J. Austral. Math. Soc. Ser. B 32(1991), 348–364

POISSON APPROXIMATIONS FOR TELECOMMUNICATIONS NETWORKS

T. C. BROWN[1] and P. K. POLLETT[2]

(Received 26 January 1990; revised 25 May 1990)

Abstract

In this paper, we review some techniques for studying traffic processes in telecommunications networks. The first of these allows one to identify Poisson traffic via the notion of "deterministic past-conditional arrival rate". Our approach leads to a method by which one can assess the degree of deviation of traffic processes from Poisson processes. We explain how this can be used to delimit circumstances under which traffic is *approximately* Poissonian.

1. Introduction

In the modelling of teletraffic systems, there are many instances where an assumption is made that the offered traffic is a Poisson process. This assumption arises largely in order that the models can be analysed simply, but it is certainly appropriate when it can be argued that the numbers of arrivals in a given time interval is independent of past arrivals, and has a Poisson distribution. As an immediate consequence of the Poisson assumption, the arrival rate is deterministic (non-random), conditional on the past of the process. This seemingly unremarkable property provides the key to the modern theory of traffic processes, for it actually *characterises* Poisson processes, in that any traffic process with the property must be a Poisson process. That deterministic past-conditional arrival rate *implies* Poisson traffic is a celebrated theorem of Watanabe, and in Section 2 we shall give a very elementary proof of this result.

Key words and phrases. Poisson approximations; networks; Watanabe's theorem.

[1]Department of Mathematics, The University of Western Australia, W.A. 6009.
[2]Department of Mathematics, The University of Queensland, Qld. 4072.

On many occasions, and this is particularly true of internal flow processes in message and packet-switching networks, the Poisson assumption is, at best, only a good approximation to reality. In Section 3 we shall present a simple method which allows one to assess the accuracy of this approximation. The method, introduced by the authors in the context of Markovian queueing networks (Brown and Pollett [4], Pollett [9]), involves establishing bounds on the degree of deviation from Poisson traffic. These bounds enable one to make precise predictions as to the circumstances in which the approximation to Poisson traffic is good. Further, they allow one to establish an approximate version of the loop criterion of Melamed [8] for identifying which traffic flows in Markovian networks are Poisson. We shall leave the technical details to Section 4, where some indication of the method of deriving the bounds is given and more recent work is sketched. There are surprising implications of the latter, even for the folklore of when the Poisson distribution is a good approximation to the Binomial distribution.

Some of the mathematics used may not be familiar to readers. An attempt has been made to keep the description informal, but, where this might lead to ambiguities, formal clarifications have been enclosed in square brackets.

2. Watanabe's theorem

Throughout, $\{N(t),\ t \geq 0\}$ will denote a traffic process, so that $N(t)$ is the (cumulative) number of arrivals in $[0, t]$. We shall assume that no two calls can arrive at *exactly* the same time. The assumption of a *unit* arrival rate, conditional on the past, can be written as

$$E_s(N(t) - N(s)) = t - s \tag{1}$$

where $0 \leq s \leq t$ and E_s represents expectation conditional on the past at time s. It is important to realise that the past, here, can include information *other than that given simply by arrivals*; in the context of switching networks, the past state of the the entire system, including both internal and external traffic, might be included. [Formally, E_s is expectation conditional on a σ-algebra $\mathscr{F}_s$, where $\{\mathscr{F}_s\}$ is an increasing, right-continuous family of σ-algebras, with $N(s)$ being $\mathscr{F}_s$ measurable and $\mathscr{F}_0$ being complete; see, for example, Brémaud [2], Appendices A1.5 and A1.7, for details].

Suppose that $0 < T_1 < T_2 < \ldots$ are the random variables giving the times of the 1st, 2nd, ... arrivals. These times are examples of stopping times; a *stopping time* is a *random* time, τ, such that, for each fixed time t, it is known whether or not τ has occurred before t. [Formally, $[\tau \leq t] \in \mathscr{F}_t$ for each $t \geq 0$]. For stopping times $\sigma \leq \tau$, (1) leads to the following identity:

$$E(N(\tau) - N(\sigma)) = E(\tau - \sigma). \tag{2}$$

[This can be established by considering the cases when σ and τ take a finite number of values: here (1) can be used directly. For other cases, convergence theorems are used to extend from the finite case. Equation (2) is valid even if one side is infinite, in which case both sides must be].

From (2), the familiar forward differential equations can be established. To see this, let $i \geq 1$ be fixed. We aim to get a differential equation for $p_i(t) = P(N(t) = i)$, $t > 0$, involving p_{i-1}. Now

$$p_i(t) = P(T_i \leq t) - P(T_{i+1} \leq t), \tag{3}$$

because there are exactly i arrivals up to and including t, if and only if the ith arrival is before or at t and the $(i+1)$th arrival is strictly after t. Let σ be the minimum of t and T_{i-1}, and τ be the minimum of t and T_i. Then, by considering the possibilities, $T_{i-1} \leq T_i \leq t$, $T_{i-1} \leq t < T_i$ and $t < T_{i-1} \leq T_i$, it is easy to see that

$$N(\tau) - N(\sigma) = \begin{cases} 1 & \text{if } T_i \leq t, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, in this case, (2) reads

$$P(T_i \leq t) = E\left(\int_0^t Y(s)ds\right),$$

where

$$Y(s) = \begin{cases} 1 & \text{if } N(s) = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, upon changing the order of expectation and integration (this can be justified because $Y \geq 0$), we get

$$\begin{aligned} P(T_i \leq t) &= \int_0^t EY(s)ds \\ &= \int_0^t p_{i-1}(s)ds, \end{aligned} \tag{4}$$

by the definition of Y and p_{i-1}. Combining (4) with (3) produces

$$p_i(t) = \int_0^t p_{i-1}(s)ds - \int_0^t p_i(s)ds,$$

and this can be differentiated to give

$$p_i' = p_{i-1} - p_i, \tag{5}$$

because the integrands must be continuous functions of s, being themselves integrals. In similar fashion we get, for $i = 0$,

$$p_0(t) = 1 - \int_0^t p_0(s)ds,$$

which gives

$$p_0' = -p_0.$$

This, together with (5), comprise the forward equations for the Poisson process. They can be solved in the usual way to show that $N(t)$ has a Poisson distribution with parameter t.

Of course, this conclusion is not sufficient to prove that $\{N(t)\}$ is a Poisson process. We also require the independence of $N(t) - N(s)$ $(0 \le s \le t)$ of the past at s. However, if A is a fixed event in the past at a fixed time s, and

$$M(t) = N(s + t) - N(s),$$

then (2) gives, for $0 \le t \le u$,

$$E_t'(M(u) - M(t)) = u - t,$$

where E_t' refers to expectation conditional on the past at time t, including the fixed event A. But the previous analysis then gives a Poisson(t) distribution for $M(t)$, that is, a Poisson distribution with parameter t for $N(s + t) - N(s)$, conditional on A. This establishes the required independence.

The standard textbook approach to the Poisson process (see, for example, Leon-Garcia [7]), is to assume that

$$\begin{aligned} P_s(N(s+t) - N(s) = 1) &= t + o(t) \\ P_s(N(s+t) - N(s) \ge 2) &= o(t) \\ P_s(N(s+t) - N(s) = 0) &= 1 - t + o(t), \end{aligned} \tag{6}$$

where $s, t \ge 0$, P_s refers to conditional probabilities given the past at time s, and $o(\cdot)$ refers to a function, f, such that $f(h)/h \to 0$ as $h \to 0$. The important difference with the assumption (1), apart from the fact that (6) is apparently much more detailed, is that (1) relates only to expectations (which have nice properties like linearity), whereas (6) makes assumptions about probabilities which may be harder to check. It is therefore remarkable that one can proceed from the simple statement (1) about expectations, to the detailed description of a Poisson process as having independent increments, unit rate increase of expectation *and* a Poisson distribution at each time point. We shall see, in Section 4, that the description (1) is the key to approximation results for the Poisson process.

3. Poisson approximations in teletraffic networks

Apart from those which emanate from outside, traffic processes in telecommunications networks are rarely Poisson. For example, Melamed [8] showed that in a Jacksonian queueing network, the flow from one node to another is Poisson if and only if units are "routed" in such a way that they can pass between the two nodes *at most once*. It is therefore of interest to delimit circumstances in which traffic is *approximately* Poissonian.

Suppose that $\{N(t)\}$ is a traffic process and let $\gamma(s)$ be the rate of the process at time s conditional on the past state of the network, the so-called *conditional intensity*. Since we shall be dealing only with Markovian models, $\gamma(s)$ is completely determined by the *present* state of the network and, moreover, an explicit expression can be written down for $\gamma(s)$ in terms of the state. Fix $t \geq 0$ and let A be an event detemined by N on $[0, t]$, so that, for example, A might be the event that the maximum time between arrivals in $[0, t]$ is 1, or the event that $N(s)/s \leq 2$ for all s in $[0, t]$. It is proposed to approximate $P(A)$ by $\Pi(A)$, where $\Pi(A)$ is the probability that A would have if N were Poisson with a specified rate, α. Using methods which will be outlined in the next section, it can be shown that, for arbitrary A,

$$|P(A) - \Pi(A)| \leq \int_0^t (E(\alpha - \gamma(s))^2)^{\frac{1}{2}}\, ds. \tag{7}$$

It is not proposed that this bound will give useful *numerical* estimates of the maximum difference in the probabilities, for reasons which we shall see in the next section. However, for any given model, we shall seek criteria under which the Poisson approximation is good, that is when this bound is close to 0. In the present context, we can further simplify the bound, since, because we shall always assume that the network in question is in equilibrium, N will be stationary, and so $\gamma(s)$ will have the same distribution for all s. Further, since we are at liberty to choose α in any way we please, it is sensible to choose a value which makes the bound as small as possible. This happens when $\alpha = E\gamma(s)$, and so we can use the following simple expression:

$$|P(A) - \Pi(A)| \leq t(\operatorname{Var}\gamma(s))^{\frac{1}{2}}. \tag{8}$$

Calculating the bound then simply amounts to determining the equilibrium variance of the intensity of the traffic process.

For example, consider a circuit-switched *star network*, where there are K outer nodes which communicate via a single central node. Thus there are K

links (circuit groups), and each route consists of a pair of links (i, j). We shall assume that the links have the same number of circuits, C. Assume also that calls requesting the various routes arrive according to independent Poisson streams and are offered at the same rate, $\rho = \nu/(K-1)$, that call lengths have a negative exponential distribution with mean 1 and that these are mutually independent, and independent of previous arrivals. Ziedins and Kelly [11] showed that, by keeping the total offered traffic, ν, fixed, and letting the number of switching nodes, K, become large, the traffic offered to any given *link* is approximately Poisson; recall that a call is said to be offered to link k if a call is offered to some route which includes link k and all other links on that route have at least one free circuit. This conclusion was made possible because the major result of their paper established that, for large K, the links are blocked *independently*. In particular, if $m_i(s)$ is the number of circuits in use on link i at time s, then, when the network is in equilibrium,

$$P(m_1 < C, m_2 < C) - (P(m_1 < C))^2 \to 0, \tag{9}$$

as $K \to \infty$. Their argument for the Poisson approximation was as follows: Let $N_k(t)$ be the number of calls offered to link k in the time-interval $[0, t]$. Then clearly γ_k, the conditional intensity of N_k, is given by

$$\gamma_k(s) = \sum_{j \neq k} \rho Y_j(s),$$

where

$$Y_j(s) = \begin{cases} 1, & \text{if } m_j(s) < C, \\ 0, & \text{if } m_j(s) = C. \end{cases}$$

Thus, in equilibrium,

$$E\gamma_k(s) = \nu P(m_1 < C),$$

and

$$\operatorname{Var} \gamma_k(s)$$
$$= \nu \left(\rho P(m_1 < C) + \left(\frac{K-1}{K-2} \right) \nu P(m_1 < c, m_2 < C) - \nu P(m_1 < C)^2 \right)$$

and so the result follows immediately from (8).

A simple extension of (8) involves considering a number of traffic processes simultaneously. Suppose that $N_1, N_2, \ldots, N_l$ are l such processes and $\gamma_1(s), \gamma_2(s), \ldots, \gamma_l(s)$ are their conditional intensities at time s. Then, if A is an event which is determined by these processes on the interval $[0, t]$, and $\Pi(A)$ is the probability that A would have if these processes were *independent* Poisson processes with rates $\alpha_1, \alpha_2, \ldots, \alpha_l$, then

$$|P(A) - \Pi(A)| \leq \sum_{k=1}^{l} \int_0^t (E(\alpha_k - \gamma_k(s))^2)^{\frac{1}{2}} \, ds.$$

If the network is in equilibrium and $\alpha_1, \alpha_2, \ldots, \alpha_l$ are chosen such that $\alpha_k = E\gamma_k(s)$, then a simpler bound is given by

$$|P(A) - \Pi(A)| \leq t \sum_{k=1}^{l} (\operatorname{Var} \gamma_k(t))^{\frac{1}{2}}. \tag{10}$$

The significance of this result in the analysis of the star network is clear: the traffic offered jointly to any l of the links, where $l < K$ is fixed, are *asymptotically independent* Poisson processes. This is consistent with Ziedins and Kelly's asymptotic result, (9), that links are blocked independently.

Let us now turn our attention to message and packet-switching networks. In contrast to circuit-switching networks, where all the links along a given path are used simultaneously, only one link is used at any one time in any given transmission, and transmissions must be received in their entirety at a given node before being transmitted along the next link in their route. If the link cannot transmit immediately, the message (or packet) is stored in a buffer until a circuit becomes available, hence the prevalent usage of the term *store-forward*. Packet-switching differs from message-switching in that messages are broken down into packets, which are transmitted individually in a store-forward fashion. Consequently, different parts of a given message can be transmitted simultaneously on successive links. Since packets are stored and forwarded in the same manner as messages, we shall henceforth use "message" as the generic term for both.

We shall label the links $1, 2, \ldots, J$ and we shall make the usual simplifying assumptions governing the way in which they operate. The links are perfectly reliable and are not subject to noise, so that message transmission times are determined by their length. The time taken by the nodes to switch, and, if necessary, buffer, reassemble and acknowledge messages, is negligible in comparison with their transmission times. Traffic entering the network from external sources is Poisson, and message lengths are mutually independent and exponentially distributed with mean 1. It will be convenient to suppose that the links function in the following way, conceived by Kelly [5]: A total effort (or capacity) of $\phi_j(n)$ is provided by link j when there are n messages whose transmission is incomplete. The message buffer for link j has distinct positions labelled $l = 1, 2, \ldots$. When there are n messages present, a proportion, $\eta_j(l, n)$, of the total capacity is offered to position l; when the transmission of this message is complete, messages previously occupying buffer positions $l+1, l+2, \ldots, n$ now move into positions $l, l+1, \ldots, n-1$, respectively. This apparent "shunting" of messages will not occur in practice, but rather one might imagine a list of free positions is maintained for the purpose of allocating free buffer space, whence l is merely the index of an occupied position. In the model, such indexing is superfluous and leads to unnecessary complications. If, for example,

$$\phi_j(n) = \begin{cases} \phi_j, & \text{for } n = 1, 2\ldots, \\ 0, & \text{for } n = 0, \end{cases} \tag{11}$$

and

$$\eta_j(l, n) = \begin{cases} \frac{1}{n}, & \text{for } l = 1, 2\ldots, n, \ \ n = 1, 2, \ldots, K_j - 1, \\ \frac{1}{K_j}, & \text{for } l = 1, 2\ldots, K_j, \ \ n = K_j, K_j + 1, \ldots, \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

then a total capacity, ϕ_j, is apportioned equally among, at most, the first K_j positions; this is termed the *processor-sharing* discipline. If K_j is infinite, the capacity is apportioned equally among *all* messages present.

Yet another purely conceptual assumption is made by supposing that, when a message which is required to be transmitted along link j arrives to find n buffer positions occupied, it moves into position l, where $l = 1, 2, \ldots, n + 1$, with probability $\delta_j(l, n+1)$, and messages in positions $l, l+1, \ldots, n$ now move into positions $l+1, l+2, \ldots, n+1$. Commonly, δ is defined by

$$\delta_j(l, n) = \begin{cases} 1, & \text{for } l = n, \ \ n = 1, 2, \ldots, \\ 0, & \text{otherwise,} \end{cases}$$

so that the new message "joins the end of the queue", but another useful prescription for δ_j is

$$\delta_j(l, n) = \begin{cases} \frac{1}{n}, & \text{for } l = 1, 2\ldots, n, \ \ n = 1, 2, \ldots, K_j - 1, \\ \frac{1}{K_j}, & \text{for } l = 1, 2\ldots, K_j, \ \ n = K_j, K_j + 1, \ldots, \\ 0, & \text{otherwise,} \end{cases}$$

which, together with (11) and (12), defines a *preemptive-resume processor-sharing* discipline. It should be stressed that the prescription of δ_j does not have precisely to mirror the actual engineering of the buffer, but rather, a regime is prescribed which has the same effect. Clearly it is necessary that $\phi_j(0) = 0$, $\phi_j(n) > 0$, for $n > 0$, and that

$$\sum_{l=1}^{n} \eta_j(l, n) = \sum_{l=1}^{n} \delta_j(l, n) = 1,$$

for all $n > 0$. The advantage of the model we have described is that the parameters η_j, δ_j and ϕ_j can be specified to model most of the usual service disciplines. For example, in addition to the abovementioned disciplines, we can accommodate *first-come-first-served* and *preemptive-resume last-come-first-served*.

https://doi.org/10.1017/S0334270000006913 Published online by Cambridge University Press

We shall identify messages by type, where the *type* of a message is a unique collection, $\mathbf{r}$, of links used by that message; we shall write $\mathbf{r} = (r_1, r_2, \ldots, r_{w(\mathbf{r})})$, where $w(\mathbf{r})$ is the total number of stages. We shall denote by $\mathscr{R}$ the set of all types, and we shall suppose that type-$\mathbf{r}$ messages arrive as a Poisson process with rate $\nu(\mathbf{r})$ and that $\mathscr{R}$ indexes *independent* Poisson processes. This setup is known as *fixed routing*, but note that *alternate routing* can be accommodated within the framework described (see Pollett [10]).

Let us first examine the net traffic, $\{N_k(t)\}$, offered to a given link k, on the interval $[0, t]$. Since all messages have unit mean length, the rate at which messages are transmitted by link j is $\phi_j(n)$ when there are n messages present. Thus, if $n_j(s)$ is the number of messages present at link j at time s, then γ_k, the conditional intensity of N_k, is given by

$$\gamma_k(s) = \nu_k + \sum_{j=1}^{J} \phi_j(n_j(s))\lambda_{jk},$$

where ν_k is the rate of externally offered traffic (irrespective of type), and λ_{jk} is the proportion of messages emanating from link j which next use link k (again, irrespective of type); all of these quantities can be written down explicitly in terms of $\nu(\mathbf{r})$, $\mathbf{r} \in \mathscr{R}$. For simplicity we shall suppose that $\Lambda = (\lambda_{jk})$ is an irreducible matrix and that, for each j,

$$b_j^{-1} = \sum_{n=0}^{\infty} \frac{\alpha_j^n}{\prod_{l=1}^{n} \phi_j(l)} < \infty, \tag{13}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_J$ is the unique strictly positive solution to the system of equations

$$\alpha_k = \nu_k + \sum_{j=1}^{J} \alpha_j \lambda_{jk}, \qquad k = 1, 2, \ldots, J.$$

Under these conditions, an equilibrium distribution exists for the model. Indeed, in equilibrium, the states of the individual links are independent, and the probability that there are n messages at link j is given by

$$\pi_j(n) = \frac{b_j \alpha_j^n}{\prod_{l=1}^{n} \phi_j(l)}.$$

Further, given the number of messages at link j, the message occupying buffer position l is of type $\mathbf{r}$ with probability

$$q_j^{\mathbf{r}} = \alpha_j^{\mathbf{r}} / \alpha_j,$$

where

$$\alpha_j^{\mathbf{r}} = \begin{cases} \nu(\mathbf{r}), & \text{if } r_s = j \text{ for some } s, \\ 0, & \text{otherwise,} \end{cases}$$

the message types being, themselves, conditionally independent (see, for example, Kelly [6]).

This result allows one to show that, when the network is in equilibrium,

$$E\gamma_k(s) = \alpha_k$$

and

$$\operatorname{Var}\gamma_k(s) = \sum_{j=1}^{J}\alpha_j\left(\sum_{n=0}^{\infty}\phi_j(n+1)\pi_j(n) - \alpha_j\right)\lambda_{jk}^2,$$

which, together with (8), establishes a bound on the degree of deviation of N_k from a Poisson process with rate α_k. Once the functions $\phi_1, \phi_2, \dots, \phi_J$ have been specified, the bound can be calculated explicitly. If, for each j, ϕ_j is given by (11), for example in the processor-sharing discipline, then

$$|P(A) - \Pi(A)| \le t\left(\sum_{j=1}^{J}\phi_j^2\lambda_{jk}^2\rho_j(1-\rho_j)\right)^{1/2}, \qquad (14)$$

where $\rho_j = \alpha_j/\phi_j$ (note that $\phi_j > \alpha_j$ in order that (13) be satisfied). Thus, if for each j, $\phi_j\lambda_{jk} = O(1)$ and either

(i) $\phi_j\lambda_{jk}$ is $o(J^{-1/2})$, or

(ii) ρ_j is $o(J^{-1})$, or

(iii) $1-\rho_j$ is $o(J^{-1})$,

then the approximation will be good. Circumstance (ii) corresponds to light traffic at link j, and therefore light input to link k from link j. Circumstance (iii) corresponds to heavy traffic at link j, and therefore link j will mostly be busy; note that while the link is busy, messages are transmitted to link k at a constant rate, $\phi_j\lambda_{jk}$. Circumstance (i) is perhaps the most interesting in that, if the network is large and the routing through the network and to the outside is relatively even, the probabilities λ_{jk} will be of the order of J^{-1}. Provided, then, that the transmission rates are moderate, (i) will hold for each j. In other words, the bound tells us that the Poisson approximation will be good in large networks with moderate transmission rates. This intuitive fact does not seem to be easily derivable by other means.

If $\phi_j(n) = n\phi_j$, that is, link j has infinite capacity, then the bound (14) is changed simply by removing the factor $(1-\rho_j)$ in each term of the sum, so that (i) and (ii) above continue to apply. If $\phi_j(n) = \phi_j\min\{n, C_j\}$, that is, link j has a maximum capacity of C_j, then a simple bound is obtained by replacing the factor $(1-\rho_j)$ by $(C_j-\rho_j)$. Thus, (ii) is unchanged but (i) becomes

(i)′ $\phi_j\lambda_{jk}$ is $o(C_jJ^{-\frac{1}{2}})$

https://doi.org/10.1017/S0334270000006913 Published online by Cambridge University Press

and the intuitive interpretation which arises is that the Poisson approximations will be good provided the network is large, the routing is even and the transmission rates are moderate in relation to the maximum capacity of each link.

Finally, we shall examine individually the components of the traffic offered to link k. Let $\mathscr{R}_k$ be the collection of routes that incorporate link k, that is $\mathscr{R}_k = \{\mathbf{r} \in \mathscr{R} : r_s = k \text{ for some } s = 1, 2, \ldots, w(\mathbf{r})\}$, so that if $N_k^{\mathbf{r}}(t)$ is the number of type-$\mathbf{r}$ messages to arrive at link k on $[0, t]$, then $N_k(t) = \sum_{\mathbf{r}\in\mathscr{R}_k} N_k^{\mathbf{r}}(t)$. Now fix $\mathbf{r}$ in $\mathscr{R}_k$ and let w be the (unique) stage such that $r_w = k$. Set

$$\nu_k^{\mathbf{r}} = \begin{cases} \nu(\mathbf{r}), & \text{if } w = 1, \\ 0, & \text{otherwise.} \end{cases}$$

If $u_j(l; s)$ is the type of the message in buffer position l on link j at time s, then $\gamma_k^{\mathbf{r}}$, the conditional intensity of $N_k^{\mathbf{r}}$, is given by

$$\gamma_k^{\mathbf{r}}(s) = \nu_k^{\mathbf{r}} + \sum_{j=1}^{J} \phi_j(n_j(s)) \sum_{l=1}^{n_j(s)} \eta_j(l, n_j(s)) Y_j(l; s), \tag{15}$$

where

$$Y_j(l; s) = \begin{cases} 1, & \text{if } u_j(l; s) = \mathbf{r}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that the sum on the right-hand side of (15) will be positive for at most one value of j. As one might expect, a simple calculation shows that $E\gamma_k^{\mathbf{r}}(s) = \alpha_k^{\mathbf{r}}$ when the network is in equilibrium. A similar calculation shows that, again in equilibrium, $\operatorname{Var}\gamma_k^{\mathbf{r}}(s) = q_k^{\mathbf{r}} \operatorname{Var}\gamma_k(s)$. Thus, individual bounds on the degree of deviation of $N_k^{\mathbf{r}}$ from a Poisson process with rate $\alpha_k^{\mathbf{r}}$ can be obtained from the previous bound by simply including a factor, $q_k^{\mathbf{r}}$. Further, and most importantly, all of the abovementioned criteria specify conditions under which $N_k^{\mathbf{r}}$, $\mathbf{r} \in \mathscr{R}$, can be approximated, jointly, by *independent* Poisson processes with rates $\alpha_k^{\mathbf{r}}$, $\mathbf{r} \in \mathscr{R}$. This conclusion is a consequence of (10) and the fact that the sum of the individual bounds is the original bound.

It should be stressed that the upper bounds of this section are just that: there is no implication that the circumstances they indicate for good Poisson approximations are the *only* such circumstances. We shall see, by analogy with simpler examples in the next section, that the bounds here are *not* in fact the best possible, although, for the relatively complicated examples of teletraffic networks, improvements await further developments in theory.

4. Derivation of the approximations

One key idea in deriving the approximations is to develop an equation analogous to (1) for an *arbitrary* process of arrivals. This can be done using the abovementioned notion of a *conditional intensity*. As before, let $\{N(t)\}$ be an arbitrary traffic process. Under regularity conditions, the conditional intensity process, $\{\gamma(s),\ s \geq 0\}$, defined by

$$\gamma(s) = \lim_{t \downarrow o} E_s(N(t+s) - N(s)), \tag{16}$$

exists, and, if $\Gamma(t) = \int_0^t \gamma(s)\,ds$, then

$$E_s(N(t) - N(s)) = E_s(\Gamma(t) - \Gamma(s)) \tag{17}$$

for $0 \leq s \leq t$. [For an *arbitrary* traffic process such that $EN(t) < \infty$ for all t, and arbitrary histories, there *always* exists an increasing, "previsible" process, $\{\Gamma(t)\}$, such that (17) holds, but, in the models described in the last section, the regularity conditions for (16) are certainly satisfied. See, for example, Brown [3] for the general case, which is called the Doob-Meyer decomposition of $\{N(t)\}$]. We then also have a general analogue, and indeed generalisation, of (2): for stopping times $\sigma \leq \tau$,

$$E_\sigma(N(\tau) - N(\sigma)) = E_\sigma(\Gamma(\tau) - \Gamma(\sigma)), \tag{18}$$

where E_σ means expectation conditional on the past at time σ [Formally, the past at time σ is the σ-field of events, A, such that $A \cap [\sigma \leq t] \in \mathscr{F}_t$ for each $t \geq 0$].

The other key idea in deriving the approximations is that of *coupling*. Recall that we wish to bound $|P(A) - \Pi(A)|$, where A is an event determined by the traffic process during $[0, t]$, and $\Pi(A)$ is the corresponding probability if N were a Poisson process of rate α. A measure-theoretic argument shows that to bound globally the error in the approximation, it suffices to consider events, A, of the form

$$A = [N(t_1) = j_1,\ N(t_2) - N(t_1) = j_2,\ \ldots,\ N(t_n) - N(t_{n-1}) = j_n],$$

where $n \geq 1$, $0 \leq t_1 < \cdots < t_n \leq t$ and $j_1, j_2, \ldots, j_n \in \{0, 1, 2, \ldots\}$. If $\{M(t),\ t \geq 0\}$ is *any* Poisson process of rate α, we are required to bound

$$|P(\mathbf{X} = \mathbf{j}) - P(\mathbf{Y} = \mathbf{j})|,$$

where

$$\mathbf{X} = (N(t_1),\ N(t_2) - N(t_1),\ \ldots,\ N(t_n) - N(t_{n-1})),$$

$$\mathbf{Y} = (M(t_1), M(t_2) - M(t_1), \dots, M(t_n) - M(t_{n-1}))$$

and $\mathbf{j} = (j_1, \dots, j_n)$. But this quantity is bounded by

$$P(\mathbf{X} = \mathbf{j} \text{ and } \mathbf{Y} \neq \mathbf{j}) + P(\mathbf{X} \neq \mathbf{j} \text{ and } \mathbf{Y} = \mathbf{j}) \leq P(\mathbf{X} \neq \mathbf{Y}). \tag{19}$$

Thus, if a Poisson process of rate α can be *constructed* from N and the last probability is bounded globally (over all n and all the t's), then this will be the desired bound. This procedure is *called* coupling because a *couple* of random objects, in this case the traffic processes N and M, are constructed in order to carry out the calculations.

To construct M, we define $\widehat{\Gamma}(t) = \inf\{s : \Gamma(s) \geq t\}$ so that $\widehat{\Gamma}$ is a (pseudo) inverse of the increasing (random) function $s \mapsto \Gamma(s)$; if Γ is strictly monotonic, then $\widehat{\Gamma}$ is indeed its inverse. In general $\Gamma(\widehat{\Gamma}(t)) = t$, but $\widehat{\Gamma}(\Gamma(t)) \geq t$, with inequality occurring at those t for which Γ is constant around an interval containing t. Certainly, $\widehat{\Gamma}(t) \geq \widehat{\Gamma}(s)$ for $t \geq s$ and hence we may apply (18) with $\sigma = \widehat{\Gamma}(s)$ and $\tau = \widehat{\Gamma}(t)$ to conclude that

$$E_{\widehat{\Gamma}(s)}\{N(\widehat{\Gamma}(t)) - N(\widehat{\Gamma}(s))\} = E_{\widehat{\Gamma}(s)}\{\Gamma(\widehat{\Gamma}(t)) - \Gamma(\widehat{\Gamma}(s))\} = t - s.$$

Thus, by the argument presented in Section 2, $\{N(\widehat{\Gamma}(t))\}$ is a unit-rate Poisson process and so $\{M(t)\}$, where $M(t) = N(\widehat{\Gamma}(\alpha t))$ is a Poisson process of rate α.

To compute the right side of (19) note that

$$P(\mathbf{X} \neq \mathbf{Y}) \leq \sum_{i=1}^{n} P(X_i \neq Y_i) \leq \sum_{i=1}^{n} E\,|X_i - Y_i|.$$

Each term in the last sum is bounded in a similar fashion, so we shall consider only the first. For this,

$$E|X_1 - Y_1| = E|N(t_1) - N(\widehat{\Gamma}(\alpha t_1))| = E\{N(\tau_1) - N(\sigma_1) + N(\tau_2) - N(\sigma_2)\},$$

where σ_1 is t_1 if $\widehat{\Gamma}(\alpha t_1) > t_1$ and t otherwise, τ_1 is $\widehat{\Gamma}(\alpha t_1)$ if $\widehat{\Gamma}(\alpha t_1) > t_1$ and t otherwise, σ_2 is $\widehat{\Gamma}(\alpha t_1)$ if $t_1 > \widehat{\Gamma}(\alpha t_1)$ and t otherwise, and τ_2 is t_1 if $t_1 > \widehat{\Gamma}(\alpha t_1)$ and t otherwise. Thus, using (18),

$$\begin{aligned} E\,|X_1 - Y_1| &= E\{N(\tau_1) - N(\sigma_1)\} + E\{N(\tau_2) - N(\sigma_2)\} \\ &= E\{\Gamma(\tau_1) - \Gamma(\sigma_1)\} + E\{\Gamma(\tau_2) - \Gamma(\sigma_2)\} \\ &= E\,|\alpha t_1 - \Gamma(t_1)|. \end{aligned}$$

https://doi.org/10.1017/S0334270000006913 Published online by Cambridge University Press

Finally, we obtain

$$P(\mathbf{X} \neq \mathbf{Y}) \leq \sum_{i=1}^{n} E\left|\int_{t_{i-1}}^{t_i} \alpha ds - \int_{t_{i-1}}^{t_i} \gamma(s) ds\right|$$
$$\leq \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} E|\alpha - \gamma(s)| ds$$
$$= \int_0^t E\,|\alpha - \gamma(s)|\, ds. \tag{20}$$

In order to simplify computation and interpretation of bounds it is often more convenient to use a slightly weaker bound than (20), obtained by observing that the integrand is bounded above by $(E(\alpha - \gamma(s))^2)^{\frac{1}{2}}$.

To see why it is not reasonable to expect the bounds given here and in the previous section to be particularly tight, we shall consider a somewhat simpler example, using similar methods to those described above. The example is also relevant to teletraffic theory. Indeed, the foreword to Traffic Engineering Report No. 14 of the Australian Government, Postmaster-General's Department, by C.W. Pratt, states

"Under simplifying assumptions of independence and uniformity of behaviour on the part of telephone users, the Binomial distribution is a satisfactory model for traffic offered by a relatively small group of subscribers. For a large number (say, greater than 200) of subscribers, each with a relatively small probability of being engaged in a conversation, the Poisson distribution is used as a good single-parameter approximation to the Binomial distribution. It is a limiting form of the Binomial distribution, and the nature of the limiting process implies an infinite number of subscribers."

Consider a single telephone exchange and suppose that there are n subscribers, who act independently. Let us generalise to the situation in which subscriber i has probability p_i of offering a call in a particular time period, and let X_i be 1 if the subscriber offers the call and 0 otherwise. Then $N = \sum_{i=1}^{n} X_i$ is the number of calls offered in that period. If $p_1 = p_2 = \cdots = p_n = p$ (say), then N has a Binomial distribution with parameters n and p. In the general case, the distribution of N is much more complicated and, in principle, one requires a separate computation for each n and each set of probabilities $p_1, p_2, \ldots, p_n$.

To use the coupling device, let $U_1, U_2, \ldots, U_n$ be independent and identically distributed random variables with uniform distributions on $(0, 1)$. We could realise $X_1, X_2, \ldots, X_n$ by setting $X_i = 1$ if $U_i > 1 - p_i$ and 0 otherwise, for then we would have that

$$P(X_i = 1) = P(1 - p_i < U_i \leq 1) = p_i$$

and that $X_1, X_2, \ldots, X_n$ are independent. Suppose that A is an event determined by N, and $\lambda = \sum_{i=1}^n p_i$. It is proposed to approximate $P(A)$ by $\Pi(A)$, where $\Pi(A)$ is the probability of A when N has a Poisson distribution with parameter λ. By the previous coupling argument,

$$|P(A) - \Pi(A)| \leq P(N \neq M), \tag{21}$$

where M is a random variable having a Poisson distribution with parameter λ. To construct M, take $p \geq 0$ and let $F(j)$ be the cumulative Poisson probability of $0, \ldots, j$, so that $F(j) = e^{-p} + pe^{-p} + \cdots + \frac{p^j}{j!}e^{-p}$ and $F(-1) = 0$. For $0 < x \leq 1$, let $\widetilde{F}(x) = j$, where j is the unique non-negative integer for which $F(j-1) < x \leq F(j)$ [$\widetilde{F}$ is again a pseudo inverse of F, but here $\widetilde{F}(x) = \inf\{t: F(t) \geq x\}$]. Consider, for fixed i, the random variable $Y_i = \widetilde{F}(U_i)$, with p_i replacing p. We then have that

$$P(Y_i = j) = P(F(j-1) < U_i \leq F(j)) = e^{-p_i}\frac{p_i^j}{j!},$$

so that Y_i does have a Poisson distribution with parameter p_i. Hence, by standard theory, M, defined by $M = \sum_{i=1}^n Y_i$, has a Poisson distribution with parameter λ. Moreover,

$$\begin{aligned} P(N \neq M) &\leq \sum_{i=1}^n P(X_i \neq Y_i) \\ &= \sum_{i=1}^n \{P(1 - p_i < U_i \leq e^{-p_i}) + P(U_i > (1 + p_i)e^{-p_i})\} \\ &= \sum_{i=1}^n p_i(1 - e^{-p_i}) \leq \sum_{i=1}^n p_i^2 \\ &\leq \lambda p, \end{aligned} \tag{22}$$

where the fact that $1 - p_i \leq e^{-p_i}$ has been used in the 2nd and 4th steps, and p is the maximum of $p_1, p_2, \ldots, p_n$. Combining (21) and (22), we have that

$$|P(A) - \Pi(A)| \leq \lambda p. \tag{23}$$

Thus, a good Poisson approximation is certainly obtained if the *maximum* probability of a subscriber offering a call is small and the mean offered traffic is moderate. It is worth noting that n, the number of subscribers, enters in the *opposite* way to the above advice on the use of the Poisson distribution: if p is small and n is not too large, then the approximation is sure to be good. What happens if n is large, p is small and np is large? It is mentioned later in the cited Traffic Engineering Report that, in these circumstances, the approximation should be good because both N and M will, by the Central

Limit Theorem, have approximate normal distributions. In fact, using a dazzling technique originally due to Stein, Barbour and Hall [1] established that

$$|P(A) - \Pi(A)| \leq (1 - e^{-\lambda})p \tag{24}$$

for all events, A, confirming the above reasoning. Moreover, they showed that, for any set of probabilities, $p_1, p_2, \ldots, p_n$, such that $\lambda \geq 1$, there exists an event, A, for which

$$|P(A) - \Pi(A)| \geq \frac{1}{32\lambda} \sum_{i=1}^{n} p_i^2,$$

so that (24), which is an improvement on (22), is the right order of magnitude. Thus, the advice should be that a Poisson approximation is appropriate when the maximum probability of a subscriber offering traffic is small, *irrespective* of the number of subscribers *or* of their uniformity of behaviour.

It is perhaps worth remarking that the bound (7) can be used to estimate the error in approximating the whole process of calls by a Poisson process. For simplicity, suppose that the subscribers wait for an exponentially distributed amount of time before offering a call and that the time period over which the call process is observed is scaled to be $[0, 1]$. If μ_i is the parameter of the exponential time, T_i, for the i^{th} subscriber, then $\mu_i = -\ln(1 - p_i)$ in order that $P(T_i \leq 1) = p_i$. The conditional intensity of the call process is $\gamma(s) = \sum_{i=1}^{n} \mu_i I[T_i > s]$. Assuming that the subscribers act independently, we then have

$$\operatorname{Var}(\gamma(s)) = \sum_{i=1}^{n} \mu_i^2 e^{-\mu_i s}(1 - e^{-\mu_i s}),$$

so that the bound for approximating the actual call process by a non-homogeneous Poisson process with rate $\sum_{i=1}^{n} \mu_i e^{-\mu_i s}$ is

$$\int_0^1 \left(\sum_{i=1}^{n} \mu_i^2 e^{-\mu_i s}(1 - e^{-\mu_i s}) \right)^{\frac{1}{2}} ds \leq \mu \lambda^{\frac{1}{2}}, \tag{25}$$

where μ is the maximum of $\mu_1, \mu_2, \ldots, \mu_n$. For small p, μ is approximately equal to p and so this bound is better for large λ than the bound (23); of course, when applied to the process at 1, (25) is a bound for the distance from a Poisson random variable with mean $\sum_{i=1}^{n} \mu_i$, rather than λ.

5. Conclusions

It is hoped that the paper has illustrated two vital themes of modern stochastic processes:

(i) that conditional expectations and stopping times can be profitably mixed;

(ii) that *construction* can be very helpful in calculations.

The last section has illustrated how both themes can be used in bounding errors in Poisson approximations, but the bounds are not necessarily the best possible. Work is currently in progress on

(a) combining the Stein-type approach of Barbour and Hall with conditional-intensity calculations, and

(b) approximating blocking probabilities using this approach.

References

[1] A. D. Barbour and P. Hall, "On the rate of Poisson convergence", *Math. Proc. Cam. Phil. Soc.* **95** (1984) 473–480.

[2] Brémaud, *Point processes and queues, martingale dynamics*, (Springer-Verlag, New York, 1981).

[3] T. C. Brown, *The Doob-Meyer decomposition and stochastic calculus*, Research Report 1988/4, Department of Mathematics, The University of Western Australia.

[4] T. C. Brown and P. K. Pollett, "Some distributional approximations in Markovian queueing networks", *Adv. Appl. Prob.* **14** (1982) 654–671.

[5] F. P. Kelly, "Networks of queues with customers of different types", *J. Appl. Probab.* **12** (1975) 542–554.

[6] F. P. Kelly, *Reversibility and stochastic networks*, (Wiley, New York, 1979).

[7] A. Leon-Garcia, *Probability and random processes for electrical engineering*, (Addison Wesley, London, 1989).

[8] B. Melamed, "Characterization of Poisson traffic streams in Jackson queueing networks", *Adv. Appl. Prob.* **11** (1979) 422–438.

[9] P. K. Pollett, "Distributional approximations for networks of quasireversible queues", in *Stochastic analysis and applications. Lecture notes in mathematics no. 1095* (eds. D. Williams and A. Truman), (Springer-Verlag, Berlin, 1984).

[10] P. K. Pollett, "Analysis of response times and optimal allocation of resources in message and packet switched networks", *Asia-Pacific J. Opert. Res.* **3** (1986) 134–149.

[11] I. B. Ziedins and F. P. Kelly, "Limit theorems for loss networks with diverse routing", *Adv. Appl. Probab.* **21** (1989) 804–830.