Queueing Systems (2022) 100:61–91
https://doi.org/10.1007/s11134-021-09732-9

On a single server queue fed by scheduled traffic with Pareto perturbations

Victor F. Araman[1] · Hong Chen[2] · Peter W. Glynn[3] · Li Xia[4]

Received: 31 January 2021 / Revised: 21 December 2021 / Accepted: 23 December 2021 /
Published online: 20 January 2022

Abstract
A“scheduled” arrival process is one in which the nth arrival is scheduled for time n, but instead occurs at $n+\xi_n$, where the ξ_j ’s are i.i.d. We describe here the behavior of a single server queue fed by such traffic in which the processing times are deterministic. A particular focus is on perturbations with Pareto-like tails but with finite mean. We obtain tail approximations for the steady-state workload in both cases where the queue is critically loaded and under a heavy-traffic regime. A key to our approach is our analysis of the tail behavior of a sum of independent Bernoulli random variables with parameters of the form $p_n \sim c\, n^{-\alpha}$ as $n \to \infty$, for $c > 0$ and $\alpha > 1$.

Keywords Scheduled traffic · Heavy-tailed distribution · Limit theorems · Bernoulli sums · Heavy traffic · Tail asymptotics

Mathematics Subject Classification Primary 60F05 · 60F10 · 60K25 · Secondary 60G07

✉ Victor F. Araman
va03@aub.edu.lb

Hong Chen
hchen@saif.sjtu.edu.cn

Peter W. Glynn
glynn@stanford.edu

Li Xia
xiali5@sysu.edu.cn

1 Olayan School of Business, American University of Beirut, Beirut, Lebanon

2 Shanghai Advanced Institute of Finance, Shanghai Jiaotong University, Shanghai 200030, China

3 Management Science and Engineering Department, Stanford University, Stanford, CA 74305, USA

4 Business School, Sun Yat-Sen University, Guangzhou 510275, China

1 Introduction

In conventional queueing models, it is frequently assumed that the exogenous arrivals to the system are described by a renewal (counting) process. Specifically, the sequence $\chi = (\chi_n : n \geq 1)$ of inter-arrival times of successive customers is assumed to be a sequence of independent and identically distributed (i.i.d.) non-negative random variables (rv's). More complex (arrival) traffic models can be obtained by assuming that the χ_j's are Markov-dependent or form a stationary time series.

While such traffic models are frequently appropriate, there are some modeling settings in which one may seek alternatives. One such setting is that in which arrivals are scheduled in advance; for example, an outpatient clinic. Patients are typically scheduled to arrive at regular fifteen or twenty minute intervals. Of course, some patients arrive early for their appointments, and others arrive late, so that there is some random variation present. A natural traffic model to adopt here is to assume that the nth patient is scheduled to arrive at the clinic at time nh, but actually arrives at time $nh + \xi_n$, where $\xi = (\xi_n : n \geq 0)$ is a stationary sequence of rv's. We call such an arrival process a "scheduled traffic model", and we refer to the ξ_n's as the (random) *perturbations* about the schedule. For the rest of the paper and without loss of generality, we assume $h = 1$.

Such scheduled traffic occurs naturally in other setups as well. Consider a retailer that places orders for an item at the end of each week, where the number of orders placed is a time-stationary sequence. If the supplier then fulfills each of these orders with i.i.d. random delays, then the times at which the retailer's inventory is replenished with additional units of that item forms a scheduled traffic process.

To the best of our knowledge, such traffic seems to have been first analyzed by Winsten [20], in the context of a single server queue. He restricted his attention to the case of exponential service times and bounded perturbations, and argued that the number-in-system process has an equilibrium distribution that is conditional geometric (whereby, it is geometric for values of the queue larger than the bound of the perturbation). Mercer [16,17] generalized the results of Winsten [20] by analyzing various variants of such queueing system, including bulk arrivals and more general service time distributions. Loynes [14] obtained a probabilistic characterization, akin to our Eq. (4.1) for the workload, for the equilibrium waiting time of a single server queue for arrival processes that includes our scheduled arrival process. (However, this characterization does not lend itself to direct quantitative computation, and this current paper is a partial effort to address this via a study of related limit theorems.)

Doi et al. [7] considered a modified version of our scheduled model, by dropping any arrival that occurs beyond the next customer's scheduled time. Another approach to modeling appointment scheduling has been considered by Hassin and Mendel [10], who took the view that either customers arrive on time or are considered as "no-shows." Zacharias and Armony [21] modeled the process of taking an appointment, as well as the resulting in-clinic queueing that follows, by assuming that only a fraction of those scheduled will show up. However, those showing up are assumed to follow a renewal-like process. Kemper et al. [12] suggested a procedure for scheduling appointments when the i.i.d. perturbations are assumed to be substantially smaller than a typical job duration.

We also note the work of Luo et al. [15] who studied an appointment-based service system modeled by two queues in tandem, with a possibility of no-show. These assumptions lead to an arrival process in which arrivals occur in the same order than the initial appointments, unlike our scheduled arrival process. Recently, Honnappa et al. [11] considered so-called transitory queues, and introduced an arrival process in which the customers scheduled for a given day have arrival times that are i.i.d., randomly scattered over the day. This is an arrival process that, while non-renewal, is quite different from the scheduled traffic that is our main focus.

The current paper derives from the observation that scheduled traffic should generate a much more regular arrival stream than does a renewal or Markov-dependent arrival process. In particular, one expects that such queues will often have qualitative structure similar to that of a queue fed by deterministic inter-arrivals. Indeed, adopting here Kendall's notation for queues, and denoting scheduled traffic by S, Kingman [13] obtained a general heavy-traffic result for single server queues that shows that the heavy traffic limit theorem for the equilibrium distribution of an $S/G/1$ queue (with G having positive and finite variance) is identical to that of the corresponding $D/G/1$ queue. (However, we have previously shown that heavy traffic limit theory looks quite different when the perturbations have infinite mean; see Araman and Glynn [1]. In particular, the limit then involves fractional Brownian motion with a Hurst parameter $H < 1/2$, rather than Brownian motion.) Chen and Zhao [3] considered an application of the $S/D/1$ queue to the air traffic control space in the vicinity of an airport, and showed that the $S/D/1$ queue is frequently stable even when the utilization, ρ, equals one (i.e., under critical loading), unlike the renewal arrival version of the same model. Moreover, in a forthcoming work, we show that the tail asymptotic for the equilibrium waiting time in an $S/G/1$ queue mirrors that of a $D/G/1$ queue when the service times are non-degenerate and light-tailed, in the sense that the exponential decay rates are identical.

Of course, in practice, we expect that a queue fed by scheduled traffic has behavior different from that of the associated queue with deterministic arrivals. In view of these considerations, it seems mathematically natural to explore a setting in which we can study the differences between a queue fed by scheduled traffic and a queue fed by deterministic traffic. As discussed above, the presence of truly random service times in a single server system creates a mathematical environment in which the service time variability far dominates the relative difference in queueing behavior between a scheduled and deterministic arrival process. However, as we shall see in this paper, the $S/D/1$ queue having deterministic service times exhibits effects that are quite differentiated from those associated with both the $D/D/1$ queue and the $G/D/1$ queue. In other words, we see qualitatively different queueing behavior (from $D/D/1$ and $G/D/1$) for scheduled traffic in this $S/D/1$ setting.

Specifically, we show in this paper that:

(1) when the arrival rate equals the service rate, and the perturbations are Pareto distributed, the workload process $W(t)$ of an S/D/1 queue grows without bound, specifically at a rate of $\log t / \log \log t$, unlike the bounded workload associated with a $D/D/1$ queue or the square root rate associated with a $G/D/1$ queue fed by renewal arrivals; see Theorem 5.

(2) when the arrival rate is less than the service rate, and the perturbations are Pareto distributed, we derive a heavy-traffic limit theorem for the equilibrium distribution of the $S/D/1$ workload scales process that establishes that the equilibrium workload scales as $\log(1/(1-\rho))/\log\log(1/(1-\rho))$ in the utilization ρ, in contrast to the $1/(1-\rho)$ scaling associated with a $G/D/1$ queue in heavy traffic; see Theorem 6.

This paper also includes several new results on the class of point processes that are generated by our scheduled arrival model. In particular, when the perturbations are i.i.d., we establish:

(3) asymptotics for the covariance structure of the counting process $N(t)$ that counts the number of arrivals in $[0, t]$, when the tails of the perturbations display both power law decay and exponential decay; see Proposition 1.
(4) stochastic boundedness in t for the difference between the number of arrivals $N(t)$ and the number scheduled for $[0, t]$ when the perturbations have finite mean, in contrast to the square root stochastic fluctuations typically associated with renewal traffic; see Theorem 1.
(5) the tail of the difference between $N(t)$ and the number scheduled for $[0, t]$ when the perturbations have Pareto tails has a Poisson-type decay; see Theorem 3 and Proposition 2.

Along the way to proving these results, we also develop new exact and logarithmic tail asymptotics for sums of independent Bernoulli random variables with probabilities of the form $p_n \sim c\,n^{-\alpha}$ as $n \to \infty$, for $c > 0$ and $\alpha > 1$.

This paper is organized as follows: In the next section, we present some properties of scheduled traffic. In Sect. 3 we study sums of independent Bernoulli random variables, thereby allowing us to infer the tail of asymptotics of $N(t)$. We next analyze the behavior of a single server queue when fed by a scheduled traffic. Specifically, in Sects. 4 and 5, we investigate the $S/D/1$ queue, and obtain limiting results for the workload, both when the queue is critically loaded and under a heavy traffic regime.

2 Properties of scheduled traffic

Let $(\xi_j : j \in \mathbb{Z})$ be an i.i.d. sequence of perturbations. We note that independence of the ξ_j's seems plausible in many settings, given that perturbation j is typically determined by decisions or preferences that are idiosyncratic to consumer j. Given the ξ_j's, for any measurable subset of the real line, A, we define the random measure $\widetilde{N}$ via

$$\widetilde{N}(A) = \sum_j I(j + \xi_j + U \in A),$$

where U is a uniform r.v. on $[0, 1]$ independent of the ξ_j's. It is easily argued that $\widetilde{N}$ is time-stationary, in the sense that $\widetilde{N}(\cdot + t) \stackrel{D}{=} \widetilde{N}(\cdot)$ for $t \in \mathbb{R}$ (where $A + t \stackrel{\Delta}{=} \{x + t : x \in A\}$, $\stackrel{D}{=}$ denotes equality in distribution, and $\stackrel{\Delta}{=}$ denotes equality by definition.) We further define the counting process $N = \big(N(t) : t \geq 0\big)$ via $N(t) = \widetilde{N}\big((0, t]\big)$; $N(t)$ counts the cumulative number of arrivals to the system in $(0, t]$. Our focus, in this section, is on the scheduled arrival process N. We introduce the uniform r.v. U in order to obtain a version of $\tilde{N}$ that has stationary increments. (Without U, $\tilde{N}(\cdot + n) \stackrel{D}{=} \tilde{N}(\cdot)$ only for $n \in \mathbb{Z}$.) One can interpret U as random time origin chosen independently of the ξ_j's.

We start by noting that, regardless of whether ξ_0 has infinite mean or not, N is a unit intensity counting process. Specifically,

$$\begin{aligned} \mathbb{E}N(t) &= \sum_j \int_0^1 \mathbb{P}\big(j + x + \xi_j \in (0, t]\big)\mathrm{d}x \\ &= \sum_j \int_0^1 \mathbb{P}\big(j + x + \xi_0 \in (0, t]\big)\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \mathbb{P}\big(r + \xi_0 \in (0, t]\big)\mathrm{d}r \\ &= \mathbb{E} \int_{-\infty}^{\infty} I\big(r \in (-\xi_0, t - \xi_0]\big)\mathrm{d}r \\ &= t. \end{aligned}$$

In fact, regardless of the tails of the ξ_j's, the counting process N has light tails. In particular, the moment generating function of $N(t)$ is always finite-valued. Specifically, the independence of the ξ_j's ensures that, for any θ,

$$\begin{aligned} \log\left(\mathbb{E}\exp(\theta\, N(t))\right) &= \sum_j \log\left(\int_0^1 \mathbb{E}\exp\Big(\theta\, I\big(j + x + \xi_j \in (0, t]\big)\Big)\mathrm{d}x\right) \\ &= \sum_j \log\left(\int_j^{j+1} \mathbb{E}\exp\Big(\theta\, I\big(r + \xi_0 \in (0, t]\big)\Big)\mathrm{d}r\right) \\ &= \sum_j \log\left(1 + (e^\theta - 1)\int_j^{j+1} \mathbb{P}(r + \xi_0 \in (0, t])\mathrm{d}r\right) \\ &\le (e^\theta - 1)\sum_j \int_j^{j+1} \mathbb{P}(r + \xi_0 \in (0, t])\mathrm{d}r = (e^\theta - 1)\, t. \end{aligned}$$

In order to obtain insight into the dependence structure of N, we next study its covariance properties. Set $\Delta N(t) = N(t) - N(t-1)$ for $t \ge 1$, and recall that

$$\begin{aligned} &\mathbb{C}\mathrm{ov}(\Delta N(1), \Delta N(t)) \\ &\quad = \mathbb{E}\,\mathbb{C}\mathrm{ov}\big((\Delta N(1), \Delta N(t))|U\big) + \mathbb{C}\mathrm{ov}\big(\mathbb{E}(\Delta N(1)|U), \mathbb{E}(\Delta N(t)|U)\big); \end{aligned} \tag{2.1}$$

see p. 381 of Ross [19]. Noting that

$$\Delta N(t) = \sum_j I\big(j + \xi_j + U \in (t-1, t]\big)$$

$$
\begin{aligned}
&= \sum_j I\big(j - \lfloor t \rfloor + \xi_j + U \in (t - \lfloor t \rfloor - 1, t - \lfloor t \rfloor]\big) \\
&\stackrel{D}{=} \sum_k I\big(k + \xi_k + U \in (t - \lfloor t \rfloor - 1, t - \lfloor t \rfloor]\big),
\end{aligned}
$$

it is evident that $\mathbb{E}(\Delta N(t)|U)$ depends on t only through $t - \lfloor t \rfloor$. The second term in (2.1) does not decay to zero as $t \to \infty$ and it reflects the correlation due to the common random placement of the time origin associated with U. The more informative term on the right-hand side of (2.1) is $\mathbb{C}\text{ov}\big((\Delta N(1), \Delta N(t))|U\big)$. Note that, for $t \geq 2$,

$$
\begin{aligned}
&\mathbb{C}\text{ov}\big((\Delta N(1), \Delta N(t))|U\big) \\
&= \sum_{i,j} \mathbb{P}(i + \xi_i + U \in (0, 1], j + \xi_j + U \in (t - 1, t]|U) \\
&\quad - \sum_{i,j} \mathbb{P}(i + \xi_i + U \in (0, 1]|U)\, \mathbb{P}(j + \xi_j + U \in (t - 1, t]|U) \\
&= \sum_{i \neq j} \mathbb{P}(i + \xi_i + U \in (0, 1], j + \xi_j + U \in (t - 1, t]|U) \\
&\quad - \sum_{i,j} \mathbb{P}(i + \xi_i + U \in (0, 1]|U)\, \mathbb{P}(j + \xi_j + U \in (t - 1, t]|U) \\
&= -\sum_{i} \mathbb{P}(i + \xi_0 + U \in (0, 1]|U)\, \mathbb{P}(i + \xi_0 + U \in (t - 1, t]|U),
\end{aligned}
\tag{2.2}
$$

so the conditional covariance is always non-positive. This is intuitively reasonable, since scheduled traffic has the characteristic that if an abnormally large number of customers arrive in an interval, this reduces the number available to arrive in a subsequent interval. We can now use (2.2) to develop asymptotics for the conditional covariance.

Proposition 1 *(i.) Suppose that ξ_0 has a bounded density f for which there exist positive constants $c_1, c_2, \alpha_1, \alpha_2$ such that*

$$
\begin{aligned}
f(x) &\sim c_1\, x^{-\alpha_1 - 1}, \\
f(-x) &\sim c_2\, x^{-\alpha_2 - 1}
\end{aligned}
$$

as $x \to \infty$. If $\alpha_1 < \alpha_2$, then

$$\mathbb{C}ov\big((\Delta N(1), \Delta N(n))|U\big) \sim -c_1\, n^{-\alpha_1 - 1}$$

as $n \to \infty$, whereas if $\alpha_2 < \alpha_1$, then

$$\mathbb{C}ov\big((\Delta N(1), \Delta N(n))|U\big) \sim -c_2\, n^{-\alpha_2 - 1}$$

as $n \to \infty$. If $\alpha_1 = \alpha_2$, then

$$\mathbb{C}ov\big((\Delta N(1), \Delta N(n))|U\big) \sim -(c_1 + c_2)\, n^{-\alpha_1 - 1}$$

as $n \to \infty$.

(ii.) *Suppose that* ξ_0 *has a bounded density* f *for which there exist positive constants* $d_1, d_2, \beta_1, \beta_2$ *such that*

$$f(x) \sim d_1\, e^{-\beta_1 x},$$
$$f(-x) \sim d_2\, e^{-\beta_2 x}$$

as $x \to \infty$. *If* $\beta_1 < \beta_2$, *then*

$$\mathbb{C}ov\big((\Delta N(1), \Delta N(n))|U\big) \sim -e^{-\beta_1 n}\, \frac{d_1}{\beta_1}\, (e^{\beta_1} - 1) \times \sum_j e^{-\beta_1 (j-U)}\, \mathbb{P}(\xi_0 + U \in (j-1, j]|U)$$

as $n \to \infty$, *whereas if* $\beta_2 < \beta_1$, *then*

$$\mathbb{C}ov\big((\Delta N(1), \Delta N(n))|U\big) \sim -e^{-\beta_2 n}\, \frac{d_2}{\beta_2}\, (1 - e^{-\beta_2}) \times \sum_j e^{-\beta_2 (j+U)}\, \mathbb{P}(\xi_0 + U \in (j-1, j]|U)$$

as $n \to \infty$. *If* $\beta_1 = \beta_2$, *then*

$$\mathbb{C}ov\big((\Delta N(1), \Delta N(n))|U\big) \sim -n\, e^{-\beta_1 (n+1)}\, \frac{d_1\, d_2}{\beta_1^2}\, (1 - e^{-\beta_1})^2$$

as $n \to \infty$.

Proof According to (2.2), the conditional covariance is given by

$$\begin{aligned} -\sum_j &\mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \\ &= -\sum_{j>-n/2} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \\ &- \sum_{k \le n/2} \mathbb{P}(\xi_0 + U \in (k-n-1, k-n]|U)\, \mathbb{P}(\xi_0 + U \in (k-1, k]|U). \end{aligned}$$

Given our bounded density assumption, the (conditional) Bounded Convergence Theorem implies that

$$n^{\alpha_1+1}\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \to c_1$$

as $n \to \infty$, and $\big(n^{\alpha_1+1}\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]) : j > -n/2\big|U\big)$ is uniformly bounded. Another application of the (conditional) Bounded Convergence Theorem therefore implies that

$$n^{\alpha_1+1} \sum_{j>-n/2} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \to c_1$$

as $n \to \infty$. Similarly,

$$n^{\alpha_2+1} \sum_{k\leq n/2} \mathbb{P}(\xi_0 + U \in (k-n-1, k-n]|U)\, \mathbb{P}(\xi_0 + U \in (k-1, k]|U) \to c_2$$

as $n \to \infty$, proving part i.).

For part ii.), suppose first that $\beta_1 < \beta_2$ and note that

$$\sum_j e^{-\beta_1 j}\, \mathbb{P}(\xi_0 + U \in (j-1, j]|U) < \infty.$$

Furthermore, our assumption on f guarantees that

$$e^{\beta_1 n}\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \to \frac{d_1}{\beta_1}\,(e^{\beta_1} - 1)\, e^{-\beta_1 (j-U)}$$

as $n \to \infty$, and $\left(e^{\beta_1 j}\, \mathbb{P}(\xi_0 + U \in (j-1, j]|U) : j \geq 0\right)$ is uniformly bounded. Applying the Bounded Convergence Theorem, we conclude that

$$\begin{aligned}
&e^{\beta_1 n} \sum_{j>-n} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \\
&= \sum_{j>-n} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, e^{-\beta_1 j} \cdot e^{\beta_1 (n+j)}\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \\
&\to \frac{d_1}{\beta_1}\,(e^{\beta_1} - 1)\, e^{\beta_1 U} \sum_j e^{-\beta_1 j}\, \mathbb{P}(\xi_0 + U \in (j-1, j]|U)
\end{aligned}$$

as $n \to \infty$. Similarly,

$$\begin{aligned}
&e^{\beta_2 n} \sum_{k<0} \mathbb{P}(\xi_0 + U \in (k-n-1, k-n]|U)\, \mathbb{P}(\xi_0 + U \in (k-1, k]|U) \\
&\quad \to \frac{d_2}{\beta_2}\,(1 - e^{-\beta_2})\, e^{-\beta_2 U} \sum_{k<0} \mathbb{P}(\xi_0 + U \in (k-1, k]|U)\, e^{\beta_2 k}
\end{aligned} \tag{2.3}$$

as $n \to \infty$, thereby establishing that the conditional covariance satisfies

$$\begin{aligned}
&e^{\beta_1 n} \sum_j \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \\
&\quad \to \frac{d_1}{\beta_1}\,(e^{\beta_1} - 1) \sum_j e^{-\beta_1 (j-U)}\, \mathbb{P}(\xi_0 + U \in (j-1, j]|U)
\end{aligned}$$

as $n \to \infty$. The case where $\beta_2 < \beta_1$ can be handled identically.

To handle the case where $\beta_1 = \beta_2$, we write the conditional covariance as

$$
\begin{aligned}
&-\sum_{j\geq 0} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \\
&-\sum_{-n\leq j<0} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \\
&-\sum_{k<0} \mathbb{P}(\xi_0 + U \in (k-n-1, k-n]|U)\, \mathbb{P}(\xi_0 + U \in (k-1, k]|U).
\end{aligned}
$$

Relation (2.3) shows that the third term is of order $O(e^{-\beta_2 n})$ as $n \to \infty$; a similar argument proves that the first term is of order $O(e^{-\beta_1 n})$. To handle the second term, we write it as

$$
\begin{aligned}
&-\sum_{-n\leq j<-n/2} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \\
&\quad -\sum_{-n/2\leq j<0} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U) \\
&= -\sum_{0\leq k<n/2} \mathbb{P}(\xi_0 + U \in (k-n-1, k-n]|U)\, \mathbb{P}(\xi_0 + U \in (k-1, k]|U) \\
&\quad -\sum_{-n/2\leq j<0} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \mathbb{P}(\xi_0 + U \in (n+j-1, n+j]|U).
\end{aligned}
\tag{2.4}
$$

But the second term above equals

$$
-\sum_{-n/2\leq j<0} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \frac{d_1}{\beta_1}\, (e^{\beta_1} - 1)\, e^{-\beta_1 (n+j-U)} \big(1 + o(1)\big),
$$

as $n \to \infty$, where the $o(1)$ term is uniform in $-n/2 \leq j < 0$. So this sum equals

$$
-\big(1 + o(1)\big)\, e^{-\beta_1 n} \sum_{-n/2\leq j<0} \mathbb{P}(\xi_0 + U \in (j-1, j]|U)\, \frac{d_1}{\beta_1}\, (e^{\beta_1} - 1)\, e^{-\beta_1 (j-U)}.
$$

Since $\beta_1 = \beta_2$, $\mathbb{P}\big(\xi_0 + U \in (j-1, j]|U\big)\, e^{\beta_1 j} \to \frac{d_2}{\beta_1}\, (1 - e^{-\beta_1})\, e^{-\beta_1 U}$ as $j \to -\infty$. Consequently, the second term is asymptotic to $-e^{-\beta_1 (n+1)}(n/2)\, \frac{d_1 d_2}{\beta_1^2}\, (1 - e^{-\beta_1})^2$ as $n \to \infty$. A similar analysis works for the first term in (2.4), proving part ii.) for $\beta_1 = \beta_2$. □

Proposition 1 *i.)* asserts that the conditional autocorrelations are always summable, regardless of the values of $\alpha_1, \alpha_2 > 0$. In particular, this occurs even when the perturbations have infinite mean. This case is discussed in Araman and Glynn [1]. It is shown there that $N(\cdot)$ then satisfies a functional limit theorem with normalization $n^{(1-\min\{\alpha_1,\alpha_2\})/2}$ and with fractional Brownian motion with $H < 1/2$ as a limit.

We turn next to a key representation for N that holds only when $\mathbb{E}|\xi_0| < \infty$. In preparation for stating this result, let

$$\mathcal{E}(t) = \sum_{i+U>t} I(i + \xi_i + U \le t)$$
$$\mathcal{L}(t) = \sum_{i+U\le t} I(i + \xi_i + U > t).$$

The r.v. $\mathcal{E}(t)$ represents the total number of early customers at time t, who have arrived earlier than scheduled, while $\mathcal{L}(t)$ is the total number of late customers that will arrive after t but were scheduled to arrive before t. The Borel-Cantelli lemma makes clear that $\mathcal{E}(t)$ is finite-valued a.s. if and only if $\mathbb{E}\xi_0^- \triangleq \mathbb{E}\max(-\xi_0, 0) < \infty$, while $\mathcal{L}(t)$ is finite-valued a.s. if and only if $\mathbb{E}\xi_0^+ \triangleq \mathbb{E}\max(\xi_0, 0) < \infty$. Furthermore, $\big((\mathcal{E}(t), \mathcal{L}(t)) : t \in \mathbb{R}\big)$ is a time-stationary process, where for every t, $\mathcal{E}(t)$ and $\mathcal{L}(t)$ are independent random variables.

Proposition 2 *Suppose that $\mathbb{E}|\xi_0| < \infty$. Then, for $t \ge 0$,*

$$N(t) - t = \left(\sum_{i+U\in(0,t]} 1\right) - t + \big(\mathcal{E}(t) - \mathcal{L}(t)\big) - \big(\mathcal{E}(0) - \mathcal{L}(0)\big). \tag{2.5}$$

Proof Observe that

$$\begin{aligned} N(t) - t &= \sum_{i+U\in(0,t]} I(i + \xi_i + U \in (0,t]) + \sum_{i+U>t} I(i + \xi_i + U \in (0,t]) \\ &\quad + \sum_{i+U\le 0} I(i + \xi_i + U \in (0,t]) - t \\ &= \sum_{i+U\in(0,t]} \big(1 - I(i + \xi_i + U > t) - I(i + \xi_i + U \le 0)\big) - t \\ &\quad + \sum_{i+U>t} \big(I(i + \xi_i + U \le t) - I(i + \xi_i + U \le 0)\big) \\ &\quad + \sum_{i+U\le 0} \big(I(i + \xi_i + U > 0) - I(i + \xi_i + U > t)\big). \end{aligned} \tag{2.6}$$

We now combine the first indicator sum with the sixth (to obtain $-\mathcal{L}(t)$), and the second indicator sum with the fourth (to obtain $-\mathcal{E}(0)$), thereby proving the result. □

The previous result is quite intuitive. Indeed, the term $\sum_{i+U\in(0,t]} I(i + \xi_i + U \in (0,t])$ is equal to the number initially scheduled in $[0, t)$, while $\mathcal{E}(t) - \mathcal{E}(0)$ counts the number of arrivals that showed up in $(0, t]$ but were not scheduled to, while $\mathcal{L}(t) - \mathcal{L}(0)$ counts the number of arrivals that were scheduled to arrive in $(0, t]$ but did not.

We can now prove that $N(t) - t$ converges weakly as $t \to \infty$, when we let $t \to \infty$ in such a way that $t - \lfloor t \rfloor$ is constant.

Theorem 1 *Suppose that* $\mathbb{E}|\xi_0| < \infty$*, and fix* $s \in [0, 1)$*. Then,*

$$N(n+s) - (n+s) \Rightarrow -s + I(U \le s) + \big(\mathcal{E}'(s) - \mathcal{L}'(s)\big) - \big(\mathcal{E}(0) - \mathcal{L}(0)\big)$$

as $n \to \infty$*, where* $\mathcal{E}'(s), \mathcal{L}'(s), \mathcal{E}(0), \mathcal{L}(0)$ *are independent of one another given* U*, and* $\mathcal{E}'(s) \overset{D}{=} \mathcal{E}(0)$*,* $\mathcal{L}'(s) \overset{D}{=} \mathcal{L}(0)$*.*

Proof Recall that

$$N(n+s) - (n+s) = \sum_{i+U \in (0,n+s]} 1 - (n+s) + (\mathcal{E}(n+s) - \mathcal{E}(0)) - (\mathcal{L}(n+s) - \mathcal{L}(0)).$$

We start by observing that

$$\sum_{i+U \in (0,n+s]} 1 - (n+s) = -s + I(U \le s)$$

for $n \in \mathbb{Z}_+, s \in [0, 1)$. Furthermore, if k_n is an integer-valued sequence such that $k_n/n \to \nu \in (0, 1)$ as $n \to \infty$, we can write

$$\begin{aligned}
&\big(\mathcal{E}(n+s) - \mathcal{E}(0), \mathcal{L}(n+s) - \mathcal{L}(0)\big) \\
&= \Bigg(\sum_{i+U > n+s} I(i + U + \xi_i \in (0, n+s]) - \sum_{i+U \in (0,n+s]} I(i + U + \xi_i \le 0), \\
&\quad \sum_{i+U \in (0,n+s]} I(i + \xi_i + U > n+s) - \sum_{i+U \le 0} I(i + U + \xi_i \in (0, n+s]) \Bigg) \\
&= \Bigg(\sum_{i+U > n+s} I(i + U + \xi_i \in (0, n+s]) \\
&\quad - \sum_{i+U \in (0,k_n]} I(i + U + \xi_i \le 0) - \sum_{i+U \in (k_n,n+s]} I(i + U + \xi_i \le 0), \\
&\quad \sum_{i+U \in (0,k_n]} I(i + \xi_i + U > n+s) + \sum_{i+U \in (k_n,n+s]} I(i + \xi_i + U > n+s) \\
&\quad - \sum_{i+U \le 0} I(i + U + \xi_i \in (0, n+s]) \Bigg) \\
&\overset{D}{=} \Big(\mathcal{E}''(n+s) - \hat{\mathcal{E}}_n - \hat{\mathcal{E}}''_n(n+s), \hat{\mathcal{L}}''_n(n+s) + \mathcal{L}''_n(n+s) - \hat{\mathcal{L}}(n+s) \Big).
\end{aligned}$$

Note that, because $\mathbb{E}\xi_0^+ < \infty$,

$$\mathbb{E}[\hat{\mathcal{L}}_n''(n+s)|U] = \sum_{i+U\in(0,k_n]} \mathbb{P}(i+\xi_i+U > n+s|U)$$
$$\leq \sum_{j+U\leq 0} \mathbb{P}(\xi_0 > n-k_n-j-1|U) \to 0$$

as $n \to \infty$, proving that $\hat{\mathcal{L}}_n''(n+s) \Rightarrow 0$ as $n \to \infty$. Similarly, the fact that $\mathbb{E}\xi_0^- < \infty$ implies that $\hat{\mathcal{E}}_n''(n+s) \Rightarrow 0$ as $n \to \infty$. Finally, the four random variables $\big(\mathcal{E}''(n+s), \hat{\mathcal{E}}_n, \mathcal{L}_n''(n+s), \hat{\mathcal{L}}(n+s)\big)$ all involve sums over subsets in i that are disjoint, so they are conditionally independent of one another, given U. Furthermore, note that

$$\mathcal{L}_n''(n+s) = \sum_{i+U\in(k_n,n+s]} I(i+\xi_i+U > n+s)$$
$$= \sum_{j+U\in(k_n-n,s]} I(j+\xi_{n+j}+U > s)$$
$$\stackrel{D}{=} \sum_{j+U\in(k_n-n,s]} I(j+\xi_j+U > s) \to \sum_{j+U<s} I(j+\xi_j+U > s)$$

as $n \to \infty$, proving that $\mathcal{L}_n''(n+s) \Rightarrow \mathcal{L}'(s)$ as $n \to \infty$. Similarly, $\mathcal{E}_n''(n+s) \Rightarrow \mathcal{E}'(s)$, while $\hat{\mathcal{E}}_n \Rightarrow \mathcal{E}(0)$ and $\hat{\mathcal{L}}_n \Rightarrow \mathcal{L}(0)$ as $n \to \infty$, proving the theorem. □

Note that we must restrict convergence to sequences of the form $t_n = n + s$ with $n \to \infty$. In particular, weak convergence does not hold when $t \to \infty$ without any restrictions. To see this, consider the case in which $\xi_0 = 0$ $a.s.$ Then,

$$N(t) - t = \sum_{i+U\in(0,t]} 1 - t$$
$$= -(t - \lfloor t \rfloor) + I(U \leq t - \lfloor t \rfloor),$$

and observe that the distribution of the right-hand side depends on $t - \lfloor t \rfloor$, regardless of the magnitude of t.

Theorem 1 shows that $N(t) - t$ is stochastically bounded in t. This is in sharp contrast to the case in which (for example) N is a unit rate renewal counting process with finite-variance inter-arrival times, in which event $t^{-1/2}\big(N(t) - t\big)$ converges weakly to a normal r.v. (see Ross [18]), so that $N(t) - t$ exhibits stochastic fluctuations of order $t^{1/2}$.

3 Tail asymptotics for sums of Bernoulli random variables

The analysis of Sect. 2 establishes that N, $\mathcal{E}$ and $\mathcal{L}$ all can be clearly represented as sums of independent Bernoulli r.v.'s. As we will see in the next section, the tail behavior of these r.v.'s significantly affects the queueing dynamics of systems that are fed by scheduled traffic. In addition, Bernoulli sums arise in many other application settings (for example, credit risk). As a consequence, this section is focused on tail behavior for such Bernoulli sums.

Let $(I_j : j \in \mathbb{Z})$ be a family of independent r.v.'s, in which $p_j = \mathbb{P}(I_j = 1) = 1 - \mathbb{P}(I_j = 0)$.

Theorem 2 *Suppose that there exist constants $c > 0$ and $\alpha > 1$ for which $p_n \sim c\, n^{-\alpha}$ as $n \to \infty$. If $Z = \sum_{j \geq 0} I_j$, then*

$$\frac{1}{z \log z} \log \mathbb{P}(Z > z) \to -\alpha$$

as $z \to \infty$.

Proof We shall employ an argument similar to that commonly used in the theory of large deviations; see, for example, p. 44 in Dembo and Zeitouni [6]. (Note, however, that the asymptotic setting described by Theorem 2 is not covered by traditional large deviations.) We start by observing that

$$\psi(\theta) \triangleq \log \mathbb{E} \exp\big(\theta Z\big) = \sum_{j \geq 0} \log\big(p_j (e^{\theta} - 1) + 1\big)$$

(where the sum converges absolutely since $\alpha > 1$). Choose $\theta = \theta(z)$ such that $e^{\theta(z)} = r z^{\alpha}$ (where $r > 0$), and note that, for $\epsilon > 0$, $\theta > 0$, and z sufficiently large,

$$\begin{aligned} \psi\big(\theta(z)\big) &= \sum_{0 \leq j \leq \lfloor \epsilon z \rfloor} \log\big(p_j (e^{\theta(z)} - 1) + 1\big) + \sum_{j > \lfloor \epsilon z \rfloor} \log\big(p_j (e^{\theta(z)} - 1) + 1\big) \\ &\leq \sum_{0 \leq j \leq \lfloor \epsilon z \rfloor} \log\big(e^{\theta(z)} + 1\big) + \sum_{j > \lfloor \epsilon z \rfloor} \log\big((1+\epsilon) c\, j^{-\alpha} e^{\theta(z)} + 1\big) \\ &\leq (\lfloor \epsilon z \rfloor + 1) \log(1 + r z^{\alpha}) + \sum_{j > \lfloor \epsilon z \rfloor} \log\big((1+\epsilon)\, rc\, (j/z)^{-\alpha} + 1\big). \end{aligned} \tag{3.1}$$

Note that the second term in (3.1), when multiplied by $1/z$, is a Riemann sum approximation, and hence

$$\begin{aligned} &\frac{1}{z} \sum_{j > \lfloor \epsilon z \rfloor} \log\big((1+\epsilon)\, rc\, (j/z)^{-\alpha} + 1\big) \\ &\quad \to \int_{\epsilon}^{\infty} \log\big((1+\epsilon)\, rc\, x^{-\alpha} + 1\big) \mathrm{d}x \end{aligned}$$

as $z \to \infty$. (Specifically, the function $\log\big((1+\epsilon)\, rc\, x^{-\alpha} + 1\big)$ is directly Riemann integrable (see Asmussen [2]), so the Riemann approximation over $[\epsilon, \infty)$ converges.) It follows that

$$\overline{\lim}_{z \to \infty} \frac{1}{z \log z} \psi\big(\theta(z)\big) \leq \epsilon\, \alpha.$$

Markov's inequality guarantees that

$$\mathbb{P}(Z > z) \leq \exp\big(-\theta(z)z + \psi(\theta(z))\big),$$

and hence

$$\overline{\lim}_{z\to\infty} \frac{1}{z\log z} \log \mathbb{P}(Z > z) \leq -\alpha\,(1-\epsilon).$$

Since $\epsilon > 0$ can be chosen to be arbitrarily small, we conclude that

$$\overline{\lim}_{z\to\infty} \frac{1}{z\log z} \log \mathbb{P}(Z > z) \leq -\alpha. \tag{3.2}$$

To obtain the lower bound needed for Theorem 2, we apply a change-of-measure argument. For $z > 0$, put

$$\widetilde{\mathbb{P}}_z(\cdot) = \mathbb{E} I(\cdot) \exp\big(\theta(z)\, Z - \psi(\theta(z))\big),$$

and let $\widetilde{\mathbb{E}}_z(\cdot)$ be the associated expectation operator. Then,

$$\mathbb{P}(Z > z) = \widetilde{\mathbb{E}}_z I(Z > z) \exp\big(-\theta(z) Z + \psi(\theta(z))\big). \tag{3.3}$$

Of course,

$$\widetilde{\mathbb{E}}_z Z = \psi'(\theta(z)) = \sum_{j\geq 0} \frac{p_j e^{\theta(z)}}{p_j (e^{\theta(z)} - 1) + 1}. \tag{3.4}$$

So,

$$\frac{1}{z}\,\widetilde{\mathbb{E}}_z Z = \sum_{j\geq 0} \frac{p_j\, r z^{\alpha}}{p_j (r z^{\alpha} - 1) + 1} \cdot \frac{1}{z}.$$

Since $p_j\, z^{\alpha} \sim c\,(j/z)^{-\alpha}$ as $j \to \infty$, a simple adaptation of the earlier Riemann sum approximation argument proves that

$$\frac{1}{z}\,\widetilde{\mathbb{E}}_z Z \to \int_0^{\infty} \frac{c\,r}{c\,r + x^{\alpha}} \mathrm{d}x$$

as $z \to \infty$. Similarly,

$$\begin{aligned} \frac{1}{z}\,\widetilde{\mathrm{var}}_z Z = \frac{1}{z}\psi''(\theta(z)) &= \sum_{j\geq 0} \frac{p_j\,(1-p_j)\, e^{\theta(z)}}{(p_j\,(e^{\theta(z)} - 1) + 1)^2} \cdot \frac{1}{z} \\ &\to \int_0^{\infty} \frac{c\,r\,x^{\alpha}}{(c\,r + x^{\alpha})^2} \mathrm{d}x \end{aligned}$$

as $z \to \infty$. For $\epsilon > 0$, we now select $r > 0$ (uniquely) so that

$$\int_0^\infty \frac{c\,r}{c\,r + x^\alpha} \mathrm{d}x = 1 + \epsilon.$$

Observe that because $\psi(\theta(z)) > 0$,

$$\begin{aligned}\mathbb{P}(Z > z) &= \widetilde{\mathbb{E}}_z I(Z > z) \exp\big(-\theta(z)\,Z + \psi(\theta(z))\big) \\ &\geq \exp\big(-\theta(z)\,z(1+2\varepsilon) + \psi(\theta(z))\big)\,\widetilde{\mathbb{P}}_z((1+2\varepsilon)z > Z > z) \\ &\geq \exp\big(-\theta(z)\,z(1+2\varepsilon)\big)\,\widetilde{\mathbb{P}}_z((1+2\varepsilon)z > Z > z). \end{aligned} \tag{3.5}$$

But, for z large enough, we have that

$$-\varepsilon/2 \leq \frac{1}{z}\widetilde{\mathbb{E}}_z Z - (1+\varepsilon) \leq \varepsilon/2$$

so that

$$\begin{aligned}\widetilde{\mathbb{P}}_z((1+2\varepsilon)z > Z > z) &= \widetilde{\mathbb{P}}_z((1+2\varepsilon)z - \widetilde{\mathbb{E}}_z Z > Z - \widetilde{\mathbb{E}}_z Z > z - \widetilde{\mathbb{E}}_z Z) \\ &\geq \widetilde{\mathbb{P}}_z((1+2\varepsilon)z - (1+3\varepsilon/2)z > Z - \widetilde{\mathbb{E}}_z Z > z - (1+\varepsilon-\varepsilon/2)\,z) \\ &\geq 1 - \widetilde{\mathbb{P}}_z(|Z - \widetilde{\mathbb{E}}_z Z| > \epsilon\, z/2) \\ &\geq 1 - 4\frac{\widetilde{\mathrm{var}}_z Z}{z^2\epsilon^2} \to 1 \end{aligned}$$

as $z \to \infty$, where the last inequality is an application of Chebyshev's inequality. Hence, (3.5) implies that

$$\underline{\lim}_{z\to\infty} \frac{1}{z\,\log z}\,\log \mathbb{P}(Z > z) \geq -\alpha,$$

proving the theorem. □

We now turn to the tail of Z when Z is the difference of two independent Bernoulli sums, $\sum_{j\geq 0} I_j$ and $\sum_{j<0} \tilde{I}_j$.

Corollary 1 *Suppose that* $\mathbb{E}\sum_{j<0}\tilde{I}_j < \infty$ *and that there exists* $c > 0$ *and* $\alpha > 1$ *for which* $\mathbb{E}I_n \sim c\,n^{-\alpha}$ *as* $n \to +\infty$. *If* $Z = \sum_{j\geq 0} I_j - \sum_{j<0}\tilde{I}_j$, *then*

$$\frac{1}{z\log z}\log \mathbb{P}(Z > z) \to -\alpha$$

as $z \to \infty$.

Proof We note that $\mathbb{P}(Z > z) \leq \mathbb{P}(\sum_{j\geq 0} I_j > z)$, and apply Theorem 2 to conclude that

$$\overline{\lim}_{z\to\infty} \frac{1}{z\log z}\log \mathbb{P}(Z > z) \leq -\alpha.$$

For the lower bound, observe that the independence yields

$$\begin{aligned}\mathbb{P}(Z > z) &\geq \mathbb{P}\left(\sum_{j\geq 0} I_j > z + d, \sum_{j<0} \tilde{I}_j \leq d\right)\\ &= \mathbb{P}\left(\sum_{j\geq 0} I_j > z + d\right) \mathbb{P}\left(\sum_{j<0} \tilde{I}_j \leq d\right).\end{aligned}$$

Hence, we apply Theorem 2 to conclude that

$$\underline{\lim}_{z\to\infty} \frac{1}{z\log z} \log \mathbb{P}(Z > z) \geq \underline{\lim}_{z\to\infty} \frac{1}{z\log z} \log \mathbb{P}\left(\sum_{j\geq 0} I_j > z + d\right) = -\alpha,$$

proving the result. □

We can immediately apply Theorem 2 and its corollary to the tail asymptotics of $\mathcal{E}$, $\mathcal{L}$ and $N(t)$.

Theorem 3 *i.) Suppose that* ξ_0 *is such that* $\mathbb{P}(\xi_0 > x) \sim c_1 x^{-\alpha_1}$ *as* $x \to \infty$ *for* $c_1 > 0$, $\alpha_1 > 1$. *Then,*

$$\frac{1}{x\log x} \log \mathbb{P}\big(\mathcal{L}(t) > x\big) \to -\alpha_1$$

as $x \to \infty$.

ii.) Suppose that ξ_0 *is such that* $\mathbb{P}(\xi_0 < -x) \sim c_2 x^{-\alpha_2}$ *as* $x \to \infty$ *for* $c_2 > 0, \alpha_2 > 1$. *Then,*

$$\frac{1}{x\log x} \log \mathbb{P}\big(\mathcal{E}(t) > x\big) \to -\alpha_2$$

as $x \to \infty$.

iii.) Suppose that ξ_0 *has a bounded density* f *for which there exist positive constants* $c_1, c_2, \alpha_1, \alpha_2$ *such that*

$$\begin{aligned}f(x) &\sim c_1\, x^{-\alpha_1 - 1},\\ f(-x) &\sim c_2\, x^{-\alpha_2 - 1}\end{aligned}$$

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

$$\frac{1}{x\log x} \log \mathbb{P}\big(N(t) > x\big) \to -\min(\alpha_1 + 1, \alpha_2 + 1) \tag{3.6}$$

as $x \to \infty$.

Proof For part i.) we recall that $\mathcal{L}(t) \stackrel{D}{=} \mathcal{L}(0)$. Furthermore, $\mathbb{P}(\xi_{j+1} > j) \le \mathbb{P}(-j + \xi_{-j} + U > 0) \le \mathbb{P}(\xi_j > j-1)$, so that

$$\mathbb{P}\left(\sum_{j=1}^{\infty} I_j > x\right) \le \mathbb{P}(\mathcal{L}(t) > x) \le \mathbb{P}\left(\sum_{j=0}^{\infty} I_j > x\right), \tag{3.7}$$

where the I_j's are independent Bernoulli r.v.'s in which $I_j = I(\xi_j > j-1)$. Theorem 2 can then be applied to the extreme members of (3.7), yielding i.). Part ii.) follows similarly. As for iii.), suppose that $\alpha_1 \le \alpha_2$ and set $I_j = I(j + \xi_j + U \in (0, t])$. Fix an integer $d \ge 1$ and observe that

$$\begin{aligned} \mathbb{P}\left(\sum_{j\le 0} I_j > x\right) &\le \mathbb{P}\big(N(t) > x\big) \\ &\le \sum_{k=0}^{d-1} \mathbb{P}\left(\sum_{j\le 0} I_j \in x\,[k/d, (k+1)/d), \sum_{j>0} I_j > x\,(1 - (k+1)/d)\right) \\ &\le d \max_{0\le k\le d-1} \mathbb{P}\left(\sum_{j\le 0} I_j \ge x\,k/d\right) \mathbb{P}\left(\sum_{j>0} I_j \ge x\,(1-(k+1)/d)\right). \end{aligned} \tag{3.8}$$

Recalling our bounded density assumption, the Bounded Convergence Theorem implies that $\mathbb{P}(I_{-j} = 1) \sim c_1\, t\, j^{-\alpha_1 - 1}$ and $\mathbb{P}(I_j = 1) \sim c_2\, t\, j^{-\alpha_2 - 1}$ as $j \to \infty$.

Arguing as for i.) and ii.), we find that

$$\frac{1}{x \log x} \log \mathbb{P}\left(\sum_{j\le 0} I_j > x\right) \to -(\alpha_1 + 1) \tag{3.9}$$

and

$$\frac{1}{x \log x} \log \mathbb{P}\left(\sum_{j>0} I_j > x\right) \to -(\alpha_2 + 1) \tag{3.10}$$

as $x \to \infty$. Utilizing (3.9) and (3.10), we observe that, for $0 \le k \le d-1$, the term

$$\begin{aligned} &\frac{1}{x \log x} \log\left[\mathbb{P}\left(\sum_{j\le 0} I_j \ge x\,k/d\right) \mathbb{P}\left(\sum_{j>0} I_j \ge x\,(1-(k+1)/d)\right)\right] \\ &\to -\frac{(\alpha_1+1)\,k}{d} - \frac{(\alpha_2+1)\,(d-(k+1))}{d} \le -(\alpha_1+1)\frac{d-1}{d} \end{aligned} \tag{3.11}$$

as $x \to \infty$. Hence, letting $x \to \infty$ on the extreme terms of (3.8), followed by sending $d \to \infty$ yields iii.). A symmetric argument works for $\alpha_1 > \alpha_2$. □

The next result shows that the tail exponent of $N(t)$ given in $iii.$) of Theorem 3 is not inherited by its equilibrium limit given in Theorem 1. In other words, one cannot interchange $x \to \infty$ in the logarithmic limit in (3.6), with $t \to \infty$ in time.

For our next result, we fix $s \in [0, 1]$ and recall Theorem 1 and the quantities $\mathcal{E}'(s)$ and $\mathcal{E}'(s)$ defined there.

Proposition 3 *Suppose that ξ_0 is such that $\mathbb{P}(\xi_0 > x) \sim c_1 x^{-\alpha_1}$ and $\mathbb{P}(\xi_0 < -x) \sim c_2 x^{-\alpha_2}$ as $x \to \infty$ for $c_1, c_2 > 0$ and $\alpha_1, \alpha_2 > 1$. Then,*

$$\frac{1}{x \log x} \log \mathbb{P}\Big(\big(\mathcal{E}'(s) - \mathcal{L}'(s)\big) - \big(\mathcal{E}(0) - \mathcal{L}(0)\big) > x\Big) \to -\min(\alpha_1, \alpha_2)$$

as $x \to \infty$.

Proof Utilizing Corollary 1 and arguing as in the proof of Theorem 3, we find that

$$\frac{1}{x \log x} \log \mathbb{P}\big(\mathcal{E}'(s) - \mathcal{L}'(s) > x)\big) \to -\alpha_2,$$

and

$$\frac{1}{x \log x} \log \mathbb{P}\big(\mathcal{L}(0) - \mathcal{E}(0) > x)\big) \to -\alpha_1$$

as $x \to \infty$. We can now use the same upper bound argument as in (3.8) to conclude that

$$\overline{\lim}_{x\to\infty} \frac{1}{x \log x} \log \mathbb{P}\Big(\big(\mathcal{E}'(s) - \mathcal{L}'(s)\big) - \big(\mathcal{E}(0) - \mathcal{L}(0)\big) > x\Big) \leq -\min(\alpha_1, \alpha_2).$$

For the lower bound, suppose that $\alpha_2 \leq \alpha_1$. We find that

$$\begin{aligned} &\mathbb{P}\big(\big(\mathcal{E}'(s) - \mathcal{L}'(s)\big) - \big(\mathcal{E}(0) - \mathcal{L}(0)\big) > x\big) \\ &\quad \geq \mathbb{P}\big(\mathcal{E}'(s) - \mathcal{L}'(s) > x + d\big)\, \mathbb{P}\big(\mathcal{L}(0) - \mathcal{E}(0) \geq -d\big), \end{aligned}$$

so that

$$\underline{\lim}_{x\to\infty} \frac{1}{x \log x} \log \mathbb{P}\big(\big(\mathcal{E}'(s) - \mathcal{L}'(s)\big) - \big(\mathcal{E}(0) - \mathcal{L}(0)\big) > x\big) \geq -\alpha_2.$$

A symmetric argument holds for $\alpha_1 < \alpha_2$. □

We have already argued in Sect. 2 that for any i.i.d. sequence of perturbations, the corresponding counting process $N(t)$ has a light tail. Theorem 3 shows that when the perturbations have Pareto tails, then $N(t)$ has a tail lighter than an exponential and heavier than a normal distribution.

Because of their intrinsic interest and their importance for scheduled queues, we now provide exact tail asymptotics for Bernoulli sums.

Theorem 4 *i.) Suppose that there exists $c > 0$ and $\alpha > 1$ such that $p_n = c\, n^{-\alpha}(1 + O(1/n))$ as $n \to \infty$. Then,*

$$\mathbb{P}\left(\sum_{j\geq 0} I_j \geq n\right) \sim \frac{1}{\sqrt{2\pi\,\eta_*}}\, r_*^{-n}\, n^{-\alpha n - 1/2} \exp\big(\psi(r_* n^\alpha)\big)$$

as $n \to \infty$, where r_* satisfies

$$\int_0^\infty \frac{c\, r_*}{c\, r_* + x^\alpha} \mathrm{d}x = 1$$

and

$$\eta_* = \int_0^\infty \frac{c\, r_*\, x^\alpha}{(c\, r_* + x^\alpha)^2} \mathrm{d}x.$$

ii.) Suppose that $p_n = c(w+n)^{-\alpha}$ for $n \geq 0$, where $c, w > 0$ and $\alpha > 1$. Then,

$$\mathbb{P}\left(\sum_{j\geq 0} I_j \geq n\right) \sim \left(\frac{1}{\sqrt{2\pi}}\right)^{\alpha+1} \frac{\Gamma(w)^\alpha}{\sqrt{\eta_*}}\, (c\, r_*)^{\frac{1}{2}-w}\, n^{-\alpha n + \frac{1}{2}(\alpha-1) - w\alpha}\, e^{\gamma n}$$

as $n \to \infty$, where $\gamma = \int_0^1 \log\big(1 + \frac{1}{c\, r_*} x^\alpha\big)\mathrm{d}x + \int_1^\infty \log\big(1 + c\, r_*\, x^{-\alpha}\big)\mathrm{d}x + \alpha + \log(c)$.

Proof We start from the change-of-measure formula (3.3), with the specific choice $\theta_*(n) = \log r_* + \alpha \log n$ (so that $\exp(\theta_*(n)) = r_* n^\alpha$). Then,

$$\mathbb{P}(Z \geq n) = \exp\big(-\theta_*(n)\, n + \psi(r_* n^\alpha)\big) \cdot \widetilde{\mathbb{E}}_n I(Z \geq n) \exp\big(-\theta_*(n)(Z - n)\big).$$

We wish now to apply the local central limit theorem (CLT) to Z under $\widetilde{P}_n$. Recall (3.4) and note that

$$\begin{aligned}
\widetilde{\mathbb{E}}_n Z &= \sum_{j\geq 0} \frac{p_j\, r_* n^\alpha}{p_j\,(r_*\, n^\alpha - 1) + 1} \\
&= \frac{p_0\, r_* n^\alpha}{p_0\,(r_*\, n^\alpha - 1) + 1} + \sum_{j\geq 1} \frac{(n/j)^\alpha c\, r_*}{(n/j)^\alpha\, c\, r_* - p_j + 1}\,(1 + O(1/j)) \\
&= \frac{1}{1 + O(n^{-\alpha})} + \sum_{j\geq 1} \frac{c\, r_*}{c\, r_* + (j/n)^\alpha - (j/n)^\alpha\, p_j}\,(1 + O(1/j)) \\
&= \sum_{j\geq 0} \frac{c\, r_*}{c\, r_* + (j/n)^\alpha + O(n^{-\alpha})}\,(1 + O(1/j)) \\
&= \sum_{j\geq 0} \frac{c\, r_*}{c\, r_* + (j/n)^\alpha}(1 + O(1/j))
\end{aligned}$$

$$= \sum_{j\geq 0} \frac{c\, r_*}{c\, r_* + (j/n)^\alpha} + \sum_{0\leq j\leq k_n} O(1/j) + O(k_n^{-1}) \sum_{j>k_n} \frac{c\, r_*}{c\, r_* + (j/n)^\alpha}$$
$$= n \sum_{j\geq 0} \frac{1}{n}\, v(j/n) + O(\log k_n) + O(n\, k_n^{-1}) \sum_{j>k_n} \frac{1}{n}\, v(j/n),$$

where $v(x) = c\, r_* \,(c\, r_* + x^\alpha)^{-1}$ and k_n is selected so that $k_n/n^{2/3} \to 1$ as $n \to \infty$. But

$$n \sum_{j\geq 0} \frac{1}{n}\, v(j/n) = n \int_0^\infty v(x)\mathrm{d}x + n \sum_{j\geq 0} \int_{j/n}^{(j+1)/n} [v(j/n) - v(x)]\mathrm{d}x.$$

The defining equation for r_* implies that $\int_0^\infty v(x)\mathrm{d}x = 1$. Set

$$\omega_n(x) = \int_{j/n}^{x} [v(j/n) - v(y)]\mathrm{d}y.$$

Since ω_n is twice differentiable with $w_n'(x) = v(j/n) - v(x)$ and $w_n''(x) = -v'(x)$, there exists $x_{j,n} \in [j/n, (j+1)/n]$ such that

$$\omega_n\big((j+1)/n\big) = \omega_n(j/n) + 1/n \cdot \omega_n'(j/n) + 1/n^2 \cdot \omega_n''(x_{j,n})/2$$

so that

$$\int_{j/n}^{(j+1)/n} [v(j/n) - v(y)]\mathrm{d}y = -v'(x_{j,n}) \cdot \frac{1}{2\, n^2}$$

and hence

$$n \sum_{j\geq 0} \int_{j/n}^{(j+1)/n} [v(j/n) - v(x)]\mathrm{d}x = -1/2 \sum_{j\geq 0} v'(x_{j,n})\, \frac{1}{n}.$$

The latter sum is a Riemann sum approximation to the integral of $-\frac{1}{2}v'(\cdot)$ over $[0, \infty)$. Consequently,

$$n \sum_{j\geq 0} \int_{j/n}^{(j+1)/n} [v(j/n) - v(x)]\mathrm{d}x \to -\frac{1}{2} \int_0^\infty v'(x)\mathrm{d}x = 1/2. \tag{3.12}$$

Similarly,

$$\sum_{j>k_n} \frac{1}{n} v(j/n) - \int_{k_n/n}^\infty v(x)\mathrm{d}x \to 0$$

as $n \to \infty$. It follows that

$$\widetilde{\mathbb{E}}_n Z = n + O(n^{1/3}) \tag{3.13}$$

as $n \to \infty$. Also, as noted in the proof of Theorem 2,

$$\frac{1}{n}\, \widetilde{\text{var}}_n Z \to \eta_* \tag{3.14}$$

as $n \to \infty$.

We are now ready to apply the local CLT due to Davis and McDonald [5]. We first write $Z = \sum_{j=0}^{b_n} I_j + Y_n$, where $b_n \to \infty$ fast enough that $\widetilde{\mathbb{E}}_n Y_n^2/n \to 0$ as $n \to \infty$. It is easily verified that the conditions of the Lindeberg-Feller CLT apply to $(Z - \widetilde{\mathbb{E}}_n Z)/(\widetilde{\text{var}}_n Z)^{1/2}$; see p. 215 of Chung [4]. Furthermore, by recalling that, $\widetilde{\text{var}}_n I_j = \widetilde{\mathbb{P}}_n(I_j = 1)\, \widetilde{\mathbb{P}}_n(I_j = 0)$, we conclude that the sequence $Q_n = \sum_{j \le b_n} \min\big(\widetilde{\mathbb{P}}_n(I_j = 0), \widetilde{\mathbb{P}}_n(I_j = 1)\big)$ that appears in the hypotheses of Theorem 1.2 of Davis and McDonald [5] can be lower bounded by $\sum_{j \le b_n} \widetilde{\text{var}}_n I_j \sim n\, \eta_*$ as $n \to \infty$. Consequently, Theorem 1.2 asserts that

$$\widetilde{\mathbb{P}}_n(Z = k) = \phi\left(\frac{k - \widetilde{\mathbb{E}}_n Z}{\sqrt{n\, \eta_*}}\right) \frac{1}{\sqrt{n\, \eta_*}}\, (1 + o(1))$$

uniformly in k as $n \to \infty$, where $\phi(\cdot)$ is the density of a $\mathcal{N}(0, 1)$ r.v. Hence, in view of (3.13) and (3.14),

$$\widetilde{\mathbb{P}}_n(Z = n + k) = \phi\left(\frac{k}{\sqrt{n\, \eta_*}}\right) \frac{1}{\sqrt{n\, \eta_*}}\, (1 + o(1))$$

as $n \to \infty$, so that

$$\begin{aligned}
&\widetilde{\mathbb{E}}_n I(Z \ge n) \exp\big(-\theta_*(n)\,(Z - n)\big) \\
&= \sum_{k \ge 0} \widetilde{\mathbb{P}}_n(Z = n + k)\, \exp\big(-\theta_*(n)\, k\big) \\
&\sim \frac{1}{\sqrt{2\,\pi\, n\, \eta_*}}
\end{aligned}$$

as $n \to \infty$, proving part i.).

Part ii.) is a special case of i.). All that is needed is the development of an asymptotic for $\psi(r_*\, n^\alpha)$, to the order of $o(1)$. Denoting (as usual) the gamma function by $\Gamma(\cdot)$, we write

$$\begin{aligned}
\psi(r_*\, n^\alpha) &= \sum_{j=0}^{n-1} \log\big(c\,(j + w)^{-\alpha}\,(r_*\, n^\alpha - 1) + 1\big) \\
&\quad + \sum_{j \ge n} \log\big(c\,(j + w)^{-\alpha}\,(r_*\, n^\alpha - 1) + 1\big)
\end{aligned}$$

$$
\begin{aligned}
&= \log\left(\prod_{j=0}^{n-1}\left(\frac{n}{j+w}\right)^{\alpha}(c\,r_*)^n\right) + \sum_{j=0}^{n-1}\log\left(1+(\frac{j+w}{n}\right)^{\alpha}\frac{1}{c\,r_*}-\frac{1}{c\,r_*\,n^{\alpha}}\Big)\\
&\quad + \sum_{j\geq n}\log\left(1+c\,r_*\left(\frac{j+w}{n}\right)^{-\alpha}\left(1-\frac{1}{r_*\,n^{\alpha}}\right)\right) \qquad (3.15)\\
&= \log\left(\left(\frac{n^n\,\Gamma(w)}{\Gamma(w+n)}\right)^{\alpha}(c\,r_*)^n\right) + n\int_{w/n}^{1+w/n}\log\left(1+\frac{1}{c\,r_*}\,x^{\alpha}-\frac{1}{c\,r_* n^{\alpha}}\right)\mathrm{d}x\\
&\quad + n\int_{1+w/n}^{\infty}\log\left(1+c\,r_*\,x^{-\alpha}\left(1-\frac{1}{r_*\,n^{\alpha}}\right)\right)\mathrm{d}x\\
&\quad + n\sum_{j=0}^{n-1}\int_{j/n}^{(j+1)/n}[h_n(j/n)-h_n(x)]\mathrm{d}x + n\sum_{j\geq n}\int_{j/n}^{(j+1)/n}[\tilde{h}_n(j/n)-\tilde{h}_n(x)]\mathrm{d}x,
\end{aligned}
$$

where

$$
\begin{aligned}
h_n(x) &= \log\left(1+\left(x+\frac{w}{n}\right)^{\alpha}\frac{1}{c\,r_*}-\frac{1}{c\,r_* n^{\alpha}}\right),\\
\tilde{h}_n(x) &= \log\left(1+c\,r_*\left(x+\frac{w}{n}\right)^{-\alpha}\left(1-\frac{1}{r_*\,n^{\alpha}}\right)\right).
\end{aligned}
$$

The first term in the third equality is due to the property of the Gamma function whereby, for any $z > 0$, $\Gamma(z+1) = z\Gamma(z)$. Set $h(x) = \log\left(1+\frac{1}{c\,r_*}\,x^{\alpha}\right)$, $\tilde{h}(x) = \log\left(1+c\,r_*\,x^{-\alpha}\right)$. Arguing as in (3.12), the sum of the last two terms converges to

$$
\begin{aligned}
&-1/2\int_0^1 h'(x)\mathrm{d}x - 1/2\int_1^{\infty}\tilde{h}'(x)\mathrm{d}x\\
&= -1/2\,(h(1)-h(0)) - 1/2\,(\tilde{h}(\infty)-\tilde{h}(1)) \qquad (3.16)\\
&= -1/2\left(\log\left(1+\frac{1}{c\,r_*}\right)-\log(1+c\,r_*)\right) = 1/2\,\log(c\,r_*).
\end{aligned}
$$

Also,

$$
\begin{aligned}
&n\int_{w/n}^{1+w/n}\log\left(1+\frac{1}{c\,r_*}\,x^{\alpha}-\frac{1}{c\,r_* n^{\alpha}}\right)\mathrm{d}x\\
&= n\int_{w/n}^{1+w/n}\log\left(1+\frac{1}{c\,r_*}\,x^{\alpha}\right)\mathrm{d}x + O(n^{1-\alpha}) \qquad (3.17)\\
&= n\int_0^1\log\left(1+\frac{1}{c\,r_*}\,x^{\alpha}\right)\mathrm{d}x + w\,h(1) - w\,h(0) + o(1)
\end{aligned}
$$

as $n \to \infty$. Similarly,

$$\begin{aligned} & n \int_{1+w/n}^{\infty} \log\left(1 + c\, r_*\, x^{-\alpha}\left(1 - \frac{1}{n^\alpha}\right)\right) \mathrm{d}x \\ & = n \int_1^\infty \log\left(1 + c\, r_*\, x^{-\alpha}\right) \mathrm{d}x - w\, \tilde{h}(1). \end{aligned} \tag{3.18}$$

Finally, we use the asymptotic

$$\Gamma(w+n) \sim \sqrt{2\,\pi\, n}\left(\frac{w+n-1}{e}\right)^{n+w-1}$$

as $n \to \infty$ (see p. 63 of Feller [8]), to conclude that

$$\left(\frac{n^n\,\Gamma(w)}{\Gamma(w+n)}\right)^\alpha \sim \left(\frac{\Gamma(w)}{\sqrt{2\,\pi}}\right)^\alpha n^{(-w+1/2)\,\alpha}\, e^{\alpha\, n} \tag{3.19}$$

as $n \to \infty$. Combining (3.15) through (3.19) yields part ii.). □

With Theorem 4 at our disposal, we can now derive exact asymptotics for the r.v.'s $\mathcal{E}(t)$ and $\mathcal{L}(t)$. For example, in view of the fact that the proof of part ii.) holds uniformly in w,

$$\mathbb{P}(\mathcal{E}(0) \geq n) \sim \left(\frac{1}{\sqrt{2\pi}}\right)^{\alpha+1} \sqrt{\frac{c\, r_*}{\eta_*}}\, \mathbb{E}\left[\frac{\Gamma(U)^\alpha}{(c\, r_*\, n^\alpha)^U}\right] n^{-\alpha n + \frac{1}{2}\,(\alpha-1)}\, e^{\gamma n}$$

as $n \to \infty$, provided that $\mathbb{P}(\xi_0 \leq -x) = c\, x^{-\alpha}$ for $x > 1$ and where γ is the same constant defined in Theorem 4.

4 Behavior of the $S/D/1$ workload process under critical loading

In this section, we consider a queue that is fed by a scheduled traffic in which each customer's service time requirement is of unit duration, and in which the server has the capacity to process work at unit rate. Under these assumptions, the rate at which work arrives per unit time equals the service capacity of the system, so that the queue is subject to critical loading.

Note that the total work to arrive in $(0, t]$ is given by $N(t)$. Let $W(t)$ be the workload in the system at time t (i.e. $t + W(t)$ is the first time subsequent to t at which the system would empty if no additional work were to arrive after t.) If $W(0) = 0$, then

$$W(t) = \max_{0 \leq s \leq t}\Big[(N(t) - t) - (N(s) - s)\Big]. \tag{4.1}$$

Our goal is to analyze the behavior of $W(t)$ for t large. We note that because the service times are deterministic with unit duration, the number-in-system at time t equals $\lceil W(t) \rceil$. Therefore, our results on the workload can be easily converted into results on the number-in-system.

If $\mathbb{E}|\xi_0| < \infty$, Proposition 2 applies so that

$$W(t) = \max_{0 \le s \le t} \Big[\mathcal{E}(t) - \mathcal{E}(s) - \mathcal{L}(t) + \mathcal{L}(s)\Big] + O_p(1), \tag{4.2}$$

where $O_p(1)$ is a term that is stochastically bounded in t. Because of the stationarity of $\big((\mathcal{E}(t), \mathcal{L}(t)) : t \in \mathbb{R}\big)$, the first term in (4.2) has the same distribution as

$$\begin{aligned} M(t) &= \max_{0 \le s \le t} \Big[\mathcal{E}(0) - \mathcal{E}(-(t-s)) - \mathcal{L}(0) + \mathcal{L}(-(t-s))\Big], \\ &= \max_{0 \le r \le t} \Big[\mathcal{E}^*(r) - \mathcal{L}^*(r)\Big] - \mathcal{E}^*(0) + \mathcal{L}^*(0). \end{aligned} \tag{4.3}$$

Note that $\mathcal{E}^*(\cdot)$ is the "early customer" process for the time-reversed system in which the perturbations are given by $(-\xi_{-j} : j \in \mathbb{Z})$, and $\mathcal{L}^*(\cdot)$ is the corresponding "late customer" process. As a result, $\mathcal{E}^*(r) \stackrel{D}{=} \mathcal{L}(-r)$ and $\mathcal{L}^*(r) \stackrel{D}{=} \mathcal{E}(-r)$. As a matter of fact these equalities hold pathwise except for the fact that $\mathcal{E}^*$ and $\mathcal{L}^*$ are generated following the uniform distribution, $1-U$, instead of U. Given that $\mathcal{E}^*$ and $\mathcal{L}^*$ are non-negative processes for which $\mathcal{E}^*(r)$ is independent of $\mathcal{L}^*(r)$ for $r \in \mathbb{R}$, it is evident that the growth of $M(t)$ will be determined by $\mathcal{E}^*$ and that the left tail of $-\xi_0$ (or right tail of ξ_0) governs the large time behavior of $M(\cdot)$ (and hence $W(\cdot)$). The dominance of the right tail of ξ_0 over the left tail is perhaps explained by the fact that the left tail induces the arrival of "early customers" from the future evolution of the queue. More such early arrivals in an interval mean fewer potential customers available from which to stimulate a future burst of arrivals, so that the left tail has less influence over "growing" $M(\cdot)$ over time.

Theorem 5 *Suppose that $\mathbb{E}\xi_0^- < \infty$ and that there exists constant $c > 0$ and $\alpha > 1$ for which $\mathbb{P}(\xi_0 > x) \sim c\,x^{-\alpha}$ as $x \to \infty$. Then,*

$$\frac{W(t)}{\log t / \log\log t} \Rightarrow 1/\alpha$$

as $t \to \infty$.

Proof Clearly,

$$\begin{aligned} \max_{0 \le r \le t} \big[\mathcal{E}^*(r) - \mathcal{L}^*(r)\big] &\le \max_{0 \le r \le t} \mathcal{E}^*(r) \\ &\le \max_{1 \le n \le \lfloor t \rfloor + 1} \max_{0 \le s < 1} \mathcal{E}^*(n-s). \end{aligned}$$

For $0 \le s < 1$ and $n \ge 1$,

$$\begin{aligned} \mathcal{E}^*(n-s) &= \sum_{j+U \le -n+s} I(j + U + \xi_j > -n + s) \\ &\le \sum_{j+U \le -n+s} I(j + U + \xi_j > -n) \\ &= \mathcal{E}^*(n) + \sum_{-n < j+U \le -n+s} I(j + U + \xi_j > -n) \end{aligned}$$

$$
\begin{aligned}
&\le \mathcal{E}^*(n) + \sum_{-n<j+U\le -n+s} 1 \\
&= \mathcal{E}^*(n) + 1
\end{aligned} \tag{4.4}
$$

So,

$$
\max_{0\le r\le t} \left[\mathcal{E}^*(r) - \mathcal{L}^*(r)\right] \le 1 + \max_{1\le n\le \lfloor t\rfloor+1} \mathcal{E}^*(n).
$$

Hence, for $\varepsilon > 0$ and t sufficiently large,

$$
\begin{aligned}
&\mathbb{P}\left(\max_{0\le r\le t}\left[\mathcal{E}^*(r) - \mathcal{L}^*(r)\right] > \frac{1+3\,\varepsilon}{\alpha}\frac{\log t}{\log\log t}\right) \\
&\le \mathbb{P}\left(\max_{0\le n\le \lfloor t\rfloor+1} \mathcal{E}^*(n) > \frac{1+2\,\varepsilon}{\alpha}\frac{\log t}{\log\log t}\right) \\
&\le \sum_{n=1}^{\lfloor t\rfloor+1} \mathbb{P}\left(\mathcal{E}^*(n) > \frac{1+2\,\varepsilon}{\alpha}\frac{\log t}{\log\log t}\right) \\
&\le (\lfloor t\rfloor+1)\,\mathbb{P}\left(\mathcal{E}^*(0) > \frac{1+2\,\varepsilon}{\alpha}\frac{\log t}{\log\log t}\right) \\
&= (\lfloor t\rfloor+1)\,\exp\left(\log\left(\mathbb{P}\left(\mathcal{L}(0) > \frac{1+2\,\varepsilon}{\alpha}\frac{\log t}{\log\log t}\right)\right)\right) \\
&\le (\lfloor t\rfloor+1)\,\exp\big(-(1+\varepsilon)\,\log t\big) \\
&= O(t^{-\varepsilon}) \to 0
\end{aligned}
$$

as $t \to \infty$, where we used Theorem 3 for the final inequality.

To obtain the necessary lower bound, fix $\varepsilon \in (0, 1/8\alpha)$ and note that, for such ε, $1-2\,\varepsilon+\varepsilon^2 < 1-2\,\varepsilon-\varepsilon^2+\varepsilon/2\alpha\ (<1)$. Choose τ in the interval $(1-2\,\varepsilon+\varepsilon^2, 1-2\,\varepsilon-\varepsilon^2+\varepsilon/2\alpha)$. Put $b(t) = (1/\alpha)(\log t/\log\log t)$, $c(t) = (1-2\varepsilon)^2\, b(t)^2$, and $k(t) = \lceil t^\tau \rceil$. As in the proof of Theorem 2, we find that, for $\theta > 0$,

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}(0) \ge n) &\le \mathbb{P}\left(\sum_{j\ge -1} I(j+\xi_j \le 0) \ge n\right) \\
&\le \exp\left(-\theta n + \sum_{j\ge -1} \log\left(\mathbb{P}(j+\xi_0 \le 0)(e^\theta - 1) + 1\right)\right) \\
&\le \exp\left(-\theta n + \sum_{j\ge -1} \mathbb{P}(j+\xi_0 \le 0)(e^\theta - 1)\right) \\
&= \exp\left(-\theta n + (e^\theta - 1)\,\mathbb{E}\sum_{j=-1}^{\lceil -\xi_0\rceil} 1\right)
\end{aligned}
$$

$$\leq \exp\left(-\theta n + (e^{\theta} - 1)(\mathbb{E}\xi_0^- + 3)\right).$$

By setting $\theta = \log n$, we conclude that

$$\mathbb{P}\big(\mathcal{E}(0) \geq n)\big) \leq \exp\left(- n \log n + O(n)\right)$$

so that

$$\mathbb{P}\big(\mathcal{E}(0) \geq \varepsilon\, b(t)\big) \leq t^{-\varepsilon/2\alpha} \tag{4.5}$$

for t sufficiently large. In addition, an examination of the proof of Theorem 2 shows that, under the conditions stated there,

$$\frac{1}{z \log z} \log \mathbb{P}\left(\sum_{j=0}^{\lceil z^2 \rceil} I_j > z\right) \to -\alpha \tag{4.6}$$

as $z \to \infty$.

We now subdivide the interval $[-t, 0]$ into $k(t)$ subintervals of equal length, and let $r_1, r_2, ..., r_{k(t)}$ be the right endpoints of the $k(t)$ subintervals.

Then,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{0\leq r\leq t} \big[\mathcal{E}^*(r) - \mathcal{L}^*(r)\big] > (1-3\,\varepsilon)\, b(t)\Big)\\
&\quad \geq \mathbb{P}\Big(\max_{1\leq i\leq k(t)} \big[\mathcal{L}(r_i) - \mathcal{E}(r_i)\big] > (1-3\,\varepsilon)\, b(t)\Big)\\
&\quad \geq \mathbb{P}\Big(\max_{1\leq i\leq k(t)} \big[(\mathcal{L}(r_i) - \mathcal{E}(r_i))\, I\big(\mathcal{L}^*(r_i) \leq \varepsilon\, b(t)\big)\big] > (1-3\,\varepsilon)\, b(t)\Big)\\
&\quad \geq \mathbb{P}\Big(\max_{1\leq i\leq k(t)} \big[\mathcal{L}(r_i)\, I\big(\mathcal{E}(r_i) \leq \varepsilon\, b(t)\big)\big] > (1-2\,\varepsilon)\, b(t)\Big)
\end{aligned}$$

Let i^* be the maximizer of $\mathcal{L}(r_i)$. We can then write that

$$\begin{aligned}
&\mathbb{P}\Big(\max_{0\leq r\leq t} \big[\mathcal{E}^*(r) - \mathcal{L}^*(r)\big] > (1-3\,\varepsilon)\, b(t)\Big)\\
&\quad \geq \mathbb{P}\Big(\max_{1\leq i\leq k(t)} \mathcal{L}(r_i) > (1-2\,\varepsilon)\, b(t)\,, \mathcal{E}(r_{i^*}) \leq \varepsilon\, b(t)\Big)\\
&\quad = \mathbb{P}\Big(\max_{1\leq i\leq k(t)} \mathcal{L}(r_i) > (1-2\,\varepsilon)\, b(t)\Big)\\
&\qquad - \mathbb{P}\Big(\max_{1\leq i\leq k(t)} \mathcal{L}(r_i) > (1-2\,\varepsilon)\, b(t)\,, \mathcal{E}(r_{i^*}) > \varepsilon\, b(t)\Big)\\
&\quad \geq \mathbb{P}\Big(\max_{1\leq i\leq k(t)} \mathcal{L}(r_i) > (1-2\,\varepsilon)\, b(t)\Big)\\
&\qquad - \mathbb{P}\Big(\max_{1\leq i\leq k(t)} \mathcal{L}(r_i) I\big(\mathcal{E}(r_i) > \varepsilon\, b(t)\big) > (1-2\,\varepsilon)\, b(t)\Big)\\
&\quad \geq \mathbb{P}\Big(\max_{1\leq i\leq k(t)} \mathcal{L}(r_i) > (1-2\,\varepsilon)\, b(t)\Big)
\end{aligned}$$

$$- \sum_{i=1}^{k(t)} \mathbb{P}(\mathcal{L}(r_i) I(\mathcal{E}(r_i) > \varepsilon\, b(t)) > (1-2\,\varepsilon)\, b(t)).$$

Recall that $\mathcal{L}(r_i) \stackrel{D}{=} \mathcal{L}(0)$, $(\mathcal{E}(r_i) \stackrel{D}{=} \mathcal{E}(0))$. We conclude that

$$\begin{aligned}
&\mathbb{P}\Big(\max_{0\le r\le t} \big[\mathcal{E}^*(r) - \mathcal{L}^*(r)\big] > (1-3\,\varepsilon)\, b(t)\Big)\\
&\quad \ge \mathbb{P}\Big(\max_{1\le i\le k(t)} \mathcal{L}(r_i) > (1-2\,\varepsilon)\, b(t)\Big)\\
&\qquad - k(t)\, \mathbb{P}\Big(\mathcal{L}(0) > (1-2\,\varepsilon)\, b(t), \mathcal{E}(0) > \varepsilon\, b(t)\Big)\\
&\quad \ge \mathbb{P}\Big(\max_{1\le i\le k(t)} \sum_{r_i - c(t)\le j\le r_i - 1} I(j+\xi_j > r_i) > (1-2\,\varepsilon)\, b(t)\Big)\\
&\qquad - k(t)\, \mathbb{P}\Big(\mathcal{L}(0) > (1-2\,\varepsilon)\, b(t)\Big)\, \mathbb{P}\Big(\mathcal{E}(0) > \varepsilon\, b(t)\Big)\\
&\quad = 1 - \left(1 - \mathbb{P}\left(\sum_{-c(t)\le j\le -1} I(j+\xi_j > 0) > (1-2\,\varepsilon)\, b(t)\right)\right)^{k(t)}\\
&\qquad - k(t)\, \mathbb{P}\Big(\mathcal{L}(0) > (1-2\,\varepsilon)\, b(t)\Big)\, \mathbb{P}\Big(\mathcal{E}(0) > \varepsilon\, b(t)\Big),
\end{aligned} \tag{4.7}$$

where we used again the independence of $\mathcal{E}(t)$ and $\mathcal{L}(t)$, and that of disjointly indexed indicator r.v.'s for both of the last two lines displayed above.

Given (4.6), it follows that

$$\mathbb{P}\left(\sum_{-c(t)\le j\le -1} I(j+\xi_j > 0) > (1-2\,\varepsilon)\, b(t)\right) \ge t^{-(1-2\,\varepsilon)-\varepsilon^2}$$

for t sufficiently large. In view of the choice of τ, we conclude that

$$\left(1 - \mathbb{P}\left(\sum_{-c(t)\le j\le -1} I(j+\xi_j > 0) > (1-2\,\varepsilon)\, b(t)\right)\right)^{k(t)} \to 0 \tag{4.8}$$

as $t \to \infty$. On the other hand,

$$\mathbb{P}\Big(\mathcal{L}(0) > (1-2\,\varepsilon)\, b(t)\Big) \le t^{-(1-2\,\varepsilon)+\varepsilon^2}$$

for t sufficiently large. Given (4.5) and our choice of τ, we find that

$$k(t)\, \mathbb{P}\Big(\mathcal{L}(0) > (1-2\,\varepsilon)\, b(t)\Big)\, \mathbb{P}\Big(\mathcal{E}(0) > \varepsilon\, b(t)\Big) \to 0 \tag{4.9}$$

as $t \to \infty$. Relations (4.5), (4.7), (4.8) and (4.9) prove the theorem. □

Theorem 5 shows that the workload of the $S/D/1$ queue under critical loading increases very slowly (at $\log t/\big(\log\log t\big)$ rate), even in the presence of "heavy tailed" perturbations. This is in sharp contrast to the $t^{1/2}$ increase in workload that occurs under critical loading for a $G/D/1$ queue, in which the arriving traffic is described by a renewal process with finite positive variance (see Glynn [9]). This result makes clear the significant positive impact that scheduling can have upon queue performance.

5 Behavior of the $S/D/1$ workload process in heavy traffic

We now turn to the analysis of the $S/D/1$ queue when the system has more service capacity than is needed. We assume, as in Sect. 4, that work is arriving at unit rate (on average) via deterministic service time requirements of unit size, but give the server a capacity to process work at the rate $1/\rho$ with $\rho < 1$ (so that the queue's utilization factor is ρ). Let $W_\rho(\cdot)$ be the associated workload process. Then,

$$W_\rho(t) = \max_{0\le s\le t}\left[\mathcal{E}(t) - \mathcal{E}(s) - \mathcal{L}(t) + \mathcal{L}(s) + a(t) - a(s) - \frac{1-\rho}{\rho}(t-s)\right],$$

where $a(t) \overset{\Delta}{=} -(t - \lfloor t\rfloor) + I(U \le t - \lfloor t\rfloor)$. As argued in Sect. 4, $W_\rho(t) \overset{D}{=} M_\rho(t)$, where

$$M_\rho(t) = \max_{0\le r\le t}\left[\mathcal{E}^*(r) - \mathcal{L}^*(r) - \frac{1-\rho}{\rho}r + a(0) - a(-r)\right] + \mathcal{L}^*(0) - \mathcal{E}^*(0). \tag{5.1}$$

Since $M_\rho(t) \nearrow M_\rho(\infty)$ a.s. as $t\to\infty$, it follows that $W_\rho(t) \Rightarrow W_\rho(\infty)$ as $t\to\infty$, where $W_\rho(\infty) \overset{D}{=} M_\rho(\infty)$. Our key result in this section describes the "heavy traffic" behavior of $W_\rho(\infty)$ as $\rho \nearrow 1$.

Theorem 6 *Suppose that* $\mathbb{E}\xi_0^- < \infty$ *and that there exist constants* $c > 0$ *and* $\alpha > 1$ *for which* $\mathbb{P}(\xi_0 > x) \sim c\,x^{-\alpha}$ *as* $x \to \infty$*. Then,*

$$\frac{\log\log\left(\frac{1}{1-\rho}\right)}{\log\left(\frac{1}{1-\rho}\right)}\,W_\rho(\infty) \Rightarrow \frac{1}{\alpha} \tag{5.2}$$

as $\rho \nearrow 1$.

Proof Note that, for $1/2 < \rho < 1$,

$$\begin{aligned}&\max_{r\ge 0}\left[\mathcal{E}^*(r) - \mathcal{L}^*(r) - \frac{1-\rho}{\rho}r\right]\\ &\quad\ge \max_{0\le r\le 1/(1-\rho)}\left[\mathcal{E}^*(r) - \mathcal{L}^*(r) - \frac{1-\rho}{\rho}r\right]\end{aligned}$$

$$\geq \max_{0\leq r\leq 1/(1-\rho)} \Big[\mathcal{E}^*(r) - \mathcal{L}^*(r)\Big] - 2.$$

Of course, Theorem 5 establishes that

$$\frac{\log\log\left(\frac{1}{1-\rho}\right)}{\log\left(\frac{1}{1-\rho}\right)} \max_{0\leq r\leq 1/(1-\rho)} \Big[\mathcal{E}^*(r) - \mathcal{L}^*(r)\Big] \Rightarrow \frac{1}{\alpha} \tag{5.3}$$

as $\rho \nearrow 1$, proving the required lower bound for (5.2).

To prove the upper bound, observe that

$$\begin{aligned} &\max_{r\geq 0}\left[\mathcal{E}^*(r) - \mathcal{L}^*(r) - \frac{1-\rho}{\rho}\, r\right] \\ &\leq \max_{0\leq r\leq \left(\frac{1}{1-\rho}\right)^{1+\varepsilon}} \Big[\mathcal{E}^*(r) - \mathcal{L}^*(r)\Big] + \max_{r\geq \left(\frac{1}{1-\rho}\right)^{1+\varepsilon}} \left[\mathcal{E}^*(r) - \frac{1-\rho}{\rho}\, r\right]. \end{aligned} \tag{5.4}$$

Application of Theorem 5 proves that

$$\frac{\log\log\left(\frac{1}{1-\rho}\right)}{\log\left(\frac{1}{1-\rho}\right)} \max_{0\leq r\leq \left(\frac{1}{1-\rho}\right)^{1+\varepsilon}} \Big[\mathcal{E}^*(r) - \mathcal{L}^*(r)\Big] \Rightarrow \frac{1+\varepsilon}{\alpha} \tag{5.5}$$

as $\rho \nearrow 1$. On the other hand,

$$\begin{aligned} &\mathbb{P}\left(\max_{r\geq \left(\frac{1}{1-\rho}\right)^{1+\varepsilon}} \left[\mathcal{E}^*(r) - \frac{1-\rho}{\rho}\, r\right] \geq 1\right) \\ &\leq \mathbb{P}\left(\max_{n\geq 0}\left[\max_{0\leq s\leq 1} \mathcal{E}^*\left(\left(\frac{1}{1-\rho}\right)^{1+\varepsilon} + n + s\right) - \left(\frac{1}{1-\rho}\right)^{\varepsilon}\frac{1}{\rho} - \frac{1-\rho}{\rho}\, n\right] \geq 1\right) \\ &\leq \mathbb{P}\left(\max_{n\geq 1}\left[\mathcal{E}^*\left(\left(\frac{1}{1-\rho}\right)^{1+\varepsilon} + n\right) - \frac{1}{(1-\rho)^{\varepsilon}} - (1-\rho)\, n\right] \geq 0\right), \end{aligned} \tag{5.6}$$

where we used (4.4) for the last inequality. The quantity (5.6) can, in turn, be upper bounded by

$$\sum_{n=0}^{\infty} \mathbb{P}\left(\mathcal{E}^*(0) \geq \frac{1}{(1-\rho)^{\varepsilon}} + (1-\rho)\, n\right).$$

Theorem 3 proves that

$$\mathbb{P}(\mathcal{E}^*(0) \geq t) \leq \exp(-\alpha\, t)$$

for t sufficiently large, and hence the above sum is dominated by

$$\begin{aligned}
&\sum_{n=0}^{\infty} \exp\left(-\alpha\left(\frac{1}{(1-\rho)^{\varepsilon}} + (1-\rho)\, n\right)\right) \\
&= \exp\left(-\frac{\alpha}{(1-\rho)^{\varepsilon}}\right)(1-\exp(-\alpha(1-\rho)))^{-1} \\
&\sim \exp\left(-\frac{\alpha}{(1-\rho)^{\varepsilon}}\right)\left(\frac{1}{\alpha(1-\rho)}\right) \to 0
\end{aligned}$$

as $\rho \nearrow 1$. Relations (5.3), (5.4), (5.5), and (5.6) then prove the theorem, in view of the fact that ε can be made arbitrarily small. □

This $S/D/1$ heavy traffic limit theorem should be contrasted against the analogous $G/D/1$ limit theorem, for which the steady-state r.v. $W_\rho(\infty)$ scales as $1/(1-\rho)$ as $\rho \nearrow 1$; see Glynn [9]. For the $G/D/1$ queue, time scales of order $1/(1-\rho)^2$ are needed in order that fluctuations of order $1/(1-\rho)$ are exhibited (when $W_\rho(0) = 0$) (see again Glynn [9]). The proof of Theorem 6 shows that the time scale needed for W_ρ to reach equilibrium is of order $1/(1-\rho)$, so that the $S/D/1$ queue equilibrates more quickly than does the $G/D/1$ queue.

Acknowledgements The authors are very grateful to the Associate Editor and the two referees for their careful reading of the paper and for their helpful and constructive comments.

References

1. Araman, V.F., Glynn, P.W.: Fractional Brownian motion with $H<1/2$ as a limit of scheduled traffic. J. Appl. Probab. **49**(3), 1169–1188 (2012)
2. Asmussen, S.: Applied Probability and Queues. Springer, New York (2003)
3. Chen, H., Zhao, Y.J.: A new queueing model for aircraft landing process (1997). https://doi.org/10.2514/6.1997-3737
4. Chung, K.L.: A Course in Probability Theory. Academic Press, San Diego (1974)
5. Davis, B., McDonald, D.: An elementary proof of the local central limit theorem. J. Theor. Probab. **8**(3), 693–701 (1995)
6. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications. Springer, New York (1998)
7. Doi, M., Chen, Y.-M., Ōsawa, H.: A queueing model in which arrival times are scheduled. Oper. Res. Lett. **21**(5), 249–252 (1997)
8. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. II. Wiley, New York (1971)
9. Glynn, P.W.: Diffusion approximations. In: Heyman, D., Sobel, M. (eds.) Stochastic Models: Handbooks of OR & MS, vol. 2. Elsevier Science, Amsterdam (1990)
10. Hassin, R., Mendel, S.: Scheduling arrivals to queues: a single-server model with no-shows. Manag. Sci. **54**(3), 565–572 (2008)
11. Honnappa, H., Jain, R., Ward, A.: On transitory queueing. Working paper, Purdue University, West Lafayette, IN (2018)
12. Kemper, B., Klaassen, C.K.J., Mandjes, M.: Optimized appointment scheduling. EJOR **239**(1), 243–255 (2014)
13. Kingman, J.F.C.: On queues in heavy traffic. J. R. Stat. Soc. Ser. B (Methodological) **24**(2), 383–392 (1962)
14. Loynes, R.M.: The stability of a queue with non-independent inter-arrival and service times. Math. Proc. Camb. Philos. Soc. **58**(3), 497–520 (1962). https://doi.org/10.1017/S0305004100036781

15. Luo, J., Kulkarni, V.G., Ziya, S.: A tandem queueing model for an appointment-based service system. Queueing Syst. **79**(1), 53–85 (2015)
16. Mercer, A.: A queueing problem in which the arrival times of the customers are scheduled. J. R. Stat. Soc. Ser. B (Methodological) **22**(1), 108–113 (1960)
17. Mercer, A.: Queues with scheduled arrivals: a correction, simplification and extension. J. R. Stat. Soc. Ser. B (Methodological) **35**(1), 104–116 (1973)
18. Ross, S.M.: Stochastic Processes, 2nd Edition John Wiley and Sons (1996)
19. Ross, S.M.: A First Course in Probability. Pearson Prentice Hall, Upper Saddle River (2015)
20. Winsten, C.B.: Geometric distributions in the theory of queues. J. R. Stat. Soc. Ser. B (Methodological) **21**(1), 1–35 (1959)
21. Zacharias, C., Armony, M.: Joint panel sizing and and appointment scheduling in outpatient care. Manag. Sci. **63**(11), 3978–3997 (2017)