Top

Volume 11, Number 1, 131-140
June 2003

REPRINT

M.F. Neuts and J. Guo

On Simulating Arbitrary Events

Published by
Sociedad de Estadística e Investigación Operativa
Madrid, Spain

Top

Volume 11, Number 1
June 2003

Published by
Sociedad de Estadística e Investigación Operativa
Madrid, Spain

Sociedad de Estadística e Investigación Operativa
Top (2003) Vol. 11, No. 1, pp. 131–140

On Simulating Arbitrary Events

Marcel F. Neuts
Systems and Industrial Engineering, The University of Arizona
Tucson, AZ, 85721, USA.
email: marcel@mindspring.com
Jun Guo
Department of Electrical and Electronic Engineering,
The University of Melbourne, Victoria 3010, Australia
email: j.guo@ee.mu.oz.au

Abstract

In simulation studies, the distributional properties of random variables associated with *arbitrary* events in point processes, queues, and other stochastic models are to be understood as appropriate averages over long simulation runs. We caution against trying to generate arbitrary events explicitly by some randomized selection. Because of the likelihood of hidden selection biases, that easily results in significant errors. The point is illustrated by an example for which explicit formulas yield computational results that allow comparisons with the simulation estimates.

Key Words: Simulation, arbitrary events, selection bias.

AMS subject classification: 68U20.

1 Introduction

Since the familiar paradoxes of Bertrand (1907), we know that in probability, the word *arbitrary* must be given a precise meaning. In simulations of service systems, one is often interested in waiting times experienced by a typical job. The issue again arises what is meant by an *arbitrary*, or typical job. On several occasions, we have heard proposals to tag the arbitrary job by some *random selection* made in the course of the simulation. The purpose of this paper is to give an informal discussion of the meaning of *arbitrary* and to caution against ad hoc *random selections.* The latter almost always induce a selection bias.

This work was supported by the Australian Research Council and was done at the ARC Special Research Centre for Ultra-Broadband Information Networks (CUBIN), Department of Electrical and Electronic Engineering, The University of Melbourne, Victoria 3010, Australia.

The research of M.F. Neuts was supported in part by NSF Grant Nr. DMI-9988749.

Manuscript received: September 2002. Final version accepted: April 2003.

In the theory of stationary point processes, the mathematical definition of *arbitrary* is the formalization of an idealized averaging process, yet no explicit random selection is involved therein. Rigorous definitions of arbitrary time epochs and arbitrary arrivals or departures require substantial discussions in major treatises on point processes or queues, see e.g., Daley et al. (2002), Franken et al. (1983), and Karr (1991).

In our partially expository discussion, we use a simplified version of the simulation of an actual telecommunications model, see Neuts et al. (2003). We first describe the notion of an arbitrary admitted message and the estimation of related empirical quantities. Next, we discuss a plausible random selection which might be claimed to produce an arbitrary message. We shall argue that random procedures to generate events believed to be *arbitrary* typically have hidden selection biases. These result in erroneous simulation results. While we only compare the valid simulation method to one example of random selection, selection bias is to be expected in other examples. Moreover, in an elaborate simulation we typically do not have explicit analytic results for comparisons. Biased numerical results may still appear plausible yet they are based on an erroneous method, so we have no grounds for confidence in that plausibility.

We intentionally chose a greatly simplified version of our original telecommunications model. That way, we are able to derive explicit analytic results which permit a full comparison of the two alternative simulation methods to substantiate the resulting numerical differences.

2 The model

Consider a positive integer K and a Poisson process of rate λ on the time interval $(0, 1)$. At time 0, we draw an integer i, $0 \leq i \leq K - 1$, with the discrete probability density $\{p(i)\}$. Up to $K - i$ of the earlier Poisson events in $(0, 1)$, are retained (admitted); all later arrivals are discarded. If no Poisson events occur in $(0, 1)$, we repeat the simulation until arrivals occur. Successive independent replications of the experiment are performed. We want to estimate the expected arrival time $\mu = E(Y)$ of an *arbitrary* admitted Poisson event.

In the theory of stationary point processes, the definition of an *arbitrary* event has received much attention. Our first method is inspired by that theory. In a first approach, we do not actually try to generate *arbitrary* job. Instead, the distributional properties of such a job are studied by averaging over a long simulation run. The second approach is based on the suggestion that the arbitrary arrival can be chosen at random from among the jobs admitted during each replication. These approaches yield different results. By deriving analytically explicit expressions for the mean arrival times in both cases, we show why the differences in the simulation results are to be expected. We clarify the hidden *selection bias* inherent in the second method and argue that the first method is the valid one.

In each replication of the first method, *Selection 1*, we generate the initial condition i and the times of the Poisson arrivals during $(0,1)$. Up to $K-i$ of these arrivals are designated as the *admitted* arrivals. Let their number be j. If $j=0$, we regenerate the Poisson process until at least one arrival occurs ($j>0$). Replications with at least one arrival are indexed by r. We form the sum J^* of the successive numbers j_r of admitted arrivals and the sum $S^*(Y)$ of the successive arrival times $Y_{r,\nu}$, $1 \le \nu \le j_r$, over all replications.

The mean arrival time is estimated by $S^*(Y)/J^*$. When the distribution $F(\cdot)$ of an arbitrary arrival time is of interest, we form the empirical distribution of the times of admissions or, alternatively, we divide $(0,1)$ into a large number of subintervals of equal size that serve as the classes of a histogram of the observed arrival times.

In each replication of the second method, *Selection 2*, we again generate i and the times of up to $K-i$ arrivals in the Poisson process during $(0,1)$, repeating if necessary until there is at least one arrival. If j arrivals are *admitted*, one of these is chosen at random. It is treated as an arbitrary arrival and its arrival time Y_r is the variate generated by the rth replication. The mean arrival time is estimated by the sample mean of the Y_r and the empirical distribution or a histogram can serve to estimate the distribution of Y.

2.1 Analysis of the model Selection 1

The mean number E^* of items admitted during $(0,1)$ is given by

$$E^* = \sum_{i=0}^{K-1} p(i) \left[K - i - \sum_{j=0}^{K-i-1} (K-i-j) e^{-\lambda} \frac{\lambda^j}{j!} \right]. \tag{2.1}$$

By direct calculation, that formula is readily obtained from the defining formula

$$E^* = \sum_{i=0}^{K-1} p(i) \left\{ \sum_{\nu=1}^{K-i-1} \nu e^{-\lambda} \frac{\lambda^\nu}{\nu!} + (K-i) \left[1 - \sum_{r=0}^{K-i-1} e^{-\lambda} \frac{\lambda^r}{r!} \right] \right\}.$$

The fraction of admitted Poisson events in the interval $(u, u+du]$, $0 < u < 1$, is

$$\frac{1}{E^*} \sum_{i=0}^{K-1} p(i) \sum_{r=i+1}^{K} e^{-\lambda u} \frac{(\lambda u)^{r-i-1}}{(r-i-1)!} \lambda du, \tag{2.2}$$

since, to be admitted in $(u, u+du]$, the Poisson arrival must be the rth, $i+1 \leq r \leq K$, and it must occur between u and $u+du$. There is no selection of the admitted points. Therefore, the factor of du in (2.2) is also the probability density $\psi_1(u)$ of an *arbitrary* admitted item in $(0,1)$. Using (2.1), we readily verify that, over $(0,1)$, the function $\psi_1(\cdot)$ integrates to one.

We denote the mean of the density $\psi_1(\cdot)$ by μ_1. By an elementary integration, using (2.2), we obtain that

$$\mu_1 = \frac{1}{2\lambda E^*} \sum_{i=0}^{K-1} p(i) \left\{ (K-i)(K-i+1)[1 - \sum_{\nu=0}^{K-i} e^{-\lambda} \frac{\lambda^\nu}{\nu!}] + \lambda^2 \sum_{\nu=0}^{K-i-2} e^{-\lambda} \frac{\lambda^\nu}{\nu!} \right\}. \tag{2.3}$$

2.2 Analysis of the model Selection 2

Let us write $\beta(u;a;b)$ for the beta density

$$\beta(u;a;b) = \frac{1}{B(a,b)} u^{a-1}(1-u)^{b-1}, \tag{2.4}$$

with parameters a and b on $(0,1)$.

We now derive the probability density $\psi_2(\cdot)$ and the mean μ_2 of the admission time of the point selected by the second method.

Theorem 2.1. *For $0 \leq u \leq 1$, the probability density $\psi_2(\cdot)$ is given by*

$$(1 - e^{-\lambda})\psi_2(u) = \sum_{i=0}^{K-1} \psi_2(u;i), \tag{2.5}$$

where for $0 \le i \le K-1$,

$$\psi_2(u;i) = \sum_{j=2}^{K-i-1} \sum_{r=1}^{j-1} e^{-\lambda} \frac{\lambda^j}{j!} \frac{1}{j} \beta(u;r;j-r+1)$$

$$+ \frac{1}{K-i} \sum_{r=1}^{K-i-1} \left\{ \frac{e^{-\lambda u}(\lambda u)^{r-1}\lambda}{(r-1)!} - \sum_{\nu=0}^{K-i-r-1} e^{-\lambda} \frac{\lambda^{r+\nu}}{(r+\nu)!} \beta(u;r;\nu+1) \right\}$$

$$+ \sum_{j=1}^{K-i-1} \frac{1}{j} e^{-\lambda} \frac{(\lambda u)^{j-1}}{(j-1)!} \lambda + \frac{1}{K-i} e^{-\lambda u} \frac{(\lambda u)^{K-i-1}}{(K-i-1)!} \lambda.$$

The mean μ_2 *is given by*

$$\mu_2 = \frac{1}{2(1-e^{-\lambda})} \sum_{i=0}^{K-1} p(i) \left\{ \sum_{j=1}^{K-i-1} e^{-\lambda} \frac{\lambda^j}{j!} + \frac{K-i+1}{\lambda} \left[1 - \sum_{\nu=0}^{K-i} e^{-\lambda} \frac{\lambda^\nu}{\nu!} \right] \right\}. \tag{2.6}$$

Proof. Let the initial condition i be fixed. We consider the various possible positions in the interval $(0,1)$ of the rth (the one to be selected) and the jth (and last) admitted arrival. To avoid repeated use of the common factor $(1-e^{-\lambda})^{-1}$, the probabilities and expected values that follow are conditional on there being at least one Poisson arrival in $(0,1)$.

(a) For $1 \le r < j < K-i$ and $0 < u < v < 1$, the elementary probability that the rth arrival occurs in $(u, u+du]$ and that the jth and *last* arrival occurs in $(v, v+dv]$ is given by

$$e^{-\lambda u} \frac{(\lambda u)^{r-1}}{(r-1)!} \lambda du \cdot e^{-\lambda(v-u)} \frac{[\lambda(v-u)]^{j-r-1}}{(j-r-1)!} \lambda dv \cdot e^{-\lambda(1-v)}$$
$$= e^{-\lambda} \frac{\lambda^j}{(r-1)!(j-r-1)!} u^{r-1}(v-u)^{j-r-1} du\, dv.$$

(b) For $1 \le r < j = K-i$ and $0 < u < v < 1$, the elementary probability that the rth arrival occurs in $(u, u+du]$ and that the $(K-i)$th and *last* arrival occurs in $(v, v+dv]$ is given by

$$e^{-\lambda u} \frac{(\lambda u)^{r-1}}{(r-1)!} \lambda du \cdot e^{-\lambda(v-u)} \frac{[\lambda(v-u)]^{K-i-r-1}}{(K-i-r-1)!} \lambda dv$$
$$= e^{-\lambda v} \frac{\lambda^{K-i}}{(r-1)!(K-i-r-1)!} u^{r-1}(v-u)^{K-i-r-1} du\, dv.$$

(c) For $1 \le r = j < K - i$ and $0 < u < 1$, the elementary probability that the rth arrival and *last* arrival occurs in $(u, u + du]$ is given by

$$e^{-\lambda u}\frac{(\lambda u)^{r-1}}{(r-1)!}\lambda du \cdot e^{-\lambda(1-u)} = e^{-\lambda}\frac{\lambda^r u^{r-1}}{(r-1)!}du.$$

(d) For $1 \le r = j = K - i$ and $0 < u < 1$, the elementary probability that the $(K-i)$th arrival and *last* arrival occurs in $(u, u + du]$ is given by

$$e^{-\lambda u}\frac{(\lambda u)^{K-i-1}}{(K-i-1)!}\lambda du = e^{-\lambda u}\frac{\lambda^{K-i}u^{K-i-1}}{(K-i-1)!}du.$$

In the expressions where v occurs, we integrate over v from u to 1, to obtain from Case a that for $1 \le r < j < K - i$ and $0 < u < 1$, the elementary probability that there are j admitted arrivals and that the rth arrival occurs in $(u, u + du]$ equals

$$e^{-\lambda}\frac{\lambda^j}{j!}\beta(u; r; j - r + 1)du.$$

From b, it follows that for $1 \le r < j = K - i$ and $0 < u < 1$, the elementary probability that there are $j = K - i$ admitted arrivals and that the rth arrival occurs in $(u, u + du]$ equals

$$\begin{aligned}
&\frac{e^{-\lambda u}(\lambda u)^{r-1}\lambda}{(r-1)!}du \int_0^{1-u} e^{-\lambda w}\frac{(\lambda w)^{K-i-r-1}}{(K-i-r-1)!}\lambda dw \\
&= \frac{e^{-\lambda u}(\lambda u)^{r-1}\lambda}{(r-1)!}\left[1 - \sum_{\nu=0}^{K-i-r-1} e^{-\lambda(1-u)}\frac{[\lambda(1-u)]^\nu}{\nu!}\right] du \\
&= \left\{\frac{e^{-\lambda u}(\lambda u)^{r-1}\lambda}{(r-1)!} - \sum_{\nu=0}^{K-i-r-1} e^{-\lambda}\frac{\lambda^{r+\nu}}{(r+\nu)!}\beta(u; r; \nu + 1)\right\} du.
\end{aligned}$$

To obtain the elementary probability $\psi_2(u; i)du$ that the selected is admitted in $(u, u + du]$, we multiply the term deduced from Case a by j^{-1} and sum over all allowable values of r and j, we add the term deduced from Case b multiplied by $(K - i)^{-1}$, we sum the terms in Case c multiplied by r^{-1} (recall that $r = j$) and, finally, we add the term in Case d multiplied by $(K - i)^{-1}$. That sum is multiplied by $(1 - e^{-\lambda})^{-1}$ to account for the fact that the generation of the Poisson process is repeated until there are some arrivals. There are no useful simplifications in the resulting expressions.

To evaluate μ_2, it is more convenient to calculate the mean arrival time of the selected item over the set where that item is the rth out of j admitted items to arrive. From a, by integrating with respect to u after multiplication by u, we get that for $1 \leq r < j < K - i$, that expected value equals

$$\frac{r}{j+1} e^{-\lambda} \frac{\lambda^j}{j!}.$$

Similarly, after a little calculation, it follows from b that, for $1 \leq r < j = K - i$, that expectation equals

$$\frac{r}{\lambda} \left[1 - \sum_{\nu=0}^{K-i} e^{-\lambda} \frac{\lambda^\nu}{\nu!} \right].$$

The cases c and d contribute the expectations

$$\frac{r}{\lambda} e^{-\lambda} \frac{\lambda^{r+1}}{(r+1)!},$$

and

$$\frac{K-i}{\lambda} \left[1 - \sum_{\nu=0}^{K-i} e^{-\lambda} \frac{\lambda^\nu}{\nu!} \right].$$

So, we see that the mean arrival time of the selected item over the set where there are j admitted items is equal to

$$\frac{1}{j} [\sum_{r=1}^{j-1} \frac{r}{j+1} e^{-\lambda} \frac{\lambda^j}{j!} + \frac{j}{\lambda} e^{-\lambda} \frac{\lambda^{j+1}}{(j+1)!}] = \frac{1}{2} e^{-\lambda} \frac{\lambda^j}{j!},$$

for $1 \leq j < K - i$, from the cases a and c. For $j = K - i$, the cases b and d yield that expected value as

$$\frac{K-i+1}{2\lambda} \left[1 - \sum_{\nu=0}^{K-i} e^{-\lambda} \frac{\lambda^\nu}{\nu!} \right].$$

Finally, we add the contributions of all values of j, we take into account the fact that we repeat generating the Poisson process until there is at least one admitted arrival, and we uncondition on the initial condition i. After routine calculations, we obtain the stated expression for the expected value.

□

3 A numerical example

We report numerical results for one example chosen from among many. To illustrate that simulation results may appear plausible and not bring out the selection bias, we intentionally chose a case where the differences between the computational results for the two models are not too large.

We implemented the formulas (2.3) and (2.6) for $\lambda = 10$, for $p(0) = 0.5$, p(1) $= 0.1, p(2) = 0.1$, $p(3) = 0$, and $p(4) = 0.3$, and for $1 \leq K \leq 12$. Whenever $K \leq 3$, the relevant probabilities $p(i)$ were normalized to sum to one. Table 1 gives the computed results and the simulation estimates, respectively based on 1,000,000 and 10,000,000 variates, for the models Selection 1 and Selection 2.

K	**comput**	**simul** 10^6	**simul** 10^7	**comput**	**simul** 10^6	**simul** 10^7
1	0.09995	0.09980	0.09993	0.09995	0.09984	0.09996
2	0.14521	0.14537	0.14520	0.14151	0.14147	0.14143
3	0.18800	0.18786	0.18802	0.17804	0.17794	0.17795
4	0.23342	0.23345	0.23340	0.22691	0.22691	0.22684
5	0.26277	0.26289	0.26277	0.22195	0.22203	0.22200
6	0.29732	0.29747	0.29727	0.26817	0.26832	0.26817
7	0.33199	0.33203	0.33192	0.31153	0.31164	0.31146
8	0.36446	0.36441	0.36441	0.35102	0.35154	0.35101
9	0.39351	0.39335	0.39351	0.38580	0.38538	0.38570
10	0.41859	0.41875	0.41859	0.41532	0.41483	0.41519
11	0.43952	0.43970	0.43953	0.43939	0.43937	0.43948
12	0.45644	0.45655	0.45643	0.45821	0.45809	0.45819

Table 1: *Computational results and simulation estimates based resp. on* 10^6 *and* 10^7 *variates. The leftmost computational results use formula (2.3); the rightmost are based on (2.6).*

4 Discussion

The analytic expressions for μ_1 and μ_2 are clearly different. Depending on the values of the various parameters, the magnitudes of these means are usually clearly distinct; in many other cases, their difference is sufficiently small that, if only simulation results are available, it could be attributed to random variation. Although all simulated variates take values between zero and one, for many choices of the parameters, they exhibit much random variation. Based on runs of 10^6 or 10^7 variates, our simulation estimates generally agreed with the computed means to three or four decimal places.

We shall argue that the first simulation method correctly estimates the mean of an *arbitrary* admitted Poisson event. The second method has a hidden selection bias and should be avoided. An obvious source of selection bias is the choice of the initial i which introduces significant variability in the number of Poisson arrivals that may be admitted. If i is large, few arrivals can be admitted and these typically will occur early in $(0, 1)$. When i is small and the value of λ is such that many arrivals are admitted, the arrival time of one that is randomly chosen may be all over the interval $(0, 1)$.

However, even if i is fixed ($p(i) = 1$ for some i), we see that μ_1 and μ_2 still differ. The choice of λ and K strongly affects the magnitude of that difference. The selection bias is now induced by the variability of the Poisson counts in successive replications.

In spite of a superficial arbitrariness in their random selection, the points produced by the second method can no way be viewed as *globally* arbitrary. In the informal discussion that follows, we imagine that an unlimited sequence of intervals of unit length are concatenated. In each interval, we generate the epochs of admissions as described in the Introduction. By choosing the time origin at random in one of these unit intervals, that concatenation can be viewed as a realization of a stationary point process.

If it were essential to generate independent and *arbitrary* points from such a process, we ought to perform repeated long simulations of the point process (say, with a million or more events) and for each such simulation one event could be chosen at random. That offers a valid, but impractical approximation to the explicit generation of *arbitrary* points.

However, to accomplish what we set out to do - to estimate the mean distance μ between a point and the nearest lattice point to the left of it - *arbitrary* points are not necessary. For an event chosen at random from a long simulation, that sample mean is an excellent estimate of the theoretical mean. That natural estimate of μ is precisely the average used in the first method Selection 1. In exactly the same manner we obtain an estimate of the probability distribution of the distance between an arbitrary point and the nearest lattice point to its left. For each event, we define an indicator variable which equals one if that distance is at most x and zero otherwise. The empirical frequency of ones in that sequence of indicator variables constructed from a single simulation is the natural estimator of the mean of the indicator associated with a randomly chosen (and therefore *arbitrary*) point in a single long simulation run. Of course, that mean is the value of the probability distribution at x.

The conclusion is clear. In estimating statistical descriptors of *arbitrary* events from simulations one should interpret "*arbitrary*" in the sense that we have explained. The alternative of trying to generate arbitrary variates by some ad hoc random selection is dangerous. It nearly always suffers from some hidden selection bias.

Acknowledgements

The authors thank Professors Moshe Zukerman, Peter Taylor and Dr. Hai Le Vu for many helpful discussions on the subject of this paper and Professors Sigrun Andradottir, George Fishman and Bruce Schmeiser on its relation to the simulation literature. Neuts expresses deep appreciation to the Department of Electronic and Electrical Engineering, The University of Melbourne, for its cordial hospitality during his six-month visit in 2002.

References

Bertrand J. (1907). *Calcul des probabilités.* Gauthier-Villars; 2nd edition, 1907, Reprinted, 1972; Chelsea.

Daley D.J., Vere-Jones D. and Smirnov B. M. (2002). *An Introduction to the Theory of Point Processes.* Springer Verlag, 2nd edition.

Franken P., König D., Arndt U. and Schmidt V. (1983). *Queues and Point Processes.* John Wiley.

Karr A.F. (1991). *Point Processes and Their Statistical Inference.* Marcel Dekker.

Neuts M.F., Guo J., Zukerman, M. and Vu H.L. (2003). The Waiting Time Distribution for a TDMA Model with a Finite Buffer. Proceedings of INFOCOM 2003, San Francisco, April 2003.

Top

Volume 11, Number 1
June 2003

CONTENTS

Page