*The Annals of Applied Probability*
2011, Vol. 21, No. 2, 645–668
DOI: 10.1214/10-AAP707
© Institute of Mathematical Statistics, 2011

# ON THE TRANSITION FROM HEAVY TRAFFIC TO HEAVY TAILS FOR THE M/G/1 QUEUE: THE REGULARLY VARYING CASE

BY MARIANA OLVERA-CRAVIOTO, JOSE BLANCHET AND PETER GLYNN

*Columbia University, Columbia University and Stanford University*

Two of the most popular approximations for the distribution of the steady-state waiting time, $W_\infty$, of the M/G/1 queue are the so-called heavy-traffic approximation and heavy-tailed asymptotic, respectively. If the traffic intensity, $\rho$, is close to 1 and the processing times have finite variance, the heavy-traffic approximation states that the distribution of $W_\infty$ is roughly exponential at scale $O((1-\rho)^{-1})$, while the heavy tailed asymptotic describes power law decay in the tail of the distribution of $W_\infty$ for a fixed traffic intensity. In this paper, we assume a regularly varying processing time distribution and obtain a sharp threshold in terms of the tail value, or equivalently in terms of $(1-\rho)$, that describes the point at which the tail behavior transitions from the heavy-traffic regime to the heavy-tailed asymptotic. We also provide new approximations that are either uniform in the traffic intensity, or uniform on the positive axis, that avoid the need to use different expressions on the two regions defined by the threshold.

**1. Introduction.** A substantial literature has been developed over the last forty years that recognizes the simplifications that arise in the analysis of queueing systems in the presence of "heavy traffic." The earliest such "heavy traffic" approximation was that obtained by Kingman (1961, 1962) for the steady-state waiting time $W_\infty$ for the G/G/1 queue. In particular, let $W_n$ be the waiting time (exclusive of service) of the $n$th customer for a first-in first-out (FIFO) single-server queue (with an infinite capacity waiting room) fed by a renewal arrival process [with i.i.d. inter-arrival times $(\chi_n : n \geq 1)$] and an independent stream of i.i.d. processing times $(V_n : n \geq 0)$. If $\rho \triangleq EV_1/E\chi_1 < 1$, then $W_n \Rightarrow W_\infty$ as $n \to \infty$, where $W_\infty$ can be approximated via

$$W_\infty \overset{\mathcal{D}}{\approx} \frac{\operatorname{Var}\chi_1 + \operatorname{Var}V_1}{2(E\chi_1 - EV_1)} \operatorname{Exp}(1) \tag{1.1}$$

when $\rho$ is close to 1. Here, Exp(1) is an exponential r.v. with mean one and $\overset{\mathcal{D}}{\approx}$ denotes "has approximately the same distribution as." A precise statement of the limit theorem supporting the heavy traffic approximation (1.1) is given by (2.4)

Received January 2009; revised September 2010.
*MSC2010 subject classifications.* Primary 41A60; secondary 60F10, 60F05, 60G10, 60G50.
*Key words and phrases.* M/G/1 queue, heavy traffic, heavy tails, uniform approximations, large deviations.

below. The term "heavy traffic" arises as a consequence of the fact that long queues tend to form when $EV_1$ and $E\chi_1$ are roughly balanced.

The modern approach to justifying (1.1) involves first showing that $(W_n : n \geq 0)$ can be approximated in heavy traffic by a one-dimensional reflecting Brownian motion (RBM) [see, e.g., Iglehart and Whitt (1970a, 1970b)] and then verifying that the steady-state r.v. $W_\infty$ can be approximated by that of the RBM [Szczotka (1990, 1999)]. Similar methods apply, in significant generality, to multi-station queueing networks. For example, Reiman (1984) proves a functional limit theorem that justifies approximating a single class multi-station queueing network by multi-dimensional RBM. Recent work of Gamarnik and Zeevi (2006) establishes the associated steady-state convergence. Harrison and Williams (1987) analyze the multi-dimensional RBM and show that it has exponential tails.

On the other hand, if the processing times are heavy-tailed (e.g., regularly varying), there is a significant literature that establishes, for various models, that the associated queueing system possesses a heavy-tailed steady-state. A representative result of this type states that when $\rho < 1$ for the G/G/1 FIFO queue described above (with regularly varying processing times), we have

$$P(W_\infty > x) \sim \frac{\lambda}{1-\rho} \int_x^\infty P(V_1 > y)\, dy \tag{1.2}$$

as $x \to \infty$, where $\lambda \stackrel{\triangle}{=} 1/E\chi_1$ [see, e.g., Embrechts and Veraverbeke (1982)]. Corresponding heavy-tailed steady-state asymptotics also exist in the context of queueing networks [see, e.g., Baccelli, Schlegel and Schmidt (1999) and Baccelli and Foss (2004)].

At first, it may seem contradictory that the heavy-traffic theory typically predicts exponential tails for the steady-state distribution, whereas regularly varying heavy-tailed asymptotics predict power-law decay in the steady-state tail. Of course, the key is to note that the two families of results involve different types of limits, one as $\rho \to 1$ (heavy traffic) and the other as $x \to \infty$ (heavy tails). The interesting mathematical issue here is therefore to send $\rho$ to 1 and $x \to \infty$ simultaneously, and to determine the $x$-value (as a function of $\rho$) at which the steady-state distribution begins to "feel" the presence of the heavy tails in the processing times. In particular, this paper develops a very explicit description, in the setting of the M/G/1 queue (in which the arrival process is assumed Poisson), of where the transition from the exponential heavy-traffic approximation (1.1) to the heavy-tailed approximation (1.2) occurs. As a corollary to our main results (Corollary 2.3) we find that when the processing times are regularly varying, then the tail probability $P(W_\infty > x)$ sharply transitions at

$$x^* \approx \frac{1}{1-\rho} \log\left(\frac{1}{1-\rho}\right) \frac{EV_1^2}{2EV_1} (\alpha - 2) \tag{1.3}$$

from the approximation (1.1) to the approximation (1.2) (where $\alpha$ is the tail index of the regularly varying $V_1$). Roughly speaking, to the left of $x^*$, (1.1) is valid whereas to the right of $x^*$, (1.2) is appropriate. A companion paper [Olvera-Cravioto and Glynn (2010)] provides uniform approximations for $P(W_\infty > x)$ in the general subexponential case, and shows how in the setting of Weibullian tails one can identify an intermediate zone in which neither the heavy-traffic asymptotic nor the heavy-tailed asymptotic hold.

This result ties together two significant queueing theory literatures, namely heavy traffic theory and heavy-tailed approximations. As the first such result describing the transition from the heavy traffic regime to the heavy-tailed asymptotic, it suggests the possibility of similar such results for more complex systems and networks. Furthermore, one of our main results, Theorem 2.1, provides an approximation for the tail probability $P(W_\infty > x)$ that is uniform across all values of $\rho$, and that in numerical experiments seems to perform very well. This new uniform approximation, which takes advantage of the Pollaczek–Khintchine formula for the M/G/1 queue, provides a significant numerical improvement over the existing heavy-traffic and heavy-tail approximations that are commonly used to approximate the tail of the r.v. $W_\infty$.

**2. The main results.** Let $(W_n(\rho) : n \geq 0)$ be the waiting time sequence for an M/G/1 FIFO queue that is fed by a Poisson arrival process having arrival rate $\lambda = \rho/EV_1$ and independent i.i.d. processing times $(V_n : n \geq 0)$. We assume throughout the remainder of this paper (unless otherwise noted) that $V_1$ has a regularly varying distribution with tail index $\alpha > 2$, so that

$$P(V_1 > x) \sim x^{-\alpha} L(x)$$

as $x \to \infty$, where $L(\cdot)$ is slowly varying [see page 412 of Asmussen (2003)].

If $\rho < 1$, $W_n(\rho) \Rightarrow W_\infty(\rho)$ as $n \to \infty$, where the Pollaczek–Khintchine formula [see, e.g., page 237 of Asmussen (2003)] guarantees that

$$P\bigl(W_\infty(\rho) > \cdot\bigr) = \sum_{n=0}^{\infty} (1 - \rho)\rho^n P(S_n > \cdot). \tag{2.1}$$

Here, $S_n = X_1 + \cdots + X_n$ (with $S_0 = 0$), where the $X_j$'s are i.i.d. with common density $g(\cdot) = P(V_1 > \cdot)/EV_1$. The heavy-tail result (1.2) translates, in the M/G/1 setting, into the asymptotic

$$P\bigl(W_\infty(\rho) > x\bigr) \sim \frac{\rho}{1 - \rho} P(X_1 > x), \tag{2.2}$$

as $x \to \infty$. It is straightforward [see, e.g., page 404 of Asmussen (2003)] to show that (2.2) in turn implies that

$$P\bigl(W_\infty(\rho) > x\bigr) \sim \frac{\lambda}{1 - \rho} \cdot \frac{x^{1-\alpha}}{\alpha - 1} L(x) \tag{2.3}$$

as $x \to \infty$.

Turning next to the heavy traffic limit theorem for $W_\infty(\rho)$ [due to Kingman (1961)], its precise statement (in our M/G/1 setting) is that

$$(1-\rho)W_\infty(\rho) \Rightarrow \frac{EV_1^2}{2EV_1}\operatorname{Exp}(1) \tag{2.4}$$

as $\rho \nearrow 1$, providing theoretical support for the approximation

$$P\bigl(W_\infty(\rho) > x\bigr) \approx \exp\bigl(-2(1-\rho)EV_1 x/EV_1^2\bigr) \tag{2.5}$$

when $\rho$ is close to 1. To get a sense of the point $x^* = x^*(\rho)$ at which the heavy traffic approximation (2.5) transitions into the heavy-tail approximation (2.3), note that the point $x^*$ at which the exponential (2.5) crosses the power law tail (2.3) must satisfy

$$2x^*(1-\rho)\frac{EV_1}{EV_1^2} \approx \log(1-\rho) + (\alpha-1)\log x^*. \tag{2.6}$$

This implies that $x^* \approx \kappa(1-\rho)^{-1}\log((1-\rho)^{-1})$, where $\kappa = (\alpha-2)EV_1^2/(2EV_1)$.

To make the above heuristic rigorous we look more closely at the Pollaczek–Khintchine formula. First we note that the heavy-tail asymptotic (2.2) can be obtained by simply substituting $P(S_n > \cdot)$ by $nP(X_1 > \cdot)$, that is, by using the so-called subexponential asymptotic for $P(S_n > x)$. Such asymptotics are typically stated for fixed values of $n$, but can be shown to hold for $n \to \infty$ provided $n$ grows slowly enough compared to $x$ [see, e.g., Borovkov (2000); Rozovskiĭ (1989)]. In other words, we can obtain the heavy-tail asymptotic from the first terms of (2.1),

$$\sum_{n=1}^{N(x)}(1-\rho)\rho^n P(S_n > x) \approx \sum_{n=1}^{N(x)}(1-\rho)\rho^n nP(X_1 > x) \sim \frac{\rho}{1-\rho}P(X_1 > x)$$

for some appropriately defined $N(x)$. This raises the question of whether we can also obtain the heavy-traffic asymptotic directly from (2.1), and the answer is yes. For large $n$, say $n \ge x/EX_1$, $P(S_n > x) = O(1)$, so by simply replacing $P(S_n > x)$ by one we obtain

$$\sum_{n=[x/EX_1]}^{\infty}(1-\rho)\rho^n P(S_n > x) \approx \sum_{n=[x/EX_1]}^{\infty}(1-\rho)\rho^n = \rho^{[x/EX_1]}.$$

Since as $\rho \nearrow 1$, $\rho^{[x/EX_1]} \sim e^{-(1-\rho)x/EX_1} = e^{-2(1-\rho)EV_1 x/EV_1^2}$, we can recover the heavy-traffic asymptotic from the last terms of (2.1).

This reasoning leads us to the observation that the transition of $P(W_\infty > x)$ occurs at the level of the partial sums $P(S_n > x)$. For the regularly varying case, the transition from the subexponential asymptotic $nP(X_1 > x - nEX_1)$ to the CLT approximation $1 - \Phi((x - nEX_1)/\sqrt{\operatorname{Var}(X_1)})$ [or its stable law counterpart when $\operatorname{Var}(X_1) = \infty$] occurs smoothly, which allows us to approximate the Pollaczek–Khintchine formula directly and obtain an expression that does not require $\rho$ to be close to one. Theorem 2.1 below describes this (uniform in $\rho$) approximation, and Theorem 2.2 gives an equivalent formulation in terms of more familiar asymptotic expressions. As corollaries, we obtain the result regarding the transition from heavy-traffic to heavy-tail of $P(W_\infty > x)$, both in terms of $\rho$ as a function of $x$ and $x$ as a function of $\rho$.

We also point out that similar versions of our results should also hold for the GI/GI/1 case. The added difficulty lies in the fact that although $W_\infty(\rho)$ still has a representation of the form

$$W_\infty(\rho) = \sum_{n=1}^{\infty}(1-\theta)\theta^n P(Y_1 + \cdots + Y_n > x),$$

where the $Y_i$'s i.i.d regularly varying random variables [see Asmussen (2003), Chapter X.9], the distribution of the $Y_i$'s and the geometric parameter $\theta$ are not explicitly known. In particular, both of them depend on $\rho$, so a uniform in $\rho$ version of Theorem 3.1 and an asymptotic expression for $\theta(\rho)$ are required. Such uniform in $\rho$ results have been recently developed in Blanchet, Glynn and Lam (2010). Proof techniques very similar to those given here can then be used to obtain the GI/GI/1 equivalents of our results.

THEOREM 2.1. *Suppose* $P(V_1 > x) \sim L(x)x^{-\alpha}$ *with* $\alpha > 2$ *and let* $\mu = EX_1 = EV_1^2/(2EV_1)$. *Define* $\beta = (2 \wedge (\alpha - 1))^{-1}$, $M(x) = \lfloor (x - x^\beta)/\mu \rfloor$, *and*

$$S(\rho, x) = \sum_{n=1}^{M(x)} (1-\rho)\rho^n n P\big(X_1 > x - (n-1)\mu\big).$$

*Then*,

$$\sup_{0<\rho<1}\left|\frac{P(W_\infty(\rho) > x)}{S(\rho,x) + \rho^{x/\mu}} - 1\right| \to 0$$

*as* $x \to \infty$. *Alternatively*,

$$\sup_{x\geq 0}\left|\frac{P(W_\infty(\rho) > x)}{S(\rho,x) + \rho^{x/\mu}} - 1\right| \to 0$$

*as* $\rho \nearrow 1$.

THEOREM 2.2. *Suppose* $P(V_1 > x) \sim L(x)x^{-\alpha}$ *with* $\alpha > 2$ *and let* $\mu = EX_1 = EV_1^2/(2EV_1)$ *and* $\gamma(x,\rho) = 1 - \rho^{x/\mu} - \rho^{x/\mu}(1-\rho)x/\mu$. *Then*,

$$\sup_{0<\rho<1}\left|\frac{P(W_\infty(\rho) > x)}{(\rho/(1-\rho))\gamma(x,\rho)P(X_1 > x) + \rho^{x/\mu}} - 1\right| \to 0$$

*as* $x \to \infty$. *Alternatively*,

$$\sup_{x\geq 0}\left|\frac{P(W_\infty(\rho) > x)}{(\rho/(1-\rho))\gamma(x,\rho)P(X_1 > x) + \rho^{x/\mu}} - 1\right| \to 0$$

*as* $\rho \nearrow 1$.

From Theorem 2.2 we can derive the following corollary stating the different regions where either the heavy-traffic approximation or the heavy-tail asymptotic govern the tail behavior of the steady-state waiting time. Corollary 2.3 describes the shape of the distribution of $W_\infty(\rho)$ for a fixed value of $\rho$. On the other hand, Corollary 2.4 can be of practical use in understanding the sensitivity of a system to the traffic intensity, since for a fixed value of $x$ it tells us how $P(W_\infty(\rho) > x)$ changes as $\rho$ gets closer to one.

COROLLARY 2.3. *Suppose* $P(V_1 > x) \sim L(x)x^{-\alpha}$ *with* $\alpha > 2$ *and let* $\kappa = (\alpha - 2)EV_1^2/(2EV_1)$. *Suppose that* $y = y(\rho)$ *satisfies*

$$y(\rho) = c\kappa(1-\rho)^{-1}\log\bigl((1-\rho)^{-1}\bigr)$$

*for* $\rho < 1$.

(a) *If* $0 < c < 1$, *then*

$$\sup_{0\le x\le y}\left|\frac{P(W_\infty(\rho) > x)}{\exp(-2(1-\rho)EV_1x/EV_1^2)} - 1\right| \to 0 \tag{2.7}$$

*as* $\rho \nearrow 1$. *Relation* (2.7) *continues to hold when* $c = 1$, *provided that* $L(x)/(\log x)^{\alpha-1} \to 0$ *as* $x \to \infty$.

(b) *If* $c > 1$, *then*

$$\sup_{x\ge y}\left|\frac{P(W_\infty(\rho) > x)}{(\rho/(1-\rho))P(X_1 > x)} - 1\right| \to 0 \tag{2.8}$$

*as* $\rho \nearrow 1$. *Relation* (2.8) *continues to hold when* $c = 1$, *provided that* $L(x)/(\log x)^{\alpha-1} \to \infty$ *as* $x \to \infty$.

The corresponding version in terms of $\rho$ as a function of $x$ is given below.

COROLLARY 2.4. *Suppose* $P(V_1 > x) \sim L(x)x^{-\alpha}$ *with* $\alpha > 2$ *and let* $\kappa = (\alpha - 2)EV_1^2/(2EV_1)$. *Suppose that* $\hat\rho = \hat\rho(x)$ *satisfies*

$$\hat\rho(x) = 1 - c\kappa(\log x)/x.$$

(a) *If* $0 < c < 1$, *then*

$$\sup_{\hat\rho\le\rho<1}\left|\frac{P(W_\infty(\rho) > x)}{\exp(-2(1-\rho)EV_1x/EV_1^2)} - 1\right| \to 0 \tag{2.9}$$

*as* $x \to \infty$. *Relation* (2.9) *continues to hold when* $c = 1$, *provided that* $L(x)/\log x \to 0$ *as* $x \to \infty$.

(b) *If* $c > 1$, *then*

$$\sup_{0<\rho\le\hat{\rho}} \left| \frac{P(W_\infty(\rho) > x)}{(\rho/(1-\rho))P(X_1 > x)} - 1 \right| \to 0 \tag{2.10}$$

*as* $x \to \infty$. *Relation* (2.10) *continues to hold when* $c = 1$, *provided that* $L(x)/\log x \to \infty$ *as* $x \to \infty$.

Note that Theorems 2.1 and 2.2 suggest different approximations for $P(W_\infty(\rho) > x)$. We tested both approximations and found that

$$H(\rho, x) = S(\rho, x) + \rho^{x/\mu}$$

is better than its asymptotic counterpart and performs very well for most values of $x$ and $\rho$. In Section 4 we analyze how this approximation compares to using the simpler heavy-traffic and heavy-tail asymptotics in the regions where they are valid, and we give a couple of numerical examples.

It is instructive to contrast the behavior obtained in the above regularly varying setting with what occurs in the light-tailed setting. Suppose, in particular, that $E\exp(\theta V_1) < \infty$ for some $\theta > 0$, and define $\theta^*(\rho)$ as the root of $\rho E\exp(\theta^*(\rho)V_1) = 1$.

THEOREM 2.5. *Suppose that* $E\exp(\theta V_1) < \infty$ *for some* $\theta > 0$.

(a) *If* $y = y(\rho) = o((1-\rho)^{-2})$, *then*

$$\sup_{0\le x\le y} \left| \frac{P(W_\infty(\rho) > x)}{\exp(-2(1-\rho)EV_1 x/EV_1^2)} - 1 \right| \to 0$$

*as* $\rho \nearrow 1$.

(b) *For* $x \ge 0$,

$$P\big(W_\infty(\rho) > x(1-\rho)^{-2}\big) \sim \exp\biggl(-2x(1-\rho)^{-1}\frac{EV_1}{EV_1^2} + x\frac{EV_1^3}{3EV_1^2} - \frac{x}{4}\frac{EV_1^2}{EV_1}\biggr)$$

*as* $\rho \nearrow 1$.

(c) *As* $\rho \nearrow 1$,

$$\sup_{x\ge 0} \left| \frac{P(W_\infty(\rho) > x)}{\exp(-\theta^*(\rho)x)} - 1 \right| \to 0.$$

Note that in contrast to the heavy-tailed setting, the heavy traffic approximation is now valid over a larger range, namely up to tail values of order $o((1-\rho)^{-2})$. At tail values of order $(1-\rho)^{-2}$, the third moment of $V_1$ enters the asymptotic for $P(W_\infty(\rho) > x)$ [see also Abate, Choudhury and Whitt (1995) and Blanchet and Glynn (2007)]. Finally, part (c) shows that the Cramér–Lundberg tail asymptotic [see, e.g., pages 365–369 of Asmussen (2003)] is globally valid in heavy traffic, showing the clear superiority of the Cramér–Lundberg asymptotic over the heavy traffic approximation when $\rho$ is close to 1. On the other hand, for regularly varying tails, any global approximation to $P(W_\infty(\rho) > \cdot)$ must utilize both the heavy traffic approximation and the appropriate tail asymptotic.

We close this section with a brief discussion of how the theory described in this paper extends to the more general setting of geometric random sums. Specifically, consider the random variable

$$Z(p) = \sum_{i=1}^{N(p)} Y_i,$$

where $(Y_i : i \geq 1)$ is a sequence of nonnegative nonlattice i.i.d. random variables independent of the geometric r.v. $N(p)$ having mass function

$$P\big(N(p) = k\big) = (1-p)p^{k-1}$$

for $k \geq 1$ [see Kalashnikov (1997) for various applied settings in which such geometric random sums arise]. We assume that $Y_1$ is regularly varying with finite variance, so that there exists $\beta > 2$ and a slowly varying function $L(\cdot)$ for which

$$P(Y_1 > x) \sim x^{-\beta} L(x)$$

as $x \to \infty$. Put $\tau = (\beta - 1)EY_1$.

THEOREM 2.6. *Let* $\mu = EY_1$ *and* $\gamma(x,p) = 1 - (1-p)^{x/\mu} - (1-p)^{x/\mu} px/\mu$. *Then,*

$$\sup_{0<p<1} \left| \frac{P(Z(p) > x)}{(1-p)\gamma(x,p)P(Y_1 > x)/p + (1-p)^{x/\mu}} - 1 \right| \to 0$$

*as* $x \to \infty$. *Alternatively,*

$$\sup_{x>0} \left| \frac{P(Z(p) > x)}{(1-p)\gamma(x,p)P(Y_1 > x)/p + (1-p)^{x/\mu}} - 1 \right| \to 0$$

*as* $p \downarrow 0$.

COROLLARY 2.7. *Suppose that* $y = y(p)$ *satisfies*

$$y(p) = c\tau p^{-1} \log(1/p)$$

(a) *If* $0 < c < 1$, *then*

$$\sup_{0 \leq x \leq y} \left| \frac{P(Z(p) > x)}{\exp(-px/EY_1)} - 1 \right| \to 0 \tag{2.11}$$

*as* $p \downarrow 0$. *Relation* (2.11) *continues to hold when* $c = 1$, *provided that* $L(x)/(\log x)^\beta \to 0$ *as* $x \to \infty$.

(b) *If* $c > 1$, *then*

$$\sup_{x \geq y} \left| \frac{P(Z(p) > x)}{P(Y_1 > x)/p} - 1 \right| \to 0 \tag{2.12}$$

*as* $p \downarrow 0$. *Relation* (2.12) *continues to hold when* $c = 1$, *provided that* $L(x)/(\log x)^{\beta} \to \infty$ *as* $x \to \infty$.

**3. Proofs.** In this section, we prove Theorems 2.1, 2.2 and Corollary 2.3; the proofs of Theorem 2.6 and Corollary 2.7 are essentially identical to those of Theorem 2.2 and Corollary 2.3. The proof of Corollary 2.4 is very similar in spirit to that of Corollary 2.3, the difference being that it follows from the uniform in $0 < \rho < 1$ statement of Theorem 2.2 instead of the uniform in $x > 0$. Theorem 2.5 follows directly from Theorem 2 in Blanchet and Glynn (2007).

We now turn our attention to the proof of Theorem 2.1. Recall that

$$P\bigl(W_\infty(\rho) > x\bigr) = \sum_{n=0}^{\infty} (1-\rho)\rho^n P(S_n > x), \tag{3.1}$$

where the $X_i$'s are i.i.d. with common density $g(\cdot) = P(V_1 > \cdot)/EV_1$ and $S_n = X_1 + \cdots + X_n$.

Our analysis is based on the principle that we can approximate $P(S_n > x)$ by the heavy tail asymptotic $nP(X_1 > x - (n-1)E[X])$ uniformly in $n$ throughout the region of large deviations of $S_n$. Early results of this kind are due to Nagaev (1981), Rozovskiĭ (1989), Mikosch and Nagaev (1998), Borovkov (2000), and more recently, Denisov, Dieker and Shneer (2008). The statement we present below is taken from Borovkov and Borovkov (2008), Theorems 3.4.1 and 4.4.1.

THEOREM 3.1 (Borovkov). *Let* $Y_1, Y_2, \ldots$ *be i.i.d. random variables having* $EY = 0$, $\overline{F}(t) = P(Y_1 > t)$ *and* $\overline{F}(t) = t^{-\beta} L(t)$ *where* $L(\cdot)$ *is slowly varying. Set* $S_n = Y_1 + \cdots + Y_n$, $n \geq 1$.

(a) *If* $\beta > 2$ *and* $EY^2 < \infty$, *define* $\sigma(n) = \sqrt{(\beta - 2)n \log n}$.
(b) *If* $\beta \in (1, 2)$ *and* $F(-t) \leq c\overline{F}(t)$ *for* $t > 0$ *and some constant* $c > 0$, *define* $\sigma(n) = \overline{F}^{-1}(1/n)$.

*Then, there exists a function* $\varphi(t) \downarrow 0$ *as* $t \uparrow \infty$ *such that*

$$\sup_{y \geq t\sigma(n)} \left| \frac{P(S_n > y)}{nP(Y_1 > y)} - 1 \right| \leq \varphi(t)$$

*uniformly in* $n$.

Below we give an application of Borovkov's result to our particular setting.

LEMMA 3.2. *Let $X_1, X_2, \dots$ be i.i.d. nonnegative random variables with $\mu = E[X] < \infty$, and $P(X_1 > t) = t^{-\alpha+1}L(t)$ where $L(\cdot)$ is slowly varying and $\alpha > 2$. Set $S_n = X_1 + \cdots + X_n$, $n \geq 1$. For any $(2 \wedge (\alpha - 1))^{-1} < \gamma < 1$ define $M_\gamma(x) = \lfloor (x - x^\gamma)/\mu \rfloor$. Then, there exists a function $\varphi(t) \downarrow 0$ as $t \uparrow \infty$ such that*

$$\sup_{1 \leq n \leq M_\gamma(x)} \left| \frac{P(S_n > x)}{nP(X_1 > x - (n-1)\mu)} - 1 \right| \leq \varphi(x).$$

PROOF. Suppose first that $\alpha > 3$ and let $\sigma(n) = \sqrt{(\alpha - 2)n \log n}$. Since

$$\frac{P(S_n > x)}{nP(X_1 > x - (n-1)\mu)} = \frac{P(S_n^* > x - n\mu)}{nP(Y_1 > x - n\mu)},$$

where $Y_i = X_i - \mu$ and $S_n^* = Y_1 + \cdots + Y_n$. Then the result will follow from Theorem 3.1(a) once we show that $(x - n\mu)/\sigma(n) \to \infty$ uniformly for $1 \leq n \leq M_\gamma(x)$. To see this simply note that

$$\frac{x - n\mu}{\sigma(n)} \geq \frac{x - M_\gamma(x)\mu}{\sigma(M_\gamma(x))} \sim \sqrt{\frac{\mu}{\alpha - 2}} \cdot \frac{x^{\gamma - 1/2}}{\sqrt{\log x}}.$$

Since $\gamma > 1/2$, the above converges to infinity.

Suppose now that $\alpha \in (2, 3)$ and note that $P(Y_1 \leq -t) = 0$ for $t \geq \mu$. Note also that since $\overline{F}(t) = P(Y_1 > t)$ is regularly varying with index $\alpha - 1$, then $\sigma(n) = \overline{F}^{-1}(1/n) = n^{1/(\alpha-1)}\tilde{L}(n)$ for some slowly varying function $\tilde{L}(\cdot)$ [see Bingham, Goldie and Teugels (1987)]. Then the result will follow from Theorem 3.1(b) once we show that $(x - n\mu)/\sigma(n) \to \infty$ uniformly for $1 \leq n \leq M_\gamma(x)$. To see this note that

$$\frac{x - n\mu}{\sigma(n)} \geq \frac{x - M_\gamma(x)\mu}{\sigma(M_\gamma(x))} \sim \frac{x^\gamma}{\sigma(x/\mu)} \sim \frac{x^{\gamma - 1/(\alpha-1)}}{\mu^{-1/(\alpha-1)}\tilde{L}(x)},$$

and since $\gamma > 1/(\alpha - 1)$ the above converges to infinity.

The case $\alpha = 3$ is rather technical and does not provide additional insights. We refer the reader to the internet supplement Olvera-Cravioto, Blanchet and Glynn (2010) for the details. □

We now give a lemma that will allow us to transform the statements of the main results from being uniform in $0 < \rho < 1$ to being uniform in $x > 0$, under the limiting regimes $x \to \infty$ and $\rho \nearrow 1$, respectively.

LEMMA 3.3. *Suppose that*

$$\sup_{0<\rho<1} \left| \frac{P(W_\infty(\rho) > x)}{A(\rho, x)} - 1 \right| \to 0$$

*as $x \to \infty$, where $A(\rho, x)$ satisfies*

$$\sup_{0<x<(1-\rho)^{-\eta}} |A(\rho, x) - 1| \to 0$$

*as* $\rho \nearrow 1$ *for some* $0 < \eta < 1$. *Then*,

$$\sup_{x>0}\left|\frac{P(W_\infty(\rho) > x)}{A(\rho, x)} - 1\right| \to 0$$

*as* $\rho \nearrow 1$.

PROOF. We argue by contradiction. Suppose that there exists an $\varepsilon > 0$ and a function $x:(0,1) \to (0,\infty)$ such that $x(\phi) \geq (1-\phi)^{-\eta}$ and

$$\left|\frac{P(W_\infty(\phi) > x(\phi))}{A(\phi, x(\phi))} - 1\right| > \varepsilon$$

for all $0 < \phi < 1$. Then,

$$\sup_{0<\rho<1}\left|\frac{P(W_\infty(\rho) > x(\phi))}{A(\rho, x(\phi))} - 1\right| \geq \left|\frac{P(W_\infty(\phi) > x(\phi))}{A(\phi, x(\phi))} - 1\right| > \varepsilon.$$

But this cannot be since by assumption,

$$\lim_{\phi\nearrow 1}\sup_{0<\rho<1}\left|\frac{P(W_\infty(\rho) > x(\phi))}{A(\rho, x(\phi))} - 1\right| = 0.$$

It follows that

$$\sup_{x\geq(1-\rho)^{-\eta}}\left|\frac{P(W_\infty(\rho) > x)}{A(\rho, x)} - 1\right| \to 0$$

as $\rho \nearrow 1$. For $0 < x < (1-\rho)^{-\eta}$ note that

$$\begin{aligned}
&\lim_{\rho\nearrow 1}\sup_{0<x<(1-\rho)^{-\eta}}\left|\frac{P(W_\infty(\rho) > x)}{A(\rho, x)} - 1\right| \\
&\quad \leq \lim_{\rho\nearrow 1}\sup_{0<x<(1-\rho)^{-\eta}}\frac{|P(W_\infty(\rho) > x) - 1|}{A(\rho, x)} + \lim_{\rho\nearrow 1}\sup_{0<x<(1-\rho)^{-\eta}}\frac{|A(\rho, x) - 1|}{A(\rho, x)} \\
&\quad = \lim_{\rho\nearrow 1}\sup_{0<x<(1-\rho)^{-\eta}}\bigl|P\bigl(W_\infty(\rho) > x\bigr) - 1\bigr|.
\end{aligned}$$

The last limit is zero by the standard heavy traffic limit. □

Throughout the rest of this section let $\mu = EX_1$, $\beta = (2 \wedge (\alpha - 1))^{-1}$, $M(x) = \lfloor (x - x^\beta)/\mu \rfloor$, and

$$S(\rho, x) = \sum_{n=1}^{M(x)} (1-\rho)\rho^n n P\bigl(X_1 > x - (n-1)\mu\bigr).$$

PROOF OF THEOREM 2.1. We will prove the uniform in $0 < \rho < 1$ asymptotic, since the statement regarding the uniformity in $x > 0$ will follow from

Lemma 3.3 by noting that

$$\sup_{0<x<(1-\rho)^{-1/4}} S(\rho,x) \le \sup_{0<x<(1-\rho)^{-1/4}} \frac{(1-\rho)M(x)(M(x)+1)}{2} \le \frac{(1-\rho)^{1/2}}{\mu^2},$$

which clearly converges to zero. Throughout the proof $C>0$ is a generic constant.

Fix $\beta<\gamma<1\wedge\beta(\alpha-1)$ and define $M_\gamma(x)=\lfloor(x-x^\gamma)/\mu\rfloor$. Then, by Lemma 3.2, there exists a function $\varphi_1(t)\to 0$ as $t\to\infty$ such that

$$\sup_{1\le n\le M_\gamma(x)}\left|\frac{P(S_n>x)}{nP(X_1>x-(n-1)\mu)}-1\right|\le\varphi_1(x).$$

By (3.1) we have

$$\begin{aligned}
&\left|P\big(W_\infty(\rho)>x\big)-S(\rho,x)-\rho^{x/\mu}\right|\\
&\quad\le\varphi_1(x)\sum_{n=1}^{M_\gamma(x)}(1-\rho)\rho^n nP\big(X_1>x-(n-1)\mu\big)\\
&\qquad+\sum_{n=M_\gamma(x)+1}^{M(x)}(1-\rho)\rho^n nP\big(X_1>x-(n-1)\mu\big)\\
&\qquad+\sum_{n=M_\gamma(x)+1}^{\lfloor x/\mu\rfloor}(1-\rho)\rho^n P(S_n>x)\\
&\qquad+\left|\sum_{n=\lfloor x/\mu\rfloor+1}^{\infty}(1-\rho)\rho^n P(S_n>x)-\rho^{x/\mu}\right|.
\end{aligned}$$

Clearly,

$$\varphi_1(x)\sum_{n=1}^{M_\gamma(x)}(1-\rho)\rho^n nP\big(X_1>x-(n-1)\mu\big)\le\varphi_1(x)S(\rho,x). \tag{3.2}$$

Fix $0<\varepsilon<\min\{\alpha-2,(\beta(\alpha-1)-\gamma)/\beta\}$. The second term is bounded by

$$\begin{aligned}
&\sum_{n=M_\gamma(x)+1}^{M(x)}(1-\rho)\rho^n nP\big(X_1>x-(n-1)\mu\big)\\
&\quad\le Cx(1-\rho)\sum_{n=M_\gamma(x)+1}^{M(x)}\rho^n\big(x-(n-1)\mu\big)^{-\alpha+1+\varepsilon}.
\end{aligned}$$

Since $g(n) = \rho^n (x - (n-1)\mu)^{-\alpha+1+\varepsilon}$ is convex in $n$,

$$
\begin{aligned}
x(1-\rho) & \sum_{n=M_\gamma(x)+1}^{M(x)} g(n) \\
& \le x(1-\rho)\big(M(x) - M_\gamma(x)\big) \max\big\{g\big(M_\gamma(x)+1\big), g(M(x))\big\} \\
& \le C(1-\rho)x^{1+\gamma} \max\big\{\rho^{(x-x^\gamma)/\mu} x^{-\gamma(\alpha-1-\varepsilon)}, \rho^{(x-x^\beta)/\mu-1} x^{-\beta(\alpha-1-\varepsilon)}\big\} \\
& \le C(1-\rho)x^{1+\gamma-\beta(\alpha-1-\varepsilon)} \rho^{(x-x^\gamma)/\mu},
\end{aligned} \tag{3.3}
$$

where our choice of $\varepsilon$ guarantees that $\gamma - \beta(\alpha - 1 - \varepsilon) < 0$. Also, we have

$$
\begin{aligned}
\sum_{n=M_\gamma(x)+1}^{\lfloor x/\mu \rfloor} (1-\rho)\rho^n P(S_n > x) & \le \rho^{M_\gamma(x)+1}\big(1 - \rho^{\lfloor x/\mu \rfloor - M_\gamma(x)}\big) \\
& \le C\rho^{(x-x^\gamma)/\mu} x^\gamma |\log \rho|.
\end{aligned} \tag{3.4}
$$

To derive the last bound let $K(x) = \lfloor (x + x^\gamma)/\mu \rfloor$. Then,

$$
\begin{aligned}
& \left| \sum_{n=\lfloor x/\mu \rfloor+1}^{\infty} (1-\rho)\rho^n P(S_n > x) - \rho^{x/\mu} \right| \\
& \quad = \rho^{x/\mu} - \rho^{\lfloor x/\mu \rfloor + 1} + \sum_{n=\lfloor x/\mu \rfloor+1}^{\infty} (1-\rho)\rho^n P(S_n \le x) \\
& \quad \le \rho^{x/\mu}(1-\rho) + \sum_{n=\lfloor x/\mu \rfloor+1}^{K(x)} (1-\rho)\rho^n + \sum_{n=K(x)+1}^{\infty} (1-\rho)\rho^n P(S_n \le x).
\end{aligned}
$$

It is easy to check that

$$
\sum_{n=\lfloor x/\mu \rfloor+1}^{K(x)} (1-\rho)\rho^n \le C\rho^{x/\mu} x^\gamma |\log \rho|.
$$

For the tail of the sum let $Y_i = \mu - X_i$ and $S_n^* = Y_1 + \cdots + Y_n$. Let $b_n$ be the scaling for which $Z_n = S_n^*/b_n \Rightarrow Z$, where $Z$ is a stable random variable. Note that $b_n = n^\beta L_0(n)$ for some slowly varying $L_0(\cdot)$. It follows that for all $n > K(x)$,

$$
P(S_n \le x) = P\left(Z_n \ge \frac{n\mu - x}{b_n}\right) \le P\left(Z_n \ge \frac{(K(x)+1)\mu - x}{b_{K(x)+1}}\right),
$$

where

$$
\frac{(K(x)+1)\mu - x}{b_{K(x)+1}} \ge \frac{\mu^\beta x^\gamma}{(x + x^\gamma)^\beta L_0((x - x^\gamma)/\mu)} \ge \frac{c x^{\gamma-\beta}}{L_0(x)}
$$

for some constant $c > 0$. It follows that

$$\sup_{n>K(x)} P(S_n \le x) \le \sup_{n>K(x)} P\big(Z_n \ge cx^{\gamma-\beta}/L_0(x)\big) \le \varphi_2(x)$$

for some $\varphi_2(t) \to 0$ as $t \to \infty$. Hence,

$$\sum_{n=K(x)+1}^{\infty} (1-\rho)\rho^n P(S_n \le x) \le \varphi_2(x) \sum_{n=K(x)+1}^{\infty} (1-\rho)\rho^n \le \varphi_2(x)\rho^{x/\mu}.$$

We thus have that

$$\begin{aligned} &\left| \sum_{n=\lfloor x/\mu \rfloor+1}^{\infty} (1-\rho)\rho^n P(S_n > x) - \rho^{x/\mu} \right| \\ &\qquad \le \varphi_2(x)\rho^{x/\mu} + C\rho^{x/\mu}|\log\rho| x^{\gamma}. \end{aligned} \tag{3.5}$$

Combining (3.2)–(3.5) gives

$$\begin{aligned} &\big|P\big(W_\infty(\rho) > x\big) - S(\rho, x) - \rho^{x/\mu}\big| \\ &\qquad \le \varphi_1(x) S(\rho, x) + \varphi_2(x)\rho^{x/\mu} + C\rho^{(x-x^{\gamma})/\mu}|\log\rho| x^{\upsilon}, \end{aligned}$$

where $\upsilon = \max\{1 + \gamma - \beta(\alpha - 1 - \varepsilon), \gamma\} \in (0, 1)$. It only remains to show that $\rho^{(x-x^{\gamma})/\mu}|\log\rho| x^{\upsilon} = o(S(\rho, x) + \rho^{x/\mu})$ uniformly in $0 < \rho < 1$. To see this let $\rho(x) = 1 - (x^{\upsilon}\log x)^{-1}$, then

$$\begin{aligned} \sup_{\rho(x)\le\rho<1} \frac{\rho^{(x-x^{\gamma})/\mu}|\log\rho| x^{\upsilon}}{S(\rho, x) + \rho^{x/\mu}} &\le \sup_{\rho(x)\le\rho<1} e^{|\log\rho| x^{\gamma}/\mu}|\log\rho| x^{\upsilon} \\ &= e^{|\log\rho(x)| x^{\gamma}/\mu}|\log\rho(x)| x^{\upsilon} \\ &\le \frac{C}{\log x} \to 0, \end{aligned}$$

and since $S(\rho, x) \ge P(X_1 > x)\sum_{n=1}^{\lfloor x/\mu \rfloor}(1-\rho)\rho^n$,

$$\begin{aligned} &\sup_{0<\rho<\rho(x)} \frac{\rho^{(x-x^{\gamma})/\mu}|\log\rho| x^{\upsilon}}{S(\rho, x) + \rho^{x/\mu}} \\ &\quad \le \sup_{0<\rho<\rho(x)} \frac{\rho^{(x-x^{\gamma})/\mu}|\log\rho| x^{\upsilon}}{L(x)x^{-\alpha+1}\rho(1-\rho^{\lfloor x/\mu \rfloor})} \\ &\quad \le C \sup_{0<\rho<\rho(x)} \rho^{(x-x^{\gamma})/\mu - 1}|\log\rho| x^{\upsilon+\alpha-1+\varepsilon} \\ &\quad \le C \sup_{t>(x^{\upsilon}\log x)^{-1}} \exp\left(-\left(\frac{x-x^{\gamma}}{\mu} - 1\right)t + \log t + (\upsilon + \alpha - 1 + \varepsilon)\log x\right) \end{aligned}$$

$$= C \exp\biggl(-\frac{x^{1-\upsilon}}{\mu \log x}\biggl(1 - \frac{1}{x^{1-\gamma}} - \frac{\mu}{x}\biggr) - \log\log x + (\alpha - 1 + \varepsilon)\log x\biggr)$$
$$\to 0. \qquad \square$$

We now prove Theorem 2.2.

PROOF OF THEOREM 2.2. Again, we only prove the statement regarding the uniformity in $0 < \rho < 1$, since the statement for $x > 0$ follows from Lemma 3.3 and the observation that, as $\rho \nearrow 1$,

$$\begin{aligned}
&\sup_{0<x<(1-\rho)^{-1/4}} \frac{\rho}{1-\rho}\biggl(1 - \rho^{x/\mu} - \frac{(1-\rho)x}{\mu}\rho^{x/\mu}\biggr) P(X_1 > x) \\
&\quad \le \sup_{0<x<(1-\rho)^{-1/4}} \frac{\rho}{1-\rho}\biggl(|\log \rho|\frac{x}{\mu} - \frac{(1-\rho)x}{\mu}\rho^{x/\mu}\biggr) \\
&\quad = \frac{\rho}{\mu(1-\rho)^{5/4}}\bigl(|\log\rho| - (1-\rho)e^{-|\log\rho|(1-\rho)^{-1/4}/\mu}\bigr) \\
&\quad = O\bigl((1-\rho)^{1/2}\bigr).
\end{aligned}$$

By Theorem 2.1 we only need to show that

$$\sup_{0<\rho<1}\left|\frac{S(\rho,x) - (\rho/(1-\rho))\gamma(x,\rho)P(X_1>x)}{(\rho/(1-\rho))\gamma(x,\rho)P(X_1>x) + \rho^{x/\mu}}\right| \to 0$$

as $x \to \infty$. We start by noting that

$$\begin{aligned}
&S(\rho,x) - \frac{\rho}{1-\rho}\gamma(x,\rho)P(X_1>x) \\
&\quad = \sum_{n=1}^{M(x)} (1-\rho)\rho^n n\bigl(P\bigl(X_1 > x-(n-1)\mu\bigr) - P(X_1>x)\bigr) \\
&\qquad + P(X_1>x)\Biggl(\sum_{n=1}^{M(x)}(1-\rho)\rho^n n - \frac{\rho}{1-\rho}\gamma(x,\rho)\Biggr).
\end{aligned}$$

Then, since $\frac{\rho}{1-\rho}\gamma(x,\rho) \ge \sum_{n=1}^{\lfloor x/\mu\rfloor-1}(1-\rho)\rho^n n$,

$$\begin{aligned}
&\left|S(\rho,x) - \frac{\rho}{1-\rho}\gamma(x,\rho)P(X_1>x)\right| \\
&\quad \le P(X_1>x)\sum_{n=1}^{M(x)}(1-\rho)\rho^n n\biggl(\frac{P(X_1>x-(n-1)\mu)}{P(X_1>x)} - 1\biggr) \\
&\qquad + P(X_1>x)\Biggl(\frac{\rho}{1-\rho}\gamma(x,\rho) - \sum_{n=1}^{M(x)}(1-\rho)\rho^n n\Biggr).
\end{aligned}$$

The second term can be bounded as follows:

$$
\begin{aligned}
&P(X_1 > x)\left(\frac{\rho}{1-\rho}\gamma(x,\rho) - \sum_{n=1}^{M(x)} (1-\rho)\rho^n n\right) \\
&\qquad \le \frac{\rho}{1-\rho} P(X_1 > x)\big(\rho^{M(x)} - \rho^{x/\mu}\big)\left(1 + \frac{(1-\rho)x}{\mu}\right) \qquad (3.6)\\
&\qquad \le \frac{C}{1-\rho} P(X_1 > x)\big(\rho^{(x-x^\beta)/\mu} - \rho^{x/\mu+1}\big)\big(1 + (1-\rho)x\big),
\end{aligned}
$$

where $C > 0$ is a generic constant. Fix $0 < \varepsilon < \beta(\alpha-2)/(\alpha-2+\beta)$ and set $N(x) = \lfloor (1-\varepsilon)x/\mu \rfloor$. Then, for $1 \le n \le N(x)$,

$$
\begin{aligned}
\frac{P(X_1 > x-(n-1)\mu)}{P(X_1 > x)} &\le \left(\frac{x-(n-1)\mu}{x}\right)^{-\alpha+1} \sup_{1\le n\le N(x)} \frac{L(x-(n-1)\mu)}{L(x)} \\
&\le \left(1 + (\alpha-1)\varepsilon^{-\alpha-2}\frac{(n-1)\mu}{x}\right)\big(1+\varphi_1(x)\big),
\end{aligned}
$$

where $\varphi_1(x) = \sup_{\varepsilon \le t\le 1} \frac{L(tx)}{L(x)} - 1 \to 0$ by properties of slowly varying functions. Therefore, for $1 \le n \le N(x)$,

$$
\frac{P(X_1 > x-(n-1)\mu)}{P(X_1 > x)} - 1 \le C\left(\frac{n-1}{x} + \varphi_1(x)\right).
$$

It follows that

$$
\begin{aligned}
&P(X_1 > x)\sum_{n=1}^{N(x)} (1-\rho)\rho^n n\left(\frac{P(X_1 > x-(n-1)\mu)}{P(X_1 > x)} - 1\right) \\
&\qquad \le C P(X_1 > x)\left(\frac{1}{x}\sum_{n=1}^{N(x)} (1-\rho)\rho^n n(n-1) + \varphi_1(x)\frac{\rho}{1-\rho}\gamma(x,\rho)\right) \\
&\qquad \le C P(X_1 > x)\bigg(\frac{2\rho^2}{x(1-\rho)^2}\big(1-\rho^{N(x)} - (1-\rho)N(x)\rho^{N(x)}\big) \qquad (3.7)\\
&\qquad\qquad\qquad + \varphi_1(x)\frac{\rho}{1-\rho}\gamma(x,\rho)\bigg) \\
&\qquad \le \frac{C\rho}{1-\rho} P(X_1 > x)\gamma(x,\rho)\left(\frac{1}{x(1-\rho)} + \varphi_1(x)\right).
\end{aligned}
$$

For the terms $N(x) < n \le M(x)$ we have

$$
\begin{aligned}
&P(X_1 > x)\sum_{n=N(x)+1}^{M(x)} (1-\rho)\rho^n n\left(\frac{P(X_1 > x-(n-1)\mu)}{P(X_1 > x)} - 1\right) \\
&\qquad \le Cx(1-\rho)\rho^{N(x)+1}\sum_{n=N(x)+1}^{M(x)} P\big(X_1 > x-(n-1)\mu\big) \qquad (3.8)
\end{aligned}
$$

$$
\begin{aligned}
&\le Cx(1-\rho)\rho^{(1-\varepsilon)x/\mu}\int_{x-\mu M(x)}^{x-\mu N(x)} P(X_1>t)\,dt \\
&\le Cx(1-\rho)\rho^{(1-\varepsilon)x/\mu}\big(x-\mu M(x)\big)P\big(X_1>x-\mu M(x)\big) \\
&\le C(1-\rho)\rho^{(1-\varepsilon)x/\mu}x^{1+\beta}P(X_1>x^\beta),
\end{aligned}
$$

where for the third inequality we used Proposition 1.5.10 in Bingham, Goldie and Teugels (1987). Combining (3.6)–(3.8) gives

$$
\begin{aligned}
&\left|S(\rho,x)-\frac{\rho}{1-\rho}(1-\rho^{x/\mu})P(X_1>x)\right| \\
&\quad\le \frac{C}{1-\rho}P(X_1>x)\big(\rho^{(x-x^\beta)/\mu}-\rho^{x/\mu+1}\big)\big(1+(1-\rho)x\big) \\
&\qquad+\frac{C\rho}{1-\rho}P(X_1>x)\gamma(x,\rho)\left(\frac{1}{x(1-\rho)}+\varphi_1(x)\right) \\
&\qquad+C(1-\rho)\rho^{(1-\varepsilon)x/\mu}x^{1+\beta}P(X_1>x^\beta).
\end{aligned}
$$

Let $A(\rho,x)=\frac{\rho}{1-\rho}\gamma(x,\rho)P(X_1>x)+\rho^{x/\mu}$ and define $\rho(x)=1-c\mu(\alpha-2)\log x/x$, with $\frac{(1-\beta)(\alpha-2+\varepsilon)}{(\alpha-2)(1-\varepsilon)}<c<1$. Note that $\gamma(x,\rho)\sim 1$ as $x\to\infty$ uniformly for $0<\rho\le\rho(x)$. Then,

$$
\begin{aligned}
&\sup_{0<\rho\le\rho(x)}\frac{1}{A(\rho,x)}\left|S(\rho,x)-\frac{\rho}{1-\rho}(1-\rho^{x/\mu})P(X_1>x)\right| \\
&\quad\le C\sup_{0<\rho\le\rho(x)}\bigg\{\frac{1}{\gamma(x,\rho)}\big(\rho^{(x-x^\beta)/\mu-1}-\rho^{x/\mu}\big)\big(1+(1-\rho)x\big)+\frac{1}{x(1-\rho)} \\
&\qquad\qquad+\varphi_1(x)+\frac{(1-\rho)^2x^{1+\beta}P(X_1>x^\beta)}{\rho\gamma(x,\rho)P(X_1>x)}\rho^{(1-\varepsilon)x/\mu}\bigg\} \\
&\quad\le C\bigg\{\rho(x)^{(x-x^\beta)/\mu-1}\big(1+\big(1-\rho(x)\big)x\big)+\frac{1}{x(1-\rho(x))}+\varphi_1(x) \\
&\qquad\qquad+\frac{P(X_1>x^\beta)}{P(X_1>x)}x^{1+\beta}\big(1-\rho(x)\big)^2\rho(x)^{(1-\varepsilon)x/\mu-1}\bigg\} \\
&\quad\le C\left\{\frac{\log x}{x^{c(\alpha-2)}}+\frac{1}{\log x}+\varphi_1(x)+\frac{P(X_1>x^\beta)}{P(X_1>x)}\cdot\frac{(\log x)^2}{x^{1-\beta+c(\alpha-2)(1-\varepsilon)}}\right\}.
\end{aligned}
$$

The first three terms in the expression above clearly converge to zero. To see that the fourth one does as well use Potter's theorem [Bingham, Goldie and Teugels

(1987), page 25] to obtain

$$\frac{P(X_1 > x^\beta)}{P(X_1 > x)} \cdot \frac{(\log x)^2}{x^{1-\beta+c(\alpha-2)(1-\varepsilon)}}$$
$$\le A_\varepsilon x^{(1-\beta)(\alpha-1+\varepsilon)} \cdot \frac{(\log x)^2}{x^{1-\beta+c(\alpha-2)(1-\varepsilon)}}$$

for some constant $A_\varepsilon > 1$. Our choice of $\varepsilon$ and $c$ guarantees that $1-\beta+c(\alpha-2)(1-\varepsilon) > (1-\beta)(\alpha-1+\varepsilon)$.

To analyze the supremum over $\rho(x) \le \rho < 1$ we first note that

$$\frac{\rho}{(1-\rho)^2}\gamma(x,\rho) \le \sum_{n=1}^{\lfloor x/\mu \rfloor+1} \rho^n n.$$

Then,

$$\begin{aligned}
\sup_{\rho(x)\le\rho<1} \frac{1}{A(\rho,x)}&\left|S(\rho,x) - \frac{\rho}{1-\rho}(1-\rho^{x/\mu})P(X_1 > x)\right| \\
&\le C \sup_{\rho(x)\le\rho<1} \Biggl\{ \frac{P(X_1 > x)}{1-\rho}\bigl(\rho^{-x^\beta/\mu} - \rho\bigr)\bigl(1+(1-\rho)x\bigr) \\
&\qquad + \frac{\rho\gamma(x,\rho)P(X_1 > x)}{(1-\rho)\rho^{x/\mu}}\left(\frac{1}{x(1-\rho)} + \varphi_1(x)\right) \\
&\qquad + (1-\rho)\rho^{-\varepsilon x/\mu}x^{1+\beta}P(X_1 > x^\beta)\Biggr\} \\
&\le C \sup_{\rho(x)\le\rho<1} \Biggl\{ \frac{P(X_1 > x)}{1-\rho}|\log\rho| x^\beta \log x + \rho^{-\varepsilon x/\mu}x^\beta \log x\, P(X_1 > x^\beta) \\
&\qquad + \frac{P(X_1 > x)}{x}\sum_{n=1}^{\lceil x/\mu\rceil+1} \rho^{n-x/\mu} n\bigl(1+x(1-\rho)\bigr)\Biggr\} \\
&\le C\Biggl\{ P(X_1 > x)x^\beta \log x + \rho(x)^{-\varepsilon x/\mu}x^\beta \log x\, P(X_1 > x^\beta) \\
&\qquad + P(X_1 > x)\sum_{n=1}^{\lceil x/\mu\rceil+1} \rho(x)^{n-x/\mu}(1+\log x)\Biggr\}.
\end{aligned}$$

The first term clearly converges to zero. To see that the second and third terms converge to zero as well note that

$$\begin{aligned}
\rho(x)^{-\varepsilon x/\mu}x^\beta \log x\, P(X_1 > x^\beta) &\le C x^{\varepsilon c(\alpha-2)+\beta-\beta(\alpha-1-\varepsilon)} \log x \\
&\le C x^{\varepsilon(\alpha-2)-\beta(\alpha-2-\varepsilon)} \log x
\end{aligned}$$

and

$$P(X_1 > x) \sum_{n=1}^{\lceil x/\mu \rceil + 1} \rho(x)^{n - x/\mu} \log x \le C x^{-\alpha+\varepsilon} \frac{(1 - \rho(x)^{\lceil x/\mu \rceil + 1})}{\rho(x)^{x/\mu}(1 - \rho(x))} \log x$$
$$\le C x^{-\alpha+\varepsilon+1+c(\alpha-2)}$$
$$\le C x^{-1+\varepsilon}.$$

Our choice of $\varepsilon$ guarantees that both expressions above converge to zero. This completes the proof. $\square$

We end this section with the proof of Corollary 2.3.

PROOF OF COROLLARY 2.3. Let

$$y = y(\rho) = c\mu(\alpha - 2)(1 - \rho)^{-1} \log\bigl((1 - \rho)^{-1}\bigr).$$

We start with the proof of part (a). We need to verify that for $0 < c \le 1$

$$\sup_{0 \le x \le y} \left| \frac{(\rho/(1-\rho))\gamma(x,\rho)P(X_1 > x) + \rho^{x/\mu}}{\exp(-(1-\rho)x/\mu)} - 1 \right| \to 0$$

as $\rho \nearrow 1$. Note that

$$\sup_{0 \le x \le y} \left| \frac{(\rho/(1-\rho))\gamma(x,\rho)P(X_1 > x) + \rho^{x/\mu}}{\exp(-(1-\rho)x/\mu)} - 1 \right|$$

$$\le \sup_{0 \le x \le y} \left| \exp\left( \frac{x \log \rho}{\mu} + \frac{(1-\rho)x}{\mu} \right) - 1 \right| \tag{3.9}$$

$$+ \sup_{0 \le x \le y} \frac{\rho\gamma(x,\rho)}{(1-\rho)} P(X_1 > x) \exp\bigl((1-\rho)x/\mu\bigr). \tag{3.10}$$

We can bound (3.9) as follows:

$$\sup_{0 \le x \le y} \left| \exp\left( \frac{x \log \rho}{\mu} + \frac{(1-\rho)x}{\mu} \right) - 1 \right| \le \sup_{0 \le x \le y} \frac{x|\log \rho + 1 - \rho|}{\mu}$$
$$\le C y (1-\rho)^2$$
$$\le C t^{-1} \log t,$$

where $t = (1 - \rho)^{-1}$. Also, note that since $\rho^{x/\mu} \ge 1 - |\log \rho| x/\mu$ and $|\log \rho| = 1 - \rho + O((1-\rho)^2)$ as $\rho \nearrow 1$,

$$\gamma(x,\rho) = 1 - \rho^{x/\mu} - \rho^{x/\mu}(1-\rho)x/\mu$$
$$\le \bigl(|\log \rho| - (1-\rho)\bigr)x/\mu + |\log \rho|(1-\rho)(x/\mu)^2$$
$$\le C(1-\rho)^2 x^2.$$

Then (3.10) is bounded by

$$\begin{aligned}
&\sup_{0\le x\le y} \frac{\rho\gamma(x,\rho)}{(1-\rho)} P(X_1 > x)\exp\bigl((1-\rho)x/\mu\bigr) \\
&\quad \le C \sup_{0\le x\le (1-\rho)^{-1/4}} \frac{\gamma(x,\rho)}{1-\rho} \\
&\qquad + \sup_{(1-\rho)^{-1/4}\le x\le y} \frac{1}{1-\rho} P(X_1 > x)\exp\bigl((1-\rho)x/\mu\bigr) \\
&\quad \le \sup_{0\le x\le (1-\rho)^{-1/4}} C(1-\rho)x^2 \\
&\qquad + \sup_{(1-\rho)^{-1/4}\le x\le y} \frac{L(x)}{1-\rho} \exp\bigl((1-\rho)x/\mu - (\alpha-1)\log x\bigr) \\
&\quad \le C(1-\rho)^{1/2} + \frac{L(y)}{1-\rho}\exp\bigl((1-\rho)y/\mu - (\alpha-1)\log y\bigr) \\
&\quad \le Ct^{-1/2} + CL(t\log t)\exp\bigl(-(1-c)(\alpha-2)\log t - (\alpha-1)\log\log t\bigr).
\end{aligned}$$

Clearly, if $0 < c < 1$, then the two expressions above converge to zero as $t \to \infty$. If $c = 1$ and $L(x)/(\log x)^{\alpha-1} \to 0$ as $x \to \infty$, then

$$\begin{aligned}
&\lim_{\rho\nearrow 1} \sup_{0\le x\le y} \left| \frac{(\rho/(1-\rho))\gamma(x,\rho)P(X_1 > x) + \rho^{x/\mu}}{\exp(-(1-\rho)x/\mu)} - 1 \right| \\
&\quad \le C \lim_{t\to\infty} L(t\log t)\exp\bigl(-(\alpha-1)\log\log t\bigr) \\
&\quad = C \lim_{t\to\infty} \frac{L(t\log t)}{(\log(t\log t))^{\alpha-1}} \cdot \left(\frac{\log(t\log t)}{\log t}\right)^{\alpha-1} = 0.
\end{aligned}$$

We now move to part (b). We need to verify that for $c \ge 1$

$$\sup_{x\ge y} \left| \frac{(\rho/(1-\rho))\gamma(x,\rho)P(X_1 > x) + \rho^{x/\mu}}{(\rho/(1-\rho))P(X_1 > x)} - 1 \right| \to 0$$

as $\rho \nearrow 1$. Note that

$$\begin{aligned}
&\sup_{x\ge y} \left| \frac{(\rho/(1-\rho))\gamma(x,\rho)P(X_1 > x) + \rho^{x/\mu}}{(\rho/(1-\rho))P(X_1 > x)} - 1 \right| \\
&\quad \le \sup_{x\ge y} |\gamma(x,\rho) - 1| + \sup_{x\ge y} \frac{(1-\rho)\rho^{x/\mu}}{\rho P(X_1 > x)} \\
&\quad \le \rho^{y/\mu}\bigl(1 + (1-\rho)y/\mu\bigr) + C\sup_{x\ge y} \frac{1-\rho}{L(x)} \exp\left(-\frac{x}{\mu}(1-\rho) + (\alpha-1)\log x\right)
\end{aligned}$$

$$\begin{aligned}
&\le C\rho^{y/\mu}(1-\rho)y + C\frac{1-\rho}{L(y)}\exp\left(-\frac{y}{\mu}(1-\rho)+(\alpha-1)\log y\right)\\
&\le Ct^{-c(\alpha-2)}\log t\\
&\quad + \frac{C}{L(t\log t)}\exp\bigl(-(c-1)(\alpha-2)\log t+(\alpha-1)\log\log t\bigr),
\end{aligned}$$

where $t=(1-\rho)^{-1}$. Clearly, if $c>1$ the above converges to zero. If $c=1$ and $L(x)/(\log x)^{\alpha-1}\to\infty$ as $x\to\infty$, then

$$\begin{aligned}
&\lim_{\rho\nearrow 1}\sup_{0\le x\le y}\left|\frac{(\rho/(1-\rho))\gamma(x,\rho)P(X_1>x)+\rho^{x/\mu}}{\exp(-(1-\rho)x/\mu)}-1\right|\\
&\qquad\le C\lim_{t\to\infty}\frac{1}{L(t\log t)}\exp\bigl((\alpha-1)\log\log t\bigr)\\
&\qquad= C\lim_{t\to\infty}\frac{(\log(t\log t))^{\alpha-1}}{L(t\log t)}\cdot\left(\frac{\log t}{\log(t\log t)}\right)^{\alpha-1}=0. \qquad\square
\end{aligned}$$

**4. Numerical approximations.** Theorems 2.1 and 2.2 suggest approximating $P(W_\infty(\rho)>x)$ either with

$$H(\rho,x)\triangleq S(\rho,x)+\rho^{x/\mu}=\sum_{n=1}^{M(x)}(1-\rho)\rho^n nP\bigl(X_1>x-(n-1)\mu\bigr)+\rho^{x/\mu}$$

or with

$$J(\rho,x)\triangleq\frac{\rho}{1-\rho}\gamma(\rho,x)P(X_1>x)+\rho^{x/\mu},$$

respectively.

We compared both approximations to simulated values of $P(W_\infty(\rho)>x)$ and found that $H(\rho,x)$ tends to be better than $J(\rho,x)$ and seems to perform very well across all values of $x$ for different choices of $\rho$. This is not surprising given that $H(\rho,x)$ more closely resembles the Pollaczek–Khintchine formula than $J(\rho,x)$.

When $\sigma^2=\mathrm{Var}(X_1)<\infty$, the central limit theorem can be used to approximate the tail of the Pollaczek–Khintchine formula in a way that $\sigma^2$ is incorporated into the approximation. The term $\rho^{x/\mu}$ appearing in the definitions of $H(\rho,x)$ and $J(\rho,x)$ can be replaced by

$$T(\rho,x)\triangleq\sum_{n=1}^{\infty}(1-\rho)\rho^n\bigl(1-\Phi\bigl((x-n\mu)/\sqrt{\sigma^2 n}\bigr)\bigr),$$

which can alternatively be written as $T(\rho,x)=E[\rho^{M(x,Z)}]$, where $Z\sim N(0,1)$ and $M(x,z)=\lfloor(\sqrt{x/\mu+(\sigma z)^2/(2\mu)^2}-(\sigma z)/(2\mu))^2\rfloor$. We do not give proofs

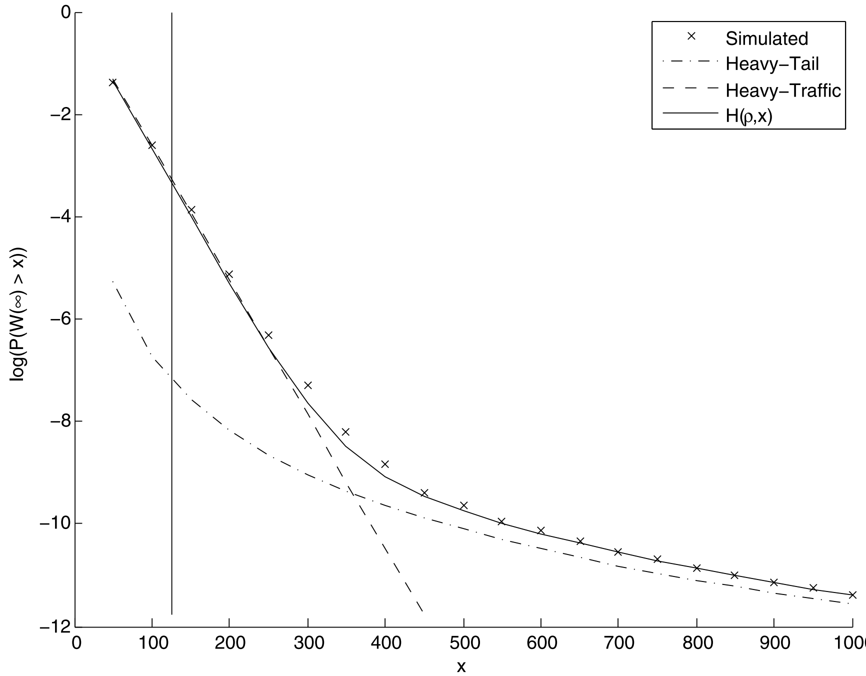


FIG. 1. *Pareto integrated tail with $\rho = 0.95$ and $\alpha = 3.1$.*

here, but it can be shown that provided $\sigma^2 < \infty$, Theorems 2.1 and 2.2 continue to hold with $\rho^{x/\mu}$ replaced by $T(\rho, x)$. This is relevant from the numerical standpoint since the resulting approximations tend to perform better than those with the simpler $\rho^{x/\mu}$.

We plotted approximation $H(\rho, x)$ against simulated values of $P(W_\infty(\rho) > x)$. Figures 1 and 2 correspond to queues having Pareto integrated tail distribution, that is, $P(X_1 > x) = x^{-\alpha+1}$ for $x \geq 1$. For comparison purposes we also plotted the heavy-traffic approximation,

$$\textit{Heavy-Traffic} = \exp\bigl(-(1-\rho)x/\mu\bigr),$$

and the heavy-tail asymptotic,

$$\textit{Heavy-Tail} = \frac{\rho}{1-\rho} P(X_1 > x).$$

The vertical line corresponds to the value

$$\hat{x}(\rho) = \mu(\alpha - 2)(1-\rho)^{-1} \log\bigl((1-\rho)^{-1}\bigr).$$

The simulated values of $P(W_\infty(\rho) > x)$ were obtained using the conditional Monte Carlo algorithm from Asmussen and Kroese (2006), and each point was estimated using enough simulation runs to obtain a relative error of at most 0.05 with approximately 99% confidence.

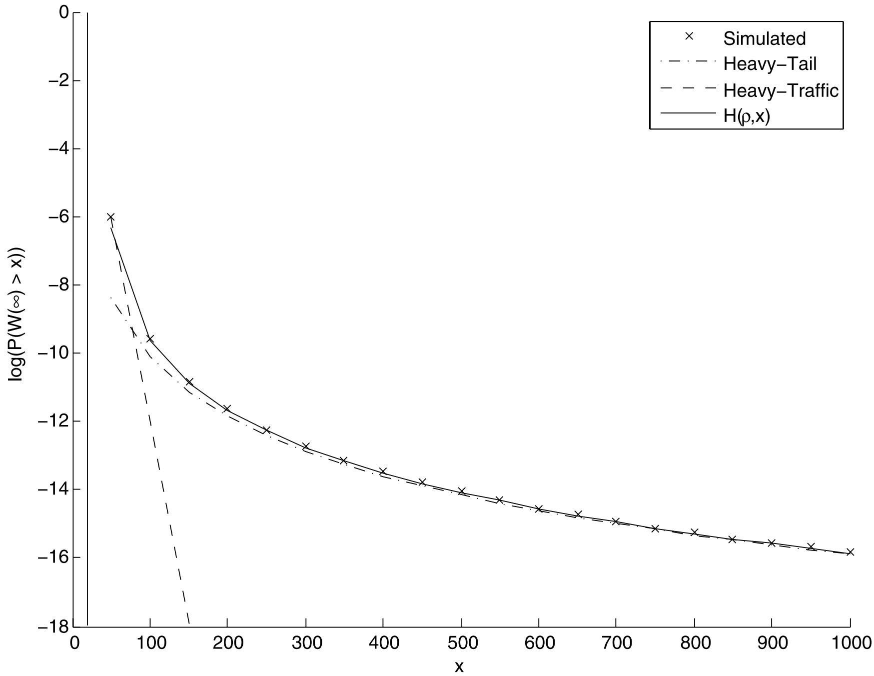


FIG. 2. *Pareto integrated tail with* $\rho = 0.8$ *and* $\alpha = 3.5$.

## REFERENCES


ABATE, J., CHOUDHURY, G. L. and WHITT, W. (1995). Exponential approximations for tail probabilities in queues. I. Waiting times. *Oper. Res.* **43** 885–901. MR1361346

ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd ed. *Applications of Mathematics* (*New York*) **51**. Springer, New York. MR1978607

ASMUSSEN, S. and KROESE, D. P. (2006). Improved algorithms for rare event simulation with heavy tails. *Adv. in Appl. Probab.* **38** 545–558. MR2264957

BACCELLI, F. and FOSS, S. (2004). Moments and tails in monotone-separable stochastic networks. *Ann. Appl. Probab.* **14** 612–650. MR2052896

BACCELLI, F., SCHLEGEL, S. and SCHMIDT, V. (1999). Asymptotics of stochastic networks with subexponential service times. *Queueing Systems Theory Appl.* **33** 205–232. MR1748644

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. *Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR898871

BLANCHET, J. and GLYNN, P. (2007). Uniform renewal theory with applications to expansions of random geometric sums. *Adv. in Appl. Probab.* **39** 1070–1097. MR2381589

BLANCHET, J., GLYNN, P. and LAM, H. (2010). Cramer–Lundberg approximations in the absence of exponential moments: A heavy-traffic perspective. To appear.

BOROVKOV, A. A. (2000). Estimates for the distribution of sums and maxima of sums of random variables when the Cramér condition is not satisfied. *Sibirsk. Mat. Zh.* **41** 997–1038. MR1803562

BOROVKOV, A. A. and BOROVKOV, K. A. (2008). *Asymptotic Analysis of Random Walks*. *Encyclopedia of Mathematics and Its Applications* **118**. Cambridge Univ. Press, Cambridge. MR2424161

DENISOV, D., DIEKER, A. B. and SHNEER, V. (2008). Large deviations for random walks under subexponentiality: The big-jump domain. *Ann. Probab.* **36** 1946–1991. MR2440928

EMBRECHTS, P. and VERAVERBEKE, N. (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance Math. Econom.* **1** 55–72. MR652832

GAMARNIK, D. and ZEEVI, A. (2006). Validity of heavy traffic steady-state approximation in generalized Jackson networks. *Ann. Appl. Probab.* **16** 56–90. MR2209336

HARRISON, J. M. and WILLIAMS, R. J. (1987). Brownian models of open queueing networks with homogeneous customer populations. *Stochastics* **22** 77–115. MR912049

IGLEHART, D. L. and WHITT, W. (1970a). Multiple channel queues in heavy traffic. I. *Adv. in Appl. Probab.* **2** 150–177. MR0266331

IGLEHART, D. L. and WHITT, W. (1970b). Multiple channel queues in heavy traffic. II. Sequences, networks, and batches. *Adv. in Appl. Probab.* **2** 355–369. MR0282443

KALASHNIKOV, V. (1997). *Geometric Sums: Bounds for Rare Events with Applications. Mathematics and Its Applications* **413**. Kluwer, Dordrecht. MR1471479

KINGMAN, J. F. C. (1961). The single server queue in heavy traffic. *Proc. Cambridge Philos. Soc.* **57** 902–904. MR0131298

KINGMAN, J. F. C. (1962). On queues in heavy traffic. *J. Roy. Statist. Soc. Ser. B* **24** 150–177. MR0148146

MIKOSCH, T. and NAGAEV, A. V. (1998). Large deviations of heavy-tailed sums with applications in insurance. *Extremes* **1** 81–110. MR1652936

NAGAEV, S. V. (1982). On the asymptotic behavior of one-sided large deviations. *Theory Probab. Appl.* **26** 362–366. MR616627

OLVERA-CRAVIOTO, M., BLANCHET, J. and GLYNN, P. W. (2010). On the transition from heavy traffic to heavy tails for the M/G/1 queue: The regularly varying case, internet supplement. Available at http://www.columbia.edu/~mo2291/.

OLVERA-CRAVIOTO, M. and GLYNN, P. W. (2010). Uniform approximations for the M/G/1 queue with subexponential processing times. To appear. Available at arXiv:1009.5327.

REIMAN, M. I. (1984). Open queueing networks in heavy traffic. *Math. Oper. Res.* **9** 441–458. MR757317

ROZOVSKIǏ, L. V. (1989). Probabilities of large deviations of sums of independent random variables with a common distribution function that belongs to the domain of attraction of a normal law. *Theory Probab. Appl.* **34** 625–644. MR1036709

SZCZOTKA, W. (1990). Exponential approximation of waiting time and queue size for queues in heavy traffic. *Adv. in Appl. Probab.* **22** 230–240. MR1039386

SZCZOTKA, W. (1999). Tightness of the stationary waiting time in heavy traffic. *Adv. in Appl. Probab.* **31** 788–794. MR1742694

M. OLVERA-CRAVIOTO
DEPARTMENT OF INDUSTRIAL ENGINEERING
AND OPERATIONS RESEARCH
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027
USA
E-MAIL: molvera@ieor.columbia.edu

J. BLANCHET
DEPARTMENT OF INDUSTRIAL ENGINEERING
AND OPERATIONS RESEARCH
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027
USA
E-MAIL: jose.blanchet@columbia.edu

P. GLYNN
DEPARTMENT OF MANAGEMENT SCIENCE
AND ENGINEERING
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305
USA
E-MAIL: glynn@stanford.edu