Queueing Systems 2(1986)191 – 215

A CENTRAL-LIMIT-THEOREM VERSION OF $L = \lambda W$

P.W. GLYNN*

Department of Industrial Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

and

W. WHITT

AT & T Bell Laboratories, Holmdel, New Jersey 07733, USA

Received 10 September 1985
(Revised 2 May 1986)

Abstract

Underlying the fundamental queueing formula $L = \lambda W$ is a relation between cumulative processes in continuous time (the integral of the queue length process) and in discrete time (the sum of the waiting times of successive customers). Except for remainder terms which usually are asymptotically negligible, each cumulative process is a random time-transformation of the other. As a consequence, in addition to the familiar relation between the with-probability-one limits of the averages, roughly speaking, the customer-average wait obeys a central limit theorem if and only if the time-average queue length obeys a central limit theorem, in which case both averages, properly normalized, converge in distribution jointly, and the individual limiting distributions are simply related. This relation between the central limit theorems is conveniently expressed in terms of functional central limit theorems, using the continuous mapping theorem and related arguments. The central limit theorems can be applied to compare the asymptotic efficiency of different estimators of queueing parameters. For example, when the arrival rate λ is known and the interarrival times and waiting times are negatively correlated, it is more asymptotically efficient to estimate the long-run time-average queue length L indirectly by the sample-average of the waiting times, invoking $L = \lambda W$, than it is to estimate it by the sample-average of the queue length. This variance-reduction principle extends a corresponding result for the standard GI/G/s model established by Carson and Law [2].

Keywords

Queueing theory, conservation laws, Little's law, limit theorems, statistical estimation, simulation.

*Supported by the National Science Foundation under Grant No. ECS-8404809 and by the U.S. Army under Contract No. DAAG29-80-C-0041.

1. Introduction and summary

A fundamental principle of queueing theory is expressed by the formula $L = \lambda W$, which states in part that the time-average queue length L is equal to the product of the arrival rate λ and the customer-average wait in queue W. The principle also provides conditions for these averages to exist as limits; see [6,9,10] and references therein. The primary purpose of this paper is to provide a CLT (central limit theorem) refinement. Roughly speaking, we show that the customer-average wait obeys a CLT if and only if the time-average queue length obeys a CLT, and we relate the two limits. Just as the classical CLTs can be regarded as refinements of the classical LLNs (laws of large numbers), so our result is a refinement of the standard relation between the w.p.1 (with probability one) limits of the averages.

In order to relate the CLT behavior of the time averages and customer averages, in this paper we work with FCLTs (functional central limit theorems) instead of ordinary CLTs, using the theory of weak convergence of probability measures on the function space $D \equiv D[0, \infty)$; see [1,8,14] and references therein. Relations among the FCLTs are established using the continuous mapping theorem (theorem 5.1 of [1]) and related mapping arguments for basic functions on D or $D \times D$ such as composition, addition, composition plus translation, supremum and inverse [14].

A key idea here is to replace specific conditions on the stochastic process of interest (such as finite moments, stationarity, metric transitivity, independence, ϕ-mixing, the Markov property or the regenerative property) by the existence of other related limits. The FCLT setting is especially convenient for this purpose. It facilitates the same kind of sample-path analysis used to establish $L = \lambda W$ in [9] and [10]. This same approach has been used to establish heavy-traffic limit theorems for queues given FCLTs for counting processes associated with the arrivals and service times; theorem 1 of [5]. It has also been used to relate limits for stochastic processes to limits for their discounted counterparts [13]. Of particular relevance here, this approach has been used to establish CLTs and FCLTs for random sums of random variables; sect. 17 of [1], [4,8,14].

The FCLTs here are useful, not only to describe the averages (which are congestion measures of interest in their own right), but also to do statistical estimation of queueing system parameters by computer simulation or by direct system measurement. For example, if we want to estimate the expected waiting time, then we can either estimate it in the natural way using the usual customer average or we can estimate it indirectly via the time average of the queue-length process, invoking $L = \lambda W$. It is natural to ask which way provides a more efficient estimator. This question has been investigated for the GI/G/s queue by Carson and Law [2] using the regenerative property. The FCLTs here provide a basis for making such comparisons for more general systems.

The rest of the paper is organized as follows. In sect. 2 we review the basic relationship between cumulative processes underlying $L = \lambda W$. We also present a new variant of $L = \lambda W$, the relation between the (strong) SLLNs, aimed at producing something closer to a symmetric version. The existing versions go from λ and W to L. We indicate how to go the other way, by adding an extra condition. (It is well known that an extra condition is needed.) We also relate $L = \lambda W$ to limits for other processes. From theorem 2 it is easy to see that the FCLT versions of $L = \lambda W$ in sects. 3 and 4 can be viewed as direct FCLT analogs of the standard SLLN versions of $L = \lambda W$.

In sect. 3 we establish the equivalence of the customer-average and time-average FCLTs under the condition that a remainder term is asymptotically negligible. In sect. 4 we show, paralleling the standard version of $L = \lambda W$, that a joint FCLT for the averages related to λ and W by itself knocks out the remainder term. Our main result, theorem 4, establishes joint convergence for 15 related processes. In sect. 5 we discuss implications for parameter estimation from simulation or system measurements. We conclude in sect. 6 by discussing functional-law-of-the-iterated logarithm (FLIL) versions of $L = \lambda W$, which can be established by almost the same arguments as for the FCLT versions.

We have also established related results that will be presented elsewhere. We have established ordinary-central-limit-theorem (CLT) versions of $L = \lambda W$ by very different arguments. Since both the conditions and the conclusions are weaker than the FCLTs here, neither set of results contains the other. We have also established ordinary law-of-the-iterated-logarithm (LIL) versions of $L = \lambda W$. We have also extended the FCLT results here to the setting of $H = \lambda G$ and generalizations such as [7]. Finally, we have further investigated the statistical applications discussed in sect. 5.

2. A basic relationship between cumulative processes

The standard $L = \lambda W$ framework is a sequence of ordered pairs of random variables $\{(A_k, D_k), k = 1, 2, \ldots\}$, where $0 \leqslant A_k \leqslant A_{k+1}$ and $A_k \leqslant D_k$ for all k w.p.1. As in [9] and [10], the results in this section are for individual sample paths of the stochastic process $\{(A_k, D_k), k \geqslant 1\}$. In applications, the limits will typically hold for a set of sample paths having probability one.

We usually interpret A_k and D_k as the arrival and departure epochs of the kth arriving customer. (However, arrival and departure should be interpreted with respect to the system under consideration. For example, if the system refers to a queue, excluding the servers, then D_k is the epoch when the kth customer leaves the queue, which usually occurs when the customer begins service.) As regularity conditions, we assume that D_k is finite and that there are only finitely many arrivals in finite time, i.e. $A_k \to \infty$ as $k \to \infty$. For many models, such as one single-server queue with the first-come first-served discipline, we also have $D_k \leqslant D_{k+1}$ for all k, but we do *not* assume that the sequence $\{D_k\}$ is nondecreasing. We think of the system as initially empty, but other initial conditions can be introduced by letting $A_j = 0$, $1 \leqslant j \leqslant k$, for some k. The sequence $\{(A_k, D_k)\}$ is our complete model specification.

We think of the kth customer as being in the system during the interval $[A_k, D_k]$, so that we let the queue length at time t, $Q(t)$, be the number of k with $A_k \leqslant t \leqslant D_k$, and we let the waiting time of the kth customer be $W_k = D_k - A_k$. It is convenient to work with the indicator function of the interval $[A_k, D_k]$, defined by $I_k(t) = 1$ if $A_k \leqslant t \leqslant D_k$ and 0 otherwise. Then $Q(t)$ and W_k are defined by

$$Q(t) = \sum_{k=1}^{\infty} I_k(t), t \geqslant 0, \quad \text{and} \quad W_k = \int_0^{\infty} I_k(t)\, dt, k \geqslant 1. \tag{2.1}$$

Let $N(t)$ be the arrival counting process, defined from the arrival sequence $\{A_k\}$ by $N(t) = \max\{k \geqslant 0: A_k \leqslant t\}$, $t \geqslant 0$, where we understand $A_0 = 0$ without having a 0*th* customer. The sequence $\{A_k, k \geqslant 1\}$ and the process $\{N(t), t \geqslant 0\}$ can be regarded as inverse processes because they satisfy the basic relation $A_k \leqslant t$ if and only if $N(t) \geqslant k$.

We focus on the *cumulative processes* associated with W_k and $Q(t)$, namely, $\{\Sigma_{j=1}^k W_j, k \geqslant 1\}$ and $\{\int_0^t Q(s)\, ds, t \geqslant 0\}$, and the closely related *customer average* $\{k^{-1} \Sigma_{j=1}^k W_j, k \geqslant 1\}$ and *time average* $\{t^{-1} \int_0^t Q(s)\, ds, t > 0\}$. There is a basic relationship among the four processes $\{\Sigma_{j=0}^k W_j, k \geqslant 1\}$, $\{\int_0^t Q(s)\, ds, t \geqslant 0\}$, $\{A_k, k \geqslant 1\}$ and $\{N(t), t \geqslant 0\}$ that underlies Little's formula and its refinements. The basic relationship indicates that the cumulative process of $Q(t)$ (W_k) is equal to the cumulative process of W_k ($Q(t)$) evaluated at the random time $N(t)$ (A_k), plus a remainder term $R(t)$ (S_k). This is perhaps best understood by a picture (see fig. 1). There, time is on the x-axis and the customer index is on the y-axis. A bar appears in $[k-1, k] \times [A_k, D_k]$ for customer k, so that W_k measures the length of the kth bar and $Q(t)$ counts the number of bars intersecting the line $x = t$. Both cumulative processes thus measure the area of a set of bars and partial bars. As t or k become large, the partial bars typically become relatively negligible, so that the two cumulative processes are measuring approximately the same thing. Of course, we need to relate t and k, the obvious way being to let k be $N(t)$ when t is given, and to let t be A_k when k is given. In fig. 1, $\int_0^{t_0} Q(s)\, ds$ and $\Sigma_{k=1}^{17} W_k$ are depicted where $N(t_0) = 17$.

It is also possible and useful to *bound* each cumulative process by the other cumulative process evaluated at random times related to the arrivals and departures. For this purpose, let $O(t)$ be the output counting process, depicting the number of k for which $D_k \leqslant t$, and let $D'_k = \max_{1 \leqslant j \leqslant k} D_j$. We omit the proof, which is similar to those in [9] and [10]. (More details appear in a previous draft.)

THEOREM 1

The cumulative processes are related by

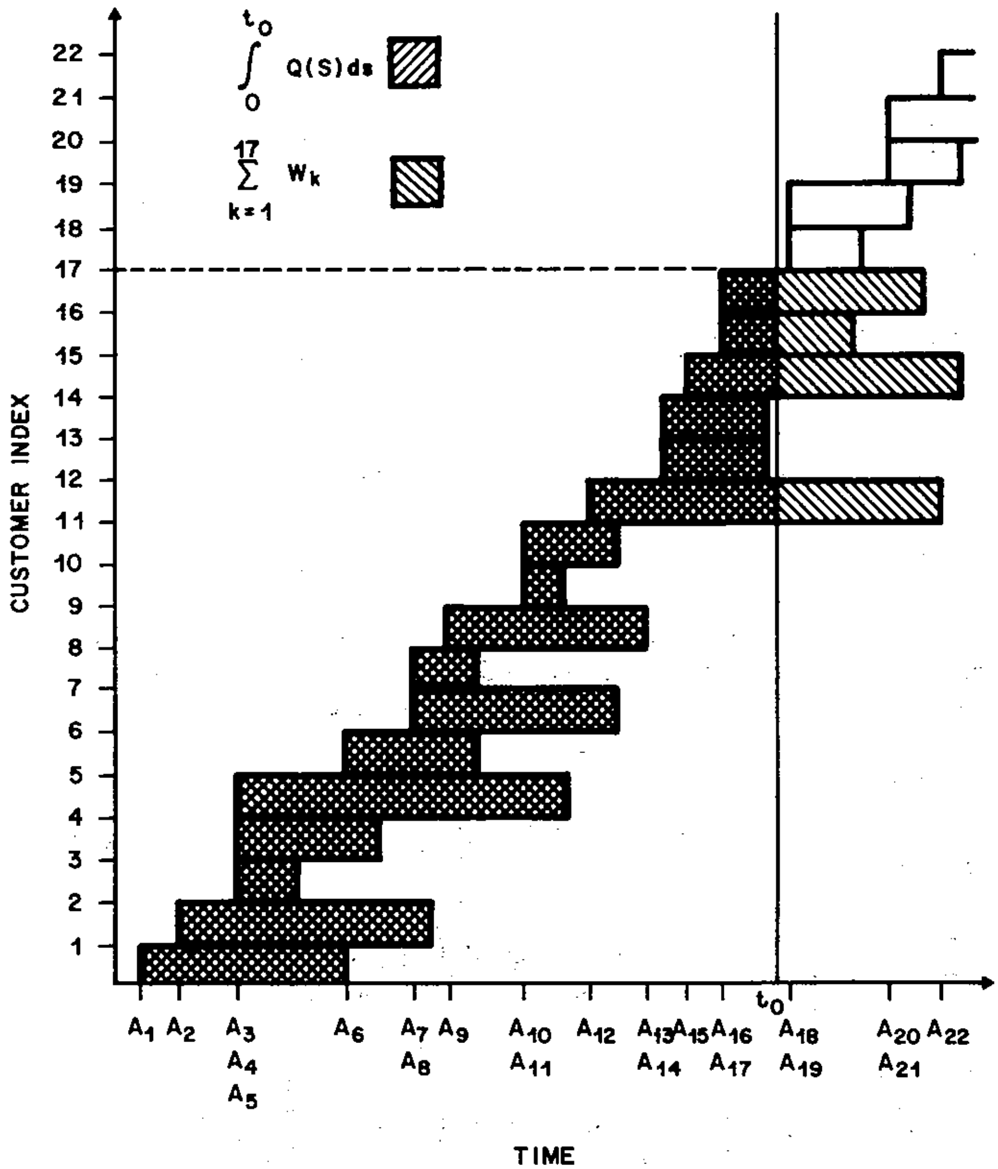

Fig. 1. The cumulative processes associated with $Q(t)$ and W_k.

(a) $$\sum_{k=1}^{O(t)} W_k \leqslant \sum_{k=1}^{N(t)} W_k - R(t) = \int_0^t Q(s)\,\mathrm{d}s \leqslant \sum_{k=1}^{N(t)} W_k \qquad \text{for all } t \geqslant 0,$$

$$\int_0^{A_k} Q(s)\,\mathrm{d}s \leqslant \sum_{j=1}^{k} W_j = \int_0^{A_k} Q(s)\,\mathrm{d}s + S_k \leqslant \int_0^{D_k'} Q(s)\,\mathrm{d}s \qquad \text{for all } k \geqslant 1,$$

where

$$R(t) = \sum_{k=1}^{N(t)} \int_t^{\infty} I_k(s)\,\mathrm{d}s, \; t \geqslant 0, \quad \text{and} \quad S_k = \sum_{j=1}^{k} \int_{A_k}^{\infty} I_j(s)\,\mathrm{d}s, \; k \geqslant 1; \tag{2.2}$$

(b) $S_k \leqslant R(A_k)$ and $R(t) \leqslant S_{N(t)}$ for all k and t;

(c) If $\{A_k, k \geqslant 1\}$ is strictly increasing, then $R(A_k) = S_k$.

An elementary consequence of theorem 1 is that the remainder terms vanish whenever the system is empty (where $N(t) = O(t)$ and $D_k' \leqslant t \leqslant A_{k+1}$). This yields inequalities for the lim infs and lim sups when the system is empty infinitely often.

COROLLARY 1.1

If there exists a sequence $\{t_k : k \geqslant 1\}$ with $t_k \to \infty$ as $k \to \infty$ such that $Q(t_k) = 0$ for all k, then

$$\varliminf_{t \to \infty} t^{-1} \int_0^t Q(s)\,\mathrm{d}s \leqslant \left(\varlimsup_{t \to \infty} t^{-1} N(t)\right)\left(\varliminf_{n \to \infty} n^{-1} \sum_{k=1}^{n} W_k\right)$$

$$\varlimsup_{t \to \infty} t^{-1} \int_0^t Q(s)\,\mathrm{d}s \geqslant \left(\varliminf_{t \to \infty} t^{-1} N(t)\right)\left(\varlimsup_{n \to \infty} n^{-1} \sum_{k=1}^{n} W_k\right).$$

There are two aspects to $L = \lambda W$: (i) establishing that the normalized remainder terms $t^{-1}R(t)$ and $k^{-1}S_k$ become negligible as $t \to \infty$ and $k \to \infty$, and (ii) exploiting the remaining relation when the remainder terms are neglected. The practical value is almost entirely in the second aspect; the first aspect is primarily a technical detail. It is customary to treat the two aspects together [9], but this combined treatment causes a certain unpleasant asymmetry to appear because the very existence of one pair of limits is sufficient to knock out the remainder term, while the existence of the other pair is not. The following elementary example shows that going from limits related to λ and L to a limit for W is not possible without adding an extra condition.

Example 1

Let $A_k = k$ for all k, corresponding to deterministic arrivals. Let $D_{2^k} = 2^{k+1}$ and $D_j = j$ for $j \neq 2^k, j \geq 1$ and $k \geq 0$. All customers depart immediately upon arrival except the customers arriving at epochs 2^k. The customer arriving at 2^k replaces the arrival at 2^{k-1}, so that there is always one customer in the system (not counting instants of arrival and departure). Obviously,

$$\lim_{k \to \infty} k^{-1} A_k = \lim_{t \to \infty} t^{-1} N(t) = 1 = \lim_{t \to \infty} t^{-1} \int_0^t Q(s)\,\mathrm{d}s,$$

but

$$\sum_{j=1}^{2^k - 1} W_j = 2^k \quad \text{and} \quad \sum_{j=1}^{2^k} W_j = 2^{k+1}$$

so that

$$\left\{ n^{-1} \sum_{j=1}^{n} W_j,\ n \geq 1 \right\}$$

does not converge as $n \to \infty$. It has every point in the interval [1,2] as a limit of some subsequence. However, we can apply corollary 1.1 to see that

$$\varliminf_{n \to \infty} n^{-1} \sum_{k=1}^{n} W_k \leq \lim_{t \to \infty} t^{-1} \int_0^t Q(s)\,\mathrm{d}s = 1 \leq \varlimsup_{t \to \infty} n^{-1} \sum_{k=1}^{n} W_k. \qquad \square$$

We now present a new result that is closer to a symmetric statement. We also relate more of the relevant limits. For this purpose, consider the following limits, all of which apply to individual sample paths of $\{(A_k, D_k)\}$. At this point, we change the notation somewhat. We use q and w for the limits of the averages instead of L and W.

$$
\begin{array}{llll}
\text{(i)} & \lim_{n\to\infty} n^{-1}A_n = \lambda^{-1},\ 0 < \lambda^{-1} < \infty, & \text{(ii)} & \lim_{t\to\infty} t^{-1}N(t) = \lambda,\ 0 < \lambda < \infty, \\
\text{(iii)} & \lim_{n\to\infty} n^{-1}D_n = \lambda^{-1}, & \text{(iv)} & \lim_{n\to\infty} n^{-1}W_n = 0, \\
\text{(v)} & \lim_{n\to\infty} n^{-1}D'_n = \lambda^{-1}, & \text{(vi)} & \lim_{t\to\infty} t^{-1}O(t) = \lambda, \\
\text{(vii)} & \lim_{t\to\infty} t^{-1}Q(t) = 0, & \text{(viii)} & \lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} W_k = w, \\
\text{(ix)} & \lim_{t\to\infty} t^{-1}\int_0^t Q(s)\,\mathrm{d}s = q, & \text{(x)} & \lim_{n\to\infty} n^{-1}S_n = 0, \\
\text{(xi)} & \lim_{t\to\infty} t^{-1}R(t) = 0, & \text{(xii)} & \lim_{n\to\infty} n^{-1}\int_0^{A_n} Q(s)\,\mathrm{d}s = w, \\
\text{(xiii)} & \lim_{t\to\infty} t^{-1}\sum_{k=1}^{N(t)} W_k = q, & \text{(xiv)} & \lim_{n\to\infty} n^{-1}\int_0^{D'_n} Q(s)\,\mathrm{d}s = w, \\
\text{(xv)} & \lim_{t\to\infty} t^{-1}\sum_{k=1}^{O(t)} W_k = q. & &
\end{array}
\tag{2.3}
$$

The standard statement is that limits (i) and (viii) imply (ix), and if the limits hold, then $q = \lambda w$. Usually, q, λ and w are constants (nonrandom), but they could be random variables; then the relationship $q = \lambda w$ holds for each sample path. Part (e) below offers something close to a symmetric version.

THEOREM 2

The limits in (2.3) are related as follows:

(a) (i) holds if and only if (ii) holds. Henceforth, assume that they hold.

(b) (iii), (iv) and (v) are equivalent.

(c) (vi) and (vii) are equivalent.

(d) (iii) implies (vi), but (vi) does not imply (iii).

(e) (viii) holds if and only if (iii) and (ix) hold, in which case $q = \lambda w$.

(f) If (viii) holds, then so do limits (x)–(xv).

(g) (x) and (xi) are equivalent.

(h) Under (ix), any one of (iii)–(v), (x) and (xi) implies all other limits.

Proof

Part (a) is an elementary well-known consequence of the inverse relationship. Given (i), (iii) is equivalent to (iv) because $W_n = D_n - A_n$. Obviously, (i) and (v) imply (iii) because $A_n \leqslant D_n \leqslant D_n'$ for all n. We show that (i) and (iii) imply (v) by showing that, for any positive ϵ, there is an $n_0(\epsilon)$ such that $D_n' \leqslant (1 + \epsilon) A_n$ for $n \geqslant n_0(\epsilon)$. For ϵ given, we obtain $n_0(\epsilon)$ from $A_n^{-1} W_n \to 0$ as $n \to \infty$, which holds by virtue of (i) and (iii). Hence, there is $n_0(\epsilon)$ such that $W_n \leqslant \epsilon A_n$ for $n \geqslant n_0(\epsilon)$. This in turn implies that $D_n \leqslant (1 + \epsilon)A_n$ for $n \geqslant n_0(\epsilon)$. Obviously, $n^{-1} D'_{n_0(\epsilon)} \to 0$ as $n \to \infty$, so that there exists $n_1(\epsilon) \geqslant n_0(\epsilon)$ such that $D_n' \leqslant (1 + \epsilon)A_n$ for $n \geqslant n_1(\epsilon)$. Since ϵ is arbitrary (v) indeed follows from (iii). Part (c) is immediate since $Q(t) = N(t) - O(t)$.

For part (d), (iii) $\to$ (vi) is similar to (iii) $\to$ (v): Again, by (i) and (iii), for arbitrary $\epsilon > 0$, there exists $n_0(\epsilon)$ such that $W_n < \epsilon A_n$ for $n \geqslant n_0(\epsilon)$. Using this ϵ and $n_0(\epsilon)$, we have the set inclusions

$$\begin{aligned}\{n: D_n \leqslant t\} &\supseteq \{n: A_n + W_n \leqslant t\} \cap \{n: n \geqslant n_0(\epsilon)\}\\ &\supseteq \{n: A_n \leqslant t/(1+\epsilon)\} \cap \{n: n \geqslant n_0(\epsilon)\}\\ &\supseteq \{n: N(t/(1+\epsilon)) \geqslant n\} \cap \{n: n \geqslant n_0(\epsilon)\},\end{aligned}$$

so that $O(t) \geqslant N(t/(1 + \epsilon)) - n_0(\epsilon)$ for all $t \geqslant 0$. Since $O(t) \leqslant N(t)$ for all t and $t^{-1} N(t) \to \lambda$ as $t \to \infty$,

$$\lim_{t\to\infty} t^{-1} N(t) \geqslant \overline{\lim_{t\to\infty}}\, t^{-1} O(t) \geqslant \underline{\lim_{k\to\infty}}\, t^{-1} O(t) \geqslant \frac{1}{1+\epsilon} \lim_{t\to\infty} t^{-1} N(t).$$

Since ϵ was arbitrary, $t^{-1} O(t) \to \lambda$ as $t \to \infty$.

For the second half of (d), apply example 1 to see that (i) and (vi) do not imply (iii): $W_{2k} = 2^k$ for all k there.

We obtain (e) and (f) together using theorem 1. The implication (viii) $\to$ (ix) is provided by applying theorem 1 (a) plus (c) above:

$$\lambda w = \lim_{t \to \infty} \left(\frac{O(t)}{t}\right) \frac{1}{O(t)} \sum_{k=1}^{O(t)} W_k \leqslant \varliminf_{t \to \infty} t^{-1} \int_0^t Q(s)\, \mathrm{d}s$$

$$\leqslant \varlimsup_{t \to \infty} t^{-1} \int_0^t Q(s)\, \mathrm{d}s \leqslant \lim_{t \to \infty} \left(\frac{N(t)}{t}\right) \frac{1}{N(t)} \sum_{k=1}^{N(t)} W_k = \lambda w.$$

Since (viii) → (iv), ((i), (viii)) → (iii) too by part (b). Finally, ((i), (iii), (ix)) → (viii) is proved by applying (b) above and theorem 1(a). From (i) and (iii), we obtain (v), so that

$$\lambda^{-1} q = \lim_{k \to \infty} \left(\frac{A_k}{k}\right) A_k^{-1} \int_0^{A_k} Q(s)\, \mathrm{d}s \leqslant \varliminf_{k \to \infty} k^{-1} \sum_{j=1}^{k} W_j$$

$$\leqslant \varlimsup_{k \to \infty} k^{-1} \sum_{j=1}^{k} W_j \leqslant \lim_{k \to \infty} \left(\frac{D_k'}{k}\right) (D_k')^{-1} \int_0^{D_k'} Q(s)\, \mathrm{d}s = \lambda^{-1} q.$$

For (g), apply (a) and theorem 1(b). For (h), apply theorem 1(a) to show that (ix) and (x) imply (viii). The rest follows from (b)–(f). □

3. A central-limit-theorem refinement

We work in the setting of [1] and [14], which means weak convergence (convergence in distribution), denoted by $\Rightarrow$. We often write $X_n \Rightarrow X$, omitting "as $n \to \infty$" when this is obvious. We consider random elements of $D \equiv D[0, \infty)$, the space of all real-valued functions on $[0, \infty)$ which are right-continuous with left limits. Let the space D be endowed with the standard Skorohod (J_1) topology and let product spaces D^k be endowed with the usual product topology. Let $C \equiv C[0, \infty)$ be the subset of continuous functions in D. Convergence $x_n \to x$ in D with the Skorohod topology reduces to uniform convergence on compact subsets when $x \in C$.

We define the following random functions in D. Let $[x]$ be the integer part of x.

$$A_n(t) = n^{-1/2} [A_{[nt]} - \mu n t], \qquad N_n(t) = n^{-1/2} [N(nt) - \lambda n t],$$

$$W_n(t) = n^{-1/2} \left[\sum_{k=1}^{[nt]} W_k - w n t \right], \quad Q_n(t) = n^{-1/2} \left[\int_0^{nt} Q(s)\, \mathrm{d}s - q n t \right],$$

$$(QA)_n(t) = n^{-1/2}\left[\int_0^{A[nt]} Q(s)\,ds - \lambda^{-1} qnt\right],$$

$$(WN)_n(t) = n^{-1/2}\left[\sum_{k=1}^{N(nt)} W_k - \lambda wnt\right], \tag{3.1}$$

$$R_n(t) = n^{-1/2}R(nt),\ \ S_n(t) = n^{-1/2}S_{[nt]},\ \ \boldsymbol{\theta}(t) = 0 \text{ and } e(t) = t$$

for $t \geqslant 0$, where here μ, λ, w and q are positive real numbers. Note that the process $Q(t)$ defined in sect. 2 need not be a random element of D, but the associated cumulative process $\int_0^t Q(s)\,ds$ is, as are all random functions in (3.1). Let $\stackrel{d}{=}$ denote equality in distribution.

THEOREM 3

Suppose that $\mu = \lambda^{-1}$, $q = \lambda w$, and one of the remainder terms is asymptotically negligible, i.e. that either $R_n \Rightarrow \boldsymbol{\theta}$ or $S_n \Rightarrow \boldsymbol{\theta}$. If any one of the following four weak convergence limits in $D \times D$ holds:

$(W_n, A_n) \Rightarrow (W, A)$ where $P(A \in C) = 1$,

$(W_n, N_n) \Rightarrow (W, N)$ where $P(N \in C) = 1$,

$(Q_n, A_n) \Rightarrow (Q, A)$ where $P(A \in C) = 1$, or

$(Q_n, N_n) \Rightarrow (Q, N)$ where $P(N \in C) = 1$,

then the other three do too and, in addition, there is the joint convergence

$$(W_n, Q_n, A_n, N_n, (QA)_n, (WN)_n, R_n, S_n) \Rightarrow (W, Q, A, N, W, Q, \boldsymbol{\theta}, \boldsymbol{\theta}) \text{ in } D^8,$$

in which case

$$N(t) = -\lambda A(\lambda t) \stackrel{d}{=} -\lambda^{3/2} A(t)$$

$$Q(t) = W(\lambda t) - qA(\lambda t) \stackrel{d}{=} \lambda^{1/2}(W(t) - qA(t)), \quad t \geqslant 0. \tag{3.2}$$

If, in addition, one of (i)–(iv) holds with two-dimensional Brownian motion as a limit, then the joint limit process (A, W, N, Q) is also Brownian motion; see (3.6) and (3.7) for the covariances.

Proof

First, from theorem 7.3 plus the corollary to lemma 7.6 of [14], if either $A_n \Rightarrow A$ with $P(A \in C) = 1$ or $N_n \Rightarrow N$ with $P(N \in C) = 1$, then both hold and, moreover, there is the joint convergence $(A_n, N_n) \Rightarrow (A, N)$, where $N(t) = -\lambda A(\lambda t)$, $t \geq 0$. Moreover, $\Phi_n \Rightarrow \lambda^{-1} e$ and $\Psi_n \Rightarrow \lambda e$, where

$$\Phi_n(t) = n^{-1} A_{[nt]} \quad \text{and} \quad \Psi_n(t) = n^{-1} N(nt), \qquad t \geq 0. \tag{3.3}$$

To apply the corollary to lemma 7.6 in [14], make the translation

$$x_n(t) = \Phi_n(t), x_n^{-1}(t) = \Psi_n(t) + n^{-1}, \ c_n = n^{-1/2}, \ a = \lambda^{-1} .$$

The extra n^{-1} in $x_n^{-1}(t)$ above is asymptotically negligible; it can be neglected by theorem 4.1 of [1].

From theorem 1(b), $S_n(t) \leq (R_n \circ \Phi_n)(t)$ and $R_n(t) \leq (S_n \circ \Psi_n)(t)$ for all t, where $\circ$ denotes the composition map, i.e. $(x \circ y)(t) = x(y(t))$, $t \geq 0$. Hence, if either $R_n \Rightarrow \theta$ or $S_n \Rightarrow \theta$, then both do. To justify this, apply theorem 4.4 of [1] to get joint convergence $(R_n, \Phi_n) \Rightarrow (\theta, \lambda^{-1} e)$ or $(S_n, \Psi_n) \Rightarrow (\theta, \lambda e)$ and then apply the continuous mapping theorem with composition, theorem 3.1 of [14]. Since $R_n \Rightarrow \theta$ or $S_n \Rightarrow \theta$ by hypothesis, both do and we can ignore the remainder term in the convergence; we invoke theorem 4.1 of [1].

To treat the random sums $(WN)_n$ and $(QA)_n$, we apply theorem 5.1(i) of [14] twice. (Note that there part (ii) does not apply because Φ_n and Ψ_n in (3.3) do not have continuous paths and part (iii) involves a complicated condition on the fluctuations via the function $F(x_n, y_n; a, b)$.) For the details, first suppose that we start with $(W_n, A_n) \Rightarrow (W, A)$ or $(W_n, N_n) \Rightarrow (W, N)$. Then, using the inverse map as described at the beginning of this proof and theorem 4.4 of [1], we obtain $(W_n, A_n, N_n, \Psi_n) \Rightarrow (W, A, N, \lambda e)$ in D^4. We then get

$$(W_n, A_n, N_n, \Psi_n, (WN)_n) \Rightarrow (W, A, N, \lambda e, Q) \text{ in } D^5$$

from theorem 5.1(i) of [14]. To make the translation, let

$$x_n(t) = w^{-1} n^{-1/2} \sum_{k=0}^{[nt]} W_k, \ c_n = n^{1/2}, \ y_n(t) = \Psi_n(t), \ y(t) = \lambda t, \tag{3.4}$$

$$b_n = \lambda, x(t) = w^{-1} W(t), \ z(t) = N(t).$$

Afterwards, multiply by w in (3.4) to get the limit for $(WN)_n$ in (3.1). Since $R_n \Rightarrow \mathbf{\theta}$, $\rho(Q_n^c, (WN)_n) \Rightarrow 0$, where ρ is a metric on D inducing uniform convergence on compact subsets, so that $Q_n \Rightarrow W \circ \lambda e + wN$ by theorem 4.1 of [1]. More generally,

$$\rho((W_n, A_n, N_n, \Psi_n, (WN)_n), (W_n, A_n, N_n, \Psi_n, Q_n)) \Rightarrow 0,$$

where ρ is a corresponding metric on D^5 inducing the product topology. Hence, we have the claimed joint convergence. The argument is essentially the same starting with $(Q_n, A_n) \Rightarrow (Q, A)$ or $(Q_n, N_n) \Rightarrow (Q, N)$.

To check consistency of the limits in (3.2), observe that

$$\begin{aligned} W &= Q \circ \lambda^{-1} e + qA \\ &= [W \circ \lambda e + wN] \circ \lambda^{-1} e + qA = W \circ \lambda e \circ \lambda^{-1} e + wN \circ \lambda^{-1} e + qA \\ &= W + w[-\lambda A \circ \lambda e] \circ \lambda^{-1} e + qA = W - \lambda w A + qA = W. \end{aligned} \tag{3.5}$$

Finally, to verify the equality of distribution in (3.2), just change the time scale in (3.1): If $W_n \Rightarrow W$, then $W_{cn} \Rightarrow W$ and $W_n \circ ce \Rightarrow W \circ ce$, but $W_{cn} = c^{-1/2} W_n \circ ce$, so that indeed $W \stackrel{d}{=} c^{-1/2} W \circ ce$. □

COROLLARY 3.1

If, in addition to the assumptions of theorem 3, the limit processes $W(t)$ and $Q(t)$ have no jumps at $t = 1$ w.p.1, then there is the ordinary CLT

$$\begin{aligned} &(W_n(1), Q_n(1), A_n(1), N_n(1), (QA)_n(1), (WN)_n(1), R_n(1), S_n(1)) \\ &\Rightarrow (W(1), Q(1), A(1), N(1), W(1), Q(1), 0, 0) \text{ in } R^8. \end{aligned}$$

Furthermore, the joint limit $(A(1), W(1))$ has a bivariate normal distribution with zero means and covariance matrix Σ if and only if $(N(1), Q(1))$ has a bivariate normal distribution with zero means and covariance matrix $\hat{\Sigma}$ where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 \end{pmatrix}, \text{ and } \begin{aligned} \hat{\sigma}_1^2 &= \lambda^3 \sigma_1^2 \\ \hat{\sigma}_{12} &= \lambda^2 (q\sigma_1^2 - \sigma_{12}) \\ \hat{\sigma}_2^2 &= \lambda(q^2 \sigma_1^2 - 2q\sigma_{12} + \sigma_2^2), \end{aligned} \tag{3.6}$$

in which case $(A(1), W(1), N(1), Q(1))$ has a multivariate normal distribution with

$$\mathrm{Cov}(A(1), N(1)) = -(\lambda \wedge 1)\,\lambda\sigma_1^2,$$

$$\mathrm{Cov}(A(1), Q(1)) = (\lambda \wedge 1)\,(q^2\sigma_1^2 - 2q\sigma_{12} + \sigma_2^2),$$

$$\mathrm{Cov}(W(1), N(1)) = -(\lambda \wedge 1)\,\lambda\sigma_{12}, \quad \text{and}$$

$$\mathrm{Cov}(W(1), Q(1)) = (\lambda \wedge 1)\,(\sigma_2^2 - q\sigma_{12}), \tag{3.7}$$

where $\lambda \wedge 1 = \min\{\lambda, 1\}$.

Proof

Apply the continuous mapping theorem with the projection map $\pi_1 : D^8 \to R^8$, defined by $\pi_1((x_1, \ldots, x_8)) = (x_1(1), \ldots, x_8(1))$. □

COROLLARY 3.2

Under the assumptions of theorem 3, there is convergence in probability:

$$n^{-1}A_n \xrightarrow{\mathrm{p}} \lambda^{-1} \text{ as } n \to \infty, \; t^{-1}N(t) \xrightarrow{\mathrm{p}} \lambda \text{ as } t \to \infty,$$

$$n^{-1}\sum_{j=1}^{n} W_j \xrightarrow{\mathrm{p}} w \text{ as } n \to \infty, \; t^{-1}\int_0^t Q(s)\,\mathrm{d}s \xrightarrow{\mathrm{p}} q \text{ as } t \to \infty,$$

$$t^{-1}\sum_{j=1}^{N(t)} W_j \xrightarrow{\mathrm{p}} q \text{ as } t \to \infty, \; n^{-1}\int_0^{A_n} Q(s)\,\mathrm{d}s \xrightarrow{\mathrm{p}} w \text{ as } n \to \infty,$$

$$t^{-1}R(t) \xrightarrow{\mathrm{p}} 0 \text{ as } t \to \infty \quad \text{and} \quad n^{-1}S_n \xrightarrow{\mathrm{p}} 0 \text{ as } n \to \infty.$$

Remarks

(3.1) The relations $\mu = \lambda^{-1}$ and $q = \lambda w$ have been assumed in theorem 3. By the convergence-of-types theorem, ([3], p. 253), no other translation constants can be used.

(3.2) Neither the almost-sure convergence in theorem 2 nor the weak convergence in theorem 3 and corollary 3.1 directly imply the other. Of course, theorem 2 yields corollary 3.2 too.

(3.3) As explained in [14], even though the FCLT statements are expressed in terms of weak convergence, both the statements and the proofs can be expressed in terms of limits for individual sample paths. Thus, the FCLT relations can be regarded as natural extensions of the sample-path analysis underlying sect. 2.

(3.4) Despite the fact that (A_k, W_k) represents a discrete-time process while $(N(t), Q(t))$ represents a continuous-time process, there is a natural symmetry between the two, as far as the asymptotic behavior is concerned. In particular, let f be the function mapping the basic parameter five-tuple $(\lambda^{-1}, w, \sigma_1^2, \sigma_2^2, \sigma_{12})$ associated with the asymptotic behavior of (A_n, W_n) into the corresponding parameter five-tuple $(\lambda, q, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_{12})$ associated with the asymptotic behavior of (N_n, Q_n). From (3.6), it is easy to see that f is invertible and $f^{-1} = f$ provided that $0 < \lambda < \infty$; i.e. $f \circ f$ is the identity map. From (3.2) we see that this is also true for the function taking (λ^{-1}, w, A, W) into (λ, q, N, Q).

(3.5) Theorem 3 is an FCLT analog of part of theorem 2. The corresponding w.p.1 statement is: Suppose that either limit (x) or limit (xi) in (2.8) holds. Also suppose that either limit (i) or limit (ii) holds. Then limit (viii) there holds if and only if limit (ix) holds. This statement is a corollary to theorem 2. Theorem 4 in sect. 4 is an FCLT analog to the rest of theorem 2.

(3.6) As with classical CLTs, convergence in distribution may hold for the cumulative processes of W_k and $Q(t)$ with different normalization; e.g. they may converge to a stable law other than the normal distribution (as often occurs when the second moments are not finite). Instead of (3.1), we could have the random functions

$$W_n'(t) = n^{-\alpha}\left[\sum_{k=1}^{[nt]} W_k - wnt\right] \quad \text{and} \quad Q_n'(t) = n^{-\alpha}\left[\int_0^{nt} Q(s)\,\mathrm{d}s - qnt\right], \qquad (3.8)$$

where $0 < \alpha < 2$. Theorem 3 remains valid with (W_n', Q_n') in (3.8) substituted for (W_n, Q_n) in (3.1), but now the limit processes are related by $Q(t) = W(\lambda t)$, $t \geqslant 0$. The special case of $\alpha = 1$ in (3.8) also can be used to establish the relation between the FWLLNs (functional weak laws of large numbers) paralleling the relation between the CLTs in theorem 3. Then the limit processes W, A, N and Q are all $\mathbf{0} \equiv \mathbf{0}(t) = 0$, $t \geqslant 0$.

(3.7) If we wish to keep the standard Skorohod J_1 topology on D, then the condition $P(N \in C) = 1$ (equivalently, $P(A \in C) = 1$) in theorem 3 is necessary; see theorem 7.4 of [14]. If we wish to treat limits N and A with discontinuous paths, then we can do so with the Skorohod M_1 topology; see theorem 7.5 of [14]. (Theorem

5.1(i) of [14] does extend to the M_1 topology, but to apply theorem 5.1(i) we need to have some control of the discontinuities in the limit processes, just as with the J topologies; the condition is just as specified for J_1 in [14].)

4. Knocking out the remainder term

Paralleling theorem 2, the mere existence of a weak convergence limit for one pair of processes, either (N_n, W_n) or (A_n, W_n), is enough by itself to knock out the remainder terms R_n and S_n and yield the desired joint convergence in theorem 3. For this step we assume that the limit process W also has continuous paths, i.e. $P(W \in C) = 1$.

We now introduce the following additional random functions:

$$D_n(t) = n^{-1/2}[D_{[nt]} - \lambda^{-1}nt], \; D_n'(t) = n^{-1/2}[D'_{[nt]} - \lambda^{-1}nt],$$

$$O_n(t) = n^{-1/2}[O(nt) - \lambda nt], \quad \overline{W}_n(t) = n^{-1/2}W_{[nt]}, \; \overline{Q}_n(t) = n^{-1/2}Q(nt) \quad (4.1)$$

$$(WO)_n(t) = n^{-1/2}\left[\sum_{k=1}^{O(nt)} W_k - \lambda w n t\right], \; (QD')_n(t) = n^{-1/2}\left[\int_0^{D'_{[nt]}} Q(s)\,\mathrm{d}s - \lambda^{-1} q n t\right]$$

for $t \geqslant 0$.

The random functions (4.1) and (3.1) together correspond to the different limits in (2.3). The following result is the FCLT-analog of theorem 2. As before, part (e) is of primary interest. It produces a statement of the FCLT refinement of $L = \lambda W$ that is nearly symmetric.

THEOREM 4

(a) $A_n \Rightarrow A$ with $P(A \in C) = 1$ if and only if $N_n \Rightarrow N$ with $P(N \in C) = 1$, in which case $(A_n, N_n) \Rightarrow (A, N)$, where A and N are related as in (3.2). Henceforth, assume that these limits hold.

(b) The limits $(A_n, D_n) \Rightarrow (A, A)$, $(A_n, \overline{W}_n) \Rightarrow (A, \boldsymbol{\theta})$ and $(A_n, D_n') \Rightarrow (A, A)$ are all equivalent, in which case $(A_n, D_n, \overline{W}_n, D_n') \Rightarrow (A, A, \boldsymbol{\theta}, A)$.

(c) The limits $(N_n, O_n) \Rightarrow (N, N)$ and $(N_n, \overline{Q}_n) \Rightarrow (N, \boldsymbol{\theta})$ are equivalent, in which case $(N_n, O_n, \overline{Q}_n) \Rightarrow (N, N, \boldsymbol{\theta})$ in D^3.

(d) The limit $(A_n, D_n) \Rightarrow (A, A)$ implies $(A_n, D_n, \overline{W}_n, D_n', N_n, O_n, \overline{Q}_n) \Rightarrow (A, A, \boldsymbol{\theta}, A, N, N, \boldsymbol{\theta})$, but the limit $(N_n, O_n) \Rightarrow (N, N)$ does not imply that $\overline{W}_n \Rightarrow \boldsymbol{\theta}$.

(e) The limit $(A_n, W_n) \Rightarrow (A, W)$ with $P(W \in C) = 1$ holds if and only if the limit $(A_n, D_n, Q_n) \Rightarrow (A, A, Q)$ holds with $P(Q \in C) = 1$, in which case $(V_n, D_n, \overline{W}_n, D_n', O_n, \overline{Q}_n, (WO)_n, (QD')_n) \Rightarrow (V, A, \mathbf{0}, A, N, \mathbf{0}, Q, W)$ in D^{15}, where V_n and V are the processes in D^8 in theorem 3.

Proof

As noted in the proof of theorem 3, (a) follows from section 7 of [14]. For (b), limits for D_n and $\overline{W}_n$ are equivalent because $\overline{W}_n = A_n - D_n$. Since $A_n(t) \leqslant D_n(t) \leqslant D_n'(t)$ for all n and t, the limit for (A_n, D_n') implies the others. The hard case, obtaining the limit for D_n' from the limit for D_n, is obtained from theorem 6.2(ii) of [14]. (The proof there is not unlike the corresponding part of the proof of theorem 2 here.) For [14], make the translation $x_n(t) = n^{-1/2}D_{[nt]}$ and $c_n = \lambda^{-1}n^{1/2}$. The joint convergence is obtained too because [14] is applied via the w.p.1 representation of weak convergence; i.e. starting with the weak convergence $(A_n, D_n) \Rightarrow (A, A)$, we apply the Skorohod representation theorem ([14], p. 68) to obtain the w.p.1 convergent version $(\tilde{A}_n, \tilde{D}_n) \to (\tilde{A}, \tilde{A})$ from which we get $(\tilde{A}_n, \tilde{D}_n, (\tilde{D}_n)') \to (\tilde{A}, \tilde{A}, \tilde{A})$ w.p.1 by applying theorem 6.2(ii) of [14]. Since $(\tilde{A}_n, \tilde{D}_n, (\tilde{D}_n)')$ is distributed as (A_n, D_n, D_n'), $(A_n, D_n, D_n') \Rightarrow (A, A, A)$.

Part (c) is immediate since $\overline{Q}_n = N_n - O_n$. They key to part (d) is the fact that $O(t)$ is sandwiched between the inverse processes of A_n and D_n'. Let $M(t) = \max\{k \geqslant 0: D_k' \leqslant t\}$, $t \geqslant 0$, where we understand $D_0' = 0$ without having a $0th$ customer. It is easy to see that $M(t) \leqslant O(t) \leqslant N(t)$ for all t. The FCLT for $M(t)$ is obtained from the FCLT for D_k' just as the FCLT for $N(t)$ is obtained from the FCLT for A_k, from the inverse mapping in section 7 of [14]. Since $M(t) \leqslant O(t) \leqslant N(t)$ for all t, $(A_n, N_n, D_n', O_n) \Rightarrow (A, N, A, N)$ and the more general joint convergence in (d) holds if $(A_n, D_n') \Rightarrow (A, A)$. Example 1 in sect. 2 yields $(N_n, O_n) \Rightarrow (N, N)$, where $N = \mathbf{0}$ without having $\overline{W}_n \Rightarrow \mathbf{0}$.

(e) Suppose that $(A_n, W_n) \Rightarrow (A, W)$, where $P(W \in C) = 1$. Then we have $\overline{W}_n \Rightarrow \mathbf{0}$ by an application of the continuous mapping theorem with the *maximum jump functional*

$$J_T(x) = \sup_{0 \leqslant t \leqslant T} \{|x(t) - x(t-)|\}$$

for any T, which is measurable on D and continuous at any $x \in C$. By theorem 4.4 of [1], we have $(A_n, W_n^c, \overline{W}_n) \Rightarrow (A, W^c, \mathbf{0})$. By (a)–(d), we have

$$(A_n, N_n, D_n, \overline{W}_n, D_n', O_n, \overline{Q}_n, W_n) \Rightarrow (A, N, A, \mathbf{0}, A, N, \mathbf{0}, W) \text{ in } D^8.$$

We then add on the limits $(WN)_n \Rightarrow Q$ for $(WN)_n$ in (3.1) and $(WO)_n \Rightarrow Q$ for $(WO)_n$ in (4.1) by applying theorem 5.1(i) of [14], as in the proof of theorem 3. The final joint convergence in D^{15} is obtained by applying theorem 1(a) here and theorem 4.1 of [1].

Example 1 in sect. 2 shows that we can not obtain the desired joint convergence in D^{15} starting with $(A_n, Q_n) \Rightarrow (A, Q)$. Since $P(Q \in C) = 1$, we can apply the continuous mapping theorem again with the maximum jump functional in (4.3) to get $\bar{Q}_n \Rightarrow \mathbf{0}$, but we do not get $D_n' \Rightarrow A$. However, if we start with $(A_n, D_n, Q_n) \Rightarrow (A, A, Q)$, then we get $(A_n, N_n, D_n, D_n', \bar{W}_n, \bar{Q}_n, O_n, Q_n) \Rightarrow (A, N, A, A, \mathbf{0}, \mathbf{0}, N, Q)$ in D^8 from (a)–(c) and the maximum jump functional. We then get the rest from theorem 1(a). □

5. Statistical estimation

We evaluate estimators here in terms of their *asymptotic efficiency.* For each j, let $\hat{Q}_j(t)$ be an estimator for a parameter q based on data from the interval $[0, t]$. We assume that $\hat{Q}_j(t)$ obeys a CLT, i.e. $t^{1/2}(\hat{Q}_j(t) - q) \Rightarrow Q_j$ in R at $t \to \infty$, where $EQ_j = 0$ and $\mathrm{Var}(Q_j) = E(Q_j^2) = c_j^2$. Of course, usually the limiting random variable Q_j will be normally distributed, but we do not require it. We call c_j^2 the *efficiency parameter* of the estimator $\hat{Q}_j(t)$. We say that one estimator of q, $\hat{Q}_1(t)$, is a more (asymptotically) efficient estimator than another, $\hat{Q}_2(t)$, if $c_1^2 < c_2^2$.

Remark

(5.1) When the limiting random variables Q_1 and Q_2 need not be normally distributed with mean zero, we might compare asymptotic efficiency by using a variability partial ordering; see [11] and references therein. For example, we could say that $\hat{Q}_1(t)$ is more efficient than $\hat{Q}_2(t)$ if $Ef(Q_1) \leqslant Ef(Q_2)$ for all convex real-valued functions f. This stronger partial ordering is equivalent to the above definition in the special case of normal random variables with mean zero.

Now suppose that the parameter of interest q is in fact the limiting time-average queue length in the general framework of sect. 2, obtained via (ix) in (2.3). (Similar results hold for w, but we only discuss q in detail; see remark 5.6.) The *natural estimator* of q is

$$\hat{Q}_1(t) = t^{-1} \int_0^t Q(s)\, ds.$$

However, we can also use $L = \lambda W$ to obtain other estimators for q. There are two important cases: (i) when the arrival rate λ is estimated, and (ii) when the arrival rate λ

is known and used directly. We might elect to estimate λ even when λ is known, or we might be forced to estimate λ because it is unknown. For an open network of queues, λ is typically known when we estimate from simulation but not when we estimate from system measurements. For a closed network of queues, λ is typically unknown even when we estimate from simulation.

When λ is estimated, the standard estimator based on data from the interval $[0, t]$ is $\hat{\Lambda}_1(t) = t^{-1}N(t)$. Alternative estimators of λ based on data from $[0,t]$ are, for example, $\hat{\Lambda}_2(t) = t^{-1}O(t)$ and $\hat{\Lambda}_3(t) = N(t)/A_{N(t)}$. Obviously, all the estimators $\hat{\Lambda}_j(t)$ converge to λ as $t \to \infty$ if the limits in (2.3) hold.

Associated with each of the two above cases are alternative estimators of q using estimators of w based on data over the interval $[0, t]$:

$$\text{(i)}\ \ \hat{Q}_{jk}(t) = \hat{\Lambda}_j(t)\hat{W}_k(t) \quad \text{and} \quad \text{(ii)}\ \ \hat{Q}_{k+1}(t) = \lambda \hat{W}_k(t)\,, \tag{5.1}$$

where $\hat{\Lambda}_j(t)$ is an estimator of λ and $\hat{W}_k(t)$ is an estimator of w, both using data over the interval $[0, t]$. Standard estimators of w using data in $[0, t]$ are

$$\hat{W}_1(t) = (N(t))^{-1} \sum_{k=1}^{N(t)} W_k,\ \hat{W}_2(t) = (O(t))^{-1} \sum_{k=1}^{O(t)} W_k \quad \text{and}$$

$$\hat{W}_3(t) = (N(t))^{-1} \sum_{k=1}^{O(t)} W_k.$$

We might be forced to use $\hat{W}_2(t)$ or $\hat{W}_3(t)$ instead of $\hat{W}_1(t)$ because W_k may not be observable before the customer departs.

We call estimators of the form (i) and (ii) in (5.1), respectively, *direct estimators* and *indirect estimators* of q via $L = \lambda W$. The CLTs provide a basis for comparing the asymptotic efficiency of the indirect and direct estimators in (5.1) with the *natural estimator* $\tilde{Q}_1(t)$.

FCLTs for these estimators and related ones follow from the FCLT for W_n or Q_n in (3.1) and an FCLT for random sums, which is a variant of theorem 17.1 of [1]. To state this result, let $T_n(t)$ and $U_n(t)$ be arbitrary random elements of D and let

$$\bar{T}_n(t) = n^{-1} T_n(t),\ U_n(t) = n^{1/2}[U_n(t) - \gamma]\,,$$

$$\hat{W}_n(t) = n^{1/2} U_n(t)\,[(T_n(t))^{-1} \sum_{k=1}^{T_n(t)} W_k - w]\,, \quad \text{and}$$

$$\hat{Q}_n(t) = n^{1/2} U_n(t)\,[(T_n(t))^{-1} \int_0^{T_n(t)} Q(s)\,\mathrm{d}s - q] \tag{5.2}$$

for $t \geqslant 0$. Let $T_n(t)$ be non-negative and nondecreasing and let γ and τ be positive constants. For the following result, W_k and $Q(s)$ need not be non-negative.

LEMMA 1

(a) If $W_n \Rightarrow W$, $\bar{T}_n \Rightarrow \tau e$ and $U_n \Rightarrow \mathbf{0}$, then $\hat{W}_n \Rightarrow \hat{W} = \gamma\tau^{-1}(W \circ \tau e)$.

(b) If $Q_n \Rightarrow Q$, $\bar{T}_n \Rightarrow \tau e$ and $U_n \Rightarrow \mathbf{0}$, then $\hat{Q}_n \Rightarrow \hat{Q} = \gamma\tau^{-1}(Q \circ \tau e)$.

Proof

We only prove (a) because (b) is similar. By theorem 4.4 of [1], $(W_n, U_n, \bar{T}_n) \Rightarrow (W, U, \bar{T})$ in D^3. From the definition of $\hat{W}_n$ in (5.2),

$$\hat{W}_n(t) = n^{1/2}\gamma[(T_n(t))^{-1} \sum_{k=1}^{T_n(t)} (W_k - w)] + U_n(t)\,[(T_n(t))^{-1} \sum_{k=1}^{T_n(t)} (W_k - w)]$$

$$= (T_n(t))^{-1}\,\gamma(W_n \circ T_n)(t) + (T_n(t))^{-1}\,U_n(t)\,(W_n \circ T_n)(t),$$

so that $\hat{W}_n \Rightarrow \tau^{-1}\gamma(W \circ \tau e) + \tau^{-1}\mathbf{0}(W \circ \tau e) = \tau^{-1}\gamma(W \circ \tau e)$. □

Remarks

(5.2) If W is Brownian motion in lemma 1, then $\hat{W}$ is distributed as $\gamma\tau^{-1/2}W$, and similarly for $\hat{Q}$.

(5.3) In contrast to the limits for random sums in theorems 3 and 4, e.g. $(WN)_n$, the random index $T_n(t)$ affects the limits $\hat{W}$ and $\hat{Q}$ in lemma 1 only via the parameter τ; i.e. the limits $\hat{W}$ and $\hat{Q}$ are unchanged if we let $T_n(t)$ be deterministic: $T_n(t) = \tau nt$. Correspondingly, in lemma 1 we only assume a functional weak law of large numbers (FWLLN) for $T_n(t)$ instead of an FCLT.

Our next result shows that there is no advantage or disadvantage to direct estimation of q via (i) in (5.1) compared with the natural estimation by $\hat{Q}_1(t)$ as far as asymptotic efficiency is concerned *when λ is estimated.* Let c_{jk}^2 be the efficiency parameter associated with $\hat{Q}_{jk}(t)$ in (5.1). Let $\hat{Q}_{jk}^n$ be the associated random function in D, defined by

$$\hat{Q}_{jk}^n(t) = n^{1/2}[\hat{Q}_{jk}(nt) - qnt], \quad t \geqslant 0. \tag{5.3}$$

THEOREM 5

If $(A_n, W_n) \Rightarrow (A, W)$ with $P((A, W) \in C^2) = 1$, so that the joint convergence in theorem 4(e) holds, then $\hat{Q}^n_{jk} \Rightarrow Q$ in D, so that $Q_{jk}(t)$ obeys a CLT with $c^2_{jk} = c^2_1$ for $j = 1, 2, 3$ and $k = 1, 2, 3$.

Proof

The CLTs follow from the associated FCLTs by applying the continuous mapping theorem with the projection map, as in corollary 3.1. The FCLTs in turn follow either directly from theorem 4 or by lemma 1. For example, the FCLTs associated with $\hat{Q}_{11}(t)$ and $\hat{Q}_{22}(t)$ in (5.1) correspond exactly to the FCLTs for $(WN)_n$ in (3.1) and $(WO)_n$ in (4.1), respectively. The FCLT for $\hat{Q}_{12}(t)$ follows from lemma 1 by letting $T_n(t) = O(nt)$ and $U_n(t) = O(nt)/N(nt)$. Since $(O_n, N_n) \Rightarrow (N, N)$ by virtue of theorem 4, $U_n(t) = n^{1/2}(U(nt) - 1) = \Psi_n^{-1}(t)\,[O_n(t) - N_n(t)]$, so that $U_n \Rightarrow \lambda^{-1} e(N - N) = \boldsymbol{\theta}$. □

The more interesting and difficult case is when λ is known and is applied in the indirect estimation, so that we compare the direct efficiency parameter c^2_1 with the indirect efficiency parameters c^2_{k+1} associated with $\hat{Q}_{k+1}(t)$ in (5.1) for $k = 1, 2, 3$. The following theorem gives necessary and sufficient conditions for the indirect estimators of q to be more efficient than the direct estimator. The efficiency parameters are easy to identify, but difficult to evaluate in detail because they involve a complicated covariance in a two-dimensional limit vector. Let $\hat{Q}^n_{k+1}$ be the random function

$$\hat{Q}^n_{k+1}(t) = n^{1/2}[\hat{Q}_{k+1}(nt) - qnt], \quad t \geq 0. \tag{5.4}$$

THEOREM 6

(a) If $(A_n, W_n) \Rightarrow (A, W)$ with $P((A, W) \in C^2) = 1$, so that all the limits in theorem 4(e) hold, then $\hat{Q}^n_{k+1} \Rightarrow W \circ \lambda e$ and $\hat{Q}_{k+1}(t)$ obeys a CLT with $c^2_{k+1} = \mathrm{Var}(W(\lambda))$, $k = 1, 2, 3$.

(b) If, in addition, (A, W) is Brownian motion with zero drift and diffusion coefficient matrix Σ in (3.6), then

$$c^2_{k+1} = \lambda\sigma^2_2 = c^2_1 + 2\lambda q\sigma_{12} - \lambda q^2\sigma^2_1 \tag{5.5}$$

for $k = 1, 2, 3$, so that $c^2_{k+1} < c^2_1$ if and only if $q\sigma^2_1 - 2\sigma_{12} > 0$.

Proof

We consider only the case $k = 1$; the other cases can be treated similarly. For part (a), apply lemma 1 with $T_n(t) = N(nt)$ and $U_n(t) = \gamma = \lambda$, so that $\hat{Q}_2^n \Rightarrow \lambda\lambda^{-1}(W \circ \lambda e) = W \circ \lambda e$. For part (b) under the additional assumption, $\{W(\lambda t), t \geq 0\}$ is equal in distribution to $\{\lambda^{1/2} W(t), t \geq 0\}$, so that $\text{Var}(W(\lambda)) = \lambda \text{Var}(W(1)) = \lambda\sigma_2^2$. □

The following corollary to theorem 6 extends the GI/G/s estimator comparison in theorem 5 by Carson and Law [2]. We show that in considerable generality, it is more efficient to estimate q indirectly using known λ and a direct estimator for w.

COROLLARY 6.1

Suppose that $(A_n, W_n) \Rightarrow (A, W)$ with $P((A, W) \in C^2) = 1$ and

$$\sigma_{12} = \lim_{n\to\infty} \text{Cov}(A_n(1), W_n(1)).$$

If $\text{Cov}(W_k, U_j) \leq 0$ for all j and k, where $U_j = A_j - A_{j-1}$, then $\sigma_{12} \leq 0$ and $c_{k+1}^2 \leq c_1^2$, so that the indirect estimators of q are more efficient than the direct estimator.

Proof

Note that

$$\text{Cov}(A_n(1), W_n(1)) = n^{-1} \sum_{k=1}^{n} \sum_{j=1}^{n} \text{Cov}(W_k, U_j),$$

and apply theorem 6. □

Remarks

(5.4) Given the convergence of $(A_n(1), W_n(1))$, the covariance of the limit σ_{12} is indeed the limit of the covariances as assumed in corollary 6.1 under appropriate uniform integrability conditions ([1], p. 32). It suffices for $E(|A_n(1)|^{2+\delta})$ and $E(|W_n(1)|^{2+\delta})$ to be uniformly bounded. By Hölder's inequality,

$$E(|A_n(1)\, W_n(1)|^{1+\delta/2}) \leq E(|A_n(1)|^{2+\delta})^{1/2}\, E(|W_n(1)|^{2+\delta})^{1/2}.$$

(5.5) By essentially the same reasoning as in corollary 6.1, we can obtain a sufficient condition for $\hat{\sigma}_{12} \leqslant 0$, which implies that $c_1^2 \leqslant c_{k+1}^2$, so that the natural estimator of q is more efficient than the indirect estimator. Assuming that

$$\hat{\sigma}_{12} = \lim_{n \to \infty} \mathrm{Cov}\,(N_n(1), Q_n(1)),$$

it suffices to have Cov $(N(t_2) - N(t_1), Q(t)) \leqslant 0$ for all t and all $t_1 < t_2$. However, it seems that $\sigma_{12} \leqslant 0$ is much more likely than $\hat{\sigma}_{12} \leqslant 0$ in actual applications.

(5.6) Results paralleling theorems 5 and 6 hold for estimators of w based on data over the interval $[0, A_n]$ or one of the related intervals $[0, D_n]$ and $[0, D_n']$. The direct estimator for w is obviously $n^{-1} \Sigma_{k=1}^{n} W_k$. An indirect estimator of w paralleling $\hat{Q}_{11}(t)$ in (5.1) is

$$n^{-1} \int_0^{A_n} Q(s)\,\mathrm{d}s = (n^{-1}A_n)(A_n)^{-1} \int_0^{A_n} Q(s)\,\mathrm{d}s.$$

As in theorem 5, there is no advantage or disadvantage to using Little's law in the estimation of w when λ or λ^{-1} is to be estimated too. In the setting of theorem 6(a), the indirect estimators of w have efficiency parameters Var $(Q(\lambda^{-1}))$. In the setting of theorem 6(b) with λ known, the direct estimator has efficiency parameter σ_2^2, while the indirect estimator has efficiency parameter $\lambda^{-1}\hat{\sigma}_2^2 = (\sigma_2^2 - 2q\sigma_{12} + q^2\sigma_1^2)$, so that the natural estimator of w is more efficient than the indirect estimator if and only if the indirect estimator of q is more efficient than the natural estimator, i.e. if and only if $q\sigma_1^2 - 2\sigma_{12} > 0$. The issue is simply whether or not the known value of λ should be used if it is available.

6. Functional laws of the integrated logarithm

The functional-limit-theorem setting is convenient because virtually identical arguments yield FSLLNs, FWLLNs and FLILs as well as FCLTs. We mentioned FWLLNs in remark (3.6). We have not discussed FSLLNs because they are equivalent to ordinary SLLNs. Now we briefly discuss FLILs.

The FLILs involve the random functions in (3.1) and (4.1) with a different normalization; we multiply by $\phi(n)$ instead of $n^{-1/2}$, where usually $\phi(n) = (2n \log\log n)^{-1/2}$, e.g.

$$W_n(t) = \phi(n) \left[\sum_{k=1}^{[nt]} W_k - wnt \right], \quad t \geqslant 0. \tag{6.1}$$

We say that W_n in (6.1) obeys an FLIL and write $W_n \sqrt{\to} K_W$ if w.p.1 the sequence $\{W_n : n \geq 3\}$ is relatively compact in D (if every subsequence has a convergent subsubsequence) and the set of all limit points is the compact set K_W. The specific structure of the compact limit sets is not of major concern here, but we describe what usually occurs. Let $|\cdot|$ be the Euclidean norm in R^k. When we consider the product space D^k, we identify it with the space $D[0, \infty)$ in which the functions take values in R^k. The standard compact limit sets in D^k are the sets of all R^k-valued functions $\{x(t): t \geq 0\}$ that are absolutely continuous with respect to Lebesgue measure with derivative $x'(t)$ satisfying $\int_0^\infty |x'(t)|^2 \, dt \leq \delta$; see Strassen [12]. (A different compact limit set is defined and used by Wichura [15].) These limit sets are convex as well as compact. Compactness in one of the Skorohod topologies on D^k reduces to compactness with the topology of uniform convergence on compact subsets because all the functions are continuous.

With this framework, theorems 3 and 4 here have obvious FLIL analogs by virtually identical proofs. We replace weak convergence $W_n \Rightarrow W$ by the relative compactness $W_n \sqrt{\to} K_W$ defined above and we use the normalization by $\phi(n)$ as in (6.1). For the analog of theorem 3 (theorem 4), we assume that the compact limit set K_A (K_{AW}) contains only continuous functions. We only state the analog of theorem 4(e).

THEOREM 7

Let the random functions be defined in (3.1) and (4.1) with the normalization by $\phi(n)$. If $(A_n, W_n) \sqrt{\to} (K_{AW})$ in D^2, where $(K_{AW}) \subseteq C^2$ w.p.1, then

$$(V_n, D_n, \bar{W}_n, D'_n, O_n, \bar{Q}_n, (WO)_n, (QD')_n) \sqrt{\to} (K_{V,A,\theta,A,N,\theta,Q,W}) \text{ in } D^{15}.$$

Just as the limit process (A, W) in theorems 3 and 4 in general has a complicated joint distribution, so the compact limit set K_{AW} has a form not determinable by the marginal compact limit sets K_A and K_W separately. Even if the original sequences $\{A_n\}$ and $\{W_n\}$ were independent, the compact limit set K_{AW} associated with the FLIL would usually *not* be the product set $K_A \times K_W$. The separate sets K_A and K_W are of course identifiable as projections of the limit set K_{AW}.

Just as with the FCLTs, the FLIL-analogs of theorems 3 and 4 establish that the limit sets are essentially two-dimensional. For example, the limit set $(K_{W,Q,A,N,W,Q,\theta,\theta})$ in D^8 for the FLIL version of theorem 3 is to be interpreted as the "two-dimensional" subset $\{w, q, a, n, w, q, \theta, \theta) : (w, a) \in (K_{WA})\}$, where $q(t) = w(\lambda t) - qa(\lambda t)$, $n(t) = -\lambda a(\lambda t)$ and $\theta(t) = 0$ for all $t \geq 0$. This arises because essentially we have a continuous mapping. We thus obtain interesting ordinary LILs in R^k that are more descriptive than what is provided by the ordinary LILs directly.

COROLLARY 7.1

Under the assumptions of theorem 7,

$$\phi(n)\left(A_{[\lambda n]} - n,\ \sum_{j=1}^{[\lambda n]} W_j - \lambda w n,\ N(n) - \lambda n,\ \int_0^n Q(s)\,\mathrm{d}s - qn\right) \surd\!\!\to K_{AWNQ} \quad \text{in } R^4,$$

where

$$K_{AWNQ} = \left\{x \in R^4 : (x_1, x_2) \in \pi_1(K_{AW}),\ x_3 = -\lambda x_1 \ \text{ and } \ x_4 = x_2 - q x_1\right\}$$

with $\pi_1 : D^2 \to R^2$ the projection map, defined by $\pi_1(x) = x(1)$.

Proof

Apply theorem 7 to get $(A_n, W_n, N_n, Q_n) \surd\!\!\to K_{AWNQ}$ in D^4 and then the continuous map $f: D^4 \to R^4$ defined by $f(x) = (x_1(\lambda), x_2(\lambda), x_3(1), x_4(1))$ for $x = (x_1, x_2, x_3, x_4) \in D^4$.

References

[1] P. Billingsley, *Convergence of Probability Measures* (Wiley, New York, 1968).

[2] J.S. Carson and A.M. Law, Conservation equations and variance reduction in queueing simulations, Oper. Res. 28(1980)535.

[3] W. Feller, *An Introduction to Probability Theory and Its Applications,* Vol. II, 2nd edition (Wiley, New York, 1971).

[4] D.L. Iglehart and D.P. Kennedy, Weak convergence of the average of flag processes, J. Appl. Prob. 7(1970)747.

[5] D.L. Iglehart and W. Whitt, Multiple channel queues in heavy traffic, II: Sequences, networks and batches, Adv. Appl. Prob. 2(1970)355.

[6] J.D.C. Little, A proof for the queueing formula: $L = \lambda W$, Oper. Res. 9(1961)383.

[7] T. Rolski and S. Stidham, Jr., Continuous versions of the queueing formulas $L = \lambda W$ and $H = \lambda G$, Oper. Res. Lett. 2(1983)211.

[8] R.F. Serfozo, Functional limit theorems for stochastic processes based on embedded processes, Adv. Appl. Prob. 7(1975)123.

[9] S. Stidham, Jr., A last word on $L = \lambda W$, Oper. Res. 22(1974)417.

[10] S. Stidham, Jr., Sample-path analysis of queues, in: *Applied Probability – Computer Science: The Interface,* ed. R. Disney and T. Ott (Birkhauser, Boston, 1982).

[11] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models,* ed. D.J. Daley (Wiley, New York, 1983).

[12] V. Strassen, An invariance principle for the law of the iterated logarithm, Z. Wahrscheinlichkeitsth. 3(1964)211.

[13] W. Whitt, Stochastic Abelian and Tauberian theorems, Z. Wahrscheinlichkeitsth. 22(1972) 251.

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

[15] M.J. Wichura, Functional laws of the iterated logarithm for the partial sums of i.i.d. random variables in the domain of attraction of a completely asymmetric stable law, Ann. Probab. 2 (1974)1108.