\inf_{\phi \in \mathcal{V}_{0,1}^+} \left\{ - \int_0^1 \log(\dot{\phi}^{(a)}(s)) \, ds : \phi(t) \geq x \right\} = -t \log\left(\frac{x}{t}\right) - (1-t) \log\left(\frac{1-x}{1-t}\right),$$

and this infimum is attained at $\phi(s) := sx/t$ if $s < t$ and $\phi(s) := x + (s-t)(1-x)/(1-t)$ if $s \in [t, 1]$. Let $\rho \in (0, 1)$ be sufficiently small so that, for all $t \in [1-\rho, 1]$, $x_F(t) > t$ and inequality (15) holds. Then the function

$$t \rightarrow -t \log\left(\frac{x_F(t)}{t}\right) - (1-t) \log\left(\frac{1-x_F(t)}{1-t}\right), \quad t \in [1-\rho, 1],$$

is increasing. Thus, we have

$$\begin{aligned} \inf\{J^F(\chi) : \chi \in \bar{A}_\varepsilon\} & \geq \inf_{t \in [1-\rho, 1]} \inf_{\phi \in \mathcal{V}_{0,1}^+} \left\{ - \int_0^1 \log(\dot{\phi}^{(a)}(s)) \, ds : \phi(t) \geq x_F(t) \right\} \\ & = \inf_{t \in [1-\rho, 1]} \left(-t \log\left(\frac{x_F(t)}{t}\right) - (1-t) \log\left(\frac{1-x_F(t)}{1-t}\right) \right) \\ & \geq -(1-\rho) \log\left(\frac{F(\exp(\rho^{-\beta}) + 1)}{1-\rho}\right) - \rho \log\left(\frac{1-F(\exp(\rho^{-\beta}) + 1)}{\rho}\right), \end{aligned}$$

and this latter term tends to $+\infty$ as $\rho \rightarrow 0$ by assumption (14). So (16) is satisfied and the sequences $\{g_\varepsilon(X_n)\}$ are exponentially good approximations of $\{g(X_n)\}$.

To establish (17), reasoning by contradiction, assume that there exist a $\delta > 0$ and a sequence $\{\varepsilon_n\}$ such that $\varepsilon_n \downarrow 0$ and

$$\sup_{\{\chi \in \mathcal{V}_{a,b}^+ : J^F(\chi) \leq \alpha\}} \int_{1-\varepsilon_n}^1 \log(\chi(s)) \, ds \geq \delta.$$

The function $\chi \rightarrow \int_{1-\varepsilon}^1 \log(\chi(s)) ds$ is continuous for any $\varepsilon > 0$. Therefore, by the goodness of J^F , there exists $\chi_{\varepsilon_n, \alpha}$ which attains the supremum and $J^F(\chi_{\varepsilon_n, \alpha}) \leq \alpha$. Thus, we have a contradiction because

$$\alpha \geq J^F(\chi_{\varepsilon_n, \alpha}) \geq -(1 - \varepsilon_n) \log\left(\frac{F(\exp(\varepsilon_n^{-\beta}) + 1)}{1 - \varepsilon_n}\right) - \varepsilon_n \log\left(\frac{1 - F(\exp(\varepsilon_n^{-\beta}) + 1)}{\varepsilon_n}\right),$$

and, using the hypothesis in (14), this final term tends to $+\infty$ as $n \rightarrow \infty$.

Example 7. (Every law that is ultimately Weibull satisfies the conditions of Theorem 5.) Consider a law that is ultimately Weibull, $F(x) = 1 - e^{-x^\alpha}$ for some $\alpha > 0$ and all sufficiently large x . For any $\beta \in (0, 1)$,

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log(1 - F(\exp(\varepsilon^{-\beta}))) = - \lim_{\varepsilon \rightarrow 0} \varepsilon \exp(\alpha \varepsilon^{-\beta}) = -\infty,$$

and, thus, (14) is satisfied. Define

$$x_F(t) = F(l(t)),$$

where $l(t) = \exp((1 - t)^{-\beta}) + 1$ for some $\beta \in (0, 1)$. It is easy to check that $x_F(t) > t$ for all t sufficiently close to 1. Equation (15) is equivalent to

$$\alpha\beta(1 - t)^{-(\beta+1)}(l(t) - 1)(l(t))^{\alpha-1} \geq \frac{1 - e^{-l(t)^\alpha}}{1 - e^{-l(t)^\alpha} - t} \log\left(\frac{1 - t}{t} \frac{1 - e^{-l(t)^\alpha}}{e^{-l(t)^\alpha}}\right) \tag{18}$$

for all t sufficiently close to 1. Equation (18) holds if we can show that

$$\lim_{t \rightarrow 1} \left\{ [\alpha\beta(1 - t)^{-(\beta+1)}(l(t) - 1)(l(t))^{\alpha-1}(1 - e^{-l(t)^\alpha} - t)] \times \left[(1 - e^{-l(t)^\alpha}) \log\left(\frac{1 - t}{t} \frac{1 - e^{-l(t)^\alpha}}{e^{-l(t)^\alpha}}\right) \right]^{-1} \right\} = +\infty.$$

For this, note that, for all t close to 1, we have

$$(1 - e^{-l(t)^\alpha}) \log\left(\frac{1 - t}{t} \frac{1 - e^{-l(t)^\alpha}}{e^{-l(t)^\alpha}}\right) \leq (1 - e^{-l(t)^\alpha}) \log(e^{l(t)^\alpha} - 1) \leq (1 - e^{-l(t)^\alpha})(l(t))^\alpha,$$

and so

$$\begin{aligned} & [\alpha\beta(1 - t)^{-(\beta+1)}(l(t) - 1)(l(t))^{\alpha-1}(1 - e^{-l(t)^\alpha} - t)] \\ & \times \left[(1 - e^{-l(t)^\alpha}) \log\left(\frac{1 - t}{t} \frac{1 - e^{-l(t)^\alpha}}{e^{-l(t)^\alpha}}\right) \right]^{-1} \\ & \geq \frac{\alpha\beta(1 - t)^{-(\beta+1)}(l(t) - 1)(1 - e^{-l(t)^\alpha} - t)}{(1 - e^{-l(t)^\alpha})l(t)}. \end{aligned} \tag{19}$$

The claim follows noting that the term in (19) goes to $+\infty$ as $t \rightarrow 1$ because it is asymptotically equivalent to $\alpha\beta(1 - t)^{-\beta}$.

Example 8. (*Truncated Pareto emulating Pareto.*) As an application of Theorem 5, we determine estimates on the likelihood that the Hill plot misclassifies the distribution function F as having Pareto tails when it does not. For certain financial objects, it has been suggested that, while on short time scales, fluctuations in value are large, in the longer term they are not; see, e.g. [18]. Similar observations have been made in ground-water hydrology, atmospheric science, and many other fields; for examples, see [1] and the references therein. This has led to the proposal of, e.g. financial market models based on random walks whose increments have apparent power-tail behaviour near the centre of their support, but whose tails decay at least as fast as an exponential distribution. Truncated Lévy distributions have been used with either a sudden truncation [18] or a transition to an exponential distribution beyond a given cutoff [14]. Similarly, truncated Pareto distributions have also been proposed. Consider a truncated Pareto distribution with parameter $\gamma > 0$ supported on $[1, K]$ that changes into an exponential distribution on $[K, \infty)$ with rate λ :

$$F(x) = \begin{cases} 1 - x^{-\gamma} & \text{if } x \in [1, K), \\ 1 - K^{-\gamma} e^{-\lambda(x-K)} & \text{if } x \in [K, \infty), \end{cases}$$

and, therefore, its quantile function is

$$F^{-1}(u) = \begin{cases} (1 - u)^{-1/\gamma} & \text{if } u \in [0, 1 - K^{-\gamma}), \\ K - \lambda^{-1} \log(K^\gamma(1 - u)) & \text{if } u \in [1 - K^{-\gamma}, 1]. \end{cases}$$

By the preceding example, the conditions of Theorem 5 are met.

Consider $L^F(\hat{\chi})$, where $\hat{\chi}$ corresponds to the Hill plot of the Pareto(α) distribution. Then, for the quantile function $\hat{\phi}(s) = (1 - s)^{-1/\alpha}$, we have

$$\hat{\chi}(t) = \frac{1}{t} \int_{1-t}^1 \log(\hat{\phi}(s)) ds - \log(\hat{\phi}(1 - t)) = \frac{1}{\alpha}.$$

Referring to Corollary 1, the function f is defined by $f(x) := \hat{F}(x) = \gamma x^{-\gamma-1}$ if $x \in [1, K]$ and $f(x) := \hat{F}(x) = \lambda K^{-\gamma} \exp(-\lambda(x - K))$ if $x \in (K, \infty)$. Thus, by the expression of L^F in Theorem 5 and (9), we have

$$\begin{aligned} L^F(\hat{\chi}) &\leq J^F(\hat{\phi}) \\ &= - \int_0^1 (\log(f(\hat{\phi}(t))) + \log(\hat{\phi}(t))) dt \\ &= - \int_0^{1-K^{-\alpha}} \log(\gamma((1 - t)^{-1/\alpha})^{-\gamma-1}) dt \\ &\quad - \int_{1-K^{-\alpha}}^1 \log(\lambda K^{-\gamma} \exp(-\lambda((1 - t)^{-1/\alpha} - K))) dt \\ &\quad - \int_0^1 \log\left(\frac{(1 - t)^{-1/\alpha-1}}{\alpha}\right) dt. \end{aligned}$$

The second term in this equation, corresponding to the exponential part of the distribution emulating the quantile function of a Pareto(α) law, leads the integral to be infinite if $\alpha \in (0, 1]$ and finite if $\alpha > 1$. That is, if the real distribution is a Pareto(γ) distribution truncated by an Exponential(λ), then, with finite rate, one can observe a Pareto(α) Hill plot so long as $\alpha > 1$.

Acknowledgements

We thank the anonymous reviewers for their careful reading of the paper and, in particular, one of them for pointing out a topological error in the original submission.

References

- [1] ABAN, I. B., MEERSCHAERT, M. M. AND PANORSKA, A. K. (2006). Parameter estimation for the truncated Pareto distribution. *J. Amer. Statist. Assoc.* **101**, 270–277.
- [2] ALESHKYAVICHENE, A. K. (1991). Large and moderate deviations for L -statistics. *Litovsk. Mat. Sb.* **31**, 227–241.
- [3] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd edn. John Wiley, New York.
- [4] BOISTARD, H. (2007). Large deviations for L -statistics. *Statist. Decisions* **25**, 89–125.
- [5] CALLAERT, H., VANDEMAELE, M. AND VERAVERBEKE, N. (1982). A Cramér type large deviation theorem for trimmed linear combinations of order statistics. *Commun. Statist. Theory Meth.* **11**, 2689–2698.
- [6] DE HAAN, L. AND RESNICK, S. (1998). On asymptotic normality of the Hill estimator. *Commun. Statist. Stoch. Models* **14**, 849–866.
- [7] DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviation Techniques and Applications*. Springer, Berlin.
- [8] DREES, H., DE HAAN, L. AND RESNICK, S. (2000). How to make a Hill plot. *Ann. Statist.* **28**, 254–274.
- [9] FRISTEDT, B. AND GRAY, L. (1997). *A Modern Approach to Probability Theory*. Birkhäuser, Boston, MA.
- [10] GROENEBOOM, P. AND SHORACK, G. R. (1981). Large deviations of goodness of fit statistics and linear combinations of order statistics. *Ann. Prob.* **9**, 971–987.
- [11] GROENEBOOM, P., OOSTERHOFF, J. AND RUYMGAART, F. H. (1979). Large deviation theorems for empirical probability measures. *Ann. Prob.* **7**, 553–586.
- [12] HAEUSLER, E. AND SEGERS, J. (2007). Assessing confidence intervals for the tail index by Edgeworth expansions for the Hill estimator. *Bernoulli* **13**, 175–194.
- [13] HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.
- [14] KOPONEN, I. (1995). Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Phys. Rev. E* **52**, 1197–1199.
- [15] LÉONARD, C. AND NAJIM, J. (2002). An extension of Sanov's theorem: application to the Gibbs conditioning principle. *Bernoulli* **8**, 721–743.
- [16] LEWIS, J. T., PFISTER, C.-E. AND SULLIVAN, W. G. (1995). Entropy, concentration of probability and conditional limit theorems. *Markov Process. Relat. Fields* **1**, 319–386.
- [17] LYNCH, J. AND SETHURAMAN, J. (1987). Large deviations for processes with independent increments. *Ann. Prob.* **15**, 610–627.
- [18] MANTEGNA, R. N. AND STANLEY, H. E. (1994). Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight. *Phys. Rev. Lett.* **73**, 2946–2949.
- [19] MOGUL'SKIĬ, A. A. (1976). Large deviations for trajectories of multi-dimensional random walks. *Theoret. Prob. Appl.* **21**, 300–315.
- [20] MOGUL'SKIĬ, A. A. (1993). Large deviations for processes with independent increments. *Ann. Prob.* **21**, 202–215.
- [21] NAJIM, J. (2002). A Cramér type theorem for weighted random variables. *Electron. J. Prob.* **7**, 32 pp.
- [22] PUHALSKII, A. (1995). Large deviation analysis of the single server queue. *Queueing Systems* **21**, 5–66. (Correction: **23** (1996), 337.)
- [23] PUHALSKII, A. (2001). *Large Deviations and Idempotent Probability* (Monogr. Surveys Pure Appl. Math. **119**). Chapman and Hall/CRC, Boca Raton, FL.
- [24] PUHALSKII, A. AND WHITT, W. (1997). Functional large deviation principles for first-passage-time processes. *Ann. Appl. Prob.* **7**, 362–381.
- [25] PUHALSKII, A. A. AND WHITT, W. (1998). Functional large deviation principles for waiting and departure processes. *Prob. Eng. Inf. Sci.* **12**, 479–507.
- [26] RESNICK, S. AND STÄRICĂ, C. (1998). Tail index estimation for dependent data. *Ann. Appl. Prob.* **8**, 1156–1183.
- [27] SANOV, I. N. (1957). On the probability of large deviations of random magnitudes. *Mat. Sb. N. S.* **42**, 11–44.
- [28] SEGERS, J. (2002). Abelian and Tauberian theorems on the bias of the Hill estimator. *Scand. J. Statist.* **29**, 461–483.
- [29] SHORACK, G. R. AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley, New York.
- [30] SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Teor. Veroyat. Primen.* **1**, 289–319.
- [31] STIGLER, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **2**, 676–693.

- [32] VANDEMAELE, M. AND VERAVERBEKE, N. (1982). Cramér type large deviations for linear combinations of order statistics. *Ann. Prob.* **10**, 423–434.
- [33] VARADHAN, S. R. S. (1966). Asymptotic probabilities and differential equations. *Commun. Pure Appl. Math.* **19**, 261–286.
- [34] VOL'PERT, A. I. (1967). Spaces BV and quasilinear equations. *Mat. Sb. N. S.* **73**, 255–302.
- [35] WHITT, W. (1980). Some useful functions for functional limit theorems. *Math. Operat. Res.* **5**, 67–85.
- [36] WHITT, W. (2002). *Stochastic-Process Limits*. Springer, New York.