Large finite population queueing systems. The single-server model

G. Louchard

Université Libre de Bruxelles, Département d’Informatique-CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium

Received 7 March 1991; revised 29 October 1993

Abstract

Stochastic variables associated to a single-server queueing system with finite population are shown to weakly converge, on some time regions, to Gaussian processes, Brownian motions or stochastic integrals on such when the population size increases. Queue length, unfinished work, storage occupied (in a computer system model) and idle time show different limiting behaviour, depending on the arrival and service distribution.

1. Introduction

In Louchard (1988), we have considered the large finite population, infinite-server model. Here we investigate the single-server model in which a finite number of $n$ customers arrive at some facility, with given arrival and service time distributions. For large $n$, we obtain, on particular time regions, Gaussian limiting processes and diffusion approximations (in particular Brownian motions or stochastic integrals on such) for a wide variety of stochastic variables of interest. An overview of diffusion approximations can be found in Glynn (1990).

Several authors (for example Iglehart and Whitt, 1970; Reiman, 1983) have investigated heavy traffic limits in queues where $\rho \to 1$ in some specified sense. Other related papers are discussed in Section 2. In some aspects, our model demonstrates more diversified behaviour and can be generalized in several directions. We have also investigated several transition (in the sense of Newell, 1982) and light traffic situations.

This paper is organized as follows: Section 2 summarizes the model and the basic notations we need, Section 3 describes the main typical queueing cases, Section 4 considers the queue length $Q_n(t)$ at time $t$, Section 5 analyses the queue size maximum, Section 6 is devoted to the total unfinished work $U_n(t)$ at $t$, Section 7 examines the total storage $M_n(t)$ occupied at $t$ in a computer system and Section 8 considers some
other variables of interest such as idle time, busy period cost and distribution. Section 9 concludes the paper. Two appendices provide some particular technical results used in the paper.

The reader will notice that, in contrast to our previous paper, we have not written down all covariance matrices for our variables: here they can easily be deduced from our explicit Brownian stochastic integrals.

2. Model and basic notations

2.1. The model

Let us consider a large finite population of \( n \) customers. Each customer applies for a single-server facility. All customers have the same arrival time distribution function \( F \) (and not inter-arrival time as usually used in classical queueing models). Such model was first introduced by Newell (1982, p. 32, 112) and partially analysed in Newell (1968a–c) in the framework of queues with time-dependent arrival rates. Many applications of the model are considered in Gaver et al. (1975). This system can be used, for instance, to model a computing centre, where programs are run once a day, the times of submission being random. We also assume that each customer asks for some storage \( M \). As in Coffman and Reiman (1983), the storage demand can depend on the service (processor time) required by the customer. Given \( F \), we are interested in the asymptotic \((n \to \infty)\) distributions of \( Q_n(t), U_n(t), M_n(t) \). Several regions of the time axis will be considered, leading to oversaturated, undersaturated and transition behaviours.

Note that our model generally leads to non-Markovian properties: we do not have exponential inter-arrival times and we consider a general service time. In some particular short time intervals, the processes can be locally Markovian: that is where we can obtain diffusion approximations. Let us mention a few related results: the simple case where the distribution function of each customer's arrival time is uniform on \([0,1]\) has been analysed in Iglehart and Whitt (1970), Example 3. The number of servers is finite; the service rate depends on \( n \) in such a way that the traffic intensity \( \rho_n \) converges to \( \rho \geq 1 \) as the population size \( n \) grows to infinity.

This could of course be transformed into our general case but only with some time distortion, which would lead to time-dependent service time: it is not possible to transform the results for the uniform distribution to a more general arrival time distribution.

Giorno et al. (1987) consider a Markovian non-homogeneous queueing system. In particular, they analyse a diffusion approximation with linear drift and periodically time-varying infinitesimal variance. Massey (1985), analyses a time-dependent M/M/1 queue: mean and variance are considered when a time-scale change is applied (the paper's parameter \( \varepsilon \) is equivalent to \( n^{-\alpha} \) in our model). Asymptotic
distributions for the same model have been obtained recently by Mandelbaum and Massey (1993).

2.2. Basic notations

The following notations will be used throughout this paper.

- \( n := \) the total input population. All customers have independent identically distributed characteristics.
- \( F := \) the distribution function (DF) of each customer arrival time, (same DF for all customers) with density \( f \) (initial time of the system is 0). Each customer applies for service only once. The function \( \tau := F/(1 - F) \) will be used later. When defined, \( \gamma(t) := f'(t)/2 \).
- \( G := \) the DF of each customer service time \( S \), with density \( g \), finite mean \( m_S \), variance \( \sigma_S^2 \). Let \( \mu_S := 1/m_S \), \( \chi^2_S := \sigma_S^2/m_S^2 \). To obtain realistic limits, each service time will be normalized by \( n \) (i.e. divided by \( n \)).
- \( \text{VAR}(X), \text{COV}(X, Y) := \), respectively, variance \( [X] \), covariance \( [X, Y] \) for any random variables (RV) \( X, Y \).
- \( N(m, V) := \) the normal (or Gaussian) RV with mean \( m \) and \( \text{Var} V \). The standard Gaussian DF will be denoted by \( \Phi \).
- \( \text{the FIFO rule is applied to all queueing systems.} \)
- \( \text{For the computer system model, the storage demand } M \text{ of each customer is a random variable (RV), depending on } S. \text{ This model has been introduced by Coffman and Reiman (1983). Let} \)

\[
m_M(\xi) := E[M | S = \xi],
\]

\[
\sigma_M(\xi) := \text{VAR} [M | S = \xi],
\]

\( m_M, \sigma_M := \text{unconditional mean and VAR of } M. \)

- \( E_a[B(X)] := \text{Pr}[B | X(0) = a] \) for any event \( B(X) \) belonging to the Borel field generated by a process \( X(t) \).
- \( A_n(t) := \) total number of customers who have applied for service by time \( t \).
- \( S_n(t) := \) the renewal processes associated with the RV \( (S/n) \). Note that \( S_n(t)/n \Rightarrow \mu_S t. \)
- \( Q_n(t) := \) total number of customers in the system at time \( t \).
- \( U_n(t) := \) total unfinished work (backlog) at time \( t \).
- \( M_n(t) := \) total storage occupied at time \( t \) (in the computer system model).
- \( \text{For any random variable } X_n(t), \text{ we define } \tilde{X}_n(t) := [X_n(t) - E[X_n(t)]]/\sqrt{n}. \)
- \( \mathcal{R}(f(t)) := f(t) - \inf_{0 \leq s \leq t} f(s). \text{ This is the continuous reflection mapping. (see Harrison, 1985, Section 2.2 for more details).} \)
- \( \text{For } i = 0, 1, \ldots, \text{let} \)

\( B(t) := \) a copy of the standard Brownian motion (BM)

\( BB_t(t), t \in [0, 1] := \) a copy of the standard Brownian bridge (BB) on \([0, 1]\).
Note that $BB_i(t) \equiv (1 - t)B_i(t/(1 - t))$ for some BM $B_i$
$BT_i(s, t) := \text{a copy of the two-parameter BM (see Cs"org"o and Revesz (1981) for a detailed description)}$
$BR_i(q(t), \sigma, t) := \text{a copy of the reflected BM (RBM) with local trend } q(t) \text{ and } \VAR \sigma^2$. Note that
$BR(q(t), \sigma, t) \overset{d}{=} \sigma \mathcal{R} [B(t) + (\int_0^t q(u) \, du)/\sigma]$
$x(t), t \in [0, 1] := \text{a copy of the standard Brownian excursion (BE) (see Chung (1976) for more details on this process).}$
All these copies are independent.

- $\Rightarrow := \text{weak convergence of random functions in the space of all right-continuous functions having left limits and endowed with the Skorohod metric, } d$ (see Billingsley, 1968). The limits in the paper are always defined for $n \to \infty$.
- To each customer, we associate a variable $v := (i/n)$, where $i :=$ the rank number (i.e. the order of arrival) of this customer. $s$ is then defined by $v = F(s)$. $s$ can be seen as a deterministic arrival time associated with customer $i$.

3. Preliminary analysis

In this section, we describe the queueing system we study and analyse the various typical time intervals we have to investigate.

3.1. The queueing model

Let $J_n(t) := A_n(t) - S_n(t)$ and $Q_n^r(t) = J_n(t) - \inf_{s \leq t} J_n(s)$ (this is the Borovkov modified system). As shown in Iglehart and Whitt (1970), Section 3, the Skorohod distance between $Q_n$ and $Q_n^r$ converges to 0 in heavy traffic. Our following investigations for heavy traffic will thus be based on $Q_n^r$, written as
$Q_n^r(t) = \mathcal{R} [J_n(t)]. \quad (1)$

As in Louchard (1988), we use the well-known fact that
$\tilde{A}_n(t) := \frac{A_n(t) - nF(t)}{\sqrt{n}} \Rightarrow \tilde{A}(t) = BB_0[F(t)] = [1 - F(t)]B_0[\tau(t)] \quad (2)$
for some BB and BM, $BB_0$ and $B_0$ with $\tau := F/(1 - F)$. This can also be written as
$\tilde{A}(t) = [1 - F(t)] \int_0^t \sqrt{\tau(u)} \, dB_1(u) \quad (3)$
for some BM, $B_1$.

This is a classical time substitution: see McKean (1969) Section 2.5 for details. Such substitutions will be frequently used in the sequel. It is also well-known that, for fixed $t_0$, $A_n(t) - A_n(t_0)$, with $t - t_0 = O(1/n)$, behaves like a Poisson process with parameter
We shall also use the classical weak convergence (see Iglehart, 1974, Theorem 4.1)

\[ \frac{S_n(t) - n \mu_S t}{\sqrt{n \chi_S}} \Rightarrow B_2(t) \quad \text{for some BM, } B_2 \]

and \( \chi_S^2 := \sigma_S^2 / m_S^3 \).

An equivalent convergence for \( S_n(t) \) is given by (see (A.3))

\[ \frac{S_n(t) - n \mu_S t}{\sqrt{n}} \Rightarrow - \int_0^{\mu_S t} \sigma_S \mu_S \, dB_3(v) \quad \text{for some BM, } B_3, \]

where the variable \( v \) corresponds to the (normalized) deterministic rank number of each customer in the renewal process and \( \sigma_S \, dB_3(v) \) corresponds to its service time.

### 3.2. An enumeration of cases

A first analysis of our queue behaviour leads to a decomposition into several typical cases (see Fig. 1). We consider different parts of the time axis, for a given DF, \( F(t) \).

These cases can be formally defined as follows. For simplicity, we assume that \( f \) is continuous and that, where \( f(t) = \mu_S, \ f'(t)/2 \) exists = \( \gamma(t) \) (say) with \( \gamma(t) \neq 0 \) (first-order contact between the density curve and the line parallel to \( \mu_S t \)). The case where \( \gamma(t) = 0 \) and \( f''(t) < 0 \) can be analysed similarly to the "mild rush hour" in Newell (1968c). This is noted by case 10 in Table 1, but we will not pursue the detailed analysis here. Other cases are more complicated but can be treated similarly. Let \( A^i \) be disjoint time intervals: \( [t_i, t_i'] \) such that

- the first one, \( A^0 \), exists if either \( f(0) > \mu_S \) or \( f(0) = \mu_S \) and \( \gamma(0) > 0 \). \( t_0' = 0, \ t_0^0 = \min(t: F(t) = \mu_S t) \),
- the other ones: \( A^i (i \geq 1) \) are defined by \( f(t_1^i) = \mu_S, \ \gamma(t_1^i) > 0, \ t_1'^i : = \min(s > t_1^i: F(s) - F(t_1^i) = \mu_S(s - t_1^i)) \).

Some time-rescaling will be frequently needed to refine our processes: the time-window of observation will be differently chosen to insure weak convergence. We denote it by \( t = u/n^2 \), where \( u \) will be appropriately defined in each case. The cases can now be summarized in Table 1, where \( 0 < \alpha \leq 1 \).

### 4. Approximations for \( Q_n(t) \)

We first remark that heavy traffic limit theorems for multiple channels queues have been obtained by Iglehart and Whitt (1970), where they let \( \rho_n \to 1 \) with \( \sqrt{n} (\lambda_{n,A} - \lambda_{n,S}) \to C (-\infty \leq C \leq \infty), \ \lambda_{n,A} \) and \( \mu_{n,S} \) being the total arrival and service rates (see also Reiman (1983) for several generalizations).
Fig. 1
Table 1

<table>
<thead>
<tr>
<th>Case</th>
<th>( f(\text{or } F) ) property</th>
<th>Observation interval for ( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( f(0) &gt; \mu_s )</td>
<td>( t = O(1/n^4) )</td>
</tr>
<tr>
<td>2</td>
<td>( f(0) = \mu_s, \gamma(0) &gt; 0 )</td>
<td>( t = O(1/n^4) )</td>
</tr>
<tr>
<td>3</td>
<td>( f(0) = \mu_s, \gamma(0) &lt; 0 )</td>
<td>( t = O(1/n^4) )</td>
</tr>
<tr>
<td>4</td>
<td>( f(0) &lt; \mu_s )</td>
<td>( t = O(1/n^4) )</td>
</tr>
<tr>
<td>5</td>
<td>( \exists A' (i \geq 0) )</td>
<td>( t \in A' ) for some ( i )</td>
</tr>
<tr>
<td>6</td>
<td>( \exists A' (i \geq 0) )</td>
<td>( t - t_1^* = O(1/n^4) )</td>
</tr>
<tr>
<td>7</td>
<td>( f(t) &lt; \mu_s )</td>
<td>( t \not\in A' )</td>
</tr>
<tr>
<td>8</td>
<td>( \exists A' (i &gt; 0) )</td>
<td>( t - t_1^* = O(1/n^4) )</td>
</tr>
<tr>
<td>9</td>
<td>( \tilde{t} := t_1(A' t_{i-1} \equiv i \bar{A}') ) for some ( i \geq 0 )</td>
<td>( t - \tilde{t} = O(1/n^4) )</td>
</tr>
<tr>
<td>10</td>
<td>( \exists \tilde{t}; \gamma(\tilde{t}) = 0, f'(\tilde{t}) &lt; 0 )</td>
<td>( t - \tilde{t} = O(1/n^{3.5}) )</td>
</tr>
</tbody>
</table>

Our system is, in some aspects, rather different and a large variety of limiting processes will be investigated. We shall analyse the different cases of Section 3 starting with the easiest case and going to the more complicated. For convenience of notations, we define each \( A' \) by \([t_1, t_2]\) (without indices), and denote \( v := f(t_1) \), \( \gamma := f'(t_1)/2 \). Of course, when \( i \geq 1 \), some queue backlog can exist at \( t_1 \) and it must be taken into account. Different time-scale changes will lead to various asymptotic processes.

We obtain, for instance, non-Markovian Gaussian processes, \( M/G/1 \) light traffic queue, ordinary and reflected BM, diffusions.

Case 5: This is the easiest situation. Here, \((Q_n(t) - J_n(t))/\sqrt{n} \rightarrow 0 \) and no reflection must be applied as \( E[ J_n(t)] = n[F(t) - \mu_s t] \), with \( F(t) - \mu_s t > 0 \). The argument is well-known, see, for instance, Iglehart (1974) Section 5 or Iglehart and Whitt (1970) Theorem 2.2). From (3) and (4), we obtain a non-Markovian–Gaussian process

\[
\frac{Q_n(t) - n[F(t) - \mu_s t]}{\sqrt{n}} = \tilde{A}(t) - \tilde{\chi}_s B_2(t). \tag{6}
\]

Case 7: Let \( t = u/n \). Locally, the queue behaves, in the \( u \) time-scale, like a non-Markovian light traffic \( M/G/1 \) standard queue with input rate \( f(t) \). Diffusion approximations may also be useful in this situation but caution must be taken and refinements are usually necessary (see Whitt (1982) for more details). The generating function of the stationary distribution of the \( M/G/1 \) queue is well-known: see Cohen (1982). From this result, it is easy to check that, under some regularity conditions on \( G \), the tail of the stationary distribution is geometric. The transient behaviour of the \( M/M/1 \) queue has been analysed in Abate and Whitt (1987, 1988). The time-dependent light traffic \( M/M/1 \) queue is investigated in Mandelbaum and Massey (1993, Theorem 3.3).
Case 1: Let \( t = \frac{u}{n^a} \) and \( v := f(0) \). Eqs. (2) and (3) are now simplified into
\[
\frac{\left[ A_n(u) - \nu u n^{1-x} \right] n^{(1-x)/2}}{n^{(1-a)/2} n^{x/2}} = \sqrt{v} B_1(u), \quad 1 > x > \frac{1}{3},
\]
where \( B_1 \) and \( B_2 \) are the same BM as in (3) and (4), in the \( u \)-scale, and the condition on \( x \) is imposed such that the quadratic term in the \( F(t) \) expansion is negligible with respect to the Brownian contribution. From (4) we thus find that
\[
\left[ Q_n(u) - n^{1-x}(v - \mu S) u \right]/n^{(1-x)/2} = \delta_1 B_3(u), \quad 1 > x > \frac{1}{3},
\]
where \( \delta_1^2 := v + \xi^2 \), \( \delta_1 B_3(u) := \sqrt{v} B_1(u) - \xi S B_2(u) \). Again no reflection is needed as \( (1 - x) > (1 - a)/2 \) (the trend is much larger than BM contribution).

Case 4: Let \( t = u/n \). As in case 7 the queue behaves now, in the \( u \)-scale, like a light-traffic \( M/G/1 \) queue with input rate \( v \).

Case 2: \( \gamma > 0 \): Let \( t = \frac{u}{n^{a_0}} \). (2) and (3) give now (\( v = \mu S \)):
\[
\frac{\left[ A_n(u) - \mu S n^{1-x} u - \gamma u^2 n^{1-2a} \right] n^{(1-x)/2}}{n^{(1-a)/2} a^{x/2}} = s B_1(u), \quad 1 > x > \frac{1}{3}.
\]
The condition on \( \mu \) is now related to the cubic term in the \( F(t) \) expansion. From (4) we find that
\[
\left[ J_n(u) - \gamma u^2 n^{1-2a} \right]/n^{(1-a)/2} = \delta_2 B_3(u), \quad 1 > x > \frac{1}{3}.
\]
where \( \delta_2^2 := \mu S + \xi^2 \), and \( \delta_2 B_3(u) := \sqrt{\mu S} B_1(u) - \xi S B_2(u) \). From (1), the detailed analysis of the queue, in the \( u \)-scale, depends obviously on \( x \). Thus,

(a) if \( 1 > x > \frac{1}{3} \), the trend becomes asymptotically negligible with respect to the BM contribution,

(b) if \( x = \frac{1}{3} \), the trend is given by \( 2\gamma u \) and

(c) if \( \frac{1}{3} < x < \frac{1}{3} \), no reflection is needed as \( (1 - 3x)/2 > 0 \).

The different subcases are thus described as follows, stating from the coarsest time-scale:
(2.1) \( \gamma > 1 \). The queue behaves like a \( M/G/1 \) standard queue with input rate \( \mu S \) and traffic intensity 1.
(2.2) \( \frac{1}{3} < x < 1 \). We have the weak convergence:
\[
\frac{Q_n(u)}{n^{(1-a)/2}} \Rightarrow BR_0(0, \delta_2, u) \quad \text{for some RBM, } BR_0 \quad \text{(no trend)}.
\]
(2.3) \( x = \frac{1}{3} \). This gives
\[
\frac{Q_n(u)}{n^{1/3}} \Rightarrow BR_1(2\gamma u, \delta_2, u) \quad \text{for some RBM, } BR_1.
\]
(2.4) \( \frac{1}{3} < x < \frac{1}{3} \). We obtain
\[
\left[ Q_n(u) - \gamma u^2 n^{1-2a} \right]/n^{(1-a)/2} = \delta_2 B_3(u) \quad \text{(no reflection)}.
\]
Case 3 (\( \gamma < 0 \)): Let \( t = u/n^a \). The analysis is similar to that of case 2 for the first three subcases. We observe the following situations:
(3.1) $x \to 1$. Same M/G/1 behaviour as in (2.1).

(3.2) $\frac{1}{2} < x < 1$. Same weak convergence as in (2.2).

(3.3) $x = \frac{1}{2}$. Same weak convergence as in (2.3) (note that here $\gamma < 0$).

(3.4) $\frac{1}{2} < x < \frac{1}{4}$. Keep $u$ fixed, at $u_0$ say. Another local scale change will simplify the queue description. Let $u - u_0 = w/n^\beta$, $0 < \beta < 1$. From (9) we deduce the local trend (in the $w$-scale): $2\gamma u_0 n^{(1-3\alpha)/2} dw/n^\beta$, and the stochastic increment: $(\delta_2/n^{\beta/2}) dB_\beta(w)$ for some BM, $B_\beta$. To equalize the $n$ exponents, we must have $\beta = 1 - 3\alpha$ (hence $dt = dw/(n^{(1-2\alpha)})$, and this gives the weak convergence, in the $w$-scale:

$$\frac{Q_n(w)}{n^x} \Rightarrow BR_2(2\gamma u_0, \delta_2, w)$$

for some RBM, $BR_2$.

The trend is locally constant in the $w$-scale. It is well-known that the stationary density is given by (see also (43) below)

$$\Pr[BR_2 \in dx] = \exp \left[ -\frac{4\gamma u_0 x}{\delta_2^2} \right] \frac{4\gamma u_0}{\delta_2^2} dx$$

(10)

**Remark 1.** We are (locally) back to the case mentioned at the beginning of this section and treated in Iglehart and Witt (1970).

**Remark 2.** We could use the same local scale change in subcase (2.4), without reflection of course.

**Case 6:** A simple case where the input rate is quadratic in time has been briefly analysed in Newell (1968). Our general model leads to more diversified behaviours. The essential point is to determine the time $t^*$ at which the queue becomes empty (after that, the non-Markovian case 7 prevails). Of course, $t^* \sim t_2$, but we must investigate their asymptotic difference.

(6.1) Let us first consider $A^0$. Fix a time $t_3$, just before $t_2$ such that $\Pr[Q_n(t_3) \leq 0] \to 0$. For instance, let $t - t_2 = u/n^{1/2-\epsilon}$, ($\epsilon$ small positive) and let $t_3$ correspond to $u = -1$.

Case 5, easily leads, by (6), to

$$[Q_n(u) - 1 + \kappa n^{1/2+\epsilon}]/(\sqrt{n}\delta_3) \Rightarrow y,$$

(in the $u$-scale) where $\kappa := f(t_2) - \mu_5(\kappa < 0)$, $\delta_3 := F(t_2)[1 - F(t_2)] + \gamma^2_t t_2$ and $y$ is $\mathcal{N}(0, 1)$. From time $-1$ on, we now clearly have

$$[Q_n(u) - \sqrt{n}\delta_3 - \kappa u n^{1/2+\epsilon}]/(\delta_4 n^{1/4+\epsilon/2}) \Rightarrow B_\gamma(u + 1)$$

for some BM, $B_\gamma$, where $\delta_4 := f(t_2) + \gamma^2_S$.

We easily find that the time $u^*$ at which $Q_n(u)$ returns to zero is asymptotically given by $u^* \sim y\delta_3/(n^\epsilon \kappa)$ which leads to $\sqrt{n}(t^* - t_2) \Rightarrow y\delta_3/\kappa$, independent of $\epsilon$ (as it should).
(6.2) Let us now consider some \( A^i (i > 0) \), starting with case 8. The limiting process depends on whether or not we know the number of customers who have not yet arrived at \( t_1 \).

- If the input population \( \bar{n}(t_1) \) remaining at \( t_1 \), is unknown, the analysis proceeds as in subcase (6.1), with \( \delta_5^2 \) replaced by

\[
\delta_5^2 := \Delta F (1 - \Delta F) + \chi_5^2 |A^i|,
\]

where \( \Delta F := F(t_2) - F(t_1) \).

We obtain

\[
\sqrt{n(t^* - t_2)} \Rightarrow \nu \delta_5 / \kappa.
\]

- If \( \bar{n}(t_1) \) is known, then let

\[
n^* = \bar{n}(t_1) / [1 - F(t_1)].
\]

From (2) we see that \((n^* - n) / \sqrt{n} \Rightarrow -B_0[\tau(t_1)]\).

Let

\[
z := B_0[\tau(t_1)].
\]

Knowing \( \bar{n}(t_1) \) amounts to fix \( z \). Proceedings as in subcase (6.1), we verify that, conditioned on \( z \), \([Q_n(-1) + \kappa n^{1/2 + \epsilon} - \sqrt{n \Delta F}]/(\sqrt{n} \delta_5 \Rightarrow \nu \) (in the \( u \)-scale) with \( \delta_5^2 := \Delta F [1 - \Delta F/(1 - F(t_1))] + \chi_5^2 |A^i|, \nu = \mathcal{N}(0,1) \) and also that

\[
\sqrt{n(t^* - t_2)} \Rightarrow \frac{\nu \delta_5 + z \Delta F}{\kappa}.
\]

Finally we must consider the queue backlog at \( t_1 \). It will be seen in case 8 that this backlog is \( O(n^{1/3}) \) and it is asymptotically negligible in \( Q_n(-1) \).

(6.3) Let us now analyse some \( A^i (i > 0) \), starting with case 9. The backlog at \( t_1 \) can take two forms: either subcase (9.1.1) and the contribution is \( O(n^{1/4}) \) (negligible in \( Q_n(-1) \)), or (9.1.2), with a contribution \( z \delta_{5,i-1} / \kappa \) (\( z = \mathcal{N}(0,1) \) and \( \delta_{5,i-1} \) is given by (11) for \( A^i \)). In the latter subcase, we must add \( z \delta_{5,i-1} / \kappa \) to (12) or (15). For the more classical time-dependent M/M/1 queue, see Mandelbaum and Massey (1993, Theorem 3.7) for an "end of overloading" analysis.

Case 8: \( \gamma > 0 \).

(8.1) Let us first start in the region defined by

\[t - t_1 = u/n^\alpha, \quad u < 0, \quad 1/5 < \alpha < 1/3.\]

We are clearly in the same situation as in subcase (3.4); a local time-scale change leads to a stationary reflecting BM.

(8.2) If \( (t - t_1) = O(1/n^{1/3}) \), let \( t = v/n \). In the \( v \)-scale, the trend is given by \( 2\gamma v/n \) and the infinitesimal variance by \( \delta_5^2 \). As \( n \) is large, we are clearly in the situation called "transition behaviour" in Newell (1968a) with \( b(0) = \delta_5^2 \) and \( \tilde{x} = 2\gamma v/n \) (\( \tilde{x} \) is written for
\(\alpha\) in Newell’s notation). Newell (1968a, b) defines two new units of time and queue length:

\[
T = b(0)^{1/3} / \bar{t}^{2/3} = \delta_2^{2/3} n^{2/3} / (2\gamma)^{2/3},
\]

\[
L = \delta_2^{4/3} n^{1/3} / (2\gamma)^{1/3}, \tag{16}
\]

which, here, amounts to the time-scale change: \(t = u \delta_2^{2/3} n^{1/3} (2\gamma)^{2/3}\). In these new scales, the density satisfies the diffusion equation

\[
\frac{\partial f(x, u)}{\partial u} = -u \frac{\partial f}{\partial x} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}, \tag{17}
\]

with the following constraints on the DF \(F(x, u)\): \(F(0, u) = 0, \forall u\) and \(F(x, 0) = 1, \forall x\). Newell (1968a) has analysed (17) with a numerical discrete simulation and gives in Fig. 2 the mean queue length, which goes smoothly from the stationary mean related to (10) (case (3.4)) to the Gaussian mean given in our subcase (2.4), augmented by 0.95 (in \(L\) unit). Newell (1968a) in Fig. 3, gives the variance related to (17).

\(k\)-order contact \((k \geq 2)\) for \(M/M/1\) is analysed in Massey (1985) and in Mandelbaum and Massey (1993).

(8.3) With a small probability, the queue characterized by the density (17) could return to zero in the transition region (8.2). Accordingly, we would then return to subcase (2.1) or (2.2), depending on the hitting time value.

(8.4) As soon as we are in the region defined by \(t - t_1 = u/n^2, u > 0, 1/5 < \alpha < 1/3,\) we are back in subcase (2.4) and later on, in case 5.

Case 9 \((\gamma > 0)\): We must first determine either the time \(t^* < t_1^\prime\) at which the queue becomes empty, if such \(t^*\) exists, or else analyse the value of \(Q_n(t_1^\prime)\). The conditioning on \(n^s\) will then be considered.

(9.1) Let \(t - t_2 = u/n^{(1/4 - \epsilon)}\) (\(\epsilon\) small positive) and start from \(u = -1\). This ensures that, asymptotically, the queue is strictly positive. We can easily deduce that

\[
[Q_n(1) - \gamma n^{1/2 + 2\epsilon}] / (\sqrt{n} \delta_5) \Rightarrow \gamma,
\]

where \(\delta_5\) is defined by (11) and \(\gamma\) is \(\mathcal{N}(0, 1)\). Two subcases will be considered.

(9.1.1) If \(\gamma < 0\), this leads, after a few manipulations, to

\[
n^{1/4}(t_2 - t^*) \Rightarrow [ - y \delta_5 / \gamma ]^{1/2}, \quad t^* < t_2.
\]

From that time on, we are in case (8.1) (with \(z = 1/4\)).

(9.1.2) If \(\gamma > 0\), this leads to \(Q_n(0) / \sqrt{n} \Rightarrow y \delta_5\) and from time \(u = 0\), we start the queue like in case (2.4), with \(z = 1/4\) and \(Q_n(0)\) as initial value (no reflection).

(9.2) As in case (6.2), assume that \(n^*\) (i.e. \(z\)) is known at \(t_1^\prime\). Leaving out the details, we conclude that it is enough to replace, in case (9.1), \(y \delta_5\) by \(y \delta_6 + z \Delta F\).

Our most interesting results can be summarized in the following theorem.
Table 2

<table>
<thead>
<tr>
<th>Case</th>
<th>$\alpha$ range</th>
<th>$E Q_n(u)$</th>
<th>$a$</th>
<th>$Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1 &gt; \alpha &gt; 1/3$</td>
<td>$u[F(t) - \mu_1^1]$</td>
<td>$(1 - z)/2$</td>
<td>$\delta_1 B_4(u)$</td>
</tr>
<tr>
<td>2.2</td>
<td>$1/3 &lt; \alpha &lt; 1$</td>
<td>$u^{1 - x}(v - \mu_1^1)u$</td>
<td>$(1 - \alpha)/2$</td>
<td>$BR_0(0, \delta_1, u)$</td>
</tr>
<tr>
<td>2.3</td>
<td>$\alpha = 1/3$</td>
<td>0</td>
<td>$1/2$</td>
<td>$BR_0(2; u, \delta_2, u)$</td>
</tr>
<tr>
<td>2.4</td>
<td>$1/5 &lt; \alpha &lt; 1/3$</td>
<td>$2x^2 y^{2u^2}$</td>
<td>$(1 - \alpha)/2$</td>
<td>$\delta_1 B_4(u)$</td>
</tr>
<tr>
<td>3.2</td>
<td>$1/5 &lt; \alpha &lt; 1/3$</td>
<td>0</td>
<td>$1/2$</td>
<td>$BR_0(2; u, \delta_2, u)$</td>
</tr>
<tr>
<td>3.3</td>
<td>$\alpha = 1/3$</td>
<td>0</td>
<td>$1/3$</td>
<td>$BR_1(2; u, \delta_2, u)$</td>
</tr>
<tr>
<td>3.4</td>
<td>$1/5 &lt; \alpha &lt; 1/3$</td>
<td>0</td>
<td>$1/4$</td>
<td>$BR_2(2; u, \delta_2, u)$</td>
</tr>
</tbody>
</table>

Theorem 4.1. Let $t = u/n^2$. The main cases of Table 1 lead to $[Q_n(u) - E Q_n(u)]/n^\alpha \sim Z(u)$, where the different parameters are given by Table 2 (with details given above).

5. Queue size maximum

5.1. Basic results

This random variable is basic to resource allocation, specially in the computer centre model. Case 5 obviously leads to the main contribution. From (6) we see that we must deal with a non-Markovian–Gaussian process.

We shall use here a technique based on Daniels’ results (1985, 1989). Consider a Gaussian process $Z(t)$ superimposed on a curve $\tilde{y}(t)$. If we look for $\mathcal{M} := \max[Z(t) + \tilde{y}(t)]$ and the time $t^*$ at which this maximum occurs, it is equivalent to search for the hitting time of $Z(t)$ to the absorbing boundary $\mathcal{M} - \tilde{y}(t)$. It is well known (see Durbin (1985)), that, near the crossing point, $Z(t)$ behaves locally like a BM (or a variant of it, such as a Brownian bridge BB). It is also known that the hitting time and place densities for a BB can be deduced from the hitting time density for a BM (see for instance Louchard (1984) for a constant boundary and Csaki et al. (1987) for a general proof). Assume that $\tilde{y}(t)$ is given by

$$\tilde{y}(t) = \sqrt{n} y(t), \quad n \geq 1$$

and that it has a unique maximum at $\tilde{t}$, with $y(\tilde{t}) = 0$. Daniels and Skyrme (1985) have computed the asymptotic densities of the hitting time and hitting place.

In the Gaussian process case, with covariance $C(s, t)$, $s \leq t$, Daniels (1989) has matched the local behaviour of $C(s, t)$ with the BM (or one of his variants) covariance near $\tilde{t}$. In the BB match, we have

$$[Z(t) + \sqrt{n} y(t)] \sim \sqrt{A} [BB(t - t_0) + \sqrt{n} \tilde{y}(t - t_0)] \quad \text{on} \quad t \in (t_0, t_0 + T),$$

on $t \in (t_0, t_0 + T)$,
with $BB(T) = 0$, $y(t) \equiv \sqrt{A} \gamma(t - t_0)$ and $A$ is some constant. We can deduce the distribution of the maximum $\mathcal{M}$ and the time $t^*$ of the maximum from Daniels (1989) (3.8) and Daniels and Skyrme (1985) (5.9). They have obtained the following results:

Let

$$c_1 := \left[ \partial_s C(s,t) \right]_t \geq 0, \quad c_2 := \left[ \partial_t C(s,t) \right]_t \leq 0, \quad c := C(\bar{t}, \bar{t}),$$

$$A := c_1 + |c_2|,$$

$$t_0 := \bar{t} - c/c_1, \quad t_0 + T := \bar{t} + c/|c_2|,$$

$$T := cA/(c_1 |c_2|),$$

$$B := -y''(\bar{t}),$$

$$u := n^{1/3} A^{-1/3} B^{2/3} (t^* - t),$$

$$F(x) := \exp(x^3/6) G(x),$$

$$G(x) := \frac{1}{2\pi i} \int_{C} e^{sx} \frac{ds}{Ai(-2^{1/3}s)},$$

$$\lambda := \int_{-\infty}^{\infty} [F(x)/x^+] dx = 0.99615. A direct new justification of (20) is given in Louchard et al. (1991).

Then, $\mathcal{M}$ is asymptotically Gaussian with mean

$$E(\mathcal{M}) = \lambda n^{-1/6} A^{2/3} B^{-1/3} + O(n^{-1/3})$$

and variance

$$\sigma^2 \mathcal{M} = c + O(n^{-1/3}).$$

The conditioned maximum $\mathcal{M} | t^*$ is asymptotically Gaussian with mean

$$E(\mathcal{M} | t^*) = n^{-1/6} A^{-1/3} [c_1 v(-u) + |c_2| v(u)] B^{-1/3} + O(n^{-1/3}),$$

$$\sigma^2[\mathcal{M} | t^*] = c + O(n^{-1/3}).$$

The joint density of $\mathcal{M}$ and $t^*$ is given by

$$\varphi(\mathcal{M}, u) du$$

$$= \sqrt{2} \exp\left(-m^2/(2c)\right) \left[ f(u) \frac{1}{2} + n^{-1/6} B^{-1/3} A^{-1/3} \mathcal{M} \left[ -\frac{1}{2} u^2 f(u) \frac{A}{c} \right]ight.$$\n
$$+ F'(u) F(u) \frac{c_1}{c} + F(-u) F'(u) \frac{|c_2|}{c} \left] + O(n^{-1/3}) \right\} d\mathcal{M} du$$

while $u$ has density

$$f(u)(1 + O(n^{-1/3})).$$
5.2. Queue size maximum

Let \( X(t) := A(t) - \chi_2B_2(t) \). Let us first remark that the order of the error term in (6) is well known: the relative error in the density is \( O(1/\sqrt{n}) \) (non-uniform in \( X \)). Let \( z(t) := F(t) - \mu_st \) and \( \tilde{t} \) such that \( z'(\tilde{t}) = 0 \), i.e. \( f(\tilde{t}) = \mu_s \) (assume for simplicity that \( \tilde{t} \) is unique). Rewrite (6) as

\[
Q_n(t) \sim \sqrt{n} \left\{ \sqrt{n}z(\tilde{t}) + X(t) + \sqrt{n}[z(t) - z(\tilde{t})] \right\} + O(1).
\]

The \( O(1) \) term is non-uniform in \( X \) but it is easy to check that \( MQ := \max_{0 \leq t \leq t_2} Q_n(t) \) is only affected by an \( O(1) \) term.

Comparing with (18) we must identify \( y(t) \) with \( z(t) - z(\tilde{t}) \). We can now compute, from (2) and (6)

\[
C(s, t) = F(s)[1 - F(t)] + \chi_s, \quad s \leq t
\]

This readily gives

\[
c = C(\tilde{t}, \tilde{t}) = F(\tilde{t})[1 - F(\tilde{t})] + \chi_S\tilde{t},
\]

\[
c_1 = f(\tilde{t})[1 - F(\tilde{t})] + \chi_S > 0,
\]

\[
c_2 = -f(\tilde{t})F(\tilde{t}) < 0,
\]

\[
A = c_1 + |c_2|, \quad B = -f'(\tilde{t}).
\]

The basic results of Section 5.1 leads to the following theorem.

**Theorem 5.1.**

\[
MQ := \max_{0 \leq t \leq t_2} Q_n(t) \sim nz(\tilde{t}) + \sqrt{n}\mathcal{M} + O(n^{1/6}),
\]

where \( \mathcal{M} \) is a RV characterized by (21) – (24). \( t^* \) is a random variable characterized by (19), (25) and (26). The constants are given in (27).

5.3. Other subcases

In subcase (5.1), we consider an interval \( A', i > 0 \), starting with case 8. By Newell (1968a) (see our subcase (8.2)) we know that we must add to \( MQ \) a constant given by \( L \) is given in (16))

\[
0.95L = 0.95\delta_2^{1/3}n^{1/3}/(2\gamma)^{1/3},
\]

which is of the same order as \( \sqrt{n}E(\mathcal{M}) \).

In the subcase (5.2), we consider an interval \( A' (i > 0) \), starting with case 9. Two subcases have been analysed. subcase (9.1.1) leads to the same correction as (28). Subcase (9.1.2) shows that we must add to \( MQ \) a constant given by \( y\delta_3\sqrt{n} \).
6. Approximations for the total unfinished work at $t$, $U_n(t)$

Two different expressions for $U_n(t)$ will lead to sums of stochastic integrals, BM or multiples of $Q_n(t)$. The first approach uses the reflection mapping on the total service time, the other one is related to the customers-in-queue service times.

6.1. Basic models

We will use two approaches ($\mathcal{A}$ and $\mathcal{B}$) in our analysis, each shedding some particular light on the stochastic behaviour of $U$. The following expressions are readily obtained

$\bullet \mathcal{A}$

$$U_n(t) = \mathcal{A} \left[ J_{U,n}(t) \right], \quad \mbox{(29)}$$

where

$$J_{U,n}(t) := \sum_{i=1}^{A_n(t)} S_i - nt,$$

$S_i$ denotes the service time of customer $i$.

$\bullet \mathcal{B}$

$$U_n(t) = \sum_{i=A_n(t) - [Q_n(t) - 1]}^{A_n(t)} \left[ S_i - m_S \right] + m_S [ Q_n(t) - 1 ]^+. \quad \mbox{(28)}$$

$$= \Sigma_1 + \Sigma_2, \quad \mbox{say}$$

We first remark that, in our diffusion approximation, we can clearly replace $[Q_n(t) - 1]^+$ by $Q_n(t)$. Also if

$$\frac{E[A_n(t)]}{n^a} \rightarrow F(t),$$

$$\frac{E[Q_n(t)]}{n^b} \rightarrow H(t),$$

$$\frac{Q_n(t) - E Q_n(t)}{C n^c} \Rightarrow Z(t),$$

where $a, b, c, C, F(t) > 0$ are finite, and $Z$ is some BM or RBM, we can use the random time-change theorem (see Iglehart (1974), Theorem 2.10) and Coffman and Reiman (1982, p. 9). We can prove that

$$\frac{\Sigma_1}{n^{c/2}} \Rightarrow 0 \quad \mbox{if } c > \max \{a, b\}. \quad \mbox{(30)}$$
The tightness problem for $U_n$ has been completely solved by Iglehart and Whitt (1970, Section 6), so we will not discuss it here. Finally if $c = b > a$ (no reflection in this case), and using again the random time-change theorem, we find that

$$
\frac{\Sigma_1}{n^{3/2}} \Rightarrow \int_{F(a) - H(a)}^{F(b)} \sigma_s \, dB_3(v),
$$

where $B_3$ is the same BM as in (5). Indeed, the variable $v$ corresponds to the (normalized) deterministic rank number of customers in the input process and, because of FIFO rule, each customer has the same rank number in both (input or service renewal) processes. Let $v = F(s)$, (31) becomes

$$
\frac{\Sigma_1}{n^{3/2}} \Rightarrow \int_{t^*}^t \sigma_s \sqrt{f(s)} \, dB_8(s) \quad \text{for some BM, } B_8,
$$

where $t^*(t)$ is defined by $F(t^*) = F(t) - H(t)$ and $s$ corresponds now to the (normalized) deterministic arrival time of each customer (with $\sigma_s dB_8(s)$ corresponding to its service time). Note that $t^*$ corresponds to the deterministic arrival time of the last served customer and that $B_8$ is derived from $B_3$ after a deterministic time change.

6.2. Analysis of some typical cases

We restrict our attention to some important cases: other situations can be treated by similar techniques.

Case 5

- Using the approach developed by Louchard (1988) (see in particular, Section 4.2), we obtain, from (2) (no reflection needed):

$$
\left[ U_n(t) - nmS [F(t) - \mu t] \right] / \sqrt{n} \Rightarrow mS \tilde{A}(t) + \int_0^{t^*} \sigma_s \sqrt{f(s)} \, dB_8(s),
$$

where $B_8$ is the same BM as in (32).

- Using (6), (31) with $c = b = 1, a = 1/2$ and (32) we observe that

$$
\frac{U_n(t)}{\sqrt{n}} - \sqrt{nmS [F(t) - \mu t]} \Rightarrow \int_0^{t^*} \sigma_s \sqrt{f(s)} \, dB_8(s) + mS [\tilde{A}(t) - \gamma S B_2(t)],
$$

with

$$
F(t^*) = F(t) - [F(t) - \mu t] = \mu t.
$$

Also $\gamma S B_2(t)$ is equivalent (see (5)) to $-\int_0^{\mu t} \sigma_s \mu S \, dB_3(v)$. Letting again $v = F(s)$, we obtain as in (32)

$$
\gamma S B_2(t) = -\int_0^{t^*} \sigma_s \mu S \sqrt{f(s)} \, dB_8(s),
$$

which proves the equivalence of (33) and (34).
Case 1

$\mathcal{A}$ Again let $t = u/n^\alpha$, $1 > \alpha > \frac{1}{3}$. With (7) (no reflection) we readily obtain

$$U_n(u) - m_5 v u n^{1-x} + u n^{1-x}]/n^{(1-\alpha)/2} \Rightarrow m_n \sqrt{v} B_1(u) + \sigma_n \sqrt{v} B_8(u).$$

Here $B_8$ is the same BM as in (32), seen in the $u$-scale. The right-hand side can of course be condensed into a single BM.

$\mathcal{B}$ We now deduce that

$$U_n(u)/n^{(1-\alpha)/2} - m_n (v - \mu_5) u n^{(1-x)/2} \Rightarrow$$

$$\sigma_n \sqrt{v} [B_8(u) - B_6(u*)] + m_n \sqrt{v} [B_1(u) - 2 B_2(u)].$$

Again, using (36), the identification with $\mathcal{A}$ is easy. We remark that here (35) gives

$$\gamma u^*/n^\alpha + \gamma(u*)^2/n^2 \sim \mu_5 u/n^\alpha$$

from which it follows that

$$u^* = \frac{\mu_5}{v} u + O(1/n^2). \quad (37)$$

Case 2

$\mathcal{A}$ Let $t = u/n^\alpha$, $1 > \alpha > \frac{1}{3}$. From (8), we deduce that

$$[J_{v,a}(u) - m_5 \gamma u^2 n^{1-2x}/n^{(1-x)/2} \Rightarrow m_n \sqrt{\mu_5} B_1(u) + \sigma_n \sqrt{\mu_5} B_8(u).$$

Note that $\gamma$ does not appear in the coefficient of $B_8$ since its contribution is asymptotically negligible. Now the application $U_n(u) = \mathcal{B}[J_{v,a}(u)]$ proceeds exactly as in case 2 for $Q_n(u)$.

$\mathcal{B}$ Applying (30) to the following subcases of Section 4, case 2, we find that

$$\begin{align*}
(2.2) & \quad b = (1 - x)/2, \quad a = (1 - x)/2 \quad (1/3 < x < 1) \\
(2.3) & \quad b = 1/3, \quad a - 1/3 \quad (x - 1/3) \quad (c = 1 - x).
\end{align*}$$

We see that $c = 1 - x > \max[a,b]$ in all subcases and the contribution of $\Sigma_1/n^{(1-x)/2}$ is asymptotically negligible. We are thus left with

$$[U_n(u) - m_n Q_n(u)]/n^{1-x/2} \Rightarrow 0. \quad (38)$$

But in (9), the term $-2 \mu_3 B_2(u)$ is equivalent (see (36)) to $\sigma_n \mu_5^{3/2} B_6(u*).$ By (37) (with $\mu_5 = v$), this asymptotically gives $\sigma_n \mu_5^{3/2} B_8(u),$ identification with $\mathcal{A}$ is now obvious. A similar result can be found in Reiman (1983, Theorem 4) for the particular heavy traffic situation mentioned at the beginning of Section 4.

Our results can be summarized in the following theorem.
Table 3

<table>
<thead>
<tr>
<th>Case</th>
<th>$\alpha$ range</th>
<th>$E U_n(u)$</th>
<th>$a$</th>
<th>$Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$\alpha = 0$</td>
<td>$\nu m_5[F(t) - \mu_S]$</td>
<td>1/2</td>
<td>$m_5A(t) + \int_0^t \sigma_5 \sqrt{f(s)} dB_5(s)$</td>
</tr>
<tr>
<td>1</td>
<td>$1 &gt; \alpha &gt; 1/3$</td>
<td>$n^{1/2} m_5(S - \mu_S)u$</td>
<td>$(1 - \alpha)/2$</td>
<td>$m_5 \sqrt{\nu} B_1(u) + \sigma_5 \sqrt{\nu} B_5(u)$</td>
</tr>
<tr>
<td>2</td>
<td>$1 &gt; \alpha &gt; 1/5$</td>
<td>$[U_n(u) - m_5Q_n(u)]/n^{1 - \alpha/2}$</td>
<td>$\Rightarrow 0$</td>
<td></td>
</tr>
</tbody>
</table>

Theorem 6.1. Let $t = u/n^2$. The main cases of Table 1 lead to

$$[U_n(u) - E U_n(u)]/n^a \Rightarrow Z(u)$$

where the different parameters are given by Table 3.

7. Approximation for $M_n(t)$

Suppose that, in a computer system, each customer asks for some storage $M$. The variable of interest is $M_n(t)$, the total storage occupied at time $t$.

We consider successively independent and dependent storage demand. We again obtain mixtures of BM, Gaussian process and stochastic integrals or multiple of $Q_n(t)$. The two approaches lead to different views on the related stochastic processes.

7.1. Basic models

Our two approaches become

- $A$

$$M_n(t) = \mathcal{R} \left\{ \sum_{i=1}^{A_n(t)} M_i - \sum_{i=1}^{S_n(t)} M_i \right\},$$

where $M_i$ denotes the storage demand of customer $i$.

- $B$

$$M_n(t) = \sum_{i=A_n(t)-[Q_n(t)-1]}^{A_n(t)} [M_i - m_M] + m_M [Q_n(t) - 1]^+. \quad \sum_1 + \sum_2,$$

We will first analyse independent and then dependent storage demand.

7.2. Independent storage demand

We assume here that the storage demand $M_i$ of customer $i$ is independent of $S_i$, with mean $m_M$ and $\text{VAR} \sigma_M^2$. In approach $B$, (30) still holds. (31) is also applicable, with right-hand side

$$\int_{F(t)-H(t)}^{F(t)} \sigma_M dB_5(v)$$

(39)
for some BM $B_9$ such that $\sigma_M dB_9(v)$ corresponds to the $i$th customer storage demand ($v := (i/n)$ is the same variable as in (31)). Letting $v = F(s)$ and $t^*(t)$ is defined by $F(t^*) = F(t) - H(t)$, we can reduce this, by deterministic time change, to

$$\int_0^{t^*} \sigma_M \sqrt{f(s)} dB_{10}(s)$$

(40)

for some BM $B_{10}$ (same variable $s$ as in (32) and (36)), $\sigma_M dB_{10}(s)$ corresponding to the storage demand of the customer who did arrive at time $s$.

Case 5

• $\mathcal{A}$ From (A.4) and (A.5) we obtain

$$\left[ \sum_{i=1}^{s_n(t)} M_i - n \mu_S m_Mt \right] / \sqrt{n} \Rightarrow -m_M \int_0^{\mu_S t} \sigma_S \mu_S dB_{3}(v) + \int_0^{\mu_S t} \sigma_M dB_{9}(v).$$

As in (33), it follows that

$$[M_n(t) - nm_M[F(t) - \mu_S t]]/\sqrt{n} \Rightarrow \left[ m_M \tilde{A}(t) + \int_0^{t} \sigma_M \sqrt{f(s)} dB_{10}(s) \right] \Rightarrow \left[ m_M \int_0^{\mu_S t} \sigma_S \mu_S dB_{3}(v) + \int_0^{\mu_S t} \sigma_M dB_{9}(v) \right].$$

(41)

Now, letting $v = F(s)$, we obtain as in (40) (use (35)):

$$\int_0^{\mu_S t} \sigma_M dB_{9}(v) \equiv \int_0^{t^*} \sigma_M \sqrt{f(s)} dB_{10}(s).$$

The second and fourth terms in (41) give $\int_{t^*} \sigma_M \sqrt{f(s)} dB_{10}(s)$.

• $\mathcal{A}$ From (40) we deduce that

$$\frac{M_n(t)}{\sqrt{n}} - \sqrt{nm_M[F(t) - \mu_S t]} \Rightarrow \int_{t^*}^{t} \sigma_M \sqrt{f(s)} dB_{10}(s) + m_M[\tilde{A}(t) - \chi_S B_2(t)].$$

From (5), identification with $\mathcal{A}$ is immediate.

Case 1

• $\mathcal{A}$ Proceeding as in Section 6.2, ($1 > \alpha > \frac{1}{2}$) we readily obtain (with $B_3, B_9$ and $B_{10}$ seen in the $u$-scale):

$$[M_n(u) - m_M(v - \mu_S)u_n^{1-\alpha}] / n^{(1-\alpha)/2} \Rightarrow [m_M \sqrt{v} B_1(u) + \sigma_M \sqrt{v} B_{10}(u)] + [m_M \sigma_S \mu_S B_3(\mu_S u) - \sigma_M B_9(\mu_S u)].$$

Second and fourth terms are obviously equivalent to $\sigma_M \sqrt{v}[B_{10}(u) - B_{10}(u^*)]$, with $u^* := (\mu_S/v)u$. (see (37)).
We easily deduce that
\[
\frac{[M_n(n)/n^{1-\alpha/2}] - m_M(v - \mu_3)un^{(1-\alpha)/2}}{\sigma_M \sqrt{\nu[B_10(u) - B_{10}(u^*)] + m_M[\sqrt{\nu}B_1(u) - \chi_S B_2(u)]}}
\]

From (5), identification with \(A\) is immediate.

Case 2: We can prove that (see (38)):
\[
\frac{[M_n(u) - m_M Q_n(u)]/n^{(1-\alpha)/2}}{\sigma_M} = 0, \quad 1 > \alpha > \frac{1}{2}.
\]

Note that the variance \(\sigma_M^2\) does not enter into the limit. (cf. the similar situation in Coffman and Reiman, 1983). We summarize our results on the independent case in the following theorem.

**Theorem 7.1.** Let \(t = u/n^2\). Then \([M_n(u) - EM_n(u)]/n^a \Rightarrow Z(u)\), where the different parameters are given in Table 4.

<table>
<thead>
<tr>
<th>Case</th>
<th>(\alpha) range</th>
<th>EM(_n)(u)</th>
<th>(a)</th>
<th>(Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>(\alpha = 0)</td>
<td>(nm_M[F(t) - \mu_3]&lt;))</td>
<td>1/2</td>
<td>(m_M[\hat{A}(t) - \chi_S B_2(t)])</td>
</tr>
<tr>
<td>1</td>
<td>(1 &gt; \alpha &gt; 1/3)</td>
<td>(m_M(v - \mu_3)un^{1-\alpha})</td>
<td>(1 - (\alpha))/2</td>
<td>(m_M[\sqrt{\nu}B_1(u) - \chi_S B_2(u)]) + (\sigma_M \sqrt{\nu[B_10(u) - B_{10}(u^*)]})</td>
</tr>
<tr>
<td>2</td>
<td>(1 &gt; \alpha &gt; 1/5)</td>
<td>([M_n(u) - m_M Q_n(u)]/n^{(1-\alpha)/2})</td>
<td>(0)</td>
<td></td>
</tr>
</tbody>
</table>

### 7.3. Dependent storage demand

We assume that the conditioned mean storage demand is given by \(m_N(\xi) := E[M|S = \xi]\), \(m_M\) still denoting the unconditioned mean storage demand and the conditioned variance being given by \(\sigma_M^2(\xi)\). Note that the unconditioned variance is now given by
\[
\sigma_M^2 = \int_0^\infty \left[ [m_M(\xi) - m_M]^2 + \sigma_M^2(\xi) \right] q(\xi) d\xi.
\]

Case 5

From (A.6) and (A.7) (with A.3) we obtain
\[
\left[ \sum_{i=1}^{S_n(t)} M_i - n\mu_Sm_M t \right] / \sqrt{n} \Rightarrow \int_0^{\mu_S^t} \sigma_M dB_{12}(v) - m_M \int_0^{\mu_S^t} \mu_S \sigma_S dB_3(v).
\]
The two BMs are correlated (see (A.8)). Let

\[ \int_0^{t^*} \sigma_M \, dB_{13}(v) = \int_0^{t^*} \sigma_M \sqrt{f(s)} \, dB_{13}(s) \quad \text{for some BM, } B_{13} \]

It follows that

\[
\begin{align*}
\left[ M_n(t) - nm_M[F(t) - \mu_S t] \right]/\sqrt{n} \\
\Rightarrow m_M \tilde{A}(t) + \int_0^{t^*} \sigma_M \sqrt{f(s)} \, dB_{13}(s) \\
+ m_M \int_0^{t^*} \mu_S \sigma_S dB_3(v) - \int_0^{t^*} \sigma_M dB_{12}(v).
\end{align*}
\]

Proceeding as in (40) and using (5), we find that

\[
\frac{M_n(t)}{\sqrt{n}} - \sqrt{nm_M[F(t) - \mu_S t]}
\Rightarrow \int_0^{t^*} \sigma_M \sqrt{f(s)} \, dB_{13}(s) + m_M \left[ \tilde{A}(t) + \int_0^{t^*} \mu_S \sigma_S dB_3(v) \right].
\]

The identification with \( \mathcal{A} \) is now obvious.

**Case 1:** It can be verified that (with \( 1 > \alpha > 1/3 \))

\[
\left[ M_n(u) - m_M(v - \mu_S) \mu u^{1-\alpha}/n^{(1-\alpha)/2} \right]
\Rightarrow m_M \left[ \sqrt{v} B_1(u) + \sigma_S \tilde{\mu}_S B_3(\mu_S u) \right] + \sigma_M \sqrt{v} \left[ B_{13}(u) - B_{13}(u^*) \right],
\]

where, again, \( B_3 \) and \( B_{13} \) are correlated.

**Case 2:** We now obtain

\[
\left[ M_n(u) - m_M Q_n(u) \right]/n^{(1-\alpha)/2} \Rightarrow 0, \quad 1 > \alpha > \frac{1}{2}.
\]

The following theorem summarizes our results on the dependent case:

**Theorem 7.2.** Let \( t = u/n^\alpha \). We obtain (with notations as in Theorem 7.1, the correlation between \( B_3 \) and \( B_{10} \) is given in (A.8)) (Table 5):

<table>
<thead>
<tr>
<th>Case</th>
<th>( \alpha ) range</th>
<th>( EM_n(u) )</th>
<th>( \alpha )</th>
<th>( Z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>( \alpha = 0 )</td>
<td>( nm_M[F(t) - \mu_S t] )</td>
<td>1/2</td>
<td>( m_M[\tilde{A}(t) - \sigma_S \tilde{\mu}_S B_3(\mu_S t)] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+ ( \int_0^{t^*} \sigma_M \sqrt{f(s)} , dB_{13}(s) )</td>
</tr>
<tr>
<td>1</td>
<td>( 1 &gt; \alpha &gt; 1/3 )</td>
<td>( m_M(v - \mu_S) \mu u^{1-\alpha}/n^{(1-\alpha)/2} )</td>
<td>( 1 - \alpha/2 )</td>
<td>( m_M[\sqrt{v} B_1(u) - \sigma_S \tilde{\mu}_S B_3(\mu_S u)] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+ ( \sigma_M \sqrt{v} \left[ B_{13}(u) - B_{13}(u^*) \right] )</td>
</tr>
<tr>
<td>2</td>
<td>( 1 &gt; \alpha &gt; 1/5 )</td>
<td>( [M_n(u) - m_M Q_n(u)]/n^{(1-\alpha)/2} )</td>
<td>( 0 )</td>
<td></td>
</tr>
</tbody>
</table>
8. Other variables of interest in the limiting processes

In this section, we briefly analyse some limiting results for some stochastic variables associated with the queueing model. These variables are related to RBM with constant finite trend (q say) such as those arising in subcases (2.2), (3.2), (3.4), (9.1), for \( Q_n \), in case 2 for \( U_n \) and \( M_n \), or in Iglehart and Witt (1970) and Coffman and Reiman (1983). To simplify notations, assume that each variance is unity so that each RBM in the sequel is of the form \( BR(q, 1, u) \). We consider successively the cost of a busy period, the idle time \( I_n(t) \) and the distribution of busy periods.

8.1. Cost of a busy period

The busy period of length \( L \) is the support of a Brownian excursion (BE). We remark that, in view of (A.15), for a fixed length \( l \) of the busy period, the distribution of the queue length is independent of the trend \( q \) and is exactly given by the distribution of the BE: \( \{ Y(u) \mid L = l \} \). Thus if we assume that the cost, at each time, is proportional to the queue length, the total cost is equal to the BE Area: \( \int_0^L Y(u) \, du \). This random variable has been investigated in Louchard (1984). It is easily seen that

\[
\left[ \int_0^l Y(u) \, du \mid L = l \right] \overset{d}{=} l^{3/2} \left[ \int_0^1 x(u) \, du \right],
\]

where \( x(u) \) is the standard scaled BE \( \int_0^1 x(u) \, du \) is the area of the BE. We obtained in Louchard (1984) an explicit form for the Laplace transform of the generating function of this area. Moments, numerical values for density and distribution function and asymptotic expressions for small argument were obtained by Louchard (1984). Takacs (1991) has obtained an explicit expression for the density.

8.2. Idle time \( I_n(t) \)

It is easily proved (see Coffman and Reiman (1983, p. 185) that

\[
I_n(t) \equiv - \inf_{0 \leq s \leq t} \left[ \sum_{i=1}^{A_n(t)} S_i - nt \right].
\]

Comparing this with (29) and (38), we observe that, in subcases (2.2), (3.2), (3.4), (9.1) (or even (2.3)), \( I_n(t) \), suitably normalized, converges weakly to the local time \( \widetilde{I}(t) \) of the corresponding RBM (see Chung and Williams, 1983, p. 146 for a recent proof). By (B.3), we obtain the following joint density (where the trend \( q \) is constant):

\[
E_0[BR(q, 1, u) \in dx, \tilde{I}(u) \in dy] = \frac{2(x + y)}{\sqrt{2\pi u^3}} \exp(-2yq - (x + y - qu)^2/2u) \, dx \, dy.
\]
The marginal densities for $BR(q, 1, u)$ and $\bar{T}(u)$ are, respectively, given by

$$
\phi(q, u, x) := \frac{2}{\sqrt{2\pi u}} \exp\left( - \frac{(x - qu)^2}{2u} \right) - 2qe^{2xy} \Phi\left[ - \frac{(x + qu)}{\sqrt{u}} \right]
$$

(this is well known: see Cox and Miller (1980, p. 224 or Harrison 1985, p. 15) and

$$
\psi(q, u, y) := \frac{2}{\sqrt{2\pi u}} \exp\left( - \frac{(y + qu)^2}{2u} \right) + 2qe^{-2xy} \Phi\left[ - \frac{(y - qu)}{\sqrt{u}} \right].
$$

If $u \to \infty$, (42) gives for $BR$ the stationary density $-2qe^{2x}$ if $q < 0$ and $\frac{2}{u} BR(q, 1, u) \to q$ if $q > 0$. Eq. (43) gives asymptotically for $\bar{T}$ the stationary density $2qe^{-2y}$ if $q > 0$ and $\bar{T}(u)/u \to -q$ if $q < 0$. Moments of the density (42) have been derived by Abate and Whitt (1987a, b).

8.3. Distribution of busy periods (constant trend $q$)

From (B.4), we find that

$$
t^{-1}(b) = \int_0^\infty l p([0, b] \times d l),
$$

where $t^{-1}(b) = \inf(s; \bar{T}(s) = b)$, and $p(db \times dl)$ is the Poisson measure with mean $db \exp(-q^2 l/2) dl/\sqrt{2\pi l^3}$. This decomposition ties the flat stretches of $\bar{T}$ with the open intervals $z_n (n > 1)$ of the complement of the set $z^+ = (t; BR(q, 1, t) = 0)$. As these $z_n$ are indeed the successive busy periods, we see that these busy periods are in some senses shorter when $q \neq 0$. Of course, if $q > 0$, the last busy period is infinite. From (B.6) and (42) we see that its starting time, $s$, (which is indeed the last exit time for $BR(q, 1, u)$) has the density $\phi(q, s, 0) \cdot q$. Another way to consider the busy period distribution is to write the joint density for the random variables:

$$
G(t) := \sup(s; s \leq t; BR(q, 1, s) = 0),
$$

$$
D(t) := \inf(s; s \geq t; BR(q, 1, s) = 0),
$$

$$
L(t) := D(t) - G(t).
$$

From (B.7), we deduce that

$$
E_{q}[G(t) \in ds, L(t) \in dl] = \phi(q, s, 0) ds \frac{\exp(-q^2 l/2) dl}{2\sqrt{2\pi l^3}}.
$$

For $q = 0$, this is equivalent to Lemma 3.1 of Cohen and Hooghiemstra (1981). Finally, the distribution of RBM between $G(t)$ and $t$ is equivalent to what is called a meandering process (see Chung (1968) for the trend-free case). From (B.5) we obtain by integration with respect to $u$

$$
E_{q}[G(t) \in ds, BR(q, 1, t) \in dy] \quad (s < t)
$$

$$
= \phi(q, s, 0) \cdot \frac{y}{\sqrt{2\pi(t - s)^3}} \exp\left[ - \frac{(y - q(t - s))^2}{2(t - s)} \right] ds dy. \quad (44)
$$
From (44) we can also derive the conditional density for $BR(q, 1, t)$ when $q < 0$ and $t \to \infty$ (i.e. for large busy period). This leads to

$$q^2 ye^{q^2} dy.$$

8.4. Absorbing barriers

The effect of an absorbing barrier of $BR$ has been analysed by Sweet and Hardin (1970). The density of the $BR$ before absorption and the hitting time are given by Sweet and Hardin (1970, Eqs. (2.19) and (2.23), respectively).

9. Conclusions

Using weak convergence theorems, we have shown that several stochastic variables associated with a finite population queueing system can be approximated by Brownian motions and stochastic integrals. To keep the paper reasonably short, we have not generalized (as in Louchard, 1988) our results to bulk arrivals. However, the techniques used in that paper can be adapted to this case. An other extension is to study networks and queues with more sophisticated queueing disciplines. As already noted by Newell it is astonishing that such a simple problem leads to so many developments. Here, we have for instance 7 different powers of parameter $n$ involved in our various approximations: $n, n^{1/2}, n^{1/3}, n^{1/5}, n^{1/6}, n^{1/4}, n^{3/5}$. As in our previous work, the problem of rate of convergence remains open. We intend in the future to try some simulations in order to check the quality of our approximations.

Acknowledgements

The author is indebted to W.A. Massey who kindly supplied a copy of Mandelbaum and Massey (1993) before publication. The pertinent comments of the referees led to enrichment of the bibliography and improvement in the presentation.

References


J. Abate and W. Whitt, Transient behaviour of the M/M/1 queue: Starting at the origin, Queueing system 2 (1987) 41–65.


Appendix A: Some results on renewal processes

In the Renewal process $S_n(t)$, let customer $i$ bring some random variable $V_i$ (with finite variance), depending on its service time $S_i$. More precisely, let

\[ [V|\xi] := [V|S = \xi], \]

\[ m(\xi) := E[V|\xi], \]

\[ m_V := E[V] = \int_0^\infty m(\xi) g(\xi) d\xi, \]

\[ \sigma^2(\xi) := \text{VAR}[V|\xi], \]

\[ \sigma^2_V := \text{VAR}[V] = \int_0^\infty [\sigma^2(\xi) + [m(\xi) - m_V]^2] g(\xi) d\xi, \]

\[ C_{V,S} := \text{COV}[V,S]. \]

We are interested in the random variables

\[ T_n(t) := \left( \sum_{i=1}^{S_n(t)} V_i - n\mu_S m_V t \right) / \sqrt{n}. \]

From Iglehart (1973, Lemma 2.4), we find (after slight simplification) that $T_n(t) \Rightarrow \tilde{T}(t)$, where

\[ \tilde{T}(t) := \int_0^{\text{ust}} \sigma_T dB_{11}(v) \]

for some BM, $B_{11}$ and

\[ \sigma^2_T := \int_0^\infty [\sigma^2(\xi) + [m(\xi) - m_V]^2] g(\xi) d\xi + m_V^2 \mu_S^2 \sigma^2_S - 2m_V \mu_S C_{V,S}. \]

But (A.1) is obviously not detailed enough for our purpose. To obtain a more detailed representation, we rewrite $T_n(t)$ as follows:

\[ T_n(t) = \sum_{i=1}^{S_n(t)} \left\{ [m(S_i) - m_V \mu_S S_i] + [V|S_i] - m(S_i)] \right\} / \sqrt{n} \]

\[ - \left\{ m_V \mu_S \left[ nt - \sum_{i=1}^{S_n(t)} S_i \right] \right\} / \sqrt{n}. \]

(A.2)
We see that the last term in (A.2) \( \Rightarrow 0 \) by Billingsley (1968, Theorem 4.1). Proceeding now as in Louchard (1988, Section 4.2 and Remark 5) and using the functional central limit theorem for random sums, we see after a few manipulations that the first two terms of (A.2) converge weakly to

\[
\bar{T}_1(t) + \bar{T}_2(t),
\]

where

\[
\bar{T}_1(t) := \int_{v=0}^{\mu_S t} \int_{\xi=0}^{\infty} \sqrt{g(\xi)} [m(\xi) - m_V \mu_S \xi] d_\xi \Psi(v, \xi),
\]

\[
\Psi(v, \xi) := G(\xi) \int_{s=0}^{\xi} \frac{\sqrt{g(s)}}{G(s)} d_e d_\xi BT_0(v, s),
\]

\[
\bar{T}_2(t) := \int_{v=0}^{\mu_S t} \int_{\xi=0}^{\infty} \sqrt{g(\xi)} \sigma(\xi) d_e d_\xi BT_1(v, \xi),
\]

\( BT_0, BT_1 \) are two-parameters BM and \( v \) is the normalized rank number \((i/n)\) of customer \( i \) \((S_i/n) \Rightarrow \mu_S t\). Hence \( \bar{T}_2(t) \) clearly corresponds to the (conditioned) fluctuations of \( V \) and \( \bar{T}_1(t) \) corresponds to the service fluctuations. Proceeding as in Louchard (1988, Eq. (34)), we can check that the variance of \( \bar{T}_1 + \bar{T}_2 \) is indeed given by \( \sigma_T^2 \); we now decompose \([m(\xi) - m_V \mu_S \xi]\) into \([m(\xi) - m_V]\) and \([m_V - m_V \mu_S \xi]\).

Four useful cases can be considered:

(i) \( V = 1 \). We immediately obtain

\[
\bar{T}_n(t) \equiv \bar{S}_n(t) \Rightarrow \bar{T}_1(t) = -\int_{v=0}^{\mu_S t} \int_{\xi=0}^{\infty} \sqrt{g(\xi)} \mu_S [\xi - m_S] d_\xi \Psi(v, \xi),
\]

which can obviously be condensed, when needed, as

\[
-\int_{0}^{\mu_S t} \mu_S \sigma_S dB_3(v) \quad \text{for some BM, } B_3
\]

(in this appendix, we use BM indexing as needed in the paper). Eq. (A.3) is actually another form of Iglehart (1974, Theorem 4.1).

(ii) \( V = S \). We see that \( \bar{T}_1 - \bar{T}_2 \equiv 0 \) which is, of course, what we expected.

(iii) \( V = M \), the customer’s storage demand, assumed to be independent of \( S \) so that \( m(\xi) \equiv m_V \equiv m_M \). Here \( \bar{T}_2(t) \) can be condensed as

\[
\int_{0}^{\mu_S t} \sigma_M dB_9(v) \quad \text{for some BM, } B_9
\]

and \( \bar{T}_1(t) \) is easily seen to become

\[
-m_M \int_{0}^{\mu_S t} \mu_S \sigma_S dB_3(v).
\]
(iv) \( V = M \), the customer's storage demand, assumed to depend on \( S \). The unconditioned variance of \( M \) is now given by

\[
\sigma_M^2 = \int_0^\infty \left[ [m_M(\xi) - m_M]^2 + \sigma_M^2(\xi) \right] g(\xi) \, d\xi.
\]

In \( \tilde{T}(t) \), it is sometimes convenient to separate the part, \( \tilde{T}_{M\sigma} \) (say), arising from the total variation of \( M \). We thus obtain

\[
\tilde{T}(t) = \tilde{T}_{M\sigma}(t) + \tilde{T}_{MS}(t).
\]

where

\[
\tilde{T}_{M\sigma}(t) := \int_0^{\mu_{st}} \int_0^\infty \left[ [m_M(\xi) - m_M] \, d\xi \, \Psi'(v, \xi) + \sqrt{g(\xi)} \sigma_M(\xi) \, dz \, BT_1(v, \xi) \right],
\]

(A.6)

\[
\tilde{T}_{MS}(t) := \int_0^{\mu_{st}} \int_0^\infty m_M \mu_S (m_M - \xi) \, d\xi \, \Psi'(v, \xi).
\]

(A.7)

Here \( \tilde{T}_{MS} \) can of course be written as (A.5) and \( \tilde{T}_{M}(t) \) can be expressed as

\[
\int_0^{\mu_{st}} \sigma_M \, dB_{12}(v) \quad \text{for some BM, } B_{12}
\]

which is correlated with \( \tilde{T}_{MS} \):

\[
E_0[\tilde{T}_{M\sigma}(t) \tilde{T}_{MS}(t)] = - m_M \mu_S C_{M,S}
\]

\[
= \lambda_{st} \int_0^\infty g(\xi) m_M \mu_S [m_M(\xi) - m_M] [m_S - \xi] \, d\xi.
\]

(A.8)

Appendix B: Some results on \( BR(q, 1, u) \)

To simplify notations, we keep \( q \) fixed in this section and write simply \( BR(u) \) for \( BR(q, 1, u) \). The ordinary BM with trend \( q \) will be denoted by

\[
Z(u) = B(u) + qu.
\]

Let us first recall some well-known results (see Cox and Miller, 1980, p. 221 or Harrison, 1985, p. 12). Let \( m_a := \inf\{s: Z(s) = a\} \). We have

\[
E_0[Z(u) \in dy, u < m_a] \quad (\xi > 0)
\]

\[
= \frac{1}{\sqrt{2\pi u}} \left[ \exp[-(y - qu)^2/2u] - \exp[2\xi q - (y - qu - 2\xi)^2/2u] \right] dy,
\]

(B.1)

\[
E_0[m_\xi \in du] = \frac{\xi}{\sqrt{2\pi u^3}} \exp[-(\xi - qu)^2/2u] \, du.
\]

(B.2)
We know that \( BR(u) \equiv Z(u) - \inf_{0 \leq s \leq u} Z(s) \). From (B.1), by simple calculations, we find that

\[
E_0 \left[ BR(u) \in dx, \inf_{0 \leq s \leq u} Z(s) \in dy \right] (y > 0) = \frac{2(x + y)}{\sqrt{2\pi u^3}} \exp\left[ -2yq - (x + y - qu)^2 / 2u \right] dx dy. \quad (B.3)
\]

When \( q = 0 \), we obtain Levy's classical result (see Ito and McKean, 1974, p. 45). For \( q > 0 \), the Laplace transform of (B.2) is easily seen to be \( \exp(-4\frac{q}{q^2 + 2x}) \), which can be transformed by standard calculations to

\[
\exp\left[ -\zeta \int_0^\infty (1 - e^{a_l}) \frac{e^{-q^{1/2}}}{\sqrt{2\pi l^3}} \, dl \right]. \quad (B.4)
\]

For \( q = 0 \), this reduces to the classical Levy's decomposition. If \( q < 0 \), it is easy to see that, conditioned on a finite hitting time, (B.2) leads exactly to the same decomposition.

We finally turn to the last exit time before \( t \). Let

\[
G(t) := \sup(s: s \leq t; BR(s) = 0)
\]

\[
D(t) := \inf(s: s \geq t; BR(s) = 0)
\]

Using Chung's approach (see Chung 1976; Louchard, 1984), we easily find, with \( q < 0 \):

\[
E_0 \left[ G(t) \in ds, BR(t) \in dy, D(t) \in du \right] = E_0 \left[ BR(s) \in d0 \right] \frac{y}{\sqrt{2\pi(t - s)^3}} \exp\left[ -(y - q(t - s))^2 / (2(t - s)) \right] \frac{y}{\sqrt{2\pi(u - t)^3}} \exp\left[ -(y + q(u - t))^2 / (2(u - t)) \right] ds dy du. \quad (B.5)
\]

Again, if \( q > 0 \), we obtain the same density but here the probability of returning to 0 (after \( t \)) is given by \( \exp(-2yq) \). If \( q > 0 \), the last exit time density is obtained by integrating (B.5) on \( y, u \) and letting \( t \to \infty \). The result is simply

\[
E_0 \left[ BR(s) \in d0 \right] q ds. \quad (B.6)
\]

Integration of (B.5) with respect to \( y \) leads of course to (for all \( q \))

\[
\frac{e^{-q^{1/2}}}{2\sqrt{2\pi l^3}} \quad \text{(with } l := u - s), \quad (B.7)
\]

which confirms (B.4).