

# 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  $n^{\text{th}}$  customer for a first-in first-out (FIFO) single-server queue (with an infinite capacity waiting room) fed by a renewal arrival process (with iid inter-arrival times  $(\chi_n : n \geq 1)$ ) and an independent stream of iid processing times  $(V_n : n \geq 0)$ . If  $\rho \triangleq EV_1/E\chi_1 < 1$ , then  $W_n \Rightarrow W_\infty$  as  $n \rightarrow \infty$ , where  $W_\infty$  can be approximated via

$$(1.1) \quad W_\infty \stackrel{\mathcal{D}}{\approx} \frac{\text{Var}\chi_1 + \text{Var}V_1}{2(E\chi_1 - EV_1)} \text{Exp}(1)$$

when  $\rho$  is close to 1. Here,  $Exp(1)$  is an exponential rv with mean one and  $\stackrel{\mathcal{D}}{\approx}$  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) 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, for example, Iglehart and Whitt, 1970a,b) and then verifying that the steady-state rv  $W_\infty$  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

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

as  $x \rightarrow \infty$ , where  $\lambda \triangleq 1/E\chi_1$ ; see, for example, Embrechts and Veraverbeke (1982). Corresponding heavy-tailed steady-state asymptotics also exist in the context of queueing networks; see, for example, Baccelli et al. (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  $\rho \rightarrow 1$  (heavy traffic) and the other as  $x \rightarrow \infty$  (heavy tails). The interesting mathematical issue here is therefore to send  $\rho$  to 1 and  $x \rightarrow \infty$  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

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

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, 2009) studies the transition behavior in the setting of Weibullian tails and identifies 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 rv  $W_\infty$ .

**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 iid 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  $x \rightarrow \infty$ , where  $L(\cdot)$  is slowly varying (see p. 412 of Asmussen, 2003).

If  $\rho < 1$ ,  $W_n(\rho) \Rightarrow W_\infty(\rho)$  as  $n \rightarrow \infty$ , where the Pollaczek-Khintchine formula (see, for example, p. 237 of Asmussen, 2003) guarantees that

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

Here,  $S_n = X_1 + \dots + X_n$  (with  $S_0 = 0$ ), where the  $X_j$ 's are iid 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

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

as  $x \rightarrow \infty$ . It is straightforward (see, for example, p. 404 of Asmussen, 2003) to show that (2.2) in turn implies that

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

as  $x \rightarrow \infty$ .

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

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

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

$$(2.5) \quad P(W_\infty(\rho) > x) \approx \exp\left(-2(1-\rho)EV_1x/EV_1^2\right)$$

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

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

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  $nP(X_1 > \cdot)$ , i.e., 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  $n \rightarrow \infty$  provided  $n$  grows slowly enough compared to  $x$  (see, for example, Borovkov, 2000; Rozovskii, 1989). In other words, we can obtain the heavy-tail asymptotic from the first terms of (2.1),

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

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  $n \geq x/EX_1$ ,  $P(S_n > x) = O(1)$ , so by simply replacing  $P(S_n > x)$  by one we obtain

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

Since as  $\rho \nearrow 1$ ,  $\rho^{\lceil x/EX_1 \rceil} \sim e^{-(1-\rho)x/EX_1} = e^{-2(1-\rho)EV_1x/EV_1^2}$ , 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  $1 - \Phi\left((x - nEX_1)/\sqrt{\text{Var}(X_1)}\right)$  (or its stable law counterpart when  $\text{Var}(X_1) = \infty$ ) 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, 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 et al. (2009). 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

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

Then,

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

as  $x \rightarrow \infty$ . Alternatively,

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

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,*

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

as  $x \rightarrow \infty$ . Alternatively,

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

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*

$$y(\rho) = c\kappa(1 - \rho)^{-1} \log\left((1 - \rho)^{-1}\right)$$

for  $\rho < 1$ .

a.) If  $0 < c < 1$ , then

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

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

b.) If  $c > 1$ , then

$$(2.8) \quad \sup_{x \geq y} \left| \frac{P(W_\infty(\rho) > x)}{\frac{\rho}{1-\rho}P(X_1 > x)} - 1 \right| \rightarrow 0$$

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

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

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

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

b.) If  $c > 1$ , then

$$(2.10) \quad \sup_{0 < \rho \leq \hat{\rho}} \left| \frac{P(W_\infty(\rho) > x)}{\frac{\rho}{1-\rho}P(X_1 > x)} - 1 \right| \rightarrow 0$$

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

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*

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

*as  $\rho \nearrow 1$ .*

b.) *For  $x \geq 0$ ,*

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

*as  $\rho \nearrow 1$ .*

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

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

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 Whitt et al. (1995) and Blanchet and Glynn (2007). Finally, part c.) shows that the Cramér-Lundberg tail asymptotic (see, for example, p. 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 non-negative non-lattice iid random variables independent of the geometric rv  $N(p)$  having mass function

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

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  $x \rightarrow \infty$ . 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,*

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

as  $x \rightarrow \infty$ . Alternatively,

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

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*

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

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

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

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

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

**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

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

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

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 (1982), Rozovskii (1989), Mikosch and Nagaev (1998), Borovkov (2000), and more recently, Denisov et al. (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  $Y_1, Y_2, \dots$  be iid random variables having  $EY = 0$ ,  $\bar{F}(t) = P(Y_1 > t)$  and  $\bar{F}(t) = t^{-\beta}L(t)$  where  $L(\cdot)$  is slowly varying. Set  $S_n = Y_1 + \dots + Y_n$ ,  $n \geq 1$ .*

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

*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 iid 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  $S_n = X_1 + \dots + X_n$ ,  $n \geq 1$ . 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  $S_n^* = Y_1 + \dots + Y_n$ . Then the result will follow from Theorem 3.1 a.) once we show that  $(x - n\mu)/\sigma(n) \rightarrow \infty$  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  $\bar{F}(t) = P(Y_1 > t)$  is regularly varying with index  $\alpha - 1$ , then  $\sigma(n) = \bar{F}^{-1}(1/n) = n^{1/(\alpha-1)} \tilde{L}(n)$  for some slowly varying function  $\tilde{L}(\cdot)$  (see Bingham et al., 1987). Then the result will follow from Theorem 3.1 b.) once we show that  $(x - n\mu)/\sigma(n) \rightarrow \infty$  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 et al. (2009) for the details.  $\square$

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  $x \rightarrow \infty$  and  $\rho \nearrow 1$ , respectively.

LEMMA 3.3. *Suppose that*

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

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

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

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

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

as  $\rho \nearrow 1$ .

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

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

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

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

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

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

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

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

The last limit is zero by the standard heavy traffic limit.  $\square$

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

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

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

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

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  $\varphi_1(t) \rightarrow 0$  as  $t \rightarrow \infty$  such that

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

By (3.1) we have

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

Clearly,

$$(3.2) \quad \varphi_1(x) \sum_{n=1}^{M_\gamma(x)} (1-\rho)\rho^n n P(X_1 > x - (n-1)\mu) \leq \varphi_1(x) S(\rho, x).$$

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

$$\begin{aligned} & \sum_{n=M_\gamma(x)+1}^{M(x)} (1-\rho)\rho^n n P(X_1 > x - (n-1)\mu) \\ & \leq Cx(1-\rho) \sum_{n=M_\gamma(x)+1}^{M(x)} \rho^n (x - (n-1)\mu)^{-\alpha+1+\epsilon}. \end{aligned}$$

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

$$\begin{aligned}
& x(1 - \rho) \sum_{n=M_\gamma(x)+1}^{M(x)} g(n) \\
& \leq x(1 - \rho)(M(x) - M_\gamma(x)) \max\{g(M_\gamma(x) + 1), g(M(x))\} \\
& \leq C(1 - \rho)x^{1+\gamma} \max\left\{\rho^{(x-x^\gamma)/\mu}x^{-\gamma(\alpha-1-\epsilon)}, \rho^{(x-x^\beta)/\mu-1}x^{-\beta(\alpha-1-\epsilon)}\right\} \\
(3.3) \quad & \leq C(1 - \rho)x^{1+\gamma-\beta(\alpha-1-\epsilon)}\rho^{(x-x^\gamma)/\mu},
\end{aligned}$$

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

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

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

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

It is easy to check that

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

For the tail of the sum let  $Y_i = \mu - X_i$  and  $S_n^* = Y_1 + \dots + Y_n$ . 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  $n > K(x)$ ,

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

where

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

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

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

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

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

We thus have that

$$(3.5) \quad \left| \sum_{n=\lfloor x/\mu \rfloor + 1}^{\infty} (1-\rho)\rho^n P(S_n > x) - \rho^{x/\mu} \right| \leq \varphi_2(x)\rho^{x/\mu} + C\rho^{x/\mu} |\log \rho| x^\gamma.$$

Combining (3.2), (3.3), (3.4) and (3.5) gives

$$\begin{aligned} & \left| P(W_\infty(\rho) > x) - S(\rho, x) - \rho^{x/\mu} \right| \\ & \leq \varphi_1(x)S(\rho, x) + \varphi_2(x)\rho^{x/\mu} + C\rho^{(x-x^\gamma)/\mu} |\log \rho| x^v, \end{aligned}$$

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

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

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

$$\begin{aligned} & \sup_{0 < \rho < \rho(x)} \frac{\rho^{(x-x^\gamma)/\mu} |\log \rho| x^v}{S(\rho, x) + \rho^{x/\mu}} \\ & \leq \sup_{0 < \rho < \rho(x)} \frac{\rho^{(x-x^\gamma)/\mu} |\log \rho| x^v}{L(x)x^{-\alpha+1}\rho(1-\rho^{\lfloor x/\mu \rfloor})} \\ & \leq C \sup_{0 < \rho < \rho(x)} \rho^{(x-x^\gamma)/\mu-1} |\log \rho| x^{v+\alpha-1+\epsilon} \\ & \leq C \sup_{t > (x^v \log x)^{-1}} \exp\left(-\left(\frac{x-x^\gamma}{\mu} - 1\right)t + \log t + (v+\alpha-1+\epsilon)\log x\right) \\ & = C \exp\left(-\frac{x^{1-v}}{\mu \log x} \left(1 - \frac{1}{x^{1-\gamma}} - \frac{\mu}{x}\right) - \log \log x + (\alpha-1+\epsilon)\log x\right) \rightarrow 0. \end{aligned}$$

□

We now prove Theorem 2.2.

PROOF OF THEOREM 2.2. Again, we only proof 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  $\rho \nearrow 1$ ,

$$\begin{aligned} & \sup_{0 < x < (1-\rho)^{-1/4}} \frac{\rho}{1-\rho} \left( 1 - \rho^{x/\mu} - \frac{(1-\rho)x}{\mu} \rho^{x/\mu} \right) P(X_1 > x) \\ & \leq \sup_{0 < x < (1-\rho)^{-1/4}} \frac{\rho}{1-\rho} \left( |\log \rho| \frac{x}{\mu} - \frac{(1-\rho)x}{\mu} \rho^{x/\mu} \right) \\ & = \frac{1}{\mu(1-\rho)^{5/4}} \left( |\log \rho| - (1-\rho)e^{-|\log \rho|(1-\rho)^{-1/4}/\mu} \right) \\ & = O\left((1-\rho)^{1/2}\right). \end{aligned}$$

By Theorem 2.1 we only need to show that

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

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

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

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

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

The second term can be bounded as follows,

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

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

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

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

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

It follows that

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

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

$$\begin{aligned}
& P(X_1 > x) \sum_{n=N(x)+1}^{M(x)} (1-\rho)\rho^n n \left( \frac{P(X_1 > x - (n-1)\mu)}{P(X_1 > x)} - 1 \right) \\
& \leq Cx(1-\rho)\rho^{N(x)+1} \sum_{n=N(x)+1}^{M(x)} P(X_1 > x - (n-1)\mu) \\
& \leq Cx(1-\rho)\rho^{(1-\epsilon)x/\mu} \int_{x-\mu M(x)}^{x-\mu N(x)} P(X_1 > t) dt \\
& \leq Cx(1-\rho)\rho^{(1-\epsilon)x/\mu} (x - \mu M(x)) P(X_1 > x - \mu M(x)) \\
(3.8) \quad & \leq C(1-\rho)\rho^{(1-\epsilon)x/\mu} x^{1+\beta} P(X_1 > x^\beta),
\end{aligned}$$

where for the third inequality we used Proposition 1.5.10 in Bingham et al. (1987). Combining (3.6), (3.7) and (3.8) gives

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

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  $\frac{(1-\beta)(\alpha-2+\epsilon)}{(\alpha-2)(1-\epsilon)} < c < 1$ . Note that  $\gamma(x, \rho) \sim 1$  as  $x \rightarrow \infty$  uniformly for  $0 < \rho \leq \rho(x)$ . Then,

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

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 et al. (1987), p. 25) to obtain

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

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

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

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

Then,

$$\begin{aligned} & \sup_{\rho(x) \leq \rho < 1} \frac{1}{A(\rho, x)} \left| S(\rho, x) - \frac{\rho}{1-\rho} (1 - \rho^{x/\mu}) P(X_1 > x) \right| \\ & \leq C \sup_{\rho(x) \leq \rho < 1} \left\{ \frac{P(X_1 > x)}{1-\rho} (\rho^{-x^\beta/\mu} - \rho) (1 + (1-\rho)x) \right. \\ & \quad \left. + \frac{\rho P(X_1 > x) \gamma(x, \rho)}{x(1-\rho)^2 \rho^{x/\mu}} + \varphi_1(x) + (1-\rho) \rho^{-\epsilon x/\mu} x^{1+\beta} P(X_1 > x^\beta) \right\} \\ & \leq C \sup_{\rho(x) \leq \rho < 1} \left\{ \frac{P(X_1 > x)}{1-\rho} |\log \rho| x^\beta \log x + \rho^{-\epsilon x/\mu} x^\beta \log x P(X_1 > x^\beta) \right. \\ & \quad \left. + x^{-1} P(X_1 > x) \left( \sum_{n=1}^{\lfloor x/\mu \rfloor + 1} \rho^{n - \lfloor x/\mu \rfloor - 1} n + (1-\rho)^2 x \right) + \varphi_1(x) \right\} \\ & \leq C \left\{ P(X_1 > x) x^\beta \log x + \rho(x)^{-\epsilon x/\mu} x^\beta \log x P(X_1 > x^\beta) \right. \\ & \quad \left. + P(X_1 > x) \sum_{n=1}^{\lfloor x/\mu \rfloor + 1} \rho(x)^{n - \lfloor x/\mu \rfloor - 1} + \frac{(\log x)^2}{x^2} P(X_1 > x) + \varphi_1(x) \right\}. \end{aligned}$$

The first, fourth and fifth terms clearly converge to zero. To see that the second and third terms converge to zero as well note that

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

and

$$\begin{aligned} P(X_1 > x) \sum_{n=1}^{\lfloor x/\mu \rfloor + 1} \rho(x)^{n - \lfloor x/\mu \rfloor - 1} &\leq Cx^{-\alpha + \epsilon} \frac{(1 - \rho(x))^{\lfloor x/\mu \rfloor + 1}}{\rho(x)^{\lfloor x/\mu \rfloor} (1 - \rho(x))} \\ &\leq Cx^{-\alpha + \epsilon + 1 + c(\alpha - 2)} (\log x)^{-1} \\ &\leq Cx^{-1 + \epsilon} (\log x)^{-1}. \end{aligned}$$

Our choice of  $\epsilon$  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

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

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

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

as  $\rho \nearrow 1$ . Note that

$$\begin{aligned} &\sup_{0 \leq x \leq y} \left| \frac{\frac{\rho}{1-\rho} \gamma(x, \rho) P(X_1 > x) + \rho^{x/\mu}}{\exp(-(1-\rho)x/\mu)} - 1 \right| \\ (3.9) \quad &\leq \sup_{0 \leq x \leq y} \left| \exp\left(\frac{x \log \rho}{\mu} + \frac{(1-\rho)x}{\mu}\right) - 1 \right| \end{aligned}$$

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

We can bound (3.9) as follows

$$\begin{aligned} \sup_{0 \leq x \leq y} \left| \exp\left(\frac{x \log \rho}{\mu} + \frac{(1-\rho)x}{\mu}\right) - 1 \right| &\leq \sup_{0 \leq x \leq y} \frac{x |\log \rho| + 1 - \rho}{\mu} \\ &\leq Cy(1-\rho)^2 \\ &\leq Ct^{-1} \log t, \end{aligned}$$

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

$$\begin{aligned} \gamma(x, \rho) &= 1 - \rho^{x/\mu} - \rho^{x/\mu}(1-\rho)x/\mu \\ &\leq (|\log \rho| - (1-\rho))x/\mu + |\log \rho|(1-\rho)(x/\mu)^2 \\ &\leq C(1-\rho)^2 x^2. \end{aligned}$$

Then (3.10) is bounded by

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

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

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

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

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

as  $\rho \nearrow 1$ . Note that

$$\begin{aligned}
& \sup_{x \geq y} \left| \frac{\frac{\rho}{1-\rho} \gamma(x, \rho) P(X_1 > x) + \rho^{x/\mu}}{\frac{\rho}{1-\rho} P(X_1 > x)} - 1 \right| \\
& \leq \sup_{x \geq y} |\gamma(x, \rho) - 1| + \sup_{x \geq y} \frac{(1-\rho) \rho^{x/\mu}}{\rho P(X_1 > x)} \\
& \leq \rho^{y/\mu} (1 + (1-\rho)y/\mu) + C \sup_{x \geq y} \frac{1-\rho}{L(x)} \exp\left(-\frac{x}{\mu}(1-\rho) + (\alpha-1) \log x\right) \\
& \leq C \rho^{y/\mu} (1-\rho)y + C \frac{1-\rho}{L(y)} \exp\left(-\frac{y}{\mu}(1-\rho) + (\alpha-1) \log y\right) \\
& \leq C t^{-c(\alpha-2)} \log t + \frac{C}{L(t \log t)} \exp(-(c-1)(\alpha-2) \log t + (\alpha-1) \log \log t),
\end{aligned}$$

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

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

□

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

$$\begin{aligned}
H(\rho, x) & \triangleq S(\rho, x) + \rho^{x/\mu} \\
& = \sum_{n=1}^{M(x)} (1-\rho) \rho^n n P(X_1 > x - (n-1)\mu) + \rho^{x/\mu}
\end{aligned}$$

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 the Pollaczek-Khintchine formula than  $J(\rho, x)$ .

When  $\sigma^2 = \text{Var}(X_1) < \infty$ , 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

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

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

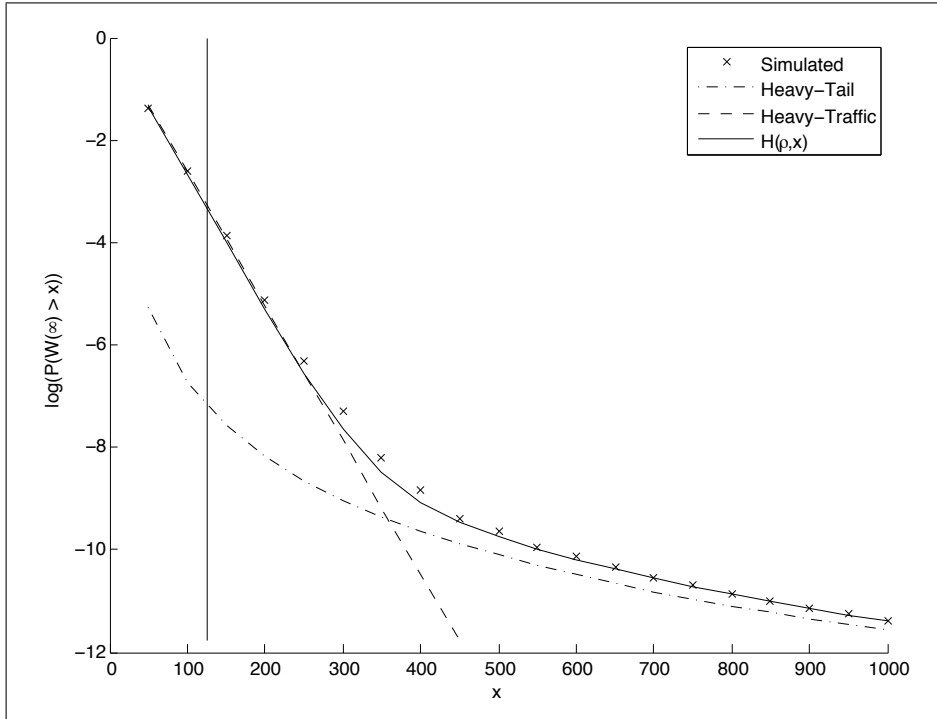


FIG 1. Pareto integrated tail with  $\rho = 0.95$  and  $\alpha = 3.1$ .

We plotted approximation  $H(\rho, x)$  against simulated values of  $P(W_\infty(\rho) > x)$

$x$ ). Figures 1 and 2 correspond to queues having Pareto integrated tail distribution, i.e.,  $P(X_1 > x) = x^{-\alpha+1}$  for  $x \geq 1$ . For comparison purposes we also plotted the heavy-traffic approximation,

$$\text{Heavy-Traffic} = \exp(-(1 - \rho)x/\mu)$$

and the heavy-tail asymptotic,

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

The vertical line corresponds to the value

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

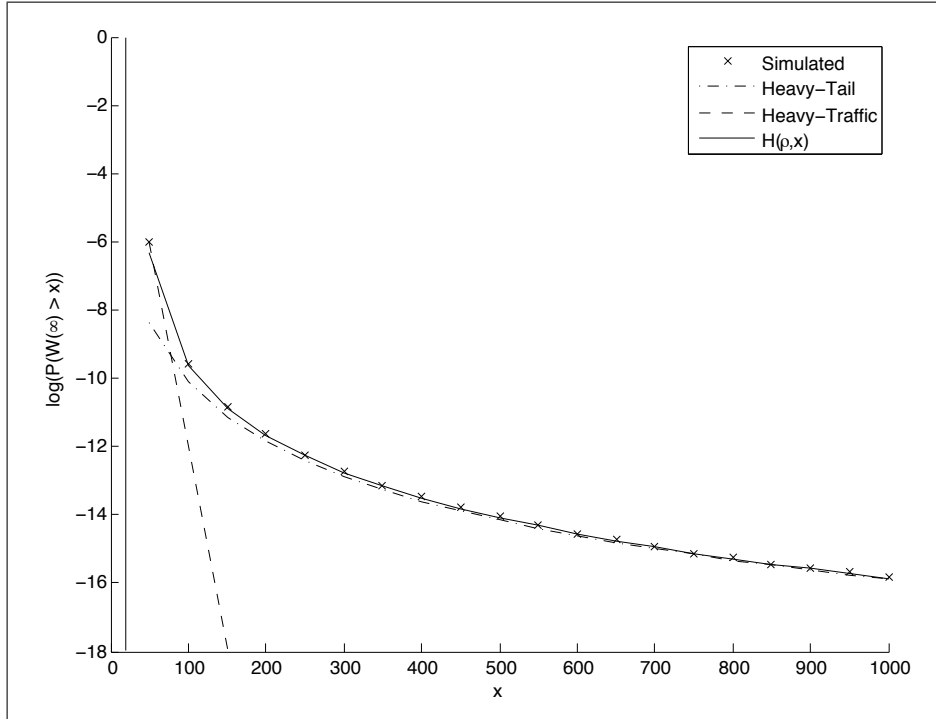


FIG 2. Pareto integrated tail with  $\rho = 0.8$  and  $\alpha = 3.5$ .

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.

## References.

- Asmussen, S. (2003). *Applied Probability and Queues*. Springer, New York.
- Asmussen, S. and Kroese, D. (2006). Improved algorithms for rare event simulation with heavy tails. *Adv. Appl. Prob.*, 38:545–558.
- Baccelli, F. and Foss, S. (2004). Moments and tails in monotone separable stochastic networks. *Ann. Appl. Prob.*, 14:612–650.
- Baccelli, F., Schlegel, S., and Schmidt, V. (1999). Asymptotics of stochastic networks with subexponential service times. *Queueing Systems*, 33:205–232.
- Bingham, N. H., Goldie, C., and Teugels, J. (1987). *Regular Variation*. Cambridge University Press, Cambridge.
- Blanchet, J. and Glynn, P. (2007). Uniform renewal theory with applications to expansions of random geometric sums. *Adv. Appl. Prob.*, 39(4):1070–1097.
- Blanchet, J., Glynn, P., and Lam, H. (2009). Cramer-lundberg approximations in the absence of exponential moments: a heavy-traffic perspective. In preparation.
- Borovkov, A. (2000). Estimates for the distribution of sums and maxima of sums of random variables without the Cramer condition. *Siberian Math J.*, 41(5):997–1038.
- Borovkov, A. and Borovkov, K. (2008). *Asymptotic Analysis of Random Walks*. Cambridge University Press, New York.
- Denisov, D., Dieker, A., and Shneer, V. (2008). Large deviations for random walks under subexponentiality: the big-jumb domain. *Ann. Probab.*, 36.
- Embrechts, P. and Veraverbeke, N. (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics*, 1:55–72.
- Gamarnik, D. and Zeevi, A. (2006). Validity of heavy traffic steady-state approximations in generalized Jackson networks. 16(1):56–90. *Ann. Appl. Probab.*
- Harrison, J. and Williams, R. (1987). Brownian models for open queueing networks with homogeneous customer population. *Stochastics*, 22:77–115.
- Iglehart, D. and Whitt, W. (1970a). Multiple channel queues in heavy traffic, I. *Adv. Appl. Prob.*, 2:150–177.
- Iglehart, D. and Whitt, W. (1970b). Multiple channel queues in heavy traffic, II. *Adv. Appl. Prob.*, 2:355–369.
- Kalashnikov, V. (1997). *Geometric Sums: Bounds for Rare Events with Applications: Risk Analysis, Reliability, Queueing*. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Kingman, J. (1961). The single server queue in heavy traffic. *Proc. Camb. Phil.Soc.*, 57:902–904.
- Kingman, J. (1962). On queues in heavy traffic. *J. Roy. Statist. Soc. Ser. B*, 24:150–177.
- Mikosch, T. and Nagaev, S. (1998). Large deviations of heavy-tailed sums with applications in insurance. *Extremes*, 1(1):81–110.
- Nagaev, S. (1982). On the asymptotic behavior of one-sided large deviation probabilities. *Theory Probab. Appl.*, 26(2):362–366.
- Olvera-Cravioto, M., Blanchet, J., and Glynn, P. (2009). On the transition from heavy traffic to heavy tails for the M/G/1 queue: The regularly varying case, internet supplement. <http://www.columbia.edu/~mo2291/>.

- Olvera-Cravioto, M. and Glynn, P. (2009). On the transition from heavy traffic to heavy tails for the M/G/1 queue: The semiexponential case. Submitted for publication.
- Reiman, M. (1984). Open queueing networks in heavy traffic. *Math. Oper. Res.*, 9:441–458.
- Rozovskii, L. (1989). Probabilities of large deviations of sums of independent random variables with common distribution function in the domain of attraction of the normal law. *Theory Probab. Appl.*, 34:625–644.
- Szczotka, W. (1990). Exponential approximation of waiting time and queue size for queues in heavy traffic. *Adv. Appl. Prob.*, 22:230–240.
- Szczotka, W. (1999). Tightness of the stationary waiting time in heavy traffic. *Adv. Appl. Prob.*, 31:788–794.
- Whitt, W., Abate, J., and Choudhury, G. L. (1995). Exponential approximations for tail probabilities in queues, I: Waiting times. *Operations Research*, 43:885–901.

E-MAIL: molvera@ieor.columbia.edu

E-MAIL: jose.blanchet@columbia.edu

E-MAIL: glynn@stanford.edu