Nonlife Actuarial Models

Chapter 14

Basic Monte Carlo Methods

Learning Objectives

1. Generation of uniform random numbers, mixed congruential method
2. Low discrepancy sequence
3. Inversion transformation and acceptance-rejection methods
4. Generation of specific discrete and continuous random variates
5. Generation of correlated normal random variables
6. Variance reduction techniques
7. Antithetic variable, control variable and importance sampling

14.1 Monte Carlo Simulation

- Suppose $h(\cdot)$ is a smooth integrable function over the interval $[0, 1]$, and it is desired to compute the integral

$$\int_0^1 h(x)\, dx. \tag{14.1}$$

- Let $U \sim \mathcal{U}(0, 1)$. The integral in (14.1) is equal to $\mathrm{E}[h(U)]$.

- If the solution of (14.1) is difficult to obtain analytically, we may consider the stochastic solution of it as the mean of $h(U)$.

- The stochastic solution can be estimated by drawing a random sample of n observations $u_1, \cdots, u_n$ from U, and the computed estimate is given by

$$\hat{\mathrm{E}}\left[h(U)\right] = \frac{1}{n} \sum_{i=1}^{n} h(u_i). \tag{14.2}$$

- $\hat{\mathrm{E}}[h(U)]$ converges to $\mathrm{E}[h(U)]$ when n tends to ∞.
- Von Neumann and Ulam coined the use of the term **Monte Carlo method** to describe this technique.
- This technique can also be extended to study the solution of any simulated stochastic process (not necessarily with a deterministic counterpart), called statistical **simulation**.

14.2 Uniform Random Number Generators

- Independent random variates from the $\mathcal{U}(0,1)$ distribution can be generated in the computer by dividing random integers in the interval $[0, m)$ by m, where m is a large number.

- An important method for generating sequences of random integers is the use of the congruential algorithm.

- We first define the expression

$$y \equiv z \pmod{m}, \tag{14.3}$$

where m is an integer, and y and z are integer-valued expressions, to mean that there exists an integer k, such that

$$z = mk + y. \tag{14.4}$$

- This also means that y is the remainder when z is divided by m.
- The **mixed-congruential method** of generating a sequence of random integers x_i is defined by the equation

$$x_{i+1} \equiv (ax_i + c) \pmod{m}, \qquad \text{for } i = 0, 1, 2, \cdots, \tag{14.5}$$

  where a is the **multiplier**, c is the **increment** and m is the **modulus**. The mixed-congruential method requires the restrictions: $m > 0$, $0 < a < m$ and $0 \leq c < m$.
- When $c = 0$, the method is said to be **multiplicative-congruential**.
- To start the sequence of x_i, we need a **seed** x_0. Given the seed x_0, the sequence of numbers x_i are completely determined.
- Random numbers generated by computer algorithms usually follow deterministic sequences, and are called **pseudo-random numbers**.

- Given a seed x_0, when the algorithm produces a value $x_k = x_h$ for certain integers h and k, such that $k > h \geq 0$, the sequence will start to repeat itself.

- We define the **period of the seed** as the shortest subsequence of numbers, which, by repeating itself, forms the complete sequence generated.

- The **period of the generator** is the *largest* period among all seeds.

Example 14.1: Consider the following mixed-congruential generator

$$x_{i+1} \equiv (3x_i + 1) \pmod{8}.$$

What is the period of (a) $x_0 = 2$, and (b) $x_0 = 4$? What is the period of the generator?

Solution: For (a) it is easy to show that the sequence of numbers generated are

$$2, 7, 6, 3, 2, \cdots$$

and this sequence repeats itself. Thus, the period of $x_0 = 2$ is 4. For (b), we have the sequence

$$4, 5, 0, 1, 4, \cdots$$

Hence, the period of $x_0 = 4$ is again 4. To summarize, for given seed values $x_{0,}$ the values of x_1 are given as follows

x_0	0	1	2	3	4	5	6	7
x_1	1	4	7	2	5	0	3	6

All seeds have period 4, and the generated sequences belong to one of the two sequences above. Thus, the period of the generator is 4. □

14.3 General Random Number Generators

- In many practical applications we may be required to generate random numbers from distributions other than $\mathcal{U}(0,1)$.
- It turns out that the generation of random numbers following an arbitrary distribution can be done using uniform random numbers via the inversion transformation.
- We first define the important **probability integral transform**, which is basically the transformation of a random variable using its distribution function.

Definition 14.1: Let X be a random variable with df $F(\cdot)$. The probability integral transform Y of X is a random variable defined by $Y = F(X)$.

- Thus, the probability integral transform is just a df, where the argument is a random variable rather than a fixed number.

- It turns out that through the probability integral transform we can obtain a random variable that is distributed as $\mathcal{U}(0,1)$.

Theorem 14.1 **(a) Probability integral transform theorem:** If X is a random variable with continuous df $F(\cdot)$, then the random variable $Y = F(X)$ is distributed as $\mathcal{U}(0,1)$. **(b) Quantile function theorem:** Let $F(\cdot)$ be a df, and define $F^{-1}(\cdot)$ as $F^{-1}(y) = \inf\{x : F(x) \geq y\}$, for $0 < y < 1$. If $U \sim \mathcal{U}(0,1)$, then the df of $X = F^{-1}(U)$ is $F(\cdot)$.

Proof: For Part (a), if $F(\cdot)$ is strictly increasing, the proof is quite straightforward. In this case, for $0 < y < 1$, there exists a unique x such that $F(x) = y$. Furthermore, $Y \leq y$ if and only if $X \leq x$. Thus, if $G(\cdot)$ is

the df of Y, then

$$G(y) = \Pr(Y \leq y) = \Pr(X \leq x) = F(x) = y.$$

Hence, $G(y) = y$, which implies $Y \sim \mathcal{U}(0, 1)$. For a general proof requiring $F(\cdot)$ to be continuous only, see Angus (1994).

For Part (b), we note that $X \leq x$ if and only if $U \leq F(x)$. Thus, we conclude

$$\Pr(X \leq x) = \Pr(U \leq F(x)) = F(x).$$

The last equality above is due to the fact that $U \sim \mathcal{U}(0, 1)$. Hence, the df of X is $F(\cdot)$, as required by the theorem. □

14.3.1 Inversion Method

- Provided we can invert the function $F(\cdot)$ to obtain $F^{-1}(\cdot)$, $F^{-1}(U)$ will be a random variable with df $F(\cdot)$.
- This is called the **inversion method** for generating a random number for an arbitrary distribution.

Example 14.3: Derive algorithms to generate random numbers from the following distributions: (a) $\mathcal{W}(\alpha, \lambda)$, and (b) $\mathcal{P}(\alpha, \gamma)$.

Solution: For (a), from equation (2.36), the df of $\mathcal{W}(\alpha, \lambda)$ is

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\lambda}\right)^{\alpha}\right].$$

Inverting the df, we generate X using the formula

$$X = \lambda\left[-\log\left(1 - U\right)\right]^{\frac{1}{\alpha}}.$$

As $1 - U$ is also distributed as $\mathcal{U}(0, 1)$, we can use the simplified formula

$$X = \lambda \left[-\log U \right]^{\frac{1}{\alpha}}$$

to generate $\mathcal{W}(\alpha, \lambda)$.

For (b), from equation (2.38), the df of $\mathcal{P}(\alpha, \gamma)$ is

$$F(x) = 1 - \left(\frac{\gamma}{x + \gamma} \right)^{\alpha}.$$

Thus, random numbers from $\mathcal{P}(\alpha, \gamma)$ may be generated using the equation

$$X = \gamma (U^{-\frac{1}{\alpha}} - 1).$$

□

- The above examples illustrate the use of the inverse transform of the df to generate continuous random numbers.

- The inversion method can also be used to generate discrete or mixed-type variables.

Example 14.4: The ground-up loss X of an insurance policy is distributed as $\mathcal{W}(0.5, 5)$. There is a deductible of $d = 1$ and maximum covered loss of $u = 8$. Derive an algorithm to generate the loss in a loss event variable X_L using a $\mathcal{U}(0, 1)$ variate U. What are the values of X_L generated when $U = 0.8, 0.25$ and 0.5?

Solution: $X_L = 0$ when $X \leq 1$. Thus,

$$F_{X_L}(0) = \Pr(X \leq 1) = \Pr(\mathcal{W}(0.5, 5) \leq 1) = 1 - \exp\left[-\left(\frac{1}{5}\right)^{0.5}\right] = 0.3606.$$

X_L is also right censored at point 7, with

$$\Pr(X_L = 7) = \Pr(X \geq 8) = \exp\left[-\left(\frac{8}{5}\right)^{0.5}\right] = 0.2823.$$

Hence, $\Pr(X_L < 7) = 1 - 0.2823 = 0.7177$, and the df of X_L is

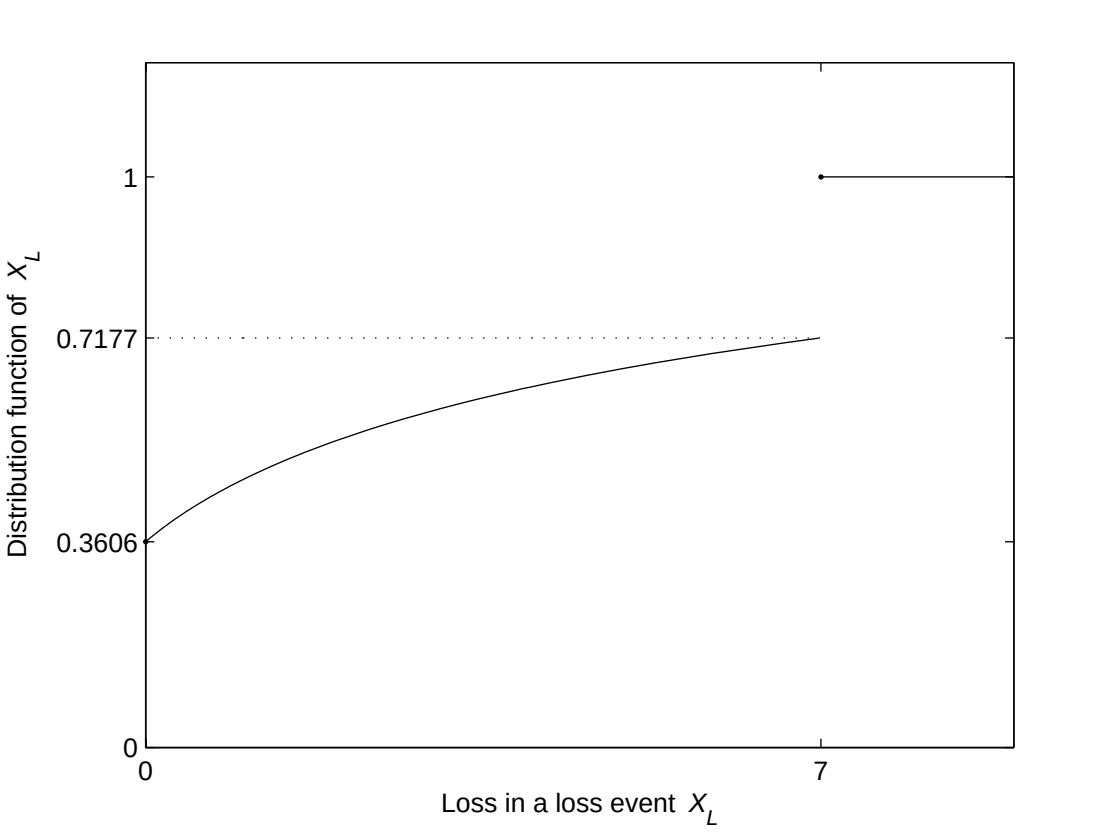
$$F_{X_L}(x) = \begin{cases} 0.3606, & \text{for } x = 0, \\ 1 - \exp\left[-\left(\dfrac{x+1}{5}\right)^{0.5}\right], & \text{for } 0 < x < 7, \\ 1, & \text{for } x \geq 7. \end{cases} \tag{14.9}$$

Thus, X_L is a mixed-type random variable, and its df is plotted in Figure 14.1.

We may invert $F_{X_L}(x)$ as follows to generate a random variate of X_L given a $\mathcal{U}(0,1)$ variate U

$$X_L = \begin{cases} 0, & \text{for } 0 \leq U < 0.3606, \\ 5\left[-\log\left(1-U\right)\right]^2 - 1, & \text{for } 0.3606 \leq U < 0.7177, \\ 7, & \text{for } 0.7177 \leq U < 1. \end{cases} \tag{14.10}$$

When $U = 0.8$, $X_L = 7$. When $U = 0.25$, $X_L = 0$. Finally, when $U = 0.5$,


1
0.7177
0.3606
0
0
7
Distribution function of X_L
Loss in a loss event X_L

X_L is computed as

$$X_L = 5\left[-\log(1-0.5)\right]^2 - 1 = 1.4023.$$

Note that X_L can also be generated by left-truncating and right-censoring a Weibull variate computed using the inversion method. □

14.3.2 Acceptance-Rejection Method

- Let $f(\cdot)$ be the pdf of a random variable X, the df of which cannot be easily inverted, and let Y be another random variable with pdf $q(\cdot)$, for which an easy and efficient generator is available.

- Assume X and Y have the same support $[a, b]$, and there exists a constant c such that $M(x) \equiv cq(x) \geq f(x)$ for $x \in [a, b]$.

- The steps of the **acceptance-rejection procedure** are as follows

1. Generate a number x from the distribution with pdf $q(\cdot)$.

2. Generate a number u independently from the $\mathcal{U}(0,1)$ distribution.

3. If $u \leq f(x)/M(x)$, assign $z = x$, otherwise return to Step 1.

- It turns out that the sequence of numbers z obtained from the above procedure have pdf $f(\cdot)$. To prove this statement we consider the df of the random variable Z generated, which is given by

$$\begin{aligned} \Pr(Z \leq z) &= \Pr\left(Y \leq z \,|\, U \leq \frac{f(Y)}{M(Y)}\right) \\ &= \frac{\int_a^z \int_0^{\frac{f(x)}{M(x)}} q(x)\, du\, dx}{\int_a^b \int_0^{\frac{f(x)}{M(x)}} q(x)\, du\, dx} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\int_a^z q(x) \left(\int_0^{\frac{f(x)}{M(x)}} du \right) dx}{\int_a^b q(x) \left(\int_0^{\frac{f(x)}{M(x)}} du \right) dx} \\
&= \frac{\int_a^z q(x) \frac{f(x)}{M(x)} \, dx}{\int_a^b q(x) \frac{f(x)}{M(x)} \, dx} \\
&= \frac{\int_a^z f(x) \, dx}{\int_a^b f(x) \, dx} \\
&= \int_a^z f(x) \, dx. \qquad (14.11)
\end{aligned}
$$

- The pdf $q(\cdot)$ is called the **majorizing density**, and the function $M(x) = cq(x)$ is called the **majorizing function**.

- The principle is to find a majorizing function that *envelopes* the pdf

$f(\cdot)$ as closely as possible.

- For a given majorizing density $q(\cdot)$, c should be chosen to tighten the enveloping of $M(x)$ over $f(x)$, i.e, the optimum c should be

$$c = \inf \{r : rq(x) \geq f(x) \text{ for } x \in [a, b]\} . \tag{14.12}$$

- However, even if the optimum c is not used the acceptance-rejection procedure stated above remains valid, albeit there is loss in efficiency.

Example 14.5: Let the pdf of X be

$$f(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \qquad \text{for } x \geq 0.$$

Suppose the majorizing density is selected to be

$$q(x) = e^{-x}, \qquad \text{for } x \geq 0.$$

Discuss the use of the acceptance-rejection procedure for the generation of random numbers of X.

Solution: X is obtained as the absolute value of the standard normal random variable. Inverse transformation method is intractable for this distribution. Figure 14.2 plots the pdf $f(\cdot)$ and $q(\cdot)$. The two functions cross each other. To create the optimum $cq(\cdot)$ the value of c is $\sqrt{2e/\pi} = 1.3155$. However, any value of $c \geq 1.3155$ may be used to compute the majorizing function and appropriate random numbers will be produced. Figure 14.2 also shows the majorizing function with $c = 1.5$, which is not optimum.

The acceptance-rejection procedure for generating X is summarized as follows

1. Generate a number x with pdf e^{-x}. This can be done by computing

$x = -\log v$, where v is a random number from $\mathcal{U}(0,1)$.

2. Generate a number u independently from $\mathcal{U}(0,1)$.

3. For a selected value of $c \geq 1.3155$, if

$$u \leq \frac{\frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)}{c \exp(-x)} = \frac{1}{c}\sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2} + x\right) \equiv R(x),$$

assign $Z = x$. Otherwise, return to Step 1.

Table 14.1 shows a sample of values of random numbers generated using $c = 1.5$. The last row of values Z are the random numbers having pdf $f(\cdot)$.

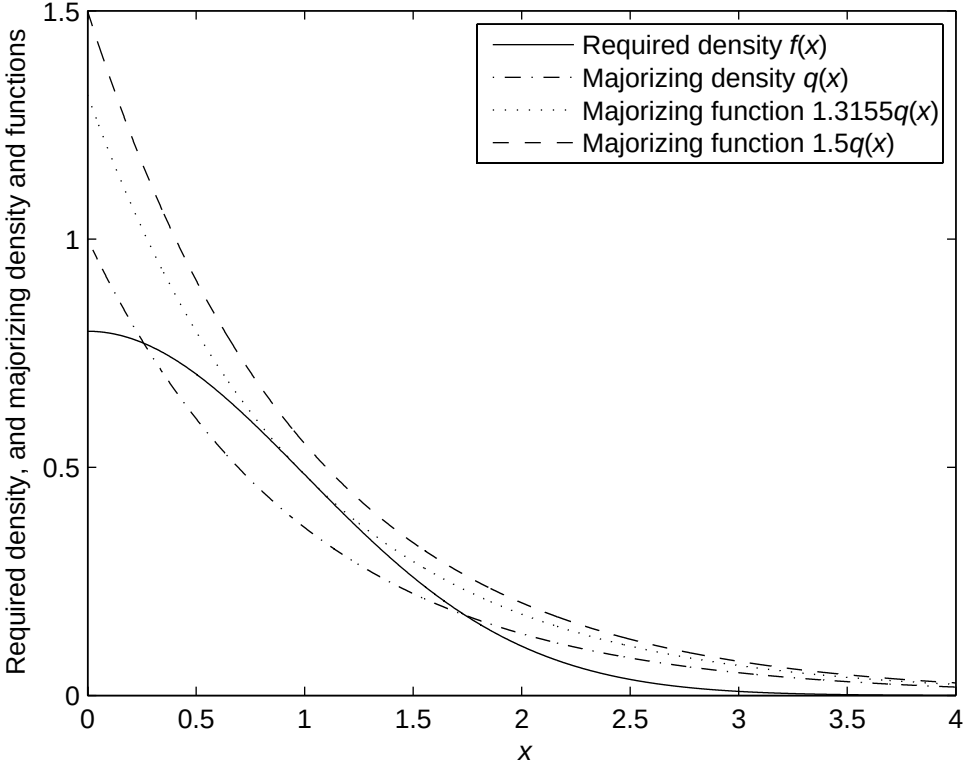

Required density f(x)
Majorizing density q(x)
Majorizing function 1.3155q(x)
Majorizing function 1.5q(x)
Required density, and majorizing density and functions
x
0
0.5
1
1.5
2
2.5
3
3.5
4
0
0.5
1
1.5

Table 14.1: Illustrative results for Example 14.5

i	1	2	3	4
u	0.3489	0.9236	0.5619	0.4581
v	0.4891	0.0910	0.5047	0.9057
x	0.7152	2.3969	0.6838	0.0990
$R(x)$	0.8421	0.3306	0.8342	0.5844
Z	0.7152	reject	0.6838	0.0990

□

- The probability of acceptance in Step 3 of the acceptance-rejection procedure is given by

$$\begin{aligned} \Pr\left(U \leq \frac{f(X)}{M(X)}\right) &= \int_a^b \int_0^{\frac{f(x)}{M(x)}} q(x)\, du\, dx \\ &= \int_a^b q(x) \frac{f(x)}{M(x)}\, dx \end{aligned}$$

$$= \frac{1}{c}. \tag{14.13}$$

- Thus, we may use $1/c$ as a measure of the efficiency of the procedure.

14.3.3 Generation of Correlated Random Variables

- We consider the problem of generating samples of normal random variables that are correlated.

- We first discuss the main properties of a multivariate normal distribution, followed by methods of generating correlated multivariate normal variates.

- Let $\boldsymbol{X} = (X_1, \cdots, X_k)'$ be a k-element random variable. If $\boldsymbol{X}$ has a multivariate normal distribution, its joint df is completely deter-

mined by its mean vector $\mu = (\mu_1, \cdots, \mu_k)'$ and its variance matrix

$$\Omega = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdot & \cdot & \sigma_{1k} \\ \sigma_{12} & \sigma_2^2 & \cdot & \cdot & \sigma_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{1k} & \cdot & \cdot & \cdot & \sigma_k^2 \end{bmatrix}, \tag{14.14}$$

where

$$\mu_i = \mathrm{E}(X_i) \quad \text{and} \quad \sigma_i^2 = \mathrm{Var}(X_i), \qquad \text{for } i = 1, \cdots, k, \tag{14.15}$$

and

$$\sigma_{ij} = \mathrm{Cov}(X_i, X_j), \qquad \text{for } i, j = 1, \cdots, k. \tag{14.16}$$

- We will then write

$$\boldsymbol{X} \sim \mathcal{N}(\mu, \Omega). \tag{14.17}$$

- If $\boldsymbol{X}$ has a nondegenerate distribution, there exists a lower triangular $k \times k$ matrix C (i.e., the elements in the upper triangle of the matrix are all zero), denoted by

$$C = \begin{bmatrix} c_{11} & 0 & 0 & \cdot & 0 \\ c_{21} & c_{22} & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{k1} & \cdot & \cdot & \cdot & c_{kk} \end{bmatrix}, \tag{14.18}$$

  such that

$$\Omega = CC'. \tag{14.19}$$

- The equation above is called the **Choleski decomposition** of Ω. The lower triangular matrix C is obtainable in many statistical packages.

- The multivariate normal distribution has some very convenient properties. Let A be a $m \times k$ $(m \leq k)$ constant matrix and b be a $m \times 1$ constant vector. Then

$$A\boldsymbol{X} + b \sim \mathcal{N}(A\mu + b, A\Omega A'). \tag{14.20}$$

- If $\boldsymbol{Y} = (Y_1, \cdots, Y_k)'$ has a multivariate normal distribution with mean vector $\mu = (0, \cdots, 0)' = \mathbf{0}$ and variance matrix $\Omega = \mathbf{I}$ (i.e., the $k \times k$ identity matrix), we write

$$\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{14.21}$$

- $Y_1, \cdots, Y_k$ are iid standard normal variates. Furthermore, if we define

$$\boldsymbol{X} = C\boldsymbol{Y} + \mu, \tag{14.22}$$

then from equation (14.20) we conclude

$$\boldsymbol{X} \sim \mathcal{N}(C\mathbf{0} + \mu, C\mathbf{I}C') \equiv \mathcal{N}(\mu, \Omega). \tag{14.23}$$

- To generate random numbers of the multivariate normal distribution $\boldsymbol{X} \sim \mathcal{N}(\mu, \Omega)$, we first generate k iid standard normal variates $\boldsymbol{Y} = (Y_1, \cdots, Y_k)'$. Then using equation (14.22), we obtain the required random numbers for $\boldsymbol{X}$.

Example 14.6: Let X_1 and X_2 be jointly normally distributed with means μ_1 and μ_2, respectively, variances σ_1^2 and σ_2^2, respectively, and covariance σ_{12}. How would you generate random numbers of X_1 and X_2 given independent random numbers of the standard normal distribution?

Solution: We first solve for the Choleski decomposition of

$$\Omega = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

It can be easily checked that

$$C = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{bmatrix},$$

where ρ is the correlation coefficient, i.e.,

$$\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}.$$

Hence, if Z_1 and Z_2 are independently distributed $\mathcal{N}(0,1)$ variates, then we can generate X_1 and X_2 from the equation

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \sigma_1 Z_1 + \mu_1 \\ \rho\sigma_2 Z_1 + \sigma_2\sqrt{1-\rho^2}\, Z_2 + \mu_2 \end{bmatrix}.$$

It is easy to verify that $\mathrm{E}(X_1) = \mu_1$, $\mathrm{E}(X_2) = \mu_2$, $\mathrm{Var}(X_1) = \sigma_1^2$,

$$\mathrm{Var}(X_2) = \rho^2\sigma_2^2 + \sigma_2^2\left(1-\rho^2\right) = \sigma_2^2,$$

and

$$\mathrm{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2 = \sigma_{12}.$$

□

14.4 Specific Random Number Generators

14.4.1 Some Continuous Distributions

Normal distribution

- The **Box-Muller method** generates pairs of standard normal variates from pairs of uniform variates using trigonometric transformation.
- The **Marsaglia-Bray method** uses a mixture distribution together with the acceptance-rejection method.

Gamma distribution

- A $\mathcal{G}(\alpha, \beta)$ variate has the same distribution as a $\beta\mathcal{G}(\alpha, 1)$ variate, and we only need to consider standard gamma distributions.

- If α is an integer (Erlang distribution), $X \sim \mathcal{G}(\alpha, 1)$ can be generated by the equation

$$X = -\sum_{i=1}^{\alpha} \log U_i, \tag{14.25}$$

where $U_i \sim$ iid $\mathcal{U}(0, 1)$, for $i = 1, \cdots, \alpha$.

- $\mathcal{G}(\alpha, 1)$ is the sum of two gamma variates, one of which has an Erlang distribution and the other has a standard gamma distribution with parameter in the interval $(0, 1)$.

- We now consider the case of generating a $\mathcal{G}(\alpha, 1)$ variate with $\alpha \in (0, 1)$.

- The **Ahrens method** provides an efficient procedure to generate a $\mathcal{G}(\alpha, 1)$ variate with $\alpha \in (0, 1)$ using the acceptance-rejection ap-

proach. The required pdf is

$$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}, \qquad \text{for } \alpha \in (0,1),\ x \geq 0. \tag{14.26}$$

The majorizing frequency consists of two segments defined as follows

$$q(x) = \begin{cases} \dfrac{e}{\alpha + e}\, \alpha x^{\alpha-1}, & \text{for } 0 \leq x \leq 1, \\ \dfrac{\alpha}{\alpha + e}\, e^{1-x}, & \text{for } 1 < x. \end{cases} \tag{14.27}$$

The df of this density, denoted by $Q(\cdot)$, is

$$Q(x) = \begin{cases} \dfrac{e}{\alpha + e}\, x^{\alpha}, & \text{for } 0 \leq x \leq 1, \\ 1 - \dfrac{\alpha}{\alpha + e}\, e^{1-x}, & \text{for } 1 < x. \end{cases} \tag{14.28}$$

Using the inverse transformation we can generate a random number X with df $Q(\cdot)$ from a $\mathcal{U}(0,1)$ variate U as follows

$$X = \begin{cases} \left[\dfrac{(\alpha+e)\,U}{e}\right]^{\frac{1}{\alpha}}, & \text{for } 0 \le U \le \dfrac{e}{\alpha+e}, \\ 1 - \log\left[\dfrac{(1-U)(\alpha+e)}{\alpha}\right], & \text{for } \dfrac{e}{\alpha+e} < U < 1. \end{cases} \tag{14.29}$$

To envelope the pdf $f(\cdot)$ we use the majorizing function $M(x) = cq(x)$, where c is given by

$$c = \frac{\alpha+e}{\Gamma(\alpha)\alpha e}. \tag{14.30}$$

Thus, the majorizing function is

$$M(x) = cq(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)}\, x^{\alpha-1}, & \text{for } 0 \le x \le 1, \\ \\ \dfrac{1}{\Gamma(\alpha)}\, e^{-x}, & \text{for } 1 < x. \end{cases} \tag{14.31}$$

We further note that

$$\frac{f(x)}{M(x)} = \begin{cases} e^{-x}, & \text{for } 0 \le x \le 1, \\ \\ x^{\alpha-1}, & \text{for } 1 < x. \end{cases} \tag{14.32}$$

- The Ahrens method for the generation of the random variable X with pdf $f(\cdot)$ is as follows

1. Generate a random number u_1 from $\mathcal{U}(0,1)$. If $u_1 > e/(\alpha + e)$, go to Step 3. Otherwise, continue with Step 2.

2. Set $z = [(\alpha + e)\, u_1/e]^{\frac{1}{\alpha}}$ and generate independently another random number u_2 from $\mathcal{U}(0,1)$. If $u_2 > e^{-z}$, go to Step 1, otherwise assign $X = z$.

3. Set $z = 1 - \log[(1 - u_1)(\alpha + e)/\alpha]$ and generate independently another random number u_2 from $\mathcal{U}(0,1)$. If $u_2 > z^{\alpha-1}$, go to Step 1, otherwise assign $X = z$.

14.4.2 Some discrete distributions

Binomial distribution

- We can use a simple **table look-up method** to generate the random numbers.

- Alternatively X can be generated as the number of successes in n independent trials where the probability of success in each trial is θ.

- Thus, we generate n random variates U_i, $i = 1, \cdots, n$, from $\mathcal{U}(0, 1)$ and compute X as the number of U_i that are less than θ.

Poisson distribution

- As the Poisson distribution has an infinite support, the table look-up method does not work.
- We may make use of the relationship between the exponential distribution and the Poisson distribution to derive an algorithm.
- Let $X \sim \mathcal{PN}(\lambda)$ be the number of arrivals of a certain event in a unit time interval. Then the inter-arrival time Y of the events follows an exponential distribution $\mathcal{E}(\lambda)$, i.e., an exponential distribution with mean waiting time $1/\lambda$.

- Thus, we can generate Y_i from $\mathcal{E}(\lambda)$ and accumulate them to obtain the *total* waiting time. We then set X to be the largest number of Y_i accumulated such that their total is less than 1, i.e.,

$$X = \min\{n : \sum_{i=1}^{n+1} Y_i > 1\}. \tag{14.35}$$

- As Y_i can be generated by $(-\log U_i)/\lambda$, where $U_i \sim \mathcal{U}(0,1)$, we re-write the above as

$$\begin{aligned} X &= \min\{n : \sum_{i=1}^{n+1} \frac{1}{\lambda}(-\log U_i) > 1\} \\ &= \min\{n : \sum_{i=1}^{n+1} \log U_i < -\lambda\} \\ &= \min\{n : \prod_{i=1}^{n+1} U_i < e^{-\lambda}\}. \end{aligned} \tag{14.36}$$

14.5 Accuracy and Monte Carlo Sample Size

- We may use the Monte Carlo sample to estimate the standard error of the estimated solution and obtain a confidence interval for the solution.

- The standard error may also be used to estimate the required sample size to produce a solution within a required accuracy given a certain probability level.

Example 14.7: The specific damages X covered by a liability insurance policy are distributed as $\mathcal{G}(4, 3)$. The total damages, inclusive of punitive damages, are given by cX, where $c > 1$. The policy has a maximum covered loss of u. Using Monte Carlo methods or otherwise, determine the expected loss of the insured and the probability that the total damages do

not exceed u. Discuss the accuracy of your solutions. Consider the case of $c = 1.1$ and $u = 20$.

Solution: The pdf of X is

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\, x^{\alpha-1}e^{-\frac{x}{\beta}}, \qquad x \geq 0.$$

We denote the df of $X \sim \mathcal{G}(\alpha, \beta)$ by

$$\gamma_x(\alpha, \beta) = \Pr(X \leq x),$$

and note that

$$\gamma_x(\alpha, \beta) = \gamma_{\frac{x}{\beta}}(\alpha, 1) \equiv \gamma_{\frac{x}{\beta}}(\alpha).$$

The function $\gamma_x(\alpha)$ is also called the (lower) incomplete gamma function.

The expected loss is

$$\mathrm{E}\left[(cx) \wedge u\right] \;=\; \int_0^{\frac{u}{c}} cx\, \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}\, dx + u \int_{\frac{u}{c}}^{\infty} \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}\, dx$$

$$
\begin{aligned}
&= \frac{c\Gamma(\alpha+1)\beta}{\Gamma(\alpha)} \int_0^{\frac{u}{c}} \frac{x^{(\alpha+1)-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha+1)\beta^{\alpha+1}}\, dx + u\left[1 - \gamma_{\frac{u}{c\beta}}(\alpha)\right] \\
&= c\alpha\beta\gamma_{\frac{u}{c\beta}}(\alpha+1) + u\left[1 - \gamma_{\frac{u}{c\beta}}(\alpha)\right].
\end{aligned}
$$

Thus, the expected loss can be computed using the incomplete gamma function. Similarly, we can derive the second moment of the loss as

$$
\mathrm{E}\left[((cx) \wedge u)^2\right] = c^2(\alpha+1)\alpha\beta^2\gamma_{\frac{u}{c\beta}}(\alpha+2) + u^2\left[1 - \gamma_{\frac{u}{c\beta}}(\alpha)\right],
$$

from which we can compute $\mathrm{Var}[(cx) \wedge u]$.

Now with the given values of $c = 1.1$ and $u = 20$ we obtain

$$
\mathrm{E}\left[(cx) \wedge u\right] = 12.4608 \qquad \text{and} \qquad \mathrm{Var}\left[(cx) \wedge u\right] = 25.9197.
$$

Using a Monte Carlo sample of 10,000 observations, we obtain estimates of the mean and variance of the loss as (these are the sample mean and

the sample variance of the simulated losses)

$$\hat{\mathrm{E}}\left[(cx)\wedge u\right]=12.5466\qquad\text{and}\qquad\widehat{\mathrm{Var}}\left[(cx)\wedge u\right]=25.7545.$$

The Monte Carlo standard error of $\hat{\mathrm{E}}[(cx)\wedge u]$ is

$$\sqrt{\frac{\widehat{\mathrm{Var}}\left[(cx)\wedge u\right]}{10{,}000}}=\sqrt{\frac{25.7545}{10{,}000}}=0.0507.$$

Thus, using normal approximation, the Monte Carlo estimate of the 95% confidence interval of the expected loss is

$$12.5466\pm(1.96)(0.0507)=(12.4471,\,12.6461),$$

which covers the true value of 12.4608.

The probability of the total damages not exceeding u is

$$\Pr\left(cX\leq u\right)=\int_0^{\frac{u}{c}}\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}\,dx=\gamma_{\frac{u}{c\beta}}(\alpha),$$

and we have

$$\gamma_{\frac{20}{(1.1)(3)}}(4) = 0.8541.$$

The Monte Carlo estimate of this probability is the sample proportion of $1.1X \leq 20$, which was found to be 0.8543. The 95% confidence interval of the true probability is

$$0.8543 \pm 1.96\sqrt{\frac{(0.8543)(1-0.8543)}{10{,}000}} = (0.8474,\ 0.8612),$$

which again covers the true probability. □

14.6 Variance Reduction Techniques

- Consider a deterministic problem with solution equal to $\mathrm{E}[h(X)]$, where X is a random variable (not necessarily uniformly distributed) and $h(\cdot)$ is an integrable function over the support of X.

- The **crude Monte Carlo** estimate of the solution is

$$\hat{\mathrm{E}}[h(X)] = \frac{1}{n}\sum_{i=1}^{n} h(x_i), \tag{14.37}$$

  where $x_1, \cdots, x_n$ are a random sample of X.

- The accuracy of this stochastic solution depends on the variance of the estimator.

14.6.1 Antithetic Variable

- Consider a Monte Carlo sample of two observations X_1 and X_2. If X_1 and X_2 are iid, The variance of the Monte Carlo estimator is

$$\mathrm{Var}\left(\hat{\mathrm{E}}[h(X)]\right) = \frac{\mathrm{Var}[h(X)]}{2}. \tag{14.38}$$

- If X_1 and X_2 are identically distributed as X, but *not independent*, then the variance of the Monte Carlo estimator is

$$\mathrm{Var}\left(\hat{\mathrm{E}}[h(X)]\right) = \frac{\mathrm{Var}[h(X)] + \mathrm{Cov}(h(X_1), h(X_2))}{2}. \tag{14.39}$$

- Now if $\mathrm{Cov}(h(X_1), h(X_2)) < 0$, the variance of the Monte Carlo estimator is reduced.

- Random numbers generated in such a way that the functional evaluations at these numbers are negatively correlated are said to be antithetic variables.

- If $X_1 \sim \mathcal{U}(0,1)$, then $X_2 = 1 - X_1 \sim \mathcal{U}(0,1)$ and is negatively correlated with X_1.

- It should be noted, however, that for the antithetic variable technique to work well, it is the negative correlation between $h(X_1)$ and $h(X_2)$ that is required.

14.6.2 Control Variable

- To estimate $\mathrm{E}[h(X)]$ using control variable, we consider an auxiliary function $g(\cdot)$ and the associated expectation $\mathrm{E}[g(X)]$. We select the function $g(\cdot)$ so that it is close to $h(\cdot)$ and yet $\mathrm{E}[g(X)]$ is *known*.

- Now a Monte Carlo estimate of $\mathrm{E}[h(X)]$ can be computed as

$$\hat{\mathrm{E}}_{\mathrm{CV}}[h(X)] = \mathrm{E}[g(X)] + \frac{1}{n}\sum_{i=1}^{n}\left[h(X_i) - g(X_i)\right]. \tag{14.40}$$

- $\hat{\mathrm{E}}_{\mathrm{CV}}[h(X)]$ is an unbiased estimate of $\mathrm{E}[h(X)]$. The variance of this estimator is

$$\mathrm{Var}(\hat{\mathrm{E}}_{\mathrm{CV}}[h(X)]) = \frac{\mathrm{Var}\left[h(X) - g(X)\right]}{n}, \tag{14.41}$$

  which is smaller than the variance of $\hat{\mathrm{E}}_{\mathrm{CR}}\left[h(X)\right]$ if

$$\mathrm{Var}\left[h(X) - g(X)\right] < \mathrm{Var}\left[h(X)\right]. \tag{14.42}$$

Example 14.10: Consider the distribution of the loss in a loss event variable X_L in Example 14.4. Estimate $\mathrm{E}(X_L)$ using a Monte Carlo simulation with control variable.

Solution: To estimate $\mathrm{E}(X_L)$ using control variable, we consider a

random variable $\tilde{X}_L$ with the following df

$$F_{\tilde{X}_L}(x) = \begin{cases} 0.3606, & \text{for } x = 0, \\ 0.3606 + 0.0510x, & \text{for } 0 < x < 7, \\ 1, & \text{for } x \geq 7, \end{cases}$$

where

$$0.0510 = \frac{0.7177 - 0.3606}{7} = \frac{0.3571}{7}$$

is the slope of the line joining the points (0, 0.3606) and (7, 0.7177). Comparing the above with equation (14.9) we can see that the df of X_L in the interval $[0.3606, 0.7177)$ is now *linearized*. The mean of $\tilde{X}_L$ is

$$\mathrm{E}(\tilde{X}_L) = (0.3571)(3.5) + (0.2823)(7) = 3.2260.$$

Given a $\mathcal{U}(0, 1)$ variate U_i, X_L can be generated using equation (14.10),

and we denote this by $X_L(U_i)$. Now the inverse transformation of $\tilde{X}_L$ is

$$\tilde{X}_L = \begin{cases} 0, & \text{for } 0 \le U < 0.3606, \\ (U - 0.3606)/0.0510, & \text{for } 0.3606 \le U < 0.7177, \\ 7, & \text{for } 0.7177 \le U < 1. \end{cases}$$

Hence, the Monte Carlo estimate of $\mathrm{E}(X_L)$ using the control variable $\tilde{X}_L$ is computed as

$$3.2260 + \frac{1}{n} \sum_{i=1}^{n} \left[X_L(U_i) - \tilde{X}_L(U_i) \right].$$

Note that $X_L(U_i) - \tilde{X}_L(U_i)$ is nonzero only when $U_i \in [0.3606, 0.7177)$, in which case we have

$$X_L(U_i) - \tilde{X}_L(U_i) = 5 \left[-\log (1 - U_i) \right]^2 - 1 - \frac{U_i - 0.3606}{0.0510}.$$

We performed a Monte Carlo simulation with $n = 10{,}000$. The sample

mean and sample variance are

$$\hat{\mathrm{E}}_{\mathrm{CV}}(X_L) = 2.8650, \qquad s^2_{\mathrm{CV}} = 0.3150.$$

Thus, there is a substantial increase in efficiency versus the crude Monte Carlo and Monte Carlo with antithetic variable.

14.6.3 Importance Sampling

- Consider the following integral of a smooth integrable function $h(\cdot)$ over the interval $[a, b]$

$$\int_a^b h(x)\, dx, \tag{14.43}$$

which can be re-written as

$$\int_a^b [(b-a)h(x)] \frac{1}{b-a}\, dx. \tag{14.44}$$

- Thus, the integral can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n}(b-a)h(X_i), \tag{14.45}$$

  where X_i are iid $\mathcal{U}(a,b)$.

- If $\tilde{X}$ is a random variable with support $[a,b]$ and pdf $q(\cdot)$, the integral in equation (14.43) can be written as

$$\int_a^b h(x)\,dx = \int_a^b \left[\frac{h(x)}{q(x)}\right] q(x)\,dx, \tag{14.46}$$

  which can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{h(\tilde{X}_i)}{q(\tilde{X}_i)}, \tag{14.47}$$

where $\tilde{X}_i$ are iid as $\tilde{X}$. The estimator in equation (14.47) has a smaller variance than the estimator in equation (14.45) if

$$\text{Var}\left[\frac{h(\tilde{X}_i)}{q(\tilde{X}_i)}\right] < \text{Var}\left[(b-a)h(X_i)\right]. \tag{14.48}$$

- The advantage of importance sampling is likely to be large if the variation in the ratio $h(\cdot)/q(\cdot)$ is small over the interval $[a, b]$ (i.e., the two functions are close to each other).

Example 14.11: Consider the distribution of the loss in a loss event variable X_L in Example 14.4. Estimate $\text{E}(X_L)$ using a Monte Carlo simulation with importance sampling.

Solution: Defining $h(U)$ as $X_L(U)$ in equation (14.10), we have

$$\text{E}(X_L) = \int_0^1 h(x)\,dx$$

$$
\begin{aligned}
&= \int_{0.3606}^{0.7177} \left(5\left[-\log(1-x)\right]^2 - 1\right) dx + \int_{0.7177}^{1} 7\, dx \\
&= \int_{0.3606}^{0.7177} \left(5\left[-\log(1-x)\right]^2 - 1\right) dx + 1.9761.
\end{aligned}
$$

The integral above is the expected value of

$$(0.7177 - 0.3606)(5[-\log(1-\tilde{U})]^2 - 1),$$

where $\tilde{U} \sim \mathcal{U}(0.3606,\, 0.7177)$. Thus, we estimate $\mathrm{E}(X_L)$ by

$$1.9761 + \frac{0.3571}{n} \sum_{i=1}^{n} (5[-\log(1-\tilde{U}_i)]^2 - 1),$$

where

$$\tilde{U}_i = 0.3606 + 0.3571 U_i,$$

with $U_i \sim$ iid $\mathcal{U}(0, 1)$.

We performed a Monte Carlo simulation with 10,000 observations, and obtained $\hat{\mathrm{E}}(X_L) = 2.8654$, with a sample variance of 0.4937. Thus, the importance sampling method produced a big gain in efficiency over the crude Monte Carlo method, although the gain is not as much as that from the control variable method in Example 14.10. □