应用概率统计　第 32 卷
第 3 期　2016 年 6 月

Chinese Journal of Applied Probability and Statistics
Jun., 2016, Vol. 32, No. 3, pp. 221-260
doi: 10.3969/j.issn.1001-4268.2016.03.001

Survey

《应用概率统计》版权所有

Some Recent Advances in Stochastic Simulation

WANG Chia-Li

(*Department of Applied Mathematics, National Dong Hwa University, Hualien, Taiwan*)

Abstract: This review article introduces two recent advances in stochastic simulation: the construction of efficient algorithms for estimating rare events and the generation of samples from a stationary distribution that has no closed form.

Estimating a very small quantity requires extreme accuracy to form a useful confidence interval. This makes the slowly convergent rare-event simulation a challenge task in both efficiency and accuracy. In this report, we introduce the examples of rare events of interest and the difficulties in estimating them. Various approaches to pursue robust and efficient estimators along the development are discussed and evaluated. Numerical experiments on estimating ruin probability are provided to show the quality of these approaches.

In steady-state simulation, how to generate samples from a stationary stochastic process has long been the key subject. The common practice is to discard the data gathered during the initial transient period. However, how long the warm-up period must be raises another problem that has no satisfactory answer. Fortunately, by the development in the past two decades, exact simulation has become possible for certain stochastic models. In this report, we will introduce two important methods and related applications.

Keywords: ruin probability; logarithmic efficiency; exponential change of measure; heavy tail; subexponential; importance sampling; control variates; conditional estimators; Markov chain; stationary distribution; regenerative process; coupling from the past

2010 Mathematics Subject Classification: 60K25

§1. Examples of Rare Event

Suppose we want to estimate $\alpha = \mathsf{P}(A)$ where α is small, say of order 10^{-3} or less, that is, A is a *rare event*. Interesting examples occur in telecommunication (α is the probability of cell loss or buffer overflow), reliability (α is the probability of failure before some fixed time), insurance risk (α is the ruin probability), etc. We give more details of these applications below.

Received March 1, 2016.

《应用概率统计》版权所有

Example 1 (Performance Measure in Telecommunication Networks) In asynchronous transfer mode (ATM) networks, the connection admission control (CAC) algorithm decides in real time whether a new connection can be admitted into the network without violating quality of service (QoS) measures, such as the cell loss probability, for the new connection or the existing connections. The complexity of the ATM architectures, the low cell loss probabilities required by ATM networks (10^{-6} to 10^{-12}), and the unwieldiness of matching statistical traffic models to the traffic descriptors used by the CAC algorithm combined make the test of CAC algorithms difficult.

Monte Carlo (MC) simulation can be used to obtain accurate estimates of performance measures of a complex ATM, but is too slow to be used for CAC in real time. Even in non-real time environments, such as to test a CAC algorithm, MC simulation can be too slow for very low cell loss probabilities because of the long run times required to obtain accurate estimates. Hence, efficient estimators that would speed up simulations involving rare events of network (queueing) systems are in great needs.

Example 2 (Reliability of Markovian Systems) Consider a highly reliable Markovian system of components with small failure rates. Complex system interdependencies can be easily modeled in the Markov framework. These interdependencies may include failure propagation, i.e., failure of one component with certain probability leads to failure of other components, different modes of component failure, repair and operational dependence, component switch-over times, etc.

Suppose that the system has n distinct component-types. Type i has m_i identical components for functional and spare requirements. Let λ_i denote the failure rate for each of these components. That the system is highly reliable is modeled by letting

$$C_1\epsilon^{r_i} \leqslant \lambda_i \leqslant C_2\epsilon^{r_i}$$

for sufficiently small ϵ, where $r_i \geqslant 1$ and C_i's are positive constants for all i. The system is then analyzed as $\epsilon \to 0$.

A suitable mathematical model can be built as: Let $\{Y(t) : t \geqslant 0\}$ be a continuous-time Markov chain, where

$$Y(t) = (Y_1(t), Y_2(t), \ldots, Y_n(t), R(t))$$

with $Y_i(t)$ being the number of failed type-i components at time t, and $R(t)$ contains all information required to make $\{Y(t) : t \geqslant 0\}$ a Markov chain.

Let $\mathscr{A}$ be the state when all components are "up" and $\mathscr{R}$ be the set of failed states. The probability that the system starting from $\mathscr{A}$, hits $\mathscr{R}$ before returning to $\mathscr{A}$ is critical to efficient estimation of performance measures such as system unavailability and mean time to failure. To estimate the probability, one can simulate an embedded discrete-time Markov chain in $\{Y(t) : t \geqslant 0\}$ as both give identical result. Set $S_0 = \mathscr{A}$. Then, the process $\{S_0, S_1, \dots, S_T\}$ may be observed where $T = \inf\{n \geqslant 1 : S_n \in \mathscr{A} \cup \mathscr{R}\}$, and the probability of interest is $\mathsf{P}(S_T \in \mathscr{R})$.

Example 3 (Ruin Probability of Risk Processes) Consider a classic problem in insurance. Let u be the initial reserve, c be the premium rate, $X_1, X_2, \dots$ be i.i.d. claims with distribution G, and $\Lambda(t)$ be the number of claims by time t, where $\{\Lambda(t)\}$ is a Poisson process with rate λ. The *insurance surplus* at time t is

$$U(t) = u + ct - \sum_{i=1}^{\Lambda(t)} X_i, \qquad t \geqslant 0.$$

We call $\{U(t)\}$ a *risk process*, and define the *probability of ruin* as

$$\phi(u) = \mathsf{P}\Big(\inf_{t \geqslant 0}\{U(t)\} < 0\Big).$$

In both theory and practice, $\phi(u)$ is very small, the ruin is a rare event.

With a normalization, i.e., $c = 1$, an alternative expression of $\phi(u)$ that is useful in estimation is derived from a corresponding $M/G/1$ queue with arrival process $\{\Lambda(t)\}$ and service times X_i ([1; p. 399]). Let S_N denote the stationary work in the queue with $S_N = \sum_{i=1}^{N} X_{ei}$, where X_{ei} have the *equilibrium* distribution of X, and are i.i.d. and independent of N, the stationary number of customers in system under, e.g., preemptive last-come-first-serve and geometrically distributed. It is shown $\phi(u) = \mathsf{P}(S_N > u)$.

Then, $\phi(u)$ can be estimated by generating replicates of N and i.i.d. X_e's to form J i.i.d. replicates S_{Nj} of S_N. The estimator of $\phi(u)$ is thus

$$\frac{\sum_{j=1}^{J} I(S_{Nj} > u)}{J}.$$

In any of the above examples, if we estimate the probability of rare event α by Monte Carlo simulation, i.e., $Z = I\{A\}$, then we have Bernoulli sampling with variance $\sigma_Z^2 = \alpha(1-\alpha)$ that approaches 0 as $\alpha \to 0$. In other words, we have a small absolute error $\sigma_Z \sim \sqrt{\alpha}$ (We write $a(y) \sim b(y)$ if $a(y)/b(y) \to 1$ as $y \to \infty$).

However, the *relative* error

$$\frac{\sigma_Z}{\alpha} \sim \frac{1}{\sqrt{\alpha}} \to \infty \qquad \text{as} \ \ \alpha \to 0,$$

《应用概率统计》版权所有

《应用概率统计》版权所有

which is the relevant performance measure of the estimation. The reason why that the relative error is so important is the following:

Suppose we obtain an estimator of α of order 10^{-5} and a confidence interval of half-width 10^{-4}. The confidence interval may look narrow, but it does not tell whether α is of the magnitude 10^{-4}, 10^{-5} or even smaller. Another way to illustrate this point is by considering the sample size n that meets the required relative precision, say, 10%, in terms of the half-width of the 95% confidence interval, i.e., $1.965\sigma_Z/(\alpha\sqrt{n}) = 0.1$. One gets $n \sim 100 \times 1.965^2/\alpha$, inversely proportional to α, which increases without bound as $\alpha \to 0$. Therefore, the concern of rare-event simulation is on the efficiency and, consequently, on the accuracy.

§2. Criterion of Efficiency

Here we set up formal criterion for efficiency concepts.

Let $A(x)$ be a rare event with parameter x, $x \in (0, \infty)$ or $x \in N$. Assume that $\alpha(x) = \mathsf{P}\{A(x)\} \to 0$ as $x \to \infty$, and for each x, $Z(x)$ is an unbiased estimator for $\alpha(x)$.

We note that Jensen's inequality implies

$$\mathsf{E}[Z^2(x)] \geqslant \mathsf{E}^2[Z(x)] = \alpha^2(x),$$

a lower bound on the second moment. Hence, $Z(x)$ is said to be *asymptotically optimal*, or has *vanishing relative error* if

$$\lim_{x\to\infty} \frac{\mathsf{E}[Z^2(x)]}{\alpha^2(x)} = 1. \tag{1}$$

The second best performance that can be achieved in rare-event simulation is *bounded relative error*, i.e.,

$$\limsup_{x\to\infty} \frac{\mathsf{Var}\,(Z(x))}{\alpha^2(x)} < \infty. \tag{2}$$

Thus, an algorithm that meets such criterion requires bounded sample size even as $x \to \infty$.

A slightly weaker efficiency concept, namely, *logarithmic efficiency*, is $\mathsf{Var}\,(Z(x)) \to 0$ fast enough to have

$$\limsup_{x\to\infty} \frac{\mathsf{Var}\,(Z(x))}{\alpha^{2-\epsilon}(x)} = 0 \tag{3}$$

for all $\epsilon > 0$, or, equivalently,

$$\liminf_{x\to\infty} \frac{|\ln \mathsf{Var}\,(Z(x))|}{|\ln \alpha^2(x)|} \geqslant 1. \tag{4}$$

To see that (3) is weaker than (2), suppose $\alpha(x) \sim C\mathrm{e}^{-\gamma x}$. Estimator $Z(x)$ that meets (3) allows $\mathsf{Var}\,(Z(x))$ to decrease like $x^p\mathrm{e}^{-2\gamma x}$ for any $p > 0$.

Advantages of working with logarithmic efficiency rather than bounded relative error are: the difference is minor from a practical point of view, in some examples, logarithmic efficient estimators exist, whereas estimators with bounded relative error neither exist nor have been developed, and logarithmic efficiency is often much easier to verify.

Remark 4 To put the criterion in a common limit form, we say $\mathsf{Var}\,(Z(x)) = o(\alpha^2(x))$ if $Z(x)$ has vanishing relative error, $O(\alpha^2(x))$ if it has bounded relative error, or $o(\alpha^{2-\epsilon}(x))$ if it is logarithmic efficiency.

In practice, requirements (1)–(4) are often verified for $\mathsf{E}[Z^2(x)]$, the upper bound, instead of $\mathsf{Var}\,(Z(x))$.

Example 5 Let $N \sim \mathrm{Geo}(p)$ so that $\mathsf{P}(N = n) = p(1-p)^{n-1}$. Consider

$$\alpha = \mathsf{P}(N \leqslant m) = \sum_{n=1}^{m} p(1-p)^{n-1} = 1-(1-p)^m.$$

Now, if $p \to 0$, $\alpha \to 0$ for finite m, and $\alpha \sim mp$.

To estimate α, we use *importance sampling*, that is, with success probability $\widetilde{p}$ instead of p to form estimator

$$Z = I\{N \leqslant m\}\frac{p(1-p)^{N-1}}{\widetilde{p}(1-\widetilde{p})^{N-1}}.$$

Thus,

$$\begin{aligned}\mathsf{E}_{\widetilde{p}}(Z^2) &= \frac{p^2}{\widetilde{p}^2}\mathsf{E}_{\widetilde{p}}\Big[I\{N \leqslant m\}\Big(\frac{1-p}{1-\widetilde{p}}\Big)^{2(N-1)}\Big] = \frac{p^2}{\widetilde{p}}\sum_{n=1}^{m}\frac{(1-p)^{2(n-1)}}{(1-\widetilde{p})^{n-1}} \\ &= \frac{p^2}{\widetilde{p}}\frac{1-(1-p)^{2m}/(1-\widetilde{p})^m}{1-(1-p)^2/(1-\widetilde{p})}.\end{aligned}$$

A good candidate for $\widetilde{p}$ is obtained by the saddle-point argument, namely, equating $\mathsf{E}_{\widetilde{p}}(N)$ to m (Will be introduced in the next section). As $\mathsf{E}(N) = 1/p$, this means $\widetilde{p} = 1/m$. From $p \sim \alpha/m$ and the above, we then get

$$\mathsf{E}_{\widetilde{p}}(Z^2) \sim \frac{\alpha^2/m^2}{1/m}\frac{1-1/\mathrm{e}^{-1}}{1-1/(1-1/m)} \sim \frac{\alpha^2}{m}\frac{1/\mathrm{e}^{-1}-1}{1/m} = \alpha^2(\mathrm{e}-1).$$

So, this estimator has a bounded relative error.

More generally, the bounded relative error holds also by taking $\widetilde{p} = c/m$ because $\mathsf{E}_{\widetilde{p}}(Z^2) \sim \alpha^2(\mathrm{e}^c-1)/c^2$. The optimal $c = c^*$ is obtained by minimizing $(\mathrm{e}^c-1)/c^2$, which yields $c^* = 1.59$. However, the corresponding variance-reduction factor $(\mathrm{e}-1)c^{*2}/(\mathrm{e}^{c^*}-1) = 1.12$ is very small.

《应用概率统计》版权所有

《应用概率统计》版权所有

Let $X_1, X_2, \ldots$ be independent random replicates of X with $\mathsf{P}(X>0)>0$ and $\mathsf{E}(X)<0$. Also, let $S_n=\sum\limits_{i=1}^{n} X_i$ be the partial sum. To deal with S_n, the following simple lemma will often be useful:

Lemma 6 Let X_i have density f. If X_i's are simulated i.i.d. with density $\widetilde{f}$, L is the likelihood ratio and $Z(x)=LI\{S_n>x\}$, then

$$\widetilde{\mathsf{E}}[Z^2(x)]=\widehat{c}^{\,-n}\widehat{\mathsf{P}}(S_n>x), \qquad \text{where } \widehat{c}^{\,-1}=\int f^2/\widetilde{f} \text{ and } \widehat{f}=\widehat{c}f^2/\widetilde{f}.$$

Proof

$$\begin{aligned}
\widetilde{\mathsf{E}}[Z^2(x)] &= \widetilde{\mathsf{E}}\Big[\Big(\prod_{i=1}^{n}\frac{f(X_i)}{\widetilde{f}(X_i)}\Big)^2 I\{S_n>x\}\Big] \\
&= \int\cdots\int_{x_1+x_2+\cdots+x_n>x}\frac{f^2(x_1)}{\widetilde{f}^2(x_1)}\cdots\frac{f^2(x_n)}{\widetilde{f}^2(x_n)}\widetilde{f}(x_1)\cdots\widetilde{f}(x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \widehat{c}^{\,-n}\int\cdots\int_{x_1+x_2\cdots+x_n>x}\widehat{f}(x_1)\cdots\widehat{f}(x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \widehat{c}^{\,-n}\widehat{\mathsf{P}}(S_n>x). \qquad \Box
\end{aligned}$$

The next example investigates the efficiency of estimating $\alpha(x)=\mathsf{P}(S_n>x)$ for large x, that is, $\alpha(x)$ is small.

Example 7 Suppose that X_i's are independent and $\sim\exp(1)$. Since $S_n=\sum\limits_{i=1}^{n}X_i$ is n-Erlang, i.e., $\mathrm{Gamma}(n,1)$, we have

$$\alpha(x)\sim\frac{x^{n-1}\mathrm{e}^{-x}}{(n-1)!} \qquad \text{as } x\to\infty. \tag{5}$$

Consider the important density $\widetilde{f}(x)=\lambda\mathrm{e}^{-\lambda x}$. Then,

$$\widehat{c}^{\,-1}=\frac{1}{\lambda}\int_0^\infty \mathrm{e}^{-(2-\lambda)y}\mathrm{d}y=\frac{1}{\lambda(2-\lambda)}, \qquad \widehat{f}(x)=\widehat{\lambda}\mathrm{e}^{-\widehat{\lambda}x},$$

where $\widehat{\lambda}=2-\lambda$. It seems reasonable to choose λ such that $\widetilde{\mathsf{E}}(S_n)=n/\lambda$ is of order x. So we let $\lambda=c/x$ and get from Lemma 6

$$\begin{aligned}
\widetilde{\mathsf{E}}[Z^2(x)] &= \widehat{c}^{\,-n}\widehat{\mathsf{P}}(S_n>x)=\widehat{c}^{\,-n}\int_x^\infty\frac{(2-\lambda)^n y^{n-1}}{(n-1)!}\mathrm{e}^{-(2-\lambda)y}\mathrm{d}y \\
&= \widehat{c}^{\,-n}\int_{(2-\lambda)x}^\infty\frac{z^{n-1}}{(n-1)!}\mathrm{e}^{-z}\mathrm{d}z\sim\widehat{c}^{\,-n}\frac{((2-\lambda)x)^n}{(n-1)!}\mathrm{e}^{-(2-\lambda)x} \\
&= \frac{\mathrm{e}^{c}x^{2n-1}}{(2-\lambda)c^n(n-1)!}\mathrm{e}^{-2x},
\end{aligned}$$

which together with (5) show that the important sampling yields logarithmic efficiency but not bounded relative error (then the power of x should have been $2n-2$).

《应用概率统计》版权所有

In this report, we will demonstrate the rare-event simulation mainly via the estimation of $\alpha(x) = \mathsf{P}(S_n > x)$. Notice that not only $\alpha(x)$, but also the estimator's asymptotic variance depend on the tail distribution of X. That is to say, to estimate $\alpha(x)$ efficiently, we need to first investigate the tail property of X. A common approach is to classify X as having either a *light* or a *heavy* tail. We will discuss estimators developed for either case.

§3. Rare Events with Light Tails

We say that the tail of a distribution is light if it decays at an exponential rate or faster. More precisely, a random variable X has a light-tail distribution F if moment generating function (m.g.f.)

$$\widehat{F}(s) = \mathsf{E}(\mathrm{e}^{sX}) = \int \mathrm{e}^{sx} F(\mathrm{d}x)$$

is finite for some $s > 0$. The most efficient estimator for a light-tail rare event is based on *exponential change of measure* (ECM). It deserves a brief review.

For a distribution F and its m.g.f. $\widehat{F}$, let $\kappa(\theta) \equiv \ln \widehat{F}(\theta)$ be the cumulant generating function (c.g.f), i.e., the logarithm of the m.g.f. Then, the ECM of F is defined as

$$F_\theta(\mathrm{d}x) \equiv \frac{\mathrm{e}^{\theta x}}{\widehat{F}[\theta]} F(\mathrm{d}x) = \mathrm{e}^{\theta x - \kappa(\theta)} F(\mathrm{d}x),$$

which is a distribution and, most importantly, in many cases reserves the form as that of F only except the parameter. A few one-dimensional examples are below:

- If $F \sim \exp(\lambda)$, then $F_\theta \sim \exp(\lambda - \theta)$, where $-\infty < \theta \leqslant \lambda$;
- If $F \sim N(\mu, \sigma^2)$, then $F_\theta \sim N(\mu + \theta\sigma^2, \sigma^2)$;
- If $F \sim \mathrm{binomial}(N, p)$, then $F_\theta \sim \mathrm{binomial}(N, p\mathrm{e}^\theta/(1 - p + p\mathrm{e}^\theta))$;
- If $F \sim \mathrm{Poisson}(\lambda)$, then $F_\theta \sim \mathrm{Poisson}(\lambda \mathrm{e}^\theta)$.

In the change, likelihood ratio $(\mathrm{d}\mathsf{P}/\mathrm{d}\mathsf{P}_\theta)_n$ for i.i.d. $X_1, X_2, \ldots, X_n \sim X$ is

$$L_{n,\theta} \equiv \prod_{k=1}^{n} \frac{\widehat{F}[\theta]}{\mathrm{e}^{\theta X_k}} = \mathrm{e}^{-\theta S_n} \widehat{F}[\theta]^n = \mathrm{e}^{-\theta S_n + n\kappa(\theta)}.$$

Thus, if Y_n is $\sigma(X_1, X_2, \ldots, X_n)$-measurable, we have

$$\mathsf{E}(Y_n) = \mathsf{E}_\theta(Y_n L_{n,\theta}) = \mathsf{E}_\theta(Y_n \mathrm{e}^{-\theta S_n + n\kappa(\theta)}).$$

Note that the minimizer of $\widehat{F}[\theta]$ and/or $\kappa(\theta)$ is the solution of $\widehat{F}'[\theta] = \kappa'(\theta) = 0$, called γ_0, and the nonzero solution of $\kappa(\theta) = 0$ is also the one of $\widehat{F}[\theta] = 1$, called γ.

《应用概率统计》版权所有

When performing ECM, the changed drift is

$$\mu_\theta \equiv \mathsf{E}_\theta(X) = \mathsf{E}(X\mathrm{e}^{\theta X}/\widehat{F}[\theta]) = \frac{\widehat{F}'[\theta]}{\widehat{F}[\theta]} = \kappa'(\theta). \tag{6}$$

Hence, $\mu_\theta < 0$ when and only when $\theta < \gamma_0$, and $\mu_{\gamma_0} = 0$.

In particular, when ECM is used to make $I_\theta\{S_n \geqslant x\}$ more likely to occur than $I\{S_n \geqslant x\}$, a common choice of θ is to let $\mathsf{E}_\theta(S_n) = x$ (analytically, $\mathsf{E}_\theta(S_n) = n\kappa'(\theta)$). This θ is often referred to as a *saddle point*.

Remark 8 The ECM also appears in statistical hypothesis testing for the *estimation of p-values* (but with a different name: one-parameter exponential family F_θ): Define

$$F_\theta(\mathrm{d}x) = \mathrm{e}^{\theta x}F(\mathrm{d}x)/\widehat{F}(\theta),$$

and consider testing $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$.

If we observe $X_i = x_i$, $i = 1, 2, \ldots, n$, then the p-value is given by $\mathsf{P}_0(S_n > s_n)$ where $S_n = \sum\limits_{i=1}^{n} X_i$ and $s_n = \sum\limits_{i=1}^{n} x_i$. Here $\mathsf{P}_0(S_n > s_n)$ may be approximated by the saddle-point method.

3.1 Siegmund's Algorithm

Assume that F is not concentrated on $(-\infty, 0]$, i.e., $\overline{F}(0) > 0$, and $\mathsf{E}(X) < 0$. Consider the problem of estimating

$$\alpha(x) = \mathsf{P}\{\tau(x) < \infty\}, \qquad \text{where} \ \ \tau(x) = \inf\{n : S_n > x\}.$$

Since $\mathsf{E}(X) < 0$, we can employ importance sampling via ECM to get

$$\alpha(x) = \mathsf{E}_\theta[I_{\tau(x)<\infty}L_{\tau(x),\theta}] = \mathsf{E}_\theta[I_{\tau(x)<\infty}\mathrm{e}^{-\theta S_{\tau(x)}+\tau(x)\kappa(\theta)}].$$

To choose appropriate θ, we need to ensure that $\mathsf{P}_\theta(\tau(x) < \infty) = 1$, i.e., $\mathsf{E}_\theta(X) > 0$. By (6), we have $\theta \geqslant \gamma_0$, and with such a θ, $\alpha(x) = \mathsf{E}_\theta[L_{\tau(x),\theta}]$ such that a Monte Carlo simulation can be conducted with $Z(x) = L_{\tau(x),\theta}$.

Furthermore, γ is the optimal parameter provided the existence, which is assured by $\widehat{F}[\theta] = 1$ having a solution and $\widehat{F}'[\gamma] < \infty$, which, in turn, is assured by the existence of exponential moments. Then, we have

$$\alpha(x) = \mathsf{E}_\gamma[\mathrm{e}^{-\gamma S_{\tau(x)}}] = \mathrm{e}^{-\gamma x}\mathsf{E}_\gamma[\mathrm{e}^{-\gamma\epsilon(x)}],$$

where $\epsilon(x) = S_{\tau(x)} - x$ is the overshoot.

Two examples below indicate what the change of measure looks like.

Example 9 Let $F \sim N(-\mu, 1)$ with $\mu > 0$. Then $\widehat{F}[s] = \exp\{-\mu s + s^2/2\}$ so that γ is the solution of $\gamma^2/2 - \mu\gamma = 0$, which in view of $\gamma > 0$ implies $\gamma = 2\mu$. Consequently,

$$\widehat{F}_\gamma[s] = \widehat{F}[s+\gamma] = \exp\{\mu s + s^2/2\},$$

which shows that $F_\gamma \sim N(\mu, 1)$.

Example 10 Consider the $M/M/1$ queue with arrival rate λ and service rate μ, $\lambda < \mu$. Let $X = S - T$ be independent difference of service and interarrival times. Then, $\widehat{F}[\theta] = 1$ means

$$1 = \mathsf{E}(\mathrm{e}^{\gamma S})\mathsf{E}(\mathrm{e}^{-\gamma T}) = \frac{\lambda}{\lambda+\gamma}\frac{\mu}{\mu-\gamma},$$

which has the positive solution $\gamma = \mu - \lambda$. So, we get

$$\widehat{F}_\gamma[s] = \widehat{F}[\gamma+s] = \frac{\lambda}{\lambda-s}\frac{\mu}{\mu+s},$$

which shows that the changed measure corresponds to an $M/M/1$ queue with arrival rate μ and service rate λ.

We are now ready to present the main results.

Theorem 11 The algorithm given by $Z(x) = \mathrm{e}^{-\gamma x}\mathrm{e}^{-\gamma\epsilon(x)}$ (simulated from F_γ) has bounded relative error.

Proof We first note that the process $\{\epsilon(x), x \geqslant 0\}$ is regenerative with regeneration occurs at each partial maximum of $\{S_n\}$. Assume that F is aperiodic in the lattice case or nonlattice otherwise. Then, $\epsilon(x) \to \epsilon(\infty)$ and

$$\mathsf{E}_\gamma[\mathrm{e}^{-\gamma\epsilon(x)}] \to \mathsf{E}_\gamma[\mathrm{e}^{-\gamma\epsilon(\infty)}] \equiv C, \qquad \text{as } x \to \infty.$$

It follows that

$$\alpha(x) \sim C\mathrm{e}^{-\gamma x},$$

a celebrated result referred to as the *Cramér-Lundberg approximation.*

Now, we have

$$\mathsf{E}_\gamma[Z^2(x)] = \mathrm{e}^{-2\gamma x}\mathsf{E}_\gamma(\mathrm{e}^{-2\gamma\epsilon(x)}) \sim C_1\mathrm{e}^{-2\gamma x},$$

where $C_1 = \mathsf{E}_\gamma(\mathrm{e}^{-2\gamma\epsilon(\infty)})$, and hence,

$$\mathsf{Var}_\gamma(Z(x)) \sim C_1\mathrm{e}^{-2\gamma x} - (C\mathrm{e}^{-\gamma x})^2 \sim C_2\mathrm{e}^{-2\gamma x},$$

where $C_2 = C_1 - C^2 > 0$ from Jensen's inequality.

《应用概率统计》版权所有

《应用概率统计》版权所有

Thus, the relative error is

$$\frac{\sqrt{\mathsf{Var}_{\gamma}(Z(x))}}{\alpha(x)} \sim \frac{\mathrm{e}^{-\gamma x}\sqrt{C_2}}{C\mathrm{e}^{-\gamma x}} = \frac{\sqrt{C_2}}{C},$$

which does not increase in x. □

We note that the expected time to generate one replication by the algorithm $\mathsf{E}_{\gamma}[\tau(x)]$ is $O(x)$. Also, the change of measure in Theorem 11 is unique in yielding at least logarithmic efficiency. For a proof, see [5; p. 166].

3.2 Efficient Simulation of $\mathsf{P}\{S_n > n(\mu+\epsilon)\}$

Consider the random walk $S_n = \sum\limits_{i=1}^{n} X_i$ again, but the sign of $\mu = \mathsf{E}(X)$ is unimportant because the rare event of interest now is $A(n) = \{S_n > n(\mu+\epsilon)\}$ with $\epsilon > 0$. That is, the parameter of the problem is the discrete n.

Since $\alpha(n) = \mathsf{P}\{A(n)\} \to 0$ as $n \to \infty$ from the LLN, the event $A(n)$ is rare indeed. By ECM, we have

$$Z(n) = \mathrm{e}^{-\theta S_n + n\kappa(\theta)} I\{S_n > n(\mu+\epsilon)\}.$$

The appropriate choice of θ is by the saddle-point method as

$$\mathsf{E}_{\theta}X = \kappa'(\theta) = \mu + \epsilon, \tag{7}$$

which implies $\theta > 0$ owing to the strictly convexity and then increasing of κ. Moreover, we have $I \equiv \theta(\mu+\epsilon) - \kappa(\theta) > 0$.

Theorem 12 The ECM given by (7) is logarithmically efficient, and is the only importance distribution $\widetilde{F}$ with this property.

As for Siegmund's algorithm, the proof is a small variant of the standard estimates for obtaining the asymptotic of $\alpha(n)$ itself. So, we omit it.

This change of measure yields the following famous bound:

Lemma 13 (Chernoff Bound) $\alpha(n) \leqslant \mathrm{e}^{-nI}$.

Proof Using the basic likelihood ratio identity and $\theta > 0$, we have

$$\begin{aligned}\alpha(n) &= \mathsf{E}_{\theta}[I\{A(n)\}L_{n,\theta}] = \mathsf{E}_{\theta}[I\{S_n > n(\mu+\epsilon)\}\mathrm{e}^{-\theta S_n + n\kappa(\theta)}] \\ &= \mathrm{e}^{-nI}\mathsf{E}_{\theta}[I\{S_n > n(\mu+\epsilon)\}\mathrm{e}^{-\theta(S_n - n(\mu+\epsilon))}] \leqslant \mathrm{e}^{-nI}.\end{aligned}$$ □

《应用概率统计》版权所有

3.3 Compound Poisson Sums

Consider the random sum $S_N = \sum\limits_{i=1}^{N} X_i$, where X_i's are i.i.d. and nonnegative with distribution F, and N is an independent Poisson random variable with mean λ. As discussed in Example 3, we often need to evaluate $\mathsf{P}(S_N > x)$ when x is large in insurance risk and queueing theory.

By a variant of the previous analysis, we shall show that ECM is again logarithmically efficient given that F is light-tail and satisfies some regularity conditions. Let $f(x)$ denote the density of F. The regularity conditions are:

A. f is gamma-like, i.e., $f(x) \sim c_1 x^{\alpha-1}\mathrm{e}^{-\delta x}$, as $x \to \infty$.

The ones that meet A are, for examples, exponential distributions, phase-type distributions and inverse Gaussian distributions.

B. f is log-concave, or, more generally, $f(x) = q(x)\mathrm{e}^{-h(x)}$, where $q(x)$ is bounded away from 0 and ∞, and $h(x)$ is concave in $[x_0, x^*)$, where $x^* = \sup\{x : f(x) > 0\}$. Furthermore, $\int_0^\infty f(x)^a \mathrm{d}x < \infty$ for some $1 < a < 2$.

The distributions that satisfy B have finite support or with a density not too far from e^{-x^α} with $\alpha > 1$.

Define the c.g.f. of S_N be $\varphi(\beta) = \ln \mathsf{E}(\mathrm{e}^{\beta S_N})$. Then, by conditioning on N, we get

$$\varphi(\beta) = \lambda(\widehat{F}[\beta] - 1),$$

where $\widehat{F}[\beta]$ is the m.g.f. of F. Under ECM, the c.g.f. becomes

$$\varphi_\theta(\beta) = \varphi(\beta+\theta) - \varphi(\theta) = \lambda(\widehat{F}[\beta+\theta] - \widehat{F}[\theta]) = \lambda\widehat{F}[\theta](\widehat{F}_\theta[\beta] - 1),$$

where $F_\theta(\mathrm{d}x) = \mathrm{e}^{\theta x}F(\mathrm{d}x)/\widehat{F}[\theta]$ and θ is again determined by the saddle-point argument as the solution of $\mathsf{E}_\theta(S_N) = x$, i.e., by $x = \varphi'(\theta) = \lambda\widehat{F}'[\theta]$.

Theorem 14 If either of A or B holds, the estimator (simulated from F_θ)

$$Z(x) = \mathrm{e}^{-\theta S_N + \varphi(\theta)} I\{S_N > x\}$$

for $\alpha(x)$ is logarithmically efficient.

The proof can be found in [5; p. 171].

《应用概率统计》版权所有

§4. Rare Events with Heavy Tails

We say that F (or X) is *heavy-tail* if for all $x \geqslant 0$, $\overline{F}(x) > 0$ and

$$\lim_{y\to\infty} \mathsf{P}(X > x + y \,|\, X > y) = \lim_{y\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(y)} = 1.$$

Intuitively, this means that if X ever exceeds a large value, then it is likely to exceed any larger value as well. Its definition may also be expressed as

$$\overline{F}(x+y) \sim \overline{F}(y), \qquad \text{for all } \ x \geqslant 0.$$

A popular class of heavy-tail distributions is the so-call *subexponential*, which means that

$$\frac{\mathsf{P}(X_1 + X_2 > x)}{\mathsf{P}(X_1 > x)} \to 2, \qquad x \to \infty,$$

where X_1, X_2 are i.i.d. with distribution F. This can be shown by induction to be equivalent to

$$\frac{\mathsf{P}(X_1 + X_2 + \cdots + X_n > x)}{\mathsf{P}(X_1 > x)} \to n, \qquad x \to \infty \tag{8}$$

for all $n \geqslant 2$.

In this report, we will concentrate on the following two examples in the subexponential class:

Regular Variation: $\overline{F}(x) = L(x)/x^{\delta}$, where $\delta > 0$ and L is slowing varying, i.e., $L(tx)/L(x) \to 1$ as $x \to \infty$ for any fixed $t > 0$. The most prominent example is the Pareto distribution, with tail $1/(1+x)^{\delta}$, $0 < \delta$.

This class of distributions is initially used to model the distribution of wealth, and has become an important tool for telecommunication model. In fact, there is a growing interest nowadays in providing internet to mobile users, and internet traffic statistics show that the session duration of call holding times are Pareto distributed.

Weibull Distribution: $\overline{F}(x) = \mathrm{e}^{-cx^{\beta}}$, $0 < \beta < 1$. If $\beta = 1$, it becomes exponential.

This distribution is often used in survival analysis, as an excellent model choice for describing the life of manufactured objects, and in insurance risk analysis for modeling excessive claim sizes. Indeed, subexponentiality is considered as a synonym for heavy tail in insurance mathematics. Furthermore, among the class of subexponential distributions, Weibull with $\beta \in (0, 1)$ provides adequate candidates for modeling large claims as it allows greater flexibility for data fitting.

《应用概率统计》版权所有

The development of rare-event simulation with heavy tails was postponed until recently. It is because that the main ideas of efficient simulation from light tails cannot be applied to. Hence, new ideas are presented in the following subsections and in need for more.

4.1 Conditional Estimators

To estimate $\alpha(x) = \mathsf{P}(S_n > x)$ as before, a direct MC estimator is

$$Z_1(x) = I\{S_n > x\},$$

and a *conditional estimator* is $\mathsf{P}(S_n > x \,|\, \mathscr{F})$ with $\mathscr{F} \subset \sigma(X_1, X_2, \ldots, X_n)$. Although, by the conditional variance formula, this estimator always yields variance reduction, our task is to find a proper $\mathscr{F}$ so that the reduction is so substantial that the estimator is, hopefully, logarithmically efficient, or even has a bounded relative error.

The first and obvious approach is to condition on $X_1, X_2, \ldots, X_{n-1}$, which leads to

$$Z_2(x) = \mathsf{P}(S_n > x \,|\, X_1, X_2, \ldots, X_{n-1}) = \overline{F}(x - S_{n-1}),$$

that is, only $X_1, X_2, \ldots, X_{n-1}$ are generated.

Having smaller variance than $Z_1(x)$, $Z_2(x)$ does not, however, yield meaningful improvement; its variance is of the same order as that of $\overline{F}(x)$:

$$\mathsf{E}[Z_2^2(x)] \geqslant \mathsf{E}[\overline{F}^2(x - S_{n-1}) I\{X_1 > x\}] = \mathsf{P}(X_1 > x) = \overline{F}(x).$$

Consequently,

$$\liminf_{x\to\infty} \frac{|\ln \mathsf{Var}\,(Z_2(x))|}{|\ln \alpha^2(x)|} = \lim_{x\to\infty} \frac{|\ln \overline{F}(x)|}{2|\ln n + \ln \overline{F}(x)|} = \frac{1}{2}.$$

The reason that it does not work well is that the probability of one single large X_i is relatively too big, which leads to the idea of discarding the largest of the X_i's and considering only the remaining ones. Thus, we generate $X_1, X_2, \ldots, X_n$ and form the order statistics $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$. We then throw away $X_{(n)}$, let $S_{(n-1)} = X_{(1)} + X_{(2)} + \cdots + X_{(n-1)}$ and have

$$\begin{aligned} Z_3(x) &= \mathsf{P}(S_n > x \,|\, X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}) = \mathsf{P}(X_{(n)} + S_{(n-1)} > x \,|\, X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}) \\ &= \frac{\overline{F}((x - S_{(n-1)}) \vee X_{(n-1)})}{\overline{F}(X_{(n-1)})}. \end{aligned}$$

Theorem 15 When the tail is regularly varying, $Z_3(x)$ is logarithmically efficient.

《应用概率统计》版权所有

Sketch of the proof We first bound the density of $X_{(n-1)}$ as

$$f_{(n-1)}(y) = n(n-1)F^{n-2}(y)\overline{F}(y)f(y) \leqslant c\overline{F}(y)f(y).$$

Then, we evaluate $\mathsf{E}[Z_3^2(x)]$ separately over the region:

$$X_{(n-1)} \leqslant \frac{x}{n}, \quad \frac{x}{n} < X_{(n-1)} \leqslant \frac{x}{2} \quad \text{and} \quad X_{(-1)} > \frac{x}{2}. \qquad \square$$

In a short time, an equally simple conditional estimator is proposed in [4] that improves $Z_3(x)$. Because a major portion of $Z_3(x)$'s variance is contributed from its denominator, the idea, hence, is to avoid having the probability of any given event in the denominator. It is achieved by partition according to which X_i is the largest, i.e., $X_i = X_{(n)}$, and condition on other X_j's. Since, by symmetry,

$$\mathsf{P}(S_n > x) = n\mathsf{P}(S_n > x, X_n = X_{(n)}),$$

they construct the estimator in [4] as

$$Z_4(x) = n\mathsf{P}(S_n > x, X_n = X_{(n)} \,|\, X_1, X_2, \ldots, X_{n-1}) = n\overline{F}(X_{(n-1)} \vee (x - S_{n-1})).$$

Theorem 16 Estimator $Z_4(x)$ has bounded relative error in the regular varying case, and is logarithmically efficient in the Weibull case for $\beta < \ln(3/2)/\ln 2 = 0.585$.

Proof Here we only provide the proof for the regularly varying case, and refer to [4] for the Weibull case.

If $X_{(n-1)} \leqslant x/n$, then $S_{(n-1)} \leqslant (n-1)x/n$ and, consequently, $X_{(n-1)} \vee (x - S_{n-1}) \geqslant x/n$. Thus,

$$\frac{\mathsf{E}[Z_4^2(x)]}{\overline{F}^2(x)} \leqslant \frac{n^2\overline{F}^2(x/n)}{\overline{F}^2(x)} = \frac{n^2L^2(x/n)/(x/n)^{2\delta}}{L^2(x)/x^{2\delta}} = \frac{n^{2+2\delta}L^2(x/n)}{L^2(x)} \sim n^{2+2\delta}.$$

Now, from (8), $\alpha(x) \sim n\overline{F}(x)$ and the proof is complete. $\square$

It is remarked that $Z_4(x)$ is the first rare-event estimator that achieves the criterion of bounded relative error.

The not-so-good performance for not-so-heavy tails is due to the possibility of large sample size required. We illustrate the reason by an extreme case: the minimum of $\max\{X_{(n-1)}, x - S_{(n-1)}\}$ occurs when $X_1 = X_2 = \cdots = X_{n-1} = x/n$, and consequently,

$$\max_{X_1, X_2, \ldots, X_{n-1}} Z_4(x) = n\overline{F}(x/n).$$

Thus, for large n the estimator can be large.

Finally, for the case of random sum S_N, it is suggested in [4] that one can either use N as a control variate or stratify N to obtain further substantial variance reduction. See also [8] for a theoretical treatment on this issue.

《应用概率统计》版权所有

4.2 Control Variate Estimators

Recall that the variance-reduction estimator for $\mathsf{E}(X)$ with a *control variate* Y, with $\mathsf{E}(Y)$ known, is $X - c[Y - \mathsf{E}(Y)]$, where optimal $c = \mathsf{Cov}\,(X, Y)/\mathsf{Var}\,(Y)$.

From (8), i.e., the occurrence of $S_n > x$ is likely due to the occurrence of $X_{(n)} > x$, it suggests to use $I\{X_{(n)} > x\}$ as a control variate for estimating $\mathsf{P}(S_n > x)$ when X_i is subexponential.

The most straightforward one is

$$Z_5(x) = I\{S_n > x\} - c[I\{X_{(n)} > x\} - \mathsf{P}\{X_{(n)} > x\}].$$

One would naturally expect that the heavier the tail is, the better its performance is. Indeed, we have the following result from [9]:

Theorem 17 When F is regularly varying with $0 < \delta < 1$, $Z_5(x)$ has a bounded relative error.

We first state a few lemmas that are useful in proving the theorem.

Lemma 18 ([11]) For $\overline{F}$ being regularly varying with $0 < \delta \leqslant 1$,

$$\lim_{x\to\infty} \frac{\mathsf{P}(S_n > x) - n\overline{F}(x)}{f(x)\displaystyle\int_0^x \overline{F}(y)\mathrm{d}y} = \begin{cases} c_\delta \dfrac{n(n-1)}{2}, & \text{if } 0 < \delta < 1; \\ n(n-1), & \text{if } \delta = 1, \end{cases}$$

where c_δ is a constant depending on δ.

Lemma 19 ([16; p. 62]) If $L_1(x)$ is a slowly varying function and locally bounded in $[x_0, \infty)$ for some $x_0 > 0$, then for $\delta > 0$

$$\int_x^\infty y^{-(\delta+1)} L_1(y)\mathrm{d}y = x^{-\delta} L_2(x),$$

where $L_2(x)$ is a slowly varying function of x at ∞ and $\lim\limits_{x\to\infty} L_1(x)/L_2(x) = \delta$. If $L_1(y)/y$ is integrable, then the result also holds for $\delta = 0$.

Lemma 20 ([15; p. 25]) Suppose that $\overline{F}$ is regularly varying with parameter $\delta \leqslant 1$. Then $\int_0^x \overline{F}(t)\mathrm{d}t$ is also regularly varying with parameter $1 - \delta$ and

$$\lim_{x\to\infty} \frac{x\overline{F}(x)}{\displaystyle\int_0^x \overline{F}(t)\mathrm{d}t} = 1 - \delta.$$

Proof of Theorem 17 We first write

$$\mathsf{Var}\,(Z_5(x)) = \mathsf{Var}\,(I\{S_n > x\}) - \frac{\mathsf{Cov}^{\,2}(I\{S_n > x\}, I\{X_{(n)} > x\})}{\mathsf{Var}\,(I\{X_{(n)} > x\})}$$

《应用概率统计》版权所有

$$\begin{aligned} &= \mathsf{P}(S_n > x)\mathsf{P}(S_n < x) - \frac{\mathsf{P}^2(X_{(n)} > x)\mathsf{P}^2(S_n < x)}{\mathsf{P}(X_{(n)} > x)\mathsf{P}(X_{(n)} < x)} \\ &= \mathsf{P}(S_n < x)\Big[\mathsf{P}(S_n > x) - \frac{\mathsf{P}(X_{(n)} > x)\mathsf{P}(S_n < x)}{\mathsf{P}(X_{(n)} < x)}\Big]. \end{aligned}$$

Furthermore, with $\mathsf{P}(S_n < x) = \mathsf{P}(X_{(n)} < x) - \mathsf{P}(S_n > x, X_{(n)} < x)$ and some algebraic manipulation, we derive

$$\mathsf{Var}\,(Z_5(x)) = \mathsf{P}(S_n > x, X_{(n)} < x)\frac{\mathsf{P}(S_n < x)}{\mathsf{P}(X_{(n)} < x)}. \tag{9}$$

Now, with $\mathsf{P}(S_n > x, X_{(n)} < x) = \mathsf{P}(S_n > x) - \mathsf{P}(X_{(n)} > x)$ and

$$\mathsf{P}(X_{(n)} > x) = n\overline{F}(x) - \frac{n(n-1)}{2}\overline{F}^2(x) + o\big(\overline{F}^2(x)\big),$$

we have

$$\mathsf{P}(S_n > x, X_{(n)} < x) = \mathsf{P}(S_n > x) - n\overline{F}(x) + \frac{n(n-1)}{2}\overline{F}^2(x) + o\big(\overline{F}^2(x)\big). \tag{10}$$

Let $\overline{F}(x) = L_2(x)/x^{\delta}$ and $f(x) = L_1(x)/x^{\delta+1}$. We get $\lim\limits_{x\to\infty} xf(x)/\overline{F}(x) = \delta$ from Lemma 19. Thus, by Lemmas 18 and 20,

$$\begin{aligned} \lim_{x\to\infty}\frac{\mathsf{P}(S_n > x) - n\overline{F}(x)}{n^2\overline{F}^2(x)} &= \lim_{x\to\infty}\frac{\mathsf{P}(S_n > x) - n\overline{F}(x)}{n^2 f(x)\int_0^x\overline{F}(y)\mathrm{d}y}\,\frac{xf(x)}{\overline{F}(x)}\,\frac{\displaystyle\int_0^x\overline{F}(y)\mathrm{d}y}{x\overline{F}(x)} \\ &= \frac{c_\delta n(n-1)}{2n^2}\,\frac{\delta}{1-\delta}. \end{aligned}$$

Using (9) and (10), we have

$$\begin{aligned} &\lim_{x\to\infty}\frac{\mathsf{Var}\,(Z_5(x))}{\mathsf{P}^2(S_n > x)} = \lim_{x\to\infty}\frac{\mathsf{P}(S_n > x, X_{(n)} < x)\mathsf{P}(S_n < x)}{\mathsf{P}^2(S_n > x)\mathsf{P}(X_{(n)} < x)} \\ &= \lim_{x\to\infty}\frac{\Big[\mathsf{P}(S_n > x) - n\overline{F}(x) + \dfrac{n(n-1)}{2}\overline{F}^2(x) + o\big(\overline{F}^2(x)\big)\Big]}{n^2\overline{F}^2(x)}\,\frac{n^2\overline{F}^2(x)}{\mathsf{P}^2(S_n > x)}\,\frac{\mathsf{P}(S_n < x)}{\mathsf{P}(X_{(n)} < x)} \\ &= \frac{1-\delta+c_\delta\delta}{1-\delta}\,\frac{n-1}{2n}, \end{aligned}$$

because the limits of $n^2\overline{F}^2(x)/\mathsf{P}^2(S_n > x)$ and $\mathsf{P}(S_n < x)/\mathsf{P}(X_{(n)} < x)$ are both 1 as $x\to\infty$. So, $Z_5(x)$ has a bounded relative error. □

Hence, for regularly varying with $0 < \delta < 1$, $Z_5(x)$ is as efficient as $Z_4(x)$ (unfortunately, only in theory). However, for $\delta \geqslant 1$, the relative error of $Z_5(x)$ is no longer bounded. Nevertheless, the result is obtained in [9]:

《应用概率统计》版权所有

Theorem 21 When F is regularly varying with $\delta = 1$, estimator $Z_5(x)$ is logarithmic efficient.

Also shown in [9] is the following:

Corollary 22 For N being a random variable with finite $\mathsf{E}(N^2)$ and F being regularly varying, $Z_5(x)$ has bounded relative error with $0 < \delta < 1$, and is logarithmically efficient with $\delta = 1$.

For regularly varying with $\delta > 1$, estimator $Z_5(x)$ is not logarithmic efficient. Neither it is for the Weibull claim size.

We now present the numerical performance of these estimators for S_N, where $N \sim \text{Geo}(1-\rho)$. The magnitudes of the ruin probability considered are 10^{-2}, 10^{-3} and 10^{-4}, and the sample size is 10^7. In Tables 1–3 below, we show the relative error of estimators $Z_3(x)$, $Z_4(x)$ and $Z_5(x)$.

Table 1 Results for the Pareto claim size with $\delta = 0.5$

ρ	$\alpha(x)$	$Z_3(x)$	$Z_4(x)$	$Z_5(x)$
0.25	9.995×10^{-3}	1.140	0.1272	0.4857
0.25	1.000×10^{-3}	1.569	0.0134	0.5237
0.25	1.000×10^{-4}	1.927	0.0014	0.4999
0.5	9.997×10^{-3}	1.785	0.0449	0.6948
0.5	3.020×10^{-3}	2.078	0.0446	0.7172
0.5	9.999×10^{-5}	2.618	0.0005	0.6159
0.75	9.998×10^{-3}	1.611	0.0149	0.8587
0.75	1.000×10^{-3}	2.186	0.0015	0.9337
0.75	9.998×10^{-5}	2.459	0.0001	0.7501

Table 2 Results for the Pareto claim size with $\delta = 1.0$

ρ	$\alpha(x)$	$Z_3(x)$	$Z_4(x)$	$Z_5(x)$
0.25	1.068×10^{-2}	1.4255	0.3954	1.3048
0.25	1.011×10^{-3}	1.8422	0.0766	1.7344
0.25	1.001×10^{-4}	2.1199	0.0082	1.9813
0.5	1.0955×10^{-2}	2.1809	0.2336	2.0933
0.5	1.0136×10^{-3}	2.7115	0.0262	2.6679
0.5	1.0015×10^{-4}	2.9479	0.0028	2.8243
0.75	1.0816×10^{-2}	2.7739	0.0848	2.8266
0.75	1.0162×10^{-3}	3.4228	0.0090	3.5411
0.75	1.0015×10^{-4}	3.9802	0.0010	3.5914

Table 3 Results for the Pareto claim size with $\delta = 1.5$

ρ	$\alpha(x)$	$Z_3(x)$	$Z_4(x)$	$Z_5(x)$
0.25	1.258×10^{-2}	1.8169	0.5667	2.0458
0.25	1.010×10^{-3}	2.3938	0.3645	3.7140
0.25	1.013×10^{-4}	2.6637	0.0480	5.5750
0.5	1.514×10^{-2}	2.7867	0.5742	3.3631
0.5	1.038×10^{-3}	3.5447	0.1324	6.2539
0.5	1.017×10^{-4}	3.6408	0.0208	9.2826
0.75	2.103×10^{-2}	3.5296	0.3283	4.3116
0.75	1.032×10^{-3}	4.4880	0.0524	9.2254
0.75	1.026×10^{-4}	4.8918	0.0096	13.675

Recall that $Z_4(x)$ is logarithmically efficient for the Weibull claim size only when $\beta < 0.585$. One may wonder how much more reduction can be made by combining control variate with $Z_4(x)$. So, we define

$$Z_6(x) = Z_4(x) - c[I\{X_{(n-1)} > x/\mathrm{e}\} - \mathsf{P}\{X_{(n-1)} > x/\mathrm{e}\}],$$

where the lower bound of $X_{(n-1)}$ is chosen empirically. Nevertheless, the improvement is not substantial. The following is shown in [9]:

Theorem 23 When $\overline{F}(x) \sim x^{1-\beta}\mathrm{e}^{-x^{\beta}}$, $Z_6(x)$ is logarithmically efficient for $\beta < 0.65$.

A recent efficient estimator using a control variate which targets on lighter heavy-tails is proposed in [3]. With the fact that for large x, the major variance of $Z_4(x)$ comes from $S_{n-1}f(x)$, it is natural to use the term as a control variate, which leads to the estimator

$$Z_7(x) = Z_4(x) + n[\mathsf{E}(S_{n-1}) - S_{n-1}]f(x).$$

Theorem 24 Assume that $0 < \beta < 0.585$. Then $Z_7(x)$ has vanishing relative error. More precisely,

$$\mathsf{Var}\,(Z_7(x)) \sim \frac{n^2}{4}\mathsf{Var}\,(S_{n-1}^2)f'(x)^2.$$

Remark 25 Since $Z_7(x)$ has the form $Z_4(x) + c[S_{n-1} - \mathsf{E}(S_{n-1})]$, it uses S_{n-1} as a control for $Z_4(x)$. It is then a question whether the $c = -nf(x)$ at least asymptotically coincides with the optimal $c^* = -\mathsf{Cov}\,(Z_4(x), S_{n-1})/\mathsf{Var}\,(S_{n-1})$. The answer is yes by the fact proved in [3] that

$$\mathsf{Cov}\,(Z_4(x), S_{n-1}) = n\mathsf{Var}\,(S_{n-1})f(x) + o(f(x)).$$

《应用概率统计》版权所有

《应用概率统计》版权所有

4.3 Importance Sampling Algorithms

Although the ECM is impossible for a heavy-tail distribution, other means of importance sampling can be considered. Suppose that the density f of X_i can be changed to $\widetilde{f}$. If $\widetilde{f}$ does not depend on x, we have the result below:

Theorem 26 Let $\widehat{F}$ be the distribution with the density proportional to $f^2/\widetilde{f}$. If $\widehat{F}$ is subexponential and satisfies

$$\liminf_{n\to\infty}\frac{|\ln(1-\widehat{F}(x))|}{2|\ln(1-F(x))|}\geqslant 1,$$

then the importance sampling algorithm given by $\widetilde{f}$ is logarithmically efficient.

Proof Let $\widehat{c}=\int_0^\infty f^2/\widetilde{f}$. Then, by Lemma 6, the second moment of the estimator is

$$\widetilde{\mathsf{E}}\Big[\Big(\prod_{i=1}^{n}\frac{f(X_i)}{\widetilde{f}(X_i)}\Big)^2 I\{S_n>x\}\Big]=\widehat{c}^{\,-n}\widehat{\mathsf{P}}(S_n>x)\sim\widehat{c}^{\,-n}n(1-\widehat{F}(x)),$$

where the last step is due to the subexponentiality of $\widehat{F}$. Since $\alpha(x)\sim n(1-F(x))$ and $\widehat{c}$ does not depend on x, the second assumption on $\widehat{F}$ implies that Eq. (4) holds. □

Example 27 If F is regularly varying with $\delta>1$, one can take the tail of $\widetilde{F}$ as, e.g., $1/\ln(\mathrm{e}+x)$ (then $\widetilde{F}$ is a regularly varying distribution with $\delta=1$). Indeed, $\widetilde{f}=L_1(x)/x$ and Karamata's theorem:

$$\int_x^\infty\frac{L(y)}{y^{\delta-1}}\mathrm{d}y\sim\frac{L(x)}{(\delta-2)x^{\delta-2}}$$

for a slowly varying function L imply

$$1-\widehat{F}(x)=c_1\int_x^\infty\frac{L^2(x)/x^{2\delta}}{L_1(x)/x}\mathrm{d}x\sim c_2\frac{L_2(x)}{x^{2\delta-2}},$$

where L_1 and $L_2=L^2/L_1$ are slowly varying.

Theorem 26 is, however, one of the notorious reminders that a limit theorem does not always tell the truth on how an algorithm performs for a given set of parameters. All numerical experiments by the estimator show poor performance.

Finally, it is worth mentioning that two recent papers propose algorithms in the setting of dynamic importance sampling, [6] and [7]. Both of them have vanishing relative error. However, these algorithms are quite complicated than those introduced here.

§5. Exact Simulation

We now turn to the second topic of this report.

《应用概率统计》版权所有

Let $\{X(t), t \geqslant 0\}$ (or $\{X_n, n \in Z^+\}$) be a stochastic process in continuous (discrete) time. Suppose that the time-average limit, α, exists and is to be estimated.

In most applications and our study here, $X(t)$ is a Markov process. A stronger condition than the existence of the time-average limit is the existence of the *stationary distribution* (or equilibrium, steady-state distribution), π, say, by the ergodic theorem on Markov processes.

Generally, π is not known explicitly, and the commonly used algorithm for computing such steady-state quantities would be

$$\alpha_t = \frac{1}{t}\int_0^t X(s)\mathrm{d}s \to \alpha,$$

as $t \to \infty$ that is based on the LLN.

However, there are two principle difficulties in the approach:

(i) α_t is generally biased due to the initial effect as it usually starts from a nonequilibrium distribution; in other words, the data gathered during the initial transient can not represent the steady-state behavior of the system and, thus, is biased.

(ii) since α_t is produced from a single realization of the process, embedded dependence prevents our using C.L.T. to construct confidence intervals.

The common practice to deal with difficulty (i) is to discard the data gathered during this period. One would let the simulation warm up before collecting any data. However, how long the warm-up period must be is also a problem that has no satisfactory answer. Thus, how to generate samples from a stationary stochastic process has long been the key subject in steady-state simulation.

Clearly, if it were possible to generate the random variable with the *stationary* distribution, we could avoid the bias by using the stationary distribution as the initial one to have a stationary Markov chain (MC) to begin with. Such generation is called *exact sampling*, or *perfect simulation*. By the development in the past two decades, exact simulation has become possible for certain stochastic models.

The first algorithm of exact simulation was proposed in 1992 by Asmussen, Glynn & Thorisson[2], for a class of finite MC's. However, it was prohibitively inefficient in terms of computer time. A few years later, Propp & Wilson[14] used the idea of coupling from the past (CFTP) to construct an algorithm for perfect sampling. In this report, we will introduce both methods and related applications. We shall also note here that this topic remains an active research area.

《应用概率统计》版权所有

§6. Detection of Stationary

6.1 Regenerative Processes

A stochastic process $\{Y(t) : t \geqslant 0\}$ is called *regenerative* if there exists a subsequence $0 = T(0) < T(1) < T(2) < \cdots < \infty$ of random times (the *regenerative points*) such that the cycles

$$\{Y(T(k) + s) : 0 \leqslant s < T(k+1) - T(k)\}$$

are independent for different k and have the same distribution for $k \geqslant 1$. For expositions of the theory of regenerative processes, see [19].

The regenerative process can be viewed as a function of a MC $\{X(t)\}$: define $A(t) = t - T(k)$ as the *age* of the cycle when $T(k) \leqslant t < T(k+1)$, and let $X(t) = (A(t), Y(t))$.

The existence of limiting time average of functions f of a regenerative process holds without further conditions, provided only that $\mathsf{E}[T(2) - T(1)] < \infty$. The existence of a limiting distribution, in the sense of weak convergence, is a more technical topic and involves conditions on the distribution of $T(2)-T(1)$. However, for the purpose of stationary simulation it is the existence of time average that matters, not of a limiting distribution.

Example 28 Let $Y(t)$ be the workload at time t of a $GI/G/1$ queue, and $T(0), T(1)$, $\ldots$ be the epochs on customers entering an empty system. Thus, cycles $1, 2, \ldots$ are simply the conventional busy cycles (a busy period followed by an idle period).

If the arrival process is Poisson, an $M/G/1$ queue, one can alternatively let $T(0), T(1), \ldots$ be the epochs on customers departing to leave the server idle. Thus, a cycle becomes an idle period followed by a busy period. This definition works by the memoryless property of the exponential distribution. Otherwise, the evolution of the process after the start of an idle period depends on the residual arrival time at that instant.

Example 29 Reflected Brownian $\{Y(t)\}$ can be viewed as a continuous-state queueing or storage model. So, it is tempting to copy the busy period construction from the previous example by letting $T(0) = 0$, $T(1) = \inf\{t > 0 : Y(t) = 0\}$, $T(2) = \inf\{t > T(1) : Y(t) = 0\}$, and so on. However, sample path properties of Brownian motion imply that this definition yields the unusable $0 = T(0) = T(1) = \cdots$. Thus, we should define

$$T(k+1) \equiv \inf\Big\{t > T(k) : Y(t) = 0 \ \text{ and } \sup_{T(k) \leqslant s \leqslant t} Y(s) \geqslant 1\Big\}.$$

《应用概率统计》版权所有

6.2 Asmussen-Glynn-Thorisson Algorithm

Consider a continuous-time regenerative process $\{X(t), t \geqslant 0\}$. Define the initial cycle $C_0 = \{X(t), 0 \leqslant t < T(0)\}$ with a distribution possibly differing from the remaining cycles

$$C_k = \{X_{T(k-1)+t}, 0 \leqslant t < T(k) - T(k-1)\}, \qquad k = 1, 2, \ldots,$$

where $C_1, C_2, \ldots$ are i.i.d. We write $X = (C_0, C_1, \ldots)$ and C for generic cycle, τ for its length.

We start with the construction of a stationary regenerative process.

Proposition 30 (a) Assume that the distribution of C_0 has density

$$\mathsf{P}(C_0 \in \cdot) = \frac{1}{\mathsf{E}(\tau)} \mathsf{E}(\tau I\{C \in \cdot\}).$$

Let $U \sim U(0,1)$ and define $X^*(t) = X(t + UT(0))$. Then, X^* is a stationary version of X.

(b) Assume that the distribution of $T(0)$ has density $\mathsf{P}(\tau > t)/\mathsf{E}(\tau)$ and

$$\mathsf{P}(C_0 \in \cdot \,|\, T(0) = t) = \mathsf{P}(\theta_t C \in \cdot \,|\, \tau > t),$$

where $\theta_t C = \{X(s+t), 0 \leqslant s < \tau - t\}$. Then X is stationary.

Part (a) is intuitive that a stationary version of X can be obtained by constructing C_0 from the generic cycle C by first length-biasing with τ and next placing the time origin uniformly within the cycle. Part (b) leads to the following algorithm, in which the stationary cycle random variable τ^* defined as having the *equilibrium* density function $\mathsf{P}(\tau > x)/\mathsf{E}(\tau)$:

Proposition 31 For a regenerative process with simulatable τ^*, its stationary version can be generated by simulating C_0 as follows: Generate τ^* and successive cycles $C'_1, C'_2, \ldots$. Let σ be the first k with $\tau'_k > \tau*$, and take $C_0 = \theta_{\tau*} C'_\sigma$.

Example 32 To generate a stationary version of an $M/G/1$ queue is an easy problem under the FIFO discipline, since the stationary distribution of the workload (and hence the delay) can be generated by the Pollaczeck-Khinchine formula (see (14) in Section 6.3). For other service disciplines that are work-conserving, i.e., the workload process has the same distribution as under FIFO, the simulation of a stationary workload process does not present new problems.

Suppose that we want to simulate a stationary version $D_0^*, D_1^*, D_2^*, \ldots$ of the sequence of delays of customers $0, 1, 2, \ldots$ in the *processor-sharing* queue (this sequence is stochastically

《应用概率统计》版权所有

different than the one under FIFO). We take customers arriving to an empty system as regeneration points which are the same as under FIFO. Therefore, we can generate the stationary FIFO delay by the Pollaczeck-Khinchine formula and construct the FIFO delay sequence by the Lindley recursion (see [19; p.406]) until a customer sees an empty system upon arrival. The desired stationary cycle length τ^* is then the number of customers served until then, and to complete the construction of C_0, one simulates $C_1', C_2', \ldots$ using the processor-sharing discipline.

A general criterion for the ability to simulate a stationary version of a regenerative process is the availability of an integrable function $b(t)$ of explicit form such that $\mathsf{P}(\tau > t) \leqslant b(t)$, complemented with a (typically more easily available!) lower bound $\mathsf{P}(\tau > t) \geqslant q(t) > 0$.

Theorem 33 An algorithm for generating τ^* is the following:

1. Generate T from the density $h = b/\int b$. Fix $\delta > 1$ and write $t = T$, $g = g(t) = \delta b(t)$, $q = q(t)$, $p = \mathsf{P}(\tau > t)$.
2. Choose $n \geqslant 1$ such that $\delta(1 - g^n) \geqslant 1$ when $g < 1$, and $g \leqslant q/(1-q)^n$ when $g > 1$.
3. From the simulated values of T, generate a Bernoulli(p/g) random variable V using the Keane-O'Brien algorithm (see [10]) with n as in Step 2.
4. If $V = 1$, return $\tau^* = t$. Otherwise, return to Step 1.

This algorithm is an acceptance-rejection algorithm, accepting a random variable from $h(t)$ with probability $\mathsf{P}(\tau > t)/g(t)$. Since both $h(t)$ and $g(t)$ are proportional to $b(t)$, the output τ^* therefore has density proportional to $\mathsf{P}(\tau > t)$ as desired.

Proof We only need to prove that Step 3 is feasible, this means that we should have

$$\min\{p/g, 1 - p/g\} \geqslant (\min\{p, 1-p\})^n.$$

We first show that p/g is at least either p^n or $(1-p)^n$. If $g \leqslant 1$, then $p/g \geqslant p^n$ because of $n \geqslant 1$. If $g > 1$, then

$$\frac{p}{g} = \frac{p}{(1-p)^n}\frac{1}{g}(1-p)^n \geqslant \frac{p}{(1-p)^n}\frac{(1-q)^n}{q}(1-p)^n \geqslant (1-p)^n$$

since $1 \geqslant p \geqslant q$.

For the similar lower bound of $1 - p/g$, note first that $1 - p/g \geqslant 1 - p \geqslant (1-p)^n$ when $g \geqslant 1$. If $g < 1$, we will show that $1 - p/g \geqslant p^n$. This will hold if $1 - p/g \geqslant g^n$, which in turn is equivalent to $p \leqslant g - g^{n+1}$. The truth of this follows from

$$g - g^{n+1} = \delta b(t)(1 - g^n) \geqslant b(t) \geqslant p. \qquad \square$$

《应用概率统计》版权所有

Example 34 For a simple yet interesting example in which an upper bound on $\mathsf{P}(\tau > t)$ can be obtained, consider an (s, S)-policy with state space $\{0, 1, \ldots, S\}$. Items are removed at Poisson (λ) times, and when the inventory level makes a transition from $s+1$ to s, a supplier is called and arrives after a random time Z (the *lead* time) to replenish the inventory to S. As regeneration times, we take the times of transition $s+1 \to s$. If Z is stochastically smaller than the $\Gamma(\alpha, \lambda')$ distribution for some α and some $\lambda' \geqslant \lambda$, an upper bound on $\mathsf{P}(\tau > t)$ is then the tail probability $b(t)$ of a $\Gamma(\alpha + S - s, \lambda)$ distribution. A (trivial) lower bound is $q(t) = \mathsf{P}(Z > t)$. If instead $\lambda' \leqslant \lambda$, use the $\Gamma(\alpha + S - s, \lambda')$ distribution as an upper bound.

As a final remark for this approach, we note that in the final step of exact simulation (say, in Proposition 31), it may require to generate many cycles before obtaining one that has larger cycle length than τ^*, i.e., σ has a heavy tail. In fact, suppose $\tau \sim \exp$ so that $\tau^* \overset{\mathscr{D}}{=} \tau$. We have

$$\mathsf{E}(\sigma) = \sum_{i=1}^{\infty} \mathsf{P}(\sigma \geqslant i) = \sum_{i=1}^{\infty} \frac{1}{i} = \infty.$$

It is a drawback and invites for improvement.

6.3 An Application by [17]

Suppose that $\{X_n : n \geqslant 0\}$ is a positive recurrent discrete-time regenerative process, with i.i.d. cycle lengths generically denoted as $T \sim F$ with $\mathsf{E}(T) = \tau < \infty$. A generic cycle is thus $C = \{X_n : 0 \leqslant n < T\}$. From regenerative process theory, the (marginal) stationary distribution π is given by

$$\pi(\cdot) = \frac{1}{\tau} \mathsf{E}\Big(\sum_{n=0}^{T-1} I\{X_n \in \cdot\}\Big) = \frac{1}{\tau} \mathsf{E}\Big(\sum_{n=1}^{T} I\{X_n \in \cdot\}\Big).$$

The result below, presented in discrete time, is from Proposition 31.

Proposition 35

1. Suppose we can and do simulate i.i.d. copies of $C = \{X_n : 0 \leqslant n < T\}$, denoted by $C_j = \{X_n(j) : 0 \leqslant n < T_j\}$, $j \geqslant 1$, having i.i.d. cycle lengths $\{T_j\}$ distributed as F.
2. Suppose further that we can and do simulate (independently) one copy T^e distributed as $\mathsf{P}(T^e = n) = \mathsf{P}(T \geqslant n)/\tau$, $n \geqslant 1$.
3. Let $\sigma = \min\{j \geqslant 1 : T_j \geqslant T^e\}$.
4. Use cycle C_σ to construct $X^* = X_{T^e}(\sigma)$ (e.g., if $T^e = n$ and $\sigma = j$, then $X^* = X_n(j)$).

Then the simulated random element X^* is distributed as π.

《应用概率统计》版权所有

Proof Conditional on $T^e = n$, it holds that $\sigma = \min\{j \geqslant 1 : T_j \geqslant n\}$, and thus C_σ simply has the distribution of a first cycle given that its length is greater than or equal to n:

$$\mathsf{P}(X^* \in \cdot \,|\, T^e = n) = \mathsf{P}(X_n \in \cdot \,|\, T \geqslant n) = \frac{\mathsf{P}(X_n \in \cdot, T \geqslant n)}{\mathsf{P}(T \geqslant n)}.$$

Since $\mathsf{P}(T^e = n) = \mathsf{P}(T \geqslant n)/\tau$, we obtain

$$\mathsf{P}(X^* \in \cdot) = \sum_{n=1}^{\infty} \frac{\mathsf{P}(X_n \in \cdot, T \geqslant n)}{\mathsf{P}(T \geqslant n)} \frac{\mathsf{P}(T \geqslant n)}{\tau} = \frac{1}{\tau}\mathsf{E}\Big(\sum_{n=1}^{T} I\{X_n \in \cdot\}\Big) = \pi(\cdot). \qquad \square$$

Now, consider a FIFO $M/G/c$ queue, $c \geqslant 2$, with Poisson arrival times $\{t_n : n \geqslant 1\}$ at rate λ, and i.i.d. service times $\{S_n : n \geqslant 0\}$ that have distribution function G with finite mean $\mathsf{E}(S) = 1/\mu$.

With i.i.d. interarrival times $A_n = t_{n+1} - t_n$, let $\boldsymbol{W}_n = (W_n(1), W_n(2), \ldots, W_n(c))$ denote the *Kiefer-Wolfowitz workload vector* (see [19; p. 494]). It satisfies the recursion:

$$\boldsymbol{W}_{n+1} = R(\boldsymbol{W}_n + S_n\boldsymbol{e} - A_n\boldsymbol{f})^+, \tag{11}$$

where $\boldsymbol{e} = (1, 0, \ldots, 0)$, $\boldsymbol{f} = (1, 1, \ldots, 1)$, and R places a vector in ascending order. Notice that $D_n = W_n(1)$ is then the delay in queue of the nth customer.

This recursion defines a MC due to the given i.i.d. assumptions, and whenever $\boldsymbol{W}_n = \boldsymbol{0}$, the chain regenerates with initial condition $\boldsymbol{W}_0 = \boldsymbol{0}$. The event $\boldsymbol{W}_n = \boldsymbol{0}$ is equivalent to “the nth arrival finds the system empty”.

With $\rho = \lambda/\mu < c$, it is well known that $\boldsymbol{W}_n$ converges in distribution to a proper stationary distribution, denoted as π. In this section, we show an algorithm for sampling exactly from π. Our only assumption is that we can simulate from both G and its equilibrium version G_e.

Given a c-server queueing model, the *random assignment* (RA) is when each of the c servers forms its own FIFO single-server queue, and each arrival to the system, independent of the past, equally likely to join any queue.

Let $Q_{\rm F}(t)(Q_{\rm RA}(t))$ denote total number of customers in system at time t for the FIFO (RA) $M/G/c$ queue, where both models are initially empty and fed exactly the same input of Poisson arrivals $\{t_n\}$ and i.i.d. service times $\{S_n\}$. Assume further that for both models the service times are used by the servers in the order in which service initiations occur (S_n is the service time used for the nth such initiation). Then, it is shown (see [1; p. 342])

$$\mathsf{P}\{Q_{\rm F}(t) \leqslant Q_{\rm RA}(t), \,\forall\, t \geqslant 0\} = 1. \tag{12}$$

《应用概率统计》版权所有

Hence, we can jointly simulate versions of two stochastic processes $\{Q_{\rm F}(t)\}$ and $\{Q_{\rm RA}(t)\}$ while achieving a coupling such that (12) holds. In particular, *whenever an arrival finds the RA model empty, the FIFO model is found empty as well.*

For the RA model, let $\boldsymbol{Q}(t) = (Q_1(t), Q_2(t), \ldots, Q_c(t))$, where $Q_i(t)$ denotes the number of customers in the ith queue at time t (including the number in service, if any), and let $\boldsymbol{Q}_n = \boldsymbol{Q}(t_n^-)$ denote the number in system at the nodes found by the nth arriving customer. Thus, we can simulate the discrete-time process $\boldsymbol{Q}_n$, starting $\boldsymbol{Q}_0 = \boldsymbol{0}$, until it empties again. Consecutive visits of $\boldsymbol{Q}_n$ to $\boldsymbol{0}$ constitute positive recurrent regeneration points for the RA models. These also serve as positive recurrent regeneration points for the FIFO model due to (12), i.e., if $\boldsymbol{Q}_n = \boldsymbol{0}$, then $\boldsymbol{W}_n = \boldsymbol{0}$.

So, a generic cycle length T is defined by initializing $\boldsymbol{Q}_0 = \boldsymbol{0}$ and setting

$$T = \min\{n \geqslant 1 : \boldsymbol{Q}_n = \boldsymbol{0}\}. \tag{13}$$

To generate a sample of T requires a standard discrete-event simulation of $\{\boldsymbol{Q}(t)\}$, where the events are an *arrival* versus a *service completion*, and a service time S is generated only when it is needed for processing by a server to ensure that (12) applies.

The sequential generated input random variables are i.i.d. service times $S_n \sim G$, i.i.d. interarrival times $A_n \sim \exp(\lambda)$, and i.i.d. random selections $U_n \sim$ discrete uniform distribution over $\{1, 2, \ldots, c\}$. If $U_n = i$, then the nth arrival joins the ith queue.

At time $t_0 = 0$, U_0 is generated, and a server is randomly selected according to U_0 and begins service for a generated service time S_0 (e.g., the system is found empty by an initial customer who starts the cycle). The number in system at queue U_0 is increased to 1. A_0 is then generated so as to schedule the next arrival. The simulation continues into the future analogously until an arriving customer finds the entire system empty, thus ending the RA cycle.

We do not simulate the FIFO model until the RA cycle is complete, at which time we use the input that was used for the RA cycle to construct the FIFO cycle for the workload vector in (11): store the T service times $\{S_0, S_1, \ldots, S_{T-1}\}$ as well as the T interarrival times $\{A_0, A_1, \ldots, A_{T-1}\}$ so they can be used to construct the FIFO cycle $C = \{\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_T\}$ by using recursion (11) with $\boldsymbol{W}_0 = 0$, from $n = 0$ up to $T - 1$.

To employ Proposition 35, we need to be able to simulate a copy of T^e. We shall utilize the fact that T^e has the stationary excess distribution (stationary forward recurrence time distribution) of the (discrete-time) renewal process of visits of the RA model to the empty state (this renewal process has i.i.d. cycle lengths distributed as T). By the definition of the RA regeneration points, if we take a stationary version of $\{\boldsymbol{Q}_n : n \geqslant 0\}$, denoted by $\{\boldsymbol{Q}^*_n : n \geqslant 0\}$, then $T^e = \min\{n \geqslant 1 : \boldsymbol{Q}^*_n = 0\}$.

Since to obtain a stationary version of the RA model we only need to consider the RA model by itself, independently of the FIFO model (recall Step 2 of Proposition 35), we can just assign service times upon arrival. Also, because arrivals are Poisson, and we randomly partition them into c independent Poisson processes each at rate λ/c, we can simply treat each server as its own independent stable FIFO $M/G/1$ queue with Poisson arrivals at rate $\lambda/c < \mu$. Moreover, we can model workload instead of number in system since they both empty together and thus share the same regeneration times. The stationary workload distribution at each queue is thus given by the Pollaczeck-Khinchine formula. The stationary delay has the representation

$$D = \sum_{j=1}^{L} Y_j, \tag{14}$$

where the $\{Y_j\}$ are i.i.d., $\sim G_e(x) = \mu \int_0^x \mathsf{P}(S > y)\mathrm{d}y$, $x \geqslant 0$, and $L \sim \mathrm{Geo}(1 - \lambda/\mu)$ that is independent with $\{Y_j\}$.

To put this to use: Letting $\boldsymbol{V}_n = (V_n(1), V_n(2), \ldots, V_n(c))$ denote workload (at each node) as found by the nth arriving customer to the RA model, we have, for each node $i \in \{1, 2, \ldots, c\}$,

$$V_{n+1}(i) = (V_n(i) + S_n I\{U_n = i\} - A_n)^+, \qquad n \geqslant 0.$$

Thus, $\{\boldsymbol{V}_n : n \geqslant 0\}$ forms a Markov process due to the i.i.d. assumptions on the input. Denote the corresponding continuous-time process by $\boldsymbol{V}(t) = (V(t, 1), V(t, 2), \ldots, V(t, c))$, where $V(t, i)$ denotes the workload at the ith node at time $t \geqslant 0$, and $\boldsymbol{V}_n = \boldsymbol{V}(t^-_n)$, $n \geqslant 1$. From Poisson arrivals see time averages (PASTA) (see [19; p. 293]), the limiting stationary distribution of $\boldsymbol{V}_n$, as $n \to \infty$, is identical with that of $\boldsymbol{V}(t)$, as $t \to \infty$. But the coordinates of $\boldsymbol{V}(t)$, namely $V(t, 1), V(t, 2), \ldots, V(t, i)$, are i.i.d. copies of workload for the $M/G/1$ queue. Thus, the *joint* time-stationary distribution of workload is given by

$$(D(1), D(2), \ldots, D(c)), \tag{15}$$

where the $D(i)$ here are i.i.d. distributed as D in 14.

We conclude that the stationary distribution for $\{\boldsymbol{V}_n : n \geqslant 0\}$ is the same as (15) and thus *the proportion of arrivals who find the RA system empty is given by* $\mathsf{P}^c(D = 0) = (1 - \rho_1)^c > 0$*; visits to the empty state constitute positive recurrent regeneration points*; $\mathsf{E}(T) < \infty$.

《应用概率统计》版权所有

《应用概率统计》版权所有

With $\boldsymbol{V}_0 = 0$, we have an identically distributed version of a cycle length (13) given by $T = \min\{n \geqslant 1 : \boldsymbol{V}_n = 0\}$. So, if we start it off with $\boldsymbol{V}_0$ distributed as in (15), then the process will be a stationary version, denoted by $\{\boldsymbol{V}_n^* : n \geqslant 0\}$. We conclude that $T^e = \min\{n \geqslant 1 : V_n^* = 0\}$.

Algorithm for simulating T^e:

1. Initializing $\boldsymbol{V}_0 = (D(1), D(2), \ldots, D(c))$.
2. Simulate sequentially $\{\boldsymbol{V}_n\}$ until time $T^e = \min\{n : \boldsymbol{V}_n = 0\}$.

We conclude the approach below, in which we only need to assume that we can simulate from both G and G_e:

Algorithm for simulating stationary $\boldsymbol{W}$:

1. Simulate a copy of T^e. Set $k = T^e$.
2. Independently generate T.
3. If $T < k$, then go back to Step 2.
4. Construct the FIFO cycle $C = \{\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_T\}$. Set $\boldsymbol{W} = \boldsymbol{W}_k$.
5. Output $\boldsymbol{W}$.

§7. Coupling from the Past

7.1 Coupling Method

Suppose that some comparison of probability measures on a measurable space is to be carried out. For that purpose, it is sometimes possible, and then often rewarding, to construct random variables on a common probability space, with these measures as distributions, in such a way that the comparison may be carried out in terms of the random variables. Such a construction is called a *coupling*.

In this section, we briefly review the coupling of MC's and its asymptotic properties such as asymptotic stationarity.

Let $\{X_n\}$ be an ergodic (which in the finite case means irreducible and aperiodic) and positive recurrent MC with a countable state space $\boldsymbol{S}$ with transition probability matrix $\boldsymbol{P}$. It is a classical result that X_n approaches stationarity as $n \to \infty$, regardless of the initial distribution $\{\lambda_i\}$:

$$\mathsf{P}(X_n = j) = \sum_i \lambda_i p_{ij}^{(n)} \to \pi_j \quad \text{as} \quad n \to \infty, \tag{16}$$

《应用概率统计》版权所有

where $\{\pi_i\}$ is the unique stationary distribution. The coupling proof of (16) is the following:

We first let $\{X_n'\}$ be a copy of $\{X_n\}$, independent of X_n, governed by $\boldsymbol{P}$ and stationary; the latter property is achieved by letting $X_0' \sim \pi$. Next, define $\{X”_n\}$ by

$$X_n'' = \begin{cases} X_n & \text{if } \ n < \tau; \\ X_n' & \text{if } \ n \geqslant \tau, \end{cases}$$

where $\tau = \min\{k : X_k = X_k'\}$. We thus construct a coupling of X_n and X_n' at τ, the coupling time, such that

$$\begin{aligned} |\mathsf{P}(X_n = j) - \pi_j| &= |\mathsf{P}(X_n'' = j) - \mathsf{P}(X_n' = j)| \\ &= |\mathsf{P}(X_n'' = j, \tau \leqslant n) + \mathsf{P}(X_n'' = j, \tau > n) \\ &\quad - \mathsf{P}(X_n' = j, \tau \leqslant n) - \mathsf{P}(X_n' = j, \tau > n)| \\ &= |\mathsf{P}(X_n'' = j, \tau > n) - \mathsf{P}(X_n' = j, \tau > n)|. \end{aligned}$$

From above, we have $|\mathsf{P}(X_n = j) - \pi_j| \leqslant \mathsf{P}(\tau > n)$, and thus, (16) follows if the coupling is successful, i.e., if $\tau < \infty$ almost surely.

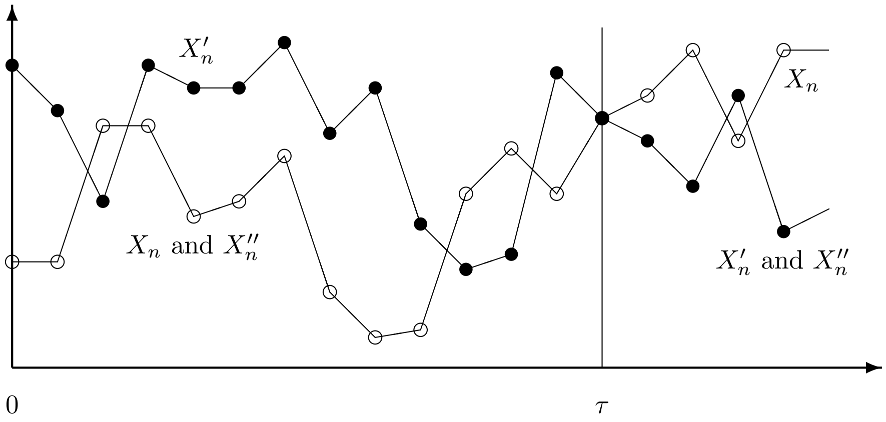

Figure 1 Classical coupling of two discrete-time MC's

7.2 Propp-Wilson Algorithm

Consider a MC with state space $\boldsymbol{S} = \{1, 2, \ldots, k\}$ and transition probabilities p_{ij}, $i, j \in \boldsymbol{S}$. We assume that the MC is ergodic.

We will use an *updating rule* to represent the MC simulation, which involves a random mapping

$$\boldsymbol{E} = (E(1), E(2), \ldots, E(k))$$

《应用概率统计》版权所有

such that $E(i) \in \boldsymbol{S}$ with distribution $\boldsymbol{p}_i$, i.e., $\mathsf{P}\{E(i) = j\} = p_{ij}$. Note that the k components of $\boldsymbol{E}$ are not necessarily independent. We shall return to this point later.

From $\boldsymbol{E}$, we then construct $\{X_n, n = 0, 1, \ldots\}$ recursively by $X_{n+1} = E(X_n)$ and let $X_0 = i$, the initial state. More generally, we can define a version $\{X_n^N(i), n = N, N+1, \ldots\}$ of $\{X_n\}$ starting from i at time N by

$$X_N^N(i) = i, X_{N+1}^N(i) = E(i) = E(X_N^N(i)), \ldots, X_{n+1}^N(i) = E(X_n^N(i)).$$

The *forward coupling time* is defined as

$$\tau_1 = \inf\{n \geqslant 1 : X_n^0(1) = X_n^0(2) = \cdots = X_n^0(k)\},$$

which is the first time at which the MC

$$\{X_n^0(1), n \geqslant 0\}, \{X_n^0(2), n \geqslant 0\}, \ldots, \{X_n^0(k), n \geqslant 0\}$$

started at time 0 from k different states coalesce. See Figure 2.

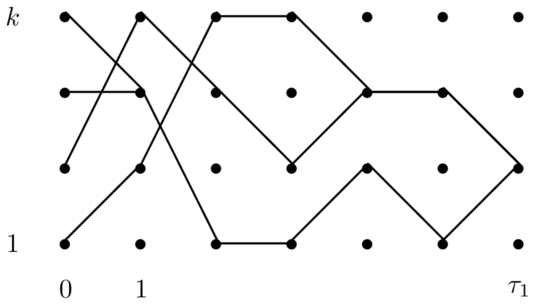

Figure 2 The forward coupling

Whether the forward coupling time is finite almost surely depends on the updating rule, that is, the specific dependence between the k components of $\boldsymbol{E}$. Call the updating rule *independent* if these components are independent.

Proposition 36 For independent updating, $\mathsf{P}(\tau_1 < \infty) = 1$.

Proof Let $p_{ij}^{(n)} = \mathsf{P}(X_n = j \,|\, X_0 = i)$, the n-step transition probability. Since $p_{ij}^{(n)} \to \pi_j > 0$, there exists an N and positive ϵ such that $p_{ij}^{(N)} > \epsilon > 0$ for all $i \in \boldsymbol{S}$. Hence, the probability that k independent MC's starting at time 0 from k different states will all be in state j at time N is $\mathsf{P}(\tau_1 \leqslant N) \geqslant \epsilon^k$. Similarly, $\mathsf{P}(\tau_1 \leqslant 2N \,|\, \tau_1 > N) \geqslant \epsilon^k$ so that

$$\mathsf{P}(\tau_1 > N) \leqslant 1 - \epsilon^k,\ \mathsf{P}(\tau_1 > 2N) \leqslant (1 - \epsilon^k)^2,\ \ldots,$$

《应用概率统计》版权所有

which implies that $\tau_1 < \infty$ a.s. □

Instead of forward coupling, Propp and Wilson propose an algorithm that uses *coupling from the past* (CFTP). This method is based on the principle that a MC that has already been running for an infinitely long time has reached its stationary distribution. To obtain a random sample, CFTP “figures out” what state the MC is in at a given time, by looking at a finite but unbounded number of randomizing operations used prior to that time.

In particular, the *backward coupling time* is defined as

$$\tau_2 = \inf\{n \geqslant 1 : X_0^{-n}(1) = X_0^{-n}(2) = \cdots = X_0^{-n}(k)\},$$

the first time at which the MC

$$\{X_0^{-n}(1), n \geqslant 0\}, \{X_0^{-n}(2), n \geqslant 0\}, \ldots, \{X_0^{-n}(k), n \geqslant 0\}$$

started at time $-n$ from k different states coalesce. It will remain true from τ_2 onward, that is, from all earlier $-n$'s. Equivalently, coalescence means that the set $\{X_0^{-n}(1), X_0^{-n}(2), \ldots, X_0^{-n}(k)\}$ contains only one point. Note that the cardinality of this set is a nonincreasing function of $-n$. See Figure 3.

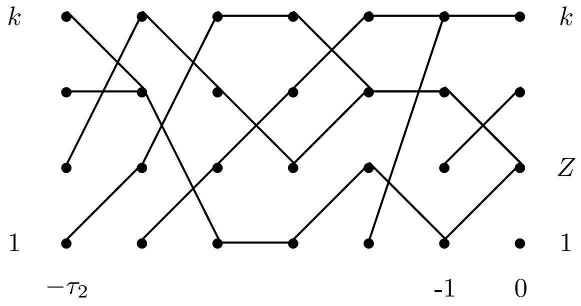

Figure 3 The backward coupling

Theorem 37 For the updating rule such that $\tau_1 < \infty$ a.s., $\tau_2 < \infty$ a.s. as well. Furthermore, $X_0^{-\tau_2}(i)$ does not depend on i and has distribution π.

Proof The first statement follows since $\mathsf{P}(\tau_2 \leqslant k) = \mathsf{P}(\tau_1 \leqslant k) \to 1$ as $k \to \infty$. That $X_0^{-\tau_2}(i)$ does not depend on i is from the definition of τ_2.

Now consider $X_0^{-n}(i)$ for some fixed i. On $\tau_2 \leqslant n$, we have $X_0^{-n}(i) = X_0^{-\tau_2}(i)$ and hence $\mathsf{P}(X_0^{-n}(i) = j) \to \mathsf{P}(X_0^{-\tau_2}(i) = j)$ as $n \to \infty$ for all i. On the other hand, $\mathsf{P}(X_0^{-n}(i) = j) = p_{ij}^n \to \pi_j$. Hence, $\mathsf{P}(X_0^{-\tau_2}(i) = j) = \pi_j$ as desired. □

The introduction of forward coupling time is only for showing that the backward coupling time is finite. It is important to note that $X^0_{\tau_1}(i)$ is *not* stationarily distributed.

The next result is for general MC's.

Corollary 38 For independent updating, $\tau_2 < \infty$ a.s., $X_0^{-\tau_2}(i)$ does not depend on i and has distribution π.

Now assume that there is some partial order $\preceq$ on $\boldsymbol{S}$ such that 1 is the minimal element and k is the maximal one. Also, we say that $\{X_n\}$ is *stochastic monotone* if $i \preceq j$ implies $X_1^0(i) \preceq X_1^0(j)$ in stochastic order, which, to be put in terms of transition probability, becomes

$$\sum_{m\geqslant l} p_{im} \leqslant \sum_{m\geqslant l} p_{jm} \qquad \text{for all } \ l \ \text{ if } \ i \preceq j.$$

Example 39 A random walk reflected at the barrier 0 and k,

$$X_{n+1} = \min(k, \max(0, X_n + B_n)),$$

is a monotonic MC, where B_i's are i.i.d. integers. Such chains appear in many finite buffer queueing problems, or dam models.

Under the monotonicity assumption, a variant of the Propp-Wilson algorithm is often more efficient. It is defined by *monotone updating* that requires

$$E(i) \preceq E(j) \qquad \text{if } \ i \preceq j.$$

It implies $X_n^N(i) \preceq X_n^N(j)$ for all N and all $n \geqslant N$. In particular, for all i,

$$X_n^N(1) \preceq X_n^N(i) \preceq X_n^N(k). \tag{17}$$

For instance, in Example 39 the natural monotone updating rule is $E(i) = \min(k, \max(0, i + B))$ with the same B for all i. As for the independent updating, one would need to take the B's to be independent for different i.

Define

$$\tau_3 = \inf\{n \geqslant 1 : X_0^{-n}(1) = X_0^{-n}(k)\}.$$

Figure 4 below depicts the monotone coupling.

《应用概率统计》版权所有

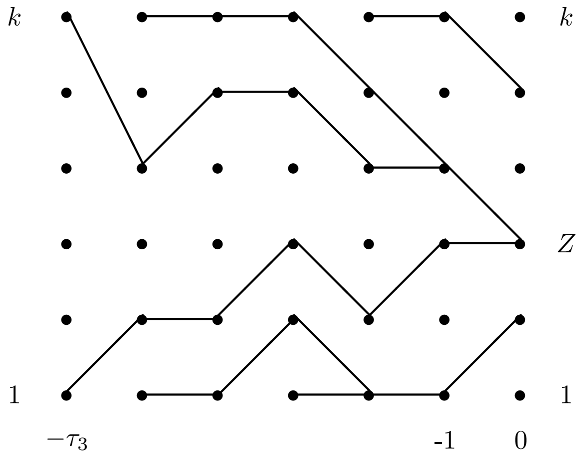

Figure 4 The monotone coupling

The following result is for stochastically monotone MC's:

Corollary 40 For monotone updating, $\tau_3 \leqslant \tau_2 < \infty$ a.s., $X_0^{-\tau_3}(i)$ does not depend on i and has distribution π.

Proof Let $\tau(k,1) = \inf\{n \geqslant 1 : X_n^0(k) = 1\}$. By the recurrence, $\tau(k,1)$ is finite and $X^0_{\tau(k,1)}(i) = 1$ for all i by (17). Clearly, by the definition, $\tau_3 \leqslant \tau_2$. On the other hand, $X_0^{-\tau_3}(i) = X_0^{-\tau_3}(1) = X_0^{-\tau_3}(k)$ for all i by (17). □

7.3 Read-Once CFTP and PASTA

A few years after the birth of CFTP, Wilson[18] developed a variant of CFTP which only runs the MC forwards in time and never restarts it at previous times in the past. Because the method can be run using a read-once stream of randomness, it is called *read-once*, or *forward*, CFTP. The memory and time requirements of read-once CFTP are on par with the requirements of the usual form of CFTP, and for a variety of applications the requirements may be noticeably less. Here, we introduce the version of read-once CFTP that incorporates with PASTA.

Generally speaking, read-once CFTP may be viewed as a retroactive stopping rule. It applies random transitions going forward in time, and then at some point it decides to stop, and then returns not the current state, but some previous state. The engine of read-once CFTP is a composite random transition procedure (see [18; Figure 3 on p. 92]) that generates a *random map* and determines whether or not the map is coalescent (i.e. whether or not it maps all states to one state). Then, it evaluates the map at a given input state to obtain an output state. If the procedure determines (by examining the representation

《应用概率统计》版权所有

《应用概率统计》版权所有

of the superset of the image of the map) that the random map is coalescent, then we say that the map is “officially coalescent”. Otherwise the map is not officially coalescent.

Intuitively, suppose the composite transition procedure gave us the entire random map rather than just evaluating it at one state. Then we could do CFTP, composing new composite maps going back in time. Let f_{-T} be the first (the smallest $T > 0$) composite map that is officially coalescent. As being conditioned to be officially coalescent, f_{-T} is independent of T. Let S be the state in the image of f_{-T}. CFTP would then apply the composite maps $f_{-T+1}, f_{-T+2}, \ldots, f_{-1}$ to S, and return the result. The composite maps $f_{-T+1}, f_{-T+2}, \ldots, f_{-1}$ are i.i.d. random composite maps conditioned not to be officially coalescent, and are independent of S. So we could equivalently generate $T-1$ fresh random composite maps conditioned not to be officially coalescent, and apply them to S. We can simply update S using fresh composite random maps, until one of the maps is officially coalescent.

In read-once CFTP, an event occurs when a composite map is officially coalescent. Imagine first randomly picking those integral times at which events occur. If there is an event at a given time, then the MC is updated by a random composite map conditioned to be officially coalescent, otherwise it is updated by a random composite map conditioned not to be officially coalescent.

Furthermore, a discrete-time version of PASTA states that the distribution of the MC sampled at times just prior to when events occur will be identical to the steady-state distribution of the MC. Thus, we *draw random samples from the MC at times just prior to when the composite maps are officially coalescent* so that the steady-state distribution of the sample will be the steady-state distribution of the MC.

While PASTA is a statement about the steady-state behavior of the draws; in read-once CFTP the first several draws taken at positive times will be out of equilibrium. In this particular application of PASTA, since there is a coalescent map between draws, not only are draws after the first one easy to compute, but they also must be independent of one another. Since the draws are independent, any particular draw is already in the steady-state distribution. Read-once CFTP ignores the first draw (since it is neither in equilibrium nor easy to compute), and outputs the subsequent draws until the desired number of independent perfectly random samples are generated.

An advantage of read-once CFTP over CFTP is that one does not need to keep track of pseudorandom number generator seeds. When many independent samples are desired, CFTP typically keeps track of seeds for a number of independent streams of pseudorandom numbers, whereas read-once CFTP needs only one good-quality stream of pseudorandom numbers to produce a large number of samples.

As a remark, one may perceive that CFTP and PASTA are not completely unrelated ideas. The “time zero sees time averages” principle behind CFTP can be used to derive the “Poisson arrivals see time averages” in both the continuous and discrete settings.

7.4 An Application by [13]

In a queueing system, the fraction of arrivals finding the queue in some state, the *customer* average, is not necessarily the same as the fraction of time the queue in that state, the *time* average. In this subsection, new proofs relating time averages and customer averages are presented, which is based on the read-once CFTP idea for discrete-time MC’s.

Consider a queueing system having a renewal customer arrival process, and let $Q(t)$ be the state of the system at time t, immediately prior to any arrival which may occur at that time. Let $X_1, X_2, \ldots$ be i.i.d. customer interarrival times and $T_n = \sum\limits_{i=1}^{n} X_i$ be the arriving time of the nth customer. We suppose that the state of the system encodes the amount of time each customer has been in service, along with their positions in the system, so that $Q(T_n)$, the state seen by arrival n, is a MC. Let $\boldsymbol{S}$ denote the state space and $0 \in \boldsymbol{S}$ correspond to the system being empty.

Suppose for some $p > 0$ that

$$P_{s,0} = \mathsf{P}\{Q(T_{n+1}) = 0 \,|\, Q(T_n) = s\} \geqslant p \qquad \text{for all } \ s,$$

and construct $\{J_n\}$ as an i.i.d. Bernoulli sequence with parameter p such that whenever $J_n = 1$ then $Q(T_n) = 0$. Specifically, let $\{U_i\}$ be i.i.d. $U(0,1)$ random variables, and

$$J_n = \max_{s \in \boldsymbol{S}} I\{Q(T_{n-1}) = s,\, Q(T_n) = 0,\, U_n < p/P_{s,0}\}.$$

That is, $J_n = 1$ if, for some s, the state seen by arrival $n-1$ is s, the state seen by arrival n is 0, and the corresponding uniform is less than $p/P_{s,0}$.

We say that random variable Z taking value in $\boldsymbol{S}$ has the time average steady-state (stationary) distribution if

$$\mathsf{P}(Z \in A) = \lim_{t \to \infty} \frac{1}{t} \int_0^t I\{Q(s) \in A\} \mathrm{d}s$$

for any set of states A. For $X \sim F$, we say that X_e has the *equilibrium* distribution if

$$\mathsf{P}(X_e \leqslant x) = \frac{1}{\mathsf{E}(X)} \int_0^x \mathsf{P}(X > s) \mathrm{d}s.$$

Its density is $f_e(x) = \mathsf{P}(X > x)/\mathsf{E}(X)$.

《应用概率统计》版权所有

《应用概率统计》版权所有

Suppose that the system is empty at time 0 and say that a new cycle begins at time T_N, where $N = \min\{n > 0 : J_n = 1\}$ denotes the number of arrivals in the first cycle. Also, let $Q_n(t)$ be a random variable, independent of all else, with distribution

$$\mathsf{P}\{Q_n(t) \leqslant m\} = \mathsf{P}\{Q(T_n + t) \leqslant m \,|\, T_n + t < T_{n+1},\, n \leqslant N\}.$$

Note that this means that $Q_n(t)$ is independent of both N and $Q(t)$, though its distribution is defined in terms of the distribution of $Q(t)$ and N.

We now state following result, which is an extension to the read-once CFTP. We refer readers to [13] for the detailed proof.

Proposition 41 With the above definitions, the variable $Q_{N-1}(X_e)$ has the time average steady-state distribution.

Sketch of the proof Let $N = \min\{n>0 : J_n=1\}$ and R_A be the amount of time spent in A during the first cycle, then

$$R_A = \int_0^{T_N} I\{Q(s) \in A\}\mathrm{d}s.$$

By the renewal reward theorem, we have

$$\mathsf{P}(Z \in A) = \frac{\mathsf{E}(R_A)}{\mathsf{E}(T_N)}.$$

Because N is a stopping time for the sequence of interarrival times, Wald's equation gives

$$\mathsf{E}(T_N) = \mathsf{E}(X)\mathsf{E}(N) = \mathsf{E}(X)/p.$$

Now, the desired result will be shown by proving

$$\mathsf{E}(R_A) = \frac{\mathsf{P}\{Q_{N-1}(X_e) \in A\}\mathsf{E}(X)}{p}. \qquad \square$$

The variable $Q_{N-1}(X_e)$ can be simulated as follows. Simulate the queue until a cycle ends, let $s = Q(T_{N-1})$, generate X_e, start a completely independent simulation of the queueing system from state s at time 0 without any arrival, and output the state of this independent simulation at time X_e. Note that this is not the same as simply outputting the state $Q(T_{N-1} + X_e)$, since $Q_{n-1}(t)$ is defined to be independent of all else (including N) and, thus, N does not indicate the end of a cycle in the system $Q_{n-1}(t)$.

On the other hand, we say that W taking value in $\boldsymbol{S}$ has the customer average steady-state distribution if

$$\mathsf{P}(W \in A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I\{Q(T_i) \in A\}.$$

《应用概率统计》版权所有

Fix a parameter $p \geqslant 0$, and suppose that each customer, independently of all else, with probability p is labeled as an *exploding* customer. Immediately after the arrival of such a customer, the system is instantly cleared of all customers and enters state 0. Now let J_n be the indicator variable equal to 1 if customer n is an exploding customer. If $J_n = 1$, then $Q(T_n)$ is the system seen by an exploding arrival. We consider $p > 0$ and $= 0$ separately.

The result below is also a version of the 'forward' CFTP, and allows exact simulation of the stationary distribution of a discrete-time MC.

Proposition 42 If $p > 0$, then $Q(T_N)$ has the customer average steady-state distribution.

Proof Let N_A denote the number of time periods when the MC is in state A during the cycle, i.e.,

$$N_A = \sum_{i=1}^{N} I\{Q(T_i) \in A\}.$$

If we suppose that a reward 1 is earned each time when the state of the chain is in A then

$$\pi_A = \mathsf{E}(N_A)/\mathsf{E}(N) = p\mathsf{E}(N_A).$$

Define event B_k be that a new cycle begins on the transition following the kth visit to A. Then

$$\sum_{k=1}^{N_A} I\{B_k\} = I\{Q(T_N) \in A\}.$$

Because $I\{B_1\}, I\{B_2\}, \ldots$ are i.i.d. and the event $\{N_A = n\}$ is independent of $I\{B_{n+1}\}$, $I\{B_{n+2}\}, \ldots$, it follows that

$$\mathsf{P}\{Q(T_N) \in A\} = \mathsf{E}(N_A)\mathsf{E}(I\{B_1\}) = p\mathsf{E}(N_A).$$

This completes the proof. □

The next result relates time and customer averages for queues with exploding customers.

Corollary 43 Suppose that $p > 0$.

1. If X is exponentially distributed, then $W \overset{\mathscr{D}}{=} Z$. In other words, if arrivals are from a Poisson process, then the customer average steady-state distribution is the same as the time average steady-state distribution.
2. Let $G(s)$, $s \in \boldsymbol{S}$, be a function of the system state having the property that, for some partial ordering, $G(Q(T))$ decreases with respect to that partial ordering at all times

《应用概率统计》版权所有

at which customers do not arrive. Then $G(W) \leqslant_{\rm st} G(Z)$ ($G(W) \geqslant_{\rm st} G(Z)$) when the interarrival distribution X is NBUE (NWUE). In other words, the customer average steady-state distribution of $G(Q)$ is stochastically smaller (larger) than the time average steady-state distribution of $G(Q)$ when the interarrival distribution is NBUE (NWUE).

Proof It follows from Porposition 42 that the customer average steady-state distribution is the distribution of the system state seen by exploding arrivals, which is distributed as the state seen by customers immediately preceding an exploding arrival after a random interarrival time during which there are no new arrivals.

On the other hand, from Proposition 41, the time average steady-state distribution is the distribution of the system state seen by customers immediately preceding an exploding arrival after a random time, distributed according to the equilibrium distribution, during which there are no new arrivals.

In the exploding customers model, each interarrival time is independent of the type of customer (exploding or nonexploding). Then, because the equilibrium distribution is the interarrival distribution for a Poisson process, the famous PASTA result, part 1, is proven. Also, becuase the equilibrium distribution is stochastically smaller than F when F is NBUE and is stochastically larger when F is NWUE, part 2 follows. □

Finally, the main result below relates customer and time averages for general queues.

Corollary 44 Corollary 43 holds if $p = 0$, under the stability condition that state 0 is positive recurrent.

We now use a simple example to illustrate this result.

Example 45 Consider a $D/D/1$ queue with interarrival and service times being 1 and 0.9, respectively. Then, all customers see an empty system upon arrival, i.e., the customer-average stationary distribution of the number in system being 0 is 1. Equivalently, from Proposition 42, $N = 1, 2, \ldots$ such that $W = Q(T_N) = 0$ surely.

On the other hand, the system is empty for 10% of the time and has one customer 90% of the time. Indeed, as

$$\mathsf{P}(X_e \leqslant x) = \frac{1}{\mathsf{E}(X)} \int_0^x \mathsf{P}(X > s){\rm d}s = x, \qquad 0 < x < 1,$$

the equilibrium interarrival time $\sim U(0, 1)$. That $Z = Q_{N-1}(X_e) = 0$ with probability 0.1 and 1 with probability 0.9 is equal to the time-average stationary distribution as Proposition 41 says.

Finaly, $W \leqslant_{\rm st} Z$ is consistent with Corollary 44 as the constant is NBUE.

Acknowledgments This report is based on the lecture notes of a short course given at Beijing Normal University in the Summer of 2015. The author is grateful to the Stochastics Research Center for the hospitality and support received there.

《应用概率统计》版权所有

References

[1] Asmussen S. *Applied Probability and Queues* [M]. 2nd ed. New York: Springer-Verlag, 2003.

[2] Asmussen S, Glynn P W, Thorisson H. Stationarity detection in the initial transient problem [J]. *ACM TOMACS*, 1992, **2(2)**: 130–157.

[3] Asmussen S, Kortschak D. Error rates and improved algorithms for rare event simulation with heavy Weibull tails [J]. *Methodol. Comput. Appl. Probab.*, 2015, **17(2)**: 441–461.

[4] Asmussen S, Kroese D P. Improved algorithms for rare event simulation with heavy tails [J]. *Adv. in Appl. Probab.*, 2006, **38(2)**: 545–558.

[5] Asmussen S, Glynn P W. *Stochastic Simulation: Algorithms and Analysis* [M]. New York: Springer-Verlag, 2007.

[6] Blanchet J, Glynn P. Efficient rare-event simulation for the maximum of heavy-tailed random walks [J]. *Ann. Appl. Probab.*, 2008, **18(4)**: 1351–1378.

[7] Dupuis P, Leder K, Wang H. Importance sampling for sums of random variables with regularly varying tails [J]. *ACM TOMACS*, 2007, **17(3)**: Article No. 14, 1–21.

[8] Hartinger J, Kortschak D. On the efficiency of the Asmussen - Kroese-estimator and its application to stop-loss transforms [J]. *Blätter der DGVFM*, 2009, **30(2)**: 363–377.

[9] Jhou Y S. Simulation of ruin probability with heavy-tail claims by control variates [D]. Taiwan: National Dong Hwa University, 2010.

[10] Keane M S, O'Brien G L. A Bernoulli factory [J]. *ACM TOMACS*, 1994, **4(2)**: 213–219.

[11] Omey E, Willekens E. Second order behaviour of the tail of a subordinated probability distribution [J]. *Stochastic Process. Appl.*, 1986, **21(2)**: 339–353.

[12] Omey E, Willekens E. Second-order behaviour of distributions subordinate to a distribution with finite mean [J]. *Comm. Statist. Stochastic Models*, 1987, **3(3)**: 311–342.

[13] Peköz E A, Ross S M. Relating time and customer averages for queues using 'forward' coupling from the past [J]. *J. Appl. Probab.*, 2008, **45(2)**: 568–574.

[14] Propp J G, Wilson D B. Exact sampling with coupled Markov chains and applications to statistical mechanics [J]. *Random Structures Algorithms*, 1996, **9(1-2)**: 223–252.

[15] Resnick S I. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling* [M]. New York: Springer, 2007.

[16] Rolski T, Schmidli H, Schmidt V, et al. *Stochastic Processes for Insurance and Finance* [M]. Chichester: Wiley, 1999.

[17] Sigman K. Exact simulation of the stationary distribution of the FIFO $M/G/c$ queue: the general case for $\rho < c$ [J]. *Queueing Syst.*, 2012: **70(1)**: 37–43.

[18] Wilson D B. How to couple from the past using a read-once source of randomness [J]. *Random Structures Algorithms*, 2000, **16(1)**: 85–113.

[19] Wolff R W. *Stochastic Modeling and the Theory of Queues* [M]. Englewood Cliffs, NJ: Prentice Hall, 1989.

随机模拟的一些新进展

王 家 礼

(国立东华大学应用数学系, 花莲, 台湾)

摘 要: 此综述文章介绍随机模拟方面的两个新进展: 构造小概率事件估计的有效算法, 产生形式不封闭的平稳分布的样本.

估计一个非常小的量, 需要极其准确地取定一个有用的置信区间. 这使得慢收敛的小概率事件模拟在有效性和准确性两方面都成为具有挑战性的任务. 在此文中, 我们介绍一些有趣的小概率事件例子以及在估计它们时的困难所在. 然后沿着发展脉络, 寻求稳健且有效的估计量的各种方法将被讨论和评估. 估计破产概率的数值实验则用来显示这些方法的质量.

在稳定态模拟中, 如何产生平稳随机过程的样本长期以来是一个关键性课题. 通常的做法是在初始的短暂时期内丢弃掉所得数据. 然而, 热身准备必须多长时间则成为另一个没有满意答案的问题. 幸运地, 经过过去二十年的发展, 对一些特定的随机模型, 精确模拟已经成为可能. 在此文中, 我们将介绍两个重要的方法及其相关的应用.

关键词: 破产概率; 对数效率; 测度的指数变换; 重尾; 次指数; 重要性抽样; 控制变量; 条件估计量; 马氏链; 平稳分布; 再生过程; 从过去出发的耦合

中图分类号: O226

《应用概率统计》版权所有