IEOR 8100: Spring 2012, Professor Whitt

More on Single-Server Queues, Wednesday, February 1, 2012

1 Point Processes and Partial Sums: Inverse Processes

1.1 The Basic Inverse Relation

Point processes on the positive half line $[0, \infty)$ can be represented in three related ways: (i) in terms of $N(t)$, the number of points in the interval $[0, t]$, as a function of $t \geq 0$, (ii) in terms of S_n, the location of point n, as a function of $n \geq 1$, with $S_0 \equiv 0$, and (iii) in terms of X_k, the interval between point $k-1$ and point k, for $k \geq 1$. In terms of the intervals X_k, we write

$$S_k \equiv X_1 + \cdots + X_k, \quad k \geq 1, \quad \text{and} \quad N(t) \equiv \max\{k \geq 0 : S_k \leq t\}, \quad t \geq 0. \tag{1}$$

In this setting, the two stochastic processes $\{N(t) : t \geq 0\}$ and $\{S_n : n \geq 0\}$ are inverse processes. This is easily seen by looking at a sample path of a counting process, such as a Poisson process. In such a plot, we show, $N(t)$, the number of points in the interval $[0, t]$ as a function of t. The sample path is a nondecreasing nonnegative integer-valued function of time t. We represent time t on the horizontal "x" axis, and the values of $N(t)$ on the vertical "y" axis. If instead, we look at the plot regarding the y axis as the domain and the x axis as the range, we see a plot of the locations of the successive points, S_n. (We allow multiple points, but a common case is unit jumps in $N(t)$.)

1.2 CLT and FCLT Equivalence

As a consequence of the inverse relation, there is a CLT and FCLT equivalence the two stochastic processes $\{N(t) : t \geq 0\}$ and $\{S_n : n \geq 0\}$. See §§6.3, 7.3 and 13.6-13.8 of [9], to which we refer below.

(a) Preservation of convergence of the inverse function with centering, Theorem 13.7.1 (b). This is a preservation of convergence for a sequence of functions associated with the inverse map mapping a subset of the function space D into itself. Consider the case of continuous limit functions; see Lemma 13.7.1. Let e be the identity map in the functions space D, i.e., $e(t) \equiv t$, $t \geq 0$. Let D_u be the subspace of functions x in D that are unbounded above and satisfy $x(0) \geq 0$. For $x \in D_u$, let the inverse function be defined by

$$x^{-1}(t) \equiv \inf\{s \geq 0 : x(s) > t\}, \quad t \geq 0. \tag{2}$$

Let C be the subset of continuous functions in D. The following is a corollary of Theorem 13.7.1 and Lemma 13.7.1 of [9].

Theorem 1.1 *Suppose that $x_n \in D_u$ for all $n \geq 1$, $y \in C$ and $y(0) = 0$. Let c_n be positive numbers such that $c_n \to \infty$ as $n \to \infty$. If $c_n(x_n - e) \to y$ in D, then $c_n(x_n^{-1} - e) \to y$ in D.*

There is an equivalence in these two limits if x_n is also nondecreasing. That can be obtained by first applying the supremum function; see Theorem 13.4.2.

(b) CLT equivalence of counting processes and associated partial sums, Theorem 7.3.1. Let S_n be the n^{th} partial sum of nonnegative real-valued random variables X_n, with $S_0 \equiv 0$, and let $N(t)$ be the associated counting process, i.e.,

$$N(t) \equiv \max\{k \geq 0 : S_k \leq t\}, \quad t \geq 0. \tag{3}$$

Let $\Rightarrow$ denote convergence in distribution. The following is a corollary to Theorem 7.3.1 of [9].

Theorem 1.2 *Let $m > 0$ and $0 < p < 1$. Then $(S_n - mn)/n^p \Rightarrow L \quad in \quad \mathbb{R}$ as $n \to \infty$ if and only if $(N(t) - m^{-1}t)/t^p \Rightarrow -m^{-(1+p)}L \quad in \quad \mathbb{R}$ as $t \to \infty$.*

The standard case is $p = 1/2$, but other cases arise. Counting processes are not exactly inverse processes of partial sums in the sense of the map in (2), but the difference tends to be asymptotically negligible; see §13.8 of [9].

(c) FCLT equivalence of counting processes and associated partial sums, Theorem 7.3.2. We now assume that we have a double sequence of nonnegative random variables $X_{n,k}$ and let $S_{n,k} = X_{n,1} + \cdots + X_{n,k}$, $k \geq 1$, $n \geq 1$. For each n, let $N_n(t)$ be the associated counting process. Define the usual random elements of D:

$$\begin{aligned} \mathbf{S_n}(\mathbf{t}) &\equiv c_n^{-1}(S_{n,\lfloor nt \rfloor} - m_n nt), \\ \mathbf{N_n}(\mathbf{t}) &\equiv c_n^{-1}(N_n(nt) - m_n^{-1}nt), \quad t \geq 0. \end{aligned} \tag{4}$$

Theorem 1.3 *Suppose that $m_n \to m > 0$, $c_n \to \infty$ and $n/c_n \to \infty$ as $n \to \infty$. Suppose that $\mathbf{S}$ almost surely has continuous sample paths with $\mathbf{S}(0) = 0$. Then $\mathbf{S_n} \Rightarrow \mathbf{S}$ in D as $n \to \infty$ if and only if $\mathbf{N_n} \Rightarrow \mathbf{N}$ in D as $n \to \infty$, in which case $(\mathbf{S_n}, \mathbf{N_n}) \Rightarrow (\mathbf{S}, \mathbf{N})$ in D^2, where*

$$\mathbf{N}(t) = -m^{-1}\mathbf{S}(m^{-1}t), \quad t \geq 0. \tag{5}$$

2 Superpositions of Point Processes

2.1 Dependence in Superpositions of Non-Poisson Point Processes

Even if the intervals between points are mutually independent in each component process in a superposition process, the intervals between points are dependent in the superposition process, unless all processes are Poisson process.

Recall that a stationary renewal process is a delayed renewal process in which the distribution until the first point has the equilibrium excess distribution. Give that the other intervals are i.i.d. with cdf F with mean $m < \infty$, the equilibrium excess cdf is

$$F_e(t) \equiv \frac{1}{m}\int_0^t (1 - F(s))\, ds, \quad t \geq 0. \tag{6}$$

A stationary renewal process is a special case of a stationary point process. For a stationary point process, the counting process has stationary increments An increment is $N(t+s) - N(s)$. Stationary increments means that the distribution of $N(t+s) - N(s)$ is independent of s. More generally, any k-tuple of increments $(N(t_2) - N(t_1), \ldots, N(t_{2k}) - N(t_{2k-1}))$ for $t_1 < t_2 < \cdots < t_{2k}$ is independent of a time shift. That means that it has the same distribution as $(N(t_2 + h) - N(t_1 + h), \ldots, N(t_{2k} + h) - N(t_{2k-1} + h))$, where all times have been shifted by $h > 0$, for all $h > 0$.

In general, the rate of a superposition process is the sum of the component rates. The following theorem summarizes both positive and negative properties of superposition processes.

Theorem 2.1 (*superpositions of renewal processes*) *The superposition of n independent stationary (ordinary) renewal processes is itself a stationary or ordinary renewal process if and only if all the component processes and the superposition process are homogeneous Poisson processes.*

On the positive side, Theorem 2.1 states that the supperposition of independent Poisson processes is itself a Poisson process. On the other hand, if any of the component processes are non-Poisson renewal processes, either stationary or ordinary, then the superposition process is neither a stationary renewal process nor an ordinary renewal process (and thus also not a Poisson process). That means that the successive intervals between points in the superposition process necessarily have dependence. Thus, dependence often plays a role in superposition processes.

2.2 Fundamental Limits for Superposition Processes

There are two fundamental limit theorems for superposition processes, but the two iterated limits do not coincide; see §9.8 of [9]. A basic case involves the sum of n i.i.d. stationary renewal processes, i.e.,

$$N(t) \equiv N_1(t) + \cdots + N_n(t), \quad t \geq 0, \tag{7}$$

where the component stochastic processes $\{N_i(t) : t \geq 0\}$ are mutually independent stationary renewal processes. However, the setting can be greatly generalized. The component processes need not be i.i.d. and each individual component process need not be a renewal process.

(a) The Poisson limit as $n \to \infty$, with the spacing between points growing

(b) The CLT and FCLT as $t \to \infty$

(c) The double limit as $n \to \infty$ and $t \to \infty$.

3 Heavy-Traffic Limits for the Single-Server Queue

See Chapters 5 and 9 in [9].

(a) the one-dimensional reflection map.

(b) The importance of the joint CLT of the arrival and service processes

(c) Heavy-traffic limit for a queue with a superposition arrival process. See §9.4 of [9].

4 Approximating a Point Process by a Renewal Process

See [7] and §§9.4, 9.6, 9.8 and 9.9 in [9].

(a) The asymptotic method (the CLT and heavy-traffic perspective))

(b) The stationary interval method (the light traffic perspective)

5 The Impact of Dependence upon Queue Performance

Even though offered traffic in telephone networks can often be regarded as Poisson, dependence played an important role in teletraffic networks in the period 1950-1970, because the networks allowed overflows from one congested resource to other resources with available capacity. Overflow process are more bursty than original offered traffic, occurring only when the initial resource is congested. The importance of dependence in the traffic of communication networks grew in the period 1980-2000 when the networks evolved from circuit-switched networks to packet networks; see [1, 2, 3, 4, 5, 6], plus citations and references to these papers.

(a) Indices of Dispersion

(b) The Important Role of Measurements.

(c) The Important Role of time scales when considering the performance at a queue (e.g., switch or router in a communication network).

6 Packetized Voice, [6]

Starting around 1960, communication networks started to shift from dedicated analog circuits to digital packet-switched communication, in which messages are broken up into smaller packets. The individual packets are transmitted independently and then reassembled into the messages at the destination. Given the technology circa 1985, when converting voice to packets, the speech traffic from each voice source was divided into alternating talk spurts and silence periods. Packets were generated during the talk spurts, but to avoid useless packets, nothing at all was transmitted during the idle periods (which could be detected).

The specific model for one voice source is depicted in Figures 1 and 2 of [6]. The times between the beginning of successive talk spurts are approximately i.i.d. Each talk spurt contains a random number N of packets approximately evenly spaced, T ms apart. The length of each idle period I is approximately exponentially distributed. Measurements indicated that N was approximately geometrically distributed with mean 22, while $T = 16$ ms. The mean idle period was $EI = 650ms$. Thus the interval Y between successive talk spurts has mean

$$E[Y] = E[N]T + E[I] = 22(16) + 365 = 352 + 650 = 1002ms \tag{8}$$

Thus the arrival rate of packets is

$$\lambda \equiv \frac{E[N]}{E[Y]} = \frac{E[N]}{E[N]T + E[I]} = \frac{22}{1002} \approx 0.022 \quad \text{per ms} \tag{9}$$

Since N is approximately geometric and X is approximately exponential, the packet arrival process can be modeled as a renewal process. The times between successive renewals is a mixture, assuming the deterministic value T with probability $21/22$, and being an exponential random variable with mean 650 with probability $1/22$. The interarrival time, say U, has mean

$$E[U] = \lambda^{-1} = \frac{1002}{22} \approx 45.5 \quad ms. \tag{10}$$

The squared coefficient of variation (scv) of an interarrival time is

$$c_a^2 \equiv \frac{Var(U)}{E[U]^2} = 18.1. \tag{11}$$

The actual traffic consists of a superposition of separate packet streams from i.i.d. voice sources plus also the possibility of data traffic, which is regarded as Poisson in [6]. As noted in §2, the superposition of independent voice traffic streams is not a renewal process; it has complicated dependence among successive interarrival times.

7 Traffic Model for an X.25 Packet Network, [1] and §9.6 in [9]

There was a period in the development of packet networks, in which the network carried several different kinds of traffic, each with its own character. The packet sizes were allowed to differ. The packet sizes could even vary for a single source.

7.1 The Batch Poisson Model

This is the $\sum(M^{B_i}/G_i)/1$ model defined in §III. There are k customer classes. Packets arrive in batches. Batches of type i arrive at a rate λp_i, where $p_1 + \cdots + p_k = 1$, so that λ is the total arrival rate of batches. As a first simple approximation, we assume that each arrival process of batches is a Poisson process, and so has scv $c^2 = 1$. The number of packets for class i per batch has mean m_i and SCV $c_{b,i}^2$. The packet lengths determine the service time, but the packet lengths are random. The service time of a class-i packet has mean τ_i and scv $c_{s,i}^2$. Overall this model has parameter vector

$$(\lambda, [p_i, m_i, c_{b,i}^2, \tau_i, c_{s,i}^2]; 1 \leq i \leq k). \tag{12}$$

If the packet sizes from a source were of constant size, then we would have $c_{s,i}^2 = 0$. That was a possibility, but it did not always happen. Moreover, even if $c_{s,i}^2 = 0$, τ_i could vary by source i, so that the traffic would be complicated.

7.2 A More Complex Model with Spacing within the Batches

This more complex model is described in §VI.C. It adds a random time L_i between the arrival of batches of type i. The interval between successive batch arrivals of type i is L_i, where

$$L^i = (\sum_{j=1}^{B_i} T_j^i) + I_i, \tag{13}$$

where T_j^i are i.i.d. spacings between the arrivals of packets within the batch, all of which are independent of the following idle period I_i The random variables T_j^i are partially characterized by their means $E[T^i]$ and scv's $c_{T^i}^2$, while the random variable I_i is partially characterized by its mean $E[I^i]$ and scv $c_{I^i}^2$. We have

$$\lambda p_i = \frac{1}{E[L^i]}, \quad \text{where} \quad E[L^i] = m_i E[T^i] + E[I^i]. \tag{14}$$

We now have the parameter vector

$$(E[T^i], c_{T^i}^2, E[I^i], c_{I^i}^2, m_i, c_{b,i}^2, \tau_i, c_{s,i}^2; 1 \leq i \leq k). \tag{15}$$

where λp_i is defined as in (14) and

$$\lambda \equiv \sum_{i=1}^{k} \lambda p_i = \sum_{i=1}^{k} [1/(m_i E[T^i] + E[I^i])]. \tag{16}$$

7.3 Performance at the Multiplexer Queue, see §§9.4 and 9.8 of [9]

In very light traffic (with low traffic intensity at the queue), the superposition arrival process with a large number of component streams would be approximately a Poisson process. However, very different behavior would occur at higher loads. It is significant that the heavy-traffic limit (as $\rho \uparrow 1$) for the steady-state workload at the multiplexer queue can be computed in terms of the model parameters. The mean steady-state workload is typically much greater than would be the case for a Poisson arrival processes. The indices of dispersion can be used to determine what should happen at typical intermediate loads, which are neither extremely light not extremely heavy.

References

[1] K. W. Fendick, V. R. Saksena, W. Whitt, Dependence in Packet Queues. *IEEE Transactions on Communications*, vol. 37, No. 11, 1989, pp. 1173-1183.

[2] K. W. Fendick, V. R. Saksena, W. Whitt, Investigating Dependence in Packet Queues with the Index of Dispersion for Work. *IEEE Transactions on Communications*, vol. 39, No. 8, 1991, pp. 1231-1244

[3] K. W. Fendick, W. Whitt, Measurements and Approximations to Describe the Offered Traffic and Predict the Average Workload in a Single-Server Queue. *Proceedings of the IEEE*, vol. 77, No. 1, 1989, pp. 171-194,

[4] Leland, W. E., M. S. Taqqu, W. Willinger, D. Wilson. On the self-similar nature of Ethernet traffic (extended version) *IEEE/ACM Transactions on Networking*, vol. 2, 1994, 1–15.

[5] Paxson, V., S. Floyd. Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, vol. 3, 1995, 226-224.

[6] K. Sriram, W. Whitt, Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data. *IEEE Journal on Selected Areas in Communications*, vol. SAC-4, No. 6, September 1986, pp. 833-846.

[7] W. Whitt, Approximating a Point Process by a Renewal Process: Two Basic Methods. *Operations Research*, vol. 30, No. 1, January-February 1982, pp. 125-147.

[8] W. Whitt, The Queueing Network Analyzer. *Bell System Technical Journal*, vol. 62, No. 9, November 1983, pp. 2779-2815.

[9] W. Whitt, *Stochastic-Process Limits*, Springer, New York, 2002. Available at: http://www.columbia.edu/ ∼ww2040/jumps.html