HOW TO ESTIMATE THE RATE FUNCTION OF A CUMULATIVE PROCESS

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

ANTHONY P. METCALFE,** *University College Cork*

Abstract

Consider two sequences of bounded random variables, a value and a timing process, that satisfy the large deviation principle (LDP) with rate function $J(\cdot, \cdot)$ and whose cumulative process satisfies the LDP with rate function $I(\cdot)$. Under mixing conditions, an LDP for estimates of I constructed by transforming an estimate of J is proved. For the case of a cumulative renewal process it is demonstrated that this approach is favourable to a more direct method, as it ensures that the laws of the estimates converge weakly to a Dirac measure at I.

Keywords: Estimating large deviations; cumulative process; renewal process

2000 Mathematics Subject Classification: Primary 60F10
Secondary 60K05

1. Introduction, setup, and estimation schemes

Let $\{(X_n, \tau_n), n \in \mathbb{N}\}$ be a stationary process of (not necessarily independent) bounded random variables taking values in $[a, b] \times [\alpha, \beta] \subset \mathbb{R} \times (0, \infty)$. Defining $S_n := X_1 + \cdots + X_n$ and $T_n := \tau_1 + \cdots + \tau_n$, we assume that $\{(S_n, T_n)/n, n \in \mathbb{N}\}$ satisfies the large deviation principle (LDP) on the scale $1/n$, with rate function J that is the Legendre–Fenchel transform of its scaled cumulant generating function (sCGF):

$$J(x_1, x_2) = \sup_{(\theta_1, \theta_2) \in \mathbb{R}^2} (\theta_1 x_1 + \theta_2 x_2 - M(\theta_1, \theta_2)), \tag{1}$$

where

$$M(\theta_1, \theta_2) = \lim_{n \to \infty} \frac{1}{n} \log \mathrm{E}[\exp(\theta_1 S_n + \theta_2 T_n)].$$

Treating τ_i as the interarrival time between volumes X_{i-1} and X_i, the (pure, zero-delayed) cumulative process $\{A_t, t \in \mathbb{R}^+\}$ is defined by

$$A_t := \sum_{i=1}^{N(t)} X_i,$$

where $N(t) := \sup\{n : T_n \le t\}$ is the counting process associated with $\{T_n, n \in \mathbb{N}\}$ and the empty sum with $N(t) = 0$ is defined to have value zero. Cumulative processes arise naturally in many applications, including queueing theory and risk theory. Assume that $\{A_t/t, t \in \mathbb{R}^+\}$

satisfies the LDP on the scale $1/t$, with rate function I that is the Legendre–Fenchel transform of its sCGF:

$$I(x) = \sup_{\theta \in \mathbb{R}} (\theta x - \lambda(\theta)), \quad \text{where } \lambda(\theta) = \lim_{t \to \infty} \frac{1}{t} \log \mathrm{E}[\exp(\theta A_t)]. \tag{2}$$

If we are given the observations $(X_1, \tau_1), (X_2, \tau_2), \dots$, but the statistics of the process $\{(X_n, \tau_n),\ n \in \mathbb{N}\}$ are unknown, how would we estimate the rate function I? The purpose of this work is to consider the large deviations of estimating I by transforming an estimate of the sCGF M. Although a more direct approach based on estimating λ and taking its Legendre–Fenchel transform is possible, this scheme performs better in certain typical circumstances; for instance in the case when $\{A_t,\ t \in \mathbb{R}^+\}$ is a cumulative renewal process, that is, $\{(X_n, \tau_n),\ n \in \mathbb{N}\}$ forms an independent and identically distributed (i.i.d.) sequence. This is demonstrated in Section 3.

The transformation is based on the following observation. For a broad class of processes it is known that J and I from (1) and (2) are related by

$$I(x) = \inf_{y>0} y J\left(\frac{x}{y}, \frac{1}{y}\right) \tag{3}$$

(see [11], [19], [26]). Having estimated J, our estimate of I is constructed by transforming the estimates through (3).

In order to discuss the LDP for estimates of I and J, we must select topological spaces for I and J and their estimates. The spaces of lower-semicontinuous (LSC) convex functions from $[a/\beta, b/\alpha]$ and $[a, b] \times [\alpha, \beta]$, taking values in $\mathbb{R} \cup \{+\infty\}$, and respectively denoted by $\mathrm{conv}[a/\beta, b/\alpha]$ and $\mathrm{conv}([a, b] \times [\alpha, \beta])$ are natural choices. However, the usual topologies, such as uniform convergence on bounded subsets, are not suitable. For example, we want a topology in which estimates such as $I_n = n|x|$ converge to I, which is 0 at $x = 0$ and ∞ otherwise.

The topology we employ is the Attouch–Wets topology (see [1] and [2]), which is a metrizable topology that identifies the convergence of functions with the convergence of their epigraphs. For an extended real-valued function $f : Z \to \bar{\mathbb{R}}$, where $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$, on a metric space (Z, d), the epigraph of f is defined to be $\mathrm{epi}\, f := \{(z, \alpha) \in Z \times \mathbb{R} : \alpha \geq f(z)\}$. Equipping $Z \times \mathbb{R}$ with the box metric D, f is LSC if and only if $\mathrm{epi}\, f$ is a closed subset of $Z \times \mathbb{R}$ (see, for example, [25, Theorem 1.6]). Therefore, a topology for LSC functions is inherited from a topology on the closed subsets of the metric space.

Let $\mathrm{CL}(Z)$ denote the collection of nonempty closed subsets of Z. A sequence $\{A_n,\ n \in \mathbb{N}\} \subset \mathrm{CL}(Z)$ converges in the Attouch–Wets topology to $A = \tau_{\mathrm{AW}d} - \lim_{n \to \infty} A_n \in \mathrm{CL}(Z)$ if, given any bounded set $B \subset Z$ and any $\varepsilon > 0$, there exists an N_ε such that

$$\sup_{z \in B} |d(z, A_n) - d(z, A)| < \varepsilon, \quad \text{for all } n > N_\varepsilon,$$

where $d(z, A) := \inf_{x \in A} d(z, x)$. Letting $\mathrm{conv}(Z)$ denote the set of convex LSC functions from Z to $\mathbb{R} \cup \{+\infty\}$, a sequence $\{f_n,\ n \in \mathbb{N}\} \subset \mathrm{conv}(Z)$ converges to $f \in \mathrm{conv}(Z)$ if and only if $\{\mathrm{epi}\, f_n,\ n \in \mathbb{N}\}$ converges to $\mathrm{epi}\, f$ in the Attouch–Wets topology on $\mathrm{CL}(Z \times \mathbb{R})$, i.e. $f = \tau_{\mathrm{AW}D} - \lim_{n \to \infty} f_n$ if and only if $\mathrm{epi}\, f = \tau_{\mathrm{AW}D} - \lim_{n \to \infty} \mathrm{epi}\, f_n$.

Whenever Z is a subset of $\mathbb{R}^n$ (for any finite n) equipped with the Euclidean distance d, Theorem 3.3.3 of [3] can be used to show that the topologies $\tau_{\mathrm{AW}D}$ and $\tau_{\mathrm{AW}d}$ on $\mathrm{CL}(Z \times \mathbb{R})$

coincide. Thus, either can be used to form the Attouch–Wets topology on conv(Z), and we denote this topology simply by τ_{AW} from now on.

A necessary and sufficient condition for checking convergence in CL(Z) is the following theorem, which we shall use in our proofs.

Theorem 1. ([3, Theorem 3.1.7]) *Let* (Z, d) *be a metric space, and let* $A, A_1, A_2, \ldots$ *be a sequence in* CL(Z). *Then* $A = \tau_{AW_d} - \lim_{n \to \infty} A_n$ *if and only if for each bounded subset* B *of* Z *we have both* $\lim_{n \to \infty} e_d(A \cap B, A_n) = 0$ *and* $\lim_{n \to \infty} e_d(A_n \cap B, A) = 0$, *where, for any two nonempty subsets* E *and* F, $e_d(E, F)$, *the excess of* E *over* F, *is given by* $e_d(E, F) = \sup\{d(e, F), e \in E\}$.

A good reference for the Attouch–Wets topology is [3]. Its suitability for our purposes will be demonstrated by the continuity of the transformation in (3).

The sCGF estimation scheme we adopt for M was proposed by A. Dembo in a private communication to Duffield *et al.* [9]. The scheme is this: we select a block length $B \in \mathbb{N}$ sufficiently large that we believe the blocked sequence $\{Y_n, n \in \mathbb{N}\}$, where $Y_n := (X_{(n-1)B+1}, \tau_{(n-1)B+1}) + \cdots + (X_{nB}, \tau_{nB})$, can be treated as approximately i.i.d., and then use the estimator

$$M_n(\boldsymbol{\theta}) = \frac{1}{B} \log \frac{1}{n} \sum_{i=1}^{n} \exp(\langle \boldsymbol{\theta}, Y_i \rangle), \tag{4}$$

where $\boldsymbol{\theta} \in \mathbb{R}^2$ and $\langle \cdot, \cdot \rangle$ is the standard inner product. We propose taking the Legendre–Fenchel transform of M_n to form an estimate J_n of J, where

$$J_n(\boldsymbol{x}) = \sup_{\boldsymbol{\theta}}(\langle \boldsymbol{\theta}, \boldsymbol{x} \rangle - M_n(\boldsymbol{\theta})), \tag{5}$$

and then transforming J_n by (3) to form an estimate I_n of I, where

$$I_n(x) = \inf_{y > 0} y J_n\left(\frac{x}{y}, \frac{1}{y}\right). \tag{6}$$

In Theorem 3, below, the LDP for $\{I_n, n \in \mathbb{N}\}$ in conv$[a/\beta, b/\alpha]$ is established via the contraction principle. In Section 3, it is explained why this scheme can be preferable to the more direct approach of estimating I through the Legendre–Fenchel transform of estimates of λ. In particular, using this scheme if $\{A_t, t \in \mathbb{R}^+\}$ is a cumulative renewal process, we prove that the laws of $\{I_n, n \in \mathbb{N}\}$ converge weakly to a Dirac measure at I from (3), which is not typically the case using the direct method. In Section 4, we present a guide to related work and an indication of the practical significance of the proposed estimation scheme.

2. Main results

Let $\mathcal{M}_1([Ba, Bb] \times [B\alpha, B\beta])$ denote the space of probability measures on $[Ba, Bb] \times [B\alpha, B\beta]$ equipped with the τ topology. In [4], Bryc and Dembo introduced a mixing condition, condition (S), for stationary processes. For example, it is satisfied by hypercontractive Markov chains. If $\{(X_n, \tau_n), n \in \mathbb{N}\}$ satisfies (S) then, by inclusion of σ-algebras, the blocked process $\{Y_n, n \in \mathbb{N}\}$ also does. Under condition (S), Theorem 1 of [4] stated that the empirical laws

$$L_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i} \tag{7}$$

satisfy the LDP in $\mathcal{M}_1([Ba, Bb] \times [B\alpha, B\beta])$ with good, convex rate function H. The rate function H can be calculated by using the following representation. For $\nu \in \mathcal{M}_1([Ba, Bb] \times [B\alpha, B\beta])$,

$$H(\nu) = \sup_{f \in M([Ba,Bb] \times [B\alpha, B\beta], \mathbb{R})} \left\{ \int f \, \mathrm{d}\nu - \Lambda(f) \right\},$$

where $M([Ba, Bb] \times [B\alpha, B\beta], \mathbb{R})$ is the set of bounded, measurable functions from $[Ba, Bb] \times [B\alpha, B\beta]$ to $\mathbb{R}$ and

$$\Lambda(f) = \lim_{n \to \infty} \frac{1}{n} \log \mathrm{E} \left[\exp \left(\sum_{i=1}^{n} f(\mathbf{Y}_i) \right) \right].$$

For each $\nu \in \mathcal{M}_1([Ba, Bb] \times [B\alpha, B\beta])$, define M_ν by

$$M_\nu(\boldsymbol{\theta}) = \frac{1}{B} \log \int_{\mathbf{x} \in [Ba, Bb] \times [B\alpha, B\beta]} \mathrm{e}^{\langle \boldsymbol{\theta}, \mathbf{x} \rangle} \nu(\mathrm{d}\mathbf{x}),$$

and define J_ν to be the Legendre–Fenchel transform of M_ν. Note that when $\nu = L_n$, the empirical law of (7), then $J_\nu = J_n$, which is defined in (5). For a one-dimensional process, assuming that condition (S) holds, Duffy and Metcalfe [10] proved that rate function estimates satisfy the LDP. The arguments in [10] generalize verbatim to processes taking values in $\mathbb{R}^d$, for any finite d, giving us the following theorem.

Theorem 2. ([10, Theorem 1]) *If* $\{(X_n, \tau_n), n \in \mathbb{N}\}$ *satisfies condition (S) of [4], then the estimates* $\{J_n, n \in \mathbb{N}\}$ *satisfy the LDP in* $\mathrm{conv}([a, b] \times [\alpha, \beta])$ *with good rate function*

$$K(\phi) = \begin{cases} H(\nu) & \textit{if } \phi = J_\nu, \textit{ where } \nu \in \mathcal{M}_1([Ba, Bb] \times [B\alpha, B\beta]), \\ \infty & \textit{otherwise.} \end{cases}$$

The arguments presented in [10] to prove Theorem 2 rely on the assumption that the random variables are bounded. Based on evidence from numerical experiments, we conjecture that Theorem 2 also holds for (an appropriate class of) processes with unbounded random variables, but proving this is a mathematically challenging problem. As there is no constraint on the value of the bound, for practical applications we believe the boundedness assumption poses no serious difficulty.

In light of Theorem 2, in order to prove the LDP for $\{I_n, n \in \mathbb{N}\}$, where I_n is as defined in (6), it suffices to prove that the transformation (3) from $\mathrm{conv}([a, b] \times [\alpha, \beta]) \to \mathrm{conv}[a/\beta, b/\alpha]$ is continuous, and invoke the contraction principle (see [6, Theorem 4.2.1]). To prove continuity, first we prove a proposition and a corollary that provide a checkable condition for the continuity of a function $f : \mathrm{conv}(Z_1) \to \mathrm{conv}(Z_2)$, where (Z_1, d_1) and (Z_2, d_2) are metric spaces and both $\mathrm{conv}(Z_1)$ and $\mathrm{conv}(Z_2)$ are equipped with the Attouch–Wets topology. Proposition 1, below, which may be of independent interest, establishes conditions on a function $g : Z_2 \to Z_1$ which ensure that its pullback g^*, where $g^*(B_1) := \{x \in Z_2 : g(x) \in B_1\}$, is continuous in the Attouch–Wets topology.

Proposition 1. *Let* (Z_1, d_1) *and* (Z_2, d_2) *be metric spaces. Let* $g : Z_2 \to Z_1$ *be a bijective, continuous function that maps bounded sets to bounded sets and whose inverse* g^{-1} *is uniformly continuous on bounded sets. Then* $g^* : \mathrm{CL}(Z_1) \to \mathrm{CL}(Z_2)$ *is continuous.*

Proof. Let $A, A_1, A_2, \ldots$ denote a convergent sequence in $\mathrm{CL}(Z_1)$ with $A = \tau_{\mathrm{AW}d_1} - \lim_{n \to \infty} A_n$. Fix $\gamma > 0$, and let B_2 be a bounded subset of Z_2. As g maps bounded subsets

to bounded subsets, $g(B_2) := \{g(x) \colon x \in B_2\} =: B_1 \subset Z_1$ is bounded. Then, as $A = \tau_{\mathrm{AW}d_1} - \lim_{n\to\infty} A_n$, there exists an N such that $e_{d_1}(B_1 \cap A_n, A) < \gamma$ for $n > N$, so that $d_1(x, A) < \gamma$ for all $x \in B_1 \cap A_n$. Fix $n > N$ and $y \in B_2 \cap g^*(A_n)$. Then $y \in g^*(B_1) \cap g^*(A_n)$, and so $d_1(g(y), A) < \gamma$. Thus, since g is surjective, there must exist a $y' \in g^*(A)$ with $d_1(g(y), g(y')) < \gamma$.

Fix $\varepsilon > 0$. Since $B_2 \cap g^*(A_n)$ is bounded and g^{-1} is uniformly continuous on bounded subsets, we can choose γ sufficiently small that $d_2(y, y') < \varepsilon$. Thus, since $y' \in g^*(A)$, we have $d_2(y, g^*(A)) < \varepsilon$. This is true for all $y \in B_2 \cap g^*(A_n)$ with $n > N$, and so $e_{d_2}(B_2 \cap g^*(A_n), g^*(A)) < \varepsilon$. Thus, $\lim_{n\to\infty} e_{d_2}(B_2 \cap g^*(A_n), g^*(A)) = 0$. Similarly, $\lim_{n\to\infty} e_{d_2}(B_2 \cap g^*(A), g^*(A_n)) = 0$, and so $g^*(A) = \tau_{\mathrm{AW}d_2} - \lim_{n\to\infty} g^*(A_n)$.

The following corollary states, in particular, that if we wish to check the continuity of $f \colon \mathrm{conv}(Z_1) \to \mathrm{conv}(Z_2)$ and epi $f(\psi) = g^*(\mathrm{epi}\,\psi)$, then it suffices to check the conditions on g in Proposition 1.

Corollary 1. *Let (Z_1, d_1) and (Z_2, d_2) be metric spaces and let $f \colon \mathrm{conv}(Z_1) \to \mathrm{conv}(Z_2)$ be a function. Suppose that* epi $f(\psi) = g^*(\mathrm{epi}\,\psi)$ *for all $\psi \in \mathrm{conv}(Z_1)$, where $g \colon Z_2 \times \mathbb{R} \to Z_1 \times \mathbb{R}$ is a function that satisfies the conditions of Lemma 1. Then f is continuous.*

We prove that the transformation (3) is continuous by considering it as the composition of two maps. The first takes $J(x, y)$ to $y J(x/y, 1/y)$ and is proved via Corollary 1; the second takes $y J(x/y, 1/y)$ to $\inf_y y J(x/y, 1/y)$ and is proved by direct arguments.

Lemma 1. *For any $\phi \in \mathrm{conv}([a, b] \times [\alpha, \beta])$, define its transform $f_1(\phi)$ by*

$$(f_1(\phi))(x, y) := y\phi\left(\frac{x}{y}, \frac{1}{y}\right), \quad \text{for all } (x, y) \in \left[\frac{a}{\beta}, \frac{b}{\alpha}\right] \times \left[\frac{1}{\beta}, \frac{1}{\alpha}\right].$$

Then $f_1(\phi) \in \mathrm{conv}([a/\beta, b/\alpha] \times [1/\beta, 1/\alpha])$ and f_1 is continuous.

Proof. Let $\phi \in \mathrm{conv}([a, b] \times [\alpha, \beta])$ and fix $p \in [0, 1]$ and $(x_1, y_1), (x_2, y_2) \in [a, b] \times [\alpha, \beta]$. Consider $(f_1(\phi))(px_1 + (1 - p)x_2, py_1 + (1 - p)y_2)$. Noting that

$$\left(\frac{px_1 + (1 - p)x_2}{py_1 + (1 - p)y_2}, \frac{1}{py_1 + (1 - p)y_2}\right) = q\left(\frac{x_1}{y_1}, \frac{1}{y_1}\right) + (1 - q)\left(\frac{x_2}{y_2}, \frac{1}{y_2}\right)$$

and $q = py_1/(py_1 + (1 - p)y_2) \in [0, 1]$, as $y_1, y_2 > 0$, the convexity of $f_1(\phi)$ follows from the convexity of ϕ. To show that $f_1(\phi)$ is LSC, note that epi $f_1(\phi) = g^*(\mathrm{epi}\,\phi)$, where $g \colon [a/\beta, b/\alpha] \times [1/\beta, 1/\alpha] \times \mathbb{R} \to [a, b] \times [\alpha, \beta] \times \mathbb{R}$ is given by $g(x, y, z) = (x/y, 1/y, z/y)$. Since epi ϕ is closed and g is continuous, epi $f_1(\phi)$ is closed. Thus, $f_1(\phi)$ is LSC and $f_1(\phi) \in \mathrm{conv}([a/\beta, b/\alpha] \times [1/\beta, 1/\alpha])$.

The function g is bijective with inverse $g^{-1} \colon [a, b] \times [\alpha, \beta] \times \mathbb{R} \to [a/\beta, b/\alpha] \times [1/\beta, 1/\alpha] \times \mathbb{R}$ given by $g^{-1}(x, y, z) = (x/y, 1/y, z/y)$. Both g and g^{-1} are uniformly continuous on bounded sets. Therefore g satisfies the conditions of Proposition 1 and, thus, Corollary 1 implies that f_1 is continuous.

Lemma 2. *For any function $\phi \in \mathrm{conv}([a/\beta, b/\alpha] \times [1/\beta, 1/\alpha])$, define its transform $f_2(\phi)$ by*

$$(f_2(\phi))(x) := \inf_{y \in [1/\beta, 1/\alpha]} \phi(x, y), \quad \text{for all } x.$$

Then $f_2(\phi) \in \mathrm{conv}([a/\beta, b/\alpha])$ and f_2 is continuous.

Proof. As $\phi \in \text{conv}([a/\beta, b/\alpha] \times [1/\beta, 1/\alpha])$, its global infimum is attained, which ensures that the conditions of Corollary 3.32 of [25] hold and $f_2(\phi) \in \text{conv}([a/\beta, b/\alpha])$.

As ϕ is convex and LSC, it is finite on a convex closed subset of $[a/\beta, b/\alpha] \times [1/\beta, 1/\alpha]$ and is continuous on the interior of this set. Therefore, for any fixed x, $\inf_{y \in [1/\beta, 1/\alpha]} \phi(x, y)$ is attained at some point in $[1/\beta, 1/\alpha]$. Thus, $z \geq \inf_{y \in [1/\beta, 1/\alpha]} \phi(x, y)$ if and only if there exists a $y \in [1/\beta, 1/\alpha]$ with $z \geq \phi(x, y)$, i.e. $(x, z) \in \text{epi } f_2(\phi)$ if and only if there exists a $y \in [1/\beta, 1/\alpha]$ with $(x, y, z) \in \text{epi } \phi$.

Let $\phi, \phi_1, \phi_2, \ldots$ be a sequence in $\text{conv}([a/\beta, b/\alpha] \times [1/\beta, 1/\alpha])$. Let $B \subset [a/\beta, b/\alpha] \times \mathbb{R}$ be the bounded set $B = [a/\beta, b/\alpha] \times [-c, c]$. Fix n and $(x, z) \in B \cap \text{epi } f_2(\phi_n)$. Then there must exist a $y \in [1/\beta, 1/\alpha]$ with $(x, y, z) \in B' \cap \text{epi } \phi_n$, where $B' = [a/\beta, b/\alpha] \times [1/\beta, 1/\alpha] \times [-c, c]$. Thus, since $(x', z') \in \text{epi } f_2(\phi)$ if and only if there exists a y' with $(x', y', z') \in \text{epi } \phi$, and since $d((x, z), (x', z')) \leq d((x, y, z), (x', y', z'))$, we have

$$\inf_{(x', z') \in \text{epi } f_2(\phi)} d((x, z), (x', z')) \leq \inf_{(x', y', z') \in \text{epi } \phi} d((x, y, z), (x', y', z')),$$

that is,

$$d((x, z), \text{epi } f_2(\phi)) \leq d((x, y, z), \text{epi } \phi).$$

This is true for any $(x, z) \in B \cap \text{epi } f_2(\phi)$, and so

$$e_d(B \cap \text{epi } f_2(\phi_n), \text{epi } f_2(\phi)) \leq e_d(B' \cap \text{epi } \phi_n, \text{epi } \phi), \quad \text{for all } n.$$

Similarly, $e_d(B \cap \text{epi } f_2(\phi), \text{epi } f_2(\phi_n)) \leq e_d(B' \cap \text{epi } \phi, \text{epi } \phi_n)$ for all n. Thus, $\phi = \tau_{\text{AW}} - \lim_{n \to \infty} \phi_n$ implies that $f_2(\phi) = \tau_{\text{AW}} - \lim_{n \to \infty} f_2(\phi_n)$ and, so, f_2 is continuous.

The transformation in (3) is the composition $f = f_2 \circ f_1$, and so Lemmas 1 and 2 demonstrate its continuity. As we defined the estimates I_n of I in (5) by $I_n = f(J_n)$, and Theorem 2 proves that $\{J_n, n \in \mathbb{N}\}$ satisfies an LDP in $\text{conv}([a, b] \times [\alpha, \beta])$, our next theorem follows from an application of the contraction principle.

Theorem 3. *The sequence $\{I_n, n \in \mathbb{N}\}$ satisfies an LDP in $\text{conv}([a/\beta, b/\alpha])$ with good rate function*

$$K'(\phi) = \inf \left\{ K(\psi) : \psi \in \text{conv}([a, b] \times [\alpha, \beta]) \text{ and } \inf_{y > 0} y\psi\left(\frac{x}{y}, \frac{1}{y}\right) = \phi(x) \text{ for all } x \right\},$$

where K is as defined in the statement of Theorem 2.

3. The case of cumulative renewal processes

In Duffy and Metcalfe [10], it was shown that whenever the sequence $\{Y_n, n \in \mathbb{N}\}$ is i.i.d., the laws of the sequence of empirical estimates $\{J_n, n \in \mathbb{N}\}$ converge weakly to the Dirac measure δ_J. An analogous result holds for the sequence $\{I_n, n \in \mathbb{N}\}$ of estimates of I.

Theorem 4. *Assume that there exists a unique $\mu \in \mathcal{M}_1([Ba, Bb] \times [B\alpha, B\beta])$ with $H(\mu) = 0$. Then the sequence of laws of $\{I_n, n \in \mathbb{N}\}$ converges weakly to the Dirac measure $\delta_{f(J_\mu)}$.*
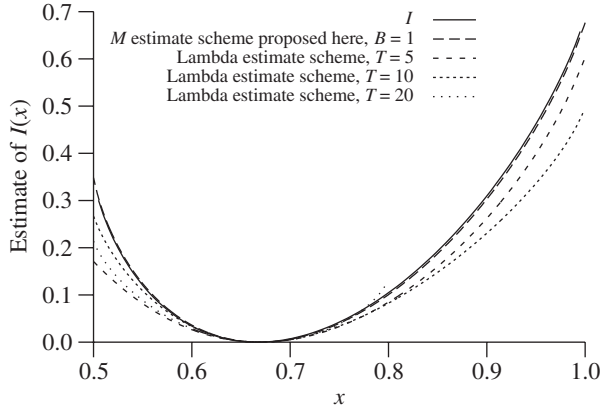
FIGURE 1: Estimating I via the scheme advocated in this paper with a block size of 1 and via the more direct scheme for a range of block sizes. Here, $X_n = 1$ for all n and $\{\tau_n, n \in \mathbb{N}\}$ is Bernoulli$\{1, 2\}$.

Proof. Clearly $K'(f(J_\mu)) = 0$. A simple adaption of Corollary 27.2.2 of [24] ensures that $K'(\psi) > 0$ for any $\psi \neq f(J_\mu)$. The result now follows from Theorems 2.1 and 2.2 of Lewis *et al.* [20].

Thus, for a cumulative renewal process, the blocked sequence $\{Y_n, n \in \mathbb{N}\}$ is i.i.d. when $B = 1$ and, by Sanov's theorem (see [6, Theorem 6.2.10]), the conditions of Theorem 4 are satisfied. In this case, H is the relative entropy $H(\nu) = H(\nu \mid \mu)$, where μ is the common distribution of $\{(X_n, \tau_n), n \in \mathbb{N}\}$. Thus, Theorem 4 implies that the laws of the sequence of estimates $\{I_n, n \in \mathbb{N}\}$ converge weakly to the Dirac measure $\delta_{f(J_\mu)} = \delta_{f(J)} = \delta_I$, which is clearly desirable.

The direct adaption of the estimation scheme from [9] to estimate I would be to select a time scale $T \in \mathbb{R}^+$ such that we believe $\{A_{Tn} - A_{T(n-1)}, n \in \mathbb{N}\}$ to form an i.i.d. sequence, estimate λ in (2) by

$$\lambda_t(\theta) = \frac{1}{T} \log \frac{1}{\lfloor t/T \rfloor} \sum_{i=1}^{\lfloor t/T \rfloor} \exp(\theta(A_{Ti} - A_{T(i-1)})),$$

where $\lfloor \cdot \rfloor$ is the integer-part function, and take its Legendre–Fenchel transform. However, if $\{(X_n, \tau_n), n \in \mathbb{N}\}$ is i.i.d., then $\{N_{Tn} - N_{T(n-1)}, n \in \mathbb{N}\}$ is typically not i.i.d. for any T, leading to bias in the estimate of λ. See [9] and [13] for a discussion on bias in sCGF estimation.

In the simplest scenario possible, we illustrate the advantage of using the apparently more complex approach proposed in this paper. If $X_i = 1$ with probability 1, then the relationship in (3) reduces to $I(x) = xJ(1, 1/x)$, relating the large deviations of $\{T_n/n, n \in \mathbb{N}\}$ and $\{N(t)/t, t \in \mathbb{R}^+\}$; a relationship known to hold for a broad class of processes (see [8], [11], [16], [23], and [26]). Setting $\{\tau_n, n \in \mathbb{N}\}$ to be i.i.d. and to take the value 1 or 2 each with probability $\frac{1}{2}$, one thousand (X_n, τ_n) pairs were generated and both schemes were implemented and run on the same data set. The results, shown in Figure 1, are typical for a broad class of cumulative renewal process. The real I was determined explicitly and is plotted for reference. Although the results of only a single data set are shown, they are representative. The scheme described in this paper nearly perfectly overlays I. The more direct scheme is shown for three block sizes of 5, 10, and 20. When the block size is 5 or 10, underestimation occurs. When the

block size is 20, the estimate does not span the whole x range of I, as the full range of possible $A_{Tn} - A_{T(n-1)}$ blocks has not been observed.

4. A rough guide to related work

The estimation scheme in (4) for a one-dimensional version of M was proposed by A. Dembo and used by Duffield *et al.* [9] for a problem in asynchronous transfer mode (ATM) networking. When combined with theorems of [17], it provided an online measurement-based mechanism for estimating the tail of queue-length distributions. For the success of this approach see, for example, [5] and [21].

In ATM networking, discrete-time models are appropriate. This makes the need for rate function estimators of continuous-time processes unnecessary. In IP (internet protocol) networks, which are more prevalent, it is more appropriate to model data traffic by a cumulative process, where X_n is a packet size and τ_n is an interpacket arrival time. Thus, to adapt for IP networking the successful connection admission control framework based on rate function estimation, we propose the estimator described in this paper.

For fixed θ, the estimator in (4) is known to be biased for correlated data, with the bias tending to zero as block size B increases. In [13], Ganesh analysed this bias for autoregressive and finite-state Markovian sources. He showed that it is of a similar form for both classes and demonstrated how to compensate for the bias.

If extra structure of the input process is known a priori, more accurate estimates can be made for $M(\theta)$ at fixed θ. If the source is known to be Markovian, for example, see [22], [12], and references therein.

In other analyses of the estimator (4), the existence of a B such that $\{Y_n, n \in \mathbb{N}\}$ is i.i.d. is usually assumed. See [18] for distribution-free confidence intervals for the measurement of $\lambda(\theta)$ for fixed θ. For a related question, in the Bayesian context, see [15], [14], and references therein. For a large deviations analysis of a connection admission control algorithm based on estimating sCGFs, see [7].

References

[1] ATTOUCH, H. AND WETS, R. J.-B. (1983). A convergence theory for saddle functions. *Trans. Amer. Math. Soc.* **280,** 1–41.

[2] ATTOUCH, H. AND WETS, R. J.-B. (1986). Isometries for the Legendre–Fenchel transform. *Trans. Amer. Math. Soc.* **296,** 33–60.

[3] BEER, G. (1993). *Topologies on Closed and Closed Convex Sets*. Kluwer, Dordrecht.

[4] BRYC, W. AND DEMBO, A. (1996). Large deviations and strong mixing. *Ann. Inst. H. Poincaré Prob. Statist.* **32,** 549–569.

[5] CROSBY, S. *et al.* (1997). Statistical properties of a near-optimal measurement-based CAC algorithm. In *Proc. IEEE ATM '97* (Lisbon, June 1997), pp. 103–112.

[6] DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviation Techniques and Applications* (Appl. Math. (NY) **38**), 2nd edn. Springer, New York.

[7] DUFFIELD, N. G. (2000). A large deviation analysis of errors in measurement based admission control to buffered and bufferless resources. *Queueing Systems Theory Appl.* **34,** 131–168.

[8] DUFFIELD, N. G. AND WHITT, W. (1998). Large deviations of inversely related processes with non-linear scalings. *Ann. Appl. Prob.* **8,** 995–1026.

[9] DUFFIELD, N. G. *et al.* (1995). Entropy of ATM traffic streams: a tool for estimating QoS parameters. *IEEE J. Selected Areas Commun.* **13,** 981–990.

[10] DUFFY, K. AND METCALFE, A. P. (2005). The large deviations of estimating rate functions. *J. Appl. Prob.* **42,** 267–274.

[11] DUFFY, K. AND RODGERS-LEE, M. (2004). Some useful functions for functional large deviations. *Stoch. Stoch. Reports* **76,** 267–279.

[12] EICHELSBACHER, P. AND GANESH, A. (2002). Bayesian inference for Markov chains. *J. Appl. Prob.* **39,** 91–99.

[13] GANESH, A. (1996). Bias correction in effective bandwidth estimation. *Performance Evaluation* **27,** 319–330.

[14] GANESH, A. AND O'CONNELL, N. (1999). An inverse of Sanov's theorem. *Statist. Prob. Lett.* **42,** 201–206.

[15] GANESH, A., GREEN, P., O'CONNELL, N. AND PITTS, S. (1998). Bayesian network management. *Queueing Systems Theory Appl.* **28,** 267–282.

[16] GLYNN, P. W. AND WHITT, W. (1994). Large deviations behaviour of counting processes and their inverses. *Queueing Systems* **17,** 107–128.

[17] 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**), Applied Probability Trust, Sheffield, pp. 413–430.

[18] GYÖRFI, L. *et al.* (2000). Distribution-free confidence intervals for measurement of effective bandwidths. *J. Appl. Prob.* **37,** 1–12.

[19] KUCZEK, T. AND CRANK, K. N. (1991). A large-deviation result for regenerative processes. *J. Theoret. Prob.* **4,** 551–561.

[20] LEWIS, J. T., PFISTER, C.-E. AND SULLIVAN, W. G. (1995). Entropy, concentration of probability and conditional limit theorems. *Markov Process. Relat. Fields* **1,** 319–386.

[21] LEWIS, J. T. *et al.* (1998). Practical connection admission control for ATM networks based on on-line measurements. *Comput. Commun.* **21,** 1585–1596.

[22] PASCHALIDIS, I. CH. AND VASSILARAS, S. (2001). On the estimation of buffer overflow probabilities from measurements. *IEEE Trans. Inf. Theory* **47,** 178–191.

[23] PUHALSKII, A. AND WHITT, W. (1997). Functional large deviation principles for first-passage-time processes. *Ann. Appl. Prob.* **7,** 362–381.

[24] ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton University Press.

[25] ROCKAFELLAR, R. T. AND WETS, R. J.-B. (1998). *Variational Analysis*. Springer, Berlin.

[26] RUSSELL, R. (1997). The large deviations of random time changes. Doctoral Thesis, Trinity College Dublin.