Adv. Appl. Prob. **36,** 198–226 (2004)
Printed in Northern Ireland

ESTIMATING TAIL DECAY FOR STATIONARY SEQUENCES VIA EXTREME VALUES

ASSAF ZEEVI,* *Columbia University*
PETER W. GLYNN,** *Stanford University*

Abstract

We study estimation of the tail-decay parameter of the marginal distribution corresponding to a discrete-time real-valued stationary stochastic process. Assuming that the underlying process is *short-range dependent*, we investigate properties of estimators of the tail-decay parameter which are based on the maximal extreme value of the process observed over a sampled time interval. These estimators only assume that the tail of the marginal distribution is roughly exponential, plus some modest 'mixing' conditions. Consistency properties of these estimators are established, as well as minimax convergence rates. We also provide some discussion on estimating the pre-exponent, when a more refined tail asymptotic is assumed. Properties of a certain moving-average variant of the extremal-based estimator are investigated as well. In passing, we also characterize the precise dependence (mixing) assumptions that support almost-sure limit theory for normalized extreme values and related first-passage times in stationary sequences.

Keywords: Extreme values; estimation; tail behavior; strong mixing; uniform mixing; minimax; almost-sure convergence

2000 Mathematics Subject Classification: Primary 60G70
Secondary 62G32; 90B20

1. Introduction

Consider a discrete-time real-valued stationary stochastic process $\mathbb{X} = (X_n : n \in \mathbb{Z}_+)$. In many applications, we are interested in the likelihood that this process takes on very large (or small) values, and desire methods to estimate this probability from a sequence of observations. Examples of the process $\mathbb{X}$ include the number of packets that await transmission in a switch or network router, backlogged demand for a certain product, or aggregate financial reserves in an insurance firm. In the case of insurance, the firm faces the risk of not meeting its obligations to policy holders if its financial reserves drop below a certain level. Excess backlog in the other two examples typically translates into reduced quality of service, viz. dropped packets and re-transmit requests in the former and potential due-date violations in the latter.

To fix ideas, let us consider the data network example. In this case $\mathbb{X}$ takes on nonnegative values, and in order to maintain smooth network operation, the fraction of dropped packets at a given switch should be kept below a certain threshold, say δ. Thus, for a given buffer size b, the constraint could be in the form $\Pr(X > b) \le \delta$. In practice, the probability distribution

Received 20 November 2001; revision received 21 August 2003.
* Postal address: Graduate School of Business, Columbia University, 3022 Broadway, New York, NY 10027, USA. Email address: assaf@gsb.columbia.edu
** Postal address: Management Science and Engineering, Terman Engineering Center, Stanford University, Stanford, CA 94305, USA.

is not known *a priori*; thus, we are faced with the task of *estimating buffer overflows* based on the observed traces of $\mathbb{X}$. The problem of estimating tail probabilities is quite important when we consider admission control schemes so as to ensure certain (probabilistic) service-level guarantees. (See e.g. the work of Hsu and Walrand [23] and Courcoubetis *et al.* [10] on dynamic bandwidth allocation in data networks, and the recent paper by Bertsimas and Paschalidis [6] on a similar problem in the context of make-to-stock manufacturing systems.)

It turns out that, under very general conditions on the primitive processes and queueing dynamics in the data network context, a rough exponential-like model for the tail probability can be derived (see e.g. [19] and [14] for single-server stations, and a network extension in [7]). In particular, this tail asymptotic is of the form

$$\log \Pr(X > x) \sim -\theta^* x, \tag{1}$$

where $\log(\cdot)$ denotes the natural logarithm and $f(x) \sim g(x)$ if and only if $f(x)/g(x) \to 1$ as $x \to \infty$. We note that (1), unlike the celebrated Cramér–Lundberg asymptotic (see e.g. [15, Section 1]), captures the behavior of the tail probability only as a first-order term in the exponent, via the *tail-decay parameter* θ^*. In many instances, deriving a more refined characterization is quite complicated or potentially intractable, in particular when we consider as a primitive the complex traffic in modern data networks.

The main goal of this paper is to study the problem of estimating θ^* in (1), based on a sequence of observations $X_1, X_2, \dots, X_n$ from the process $\mathbb{X}$. Note that (1) does not restrict the distribution in any meaningful manner except for the tail decay. In particular, it is not possible to employ simple and efficient parametric estimators if consistency is desired. To that end, extreme-value theory suggests that, under (1), the sample maximum $M_n := \max\{X_1, \dots, X_n\}$ exhibits logarithmic growth in the sample size; in particular, $M_n/\log n \to 1/\theta^*$ in a suitable sense. This, in turn, suggests that an extremal-based estimator,

$$\hat{\theta}_n := \frac{\log n}{\max\{1, M_n\}}, \tag{2}$$

can be used to construct a consistent estimate of θ^*.

The main contributions of this paper are the following.

1. We determine sufficient and (where possible) necessary conditions on the dependence structure of $\mathbb{X}$ under which $\hat{\theta}_n$ converges almost surely to θ^* (Theorem 1). As a corollary, we obtain an almost-sure limit theory for first-passage times of 'high' level sets (Corollary 1). We also show that, if the marginals are 'heavy tailed' in a suitable sense analogous to (1), then a simple variant of the extremal-based estimator (2) can be used to consistently estimate the *polynomial* tail-decay parameter (Theorem 2).

2. Regarding convergence rates, we show that, as expected, the rates of convergence of the extremal-based estimator are at best logarithmic in the sample size. This rate of convergence is shown to be minimax (Theorem 3). In addition, if no rate of convergence is assumed in (1), then there is no rate of convergence for the extremal-based estimator that holds for all distributions with the above tail behavior (Proposition 1).

3. We examine a variant of the extremal-based estimator which involves local averaging. This *moving-average* estimator is shown to be consistent (Proposition 2) and has the potential for certain variance reduction. The associated rates of convergence are still slow (Proposition 3).

4. When the tail behavior is assumed to be of the form $\Pr(X > x) \sim \eta \exp(-\theta^* x)$, we discuss how extremal-based estimators can be used to estimate the pre-exponent η. For a particular dependence structure, we provide necessary and sufficient conditions for consistency of these estimators (Proposition 4).

These results indicate that, if all that we are willing to assume is (1) along with some reasonable degree of mixing, then extremal-based estimators are almost optimal. But, perhaps the more important message, punctuated by the logarithmic minimax rates, is that estimating tail behavior may not be altogether a realistic undertaking in this set up.

In terms of methodology, this paper shares several common themes with two other papers. The first is the work of Hall *et al.* [22] who considered the closely related problem of estimating the abscissa of convergence of the Laplace transform of a distribution function P, based on a sequence of independent and identically distributed (i.i.d.) observations drawn according to P. Specifically, suppose that the Laplace transform of P converges for all $\theta > -\theta^*$ and diverges for all $\theta < -\theta^*$. Hall *et al.* [22] considered the normalized maximum value and related quantities as potential estimators of θ^*. It turns out, however, that the convergence of the Laplace transform is not sufficient to obtain consistency of the estimator (2); see Theorem 1 in [22] and the discussion following it. The idea of using 'extremal-based' estimators was also exploited in the recent work of Berger and Whitt [4] in the context of extrapolating buffer loss probabilities. The theory they developed requires a more refined structure on the tails of the marginals. In particular, Berger and Whitt [4] focused on a more refined (and consequently more restrictive) analysis in which weak convergence to an extremal limit law plays the key role. The recent paper [31] by Paschalidis and Vassilaras considered the problem of estimating buffer losses; however, their approach is based on a specific stochastic structure that involves Markov modulation of the input process which supports the use of parametric estimators. We should also mention that, in a separate paper [20], properties of certain extremal-based plug-in tail probability estimators are investigated in the context of queueing models that have regenerative structure. Finally, this paper ultimately deals with extreme-value theory, general expositions of which can be found, for example, in the books by Leadbetter *et al.* [26] and Resnick [32] and the more recent book by Embrechts *et al.* [15]. In particular, almost-sure limit theory in this context was discussed extensively by Galambos [16], and is also summarized in [15, Section 3.5]. Some applications in the queueing context can be found in the recent paper by Asmussen [1].

The paper is organized as follows. Section 2 gives some necessary background and preliminaries, while Section 3 contains the consistency results for the extremal-based estimator and discusses convergence rates. Section 4 shows that the logarithmic rates of convergence are the best possible in a minimax sense. Section 5 discusses a moving-average variant of the extremal-based estimator, and Section 6 contains some discussion on estimating the pre-exponent. Finally, Section 7 contains some concluding remarks. Proofs of the main results are relegated to Appendix A for continuity of ideas. Auxiliary results and proofs are collected in Appendix B.

2. Preliminaries

Let $\mathbb{X} = (X_n : n \in \mathbb{Z}_+)$ denote a real-valued discrete-time stationary stochastic process which has the following two particular features: it is *weakly dependent*; and the tail of its stationary marginal distribution admits a rough, logarithmic-scale asymptotic such as the tail condition (1) or its Pareto-like analogue $\log \Pr(X > x) \sim -\theta^* \log x$ as $x \to \infty$. To quantify the dependence structure, so-called mixing assumptions are typically introduced. To this end, let $\sigma(X_1, X_2, \dots)$ denote the sigma field generated by the corresponding random variables. Let $\mathcal{B}_1^m = \sigma(X_1, \dots, X_m)$ and $\mathcal{B}_{m+k}^\infty = \sigma(X_{m+k}, X_{m+k+1}, \dots)$. Then the strong mixing (or α-mixing) coefficient (of lag k) is defined as follows:

$$\alpha(k) = \sup_{A \in \mathcal{B}_1^m,\, B \in \mathcal{B}_{m+k}^\infty} |\Pr(A \cap B) - \Pr(A)\Pr(B)|,$$

where Pr is the underlying probability measure. The process $\mathbb{X}$ is then said to be *strong mixing* (or α-mixing) if $\alpha(k) \to 0$ when $k \to \infty$. This form of mixing is the weakest among standard mixing conditions (see [9]), and is exhibited by many commonly used stochastic processes under mild conditions. Examples include stationary ARMA processes with innovations that are absolutely continuous with respect to Lebesgue measure, stationary Markov chains on general state spaces that are Harris recurrent, and certain regenerative processes with finite cycle time moments (see e.g. [28], [2], [18], and the examples in [12, Sections 1.3.2, 2.4]).

A more stringent dependence structure is *uniform mixing* or ϕ-mixing. Let

$$\phi(k) = \sup_{A \in \mathcal{B}_1^m,\, B \in \mathcal{B}_{m+k}^\infty} |\Pr(B \mid A) - \Pr(B)| \tag{3}$$

denote the ϕ-mixing coefficient (of lag k), where the supremum is restricted to all $A \in \mathcal{B}_1^m$ such that $\Pr(A) > 0$. A process $\mathbb{X}$ is then said to be *uniform mixing* if $\phi(k) \to 0$ as $k \to \infty$. It is easily seen that $\alpha(k) \le \phi(k)$. Examples of uniform mixing processes include stationary autoregressive and ARMA processes with bounded spread-out innovations [2], Gaussian processes with spectral densities that are polynomial in $\exp\{\mathrm{i}\lambda\}$, and Doeblin recurrent Markov chains (see [9]). Examples of uniform mixing processes with mixing constants that decay polynomially are given in [25]. For more discussion of various mixing conditions and their relationships the reader is referred to the monograph by Doukhan [12] and the review paper by Bradley [9].

In this paper, we treat processes which exhibit *short-range dependence* and in general do not make any specific structural assumptions with the exception of the asymptotic tail behavior of the marginals. By short-range dependent we mean that $\sum_k \alpha(k), \sum_k \phi(k) < \infty$. Many storage processes exhibit short-range dependence under fairly mild conditions, e.g. a single-server queue fed by a renewal process or a Markov-modulated arrival process (with the underlying Markov chain being finite state and irreducible) gives rise to a queue length process that is short-range dependent. We should be aware, though, that in the domain of communication networks, traffic patterns often exhibit more complicated structures, and the buffer occupancy process is often no longer short-range dependent (see e.g. [3]). For some results in the context of estimating the tail-decay parameter in the case of a queue fed by a long-range-dependent source modeled as fractional Brownian motion, see [34].

The tail asymptotic (1) corresponds to the following class of marginal distributions:

$$\mathcal{F} := \{F : \bar{F}(x) = \mathrm{e}^{-\theta^* x + o(x)},\ \theta^* > 0\},$$

where $\bar{F}(x) := \Pr(X > x)$. Here, and in what follows, we write $f(x) = o(x)$ if $f(x)/x \to 0$ as $x \to \infty$. This condition is refined in various places where more specific structure is needed. We note that distributions in $\mathcal{F}$ are rapidly varying (see [15, Appendix A3]). However, the class $\mathcal{F}$ also contains distributions that are not of the von Mises class or in the domain of attraction of a Gumbel limit law (see [15, pp. 141–143]). Thus, this class of distributions does not coincide with more standard classes that are often used in the context of extreme-value theory.

We should also point out that many of the results we obtain extend with a simple modification to the case where we assume Pareto-like tail decay, i.e. $\log \Pr(X > x) \sim -\theta^* \log x$. More generally, if there exists an increasing function g such that $y^{-1} \log \Pr(g(X) > y) \to -\theta^*$, then $(g(x))^{-1} \log \Pr(X > x) \to -\theta^*$ with $y = g(x)$, so $g(M_n)/\log n \to 1/\theta^*$. We revisit this point later.

3. Strong consistency and ramifications

3.1. Rate of growth of maxima and strong consistency of extremal-based estimators

The first issue we address is whether $\hat{\theta}_n$, the extremal-based estimator given in (2), is a consistent estimator of θ^*. This is a direct consequence of the growth properties of the maximal extreme value in the class of distributions $\mathcal{F}$ with appropriate weak-dependence conditions imposed. The next theorem states that a modest polynomial decay condition is enough to ensure almost-sure convergence of the normalized maxima in the ϕ-mixing context. In contrast, for the strong mixing case, we require exponential decay of the mixing coefficients. Thus, we trade off a weaker measure of dependence with a more stringent assumption on the rate of 'memory decay'.

Theorem 1. *Suppose that $\mathbb{X}$ is a stationary process with marginal distribution $F \in \mathcal{F}$, which is either:*

(i) *uniform mixing with $\phi(k) = O(k^{-1-\varepsilon})$ for some $\varepsilon > 0$; or*

(ii) *strong mixing with $\alpha(k) = O(\mathrm{e}^{-ck})$ for some $c \in (0, \infty)$.*

Then

$$\frac{M_n}{\log n} \to \frac{1}{\theta^*} \quad \text{as } n \to \infty \tag{4}$$

almost surely and in L^p for any $p \in [1, \infty)$.

Here the notation $a_k = O(b_k)$ is used if there exists a $C < \infty$ such that $\limsup_{k\to\infty} a_k/b_k \le C$. As a simple corollary we also obtain almost-sure limits for normalized hitting times.

Corollary 1. *Let $T(b) := \inf\{n \ge 0 : X_n \ge b\}$. Under condition (i) or (ii) of Theorem 1,*

$$\frac{\log T(b)}{b} \to \theta^* \quad \text{as } b \to \infty$$

almost surely.

Regarding the dependence structure we impose in Theorem 1, it is somewhat surprising that the strong mixing condition, requiring exponential memory decay, is necessary and sufficient. We show this via a counterexample. For each $p > 2$ we construct a stationary strong mixing process $\mathbb{X}$, taking values in $\mathbb{R}_+$, with $\sum_k \alpha(k)k^{p-2} < \infty$ and $\sum_k \alpha(k)k^{p-1}$ diverging to infinity. In addition, this process has marginals in $\mathcal{F}$; however, $M_n/\log n \to c \neq 1/\theta^*$.

Example 1. We will construct a classically regenerative process, with regeneration set $\{0\}$, and which is piecewise constant over regenerative cycles. Let $\tilde{T}(k) = \inf\{n > k-1 : X_n = 0\}$ and set $\tilde{T}(0) = 0$, denoting by $\tau_k = \tilde{T}(k) - \tilde{T}(k-1)$ the cycle lengths.

Fix $p > 2$. The explicit construction is as follows. Let Y_1 be a random variable which is exponentially distributed with mean 1 and, conditional on Y_1, set $\tau_1 = \exp(Y_1/p)$, i.e. a point mass at $\exp(y/p)$ conditional on $Y_1 = y$. Let $\tilde{T}(1) = \tilde{T}(0) + \tau_1$. Set $X_0 = 0$, put $(X_n : 1 \le n < \tau)$ equal to Y_1, and set $Y_{\tilde{T}(1)} = 0$. Repeat this construction inductively to generate the remaining cycles, with $\{Y_k\}$ being i.i.d. exponential with mean 1 and $\tilde{T}(k) = \tilde{T}(k-1) + \tau_k$, with $\tau_k = \exp(Y_k/p)$. Clearly the resulting process is regenerative, with regeneration set equal to $\{0\}$. Moreover, $\mathrm{E}\,\tau^q$ is finite for all $q < p$ and diverges for the pth power. Now, since this process is classically regenerative aperiodic with $\mathrm{E}\,\tau_1 < \infty$, it follows that a stationary version of X, say $X^* = (X_n^* : n \ge 0)$, exists, with $X_n^* \overset{\mathrm{D}}{=} X_\infty$, where the distribution of X_∞ is given by the regenerative ratio formula (see [1] for details). Specializing this argument, the tails of X_∞ are found to be

$$\begin{aligned}\Pr(X_\infty \ge x) &:= \frac{1}{\mathrm{E}\,\tau}\int_{y=x}^{\infty} \mathrm{E}\Bigg[\sum_{i=0}^{\tau-1} \mathbf{1}_{\{X_i \ge x\}} \,\Bigg|\, Y = y\Bigg] P_Y(\mathrm{d}y)\\ &= \frac{1}{\mathrm{E}\,\tau}\int_{y=x}^{\infty} \mathrm{E}[\tau \mid Y = y]\mathrm{e}^{-y}\,\mathrm{d}y\\ &= \frac{1}{\mathrm{E}\,\tau}\int_{y=x}^{\infty} \mathrm{e}^{y/p}\mathrm{e}^{-y}\,\mathrm{d}y\\ &= \frac{1}{(1-1/p)\,\mathrm{E}\,\tau}\mathrm{e}^{-(1-1/p)x}.\end{aligned}$$

Thus, $\log \Pr(X_\infty \ge x)/x \to -\theta^* = -(1-1/p)$ as $x \to \infty$. On the other hand, it is evident that

$$\Pr(M_\tau > x) = \mathrm{e}^{-x}$$

with

$$M_\tau := \max\{X_0, X_1, \dots, X_{\tau_1 - 1}\}$$

denoting the maximum of the process over a regenerative cycle. Consequently, for $M_n := \max\{X_1, \dots, X_n\}$, we have

$$\frac{M_n}{\log n} \to 1$$

almost surely as $n \to \infty$. To see why this convergence holds, note that M_n can be roughly expressed as the maximum over all consecutive cycle maxima up to time n. Since X is regenerative, starting from the second cycle, the latter are independent random variables each having the distribution of M_τ, and the above assertion follows from the rate of growth of the maximum of i.i.d. exponential random variables (for a rigorous proof, see e.g. [17] or [20]). Lemma 3 in Appendix B asserts that, for a large class of regenerative processes, polynomial tails on the cycle lengths are essentially equivalent to the process being strong mixing with $\alpha(n)$ decaying polynomially. The constructed process is amenable to Lemma 3 and thus has a polynomial strong mixing rate. By construction, it has marginals in $\mathcal{F}$ with $\theta^* = (1-1/p)$. Finally, the asserted convergence in Theorem 1 fails to hold since $M_n/\log n$ converges to 1 and not to $1/\theta^*$. Note that we can repeat this construction for arbitrarily large values of p, i.e. there exist strong mixing processes that have mixing coefficients decaying as fast as that power for which Theorem 1 fails.

The main results in Theorem 1 and Corollary 1 carry over straightforwardly if we consider the class of distributions with Pareto-like tails.

Theorem 2. *Suppose that, for some $\theta^* > 1$,*

$$\log \Pr(X > x) \sim -\theta^* \log x \quad \text{as } x \to \infty.$$

Then, under condition (i) or (ii) of Theorem 1, we have

$$\frac{\log M_n}{\log n} \to \frac{1}{\theta^*} \quad \text{as } n \to \infty$$

almost surely and in L^p, and

$$\frac{\log T(b)}{\log b} \to \theta^* \quad \text{as } b \to \infty$$

almost surely.

The core of extreme-value theory links tail behavior to growth of extreme values; rates of convergence in (1) imply convergence rates in Theorem 1. To give a simple illustration, suppose that we restrict attention to distributions with $\bar{F}(x) = \exp\{-\theta^* x + o(\log x)\}$. Then, under condition (i) of Theorem 1,

$$\limsup_{n\to\infty} \frac{\log n}{\log\log n} \left| \frac{M_n}{\log n} - \frac{1}{\theta^*} \right| \le 1 \quad \text{a.s.} \tag{5}$$

(The proof of this statement amounts to repeating steps in the proof of Theorem 1 but with the refined tail condition in place.) In light of this, a natural question is whether restricting the class $\mathcal{F}$ by imposing some rate of convergence in (1) is *necessary* in order to get rates of convergence of the extremal-based estimator. To that end, we have the following result.

Proposition 1. *For any sequence of positive real numbers $r_n \uparrow \infty$ there exists an i.i.d. process with marginal $F \in \mathcal{F}$ and corresponding probability measure $\Pr_F\{\cdot\}$ such that, for all $C > 0$,*

$$\lim_{n\to\infty} \Pr_F\left\{ |\hat{\theta}_n - \theta^*| \ge \frac{C}{r_n} \right\} = 1.$$

To recapitulate, in the absence of a rate of convergence in the tail assumption (1), the extremal-based estimator may converge to θ^* at an arbitrarily slow rate. We now turn to several remarks that pertain to the results established in this section.

Remark 1. The tail asymptotic (1) is, in some sense, the 'minimal' amount of structure that supports consistency results such as (4); see [22, Theorem 1] where it is shown, for example, that $\limsup_{x\to\infty}(\log(1 - F(x))/x = -\theta^*$ is not sufficient even for weak convergence of the normalized sample maximum.

Remark 2. Theorem 1 and Corollary 1 can also be viewed as providing general conditions on the dependence structure that ensure that the almost-sure growth rates of maximal values are the same as in the i.i.d. case. Given Example 1, Theorems 1 and 2 are close to providing necessary and sufficient conditions. For further results on almost-sure limit theory for extreme values under various dependence assumptions, see [5], [17], [20], [21], [30], [33] as well as [16, Section 4] and [15, Section 3.5] and the references therein. The so-called D and D' conditions (see [26, Section 3.7], [15, Section 4.4], and [16]) are often used when weak convergence of the centered and normalized maxima to a limit extremal distribution needs to be established.

4. Minimax rates of convergence

We adopt a nonparametric minimax framework, in which the focus is on the worst-case error of an estimator over a class of distributions. We start with some definitions. Let P denote a stationary probability distribution with marginal $F(x) := P(X \le x)$. For some $C > 0$, set

$$\mathcal{F}(C) := \left\{ F : \bar{F}(x) = \mathrm{e}^{-\theta^* x + \psi(x)},\ C^{-1} \le \theta^* \le C,\ \limsup_{x\to\infty} \frac{|\psi(x)|}{\log x} \le C \right\},$$

where $\{\psi(x)\}$ is a family of functions that are bounded on compact sets uniformly over the class $\mathcal{F}(C)$. Note that $\mathcal{F}(C)$ includes the class of scale changes of gamma distributions, with magnitude of scale and shape parameter bounded by C, and obviously $\mathcal{F}(C) \subseteq \mathcal{F}$ which is associated with (1). Finally, let us define the class of admissible probability distributions to be

$$\mathcal{P}(C) := \{P : F \in \mathcal{F}(C) \text{ and either } \alpha(n) \le \exp(-C^{-1}n) \text{ or } \phi(n) \le n^{-1-C^{-1}} \text{ for all } n \ge 1\},$$

where $\alpha(\cdot)$ and $\phi(\cdot)$ are the strong and uniform mixing coefficients corresponding to a probability distribution P. Let $\bar{\theta}_n$ be a measurable function from $\mathbb{R}^n_+$ to $\mathbb{R}_+$ and set $r(\bar{\theta}_n, \theta^*) := \mathrm{E}_P\, |\bar{\theta}_n - \theta^*|^2$, where $\mathrm{E}_P\{\cdot\}$ is expectation with respect to a probability distribution $P \in \mathcal{P}(C)$.

We will measure the worst-case risk over the class $\mathcal{P}(C)$ as follows:

$$\mathcal{R}(\bar{\theta}_n, \mathcal{P}(C)) = \sup_{P\in\mathcal{P}(C)} r(\bar{\theta}_n, \theta^*).$$

Ideally, we would like to assess the *minimax risk*

$$\mathcal{R}^*(n, \mathcal{P}(C)) = \inf_{\bar{\theta}_n} \mathcal{R}(\bar{\theta}_n, \mathcal{P}(C))$$

and construct estimators that achieve this risk, so-called *minimax optimal estimators*. Unfortunately, the evaluation of $\mathcal{R}^*(n, \mathcal{P}(C))$ is usually impossible. Thus, we will focus on establishing lower bounds on this quantity, and subsequently evaluate how 'close' the extremal-based estimators are to achieving these bounds.

If $\gamma_n \uparrow \infty$ is such that

$$\liminf_{n\to\infty} (\gamma_n)^2 \mathcal{R}^*(n, \mathcal{P}(C)) > C_{\mathrm{l}}$$

for some positive constant C_{l}, then we say that $1/\gamma_n$ is the *lower rate of convergence*. If we can establish that, for some $\hat{\theta}^*_n$, there exists a $C_{\mathrm{u}} < \infty$ such that

$$\limsup_{n\to\infty} (\gamma_n)^2 \mathcal{R}(\hat{\theta}^*_n, \mathcal{P}(C)) \le C_{\mathrm{u}},$$

then we say that $\hat{\theta}^*_n$ is *asymptotically minimax optimal*. The following theorem establishes that the extremal-based estimator $\hat{\theta}_n$ is asymptotically nearly minimax optimal (i.e. the upper and lower rates of convergence differ only by a lower order factor that is logarithmic in this rate). Note that the lower bound is essentially an immediate consequence of the results in [22].

Theorem 3. *There exist constants* $C_{\mathrm{l}}, C_{\mathrm{u}} \in (0, \infty)$ *such that*

$$\liminf_{n\to\infty} (\log n)^2 \mathcal{R}^*(n, \mathcal{P}(C)) \ge C_{\mathrm{l}},$$

$$\limsup_{n\to\infty} \frac{(\log n)^2}{(\log\log n)^2} \mathcal{R}(\hat{\theta}_n, \mathcal{P}(C)) \le C_{\mathrm{u}}.$$

On a final note, if we consider the class of distributions $\tilde{\mathcal{P}}(C)$ indexed by

$$\tilde{\mathcal{F}}(C) := \Big\{ F \in \mathcal{F} : \bar{F}(x) = \mathrm{e}^{-\theta^* x + \psi(x)},\ C^{-1} \le \theta^* \le C,\ \limsup_{x\to\infty} |\psi(x)| \le C \Big\}$$

instead of $\mathcal{F}(C)$, then it is not difficult to verify that

$$\lim_{K\uparrow\infty} \limsup_{n\to\infty} \sup_{P\in\tilde{\mathcal{F}}} P\bigg\{ |\hat{\theta}_n - \theta^*| > \frac{K}{\log n} \bigg\} = 0.$$

Thus, for this class of distributions with further restrictions on the marginals, $\hat{\theta}_n$ is minimax optimal in probability over $\tilde{\mathcal{P}}(C)$. For recent work on minimax bounds in estimating the extreme-value index under zero–one loss, see [13].

5. A moving-average extremal-based estimator

In this section we introduce and study some properties of an estimator of the tail parameter based on a moving average (MA) of block-based estimators. To be specific, fix a sequence of increasing positive integers a_n, and let $m_n = \lfloor n/a_n \rfloor$. Let

$$M_{a_n}(i) := \max\{X_j : j = ia_n + 1, \dots, (i+1)a_n\} \quad \text{for } i = 0, \dots, m_n$$

and define

$$\widehat{\bigg(\frac{1}{\theta}\bigg)}_n := \frac{1}{m_n} \sum_{i=0}^{m_n - 1} \frac{M_{a_n}(i)}{\log a_n}.$$

As we shall see in what follows, this estimator has essentially the same consistency properties of the normalized global maximum. However, on a somewhat more heuristic level, the MA estimator has the important property that it is not as biased by initial large observations as the global-max estimator is. Another potential advantage of the MA estimator is that it is less sensitive to the stationarity assumption which we invoke. Moreover, if we focus on the mean squared error, then for the global-max estimator we have

$$\mathrm{E}\bigg[\frac{M_n}{\log n} - \frac{1}{\theta^*}\bigg]^2 = \bigg(\frac{1}{\theta^*} - \mathrm{E}\,\frac{M_n}{\log n}\bigg)^2 + \operatorname{var}\frac{M_n}{\log n},$$

where, as for the MA estimator,

$$\mathrm{E}\bigg[\widehat{\bigg(\frac{1}{\theta}\bigg)}_n - \frac{1}{\theta^*}\bigg]^2 = \bigg(\frac{1}{\theta^*} - \mathrm{E}\,\frac{M_{a_n}}{\log a_n}\bigg)^2 + \operatorname{var}\widehat{\bigg(\frac{1}{\theta}\bigg)}_n$$

using the standard bias-variance decomposition. Then, from the analysis of Section 4 we have that the bias term is $O(\log\log n/\log n)$ and $O(\log\log a_n/\log a_n)$ for the global-max and MA estimators respectively, for marginals that have gamma-like tails. Thus, if we take a block size that is $a_n = n^\gamma$ for some $\gamma \in (0, 1)$, the bias term in both estimators is asymptotically of the same order. Now, if the observed process $\mathbb{X}$ is i.i.d., then clearly the variance term of the MA estimator is $m_n^{-1} \operatorname{var}(M_{a_n}/\log a_n)$ and, since $a_n = n^\gamma$, roughly $\sigma^2_{\mathrm{MA}} \approx n^{-(1-\gamma)}\sigma_n^2$, where σ_n^2 corresponds to the variance of the normalized global-max estimator. In a more realistic scenario, suppose that $\mathbb{X}$ is mixing concurring with the restrictions in Theorem 1. To fix ideas, say that it is strongly mixing with $\alpha(k) = O(\exp\{-ck\})$ for some $c < \infty$. Then, using the standard covariance inequalities (see [12, Section 1.2.2]), we have

$$\left|\operatorname{cov}\left(\frac{M_{a_n}(i)}{\log a_n}, \frac{M_{a_n}(j)}{\log a_n}\right)\right| \le 8\alpha^{1/r}(|i-j|a_n)\left(\mathrm{E}\left[\frac{M_{a_n}(i)}{\log a_n}\right]^p\right)^{1/p}\left(\mathrm{E}\left[\frac{M_{a_n}(j)}{\log a_n}\right]^q\right)^{1/q}$$

with $p, q, r \ge 1$ such that $1/p + 1/q + 1/r = 1$. Since the normalized maximum converges also in L^p for any p, and since $\{\alpha^{1/r}(k)\}$ is summable, we have

$$\begin{aligned}
\operatorname{var}\left(\frac{\hat{1}}{\theta}\right)_n &= \frac{1}{m_n^2}\sum_{i=0}^{m_n-1}\sum_{j=0}^{m_n-1}\operatorname{cov}\left(\frac{M_{a_n}(i)}{\log a_n}, \frac{M_{a_n}(j)}{\log a_n}\right) \\
&= \frac{1}{m_n}\operatorname{var}\left(\frac{M_{a_n}}{\log a_n}\right) + \frac{1}{m_n}\sum_{i=1}^{m_n-1}\operatorname{cov}\left(\frac{M_{a_n}(0)}{\log a_n}, \frac{M_{a_n}(j)}{\log a_n}\right) \\
&\le \frac{1}{m_n}\sigma_n^2 + \frac{C}{m_n}.
\end{aligned}$$

Thus, the conclusions of the i.i.d. analysis are still valid in this set up. A similar derivation holds in the case of uniform mixing. Our first theorem gives conditions that ensure the strong consistency of the MA estimator.

Proposition 2. *Let $\mathbb{X}$ be a stationary process which satisfies either condition (i) or (ii) of Theorem 1. Then*

$$\left(\frac{\hat{1}}{\theta}\right)_n \to \frac{1}{\theta^*} \quad \text{as } n \to \infty$$

almost surely and in L^1.

The next theorem establishes a central limit theorem for the MA estimator under an i.i.d. assumption for the process $\mathbb{X}$. (We note that this assumption is put in place to avoid technicalities in the proof; for an extension to certain Markov processes see the technical report version of this paper [35].)

Proposition 3. *Suppose that $\mathbb{X}$ is an i.i.d. process with marginals in the set $\{F : \bar{F}(x) = \mathrm{e}^{-\theta^* x + O(1)},\ \theta^* > 0\}$. Then*

$$Z_n := \frac{\sqrt{m_n}}{\sigma_n}\left[\left(\frac{\hat{1}}{\theta}\right)_n - \mathrm{E}\,\frac{M_{a_n}(0)}{\log a_n}\right] \Rightarrow \mathcal{N}(0, 1),$$

where a_n, m_n are the two sequences defining the MA estimator, chosen so that

(i) $a_n, m_n \uparrow \infty$,

(ii) $a_n m_n \sim n$, *and*

(iii) $m_n/(\log a_n)^4 \to \infty$,

and

$$\sigma_n^2 := \operatorname{var}\frac{M_{a_n}(0)}{\log a_n}.$$

We point out that Proposition 3 should be viewed in some sense as a negative result. Roughly speaking, it asserts that

$$\left(\frac{\hat{1}}{\theta}\right)_n - \mathrm{E}\,\frac{M_{a_n}(0)}{\log a_n} \approx \frac{\sigma_n}{\sqrt{m_n}}\,\mathcal{N}(0, 1). \tag{6}$$

However, standard rates of convergence in extreme-value theory under the tail condition we impose, together with the uniform integrability results in Lemma 5, indicate that

$$\mathrm{E}\,\frac{M_{a_n}(0)}{\log a_n} - \frac{1}{\theta^*} \approx \frac{1}{\log a_n}. \tag{7}$$

If we view (6) as characterizing the 'stochastic error', and (7) as characterizing the 'deterministic error', then it is clear that the latter dominates for the feasible choices of a_n, m_n. The central limit theorem is therefore not useful in characterizing the fluctuations of the MA estimator around the tail parameter $1/\theta^*$.

6. Estimating the pre-exponent

In this section we impose a more stringent condition on the tail behavior, which in turn allows us to tackle the problem of estimating the pre-exponent. To fix ideas, we restrict the analysis here to a particular example which can be easily motivated. Consider a system in which random purchase requests $(V_n : n \geq 1)$ arrive according to a discrete-time renewal process with i.i.d. interarrival times $(U_n : n \geq 1)$. These sequences are independent of each other. The service facility answers the demand requests at a constant (unit) rate whenever purchase orders are present. Let $Z_n = V_n - U_n$, and assume that $\mathrm{E}\,Z_i < 0$ corresponding to the traffic intensity $\rho := \mathrm{E}\,V/\mathrm{E}\,U < 1$. Assume further that the Z_n are nonlattice random variables and let $\varphi(\theta) = \mathrm{E}\exp(\theta Z_i)$. Suppose that there exists a positive root θ^* of the equation $\varphi(\theta) = 1$ such that $\varphi(\theta)$ converges in a neighborhood of θ^*. Let $\mathbb{X} = (X_n : n \geq 0)$ be defined via the Lindley recursion $X_{n+1} = \max\{X_n + Z_{n+1}, 0\}$. That is, X_n measures the delay incurred by the nth request. It can be easily shown that, under the above conditions, there exists a stationary version of the delay sequence, which, with some abuse of notation, we continue to denote by $\mathbb{X}$. The Cramér–Lundberg approximation states that, for this stationary process,

$$\Pr(X > x) \sim \eta \mathrm{e}^{-\theta^* x} \quad \text{as } x \to \infty.$$

We refer to η as the *pre-exponent* and focus our analysis on estimating η. The Cramér–Lundberg approximation is known to hold in several queueing models (see [4] and the references therein), and is also quite common in insurance models and risk theory (see [15] for details and further references).

Let $\kappa_n \uparrow \infty$ be a sequence of positive real numbers and define

$$\hat{p}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i > x\}},$$

$$\hat{\theta}_n := \frac{\log n}{\max\{M_n, 1\}},$$

and

$$\hat{\eta}_n := \hat{p}_n(\kappa_n)\mathrm{e}^{\hat{\theta}_n \kappa_n}.$$

Our main result gives a precise characterization of consistency for $\hat{\eta}_n$.

Proposition 4. *Let the process $\mathbb{X}$ be a stationary version of the delay process. Then*

(i) *if* $\kappa_n = o(\log n/\log\log n)$,

$$\frac{\hat{\eta}_n}{\eta} \to 1 \quad \text{as } n \to \infty$$

almost surely;

(ii) *if* $\kappa_n = o(\log n)$,

$$\frac{\hat{\eta}_n}{\eta} \Rightarrow 1 \quad \text{as } n \to \infty;$$

(iii) *if* $\kappa_n = c \log n$ *for* $c \in (0, 1/2\theta^*)$,

$$\frac{\hat{\eta}_n}{\eta} \Rightarrow \zeta,$$

where $\zeta \stackrel{\mathrm{D}}{=} \exp\{c\log(\phi\eta) + c\theta^* Z\}$ *with Z having the normalized Gumbel (or type I) extreme-value distribution, and* $\phi \in (0, 1)$ *is the so-called extremal index of* $\mathbb{X}$.

For weak convergence of the centered and normalized maximal value in this context, see [24], and for point process weak limits see [33].

Remark 3. Note that the estimator $\hat{\eta}_n$ utilizes the extremal-based estimator of θ^*. As discussed previously, this estimator has slow (logarithmic) convergence rate. In the particular context we are considering here, the process $\mathbb{X}$ is essentially a reflected random walk. This allows for estimating θ^* with much faster (parametric) rates as we sketch in the following arguments. Let $R(\theta) := \mathrm{E}\exp\{\theta Z\}$, $\psi(\theta) = \log R(\theta)$ and set $R_n(\theta) := n^{-1}\sum_{i=1}^n \exp\{\theta Z_i\}$. Now, θ^* is the unique positive root of $\psi(\theta)$, and let us assume that $\psi'(\theta) > 0$ in a small neighborhood around θ^*. Set $\tilde{\theta}_n$ to be a positive root of the equation $R_n(\theta) = 1$. Then, using the mean value theorem we can write

$$R_n(\theta) = R_n(\tilde{\theta}_n) + (\theta - \tilde{\theta}_n) R_n'(\bar{\theta}_n)$$

with $\bar{\theta}_n$ a point on the line segment between θ and $\tilde{\theta}_n$. Taking $\theta := \theta^*$ and rearranging, we have

$$\tilde{\theta}_n - \theta^* = \frac{n^{-1}\sum_{i=1}^n(\exp\{\theta^* Z_i\} - 1)}{n^{-1}\sum_{i=1}^n Z_i \exp\{\bar{\theta}_n Z_i\}}.$$

Now, $\sum_{i=1}^n Z_i \exp\{\theta Z_i\} \to R'(\theta) = \mathrm{E}[Z\exp\{\theta Z\}]$ almost surely and uniformly on any interval containing θ^* such that $\mathrm{E}[Z\exp\{\theta Z\}]$ is finite over that interval. The continuity of $R'(\theta)$ together with the above establishes that $\sum_{i=1}^n Z_i \exp\{\bar{\theta}_n Z_i\} \to R'(\theta^*)$ and, consequently, $\sqrt{n}(\tilde{\theta}_n - \theta^*) \Rightarrow \sigma\mathcal{N}(0, 1)$. This derivation is only made possible given an i.i.d. structure, while the extremal-based estimator applies under more general dependence assumptions.

7. Concluding remarks

In many practical situations, the tail behavior of the marginal distribution admits only a rough characterization, for example logarithmic asymptotics. Consequently, the use of parametric estimators for estimating parameters governing the tail behavior is not appropriate, and so semiparametric and nonparametric estimators are called for. The extremal-based estimators studied here fall exactly in that category.

These estimators have several potential advantages. In particular, they are: (i) consistent in an almost-sure sense; (ii) nearly optimal in a minimax sense; and (iii) we can employ moving-average variants which are more suitable for applications that involve transients. In addition, it is possible to show that, in the context of estimating tail probabilities, extremal-based estimators are superior to simple nonparametric counterparts in a well-defined sense; see e.g. [20]. An obvious drawback that these estimators suffer from are the slow (logarithmic) rate of convergence characteristic of extreme values. However, as opposed to certain nonparametric variants, one can extrapolate rare-event probabilities (such as buffer overflows) beyond the given sample without actually observing the rare events in question. (For more on this point see the discussion in [20].)

Appendix A. Proofs of the main results

A.1. Proof of Theorem 1

The upper bound follows straightforwardly from Lemma 1 in Appendix B, which asserts that

$$\limsup_{n\to\infty} \frac{M_n}{\log n} \le \frac{1}{\theta^*} \quad \text{a.s.}$$

To prove the lower bound, consider first a set up with condition (i) of the theorem invoked.

Step 1. The first step consists of reducing the problem to dealing with an i.i.d. sequence. Fix $\delta \in (0, \varepsilon/6)$, with ε as in the definition of the ϕ-mixing sequence. We now proceed by 'chopping up' the sequence $(X_1, \dots, X_n)$ into blocks of length $a_n = n^{1-2\delta}$, altogether $2m_n = \lfloor n^{2\delta} \rfloor$ blocks and a remainder of length $r_n = ca_n$ with $c \in [0, 1)$. Let

$$Y_i := \bigvee_{j=2(i-1)a_n+1}^{(2i-1)a_n} X_i,$$

where $\bigvee_{i=1}^{n} X_i := \max\{X_1, \dots, X_n\}$. Thus, $\{Y_i\}_{i=1}^{m_n}$ is the sequence of block maxima over odd-numbered blocks. Let $y_n = [(1-\delta)\log n]/\theta^*$. Then

$$\begin{aligned} \Pr(M_n \le y_n) &\le \Pr\left(\bigvee_{i=1}^{m_n} Y_i \le y_n\right) \\ &\le [\Pr(Y_i \le y_n)]^{m_n} + m_n\phi(a_n) \\ &\le \mathrm{e}^{-m_n \Pr(Y > y_n)} + m_n\phi(a_n), \end{aligned} \tag{8}$$

where the first equality follows since, obviously, $M_n \le \bigvee_{i=1}^{m_n} Y_i$ almost surely, and the second inequality follows by the mixing assumption and the definition of the mixing coefficients. Thus, it suffices to show that both terms on the right-hand side of (8) are summable.

Step 2. Controlling the tail behavior of the marginal of Y. Set $\tau_n := m_n \Pr(Y > y_n)$. By the assumed rate of decay of the mixing coefficients, there must exist a natural number p such that $\sum_k \phi(pk) \le \frac{1}{4}$, say.

The key is to replace $Y_0 = \bigvee_{i=1}^{a_n} X_i$ with a p-spaced maximum,

$$\tilde{Y}_0 = \bigvee_{i=1}^{\lfloor a_n/p \rfloor} X_{ip+1}$$

and define $\tilde{Y}_i$ for $i = 1, 2, \dots, m_n$ in the obvious way. It now follows that

$$\Pr(Y > y_n) \ge \Pr(\tilde{Y} > y_n) \ge I_n - J_n,$$

where

$$I_n := \sum_{i=1}^{\lfloor a_n/p \rfloor} \Pr(X_{ip+1} > y_n), \qquad J_n := \sum_{i=1}^{\lfloor a_n/p \rfloor} \sum_{j=i+1}^{\lfloor a_n/p \rfloor} \Pr(X_{ip+1} > y_n, X_{jp+1} > y_n).$$

Now,

$$J_n = \left\lfloor \frac{a_n}{p} \right\rfloor \sum_{j=1}^{\lfloor a_n/p \rfloor - 1} \Pr(X_1 > y_n, X_{jp+1} > y_n)$$

and it follows that

$$\begin{aligned}
\tau_n &:= m_n \Pr(Y > y_n) \\
&\geq m_n \left\lfloor \frac{a_n}{p} \right\rfloor \Pr(X > y_n) - \left\lfloor \frac{a_n}{p} \right\rfloor \sum_{j=1}^{\lfloor a_n/p \rfloor - 1} \Pr(X_1 > y_n, X_{jp+1} > y_n) \\
&= I_n^{(1)}(1 - I_n^{(2)}),
\end{aligned}$$

where

$$I_n^{(1)} := m_n \left\lfloor \frac{a_n}{p} \right\rfloor \Pr(X > y_n), \qquad I_n^{(2)} := \sum_{j=1}^{\lfloor a_n/p \rfloor - 1} \frac{\Pr(X_1 > y_n, X_{jp+1} > y_n)}{\Pr(X > y_n)}.$$

Therefore we need: (a) $I_n^{(1)} \to \infty$ such that $\sum_n \exp\{-I_n^{(1)}\} < \infty$, and (b) $\limsup_n I_n^{(2)} \leq \frac{1}{2}$, say.

Step 3. We verify properties (a) and (b) above. First, by the choice of y_n,

$$\Pr(X > y_n) = \mathrm{e}^{-\theta^* y_n + \psi(y_n)} \geq \frac{c}{n^{1-\delta}}$$

for some constant $c > 0$ and for all but finitely many n. Now, by construction, $m_n a_n \geq n/4$ for sufficiently large n. Thus,

$$I_n^{(1)} \geq \frac{n^\delta}{4}$$

for all but finitely many n. Consequently, $\sum_n \exp\{-I_n^{(1)}\} < \infty$. To verify (b),

$$\begin{aligned}
I_n^{(2)} &= \sum_{j=1}^{\lfloor a_n/p \rfloor - 1} \frac{\Pr(X_1 > y_n, X_{jp+1} > y_n)}{\Pr(X > y_n)} \\
&\leq \sum_{j=1}^{\lfloor a_n/p \rfloor} \Pr(X > y_n) + \sum_{j=1}^{\lfloor a_n/p \rfloor - 1} \phi(jp),
\end{aligned}$$

where the inequality follows from the definition of ϕ-mixing (3). Now,

$$\begin{aligned}
a_n \Pr(X > y_n) &= n^{-(1-2\delta)} \mathrm{e}^{-(1-\delta)\log n + \psi(\log n)} \\
&\downarrow 0 \quad \text{as } n \to \infty
\end{aligned}$$

by the assumption on $\psi(x)$ and the choice of a_n. For the mixing term, we have

$$\sum_{j=1}^{\lfloor a_n/p\rfloor-1} \phi(jp) \le \frac{1}{4}.$$

Thus, $I_n^{(2)} \le \frac{1}{2}$ eventually, which implies that $1 - I_n^{(2)} \ge \frac{1}{2}$ for all but finitely many n. Combining these steps we have established that

$$\tau_n \ge \tfrac{1}{8}n^{\delta}$$

for all but finitely many n; thus $\sum_n \mathrm{e}^{-\tau_n} < \infty$.

Step 4. The summability of the mixing term in (8) follows from the choice of $\delta \in (0, \varepsilon/6)$, so that there exists some $\varepsilon' > 0$ for which $m_n\phi(a_n) \le c/n^{1+\varepsilon'}$ for all but finitely many n. Consequently, $\sum_n m_n\phi(a_n) < \infty$. This concludes the proof under condition (i) as we have the bound in (8) summable. Thus, by the Borel–Cantelli lemma,

$$\liminf_{n\to\infty} \frac{M_n}{\log n} \ge \frac{1-\delta}{\theta^*}$$

and, since δ is arbitrary, the result follows.

We now prove the result in the theorem when condition (ii) is invoked. The first thing is to consider a sequence $\{Y_i\}_{i=1}^{m_n}$ of random variables which are obtained by equally spaced sampling from the original X sequence. That is, $Y_1 = X_1$, $Y_2 = X_{1+a_n}, \ldots, Y_{m_n} = X_{1+m_n a_n}$, with a_n and $m_n = \lfloor n/a_n \rfloor$ two sequences of increasing positive real numbers which will be specified in what follows. To this extent, the equivalent of (8) is now

$$\Pr(M_n \le y_n) \le \mathrm{e}^{-\tau_n} + m_n\alpha(a_n)$$

with $\tau_n := m_n \Pr(Y > y_n)$. Fix $\varepsilon > 0$ and this time let $a_n = c_1 \log n$ with $c_1(\varepsilon)$ a constant chosen so that $\alpha(c_1 \log n) \le n^{-(2+\varepsilon)}$. Then, clearly

$$m_n\alpha(a_n) \le n^{-(1+\varepsilon)},$$

which is summable. Also,

$$\tau_n \ge \frac{n^{\delta}}{a_n},$$

which, by the choice of a_n, implies that $\sum_n \mathrm{e}^{-\tau_n} < \infty$. The proof is complete by appealing to Lemma 4 in Appendix B, which establishes the uniform integrability necessary for the L^p convergence.

A.2. Proof of Corollary 1

The proof follows from the relation $\{M_n \ge b\} = \{T(b) \le n\}$ and by taking a sequence

$$n_b := \left\lfloor \exp\left\{b\frac{\theta^*}{1+\delta}\right\}\right\rfloor \tag{9}$$

so that $n_b \to \infty$ as $b \uparrow \infty$. Then, by Theorem 1,

$$\Pr\left(M_n > \frac{1+\delta}{\theta^*}\log n\right) \to 0 \tag{10}$$

and the convergence holds also along the subsequence n_b. In particular, substituting (9) into (10), we have

$$\Pr\left(\frac{\log T(b)}{\theta^* b} \leq 1 - \delta'\right) \to 0$$

with $\delta' := \delta/(1+\delta)$. The upper bound follows similarly.

A.3. Proof of Theorem 2

The proof is a straightforward consequence of the proof of Theorem 1. Let $Y_i := \log(X_i)_+$. Then it is clear that $\mathbb{Y} = (Y_n : n \in \mathbb{Z}_+)$ is a stationary sequence satisfying the same mixing conditions as in Theorem 1. In addition, the tail conditions on the marginals of $\mathbb{X}$ translate into

$$\Pr(Y > x) \sim -\theta^* \log x,$$

which is exactly the tail condition (1) assumed in Theorem 1. Thus, its conclusions apply to the process $\mathbb{Y}$, proving Theorem 2.

A.4. Proof of Proposition 1

Since the fastest rate of convergence possible for the extremal-based estimator is $1/\log n$, we restrict attention to sequences that exhibit logarithmic or slower growth at infinity. Let $r(n) \uparrow \infty$ be a sequence of positive real numbers such that $\limsup_n r(n)/\log n < \infty$. Let $F \in \mathcal{F}$ be such that $\psi(x) = x/r(\lfloor x \rfloor)$, and consider a sequence of i.i.d. random variables with marginal F. Fix $C > 0$. Then

$$\begin{aligned}\Pr_F\{|\hat{\theta}_n - \theta^*| \geq C/r(n)\} &\geq \Pr_F\left\{\frac{\log n}{M_n} \leq \theta^* - \frac{C}{r(n)}\right\} = \Pr_F\left\{M_n \geq \frac{\log n}{\theta^* - C/r(n)}\right\} \\ &\geq 1 - \mathrm{e}^{-n\bar{F}(u_n)}\end{aligned}$$

with $u_n := (\log n)/(\theta^* - C/r(n))$. We now show that for the choice of $\psi(x)$ we have $n\bar{F}(u_n) \geq \log n/\log\log n$ for all but finitely many n. By the choice of $\psi(x)$ and the definition of the class $\mathcal{F}$, we have

$$\begin{aligned}\log n + \log \bar{F}(u_n) &= \psi(u_n) - \frac{C/r(n)}{\theta^* - C/r(n)} \log n \\ &= \frac{\log n}{\theta^* - C/r(n)}\left(\frac{1}{r(u_n)} - \frac{C}{r(n)}\right) \\ &\geq c \frac{\log n}{\log\log n}\end{aligned}$$

for all but finitely many n, where the last step follows from the monotonicity of $r(\cdot)$ and since $u_n \sim a \log n$. Thus, $n\bar{F}(u_n) \to \infty$, which concludes the proof.

A.5. Proof of Theorem 3

We first prove the lower bound which follows closely [22]. Hall *et al.* [22] considered the problem of discriminating between densities based on an i.i.d. sample drawn according to one of the following:

$$\begin{aligned} f_1(x) &= \theta \mathrm{e}^{-\theta x}, \\ f_2(x) &= \begin{cases} \theta \mathrm{e}^{-\theta x} & \text{for } 0 < x \leq x_0, \\ \theta \mathrm{e}^{C_1}(1 + C_1\varepsilon)\mathrm{e}^{-\theta(1+C_1\varepsilon)x} & \text{for } x > x_0, \end{cases}\end{aligned}$$

with C_1 a properly chosen constant, $\varepsilon \sim 1/\log n$ and $x_0 \sim \log n$. Write θ_1 and θ_2 for the values assumed by θ^* when F is the distribution function associated with f_1 and f_2. Then Hall *et al.* showed that

$$\liminf_{n\to\infty} \max_{j=1,2} P_j(|\hat{\theta}_n - \theta_j| > \tfrac{1}{2}|\theta_1 - \theta_2|) \geq 1 - \Phi(|C_1|/2)$$

for any estimator (i.e. measurable function $\hat{\theta}_n : \mathbb{R}_+^n \to \mathbb{R}_+$), with Φ the standard normal distribution function. It is not difficult to see that, for given C defining the class $\mathcal{F}(C)$, we can choose θ^* and C_1 such that, for n sufficiently large, the densities f_1 and f_2 have associated distribution functions in $\mathcal{F}(C)$, and thus $P_1, P_2 \in \mathcal{P}(C)$. Then, since $\theta_1 = \theta$ and $\theta_2 = \theta(1 + C_1\varepsilon)$, there exist constants $c_1, c_2 > 0$ such that

$$\liminf_{n\to\infty} \sup_{P\in\mathcal{P}(C)} P\left(|\hat{\theta}_n - \theta^*| > \frac{c_2}{\log n}\right) \geq c_1.$$

Consequently, using the Markov inequality we have

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{P\in\mathcal{P}(C)} (\log n)^2 \operatorname{E}_P |\hat{\theta}_n - \theta^*|^2 \geq c_1 c_2 > 0.$$

In particular, there exists some $C_1 > 0$ such that

$$\liminf_{n\to\infty} (\log n)^2 \mathcal{R}^*(n, \mathcal{P}(C)) > C_1,$$

which establishes the lower bound.

We divide the proof of the upper bound into steps.

Step 1. To simplify notation we write $\tilde{M}_n = M_n \vee 1$. Then

$$\operatorname{E} \frac{(\log n)^2}{(\log\log n)^2} |\hat{\theta}_n - \theta^*|^2 = \operatorname{E}\left[\left|\frac{\log n - \theta^* \tilde{M}_n}{\tilde{M}_n}\right|^2 \frac{(\log n)^2}{(\log\log n)^2}\right] \leq I_n J_n,$$

where

$$I_n = \sqrt{\operatorname{E}\left(\frac{\log n}{\tilde{M}_n}\right)^4}, \qquad J_n = \sqrt{\operatorname{E} \frac{(\log n - \theta^* \tilde{M}_n)^4}{(\log\log n)^4}}.$$

Using Lemma 5 in Appendix B, there exists a $C < \infty$ such that

$$I_n \leq \sup_n \operatorname{E}\left(\frac{\log n}{\tilde{M}_n}\right)^4 \leq C.$$

Step 2. To bound J_n, it suffices to show that

$$\sup_n \sup_{\mathcal{F}(C)} \sum_{k=1}^{\infty} \Pr\left(\left|\frac{\log n - \theta^* \tilde{M}_n}{\log\log n}\right|^4 > k\right) < \infty.$$

Start with

$$\begin{aligned}
&\sum_{k=1}^{\infty} \Pr\left(\tilde{M}_n \geq \frac{1}{\theta^*}\log n + \frac{1}{\theta^*}k^{1/4}\log\log n\right) \\
&\quad \leq C^4 + \sum_{k=C^4}^{\infty} n\exp(-\log n - k^{1/4}\log\log n + \psi(\log n)) \\
&\quad = C^4 + \sum_{k=C^4}^{\infty} \exp\left(-\log\log n\left[k^{1/4} - \frac{\psi(\log n)}{\log\log n}\right]\right) \\
&\quad \leq C^4 + \sum_{k=C^4}^{\infty} \exp(-\log\log n(k^{1/4} - C)) \\
&\quad \leq C^4 + C_1 \sum_{k=C^4}^{\infty} \mathrm{e}^{-k^{1/4}},
\end{aligned}$$

where the last inequality holds for all but finitely many n. Thus, since the family of functions $\{\psi\}$ are bounded on compact uniformly over $\mathcal{F}$,

$$\sup_n \sup_{\mathcal{F}(C)} \sum_{k=1}^{\infty} \Pr\left(\tilde{M}_n \geq \frac{1}{\theta^*}\log n + \frac{1}{\theta^*}k^{1/4}\log\log n\right) < \infty.$$

Note that we made no use of the dependence structure in this bound.

Now, for the other side we have

$$\begin{aligned}
&\sum_{k=1}^{\infty} \Pr\left(M_n \leq \frac{1}{\theta^*}\log n - \frac{1}{\theta^*}k^{1/4}\log\log n\right) \\
&\quad = \sum_{k=1}^{K_n} \Pr\left(M_n \leq \frac{1}{\theta^*}\log n - \frac{1}{\theta^*}k^{1/4}\log\log n\right) \\
&\quad \leq \sum_{k=1}^{\infty} \exp\left(-cn\Pr\left(X > \frac{1}{\theta^*}\log n - \frac{1}{\theta^*}k^{1/4}\log\log n\right)\right) + K_n r_n \\
&\quad \leq \sum_{k=1}^{\infty} \exp(-(\log n)^{k^{1/4}}) + C \\
&\quad \leq C' \sum_{k=1}^{\infty} \mathrm{e}^{-c'k^{1/4}},
\end{aligned}$$

for all but finitely many n, where $K_n = (\log n/\log\,\log n)^4$ in the first equality since, for $k > K_n$,

$$\frac{1}{\theta^*}\log n - \frac{1}{\theta^*}k^{1/4}\log\log n < 0,$$

and r_n is either $m_n\phi(a_n)$ or $m_n\alpha(a_n)$. In addition, we used the fact that, for sufficiently large n,

$$n\Pr\left(X > \frac{1}{\theta^*}\log n - \frac{1}{\theta^*}k^{1/4}\log\log n\right) \geq (\log n)^{k^{1/4}}.$$

By the definition of K_n and the proof of Theorem 1, $K_n r_n = o(1)$. Thus,

$$\sup_n \sup_{\mathcal{P}(C)} \sum_{k=1}^{\infty} \Pr\left(M_n \le \frac{1}{\theta^*} \log n - \frac{1}{\theta^*} k^{1/4} \log\log n \right) < \infty.$$

Combining the two bounds in Steps 1 and 2 we have established the result.

A.6. Proof of Proposition 2

The proof follows straightforwardly from the results in Theorem 1. For a set $\omega \in \Omega' \subseteq \Omega$ with $\Pr(\Omega') = 1$,

$$\frac{M_n(\omega)}{\log n} \to \frac{1}{\theta^*}.$$

Thus, since $a_n \uparrow \infty$, the same holds for $M_{a_n}/\log a_n$, and by the Césaro sum property the result follows for each $\omega \in \Omega'$. The L^1 convergence follows immediately.

A.7. Proof of Proposition 3

The proof will be based on the Lindberg–Feller central limit theorem (CLT) for triangular arrays. First, we can write

$$Z_n = \frac{1}{\sqrt{m_n}} \frac{\sum_{i=0}^{m_n-1} (M_{a_n}(i) - \mathrm{E}\, M_{a_n}(0))}{\sqrt{\mathrm{var}\, M_{a_n}(0)}}.$$

Let $Y_n(i) := M_{a_n}(i) - \mathrm{E}\, M_{a_n}(0)$. Then clearly $\mathrm{E}\, Y_n(i) = 0$ and $\{Y_n(i)\}_{i=0}^{m_n-1}$ is a sequence of independent random variables for each n. Let

$$\begin{aligned} s_n^2 &:= \sum_{i=0}^{m_n-1} \mathrm{E}\, Y_n^2(i) = m_n \,\mathrm{var}\, M_{a_n}(0), \\ S_n &:= \sum_{i=0}^{m_n-1} Y_n(i), \\ R_n &= \frac{1}{s_n^2} \sum_{i=0}^{m_n-1} \mathrm{E}[Y_n^2(i); |Y_n(i)| > \varepsilon s_n]. \end{aligned}$$

Then, according to the Lindberg–Feller CLT for triangular arrays (see [8, Theorem 27.2]), if $R_n \to 0$ for all $\varepsilon > 0$, then

$$\frac{S_n}{s_n} \Rightarrow \mathcal{N}(0, 1).$$

To verify the tail negligibility condition, proceed as follows. First,

$$\begin{aligned} \mathrm{E}[Y_n^2(i); |Y_n(i)| > \varepsilon s_n] &\le \sqrt{\mathrm{E}\, Y_n^4(i)}\, \sqrt{\Pr(|Y_n(i)| > \varepsilon s_n)} \\ &\le \sqrt{\mathrm{E}\, Y_n^4(i)}\, \sqrt{\frac{\mathrm{E}\, Y_n^2(i)}{\varepsilon^2 s_n^2}} \\ &= \sqrt{\mathrm{E}\, Y_n^4(i)}\, \frac{1}{\varepsilon \sqrt{m_n}}; \end{aligned}$$

thus,

$$R_n \leq \frac{1}{\operatorname{var} M_{a_n}(0)} \sqrt{\mathrm{E}\, Y_n^4(i)} \frac{1}{\varepsilon \sqrt{m_n}}.$$

Observe that, for some $C_u < \infty$,

$$\begin{aligned}
\frac{\mathrm{E}\, Y_n^4(i)}{(\log a_n)^4} &= \mathrm{E}\left[\frac{M_{a_n}(i)}{\log a_n} - \mathrm{E}\,\frac{M_{a_n}(0)}{\log a_n}\right]^4 \\
&\leq 16\,\mathrm{E}\left(\frac{M_{a_n}(0)}{\log a_n}\right)^4 \\
&\leq C_u
\end{aligned}$$

by Lemma 4 in Appendix B. Fix $\gamma \in \mathbb{R}_+$. Now,

$$\begin{aligned}
\operatorname{var} M_{a_n}(0) &= \mathrm{E}[M_{a_n}(0) - \mathrm{E}\, M_{a_n}(0)]^2 \\
&\geq \gamma^2 \Pr(|M_{a_n}(0) - \mathrm{E}\, M_{a_n}(0)| > \gamma) \\
&\geq \gamma^2 \Pr(M_{a_n}(0) - \mathrm{E}\, M_{a_n}(0) > \gamma)
\end{aligned}$$

by the Markov inequality. But, for n sufficiently large,

$$\begin{aligned}
\mathrm{E}\left[\frac{\theta^* M_{a_n}(0)}{\log a_n}\right] &= \int_0^\infty \Pr(\theta^* M_{a_n}(0) \geq x \log a_n)\,\mathrm{d}x \\
&\leq a_n \int_1^\infty \Pr(\theta^* W^* > x \log a_n)\,\mathrm{d}x + 1 \\
&\leq \int_1^\infty C \mathrm{e}^{-(x-1)\log a_n}\,\mathrm{d}x + 1 \\
&= C \int_0^\infty \mathrm{e}^{-x \log a_n}\,\mathrm{d}x + 1 \\
&= \frac{C}{\log a_n} + 1,
\end{aligned}$$

where the first inequality follows from the union bound and the second follows from the definition of the class of marginal distributions. Consequently,

$$\begin{aligned}
\Pr(M_{a_n}(0) - \mathrm{E}\, M_{a_n}(0) > \gamma) &\geq \Pr\left(M_{a_n}(0) > \gamma + \frac{\log a_n + C}{\theta^*}\right) \\
&= 1 - \Pr^{a_n}\left(X < \gamma + \frac{\log a_n + C}{\theta^*}\right) \\
&\geq 1 - \exp\left\{-a_n \Pr\left(X > \gamma + \frac{\log a_n + C}{\theta^*}\right)\right\}
\end{aligned}$$

and

$$a_n \Pr\left(X > \gamma + \frac{\log a_n + C}{\theta^*}\right) \geq \mathrm{e}^{-\theta^* \gamma - C}$$

by the tail condition for all sufficiently large n. Thus, we can choose $\gamma > 0$ such that, for some $C_l > 0$,

$$\liminf_{n\to\infty} \Pr(M_{a_n}(0) - \mathrm{E}\, M_{a_n}(0) > \gamma) \geq C_l.$$

Consequently, $\operatorname{var} M_{a_n}(0) \ge \gamma^2 C_l$ for all but finitely many n. It follows that

$$R_n \le \frac{(\log a_n)^2}{\gamma^2 C_l} C_u \frac{1}{\varepsilon \sqrt{m_n}},$$

so we can choose a_n, m_n so that $m_n^{-1/2}(\log a_n)^2 \to 0$. In particular, we can choose $m_n = n^\gamma$ for some $\gamma \in (0, 1)$, and we have $R_n \to 0$ as $n \to \infty$, which concludes the proof.

A.8. Proof of Proposition 4

We divide the proof into steps.

Step 1. (*Preliminaries.*) Under the conditions of the proposition, $(X_n : n \ge 0)$ forms a stationary Markov chain that is geometrically ergodic; thus (see [28, Theorem 1′]) it is β-mixing with exponential rate (for a definition of β-mixing, and properties, the reader is referred to [12]). Consequently, since $\beta(k) \ge \alpha(k)$, it is also strong mixing with the same rate, and therefore Theorem 1 applies. Now, using the version of the Glivenko–Cantelli theorem for β-mixing processes given in Lemma 2 in Appendix B, for any sequence κ_n of real numbers,

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le \kappa_n\}} - P(X \le \kappa_n)\right| &\le \sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le x\}} - P(X \le x)\right| \\ &= O\left(\sqrt{\frac{(\log n)^2}{n}}\right) \quad \text{a.s.}\end{aligned}$$

In particular,

$$|\hat{p}_n(\kappa_n) - \bar{F}(\kappa_n)| = O\left(\sqrt{\frac{(\log n)^2}{n}}\right) \quad \text{a.s.,}$$

and consequently

$$Z_n := \left|\frac{\hat{p}_n(\kappa_n)}{\bar{F}(\kappa_n)} - 1\right| = O\left(\sqrt{\frac{(\log n)^2}{n}}\frac{1}{\bar{F}(\kappa_n)}\right) \quad \text{a.s.}$$

Step 2. We write

$$\begin{aligned}\left|\frac{\hat{\eta}_n}{\eta} - 1\right| &= \left|\frac{\hat{p}_n(\kappa_n)\mathrm{e}^{\hat{\theta}_n\kappa_n}}{\bar{F}(\kappa_n)\mathrm{e}^{\theta^*\kappa_n}} - 1\right| \\ &= \left|\left[\left(\frac{\hat{p}_n(\kappa_n)}{\bar{F}(\kappa_n)} - 1\right)\mathrm{e}^{(\hat{\theta}_n-\theta^*)\kappa_n} + \mathrm{e}^{(\hat{\theta}_n-\theta^*)\kappa_n}\right](1 + o(1)) - 1\right| \\ &\le I_n + J_n + K_n, \end{aligned} \tag{11}$$

where

$$I_n = |Z_n \mathrm{e}^{(\hat{\theta}_n-\theta^*)\kappa_n}|(1 + o(1)), \qquad J_n = |\mathrm{e}^{(\hat{\theta}_n-\theta^*)\kappa_n} - 1|, \qquad K_n = |\mathrm{e}^{(\hat{\theta}_n-\theta^*)\kappa_n}|o(1).$$

Step 3. (*Proofs for the separate cases.*) (i) By the condition on κ_n and the rate of convergence given in (5),

$$\limsup_{n\to\infty} |(\hat{\theta}_n - \theta^*)\kappa_n| \le \limsup_{n\to\infty}\left[|\hat{\theta}_n - \theta^*|\frac{\log n}{\log\log n}\right]\limsup_{n\to\infty}\left[\kappa_n \frac{\log\log n}{\log n}\right] = 0.$$

Now, by assumption, $\bar{F}(\kappa_n) \sim \eta \mathrm{e}^{-\theta^*\kappa_n}$. Thus, since $\kappa_n = o(\log n)$,

$$\sqrt{\frac{(\log n)^2}{n}}\frac{1}{\bar{F}(\kappa_n)} \sim \sqrt{(\log n)^2}\,\mathrm{e}^{\theta^*\kappa_n-(1/2)\log n} \to 0$$

as $n \to \infty$. Thus, $\limsup_n |Z_n| = 0$ almost surely, and going back to (11) it is clear that

$$\left|\frac{\hat{\eta}_n}{\eta} - 1\right| \to 0$$

as $n \to \infty$ almost surely.

(ii) Identical to case (i) except that now we only have $|(\hat{\theta}_n - \theta^*)\kappa_n| = o_p(1)$.

(iii) From Lemma 2 it is clear that $Z_n \Rightarrow 0$ for $c < 1/2\theta^*$ and for $c \ge 1/2\theta^*$ the method of proof no longer yields a convergence result. Thus,

$$\begin{aligned}\log\hat{\eta}_n - \log\eta &= \log\hat{p}_n(\kappa_n) + \hat{\theta}_n\kappa_n - \log\eta\\ &= I_n + J_n + K_n,\end{aligned}$$

where

$$I_n = \log\frac{\hat{p}_n(\kappa_n)}{\bar{F}(\kappa_n)}, \qquad J_n = (\hat{\theta}_n - \theta^*)\kappa_n, \qquad K_n = \log\bar{F}(\kappa_n) + \theta^*\kappa_n - \log\eta.$$

Here $I_n = o_p(1)$ for $c < 1/2\theta^*$ and $K_n = o(1)$ by assumption. Since the process is exponentially mixing, we have by a result of Loynes [27] that $\theta^* M_n - \log n - \log\phi\eta \Rightarrow Z$, where $\phi \in (0, 1)$ and Z has the standard Gumbel distribution. Then

$$\begin{aligned}(\hat{\theta}_n - \theta^*)\log n &= c\log n\frac{\theta^* M_n - \log n}{M_n}\\ &\Rightarrow c\theta^*(Z + \log\phi\eta)\end{aligned}$$

by the continuous mapping theorem. Putting everything together, and using the converging together principle, we have the result.

This concludes the proof.

Appendix B. Auxiliary results and proofs

Lemma 1. *Let $\{X_i\}_{i=1}^n$ be a sequence of random variables with common marginal distribution F, and let $\theta^* = \sup\{\theta : \mathrm{E}\,\mathrm{e}^{\theta X} < \infty\}$. Suppose that $\theta^* < \infty$. Then*

$$\limsup_{n\to\infty}\frac{M_n}{\log n} \le \frac{1}{\theta^*} \quad a.s.$$

Proof. Fix $\delta > 0$. Then

$$\sum_{n=1}^{\infty} \Pr\left(X_n > \frac{(1+\delta)\log n}{\theta^*}\right) = \sum_{n=1}^{\infty} \Pr\left(X > \frac{(1+\delta)\log n}{\theta^*}\right)$$
$$\leq \mathrm{E}\, \mathrm{e}^{\theta^* X/(1+\delta)}$$
$$< \infty.$$

Thus,

$$\limsup_{n\to\infty} \frac{X_n}{\log n} \leq \frac{1+\delta}{\theta^*} \quad \text{a.s.}$$

and, since δ is arbitrary, we have $X_n/\log n \leq 1/\theta^*$ eventually, almost surely. Now, for each ω for which the above convergence holds, there exists an $N(\omega)$ such that $X_n(\omega)/\log n \leq 1/\theta^*$ for all $n > N(\omega)$. Then

$$\begin{aligned}
\frac{M_n}{\log n} &= \frac{\max\{(X_1, X_2, \dots, X_N), (X_{N+1}, \dots, X_n)\}}{\log n} \\
&= \max\left\{\bigvee_{i=1}^{N} \frac{X_i}{\log n}, \bigvee_{i=N+1}^{n} \frac{X_i}{\log n}\right\} \\
&\leq \frac{\bigvee_{i=1}^{N} X_i}{\log n} + \bigvee_{i=N+1}^{n} \frac{X_i}{\log i} \\
&\leq \frac{\bigvee_{i=1}^{N} X_i}{\log n} + \frac{1}{\theta^*},
\end{aligned}$$

where the first inequality follows since $\log i \leq \log n$ for $i = N+1, \dots, n$ and the second since $X_i/\log i \leq 1/\theta^*$ for $i > N$. Thus,

$$\limsup_{n\to\infty} \frac{M_n}{\log n} \leq \frac{1}{\theta^*}$$

as

$$\lim_{n\to\infty} \frac{\bigvee_{i=1}^{N} X_i}{\log n} = 0,$$

which concludes the proof.

Lemma 2. *If* $(Y_n : n \geq 0)$ *is* β*-mixing such that* $\beta(k) = O(k^{-(2+\varepsilon)})$ *for some* $\varepsilon > 0$*, then there exists an* $\varepsilon' \in (0, \min\{\varepsilon/3, 1\})$ *such that*

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \geq x\}} - \Pr(Y > x) \right| = O\left(\sqrt{\frac{\log n}{n^{\varepsilon'}}}\right) \quad \textit{a.s.}$$

If $\beta(k) = O(e^{-ck})$ *for some* $c > 0$*, then*

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \geq x\}} - \Pr(Y > x) \right| = O\left(\sqrt{\frac{(\log n)^2}{n}}\right) \quad \textit{a.s.}$$

Proof. The starting point of our analysis is the following result that gives exponential bounds in the Glivenko–Cantelli theorem (for details see e.g. [11, Theorem 12.4]).

Proposition 5. *Let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d. real-valued random variables with common distribution P. Then, for all $\delta > 0$ and all n,*

$$\Pr\left\{\sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\le x\}} - P(X\le x)\right| > \delta\right\} \le 8(n+1)\mathrm{e}^{-n\delta^2/32}.$$

Note that the result is 'distribution free' in the sense that it holds for any arbitrary probability distribution, as long as the X_i are i.i.d. random variables. We now extend this to the β-mixing case, and conclude the assertions of Lemma 2.

Step 1. (*Measure-theoretic preliminaries.*) Let $X = (X_1, X_2, \dots)$ be a stationary β-mixing process. To concur with standard definitions in the literature, it will be useful to consider the two-sided stationary extension of X and, with some abuse of notation, continue referring to this process as X. Let Pr be the stationary probability measure on $(\mathbb{R}^{\mathbb{Z}}, \mathscr{B}^{\mathbb{Z}})$ associated with X, and let $\Pr_{-\infty}^{0}$, $\Pr_1^{\infty}$ denote the semi-infinite marginals of Pr. Let

$$\mathcal{F}_k^{\ell} = \sigma(X_k, X_2, \dots, X_\ell).$$

One standard definition of β-mixing (see [9, Section 2, (2.1)]) is then

$$\beta(k) = \sup\{|\Pr(A) - (\Pr_{-\infty}^{0}\times\Pr_1^{\infty})(A)| : A \in \sigma(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_k^{\infty})\}. \tag{12}$$

Let the one-dimensional marginal of Pr be denoted by P, and let $\Pr_0 = \prod_{-\infty}^{\infty} P$ denote the product probability measure generated by P. A simple consequence of the definition (12) is the following (see [29, Lemma 2]): if $A \in \sigma(X_0, X_k, \dots, X_{(m-1)k})$, then

$$|\Pr(A) - \Pr_0(A)| \le m\beta(k). \tag{13}$$

Step 2. Given a sequence $\{X_i\}_{i=1}^n$ from X, let k_n and m_n be two sequences of positive integers such that $k_n, m_n \uparrow \infty$ as $n\to\infty$ and assume for simplicity of exposition that $k_n m_n = n$. It will be clear in what follows that this assumption entails no loss of generality. Now,

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\le x\}} - \Pr(X\le x)\right| &= \sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{j=1}^{k_n}\sum_{\ell=0}^{m_n-1}[\mathbf{1}_{\{X_{\ell k_n+j}\le x\}} - \Pr(X\le x)]\right| \\ &\le \frac{1}{k_n}\sum_{j=1}^{k_n}\sup_{x\in\mathbb{R}}\left|\frac{1}{m_n}\sum_{\ell=0}^{m_n-1}\mathbf{1}_{\{X_{\ell k_n+j}\le x\}} - \Pr(X\le x)\right|\end{aligned}$$

and note that, for $\delta > 0$,

$$\left\{\sup_{x\in\mathbb{R}}\left|\frac{1}{m_n}\sum_{\ell=0}^{m_n-1}\mathbf{1}_{\{X_{\ell k_n+j}\le x\}} - \Pr(X\le x)\right| > \delta\right\} \in \sigma(X_j, X_{j+k_n}, \dots, X_{(m_n-1)k_n+j}).$$

We can now apply (13) as follows:

$$\begin{aligned}
&\Pr\Bigg(\sup_{x\in\mathbb{R}}\bigg|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\le x\}}-\Pr(X\le x)\bigg|>\delta\Bigg)\\
&\quad\le \Pr\Bigg(\frac{1}{k_n}\sum_{j=1}^{k_n}\sup_{x\in\mathbb{R}}\bigg|\frac{1}{m_n}\sum_{\ell=0}^{m_n-1}\mathbf{1}_{\{X_{\ell k_n+j}\le x\}}-\Pr(X\le x)\bigg|>\delta\Bigg)\\
&\quad\le k_n\Pr\Bigg(\sup_{x\in\mathbb{R}}\bigg|\frac{1}{m_n}\sum_{\ell=0}^{m_n-1}\mathbf{1}_{\{X_{\ell k_n+j}\le x\}}-\Pr(X\le x)\bigg|>\delta\Bigg)\\
&\quad\le k_n\Pr_0\Bigg(\sup_{x\in\mathbb{R}}\bigg|\frac{1}{m_n}\sum_{\ell=0}^{m_n-1}\mathbf{1}_{\{X_{\ell k_n+j}\le x\}}-\Pr(X\le x)\bigg|>\delta\Bigg)+k_nm_n\beta(k_n)\\
&\quad\le 8(n+1)\mathrm{e}^{-m_n\delta^2/32}+n\beta(k_n),
\end{aligned}$$

where the third inequality follows from (13) and the fourth from Proposition 5 and since $k_nm_n=n$.

Step 3. First consider the case of $\beta(k)=O(k^{-(2+\varepsilon)})$. Then take $\varepsilon'\in(0,\varepsilon)$, $\delta=(\log n/n^{\varepsilon'})^{1/2}$, and $k_n=cn^{1-\varepsilon'}$ so as to make the exponential bound summable (e.g. $c=70$ will suffice). Also,

$$n\beta(k_n)\le\frac{1}{n^{1-3\varepsilon'+\varepsilon}}.$$

Since both terms in the upper bound are summable, we can use the Borel–Cantelli lemma to conclude that

$$\sup_{x\in\mathbb{R}}\bigg|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\le x\}}-P(X\le x)\bigg|=O\bigg(\sqrt{\frac{\log n}{n^{\varepsilon'}}}\bigg)\quad\text{a.s.}$$

Similar considerations in the exponential mixing case lead to the choice $k_n=c\log n$ for some appropriate choice of c. This gives rise to the asserted convergence rate,

$$\sup_{x\in\mathbb{R}}\bigg|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\le x\}}-P(X\le x)\bigg|=O\bigg(\sqrt{\frac{(\log n)^2}{n}}\bigg)\quad\text{a.s.,}$$

which concludes the proof.

Lemma 3. *Let $\mathbb{X}=(X_n:n\ge1)$ be an aperiodic classically regenerative process, with embedded renewal sequence $(T(k):k\ge0)$. Assume that, for some $m\ge1$, the cycle lengths $\tau_k=T(k)-T(k-1)$ satisfy $\mathrm{E}\,\tau_1^m<\infty$, and $F(x)=\Pr(\tau\le x)$ has regularly varying tails. Then $\mathbb{X}$ has a stationary version which is α-mixing and, for some constant $c>0$,*

(a) $n^{m-1}\alpha(n)=o(1)$ *as* $n\to\infty$,

(b)

$$\liminf_{n\to\infty}\frac{\alpha(n)}{n(1-F(n))}\ge c.$$

Proof. First, since $\mathbb{X}$ is positive recurrent, it admits a stationary version. To make the sequence of renewal epochs stationary, the time to the first renewal must have the distribution of the *forward recurrence time*. That is, for $k \geq 0$, let S be the random variable with distribution function

$$\Pr(S > k) = \frac{1}{\mathrm{E}\,\tau_1} \int_k^{\infty} (1 - F(y))\,\mathrm{d}y,$$

where $F(y) = 1 - \Pr(\tau \leq y)$. Now, we can extend the sequence $\mathbb{X} = (X_n : n \geq 1)$ to a two-sided stationary sequence. For $m, n \in \mathbb{Z}_+$, let $N(-m, n]$ denote the number of renewals in the interval $(-m, n]$. Define the following two events:

$$A_n = \{\omega : N(-n, 0] = 0\} \in \mathcal{B}_{-\infty}^{0},$$
$$B_n = \{\omega : N(n, 2n] = 0\} \in \mathcal{B}_{n}^{\infty},$$

where $\mathcal{B}$ denotes the underlying σ-field. By the definition of the α-mixing coefficients,

$$\begin{aligned}\alpha(n) &\geq |\Pr(A_n \cap B_n) - \Pr(A_n)\Pr(B_n)| \\ &= \Pr(S > n)\left|\frac{\Pr(A_n \cap B_n)}{\Pr(S > n)} - \frac{\Pr(A_n)\Pr(B_n)}{\Pr(S > n)}\right|.\end{aligned}$$

Observe that

$$\frac{\Pr(A_n)\Pr(B_n)}{\Pr(S > n)} = \Pr(A_n),$$

and note that $\Pr(A_n) \downarrow 0$ as $n \to \infty$. Finally, using stationarity we see that

$$\begin{aligned}\frac{\Pr(A_n \cap B_n)}{\Pr(S > n)} &\geq \frac{\Pr(A_n \cap B_n \cap \{S > n\})}{\Pr(S > n)} \\ &= \frac{\Pr(S > 3n)}{\Pr(S > n)} \\ &= \frac{\int_{3n}^{\infty}(1 - F(y))\,\mathrm{d}y}{\int_{n}^{\infty}(1 - F(y))\,\mathrm{d}y}\end{aligned}$$

and, using Karamata's theorem (see [15, p. 567]), we have that

$$\frac{\Pr(S > 3n)}{\Pr(S > n)} \to 3^{\gamma}$$

as $n \to \infty$, with $\gamma < 0$ the index of variation for F, that is, $F \in \mathcal{R}_{\gamma}$. Putting all of the above together, we find that

$$\liminf_{n\to\infty} \frac{\alpha(n)}{\Pr(S > n)} \geq 3^{\gamma} > 0,$$

leading immediately to (b). The upper bound follows from [18, Proposition 6.10], which asserts that, for an aperiodic positive recurrent regenerative process, the stationary version is strong mixing with $\alpha(n) = o(n^{-m+1})$.

Lemma 4. *Let $X_1, X_2, \ldots$ be an arbitrary sequence of real-valued random variables with common marginal distribution $F \in \mathcal{F}$. Let $M_n = \max_{1\leq i\leq n}\{X_i\}$. Then, for any $p \in [1, \infty)$,*

$$\sup_{n\geq 2} \mathrm{E}\left(\frac{M_n}{\log n}\right)^p < \infty.$$

Proof. Fix $p \in [1, \infty)$ and a distribution $F \in \mathcal{F}$. Define

$$K_1 = \inf\left\{ y > 0 : \text{ such that } \frac{\log P(X_1 > x)}{x} \le -\frac{\theta^*}{2} \text{ for all } x \ge y \right\}.$$

Note that $K_1 < \infty$ follows from the definition of the class $\mathcal{F}$, i.e.

$$\limsup_{x \to \infty} \frac{\log P\{X_1 > x\}}{x} \le -\frac{\theta^*}{2},$$

and set $K := \max\{K_1, 4/\theta^*\}$. Then

$$\begin{aligned}
\mathrm{E}\left[\frac{M_n}{\log n}\right]^p &= \int_0^\infty p y^{p-1} \Pr(M_n > y \log n)\, \mathrm{d}y \\
&= \int_0^K p y^{p-1} \Pr(M_n > y \log n)\, \mathrm{d}y + \int_K^\infty p y^{p-1} \Pr(M_n > y \log n)\, \mathrm{d}y \\
&\le K^p + \int_K^\infty p y^{p-1} n \Pr(X_1 > y \log n)\, \mathrm{d}y \\
&\le K^p + \int_K^\infty p y^{p-1} \exp\left\{ \log n \left(1 + y \frac{\log \Pr(X_1 > y \log n)}{y \log n} \right) \right\} \mathrm{d}y \\
&\le K^p + \int_K^\infty p y^{p-1} \exp\left\{ \log n \left(1 - y \frac{\theta^*}{2} \right) \right\} \mathrm{d}y \\
&\le K^p + \int_K^\infty p y^{p-1} \mathrm{e}^{-(\theta^*/4) y \log n}\, \mathrm{d}y \\
&\le K^p + \left(\frac{4}{\theta^*}\right)^p \frac{p!}{(\log n)^p},
\end{aligned}$$

where the first inequality follows from the union bound, and the third and fourth inequalities follow from the definition of K, noting that $\theta^* y/2 - 1 \ge \theta^* y/4$ for $y \ge 4/\theta^*$.

Lemma 5. *Let $X = (X_n : n \ge 1)$ satisfy either condition (i) or (ii) of Theorem 1, and assume that the common marginal distribution F is in $\mathcal{F}$. Let $\tilde{M}_n = \max_{1 \le i \le n} X_i \vee 1$. Then, for any $p \in [1, \infty)$,*

$$\sup_{n \ge 2} \mathrm{E}\left(\frac{\log n}{\tilde{M}_n}\right)^p < \infty.$$

Proof. Fix $\delta > 0$ and let $a_n = ((1 - \delta)/\theta^*) \log n$. First,

$$\mathrm{E}\left(\frac{\log n}{\tilde{M}_n}\right)^p = I_n + J_n,$$

where

$$I_n = \mathrm{E}\left(\left[\frac{\log n}{\tilde{M}_n}\right]^p ; \tilde{M}_n < a_n\right), \qquad J_n = \mathrm{E}\left(\left[\frac{\log n}{\tilde{M}_n}\right]^p ; \tilde{M}_n \ge a_n\right)$$

and note that J_n can bounded as follows:

$$J_n \le \left(\frac{\theta^*}{1 - \delta}\right)^p.$$

To bound I_n use the Cauchy–Schwarz inequality,

$$I_n \le \sqrt{\mathrm{E}\left|\frac{\log n}{\tilde{M}_n}\right|^{2p}}\sqrt{\Pr(M_n \le a_n)},$$

and note that

$$\mathrm{E}\left|\frac{\log n}{\tilde{M}_n}\right|^{2p} \le (\log n)^{2p}.$$

Now, if X is i.i.d., then

$$\Pr(M_n \le a_n) \le \mathrm{e}^{-n\Pr(X>a_n)} \le \mathrm{e}^{-n^{\delta}}.$$

Thus

$$\sup_n J_n < \infty.$$

A very similar analysis goes through in the case of mixing processes, the details of which are omitted, but are obvious from the proof technique of Theorem 1.

Acknowledgement

The authors thank the referee for his careful reading and helpful comments.

References

[1] Asmussen, S. (1998). Extreme value theory for queues via cycle maxima. *Extremes* **1,** 137–168.

[2] Athreya, K. B. and Pantula, S. G. (1986). Mixing properties of Harris chains and autoregressive processes. *J. Appl. Prob.* **23,** 880–892.

[3] Beran, J., Sherman, R., Taqqu, M. and Willinger, W. (1995). Long-range dependence in variable-bit-rate video traffic. *IEEE Trans. Commun.* **43,** 1566–1579.

[4] Berger, A. W. and Whitt, W. (1995). Maximum values in queueing processes. *Prob. Eng. Inf. Sci.* **9,** 375–409.

[5] Berman, S. M. (1964). Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.* **35,** 502–516.

[6] Bertsimas, D. and Paschalidis, I. (2001). Probabilistic service level guarantees in make-to-stock manufacturing systems. *Operat. Res.* **49,** 119–133.

[7] Bertsimas, D., Paschalidis, I. and Tsitsiklis, J. N. (1998). On the large deviations behaviour of acyclic networks of G/G/1 queues. *Ann. Appl. Prob.* **8,** 1027–1069.

[8] Billingsley, P. (1986). *Probability and Measure*. John Wiley, New York.

[9] Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability and Statistics* (Progr. Prob. Statist. **11**), Birkhäuser, Boston, MA, pp. 165–192.

[10] Courcoubetis, C. *et al.* (1995). Admission control and routing in ATM networks using inference from measured buffer occupancy. *IEEE Trans. Commun.* **43,** 1778–1784.

[11] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer, New York.

[12] Doukhan, P. (1994). *Mixing. Properties and Examples*. Springer, New York.

[13] Drees, H. (2001). Minimax risk bounds in extreme value theory. *Ann. Statist.* **21,** 266–294.

[14] Duffield, N. G. and O'Connell, N. (1995). Large deviations and overflow probabilities with general single-server queue, with applications. *Math. Proc. Camb. Phil. Soc.* **118,** 363–374.

[15] Embrechts, P., Klüppelberg, T. and Mikosch, C. (1997). *Modelling Extremal Events*. Springer, Berlin.

[16] Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. John Wiley, New York.

[17] Glasserman, P. and Kou, S. (1995). Limits of first passage times to rare sets in regenerative processes. *Ann. Appl. Prob.* **5,** 424–445.

[18] Glynn, P. W. (1982). Simulation output analysis for general state space Markov chains. Doctoral Thesis, Stanford University.

[19] Glynn, P. W. and Whitt, W. (1994). Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. In *Studies in Applied Probability* (J. Appl. Prob. Spec. Vol. **31A**), eds J. Galambos and J. Gani, Applied Probability Trust, Sheffield, pp. 131–156.

[20] Glynn, P. W. and Zeevi, A. J. (2000). Estimating tail probabilities in queues via extremal statistics. In *Analysis of Communication Networks: Call Centres, Traffic and Performance* (Fields Inst. Commun. **28**), American Mathematical Society, Providence, RI, pp. 135–158.
[21] Haiman, G. and Habach, L. (1999). Almost sure behaviour of extremes of m-dependent stationary sequences. *C. R. Acad. Sci. Paris Sér. I Math.* **329,** 887–892.
[22] Hall, P., Teugels, J. L. and Vanmarcke, A. (1992). The abscissa of convergence of the Laplace transform. *J. Appl. Prob.* **29,** 353–362.
[23] Hsu, I. and Walrand, J. (1996). Dynamic bandwidth allocation for ATM switches. *J. Appl. Prob.* **33,** 758–771.
[24] Iglehart, D. L. (1972). Extreme values in the GI/G/1 queue. *Ann. Math. Statist.* **43,** 627–635.
[25] Kesten, H. and O'Brien, G. L. (1976). Examples of mixing sequences. *Duke Math. J.* **43,** 405–415.
[26] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.
[27] Loynes, R. M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.* **36,** 993–999.
[28] Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stoch. Process. Appl.* **29,** 309–315.
[29] Nobel, A. B. and Dembo, A. (1993). A note on uniform laws of averages for dependent processes. *Statist. Prob. Lett.* **17,** 169–172.
[30] O'Brien, G. L. (1987). Extreme values for stationary and Markov sequences. *Ann. Prob.* **15,** 281–291.
[31] Paschalidis, I. and Vassilaras, S. (2001). On the estimation of buffer overflow probabilities from measurements. *IEEE Trans. Inf. Theory* **47,** 178–191.
[32] Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
[33] Rootzén, H. (1988). Maxima and exceedances of stationary Markov chains. *Adv. Appl. Prob.* **20,** 371–390.
[34] Zeevi, A. and Glynn, P. W. (2000). On the maximum workload of a queue fed by fractional Brownian motion. *Ann. Appl. Prob.* **10,** 1084–1099.
[35] Zeevi, A. and Glynn, P. W. (2001). Estimating tail decay for stationary sequences via extreme values. Tech. Rep., Graduate School of Business, Columbia University.