Approximating a Point Process by a Renewal Process, I: Two Basic Methods

WARD WHITT

Bell Laboratories, Holmdel, New Jersey

(Received December 1979; accepted November 1980)

This paper initiates an investigation of simple approximations for stochastic point processes. The goal is to develop methods for approximately describing complex models such as networks of queues and multiechelon inventory systems. The proposed approach is to decouple or decompose the model by replacing all the component flows (point processes) by independent renewal processes. Here attention is focused on ways to approximate a single point process by a renewal process. This is done in two steps: First, properties of the point process are used to specify a few moments of the interval between renewals; then a convenient distribution is fit to these moments. Two different methods are suggested for specifying the moments of the renewal interval. The stationary-interval method equates the moments of the renewal interval with the moments of the stationary interval in the point process to be approximated. The asymptotic method, in an attempt to account for the dependence among successive intervals, determines the moments of the renewal interval by matching the asymptotic behavior of the moments of the sums of successive intervals. These two procedures are applied to approximate the superposition (merging) of point processes. The purpose here is to provide a better understanding of these procedures and a general framework for making new approximations. In particular, the two basic procedures can be used as building blocks to construct refined composite procedures. Composite procedures for the $\sum G_i/G/1$ queue (with a superposition arrival process) are discussed by Albin in Part II. Albin has developed a hybrid procedure for approximating the mean sequence length and other characteristics in the $\sum G_i/G/1$ queue for which the average error when compared with simulated values was 3% over a large number of test systems.

COMPLEX stochastic models such as networks of queues are necessary to capture the essence of many complex systems such as communication networks. However, complexity means that approximations are often needed. Motivated by this need, we develop a general framework and several specific procedures for approximating a point process by a renewal process characterized by a few parameters.

The approximating processes are renewal processes rather than more general processes, not because it is impossible to describe a point process

Subject classifications: 569 approximating a point process by a renewal process, 683 approximations for point processes and queues, 694 approximations for networks of queues.

Operations Research
Vol. 30, No. 1, January–February 1982

0030-364X/82/3001-0125 $01.25

in more detail, but because it is usually difficult to obtain useful descriptions of models involving more general processes. What seems desirable for the practical analysis of complex stochastic systems is a procedure for obtaining simple parametric characterizations of stochastic processes and an elementary calculus for transforming the parameters to represent the basic operations of composition, decomposition, flow through a queue, overflow, etc. In fact, such comprehensive programs have recently been proposed for analyzing networks of queues in computer systems; see Chandy and Sauer [1978], Kuehn [1979a, b], Sauer and Chandy [1975], and Sevcik et al. [1977]. These authors suggest approximating all the flows (point processes) in a network of queues by renewal processes characterized by two parameters. It was discovered that one parameter (representing the rate of the process) is usually not good enough, but two parameters (representing the rate and the variability) often are sufficient. Closely related parametric approximation procedures are the equivalent random method (Cooper [1980], Eckberg [1976], Freeman [1976], Heffes [1973], Kosten [1977], Kuczura [1973a], Wilkinson [1956, 1970]) and diffusion approximations (Borovkov [1976], Chandy and Sauer, Halachmi and Franta [1978], Halfin and Whitt [1981], Whitt [1974]). Other simple parametric approximations for queues appear in Al-Khayyal and Gross [1977], Greenberg [1980], Marchal and Harris [1976], and Morse [1958].

In this paper the parameters are the moments of the renewal interval in the approximating renewal process. It is often appropriate to consider other parameters such as distribution percentiles instead of moments; see Bux and Herzog [1977], and Lazowska [1977], but we focus on moments here. As shown by Holtzman [1973] and Eckberg [1977], there can be considerable variability in queueing characteristics such as the blocking probability in a loss system over the set of all possible (interarrival-time) distributions with two moments or two other parameters held fixed. The two-parameter approximations apparently work reasonably well, because extreme cases tend not to arise. The variability in queueing characteristics with fixed interarrival-time moments is substantially reduced if the interarrival-time distribution must also have a reasonable shape, e.g., if it is unimodal or if it has a bounded failure rate; see Klincewicz and Whitt [1981].

Our research stems from an investigation of Kuehn's [1979a] procedure for approximating the superposition (sum) of two independent renewal processes by a renewal process. Kuehn used what we call the stationary-interval method or MICRO, but it seemed that a different procedure (what we call the asymptotic method or MACRO) might be both easier to use and more accurate. With the stationary-interval method, we take a microscopic view and try to match the behavior of the point process during a relatively short time interval. In particular, with the stationary-interval method the interval between renewals in the approximating renewal process is given the stationary distribution of an interval in the superposition process. This is an approximation because the successive intervals in a renewal process are independent while the successive intervals in the superposition process are not. The asymptotic method is an attempt to take account of the dependence. With the asymptotic method, we take a macroscopic view and try to match the behavior of the point process over a relatively long time interval. In particular, with the asymptotic method, the moments of the interval between renewals in the approximating process are obtained by matching the moments of the renewal counting process over a large time interval with the corresponding moments of the point process over a large time interval. Our original idea, which is equivalent for the first two moments, was to choose the approximating moments by matching the normalization constants in the central limit theorem. As expected, the stationary-interval method and the asymptotic method yield the same rate, i.e., the same first moment of the interval between renewals, but the approximating second moments or variances can be very different.

In order to compare the two methods, we considered the $\sum GI_i/M/1$ queue: a single-server queue with an exponential service-time distribution, an infinite waiting room, a FIFO (first in, first out) queue discipline, and an arrival process which is the superposition of two or more independent renewal processes. Such a superposition process naturally arises when the arrival process is the superposition of independent overflow processes or departure processes in a network of queues and these processes are individually approximated by renewal processes, for example as in Kuczura [1973a], Rath and Sheng [1979], or Kuehn [1979a]. If all the renewal processes are Poisson processes, then the superposition process is a Poisson process and we have the elementary $M/M/1$ queue. If all but one renewal process is a Poisson process, then all the Poisson processes can be combined into a single Poisson process and we have a $GI{+}M/M/1$ queue. With some difficulty, analytical results can be obtained for this system; see Sahin [1971], Kuczura [1972], and Fujisawa [1976]; for related theory, also see Kuczura [1973b,c], and Rolski [1978]. However, we know of no analytical results for the $\sum GI_i/M/1$ queue when two or more of the component processes are general (non-Poisson) renewal processes. If the renewal-interval distribution in each component renewal process is constructed from exponential building blocks, e.g., a mixture of convolutions of exponential distributions (such as special phase-type distributions, see Neuts [1978], then the method of stages can be applied to solve this system; see Chapter 4 of Kleinrock [1975]. But in general the simple system $GI{+}GI/M/1$ is remarkably intractable. To a large extent, the difficulty occurs because the queue length process is not regenerative: there is no embedded renewal process.

Hence there is considerable interest in developing approximations for the $\sum GI_i/M/1$ queue. We used the stationary-interval method and the asymptotic method to approximate the superposition arrival process by a single renewal process. It is then easy to describe the steady-state distribution of the number of customers in the resulting $GI/M/1$ system. We then compared the two approximating steady-state distributions with the actual distribution as estimated by computer simulation. The results of our preliminary experiments appear in Whitt [1979], Section 6. Our suspicions were confirmed to a large extent. The asymptotic method is easier to use and often works better, especially for queues with heavy loads, but neither procedure dominates the other.

However, it is apparent from these preliminary experiments that the simulated values often fall between the two approximations. This suggests that better approximations might be found by combining the basic approximations. Moreover, from experience with heavy-traffic limit theorems (Whitt [1974]), we know that the asymptotic method is asymptotically correct for the $\sum GI_i/G/1$ queue as the traffic intensity ρ approaches the critical value one. Also, from experience with the convergence of superposition processes to the Poisson process (Çinlar [1972], Khintchine [1960]), we know that the stationary-interval method is asymptotically correct as the number n of component arrival processes gets large. These properties suggest using convex combinations of the two basic procedures with the weight on the asymptotic method being an increasing function of ρ and a decreasing function of n. This possibility has been investigated by Albin [1980, 1981a, b, 1982] who found that dramatic improvements can be obtained using refined composite procedures. Hence, we present the two basic procedures MICRO and MACRO here, not to determine which is best, but to provide the tools for constructing better approximations. We regard the procedures here as building blocks that the model builder can use to construct better approximations. This point of view is strongly supported by Albin's results (to be reported in Part II). In this sequel, Albin develops refined composite procedures for the $\sum G_i/G/1$ queue.

We began by investigating procedures for approximating the superposition of independent renewal processes, but it soon became clear that the basic principles underlying these procedures apply to the approximation of any point process. Hence, we present the procedures as they can be applied to approximate any point process before we discuss the specific application to superposition. Moreover, we not only specify approximation algorithms; we also try to communicate the ideas behind the algorithms. A major purpose of this paper is to make the different approximation procedures easier to understand. This in turn should make it easier to select a specific procedure in an application.

This paper is organized as follows. In Section 1 we give some background on point processes. In Section 2 we describe the basic procedures for selecting the approximating renewal-interval moments. We then describe methods for fitting distributions to these moments in Section 3. We apply the approximation procedures to the superposition of point processes in Section 4 and briefly discuss experimental results in Section 5. Additional material appears in Whitt [1979] which can be obtained from the author. In Section 5 therein we apply the approximation procedures to the decomposition or splitting of point processes. In Section 6 of the same, we present preliminary experimental results comparing the procedures using the $\sum GI_i/M/1$ queue. We make a connection to the equivalent random method in Section 7. In particular, we show that the asymptotic method developed here is the heavy-traffic limit of the equivalent random method.

1. A FEW POINTS ABOUT POINT PROCESSES

The stochastic processes we consider are point processes on the positive real line; see Daley and Vere-Jones [1972] and Cherry [1972], Çinlar, Cox and Lewis [1966], Eckberg, Haskell [1974], Lawrence [1973], Neuts [1979], Rodriguez [1976]. In the point processes we consider, the total number of points is infinite but the number of points in any bounded interval is finite. Let S_n denote the position of the nth point from the origin, $n \geq 1$. Let $S_0 = 0$, but with the understanding that S_0 does not correspond to a point. If there are k points at 0, then $S_i = 0$ for $i = 1, \cdots, k$. Let X_n denote the interval between the nth and $(n-1)$st points, i.e., $X_n = S_n - S_{n-1}$, $n > 1$. Let $\{N(t), t \geq 0\}$ be the associated counting process recording the number of points in the interval $[0, t]$, i.e.,

$$N(t) = \max\{n \geq 0 : S_n \leq t\}, \; t \geq 0. \tag{1.1}$$

Since $S_n \leq t$ if and only if $N(t) \geq n$ for all n and t, the processes $\{S_n\}$ and $\{N(t)\}$ can be regarded as inverse processes. Given $\{N(t)\}$, we can construct $\{S_n\}$ and $\{X_n\}$ by setting $S_0 = 0$,

$$S_n = \min\{t \geq 0 : N(t) \geq n\}, \; n \geq 1, \tag{1.2}$$

and $X_n = S_n - S_{n-1}$. In other words, the stochastic processes $\{S_n\}$, $\{X_n\}$ and $(N(t)\}$ are three different representations of the same point process.

We usually assume the point processes to be approximated are stationary, but it is important to note that there are two different kinds of stationarity: one for the counting process $\{N(t)\}$ and the other for the interval sequence $\{X_n\}$. We say the counting process $\{N(t)\}$ is stationary if the distribution of the increments is independent of time shifts, i.e., if the joint distribution of the k-tuple.

$$[N(t_1 + h) - N(s_1 + h), N(t_2 + h) - N(s_2 + h), \cdots, N(t_k + h) - N(s_k + h)]$$

is independent of h ($h > 0$) for all positive integers k and all k-tuples $(s_1, \cdots, s_k)$ and $(t_1, \cdots, t_k)$ with $0 \le s_i < t_i$, $i = 1, \cdots, k$. We say that the interval sequence $\{X_n\}$ is stationary if the joint distribution of the k-tuple

$$[X_{n_1+h}, X_{n_2+h}, \cdots, X_{n_k+h}]$$

is independent of h (h a positive integer) for all positive integers k and all k-tuples of nonnegative integers $(n_1, \cdots, n_k)$.

The counting process $\{N(t)\}$ and the associated interval sequence $\{X_n\}$ obtained from (1.2) are usually not both stationary at the same time. However, it is significant that under mild regularity conditions there is a one-to-one correspondence between stationary counting processes and stationary interval sequences. This one-to-one correspondence is based on Palm probabilities; see Chapter 3 of Khintchine, Section 7 of Daley and Vere-Jones, Jagers [1973], and Port and Stone [1977]. To avoid confusion, let $\{N_s(t)\}$ represent a stationary counting process with associated interval sequence $\{Y_n)$ defined by (1.2), and let $\{N(t)\}$ be the associated Palm counting process (obtained from the one-to-one correspondence) with associated stationary interval sequence $\{X_n\}$. Note that the interval sequence $\{X_n\}$ is related to the counting process $\{N(t)\}$ just as $\{Y_n\}$ is related to $\{N_s(t)\}$: by (1.1) and (1.2). The Palm counting process $\{N(t)\}$ can be viewed as the stationary counting process $\{N_s(t)\}$ conditioned on the occurrence of a point at the origin. The one-to-one correspondence is valid if the stationary counting process has finite intensity, i.e., if $EN_s(t) < \infty$, or if $\lim_{t\to\infty} N_s(t)/t < \infty$; see Port and Stone; we assume one of these conditions is satisfied.

The one-to-one correspondence between the stationary counting process $\{N_s(t)\}$ and the Palm process $\{N(t)\}$ applies to the distributions of the stochastic processes, i.e., the joint distributions, but for our approximations we use only the one-dimensional distributions. Henceforth, let F be the c.d.f. of X_1 and G be the c.d.f. of Y_1. Also let $\Psi_k(t) = P(N_s(t) = k)$ and $\Phi_k(t) = P(N(t) = k)$. Let λ be the intensity of $\{N_s(t)\}$, i.e., $\lambda = \lim_{t\to\infty} N_s(t)/t$, which equals $EN_s(1)$ if this expectation is finite. Then the one-to-one correspondence between $\{N_s(t)\}$ and $\{N(t)\}$ implies that

$$\Psi_0(t) \equiv P(N_s(t) = 0) = 1 - \lambda \int_0^t \Phi_0(u)du$$

and (1.3)

$$\Psi_k(t) \equiv P(N_s(t) = k) = \lambda \int_0^t [\Phi_{k-1}(u) - \Phi_k(u)]du;$$

see Khintchine, p. 40, or Daley and Vere-Jones, p. 358. Then

$$\Phi_0(t) = -\lambda^{-1}\Psi_0'(t)$$

and (1.4)

$$\Phi_k(t) = \Phi_{k-1}(t) - \lambda^{-1}\Psi_k'(t), \; t \ge 0.$$

Also

$$G(t) = 1 - \Psi_0(t) = \lambda \int_0^t \Phi_0(u)du = \lambda \int_0^t [1 - F(u)]du, \; t \geq 0, \quad (1.5)$$

and

$$F(t) = 1 - \Phi_0(t) = 1 + \lambda^{-1}\Psi_0'(t) = 1 - \lambda^{-1}G'(t), \; t \geq 0. \quad (1.6)$$

2. THE APPROXIMATION RECIPES

Consider a stationary counting process $\{N_s(t)\}$ with associated Palm process $\{N(t)\}$ having stationary interval sequence $\{X_n\}$ and partial sums $\{S_n\}$. Suppose we wish to approximate this point process by a renewal process. We propose to match the representations in the natural way: the renewal interval sequence is the approximation for $\{X_n\}$; the associated renewal counting process is the approximation for $\{N(t)\}$; and the associated equilibrium renewal process is the approximation for $\{N_s(t)\}$; see Ross ([1970], p. 44). With this matching convention understood, it suffices to specify the renewal interval c.d.f., which we denote by H. Our first step is to determine two or more moments of H. A specific distribution is later fit to these moments when it is needed; see Section 3.

The method proposed here is to match the first m moments of the nth partial sum in the renewal process with the first m moments of S_n. To make this match, it is convenient to work with cumulants or semi-invariants instead of the moments; see Section 15 of Gnedenko and Kolmogorov [1968]. Let $\beta_j \equiv \beta_j(Z)$ be the jth cumulant of the random variable Z, i.e., the coefficient of t^j in the power series representation of $\log Ee^{tZ}$. The cumulants have the desirable property that $\beta_j(Z_1 + \cdots + Z_n) = n\beta_j(Z_1)$ for all j and n if the sequence $\{Z_n\}$ is i.i.d. The first two cumulants are the mean and variance. In general there is a one-to-one correspondence between the first m moments and the first m cumulants; see Gnedenko and Kolmogorov (p. 65).

The approximation procedure then is to select m and n, calculate or estimate $\beta_j(S_n), j = 1, 2, \cdots, m$, and then set

$$\beta_j(H) = \beta_j(S_n)/n, \qquad j = 1, 2, \cdots, m. \quad (2.1)$$

The stationary-interval method and the asymptotic method are the extreme cases in (2.1): The stationary-interval method uses $n = 1$ and the asymptotic method uses $n = \infty$.

With the stationary-interval method, we let the cumulants and moments of the renewal c.d.f. H be precisely the cumulants and moments of the c.d.f. F of the stationary interval X_1, but we ignore the dependence between the successive intervals. If instead of the Palm process $\{N(t)\}$ we begin with the stationary counting process $\{N_s(t)\}$ and the c.d.f. G of Y_1, then we can first calculate the stationary-interval distribution F using

(1.6). However, it is possible to calculate the moments of F from the moments of G without using (1.6). If $\mu_j(F)$ and $\mu_j(G)$ are the jth moments about the origin of F and G, respectively, then

$$\mu_{j+1}(F) = \mu_j(G)\mu_1(F)(j + 1), \qquad (2.2)$$

see Cox ([1972], p. 64).

For $n \geq 2$, the cumulants $\beta_j(S_n)$ are affected by the dependence among the random variables $X_n, n \geq 1$. For example, the variance $\beta_2(S_n)$ includes the covariances, i.e.,

$$\beta_2(S_n) = \textstyle\sum_{k=0}^{n-1} (n - k) \operatorname{Cov}(X_1, X_{1+k}), \qquad (2.3)$$

where $\operatorname{Cov}(X_1, X_1) = \operatorname{Var}(X_1)$.

Note that $\beta_1(S_n) = n\beta_1(X_1)$ for all n because the expected value of the sum is the sum of the expected values, even if the random variables are dependent. Hence, the procedure in (2.1) yields $\beta_1(H) = \beta_1(X_1)$ for all n.

The asymptotic method is the limit of (2.1) as $n \rightarrow \infty$. The general idea is to fit the renewal process to the point process by matching the behavior over relatively long time intervals. We assume that the limits $\lim_{n\rightarrow\infty} \beta_j(S_n)/n$ exist. The procedure then is to set

$$\beta_j(H) = \lim_{n\rightarrow\infty} \beta_j(S_n)/n, \qquad j = 1, 2, \cdots, m. \qquad (2.4)$$

To help fix these ideas, consider the following simple example.

Example 2.1. To see how the stationary-interval (MICRO) method and asymptotic (MACRO) method differ, suppose $P\{(X_i, X_{i+1}) = (1, 2) \text{ or } (2, 1)\} = 1$ for all i and $P(X_1 = 1) = P(X_1 = 2) = 1/2$. Then the stationary-interval method gives X distributed as X_1 while the asymptotic method gives $P(X = 1.5) = 1$.

It is also possible to implement the asymptotic method directly from the counting process $\{N(t)\}$. We point out that this is particularly convenient for treating superposition processes; see Section 4. To do this, we use the cumulants of $N(t)$ instead of the cumulants of S_n, assuming that $\beta_j(N(t))/t$ converges as $t \rightarrow \infty$. We then let

$$\gamma_j(H) = \lim_{t\rightarrow\infty} \beta_j(N(t))/t, \qquad j = 1, 2, \cdots, m, \qquad (2.5)$$

where $\gamma_j(H)$ are parameters from which we can calculate the moments of H. In particular, the parameters $\gamma_j(H)$ are the functions of the moments of H that arise as limits when $\{N(t)\}$ is actually a renewal process. Smith [1959] showed that the limits in (2.5) exist for renewal processes under mild regularity conditions and determined their form. Moreover, Smith showed that for a renewal process there is a one-to-one correspondence between the first m limits $\gamma_j(H)$ and the first m moments $\mu_j \equiv \mu_j(H)$ of H; also see Chapter 4 of Cox, and Chapter 1 of Murthy [1974]. Assume that $\mu_{m+1} < \infty$ and that the k-fold convolution of F has an absolutely continuous component for some $k \geq 1$. Smith showed that if $\{N(t)\}$ is a renewal counting process with interval c.d.f. H, then

$$\beta_j(N(t)) = \gamma_j t + \delta_j + o(1), \qquad 1 \leq j \leq m. \tag{2.6}$$

Moreover, Smith indicated how to calculate γ_j as a function of $\{\mu_1, \cdots, \mu_j\}$ and δ_j as a function of $\{\mu_1, \cdots, \mu_{j+1}\}$, $1 \leq j \leq m$. By combining pp. 4, 20 and 24 of Smith and replacing μ_j with $\mu_j \mu_1^{-j}$ for each j and t with $t\mu_1^{-1}$ because of the assumption that $\mu_1 = 1$ there, we can obtain the coefficients γ_j and δ_j for any j. We list the values for $j = 1, \cdots, 4$:

$$\begin{aligned}
\gamma_1 &= \mu_1^{-1} \\
\gamma_2 &= \mu_1^{-3}(\mu_2 - \mu_1^2) \\
\gamma_3 &= \mu_1^{-5}(-\mu_3\mu_1 + 3\mu_2^2 - 3\mu_2\mu_1^2 + \mu_1^4) \\
\gamma_4 &= \mu_1^{-7}(\mu_4\mu_1^2 - 10\mu_3\mu_2\mu_1 + 15\mu_2^3 + 6\mu_3\mu_1^3 - 18\mu_2^2\mu_1^2 \\
&\quad + 7\mu_2\mu_1^4 - \mu_1^6)
\end{aligned} \tag{2.7}$$

and

$$\begin{aligned}
\delta_1 &= \mu_1^{-2}((1/2)\mu_2 - \mu_1^2) \\
\delta_2 &= \mu_1^{-4}((5/4)\mu_2^2 - (2/3)\mu_3\mu_1 - (1/2)\mu_2\mu_1^2) \\
\delta_3 &= \mu_1^{-6}((3/4)\mu_4\mu_1^2 - 5\mu_3\mu_2\mu_1 + (11/2)\mu_2^3 + 2\mu_3\mu_1^3 \\
&\quad - (15/4)\mu_2^2\mu_1^2 + (1/2)\mu_2\mu_1^4) \\
\delta_4 &= \mu_1^{-8}(-(4/5)\mu_5\mu_1^3 + (17/2)\mu_4\mu_2\mu_1^2 + (16/3)\mu_3^2\mu_1^2 \\
&\quad - 44\mu_3\mu_2^2\mu_1 + (279/8)\mu_2^4 - (9/2)\mu_4\mu_1^4 + 30\mu_3\mu_2\mu_1^3 \\
&\quad - 33\mu_2^3\mu_1^2 - (14/3)\mu_3\mu_1^5 + (35/4)\mu_2^2\mu_1^4 - (1/2)\mu_2\mu_1^6).
\end{aligned} \tag{2.8}$$

Of course, $\gamma_j \equiv \gamma_j(H)$ in (2.6) are of principal interest because they are what we use in (2.5). To get the moments of H, we invert (2.7), obtaining

$$\begin{aligned}
\mu_1 &= \gamma_1^{-1} \\
\mu_2 &= \mu_1^3\gamma_2 + \mu_1^2 \\
\mu_3 &= -\mu_1^4\gamma_3 + 3\mu_2^2\mu_1^{-1} - 3\mu_2\mu_1 + \mu_1^3 \\
\mu_4 &= \mu_1^5\gamma_4 + 10\mu_3\mu_2\mu_1^{-1} - 15\mu_2^3\mu_1^{-2} - 6\mu_3\mu_1 + 18\mu_2^2 \\
&\quad - 7\mu_2\mu_1^2 + \mu_1^4.
\end{aligned} \tag{2.9}$$

For a renewal process, the equations in (2.6)–(2.9) are exact, but for other point processes they are approximations. The Poisson case is a convenient check: then $\gamma_j = \mu_1^{-1}$, $\delta_j = 0$, and $\mu_j = \mu_1^j j!$ for $j = 1$. The expansion in (2.6) is for renewal processes, but a similar expansion holds for delayed renewal processes such as the equilibrium renewal process. The only change in (2.6) is a change in the constants δ_j in (2.8); see Chapter 1 of Murthy.

It is worth noting that with the asymptotic method the first two cumulants (the mean and variance) of the approximating renewal interval c.d.f. H could also be obtained by matching normalization constants in a central limit theorem, either for $\{S_n\}$ or $\{N(t)\}$. To be precise, let $\Rightarrow$ denote convergence in distribution and let $N(0, 1)$ be a standard normal random variable with mean 0 and variance 1. Recall that for a renewal process,

$$(S_n - n\beta_1(H))/\sqrt{n\beta_2(H)} \Rightarrow N(0, 1) \quad \text{as} \quad n \to \infty \tag{2.10}$$

and

$$(N(t) - t\gamma_1(H))/\sqrt{t\gamma_2(H)} \Rightarrow N(0, 1) \quad \text{as} \quad t \to \infty; \tag{2.11}$$

see pp. 259 and 372 of Feller [1971]. However, in order to apply (2.10) or (2.11) it is necessary to assume that the point process to be approximated satisfies such a limit theorem, which is usually a stronger assumption than assuming $\beta_2(S_n)/n$ or $\beta_2(N(t))/t$ converges. In most applications, though, it is safe to assume that all the limits—(2.4), (2.5), (2.10) and (2.11)—exist. It is important to note that these four formulas yield the same two approximating moments. Moreover, these limits are the same for a renewal process and the associated equilibrium renewal process.

In closing this section, we mention again that it may be desirable to consider other parameters besides the moments of the renewal interval, e.g., distribution percentiles.

3. FITTING A DISTRIBUTION TO THE MOMENTS

Having specified several moments of the interval between renewals, we can completely specify the approximating renewal process by fitting a convenient distribution to those moments. However, the moments themselves will be sufficient for many applications. Further steps in approximating complex systems often will involve only manipulations of these parameters. Moreover, the moments can be used to bound the distribution or its Laplace transform; see page 228 of Feller, Eckberg, and references there.

It is important to realize that since the procedures are approximations there is no guarantee that there exists any c.d.f. with the m numbers as moments. It is easy to see that the procedures here always produce nonnegative numbers for the mean and variance, so there is no problem if we work with only two moments. However, if we use more than two moments, then it is possible for the candidate moments to be inconsistent. We remark that this phenomenon actually occurs in practice, even for m

only three or four; the third moment can be inconsistent in the superposition approximation; see Whitt [1979], Example 4.2.

Hence, in programs implementing the approximation procedures with more than two moments we have included a moment consistency check. Fortunately, relatively simple necessary and sufficient conditions for moment consistency are available; see pages 106, 171 of Karlin and Studden [1966]. Let $\mu_1, \cdots, \mu_m$ be numbers that are prospective moments for a nonnegative random variable. For $m \leq 5$, these numbers are moments for some probability distribution on the positive real line if and only if for each k, $k = 1, \cdots, m$, the following hold:

$$\begin{aligned}
\mu_1 &\geq 0, \quad k = 1, \\
\mu_2 - \mu_1^2 &\geq 0, \quad k = 2, \\
\mu_1\mu_3 - \mu_2^2 &\geq 0, \quad k = 3, \qquad (3.1) \\
\mu_2\mu_4 - \mu_3^2 - \mu_1(\mu_1\mu_4 - \mu_2\mu_3) + \mu_2(\mu_1\mu_3 - \mu_2^2) &\geq 0, \quad k = 4, \\
\mu_1(\mu_3\mu_5 - \mu_4^2) - \mu_2(\mu_2\mu_5 - \mu_3\mu_4) + \mu_3(\mu_2\mu_4 - \mu_3^2) &\geq 0, \quad k = 5.
\end{aligned}$$

If the c.d.f. of the renewal interval is exponential, i.e., if $H(t) = 1 - e^{-\lambda t}$, $t \geq 0$, then the moments of H are

$$\mu_j(H) = j!\lambda^{-j}, \qquad j = 1, 2, \cdots \qquad (3.2)$$

and the coefficient of variation $c \equiv c(H)$, defined as the ratio of the standard deviation to the mean, is always one.

Even when $c \neq 1$, it is convenient to use exponential building blocks; see Morse, and Kuehn [1979a]. There are two cases, depending on whether the coefficient of variation c is less than one or greater than one. A sum of independent exponential random variables always has a coefficient of variation less than one, while a mixture of exponential distributions has a coefficient of variation greater than one (see the proposition in Section 3.1 here or p. 142 of Kleinrock for a proof of this last property). Hence, we propose using a sum of independent exponentials when $c \leq 1$ and a mixture of exponentials when $c \geq 1$. Since we usually work with only two or three moments, it suffices to consider relatively simple special cases. When $c = 1$, this approach forces us to use an exponential distribution.

3.1. High Variability: Mixtures of Exponentials

When $c \geq 1$, we let the interval distribution be a mixture of exponential distributions. A mixture of m exponential distributions, called a hyperexponential distribution and denoted by the symbol H_m, has a density

$$h(x) = \textstyle\sum_{i=1}^{m} p_i\lambda_i e^{-\lambda_i x}, \qquad x \geq 0 \qquad (3.3)$$

and jth moment

$$\mu_j = \int_0^\infty x^j h(x)dx = j! \sum_{i=1}^{m} p_i \lambda_i^{-j}, \qquad j \geq 1, \tag{3.4}$$

where p_i, $\lambda_i \geq 0$ for each i and $p_1 + \cdots + p_m = 1$. For simplicity, we usually restrict attention to the case $m = 2$. It should be noted that the hyperexponential density is strictly decreasing, which might not be desirable for an approximation in some applications.

We now discuss how to go from the moments in (3.4) back to the parameters $\{p_i, \lambda_i\}$ in (3.3). First, however, we would like to know whether there is any hyperexponential distribution with moments equal to specified numbers. Obviously the general consistency check in (3.1) is necessary but not sufficient for the hyperexponential fit. Fortunately, it is possible to check whether m numbers $\mu_1, \cdots, \mu_m$ are the moments of some hyperexponential distribution by a simple modification of the classical criteria in (3.1). Let $[x]$ be the greatest integer less than or equal to x.

PROPOSITION. *The numbers $\mu_1, \cdots, \mu_m$ are the moments of a hyperexponential distribution if and only if $\mu_1, \mu_2/2!, \cdots, \mu_m/m!$ satisfy the classical criteria* (*in* (3.1) *for $m \leq 5$*), *in which case $[m/2] + 1$ exponentials will do.*

Proof. The classical criteria are necessary and sufficient for there to exist a probability distribution concentrated on $[m/2] + 1$ points; see Karlin and Studden. Make the correspondence with (3.3) by letting probabilities p_i be attached to the m points λ_i^{-1}. Note that the jth moment of this distribution with finite support is $\sum_{i=1}^{m} p_i/\lambda_i^j$, while the jth moment of an H_m distribution with these parameters is $j!\sum p_i/\lambda_i^j$.

As a consequence of the proposition, we see that there is an H_2 distribution with μ_1 and μ_2 as the first two moments if and only if $\mu_1 \geq 0$ and $c^2 = \mu_2\mu_1^{-2} - 1 \geq 1$. If μ_3 is given too, then an H_2 distribution exists with these three moments if and only if in addition to these conditions $\mu_3\mu_1 \geq 1.5\ \mu_2^2$. If μ_3 turns out to be too small when attempting an H_2-fit, one procedure is to replace μ_3 by something slightly larger than $1.5\ \mu_2^2/\mu_1$.

For the parameter fit, we work with H_2 distributions. Since there are three parameters, we can fit to three moments; see Whitt [1979], Appendix 3:

$$\lambda_i^{-1} = \{(x + 1.5y^2 + 3\mu_1^2 y) \pm \sqrt{(x + 1.5y^2 - 3\mu_1^2 y)^2 + 18\mu_1^2 y^3}\}/(6\mu_1 y) \geq 0, \tag{3.5}$$

and

$$p_1 = (\mu_1 - \lambda_2^{-1})/(\lambda_1^{-1} - \lambda_2^{-1}) \geq 0, \qquad p_2 = 1 - p_1 \geq 0, \tag{3.6}$$

where $x = \mu_1\mu_3 - 1.5\mu_2^2$ and $y = \mu_2 - 2\mu_1^2$. Note that $x, y \geq 0$ by the Proposition.

It is much easier to obtain a two-parameter fit. For example, suppose we assume balanced means as in Morse, and Kuehn [1979a], i.e., $p_1\lambda_1^{-1} = p_2\lambda_2^{-1}$. Then

$$p_i = [1 \pm \sqrt{(c^2 - 1)/(c^2 + 1)}]/2$$

and

$$\lambda_i = 2p_i\mu_1^{-1}, \; c^2 = (\mu_2 - \mu_1^2)/\mu_1^2. \tag{3.7}$$

3.2. Low Variability: Sum of Exponentials

When the coefficient of variation is less than one, we let the interval distribution be the distribution of a sum of independent exponential random variables. The sum of m independent exponential variables each with mean λ^{-1} is an E_m (Erlang) distribution with mean m/λ and coefficient of variation $c = 1/\sqrt{m}$. Since an Erlang distribution can only have a coefficient of variation of this special form, in order to construct distributions with any coefficient of variation less than one we need to make further modifications. We consider sums of independent exponential random variables with different means. Let $M_1(\lambda_1), \cdots, M_m(\lambda_m)$ be m independent exponential random variables with means $\lambda_1^{-1}, \cdots, \lambda_m^{-1}$, and let the sum be

$$X_m \equiv X_m(\lambda_1, \cdots, \lambda_m) = M_1(\lambda_1) + \cdots + M_m(\lambda_m). \tag{3.8}$$

The mean and variance of X_m are

$$\mu = EX_m = \textstyle\sum_{i=1}^m \lambda_i^{-1} \quad \text{and} \quad \sigma^2 = \operatorname{Var} X_m = \textstyle\sum_{i=1}^m \lambda_i^{-2}. \tag{3.9}$$

The coefficient of variation of $X_m(x\lambda_1, \cdots, x\lambda_m)$ is independent of the scalar x and varies between $m^{-1/2}$ and 1, with the minimum attained when all means are equal and the maximum approached when one mean dominates all the others.

A special case of interest is the sum of two independent exponential variables. This class of distributions can produce any coefficient of variation between $1/\sqrt{2}$ and 1. Given the mean μ and coefficient of variation c, the parameters λ_1 and λ_2 can be obtained by solving the two equations in (3.9):

$$\lambda_i^{-1} = (\mu/2)(1 \pm \sqrt{2c^2 - 1}). \tag{3.10}$$

In order to achieve any coefficient of variation between zero and one, we can use the sum of an exponential random variable and a constant (the special case of $k = \infty$ in E_k). We call this distribution the shifted exponential distribution, and denote it by M^d; it has density

$$f(x) = \lambda e^{-\lambda(x-d)}, \qquad x \geq d, \tag{3.11}$$

where λ^{-1} is the mean of the exponential variable and d is the constant. The two parameters λ and d are related to the mean μ and the variance σ^2 by

$$\begin{aligned} &\mu = \lambda^{-1} + d, \qquad \sigma^2 = \mu^2 c^2 = \lambda^{-2}, \\ &\lambda = \sigma^{-1}, \qquad d = \mu - \lambda^{-1}. \end{aligned} \tag{3.12}$$

4. MERGING STREAMS: SUPERPOSITION

In this section we apply the general methods introduced in Section 2 to approximate the superposition or sum of n independent point processes by a renewal process. If $\{N_i(t)\}$, $i = 1, \cdots, n$, are n counting processes, then the superposition counting process $\{N(t)\}$ is defined as

$$N(t) = N_1(t) + \cdots + N_n(t), \qquad t \geq 0. \tag{4.1}$$

Although the component processes need not be renewal processes, we think of them as renewal processes because it is always possible to first approximate each component process by a renewal process.

It is well known that the superposition of two independent renewal processes is itself a renewal process if and only if all three processes are Poisson; see Section 2 of Çinlar. It is significant that the procedures here satisfy the obvious consistency check: they all produce the correct Poisson superposition process when the component processes are all Poisson. If one or more of the component processes is not Poisson, then the superposition process not only fails to be Poisson but is not a renewal process either. Of course, this is the reason we are interested in approximations. It is possible to give more detailed exact expressions (see Cherry, Disney [1975], Lawrence, and other papers mentioned in Section 1), but we want simple expressions.

It is also well known that superpositions of independent equilibrium renewal counting processes (and more general stationary point processes) converge to a Poisson process as the number of component processes gets large and the individual processes get sparse with the total rate fixed; see Section 3 of Çinlar. This means that a Poisson process is often a good approximation for a superposition process if many processes are being superposed (see Albin [1980b]). Our interest is in the case in which relatively few processes are being superposed; then approximations are needed.

4.1. The Stationary-Interval Method

Consider n independent stationary counting processes $\{N_{si}(t)\}$, $i = 1, \cdots, n$. We are interested in approximating the stationary superposition process $\{N_s(t)\}$, defined by

$$N_s(t) = N_{s1}(t) + \cdots + N_{sn}(t), \qquad t \geq 0. \tag{4.2}$$

To apply the stationary-interval method, we need to know the moments μ_i of the c.d.f. F of the stationary interval associated with $\{N_s(t)\}$ (using the Palm correspondence and (1.6)).

Let the stationary interval in the ith component process have c.d.f. F_i with jth moment μ_{ij} and let the intensity of the ith counting process be $\lambda_i = \mu_{i1}^{-1}$. Then the c.d.f. F and its mean μ satisfy

$$\mu^{-1} = \lambda = \lambda_1 + \cdots + \lambda_n \tag{4.3}$$

and

$$1 - F(x) = \sum_{i=1}^{n} (\lambda_i/\lambda)[1 - F_i(x)] \prod_{j, j \neq i} \lambda_j \int_x^\infty [1 - F_j(s)]ds. \tag{4.4}$$

Equation (4.4) follows from the Palm theory (see Proposition 10 of Jagers), but it also can be established in other ways. (See Whitt [1979], Appendix 4, for more discussion.)

If I is a random variable with c.d.f. F (the stationary interval), then for $k \geq 2$ (see Whitt [1979], Appendix 5)

$$\begin{aligned} EI^k &= \int_0^\infty kx^{k-1}[1 - F(x)]dx \\ &= k(k-1)\lambda^{-1} \prod_{i=1}^{n} \lambda_i \int_0^\infty \left\{ x^{k-2} \prod_{i=1}^{n} \int_x^\infty [1 - F_i(s)]ds \right\} dx. \end{aligned} \tag{4.5}$$

In order to calculate the second moment (and thus the variance and coefficient of variation), we must perform the integration in (4.5) for $k = 2$. This can always be done in any specific instance, either analytically or by numerical integration. However, following Kuehn [1979a], we suggest avoiding the integration by first approximating the c.d.f.'s F_i. In this scheme each c.d.f. F_i is assumed to be either hyperexponential (H_2) or shifted exponential (M^d); see Sections 3.1 and 3.2. This can be achieved by calculating the moments of F_i and fitting one of these distributions to these moments. Moreover, only two processes are superposed at a time. Hence, it suffices to calculate the integral in (4.5) only for three cases: when F_1 and F_2 are both H_2, both M^d, or one of each. Here are the results

(Kuehn [1979a]) for the three cases:

(1) If F_1 is $H_2(\lambda_1,\lambda_2; p_1,p_2)$ and F_2 is $H_2(\eta_1,\eta_2; q_1,q_2)$, then

$$EI^2 = 2/(\mu_{11} + \mu_{21}) \textstyle\sum_{i=1}^{2} \sum_{j=1}^{2} p_i q_j/(\lambda_i \eta_j(\lambda_i + \eta_j)), \tag{4.6}$$

where $\mu_{11} = \sum_{i=1}^{2} (p_i/\lambda_i)$. If F_1 has balanced means as in (3.7) and $F_1 = F_2$, then we have the simple relation between the coefficients of variation:

$$c^2(F) - 1 = (c^2(F_1) - 1)/2; \tag{4.7}$$

see Whitt [1979], Appendix 6. Obviously (4.7) reflects the convergence of the stationary superposition-interval distribution to the exponential distribution.

(2) If F_1 is $M^d(\lambda_1,d_1)$ and F_2 is $M^d(\lambda_2,d_2)$ where $d_1 \le d_2$, then

$$\begin{aligned} EI^2 = 2(\mu_{11} + \mu_{21})^{-1} \{&\mu_{11}\mu_{21}d_1 - d_1^2(\mu_{11} + \mu_{21})/2 + d_1^3/3 \\ &+ \mu_{21}\lambda_1^{-2}(1 - e^{-\lambda_1(d_2-d_1)}) + \lambda_1^{-3}[(1 + \lambda_1 d_2)e^{-\lambda_1(d_2-d_1)} \\ &- (1 + \lambda_1 d_1)] + \lambda_1^{-1}\lambda_2^{-1}(\lambda_1 + \lambda_2)^{-1}e^{-\lambda_1(d_2-d_1)}\}; \end{aligned} \tag{4.8}$$

see Whitt [1979], Appendix 7. Let I_n be the stationary interval associated with the superposition of n independent and identically distributed M^d renewal processes. Then

$$c^2(I_n) = 1 - (2/(n + 1))(1 - (1/(\lambda\mu)^{n+1})), \qquad n \ge 1; \tag{4.9}$$

see Whitt [1979], Appendix 8.

(3) If F_1 is $H_2(\lambda_1, \lambda_2; p_1, p_2)$ and F_2 is $M^d(\eta, q)$, then we obtain the following from (4.5):

$$\begin{aligned} EI^2 = {}& e^{-\lambda_1 d}p_1/(\lambda_1\eta(\lambda_1 + \eta)) + e^{-\lambda_2 d}p_2/(\lambda_2\eta(\lambda_2 + \eta)) - p_1\lambda_1^{-3} \\ &\cdot(1 - e^{-\lambda_1 d} - \lambda_1 d e^{-\lambda_1 d}) - p_2\lambda_2^{-3}(1 - e^{-\lambda_2 d} - \lambda_2 d e^{-\lambda_2 d}) \\ &+ ((1/\eta) + d)[p_1\lambda_1^{-2}(1 - e^{-\lambda_1 d}) + p_2\lambda_2^{-2}(1 - e^{-\lambda_2 d})]. \end{aligned} \tag{4.10}$$

To summarize, here are the steps for specifying the first two renewal-interval moments when approximating the superposition of two independent renewal processes using the stationary-interval method:

1. Begin with first two moments of the interval distributions F_i in each component process.
2. Compute the mean of F using (4.3).
3. Compute the second moment of F using (4.5) assuming that the c.d.f.'s F_i have been approximated by H_2 and M^d distributions. No actual distribution fitting need be done. Use formulas (4.6)–(4.10) depending on the coefficients of variation.
4. If desired, fit a distribution to the moments, using the methods in Section 3.

The procedure above is clear whenever two processes are being superposed, but there is some ambiguity when more than two are being superposed. The procedure then is to combine two at a time. Kuehn [1979a] suggests a recursive scheme in which the kth component process is added to the approximation of the sum of the first $k-1$ component processes for each $k > 2$, but we suggest a balanced scheme in which component processes are added and then the approximations are added. The recursive scheme has some advantage in ease of implementation, but the balanced scheme has fewer steps ($\lceil \log_2 n \rceil$ instead of $n-1$ where $\lceil x \rceil$ is the smallest integer greater than or equal to x). The different schemes and different orders give slightly different results; see p. 35 of Albin [1981a].

4.2. The Asymptotic Method

Since the superposition counting process is the sum of independent counting processes, as shown in (4.1), it is easier to work with the counting processes than the intervals or the associated partial sums. In particular, the jth cumulant of the superposition counting process is the sum of the jth cumulants of the n component processes:

$$\beta_j(N(t)) = \beta_j(N_1(t)) + \cdots + \beta_j(N_n(t)), \qquad t \geq 0. \tag{4.11}$$

Our basic assumption, as in (2.5), is that

$$\lim_{t\to\infty} \beta_j(N_i(t))/t = \gamma_{ij} \tag{4.12}$$

for each pair (i, j), $i = 1, \cdots, n$ and $j = 1, \cdots, m$. (We do not need stationarity here.) Then (4.11) and (4.12) imply that

$$\lim_{t\to\infty} \beta_j(N(t))/t = \gamma_j = \gamma_{1j} + \cdots + \gamma_{nj}. \tag{4.13}$$

To get the moments of the interval-distribution in the approximating renewal process, simply apply (2.9). Of course this step is an approximation unless the superposition process is a renewal process.

This procedure leads to very simple formulas. Let μ_{ij} be obtained from γ_{ij} in (4.12) by (2.9); let $\lambda_i = \mu_{i1}^{-1} = \gamma_{i1}$; and let $c_i^2 = (\mu_{i2} - \mu_{i1})^2/\mu_{i1}^2 = \mu_{i1}\gamma_{i2}$. Then

$$\lambda = \mu^{-1} = \sum_{i=1}^{n} \lambda_i \quad \text{and} \quad c^2 = \sum_{i=1}^{n} (\lambda_i/\lambda) c_i^2. \tag{4.14}$$

Similar formulas also hold for appropriate parameters related to higher moments. In particular, if $\alpha_j = \gamma_j/\mu_1$ where γ_j is given in (2.7), then

$$\alpha_j = \sum_{i=1}^{n} (\lambda_i/\lambda) \alpha_{ij}, \qquad j \geq 2, \tag{4.15}$$

in the superposition approximation.

If the ith component process is a renewal process initially characterized by the first m moments of the interval distribution F_i, then we first use

(2.7) to obtain the corresponding time-average limit cumulants of the counting process $\{N_i(t)\}$ needed in (4.13).

To summarize, here are the steps for approximating the superposition of n independent renewal processes by a renewal process using the asymptotic method:

1. Begin with the first m moments of the interval distribution F_i in each component renewal process.
2. Use Smith's relation (2.7) to calculate the time-average limit cumulants γ_{ij} for each process.
3. Add these time-average limit cumulants to obtain the time-average limit cumulants for the superposition process; see (4.13).
4. Use Smith's inverse relation (2.9) to calculate the moments μ_j of the interval-distribution in the approximating renewal process.
5. If desired, fit a distribution to the moments, using the methods in Section 3.

Note that if we start with independent renewal processes, the Step 4 is the first approximation. It is not exact because Smith's relations (2.7)–(2.9) apply only to renewal processes. Step 2 is exact when the component processes are renewal processes and Step 3 is exact when the component processes are independent.

A variant of this procedure of going back and forth between (2.7) and (2.9) with only two moments appears on the bottom of page 73 in Cox. The two-moment version of the asymptotic method has also been used in a two-echelon inventory application by Heyman [1978]; see Section 4.1 of Heyman [1975].

It is interesting to see how the asymptotic method compares with the stationary-interval method for approximating superposition processes. For the superposition of two independent renewal processes, the approximating coefficients of variation in 100 cases are displayed in Whitt [1979] Appendix 10. The two procedures tend to be similar when both coefficients of variation are less than or equal to one (or slightly above one), but the two procedures are quite different when one or more coefficient of variation is much bigger than one.

A special case of considerable interest arises when the n renewal processes being superposed are identically distributed. From (4.15), it is clear that the parameter α_j in the superposition process has the same value as the corresponding parameter α_{ij} in each component process; see Whitt [1979], Example 4.1, for more discussion.

If we use the asymptotic method to approximate the superposition of a large number of independent stationary point processes, then we are confronted with two different limiting operations. With the asymptotic method we let time go to infinity, but we can also let the number of processes being superposed go to infinity. Then, the superposition process converges to a Poisson process. It is important to note that the order of the two limits ($n \to \infty$ and $t \to \infty$) makes a difference; see Whitt [1979], Section 4.3.

Hence, it is clearly important to consider modifications of the asymptotic method for superposition processes as the number of processes increases. To understand what adjustments should be reasonable, it is appropriate to consider the rate of convergence to the Poisson process as $n \to \infty$. Fortunately, considerable work has been done on this problem; see Section 6 of Çinlar, and Section 6 of Dudley [1972]. It has been shown that the rate of convergence is of order n^{-1}. If the component processes are not identically distributed and may have different intensities, then the rate of convergence is of order $\sum_{i=1}^{n} (\lambda_{ni}/\lambda_n)^2$, where λ_{ni} is the intensity of the ith component process in the nth system of n processes and $\lambda_n = \lambda_{n1} + \cdots + \lambda_{nn}$.

For example, if $c_a{}^2$ is the squared coefficient of variation obtained by the asymptotic method, as a refinement it would be natural to replace it with $\hat{c}^2$ where

$$\hat{c}^2 - 1 = (c_a{}^2 - 1) \sum_{i=1}^{n} (\lambda_i/\lambda)^2. \tag{4.16}$$

5. THE PROOF OF THE PUDDING: THE $\sum G_i/G/1$ QUEUE

As indicated in the introduction, these approximation procedures have been applied to a single-server queueing model with infinite waiting room and FIFO queue discipline in which the arrival process is a superposition of independent renewal processes; see Whitt [1979], Section 6, and Albin [1980, 1981a, b, 1982]. When the service-time distribution is exponential, standard formulas are used to solve the approximating $GI/M/1$ queue after the superposition arrival process is replaced by the approximating renewal process; otherwise the approximation of $GI/G/1$ due to Krämer and Langenbach-Belz [1976] is applied. The approximations for the mean queue length and other characteristics such as the probability of delay have been compared with the actual behavior as estimated from computer simulation. For the superposition of two renewal processes, the asymptotic method performs somewhat better, especially for higher traffic intensities. However, over a representative class of queues with from 2-to-16 component renewal processes, neither the asymptotic method nor the stationary-interval method performs well (20–30% error). However, when $c > 1$ ($c < 1$) the asymptotic method tends to overestimate (underestimate) the mean queue length while the stationary-interval method tends to underestimate (overestimate) it. This of course suggests using appropriate convex combinations. Such hybrid procedures have been developed by Albin and perform very well (3% error); they will be discussed in Part II, Albin [1981b].

Of course, the $\sum G_i/G/1$ queue is just one possible application of these approximations. The approach should be useful in other contexts, e.g., in inventory and reliability.

ACKNOWLEDGMENTS

Supporting numerical comparisons were done with programs written and run by Susan Albin, Linda Dolny and Paul Kuehn. I received much help from colleagues, especially Susan Albin, Paul Burke, Shlomo Halfin, Daniel Heyman, Jack Holtzman, Paul Kuehn, Moshe Segal, and Richard Serfozo.

REFERENCES

ALBIN, S. L. 1980. Approximating Superposition Arrival Processes of Queues, Bell Laboratories Holmdel, New Jersey.

ALBIN, S. 1981a. Approximating Queues with Superposition Arrival Processes, Ph.D. dissertation, Columbia University.

ALBIN, S. 1981b. Approximating a Point Process by a Renewal Process, II: Superposition Arrival Processes of Queues. Bell Laboratories, Holmdel, N.J.

ALBIN, S. L. 1982. On Poisson Approximations for Superposition Arrival Processes in Queues forthcoming in *Mgmt. Sci.*).

AL-KHAYYAL, F. A., AND D. GROSS. 1977. On Approximating and Bounding *GI/M/c* Queues. *TIMS Stud. Mgmt. Sci.* **7**, 233–245.

BOROVKOV, A. A. 1976. *Stochastic Processes in Queueing Theory.* Springer-Verlag, New York.

BUX, W. AND U. HERZOG. 1977. The Phase Concept: Approximation of Measured Data and Performance Analysis. In *Computer Performance*, pp. 23–38, K. M. Chandy and M. Reiser (eds.). North-Holland, New York.

CHANDY, K. M., AND C. H. SAUER. 1978. Approximate Methods for Analyzing Queueing Network Models of Computer Systems. *Comput. Surv.* **10,** 281–317.

CHERRY, W. P. 1972. The Superposition of Two Independent Markov Renewal Processes, Ph.D. dissertation and Technical Report No. 72-10, Department of Industrial Engineering and Operations Research, the University of Michigan.

ÇINLAR, E. 1972. Superposition of Point Processes. In *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, pp. 549-606, P. A. W. Lewis (ed.). John Wiley & Sons, New York.

COOPER, R. B. 1980. *Introduction to Queueing Theory,* 2nd ed. North Holland, New York.

COX, D. R. 1972. *Renewal Theory.* Methuen, London.

COX, D. R., AND P. A. W. LEWIS. 1966. *The Statistical Analysis of Series of Events.* Methuen, London.

DALEY, D. J., AND D. VERE-JONES. 1972. A Summary of the Theory of Point Processes. In *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, pp. 299–383, P. A. W. Lewis (ed.). John Wiley & Sons, New York.

DISNEY, R. L. 1975. Random Flow in Queueing Networks. *A.I.I.E. Transact.* **7,** 268–288.

DUDLEY, R. M. 1972. Speeds of Metric Probability Convergence. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **22,** 323–332.

ECKBERG, A. E. 1976. A Generalization of Peakedness to Arbitrary Arrival Processes and Service Time Distributions, Bell Laboratories.

ECKBERG, A. E. 1977. Sharp Bounds on Laplace-Stieltjes Transforms, with Applications to Various Queueing Problems. *Math. Opns. Res.* **2,** 135–142.

FELLER, W. 1971. *An Introduction to Probability Theory and Its Applications,* Vol. II, Ed. 2. John Wiley & Sons, New York.

FREEMAN, A. H. 1976. The Accuracy of Overflow Traffic Models. In *Proceedings of 8th International Teletraffic Congress,* Melbourne, Vol. 2, pp. 424.1–424.6.

FUJISAWA, T. 1976. Effects of Multiplicity in Arrival Processes on Queue Lengths and Waiting Times. *J. Opns. Res. Soc. Jpn.* **19,** 125–136.

GNEDENKO, B. V., AND A. N. KOLMOGOROV. 1968. *Limit Distributions for Sums of Independent Random Variables,* Ed. 2. Addison-Wesley, Reading, Mass.

GREENBERG, I. 1980. An Approximation for the Waiting Time Distribution in Single Server Queues. *Naval Res. Logist. Quart.* **27,** 223–230.

HALACHMI, B., AND W. R. FRANTA. 1978. A Diffusion Approximation to the Multi-Server Queue. *Mgmt. Sci.* **24,** 522–529.

HALFIN, S., AND W. WHITT. 1981. Heavy-Traffic Limits for Queues with Many Exponential Servers. *Opns. Res.* **29,** 567–588.

HASKELL, R. D. 1974. Semi-Markov Generated Point Processes and the Superposition of Finite Numbers of Independent Erlang and Hyperexponential Renewal Processes, Ph.D. dissertation, Naval Postgraduate School, Monterey.

HEFFES, H. 1973. Analysis of First-Come First-Served Queueing Systems with Peaked Inputs. *Bell System Tech. J.* **52,** 1215–1228.

HEYMAN, D. P. 1975. A Two-Echelon Inventory Model with Stock Returns, Bell Laboratories.

HEYMAN, D. P. 1978. Return Policies for an Inventory System with Positive and Negative Demands. *Naval Res. Logist. Quart.* **25,** 581–596.

HOLTZMAN, J. M. 1973. The Accuracy of the Equivalent Random Method with Renewal Inputs. *Bell System Tech. J.* **52,** 1673–1679.

JAGERS, P. 1973. On Palm Probabilities. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **26,** 17–32.

KARLIN, S., AND W. J. STUDDEN. 1966. *Tchebycheff Systems: With Applications in Analysis and Statistics.* John Wiley & Sons, New York.

KHINTCHINE, A. Y. 1960. *Mathematical Methods in Queueing Theory.* Griffin, London.

KLEINROCK, L. 1975. *Queueing Systems,* Vol. 1. John Wiley & Sons, New York.

KLINCEWICZ, J., AND W. WHITT. 1981. On the Quality of Two-Moment Approximations in Queueing Bell Laboratories, Holmdel, N.J.

KOSTEN, L. 1977. Parametrisation of Overflow Traffic in Telecommunication and Data handling, Department of Mathematics, Delft University of Technology. (Presented at 7th Conference on Stochastic Processes and Their Applications, Enschede.)

KRÄMER, W., AND M. LANGENBACH-BELZ. 1976. Approximate Formulae for the Delay in the Queueing System $GI/G/1$. In *Proceedings of 8th International Teletraffic Congress,* Melbourne, No. 235, pp. 1–8.

KUCZURA, A. 1972. Queues with Mixed Renewal and Poisson Inputs. *Bell System Tech. J.* **51,** 1305–1326.

KUCZURA, A. 1973a. The Interrupted Poisson Process as an Overflow Process. *Bell System Tech. J.* **52,** 437–448.

KUCZURA, A. 1973b. Piecewise Markov Processes. *SIAM J. Appl. Math.* **24,** 169–181.

KUCZURA, A. 1973c. Loss Systems with Mixed Renewal and Poisson Inputs. *Opns. Res.* **21,** 787–795.

KUEHN, P. J. 1979a. Approximate Analysis of General Queueing Networks by Decomposition. *IEEE Trans. Comm.* **COM-27,** 113–126.

KUEHN, P. J. 1979b. Analysis of Switching System Control Structures by Decomposition. In *Proceedings of 9th International Teletraffic Congress,* Spain.

LAWRENCE, A. J. 1973. Dependency of Intervals between Events in Superposition Processes. *J. Roy. Statist. Soc., Ser. B* **35,** 306–315.

LAZOWSKA, E. D. 1977. The Use of Percentiles in Modeling CPU Service Time Distributions. In *Computer Performance,* pp. 53–66. K. M. Chandy and M. Reiser (eds.). North-Holland, New York.

MARCHAL, W. G., AND C. M. HARRIS. 1976. A Modified Erlang Approach to Approximating $GI/G/1$ Queues. *J. Appl. Prob.* **13,** 118–126.

MORSE, P. M. 1958. *Queues, Inventories and Maintenance.* John Wiley & Sons, New York.

MURTHY, V. K. 1974. *The General Point Process.* Addison-Wesley, Reading, Mass.

NEUTS, M. F. 1978. Markov Chains with Applications in Queueing Theory, Which Have a Matrix-Geometric Invariant Probability Vector. *Adv. Appl. Prob.* **10,** 185–212.

NEUTS, M. F. 1979. A Versatile Markovian Point Process. *J. Appl. Prob.* **16,** 764–779.

PORT, S. C., AND C. J. STONE. 1977. Spacing Distribution Associated with a Stationary Random Measure on the Real Line. *Ann. Prob.* **3,** 387–394.

RATH, J. H., AND D. SHENG. 1979. Approximations for Overflows from Queues with a Finite Waiting Room. *Opns. Res.* **27,** 1208–1216.

RODRIGUEZ, R. N. 1976. Superposition and Approximation in Renewal Theory, Ph.D. dissertation, Department of Statistics, University of North Carolina.

ROLSKI, T. 1978. A Rate-Conservative Principle for Stationary Piecewise Markov Processes. *Adv. Appl. Prob.* **10,** 392–410.

ROSS, S. M. 1970. *Applied Probability Models with Optimization Applications.* Holden-Day, San Francisco.

SAHIN, I. 1971. Equilibrium Behavior of a Stochastic System with Secondary Input. *J. Appl. Prob.* **8,** 252–260.

SAUER, C. H., AND K. M. CHANDY. 1975. Approximate Analysis of Central Server Models. *IBM J. Res. Dev.* **19,** 301–313.

SEVCIK, K. C., A. I. LEVY, S. K. TRIPATHI AND J. L. ZAHORJAN. 1977. Improving Aproximations of Aggregated Queueing Network Subsystems. In *Computer Performance,* pp. 1–22, K. M. Chandy and M. Reiser (eds.). North-Holland, Amsterdam.

SMITH, W. L. 1959. On the Cumulants of Renewal Processes. *Biometrika* **46,** 1–29.

WHITT, W. 1974. Heavy Traffic Limit Theorems for Queues: A Survey. In

Mathematical Methods in Queueing Theory (Lecture Notes in Econ. and Math. Systems No. 98), pp. 307–350, A. B. Clarke (ed.). Springer-Verlag, New York.

WHITT, W. 1979. Approximating a Point Process by a Renewal Process: A General Framework, Bell Laboratories.

WILKINSON, R. I. 1956. Theories for Toll Traffic Engineering in the U.S.A. *Bell System Tech. J.* **35,** 421–514.

WILKINSON, R. I. 1970. Nonrandom Traffic Curves and Tables, Traffic Studies Center, Bell Laboratories.