*Applied Probability Trust* (10 September 2004)

# THE LARGE DEVIATIONS OF ESTIMATING RATE-FUNCTIONS

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

ANTHONY P. METCALFE,** *Trinity College, Dublin*

**Dedicated to John T. Lewis [1932-2004]**

*Keywords:* Estimating Large Deviations, Estimating Queue-Length Tails

AMS 2000 Subject Classification: Primary 60F10

Secondary 60K25

**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 satisfies 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.

## 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 $\{S_n/n, n \geq 1\}$ satisfies the Large Deviation Principle (LDP) (on the scale $1/n$) with rate-function $I$ that is the Legendre-Fenchel transform of the scaled cumulant generating function (sCGF)

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

If we are given an observation $X_1, X_2, \ldots$, 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 $\lambda$ and take its Legendre-Fenchel transform.

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

** Postal address: Department of Pure and Applied Mathematics, Trinity College, Dublin, Ireland

A scheme for estimating $\lambda$ was proposed by Amir Dembo in a private communication to Neil O'Connell. The scheme is described by Duffield et al. [10] who used it for a problem in ATM networks 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, Crosby et al. [6] and Lewis et al. [18].

Their scheme is this: select a block-length $b$ sufficiently large that you believe the blocked sequence $\{Y_n, n \geq 1\}$, where $Y_n := X_{(n-1)b+1} + \cdots + X_{nb}$, can be treated as i.i.d; 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 $\lambda$, we propose taking its Legendre-Fenchel transform to form an estimate $I_n$ of $I$. We will call both $\lambda_n$ and $I_n$ empirical estimates. The purpose of this note is to consider the large deviations of estimating $\lambda$ and $I$ when the empirical laws of $\{Y_n, n \geq 1\}$ satisfy the LDP. A sufficient condition for our theorems to hold is $\{X_n, n \geq 1\}$ satisfy the mixing condition (S) of Bryc and Dembo [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 one considers estimates of a rate-function. 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 $+\infty$ otherwise. Clearly this is not the case in the topology of uniform convergence on bounded subsets, but it is in Attouch-Wets topology.

For rate-functions we consider the space of lower semi-continuous convex functions equipped with the Attouch-Wets topology [1, 2], denoted $\tau_{\text{AW}}$. A sequence $\{f_n, n \geq 1\}$ converges to $f$ in $\tau_{\text{AW}}$, $\tau_{\text{AW}} - \lim f_n = f$, if given any bounded set $A \in \mathbb{R} \times \mathbb{R}$ and any $\epsilon > 0$, there exists $N_\epsilon$ such that

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

where epi $f = \{(a, b) : b \geq f(a)\}$, the epigraph of $f$, and $d$ is the Euclidean distance. A good reference for $\tau_{\text{AW}}$ is Beer [3]. Another reason for choosing $\tau_{\text{AW}}$ 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 $\lambda_n$ 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 one over-estimates the exponent, one is likely to extremely over-estimate it.

## 2. The large deviations of estimating rate-functions

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

Given an element $\nu$ of $\mathcal{M}_1(\Sigma)$ we define its sCGF by

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

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 here on.

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

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

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

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

Paraphrasing the following theorem: the large deviations of observing an empirical sCGF or rate-function is 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* $\text{Conv}(\mathbb{R})$ *with 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* $\text{Conv}(\Sigma)$ *with 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 applying the contraction principle (see Theorem 4.2.1 of [7]) and by the uniqueness of moment generating functions (see, for example, [4]). Define the function $f : \mathcal{M}_1(\Sigma)$ 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 log is continuous $f(\nu_n)(\theta) \to f(\nu)(\theta)$. But $f(\nu_n)(\theta)$ is convex in $\theta$ so that pointwise convergence implies uniform convergence on bounded subsets.

As $f(\nu)(\theta)$ is real-valued, Lemma 7.1.2 of [3] ensures that $f(\nu_n) \to f(\nu)$ in $\text{Conv}(\mathbb{R})$ equipped with $\tau_{\text{AW}}$. Thus the second part follows applying the contraction principle as the Legendre-Fenchel transform from $\text{Conv}(\mathbb{R})$ to $\text{Conv}(\Sigma)$ is continuous (see Beer [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 Bryc and Dembo [5]. This condition ensures that $\{S_n/n, n \geq 1\}$ satisfies the LDP with good rate-function given in equation (1). Moreover, by inclusion of $\sigma$-algebras, $\{Y_n, n \geq 1\}$ also satisfies (S) so that Theorem 1 of [5] proves the LDP for $\{L_n, n \geq 1\}$ in the $\tau$ topology. As the $\tau$ 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 $\mu$, then by Sanov's theorem $H(\nu)$ is the relative entropy $H(\nu|\mu)$. As the relative entropy $H(\nu|\mu)$ has unique zero at

$\nu = \mu$, Theorems 2.1 and 2.2 of Lewis et al. [17] ensure that the laws of $\lambda_n$ converge weakly to the Dirac measure at $\lambda_\mu = \lambda$ and the laws of $I_n$ converge weakly to the Dirac measure at $I_\mu = I$.

If $\{Y_n, n \geq 1\}$ is a Markov chain that satisfies the uniformity condition (U) of Deuschel and Stroock [8], then by Theorem 4.1.43 and Lemma 4.1.45 of [8] the good rate-function $H$ has unique zero at the stationary distribution $\mu$. Thus the laws of $\lambda_n$ converge weakly to the Dirac measure at $\lambda_\mu$. This is obviously an issue if $\lambda_\mu$ and $\lambda$ do not coincide, as can be seen in the following example: let $\{X_n, n \geq 1\}$ be a Markov chain taking values $\{-1, +1\}$ with transition matrix

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

Then $\lambda$ can be calculated using techniques described in section 3.1 of Dembo and Zeitouni [7]:

$$\lambda(\theta) = \log\left(\frac{(1-\alpha)e^{-\theta} + (1-\beta)e^{\theta} + \sqrt{4\alpha\beta + ((1-\alpha)e^{-\theta} - (1-\beta)e^{\theta})^2)}}{2}\right). \tag{4}$$

Choosing $b = 1$, $\{L_n, n \geq 1\}$ satisfies the LDP and the laws of $\lambda_n$ converge weakly to the Dirac measure at the sCGF of the stationary distribution:

$$\log\left(\frac{\beta}{\alpha+\beta}e^{-\theta} + \frac{\alpha}{\alpha+\beta}e^{\theta}\right). \tag{5}$$

Note that equations (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 equation (4.1.38) of [8]. It is finite if $\nu = (1-c)\delta_{-1} + c\delta_1$, in which case

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

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 equals $H(\nu)$ if $\phi = \lambda_\nu$ where $\lambda_\nu(\theta) = \log((1-c)\exp(-\theta) + c\exp(\theta))$ and $K(\phi)$ is finite and equals $H(\nu)$ if $\phi = I_\nu$ where

$$I_\nu(x) = \frac{(1-x)}{2}\log\left(\frac{1-x}{2(1-c)}\right) + \frac{x+1}{2}\log\left(\frac{x+1}{2c}\right).$$

## 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 as the queue-length cannot be negative. Assuming $\{X_n, n \geq 1\}$ to be stationary, in famous work of Loynes [19] the existence of a stationary solution to the recursion (6) is proved. The distribution of each individual random variable in 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 has logarithmic asymptotics (for example, see [15, 14, 11]):

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

where $\delta$ is determined by the large deviations rate-function

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

The great novelty of the approach of Duffield et al. [10] was to employ the following estimator for $\delta$ based on $\lambda$ 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. We do so by contraction.

With a slightly more involved argument that is similar in spirit, the following Lemma is also true when $\mathrm{Conv}(\mathbb{R})$ is equipped with $\tau_{\mathrm{AW}}$.

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

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

*where the supremum over the empty set is defined to be zero, is continuous at all* $\phi$ *such that* $\phi(0) = 0$ *and there does not exist* $\chi > 0$ *such that* $\phi(x) = 0$ *for all* $x \in [0, \chi]$.

*Proof.* Let $\phi_n \to \phi$ in $\text{Conv}(\mathbb{R})$. There are three cases to consider: $g(\phi) = +\infty$; $0 < g(\phi) < \infty$; and $g(\phi) = 0$.

Assuming $g(\phi) = +\infty$, $\phi(t) < 0$ for all $t > 0$. Given $\alpha > 0$, let $0 < \epsilon < -\phi(\alpha)$, then as $\phi_n \to \phi$ uniformly on $[0, \alpha]$, there exists $N_\epsilon$ such that for all $n > N_\epsilon$, $\phi_n(\alpha) < \phi(\alpha) + \epsilon < 0$. Thus given $\alpha > 0$ there exists $N_\epsilon$ such that $g(\phi_n) > \alpha$ for all $n > N_\epsilon$.

Assume $g(\phi) \in (0, \infty)$, let $g(\phi) > \epsilon > 0$ and let $\gamma < \min(\phi(g(\phi) + \epsilon), -\phi(g(\phi) - \epsilon))$. As $\phi_n \to \phi$ uniformly on $[0, g(\phi) + \epsilon]$, there exists $N_\gamma$ such that, for all $n > N_\gamma$, $\phi_n(g(\phi) - \epsilon) < \phi(g(\phi) - \epsilon) + \gamma < 0$ and $\phi_n(g(\phi) + \epsilon) > \phi(g(\phi) + \epsilon) - \gamma > 0$. Thus $g(\phi_n) \in (g(\phi) - \epsilon, g(\phi) + \epsilon))$ for all $n > N_\gamma$.

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

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

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

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

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

*Proof.* By Puhalskii's extension of the contraction principle (Theorem 3.1.14 of [20]), it suffices to have continuity at $\phi$ such that $J(\phi) < \infty$. As $(-\epsilon, \epsilon) \notin \Sigma$, for $\nu \in \mathcal{M}_1(\Sigma)$, $J(\lambda_\nu) = +\infty$ if there exists $\chi > 0$ such that $\lambda_\nu(\theta) = 0$ for $\theta \in [0, \chi]$. Thus Lemma 1 ensures $g$ is sufficiently continuous 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 $\mathbb{P}[X_n = 1] = p \in (0, 1)$, the rate-function $N$ in Theorem 2 can be calculated explicitly. For $\nu = (1-c)\delta_{-1} + c\delta_{+1}$,

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

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

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

This gives a serious warning: although in [10] it was shown that $\{\delta_n, n \geq 1\}$ obeys a central limit theorem, equation (7) says that when there is an over-estimate of $\delta$, it is likely to be a large over-estimate. To see this, observe Figure 1 where the rate-function for estimating $\delta$ for Bernoulli random variables with $p = 1/4$ is plotted. Over-estimation of $\delta$ is a serious issue, as it corresponds to under-estimation 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 in equation (3). With $\alpha < \beta$, $\delta = \log((1-\alpha)/(1-\beta))$, but with block-length $b = 1$ the laws of $\delta_n$ converge weakly to the Dirac measure at $\log(\beta/\alpha)$. Matching with intuition, if $\alpha + \beta < 1$, the chain is positively correlated and the weak-law will be for an over-estimate of $\delta$; if $\alpha + \beta > 1$, the chain is negatively correlated and the weak-law will be for an under-estimate for $\delta$.

## 4. Related work

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

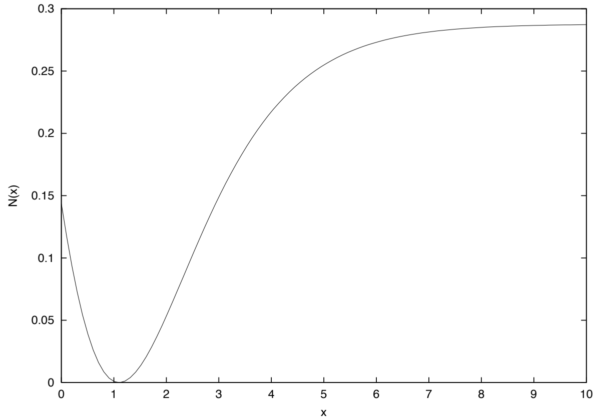


FIGURE 1: The rate-function $N(x)$ for estimating the exponent in the tail of the queue-length distribution. The arrivals less potential service is a Bernoulli process taking values in $\{-1, +1\}$ with mean $-1/2$. The rate-function is zero at the real value $\delta = \log(3)$.

## 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] H. Attouch and R.J.-B. Wets, *A convergence theory for saddle functions*, Trans. Amer. Math. Soc. **280** (1983), no. 1, 1–41.

[2] ______, *Isometries for the Legendre-Fenchel transform*, Trans. Amer. Math. Soc. **296** (1986), no. 1, 33–60.

[3] Gerald Beer, *Topologies on closed and closed convex sets*, Kluwer Academic, 1993.

[4] Patrick Billingsley, *Probability and measure*, John Willey & Sons, 1995.

[5] W. Bryc and A. Dembo, *Large deviations and strong mixing*, Ann. Inst. H. Poincaré Probab. Statist. **32** (1996), no. 4, 549–569.

[6] S. Crosby, I. Leslie, B. McGurk, J.T. Lewis, R. Russell, and F. Toomey, *Statistical properties of a near-optimal measurement-based CAC algorithm*, Proceedings of IEEE ATM '97 (Lisbon, Portugal), 1997.

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

[8] J-D. Deuschel and D.W. Stroock, *Large deviations*, Academic Press, 1989.

[9] N. G. Duffield, *A large deviation analysis of errors in measurement based admission control to buffered and bufferless resources*, Queueing Systems Theory Appl. **34** (2000), no. 1-4, 131–168.

[10] N. G. Duffield, J.T. Lewis, N. O'Connell, R. Russell, and F. Toomey, *Entropy of ATM traffic streams: a tool for estimating QoS parameters*, IEEE Journal of Selected Areas in Communications (special issue on Advances in the Fundamentals of Networking) **13** (1995), 981–990.

[11] K. Duffy, J.T. Lewis, and W.G. Sullivan, *Logarithmic asymptotics for the supremum of a stochastic process*, Ann. Appl. Probab. **13** (2003), no. 2, 430–445.

[12] A. Ganesh, P. Green, N. O'Connell, and S. Pitts, *Bayesian network management*, Queueing Systems Theory Appl. **28** (1998), no. 1-3, 267–282.

[13] A. Ganesh and N. O'Connell, *An inverse of Sanov's theorem*, Statist. Probab. Lett. **42** (1999), no. 2, 201–206.

[14] A. J. Ganesh and N. O'Connell, *A large deviation principle with queueing applications*, Stoch. Stoch. Rep. **73** (2002), no. 1-2, 25–35.

[15] P. Glynn and W. Whitt, *Logarithmic asymptotics for steady-state tail probabilities in a single-server queue*, J. Appl. Probab. **31A** (1994), 413–430.

[16] L. Györfi, A. Rácz, K. Duffy, J.T. Lewis, and F. Toomey, *Distribution-free confidence intervals for measurement of effective bandwidths*, J. Appl. Probab. **37** (2000), 1–12.

[17] J.T. Lewis, C.-E. Pfister, and W.G. Sullivan, *Entropy, concentration of probability and conditional limit theorems*, Markov Process. Related Fields **1** (1995), 319–386.

[18] J.T. Lewis, R. Russell, F. Toomey, B. McGurk, S. Crosby, and I. Leslie, *Practical connection admission control for ATM networks based on on-line measurements*, Computer Communications **21** (1998), no. 17, 1585–1596.

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

[20] A. Puhalskii, *Large deviation and idempotent probability*, Chapman & Hall/CRC, 2001.