Large deviations behavior of counting processes and their inverses

Peter W. Glynn

Department of Operations Research, Stanford University, Stanford, CA 94305-4022, USA

Ward Whitt

AT&T Bell Laboratories, Room 2C-178, Murray Hill, NJ 07974-0636, USA

We show, under regularity conditions, that a counting process satisfies a large deviations principle in $\mathbb{R}$ or the Gärtner–Ellis condition (convergence of the normalized logarithmic moment generating functions) if and only if its inverse process does. We show, again under regularity conditions, that embedded regenerative structure is sufficient for the counting process or its inverse process to have exponential asymptotics, and thus satisfy the Gärtner–Ellis condition. These results help characterize the small-tail asymptotic behavior of steady-state distributions in queueing models, e.g., the waiting time, workload and queue length.

Keywords: Large deviations, Gärtner–Ellis theorem, counting processes, point processes, cumulant generating function, waiting-time distribution, small-tail asymptotics.

1. Introduction and summary

Let $T \equiv \{T_n : n \geq 0\}$ be a nondecreasing sequence of real-valued random variables with $T_0 = 0$, and let

$$N(t) = \max\{n \geq 0 : T_n \leq t\}, \quad t \geq 0. \tag{1}$$

Then $N \equiv \{N(t) : t \geq 0\}$ is a *counting process* and T is its *inverse*. Motivated by applications to queues, see Chang [3], Chang et al. [4], Glynn and Whitt [11] and Whitt [16], we want to relate the large deviations behavior of N to the large deviations behavior of T. This is in the same spirit as previous relations between other limits for N and T, such as the law of large numbers and central limit theorem; see §7 of Whitt [15], theorem 6 of Glynn and Whitt [9] and §2 of Massey and Whitt [13].

A real-valued stochastic process $Z \equiv \{Z(t) : t \geq 0\}$ will be said to satisfy the *Gärtner–Ellis condition with decay rate function* ψ if its normalized logarithmic

moment generating function has a limit, i.e., if

$$t^{-1} \log E e^{\theta Z(t)} \to \psi(\theta) \quad \text{as } t \to \infty \text{ for all } \theta \in \mathbb{R}. \tag{2}$$

(For a discrete-time process, we let t run through the positive integers in (2).) For the queueing applications, we want to know when N and T satisfy (2) for θ in an appropriate interval. In Glynn and Whitt [11] we consider a single-server queue with unlimited waiting space and a stationary sequence of interarrival times independent of a stationary sequence of service times. By theorem 1 and proposition 2 there, if the partial sums of the interarrival times and service times each satisfy (2) with decay rate functions $\psi_a(\theta)$ and $\psi_s(\theta)$, respectively, with these decay rate functions satisfying regularity conditions, then the steady-state waiting time has logarithmic asymptotics of the form $x^{-1} \log P(W > x) \to -\theta^*$ as $x \to \infty$, where θ^* is the root of the equation $\psi(\theta) = 0$, where $\psi(\theta) = \psi_s(\theta) + \psi_a(-\theta)$. (We need (2) only in the neighborhood of θ^*.) Given this result, we want to be able to relate (2) for the interarrival-time partial sums (a process of the form T) to (2) for the corresponding arrival counting process (a process of the form N).

Since $\log E e^{\theta Z}$ is convex in θ for any random variable Z by Hölder's inequality, the decay rate function ψ in (2) is necessarily convex with $\psi(0) = 0$. For non-negative random variables Z, $\log E e^{\theta Z}$ is also nondecreasing in θ, so that $\psi(\theta)$ will be nondecreasing as well for the processes we consider. Let β^l and β^u be the limits of the region of increase of ψ i.e.,

$$\beta^l = \sup\{\theta : \psi(\theta) = \psi(-\infty)\} \quad \text{and} \quad \beta^u = \inf\{\theta : \psi(\theta) = \psi(\infty)\}. \tag{3}$$

The decay rate function ψ in (2) will be said to satisfy the *auxiliary large deviations* (LD) *regularity conditions* if (4)–(7) below hold:

$$\beta^u > 0, \tag{4}$$

$$\psi \text{ is differentiable everywhere in } (-\infty, \beta^u), \tag{5}$$

$$\lim_{\theta \uparrow \beta^u} \psi'(\theta) = +\infty \text{ if } \psi(\beta^u) < \infty \quad (\psi \text{ is steep}), \text{ and} \tag{6}$$

$$\lim_{\theta \uparrow \beta^u} \psi(\theta) = \psi(\beta^u). \tag{7}$$

Let I be the associated *large deviations (LD) rate function* (Legendre-Fenchel transform of ψ) defined by

$$I(x) = \psi^*(x) \equiv \sup\{\theta x - \psi(\theta)\} \quad \text{for } x \in \mathbb{R}. \tag{8}$$

By the Gärtner [7]–Ellis [6] theorem, under conditions (2) and (4)–(7), the *large deviations principle* (LDP) holds for Z with large deviations rate function I; i.e.,

for each Borel set A

$$-\inf_{x \in A^o} I(x) \leq \varliminf_{t \to \infty} t^{-1} \log P(t^{-1}Z(t) \in A)$$

$$\leq \varlimsup_{t \to \infty} t^{-1} \log P(t^{-1}Z(t) \in A) \leq -\inf_{x \in \bar{A}} I(x), \tag{9}$$

where A^o and $\bar{A}$ are the interior and closure of A; see § IIB of Bucklew [2], § 2.3 of Dembo and Zeitouni [5] and § 3.1 of Shwartz and Weiss [14]. Moreover, the large deviation rate function I and the decay rate function ψ are *convex conjugates*, i.e., they are closed (lower semicontinuous) convex functions related by

$$\psi(\theta) = \psi^{**}(\theta) = I^*(\theta) \equiv \sup_x \{\theta x - I(x)\} \quad \text{for } \theta \in \mathbb{R}; \tag{10}$$

see p. 183 of Bucklew [2].

A typical LD rate function I is depicted in figure 1. Assuming that ψ is nondecreasing and convex with $\psi'(0) > 0$, then I is nonnegative and convex with $I(x) = +\infty$ for $x < 0$, $I(\psi'(0)) = 0$ and $I(x) \to \infty$ as $x \to \infty$. Hence, I is nondecreasing in the interval $[\psi'(0), \infty)$ and nonincreasing in the interval $(-\infty, \psi'(0)]$. Let γ^u and γ^l be the upper and lower limits of finiteness for I, i.e.,

$$\gamma^u = \sup\{x \leq \psi'(0) : I(x) < \infty\} \quad \text{and} \quad \gamma^l = \inf\{x \geq \psi'(0) : I(x) < \infty\}. \tag{11}$$

We first determine conditions under which the Gärtner–Ellis limits (2) for N and T are equivalent. All proofs appear in section 2. Let ψ^{-1} be the inverse of ψ when ψ is finite. It will be clear for this result, and later results, that T need not be discrete-time and N need not be integer valued. It suffices for N to be nonnegative
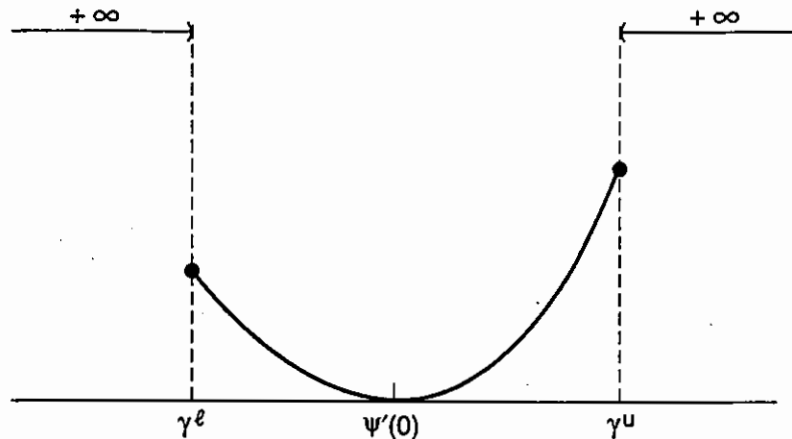

Fig. 1. A possible large deviations rate function I.

and nondecreasing; then we can relate the processes by the inverse map: $x^{-1}(t) = \inf\{s : x(s) > t\}$, $t > 0$.

THEOREM 1

If T satisfies (2) and (4)–(7), then N does too, with the possible exception of (2) for $\theta = \beta_N^u$ when $\psi_N(\beta_N^u) < \infty$. Similarly, if N satisfies (2) and (4)–(7), then T does too, with the possible exception of (2) for $\theta = \beta_T^u$ when $\psi_T(\beta_T^u) < \infty$. The decay rate functions are related by

$$\psi_N(\theta) = \begin{cases} -\beta_T^u, & \theta < \beta_N^l = -\psi_T(\beta_T^u), \\ -\psi_T^{-1}(-\theta), & \beta_N^l \leq \theta \leq \beta_N^u, \\ +\infty, & \theta > \beta_N^u = -\psi_T(\beta_T^l) = -\psi_T(-\infty) \end{cases} \tag{12}$$

and

$$\psi_T(\theta) = \begin{cases} -\beta_N^u, & \theta < \beta_T^l = -\psi_N(\beta_N^u), \\ -\psi_N^{-1}(-\theta), & \beta_T^l \leq \theta \leq \beta_T^u, \\ +\infty, & \theta > \beta_T^u = -\psi_N(\beta_N^l) = -\psi_N(-\infty) \end{cases} \tag{13}$$

for β_T^l, β_N^l, β_T^u and β_N^u defined by (3) with ψ_T and ψ_N, where $0 > \beta_T^l \geq -\infty$, $0 > \beta_N^l \geq -\infty$, $0 < \beta_T^u \leq \infty$ and $0 < \beta_N^u \leq \infty$. Moreover, the LD rate functions are related by

$$I_N(x) = \begin{cases} x I_T(1/x), & \gamma_N^l \leq x \leq \gamma_N^u, \\ +\infty, & \text{otherwise} \end{cases} \tag{14}$$

and

$$I_T(x) = \begin{cases} x I_N(1/x), & \gamma_T^l \leq x \leq \gamma_T^u, \\ +\infty, & \text{otherwise,} \end{cases} \tag{15}$$

where

$$\gamma_N^l = 1/\gamma_T^u, \quad \gamma_N^u = 1/\gamma_T^l, \tag{16}$$

$$I_N(0) = \lim_{x \to \infty} \frac{I_T(x)}{x} \quad \text{and} \quad I_N'(0) = +\infty \quad \text{if } \gamma_N^l = 0 \tag{17}$$

and

$$I_T(0) = \lim_{x \to \infty} \frac{I_N(x)}{x} \quad \text{and} \quad I_T'(0) = +\infty \quad \text{if } \gamma_T^l = 0. \tag{18}$$
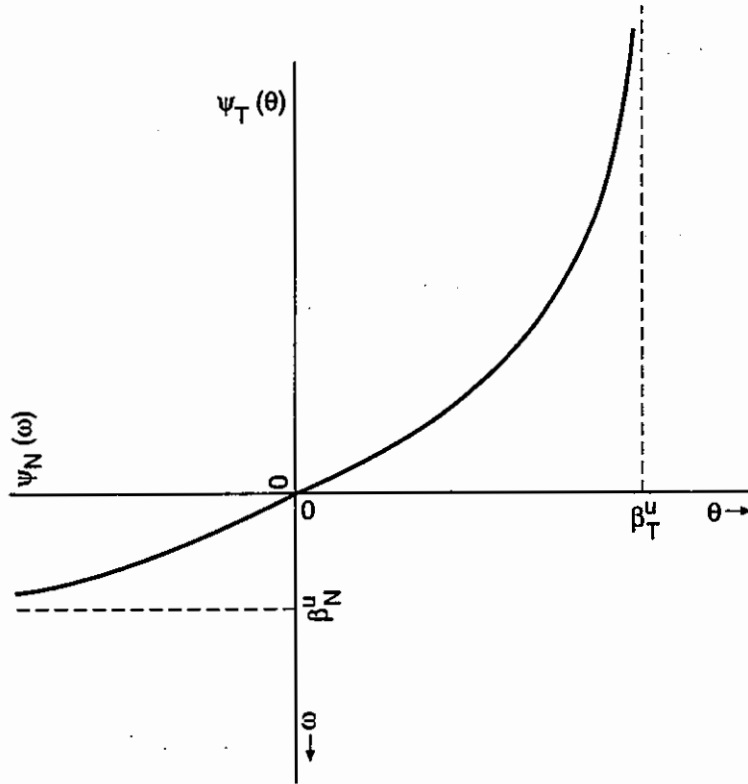
Fig. 2. Possible inverse decay rate functions ψ_T and $\psi_N = -\psi_T^{-1}(-\cdot)$ with finite asymptotes β_T and β_N.

The ambiguous behavior of (2) at the upper boundary points cannot occur if $\psi_N'(\theta) > 0$ and $\psi_T'(\theta) > 0$ for all θ in $(-\infty, 0]$. We could have included this condition with (4)–(7), but it is not required to get the LDP in (9).

The conditions of theorem 1 imply that one of the decay rate functions ψ_T and ψ_N is a closed convex function. The conclusion implies that both are. Figure 2 depicts the two *inverse decay rate functions* ψ_T and ψ_N on the same graph; ψ_T appears in the usual position, while ψ_N increases to the left with its argument ω increasing down.

To illustrate we give two simple examples. It is easy to see that the conditions of theorem 1 hold for these examples.

EXAMPLE 1

For a deterministic stationary rate-1 process, $T_n = n$ for all n, so that $\psi_T(\theta) = \psi_N(\theta) = \theta$, while $I_T(1) = I_N(1) = 0$ and $I_T(x) = I_N(x) = +\infty$ for $x \neq 1$.

EXAMPLE 2

For a rate-1 Poisson process, $\psi_T(\theta) = -\log(1 - \theta)$, $\theta < 1$, and $\psi_N(\theta) = e^\theta - 1$. Hence $I_N(x) = 1 - x + x \log x$, $x \geq 1$, and $I_N(x) = 1 - x - x \log x$, $0 \leq x \leq 1$; while $I_T(x) = x - 1 + \log x$, $0 \leq x \leq 1$, and $I_T(x) = x + \log x - 1$, $x \geq 1$.

We now give another example (suggested by a referee), which shows how we can have $\psi_T(\beta^u) < \infty$ with ψ_T steep as in (6).

EXAMPLE 3

Consider a renewal process with interrenewal-time density

$$g_p(t) = \begin{cases} C_p \lambda e^{-\lambda t}, & 0 \leq t < 1, \\ C_p t^{-p} \lambda e^{-\lambda t}, & t \geq 1, \end{cases}$$

where p is a nonnegative real number and C_p is chosen to make the total mass 1. The case $p = 0$ is the exponential density $\lambda e^{-\lambda t}$. Note that $g_{p_2}(t)/g_{p_1}(t)$ is nonincreasing in t for $p_1 < p_2$, so that the family of distributions $g_p(t)$ is stochastically decreasing in p. Moreover, $\beta^u_{T_p} = \lambda$ for all p, while $\psi_{T_p}(\beta^u_{T_p}) < \infty$ for $p > 1$ and $\psi'_{T_p}(\beta^u_{T_p}) < \infty$ for $p > 2$. Hence, for $1 < p \leq 2$, $\psi_{T_p}(\beta^u_{T_p}) < \infty$ with $\psi'_{T_p}(\beta^u_{T_p}) = \infty$, so that ψ_{T_p} is indeed steep as in (6). $\qquad\square$

Theorem 1 establishes an equivalence of LDPs in $\mathbb{R}$ for the inverse processes N and T. As in section 7 of [15], it is natural to approach this inverse property via functional or sample path LDPs in the function space $D[0,\infty)$ because in the function space setting the inverse property can be expressed directly as a continuous function with an appropriate topology on $D[0,\infty)$. A functional LDP analog of theorem 1 is established in Glynn et al. [8].

The processes N and T are easily related via their behavior in semi-infinite intervals; i.e.,

$$T_n \leq t \quad \text{if and only if} \quad N(t) \geq n. \tag{19}$$

From (19), we obtain for any $y > 0$ and $n \geq 1$,

$$n^{-1} \log P(n^{-1} T_n > y) = y(yn)^{-1} \log P((yn)^{-1} N(an) < y^{-1}). \tag{20}$$

From (20) we easily get the following equivalence result.

THEOREM 2

Let u be a nonincreasing function and let l be a nondecreasing function.
(a) There is convergence

$$n^{-1} \log P(n^{-1} T_n > y) \to u(y) \quad \text{as } n \to \infty \tag{21}$$

at all continuity points y of u if and only if

$$t^{-1} \log P(t^{-1} N(t) < y^{-1}) \to \tilde{l}(y^{-1}) \equiv y^{-1} u(y) \quad \text{as } t \to \infty \tag{22}$$

for all continuity points of y^{-1} of $\tilde{l}(y^{-1})$.
(b) There is convergence

$$n^{-1} \log P(n^{-1} T_n \leq y) \to l(y) \quad \text{as } n \to \infty \tag{23}$$

for all continuity points y of l if and only if

$$t^{-1} \log P(t^{-1} N(t) \geq y^{-1}) \to \tilde{u}(y^{-1}) \equiv y^{-1} l(y) \quad \text{as } t \to \infty \tag{24}$$

for all continuity points y^{-1} of $\tilde{u}(y^{-1})$.

As a relatively easy consequence of theorem 2, we can directly relate LDPs for N and T. For this purpose we say that the process Z satisfies a *partial LDP* if (9) holds for a proper subclass of the Borel subsets. We say that an LD rate function is *without flat spots* if for some $\bar{x}$ it is strictly decreasing where it is finite in $(-\infty, \bar{x})$ and strictly increasing where it is finite in $(\bar{x}, \infty)$.

THEOREM 3

Let I be a closed convex function on $\mathbb{R}$ without flat spots. A real-valued stochastic process Z satisfies an LDP with rate function I if and only if it satisfies a partial LDP with rate function I with respect to all semi-infinite intervals $(-\infty, y]$ and $[y, \infty)$.

We combine theorems 2 and 3 to relate the LDPs for T and N.

THEOREM 4

An LDP holds for T with lower semicontinuous rate function I_T without flat spots if and only if an LDP holds for N with lower semicontinuous rate function I_N without flat spots, where I_T and I_N are related by (14)–(18). The functions l and u

associated with T in theorem 2 are

$$u(y) = -\inf_{x \geq y} I(x) \quad \text{and} \quad l(y) = -\inf_{x \leq y} I_T(x), \tag{25}$$

and similarly for (N, I_N).

For example, theorem 4 and Cramér's theorem for partial sums of i.i.d. random variables in $\mathbb{R}$ in §2.2 of Dembo and Zeitouni [5] immediately imply that an LDP holds for the associated renewal counting process.

Given that conditions (2) and (4)–(7) for T or N directly imply that an LDP holds for T or N, theorem 4 implies that we get LDPs for both T and N under the conditions of theorem 1. Elementary convex analysis implies that the decay rate functions ψ_T and ψ_N and the LD rate functions I_T and I_N are related by (12)–(18), see section 2. The remaining step in the proof of theorem 1 (in section 2) is to prove that the Gärtner–Ellis limit (2) holds for both T and N.

So far we have shown how to relate LD asymptotics for T and N. Now we want to obtain general *sufficient conditions* for this LD asymptotics to hold for one of these processes. To do so, we will exploit regenerative structure. In particular, we will assume that $N(t)$ is a *cumulative process with respect to a sequence of regeneration times* $\{S_n : n \geq 0\}$ and $S_0 = 0$. (We could equally well start with $\{T_n\}$.) We will require that the distribution of $\tau_n = S(n) - S(n-1)$ be *spread out*; see p. 140 of Asmussen [1]. Our result is stronger than (2). It also applies to general cumulative processes. Another LD result for regenerative processes is in Kuczek and Crank [12]; they use different arguments.

Let $\phi(\theta, t)$ be the moment generating function of $N(t)$, i.e.,

$$\phi(\theta, t) = Ee^{\theta N(t)}, \quad t \geq 0. \tag{26}$$

If N is a cumulative process, then $\phi(\theta, \cdot)$ satisfies the *renewal equation*

$$\phi(\theta, t) = b(\theta, t) + \int_0^t \phi(\theta, t-s)G(\theta, ds), \tag{27}$$

where

$$b(\theta, t) = E[e^{\theta N(t)}; \tau_1 > t] \tag{28}$$

and

$$G(\theta, dt) = E[e^{\theta N(\tau_1)}; \tau_1 \in dt]. \tag{29}$$

Let $Y_i = N(S_i) - N(S_{i-1}), i \geq 1$.

THEOREM 5

Suppose that N is a cumulative process with respect to $\{S_n\}$ where τ_1 has a spread out distribution. If

(i)   there exists a root $\psi_N(\theta)$ to the equation

$$E[\exp(-\psi_N(\theta)\tau_1 + \theta Y_1)] = 1, \tag{30}$$

(ii)

$$E\left[\int_0^{\tau_1} \exp(-\psi_N(\theta)t + \theta N(t))dt\right] < \infty, \tag{31}$$

(iii)

$$\tilde{b}(\theta, t) \equiv E[\exp(-\psi_N(\theta)t + \theta N(t)); \tau_1 > t] < M \tag{32}$$

for some M and

(iv)

$$\tilde{b}(\theta, t) \to 0 \quad \text{as} \quad t \to \infty,$$

then

$$\phi_N(\theta, t) \sim \alpha_N(\theta)e^{\psi_N(\theta)t} \quad \text{as } t \to \infty, \tag{33}$$

where

$$\alpha_N(\theta) = \frac{E\left[\int_0^{\tau_1} \exp(-\psi_N(\theta)t + \theta N(t))dt\right]}{E[\tau_1 \exp(-\psi_N(\theta)\tau_1 + \theta Y_1)]}, \tag{34}$$

so that (2) holds.

In applications of theorem 5, it remains to verify conditions (i)–(iv) in theorem 5 and (4)–(7). It seems difficult to obtain good general results, but something can be said under strong conditions.

THEOREM 6

Suppose that $\{N(t)\}$ is a cumulative process with respect to $\{S_n\}$ and that τ_1 has a spread out distribution. In addition, suppose that $P(\tau_1 > K_1) = 0$ and $P(N(\tau_1) > K_2) = 0$ for some K_1 and K_2. Then a unique root $\psi_N(\theta)$ to (30) exists for all θ and assumptions (ii)–(iv) of theorem 5 hold for all θ. Moreover, ψ_N is

differentiable on $\mathbb{R}$, with derivative

$$\psi_N'(\theta) = \frac{-\dfrac{\partial}{\partial\theta} f_N(\psi_N(\theta), \theta)}{\dfrac{\partial}{\partial\gamma} f_N(\psi_N(\theta), \theta)}, \tag{35}$$

where

$$f_N(\gamma, \theta) \equiv E[\exp(-\gamma\tau_1 + \theta Y_1)]. \tag{36}$$

We can see the duality between N and T in the basic equation (30). In particular, if we switch the roles of S_n and $N(S_n)$, then $\tilde{S}_n \equiv N(S_n)$ may be regenerative times and $\tilde{T}_{\tilde{S}_n} - \tilde{T}_{\tilde{S}_{n-1}} \equiv S_n$ may be cycles associated with the inverse process T. When *both* N and T are cumulative processes this way, we call N and T *inverse cumulative processes*. Then we have versions of equation (30) for both processes, i.e., in addition to (30) for N, we have

$$E[\exp(-\psi_T(\theta) Y_1 + \theta\tau_1)] = 1. \tag{37}$$

It follows from (30) and (37) that the decay rate functions ψ_N and ψ_T must be related by $-\psi_T(-\psi_N(\theta)) = \theta$ for all θ where $\psi_N(\theta)$ is finite, i.e., which implies (12) and (13), which is consistent with theorem 1.

Chang [3] focuses on a discrete-time version of the point process N. The following comes from his example 2.2. Recall that a family of random variables $Z_1, \ldots, Z_k$ is *associated* if

$$E[f_1(Z_1) \ldots f_n(Z_n)] \geq E[f_1(Z_1)] \ldots E[f_n(Z_n)]$$

for all nondecreasing real-valued functions f_i.

THEOREM 7 (CHANG)

If N has stationary and associated increments, then $t^{-1} \log E e^{\theta N(t)}$ is nondecreasing in t and thus convergent, for each θ.

2.   Proofs

We prove the theorems in the order: 2, 3, 1, 5, 6. (No further proof is needed for theorem 4.)

Proof of theorem 2

Since $T_n > a$ if and only if $N(a) < n$,

$$P(n^{-1}T_n > a) = P(T_n > an) = P(N(an) < n),$$

from which (20) follows. From (20), we see that (21) holds if (22) holds, and (23) holds if (24) holds. To go the other way, note that, for any $\epsilon > 0$,

$$P(t^{-1}N(t) < a^{-1}) \le P(N(a\lfloor t/a \rfloor) < \lceil t/a \rceil) = P(T_{\lceil t/a \rceil} > a\lfloor t/a \rfloor)$$
$$\le P((\lceil t/a \rceil)^{-1}T_{\lceil t/a \rceil} > a - \epsilon)$$

when t is suitably large, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x and $\lceil x \rceil$ is the least integer greater than or equal to x. Hence, if (21) holds, then

$$\varlimsup_{t \to \infty} t^{-1} \log P(t^{-1}N(t) < a^{-1}) \le \varlimsup_{t \to \infty} t^{-1} \log P((\lceil t/a \rceil)^{-1}T_{\lceil t/a \rceil} > a - \epsilon)$$
$$= a^{-1}u(a - \epsilon).$$

Since ϵ was arbitrary and u is continuous at a,

$$\varlimsup_{t \to \infty} t^{-1} \log P(t^{-1}N(t) < a^{-1}) \le a^{-1}u(a).$$

Similarly,

$$P(t^{-1}N(t) < a^{-1}) \ge P(N(a\lceil t/a \rceil) < \lfloor t/a \rfloor) = P(T_{\lfloor t/a \rfloor} > a\lceil t/a \rceil)$$
$$\ge P((\lfloor t/a \rfloor)^{-1}T_{\lfloor t/a \rfloor} > a + \epsilon)$$

for t suitably large. Hence, if (21) holds, then

$$\varliminf_{t \to \infty} t^{-1} \log P(t^{-1}N(a) < a^{-1}) \ge a^{-1}u(a).$$

Hence, (22) holds. A similar argument shows that (23) implies (24). $\qquad\square$

Proof of theorem 3

We apply the characterizations of the LDP in (1.2.7) and (1.2.8) on p. 6 of Dembo and Zeitouni [5]. First we consider the upper bound. For any $\alpha < \infty$, let x_1 and x_2 be the lower and upper boundary points for the level set $\psi_I(\alpha)$ needed for the upper bound. (Here the notation $\psi_I(\alpha)$ follows [5].) By the lower

semicontinuity of I, any Γ with $\bar{\Gamma} \subseteq \psi_I(\alpha)^c$ has the property that

$$\Gamma \subseteq (-\infty, y_1] \cup [y_2, \infty),$$

where $y_1 < x_1$ and $y_2 > x_2$. Hence, for such Γ,

$$t^{-1} \log P(t^{-1}Z(t) \in \Gamma)$$
$$\leq t^{-1} \log P(t^{-1}Z(t) \in (-\infty, y_1] \cup [y_2, \infty))$$
$$\leq t^{-1} \log \max\{2P(t^{-1}Z(t) \leq y_1), 2P(t^{-1}Z(t) \geq y_2)\}$$
$$\leq (\log 2)/t + \max\{t^{-1}\log P(t^{-1}Z(t) \leq y_1), t^{-1}\log P(t^{-1}Z(t) \geq y_2)\},$$

so that

$$\varlimsup_{t \to \infty} t^{-1} \log P(t^{-1}Z(t) \in \Gamma)$$
$$\leq \max\left\{ \varlimsup_{t \to \infty} t^{-1} \log P(t^{-1}Z(t) \leq y_1), \varlimsup_{t \to \infty} t^{-1} \log P(t^{-1}Z(t) \geq y_2) \right\}$$
$$\leq \max\{-I(y_1), -I(y_2)\} \leq -\alpha,$$

since $y_1 < x_1 \leq \bar{x} \leq x_2 < y_2$.

Now we consider the lower bound. For any x in the domain of I, and any measurable Γ with $x \in \Gamma^o$, there is a neighborhood $(x - \delta_1, x + \delta_2) \subseteq \Gamma^o$. Let $\bar{x}$ be the location of the minimum of I and suppose that $x \leq \bar{x}$. (The argument when $x \geq \bar{x}$ is essentially the same.) For any ϵ given, choose δ_2 sufficiently small that $I(x + \delta_2) < I(x) + \epsilon$. Now

$$\varliminf_{t \to \infty} t^{-1} \log P(t^{-1}Z(t) \in \Gamma)$$
$$\geq \varliminf_{t \to \infty} t^{-1} \log P(t^{-1}Z(t) \in (x - \delta_1, x + \delta_2))$$
$$\geq \varliminf_{t \to \infty} t^{-1} \log(P(t^{-1}Z(t) \leq x + \delta_2) - P(t^{-1}Z(t) \leq x - \delta_1)).$$

However, for any ϵ',

$$P(t^{-1}Z(t) \leq x + \delta_2) \geq e^{-t(I(x+\delta_2)+\epsilon')}$$

and

$$P(t^{-1}Z(t) \leq x - \delta_1) \leq e^{-t(I(x-\delta_1)-\epsilon')}$$

for all suitably large t. Hence,

$$P(t^{-1}Z(t) \leq x + \delta_2) - P(t^{-1}Z(t) \leq x - \delta_1)$$
$$\geq e^{-t(I(x+\delta_2)+\epsilon')}(1 - e^{-t(I(x-\delta_1) - I(x+\delta_2) - 2\epsilon')}),$$

so that, after choosing ϵ', δ_1 and δ_2 so that $I(x - \delta_1) - I(x + \delta_2) - 2\epsilon' > 0$,

$$\varliminf_{t \to \infty} t^{-1} \log P(t^{-1}Z(t) \in \Gamma) \geq -I(x + \delta_2) - \epsilon' \geq -I(x) - \epsilon - \epsilon'.$$

Since ϵ and ϵ' were arbitrary,

$$\varliminf_{t \to \infty} t^{-1} \log P(t^{-1}Z(t) \in \Gamma) \geq -I(x). \qquad \square$$

In our proof of theorem 1 we use the following two lemmas.

LEMMA 1

For $\theta > 0$,

$$E \exp(\theta N(t)) = 1 + t\theta \int_0^\infty \exp(t\theta x) P(N(t) > tx) dx$$

and for $\theta < 0$,

$$E \exp(\theta N(t)) = -t\theta \int_0^\infty \exp(t\theta x) P(N(t) < tx) dx.$$

Proof

Note that

$$E \exp(\theta N(t)) - 1 = E \int_0^{N(t)/t} t\theta \exp(t\theta x) dx$$

$$= E t\theta \int_0^\infty \exp(t\theta x) I(N(t) > tx) dx$$

$$= t\theta \int_0^\infty \exp(t\theta x) P(N(t) > tx) dx.$$

For $\theta < 0$, observe that

$$t\theta \int_0^\infty \exp(t\theta x) P(N(t) > tx) dx = t\theta \int_0^\infty \exp(t\theta x)(1 - P(N(t) \leq tx) dx)$$

$$= \int_0^\infty t\theta \exp(t\theta x) dx - t\theta \int_0^\infty \exp(t\theta x) P(N(t) \leq tx) dx$$

$$= 1 - t\theta \int_0^\infty \exp(t\theta x) P(N(t) < tx) dx,$$

since $P(N(t) < tx) = P(N(t) \leq tx)$ almost surely with respect to Lebesgue measure. $\qquad \square$

LEMMA 2

If (2) and (4)–(7) hold for T, then

$$\varlimsup_{t \to \infty} t^{-1} \log E e^{\theta N(t)} < \infty$$

for $0 < \theta < -\psi_T(-\infty) \equiv \lim_{\theta \to -\infty} \psi_T(\theta)$.

Proof

For $0 < \theta < -\psi_T(-\infty) = \beta_N^u$, choose y so that $0 < y < \psi_T'(0)$ and $I_T(y) > \theta$. To see that this is possible, recall that I_T is continuous where it is finite and $I_T(0) = -\psi_T(-\infty)$ if $\gamma_T^l = 0$. (If $\gamma_T^l > 0$, then $I_T(y) = \infty$ for some y in this region.) Then, by the Gärtner–Ellis theorem for $\{T_n\}$,

$$\lim_{n \to \infty} n^{-1} \log P(T_n < yn) = -I(y) \equiv -I_T(y).$$

We assume that $I(y) < \infty$; a minor modification of the same argument treats the case $I(y) = \infty$. Hence, there exists n_0 such that for $n \geq n_0$

$$n^{-1} \log P(T_n < yn) \leq -I(y) + \epsilon,$$

where $\epsilon = (I(y) - \theta)/2$. Hence, for $n \geq n_0$,

$$P(T_n < yn) \leq \exp(-n(I(y) - \epsilon)). \tag{38}$$

Now

$$Ee^{\theta N(t)} = \sum_{n=0}^{\infty} e^{\theta n} P(N(t) = n)$$

$$\leq 1 + e^{\theta} \sum_{n=0}^{\infty} e^{\theta n} P(N(t) > n)$$

$$\leq 1 + e^{\theta} \sum_{n=0}^{\infty} e^{\theta n} P(T_n < t)$$

$$\leq 1 + e^{\theta} \sum_{n=0}^{\lfloor t/y \rfloor} e^{\theta n} P(T_n < t) + e^{\theta} \sum_{n=\lceil t/y \rceil}^{\infty} e^{\theta n} P(T_n < yn), \qquad (39)$$

where

$$\sum_{n=0}^{\lfloor t/y \rfloor} e^{\theta n} P(T_n < t) \leq \sum_{n=0}^{\lfloor t/y \rfloor} e^{\theta n} \leq e^{\theta \lceil t/y \rceil}/(e^{\theta} - 1) \qquad (40)$$

and, by (38), for $t > yn_0$,

$$\sum_{n=\lceil t/y \rceil}^{\infty} e^{\theta n} P(T_n < yn) \leq \sum_{n=\lceil t/y \rceil}^{\infty} \exp(\theta n - n(I(y) - \epsilon))$$

$$\leq \frac{\exp(-\lceil t/y \rceil (I(y) - \theta)/2)}{1 - \exp(-(I(y) - \theta)/2)} . \qquad (41)$$

Combining (39)–(41), we obtain the desired conclusion. □

Proof of theorem 1

We do the proof in only one direction, since the proofs in the two directions are similar. Suppose that $\{T_n\}$ satisfies (2) and (4)–(7) with decay rate function ψ_T. Then, by the Gärtner–Ellis theorem, $\{T_n\}$ obeys the LDP with LD rate function $I_T = \psi_T^*$. By theorem 4, $\{N(t)\}$ obeys the LDP with LD rate function I_N defined by (14)–(18). We then let $\psi_N = I_N^*$. Since $I_N^{**} = I_N$, it is easy to see that (12)–(18) are valid. For example, it is easy to see that ψ_N in (12) has the properties of a decay rate function (nondecreasing, convex, $\psi(0) = 0$ and (2)–(5)) if and only if ψ_T in (13) does. As indicated after theorem 1, this is easy to see from figure 2. More formally, to establish convexity, suppose that ψ_T is twice differentiable for $\alpha_T < \theta < \beta_T$

(where $\psi_T'(\theta) > 0$). Since $\psi_N(\theta) = -\psi_T^{-1}(-\theta)$, $\psi_T(-\psi_N(\theta)) = -\theta$ and

$$\psi_N'' = \frac{\psi_T''(\psi_T^{-1}(-\theta))}{\psi_T'(\psi_T^{-1}(-\theta))^3} \geq 0.$$

Then we can represent a general ψ_T as the limit of a sequence $\{\psi_{T_n} : n \geq 1\}$ where each ψ_{T_n} is strictly increasing and twice continuously differentiable in the interval (α_T, β_T). $\psi_{Nn}(\theta) \to \psi_N(\theta)$ as $n \to \infty$. Since ψ_{Nn} is convex for each n, so is ψ_N.

Given ψ_T and ψ_N in (12) and (13), it is straightforward to verify that the convex conjugates $I_T = \psi_T^*$ and $I_N = \psi_N^*$ defined by (6) have the properties (14)–(18). For example, for $\gamma_N' \leq x \leq \gamma_N^u$ and $x > 0$,

$$\begin{aligned}
\psi_N^*(x) &= \sup\{\theta x - \psi_N(\theta) : \theta \in \mathbb{R}\} \\
&= \sup\{\theta x - \psi_N(\theta) : \alpha_N \leq \theta \leq \beta_N\} \\
&= \sup\{\psi_N^{-1}(\theta)x - \theta : \alpha_N \leq \psi_N^{-1}(\theta) \leq \beta_N\} \\
&= \sup\{-\psi_T(\theta)x + \theta : \alpha_N \leq -\psi_T(\theta) \leq \beta_N\} \\
&= x\sup\{(\theta/x) - \psi_T(\theta) : \alpha_N \leq -\psi_T(\theta) \leq \beta_N\} \\
&= x\sup\{(\theta/x) - \psi_T(\theta) : \psi_T^{-1}(-\beta_N) \leq \theta \leq \psi_T^{-1}(-\alpha_N)\} \\
&= x\sup\{(\theta/x) - \psi_T(\theta) : \alpha_T \leq \theta \leq \beta_T\} \\
&= x\sup\{(\theta/x) - \psi_T(\theta) : \theta \in \mathbb{R}\} = x\psi_T^*(1/x).
\end{aligned}$$

For $\gamma_N' = 0 = x$, take the limit as $x \to 0$, obtaining

$$\psi_N^*(0) = \lim_{x \downarrow 0} \psi_N^*(x) = \lim_{y \uparrow \infty} y^{-1}\psi_T^*(y)$$

and

$$I_N'(0) = \lim_{x \to 0} \frac{\psi_N^*(x)}{x} = \lim_{y \to \infty} \psi_T^*(y) = \infty.$$

A similar argument yields $\psi_N = I_N^*$ and $\psi_T = I_T^*$ given I_N and I_T.

We now show that (2) holds for N when $0 < \theta < -\psi_T(-\infty) = \beta_N^u$. By lemma 1, it suffices to prove that

$$t^{-1} \log \int_0^\infty \exp(t\theta x) P(N(t) > tx)dx \to \psi_N(\theta). \tag{42}$$

By lemma 2, we can choose $\hat{\theta}$ with $\theta < \hat{\theta} < -\psi_T(-\infty)$ and

$$\xi \equiv \varlimsup_{t \to \infty} t^{-1} \log E e^{\hat{\theta} N(t)} < \infty.$$

For $\epsilon > 0$ given, let M be the constant

$$M = (\xi - \epsilon)/(\hat{\theta} - \theta).$$

Then, by Markov's inequality,

$$\int\limits_M^\infty \exp(t\theta x)P(N(t) > tx)dx \leq Ee^{\hat{\theta}N(t)} \int\limits_M^\infty \exp(t\theta x - t\hat{\theta}x)dx$$

$$\leq Ee^{\hat{\theta}N(t)}(e^{-t(\hat{\theta}-\theta)M})/t(\hat{\theta} - \theta)$$

$$\leq \frac{1}{t(\hat{\theta} - \theta)}\exp(-t(\hat{\theta} - \theta)M - \log Ee^{\hat{\theta}N(t)})$$

$$\leq \frac{1}{t(\hat{\theta} - \theta)}\exp(-t[(\hat{\theta} - \theta)M - \xi - \epsilon])$$

$$= \frac{1}{t(\hat{\theta} - \theta)}\exp(2\epsilon t) \tag{43}$$

for t sufficiently large. On the other hand,

$$\int\limits_0^M \exp(t\theta x)P(N(t) > tx)dx$$

$$= \sum_{i=0}^{n-1} \int\limits_{iM/n}^{(i+1)M/n} \exp(t\theta x)P(N(t) > tx)dx$$

$$\leq \sum_{i=0}^{n-1} \int\limits_{iM/n}^{(i+1)M/n} \exp(t\theta(i+1)M/n)P(N(t) > itM/n)dx$$

$$\leq \frac{M}{n}\sum_{i=0}^{n-1} \exp(t\theta(i+1)M/n)P(N(t) > itM/n)$$

$$\leq \frac{M}{n}e^{t\theta M/n}\sum_{i=0}^{n-1} \exp(t[\theta Mi/n - t^{-1}\log P(t^{-1}N(t) > iM/n)])$$

$$\leq \frac{M}{n}e^{t\theta M/n}\sum_{i=0}^{n-1} \exp\left(t[\theta Mi/n - \inf_{x \geq iM/n} I_N(x)]\right), \tag{44}$$

where n is an arbitrary positive integer. Combining (43) and (44), we obtain

$$\varlimsup_{t\to\infty} t^{-1}\log\int_0^\infty \exp(t\theta x)P(N(t)>tx)dx$$

$$\leq 2\epsilon + \theta\frac{M}{n} + \max_{0\leq i\leq n-1}\left\{\frac{\theta Mi}{n} - \inf_{x\geq iM/n} I_N(x)\right\}$$

by lemma 1.2.15 on p. 7 of Dembo and Zeitouni [5]. However,

$$\max_{0\leq i\leq n-1}\left\{\frac{\theta Mi}{n} - \inf_{x\geq iM/n} I_N(x)\right\}\leq \sup_x\{\theta x - I_N(x)\} = \psi_N(\theta).$$

Letting $\epsilon\to 0$ and $n\to\infty$, we complete the $\varlimsup$ proof.

We now turn to the lower bound. For the same θ and a new positive ϵ, choose δ and x_0 such that

$$\theta x - I_N(x) \geq \sup_y\{\theta y - I_N(y)\} - \epsilon$$

for $|x - x_0| < \delta$. Then

$$\int_0^\infty \exp(t\theta x)P(N(t)>tx)dx \geq \int_{x_0-\delta}^{x_0+\delta} \exp(t\theta x)P(N(t)>t(x_0+\delta))dx$$

$$\geq 2\delta\exp(t\theta(x_0-\delta))P(t^{-1}N(t)>x_0+\delta),$$

so that

$$\varliminf_{t\to\infty} t^{-1}\log\int_0^\infty \exp(t\theta x)P(N(t)>tx)dx$$

$$\geq \theta(x_0-\delta) - I_N(x_0+\delta) = \theta(x_0+\delta) - I_N(x_0+\delta) - 2\delta$$

$$\geq \sup_y\{\theta y - I_N(y)\} - 2\delta - \epsilon$$

$$\geq \psi_N(\theta) - 2\delta - \epsilon.$$

Finally, let $\delta\to 0$ and $\epsilon\to 0$ to complete the $\varliminf$ proof. Combining the $\varlimsup$ and $\varliminf$ proofs yields (42).

We now consider the case in which $\theta < 0$. By lemma 1, it suffices to show that

$$t^{-1} \log \int_0^\infty \exp(t\theta x) P(N(t) < tx) dx \to \psi_N(\theta) \quad \text{as } t \to \infty. \tag{45}$$

Let $\bar{x} = \psi'_N(0)$ and recall that $I_N(\bar{x}) = 0$. Let $x \vee y = \max\{x, y\}$. Then note that

$$\int_0^\infty \exp(t\theta x) P(N(t) < tx) dx$$

$$\leq \int_0^{\bar{x}} \exp(t\theta x) P(N(t) < tx) dx + \int_{\bar{x}}^\infty \exp(t\theta x) dx$$

$$\leq \sum_{i=0}^{n-1} (\bar{x}/n) \exp(t\theta\bar{x}i/n) P(N(t) < t\bar{x}(i+1)/n) - (t\theta)^{-1} \exp(t\theta\bar{x})$$

$$\leq \frac{\bar{x}}{n} \sum_{i=0}^{n-1} \exp(t\theta\bar{x}i/n - tI_N(\bar{x}(i+1)/n) + \epsilon t) - (t\theta)^{-1} \exp(t\theta\bar{x})$$

for t sufficiently large. Hence,

$$\varlimsup_{t \to \infty} t^{-1} \log \int_0^\infty \exp(t\theta x) P(N(t) < tx) dx$$

$$\leq \max_{1 \leq i \leq n-1} \{(\theta\bar{x}i/n) - I_N(\bar{x}(i+1)/n) + \epsilon\} \vee \theta\bar{x}$$

$$\leq \sup_{x \leq \bar{x}} \{\theta x - I_N(x) + \epsilon - \theta\bar{x}/n\} \vee \theta\bar{x}$$

$$\leq (\psi(\theta) + \epsilon - \theta\bar{x}/n) \vee \theta\bar{x}.$$

Now let $\epsilon \to 0$ and $n \to \infty$, and note that $\psi(\theta) > \theta\psi'(0)$ for $\theta < 0$, to complete the $\varlimsup$ proof.

We now turn to the lower bound. For $\epsilon > 0$ given, let $\delta > 0$ and x_0 be such that

$$\theta x - I_N(x) \geq \sup_y \{\theta y - I_N(y)\} - \epsilon.$$

Then

$$\int\limits_{0}^{\infty} \exp(t\theta x)P(N(t) < tx)dx \geq \int\limits_{x_0-\delta}^{x_0+\delta} \exp(t\theta x)P(N(t) < t(x_0 - \delta))dx$$

$$\geq 2\delta \exp(t\theta(x_0 + \delta))P(N(t) < t(x_0 - \delta)),$$

so that

$$\varliminf_{t\to\infty} t^{-1} \log \int\limits_{0}^{\infty} \exp(t\theta x)P(N(t) < tx)dx$$

$$\geq \theta(x_0 + \delta) - I_N(x_0 - \delta) = \theta(x_0 - \delta) - I_N(x_0 - \delta) + 2\theta\delta$$

$$\geq \sup_{y}\{\theta y - \bar{I}_N(y)\} - \epsilon + 2\theta\delta$$

$$\geq \psi_N(\theta) - \epsilon + 2\theta\delta.$$

Now let $\delta \to 0$ and $\epsilon \to 0$ to complete the proof of (50).

Finally, it remains to consider the upper boundary point β_N^u when $\beta_N^u < \infty$. Clearly,

$$\varliminf_{t\to\infty} t^{-1} \log Ee^{\beta_N^u N(t)} \geq \psi_N(\theta)$$

for any $\theta < \beta_N^u$. Thus, when $\psi_N(\beta_N^u) = \infty$,

$$\lim_{t\to\infty} t^{-1} \log Ee^{\beta_N^u N(t)} = \psi_N(\beta_N^u) = +\infty.$$

The only ambiguous case is when $\beta_N^u < \infty$ and $\psi_N(\beta_N^u) < \infty$. $\square$

Proof of theorem 5

In general, $G(\theta, \cdot)$ in (29) is not a proper probability distribution. However, our choice of $\psi(\theta)$ in (30) guarantees that

$$F(\theta, dt) \equiv \exp(-\psi(\theta)t)G(\theta, dt) \qquad (46)$$

is a probability distribution function. Furthermore, $F(\theta, dt)$ is equivalent to $P(\tau_1 \in dt)$, so that $F(\theta, \cdot)$ is spread out. Hence, we can apply Smith's key renewal

theorem, (4.4) on p. 120 of Asmussen [1], to the renewal equation

$$\tilde{\phi}(\theta, t) = \tilde{b}(\theta, t) + \int_0^t \tilde{\phi}(\theta, t - s) F(\theta, ds),$$ (47)

where $\tilde{b}(\theta, t)$ is in (32) and

$$\tilde{\phi}(\theta, t) = \exp(-\psi(\theta)t)\phi(\theta, t)$$ (48)

to obtain

$$\tilde{\phi}(\theta, t) \to \frac{\displaystyle\int_0^\infty \tilde{b}(\theta, s) ds}{\displaystyle\int_0^\infty t F(\theta, dt)} \quad \text{as } t \to \infty.$$ (49)

(Conditions (ii)–(iv) imply that $\tilde{b}(\theta, t)$ is directly Riemann integrable, using proposition 4.1(ii) of Asmussen [1]; see proposition 9 of Glynn and Whitt [10] for a related argument.) By Fubini's theorem, we see that

$$\int_0^\infty \tilde{b}(\theta, s) ds = \int_0^\infty E[\exp(-\psi(\theta)t + \theta N(t)); \tau_1 > t] dt$$

$$= E \int_0^\infty \exp(-\psi(\theta)t + \theta N(t)) I(\tau_1 > t) dt$$

$$= E\left[\int_0^{\tau_1} \exp(-\psi(\theta)t + \theta N(t)) dt\right]$$ (50)

and

$$\int_0^\infty t F(\theta, dt) = \int_0^\infty t E[\exp(-\psi(\theta)t + \theta N(t)); \tau_1 \in dt]$$

$$= E[\exp(-\psi(\theta)\tau_1 + \theta Y_1)\tau_1].$$ (51)

Combining (49)–(51), yields the desired (33) and (34). □

Proof of theorem 6

Under the boundedness assumptions, $f(\gamma, \theta)$ in (36) is bounded by $\exp(|\gamma|K_1 + |\theta|K_2)$ and infinitely differentiable in $\mathbb{R}^2$. Also, for each θ, $f(\cdot, \theta)$ is strictly decreasing with $f(\gamma, \theta) \to 0$ as $\gamma \to \infty$ and $f(\gamma, \theta) \to +\infty$ as $\gamma \to -\infty$. Hence, the root $\psi(\theta)$ of (30) exists for each θ. Moreover, it is easy to see that assumptions (ii)–(iv) hold.

To see that ψ is differentiable with derivative (35), apply the implicit function theorem with (30). Note that $\partial/\partial\gamma f(\gamma, \theta) < 0$ for all (γ, θ), so that the denominator is non-zero. $\qquad\square$

References

[1]   S. Asmussen, *Applied Probability and Queues* (Wiley, 1987).

[2]   J.A. Bucklew, *Large Deviation Techniques in Decision, Simulation and Estimation* (Wiley, 1990).

[3]   C.-S. Chang, Stability, queue length and delay of deterministic and stochastic queueing networks, IEEE Trans. Auto. Contr., to appear.

[4]   C.-S. Chang, P. Heidelberger, S. Juneja and P. Shahabuddin, Effective bandwidth and fast simulation of ATM intree networks, IBM T.J. Watson Research Center, Yorktown Heights, NY (1992).

[5]   A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications* (Jones and Bartlett, Boston, MA, 1992).

[6]   R. Ellis, Large deviations for a general class of random vectors, Ann. Prob. 12 (1984) 1–12.

[7]   J. Gärtner, On large deviations from the invariant measure, Theory Prob. Appl. 22 (1977) 24–39.

[8]   P.W. Glynn, A.A. Puhalskii and W. Whitt, A functional large deviation principle for first passage time processes, in preparation.

[9]   P.W. Glynn and W. Whitt, Ordinary CLT and WLLN versions of $L = \lambda W$, Math. Oper. Res. 13 (1988) 674–692.

[10]   P.W. Glynn and W. Whitt, Limit theorems for cumulative processes, Stoch. Proc. Appl. 47 (1993) 299–314.

[11]   P.W. Glynn and W. Whitt, Logarithmic asymptotics for steady-state tail probabilities in a single-server queue, in: *Studies in Applied Probability*, Festschrift in honor of Lajos Takács in honor of his 70th birthday, eds. J. Galambos and J. Gani, to appear.

[12]   T. Kuczek and K.N. Crank, A large-deviation result for regenerative processes, J. Theory Prob. 4 (1991) 551–561.

[13]   W.A. Massey and W. Whitt, Unstable asymptotics for nonstationary queues, Math. Oper. Res. 19 (1994) 267–291.

[14]   A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis: Queues, Communication and Computing*, in preparation.

[15]   W. Whitt, Some useful functions for functional limit theorems, Math. Oper. Res. 5 (1980) 67–85.

[16]   W. Whitt, Tail probabilities with statistical multiplexing and effective bandwidths in multiclass queues, Telecommun. Syst. 2 (1993) 71–107.