Estimating probability distributions of dynamic queues

Nicholas B. Taylor and Benjamin G. Heydecker

Centre for Transport Studies, University College London, Gower Street, London WC1E 6BT

Queues are often associated with uncertainty or unreliability, which can arise from chance or climatic events, phase changes in system behaviour, or inherent randomness. Knowing the probability distribution of the number of customers in a queue is important for estimating the risk of stress or disruption to routine services and upstream blocking, potentially leading to exceeding critical limits, gridlock or incidents. The present paper focuses on time-varying queues produced by transient oversaturation during demand peaks where there is randomness in arrivals and service. The objective is to present practical methods for estimating a probability distribution from knowledge of the mean, variance and utilisation (degree of saturation) of a queue available from computationally efficient, if approximate, time-dependent calculation. This is made possible by a novel expression for time-dependent queue variance. The queue processes considered are those commonly used to represent isolated priority (M/M/1) and signal-like (M/D/1) systems, plus some statistical variations within the common Pollaczek-Khinchin framework. Results are verified by comparison with Markov simulation based on recurrence relations.

Keywords: queue, probability distribution, dynamic, uncertainty, reliability

Introduction

Queues are often associated with uncertainty, which can be:

- exogenous, as when an incident blocks a motorway, severe weather disrupts movement, or a signal failure halts rush-hour train movements;
- contingent, as when flow breakdown occurs at a transiently overloaded motorway merge, or conflicting pedestrian or vehicle streams reduce effective capacity (the distinction with 'exogenous' is a simplification because there could be elements of both; however, queues we characterise as 'contingent' tend to recur at the same places as a result of local geometry or traffic patterns);
- endogenous to the queuing process, as a result of randomness in the arrival stream or service process at a facility with a specific capacity, such as an urban road junction, airport, call centre or hospital A&E (ER) department.

This paper focuses on the third type of process, recognising that secondary effects of random queues, such as concentration of traffic, blocking of upstream junctions, or exceeding some other critical capacity limit, can precipitate further disruption or gridlock or increase the risk of incidents. In such cases, it is desirable to be able to estimate the risk of exceeding certain values, which implies knowing probability distributions of the number of customers in the queue (queue 'size', which may include those in service, although in some cases physical length may be more important). The theory of random queues is well developed for those processes which if left running for a long time are theoretically expected

to attain equilibrium, as well as for certain aspects of transient behaviour. Equilibrium does not mean that the system becomes static, because a queue is not normally dissipative but is a dynamic process fed continually with new arrivals. The sense in which equilibrium occurs is that, under repeated realisation of conditions, the probability distribution of queue size tends towards a particular form whose moments are predictable, although they may not be immediately evident from the system description and arise only in the limit as time or the number of events tend to infinity (Machta *et al* 2013). Equilibrium queue size is defined only where demand is less than capacity, and typically depends not on the absolute traffic flow but only on the degree of saturation (ratio of demand to capacity). However, static equilibrium is theoretical in several senses:

- whatever the duration of system observation, the exact theoretical form of its probability distribution and the theoretical values of its moments will never be attained;
- only a single queue size can be observed at a time, so an empirical probability distribution can be developed only over many repeated observations;
- arrival and service rates will generally not remain constant during such an experiment, and their underlying parameter values can never be known exactly;
- distributions measured over equal (possibly overlapping) periods, however long, will never be identical, so the system will never truly settle down to a steady state. Typically variance over time is comparable to the variance of the equilibrium distribution (ergodic property).

The paper summarises properties of random queues and traffic modelling that uses them, outlines approximate but efficient closed-form methods for estimating the development of queues, extended through the use of a novel formula for 'deterministic' time-dependent variance, and finally describes fitting equilibrium and dynamic probability distributions to queue moments. Together, these enable better understanding of when and to what extent unexpectedly long queues and consequent risks and disruptions may occur.

Random queuing theory and equilibrium mean results

How relevant to the real world is random queuing theory in the light of the four points above? Its usefulness can be investigated by comparing real results with simulations based on theory, demonstrated by its ability to explain observations, and justified by normative principles such as conservation of customers. The simplest queues are generated by customers arriving randomly at a constant mean rate, so their arrival headways are exponentially distributed and their numbers in intervals of equal duration are Poisson distributed, and discharged through service opportunities occurring at some rate independent of arrivals. There is no memory in this system, apart from in the queue itself, so its development can be followed iteratively, the next state depending on the current state but not on earlier states (Markovian property), and where service is random rather than phased only transitions between adjacent states need be considered (Birth-Death property).

According to the notation devised by Kendall (1951), exponentially distributed random arrival and service processes are labelled Markovian ('M'). Any departure from this requires additional information, *e.g.* bunching or staging of arrivals or service (Erlang-*k* or 'E*k*' where *k* is an index >1), or phased service times ('D' for 'Deterministic'), which together may be labelled General ('G'), multiple service channels, or finite maximum queue size (see standard works *e.g.* Newell [1971] 1982, Kleinrock 1975, Medhi 2003, Gross *et al* 2008), and ultimately abandonment of regular statistics (Chow 2013).

In modelling road or other traffic, the simplest M/M/1 queue idealises an isolated priority process with random service, while M/D/1 reflects the stochastic element of an isolated signal-like process by assuming service in regular finite periods, with the additional implicit

assumption that no significant delay is incurred in service during the 'green' period. Webster and Cobbe (1966) use the M/D/1 mean formula as a stochastic component of their signal delay model, with a phase term for the red/green signal cycle, plus a correction that effectively accounts for different green period capacities. These models are used in many junction design and traffic assignment programs. Heidemann (1994) gives a time-dependent solution for a signal queue that is believed to be exact, while Taylor and Heydecker (2013) extend M/D/1 to account for different green periods (M/D/1[*G*]), referring to Olszewski (1990) who reviews work by several previous authors.

Conservation of customers is a deterministic constraint common to all queues, but because they can arise from different types of arrival or service process this cannot determine their equilibrium properties. These can be found only by detailed analysis of the sequence of events or state transitions, leading to the Pollaczek-Khinchin (P-K) formula for equilibrium mean queue length which, subject to some provisos, can accommodate most common arrival and service processes solely by varying its statistical parameters. We believe that the only necessary condition on the arrival and service processes is that they be temporally and mutually uncorrelated, and on this basis, the theory and methods to be described are applicable to M/G/1 and even G/G/1 processes.

Heidemann (2001) shows that there is a parallel between random queuing and both steady-state and dynamic flow-density relationships in a channel of finite capacity such as a motorway segment. He identifies degree of saturation with the ratio of actual density to jam density, while accepting there are issues with using an M/M/1 model because it is inconsistent with the observed relatively narrow range of desired travel speeds, so some form of M/G/1 process may be more appropriate.

Time-dependent traffic modelling

Time-dependent traffic and queue modelling methods are approximations that attempt to 'answer the right question' by embodying structural as well as statistical properties of the real system at appropriate levels of detail. They can be considered intermediate between microscopic modelling of individual behaviours and macroscopic objective function minimising procedures that are mostly limited to static problems. They have been used in mesoscopic traffic assignment programs such as CONTRAM (Taylor 2003) and several junction modelling tools. Typically, mesoscopic methods divide time into 'slices' during which traffic conditions are treated as constant. In principle, time-slices can be as short as desired, at the cost of increased computation time. Space too may be divided into computationally convenient segments (*e.g.* Daganzo 1994). Queues are modelled using relationships that embody statistical parameters but require only aggregate data: initial state, arrival rate and capacity, and their associated stochastic processes. Individual behaviour and random variability are subsumed by these data and the relationships, while longer-term variations are accounted for by varying the parameters.

To some extent sensitivity to data can be inferred from response to random variation or uncertainty. Stochastic User Equilibrium (SUE) traffic assignment includes a measure of the uncertainty of link travel times, the seminal paper probably being that of Sheffi and Powell (1982). A Probit method described by Maher and Hughes (1997) propagates the variances of travel costs through a network along with their means, assuming nominally Normal distributions. Gordon *et al* (2001) describe various methods and results using an extended version of CONTRAM. Zhao and Kockelman (2001) apply variance propagation to four-stage transportation models. However, sensitivity to arbitrary changes in data, especially in networks, may require evaluation of several plausible scenarios, as is common in appraisal. This is outside the scope of the methods described here.

Time-dependent queue methods and their extension to variance and distributions

Several authors have developed approaches to time-dependent queuing that combine an equilibrium relationship appropriate to the particular process with the general deterministic equation representing conservation of customers in such a way as to produce a function of time that is seamless through saturation. Various approaches and results have been described by Newell (1968), Robertson (1969), Doherty (1977), Catling (1977), Kimber and Hollis (1979), Akçelik (1980) and others. Such 'sheared' functions are necessarily approximate, and a key issue for improvement is where their inaccuracies may lie.

While these methods embrace the effect of randomness on *mean* results, they do not account either for the uncertainty in queue size or how this is distributed, including the risk that the queue exceeds some critical size. Nor can they reveal whether the distribution is highly skewed or 'heavy tailed', which can result in a long queue on some days and not others, for no apparent reason (see *e.g.* Taylor 2012). The disruption that these unpredictable variations can cause increases rapidly around saturation, so as facilities such as road junctions, hospitals, airports, borders etc, become more heavily loaded it becomes increasingly important to be able to assess and manage such events.

Many standard reference works obtain values for the *equilibrium variance* of queues (*e.g.* Kleinrock 1975). Kimber *et al* (1986) describe an empirical method for calculating the time profile of queue variance produced by Gaussian-shaped demand peaks. Arup, Bates *et al* (2004) empirically, and Addison and Heydecker (2006) and Fosgerau (2008) theoretically, show how the relationship between mean and variance of delay gives rise to hysteresis loops where the maximum variance lags the peak of the mean queue. Taylor (2005, 2007, 2013) obtains an explicit formula for the time-dependent variance of a queue that depends on the history (integral) of the mean, and uses this to develop an extension to the time-dependent queue approximation to calculate time-dependent variance together with the mean for arbitrary time-varying traffic profiles, and in the process improve the accuracy of the whole method as verified by benchmark simulations.

However, mean and variance alone are found insufficient to characterise the probability distribution. Figure 1 shows two distributions of different shape, a geometric or nested-geometric (see later) and a truncated Normal, that have similar (left) or identical (right) means and standard deviations, but visible differences between their tails up to three or four times the mean (while this example is somewhat artificial it does not seem possible to make an illustration with much larger moments using these simple distribution types). As will be described, if the probability of zero queue (or the related utilisation) is also known it appears possible to estimate the queue size probability distribution usefully. Conversely, an inherent weakness of any method that calculates only the mean is that it does not take full account of initial conditions, and the true probability of zero queue and variance at any time may not be reflected in the values defaulted or implied by the approximation.

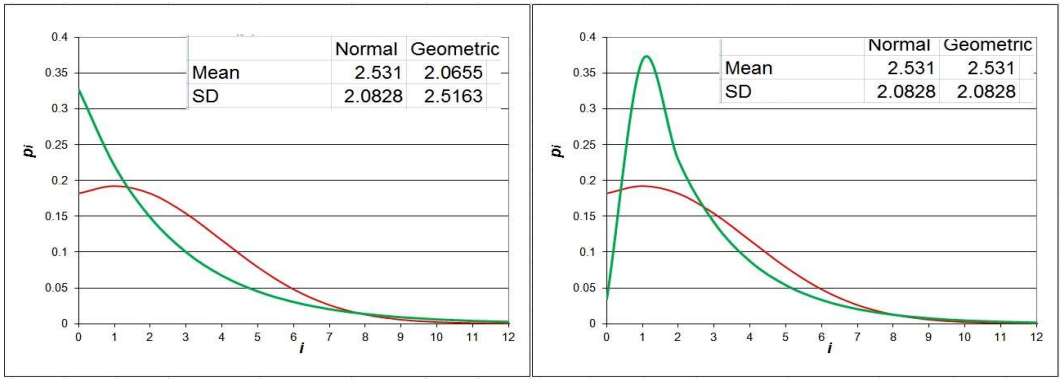

Figure 1. Distributions with similar (left) and identical (right) mean and variance

Microscopic modelling of queues and validation of predictions

Microscopic simulation of traffic has become increasingly popular because of the rising power of computers and sophistication of graphics to create realistic real-time presentations and even virtual-reality experiences. However, each simulation can generate only one sample of a process. Extended or repeated simulations are needed to build up an accurate queue size probability distribution. Furthermore, as Taylor (2013, Appendix F) shows, several orders of magnitude more events need to be simulated than would be expected on the basis of 'standard error of the mean' because of correlation between successive queue states.

The real world is equivalent to a simulation where there is considerable variability in and uncertainty about inputs, the parameters cannot be controlled and experiments cannot be repeated. Validation would therefore amount to showing broad statistical agreement between observation and theory under a range of conditions. An intermediate type of simulation is a controlled experiment, such as that carried out over a simulated peak growth period by Kimber and Daly (1986), who found a range of queue sizes consistent with theory. Validation is a major exercise outside the scope of the work on which this paper is based, whose objective is to exploit a theoretical result expressed in novel form. Therefore testing has been limited to verification against methods of known accuracy.

Verification and benchmarking by simulation

Verification of a method tests whether it produces the results expected from theory. The aim of the work described is to extend the use of efficient closed-form relationships in a field which although well-established lacked the ability to estimate probabilities. In order to verify the accuracy or otherwise of approximations, a benchmarking method of known accuracy is required. Morse (1958) gives a series formula for the probabilities of an M/M/1 queue developing over time from a specified initial size (also credited to A B Clarke). An alternative formulation is proposed by Sharma (1990), and Griffiths *et al* (2005) develop an extension for M/E*k*/1. Recurrence relations express how queue state probabilities evolve through time. Markov simulation, by animating recurrence relations in small steps (*e.g.* 0.1 second), has been used to develop distributions of M/M/1, M/D/1 and other queue types. While microscopic or 'Monte Carlo' simulation of random arrivals and service gives only a sample distribution, it can give confidence in the other methods, as shown by Figure 2.

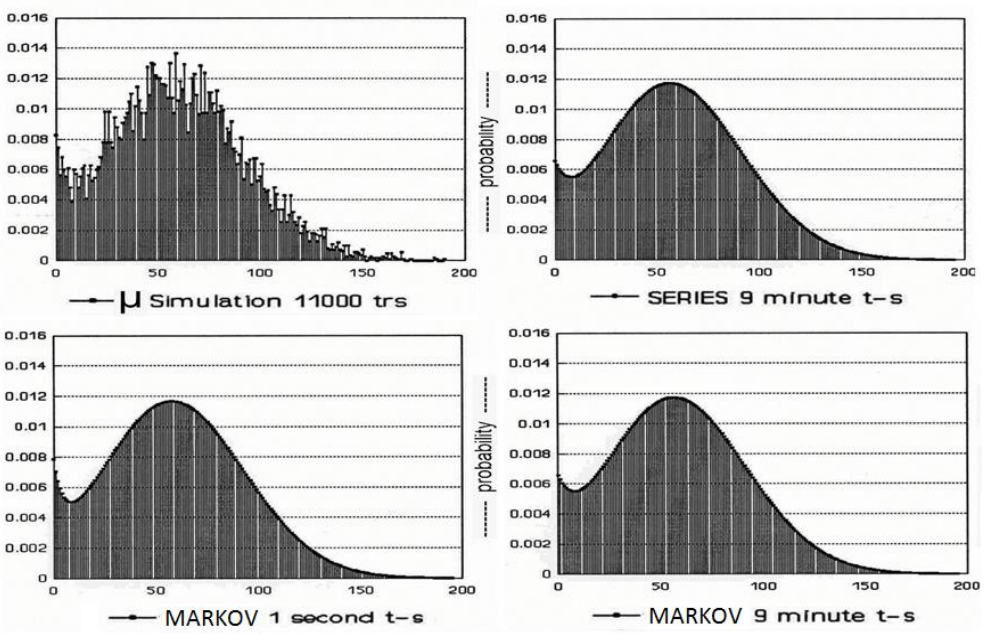

Figure 2. Post-peak distribution obtained by different benchmarking methods

Recurrence relations for M/M/1 are derived by considering arrivals and departures on an infinitesimal time scale. Extensions exist for Erlang-*k* arrivals or service, and for multiple lanes with shared arrivals and service (Taylor 2011, 2013). Those for M/D/1 can be derived on the basis of a finite service (green) period, during which numbers of arrivals are assumed to be Poisson distributed. The M/D/1[*G*] queue adds realism by taking explicit account of the absolute capacity of the service period (Taylor and Heydecker 2013). The detail of these methods is not needed here, but an example is given later to support an argument that all equilibrium distributions tend to geometric form far from an absorbing 'barrier' such as zero queue size. Various software programs to implement these methods and provide benchmark results were developed by one of us (Taylor) with the assistance of Neil H Spencer (then a sandwich student at TRL).

Queue development described in terms of mean and variance

Under constant conditions queue development is described by the deterministic formula:

$$L(t) = L_0 + (\rho - x(t))\mu t, \qquad \text{where} \qquad x(t) = \frac{1}{t}\int_0^t u(y)dy \quad \text{and} \qquad (1)$$

$L(t)$ is mean queue size at time t, L_0 is initial queue size at time t=0, ρ is the demand intensity, the ratio λ/μ of arrival rate λ to capacity μ (both in customers per unit time), u is the short-term degree of saturation or utilisation of service in the neighbourhood of t, while x represents the average degree of saturation or utilisation of service in the period $[0,t]$. In a form analogous to (1) though more complicated (Taylor 2005, 2007, 2013), time-dependent variance $V(t)$ satisfies:

$$V(t) = V_0 + L_0(L_0+1) + 2(1-\rho)(L_e(\rho) - D(t))\mu t - L(L+1), \quad \text{where} \qquad (2)$$

$$D(t) = \frac{1}{t}\int_0^t L(y)dy \qquad \text{is also average delay per unit time}$$

A significant property of both (1) and (2), provided that the queue approaches equilibrium, is that they involve the product of a quantity that tends to zero as $t\to\infty$, with time t that can increase without limit, allowing the possibility of a finite result that is formally indeterminate. This makes it possible for a queue to satisfy the conservative deterministic conditions while tending to an equilibrium state that depends only on its stochastic properties. A general expression for the steady-state equilibrium mean queue length formula is the Pollaczek-Khinchin (P-K) formula, equation (3), where I_a is the dispersion of arrivals, c_b is the coefficient of variation of service times, and I is an index variable (*i.e.* with value 0 or 1) reflecting inclusion of time spent in service as opposed to only while waiting for service:

$$L_e(\rho) = I\rho + \tfrac{1}{2}(I_a - 1)\frac{\rho}{1-\rho} + \tfrac{1}{2}(1 + c_b^2)\frac{\rho^2}{1-\rho} \qquad (3)$$

For M/M/1, I=1, c_b=1, while for M/D/1, I=0, c_b=0, with I_a=1 for both. Equation (3) is an extension of that found in most standard works, through the inclusion of the parameter I due to Kimber and Hollis (1979) and I. Summersgill, and inclusion of the dispersion I_a due to one of us (Heydecker), and can be written in an alternative form using Erlang parameters of arrivals and service, r and m respectively:

$$L_e(\rho) = \frac{\rho}{1-\rho}\left[I(1-\rho) + \frac{1-r}{2r} + \left(\frac{1+m}{2m}\right)\rho \right] \qquad (4)$$

An analogous but more complicated formula can be developed for equilibrium variance (Taylor 2013). In order to relate time-dependent deterministic and steady-state equilibrium descriptions, we adopt the 'quasi-static' approximation in which x is substituted for ρ in equation (3) or (4). Because x cannot exceed 1, the quasi-static formula is defined even when $\rho \geq 1$, allowing equations (1) and (3) to be equated and solved for x or L, giving the 'sheared' time-dependent queue approximation (for further discussion and example of application see earlier references and Taylor 2003). The method gives logical and seamless results as traffic intensity increases through saturation, but is least accurate in mid-growth and for decaying queues. Likewise, equation (2) is defined even when D does not tend to its equilibrium value L_e. Accuracy of the whole method can be improved by requiring (2) to give realistic values and the correct equilibrium variance if below saturation, constraining (1). In this way time-dependent mean and variance profiles can be estimated together for any queuing process that can be described by suitable equilibrium formulae.

The role of utilisation or probability of zero queue in queue development

Utilisation is the proportion of time that service is active, which normally means when a queue is present. At any time t, the short-term *average* probability of zero queue is related to short-term utilisation by the first of equations (5). At equilibrium (normally as $t \to \infty$) if the queue is to remain finite, equation (1) requires utilisation to tend to ρ.

$$\bar{p}_0(t) = 1 - u(t) \quad \rightarrow \quad \bar{p}_{0e} = 1 - \rho \qquad \text{at equilibrium} \qquad (5)$$

The time derivative of (1) emphasises how utilisation and $\bar{p}_0$ are intimately linked to queue dynamics, and play a critical role throughout queue development:

$$\frac{dL(t)}{dt} = (\rho - u(t))\mu = (\bar{p}_0(t) - \bar{p}_{0e})\mu \qquad (6)$$

A distinction can arise between p_0, the probability of zero queue in the distribution, and $\bar{p}_0$ its average over the service period, because the distribution typically relates to measurements *sampled* at a particular time. For queues with random service, such as M/M/1 and its Erlang-k variants, the sampling point is irrelevant, and $\bar{p}_0$ is equal to p_0. However, for any queue formulated on a finite service period, such as M/D/1, some traffic can arrive and be discharged entirely within the service period, so the probability of zero queue at the *end* of the period is necessarily greater than its average over the period, so in general $p_0 \geq \bar{p}_0$. At a signal-controlled junction the queue at the end of the green phase is of most interest as it adds directly to the following red phase queue and hence contributes to the development of queuing in oversaturated conditions. In fact the M/D/1 equilibrium probability of zero queue, sampled at the end of green, is:

$$p_{0e[M/D/1]} = e^{\rho}(1-\rho) \qquad (7)$$

Conservation of customers requires that the average $\bar{p}_{0e}$ satisfy (5) rather than (7). In time-dependent analysis it is necessary to assume a relationship between $p_0(t)$ and $\bar{p}_0(t)$, and a convenient assumption is again the 'quasi-static' one that ρ can be replaced by *x*.

'Three pillars' of the probability distribution

Given its role in queue development, and the evidence of Figure 1, it is natural to suppose that p_0 is needed as well as the mean and variance of a queue to characterise its probability distribution, so it may be styled a 'moment', but are these three moments sufficient? A reason for believing so is that, thanks to asymmetry, changing just one of the three moments p_0, *L* and *V* while holding the others fixed will directly affect skewness and higher moments too. A practical reason for stopping at three moments is that p_0 is relatively easy to obtain from the queue function via (6), and inversion of (7) if necessary, while time-dependent variance can be calculated using (2), and these values can be passed between successive time slices in which parameters may differ. Thus all three moments produced by a peaked or arbitrary demand profile can be estimated, for any process describable using the P-K formula. Figure 3 shows the close match obtainable between calculated and Markov-simulated moments, plus delay-per-unit time *D* as defined in equation (2), for the same oversaturated peak case underlying Figure 2. The only visible difference is that variance is slightly underestimated.

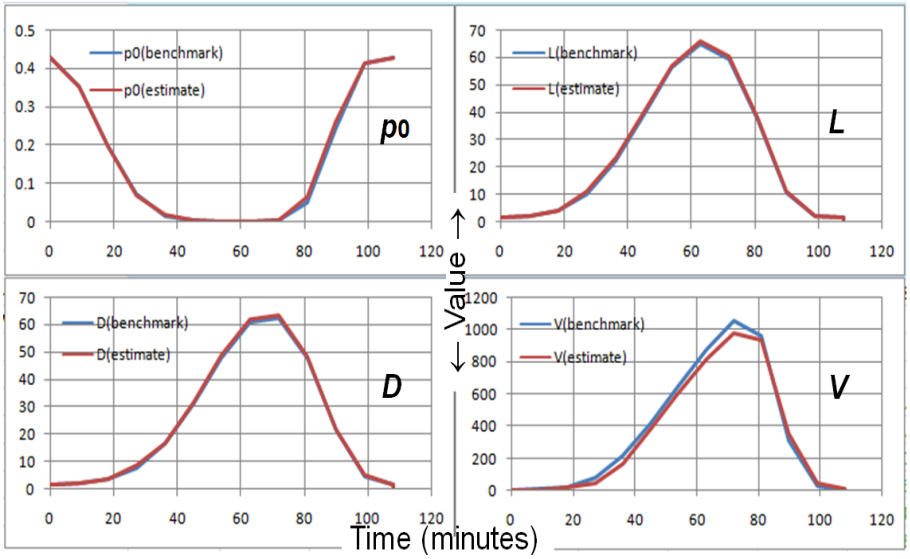

Figure 3. Estimation of oversaturated peak, compared to Markov simulated benchmark

Figure 4 gives evidence of the match between estimated and Markov-simulated results, using 408 time points from 34 peak cases tested, using both M/M/1 and M/D/1 models.

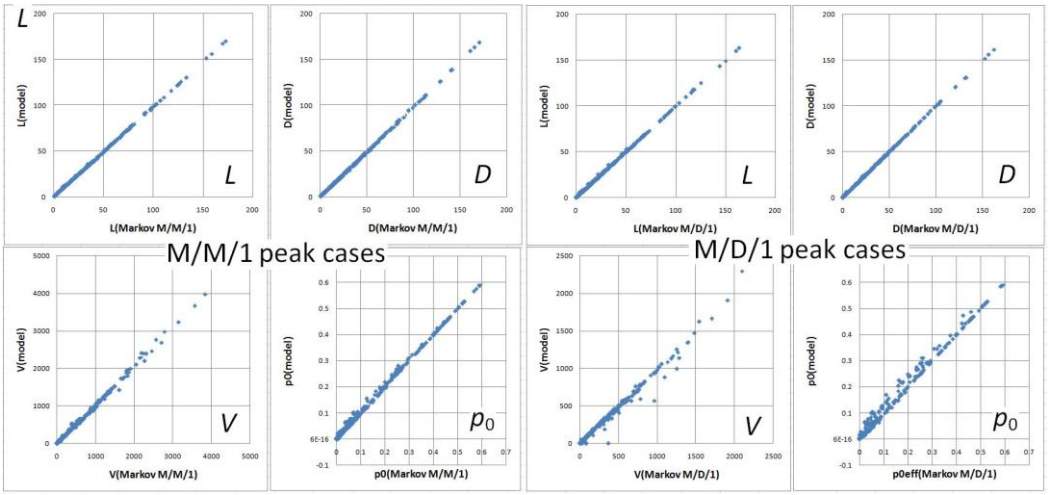

Figure 4. Plot of estimated against Markov-simulated moments of 34 peak cases

An important difference between the methods used to produce Figures 3 and 4 is in their computation time and complexity. Whereas analytical estimates can be calculated almost instantaneously by a short compiled program or by cascading sets of identical formulae in a spreadsheet representing for example 12 time slices, generating corresponding Markov-simulated profiles requires iterating recurrence relations through around 65,000 time steps.

Estimating equilibrium probability distributions

The geometric equilibrium distribution of the M/M/1 process (taller graph in Figure 1 left), is the most commonly assumed equilibrium queue distribution, and is characterised by a constant ratio ρ between the probabilities of adjacent queue states, equal to the ratio of demand to capacity, provided that ρ<1. Its components and moments are given by equations (8), standard results which serve here to contrast some important exceptions:

$$p_i = (1-\rho)\rho^i \qquad i \in [0,\infty), \qquad L_e = \frac{\rho}{1-\rho}, \qquad V_e = \frac{\rho}{(1-\rho)^2} \tag{8}$$

Other types of equilibrium distribution, such as M/D/1 or resulting from Erlang-*k* arrival or service processes, can differ from (8) in three principal ways:

- p_0 has a different value, as in the case of equation (7);
- the mode, the value if *i* for which p_i is maximum, is greater than zero;
- the effective ratio between higher state probabilities differs from ρ.

However, for higher queue states the distribution is expected to tend to geometric. This can be explained by symmetry under change of viewpoint. Far from the 'absorbing barrier' of zero queue no state is special, so all observations should be qualitatively similar. If arrivals and service can be treated as continuous, an infinitesimal time interval can be assumed and only states adjacent to the current state need be considered. Therefore, for states in equilibrium the only forcing is by arrivals and service at *effective* mean rates λ and μ respectively which are assumed to be uncorrelated. This is depicted in Figure 5:

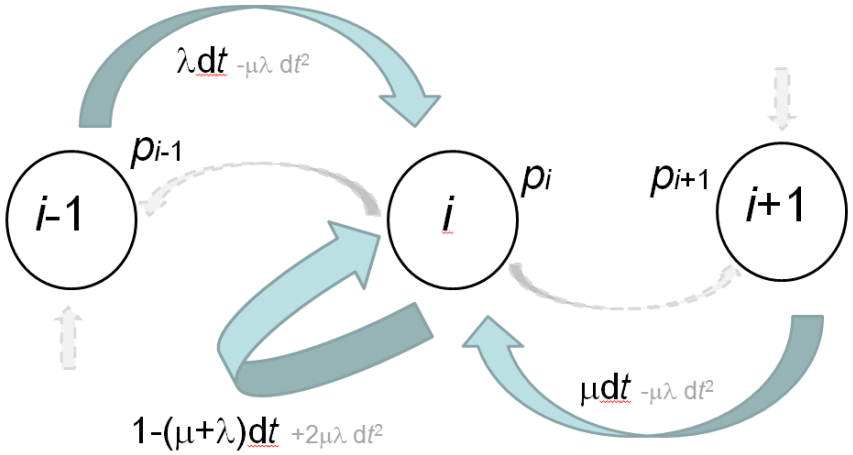

Figure 5. Infinitesimal queue state transition diagram showing 'deep' contributions to p_i

Greyed-out second order elements and other transitions are included for completeness. The former represent the probability that a customer arrives and is immediately served, so vanish in the limit as $dt \to 0$. Transitions may occur to the adjacent states, but in an infinitesimal time period dt there is no recursion so these do not affect the result. This leads to the same recurrence relation as for M/M/1, with $\tilde{\rho}$ used here in place of ρ to represent more generally the *effective* demand intensity:

$$\frac{1}{\mu}\frac{dp_i}{dt} = p_{i+1} - (1+\tilde{\rho})p_i + \tilde{\rho}p_{i-1} \qquad (= 0 \text{ at equilibrium}) \qquad (9)$$

In the equilibrium steady state this expression is zero for all i. The only equilibrium solution compatible with state-spatial symmetry is the geometric form:

$$p_i = Kf(\tilde{\rho})\tilde{\rho}^i \qquad (\text{for } i \text{ sufficiently} > 0,\ K = \text{normalising factor}) \qquad (10)$$

In the case of M/M/1, equation (9) applies to all states except zero (and the maximum queue size if there is one), resulting in the simple geometric distribution (8). Certain Erlang-k processes produce a singly-nested geometric distribution as in equations (11), of which examples may be found in standard works already cited. This can have mode > 0 as the taller graph in Figure 1 (right) earlier.

$$p_0 = 1-\rho^*, \qquad p_i = \rho^*(1-\hat{\rho})\hat{\rho}^{i-1} \ (i \geq 1) \qquad (11)$$

The equilibrium mean and variance associated with this distribution are:

$$L_e = \frac{\rho^*}{1-\hat{\rho}}, \qquad V_e = \frac{\rho^*(1+\hat{\rho}-\rho^*)}{(1-\hat{\rho})^2} \qquad (12)$$

M/M/1 is the special case where $\rho^* = \hat{\rho} = \rho$. However, if p_{0e}, L_e and V_e are *all* known, then (11) is overspecified because it has only two parameters. Therefore a doubly-nested geometric equilibrium distribution may be defined as follows:

$$p_{0e} = 1-\rho^*, \qquad p_1 = \rho^*(1-\hat{\rho}), \qquad p_i = \rho^*\hat{\rho}(1-\vec{\rho})\vec{\rho}^{i-2} \ (i \geq 2) \qquad (13)$$

whose mean and variance are:

$$L_e = \frac{\rho^*(1+\hat{\rho}-\vec{\rho})}{1-\vec{\rho}}, \qquad V_e = \frac{\rho^*(1+3\hat{\rho}-\vec{\rho}(2+\hat{\rho}-\vec{\rho}))}{(1-\vec{\rho})^2} - L_e^2 \tag{14}$$

Conversely, the three parameters are given uniquely in terms of the moments by:

$$\rho^* = 1 - p_{0e}, \quad \hat{\rho} = \frac{(V_e + L_e(L_e-1))(1-\vec{\rho})^2}{2\rho^*}, \quad \vec{\rho} = \frac{V_e + L_e(L_e-3) + 2\rho^*}{V_e + L_e(L_e-1)} \tag{15}$$

The singly-nested distribution is also the maximum entropy form where only p_0 and the mean are known (Kouvatsos 1988). Similarly, the doubly-nested distribution is the maximum entropy form where all three moments are known. This is expressible in the alternative form where the quantities in brackets are derived as Lagrange multipliers:

$$p_i = p_0 . \left(\frac{p_1 p_2}{p_0 p_3}\right)^{Q_1} . \left(\frac{p_2^2}{p_1 p_3}\right)^{Q_2} . \left(\frac{p_3}{p_2}\right)^i \tag{16}$$

In equation (16), Q_1=0 when i=0 and Q_1=1 when i>0, and Q_2=0 when $i \leq 1$ and Q_2=1 when i>1, but there is no term in i^2 as that would cause the expression to 'explode' at high values of i. The effect of variance, like the other moments, enters through the values of the first four (normalised) probabilities. Not all probability distributions, particularly dynamic ones, can be represented by (13) or (16) because the mode cannot exceed 2, and (13) fails if $\hat{\rho} \geq 1$. Evidently, the structure of (16) can be extended indefinitely provided that for some n the probabilities $\{p_0..p_n\}$ are known and for $i \geq n$ the ratio between p_i and p_{i-1} is constant. However, this is of limited use for estimating dynamic distributions where most of the mass can lie around a high modal value of i and probabilities near to i=0 can be small.

Examples of equilibrium distributions approximated by a doubly-nested geometric are shown in Figure 6, as given in Taylor and Heydecker (2013). An interesting feature of the M/D/1[G] recurrence relations, where G is actual green period capacity, is that they lead to 'raw' distributions whose origin is shifted to $-G$ and whose mode lies at mode at $G(\rho-1)$, the real value of p_0 being obtained by summing all notional components with non-positive indices. The use of Gamma distributions, as in Figure 7, is proposed as a basis for approximating these distributions. The discrete equivalent of Gamma is the Erlang family of distributions, while an alternative discrete model may be the Negative Binomial, although this would need scaling to give the correct mode.

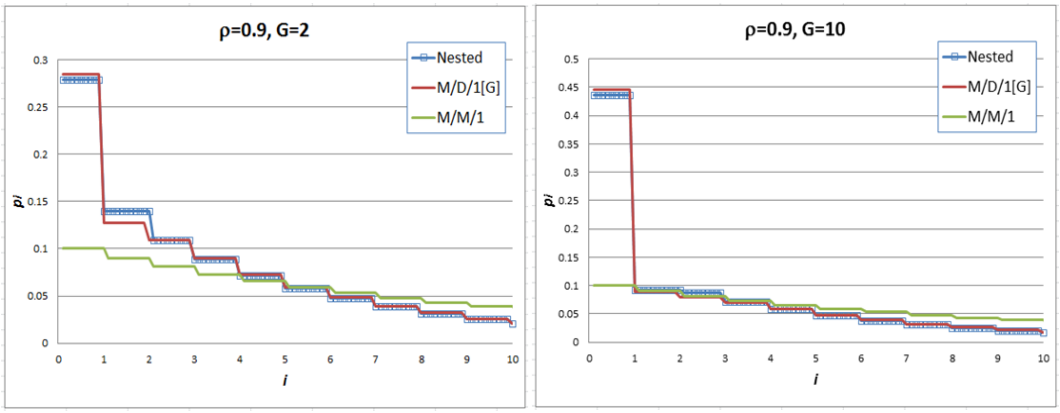

Figure 6. Doubly-nested geometric distributions fitted to Markov simulated M/D/1[*G*] distributions, with M/M/1 for comparison (lower graphs)

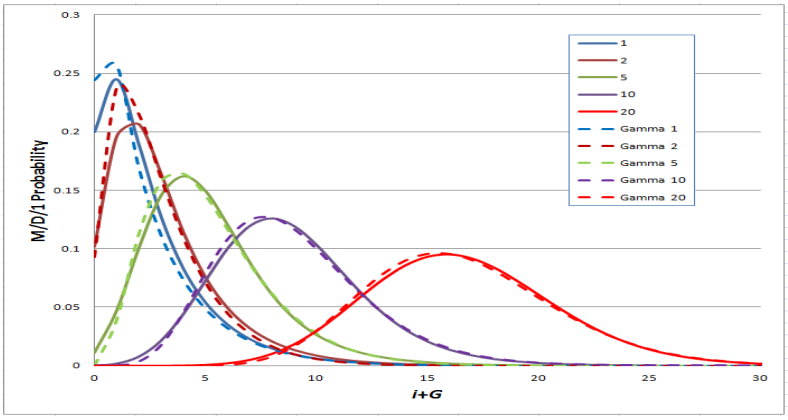

Figure 7. Gamma fits to M/D/1[*G*] extended distributions for ρ=0.8 and *G* up to 20

Continuous and diffusion approximations to dynamic distributions

Henceforth, continuous rather than discrete distribution functions will be considered as they are more convenient to work with. To approximate a dynamic distribution an asymmetrical modal distribution function is required. Such distributions include truncated Normal (as used in Figure 1, lower graphs), Gamma, Poisson and LogNormal. A dynamic approximation should tend to the equilibrium form when $t \to \infty$, so in a sense must already contain this form. The exponential function is the continuous analogue of the geometric distribution and is both a special case of Gamma and the form it approaches at higher state values. However, the LogNormal, Poisson and compound Poisson, while useful in many areas, are not asymptotically geometric, making them unsuitable for the present purpose.

The Normal distribution is the natural form of dynamic distribution arising from random drift and dispersion remote from an absorbing or reflecting barrier, analogous to Brownian motion in one dimension. When an oversaturated random queuing process is allowed to run for long enough, its probability distribution tends to Normal. This is also the limiting form of the (compound) Poisson and LogNormal when their standard deviation is small relative to the mean, but the independence of their parameters is then lost.

Figure 2 earlier, showing a simulated distribution shortly after an oversaturated peak, appears to combine exponential and Normal features. The 'duck-tail' at the left can be explained by rare queue states below the equilibrium mean that are still growing, while typically the queue is decaying at this point after the peak. This reflects the fact that a queue is a linear system whose state can be described as a superposition of independent primitive elements, as for example in the series formula given by Morse (1958).

The Kolmogorov Forward Equation or Fokker-Planck Equation (FPE), as quoted by Newell (1968), is the continuous analogue of a recurrence relation such as (9):

$$\frac{\partial p(x,t)}{\partial t} = \frac{\left(I_a \rho + I_\mu\right)}{2} \mu \frac{\partial^2 p(x,t)}{\partial x^2} + \left(\mu - \lambda\right) \frac{\partial p(x,t)}{\partial x} \qquad (17)$$

The first term on the RHS represents 'diffusion' (spreading) caused by randomness. The second term represents deterministic 'drift' at a rate determined by the difference between capacity and arrivals (-growth and +decay). Newell (1968) expects that for a pure Poisson (M/M/1) process, both the indices of dispersion, I_a for arrivals and I_μ for service, are "comparable to 1 and essentially independent of [ρ] or μ". The multiplying factor is indeed trivially correct in the case most often cited in the literature, exponentially distributed random arrival or service headways (I_a, I_μ=1) and 'heavy traffic' (ρ≈1), where the diffusion coefficient reduces to 1, but is problematic because of the way the statistics are combined.

A similar issue has arisen in relation to the P-K formula (3), where in the absence of I_a some authors have proposed replacing $(1+c_b^2)$ by $(c_a^2+c_b^2)$, where c_a is the coefficient of variation of arrivals. However, unlike the first form this does not arise naturally in any derivation. The implications of alternative statistical expressions for waiting time, equivalent to redefining the statistical parameters in equations (3) or (17), are explored by Whitt (1982), who points out that all the variations are "asymptotically correct in heavy traffic" but none is generally valid. 'Heavy traffic' cannot be assumed in the present context, and it remains to determine finally how process statistics should be incorporated into a continuous approximation.

These issues are partially avoided if an M/M/1 process is assumed, as in the analysis of bunching by Kühne and Lüdke (2013), who describe a simplified diffusion equation (18) as the continuous equivalent of the M/M/1 recurrence relations (after translating variables):

$$\frac{\partial p(x,t)}{\partial t} = \mu \frac{\partial^2 p(x,t)}{\partial x^2} + \left(\mu - \lambda\right) \frac{\partial p(x,t)}{\partial x} \qquad (18)$$

Some standard works give examples of time-dependent solutions to the diffusion equation, for example Newell (1971/1982), Kleinrock (1976), and Gross *et al* (2008), but perhaps the most useful is that of Kobayashi (1974a,b), which is stated in differential form but can be interpreted as in equation (19), where x_0 is the exact initial state at t=0:

$$p(x,t|x_0) = \left[\frac{e^{-\frac{(x-x_0+(1-\rho)\mu t)^2}{2(1+\rho)\mu t}} + e^{-\frac{2x|1-\rho|}{(1+\rho)}} e^{-\frac{(x+x_0-(1-\rho)\mu t)^2}{2(1+\rho)\mu t}}}{\sqrt{2\pi(1+\rho)\mu t}} \right] + \frac{|1-\rho|}{1+\rho} e^{-\frac{2x|1-\rho|}{(1+\rho)}} \ddot{E}(x,t|x_0) \qquad (19)$$

In this continuous interpretation $p(x,0) = \delta(x-x_0)$ (the Dirac delta at x_0), and:

$$\ddot{E}(x,t|x_0) = erfc\left(sign(1-\rho) \left(\frac{x + x_0 - (1-\rho)\mu t}{\sqrt{2(1+\rho)\mu t}} \right) \right) \qquad (20)$$

where *erfc* is the standard complementary error function, equal to twice the integral of the Normal distribution from its argument to infinity. Figure 8 shows a case of early development of a heavy undersaturated queue calculated using the diffusion method.

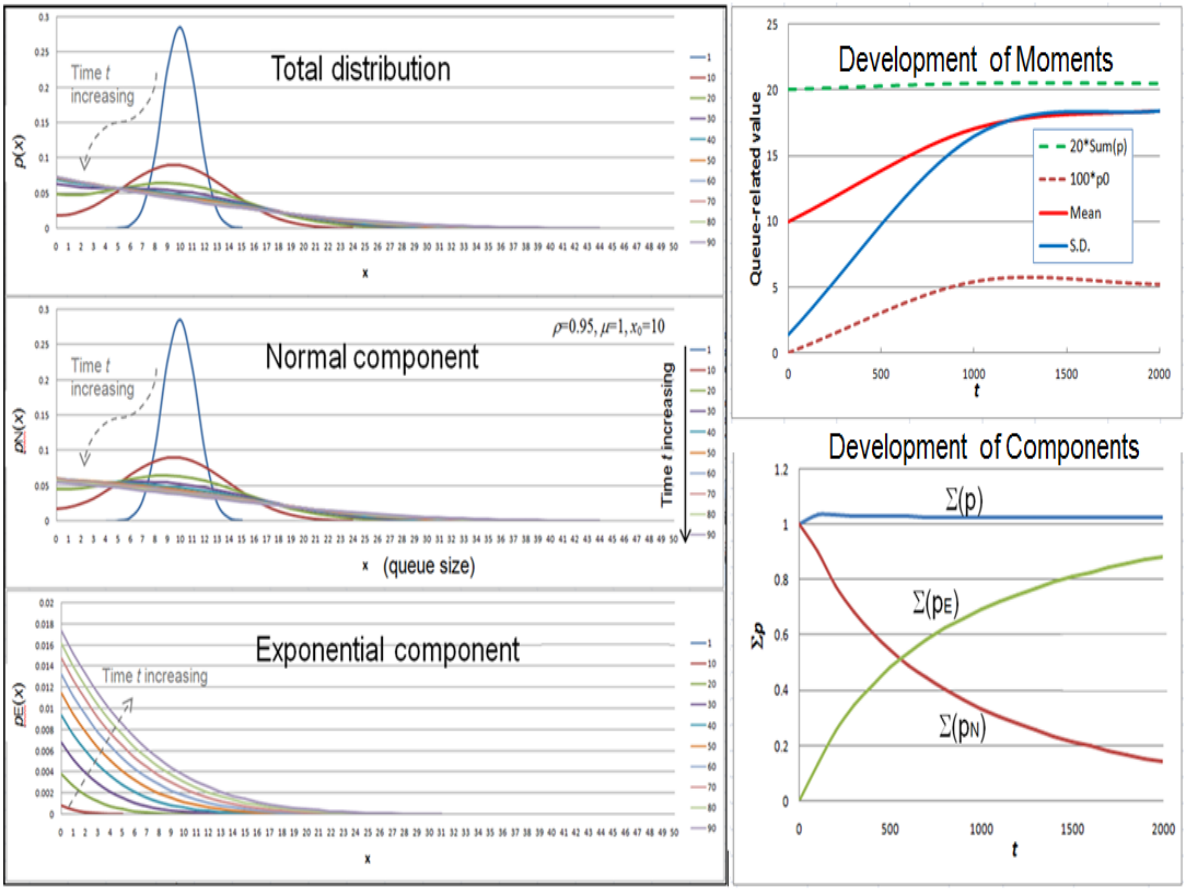

Figure 8. Evolution of Kobayashi diffusion solution, parts and moments (ρ=0.95, x_0≡10)

Figure 8 shows that to fair approximation the distribution is correctly normalised ($\Sigma p \approx 1$), and exhibits the expected convergence behaviour of a random queue with ρ=0.95, where the mean should tend to 19 and the standard deviation to about 19.5. The early development of the component distributions and total moments over 0-90 mean service time intervals is shown, where the distributions relax quite rapidly (left), while the aggregate properties relax more slowly (lower right).

Provided that the system is undersaturated, the Normal terms in (19) vanish as $t \to \infty$, and the exponential term alone then determines the equilibrium distribution. Because the diffusion solution is defined for an exact initial state, in generally it needs to be convolved with an initial probability distribution to give the distribution at a later time.

Estimating dynamic probability distributions using continuous functions

Because both statistics and time variation are already embedded in the time-dependent analytical queue approximation, an *instantaneous* probability distribution need only match the moments at the particular time to which they belong. Although it is tempting to try to use a single skewed distribution such as Poisson or LogNormal, these perform poorly as illustrated by Figure 9. The example of the diffusion solution suggests that a combination of exponential and Normal functions is more suitable.

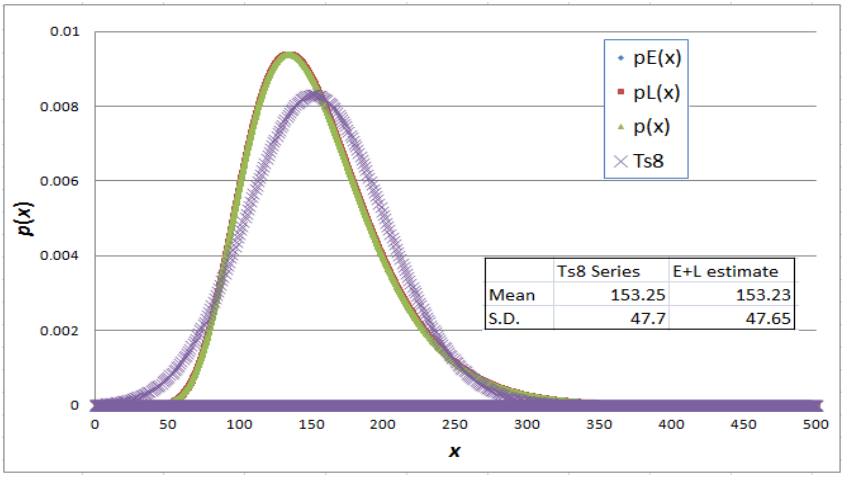

Figure 9. Inability of LogNormal (skewed) function to match a simulated distribution

The close relationship between the exponential function and the geometric distribution is shown by equations (21,22), although $p_E(0,\nu) \neq p_0(u)$ and $E[p_E] = 1/\nu \neq u/(1-u)$:

$$p_E(x,\nu) = \nu e^{-\nu x} \qquad \text{where} \qquad \nu = -\ln(u), \qquad u = 1-\bar{p}_0(t) \tag{21}$$

$$\int_0^\infty p_E(x,\nu)dx = 1\,, \qquad p_i = \int_i^{i+1} p_E(x,\nu)dx = (1-u)u^i \quad (i \geq 0) \tag{22}$$

We assume that the exponential function can be used in 'quasi-static' mode. The Normal function represents the natural drift and diffusion behaviour of a queue process effectively unconstrained by a lower (or upper) limit. It is found that a combination of exponential and Normal functions can approximate the form of a dynamic queue size distribution, at least where the equilibrium distribution is M/M/1. For other types of equilibrium distribution whose 'tail' is geometric, but which have no obvious continuous equivalent, an *ad hoc* correction to p_0 and possibly other low-state terms can be made, although this may be improved by further research. Another possibility would be use Gamma distributions to approximate more general equilibrium distributions.

Exponential and Normal functions both satisfy simple forms of the Fokker-Planck equation, ensuring that certain combinations will also be solutions, including an exponential combination which is found to be most satisfactory. This adds a relaxation parameter θ to the calculated values of $\bar{p}_0$ or p_0, L and V that have to be fitted, as in equations (23,24) where the Normal component p_N has mean m and standard deviation s:

$$p(x) = e^{-\theta x} p_E(x,\nu) + n\left(1-e^{-\theta x}\right) p_N(x,m,s) \qquad \text{or} \tag{23}$$

$$p(x) = \nu e^{-(\nu+\theta)x} + n\left(p_N(x,m,s) - e^{-\theta\left(m-\frac{\theta}{2}s^2\right)} p_N\left(x, m-\theta s^2, s\right)\right) \tag{24}$$

Equation (23) suppresses the Normal component at x=0, so $p(0)$ is determined by $\bar{p}_0$ or u alone. Normalisation factor n and moments are readily obtained by integration. In the alternative form, equation (24), the second Normal component is a factored and shifted mirror-image of the first, which can be viewed as representing the effect of the 'absorbing barrier' at x=0. The expectation of $p(x)$, equation (25), representing the mean queue size, confirms the correct values 1/ν at θ=0 (equilibrium), and m when θ=∞ (fully dynamic):

$$E[p] = \frac{\nu}{(\nu+\theta)^2} + \frac{m\theta}{\nu+\theta} + e^{-\theta\left(m-\frac{\theta}{2}s^2\right)}\left(\frac{n\theta s^2}{2}\right) erfc\left(-\frac{(m-\theta s^2)}{s\sqrt{2}}\right) \qquad (25)$$

To recover the corresponding discrete distribution, a continuity correction is employed: the target queue size used to fit the function is incremented by half a unit and the discrete probabilities are picked off at points $\{x=i+0.5\}$. Because ν is specified by (21) and n follows from normalisation of $p(x)$, the problem is underspecified, with three unknown parameters θ, m, s to be determined from two remaining constraints L and V, so numerical solution is necessary. For demonstration purposes, the Solver tool in Excel, which contains the standard GRG2 (Generalised Reduced Gradient) algorithm, has been used to estimate parameters initialised according to equations (26), where initial θ is estimated by assuming n=1 and neglecting the mirror-shifted Normal component, by minimising a simple RMS error measure given by equation (27). Solution steps are shown in Table 1, all starting from the initial parameter values (*i.e.* no there is recursion), where in practice Step 3 is found to be unnecessary.

$$m = \tilde{L} = L + 0.5, \qquad s = S = \sqrt{V}, \qquad \theta \approx \nu\left(2 \Big/ erfc\left(-\frac{\tilde{L}}{\sqrt{2V}}\right) - 1\right) \qquad (26)$$

$$Error(L,V) = \sqrt{\left(\tilde{L}_{(tar)} - L_{(dis)}\right)^2 + \left(\sqrt{V}_{(tar)} - \sqrt{V}_{(dis)}\right)^2} \qquad (27)$$

Table 1. Excel Solver estimation schemes for distribution parameters

Setup	Define criterion, *i.e.* minimise error as defined by equation (27)	
1	Set initial θ, m, s	Solve for θ, m, s
2	Set m=0 and initial θ, s	Solve for θ, s
(3)	If ρ<1, Set θ=0 and initial m, s	Solve for m, s
Output	Select solution that minimises the error criterion unless bimodal distribution is unexpected during growth phase when select solution 2	

Accuracy of estimated probability distributions

Taylor (2013) gives detailed results of tests with 34 peak cases using M/M/1 and M/D/1 models (see Figures 3, 4 earlier) as well as three 'random' profiles. Figure 10 compares fitted distributions based on estimated moments (thick traces) with Markov-simulated probability distributions at various points in one of the oversaturated peaks ('X's), also showing the exponential and Normal components (thin traces). This comparison reflects not only fitting errors but also errors in the time-dependent approximation to the peak profile queue. Smaller errors result when distributions are fitted to moments of the Markov benchmark distributions, with each time-slice treated independently, but while this can verify the fitting procedure, Figure 10 represents a more useful test of the method as a whole.

Significant features of the distributions are near-Normal shape at the peak (Ts 6), the 'duck-tail' that develops shortly after the peak (seen earlier in Figure 2), and rapid collapse later on to a complicated bi-modal form. These results show that there is no basis for assuming an equilibrium distribution, corresponding say to the instantaneous demand intensity or utilisation, and doing so post-peak could lead to underestimation of the weight of the 'tail' and hence of the likelihood of a long queue. This can also explain the occurrence sometimes of long queues for no apparent reason, and may have consequences for traffic management and for optimal facility planning.

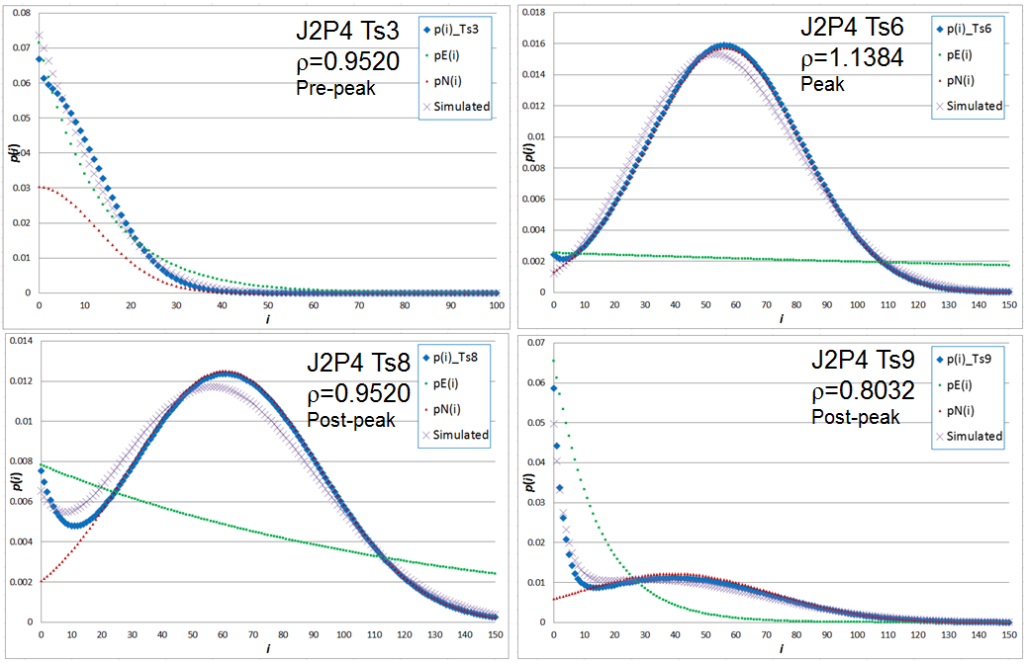

Figure 10. Distribution fits based on modelled moments of same peak case

Accuracy of estimated risk of exceeding a critical queue size

The risk of exceeding a critical queue size, set for example by the storage space available on a road section or the number of beds in an A&E (ER) department, can be estimated as the cumulative probability above that point. The accuracy of this estimate will depend both on the accuracy of the estimated probability distribution and on the chosen critical value. Where mean queue size and distribution shape vary over time, it is not obvious whether accuracy of estimated risk is best expressed in absolute or relative terms.

However, for a range of critical queue sizes, specimen absolute errors, that is differences between estimated and Markov-simulated cumulative probabilities, can be obtained for the distributions in Figure 10, as shown in Figure 11.

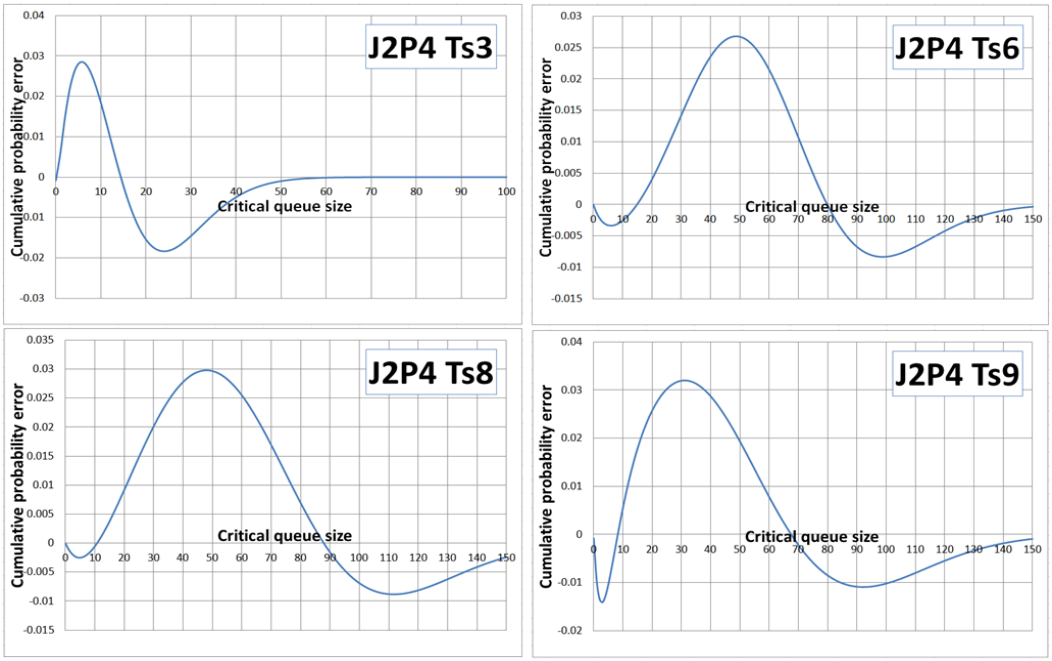

Figure 11. Absolute errors in estimated risk of queue size exceeding a critical value

The maximum absolute error in the estimated risk of exceeding critical queue size barely exceeds 3% in these cases, and is likely to be less than this pre-peak if the critical queue size is comparable to the expected peak queue. This can be taken as an indication of the practical usefulness of the estimated distributions.

Conclusion

As pointed out by Sharma (1990) there is no justification for simply assuming that a dynamic process can be represented by an equilibrium system. This paper has discussed some idealised random queue processes, that are considered to reflect types of queuing commonly occurring in practice, and sit in a framework extendable to more general processes. Examples of simulated queue size probability distributions through oversaturated peaks confirm that equilibrium distributions are inapplicable to such cases. This applies equally to undersaturated 'heavy traffic' cases where the system has not had time to approach equilibrium. At the height of a peak, a typically Normal-like distribution means that the probability of unexpected queues much longer than the mean is less than under equilibrium conditions, but soon after a peak the distribution can become very extended, making it difficult to predict the persistence of queues and consequent disruption.

Analytical time-dependent approximate queue methods are an efficient way of calculating queue development, and the quasi-static assumption they rely on can be justified broadly by structural arguments and tempered by corrections, but in the past they have not included estimation of queue variance or probability distributions. Consequently, there was a tendency to fall back on an equilibrium assumption represented by the geometric distribution, which while easy to derive from the mean applies only to the simplest M/M/1 (priority-type) equilibrium queues. This could result in the risk of transient overload being underestimated, and management measures therefore being based on invalid assessment.

Mean and variance alone are insufficient to characterise a queue size probability distribution. An approximate distribution can be reconstructed from three quantities, mean, variance, and the probability that the queue is zero, or its complement utilisation (styled as a 'moment'). This avoids the complication of estimating higher moments such as skewness. Utilisation is relatively straightforward to obtain from the time derivative of the mean queue, and is intimately linked to the dynamics of queue development. Variance is more complicated and is related to the history of queue development through the integral of the mean.

Given the three moments, equilibrium distributions of some queue processes can be calculated explicitly in doubly-nested geometric form. For dynamic queues, inspired by the diffusion approximation, a numerical method of fitting an exponentially-weighted combination of exponential and Normal functions has been described. The method has been tested using a number of oversaturated peak cases, as well as arbitrary time-dependent profiles, using either an M/M/1 or an M/D/1 process. In the example presented, the risk of exceeding a realistic critical queue size is estimated to within about 3%. Future work may be able to refine the methods to improve the accuracy of estimation of distributions associated with more general forms of queuing process.

Acknowledgments

The work described in this paper forms part of the first author's PhD research under the supervision of the second author, second supervisor Dr Taku Fujiyama, at the Centre for Transport Studies, University College London. The general support of Dr Alan Stevens then Transportation Chief Scientist at the Transport Research Laboratory (TRL) is acknowledged. Neil H Spencer assisted with development of benchmark software for simulating queue size

probability distributions. We are grateful for the helpful comments and suggestions of the Special Issue Editor and two anonymous Reviewers.

References

Addison, J. D. and Heydecker, B. G. (2006). Journey time variability on a congested link. *Proc. UTSG Conference 2006,* Dublin.

Akçelik, R. (1980). Time-dependent expressions for delay, stop rate and queue length at traffic signals. *ARRB Internal Report AIR 367–1.* Australian Road Research Board.

Arup, Bates J., Fearon J. and Black I. (2004). *Frameworks for Modelling the Variability of Journey Times on the Highway Network*. Paper dft_econappr_pdf_610439. UK Department for Transport, London.

Catling. I. (1977). A time-dependent approach to junction delays. *Traffic Engineering and Control*, 18, 520-526.

Chow, J. Y. J. (2013). On observable chaotic maps for queueing analysis. *Proc. 92nd TRB Annual Meeting,* January 2013. Transportation Research Board, Washington DC.

Daganzo, C. (1994). The Cell Transmission Model: Network Traffic. *California PATH Working Paper UCB-ITS-PWP-94-12*. University of California, Berkeley CA.

Doherty, A. R. (1977). A comprehensive junction delay formula. *LTR1 working paper*, Department of Transport, London.

Fosgerau, M. (2008). On the relation between the mean and variance of delay in dynamic queues with random capacity and demand. *MPRA Paper No 11994*/Technical University of Denmark. http://mpra.ub.uni-muenchen.de/11994/

Gordon, A., van Vuren, T., Watling, D., Polak, J., Noland, R. B., Porter, S. and Taylor. N. B. (2001). Incorporating variable travel time effects into route choice models. *Proc. European Transport Conference 2001*, AET.

Griffiths, J. D., Leonenko, G. M. and Williams, J. E. (2005). The transient solution to M/E*k*/1 queue. *Operations Research Letters*, 34(2006), 349-354.

Gross, D., Shortle. J. F., Thompson, J. M. and Harris, C. M. (2008). *Fundamentals of queueing theory*. Wiley.

Heidemann, D. (1994). Queue length and delay distributions at traffic signals. *Transportation Research B*, 24B(5), 377-389. Elsevier.

Heidemann, D. (2001). A queueing theory model of non-stationary traffic flow. *Transportation Science*, 35(4), 405-412. Informs.

Kendall, D. G. (1951). Some problems in the theory of queues. *J. Royal Statistical Society B (Methodological)*, 13(2), 151-183.

Kimber, R. M. and Daly, P. (1986). Time-dependent queuing at road junctions: observation and prediction. *Transportation Research B*, 20B(3), 187-203.

Kimber, R. M., Daly, P., Barton, J, and Giokas, C. (1986). Predicting time-dependent distributions of queues and delays for road traffic at roundabouts and priority junctions. *J. Operational Research Society,* 37(1), 87-97. Palgrave Macmillan.

Kimber. R. M. and Hollis, E. M. (1979). Traffic queues and delays at road junctions. *TRL Report LR 909.* Transport Research Laboratory, Crowthorne House.

Kleinrock. L. (1975). *Queueing systems: Volume 1 Theory*. Wiley.

Kleinrock, L. (1976). *Queueing systems: Volume 2 Computational Aspects*. Wiley.

Kobayashi, H. (1974a). Application of the diffusion approximation to queueing networks – Part I: equilibrium queue distributions. *J. Association for Computing Machinery*, 21(2), 316-328.

Kobayashi, H. (1974b). Application of the diffusion approximation to queueing networks – Part II: non-equilibrium queue distributions and applications to computer modeling. *J. Association for Computing Machinery*, 21(3), 459-469.

Kouvatsos, D. D. (1988). A maximum entropy analysis of the G/G/1 queue at equilibrium. *J. Operational Research Society*, 39(2), 183-200.

Kühne, R. and Lüdke, A. (2013). Traffic breakdowns and freeway capacity as extreme value statistics. *Transportation Research C*, 27(2013), 159-168. Elsevier.

Machta, B. B., Chachra, R., Transtrum, M. K. and Sethna, J. P. (2013). Parameter space compression underlies emergent theories and predictive models. *Science 342*, 604-607.

Maher. M. J. and Hughes, P. (1997). A probit-based stochastic user equilibrium assignment model. *Transportation Research B*, 31(4), 341-355. Elsevier.

Medhi. J.(2003). *Stochastic models in queueing theory*. Elsevier Academic Press.

Morse, P. M. (1955). Stochastic properties of waiting lines. *J. Operations Research Society of America*, 3(3), 255-261, August 1955.

Newell, G. F. (1968). Queues with time-dependent arrival rates – Part I the transition through saturation. *Journal of Applied Probability*, 5(2), 436-451.

Newell, G. F. (1971/1982). *Applications of queuing theory*. Chapman and Hall. (First edition published 1971, second edition 1982).

Olszewski, P. S. (1990). Modelling of queue probability distribution at traffic signals. *Proc. International Symposium of Traffic and Transportation Theory*, 1990.

Robertson, D. I. (1969). TRANSYT: a traffic network study tool. *Report LR253*. Transport Research Laboratory, Crowthorne House.

Sharma, O. P. (1990). *Markovian queues*. Ellis Horwood.

Sheffi, Y. and Powell, W. (1982). An algorithm for the equilibrium assignment problem with random link times. *Networks 12*, 191-207.

Taylor, N. B. (2003). The CONTRAM dynamic traffic assignment model. *J. Networks and Spatial Economics – special issue on Dynamic Traffic Assignment*, 3(2003) 297-322, Kluwer.

Taylor, N. B. (2005). Variance and accuracy of the sheared queue model. *Proc. IMA Mathematics in Transport Conference*, 7-9 September 2005, University College London.

Taylor, N. B. (2007). A new approach to modelling variability in queues. *Proc. European Transport Conference*, Leeuwenhorst, 17-19 October 2007.

Taylor, N. B. (2011). An approach to time-dependent modelling of queues in multiple lanes with turning movements. *Proc. Universities Transport Studies Group (UTSG) Conference 2011*, Open University Milton Keynes.

Taylor, N. B. (2012). A recipe for jam - can congestion be defined consistently? *Proc. Road Traffic and Control (RTIC) Conference*, September 2012, IET London.

Taylor, N. B. (2013). *Queue methods for variability in congested traffic.* PhD dissertation. Department of Civil, Environmental and Geomatic Engineering, University College London.

Taylor, N. B. and Heydecker, B. G. (2013). The effect of green time on stochastic queues at traffic signals. *Transportation Planning and Technology*, Special Issue, 37(1), February 2014.

Webster, F. V. and Cobbe, B. M. (1966). Traffic signals. *Road Research Technical Paper 56*. HMSO.

Whitt, W. (1982). Refining diffusion approximations for queues. *Operations Research Letters*, 1(2), 165-168.

Zhao, Y. and Kockelman, K. M. (2001). The propagation of uncertainty through travel demand models: an exploratory analysis. *Proc. TRB Annual Meeting, January 2001*, Transportation Research Board, Washington DC.