J. Appl. Prob. **42,** 267–274 (2005)
Printed in Israel

THE LARGE DEVIATIONS OF ESTIMATING RATE FUNCTIONS

KEN DUFFY,* *National University of Ireland, Maynooth*
ANTHONY P. METCALFE,** *Trinity College Dublin*

Abstract

Given a sequence of bounded random variables that satisfies a well-known mixing condition, it is shown that empirical estimates of the rate function for the partial sums process satisfy the large deviation principle in the space of convex functions equipped with the Attouch–Wets topology. As an application, a large deviation principle for estimating the exponent in the tail of the queue length distribution at a single-server queue with infinite waiting space is proved.

Keywords: Large deviation estimate; queue length tail estimate

2000 Mathematics Subject Classification: Primary 60F10
Secondary 60K25

Dedicated to John T. Lewis (1932–2004)

1. Introduction

Let $\{X_n, n \geq 1\}$ be a stationary process whose random variables take values in a bounded subset of $\mathbb{R}$. Define the partial sums process $\{S_n, n \geq 1\}$ by $S_n := X_1 + \cdots + X_n$ and assume that $\{S_n/n, n \geq 1\}$ satisfies the large deviation principle (LDP) (on the scale $1/n$) with a rate function I that is the Legendre–Fenchel transform of the scaled cumulant generating function (SCGF)

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

If we are given observations $X_1, X_2, \ldots,$ but the statistics of the process $\{X_n, n \geq 1\}$ are unknown, how would we estimate the rate function I? One way is to form an estimate of λ and take its Legendre–Fenchel transform.

A scheme for estimating λ was proposed to Neil O'Connell by Amir Dembo (private communication). The scheme is described by Duffield *et al.* [10], who used it for a problem in ATM networking, where, when combined with theorems of Glynn and Whitt [15], it provided an online measurement-based mechanism for estimating the tail of queue length distributions. For the success of this approach see, for example, [6] and [18].

The scheme is this: we select a block length b sufficiently large that we believe the blocked sequence $\{Y_n, n \geq 1\}$, where $Y_n := X_{(n-1)b+1} + \cdots + X_{nb}$, can be treated as being independent

Received 22 June 2004; revision received 10 September 2004.
* Postal address: Hamilton Institute, National University of Ireland, Maynooth, County Kildare, Ireland.
Email address: ken.duffy@nuim.ie
** Postal address: Department of Pure and Applied Mathematics, Trinity College Dublin, Ireland.

and identically distributed (i.i.d.), and then use the empirical estimator

$$\lambda_n(\theta) = \frac{1}{b} \log \frac{1}{n} \sum_{i=1}^{n} \exp(\theta Y_i). \tag{2}$$

After estimating λ, we propose taking its Legendre–Fenchel transform to form an estimate I_n of I. We will call both λ_n and I_n empirical estimates. The purpose of this note is to consider the large deviations in estimating λ and I when the empirical laws of $\{Y_n, n \geq 1\}$ satisfy the LDP. A sufficient condition for our theorems to hold is that $\{X_n, n \geq 1\}$ satisfy the mixing condition (S) of [5].

In Section 2, the LDP is proved for empirical estimators. As the random variables $\{Y_n, n \geq 1\}$ are assumed to be bounded, for SCGF estimates the topology of uniform convergence on compact subsets is natural, but it is not appropriate when estimates of a rate function are considered. For example, it is reasonable to desire that the rate functions $I_n(x) := n|x|$ converge to $I(x)$, which is 0 at $x = 0$ and ∞ otherwise. Clearly, this is not the case in the topology of uniform convergence on bounded subsets, but it is in the Attouch–Wets topology.

For rate functions, we consider the space of lower-semicontinuous convex functions equipped with the Attouch–Wets topology [1], [2], which we denote by τ_{AW}. A sequence $\{f_n, n \geq 1\}$ converges to f in τ_{AW}, i.e. $\tau_{\mathrm{AW}} - \lim f_n = f$, if, given any bounded set $A \in \mathbb{R} \times \mathbb{R}$ and any $\varepsilon > 0$, there exists an N_ε such that

$$\sup_{x \in A} |d(x, \operatorname{epi} f_n) - d(x, \operatorname{epi} f)| < \varepsilon \quad \text{for all } n > N_\varepsilon,$$

where $\operatorname{epi} f = \{(a, b)\colon\ b \geq f(a)\}$ denotes the epigraph of f and d is the Euclidean distance. A good reference for τ_{AW} is [3]. Another reason for choosing τ_{AW} is that the Legendre–Fenchel transform is continuous and, thus, the LDP for $\{I_n, n \geq 1\}$ can be deduced, by contraction, from the LDP for $\{\lambda_n, n \geq 1\}$.

In Section 3, as an application, the original motivation for the introduction of the estimator λ_n is considered. We prove the LDP for estimating the exponent in the tail of the queue length distribution at a single-server queue with infinite waiting space. In the simplest model, for Bernoulli random variables, it gives a serious warning: on the scale of large deviations, if the exponent is overestimated then it is likely to be extremely overestimated.

2. The large deviations of estimating rate functions

Let Σ be a closed, bounded subset of $\mathbb{R}$, and let $\mathcal{M}_1(\Sigma)$ denote the set of probability measures on Σ equipped with the weak topology induced by $C_b(\Sigma)$, the class of bounded, uniformly continuous functions from Σ to $\mathbb{R}$. With this topology, $\mathcal{M}_1(\Sigma)$ is Polish. Let $\operatorname{conv}(\mathbb{R})$ denote the set of $\mathbb{R}$-valued lower-semicontinuous convex functions over $\mathbb{R}$ equipped with the topology of uniform convergence on bounded subsets, and let $\operatorname{conv}(\Sigma)$ denote the set of $(\mathbb{R} \cup \{\infty\})$-valued lower-semicontinuous convex functions over the smallest closed interval containing Σ and equipped with τ_{AW}.

Given an element ν of $\mathcal{M}_1(\Sigma)$, we define its SCGF by

$$\lambda_\nu(\theta) := \frac{1}{b} \log \mathrm{E}_\nu[\mathrm{e}^{\theta x}] := \frac{1}{b} \log \int_\Sigma \mathrm{e}^{\theta x} \, \mathrm{d}\nu, \qquad \theta \in \mathbb{R},$$

and its rate function by

$$I_\nu(x) := \sup_{\theta \in \mathbb{R}} (\theta x - \lambda_\nu(\theta)).$$

The following assumption is in force from now on.

Assumption 1. *For fixed b, the blocked random variables* $\{Y_n, n \geq 1\}$ *take values in* Σ*, and the empirical laws* $\{L_n, n \geq 1\}$ *defined by*

$$L_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i}, \qquad n \geq 1,$$

satisfy the LDP in $\mathcal{M}_1(\Sigma)$ *with good rate function* H.

For an empirical law L_n, we define the empirical estimates $\lambda_n := \lambda_{L_n}$ and $I_n := I_{L_n}$. Note that λ_n thus defined agrees with the estimator in (2).

Theorem 1, below, can be paraphrased as follows: the large deviations of observing an empirical SCGF or rate function are just the large deviations of observing the empirical law that maps to them.

Theorem 1. (Empirical estimator LDP.) *The empirical estimators* $\{\lambda_n, n \geq 1\}$ *satisfy the LDP in* conv($\mathbb{R}$) *with the following good rate function:*

$$J(\phi) = \begin{cases} H(\nu) & \text{if } \phi = \lambda_\nu, \text{ where } \nu \in \mathcal{M}_1(\Sigma), \\ \infty & \text{otherwise}. \end{cases}$$

The empirical estimators $\{I_n, n \geq 1\}$ *satisfy the LDP in* conv(Σ) *with the following good rate function:*

$$K(\phi) = \begin{cases} H(\nu) & \text{if } \phi = I_\nu, \text{ where } \nu \in \mathcal{M}_1(\Sigma), \\ \infty & \text{otherwise}. \end{cases}$$

Proof. The first part follows by applying the contraction principle (see [7, Theorem 4.2.1]) and by the uniqueness of moment-generating functions (see, for example, [4]). Define the function $f\colon \mathcal{M}_1(\Sigma) \to \text{conv}(\mathbb{R})$ by $f(\nu) := \lambda_\nu$. Straightforward analysis shows that f is continuous. Let $\nu_n \to \nu$ in $\mathcal{M}_1(\Sigma)$. For fixed $\theta \in \mathbb{R}$, the function $x \mapsto \exp(\theta x)$ is an element of $C_b(\Sigma)$. Thus, $\nu_n(\exp(\theta x)) \to \nu(\exp(\theta x))$ and, as the logarithm is continuous, $f(\nu_n)(\theta) \to f(\nu)(\theta)$. However, $f(\nu_n)(\theta)$ is convex in θ, so pointwise convergence implies uniform convergence on bounded subsets.

As $f(\nu)(\theta)$ is real valued, [3, Lemma 7.1.2] ensures that $f(\nu_n) \to f(\nu)$ in conv($\mathbb{R}$) equipped with τ_{AW}. Thus, the second part follows by applying the contraction principle, since the Legendre–Fenchel transform from conv($\mathbb{R}$) to conv(Σ) is continuous (see [3]), and by the uniqueness of the Legendre–Fenchel transform.

Remark 1. A sufficient condition for Theorem 1 is that $\{X_n, n \geq 1\}$ satisfies the mixing condition (S) of [5]. This condition ensures that $\{S_n/n, n \geq 1\}$ satisfies the LDP with the good rate function given in (1). Moreover, by inclusion of σ-algebras, $\{Y_n, n \geq 1\}$ also satisfies (S) so that [5, Theorem 1] proves the LDP for $\{L_n, n \geq 1\}$ in the τ topology. As the τ topology is finer than the weak topology and the proof of Theorem 1 is by contraction, condition (S) suffices for it to hold.

If $\{Y_n, n \geq 1\}$ is genuinely i.i.d. with common law μ then, by Sanov's theorem, $H(\nu)$ is the relative entropy $H(\nu \mid \mu)$. As the relative entropy $H(\nu \mid \mu)$ has a unique zero at $\nu = \mu$, [17, Theorems 2.1 and 2.2] ensure that the laws of λ_n converge weakly to the Dirac measure at $\lambda_\mu = \lambda$, and the laws of I_n converge weakly to the Dirac measure at $I_\mu = I$.

https://doi.org/10.1239/jap/1110381386 Published online by Cambridge University Press

If $\{Y_n, n \geq 1\}$ is a Markov chain that satisfies the uniformity condition (U) of [8] then, by [8, Theorem 4.1.43 and Lemma 4.1.45], the good rate function H has a unique zero at the stationary distribution μ. Thus, the laws of λ_n converge weakly to the Dirac measure at λ_μ. This is obviously an issue if λ_μ and λ do not coincide, as can be seen in the following example.

Example 1. Let $\{X_n,\ n \geq 1\}$ be a Markov chain taking values $\{-1, +1\}$ with transition matrix

$$\pi = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \qquad \text{where } \alpha, \beta \in (0, 1). \tag{3}$$

Then λ can be calculated using techniques described in [7, Section 3.1]:

$$\lambda(\theta) = \log[\tfrac{1}{2}((1-\alpha)\mathrm{e}^{-\theta} + (1-\beta)\mathrm{e}^{\theta} + \sqrt{4\alpha\beta + ((1-\alpha)\mathrm{e}^{-\theta} - (1-\beta)\mathrm{e}^{\theta})^2})]. \tag{4}$$

If we choose $b = 1$ then $\{L_n, n \geq 1\}$ satisfies the LDP, and the laws of λ_n converge weakly to the Dirac measure at the SCGF of the stationary distribution

$$\log\biggl(\frac{\beta}{\alpha+\beta}\mathrm{e}^{-\theta} + \frac{\alpha}{\alpha+\beta}\mathrm{e}^{\theta}\biggr). \tag{5}$$

Note that (4) and (5) only agree if $\alpha + \beta = 1$, in which case the Markov chain is in fact Bernoulli.

For this Markov chain, the rate function H can be determined by simplifying the expression given in [8, Equation (4.1.38)]. It is finite if $\nu = (1-c)\delta_{-1} + c\delta_1$, in which case

$$H(\nu) = \begin{cases} -(1-c)\log(1-\alpha+\alpha K) - c\log\biggl(1-\beta+\dfrac{\beta}{K}\biggr) & \text{if } c \in [0, 1), \\ -\log(1-\beta) & \text{if } c = 1, \end{cases}$$

where

$$K = \frac{-\alpha\beta(1-2c) + \sqrt{(\alpha\beta(1-2c))^2 + 4\alpha\beta c(1-\alpha)(1-\beta)(1-c)}}{2\alpha(1-\beta)(1-c)}.$$

Thus, $J(\phi)$ is finite and is equal to $H(\nu)$ if $\phi = \lambda_\nu$, where $\lambda_\nu(\theta) = \log((1-c)\exp(-\theta) + c\exp(\theta))$, and $K(\phi)$ is finite and equal to $H(\nu)$ if $\phi = I_\nu$, where

$$I_\nu(x) = \tfrac{1}{2}(1-x)\log\biggl(\frac{1-x}{2(1-c)}\biggr) + \tfrac{1}{2}(x+1)\log\biggl(\frac{x+1}{2c}\biggr).$$

3. An application in queueing theory

Let X_n denote the difference, at time n, between the amount of work that arrives and the amount of work that can be processed at a discrete-time single-server queue with infinite buffer. Denote by Q_n the amount of work left to be processed by the server (the queue length) immediately after time n. The queue length evolves according to Lindley's recursion

$$Q_{n+1} = \max\{Q_n + X_{n+1}, 0\}, \tag{6}$$

where the maximum is necessary because the queue length cannot be negative. Assuming $\{X_n, n \geq 1\}$ to be stationary, the existence of a stationary solution to the recursion (6) is proved in the famous work of Loynes [19]. The distribution of each individual random variable in

https://doi.org/10.1239/jap/1110381386 Published online by Cambridge University Press

the solution is given by $Q := \max\{0, \sup_{t\geq 1} \sum_{i=1}^{t} X_i\}$. Alternatively, Q can be thought of as the supremum of a random walk, starting at 0, with increments process $\{X_n, n \geq 1\}$. Under our assumptions on $\{X_n, n \geq 1\}$, the distribution of Q has logarithmic asymptotics (see, for example, [11], [14], and [15])

$$\lim_{q\to\infty} \frac{1}{q} \log \mathrm{P}[Q > q] = -\delta,$$

where δ is determined by the large deviations rate function

$$\delta = \sup\{\theta : \lambda(\theta) \leq 0\} = \inf_{x>0} xI\left(\frac{1}{x}\right).$$

The great novelty of the approach of Duffield *et al.* [10] was to employ the following estimator for δ based on λ estimates: $\delta_n := \sup\{\theta : \lambda_n(\theta) \leq 0\}$. In [10] a central limit theorem for $\{\delta_n, n \geq 1\}$ is proved. Our aim is to prove the LDP, which we will do by contraction.

Lemma 1. *The function* $g : \operatorname{conv}(\mathbb{R}) \to [0, \infty) \cup \{\infty\}$ *defined by*

$$g(\phi) := \sup\{t \geq 0 : \phi(t) \leq 0\},$$

where the supremum over the empty set is defined to be 0*, is continuous at all* ϕ *such that* $\phi(0) = 0$ *and no* $\chi > 0$ *exists such that* $\phi(x) = 0$ *for all* $x \in [0, \chi]$.

Proof. Let $\phi_n \to \phi$ in $\operatorname{conv}(\mathbb{R})$. There are three cases to consider: $g(\phi) = \infty$, $0 < g(\phi) < \infty$, and $g(\phi) = 0$. First, assume that $g(\phi) = \infty$ and $\phi(t) < 0$ for all $t > 0$. Given $\alpha > 0$, let $0 < \varepsilon < -\phi(\alpha)$; then, as $\phi_n \to \phi$ uniformly on $[0, \alpha]$, there exists an N_ε such that $\phi_n(\alpha) < \phi(\alpha) + \varepsilon < 0$ for all $n > N_\varepsilon$. Thus, given $\alpha > 0$, there exists an N_ε such that $g(\phi_n) > \alpha$ for all $n > N_\varepsilon$.

Now, assume that $g(\phi) \in (0, \infty)$, let $g(\phi) > \varepsilon > 0$, and let

$$\gamma < \min(\phi(g(\phi) + \varepsilon), -\phi(g(\phi) - \varepsilon)).$$

As $\phi_n \to \phi$ uniformly on $[0, g(\phi) + \varepsilon]$, there exists an N_γ such that

$$\phi_n(g(\phi) - \varepsilon) < \phi(g(\phi) - \varepsilon) + \gamma < 0$$

and

$$\phi_n(g(\phi) + \varepsilon) > \phi(g(\phi) + \varepsilon) - \gamma > 0,$$

for all $n > N_\gamma$. Thus, $g(\phi_n) \in (g(\phi) - \varepsilon, g(\phi) + \varepsilon)$ for all $n > N_\gamma$.

Finally, assume that $g(\phi) = 0$. Given $\varepsilon > 0$, let $\phi(2\varepsilon) - \phi(\varepsilon) > 2\gamma > 0$. Then there exists an N_γ such that $|\phi_n(t) - \phi(t)| < \gamma$ for all $t \in [0, 2\varepsilon]$. Thus, $\phi_n(2\varepsilon) > \phi_n(\varepsilon) > 0$ for all $n > N_\gamma$ and, hence, $g(\phi_n) < \varepsilon$.

By a slightly more involved argument, which is similar in spirit, Lemma 1 is also true when $\operatorname{conv}(\mathbb{R})$ is equipped with τ_{AW}.

Remark 2. The function g has a discontinuity at $\phi(t) = 0$ for all t. This is an artifact of the estimation scheme rather than an issue with our choice of topology. For example, if $\lambda_n(\theta) = 0$ for all θ, then $Y_k = 0$, $k = 1, \ldots, n$, the queue appears perfectly balanced and, thus, $\delta_n = \infty$. However, in the closely similar situation where $Y_k = \varepsilon > 0$ for all k, the queue would appear overloaded, with $\delta_n = 0$.

https://doi.org/10.1239/jap/1110381386 Published online by Cambridge University Press

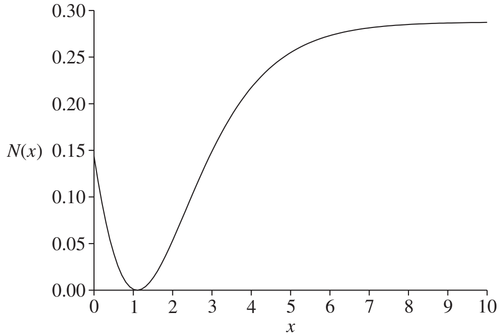

FIGURE 1: The rate function $N(x)$ for estimating the exponent in the tail of the queue length distribution. The arrivals minus potential-service is a Bernoulli process taking values in $\{-1, +1\}$ with mean $-\frac{1}{2}$. The rate function is 0 at the real value $\delta = \log(3)$.

In practice, this suggests that care must be taken with SCGF estimates around this discontinuity. For the theory, we introduce an additional assumption to avoid this discontinuity and so deduce the LDP: we assume that a small open ball around 0 is not contained in Σ.

Theorem 2. (Decay rate LDP.) *If* $(-\varepsilon, \varepsilon) \notin \Sigma$ *for some* $\varepsilon > 0$*, then the sequence* $\{\delta_n, n \geq 1\}$ *satisfies the LDP in* $[0, \infty]$ *with good rate function*

$$N(x) = \inf\{H(\nu) \colon \sup\{\theta \colon \lambda_\nu(\theta) \leq 0\} = x\}.$$

Proof. By Puhalskii's extension of the contraction principle (see [20, Theorem 3.1.14]), it suffices to have continuity at ϕ such that $J(\phi) < \infty$. As $(-\varepsilon, \varepsilon) \notin \Sigma$, $J(\lambda_\nu) = \infty$ for $\nu \in \mathcal{M}_1(\Sigma)$ if there exists a $\chi > 0$ such that $\lambda_\nu(\theta) = 0$ for $\theta \in [0, \chi]$. Thus, Lemma 1 ensures that g is sufficiently continuous for us to invoke the extended contraction principle from the LDP for $\{\lambda_n, n \geq 1\}$.

In the case where $\{X_n, n \geq 1\}$ is a Bernoulli sequence taking values in $\{-1, +1\}$ with $\mathrm{P}[X_n = 1] = p \in (0, 1)$, the rate function N in Theorem 2 can be calculated explicitly. For $\nu = (1-c)\delta_{-1} + c\delta_{+1}$,

$$H(c) := H(\nu \mid \mu) = c \log \frac{c}{p} + (1-c) \log \frac{1-c}{1-p},$$

and the rate function for $\{\delta_n, n \geq 1\}$ is

$$N(x) = \begin{cases} H\left(\dfrac{1}{1+\exp(x)}\right) & \text{if } x > 0, \\ H\left(\frac{1}{2}\right) & \text{if } x = 0 \text{ and } p \leq \frac{1}{2}, \\ 0 & \text{if } x = 0 \text{ and } p > \frac{1}{2}. \end{cases} \tag{7}$$

This presents a serious warning: although in [10] it was shown that $\{\delta_n, n \geq 1\}$ obeys a central limit theorem, (7) implies that when there is an overestimate of δ, it is likely to be a large overestimate. To see this, consider Figure 1, where the rate function for estimating δ for Bernoulli random variables with $p = \frac{1}{4}$ is plotted. Overestimation of δ is a serious issue, as it corresponds to underestimation of the likelihood of long queues.

For correlated processes $\{X_n, n \geq 1\}$, the block length b also causes problems. Consider a Markov chain on $\{-1, +1\}$ with transition matrix given by (3). If $\alpha < \beta$ then $\delta = \log((1-\alpha)/(1-\beta))$ but, with block length $b = 1$, the laws of δ_n converge weakly to the Dirac measure at $\log(\beta/\alpha)$. In accordance with intuition, if $\alpha + \beta < 1$ the chain is positively correlated and the weak law will be for an overestimate of δ; while, if $\alpha + \beta > 1$, the chain is negatively correlated and the weak law will be for an underestimate of δ.

4. Related work

In other analyses utilizing this estimator, the existence of b such that $\{Y_n, n \geq 1\}$ is genuinely i.i.d. is usually assumed. See [16] for distribution-free confidence intervals for measurement of $\lambda(\theta)$ for fixed θ. For a related question, in the Bayesian context, see [12], [13], and references therein. For a large deviations analysis of a connection admission control algorithm based on estimating SCGFs, see [9].

Acknowledgements

Ken Duffy is on secondment from the Dublin Institute of Technology. His work is supported by Science Foundation Ireland under the National Development Plan. Anthony P. Metcalfe's work is supported by an IRCSET scholarship.

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* (Math. Appl. **268**). Kluwer, Dordrecht.

[4] Billingsley, P. (1995). *Probability and Measure*. John Wiley, New York.

[5] Bryc, W. and Dembo, A. (1996). Large deviations and strong mixing. *Ann. Inst. H. Poincaré Prob. Statist.* **32,** 549–569.

[6] Crosby, S. *et al.* (1997). Statistical properties of a near-optimal measurement-based CAC algorithm. In *Proc. IEEE ATM* (Lisbon, 1997), eds A. Casaca *et al.*, IEEE Press, Piscataway, NJ, pp. 103–112.

[7] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications* (Appl. Math. **38**). Springer, New York.

[8] Deuschel, J.-D. and Stroock, D. W. (1989). *Large Deviations* (Pure Appl. Math. **137**). Academic Press, Boston, MA.

[9] 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.

[10] 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.

[11] Duffy, K., Lewis, J. T. and Sullivan, W. G. (2003). Logarithmic asymptotics for the supremum of a stochastic process. *Ann. Appl. Prob.* **13,** 430–445.

[12] Ganesh, A. and O'Connell, N. (1999). An inverse of Sanov's theorem. *Statist. Prob. Lett.* **42,** 201–206.

[13] Ganesh, A., Green, P., O'Connell, N. and Pitts, S. (1998). Bayesian network management. *Queueing Systems Theory Appl.* **28,** 267–282.

[14] Ganesh, A. J. and O'Connell, N. (2002). A large deviation principle with queueing applications. *Stoch. Stoch. Reports* **73,** 25–35.

[15] Glynn, P. and Whitt, W. (1994). Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. In *Studies in Applied Probability* (J. Appl. Prob. Special Vol. **31A**), Applied Probability Trust, Sheffield, pp. 413–430.

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

[17] 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.

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

[19] Loynes, R. M. (1962). The stability of a queue with non-independent inter-arrival and service times. *Proc. Camb. Phil. Soc.* **58,** 497–520.

[20] Puhalskii, A. (2001). *Large Deviations and Idempotent Probability* (Chapman and Hall / CRC Monogr. Surveys Pure Appl. Math. **119**). Chapman and Hall / CRC, Boca Raton, FL.

https://doi.org/10.1239/jap/1110381386 Published online by Cambridge University Press