Sample path large deviations for order statistics

Ken R. Duffy* Claudio Macci† Giovanni Luca Torrisi‡

January 28, 2009

Abstract

We consider the sample paths of the order statistics of i.i.d. 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 Skorohod (J_1) topology. If F corresponds to a discrete distribution, we prove that the sample paths satisfy the large deviation principle in the topology of weak convergence. Versions of Sanov's Theorem are deduced as a corollary to these results. A number of illustrative examples are presented, including applications to the sample paths of trimmed means and Hill Plots.

1 Introduction

Let $\{X_n : n \geq 1\}$ be a sequence of i.i.d. real-valued random variables with distribution function $F(x) = P(X_1 \leq x)$. 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 $X_{1,n}, \ldots, X_{n,n}$ denote the ascending order statistics of $X_1, \ldots, X_n$, so that $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ 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

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

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 [31]. We consider $X_n(\cdot)$ as a random element of the space of non-decreasing cádlág functions (right continuous functions with left hand

*Hamilton Institute, National University of Ireland, Maynooth, Ireland. E-mail: ken.duffy@nuim.ie

†Dipartimento di Matematica, Università di Roma "Tor Vergata", Via della Ricerca Scientifica, I-00133 Rome, Italy. e-mail: macci@mat.uniroma2.it

‡Istituto per le Applicazioni del Calcolo "Mauro Picone", Consiglio Nazionale delle Ricerche (CNR), Via della Ricerca Scientifica, I-00133 Rome, Italy. e-mail: torrisi@iac.rm.cnr.it

limits) ϕ such that $\phi(0) \geq a$ and $\phi(1) = b$. We equip this space with the Skorohod (J_1) topology [28] if F is strictly increasing and with the topology of weak convergence if F is the distribution function of a simple (i.e. finite valued) random variable. We prove that these random paths satisfy the Large Deviation Principle (LDP). That is, for all Borel sets B

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

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 semi-continuous and has compact level sets (i.e. it is a good rate function, see e.g. [8]).

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

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

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 3). This relationship suggests that one way to prove that order statistics sample paths satisfy the LDP is to begin with Sanov's Theorem [25], 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 [5] 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 strengthens the topology in Sanov's Theorem from the topology of weak convergence to the topology generated by the L_2-Wasserstein metric. This enables 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 paper differs from Boistard [5] in two significant ways: (1) the method of proof and (2) 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 Mogul'skii's Theorem [20] and then applying Puhalskii's extension of the contraction principle [22] 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. When the underlying distribution function is strictly increasing (although possibly discontinuous), this approach leads to the functional LDP holding in the Skorohod (J_1) topology. As this is a relatively fine topology, deductions using the contraction principle are stronger than with (say) the topology of weak convergence. Indeed, from this topology we can 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.

If the distribution function F is strictly increasing (but possibly discontinuous) an expression for the rate function, J^F in equation (2), is given in equation (5). If $F(x) = \int_a^x f(y)\,dy$ where $f(y) > 0$ almost everywhere, this reduces to the formula in equation (9). If F is a discrete distribution function, a formula for J^F is given in equation (14). These functional large deviation principles enable 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 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 large deviation principle for Hill Plots, which enables estimates on the likelihood that Hill's [14] 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 set-up and notation. For strictly increasing F, the functional LDP in the Skorohod (J_1) topology for order statistics is presented in Section 3. In Section 4 we temporarily restrict our attention to simple random variables and illustrate why a coarser topology is necessary. Applications of these results are presented in Section 5.

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 [10]. The metrics ρ_1 and ρ_2 are topologically equivalent when restricted to $[0,1]$. As the use of ρ_1 or ρ_2 is solely a technicality, we let ρ denote either ρ_1 or ρ_2, with the usage being dependent on if we are working with real or extended-real valued functions.

Let $D[0,1]$ denote the space of real (or extended-real) valued cádlág functions on the closed interval $[0,1]$ equipped with the Skorohod (J_1) topology [28][3][33] induced by the metric

$$d(\phi,\psi) := \inf_{\lambda\in\Lambda} \sup_{t\in[0,1]} \max\{\rho(\phi(\lambda(t)),\psi(t)), \rho(\lambda(t),t)\}, \tag{3}$$

where Λ is the set of strictly increasing functions λ from $[0,1]$ to $[0,1]$ that are continuous, have a continuous inverse, and for which $\lambda(0) = 0$ and $\lambda(1) = 1$. For each $-\infty \le a < b \le \infty$, let $\mathcal{V}_{a,b}^+ \subset D[0,1]$ denote the closed set of non-decreasing functions ϕ such that $\phi(t) \geq a$ and $\phi(1) = b$. We will equip each $\mathcal{V}_{a,b}^+$ with the Skorohod (J_1) topology, apart from in Section 4 where we shall use the topology of weak convergence. The Skorohod (J_1) topology is finer than the topology of weak convergence, but for the distributions of simple random variables, it shall prove necessary to use the latter.

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

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

where $\phi^{(a)}$ 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

$$F^{-1}(u) := \inf\{x : F(x) > u\} \text{ if } u \in [0,1) \text{ and } F^{-1}(1) := b, \tag{4}$$

is the right continuous generalized inverse of F.

3 Strictly increasing distribution functions

The following theorem is the cornerstone result. It proves the functional LDP in the Skorohod (J_1) topology for the sample paths of order statistics where the distribution function F is strictly increasing, but possibly discontinuous.

Theorem 1 (Strictly increasing distribution functions) *If F is strictly increasing, then the sample paths $\{X_n(\cdot)\}$ satisfy the LDP in $\mathcal{V}_{a,b}^+$ equipped with Skorohod (J_1) topology with rate function*

$$J^F(\chi) = \inf_{\phi \in \mathcal{V}_{0,1}^+} \left\{ -\int_0^1 \log(\dot{\phi}^{(a)}(t))dt : F^{-1}(\phi(t)) = \chi(t) \textit{ for all } t \in [0,1] \right\}. \tag{5}$$

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 $U_{1,n}, \ldots, U_{n,n}$ be the order statistics of $U_1, \ldots, U_n$, 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 $\{Y_n\}$ be i.i.d.

exponentially distributed random variables with mean 1. Define the self-normalized random functions $\{N_n\}$ by

$$N_n(t) := \left(\sum_{i=1}^{[(n+1)t]} Y_i \right) / \left(\sum_{j=1}^{n+1} Y_j \right), \text{ if } \sum_{j=1}^{n+1} Y_j > 0. \tag{6}$$

As a consequence of Proposition 8.2.1 in Shorack and Wellner [27], $N_n(\cdot)$ is equal in distribution to $U_n(\cdot)$. By an application of Mogul'skii's Theorem [20][1], the sample paths

$$S_n(t) := \frac{1}{n} \sum_{i=1}^{[(n+1)t]} Y_i \tag{7}$$

satisfy the LDP in $D[0,1]$ with rate function that is finite only for functions ϕ that are non-decreasing and of finite variation. For such a ϕ, the rate function is $I(\phi) = \int_0^1 I_l(\dot{\phi}^{(a)}(t))dt + \phi^{(s)}(1)$, 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 equation (7) and $N_n(\cdot)$ is defined in equation (6). The map g is continuous at all ϕ such that $\phi(1) > 0$ as if $\phi_n \to \phi$ in $D[0,1]$ equipped with the Skorohod (J_1) topology, then $\phi_n(1) \to \phi(1)$. As $I_l(0) = -\log(0) - 1 = \infty$, $I(\phi) < \infty$ only if $\phi(1) > 0$. Puhalskii's extension of the contraction principle, Theorem 2.1 [22], only requires g to be continuous at all limit points where the rate function is finite in order for the usual contraction principle result to hold (e.g. Theorem 4.2.1 [8]). 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:

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

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

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

[1]The version of Mogul'skii's Theorem reported in Theorem 5.1.2 [8] is insufficient for our needs, as it does not encompass the case of exponentially distributed random variables. See also [18][21][22][23].

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 equation (8). Note that, as $\log(0) = -\infty$, $J^U(\psi) = \infty$ unless the absolutely continuous component of ψ's Lebesgue decomposition is strictly increasing almost everywhere with respect to Lebesgue measure.

Consider a sequence of i.i.d. random variables $\{X_n\}$ with common distribution function $F(\cdot)$. As is well known (e.g. Theorem 14.1 [4]), with the quantile function $F^{-1}(u)$ defined in equation (4) and $\{U_n\}$ 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 non-decreasing function of u, we have that $F^{-1}(U_n(\cdot))$ is exactly the sample path of the order statistics of $F^{-1}(U_1), \ldots, F^{-1}(U_n)$. 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 (e.g. Theorem 4.2.1 [8]) can be applied.

As F is assumed to be strictly increasing (although it can have discontinuities), F^{-1} is continuous on $[0, 1]$ (e.g. Lemma 13.6.4 [33]). Note that $F^{-1}(\phi) := F^{-1} \circ \phi$ and Theorem 3.1 [32] 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 non-decreasing, so long as the range of the elements of $D[0, 1]$ is a Polish space[2]. Thus $F^{-1} : \mathcal{V}^+_{0,1} \mapsto \mathcal{V}^+_{a,b}$ is continuous and Theorem 1 follows from an application of the contraction principle (e.g. Theorem 4.2.1 [8]).

■

Comment on the requirement for F to be strictly increasing: In Section 4 we shall show that the Skorohod (J_1) topology is too strong for the LDP for the order statistics of simple random variables. In that case we give a lesser result in the topology of weak convergence.

We now state a corollary of Theorem 1 that follows from the chain rule.

Corollary 2 (Distribution functions with positive densities a.e.) *If $F(x) = \int_a^x f(y)\,dy$ and f is positive almost everywhere, so that $F(x)$ is strictly increasing and continuous, then*

[2]This is where the distinction between ρ_1 and ρ_2 matters in equation (3), as the extensions of the real-line are not complete separable metric spaces when equipped with the Euclidean norm.

$J^F(\chi) = \infty$ *unless* $F \circ \chi \in \mathcal{V}_{0,1}^+$ *is strictly increasing in which case*

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

In the next subsection we present some illustrative examples based on Theorem 1 and Corollary 2.

3.1 Examples

Example I 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 (cannot) emulate the order statistics of random variables with distribution function G if $J^F(G^{-1}) < \infty$ $(= \infty)$. Examples II to V below concern order statistics of given laws that can or cannot emulate the order statistics of other distributions. In particular, Example IV shows that the order statistics of Pareto distributions, even those with finite mean, can emulate those with infinite mean. Example V 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 Section 5.2 Example VII when we consider trimmed means.

Example I: 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 $\chi \in \mathcal{V}_{0,1}^+$. Define the set $A := \{\phi : \phi(t) \leq 1/3 \text{ for all } t \in [0, 1)\}$. Note that $X_n(\cdot) \in A$ if and only if $X_{n,n} \in [0, 1/3]$, i.e. $X_i \in [0, 1/3]$ for all $i \in \{1, \ldots, n\}$, and therefore $P(X_n(\cdot) \in A) = (1/3)^n$. The exponent in the decay of this probability is $-\log(1/3)$. 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 $J^U(\hat{\chi}) = -\log(1/3)$ and therefore $\lim n^{-1} \log P(X_n(\cdot) \in A) = -\log(1/3)$.

Note that the sample path LDP gives more information than the direct calculation. It shows that the most likely path to this event is that $X_{1,n}, \ldots, X_{n,n}$ be spread uniformly over $[0, 1/3]$, in the following sense. By, for example, Theorem 3.1 (b) of Lewis, Pfister and Sullivan [17], for any $\varepsilon > 0$

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

where $B_\varepsilon(\hat\chi)$ is the open ball of radius ε around $\hat\chi \in D[0,1]$ equipped with the Skorohod (J_1) topology. That is, conditioned on $X_{n,n} \le 1/3$, 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 $[0,1/3]$.

Example II: 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 2, $\{X_n(\cdot)\}$ satisfies the LDP in $\mathcal{V}_{0,1}^+$ with rate function

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

for any $\chi \in \mathcal{V}_{0,1}^+$. Considering $J^{\mathrm{Beta}(\alpha,\beta)}(\hat\chi)$ 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

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

This function has its minimum, $J^{\mathrm{Beta}(\alpha,\beta)}(\hat\chi) = 0$, when the $\{X_n\}$ random variables have a uniform distribution, $\alpha = \beta = 1$. Note that if $\alpha = 1$, then as $\Gamma(1+\beta) = \beta\Gamma(\beta)$, $J^{\mathrm{Beta}(\alpha,\beta)}(\hat\chi) = \beta - \log(\beta) - 1$. 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 III: rate function for the Exponential distribution. If $F(x) = 1 - \exp(-\lambda x)$ for all $x \ge 0$ so that $a = 0$ and $b = +\infty$, then $F^{-1}(u) = -\log(1-u)/\lambda$. Thus by Corollary 2

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

For example, if $\hat\chi(t) = F^{-1}(t) = -\log(1-t)/\lambda$, then $J^{\mathrm{Exp}(\lambda)}(\hat\chi) = 0$. 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 $J^{\mathrm{Exp}(\lambda)}(\hat\chi_K) = -\log(\lambda) + \lambda K/2 - \log(K)$. 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 IV: 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 2

$$J^{\text{Pareto}(\alpha)}(\chi) = -\log(\alpha) + (\alpha+1)\int_0^1 \log(\chi(t))\, dt - \int_0^1 \log(\dot{\chi}^{(a)}(t))\, dt. \qquad (11)$$

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 $K \to 0$, so that $\alpha_K \to \infty$ 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 V: Exponential and Pareto distribution. If $\hat{\chi}(t) = -\log(1-t)/\lambda + 1$ corresponding to a quantile function of an Exponential law on $[1, \infty)$, then $J^{\text{Pareto}(\alpha)}(\hat{\chi}) = \log(\lambda/\alpha) - 1 - \exp(\lambda)(\alpha+1)\text{Ei}(-\lambda)$, 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 $\alpha \le 1$ and, if $\alpha > 1$, $J^{\text{Exp}(\lambda)}(\hat{\chi}) = \log(\alpha/\lambda) + (\lambda\alpha + 1 - \alpha^2)/(\alpha(\alpha-1)) < \infty$. 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 5.2, Example VII.

3.2 Comment on Sanov's Theorem

Sanov's Theorem (e.g [8] 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 the empirical laws satisfy the LDP with relative entropy as the rate function. For some modern developments, see for example [16] and 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

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

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 non-decreasing right continuous element in the space of cádlág 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 Skorohod (J_1) topology or its natural extension if either $a = -\infty$ or $b = +\infty$ (the net $\{\phi_\iota\}$ converges to ϕ for all continuity points t of ϕ and the restrictions of ϕ_ι to $[a,t]$ converge, $d(\phi_\iota, \phi) \to 0$, to ϕ restricted to $[a,t]$).

Corollary 3 (A version of Sanov's Theorem) *Assume that $F(x) = \int_a^x f(y)\,dy$ where $f(y) > 0$ almost everywhere and $-\infty \leq a < b \leq \infty$. Then $\{F_n\}$ satisfies the LDP in $\mathcal{V}^+_{0,1}[a,b]$ with rate function $H^F(\chi) = \infty$ unless χ is absolutely continuous, in which case*

$$H^F(\chi) = \int_a^b \dot{\chi}^{(a)}(s) \log\left(\frac{\dot{\chi}^{(a)}(s)}{f(s)}\right) ds. \tag{12}$$

Proof: Note that, defining $\inf\emptyset = n+1$,

$$F_n(x) = \frac{1}{n}\left(\inf\{m \in \{1,\dots,n\} : X_{m,n} > x\} - 1\right) = \left(1 + \frac{1}{n}\right) X_n^{-1}(x) - \frac{1}{n}.$$

so that $\sup_{x\in[a,b]} |F_n(x) - X_n^{-1}(x)| \leq 2/n$. Hence the sequences $\{F_n\}$ and $\{X_n^{-1}(\cdot)\}$ are exponentially equivalent (e.g. definition 4.2.10 [8]) in the uniform topology. Thus to prove $\{F_n\}$ satisfies the LDP, it suffices to show that $\{X_n^{-1}(\cdot)\}$ does. By Theorem 7.2 [32] (with straight-forward 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 at all strictly increasing ϕ. As $J^F(\phi) = \infty$ unless ϕ is strictly increasing, the LDP for $\{X_n^{-1}(\cdot)\}$ follows from an application Puhalskii's extension of the contraction principle, Theorem 2.1 [22]. For $\chi \in \mathcal{V}^+_{0,1}[a,b]$, as $\dot{(\chi^{-1})}(t) = 1/(\dot{\chi}^{(a)}(\chi^{-1}(t)))$ Lebesgue almost everywhere and F has a density f that is positive almost everywhere, with

$\chi^{-1} \in \mathcal{V}^+_{a,b}$ being the right-continuous inverse of χ,

$$
\begin{aligned}
H^F(\chi) = J^F(\chi^{-1}) &= -\int_0^1 \left(\log(f(\chi^{-1}(t))) + \log\left(\frac{1}{\dot{\chi}^{(a)}(\chi^{-1}(t))} \right) \right) dt \\
&= -\int_{\chi^{-1}(0)}^b \left(\log(f(s)) + \log\left(\frac{1}{\dot{\chi}^{(a)}(s)} \right) \right) \dot{\chi}^{(a)}(s)\, ds \\
&= \int_a^b \dot{\chi}^{(a)}(s) \log\left(\frac{\dot{\chi}^{(a)}(s)}{f(s)} \right) ds,
\end{aligned}
$$

where the change in range in the last line follows as $\dot{\chi}(s) = 0$ for $s \in [a, \chi^{-1}(0))$ and $0 \log(0) := 0$.

■

The form for the rate function in equation (12) is analogous to that of relative entropy, but here the LDP holds in the Skorohod (J_1) topology.

4 Discrete distributions

We would like a result that is similar to Theorem 1 when F corresponds to discrete distribution taking a finite number of distinct values, but shall demonstrate that we cannot do so within the confines of the Skorohod (J_1) topology. Consider $P(X_1 = i) = p_i > 0$ for $i \in \{1, \dots, K\}$ and $K < \infty$ with $\sum_{i=1}^K p_i = 1$ and its corresponding quantile function

$$
F^{-1}(u) = \begin{cases} 1 & \text{if } u \in [0, p_1) \\ 2 & \text{if } u \in [p_1, p_1 + p_2) \\ \vdots & \vdots \\ K-1 & \text{if } u \in [p_1 + \ldots + p_{K-2}, p_1 + \ldots + p_{K-1}) \\ K & \text{if } u \in [p_1 + \ldots p_{K-1}, 1] \end{cases}
$$

The following example shows that $F^{-1} : \mathcal{V}^+_{0,1} \mapsto \mathcal{V}^+_{1,K}$, defined by $F^{-1}(\phi)(t) = F^{-1}(\phi(t))$, is not continuous when the image space is equipped with the Skorohod (J_1) topology.

Example VI: discontinuity of F^{-1} for discrete distributions in Skorohod (J_1) topology. Let $K = 3$ and $p_1 = p_2 = p_3 = 1/3$, so that

$$
F^{-1}(u) = \begin{cases} 1 & \text{if } u \in [0, 1/3) \\ 2 & \text{if } u \in [1/3, 2/3) \\ 3 & \text{if } u \in [2/3, 1]. \end{cases}
$$

Consider the following element of $\mathcal{V}^+_{0,1}$:

$$\phi(t) = \begin{cases} 2t/3 & \text{if } t \in [0, 1/2) \\ 1/3 + 2t/3 & \text{if } t \in [1/2, 1]. \end{cases}$$

The sequence $\phi_n(t) = \phi(t) + 2/(3n)1_{t\in[1/2-1/n,1/2)}$ is such that $d(\phi_n, \phi) \leq 1/n$, so that $\phi_n \to \phi$ in the Skorohod (J_1) topology. However

$$F^{-1}(\phi(t)) = \begin{cases} 1 & \text{if } t < 1/2 \\ 3 & \text{if } t \geq 1/2 \end{cases}$$

and, for $n \geq 3$,

$$F^{-1}(\phi_n(t)) = \begin{cases} 1 & \text{if } t < 1/2 - 1/n \\ 2 & \text{if } t \in [1/2 - 1/n, 1/2) \\ 3 & \text{if } t \geq 1/2. \end{cases}$$

Thus $F^{-1}(\phi_n)$ and $F^{-1}(\phi)$ have unmatched jumps so that $F^{-1}(\phi_n)$ does not converge to $F^{-1}(\phi)$ in the Skorohod (J_1) topology (e.g. pg. 80 [33]) and therefore F^{-1} is not continuous when the image space is equipped the Skorohod (J_1) topology. For discrete random variables, instead we shall employ the topology of weak convergence. Functions $\{\phi_n\}$ converge to ϕ in the topology of weak convergence if $\phi_n(t) \to \phi(t)$ at all continuity points t of ϕ. As $\mathcal{V}^+_{a,b}$ consists of non-decreasing functions, this topology coincides with the Skorohod (M_1) topology, e.g. Corollary 12.5.1 [33].

Theorem 4 (Discrete distributions) *If F is strictly discrete taking values $\{1, \ldots, K\}$, then $\{X_n(\cdot)\}$ satisfies the LDP in $\mathcal{V}^+_{1,K}$ equipped with the topology of weak convergence with a rate function that is infinite at χ unless χ is piecewise-constant taking values in $\{1, \ldots, K\}$. That is χ is such that for some subset $j_1 < \cdots < j_{K'}$ of $\{1, \ldots, K\}$ there is a non-decreasing sequence $0 = t^\chi_{j_0} < t^\chi_{j_1} < \ldots < t^\chi_{j_{K'}} = 1$ for which*

$$\chi(t) = \begin{cases} j_1 & \textit{if } t \in [t^\chi_{j_0}, t^\chi_{j_1}) \\ j_2 & \textit{if } t \in [t^\chi_{j_1}, t^\chi_{j_2}) \\ \vdots & \vdots \\ j_{K'} & \textit{if } t \in [t^\chi_{j_{K'-1}}, t^\chi_{j_{K'}}) \\ K & \textit{if } t = 1. \end{cases} \tag{13}$$

in which case

$$J^F(\chi) = \sum_{l=1}^{K'} \left(t^\chi_{j_l} - t^\chi_{j_{l-1}} \right) \log \left(\frac{t^\chi_{j_l} - t^\chi_{j_{l-1}}}{p_{j_l}} \right). \tag{14}$$

PROOF: The proofs of Theorems 1 and 4 follow the same course until the form of F^{-1} must be taken into account. Here F is discrete and $\mathcal{V}^+_{1,K}$ is equipped with the topology of weak convergence.

Assume that $d(\phi_n, \phi) \to 0$ and, as $J^U(\phi) = \infty$ if ϕ is not strictly increasing, that ϕ is strictly increasing. The function $F^{-1} : \mathcal{V}^+_{0,1} \mapsto \mathcal{V}^+_{1,K}$ is continuous when $\mathcal{V}^+_{0,1}$ is equipped with the Skorohod (J_1) topology and $\mathcal{V}^+_{1,K}$ is equipped with the topology of weak convergence if the following condition holds: $d(\phi_n, \phi) \to 0$ implies that $F^{-1}(\phi_n(t)) \to F^{-1}(\phi(t))$ at all t such that $F^{-1}(\phi(t))$ is continuous.

Define the sequence

$$t^\phi_k := \inf\{x : \phi(x) \in [p_1 + \ldots + p_{k-1}, p_1 + \ldots + p_k)\}$$

for $k \in \{1, 2, \ldots, K-1\}$ and $t^\phi_K = \inf\{x : \phi(x) \in [p_1 + \ldots + p_{K-1}, 1]\}$, where $\inf \emptyset := +\infty$. The finite collection $\{t^\phi_k : t^\phi_k < \infty\}$ is the set of discontinuities of $F^{-1}(\phi(t))$. Without loss of generality, assume that $t^\phi_k < \infty$ for all $k \in \{1, \ldots, K\}$.

Let $t_i \in (t^\phi_{i-1}, t^\phi_i)$, so that t_i is a continuity point of $F^{-1}(\phi(t))$ and $F^{-1}(\phi(t_i)) = i$. As ϕ is strictly increasing and right continuous, we can find an $\varepsilon > 0$ such that $\phi(t_i - \varepsilon) > p_1 + \ldots + p_{i-1} + \varepsilon$ and $\phi(t_i + \varepsilon) < p_1 + \ldots + p_i - \varepsilon$. As $d(\phi_n, \phi) \to 0$, we can find λ_n and N_ε such that

$$\sup_{t\in[0,1]} |\phi_n(t) - \phi(\lambda_n(t))| < \varepsilon \text{ and } \sup_{t\in[0,1]} |\lambda_n(t) - t| < \varepsilon$$

for all $n > N_\varepsilon$. The first inequality implies that $\phi_n(t_i) \in (\phi(\lambda_n(t_i)) - \varepsilon, \phi(\lambda_n(t_i)) + \varepsilon)$ for all $n > N_\varepsilon$, while the second implies that $\lambda_n(t_i) \in (t_i - \varepsilon, t_i + \varepsilon)$ for all $n > N_\varepsilon$. Thus, as ϕ is increasing, $\phi(t_i - \varepsilon) > p_1 + \ldots + p_{i-1} + \varepsilon$ and $\phi(t_i + \varepsilon) < p_1 + \ldots + p_i - \varepsilon$, $\phi_n(t_i) \in (p_1 + \ldots + p_{i-1}, p_1 + \ldots + p_i)$. Therefore $F^{-1}(\phi_n(t_i)) = i$ for all $n > N_\varepsilon$ and $F^{-1}(\phi_n(t_i)) \to F^{-1}(\phi(t_i))$. As this is true at all continuity points of $F^{-1}(\phi)$, we have that $F^{-1}(\phi_n)$ converges weakly to $F^{-1}(\phi)$ for all strictly increasing ϕ. Thus the claimed LDP follows from an application of Puhalskii's extension of the contraction principle, Theorem 2.1 [22].

The form of the rate function in (14) comes from the following argument. Note that J^F is finite only on piecewise constant functions taking values on $\{1, \ldots, K\}$. Now, let χ be the piecewise-constant function defined in (13) and $\phi \in \mathcal{V}^+_{0,1}$ such that $F^{-1}(\phi) = \chi$. Then with $p_0 := 0$

$$\phi(t) \in \begin{cases} [p_1 + \ldots + p_{j_1 - 1}, p_1 + \ldots + p_{j_1}) & \text{if } t \in [0, t^\chi_{j_1}) \\ [p_1 + \ldots + p_{j_2 - 1}, p_1 + \ldots + p_{j_2}) & \text{if } t \in [t^\chi_{j_1}, t^\chi_{j_2}) \\ \vdots & \vdots \\ [p_1 + \ldots + p_{j_{K'} - 1}, p_1 + \ldots + p_{j_{K'}}) & \text{if } t \in [t^\chi_{j_{K'-1}}, t^\chi_{j_{K'}}) \\ \{1\} & \text{if } t = 1. \end{cases}$$

Observing that $-\log(u)$ is convex for $u > 0$, Jensen's inequality shows that the infimum path is a straight line between each $t^{\chi}_{j_{i-1}}$ and $t^{\chi}_{j_i}$ with slope $p_{j_i}/(t^{\chi}_{j_i} - t^{\chi}_{j_{i-1}})$.

■

Comment on Sanov's Theorem: As in Corollary 3, the rate function J^F in equation (14) is closely related to the empirical law rate function in Sanov's Theorem for discrete random variables (e.g Theorem 2.1.24 [8]). However, as the topology of weak convergence is coarser than the Skorohod (J_1) topology, fewer functions are continuous from this space and thus this is a lesser result that in Theorem 1.

Example VII: Bernoulli random variables. If $\{X_i\}$ are Bernoulli with $P(X_i = 1) = p$ and $P(X_i = 2) = 1-p$, then $F : [1,2] \mapsto [0,1]$ is $F(x) = p$ if $x \in [1,2)$ and $F(x) = 1$ if $x = 2$. Hence $F^{-1}(u) = 1$ if $u \in [0,p)$ and $F^{-1}(u) = 2$ if $u \in [p,1]$. Thus $J^F(\chi)$ is finite if there exists $t_1 \in [0,1]$ such that $\chi(t) = 1$ for $t \in [0,t_1)$ and $\chi(t) = 2$ for $t \in [t_1,1]$, in which case $J^F(\chi) = t_1 \log(t_1/p) + (1-t_1)\log((1-t_1)/(1-p))$. This is the relative entropy for observing the symbols 1 and 2 in the proportions t_1 and $1-t_1$.

5 Applications

5.1 Sample path large deviations for L-statistics

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

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

Much is known regarding the large n behavior of the sequence of random variables $\{T_n\}$ and, in particular, its central limit behavior, e.g. Stigler [29], Vandemaele and Veraverbeke [30], Callaert, Vandemaele and Veraverbeke [6] and Aleshkyavichene [2]. Large deviation results for L-statistics can be found in Groeneboom, Oosterhoff and Ruymgaart [11], Groeneboom and Shorack [12], and Boistard [5].

Consider the empirical 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.

5.2 Trimmed means

The function K can be chosen to remove outliers. When

$$K(u) = \begin{cases} 1/(1-2\gamma) & \text{if } u \in [\gamma, 1-\gamma] \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

for $\gamma \in (0, 1/2)$, 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 $\gamma = 1/4$, it is called the interquartile mean.

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

$$T_n(t) = (1-2\gamma)^{-1} \int_\gamma^t X_{[(n+1)s],n}\, ds \text{ for } t \in [\gamma, 1-\gamma] \tag{16}$$

Theorem 5 (Trimmed means) *Assume that F is strictly increasing. If K is of the form in equation (15) with $\gamma \in (0,1/2)$, 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 the 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 equation (5). 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

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

and note that $g_\gamma(X_{[(n+1)\cdot],n}) = T_n(\cdot)$. As Theorem 1 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 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 (e.g. Theorem 11.5.1 [33]) 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 equation (5),

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

As $\psi(1-\gamma) = 1$, $\psi(1) = 1$ and ψ is non-decreasing, $\dot{\psi}^{(a)}(t) = 0$ for $t \in (1-\gamma, 1]$, $-\log(0) = \infty$ and thus $J^F(\phi) = \infty$. Hence the LDP follows from Theorem 1 and an application of Puhalskii's extension of the contraction principle Theorem 2.1 [22].

■

In contrast to Example V, Section 3.1, the following example shows that for trimmed means, Exponential Laws *can* emulate Pareto Laws with finite rate, even Pareto Laws with infinite mean.

Example VIII: 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 $J^{\mathrm{Exp}(\lambda)}(\chi)$ is given in equation (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 Section 3.1, Example V). Using the expression in Theorem 1,

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

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:

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

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

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

If $\alpha = 1$, the last term in is replaced with $\lambda(\log(1-\gamma) - \log(\gamma))$. For any $\lambda > 0$ and any $\gamma \in (0, 1/2)$, this expression is finite for $\alpha \in (0,1)$, although growing quickly as $\alpha \to 0$. Thus the paths of trimmed means of exponential distributions can mimic those of a Pareto distribution with infinite mean, even though Example V in Section 3.1 shows that this is *not* possible for untrimmed means.

5.3 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 $X_{1,n},\ldots,X_{n,n}$ 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 x sufficiently large. Hill's [14] widely-used methodology to answer this question employs the following empirical quantities: for each $1 \leq k \leq n$, define

$$H_{k,n} := \frac{1}{k+1}\sum_{i=n-k+1}^{n} \log(X_{i,n}) - \frac{k}{k+1}\log(X_{n-k,n}). \tag{17}$$

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, one creates a "Hill Plot": for a contiguous range of small k, e.g $k \in \{[n/100], [n/100]+1, \ldots, [n/10]\}$, one plots $H_{k,n}$ versus k. If the resulting plot is almost a straight line, one deduces 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 one over the height of the line.

Due to its practical importance, much is known about properties of Hill's estimator (e.g. de Hann and Resnick [7], Resnick and Stărică [24], Drees, de Hann and Resnick [9], Segers [26], Haeusler and Segers [13] and references therein). Here, as an application of Theorem 1, 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

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

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 large deviation principle 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 6 (LDP for Hill Plots) *Assume the same hypotheses of Theorem 1 with $a \geq 1$. If, in addition, either: $b < \infty$; or $F(x)$ is differentiable for all x sufficiently large, there exists $\beta \in (0, 1)$ such that*

$$\lim_{\varepsilon \to 0} \varepsilon \log \left(1 - F(\exp(\varepsilon^{-\beta}))\right) = -\infty. \tag{18}$$

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

$$\frac{d}{dt} x_F(t) \geq \frac{x_F(t)(1 - x_F(t))}{x_F(t) - t} \log\left(\frac{(1-t)x_F(t)}{t(1 - x_F(t))}\right). \tag{19}$$

Then $\{H_n(\cdot)\}$ satisfies the LDP in $D[0,1]$, equipped with the Skorohod (J_1) topology with the rate function:

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

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

$$h_1(\chi)(t) = (\chi(t), \chi(1-t)),\ h_2(\chi,\psi)(t) = (\log\chi(t), \log\psi(t)),$$
$$h_3(\chi,\psi)(t) = \left(\int_{1-t}^{1} \chi(s)ds, \psi(t)\right) \text{ and } h_4(\chi,\psi)(t) = \frac{1}{t}\chi(t) - \psi(t).$$

The function h_1 is continuous by Theorem 8.1 [32], while h_2, using the continuity of $\log(\cdot)$, is continuous by Theorem 3.1 [32]. For continuous χ, the function h_4 is continuous by Theorems 4.1 and 4.2 [32]. If $b < \infty$, then we can appeal to Theorem 11.5.1 [33] 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 [33] 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

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

that approximate the behavior of $g(\chi)$. By similar logic to that in Theorem 5, for any $\varepsilon > 0$ the function g_ε is continuous at all χ such that $J^F(\chi) < \infty$, so that Puhalskii's extension of the contraction principle can be applied, obtaining the LDP for $\{g_\varepsilon(X_n)\}$.

Thus we will show that $\{g(X_n)\}$ satisfies the LDP by applying the approximate contraction principle Theorem 4.2.23 [8]. This approach requires $\{g_\varepsilon(X_n)\}$ to be exponentially good approximations of $\{g(X_n)\}$,

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P\left(\int_{1-\varepsilon}^{1}\log(X_n(s))\,ds\geq\delta\right)=-\infty, \tag{20}$$

as well as the verification of equation (4.2.24) [8] for which it suffices to prove that

$$\limsup_{\varepsilon\to 0}\sup_{\{\chi\in\mathcal{V}_{a,b}^{+}:J^F(\chi)\leq\alpha\}}\int_{1-\varepsilon}^{1}\log(\chi(s))\,ds=0,\ \text{ for every } \alpha\in(0,\infty). \tag{21}$$

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 $\chi\in\mathcal{V}_{a,b}^{+}$,

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

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

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

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 (19) holds. Then

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

By assumption, $x_F(t) > t$ for all t sufficiently close to 1 and by equation (19) the function

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

is increasing. Thus

$$\inf\{J^F(\chi) : \chi \in \bar{A}_\varepsilon\} \geq -(1-\rho)\log\left(\frac{F(\exp(\rho^{-\beta})+1)}{1-\rho}\right) - \rho\log\left(\frac{1-F(\exp(\rho^{-\beta})+1)}{\rho}\right)$$

which tends to $+\infty$ as $\rho \to 0$ by assumption (18). Thus equation (20) is satisfied and the sequences $\{g_\varepsilon(X_n)\}$ are exponentially good approximations of $\{g(X_n)\}$.

To establish (21), reasoning by contradiction, assume that there exists $\delta > 0$ 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 $\chi \to \int_{1-\varepsilon}^1 \log(\chi(s))\,ds$ is continuous for any $\varepsilon > 0$. Therefore, by the goodness of J^F, there exist $\chi_{\varepsilon_n,\alpha}$ 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 equation (18), this final term tends to $+\infty$ as $n \to \infty$.

■

Example IX: Every Law that is ultimately Weibull satisfies the conditions of Theorem 6. Consider a law that is ultimately Weibull, $F(x) = 1 - e^{-x^\alpha}$ for some $\alpha > 0$ and for all x sufficiently large. For any $\beta \in (0,1)$,

$$\lim_{\varepsilon\to 0} \varepsilon\log\left(1 - F(\exp(\varepsilon^{-\beta}))\right) = -\lim_{\varepsilon\to 0}\varepsilon\exp\left(\alpha\varepsilon^{-\beta}\right) = -\infty.$$

and thus equation (18) 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 (19) is equivalent to

$$\alpha\beta(1-t)^{-(\beta+1)}(l(t)-1)(l(t))^{\alpha-1} \geq \frac{1-e^{-(l(t))^\alpha}}{1-e^{-(l(t))^\alpha}-t} \log\left(\frac{1-t}{t}\frac{1-e^{-(l(t))^\alpha}}{e^{-(l(t))^\alpha}}\right) \tag{22}$$

for all t sufficiently close to 1. Equation (22) holds if we can show that

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

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

$$\begin{aligned}(1-e^{-(l(t))^\alpha})\log\left(\frac{1-t}{t}\frac{1-e^{-(l(t))^\alpha}}{e^{-(l(t))^\alpha}}\right) &\leq (1-e^{-(l(t))^\alpha})\log(e^{(l(t))^\alpha}-1)\\ &\leq (1-e^{-(l(t))^\alpha})(l(t))^\alpha\end{aligned}$$

and so

$$\begin{aligned}&\frac{\alpha\beta(1-t)^{-(\beta+1)}(l(t)-1)(l(t))^{\alpha-1}(1-e^{-(l(t))^\alpha}-t)}{(1-e^{-(l(t))^\alpha})\log\left(\frac{1-t}{t}\frac{1-e^{-(l(t))^\alpha}}{e^{-(l(t))^\alpha}}\right)}\\ &\geq \frac{\alpha\beta(1-t)^{-(\beta+1)}(l(t)-1)(1-e^{-(l(t))^\alpha}-t)}{(1-e^{-(l(t))^\alpha})l(t)}.\end{aligned} \tag{23}$$

The claim follows noticing that the term in (23) goes to $+\infty$ as $t \to 1$ because it is asymptotically equivalent to $\alpha\beta(1-t)^{-\beta}$.

Example X: Truncated-Pareto emulating Pareto. As an application of Theorem 6, 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. Mantegna and Stanley [19]. Similar observations have been made in ground-water hydrology, atmospheric science and many other fields; for examples see Aban, Meerschaert and Panorska [1] and references therein. This has led to the proposal of, e.g., financial market models based on random walks whose increments have apparent power-tail behavior near the center 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 [19] or a transition to an exponential distribution beyond a given cut-off [15]. 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 6 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 2, the function f is defined by $f(x) := \dot{F}(x) = \gamma x^{-\gamma-1}$ if $x \in [1, K)$ and $f(x) := \dot{F}(x) = \lambda K^{-\gamma} \exp(-\lambda(x-K))$ if $x \in (K, \infty)$. Thus by the expression of L^F in Theorem 6 and (9) we have

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

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 $\alpha > 1$.

References

[1] Inmaculada B. Aban, Mark M. Meerschaert, and Anna K. Panorska, *Parameter estimation for the truncated Pareto distribution*, J. Amer. Statist. Assoc. **101** (2006), no. 473, 270–277.

[2] A. K. Aleshkyavichene, *Large and moderate deviations for L-statistics*, Litovsk. Mat. Sb. **31** (1991), no. 2, 227–241.

[3] Patrick Billingsley, *Convergence of probability measures*, John Wiley & Sons Inc., New York, 1968.

[4] ———, *Probability and measure*, third ed., John Wiley & Sons Inc., New York, 1995.

[5] Hélène Boistard, *Large deviations for L-statistics*, Statist. Decisions **25** (2007), no. 2, 89–125.

[6] H. Callaert, M. Vandemaele, and N. Veraverbeke, *A Cramér type large deviation theorem for trimmed linear combinations of order statistics*, Comm. Statist. A—Theory Methods **11** (1982), no. 23, 2689–2698.

[7] Laurens de Haan and Sidney Resnick, *On asymptotic normality of the Hill estimator*, Comm. Statist. Stochastic Models **14** (1998), no. 4, 849–866.

[8] A. Dembo and O. Zeitouni, *Large deviation techniques and applications*, Springer, 1998.

[9] Holger Drees, Laurens de Haan, and Sidney Resnick, *How to make a Hill plot*, Ann. Statist. **28** (2000), no. 1, 254–274.

[10] Bert Fristedt and Lawrence Gray, *A modern approach to probability theory*, Probability and its Applications, Birkhäuser Boston Inc., 1997.

[11] P. Groeneboom, J. Oosterhoff, and F. H. Ruymgaart, *Large deviation theorems for empirical probability measures*, Ann. Probab. **7** (1979), no. 4, 553–586.

[12] Piet Groeneboom and Galen R. Shorack, *Large deviations of goodness of fit statistics and linear combinations of order statistics*, Ann. Probab. **9** (1981), no. 6, 971–987.

[13] Erich Haeusler and Johan Segers, *Assessing confidence intervals for the tail index by Edgeworth expansions for the Hill estimator*, Bernoulli **13** (2007), no. 1, 175–194.

[14] Bruce M. Hill, *A simple general approach to inference about the tail of a distribution*, The Annals of Statistics **3** (1975), no. 5, 1163–1174.

[15] Ismo Koponen, *Analytic approach to the problem of convergence of truncated Lévy flights towards the gaussian stochastic process*, Phys. Rev. E **52** (1995), no. 1, 1197–1199.

[16] Christian Léonard and Jamal Najim, *An extension of Sanov's theorem: application to the Gibbs conditioning principle*, Bernoulli **8** (2002), no. 6, 721–743.

[17] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan, *Entropy, concentration of probability and conditional limit theorems*, Markov Process. Related Fields **1** (1995), no. 3, 319–386.

[18] James Lynch and Jayaram Sethuraman, *Large deviations for processes with independent increments*, Ann. Probab. **15** (1987), no. 2, 610–627.

[19] R. N. Mantegna and H. E Stanley, *Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight*, Phys. Rev. Lett. **73** (1994), no. 22, 2946–2949.

[20] A. A. Mogul′skiĭ, *Large deviations for trajectories of multi-dimensional random walks*, Th. Prob. Appl **21** (1976), 300–315.

[21] ______, *Large deviations for processes with independent increments*, Ann. Probab. **21** (1993), no. 1, 202–215.

[22] A. Puhalskii, *Large deviation analysis of the single server queue*, Queueing Systems **21** (1995), no. 1–2, 5–66.

[23] Anatolii A. Puhalskii and Ward Whitt, *Functional large deviation principles for waiting and departure processes*, Probab. Engrg. Inform. Sci. **12** (1998), no. 4, 479–507.

[24] Sidney Resnick and Catalin Stărică, *Tail index estimation for dependent data*, Ann. Appl. Probab. **8** (1998), no. 4, 1156–1183.

[25] I. N. Sanov, *On the probability of large deviations of random magnitudes*, Mat. Sb. N. S. **42 (84)** (1957), 11–44.

[26] Johan Segers, *Abelian and Tauberian theorems on the bias of the Hill estimator*, Scand. J. Statist. **29** (2002), no. 3, 461–483.

[27] Galen R. Shorack and Jon A. Wellner, *Empirical processes with applications to statistics*, John Wiley & Sons Inc., New York, 1986.

[28] A. V. Skorohod, *Limit theorems for stochastic processes*, Teor. Veroyatnost. i Primenen. **1** (1956), 289–319.

[29] Stephen M. Stigler, *Linear functions of order statistics with smooth weight functions*, Ann. Statist. **2** (1974), 676–693.

[30] M. Vandemaele and N. Veraverbeke, *Cramér type large deviations for linear combinations of order statistics*, Ann. Probab. **10** (1982), no. 2, 423–434.

[31] S. R. S. Varadhan, *Asymptotic probabilities and differential equations*, Comm. Pure Appl. Math. **19** (1966), 261–286.

[32] Ward Whitt, *Some useful functions for functional limit theorems*, Math. Oper. Res. **5** (1980), no. 1, 67–85.

[33] ______, *Stochastic-process limits*, Springer-Verlag, New York, 2002.