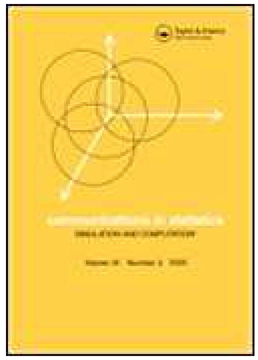
A note on skewness in regenerative simulation

Soren Asmussen, Tobias Rydén



A note on skewness in regenerative simulation

Journal:	*Communications in Statistics - Simulation and Computation*
Manuscript ID:	LSSP-2010-0154.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	17-Sep-2010
Complete List of Authors:	Asmussen, Soren; Aarhus University, Mathematics Rydén, Tobias; Royal Institute of Technology, Department of Mathematics
Keywords:	batch means, busy cycle, coverage probability, regenerative process, regenerative simulation, skewness
Abstract:	The purpose of this paper is to show, empirically and theoretically, that performance evaluation by means of regenerative simulation often involves random variables with distributions that are heavy-tailed and heavily skewed. This, in turn, leads to the variance of estimators being poorly estimated, and confidence intervals having actual coverage quite different from (typically lower than) the nominal. We illustrate these general ideas by estimating the mean occupancy and tail probabilities in M/G/1 queues, comparing confidence intervals computed from batch means to various intervals computed from regenerative cycles. In addition we provide theoretical results on skewness to support the empirical findings.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

regsim3.tex

A note on skewness in regenerative simulation

Søren Asmussen * and Tobias Rydén †

University of Aarhus and Royal Institute of Technology

Abstract

The purpose of this paper is to show, empirically and theoretically, that performance evaluation by means of regenerative simulation often involves random variables with distributions that are heavy-tailed and heavily skewed. This, in turn, leads to the variance of estimators being poorly estimated, and confidence intervals having actual coverage quite different from (typically lower than) the nominal one. We illustrate these general ideas by estimating the mean occupancy and tail probabilities in M/G/1 queues, comparing confidence intervals computed from batch means to various intervals computed from regenerative cycles. In addition we provide theoretical results on skewness to support the empirical findings.

AMS 1991 classification. Primary 00A72, 65C05. Secondary 68M20, 90B05, 90B22.

Keywords: ABC interval, batch means, busy cycle, busy period, coverage probability, Fieller test, jackknife, M/G/1 queue, regenerative process, regenerative simulation, skewness.

1 Introduction

Let $X = (X(t))_{t\geq 0}$ be a stochastic process which is *regenerative with independent cycles*. That is, there exists a renewal process $(S_k)_{k=0}^{\infty}$ such that the *regenerative cycles*

$$(X(t))_{S_{n-1}\leq t<S_n}, \quad n = 1, 2, 3, \ldots,$$

are i.i.d., and not dependent on the process up to S_0 (cf. Asmussen, 2003, p. 169). If $S_0 = 0$ the renewal process is pure, while it is delayed if $S_0 > 0$. The renewal epochs S_n are referred to as the *regeneration times* of X. Typical examples of regenerative processes are found in queueing and storage theory, with the canonical example being the GI/GI/1 queue which regenerates every time a customer arrives to an empty system.

*Department of Mathematical Sciences, Aarhus University, Ny Munkegade, DK-8000 Aarhus C, Denmark

†Corresponding author. Department of Mathematics, Royal Institute of Technology, 100 44 Stockholm, Sweden. Phone +46-8-790 8469; fax +46-8-723 1788.

Now write $\mathbf{E}_0$ for the conditional expectation given $S_0 = 0$ (i.e., given that the process is initialised by a regeneration), $\tau = S_1$ for the first subsequent regeneration time and $\mu = \mathbf{E}_0\tau$ for the mean inter-renewal time. Then, provided that μ is finite, the distribution of τ is non-lattice, and that X has right-continuous paths, it holds that $X(t)$ has an asymptotic distribution, $\mathbf{P}_\mathrm{e}$ say, defined by

$$\mathbf{E}_\mathrm{e}[f(X(\cdot))] = \frac{1}{\mu}\mathbf{E}_0 \int_0^\tau f(X(s))\,\mathrm{d}s \tag{1.1}$$

for any real measurable bounded function f of X (Asmussen, 2003, Theorem VI.1.2).

Regenerative simulation amounts to simulating i.i.d. copies of (τ, C), with C being the integral in (1.1), and using these for estimating the ratio of means on the right-hand side of (1.1). A difficulty with regenerative simulation however, that we want to highlight in this paper, is that such methods may suffer from distributions being heavy-tailed and heavily skewed. Our canonical example will be that of estimating the mean number of customers in an M/GI/1 queue; then $X(t)$ is the number of customers in the system at t, f is the identity function, and C is the area under X during a busy cycle. It has been established that for the M/M/1 queue with unit service rate, this variable has tail probabilities of the form $\mathbf{P}(C > x) \sim (1-\rho)/(\rho\sqrt{2\pi\gamma}) \times x^{-1/4}\mathrm{e}^{-\gamma\sqrt{x}}$, where ρ is the system load and $\gamma = 2\sqrt{-2(1-\rho) - (1+\rho)\log\rho}$; see Guillemin and Pinchon (1998, Eq. (5.20)) and Kearney (2004, Eq. (24)). This tail is subexponential, implying that although a central limit theorem (CLT) for the sample mean of i.i.d. copies of C holds, convergence will be slow. For extensions of these results to GI/GI/1 queues and regularly varying service times, see Kulik and Palmowski (2005). We will also show that as $\rho \uparrow 1$, the distributions of τ and C become totally unbalanced in the sense that asymptotically the probability that they exceed their means tend to zero. From a simulation perspective this means that simulating the expectation of C for instance is difficult for large ρ, since one must take a sufficient number of replications to pick up a few very large values of C before the sample mean becomes a good estimate of its expectation.

The outline of the paper is as follows. First in Section 2 we review the estimators that will be used in the paper (for recent surveys about of the general state-of-the-art in the areas, see e.g. Alexopoulos, 2007, and Law, 2007). Then in Section 3 we give some numerical examples, in which the method of batch means performs better than regenerative simulation. We also illustrate, by means of simulation, the skewness discussed above, and Section 4 provides proofs of this asymptotic unbalancedness.

2 Estimators of steady state expectations

A straightforward approach to estimating $\theta = \mathbf{E}_\mathrm{e}[f(X(\cdot))]$, is through the empirical average

$$\hat{\theta}_{\mathrm{TA}} = T^{-1}\int_0^T f(X(t))\,\mathrm{d}t \tag{2.1}$$

for some large T (TA stands for *time average*). If the simulated X is not stationary the distribution of $X(t)$ will however be different from $\mathbf{P}_{\mathrm{e}}$ for any $t > 0$, although approaching $\mathbf{P}_{\mathrm{e}}$ as $t \to \infty$. Such a transient will induce a bias in $\hat{\theta}_{\mathrm{TA}}$. The standard way of addressing this problem is to first simulate $X(t)$ during a burn-in period of length T_0, and then use the sample path over $[T_0, T_0 + T]$ to compute the empirical average. However, deciding on what is a sufficiently large value for T_0 is not always easy.

Another difficulty with time averages concerns the computation of confidence intervals. Often a CLT with rate $T^{1/2}$ holds for $\hat{\theta}_{\mathrm{TA}}$, i.e. $T^{1/2}(\hat{\theta}_{\mathrm{TA}} - \theta) \to_{\mathrm{d}} \mathrm{N}(0, \sigma^2)$. The asymptotic variance σ^2, often referred to as the *time average variance constant* (TAVC), is then given by

$$\sigma^2 = \int_{-\infty}^{\infty} \mathrm{Cov}_{\mathrm{e}}[X(0), X(s)]\,\mathrm{d}s, \tag{2.2}$$

with $\mathrm{Cov}_{\mathrm{e}}$ denoting covariance under the stationary regime. Estimating the TAVC is a non-trivial task, so that computing a confidence interval for θ is not straightforward. See Asmussen and Glynn (2007, Section IV.3) for further reading.

An alternative method is that of *batch means*. This amounts to splitting $[0, T]$ into B equally long subintervals, and letting $\hat{\theta}^{(k)} = (T/B)^{-1} \int_{(k-1)T/B}^{kT/B} f(X(t))\,\mathrm{d}t$ be the average over the k-th subinterval. If the process X is sufficiently weakly dependent, then the $\hat{\theta}^{(k)}$ will be roughly Normally distributed and independent. Thus a (two-sided) confidence interval for θ with approximate coverage $1-\alpha$, is given by $\hat{\theta}_{\mathrm{TA}} \pm t_{1-\alpha/2}(B-1)s_B/\sqrt{B}$ with s_B^2 the sample variance of the $\hat{\theta}^{(k)}$ and $t_{1-\alpha/2}(B-1)$ being the $(1-\alpha/2)$-quantile of the t-distribution with $B-1$ degrees of freedom.

We now turn to regenerative simulation, and let $((\tau_k, C_k))_{k=1}^{n}$ be i.i.d. copies of (τ, C). The standard estimator of $\theta = \mathbf{E}C/\mathbf{E}\tau$ is

$$\hat{\theta}_{\mathrm{REG}} = \frac{\sum_{k=1}^{n} C_k}{\sum_{k=1}^{n} \tau_k} = \frac{\overline{C}_n}{\overline{\tau}_n}, \tag{2.3}$$

where $\overline{\tau}_n = n^{-1}\sum_1^n \tau_k$ and $\overline{C}_n = n^{-1}\sum_1^n C_k$. That this estimator is consistent is evident from (1.1). It is not unbiased however, because $\hat{\theta}_{\mathrm{REG}}$ is a ratio estimator. Provided that τ and C both have finite second moment, the CLT $n^{1/2}(\overline{\tau}_n - \mathbf{E}\tau, \overline{C}_n - \mathbf{E}C)^\top \to_d \mathrm{N}(0, \Sigma)$ holds for some covariance matrix Σ. Using the delta method one finds that $\hat{\theta}_{\mathrm{REG}}$ is asymptotically Normal at rate $n^{1/2}$, with asymptotic variance η^2 that can be estimated as

$$\hat{\eta}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(C_k - \hat{\theta}_{\mathrm{REG}}\,\tau_k)^2 \Big/ \overline{\tau}_n^2; \tag{2.4}$$

see e.g. Asmussen and Glynn (2007, Proposition IV.4.1). Thus a (two-sided) confidence interval can be constructed as $\hat{\theta}_{\mathrm{REG}} \pm z_{1-\alpha/2}\hat{\eta}/\sqrt{n}$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard Normal distribution.

Another standard approach to constructing confidence intervals is through the jackknife (e.g. Ripley, 1987, pp. 158–160). This estimator is

$$\hat{\theta}_{\mathrm{JACK}} = n\hat{\theta}_{\mathrm{REG}} - (n-1)\hat{\theta}_{(-)} \quad \text{with} \quad \hat{\theta}_{(-)} = n^{-1}\sum_{k=1}^{n}\hat{\theta}_{(k)} \quad \text{and} \quad \hat{\theta}_{(k)} = \frac{\sum_{i\neq k} C_i}{\sum_{i\neq k}\tau_i},$$

and the corresponding variance estimator is $s^2_{\mathrm{JACK}} = (n-1)/n \times \sum_{k=1}^{n}(\hat{\theta}_{(-)} - \hat{\theta}_{(k)})^2$. Confidence intervals are usually computed as $\hat{\theta}_{\mathrm{JACK}} \pm z_{1-\alpha/2}s_{\mathrm{JACK}}$.

The appealing features of regenerative simulation are thus that the method avoids dealing with transient periods (burn-in) and dependence. The fact that $\hat{\theta}_{\mathrm{REG}}$ is a ratio estimator however hints that even if $(\overline{\tau}_n, \overline{C}_n)$ is Normal distributed with reasonable accuracy, it is not obvious that the same conclusion may be drawn about $\hat{\theta}_{\mathrm{REG}}$ itself. One purpose of this paper is indeed to compare confidence intervals computed using batch means, to intervals computed using regenerative simulation. In addition to the standard delta method intervals and jackknife intervals, we will also study three other methods described below.

One way of circumventing the ratio estimator is to construct a linear test procedure based on a t-test. Forming $Z_k(\theta) = C_k - \theta\tau_k$ it is clear from (1.1) that $\mathbf{E}Z_k(\theta) = 0$; testing $\theta = \theta_0$ can thus be done by testing if $\mathbf{E}Z_k(\theta_0) = 0$. By inverting a t-test, a (two-sided) confidence interval for θ with approximate coverage $1-\alpha$ is obtained as

$$\{\theta : |\overline{Z}_n(\theta)/s_{Z(\theta)}| \leq t_{1-\alpha/2}(n-1)\}, \tag{2.5}$$

where $\overline{Z}_n(\theta)$ and $s^2_{Z(\theta)}$ are the sample mean and variance, respectively, of $\{Z_k(\theta)\}_{k=1}^{n}$. This interval was originally derived by Fieller; see e.g. Fieller (1954) or Roy and Potthoff (1958, Section 4), who both considered it as an exact test for a bivariate Normal distribution. In the appendix we show that this procedure is also equivalent to several tests in multivariate analysis of variance (MANOVA). A point estimate can be obtained as the mid-point of the interval, or as the θ that maximises the p-value of the test. The latter is the θ such that the test statistic has minimum absolute value, which occurs when $\overline{Z}_n(\theta) = 0$ and is attained for $\hat{\theta}_{\mathrm{REG}}$. The interval is in general not symmetric around $\hat{\theta}_{\mathrm{REG}}$ however.

Further refinements of the above methods will be done by means of ABC (approximate bootstrap confidence) intervals, introduced by DiCiccio and Efron (1992). Their basic purpose is to provide confidence intervals whose end-points are correct up to order $O(n^{-1})$, rather than $O(n^{-1/2})$ as for the standard intervals, and this is done by including third-order properties (i.e., skewness) of the observed variables. In the present setting the ratio estimator $\hat{\theta}_{\mathrm{REG}} = \overline{C}_n/\overline{\tau}_n$ and the Fieller statistic $\overline{Z}(\theta) = \overline{C}_n - \theta\overline{\tau}_n$ are non-parametric estimators, and they are both functions of the sample means $\overline{\tau}_n$ and $\overline{C}_n$. The influence function for the i-th cycle (τ_i, C_i) is given by $\ell_i = C_i/\overline{\tau}_n - \overline{C}_n/\overline{\tau}_n^2 \times \tau_i$ for the ratio estimator and by $\ell_i = Z_i(\theta) - \overline{Z}_n(\theta)$ for the Fieller statistic, where $Z_i(\theta) = C_i - \theta\tau_i$ as above. The remaining steps of the ABC procedure are given e.g. in Davison and Hinkley (1997, Eq. (5.46)). For the Fieller statistic the expressions become particularly simple; the upper end-point of the one-sided ABC interval for the mean of $Z(\theta)$ with approximate coverage $1-\alpha$ is $\overline{Z}_n + (a + z_{1-\alpha})/(1 - a(a + z_{1-\alpha}))^2 \times n^{-1/2}s_{Z(\theta)}$ with $a = \widehat{\mathrm{skew}}(Z(\theta))/(6\sqrt{n})$ and $\widehat{\mathrm{skew}}(Z(\theta))$ the sample skewness of the $Z_i(\theta)$, and a two-sided confidence interval is obtained by computing end-points for approximate coverage $1 - \alpha/2$ and $\alpha/2$, respectively.

3 Numerical examples

This section contains numerical comparisons of confidence intervals computed using the methods discussed above. For the numerical results we used the M/GI/1 queue with unit mean service time and X being the number of customers in the system. The function f was chosen either as the identity, so that θ becomes the mean number of customers in stationarity, or as the indicator $I(X(t) \geq b)$, so that θ becomes the probability $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq b)$. In the following we chose $b = 3$. We studied three types of queues: the M/M/1 queue, the M/D/1 queue, and the queue with service time density

$$b(x) = \frac{\sqrt{3}}{\pi} \frac{1}{1 + x^6/27} \quad \text{for } x > 0. \tag{3.1}$$

The corresponding distribution has unit mean and variance 1/2, so that its variance is intermediate between that of the exponential distribution, with unit variance, and that of the constant unity with no variance. We will denote this queue as M/HT/1 (HT for heavy tail). The three cases are meant to represent light tail, no tail, and heavy tail, respectively, of the service time distribution. For each case we studied the two loads $\rho = 0.5$ and $\rho = 0.8$.

For the case of the number of customers in the system (i.e., f is the identity function), the Pollaczek-Khintchine formula (e.g. King, 1990, Eq (5.6)) yields the mean number of customers $\rho/(1-\rho)$, $\rho(1-\rho/2)/(1-\rho)$, and $\rho(1-\rho/4)/(1-\rho)$ respectively, for the three queues. We also remark that the TAVC σ^2 in (2.2) is finite even for the M/HT/1 queue. This follows as σ^2 is finite provided that C has finite second moment (e.g. Asmussen and Glynn, 2007, Proposition 4.2), and Daley and Jacobs (1969, Eq. (8.7)) show that so is the case if the service time distribution has finite fourth moment. Obviously this holds true for the density (3.1).

Regarding the tail probability $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq b)$, the stationary distribution of the number of customers in the M/M/1 queue is geometric with parameter ρ, so that $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq b) = \rho^b$ for this queue. For the M/D/1 and M/HT/1 queues, let π_k be the stationary probability of k customers in the system. We have $\pi_0 = 1 - \rho$, and we computed π_1 and π_2 numerically by solving the first two global balance equations of the Markov chain embedded after departures (e.g. Asmussen, 2003, Eq. (5.5)); for the M/HT/1 queue the integrals corresponding to the probabilities of 0 or 1 arrivals during a service time were evaluated using `Maple`. In this way we obtained $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq 3) = 0.0530$ and $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq 3) = 0.3655$ respectively for $\rho = 0.5$ and $\rho = 0.8$ for the M/D/1 queue, and $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq 3) = 0.0914$ and $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq 3) = 0.4564$ respectively for $\rho = 0.5$ and $\rho = 0.8$ for the M/HT/1 queue.

Tables 1–4 show empirical coverage probabilities obtained from regenerative simulation and time-average simulation. For the regenerative simulation, regeneration epochs were defined as arrivals to an empty queue, and confidence intervals were computed using (i) the estimator $\hat{\theta}_{\mathrm{REG}}$ with variance estimate (2.4) and Normal quantiles, (ii) the jackknife estimator with jackknife variance estimator and Normal quantiles, (iii) the Fieller intervals (2.5), (iv) ABC intervals derived from the ratio estimator $\hat{\theta}_{\mathrm{REG}}$ (denoted *ratio-ABC* in the tables), and (v) ABC intervals derived from the Fieller statistic $\overline{Z}_n(\theta)$ (denoted *F-ABC* in the tables). Neither the Fieller intervals nor the Fieller-ABC intervals were inverted into confidence intervals for θ however, since we were only interested in whether these tests rejected the true θ or not.

For the time-average simulation, confidence intervals were computed using the estimator $\hat{\theta}_{\mathrm{TA}}$, variance estimates obtained by splitting the total simulation time into $B = 5$, 10 or 25 batches, and t-quantiles (Asmussen and Glynn, 2007, Section IV.5, recommend 5–30 batches). Alexopoulos (2007) describes elaborate algorithms for adaptively choosing the number of batches; our aim here is however not to use the most sophisticated algorithms, but rather to show that even with a simple choice for the number of batches, this approach performs better than intervals based on regenerative simulation.

Sample sizes were $n = 500$, 1,000, and 2,000 regenerative cycles for the estimates based on regenerative simulation. Since a regenerative cycle is a busy period plus a waiting time until another arrival, its mean value is $\mathbf{E}\tau = 1/(1-\rho) + 1/\rho = 1/\rho(1-\rho)$ in a queue with unit mean service time (see e.g. King, 1990, Eq. (8.1)). For the time-average estimator the total simulation time was chosen as $T = n\mathbf{E}\tau$, preceded by a burn-in period of duration $T_0 = T/5$ for the maximal value of T (corresponding to $n = 2{,}000$). All simulations were done in `Matlab`.

From Tables 1–2, reporting empirical coverages for the mean number of customers, we can draw the general conclusions that performance is best, i.e. estimation of θ is easiest, for the M/D/1 queue and, maybe surprisingly, worst for the M/M/1 queue with the M/HT/1 queue in between. A possible explanation for this is that the M/M/1 queue has the largest variance of the service time distribution. Of the five methods based on regenerative simulation, we see that the jackknife and ABC intervals perform best; this finding is in agreement with those of Iglehart (1975, Section 5), who found the jackknife intervals to be superior to several other ones, including the Fieller and the standard delta method intervals. A closer analysis reveals that the differences towards the other intervals are statistically significant at the 5%-level (one-sided test) in most cases for $\rho = 0.8$, but in only a few cases for $\rho = 0.5$. The Fieller-ABC intervals perform the best in all cases for $\rho = 0.8$. For the M/HT/1 queue this is peculiar, as the cumulative variable C does not have finite third moment for the service time density (3.1). The ABC intervals involve the empirical skewness, whence these procedures are highly questionable from a theoretical perspective for the M/HT/1 queue, despite their better performance in the simulations.

The overall best intervals are however those computed from batch means with 5 batches. As indicated by the tables, this method has consistently better coverage than the other methods, and these differences are significant (5%-level, one-sided) in most cases. (We did however not attempt to correct for multiple testing.) However, also for this method the empirical coverage is significantly below the nominal 95% for all queues when $\rho = 0.8$, and

Table 1: *Empirical coverage probabilities for different simulation strategies and confidence intervals applied to the mean number of customers in an M/GI/1 queue with unit mean service time and load* $\rho = 0.5$. *Top part: M/M/1 queue; middle part: M/D/1 queue; bottom part: M/GI/1 queue with service time density (3.1). For the time-average estimator, the simulation periods were* $T = n/\rho(1-\rho)$, *preceded by a burn-in period of length* $T_0 = T/5$ *for the maximal* T. *All confidence intervals and tests were two-sided with nominal coverage probability 95% (size 5%). The number of replications was 8,000, giving two-sided confidence intervals for the actual coverage probabilities with half-widths of about 0.007. Figures in italics mark coverages that are significantly smaller, at the 5% level (one-sided test for difference of probabilities in two binomial distributions), than the coverage (on the same row) for the time-average estimator and 5 batches.*

	regenerative simulation					batch means		
	delta	jackknife	Fieller	ratio-ABC	F-ABC	5	10	25
$n = 500$	*0.912*	*0.914*	*0.908*	*0.918*	*0.918*	0.934	*0.925*	*0.911*
$n = 1{,}000$	*0.920*	*0.923*	*0.918*	*0.926*	*0.927*	0.945	0.941	*0.934*
$n = 2{,}000$	*0.934*	*0.935*	*0.933*	*0.932*	*0.932*	0.944	0.940	0.940
$n = 500$	*0.928*	*0.928*	*0.928*	*0.928*	*0.926*	0.946	0.942	*0.938*
$n = 1{,}000$	0.941	0.941	0.941	0.940	0.939	0.944	0.944	0.940
$n = 2{,}000$	0.945	0.946	0.944	0.942	0.942	0.947	0.946	0.944
$n = 500$	*0.920*	*0.922*	*0.918*	*0.925*	*0.923*	0.942	*0.932*	*0.923*
$n = 1{,}000$	*0.932*	0.935	*0.932*	*0.932*	*0.931*	0.941	0.937	*0.933*
$n = 2{,}000$	0.945	0.947	0.944	0.943	0.943	0.945	0.940	0.941

Table 2: *Empirical coverage probabilities for mean number of customers as in Table 1, but with load* $\rho = 0.8$.

	regenerative simulation					batch means		
	delta	jackknife	Fieller	ratio-ABC	F-ABC	5	10	25
$n = 500$	*0.832*	*0.847*	*0.808*	*0.857*	*0.864*	0.899	*0.871*	*0.841*
$n = 1{,}000$	*0.876*	*0.886*	*0.861*	*0.893*	*0.897*	0.916	*0.899*	*0.887*
$n = 2{,}000$	*0.903*	*0.908*	*0.892*	*0.912*	*0.916*	0.929	*0.918*	*0.908*
$n = 500$	*0.879*	*0.888*	*0.860*	*0.897*	*0.903*	0.914	*0.902*	*0.880*
$n = 1{,}000$	*0.903*	*0.910*	*0.893*	*0.917*	*0.921*	0.933	*0.917*	*0.904*
$n = 2{,}000$	*0.914*	*0.917*	*0.907*	*0.923*	*0.925*	0.940	*0.931*	*0.921*
$n = 500$	*0.855*	*0.868*	*0.832*	*0.878*	*0.885*	0.904	*0.884*	*0.862*
$n = 1{,}000$	*0.883*	*0.891*	*0.870*	*0.898*	*0.904*	0.926	*0.911*	*0.894*
$n = 2{,}000$	*0.907*	*0.913*	*0.898*	*0.916*	*0.920*	0.935	*0.927*	*0.917*

Table 3: *Empirical coverage probabilities for tail probability* $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq 3)$ *and load* $\rho = 0.5$*; simulation set-up is otherwise as described in Table 1.*

	regenerative simulation					batch means		
	delta	jackknife	Fieller	ratio-ABC	F-ABC	5	10	25
$n = 500$	*0.922*	*0.926*	*0.921*	0.934	0.934	0.938	0.932	*0.925*
$n = 1{,}000$	*0.932*	*0.934*	*0.930*	*0.938*	*0.938*	0.948	0.948	0.942
$n = 2{,}000$	*0.938*	*0.938*	*0.938*	*0.942*	*0.942*	0.950	0.951	0.948
$n = 500$	*0.916*	*0.917*	*0.916*	0.932	0.932	0.935	0.932	*0.926*
$n = 1{,}000$	*0.936*	0.937	*0.936*	0.942	0.942	0.942	0.942	0.939
$n = 2{,}000$	0.943	0.943	0.944	0.945	0.945	0.948	0.946	*0.939*
$n = 500$	*0.922*	*0.923*	*0.922*	0.939	0.938	0.938	0.932	*0.927*
$n = 1{,}000$	0.939	0.940	0.937	0.940	0.940	0.943	0.937	*0.936*
$n = 2{,}000$	0.946	0.946	0.946	0.949	0.949	0.946	0.944	0.941

Table 4: *Empirical coverage probabilities for tail probability* $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq 3)$ *as in Table 3, but with load* $\rho = 0.8$*.*

	regenerative simulation					batch means		
	delta	jackknife	Fieller	ratio-ABC	F-ABC	5	10	25
$n = 500$	*0.917*	*0.928*	*0.910*	*0.926*	*0.926*	0.944	*0.935*	*0.914*
$n = 1{,}000$	*0.935*	0.939	*0.930*	*0.936*	*0.936*	0.945	0.945	*0.938*
$n = 2{,}000$	0.943	0.946	0.939	0.944	0.945	0.942	0.944	0.940
$n = 500$	*0.929*	0.934	*0.919*	0.933	0.934	0.939	0.939	*0.931*
$n = 1{,}000$	*0.940*	0.943	*0.935*	0.944	0.944	0.947	0.946	*0.940*
$n = 2{,}000$	*0.935*	*0.938*	*0.935*	*0.941*	*0.941*	0.948	0.948	0.944
$n = 500$	*0.929*	*0.934*	*0.921*	*0.933*	*0.933*	0.943	*0.935*	*0.921*
$n = 1{,}000$	*0.939*	0.943	*0.930*	*0.941*	*0.941*	0.949	0.945	*0.936*
$n = 2{,}000$	0.945	0.948	*0.942*	0.947	0.947	0.950	0.951	0.948

for the M/M/1 and M/HT/1 queues also when $\rho = 0.5$, even for the largest sample size, although the the differences are small for the M/D/1 queue and for the M/M/1 queue when $\rho = 0.5$.

The results for estimating the tail probability $\mathbf{P}_{\mathrm{e}}(X(\cdot) \geq 3)$, reported in Tables 3–4, are in many respects similar to those regarding the mean number of customers, although the differences between the two loads are less pronounced. The two types of ABC intervals are indistinguishable for $\rho = 0.5$, and for $\rho = 0.8$ there are no notable differences at all between these and the jackknife intervals. Again batch means with 5 batches performs best, and often significantly better than the other methods. The coverage probabilities of all methods are better than for the mean number of customers, however.

Obviously the Fieller intervals (2.5) do not perform better than the other tests based on regenerative simulation, which are ratio estimators. This suggests that these methods perform worse than batch means not because they are ratio estimators, but because the average $(\overline{\tau}_n, \overline{C}_n)$ is not approximately Normal with sufficient accuracy. A Normal probability plot (not shown) with simulated replications of the $\overline{C}_n$ for the number of customers in an M/M/1 queue with $\rho = 0.8$ and $n = 2{,}000$, does indeed reveal some deviations from Normality. From a sample of 250,000 simulated replications of C, we obtained the sample mean 25.00 (the true value is 25), and the sample standard deviation 155.2. Moreover, the sample skewness was 20.7, and the fraction of replications larger than 25 was 0.112; these are both numbers that indicate a large degree of skewness of the distribution of C. An estimate of the density of C is plotted in Figure 1.

We also simulated, for the same setting, a sample of 250,000 replications of (τ, C) and the corresponding $Z(\theta) = C - \theta\tau$ with the true $\theta = \rho/(1-\rho) = 4$. The mean of this sample was 0.074 (the true value is 0) and its sample standard deviation was 104.2. Moreover the sample skewness was 29.3, and the fraction of replications larger than 0 was 0.046. An estimate of its density is found in Figure 1, and one should note the very sharp decrease of the density just left of zero. Our conclusion here is that splitting a sample trajectory as is done in regenerative simulation, implies a way of estimating the TAVC that involves distributions which are both heavy-tailed and heavily skewed, and that this explains the inferior performance of estimators built on regenerative simulation. Even the ABC intervals, which do adjust for skewness, are unable to fully compensate for it and achieve coverage better than for the batch means. We did also compute the BC$_a$ (bias-corrected accelerated) intervals for the Fieller statistic $\overline{Z}_n(\theta)$, which is another method of constructing second-order correct confidence intervals for the mean of this estimate (see Efron, 1987, Example 2). These intervals did however in fact perform worse than the Fieller intervals, and we do not report these results.

Regarding the variable τ, its distribution is not heavy-tailed in the M/M/1-setting, because the busy cycle distribution has an exponential tail; this follows as its Laplace transform (e.g. Asmussen, 2003, Proposition III.8.10) is finite to the left of the origin. Its distribution is however asymptotically unbalanced in a sense detailed in the next section; for the sample used here, a fraction 0.200 of the replications were larger than the true mean $1/\rho(1-\rho) = 6.25$. These observations may explain why the confidence intervals built on regenerative simulation

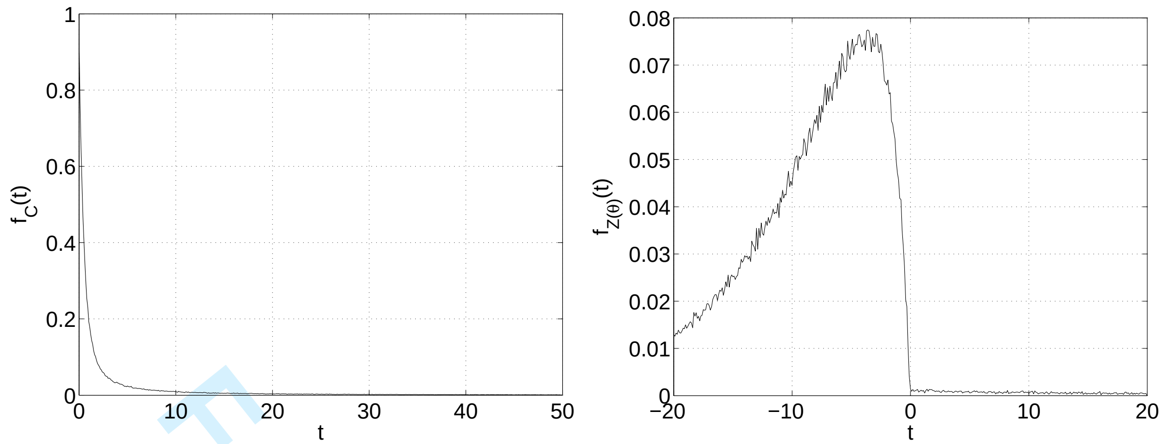


Figure 1: *Estimates of the densities of the random variables C (left) and $Z(\theta)$ (right), for the number of customers in the M/M/1 queue with $\rho = 0.8$, and θ set to its true value $\rho/(1-\rho) = 4$. The estimates are renormalised histograms computed from 250,000 replications of C and $Z(\theta)$ respectively.*

perform worse than those built on batch means also for the tail probability $\mathbf{P}_e(X(\cdot) \geq 3)$, for which the function $f(x) = I(x \geq 3)$ is bounded, but with smaller differences than for estimating the mean number of customers.

For the time-average estimator, the *distribution* of concern is the stationary distribution $\mathbf{P}_e(f(X(\cdot) \in \cdot))$. For the M/M/1 and M/D/1 queues, this is not heavy-tailed for the examples studied here. The concern is then rather about the dependence over time. This dependence is however light-tailed (exponentially decaying) both for the M/M/1 queue and for the M/D/1 queue. For these queues there are thus no heavy-tailed phenomena disturbing the time-average estimator, while there are for the estimators built on regenerative simulation. For the M/HT/1 queue we however expect heavy-tailed dependence over time and also a heavy-tailed distribution for the number of customers (although a Bernoulli distribution when using the indicator function for the tail probability). This however does not have a notable degrading effect on coverage in the simulations reported here. As noted above, for 5 batches the batch means intervals perform much better than those built on regenerative simulation. We did not explore further why coverage is better with 5 batches than with 10 or 25, but two possible explanations are that (i) with the smaller batch sizes the batch averages are correlated to an extent that the variance estimate (based on the assumption of no correlation) becomes too small, or that (ii) with the smaller batch sizes the distribution of the batch means becomes too far from Normal, so that the distribution of the t-statistic is not well approximated by the t-distribution. Of course, also both of (i) and (ii) may occur.

4 Asymptotically totally unbalanced distributions

In this section we address from a theoretical perspective the empirical observations about skewness made in the previous section. First we make the following definition.

Definition 1 *Let (ν_s) be a family of distributions on $\mathbb{R}$, for s in some index set of integers or reals, and let $m_s = \int_{\mathbb{R}} x\,\nu_s(\mathrm{d}x)$ be the mean of ν_s. We say that this family is* asymptotically totally unbalanced *as $s \uparrow a$, where a is a limit point, finite or infinite, if the mean m_s is finite for all s, and $\nu_s(\{x : x > m_s\})$ tends to either 0 or 1 as $s \uparrow a$. For $s \downarrow a$ the definition is analogous.*

Consider first the variable C when X is the number of customers in an M/M/1 queue and f is the identity function. In the Introduction we mentioned that C then has a subexponential distribution. Write $\mathbf{P}_\rho$ for the distribution of $(X(t))_{t\geq 0}$, when stationary with traffic load ρ, and $\mathbf{E}_\rho$ for the corresponding expectation.

Proposition 1 *For $X = (X(t))_{t\geq 0}$ being the number of customers in an M/M/1 queue with load ρ, $\mathbf{P}_\rho(C \in \cdot)$ and $\mathbf{P}_\rho(\tau \in \cdot)$ are asymptotically totally unbalanced as $\rho \uparrow 1$, namely*

$$\mathbf{P}_\rho(C > \mathbf{E}_\rho C) \to 0 \text{ as } \rho \uparrow 1 \quad \text{and} \quad \mathbf{P}_\rho(\tau > \mathbf{E}_\rho \tau) \to 0 \text{ as } \rho \uparrow 1.$$

Proof. We do the proof for C only, as the proof for τ is completely analogous. We first notice that C is finite $\mathbf{P}_1$-a.s., because even for $\rho = 1$ a busy cycle is of finite duration a.s.; this is since the embedded X-process, i.e. X observed at arrivals and departures only, is then a null-recurrent random walk. Second, for $\rho < \rho' \leq 1$ the $\mathbf{P}_{\rho'}$-distribution of C is stochastically larger than the $\mathbf{P}_\rho$-distribution of C. This can be established, for instance, by a coupling argument: run two queues with identical sequences of service times, with unit mean, and inter-arrival times constructed as U_k/ρ and U_k/ρ' respectively, where (U_k) is an i.i.d. sequence of standard negative exponential random variables. Letting C and C' correspond to the first busy cycle of the respective queues, it then holds that $C' \geq C$ a.s.

From this stochastic domination result we conclude that

$$\mathbf{P}_\rho(C > \mathbf{E}_\rho C) \leq \mathbf{P}_1(C > \mathbf{E}_\rho C) \quad \text{for all } 0 \leq \rho \leq 1.$$

Since C is finite $\mathbf{P}_1$-a.s., the proof is now completed by letting $\rho \uparrow 1$ and noting that then $\mathbf{E}_\rho C \to \infty$. $\square$

In the previous section we saw that even for a moderate value like $\rho = 0.8$, skewness can be severe.

The above argument also applies to M/GI/1 queues, and also to more general systems like the GI/GI/1 queue. For the M/GI/1 queue it holds that (again assuming unit mean service time) $\mathbf{E}_\rho C = 1/(1-\rho) + (\rho b_2/2)/(1-\rho)^2$ with b_2 being the second moment of the waiting time distribution (Daley and Jacobs, 1969, Eq. (8.6)). Moreover, $\mathbf{P}_1(C > x) \sim K x^{-1/3}$ as $x \to \infty$

for some constant K (Kearney, 2004, Eq. (17)). These results are however unimportant for the above argument.

We now turn to the Fieller statistic $Z(\theta) = C - \theta\tau$ in (2.5), where we again let X be the number of customers in an M/M/1 queue. Then $\theta = \mathbf{E}_e X(\cdot) = \rho/(1-\rho)$, and $\mathbf{E}Z(\theta) = 0$.

Proposition 2 *For $X = (X(t))_{t\geq 0}$ being the number of customers in an M/M/1 queue with load ρ, $\mathbf{P}_\rho(Z(\theta(\rho)) \in \cdot)$ is asymptotically totally unbalanced as $\rho \uparrow 1$, namely*

$$\mathbf{P}_\rho(Z(\theta(\rho)) > 0) \to 0 \quad \text{as } \rho \uparrow 1.$$

Proof. Consider a busy cycle during which in total $d \geq 1$ customers are served. Clearly C is then maximal when all customers arrive at the very start of the busy cycle, and the maximal value of C is $S_1 + 2S_2 + \ldots + dS_d$, where S_k is the service time of the k-th customer served during the busy cycle. The length of the regenerative cycle is $S_1 + \ldots + S_d + A$, where A is the time from the end of the busy cycle to the next arrival (then to an empty queue). Thus, for such a busy cycle we have $Z(\theta) \leq (1-\theta)S_1 + (2-\theta)S_2 + \ldots + (d-\theta)S_d$. If ρ is large enough that $\theta = \theta(\rho) \geq d$, say $\rho \geq \rho(d)$, the right-hand side of this inequality will be non-positive. Since $\theta(\rho) \to \infty$ as $\rho \uparrow 1$, such a $\rho(d)$ can always be found.

Now write M for the number of customers served during the busy period. Since the above argument does not involve the actual traffic load, we conclude that for any ρ such that $\theta(\rho) \geq d$ we have

$$\begin{aligned}\mathbf{P}_\rho(Z(\theta(\rho)) > 0) &= \mathbf{P}_\rho(Z(\theta(\rho)) > 0 \,|\, M \leq d)\mathbf{P}_\rho(M \leq d) \\ &\quad + \mathbf{P}_\rho(Z(\theta(\rho)) > 0 \,|\, M > d)\mathbf{P}_\rho(M > d) \\ &\leq 0 + \mathbf{P}_\rho(M > d) \leq \mathbf{P}_1(M > d).\end{aligned}$$

When $\rho \uparrow 1$ we can let $d \to \infty$, and hence, because M is finite $\mathbf{P}_1$-a.s. (cf. the proof of Proposition 1), the result follows. $\square$

This result explains why the Fieller intervals work poorly for ρ close to 1. We however remark that, again, the numerical results indicate that also for moderate values of ρ these intervals can suffer substantially from skewness, and that the Fieller-ABC intervals are not able to fully compensate for this skewness. We also remark that again the above argument is valid for more general systems, like M/GI/1 and GI/GI/1 queues.

A The Fieller test as a MANOVA test

Consider an $2 \times n$-matrix $\mathbf{Y}$ of random variables, where the columns of $\mathbf{Y}$ are i.i.d. random vectors with some unknown mean vector ξ and unknown covariance matrix. We shall think of each column as being an independent copy of $(\tau, C)^\top$ from a regenerative cycle. To test the hypothesis $H_0 : \theta\xi_1 - \xi_2 = 0$, we compute the quantities

$$\begin{aligned}P &= \mathbf{C}\mathbf{Y}n^{-1}\mathbf{A}^\top\mathbf{A}\mathbf{Y}^\top\mathbf{C}^\top, \\ Q &= \mathbf{C}\mathbf{Y}(\mathbf{I}_n - n^{-1}\mathbf{A}^\top\mathbf{A})\mathbf{Y}^\top\mathbf{C}^\top,\end{aligned}$$

where $\mathbf{C} = [\theta, -1]$ and $\mathbf{A}$ is an $1 \times n$-vector of ones; see Khatri (1966) or Srivastava and Carter (1983, Section 6.3.2). These expressions follow from Khatri (1966, Eqs. (17)–(18)) after some simple algebra (in Khatri's notation, $\mathbf{B} = I_{2\times 2}$, $\mathbf{V} = 1$, $\mathbf{Y}_1 = \mathbf{Y}$, $\mathbf{A}^* = \mathbf{A}$). Under the assumption that each column of $\mathbf{Y}$ is bivariate Normal, the likelihood ratio tests rejects H_0 if $Q/(Q+P)$ is small. Another test, known as Pillai's trace, rejects H_0 if $P/(P+Q)$ is large. Since P and Q are scalars in this problem and $P/(P+Q) = 1 - Q/(P+Q)$, these tests are equivalent. Pillai's trace test is known to possess robustness against non-normality (Kotz and Johnson, 1985, p. 25), which is appealing as in regenerative simulation (τ, C) is in general not bivariate Normal. It is also immediate that the LR test and the Pillai trace test are both equivalent to rejecting H_0 for large values of P/Q. However, using the notation of (2.5), we have $P = n\overline{Z}_n^2(\theta)$ and $Q = (n-1)s^2_{Z(\theta)}$, and we find that P/Q is, up to a multiplicative constant, equal to the Fieller statistic (2.5) squared. Thus the both of the MANOVA test statistics discussed here, and the Fieller t-test, are all equivalent.

References

Alexopoulos, C. (2007). Statistical analysis of simulation output: State of the art. In: Henderson, S.G., *et al.*, eds. *Proc. 2007 Winter Simulation Conference*:150–161.

Asmussen, S. (2003). *Applied Probability and Queues*, 2nd ed. New York: Springer.

Asmussen, S., Glynn, P.W. (2007). *Stochastic Simulation: Algorithms and Analysis*. New York: Springer.

Daley, D.V., Jacobs, D.R., Jr. (1969). The total waiting time in a busy period of a stable single-server queue, II. *J. Appl. Prob.* 6:565–572.

Davison, A.C., Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.

DiCiccio, T.J., Efron, B. (1992). More accurate confidence intervals in exponential families. *Biometrika* 79:231–245.

Efron, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* 82:171–185.

Fieller, E.C. (1954). Some problems in interval estimation. *J. Roy. Statist. Soc. B* 16:175–185.

Guillemin, F., Pinchon, D. (1998). On the area swept under the occupation process of an M/M/1 queue in a busy period. *Queueing Syst.* 29:383–398.

Iglehart, D.L. (1975). Simulating stable stochastic systems, V: Comparison of ratio estimators. *Naval Res. Logist. Quart.* 22:553–565.

Kotz, N.L., Johnson, S., eds. (1985). *Encyclopedia of Statistical Sciences*, vol. 6. New York: Wiley.

Kearney, M.J. (2004). On a random area variable arising in discrete-time queues and compact directed percolation. *J. Phys. A: Math. Gen.* 37:8421–8431.

Khatri, C.G. (1966). A note of a MANOVA model applied to problems in growth curve. *Ann. Inst. Statist. Math.* 18:75–86.

King, P.J.B. (1990). *Computer and Communication Systems Performance Modelling*. Hemem Hempstead: Prentice Hall.

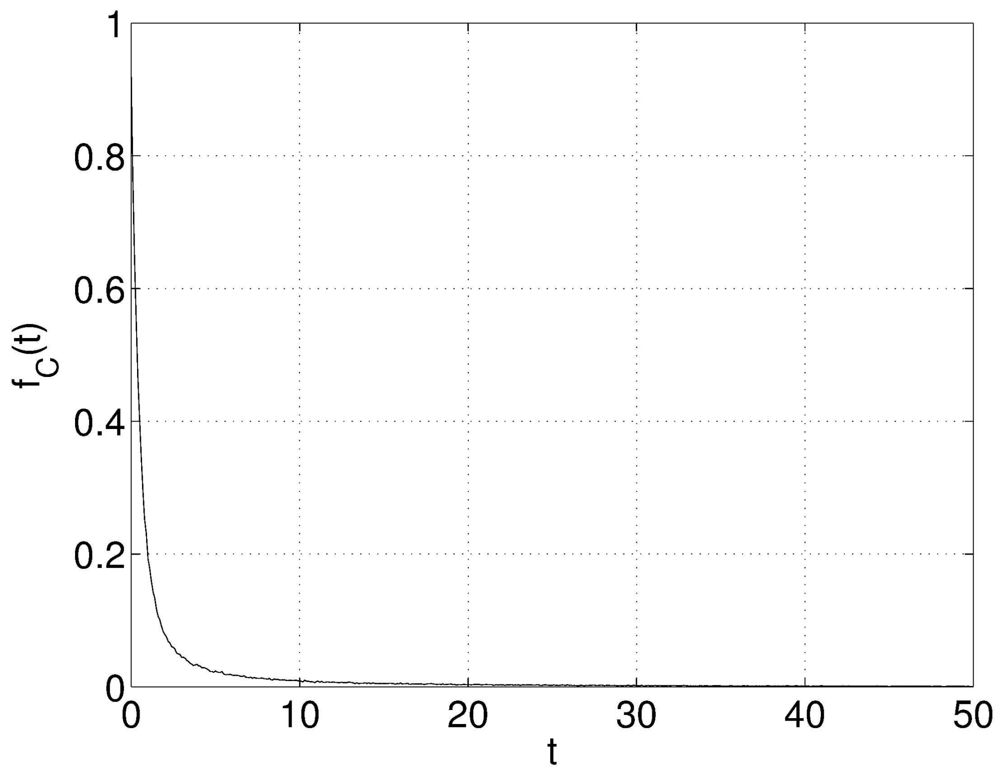
Kulik, R., Palmowski, Z. (2005). Tail behaviour of the area under the queue length process of the single server queue with regularly varying service times. *Queueing Syst.* 50:299–323.

Law, A.M. (2007). Statistical analysis of simulation output data: The practical state of the art. In: Henderson, S.G., *et al.*, eds. *Proc. 2007 Winter Simulation Conference*:77–83.

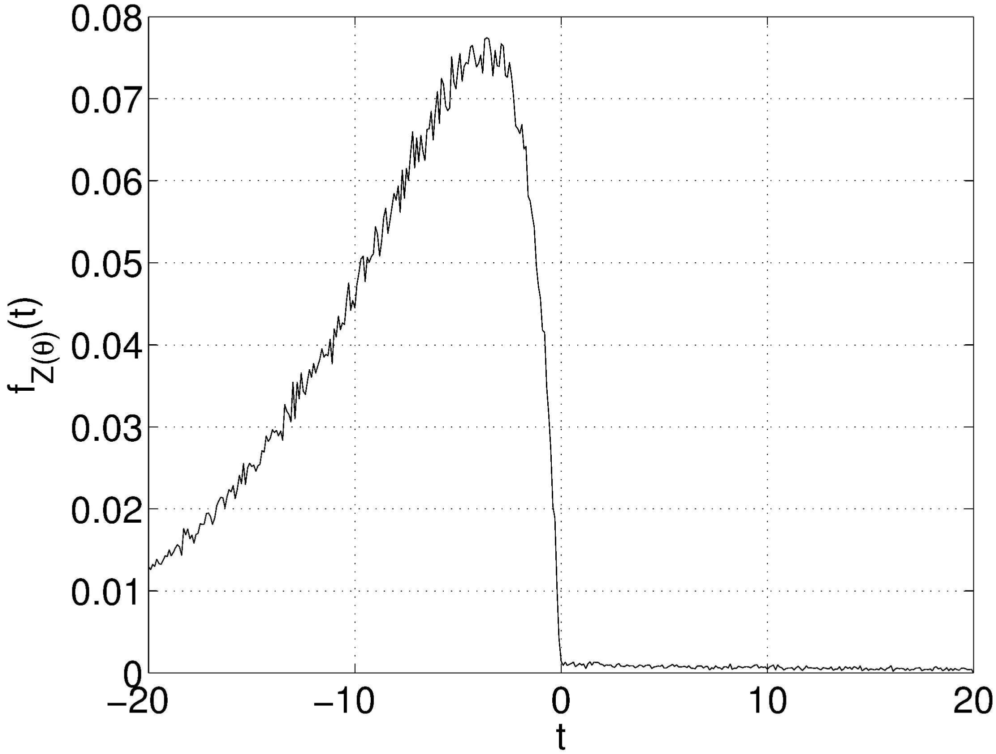
Ripley, B.D. (1987). *Stochastic Simulation*. New York: Wiley.

Roy, S.N., Potthoff, R.F. (1958). Confidence bounds on vector analogues of the "ratio of means" and the "ratio of variances" for two correlated normal variates and some associated tests. *Ann. Math. Statist.* 29:829–841.

Srivastava, M.S., Carter, E.M. (1983). *An Introduction to Applied Multivariate Statistics*. New York: North-Holland.



Estimates of the densities of the random variable C, for the number of customers in the M/M/1 queue with $\rho = 0.8$, and θ set to its true value $\rho/(1-\rho) = 4$. The estimate is a renormalised histograms computed from 250,000 replications of C.
187x146mm (600 x 600 DPI)



Estimates of the densities of the random variable Z(θ), for the number of customers in the M/M/1 queue with ρ = 0.8, and θ set to its true value ρ/(1−ρ) = 4. The estimate is a renormalised histograms computed from 250,000 replications of Z(θ).
189x146mm (600 x 600 DPI)