Math 5040–1
Stochastic Processes & Simulation

Davar Khoshnevisan
University of Utah

Module 1

Generation of Discrete Random Variables

Just about every programming language and environment has a "random-number" generator. This means that it has a routine for simulating a $\mathrm{Unif}(0\,,1)$ random variable; that is a continuous random variable with density function $f(x) = 1$ if $0 \leqslant x \leqslant 1$ and $f(x) = 0$ for other values of x. Throughout, I assume that we know how to generate a $\mathrm{Unif}(0\,,1)$ random variable U, using an existing computer program.[1] Our present goal is to use this knowledge to simulate a discrete random variable with a prescribed mass function f.

1. Bernoulli samples

Recall that a Bernoulli(p) random variable X takes the values 1 and 0 with respective probabilities p and $1-p$. We can think of this as the outcome of a coin toss: $X = 1$ means that we tossed heads; and $X = 0$ refers to tails.

Let U be a $\mathrm{Unif}(0\,,1)$ random variable, and define $X = \mathbf{1}_{\{U \leqslant p\}}$, where $\mathbf{1}_A$ denotes the indicator of an event A [i.e., it is one if A occurs; and zero otherwise]. The possible values of X are zero and one. Because $X = 1$ if and only if $U \leqslant p$,

$$P\{X = 1\} = P\{U \leqslant p\} = \int_0^p dx = p.$$

In other words, X is a Bernoulli(p) sample.

[1] One might ask, how is *that* done? That is a a good question, which is the subject for a quite different course.

In algorithmic terms, this method translates to the following: If $U \leqslant p$ then $X := 1$; else if $p < U \leqslant 1$ then $X := 0$.

2. More general discrete distributions

Now let us try to simulate a random variable with mass function

$$f(x) = \begin{cases} p_0 & \text{if } x = x_0, \\ p_1 & \text{if } x = x_1, \\ \vdots, \end{cases}$$

where $0 \leqslant p_j \leqslant 1$ for all $j \geqslant 0$, and $p_0 + p_1 + \cdots = 1$. That is, the possible values of X ought to be $x_0, x_1, \ldots$ with respective probabilities $p_0, p_1, \ldots$. We extend the algorithm of the previous subsection as follows: If $U \leqslant p_0$ then $X := x_0$; else if $p_0 < U \leqslant p_0 + p_1$ then $X := x_1$; else if $p_0 + p_1 < U \leqslant p_0 + p_1 + p_2$ then $X := x_2$, etc. At the kth stage, where $k \geqslant 1$ is integral, this algorithm performs the following query:

Is $p_0 + \cdots + p_{k-1} < U \leqslant p_0 + \cdots + p_k$?
If so, then set $X := x_k$; else, proceed to stage $k + 1$.

Here is the pseudo-code:

```
1. Generate a Unif(0,1) random variable U.
2. If U <= p0 then set X = x0 and stop.
3. Else set k=1, PrevSum = p0, and Sum = p0 + p1.
4. If PrevSum <= U < Sum then set X = xk and stop.
5. Else, set k = k+1, PrevSum = Sum, and Sum = Sum + pk;
   then goto Step 4.
```

[You should check that this does the job.]

Question 1. Could it be that this algorithm never stops?

Fortunately, the answer is "no." In order to see this let N denote the time that the algorithm stops; $N = \infty$ is a distinct possibility at this point. Now we can observe that for all integers $k \geqslant 1$,

$$\begin{aligned} P\{N > k\} &= P\{\text{the algorithm does not stop after the } k\text{th stage}\} \\ &= P\{U > p_0 + \cdots + p_{k-1}\} = 1 - P\{U \leqslant p_0 + \cdots + p_{k-1}\} \\ &= 1 - \sum_{j=0}^{k-1} p_j = \sum_{j=k}^{\infty} p_j. \end{aligned} \tag{1}$$

And so $P\{N = \infty\} = \lim_{k\to\infty} P\{N > k\} = 0$.

2.1. Discrete-uniform samples. Suppose we wish to simulate a random variable X that is distributed uniformly on $\{1, \ldots, n\}$ for a positive integer n that is fixed. For all $j = 1, \ldots, n$, the jth stage of our algorithm tells us to set

$$X = j \qquad \text{if} \qquad \frac{j-1}{n} < U \leqslant \frac{j}{n}.$$

The condition on U is equivalent to $j - 1 < nU \leqslant j$, and this means that $j = 1 + \lfloor nU \rfloor$, where $\lfloor x \rfloor$ denotes the greatest integer $\leqslant x$. The preceding can be streamlined into a one-line description as follows:

$$X := 1 + \lfloor nU \rfloor \quad \text{is distributed uniformly on } \{1, \ldots, n\}.$$

Exercise 1. Recall the Geometric(p) mass function is given by the following: $f(j) = p(1-p)^{j-1}$ if $j = 0, 1, 2, \ldots$ and $f(j) = 0$ otherwise. Prove the following:

$$X := 1 + \left\lfloor \frac{\ln U}{\ln(1-p)} \right\rfloor \qquad \text{has the Geometric}(p) \text{ mass function.}$$

2.2. Binomial samples. The Binomial(n, p) mass function is given by

$$f(x) = \begin{cases} \dbinom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, \ldots, n, \\ \\ 0 & \text{otherwise.} \end{cases}$$

In the jargon of our algorithm, $x_0 = 0, x_1 = 1, \ldots, x_n = n$; and $p_j = \binom{n}{j} p^j (1-p)^{n-j}$. Note that $p_0 = p^n$ and

$$\frac{p_{j+1}}{p_j} = \frac{n-j}{j} \times \frac{p}{1-p}.$$

Using this, check that the following alogrithm produces a Binomial(n, p) simulation:

```
1. Generate a Unif(0,1) random variable U.
2. Set M = p/(1-p), j=0, P = q^n, and Sum = P.
3. If U <= Sum, then set X = j and stop.
4. Else set j=j+1, P = ( (n-j)/j * M ) * P, and Sum = Sum + P;
   then go to Step 3.
```

Question 2. At most how long does this algorithm take before it stops?

Exercise 2. Prove that if $X = \text{Binomial}(n, p)$, then $X = I_1 + \cdots + I_n$ where the I_j's are independent Bernoulli(p)'s. Use this to prescribe an alternative algorithm for simulating X.

2.3. Poisson samples. The Poisson(λ) mass function is given by

$$f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{if } x = 0, 1, \dots, \\ \\ 0 & \text{otherwise.} \end{cases}$$

In the jargon of our algorithm, $x_0 = 0, x_1 = 1, \dots$; and $p_j = e^{-\lambda}\lambda^j/j!$. Note that $p_0 = \exp(-\lambda)$ and

$$\frac{p_{j+1}}{p_j} = \frac{\lambda}{j+1}.$$

Using this, check that the following algorithm produces a Poisson(λ) simulation:

```
1. Generate a Unif(0,1) random variable U.
2. Set j = 0, P = exp(-lambda), and Sum = P.
3. If U <= Sum, then set X = j and stop.
4. Else set P = ( lambda / (j+1) ) * P and Sum = Sum + P;
   then go to Step 3.
```

Question 3. How long before this algorithm stops?

The answer requires first the following result.

Exercise 3. Prove that if $V \geqslant 1$ is integral, then $EV = \sum_{k=1}^{\infty} P\{V \geqslant k\}$. [Hint: Start by writing the sum as $\sum_{k=1}^{\infty}\sum_{j=k}^{\infty} P\{V = j\}$ and then interchanging the double sum.]

In order to answer Question 3, recall the definition of N, and apply Exercise 3 to find that

$$EV = \sum_{k=1}^{\infty} P\{N \geqslant k\} = \sum_{k=1}^{\infty} P\{N > k-1\} = \sum_{i=0}^{\infty} P\{N > i\}.$$

Thanks to this and equation (1) on page 2,

$$EV = \sum_{i=0}^{\infty}\sum_{j=i}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} = \sum_{j=0}^{\infty}\sum_{i=0}^{j} \frac{e^{-\lambda}\lambda^j}{j!} = \sum_{j=0}^{\infty} (j+1)\,\frac{e^{-\lambda}\lambda^j}{j!}.$$

We split up $j+1$ as follows:

$$\begin{aligned} EV &= \sum_{j=0}^{\infty} j\frac{e^{-\lambda}\lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \\ &= \sum_{j=1}^{\infty} jp_j + \sum_{j=0}^{\infty} p_j = \lambda + 1. \end{aligned}$$

I have used the fact that the expectation of a Poisson(λ) is λ. In summary, we expect to stop "after $\lambda + 1$ steps."

3. Monte Carlo simulation

Now I describe the "Monte Carlo method," which is a popular application of simulation: Suppose we need to know the value of $\mathrm{E}\{h(X)\}$ where h is a function and X a random variable. There are many complicated examples wherein this expectation cannot be computed analytically. Then we can some times resort to simulation as follows: Simulate $X_1, X_2, \ldots$ [independent random variables] from the same distribution as X. By the law of large numbers,

$$\frac{1}{N}\sum_{j=1}^{N} h(X_j) \approx \mathrm{E}\{h(X)\} \qquad \text{with high probability if } N \text{ is large.}$$

Therefore, if we want to simulate $\mathrm{P}\{X \in A\}$ for a set A, then we simulate independent random variables $X_1, X_2, \ldots$ from the same distribution as that of X, and then use the preceding with $h(x) := \mathbf{1}\{x \in A\}$ to find that

$$\frac{\#\{1 \leqslant i \leqslant N : \ X_i \in A\}}{N} \approx \mathrm{P}\{X \in A\} \qquad \text{with high probability if } N \text{ is large.}$$

Exercise 4. Let $X = \text{Binomial}(10\,, \frac{1}{2})$. Simulate, using $N = 1000$ random variables, $\mathrm{P}\{X = k\}$ for all $k = 0\,, \ldots, 10$. Verify the accuracy of your simulation by comparing your simulation results to the exact results.

Exercise 5. Simulate $\mathrm{P}\{20 \leqslant X \leqslant 24\}$ where $X = \text{Poisson}(4)$.

Exercise 6 (The CLT)**.** Let $\{X_{i,j}\}_{i,j=1}^{\infty}$ be a double–array of independent random variables, all with common mass function

$$f(x) = \begin{cases} \dfrac{1}{2} & \text{if } x = 1 \text{ or } x = -1, \\ \\ 0 & \text{otherwise.} \end{cases}$$

Let $S_{n,j} := X_{1,j} + X_{2,j} + \cdots + X_{n,j}$, where n is large. Suppose m is another large integer. Explain (in clear heuristic terms) why

$$\frac{\#\left\{1 \leqslant j \leqslant m : \ a < S_{n,j} \leqslant b\right\}}{m} \approx \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,dx \qquad \text{with high probability.}$$

Use this simulation method to estimate $(2\pi)^{-1/2}\int_1^2 \exp(-x^2/2)\,dx$.

Module 2

Discrete Markov Chains

1. Introduction

A *stochastic* [*or random*] *process* is a collection $\{X_n\}_{n=0}^{\infty}$ of random variables. They might take values in very abstract spaces [e.g., $P\{X_1 = \text{cow}\} = \frac{1}{2} = P\{X_1 = \text{horse}\}$].

Let $\mathcal{S}$ denote a finite or countable set of distinct elements. Let $X := \{X_n\}_{n=0}^{\infty}$ be a stochastic process with values in $\mathcal{S}$. Frequently, we think of n as "time," and X_n as the "state" of the process at time n.

Definition 1. We say that X is a *Markov chain* on $\mathcal{S}$ if

$$P\{X_{n+1} = a_{n+1} \mid X_0 = x_0, \ldots, X_n = a_n\} = P\{X_1 = a_{n+1} \mid X_0 = a_n\}, \quad (2)$$

for all integers $n \geqslant 0$ and every $a_0, \ldots, a_{n+1} \in \mathcal{S}$. Equation (2) is called the *Markov property*.

In order to understand what this means, we should think about how we would simulate the time-evolution of the random process X:

- We start with some initial random variable X_0.
- Given that $X_0 = a_0$, we simulate X_1 according to the mass function $f(x) = P\{X_1 = x \mid X_0 = a_0\}$.
- ...
- Suppose we have simulated $X_0, \ldots, X_n$. Given that $X_0 = a_0, \ldots, X_n = a_n$, then we simulate X_{n+1} according to the *same* mass function f that was used earlier.

It should be clear that we ought to try to understand a Markov chain based on the properties its *transition probabilities*; these are the collection of probabilities

$$P_{x,y} := P\{X_1 = y \mid X_0 = x\} \qquad \text{for all } x, y \in \mathcal{S}.$$

We can think of the collection of these transition probabillities as a matrix $\mathbf{P}$ whose (x, y) coordinate is $P_{x,y}$. We call $\mathbf{P}$ the *transition matrix* of the Markov chain X. This definition makes sense also when $\mathcal{S}$ is infinite. But one has to be much more careful when one deals with infinite matrices.

1.1. A simple weather-forecasting model (Ross, p. 182). Consider the following overly-simplified model for weather changes: If it rains today, then regardless of the weather of the previous days, tomorrow will be rainy with probability α; if it is dry today, then regardless of the weather of the previous days, tomorrow will be rainy with probability β. Here, $\mathcal{S} := \{0, 1\}$ where state “0” refers to “rain” and “1” to “dry.” Our Markov chain will take the values zero and one according to the transition matrix

$$\mathbf{P} = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Exercise 1. Simulate this Markov chain. We will soon see that the following limits exist:

$$\boldsymbol{\pi}_y := \lim_{n\to\infty} P\{X_n = y \mid X_0 = x\} \qquad \text{for } y, x \in \{0, 1\},$$

and $\boldsymbol{\pi}$ does not depend on x. Apply this fact to compute $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}_1$ via Monte Carlo simulation. These are the respective probabilities of rainy–versus–dry days in the “long run.”

1.2. The simple walk. The simple [random] walk on $\mathbf{Z}$ is the Markov chain whose transitions are described as

$$P_{x,y} = \begin{cases} \dfrac{1}{2} & \text{if } y = x+1 \text{ or } y = x-1, \\ \\ 0 & \text{otherwise.} \end{cases}$$

That is, at any given time, we go to the left or right of our present position with probability $\frac{1}{2}$, independently of the how we arrived at our present position.

More generally, a simple walk on [the vertices of] a graph $\mathbf{G}$ moves to each of its current neighbors with equal probability, at all times.

Exercise 2. Let $X := \{X_n\}_{n=0}^{\infty}$ be the simple walk on $\mathbf{Z}$, and suppose $X_0 = 0$ [with probability one]. For every integer x let T_x denote the first time n such that $X_n = x$. Use Monte Carlo simulation to convince yourself that $P\{T_1 < T_{-1}\} = \frac{1}{2}$. Can you prove that this actually holds? [It does.]

Exercise 3. Let X denote the simple walk on $\mathbf{Z}$ with $X_0 = 0$.

(1) Prove that $Y_n := X_n - X_{n-1}$ $[n \geqslant 1]$ defines a collection of independent random variables with $P\{Y_n = 1\} = P\{Y_n = -1\} = \frac{1}{2}$.

(2) Prove that X_n/n goes to zero in probability; that is, for all $\lambda > 0$,
$$\lim_{n\to\infty} P\left\{\left|\frac{X_n}{n}\right| > \lambda\right\} = 0.$$

(3) Use Monte Carlo simulation to verify this phenomenon.

2. The Chapman–Kolmogorov equations

By using the transition matrix $\mathbf{P}$ of the Markov chain X, we can compute all sorts of probabilities. For instance, suppose we want the distribution of X_2, given that the initial value of the chain is $X_0 = a$ for some constant $a \in \mathcal{S}$ such that $P\{X_0 = a\} > 0$. We apply the law of total probabilities to find this as follows:
$$\begin{aligned}
&P\{X_2 = b \mid X_0 = a\}\\
&= \frac{P\{X_0 = a\,,\, X_2 = b\}}{P\{X_0 = a\}}\\
&= \frac{1}{P\{X_0 = a\}} \sum_{c\in\mathcal{S}} P\{X_0 = a\,,\, X_1 = c\,,\, X_2 = b\}\\
&= \frac{1}{P\{X_0 = a\}} \sum_{c\in\mathcal{S}} P\left\{X_2 = b \mid X_0 = a\,,\, X_1 = c\right\} \times P\{X_0 = a\,,\, X_1 = c\}\\
&= \frac{1}{P\{X_0 = a\}} \sum_{c\in\mathcal{S}} P_{c,b} \times P\{X_0 = a\,,\, X_1 = c\}.
\end{aligned}$$
Another round of conditioning shows us that
$$P\{X_0 = a\,,\, X_1 = c\} = P\{X_0 = a\} \times P_{a,c}.$$
Thus,
$$P\{X_2 = b \mid X_0 = a\} = \sum_{c\in\mathcal{S}} P_{a,c}P_{c,b} = P^2_{a,b},$$
where $P^2_{a,b}$ denotes the $(a\,,b)$ element of the matrix $\mathbf{P}^2 := \mathbf{P}\mathbf{P}$. The following is only slightly more general, and is proved similarly.

Theorem 1. *For all integers $n \geqslant 0$, and all $a, b \in \mathcal{S}$,*
$$P\{X_n = b \mid X_0 = a\} = P^n_{a,b},$$
where $P^n_{a,b}$ represents the $(a\,,b)$ coordinate of the matrix $\mathbf{P}^n := \mathbf{P}\cdots\mathbf{P}$ [n times], where $\mathbf{P}^0 := \mathbf{I}$ denotes the identity matrix.

In summary, $P^n_{a,b}$ is the probability that the Markov chain X goes from a to b in n [time] steps.

Exercise 4. Prove that for all integers $n, k \geqslant 0$, and $a, b \in \mathcal{S}$,

$$P\{X_{n+k} = b \mid X_k = a\} = P\{X_n = b \mid X_0 = a\} = P^n_{a,b}.$$

Here is a consequence of all of this:

$$P^{n+m}_{a,b} = \sum_{c \in \mathcal{S}} P^n_{a,c} P^m_{c,b}.$$

The collection of all these equations—as we vary n, m, a, and b—is called the *Chapman–Kolmogorov equations*; it turns out to have many useful consequences.

2.1. Simplified weather forecasting. Consider the simplified weather forecasting problem (§1.1, page 8), with $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{4}$. The transition matrix is given by the following:

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}.$$

Our present goal is to compute $\lim_{n\to\infty} P^n_{x,y}$; that is, the long-run probability of making a transition from state x to state y.

One can easily compute

$$\mathbf{P}^2 = \begin{pmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{pmatrix}.$$

And with enough patience $\mathbf{P}^3$. But how does one compute $\lim_{n\to\infty} \mathbf{P}^n$? Or is it even clear that the limit exists?

Exercise 5. Verify, numerically, that

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}. \tag{3}$$

Additionally, make sure that you understand the following interpretation of (3): Regardless of whether or not we start with a rainy or dry day, after a very long time any given day is rainy with probability $\approx 1/3$ and dry with probability $\approx 2/3$. A heuristic interpretation of the latter remark is that it is going to rain about one-third of the time!

And now we return to do some honest mathematics. The remainder of this discussion requires a little bit of linear algebra.

Here is a method that works quite generally: Let $\mathbf{v} := (1/\sqrt{5}, 2/\sqrt{5})$ and $\mathbf{w} := (1/\sqrt{2}, -1/\sqrt{2})$. Then, $\mathbf{vP} = \mathbf{v}$ and $\mathbf{wP} = \frac{1}{4}\mathbf{w}$. That is, $\mathbf{v}$ and $\mathbf{w}$ are the two left-eigenvectors of $\mathbf{P}$, and correspond respectively to the eigenvalues 1 and 1/4. Note that $\mathbf{v}$ and $\mathbf{w}$ have been normalized to have unit length. More significantly, $\mathbf{v}$ and $\mathbf{w}$ together span all of $\mathbf{R}^2$. That is, every 2-dimensional vector $\mathbf{u}$ can be written as follows:

$$\mathbf{u} = a\mathbf{v} + b\mathbf{w}, \tag{4}$$

with

$$a = \frac{\sqrt{5}}{3}\left(u_1 + u_2\right) \qquad \text{and} \qquad b = \frac{\sqrt{2}}{3}\left(2u_1 - u_2\right).$$

[You need to check all of these computations!] Now we can right-multiply both sides of (4) by $\mathbf{P}^n$ to find that

$$\mathbf{u}\mathbf{P}^n = a\mathbf{v}\mathbf{P}^n + b\mathbf{w}\mathbf{P}^n.$$

Induction reveals that $\mathbf{v}\mathbf{P}^n = \mathbf{v}$ and $\mathbf{w}\mathbf{P}^n = (1/4)^n\mathbf{w}$ for all integers $n \geqslant 1$. Consequently,

$$\mathbf{u}\mathbf{P}^n = a\mathbf{v} + \frac{b}{4^n}\mathbf{w},$$

whence it follows that

$$\lim_{n\to\infty} \mathbf{u}\mathbf{P}^n = a\mathbf{v} = \frac{\sqrt{5}}{3}\left(u_1 + u_2\right)\mathbf{v} = \mathbf{u}\begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

[Check by plugging in the explicit value of $\mathbf{v}$ etc.] This verifies (3) (p. 10).

Exercise 6 (Harder)**.** Prove that if

$$\mathbf{P} = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix},$$

then

$$\lim_{n\to\infty} P^n_{0,0} = \lim_{n\to\infty} P^n_{1,0} = \frac{\beta}{1-\alpha+\beta} \quad \text{and} \quad \lim_{n\to\infty} P^n_{0,1} = \lim_{n\to\infty} P^n_{1,1} = \frac{1-\alpha}{1-\alpha+\beta}.$$

Conclude, in words, that the long-run proportion of rainy days is equal to $\beta/(1-\alpha+\beta)$.

Exercise 7. Let $\mathbf{P}$ denote the transition matrix of a Markov chain on a finite state space $\mathcal{S}$.

(1) Prove that $\mathbf{P1} = \mathbf{1}$ where $\mathbf{1}$ is a vector of all ones. That is, $\lambda = 1$ is a right-eigenvalue of $\mathbf{P}$ with right-eigenvector $\mathbf{1}$.

(2) Use the preceding to prove that $\lambda = 1$ is also a left-eigenvalue of $\mathbf{P}$, and the corresponding eigenvector $\mathbf{v}$ solves $\mathbf{vP} = \mathbf{v}$.

(3) Suppose that there exists a unique left-eigenvector $\mathbf{v}$ for $\mathbf{P}$ such that $\nu_j > 0$ for all j. Prove that in that case, $\lim_{n\to\infty} P^n_{a,b}$ exists for all $b \in \mathcal{S}$ and does not depend on a.

Exercise 8. (1) Prove that the following is the transition matrix of a Markov chain with values in $\mathcal{S} := \{0\,,1\,,2\}$:

$$\mathbf{P} := \begin{pmatrix} \alpha & 1-\alpha & 0 \\ \beta & 1-\beta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where $0 < \alpha, \beta < 1$.

(2) How many left-eigenvectors correspond to the left-eigenvalue $\lambda = 1$? Do any of these eigenvectors have positive coordinates?

(3) Does $\lim_{n\to\infty} P^n_{a,b}$ exist? If so, then can you compute it for all $a, b \in \{0\,,1\,,2\}$?

Exercise 9 (Computer exercise)**.** Consider the transition matrix

$$\mathbf{P} := \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

Describe the behavior of $\mathbf{P}^n$ for large values of n.

Generation of Continuous Random Variables

1. The inverse-transform method

This is the most basic method for generating continuous random variables that have "nice" densities.

Suppose F is the distribution function of a continuous random variable with a proper inverse F^{-1}.

Theorem 1. *If* $U = \mathrm{Unif}(0\,,1)$, *then* $X := F^{-1}(U)$ *is a sample from* F.

Proof. We compute the distribution function of X:

$$P\{X \leqslant x\} = P\left\{F^{-1}(U) \leqslant x\right\} = P\{U \leqslant F(x)\},$$

since F is increasing. Note that $y = F(x)$ is a number between zero and one. Since $P\{U \leqslant y\} = y$ for such y's, it follows that $P\{X \leqslant x\} = F(x)$ for all x. This has the desired effect. □

1.1. Exponential samples. Recall that X is Exponential(λ) if its density is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

The distribution function is given, for all $x \geqslant 0$, as follows:

$$F(x) := \int_{-\infty}^{x} \lambda e^{-\lambda y}\, dy = 1 - e^{-\lambda x}.$$

If $0 < y < 1$ and $y = F(x)$, then $y = 1 - e^{-\lambda x}$. Therefore, we solve for x to find that $x = (1/\lambda)\ln(1-y)$. That is,

$$F^{-1}(y) = -\frac{1}{\lambda}\ln(1-y) \qquad \text{if } 0 < y < 1.$$

In particular, if $U = \text{Unif}(0\,,1)$, then $X := (1/\lambda)\ln(1-U)$ is an Exponential(λ) random variable. This can be simplified further slightly. Note that $V := 1-U$ is also a Unif$(0\,,1)$ sample. So, in order to generate a Exponential(λ) random variable X, we first generate a Unif$(0\,,1)$ random variable V, then set

$$X := -\frac{1}{\lambda}\ln V.$$

Our discussion implies that X has the Exponential(λ) distribution.

Exercise 1 (Gamma densities)**.** If α and λ are positive and fixed, then the following is a probability density:

$$f(x) := \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{if } x > 0, \\ \\ 0 & \text{otherwise,} \end{cases}$$

where Γ denotes the *Gamma function* which, I recall, is

$$\Gamma(t) := \int_0^\infty x^{t-1} e^{-x}\, dx \qquad \text{for all } t > 0. \tag{5}$$

Show that if $X = \text{Gamma}(k\,,\lambda)$ for an *integer* $k \geqslant 1$, then $X = T_1 + \cdots + T_k$ where the T_j's are independent Exponential(λ)'s. Use this to describe, carefully and in detail, a method for simulating Gamma$(\alpha\,,\lambda)$ in the case that $\alpha \geqslant 1$ is integral.

[Hint: Moment generating functions.]

1.2. Paréto samples. A random variable X is said to have a *Paréto* density with parameter $\alpha > 0$ if

$$P\{X > x\} = x^{-\alpha} \qquad \text{for all } x \geqslant 1,$$

and $P\{X > x\} = 1$ if $x < m$.

Exercise 2 (Easy)**.** Compute the Paréto(α) density function.

The distribution function F of X is described as follows: $F(x) = 1 - x^{-\alpha}$ if $x \geqslant 1$ and $F(x) = 0$ if $x < 1$. If $0 < y < 1$ satisfies $y = F(x)$ for some x, then x has to be at last 1 [why?] and $y = 1 - x^{-\alpha}$. Solve for x to find that $x = (1-y)^{-1/\alpha}$. That is,

$$F^{-1}(y) = (1-y)^{-1/\alpha} \qquad \text{if } 0 < y < 1.$$

Because $1 - U = \text{Unif}(0\,,1)$, it follows that $U^{-1/\alpha} = \text{Paréto}(\alpha)$.

Exercise 3 (Weibull density)**.** Compute the density of $X := T^{1/\alpha}$, where $T = \text{Exponential}(\lambda)$ and $\alpha, \lambda > 0$. Explain how you would simulate X.

Exercise 4 (Cauchy density)**.** Show that if $U = \text{Unif}(0\,,1)$, then $C := \tan\{\pi(U - \frac12)\}$ has the following density function:

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad \text{for } -\infty < x < \infty.$$

Exercise 5 (Rayleigh density)**.** A random variable X is said to have a *Rayleigh density* if $X \geqslant 0$, and

$$P\{X > x\} = e^{-x^2} \qquad \text{for } x > 0.$$

Explain in detail how you would simulate from a Rayleigh density.

Exercise 6 (Beta($\frac12\,,\frac12$) density)**.** Consider the density function

$$f(x) := \begin{cases} \dfrac{1}{\pi\sqrt{x(1-x)}} & \text{if } 0 < x < 1, \\[2em] 0 & \text{otherwise.} \end{cases}$$

Show that if $U = \text{Unif}(0\,,1)$, then $Y := \sin^2(\pi\{U - \frac12\})$ has density f.
[Suggestion: First consider $X := \sin(\pi\{U - \frac12\})$ and then $Y := X^2$.]

2. The acceptance–rejection method

Some times the distribution function, and hence its inverse, are hard to compute. In some such cases, one can employ the so-called acceptance–rejection method. This is an iterative method which frequently converges rapidly.

Consider two density functions f and g that satisfy the following "domination condition" for a finite constant c:

$$f(x) \leqslant cg(x) \qquad \text{for all } x. \tag{6}$$

We suppose we know how to sample from g; and propose to use that fact in order to simulate from f.

Exercise 7 (Easy)**.** Prove that $c \geqslant 1$.

Let $Y_1, Y_2, \dots$ be an independent sequence of random variables with density function g, and $U_1, U_2, \dots$ an independent sequence of $\mathrm{Unif}(0\,,1)$ random variables, totally independent of the Y's. Define a random variable N to be the smallest $k \geqslant 1$ such that

$$U_k \leqslant \frac{f(Y_k)}{cg(Y_k)}. \tag{7}$$

That is, if the preceding condition holds, then we accept Y_k as the simulated value of X.

Theorem 2. *We have* $P\{N < \infty\} = 1$ *and* $EN = c$*. Moreover,* $X := Y_N$ *is a continuous random variable with density function* f*.*

Proof. First, we make a few observations:

- Let A denote the collection of all points y such that $g(y) = 0$. Then
$$P\{g(Y_k) = 0\} = P\{Y_k \in A\} = \int_A g(z)\,dz = 0.$$
Therefore, (7) is well defined with probability one.
- For every integer $k \geqslant 1$,
$$\begin{aligned} P\left\{U_k \leqslant \frac{f(Y_k)}{cg(Y_k)}\right\} &= \int_{-\infty}^{\infty} P\left\{U_k \leqslant \frac{f(z)}{cg(z)}\right\} g(z)\,dz \\ &= \int_{-\infty}^{\infty} \frac{f(z)}{cg(z)} g(z)\,dz = \frac{1}{c}\int_{-\infty}^{\infty} f(z)\,dz = \frac{1}{c}. \end{aligned}$$
- Therefore, $N = \mathrm{Geometric}(1/c)$. In particular, $P\{N < \infty\} = 1$. In fact, we also have $EN = c$.

In order to conclude the proof, it suffices to show that $P\{Y_N \leqslant x\} = \int_{-\infty}^{x} f(z)\,dz$ for all x. Note that we have established already that Y_N is well-defined, since $N < \infty$.

We compute:

$$\begin{aligned} P\{Y_N \leqslant x\} &= \sum_{k=1}^{\infty} P\{Y_k \leqslant x, N = k\} \\ &= \sum_{k=1}^{\infty} \int_{-\infty}^{x} P\{N = k \mid Y_k = z\}\, g(z)\,dz. \end{aligned}$$

Given that $Y_k = z$, the conditional probability of $N = k$ is

$$\prod_{j=1}^{k-1} P\left\{U_j > \frac{f(Y_j)}{cg(Y_j)}\right\} \times P\left\{U_k \leqslant \frac{f(z)}{cg(z)}\right\} = \left(1 - \frac{1}{c}\right)^{k-1} \times \frac{f(z)}{cg(z)}.$$

Therefore,

$$\begin{aligned} P\{Y_N \leqslant x\} &= \sum_{k=1}^{\infty} \int_{-\infty}^{x} \left(1 - \frac{1}{c}\right)^{k-1} \frac{f(z)}{cg(z)} g(z)\, dz \\ &= \int_{-\infty}^{x} f(z)\, dz, \end{aligned}$$

because $\sum_{k=1}^{\infty} r^{k-1} = (1-r)^{-1}$ for $0 < r < 1$. Differentiate both sides of the preceding to find that $f_{Y_N}(x) = f_{Y_1}(x)$. □

Check that the following produces a simulation of a random variable from f, using random variables from g [and some extra uniforms]:

```
1. Generate a Unif(0,1) random variable U.
2. Generate a random variable Y from g.
3. If U <= f(Y)/cg(Y), then set X = Y and stop.
4. Else, go to to Step 1.
```

2.1. Normal samples. Suppose we wish to simulate from the standard-normal density f:

$$f(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Consider the "double-exponential" density g:

$$g(x) := \frac{1}{2} e^{-|x|}.$$

Clearly,

$$\frac{f(x)}{g(x)} = \sqrt{\frac{2}{\pi}} \exp\left\{-\frac{x^2}{2} + |x|\right\} := \sqrt{\frac{2}{\pi}} e^{h(x)}.$$

The function h is clearly symmetric about zero. Moreover, if $x > 0$, then $h'(x) = -x + 1$ and $h''(x) = -1$. Therefore, h is maximized—on the positive axis—at $x = 1$ and $\max_{x>0} h(x) = h(1) = 1/2$. By symmetry, $\max_{x<0} h(x) = h(-1) = 1/2$, as well. Because $h(0) = 0$, we find that $\max_x h(x) = 1/2$, and hence

$$\max_x \frac{f(x)}{g(x)} = \sqrt{\frac{2e}{\pi}}.$$

That is, (6) holds with $c = \sqrt{2e/c}$.

It remains to know how one can sample from g. But that is not hard. Let $G(x) := \int_{-\infty}^{x} g(z)\,dz$ define the distribution function for g. If $x \leqslant 0$, then

$$G(x) = \frac{1}{2}\int_{-\infty}^{x} e^{z}\,dz = \frac{1}{2}e^{x}.$$

And if $x > 0$, then by the symmetry of g,

$$G(x) = \frac{1}{2} + \int_{0}^{x} g(z)\,dz = \frac{1}{2}\left(1 + \int_{0}^{x} e^{-z}\,dz\right) = 1 - \frac{1}{2}e^{-x}.$$

Note:

- If $0 < y \leqslant \frac{1}{2}$ and $G(x) = y$, then $\frac{1}{2}e^{x} = y$ whence $x = \ln(2y)$. That is, $G^{-1}(y) = \ln(2y)$ if $0 \leqslant y \leqslant \frac{1}{2}$.
- If $\frac{1}{2} < y < 1$ and $G(x) = y$, then $1 - \frac{1}{2}e^{-x} = y$ whence $G^{-1}(y) = -\ln(2(1-y))$.

To summarize,

$$G^{-1}(y) = \begin{cases} \ln(2y) & \text{if } 0 < y \leqslant \frac{1}{2}, \\ \\ -\ln(2(1-y)) & \text{if } \frac{1}{2} < y < 1. \end{cases}$$

Therefore, in order to generate a random variable Y from g we first generate a Unif$(0\,,1)$ random variable U, and then define $Y = G^{-1}(U)$. The acceptance–rejection method can now be applied; this yields a sample from the standard-normal density f.

Exercise 8. Prove that if $Z = N(0\,,1)$, $a \in \mathbf{R}$, and $b > 0$, then $X := a + b^{1/2}Z$ is $N(a\,,b)$. Apply this to describe, in detail and carefully, how one can simulate a $N(\mu\,,\sigma^2)$.

2.2. Beta samples. Choose and fix two constants $\alpha, \beta > 0$. The *beta density* with parameters α and β is:

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 < x < 1, \\ \\ 0 & \text{otherwise,} \end{cases}$$

where Γ is the Gamma function; see (5) on page 14. [I will take for granted that $\int_0^1 f(x)\,dx = 1$, though proving that requires work.] You should check that the Beta$(1\,,1)$ density is the same thing as Unif$(0\,,1)$. Now suppose $\alpha, \beta \geqslant 1$ and define g to be the Unif$(0\,,1)$ density. Then

$$f(x) \leqslant \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} g(x) \qquad \text{if } 0 < x < 1.$$

That is, equation (6) on page 15 holds with $c = \Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)$.

Exercise 9 (Beta density)**.** Let f denote the Beta$(\alpha\,,\beta)$ density with $1 > \alpha, \beta > \frac{1}{2}$. Let g denote the Beta$(\frac{1}{2}\,,\frac{1}{2})$ density and recall that $\Gamma(1/2) = \sqrt{\pi}$. First verify that (6) on page 15 holds with $c := \pi\Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)$. Then explain, clearly and in detail, how one can simulate from f.

[Hint: Consult Exercise 6 on page 15.]