Sampling at subexponential times, with queueing applications

Søren Asmussen* Claudia Klüppelberg† Karl Sigman‡

Abstract

We study the tail asymptotics of the r.v. $X(T)$ where $\{X(t)\}$ is a stochastic process with a linear drift and satisfying some regularity conditions like a central limit theorem and a large deviations principle, and T is an independent r.v. with a subexponential distribution. We find that the tail of $X(T)$ is sensitive to whether or not T has a heavier or lighter tail than a Weibull distribution with tail $e^{-\sqrt{x}}$. This leads to two distinct cases, heavy-tailed and moderately heavy-tailed, but also some results for the classical light-tailed case are given. The results are applied via *distributional Little's law* to establish tail asymptotics for steady–state queue length in GI/GI/1 queues with subexponential service times. Further applications are given for queues with vacations, and M/G/1 busy periods.

Keywords Busy period, independent sampling, Laplace's method, large deviations, Little's law, Markov additive process, Poisson process, random walk, regular variation, subexponential distribution, vacation model, Weibull distribution

Classification 60J25, 60K25

*Department of Mathematical Statistics, University of Lund, Box 118, S–221 00 Lund, Sweden, email: asmus@maths.lth.se

†Center of Mathematical Sciences, Munich University of Technology, D-80290 Munich, Germany, email: cklu@ma.tum.de, http://www.ma.tum.de/stat/

‡Department of Industrial Engineering and Operations Research, Columbia University, 500 West 120th Street, MC:4704, NY, NY 10027, USA; email: sigman@ieor.columbia.edu

1 Introduction

Let $\{X(t)\}_{t\geq 0}$ be a stochastic process satisfying a LLN $X(t)/t \to m > 0$, and T an independent random time. The question we address in this paper is to derive the tail asymptotics of the r.v. $X(T)$, with particular emphasis on the case where T has a heavy–tailed (subexponential) distribution.

A main special case where $\{X(t)\}$ is a Poisson process at rate λ and T has a regularly varying tail, $P(T > x) = L(x)/x^\alpha$ with $\alpha > 0$ and $L(x)$ slowly varying, has been considered in [26] Ch. 8. The results there basically show that in this setting, the variability of T dominates that of $\{X(t)\}$ so that one can replace $X(t)$ by its expected value λt to get

$$P(X(T) > x) \ \sim\ P(\lambda T > x)\,. \tag{1.1}$$

Here and in the rest of the paper, $\sim$ means that the ratio is one in the limit $x \to \infty$. For the regularly varying case, (1.1) is also shown to hold in [39] when $\{X(t)\}$ is special Markov additive process (this is a key step in the study of a certain fluid queue). In [40], the question of subexponentiality of $X(T)$ is addressed in some special settings, but no tail estimates are given.

The results that we give are in part generalizations of (1.1), allowing a more general structure of $\{X(t)\}$ and/or more general distributions of T. However, we also find that even for the Poisson case, (1.1) does not extend to the whole class of subexponential distributions: when T has a lighter tail than the Weibull tail e^{-x^β} with $\beta = 1/2$, the deviation of $X(t)$ from λt described by the CLT makes a small but non–negligible contribution to the tail of $X(T)$. To see this, assume that $P(T > x) = \exp\{-x^\beta\}$, $\beta > 1/2$, and let $\{N(t)\}$ be a Poisson process at rate $\lambda = 1$. By the CLT, $(N(t) - t)/\sqrt{t} \to N(0,1)$ in distribution so that

$$\liminf_{x\to\infty} \inf_{y\geq x-\sqrt{x}} P(N(y) > x) \geq 1 - \Phi(-1) = \Phi(1) > 0\,,$$

($\Phi(x)$ is the standard normal d.f.). This yields

$$\begin{aligned}\liminf_{x\to\infty} \frac{P(N(T) > x)}{P(T > x)} \ &\geq\ \Phi(1) \liminf_{x\to\infty} \frac{P(T > x - \sqrt{x})}{P(T > x)} \\ &=\ \Phi(1) \lim_{x\to\infty} \exp\{x^\beta - (x - \sqrt{x})^\beta\} \ =\ \Phi(1) \lim_{x\to\infty} \exp\{\beta x^{\beta-1/2}\} \ =\ \infty\,.\end{aligned}$$

This behaviour can be seen as intermediate between (1.1) and the classical light–tailed case where $X(T)$ becomes large only when both $\{X(t)\}$ attains atypically large values and T is large at the same time; see for example [29] for a discrete time version where $\{X(t)\}$ is a random walk and T has a Poisson distribution, and Section 5 of the present paper. For this reason, we refer to a distribution with a tail like e^{-x^β} with $1/2 \le \beta < 1$ as *moderately heavy–tailed.*

Our study was motivated by a queueing problem: determining tail asymptotics of steady-state queue length L (total number in system) in a stable FIFO M/G/1 queue with Poisson arrivals $\{N(t)\}$ at rate λ and generic service time S. Most general asymptotic results known for L involve establishing geometric tails in the light–tailed case (S has finite moment generating function in a neighborhood of the origin):

$$P(L > k) \sim \beta\sigma^{-k}, \quad k \to \infty . \tag{1.2}$$

See for example [23], [32], [6], [12], [34] and [1]. The connection to the general problem outlined above is provided by *distributional Little's law* (DLL), cf. [27], which asserts that L has the same distribution as $N(W)$ where W (the steady-state sojourn time) is chosen independent of $\{N(t)\}$. Here the tail behaviour of W has been known for a long time: W has an asymptotic exponential tail in the light–tailed case (see e.g. [2], [3], [24]), whereas in the heavy–tailed case (service times S have mean $1/\mu$ and are subexponential, $\rho = \lambda/\mu$),

$$P(W > x) \sim \frac{\rho}{1-\rho} P(S_e > x), \quad x \to \infty, \tag{1.3}$$

where S_e has the *equilibrium* density $\mu P(S > x)$, cf. [13], [35], [21], [8], [9], [28], [10]. In particular, when (1.1) applies with $T = W$, we get

$$P(L > k) \ \sim \ P(\lambda W > k) \ \sim \ \frac{\rho}{1-\rho} P(\lambda S_e > k), \quad k \to \infty. \tag{1.4}$$

That is, asymptotically the tail of L is exactly like the tail of λW, a kind of "generalized Little's law". We obtain results also for more general cases, like S Weibull with $1/2 \le \beta < 1$ and more general queueing models, in particular GI/G/1, tandem, and vacation queues for which DLL holds.

The paper is organized as follows. Section 2 contains some preliminaries. In Section 3, we study $X(T)$ when T has a distribution somewhat more general than subexponential, with tails heavier than a Weibull $\overline{F}(x) \sim \exp\{-\sqrt{x}\}$. We start with the case where $\{X(t)\}$ is a Poisson process and give a first application to M/G/1 queues. Subsection 3.2 generalizes (1.1) away from the Poisson case. The main result is Theorem 3.7. Again distributions of T with a heavier tail than $e^{-\sqrt{x}}$ satisfy the required conditions. Here the link to extreme value theory becomes obvious. We present several examples for processes $\{X(t)\}$, which satisfy the conditions of Theorem 3.7. In Section 4, we give the precise asymptotics for certain moderately heavy–tailed r.v.s T, when $\{X(t)\}$ is Poisson. For the sake of completeness, in Section 5 we include some results on the light–tailed case which are in part expected but not in the literature. Finally, Section 6 gives some further queueing applications, in particular to vacation models and M/G/1 busy periods.

2 Preliminaries

2.1 Subexponential distributions

Given a non-negative random variable (r.v.) X, its distribution function (df) is denoted by $F(x) = P(X \leq x)$ and its tail by $\overline{F}(x) = 1 - F(x) = P(X > x)$. We are interested in d.f.'s that are *heavy-tailed*: $\overline{F}(x) > 0,\ x \geq 0$, and

$$\lim_{x\to\infty} P(X > x + y \mid X > x) \;=\; \lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} \;=\; 1\,, \qquad y \geq 0. \tag{2.1}$$

For our purposes we focus on a special class $\mathcal{S}$ of such distributions called *subexponential distributions* F. The reader is referred to [25] or [20] for details and further references. If F^{*n} denotes the n-fold convolution of F, $F^{*2}(x) = \int_0^x F(x-y)dF(y)$ and so on, with corresponding tail $\overline{F^{*n}}(x) = 1 - F^{*n}(x)$, then the d.f. F (or the r.v. X) is called subexponential if $\overline{F}(x) > 0,\ x \geq 0$, and for all $n \geq 2$,

$$\lim_{x\to\infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} = n. \tag{2.2}$$

(It can be shown that if the condition holds for $n = 2$ then it holds for all $n \geq 2$.)

In terms of r.v.'s, (2.2) can then be re-stated as

$$P(X_1 + \cdots + X_n > x) \sim P(\max\{X_1, \ldots, X_n\} > x),\ x \to \infty,$$

for all $n \geq 2$ where $X_1, \ldots, X_n$ are i.i.d. distributed as F. In words this means that the sum is likely to get large because one of the r.v.'s gets large. If X is subexponential then in fact $\lim_{x\to\infty} e^{\epsilon x} P(X > x) = \infty = E(e^{\epsilon X})$ for all $\epsilon > 0$, which explains why the term *subexponential* is used in the definition.

For technical reasons we sometimes restrict this class even further to the class $\mathcal{S}^* \subset \mathcal{S}$, introduced in [31] and defined by

Definition 2.1 *(The class $\mathcal{S}^*$) Let F be a distribution on $[0,\infty)$ such that $\overline{F}(x) > 0,\ x \geq 0$. We say that $F \in \mathcal{S}^*$ if F has finite first moment $1/\mu$ and*

$$\lim_{x\to\infty} \int_0^x \frac{\overline{F}(x-y)}{\overline{F}(x)} \overline{F}(y)dy = \frac{2}{\mu}. \tag{2.3}$$

$\mathcal{S}^*$ includes (when the mean is finite) the following distributions: Pareto, Burr, log-gamma, lognormal, heavy-tailed Weibull, and many others.

Of special importance to us is the *Weibull* distribution with parameter β:

$$\overline{F}(x) = e^{-x^\beta}, \quad x \geq 0, \quad 0 < \beta < 1.$$

2.2 Equilibrium distributions

For any non-negative random variable X with distribution F and finite mean $1/\mu$, the *equilibrium distribution* F_e is defined by

$$F_e(x) = \mu \int_0^x \overline{F}(y)dy, \quad x > 0. \tag{2.4}$$

We let X_e denote a r.v. distributed as F_e.

One of the important features of $\mathcal{S}^*$ for applications is the following:

Proposition 2.2 *If $F \in \mathcal{S}^*$, then both F and F_e are subexponential.*

Note that for any df F satisfying (2.1) (i.e. in particular for any $F \in \mathcal{S}^*$) the tail of F_e dominates that of F: $\overline{F}_e(x)/\overline{F}(x) \to \infty$.

2.3 Basics of the FIFO GI/GI/1 queue

Customer interarrival times $\{T_n\}$ are i.i.d. with finite mean $1/\lambda$, and service times $\{S_n\}$ are i.i.d. distributed as $G(x) = P(S \leq x)$ with finite mean $1/\mu$. The two sequences are assumed independent. $\{N(t)\}$ denotes the counting process of arrivals. We assume throughout $\rho = \lambda/\mu < 1$ (stability). Customers join the queue in the order they arrive (First In queue First

Out of queue, FIFO). The delay of the n^{th} customer (in queue) is denoted by D_n and satisfies the recursion

$$D_{n+1} = (D_n + S_n - T_n)_+, \ n \geq 0.$$

D denotes steady-state delay: $P(D \leq x) = \lim_{n\to\infty} P(D_n \leq x)$. The following is a precise rendering of result (1.3) (see Embrechts and Veraverbeke [21] (ruin probability setting) and Pakes [35] (queueing setting)):

Theorem 2.3 *D is subexponential if S_e is subexponential, and in this case*

$$P(D > x) \sim \frac{\rho}{1-\rho} P(S_e > x)\,, \quad x \to \infty. \tag{2.5}$$

Note in particular (recall Proposition 2.2) that if $G \in \mathcal{S}^*$, then (2.5) holds.

Steady-state *sojourn time* $W = D + S$ (independent sum) denotes total time spent in system and it is easily seen (from Theorem 2.3 and basic principles) that if S_e is subexponential then

$$P(W > x) \sim P(D > x)\,, \quad x \to \infty\,, \tag{2.6}$$

because the tail of S_e (and hence that of D) dominates that of S (see e.g. [20], Lemma A3.28). That's how one gets (1.3).

Remark 2.4 The M/G/1 queue is the special case when the interarrival time distribution is exponential; i.e. the arrival process is a homogenous Poisson process. In this case the implications of Theorem 2.3 become equivalences; i.e. D is subexponential if and only if S_e is subexponential if and only if (2.5) holds. The "if and only if" aspect shows how fundamental the subexponential property is in the context of applications to queues. For it implies that if S_e is heavy-tailed but not subexponential, then the asymptotic (2.5) will not hold. $\square$

2.4 Distributional Little's Law

Consider a queueing model with renewal arrivals (i.i.d. interarrival times $\{T_n\}$). Let W_n denote n^{th} customer's sojourn time (total time spent in the system from arrival to departure). Let L denote steady-state number in system, and W denote steady-state sojourn time. Finally, independent of W, let $\{N(t)\}$ denote a time stationary version of the renewal counting process (the initial arrival time is distributed as T_e (equilibrium distribution)).

The following result is from [27], and known as *distributional Little's law* (DLL):

Proposition 2.5 *If (1) and (2) below hold then $L = N(W)$ in distribution.*

(1) Customers depart the system in the same order that they arrived (first-in-first-out).

(2) W_n is independent of the future interarrival times $\{T_n, T_{n+1}, \ldots\}$, $n \geq 0$.

Some models for which DLL holds are FIFO GI/GI/1 queue, FIFO GI/GI/1 queue with server vacations, FIFO tandem queues of the form GI/GI/1 $\longrightarrow$ /GI/1 $\longrightarrow \cdots$ /GI/1. DLL does not hold for FIFO multi-server queues (such as GI/GI/c) because Condition (1) above then fails (unless service times are deterministic); but it does hold for the number of customers *waiting in the queue* (not in service) for such models (for then Condition (1) does hold); note, however, that the tail asymptotics of D or W is not at present available for GI/GI/c queues with subexponential S. Nor does DLL hold for queues with non-renewal arrivals because otherwise Condition (2) will fail (except for extremely trivial cases). It is not crucial that service times be i.i.d. , so, for example, DLL holds for GI/G/1 queues (and tandem and vacation) in which the service time sequence is stationary and independent of the renewal arrival process. (More recent references on DLL (since the classic paper of Haji and Newell [27]) are [30] and [17] for example.)

In what follows, DLL is our route to studying $P(L > k)$ due to well known asymptotics for W (such as (2.5) and (2.6)); this differs from much classical work where the approach is via transforms, cf. [23], [32], [34], [1].

2.5 Some basic lemmas

Lemma 2.6 *If $\{N(t)\}$ is Poisson with rate λ and T an independent random time with d.f. F, then*

$$P(N(T) \geq k-1) \;=\; \frac{\lambda^k}{(k-1)!}\int_{-\infty}^{\infty} e^{ku} g(u)\,du\,, \quad k \in \mathbb{N}\,, \tag{2.7}$$

where $g(u) = \overline{F}(e^u)e^{-\lambda e^u}$.

Proof By partial integration, for $k \in \mathbb{N}$,

$$P(N(T) \geq k-1) \;=\; \int_0^{\infty} P(N(t) \geq k-1)\,dF(t) \;=\; \lambda \int_0^{\infty} e^{-\lambda t}\frac{(\lambda t)^{k-1}}{(k-1)!}\overline{F}(t)\,dt$$

set $u = \ln t$

$$= \frac{\lambda^k}{(k-1)!}\int_{-\infty}^{\infty} e^{-\lambda e^u} e^{(k-1)u}\overline{F}(e^u)e^u du = \frac{\lambda^k}{(k-1)!}\int_{-\infty}^{\infty} e^{ku}\overline{F}(e^u)e^{-\lambda e^u} du$$

$$= \frac{\lambda^k}{(k-1)!}\int_{-\infty}^{\infty} e^{ku} g(u) du\,. \tag{2.8}$$

Hence (2.7) is up to a multiplicative factor the moment generating function $\widehat{g}$ of a distribution with density g. The following result is Theorem 6.6 of [14] and was proved there using Laplace's method.

Lemma 2.7 *Let g have the representation $g(u) = \gamma(u)e^{-\psi(u)}$, where γ and ψ have the properties that ψ'' exists, $\psi'' > 0$ and $\sigma = (\psi'')^{-1/2}$ is* self-neglecting, *i.e.*

$$\lim_{u\to\infty} \frac{\sigma(u + x\sigma(u))}{\sigma(u)} = 1 \quad \textit{uniformly on compact } x\textit{–sets}\,, \tag{2.9}$$

and

$$\lim_{u\to\infty} \frac{\gamma(u + x\sigma(u))}{\gamma(u)} = 1 \quad \textit{uniformly on compact } x\textit{–sets}\,. \tag{2.10}$$

Then

$$\int_{-\infty}^{\infty} e^{ku} g(u) du \sim \gamma(\psi'^{\leftarrow}(k))\sigma(\psi'^{\leftarrow}(k))\sqrt{2\pi}e^{\psi^*(k)}\,, \quad k\to\infty\,, \tag{2.11}$$

where $\psi'^{\leftarrow}$ denotes the inverse of ψ' and ψ^* is the convex conjugate of ψ.

3 The heavy-tailed case

3.1 Poisson arrivals

Proposition 3.1 *Assume that $\{N(t)\}$ is Poisson with parameter $\lambda > 0$ and $T > 0$ an independent r.v. with d.f. F satisfying*

$$\lim_{t\to\infty} \frac{\overline{F}(te^{x/\sqrt{t}})}{\overline{F}(t)} = \lim_{t\to\infty} \frac{\overline{F}(t + x\sqrt{t})}{\overline{F}(t)} = 1\,, \quad \textit{locally uniformly in } x \geq 0\,, \tag{3.1}$$

(e.g. $\overline{F}$ is flat for $\sqrt{t}$, see [14]). Then

$$P(N(T) > k) \sim \overline{F}(k/\lambda)\,, \quad k\to\infty\,. \tag{3.2}$$

Before proving this Proposition, we first point out the consequences for any F satisfying Condition (3.1), and give a quick application to the M/G/1 queue.

Lemma 3.2 *(a) If F satisfies (3.1), then F satisfies (2.1), that is, F is heavy-tailed.*
(b) If F satisfies (3.1) then F_e does.

Proof (a) For any $y > 0$ and $x \geq 0$,

$$1 = \lim_{t\to\infty} \frac{\overline{F}(t)}{\overline{F}(te^{x/\sqrt{t}})} \geq \lim_{t\to\infty} \frac{\overline{F}(t)}{\overline{F}(t+x\sqrt{t})} \geq \lim_{t\to\infty} \frac{\overline{F}(t)}{\overline{F}(t+y)} \geq 1\,. \tag{3.3}$$

(b) is an immediate consequence of l'Hospital's rule. □

Weibull-like distributions have tails like $\exp\{-x^\beta\}, 0 < \beta < 1$, and (as the reader can check) if $\beta < 1/2$, then Condition (3.1) holds. If $\beta \geq 1/2$, then Condition (3.1) does not hold. Consequently any distribution with a tail that is heavier than $\exp\{-x^\beta\}$ for some $\beta < 1/2$ will satisfy Condition (3.1), whereas any distribution with a tail that is lighter than $e^{-\sqrt{x}}$ will not satisfy Condition (3.1). In fact, we shall see in Section 4 that if the tail of F is like $e^{-\sqrt{x}}$ or lighter, then the asymptotic (3.2) *does not hold.*

Proposition 3.3 *For a stable* M/G/1 *queue, with service time distribution $G(x) = P(S \leq x)$, if $G_e \in \mathcal{S}$ and satisfies Condition (3.1), then the steady-state queue length L satisfies* (1.4).

Proof From Theorem 2.3 the tail of W is like that of S_e which is assumed to satisfy (3.1); thus so does W and the result follows from DLL (Section 2.4) and Proposition 3.1 with $T = W$. □

Remark 3.4 By Proposition 2.2, and Lemma 3.2, for any d.f. $G \in \mathcal{S}^*$ which satisfies (3.1),the result (1.4) holds. □

Proof of Proposition 3.1. Set in (2.11) $\gamma(u) = \overline{F}(e^u)$, $\psi(u) = \lambda e^u$. We check conditons (2.9) and (2.10): since $\psi^{(\nu)}(u) = \lambda e^u$ for all $\nu \geq 0$, the function $\sigma(u) = 1/\sqrt{\psi''(u)} = 1/\sqrt{\lambda}e^{-u/2}$ has derivative $\sigma'(u) = -1/(2\sqrt{\lambda})\, e^{-u/2} \to 0$, $u \to \infty$, which is sufficient for (2.9), cf. [18], Theorem 2.11.1. Furthermore, by condition (3.1),

$$\lim_{u\to\infty} \frac{\gamma(u + x\sigma(u))}{\gamma(u)} = \lim_{u\to\infty} \frac{\overline{F}(e^{u+(x/\sqrt{\lambda})e^{-u/2}})}{\overline{F}(e^u)} = 1\,, \quad \text{locally uniformly in } x\,.$$

Now let $u = \psi'^{\leftarrow}(k) = \ln(k/\lambda)$ and notice that $u \to \infty$ if and only if $k \to \infty$. Furthermore,

$$\psi^*(k) = ku - \psi(u) = k\ln(k/\lambda) - k$$

is the convex conjugate of ψ.

Then, using Stirling's formula $(k-1)! \sim e^{-(k-1)}(k-1)^{k-\frac{1}{2}}\sqrt{2\pi}$, yields in (2.7)

$$\begin{aligned} P(N(T) \geq k-1) &\sim \lambda^k e^{k-1}(k-1)^{-(k-\frac{1}{2})}\,\overline{F}\left(\frac{k}{\lambda}\right)\frac{1}{\sqrt{\lambda}}\sqrt{\frac{\lambda}{k}}\left(\frac{k}{\lambda}\right)^k e^{-k} \\ &= e^{-1}\left(\frac{k-1}{k}\right)^{-k}\left(\frac{k-1}{k}\right)^{1/2}\overline{F}\left(\frac{k}{\lambda}\right) \\ &\sim \overline{F}\left(\frac{k}{\lambda}\right), \quad k \to \infty\,. \end{aligned}$$

Noting that (2.1) can be rewritten as $\overline{F}(x+y) \sim \overline{F}(x)$ as $x \to \infty$ for $y \geq 0$, the result follows by Lemma 3.2. $\square$

Remark 3.5 If G_e satisfies Condition (3.1) (hence is heavy-tailed) but is not subexponential, then "generalized Little's law", $P(L > k) \sim P(\lambda W > k)$, remains valid. But the $(\rho/(1-\rho))P(\lambda S_e > k)$ asymptotic is no longer valid (recall Theorem 2.3, and Remark 2.4). $\square$

Remark 3.6 Richard Perline drew our attention to a classic asymptotic approximation for a mixed Poisson distribution due to Berg [16], p. 112. Notice that the integral (2.8) can be interpreted as the mixture of a Poisson distribution with mixing density $\overline{F}(t)/\int_0^\infty \overline{F}(y)dy$, $t \geq 0$. Perline [36] uses Berg's result to investigate conditions on a function χ which satisfies $\int_0^\infty e^{-t}t^{s-1}e^{\chi(t)}dt \sim \Gamma(s)e^{\chi(s)}$ as $s \to \infty$. The conditions are formulated in terms of extreme value conditions which are related to (3.1). $\square$

3.2 Generalizing away from the Poisson process

Our objective here is to obtain an analogue to Proposition 3.1 for processes more general than a Poisson process. Our methods differ, however. As for applications to queues, we present at the end of this section (as Proposition 3.13) the GI/GI/1 analogue of Proposition 3.3.

Theorem 3.7 *Let $\{X(t)\}$ be a stochastic process such that $(X(t)-mt)/\sqrt{t} \xrightarrow{\mathcal{D}} N(0,\sigma^2)$ for some $m > 0$ and $\sigma^2 > 0$. Let $T > 0$ be a r.v. independent of $\{X(t)\}$ with d.f. F and assume that for some function $a(t)$ it holds that* (*a*) $(T-t)/a(t)$ *conditional on* $\{T>t\}$ *has a limit* $V\in(0,\infty)$ *in distribution as* $t\to\infty$*;*

(*b*) $a(t)/\sqrt{t}\to\infty$ *as* $t\to\infty$*;*

(*c*) *for all* $c>0$ *it holds that* $P(X(t)>mt+ca(t))=o(P(V>t/m))$*;*

(*d*) *the sample paths of* $\{X(t)\}$ *are increasing or, more generally, there exist* $\delta,\gamma>0$ *such that* $\gamma P(X(s)>x)\le P(X(t)>x)$ *for all* $x>0$ *and all* s,t *with* $0\le s\le t-\delta$*.*

Then

$$P(X(T)>x)\ \sim\ \overline{F}(x/m)\,,\quad x\to\infty\,.$$

If $\{X(t)\}$ is Poisson and F satisfies Condition (3.1), then the conditions in the above Theorem are met.

Remark 3.8 Condition (a) of the above theorem can be rewritten as

$$\lim_{t\to\infty}P\left(\frac{T-t}{a(t)}\le x\,\middle|\,T>t\right)=P(V\le x)\,,\quad x\ge 0\,. \tag{3.4}$$

This is equivalent to T being in the maximum domain of attraction of some extreme value distribution, see e.g. [20], Section 3.4. The function a can be chosen to be absolutely continuous with Lebesgue density a'. Since T has support unbounded to the right, it must be the Fréchet or Gumbel distribution. The limit variable V has generalised Pareto distribution.

If T is in the maximum domain of attraction of the Fréchet distribution with parameter $\alpha>0$, then a can be chosen asymptotically linear with $a'(x)\to 1/\alpha$. As a Cesàro limit, $a(x)/x\to 1/\alpha$, moreover,

$$\lim_{x\to\infty}\frac{a(x+ya(x))}{a(x)}\ =\ 1+\frac{y}{\alpha}\,,\quad \text{locally uniformly in } y\,.$$

Furthermore, V has Pareto distribution.

If T is in the maximum domain of attraction of the Gumbel distribution, then $a'(x)\to 0$ as $x\to\infty$. As a Cesàro limit $a(x)/x\to 0$, moreover,

$$\lim_{x\to\infty}\frac{a(x+ya(x))}{a(x)}\ =\ 1\,,\quad \text{locally uniformly in } y\,.$$

Furthermore, V has exponential distribution. □

Proof We first note that (a) and (3.4) easily yield

$$P(T > x - \epsilon a(x)) \ \sim \ \frac{P(T > x)}{P(V > \epsilon)} .$$

Assume w.l.o.g. that $m = 1$. Write $P(X(T) > x) = f_1(x) + f_2(x) + f_3(x)$ where

$$\begin{aligned} f_1(x) &= P(X(T) > x,\, T < x - \epsilon a(x)), \\ f_2(x) &= P(X(T) > x,\, x - \epsilon a(x) < T < x + \epsilon a(x)), \\ f_3(x) &= P(X(T) > x,\, T > x + \epsilon a(x)) \end{aligned}$$

We start with f_3: For any $\epsilon > 0$ we have

$$f_3(x) \le P(T > x + \epsilon a(x)) \ \sim \ P(T > x)P(V > \epsilon) .$$

Now we use the CLT and local uniformity of the convergence, then for any $b \in \mathbb{R}$, $\eta > 0$ and x sufficiently large,

$$\begin{aligned} f_3(x) &= \int_{x+\epsilon a(x)}^{\infty} P(X(t) > x) P(T \in dt) \\ &\ge \inf_{y \in (x + b\sqrt{x}, x(1+\eta))} P\left(\frac{X(y) - y}{\sqrt{y}} > \frac{x - y}{\sqrt{y}} \right) P(x + \epsilon a(x) < T < x(1+\eta)) \\ &\ge \inf_{y \in (x + b\sqrt{x}, x(1+\eta))} P\left(\frac{X(y) - y}{\sqrt{y}} > \frac{-b\sqrt{x}}{\sqrt{x(1+\eta)}} \right) P(x + \epsilon a(x) < T < x(1+\eta)) \\ &\sim \Phi(b/(\sigma\sqrt{1+\eta}))P(x + \epsilon a(x) < T < x(1+\eta)) \\ &\ge (1 - \delta)P(x + \epsilon a(x) < T < x(1+\eta)) \end{aligned}$$

for arbitrary small δ taking b sufficiently large. Using $a(t) = O(t)$, letting $\eta \to \infty$ and combining with the lower bound above now yields $f_3(x) \sim P(T > x)P(V > \epsilon)$.

Next consider f_2. It follows from (d) that

$$\begin{aligned} f_1(x) &\le P(x - \epsilon a(x) - \delta < T < x - \epsilon a(x)) + \gamma^{-1} P(X(x - \epsilon a(x)) > x) \\ &\le o(P(T > x - \epsilon a(x)) + \gamma^{-1} P(X(x - \epsilon a(x)) > x - \epsilon a(x) + \epsilon a(x - \epsilon a(x))) \\ &= o(P(T > x - \epsilon a(x)) \ = \ o(P(T > x) , \end{aligned}$$

where we used also (c) and (a).

Finally

$$
\begin{aligned}
f_2(x) &\leq P(x - \epsilon a(x) < T < x + \epsilon a(x)) \\
&= P(T > x - \epsilon a(x)) - P(T > x + \epsilon a(x)) \\
&= \left(\frac{1 + o(1)}{P(V > \epsilon)} - P(V > \epsilon)(1 + o(1))\right) P(T > x)\,.
\end{aligned}
$$

Combining the above estimates for f_1, f_2 and f_3 and letting $\epsilon \downarrow 0$ yields the result. □

Theorem 3.9 *The conditions of Theorem 3.7 are satisfied if $\{X(t)\}$ is a Lévy process with $Ee^{sX(1)}$ defined in a neighbourhood of 0, and T is a random variable with a distribution in the maximum domain of attraction of some extreme value distribution with auxiliary function $a(\cdot)$ which is eventually monotone increasing and $a(t)/\sqrt{t} \to \infty$.*

Proof Condition (a) holds by Remark 3.8 and (d) follows easily from the increments being independent and $P(X(t) - X(s) > 0) \to 1$, $|t - s| \to \infty$ (by the CLT). Thus, it only remains to verify (c). In the regularly varying case, (c) follows from the exponential decay of $P(X(t) > (m + \epsilon)t)$ (the Chernoff bound, cf. [7], p. 260), and for the maximum domain of attraction of the Gumbel distribution, we proceed by similar methods as in *loc.cit.*. Define

$$
\kappa(\alpha) = \frac{1}{t} \log Ee^{\alpha X(t)}
$$

(note that $\kappa(\alpha)$ is independent of t by the stationary increments of Lévy processes) and let $\theta = \theta(t) > 0$ be the root of $\kappa'(\theta) = m + ca(t)/t$. Then $a(t)/t \to 0$ implies that $\theta \to 0$ as $t \to \infty$. More precisely,

$$
m + c\frac{a(t)}{t} = \kappa'(\theta) \sim m + \theta\kappa''(0)
$$

so that

$$
\theta \sim c\frac{a(t)}{t\kappa''(0)}, \quad \kappa(\theta) \sim m\theta + c^2\frac{a(t)^2}{2t^2\kappa''(0)}\,.
$$

Let now E_θ refer to the exponential change of measure defined by θ, cf. [7] Ch. XII. Then

$$
\begin{aligned}
P(X(t) > mt + ca(t)) &= E_\theta\left[e^{-\theta X(t) + t\kappa(\theta)}; X(t) > mt + ca(t)\right] \\
&\leq \exp\left\{-\theta(mt + ca(t)) + t\kappa(\theta)\right\} \\
&= \exp\left\{-t\left(m\theta + c\frac{a(t)}{t} - \kappa(\theta)\right)\right\},
\end{aligned}
$$

which can be bounded by $\exp\{-c_1 a(t)\}$. It remains to show that

$$\exp\{-c_1 a(t)\} = o(P(T > t/m)) \,. \tag{3.5}$$

By the representation theorem for a d.f. in the maximum domain of attraction of the Gumbel distribution (see e.g. [20], Section 3.3),

$$\frac{\exp\{-c_1 a(t)\}}{P(T > t/m)} \quad \sim \quad c \exp\left\{-c_1 a(t) + \int_z^{t/m} \frac{1}{a(u)} du\right\} .$$

Now

$$\frac{1}{a(t)} \int_z^{t/m} \frac{1}{a(u)} du \geq \frac{t-z}{a^2(t)} \sim \left(\frac{\sqrt{t/m}}{a(t)}\right)^2 \to 0 \,.$$

This implies (3.5). □

Remark 3.10 The conditions of the above Theorem are in particular satisfied for d.f.'s which are regularly varying, lognormal, or Weibull ($P(T > x) = \exp\{-x^\beta\}$ with $\beta < 1/2$). This is immediate by the respective auxiliary functions $a(x) = \alpha x$, $a(x) = x/\log x$ and $a(x) = x^{1-\beta}$.

□

Theorem 3.11 *The conditions of Theorem 3.7 are satisfied if $\{X(t)\}$ is an additive process on a finite Markov process $\{J_t\}$,*

$$X(t) \quad = \quad \int_0^t r(J_v)\, dv \,,$$

and T satisfies the same conditions as in Theorem 3.9.

Proof The argument closely parallels the proof of Theorem 3.9. The analogue of the Chernoff bound can be found, e.g., in [19], as well as the relevant facts on exponential change of measure which are as follows. Let $M_t(s)$ be the matrix with ijth element $E_i[e^{sX(t)}; J(t) = j]$. Then the Perron–Frobenius root can be written as $e^{t\kappa(s)}$, and the corresponding positive right eigenvector $h = (h_i)$ does not depend on t. The relevant likelihood identity yields

$$P_i(X(t) > mt + ca(t)) \quad = \quad E_{i,\theta} \frac{1}{h_i} \left[e^{-\theta X(t) + t\kappa(\theta)} h_{J(t)}; X(t) > mt + ca(t)\right] .$$

Noting that the components of h are bounded below and above, the rest of the proof is just as for Theorem 3.9. □

For the special case of regularly varying T, see [26] Ch. 8 and [39].

Theorem 3.12 *Assume that $\{X(t)\}$ is a renewal process with arbitrary delay distribution and interarrival d.f. G having mean m^{-1}. Then the conditions of Theorem 3.7 are satisfied provided T is as in Theorem 3.9.*

Proof The CLT for $X(t)$ is standard, so again we only have to verify (c). It suffices to consider the zero–delayed case since $P(X(t) > n)$ is maximized in this case.

Let S_n be a random walk with increment d.f. G and $\kappa(\theta) = \log \int_0^\infty e^{\theta x} G(dx)$. A standard identity from renewal theory states that $\{X(t) > n\} = \{S_n \le t\}$ (see e.g. [37], [38] for an application in a similar context), hence we get

$$P(X(t) > mt + ca(t)) = E_\theta \left[e^{-\theta S_{mt+ca(t)} + t\kappa(\theta)};\, S_{mt+ca(t)} \le t \right].$$

Taking $\theta = \theta(t) < 0$ as the root of $\kappa'(\theta) = (m + ca(t)/t)^{-1}$, the remaining details are just as for Theorem 3.9. □

As an analogue to, and generalization of, Proposition 3.3, we present the following

Proposition 3.13 *For a stable* GI/GI/1 *queue, with interarrival times having finite mean* $1/\lambda$ *and service time d.f. $G(x) = P(S \le x)$, if $G_e \in \mathcal{S}$ and G_e is in the maximum domain of attraction of some extreme value distribution with auxiliary function $a(\cdot)$ which is eventually monotone increasing and $a(t)/\sqrt{t} \to \infty$, then the steady-state queue length L satisfies* (1.4).

Proof Immediate from Theorem 2.3 and Theorem 3.12. □

4 The moderately heavy–tailed case

We now consider d.f.'s F that are still heavy-tailed, but with a tail at least as light as something proportional to the Weibull with $\beta = 1/2$. Note that the Weibull itself with $\beta = 1/2$ is included here which shows that this critical distribution falls into this moderately heavy–tailed case, and yields asymptotics for $N(T)$ that are different from (3.2).

Theorem 4.1 *Assume that $\{N(t)\}$ is a homogeneous Poisson process with parameter $\lambda > 0$ and $T > 0$ is a r.v. (independent of $\{X(t)\}$) with d.f. F such that $\overline{F}(t) = \widetilde{\gamma}(t)\exp\{-t^\beta\}$ for*

$1/2 \le \beta < 1$ *and* $\widetilde{\gamma}$ *is a continuous function. Assume that* $\gamma = \widetilde{\gamma} \circ \exp$ *satisfies*

$$\lim_{u\to\infty} \frac{\gamma(u + xe^{-u/2})}{\gamma(u)} = 1 \quad \textit{locally uniformly in } x\,. \tag{4.1}$$

Then

$$P(N(T) > k) \;\sim \widetilde{\gamma}\left(\frac{k}{\lambda}\right) \exp\left\{-\beta\left(\frac{k}{\lambda}\right)^{\beta} - (1-\beta)t^{\beta}\right\}, \quad k \to \infty\,,$$

where $t = t(k)$ *is the solution to the equation*

$$\beta t^{\beta} + \lambda t = k\,. \tag{4.2}$$

Remark 4.2 Condition (4.1) is a technical one. The important factor for the tail of F is $\exp\{-t^{\beta}\}$, the factor $\widetilde{\gamma}$ allows for flexibility of the model. □

For $\beta = 1/2$ equation (4.2) is a quadratic equation for $\sqrt{t}$ and can be solved explicitly.

Corollary 4.3 (Criticality of Weibull for $\beta = 1/2$)
In the situation of Theorem 4.1 we obtain for $\beta = 1/2$ *and* $\widetilde{\gamma} = 1$,

$$\begin{aligned} P(N(T) > k) \;&\sim\; \exp\left\{\frac{1}{8\lambda}\right\} \exp\left\{-\frac{1}{2}\sqrt{\frac{k}{\lambda}}\left(1 + \sqrt{1 + \frac{1}{16\lambda k}}\right)\right\} \\ &\sim\; \exp\left\{\frac{1}{8\lambda}\right\} \exp\left\{-\sqrt{\frac{k}{\lambda}}\right\}, \quad k \to \infty\,. \end{aligned}$$

Remark 4.4 If $\beta \in [1/2, 2/3)$, it follows by Taylor expansions in (4.2) that

$$\exp\left\{-\beta\left(\frac{k}{\lambda}\right)^{\beta} - (1-\beta)t^{\beta}\right\} \;\sim\; \exp\left\{-\left(\frac{k}{\lambda}\right)^{\beta} + \frac{(1-\beta)\beta^2}{\lambda}\left(\frac{k}{\lambda}\right)^{2\beta-1}\right\}.$$

If $\beta \in [2/3, 3/4)$, we get an added term of order $(k/\lambda)^{3\beta-2}$ and so on. □

Corollary 4.5 *For a stable* M/GI/1 *queue with Weibull service time d.f.* $G(x) = P(S \le x) = 1 - e^{-x^{\beta}}$ *with* $1/2 \le \beta < 1$*, the steady-state queue length* L *satisfies*

$$P(L > k) \;\sim\; \frac{1}{\beta}\left(\frac{k}{\lambda}\right)^{1-\beta} \exp\left\{-\beta\left(\frac{k}{\lambda}\right)^{\beta} - (1-\beta)t^{\beta}\right\}, \quad k \to \infty\,,$$

where $t = t(k)$ *is the solution to the equation* (4.2).

Proof Note that $P(W > x) \sim \overline{G}_e(x) \sim \beta^{-1}x^{1-\beta}e^{-x^{\beta}}$ and take $\tilde{\gamma}(x) = \beta^{-1}x^{1-\beta}$. □

Proof of Theorem 4.1

We start from (2.7), where $g(u) = \gamma(u)e^{-\psi(u)}$ with

$$\psi(u) = e^{\beta u} + \lambda e^u \tag{4.3}$$

and $\gamma(u) = \widetilde{\gamma}(e^u)$. The second derivative is positive, hence ψ is convex and

$$\sigma(u) = (\psi''(u))^{-1/2} = (\beta^2 e^{\beta u} + \lambda e^u)^{-1/2} \sim e^{-u/2}/\sqrt{\lambda}\,, \tag{4.4}$$

and, since $\sigma'(u) \to 0$, the function σ is self-neglecting, i.e. it satisfies (2.9). Furthermore, (4.1) implies (2.10). Hence (2.11) holds and we calculate the rhs. We start with $\psi'^{\leftarrow}(k)$. If we set $t = e^u$, then we need a solution to (4.2). Since $\beta < 1$ we get

$$t = \frac{k}{\lambda}(1 + r(k)) \quad \text{and} \quad r(k) \to 0\,.$$

Inserting this into (4.2) and collecting terms of smaller order yields

$$kr(k) + \beta(k/\lambda)^\beta(1 + \beta r(k)(1 + o(1)) = 0\,.$$

This implies

$$r(k) = -\left(\beta + \frac{\lambda^\beta}{\beta}k^{1-\beta}(1 + o(1))\right)^{-1}, \quad k \to \infty\,.$$

We obtain

$$e^u = t = \frac{k}{\lambda}\left(1 - \left(\beta + \frac{\lambda^\beta}{\beta}k^{1-\beta}(1 + o(1))\right)^{-1}\right)\,.$$

Now we can calculate the terms of (2.11). By continuity,

$$\gamma(\psi'^{\leftarrow}(k)) = \gamma(\ln t) = \widetilde{\gamma}(t) \sim \widetilde{\gamma}\left(\frac{k}{\lambda}\right)\,, \quad \sigma(\psi'^{\leftarrow}(k)) = \sigma(\ln t) \sim \frac{1}{\sqrt{\lambda t}} \sim \frac{1}{\sqrt{k}}\,,$$

and

$$\psi^*(k) = ku - \psi(u) = ku - e^{\beta u} - \lambda e^u\,,$$

where $e^u = t = \psi'^{\leftarrow}(k)$, hence by (4.2),

$$e^{\psi^*(k)} = t^k e^{-t^\beta - \lambda t} = t^k e^{-k-(1-\beta)t^\beta}\,.$$

$$t^k \;=\; \left(\frac{k}{\lambda}\right)^k \left(1 - \frac{\beta/\lambda^\beta}{k^{1-\beta}}(1 + o(1))\right)^{k^{1-\beta}k^\beta} \;\sim\; \left(\frac{k}{\lambda}\right)^k \exp\left\{-\beta\left(\frac{k}{\lambda}\right)^\beta\right\}\,.$$

Combining these results and using Stirling's formula yields the assertion of the theorem. In the last step we have used (2.1) giving $P(N(T) > k) \sim P(N(T) > k - 1)$. $\square$

5 The light–tailed case

The (known) classical light-tailed asymptotics of queue length probabilities is covered by

Proposition 5.1 *Assume that $\{N(t)\}$ is a Poisson process with parameter $\lambda > 0$ and $T > 0$ is independent of $\{N(t)\}$ with*

$$P(T > t) \sim ct^{\alpha-1}e^{-\delta t}, \quad t > 0, \quad c, \alpha, \delta > 0.$$

Then

$$P(N(T) > k) \sim \frac{c}{(\lambda+\delta)^\alpha}\left(\frac{\lambda}{\lambda+\delta}\right)^k k^\alpha, \quad k \to \infty.$$

Proof

$$\begin{aligned} P(N(T) > k) &\sim c\int_0^\infty P(N(t) = k)t^{\alpha-1}e^{-\delta t}\,dt \\ &= \frac{c\lambda^k}{k!}\int_0^\infty t^{k+\alpha-1}e^{-(\lambda+\delta)t}\,dt \\ &= \frac{c}{(\lambda+\delta)^\alpha}\left(\frac{\lambda}{\lambda+\delta}\right)^k \frac{\Gamma(k+\alpha)}{\Gamma(k+1)}, which gives the asymptotic as above. \end{aligned}$$

□

Here is a result on the tail of $N(T)$ covering some more light–tailed T's:

Theorem 5.2 *Assume that the conditions of Theorem 4.1 are satisfied for $\beta > 1$ and $\gamma = \tilde{\gamma}\circ\exp$ satisfies*

$$\lim_{u\to\infty} \frac{\gamma(u + xe^{-\beta u/2})}{\gamma(u)} = 1 \quad \textit{locally uniformly in } x\,. \tag{5.1}$$

Then

$$P(N(T) \geq k-1) \sim \tilde{\gamma}\left(\left(\frac{k}{\beta}\right)^{1/\beta}\right)\frac{\lambda^k\Gamma(k/\beta)}{\Gamma(k+1)}\exp\left\{-\frac{\lambda}{\beta}\left(\frac{k}{\beta}\right)^{1/\beta} - \lambda t\left(1-\frac{1}{\beta}\right)\right\}, k \to \infty,$$

where $t = t(k)$ is the solution to equation (4.2).

Proof The proof is exactly the same as the proof of Theorem 4.1 up to equation (4.4). For $\beta > 1$ we obtain

$$\sigma(u) \sim e^{-\beta u/2}/\beta, \tag{5.2}$$

and since $\sigma'(u) \to 0$, the function σ is self-neglecting. Furthermore, $\gamma = \widetilde{\gamma} \circ \exp$ satisfies condition (2.10) and hence (2.11) holds. For $\beta > 1$ the solution to (2.11) satisfies

$$t = \left(\frac{k}{\beta}\right)^{1/\beta} (1 + r(k)) \quad \text{and} \quad r(k) \to 0\,.$$

Inserting this into (4.2) and collecting terms of smaller order yields

$$\beta k r(k)\,(1 + o(1)) = -\lambda \left(\frac{k}{\beta}\right)^{1/\beta}\,.$$

This implies

$$r(k) = -\lambda/(k^{1-1/\beta}\beta^{1+1/\beta})(1 + o(1))\,.$$

We obtain

$$e^u = t = \left(\frac{k}{\beta}\right)^{1/\beta} \left(1 - \frac{\lambda}{k^{1-1/\beta}\beta^{1+1/\beta}}(1 + o(1))\right)\,.$$

Now we can calculate the terms of (2.11). By continuity,

$$\gamma(\psi'^{\leftarrow}(k)) = \gamma(\ln t) = \widetilde{\gamma}(t) \sim \widetilde{\gamma}\left((k/\beta)^{1/\beta}\right)\,,$$

$$\sigma(\psi'^{\leftarrow}(k)) \sim \frac{1}{\beta}\left(\frac{k}{\beta}\right)^{-1/2} = \frac{1}{\sqrt{\beta k}}\,,$$

and

$$e^{\psi^*(k)} = t^k e^{-t^\beta + \lambda t} = t^k e^{-k + (\beta-1)t^\beta}\,,$$

with t as above. For the first term we obtain:

$$\begin{aligned} t^k &= \left(\frac{k}{\beta}\right)^{k/\beta} \left(1 - \frac{\lambda/\beta^{1+1/\beta}}{k^{1-1/\beta}}(1 + o(1))\right)^{k^{1-1/\beta}k^{1/\beta}} \\ &\sim \left(\frac{k}{\beta}\right)^{k/\beta} \exp\left\{-\frac{\lambda}{\beta}\left(\frac{k}{\beta}\right)^{1/\beta}\right\}\,. \end{aligned}$$

Combining these results yields

$$P(N(T) \geq k-1) \;\sim\; \widetilde{\gamma}\left(\left(\frac{k}{\beta}\right)^{1/\beta}\right) \frac{\lambda^k}{(k-1)!}\sqrt{\frac{2\pi}{\beta k}}\left(\frac{k}{\beta}\right)^{k/\beta} \exp\left\{-\frac{\lambda}{\beta}\left(\frac{k}{\beta}\right)^{1/\beta} - k + (\beta-1)t^\beta\right\}\,.$$

By Stirling's formula, we may set

$$\Gamma\left(\frac{k}{\beta}\right) \sim \sqrt{2\pi}\left(\frac{k}{\beta}\right)^{k/\beta + 1/2} e^{-k/\beta}\,.$$

This gives the asymptotic form as in the assertion. $\square$

Remark 5.3 Somewhat related results in this Poisson/gamma-like regime have been derived by Adell and de la Cal [5]. Starting from the fact that by the SLLN $N_{sT}/s \to T$ a.s. as $s \to \infty$, they investigate the distance (in sup-norm) of the distribution functions of N_{sT}/s and T. □

Remark 5.4 There are of course many ligh–tailed distributions not covered by Proposition 5.1 and Theorem 5.2, for example the Raleigh case $P(T > x) = e^{-x^2/2}$. However, by now it should be clear how to use the analytic method of Lemma 2.7. After the usual lengthy calculations we obtain

$$P(N(T) \geq k-1) \sim e^{\lambda^2/4}\left(\frac{\pi}{2k}\right)^{1/4}\frac{\lambda^k}{\sqrt{(k-1)!}}\exp\left\{-\frac{\lambda}{2}\sqrt{k}\right\}, \quad k \to \infty .$$

□

6 Further queueing applications

6.1 Vacation model

Consider a stable FIFO M/G/1 queue but with *vacations*: Every time the system becomes idle, the server goes away for an amount of time V (having finite first moment). Customers who arrive while the server is away wait in the queue. If the server returns to an empty queue, he goes away yet again and so on. Vacation times $\{V_n\}$ are assumed i.i.d. and independent of all else. Letting D_v denote steady-state delay in this model and D steady-state delay in the regular (non-vacation) model, the following decomposition was established in [22]:

$$D_v = D + V_e, \tag{6.1}$$

where V_e is independent of D and has the equilibrium distribution of V. Adding an independent copy of S yields sojourn time representation

$$W_v = W + V_e, \tag{6.2}$$

where W is regular M/G/1 sojourn time and is independent of V_e.

Conditions (1) and (2) of DLL remain valid for this model yielding (in distribution)

$$L_v = N(W + V_e).$$

A variety of results follows from the general theory of the paper; we present three next as examples, leaving out the moderately heavy–tailed case.

Proposition 6.1 *For the* M/G/1 *vacation model, if the equilibrium service time* $S_e \in \mathcal{S}$ *and if* $P(V_e > x) = o(P(S_e > x))$ *as* $x \to \infty$*, then steady-state queue length* L_v *satisfies* $P(L_v > x) \sim P(L > x)$ *(vacations are negligible asymptotically). If in addition* S_e *satisfies either Condition (3.1) or the conditions on* T *in Theorem 3.9, then*

$$P(L_v > k) \ \sim\ P(\lambda W_v > k) \ \sim\ \frac{\rho}{1-\rho} P(\lambda S_e > k), \quad k \to \infty. \tag{6.3}$$

Proof Since W is subexponential with tail asymptotically proportional to that of S_e (Theorem 2.3 and (2.6)), general subexponential theory yields $P(W + V_e > x) \sim P(W > x)$. It then follows easily from Proposition 3.1 and Theorem 3.9 that $P(N(W + V_e) > x) \sim P(N(W) > x)$, yielding $P(L_v > x) \sim P(L > x)$. Equation (6.3) is then clear from (1.4). □

Proposition 6.2 *For the* M/G/1 *vacation model, if the equilibrium vacation time* $V_e \in \mathcal{S}$ *and if* $P(S_e > x) = o(P(V_e > x))$ *as* $x \to \infty$*, then steady-state queue length* L_v *satisfies* $P(L_v > x) \sim P(N(V_e) > x)$ *(vacations dominate asymptotically). If in addition* V_e *satisfies either Condition (3.1) or the conditions on* T *in Theorem 3.9, then*

$$P(L_v > k) \ \sim\ P(\lambda V_e > k), \quad k \to \infty\,. \tag{6.4}$$

Proof It suffices to show that $P(W + V_e > x) \sim P(V_e > x)$, $x \to \infty$; the rest of the proof is then as for Proposition 6.1. For $\epsilon > 0$, choose x_0 such that $P(V_e > x) \le \epsilon P(S_e > x)$ for $x \ge x_0$ and let the r.v. U have distribution given by $P(U > x) = 1$ for $x < x_0$, $P(U > x) = \epsilon P(S_e > x)$ for $x \ge x_0$. Then U is subexponential and stochastically larger than S_e. Hence by the Pollaczek-Khintchine formula, W is stochastically smaller than $\sum_1^N U_i$ where $U_1, U_2, \dots$ are i.i.d. and distributed as U and N is an independent r.v. with $P(N = k) = (1-\rho)\rho^k$. Therefore $P(W > x) \le (\rho/(1-\rho)P(U > x)$ which yields $\limsup P(W > x)/P(V_e > 0) \le \epsilon$. Letting $\epsilon \downarrow 0$ we get $P(W > x)/P(V_e > 0) \to 0$, which together with the subexponentiality of V_e yields $P(W + V_e > x) \sim P(V_e > x)$. □

Remark 6.3 (a) If $E(e^{sV}) < \infty$ for some $s > 0$ then the second condition of Proposition 6.1 holds.

(b) If $E(e^{sS}) < \infty$ for some $s > 0$ then the second condition of Proposition 6.2 holds. $\square$

Proposition 6.4 *For the* M/G/1 *vacation model, if the equilibrium service time $S_e \in \mathcal{S}$ satisfies either Condition (3.1) or the conditions on T in Theorem 3.9 and if $P(V_e > x) \sim cP(S_e > x)$ for some $c \geq 0$, then steady-state queue length L_v satisfies*

$$P(L_v > k) \;\sim\; P(\lambda W_v > k) \;\sim\; \left(c + \frac{\rho}{1-\rho}\right) P(\lambda S_e > k), \quad k \to \infty. \tag{6.5}$$

Proof By an easy variant of the proof of Proposition 6.1. $\square$

6.2 M/G/1 busy periods: heavy-tailed case

Consider the busy period B of a stable M/G/1 queue with service time S. There are two representations for B that typically are used in its analysis

$$B \;\stackrel{d}{=}\; S + \sum_{i=1}^{N(S)} B_i\,, \tag{6.6}$$

where $B_1, B_2, \ldots$ are i.i.d. with the same distribution as B, and

$$B \;=\; \sum_{i=1}^{K} S_i\,, \tag{6.7}$$

where K denotes the number of customers served during the busy period. K can be identified as the first strictly descending ladder epoch in the negative-drift random walk (starting at 0) with increments distributed as $S - T$ (service time minus interarrival time).

It is known (for $k \geq 1$) that $E(B^k) < \infty$ if and only if $E(S^k) < \infty$, and $E(e^{\epsilon B}) < \infty$ in an ϵ neighborhood of 0 if and only if $E(e^{\epsilon S}) < \infty$ in an ϵ neighborhood of 0 (cf. [41] and Theorem 4.1 in [4]). In particular, B is light-tailed if and only if S is, and the asymptotics in this case are dealt with using transform methods in Sections 7 and 8 in [4]. The asymptotics in the heavy-tailed case are not fully understood, but we present some partial results in what follows.

It was conjectured in [11] that, if the service time S is subexponential, then

$$P(B > x) \sim \frac{1}{1-\rho} P(S > (1-\rho)x)\,, \quad x \to \infty\,. \tag{6.8}$$

This conjecture seems plausible because of the result of [33], where (6.8) was proved for regularly varying S, and of the following heuristics for the general subexponential case. Considering the representation in (6.6), one expects B to be large if either S is large, in which case one expects

$$B \approx S(1 + \lambda EB) = \frac{S}{1-\rho}\,,$$

or if one of the B_i is large, which occurs with probability $EN(S)\,P(B > x) = \rho P(B > x)$. This leads to

$$P(B > x) \sim P\left(\frac{S}{1-\rho} > x\right) + \rho P(B > x)\,,$$

yielding (6.8). However, if S has a lighter tail than $\exp\{-\sqrt{x}\}$ (but is still heavy-tailed), it follows from our previous results that the tail of B is heavier than the rhs of (6.8). Indeed, let $X(t) = t + \sum_{i=1}^{N(t)} B_i$, (compound Poisson process plus linear drift at rate 1) so that $B \stackrel{d}{=} X(S)$. Then $EX(t) = t/(1-\rho)$, $\mathrm{var}X(t) = tw^2$ for some w^2, and we get

$$\begin{aligned}
\liminf_{x\to\infty} \frac{P(B > x)}{P(S > (1-\rho)x)} &\geq \liminf_{x\to\infty} \frac{P(S > (1-\rho)(x-\sqrt{x}))}{P(S > (1-\rho)x)} \inf_{y \geq (1-\rho)(x-\sqrt{x})} P(X(y) > x) \\
&\geq \Phi(c) \liminf_{x\to\infty} \frac{P(S > (1-\rho)(x-\sqrt{x}))}{P(S > (1-\rho)x)} = \infty\,.
\end{aligned}$$

Note, however, that we do not get the tail of B equally precise as in previous sections: In the specified compound Poisson process we need to know apriori what the tail of the B_i is!

The analysis does not exclude that (6.8) could be true if S is subexponential with a heavier tail than $\exp\{-\sqrt{x}\}$, say lognormal or Weibull with $\beta < 1/2$ (the regularly varying case is covered by [33]).

Acknowledgement

We take pleasure in thanking Hanspeter Schmidli, Vincent Dumas and two anonymous referees for their careful reading and valuable remarks on this paper.

References

[1] Abate, J., Choudhury, G.L., and Whitt, W. (1994) Asymptotics for steady-state tail probabilities in structured Markov queueing models. *Stochastic Models* **10**, 99-144.

[2] Abate, J., Choudhury, G.L., and Whitt, W. (1995) Exponential approximations for tail probabilities of queues, I: waiting times. *Opns. Res.*, **43**, 885-901.

[3] Abate, J., Choudhury, G.L., and Whitt, W. (1996) Exponential approximations for tail probabilities of queues, II: sojourn time and workload. *Opns. Res.*, **44**, 758-763.

[4] Abate, J. and Whitt, W. (1997) Asymptotics for M/G/1 low priority waiting-time tail probabilities. *Queueing Systems* **25**, 173-233.

[5] Adell, J.A. and de la Cal, J. (1993) On the uniform convergence of normalized Poisson mixtures to their mixing distribution. *Stat. Probab. Letters* **18**, 227-232.

[6] Asmussen, S. (1981) Equilibrium properties of the M/G/l queue. *Z. Wahrscheinlichkeitsth. verw. Geb.* **58**, 267-281.

[7] Asmussen, S. (1987). *Applied Probability and Queues.* John Wiley & Sons, Chichester New York.

[8] Asmussen, S., Fløe Henriksen, L. & Klüppelberg, C. (1994) Large claims approximations for risk processes in a Markovian environment. *Stoch. Proc. Appl.* **54**, 29–43.

[9] Asmussen, S. & Højgaard, B. (1996) Ruin probability approximations for Markov-modulated risk processes with heavy tails. *Th. Random Proc.* **2**, 96–107.

[10] Asmussen, S., Schmidli, H. & Schmidt, V. (1999) Tail probabilities for non-standard risk and queueing processes with subexponential jumps. *Adv. Appl. Probab.* **31** (to appear).

[11] Asmussen, S. & Teugels, J.L. (1996) Convergence rates for M/G/1 queues and ruin problems with heavy tails. *J. Appl. Probab.* **33**, 1181–1190.

[12] Asmussen, S. & Thorisson, H. (1988) Large deviation results for time-dependent queue length probabilities. *Stochastic Models* **4**, 99-116.

[13] von Bahr, B. (1975) Asymptotic ruin probabilities when exponential moments do not exist. *Scand. Act. J.* **1975**, 6–10.

[14] Balkema, A.A., Klüppelberg, C. & Resnick, S.I. (1993) Densities with Gaussian tails. *Proc. London Math. Soc.* **66**, 568-588.

[15] Balkema, A.A., Klüppelberg, C. & Stadtmüller, U. (1995) Tauberian results for densities with Gaussian tails. *J. London Math. Soc.* **51**, 383-400.

[16] Berg, L. (1968) *Asymptotische Darstellungen und Entwicklungen.* VEB Deutscher Verlag der Wissenschaften, Berlin.

[17] Bertsimas, D. & Mourtzinou, G. (1997) Transient laws of non-stationary systems. *Queueing Systems* **25**, 115-155.

[18] Bingham, N.H., Goldie, C.M. & Teugels, J.L. (1987) *Regular Variation.* Oxford University Press.

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

[20] Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Heidelberg.

[21] Embrechts, P. & 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.

[22] Fuhrmann, S.W.& Cooper, R.B. (1985) Stochastic decomposition in the M/G/1 queue with general vacations. *Opns. Res.* **32**, 1368-1373.

[23] Gaver, D.P., Jr. (1959) Imbedded Markov chain analysis of a waiting line process in continuous time. *Ann. Math. Statist.* **30**, 698-720.

[24] Glynn, P.W. & Whitt, W. (1994) Logarithmic asymptotics for steady–state tail probabilities in a single–server queue. *Adv. Appl. Probab.*, Takacs issue, 131–156.

[25] Goldie, C.M. & Klüppelberg, C. (1998) Subexponential distributions. In: Adler, R., Feldman, R. and Taqqu, M.S. (Eds) *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions.*, pp. 435-459. Birkhäuser.

[26] Grandell, J. (1997) *Mixed Poisson Processes.* Chapman & Hall, London.

[27] Haji, R., & G.F. Newell (1971) A relation between stationary queue and waiting time distribution, *J. Appl. Prob.*, **8**, 617-620.

[28] Jelenkovic, P.R. & Lazar, A.A. (1998) Subexponential asymptotics of a Markov–modulated random walk with queueing applications. *J. Appl. Prob.* **35**(2) (to appear).

[29] Jensen, J.L. (1995). *Saddlepoint Approximations.* Clarendon Press, Oxford.

[30] Keilson, J. & Servi, L. (1990) The distributional form of Little's law and the Fuhrmann-Cooper decomposition. *Op. Res. Lett.* **9**, 239-247.

[31] Klüppelberg, C. (1988) Subexponential distributions and integrated tails. *J. Appl. Probab.* **25**, 132-141.

[32] Le Gall, P. (1962). *Les Systemes avec ou sans Attente et les Processus Stochastiques.* Dunod, Paris.

[33] de Meyer, A. & Teugels, J.L.T. (1980) On the asymptotic behaviour of the distributions of the busy period and service time in M/G/1. *J. Appl. Probab.* **17**, 802–813.

[34] Neuts, M.F. (1986) The caudal characteristic curve of queues. *Adv. Appl. Probab.* **18**, 221–254.

[35] Pakes, A. (1975) On the tails of waiting time distributions. *J. Appl. Probab.* **12**, 555-564.

[36] Perline, R. (1998) Mixed Poisson distributions tail equivalent to their mixing distributions. *Stat. Probab. Letters* **38**, 229-233.

[37] Puhalskii, A. (1995) Large deviations analysis of the single server queue. *Queueing Systems* **21**, 5–66.

[38] Puhalskii, A. & Whitt, W. (1997) Functional large deviations principles for first–passage–time processes. *Ann. Appl. Probab.* **7**, 362-381.

[39] Rolski, T., Schlegel, S. & Schmidt, V. (1999) Asymptotics of Palm-stationary buffer content distributions in fluid flow queues. *Adv. Appl. Probab.* **31** (to appear).

[40] Schmidli, H. (1997) Compound sums and subexponentiality. *Bernoulli* (to appear).

[41] Shahabbudin, P. & Sigman, K. (1996) Finite exponential moments for branching processes. Manuscript, Columbia University.