LARGE DEVIATION ASYMPTOTICS FOR BUSY PERIODS

By Ken R. Duffy[*] and Sean P. Meyn[‡,†]

National University of Ireland Maynooth[*] and University of Florida[†]

The busy period for a queue is cast as the area swept under the random walk until it first returns to zero. Encompassing non-i.i.d. increments, the large-deviations asymptotics of the busy period B is addressed, under the assumption that the increments satisfy standard conditions, including a negative drift. The main conclusions provide insight on the probability of a large busy period, and the manner in which this occurs. The scaled probability of a large busy period has the asymptote, for any $b > 0$,

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}} \log P(B \geq bn) = -K\sqrt{b},$$

where $\quad K = 2\sqrt{-\int_0^{\lambda^*} \Lambda(\theta)\, d\theta}, \quad$ with $\lambda^* = \sup\{\theta : \Lambda(\theta) \leq 0\}$,

and with Λ denoting the scaled cumulant generating function of the increments process. The most likely path to a large swept area is found to be a simple rescaling of the path on $[0, 1]$ given by

$$\psi^*(t) = -\Lambda(\lambda^*(1 - t))/\lambda^*.$$

In contrast to the piecewise linear most likely path leading the random walk to hit a high level, this is strictly concave in general. While these two most likely paths have distinctly different forms, their derivatives coincide at the start of their trajectories, and at their first return to zero.

These results partially answer an open problem of Kulick and Palmowski [18] regarding the tail of the work done during a busy period at a single server queue. The paper concludes with applications of these results to the estimation of the busy period statistics (λ^*, K) based on observations of the increments, offering the possibility of estimating the likelihood of a large busy period in advance of observing one.

1. Introduction. Consider $\boldsymbol{S} = \{S_k : k \geq 0\}$, a random walk that starts at zero and has (not necessarily i.i.d.) increments process $\boldsymbol{X} = \{X_n : n \geq 0\}$:

$$(1) \qquad S_0 := 0 \text{ and } S_k = \sum_{i=1}^{k} X_i \text{ for } k \geq 1.$$

Define the stopping time and stopped variable:

$$(2) \qquad \tau := \inf\{k \geq 1 : S_k \leq 0\} \text{ and } B = \sum_{i=1}^{\tau} S_i.$$

The tail behavior of B (and related random variables) is of interest in several apparently distinct fields from queueing systems, to percolation, to insurance [18, 21, 12].

By simple rescaling arguments, should the following limit exist, it must have this form:

$$(3) \qquad \lim_{n \to \infty} \frac{1}{\sqrt{n}} \log P(B \geq nb) = -K\sqrt{b} \text{ for some } K \geq 0.$$

This limit is established with $K > 0$ in [2] for the particular case of the the the M/M/1 queue. By extending results in [12] from a fixed terminal point to the random terminal point τ via the infinite time-horizon sample path Large Deviation Principle setup in [14], Theorem 2 establishes the limit eq. (3) for a broad class of non-long-range dependent, light-tailed arrivals processes, providing a formula for K. The scaled Cumulant Generating Function (sCGF) associated with the scalar LDP is denoted

$$(4) \qquad \Lambda(\theta) = \lim_{k \to \infty} \frac{1}{k} \log E(\exp(\theta S_k)) \text{ for } \theta \in \mathbb{R}.$$

Identifying $\lambda^* = \sup\{\theta : \Lambda(\theta) \leq 0\}$, Theorem 2 shows that under Assumptions 1 and 2,

$$(5) \qquad K = 2\sqrt{-\int_0^{\lambda^*} \Lambda(\theta)\, d\theta}.$$

Its proof establishes that the most likely path to a large swept area is strictly concave. The most likely path that first returns to zero at $t = 1$ is identified to be

$$(6) \qquad \psi^*(t) = -\frac{1}{\lambda^*}\Lambda(\lambda^*(1 - t)) \text{ for } t \in [0, 1],$$
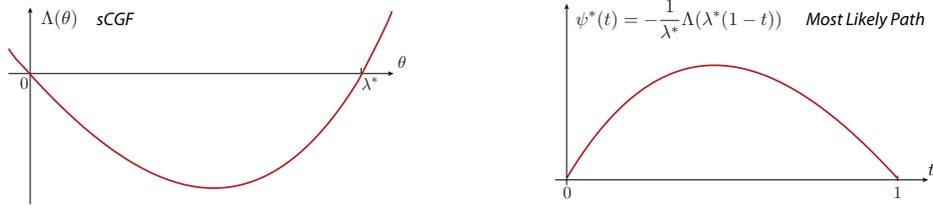
FIG 1. *Relationship between the scaled cumulant generation function of the increments of the random walk, Λ, and the most likely path to large busy period, ψ^*.*
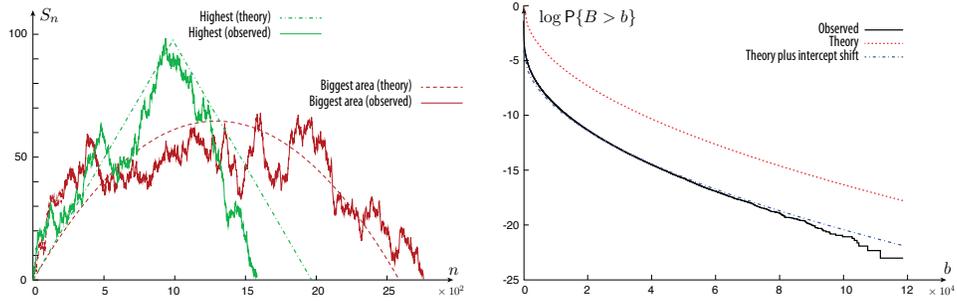

FIG 2. *I.i.d. Gaussian($-1/10$,1) increments. Simulated results from ten billion paths. (Left panel) Four paths displayed. Simulated path with largest swept area $111,524.13$ and, conditioned on that area, the most likely path to give rise to it as predicted by Theorem 2, which is concave. Also shown is the simulated path that reaches the greatest height 98.36 and, conditioned on that height, the most likely path to give rise to it as deduced from Theorem 1, which is piecewise linear. (Right panel) Logarithm of the empirical probability that the area under busy period exceeds a given level compared with theoretical prediction of $-K\sqrt{b}$ and (as the asymptote does not capture the prefactor) the linear shift $-K\sqrt{b}+\kappa$, where κ is chosen to match the offset of the empirical observation.*

with all other most likely paths being simple rescalings of ψ^*, as illustrated in Fig. 1. This is in contrast to the most likely path for the random walk to hit a high level, which is known to be piecewise linear, e.g. [1, 15].

As an illustrative example, consider a random walk with i.i.d Gaussian increments for which everything is calculated in closed form in Sec. 5.1. Ten billion busy period paths were simulated and, of these, the one with the largest swept area as well as the one that attained the greatest height were recorded. In addition to plotting these paths in Fig. 2, conditioned on these values the most likely paths to these events are shown. The distinct shapes of the paths to these two unlikely events is apparent. Note that the terminal times of the theoretically predicted paths are deductions of the conditioned area and height, respectively, and are not constrained explicitly. Also shown is the logarithm of the empirical probability with which the busy

period exceeded a given value as compared to the large deviation estimate. Up to a constant prefactor, the approximation inferred from Theorem 2 is remarkably accurate.

These results provide a partial answer to Open Problem 3.1 of [18] regarding precise asymptotics for the probability that B is large by identifying the associated rough asymptotics and showing that most likely paths are typically strictly concave. They also reveal a lacuna in [3, Theorem 4.1] where the most likely path is assumed to be piecewise linear.

We conclude the paper in Sec. 6 with a discussion of the practical utility of these results. Given observations of the increments process $\boldsymbol{X}$, estimates of the key quantities (λ^*, K) can be created that, under additional restrictions on $\boldsymbol{X}$, can be shown to satisfy a large deviation principle. This offers the possibility of accurately estimating the likelihood of a long busy period in advance of observing one.

2. A heuristic argument. In advance of proving the results surveyed in the introduction, we provide a heuristic argument along the lines suggested by an anonymous reviewer.

We are interested in asymptotics of the event

$$\left\{\sum_{i=1}^{\tau} S_i \geq bn^2\right\}$$

$$= \bigcup_{m \geq 0} \left\{\sum_{i=1}^{m} S_i \geq bn^2 : S_i > 0, \text{ for } i \in \{1, \ldots, m-1\} \text{ and } S_m \leq 0\right\}$$

$$= \bigcup_{t > 0} \left\{\sum_{i=1}^{[nt]} S_i \geq bn^2 : S_i > 0 \text{ for } i \in \{1, \ldots, [nt]-1\} \text{ and } S_{[nt]} \leq 0\right\}.$$

As the random walk has negative drift, one suspects that these sets have little probability for large values of t. Thus, under suitable conditions on the increments process, one might hope to apply a generalization of the principle of the largest term (e.g. Lemma 1.2.15 of [6]), that the exponential decay is dominated by the most likely event, to this non-finite collection, obtaining

$$\lim_{n \to \infty} \frac{1}{n} \log P\left(\sum_{i=1}^{\tau} S_i \geq bn^2\right)$$

(7)

$$= \sup_{t > 0} \lim_{n \to \infty} \frac{1}{n} \log P\left(\sum_{i=1}^{[nt]} S_i \geq bn^2, S_i > 0, i \in \{1, \ldots, [nt]-1\}, S_{[nt]} \leq 0\right),$$

and so one need only establish asymptotics of the probability of the event in the final equation (with t fixed). The difficulties that must be overcome to make such an argument rigorous mimic those for the logarithmic asymptotics of the supremum of a stochastic process, achieved in [16, 9, 19]. At the cost of additional assumptions, however, these difficulties are more readily overcome with a sample-path analysis that also reveals the form of the most likely path [15], which is the approach that we adopt in our rigorous arguments.

Continuing with the heuristic, fix $t > 0$ in eq. (7), do not fret about the positivity conditions, and consider

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\sum_{i=1}^{[nt]} S_i \geq bn^2\right) = \lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{[nt]^2}\sum_{i=1}^{[nt]} S_i \geq b\frac{n^2}{[nt]^2}\right)$$

$$= t \lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n^2}\sum_{i=1}^{n} S_i \geq \frac{b}{t^2}\right)$$

with a small leap of faith in the final equality. To identify this final expression in terms of the behavior of the random walk, it would suffice to establish that the Integrated Random Walk (i.e. the partial sums of the partial sums $\{\sum_{i=0}^{n} S_i/n^2\}$), satisfies a large deviation principle. This LDP can be established from [12, Theorem 5] giving the rate function

$$I_{\text{IRW}}(z) = \sup_{\theta}\left(\theta z - \theta \int_0^\theta \Lambda(x)dx\right) \text{ for } z \in \mathbb{R},$$

where Λ is the sCGF of the increments process found in eq. (4). Thus in conjunction with eq. (7) we have

$$(8) \quad \lim_{n\to\infty} \frac{1}{n}\log P\left(\sum_{i=1}^{\tau} S_i \geq bn^2\right) = -\inf_{t>0}\sup_{\theta}\left(\theta\frac{b}{t} - t\theta\int_0^\theta \Lambda(x)dx\right)$$

and all that remains is to simplify the expression on the right hand side. This comes about by noting that the function being optimized is convex in θ and concave in t so that the optimization order can be interchanged. The final expression is consistent with the limit eq. (3) established in this paper.

3. Functional setup. The framework for analysis used in this paper was first developed for weak convergence of probability measures [22, 4, 26] and subsequently used in the context of sample path large deviations, e.g. [7, 20, 28, 15, 14, 11]. In particular, Ganesh and O'Connell [15] employed this setup to establish an infinite time horizon version of Anantharam's result [1], proving that the most likely path to exceed a high level for a random walk with negative drift is piecewise linear on the scale of large deviations.

Let $\mathbb{C}[0, \infty)$ denote the collection of real-valued continuous functions on $[0, \infty)$. Let $\mathbb{A}[0, \infty)$ denote the collection of the integrals of functions that are elements of $\mathcal{L}^1[0, x)$ for all $x > 0$ (for example, see Riesz and Sz.-Nagy [24]). For each $r \in \mathbb{R}$, define the space

$$(9) \qquad \mathcal{Y}_r := \left\{ \phi \in \mathbb{C}[0, \infty) : \lim_{t \to \infty} \frac{\phi(t)}{(1 + t)} = r \right\}$$

and equip it with the topology induced by the norm

$$\|\phi\| = \sup_{t \geq 0} \left| \frac{\phi(t)}{1 + t} \right|.$$

Define the polygonal sample paths

$$S_n(t) := \frac{1}{n} S_{[nt]} + \left(t - \frac{[nt]}{n} \right) X_{[nt]+1}, \text{ for } t \in [0, \infty),$$

where $[t]$ is the least integer greater than t. The sample path process $\{S_n(\cdot)\}$ is known to satisfy the LDP in $\mathcal{Y}_r$ for a broad class of non-long-range dependent, non-heavy-tailed random walks. For example, see [15, Theorem 1], where [5, Theorem 2] provides general mixing and uniform tail exponent conditions under which the prerequisites of this theorem hold. This encompasses increment processes that are Harris recurrent Markov chains, subject to a Foster-Lyapunov drift condition [17]. The existence of such an LDP will be the primary assumption in our proof of eq. (3).

4. Tail asymptotics. We shall make two assumptions. The first is the existence of a sample path LDP for the random walk that ensures the walk has negative drift, but a possibility of becoming positive on the scale of large deviations.

ASSUMPTION 1. The sample path process $\{S_n(\cdot)\}$ satisfies the LDP in $\mathcal{Y}_{-\delta}$, some $\delta > 0$, with rate function

$$(10) \qquad I_\infty(\phi) = \begin{cases} \displaystyle\int_0^\infty I(\dot{\phi}(t))dt & \text{if } \phi \in \mathbb{A}[0, \infty) \cap \mathcal{Y}_{-\delta}, \\ +\infty & \text{otherwise,} \end{cases}$$

where I is the good, strictly convex rate function associated with the random walk $\{S_n(1)\}$, $I(-\delta) = 0$, and $I(x) < \infty$ for some $x > 0$.

Our identification of the most likely path to a large swept area will have a surprising relationship with the most likely path to exceed a high level, so we recall the following result of Ganesh and O'Connell.

THEOREM 1 ([15]). *Under Assumption 1, the following exist and are non-negative*

$$(11) \qquad x^* = \arg\inf_{x>0} \frac{1}{x}I(x) \ and \ \lambda^* = \frac{1}{x^*}I(x^*) = \sup\{\theta : \Lambda(\theta) \leq 0\}.$$

Moreover, for any $h > 0$,

$$(12) \qquad \lim_{n\to\infty} \frac{1}{n}\log P\left(\sup_k S_k \geq nh\right) = -h\lambda^*,$$

while the most likely path to this event is

$$(13) \qquad \varphi^*(t) = \begin{cases} x^*t & \text{if } t < h/x^*, \\ h + (t - h/x^*)(-\delta) & \text{if } t \geq h/x^*. \end{cases}$$

Note that the probability of sweeping a large area, eq. (3), decays on a slower scale than the probability of hitting a high height, eq. (12).

For the supremum of a random walk, λ^* determines the rate of decay of the probability of hitting a high level as shown in eq. (12). For busy periods, it will play a new and surprising rôle in the characterization of the most likely path to a large swept area. The inverse of ∇I will prove central to the development that follows. To that end, our second assumption is the following regularity condition.

ASSUMPTION 2. The rate function I is continuously differentiable on an interval that contains $[-\delta, x^*]$.

This assumption justifies the definition,

$$(14) \qquad \Xi(r) =: (\nabla I)^{-1}(r), \quad r \in \mathbb{R}, \text{ whenever the inverse exists.}$$

The inverse exists, so that Ξ is finite-valued, on an interval that contains $[\Xi(-\delta), \Xi(x^*)]$. That λ^* is significant here stems from the second part of the following lemma.

LEMMA 1. *The scalar λ^* defined in eq. (11) satisfies*

$$\lambda^* = \nabla I(x^*).$$

Moreover, it is the unique positive solution of

$$(15) \qquad \int_0^1 \Xi(\lambda^* s)\, ds = 0.$$

PROOF. The identity $\nabla I(x^*) = I(x^*)/x^*$ follows from the first-order optimality condition for x^*, based on its definition in eq. (11).

Clearly λ^* is positive as x^* is positive and $I(x) > 0$ for all $x > -\delta$. To see that λ^* thus defined is a solution of eq. (15), direct substitution, change of variables and integration by parts suffices. To see it is unique, note that it is equivalently characterized as any positive solution of

$$\int_0^{\lambda^*} \Xi(s)\, ds = 0.$$

As I is strictly convex, $\Xi(x)$ is strictly increasing when finite. At $x = 0$ we have that $\Xi(0) = -\delta < 0$, so this equation only has one positive solution. $\square$

We will use λ^* to define the most likely path to sweep an area over $[0, 1]$:

$$(16) \qquad \psi^*(t) = \begin{cases} \displaystyle\int_0^t \Xi\left(\lambda^*(1-s)\right) ds & \text{for } t \in [0,1] \\ (1-t)\delta & \text{for } t \geq 1. \end{cases}$$

The most likely path to sweep any other area will be a simple rescaling of this solution. This path ψ^* is strictly concave on $[0, 1]$ as I is strictly convex. On this interval it is of the form found in [12] in the analysis of simulation of queues: on differentiating each side of eq. (16), it follows that the path satisfies the simple differential equation,

$$(17) \qquad \nabla I\left(\tfrac{d}{dt}\psi^*(t)\right) = \lambda^*(1-t) \text{ for } t \in [0,1].$$

Note that, thus defined, we have that

$$(18) \qquad \tfrac{d}{dt}\psi^*(0) = \Xi(\lambda^*) = x^* \quad \text{and} \quad \tfrac{d}{dt}\psi^*(1) = \Xi(0) = -\delta.$$

That is, remarkably, the most likely paths to sweeping a large area, ψ^* in eq. (16), and to exceeding a large height, φ^* in eq. (13), both start and end with identical derivatives, but are distinct in-between. Before providing the main result, we establish the following characterizations of ψ^*.

PROPOSITION 1. *The following hold for $t \in [0,1]$:*

(i) *In contrast to eq. (16), a non-integral representation is obtained in terms of the rate function,*

$$(19) \qquad \psi^*(t) = \frac{I\left(\Xi(\lambda^*(1-t))\right)}{\lambda^*} - \Xi(\lambda^*(1-t))(1-t).$$

(ii) *In terms of the sCGF,*

$$(20) \qquad \psi^*(t) = -\frac{1}{\lambda^*}\Lambda(\lambda^*(1-t)).$$

and hence,

$$\tfrac{d}{dt}\psi^*(t) = \nabla\Lambda(\lambda^*(1-t))$$

PROOF. The identity $\nabla\Lambda(\theta) = \Xi(\theta)$ can be established for $\theta \in [0, \lambda^*]$ based on convex duality (e.g, [25, Proposition 11.3]). The representation in eq. (20) then follows by integration using $\psi^*(0) = 0$. The characterization in eq. (19) can be obtained from eq. (20) noting that $\Lambda(\theta) = \theta\,\Xi(\theta) - I(\Xi(\theta))$ for $\theta \in [0, \lambda^*]$. □

A simpler expression for λ^* is obtained when Λ is symmetric. This symmetry can be interpreted as asymptotic reversibility of the underlying walk.

PROPOSITION 2. *Recall that $\lambda^* > 0$ is a zero of Λ, $-\delta < 0$ is a zero of I, and $x^* > 0$ is the slope given in Theorem 1. Under the symmetry condition,*

$$(21) \qquad \Lambda(\theta) = \Lambda(\lambda^* - \theta) \text{ for all } \theta \in [0, \lambda^*],$$

then these parameters are related as follows,

$$(22) \qquad x^* = \delta \text{ and } \lambda^* = \nabla I(\delta) = I(\delta)/\delta = 2\nabla I(0).$$

Moreover, if eq. (21) holds for all $\theta \in \mathbb{R}$, then

$$(23) \qquad \lambda^* = \frac{I(x) - I(-x)}{x} \text{ for all } x \neq 0 \text{ such that } I(x) < \infty.$$

PROOF. If $\Lambda(\theta) = \Lambda(\lambda^* - \theta)$ for $\theta \in [0, \lambda^*]$, the path $\psi^*(t)$ is symmetric with $\psi^*(t) = \psi^*(1-t)$. Thus $\frac{d}{dt}\psi^*(t) = -\frac{d}{dt}\psi^*(1-t)$. In particular, evaluating this at $t = 0$ and using eq. (18) gives $x^* = \frac{d}{dt}\psi^*(0) = -\frac{d}{dt}\psi^*(1) = \delta$. Hence by eq. (11), $\lambda^* = \nabla I(\delta) = I(\delta)/\delta$. Equating the derivatives at $t = 1/2$ gives $2\,\Xi(\lambda^*/2) = 0$ so that $\lambda^* = 2\nabla I(0)$.

If eq. (21) holds for all θ, then

$$I(-x) = \sup_{\theta \in \mathbb{R}} \left(\theta(-x) - \Lambda(\lambda^* - \theta)\right) = \sup_{\theta \in \mathbb{R}} \left((\lambda^* - \theta)(-x) - \Lambda(\theta)\right)$$
$$= I(x) - \lambda^* x \text{ for all } x.$$

This gives eq. (23) for all $x \in \mathbb{R}$ such that $I(x) < \infty$. □

For example, eq. (21) is satisfied if $\boldsymbol{X}$ is i.i.d., $E(\exp(\theta X_1))$ is finite in a neighborhood of the origin and $P(X_1 = x)/P(X_1 = -x) = \exp(-\lambda^* x)$ for all x, as holds for X_1 Gaussian or Bernoulli-$\{-C, +C\}$. This condition, however, extends beyond i.i.d. increments processes and in Sec. 5 we present an example where eq. (21) is satisfied for a Markov chain $\boldsymbol{X}$.

Armed with these assumptions, definitions and characterizations of the path ψ^*, we now prove the main result.

THEOREM 2.   *Under the above assumptions, for any $b > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log P(B \geq n^2 b) = -K\sqrt{b}.$$

where K is given in eq. (5). Moreover,

(i) *The most likely asymptotic value of τ/n leading to $\{B \geq n^2 b\}$ is*

$$a =: \frac{2\lambda^*}{K} \sqrt{b}.$$

(ii) *The rescaled, most likely asymptotic path of $S_n(\cdot)$ is*

$$(24) \qquad \psi_b^*(t) = a\psi^*(t/a) = -2\frac{\sqrt{b}}{K} \Lambda\left(\lambda^* - \frac{t}{2}\frac{K}{\sqrt{b}}\right),$$

which is strictly concave on $[0, a]$.

PROOF. The method of proof is to construct a collection of open sets, $\{B_\epsilon\}$, and a closed set F in $\mathcal{Y}_{-\delta}$ such that for all $\epsilon > 0$ sufficiently small

$$\{S_n(\cdot) \in B_\epsilon\} \subset \{B \geq bn^2\} \subset \{S_n(\cdot) \in F\}$$

and

$$\lim_{\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log P(S_n(\cdot) \in B_\epsilon) = \limsup_{n\to\infty} \frac{1}{n} \log P(S_n(\cdot) \in F).$$

Define

$$F := \left\{ \phi \in \mathcal{Y}_{-\delta} : \int_0^\infty \max(\phi(t), 0) dt \geq b \right\}.$$

If $B \geq bn^2$, then $S_n(\cdot) \in F$ as

$$\int_0^\infty \max(S_n(t), 0) dt \geq \int_0^{n\tau} S_n(t) dt \geq \frac{1}{n^2} \sum_{i=1}^{\tau} S_i = \frac{1}{n^2} B \geq b.$$

The set F is closed as $\phi \mapsto \max(\phi, 0)$ is Lipschitz continuous from $\mathcal{Y}_{-\delta} \to \mathcal{Y}_0$ and, as for any $\phi \in \mathcal{Y}_{-\delta}$ we have $\phi(t) < 0$ for all t sufficiently large, integration is also continuous (e.g. [27, Theorem 11.5.1]). Thus we can use the LDP upper bound to obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log P(B \geq bn^2) \leq \limsup_{n\to\infty} \frac{1}{n} \log P(S_n(\cdot) \in F)$$

$$= -\inf\{I_\infty(\phi) : \phi \in F\}$$

$$= -\inf_{t>0} \inf\left\{I_\infty(\phi) : \int_0^t \max(\phi(s), 0)\, ds \ge b\right\}.$$

As $\lim_{t\to\infty} \phi(t)/(1+t) = -\delta < 0$ for all $\phi \in \mathcal{Y}_{-\delta}$, this infimum over t is attained at some finite t. Consider this inner functional infimum for fixed t:

(25)

minimize $\quad I_\infty(\phi)$

subject to $\quad \phi \in L^+[0,t]$ and $\int_0^t \max(\phi(t), 0) \ge b.$

This functional optimization problem is closely related to [12, equation (6)], where one can identify b with z. Mild alterations to Proposition 7 therein shows that with $b = a^2 \int_0^1 \psi^*(t) dt$ for some $a > 0$, with $\psi^*(t)$ defined in eq. (16), this infimum occurs at any $t \ge a$ for which it transpires that ψ_b^*, defined in eq. (24), is the optimizer. This essentially occurs as

$$\psi_b^* = \arg\inf\left\{\int_0^a I(\dot\phi(t)) dt : \int_0^a \phi(t) = a^2 \int_0^1 \psi^*(t) dt\right\}$$

The quantity λ^* in the definition of ψ^* arises as a scalar Lagrange multiplier [12]. Note that $\psi_b^*(a) = 0$, and thus the most likely path to sweep a rescaled area b satisfies $\tau/n \approx a = \sqrt{b / \int_0^1 \psi^*(t) dt}$. Using Proposition 1 we have that

$$\int_0^1 \psi^*(t)\, dt = -\frac{1}{\lambda^*} \int_0^1 \Lambda(\lambda^*(1 - t))\, dt = -\frac{1}{(\lambda^*)^2} \int_0^{\lambda^*} \Lambda(\theta)\, d\theta$$

and thus the expression for a in the statement follows.

What remains to be shown is that there is a coincident lower bound. Let ψ^* be the unique solution of eq. (16) and define ψ_b^* using eq. (24). The path $\psi_b^*(t)$ starts at 0 ends at a, and is the optimal path that sweeps an area of $\int_0^a \psi_b^*(t) dt = a^2 \int_0^1 \psi^*(t) dt = b$. Let

$$B_\epsilon = B_\epsilon^1 \cap B_\epsilon^2$$

where, with $e(t) = \epsilon(1 + t)$,

$$B_\epsilon^1 = \left\{\phi : \inf_{t\in[0,\epsilon]}(\phi(t) - t\epsilon) > 0\right\} \text{ and } B_\epsilon^2 = \{\phi : \|\phi - (\psi_b^* + e)\| < \epsilon\}$$

Both B_ϵ^1 and B_ϵ^2 are open by construction and, for ϵ sufficiently small, their intersection is non-empty. As defined, $S_n(\cdot) \notin B_\epsilon^1$ as $S_n(0) = 0$, but this is not significant as we can use an exponentially equivalent representation

with $S_n(0) := 1/n$. Thus if $S_n(\cdot) \in B_\epsilon$ with $S_n(0) := 1/n$, then $\{B > bn^2\}$. As B_ϵ is open for all ϵ, we can use the LDP lower bound

$$\liminf_{n\to\infty} \frac{1}{n} \log P(B \geq bn^2) \geq \lim_{\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log P(S_n(\cdot) \in B_\epsilon)$$

$$\geq \lim_{\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log P(S_n(\cdot) \in B_\epsilon)$$

$$= -\lim_{\epsilon\to 0} \inf\{I_\infty(\phi) : \phi \in B_\epsilon\} = -I_\infty(\psi_b^*).$$

To evaluate K, note that

$$(26) \qquad \lim_{n\to\infty} \frac{1}{n} \log P(B \geq bn^2) = -I_\infty(\psi_b^*) = -\frac{\int_0^1 I(\frac{d}{dt}\psi^*(t))dt}{\sqrt{\int_0^1 \psi^*(t)dt}} \sqrt{b}.$$

Using $\Lambda(\theta) = \theta \, \Xi(\theta) - I(\Xi(\theta))$ for $\theta \in [0, \lambda^*]$, integration by parts and the fact that $\nabla\Lambda(\lambda^*) = 0$, we have that

$$\int_0^1 I(\dot{\psi}^*(t)) \, dt = -\frac{2}{\lambda^*} \int_0^{\lambda^*} \Lambda(\theta) \, d\theta.$$

Inserting this into eq. (26) in conjunction with the expression for $\int_0^1 \psi^*(t) \, dt$ given above obtains the expression for K in eq. (5). $\quad\square$

5. Examples. Theorem 2 provides a mechanism for calculating the most likely path to a large busy period B as well as the exponent K. We shall perform this calculation for illustrative examples: with i.i.d Gaussian increments where λ^*, ψ^* and K can all be determined in closed form; with Bernoulli$\{-1, +C\}$ increments, which includes M/M/1 queue lengths, where explicit expressions of λ^* are not always possible, but ψ^* can be written in terms of it and K must always be calculated numerically; and, finally, for increments with Markovian dependencies where λ^* and ψ^* can be determined in closed form, but K must be identified numerically.

For each of the examples ten billion paths were simulated. As well as recording the logarithm of the frequency with which B exceeded b as a function of b, the largest swept area and highest paths were logged for comparison with the theoretically predicted most likely paths. For comparison with observations, Theorem 2 says that given we observe $B = n^2 b$, on the scale of large deviations the most likely time taken to generate the area is

$$\tau \approx na = n2\sqrt{b}/K = \tau_n^* \quad \text{where} \quad \tau_n^* =: \tfrac{1}{2}K/B^{-1/2}$$

and the most likely path is then

$$S_i =: nS_n(i/n) \approx \psi_{B/n^2}^*(i/n) = na\psi^*(i/(na)).$$

Thus, since $\tau_n^* = an$, for large swept area we have the approximation

$$S_i \approx \tau_n^* \psi^*(i/\tau_n^*).\tag{27}$$

This most likely path is solely parameterized from the observations by the value B. In particular, note that given B the time τ is determined, so in the comparisons for the simulation results that follow, the length of the most likely path is not explicitly fit to data.

Similarly, eq. (3) leads to the approximation

$$\log P(B \geq b) \approx -K\sqrt{b}, \text{ for large } b.\tag{28}$$

When comparing this approximation with data, the data ultimately becomes sparse for large values of b and discrepancies appear. The value of b where this occurs moves right with as a function of the number of simulated paths.

For contrast, we also record the path that reaches the highest height and use Theorem 1 for comparison. Given a path S_n that reaches a height H, $S_n(\cdot)$ reaches H/n and therefore

$$S_i = nS_n(i/n) \approx n\varphi^*(i/n) = \begin{cases} x^*i & \text{if } i < H/x^*, \\ H + (i - H/x^*)(-\delta) & \text{if } i \geq H/x^*. \end{cases}\tag{29}$$

Again, given H, the time at which this most likely path returns to zero is completely determined.

5.1. *Gaussian increments.* Let $\boldsymbol{X}$ be i.i.d. Gaussian, with X_0 having mean $-\delta < 0$ and variance σ^2. As $\boldsymbol{X}$ is i.i.d., we have that

$$\Lambda(\theta) = \frac{\sigma^2}{2}\theta^2 - \delta\theta.$$

Using Theorem 2 we obtain the following explicit expressions

$$\lambda^* = \sup\{\theta : \Lambda(\theta) \leq 0\} = \frac{2\delta}{\sigma^2}, \qquad \psi^*(t) = -\frac{1}{\lambda^*}\Lambda(\lambda^*(1-t)) = \delta t(1-t)$$

$$\text{and } K = 2\sqrt{-\int_0^{\lambda^*} \Lambda(\theta)\, d\theta} = \frac{\delta^{3/2}}{\sigma^2}\sqrt{\frac{8}{3}}.$$

The most likely path to $\{B \geq bn^2\}$ for large n can also be determined to be

$$\psi_b^*(t) = \delta t\left(1 - t\sqrt{\frac{\delta}{6b}}\right).$$

A demonstration of these results appears in Fig. 2. The highest and largest swept area paths are compared with the most likely conditioned on these quantities using the approximations in eq. (27) and eq. (29). The distinct nature of two paths to these two unlikely events is evident. The empirical likelihood of a large deviation is shown along with that from the approximation above, which gives remarkably good agreement up to a constant prefactor.

5.2. *Bernoulli* $\{-1, +C\}$ *increments.* Let $\boldsymbol{X}$ be a Bernoulli sequence taking values -1 and C with $\alpha = 1 - \mu = P\{X_0 = C\} < P\{X_0 = -1\} = \mu$, so that $-\delta = \alpha C - \mu$, and define the load $\rho = \alpha C / \mu$. For this process

$$\Lambda(\theta) = \log\left(\mu e^{-\theta} + \alpha e^{C\theta}\right)$$

and, by eq. (11), λ^* is the unique positive solution of

$$\alpha e^{\lambda^* C} + \mu e^{\lambda^*} - 1 = 0.$$

Closed form expressions for λ^* do not exist apart from in a few special cases. If $C = 1$, Λ satisfies the conditions of Proposition 2, the most likely path is symmetric and this equation has the explicit solution $\lambda^* = -\log(\rho)$, which could also have been determined by $2\nabla I(0)$ or $I(\delta)/\delta$. Given the reversibility of the M/M/1 queue, the symmetry of ψ^* established in Proposition 2 is not surprising.

If $C \geq 2$, then the corresponding reflected random walk is not reversible, and moreover the increments do not satisfy the symmetry assumptions of Proposition 2. For $C = 2$ we obtain the expression,

$$\lambda^* = \log\left(\frac{\sqrt{-3\mu^2 + 2\mu + 1} - \alpha}{2\alpha}\right),$$

while if $C = 3$

$$\lambda^* = \log\left(\frac{7z}{9} - \frac{1}{3}\right),$$

$$\text{where } z = \sqrt[3]{\sqrt{\left(\frac{\mu}{2(\mu-1)} - \frac{7}{54}\right)^2 + \frac{8}{729}} - \frac{\mu}{2(\mu-1)} + \frac{7}{54}}.$$

Using eq. (20), for arbitrary C we conclude that the most likely path on $[0,1]$ can be written in terms of λ^* as

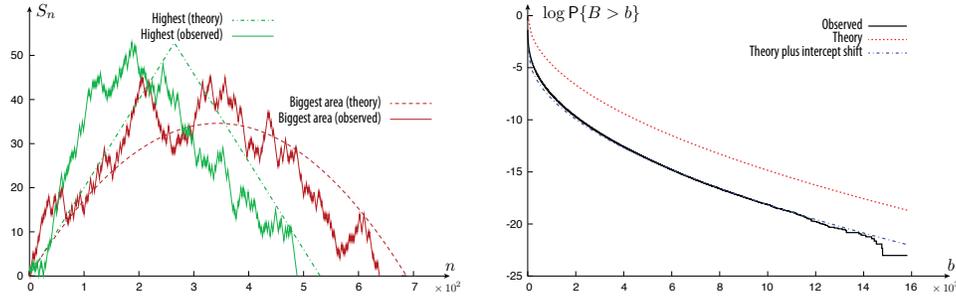$$\psi^*(t) = -\frac{1}{\lambda^*} \log\left(\mu e^{-\lambda^*(1-t)} + \alpha e^{C\lambda^*(1-t)}\right).$$

FIG 3. *Bernoulli* $\{-1, +1\}$ *increments with* $P(X_1 = -1) = 0.6$, $P(X_1 = +1) = 0.4$. *Simulated results from ten billion paths. (Left panel) Four paths displayed. Simulated path with largest swept area* 15, 825 *and, conditioned on that area, most likely path to give rise to it predicted by Theorem 2, which is concave. For contrast, also shown is the simulated path that reaches the greatest height,* 53, *and, conditioned on that height, the most likely path to give rise to it as deduced from Theorem 1, which is piecewise linear. (Right panel) Logarithm of the empirical probability that area under busy period exceeds a given level compared with* $-K\sqrt{b}$ *and, the linear shift* $-K\sqrt{b} + \kappa$ *where* κ *is chosen to match the offset of the empirical observation.*

In order to calculate K in eq. (5), we need to evaluate the integral $\int_0^{\lambda^*} \Lambda(\theta)\, d\theta$. This doesn't result in a closed form for any C, but it is simple to evaluate numerically.

In order to determine the most likely time and paths to a large busy period and a great height on the scale of large deviations we use the approximations eq. (27) and eq. (29). For the reversible Bernoulli $\{-1, +1\}$ case corresponding to M/M/1 queue-lengths, Fig. 3 compares the highest and biggest paths with those from theory. The quality of the predictions is apparent. With a numerical integration of Λ giving $K \approx 0.1485$, the asymptotic approximation is compared with the empirical probability, showing great accuracy up to a constant prefactor.

As an example of a non-reversible random walk, we consider the Bernoulli $\{-1, +10\}$ case where the most likely paths are now asymmetric (as seen in Fig. 4). For this example, $\lambda^* \approx 0.1439$ and $K \approx 0.0978$, both of which have been determined numerically.

5.3. *Markovian* $\{-1, +1\}$ *increments.* As an example beyond i.i.d. increments, assume that the increments process $\boldsymbol{X}$ forms a two-state Markov chain on the state space $\{-1, +1\}$ with transition matrix

$$\begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \text{ where } 0 < \alpha < \beta < 1.$$
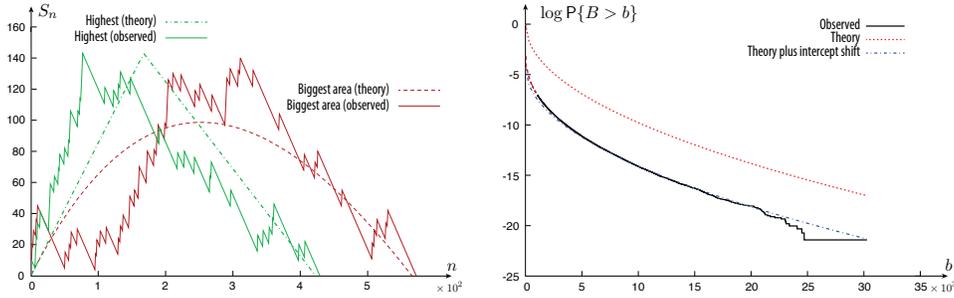
FIG 4. *Bernoulli* $\{-1, +10\}$ *increments with* $P(X_1 = -1) = 0.96$, $P(X_1 = +10) = 0.04$.
*Simulated results from ten billion paths. (Left panel) Four paths displayed. Simulated path
with largest swept area* $36,971$ *and, conditioned on that area, most likely path to give rise
to it predicted by Theorem 2, which is concave. For contrast, also shown is the simulated
path that reaches the greatest height,* 143, *and, conditioned on that height, the most likely
path to give rise to it as deduced from Theorem 1, which is piecewise linear. (Right panel)
Logarithm of the empirical probability that area under busy period exceeds a given level
compared with* $-K\sqrt{b}$ *and, the linear shift* $-K\sqrt{b} + \kappa$ *where* κ *is chosen to match the
offset of the empirical observation.*

The stationary distribution is $(\beta/(\alpha + \beta), \alpha/(\alpha + \beta))$ so we require $\alpha < \beta$
for stability. The sCGF Λ can be calculated using techniques described in
[6, Section 3.1]:

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

which satisfies the conditions of Proposition 2 and so the most likely path
is symmetric. The rate function for the associated random walk can be
calculated using methods described in [7]. For example, from [10] we have
that

$$I(x) = -\left(\frac{1-x}{2}\right)\log(1 - \alpha + \alpha\chi(x)) - \left(\frac{1+x}{2}\right)\log(1 - \beta + \beta/\chi(x))$$

where

$$\chi(x) = \frac{\alpha\beta x + \sqrt{\alpha^2\beta^2 x^2 + \alpha\beta(1+x)(1-\alpha)(1-\beta)(1-x)}}{\alpha(1-\beta)(1-x)}.$$

One can check directly that

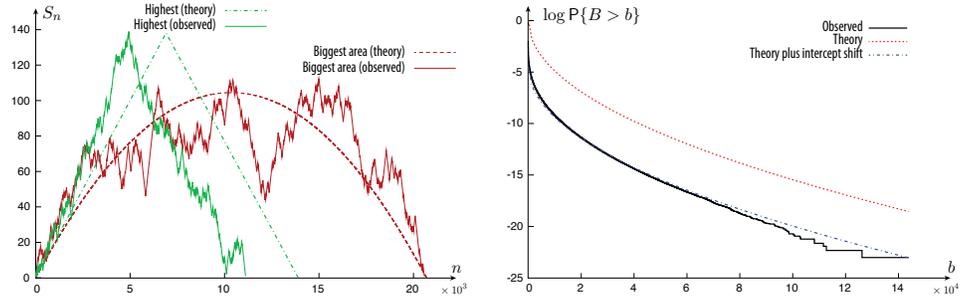$$x^* = -\rho = \delta = \frac{\beta - \alpha}{\alpha + \beta} \text{ and}$$

FIG 5. *Markov $\{-1, +1\}$ increments with $P(X_{t+1} = +1|X_t = -1) = \alpha = 2/10$, $P(X_{t+1} = -1|X_t = +1) = \beta = 3/10$. Simulated results from ten billion paths. (Left panel) Four paths displayed. Simulated path with largest swept area $143,875$ and, conditioned on that area, most likely path to give rise to it predicted by Theorem 2, which is concave. For contrast, also shown is the simulated path that reaches the greatest height, 139, and, conditioned on that height, the most likely path to give rise to it as deduced from Theorem 1, which is piecewise linear. (Right panel) Logarithm of the empirical probability that area under busy period exceeds a given level compared with $-K\sqrt{b}$ and, the linear shift $-K\sqrt{b} + \kappa$ where κ is chosen to match the offset of the empirical observation.*

$$\lambda^* = I(x^*)/x^* = \nabla I(x^*) = 2\nabla I(0) = \sup\{\theta : \Lambda(\theta) \le 0\} = \log\left(\frac{1-\alpha}{1-\beta}\right).$$

We have the following expression of the most likely path returning to 0 at $t = 1$, which can be seen directly to possess the symmetry $\psi^*(t) = \psi^*(1-t)$ for $t \in [0, 1]$,

$$\psi^*(t) = -\frac{1}{\log(\frac{1-\alpha}{1-\beta})}$$
$$\left(\log\left((1-\alpha)^t(1-\beta)^{1-t} + (1-\alpha)^{1-t}(1-\beta)^t\right.\right.$$
$$\left.\left. + \sqrt{4\alpha\beta + ((1-\alpha)^t(1-\beta)^{1-t} - (1-\alpha)^{1-t}(1-\beta)^t)^2}\right)\right) - \log(2)\right).$$

The integral $\int_0^{\lambda^*} \Lambda(\theta)\, d\theta$ does not evaluate in closed form, so again numerics must be used to determine K. For example, if $\alpha = 2/10$ and $\beta = 3/10$, then $\delta = 2/10$, $x^* = 2/10$, $\lambda^* = \log(8/7)$ and $K \approx 0.0489$. A simulation-based illustration of these results appears in Fig. 5.

6. Conclusions and discussion on estimation. As well as identifying the most likely path to a large swept area, Theorem 2 shows that in the absence of long range dependence and heavy tailed increments, in broad generality we have the approximation in eq. (28):

$$\log P(B \ge b) \approx -K\sqrt{b}, \text{ for large } b,$$

where K can be identified in terms of the sCGF Λ associated with the increments $\boldsymbol{X}$. An approximation such as this might be of value for practical purposes, but unless $\boldsymbol{X}$ is known in advance we would require a methodology to estimate K from observations of the system.

Based on thermodynamic ideas, Duffield *et al.* [8] investigated an estimation scheme for Λ based on observations of $\boldsymbol{X}$. They demonstrated empirically that the scheme has desirable properties for a large class of of increment processes. Indeed, if $\boldsymbol{X}$ consists of i.i.d. bounded random variables [10, Theorem 1] or a finite state Markov chain [13, Theorem 3], from observations of $\boldsymbol{X}$ one can construct consistent functional estimates, $\{\Lambda_n\}$, of Λ that themselves satisfy a LDP in the space of $\mathbb{R}$-valued convex functions on $\mathbb{R}$. From these, we can deduce an LDP for estimating K from observations of $\boldsymbol{X}$ as follows. If $\boldsymbol{X}$ is i.i.d, define

$$\Lambda_n(\theta) = \log\left(\frac{1}{n}\sum_{k=1}^{n} e^{\theta X_k}\right).$$

If $\boldsymbol{X}$ forms a finite state Markov chain with an irreducible transition matrix on $\{f(1), \ldots, f(M)\}$ where $f(i) \neq f(j)$ for $i \neq j$, then with $0/0$ defined to be 0 we define an empirical transition matrix with entries,

$$(P_n)_{i,j} := \left(\sum_{k=1}^{n} 1_{\{(X_{k-1}, X_k) = (f(i), f(j))\}}\right) \Big/ \left(\sum_{k=1}^{n} 1_{\{X_{k-1} = f(i)\}}\right),$$

and D_θ denote the matrix with diagonal entries $\exp(\theta f(1)), \ldots, \exp(\theta f(M))$ and all off-diagonal entries equal to zero. Then our estimate of Λ given n observations is

$$\Lambda_n(\theta) = \log \rho(P_n D_\theta),$$

where where ρ is the spectral radius. In both cases we define the estimates

$$\lambda_n^* = \sup\{\theta : \Lambda_n(\theta) \leq 0\} \text{ and } K_n = 2\sqrt{-\int_0^{\lambda_n^*} \Lambda_n(\theta)\,d\theta}.$$

Regarding the λ^* estimates, [10, Lemma 1] proves that the function $g(c) = \sup\{\theta : c(\theta) \leq 0\}$ is continuous on the space of convex functions equipped with the topology of uniform convergence on compact subsets at all c such that $c(0) = 0$ and there does not exist a $\kappa > 0$ so that $c(\theta) = 0$ for $\theta \in [0, \kappa]$. Thus from the LDP for the estimators $\{\Lambda_n\}$, if the support of X_k excludes a finite ball around the origin and if $f(k) \neq 0$ for any k in the

Markovian case, we have an LDP for $\{\lambda_n^*\}$ by Puhalskii's extension of the contraction principle [23, Theorem 2.1]. As the convex functions of interest are real-valued, the function $c \mapsto \int_0^{g(c)} c(\theta)d\theta$ is also continuous so that again Puhalskii's extension of the contraction principle applies and we have an LDP for $\{K_n\}$. Consistency of the Λ estimators ensures consistency of the λ^* and K estimators. Thus these Λ-estimators enable the estimation of K directly from observations of $\boldsymbol{X}$, offering the possibility of estimating the probability of a system experiencing an exceedingly large busy period in advance of one occurring.

Acknowledgment. We thank an anonymous reviewer for the suggestion to include a heuristic argument related to the integrated random walk, which motivated the discussion in Section 2.

REFERENCES

[1] ANANTHARAM, V. (1989). How large delays build up in a $GI/G/1$ queue. *Queueing Systems Theory Appl.* **5** 345–367. MR1030475

[2] BLANCHET, J., GLYNN, P. and MEYN, S. (2013). Large deviations for the empirical mean of an $M/M/1$ queue. *Queueing Syst.* **73** 425–446. MR3036010

[3] BOROVKOV, A. A., BOXMA, O. J. and PALMOWSKI, Z. (2003). On the integral of the workload process of the single server queue. *J. Appl. Probab.* **40** 200–225. MR1953775

[4] BOROVKOV, A. A. and SAHANENKO, A. I. (1973). Remarks on the convergence of random processes in nonseparable metric spaces and on the nonexistence of a Borel measure for processes in $C(0, \infty)$. *Teor. Verojatnost. i Primenen.* **18** 812–815. MR0328984

[5] DEMBO, A. and ZAJIC, T. (1995). Large deviations: from empirical mean and measure to partial sums. *Stoch. Proc. Appl.* **57** 191–224. MR1334969

[6] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviation Techniques and Applications.* Springer. MR1619036

[7] DEUSCHEL, J. D. and STROOCK, D. W. (1989). *Large Deviations* **137**. Academic Press Inc. MR0997938

[8] DUFFIELD, N. G., LEWIS, J. T., O'CONNELL, N., RUSSELL, R. and TOOMEY, F. (1995). Entropy of ATM traffic streams: a tool for estimating QoS parameters. *IEEE J. Sel. Area Comm.* **13** 981–990.

[9] DUFFY, K., LEWIS, J. T. and SULLIVAN, W. G. (2003). Logarithmic asymptotics for the supremum of a stochastic process. *Ann. Appl. Probab.* **13** 430–445. MR1970270

[10] DUFFY, K. and METCALFE, A. P. (2005). The large deviations of estimating rate functions. *J. Appl. Probab.* **42** 267–274. MR2144909

[11] DUFFY, K. R. and MALONE, D. (2008). Logarithmic asymptotics for a single-server processing distinguishable sources. *Math. Methods Oper. Res.* **68** 509–537. MR2457294

[12] DUFFY, K. R. and MEYN, S. P. (2010). Most likely paths to error when estimating the mean of a reflected random walk. *Perform. Evaluation* **67** 1290–1303.

[13] DUFFY, K. R. and MEYN, S. P. (2010). Estimating Loynes' exponent. *Queueing Syst.* **68** 285–293. MR2834199

[14] GANESH, A., O'CONNELL, N. and WISCHIK, D. (2004). *Big Queues. Lecture Notes in Mathematics* **1838**. Springer-Verlag, Berlin. MR2045489

[15] GANESH, A. J. and O'CONNELL, N. (2002). A large deviation principle with queueing applications. *Stoch. Stoch. Rep.* **73** 25–35. MR1914976

[16] GLYNN, P. and WHITT, W. (1994). Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *J. Appl. Probab.* **31A** 413–430. MR1274722

[17] KONTOYIANNIS, I. and MEYN, S. P. (2005). Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.* **10** no. 3, 61–123 (electronic). MR2120240

[18] KULIK, R. and PALMOWSKI, Z. (2011). Tail behaviour of the area under a random process, with applications to queueing systems, insurance and percolations. *Queueing Syst.* **68** 275–284. MR2834198

[19] LELARGE, M. (2008). Tail asymptotics for discrete event systems. *Discrete Event Dyn. Syst.* **18** 563–584. MR2443657

[20] MAJEWSKI, K. (2000). Single class queueing networks with discrete and fluid customers on the time interval $\mathbb{R}$. *Queueing Syst.* **36** 405–435. MR1823977

[21] MEYN, S. P. (2008). *Control Techniques for Complex Networks.* Cambridge University Press. MR2372453

[22] MÜLLER, D. W. (1968). Verteilungs-Invarianzprinzipien für das starke Gesetz der grossen Zahl. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **10** 173–192. MR0232428

[23] PUHALSKII, A. (1995). Large deviation analysis of the single server queue. *Queueing Syst.* **21** 5–66. MR1372048

[24] RIESZ, F. and NAGY, B. S. (1955). *Functional Analysis.* Blackie and Son Limited.

[25] ROCKAFELLAR, R. T. and WETS, R. J. B. (1998). *Variational Analysis.* Springer. MR1491362

[26] WHITT, W. (1972). Stochastic Abelian and Tauberian theorems. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **22** 251–267. MR0339326

[27] WHITT, W. (2002). *Stochastic-Process Limits.* Springer-Verlag, New York. MR1876437

[28] WISCHIK, D. J. (2001). Sample path large deviations for queues with many inputs. *Ann. Appl. Probab.* **11** 379–404. MR1843050

HAMILTON INSTITUTE
NATIONAL UNIVERSITY
  OF IRELAND MAYNOOTH
IRELAND
E-MAIL: ken.duffy@nuim.ie

DEPT. OF ELECTRICAL
  AND COMPUTER ENGINEERING
UNIVERSITY OF FLORIDA
GAINESVILLE
FL 32611-6200
U.S.A.
E-MAIL: ken.duffy@nuim.ie