1 Simulation

Question: My computer only knows how to generate a uniform random variable. How do I generate others?

1.1 Continuous Random Variables

Recall that a random variable X is *continuous* if it has a probability density function f_X so that

$$\mathbb{P}\{a \le X \le b\} = \int_a^b f_X(x)dx\,.$$

The distribution function F_X for X is defined as

$$\begin{aligned} F_X(x) &= \mathbb{P}\{X \le x\} \\ &= \int_{-\infty}^{x} f_X(s)ds\,. \end{aligned}$$

Notice that $F_X'(x) = f_X(x)$.

A uniform$[0,1]$ random variable U has density function

$$f_U(u) = \begin{cases} 1 & \text{if } 0 \le u \le 1\,, \\ 0 & \text{otherwise.} \end{cases}$$

Its distribution function is then given by

$$F_U(u) = \int_{-\infty}^{u} f_U(s)ds = \begin{cases} 0 & \text{if } u < 0 \\ u & \text{if } 0 \le u \le 1 \\ 1 & \text{if } u > 1\,. \end{cases}$$

We will assume that the computer has some function to simulate a uniform$[0,1]$ random variable. For example, in C, the standard library `rand` can be used: the line

```
U = rand()/(float) RAND_MAX
```

will simulate a uniform r.v.

Fix a random variable X with distribution function F_X we would like to simulate from. Consider the follow instructions:

1. Generate a uniform random variable U.

2. Output $F_X^{-1}(U)$

This will output a random variable with distribution function F_X. To see this, observe:

$$\begin{aligned}\mathbb{P}\left\{F_X^{-1}(U) \le x\right\} &= \mathbb{P}\left\{U \le F_X(x)\right\} \\ &= F_X(x)\,.\end{aligned}$$

As an example, consider the exponential distribution: the density of an exponential r.v. X is given by

$$f_X(x) = \begin{cases} e^{-x} & x \ge 0 \\ 0 & x < 0\,. \end{cases}$$

Thus, its distribution function is given by $F_X(x) = 1 - e^{-x}$.

Let us determine the inverse to $F_X(x)$. We need to solve for x in

$$1 - e^{-x} = u\,.$$

The solution is $x = -\log(1-u)$, that is $F_X^{-1}(u) = -\log(1-u)$.

Hence, to simulate an exponential random variable, do the following:

1. Generate a uniform r.v. U.
2. Output $-\log(1-U)$.

The next method is the *rejection* method. We want to simulate X with p.d.f. $f(x)$. Suppose we know how to simulation Y, with p.d.f. $g(y)$, and assume that there is some constant c so that

$$\frac{f(y)}{g(y)} \le c\,.$$

The rejection alogorithm is as follows:

1. Generate a uniform r.v. U and the r.v. Y.
2. If

   $$U \le \frac{f(Y)}{cg(Y)} \tag{1}$$

   then set $W = Y$ and output W. Otherwise go to step 1.

We now show that this method works. We need to show that W has the correct distribution. We only output Y when the condition (1) is met. Thus

$$\begin{aligned}\mathbb{P}\{W \le w\} &= \mathbb{P}\left\{Y \le w \,\middle|\, U \le \frac{f(Y)}{cg(Y)}\right\} \\ &= \frac{\mathbb{P}\left\{Y \le w \text{ and } U \le \frac{f(Y)}{cg(Y)}\right\}}{K},\end{aligned}$$

where $K \stackrel{\text{def}}{=} \mathbb{P}\left\{U \le \frac{f(Y)}{cg(Y)}\right\}$.

Now U and Y are independent, so the joint density function for (U, Y) is the product of the density of U and the density of Y:

$$f_{U,Y}(u, y) = \begin{cases} g(y) & \text{if } 0 \le u \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Thus

$$\begin{aligned}\mathbb{P}\left\{Y \le w \text{ and } U \le \frac{f(Y)}{cg(Y)}\right\} &= \int\int_{y \le w,\, u \le f(y)/cg(y)} f_{U,Y}(u, y)dudy \\ &= \int_{-\infty}^{w} g(y) \int_{0}^{f(y)/cg(y)} dudy \\ &= \int_{-\infty}^{w} g(y)\frac{f(y)}{cg(y)}dy \\ &= \frac{1}{c}\int_{-\infty}^{w} f(y)dy \\ &= \frac{1}{c}\mathbb{P}\{X \le w\}.\end{aligned}$$

Thus

$$\mathbb{P}\{W \le w\} = \frac{1}{cK}\mathbb{P}\{X \le w\}.$$

Letting $w \to \infty$ shows that $cK = 1$, and thus W has the same distribution function as X. Thus W has the correct distribution.

Let us see how to use this to generate a normal random variable. In this case, we will let Y be an exponential random variable. Instead of simulating a normal r.v. at first we will simulate the absolute value of a normal r.v. Such a r.v. has density

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

Notice that Y has p.d.f. $g(y) = e^{-y}$, and

$$\begin{aligned}\frac{f(y)}{g(y)} &= \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(y^2-2y\right)} \\ &= \sqrt{\frac{2e}{\pi}} e^{-\frac{1}{2}(y-1)^2} \\ &\leq \sqrt{\frac{2e}{\pi}}\,.\end{aligned}$$

Thus, we can set $c = \sqrt{2e/\pi}$ and then

$$\frac{f(y)}{cg(y)} = \exp(-\frac{1}{2}(y-1)^2)\,.$$

Thus we have the following algorithm for generating the absolute value of a normal random variable:

1. Generate two uniform random variables, U, V.
2. Set $Y = -\log(V)$.
3. If $U < \exp(-\frac{1}{2}(Y-1)^2)$ output Y. Otherwise go to step 1.

In particular cases, there can be clever ways to simulate random variables.

Example 1.1 (Two independent normals). Let (X, Y) be two independent standard normal variables. Thus, (X, Y) has a joint density

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\,.$$

We will make use of the *polar representation* of (X, Y). In particular, if $R^2 = X^2 + Y^2$, and $\Theta = \arctan(Y/X)$, then

$$(X, Y) = (R\cos\Theta, R\sin\Theta)\,.$$

What is the joint distribution of (R^2, Θ)? We use the change of variable formula. The density is given by

$$f_{R^2,\Theta}(\rho, \theta) = f_{X,Y}(\sqrt{\rho}\cos\theta, \sqrt{\rho}\sin\theta) J_T(\rho, \theta)\,,$$

where J_T is the Jacobian of the transformation $T(\rho, \theta) = (\sqrt{\rho}\cos\theta, \sqrt{\rho}\sin\theta)$. Then

$$\begin{aligned}
J_T(\rho, \theta) &= \begin{vmatrix} \frac{d}{d\rho}\sqrt{\rho}\cos\theta & \frac{d}{d\theta}\sqrt{\rho}\cos\theta \\ \frac{d}{d\rho}\sqrt{\rho}\sin\theta & \frac{d}{d\theta}\sqrt{\rho}\sin\theta \end{vmatrix} \\
&= \begin{vmatrix} \frac{1}{2\sqrt{\rho}}\cos\theta & -\sqrt{\rho}\sin\theta \\ \frac{1}{2\sqrt{\rho}}\sin\theta & \sqrt{\rho}\cos\theta \end{vmatrix} \\
&= \frac{1}{2}\cos^2\theta + \frac{1}{2}\sin^2\theta \\
&= \frac{1}{2}\,.
\end{aligned}$$

Thus

$$\begin{aligned}
f_{R^2,\Theta}(\rho, \theta) &= \frac{1}{2\pi} e^{-\frac{1}{2}\{(\sqrt{\rho}\cos\theta)^2 + (\sqrt{\rho}\sin\theta)^2\}} \frac{1}{2} \\
&= \underbrace{\frac{1}{2\pi}}_{f_\Theta(\theta)} \underbrace{\frac{1}{2} e^{-\frac{1}{2}\rho}}_{f_{R^2}(\rho)}
\end{aligned}$$

We conclude that the joint density for (R^2, Θ) factors into the product of a density involving only θ and a density involving only ρ. Thus (R^2, Θ) are independent random variables. Furthermore, from the densities we know their distributions: Θ is uniform on $[0, 2\pi]$, and R^2 is exponential(1/2).

From this we obtain the following algorithm for simulating two independent normal random variables:

1. Generate U which is uniform on $[0, 2\pi]$.
2. Generate R^2, an exponential(1/2) r.v.
3. Let $(X, Y) = (R\cos\Theta, R\sin\Theta)$, and output (X, Y).

2 Simulating Discrete Random Variables

A discrete random variable X takes values in a countable set $\Omega = \{\omega_1, \omega_2, \ldots\}$. It has an associated probability mass function

$$p_X(\omega_k) = \mathbb{P}\{X = \omega_k\}\ .$$

Here is a general algorithm for simulating a discrete random variable: Let

$$F_k = \sum_{j=1}^{k} p_X(\omega_k)\,.$$

1. Generate U a uniform random variable.
2. Initialize $k = 0$.
3. Replace k by $k + 1$.
4. If $F_{k-1} < U < F_k$ output ω_k and stop. Otherwise go to step 3

This works because

$$\mathbb{P}\left\{F_{k-1} < U < F_k\right\} = F_k - F_{k-1} = p_X(\omega_k)\,.$$

Let us see one example using this general procedure.

Example 2.1 (Geometric). A geometric(p) random variable takes values in $\{1, 2, \ldots\}$ and has mass function

$$p(k) = p(1-p)^{k-1}\,.$$

It will be convenient to work with $Q_k = 1 - P_k$:

$$\begin{aligned} Q_k &= \sum_{j=k+1}^{\infty} p(1-p)^{j-1} \\ &= p(1-p)^k \sum_{\ell=0}^{\infty} (1-p)\ell \\ &= (1-p)^k\,. \end{aligned}$$

Then using the algorithm above, we output the smallest k so that $P_{k-1} < U < P_k$. Equivalently, this is the first k so that

$$1 - P_{k-1} > 1 - U > 1 - P_k\,.$$

Notice that $1 - U$ is also uniform, so it is equivalent to generate a uniform, and output the smallest k so that $U > (1-p)^k$.

Taking logarithms, we output the smallest k so that

$$\begin{aligned}\log U &> k \log(1-p) \\ k &> \frac{\log U}{\log(1-p)}\,.\end{aligned}$$

Another way to say this is that we output

$$\left[\frac{\log u}{\log(1-p)}\right] + 1\,,$$

where $[x]$ is the largest integer smaller than x.

There are many clever tricks to simulate specific random variables, that are faster than the general algorithm above. We discuss here one example.

Example 2.2 (Binomial). A r.v. X has the Binomial(n,p) distribution if its probability mass function is

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}\,, k = 0, \ldots, n\,.$$

X can be represented as

$$X = sum_{k=1}^n I_k\,,$$

where $I_1, I_2, \ldots, I_n$ are independent Bernoulli(p) random variables. That is

$$\begin{aligned}\mathbb{P}\{I_k = 1\} &= p \\ \mathbb{P}\{I_k = 0\} &= 1-p\,.\end{aligned}$$

Hence a naive method of simulating X is:

1. Simulate $U_1, \ldots, U_n$ uniform random variables.
2. For each $k = 1, \ldots, n$, set $I_k = 1$ if $U_k < p$ and $I_k = 0$ if $U_k \geq p$.
3. Output $\sum_{k=1}^n I_k$.

This method requires we generate n uniform random variables. We now show another method which requires only a single uniform random variable.

The basic observation is the following: If U is uniform on $[0,1]$, then

(i) Given $U < p$, the distribution of U is uniform on $[0,p]$.

(ii) Given $U > p$, the distribution of U is uniform on $[p, 1]$.

We now show (i) holds. Suppose that $0 < u < p$.

$$\begin{aligned}
\mathbb{P}\{U \leq u \mid U < p\} &= \frac{\mathbb{P}\{U \leq u \text{ and } U < p\}}{\mathbb{P}\{U < p\}} \\
&= \frac{\mathbb{P}\{U \leq u\}}{p} \\
&= \frac{u}{p}\,.
\end{aligned}$$

(ii) follows by a similar argument.

This leads us to consider the following algorithm:

1. Generate a uniform U.
2. Initialize $k = 0$.
3. Let $k = k + 1$.
4. If $U < p$ do:
   (a) Set $I_K = 1$.
   (b) Replace U by $\frac{1}{p}U$.
5. If $U \geq p$ do:
   (a) Set $I_k = 0$.
   (b) Replace U by $\frac{U-p}{1-p}$.
6. If $k = n$ output $\sum_{j=1}^{n} I_k$ and stop.
7. Go to step 3

3 Markov Chains

Definition 3.1. A matrix $\boldsymbol{P}$ is *stochastic* if

$$P(i,j) \geq 0 \text{ and } \sum_j P(i,j) = 1\,.$$

Definition 3.2. A collection of random variables $\{X_0, X_1, \ldots\}$ is a *Markov chain* if there exists a stochastic matrix $\boldsymbol{P}$ so that

$$\mathbb{P}\{X_{n+1} = k \mid X_0 = j_0, X_1 = j_1, \ldots, X_n = j\} = P(j,k)\,.$$

If $\mu(k) = \mathbb{P}\{X_0 = k\}$, we can write down the probability of any event:

$$\mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} = \mu(i_0)P(i_0,i_1)\cdots P(i_{n-1},i_n)\,.$$

Notice that the future position (X_{n+1}) depends on the past ($X_0, X_1, \ldots, X_n$) only through the current position (X_n).

Example 3.3 (Random Walk). Let $\{D_k\}$ be an i.i.d. sequence of $\{-1,+1\}$-valued random variables, with

$$\begin{aligned}\mathbb{P}\{D_k = +1\} &= p\\ \mathbb{P}\{D_k = -1\} &= 1-p\,.\end{aligned}$$

Then let $S_n = \sum_{k=1}^n D_k$.

S_n takes values in $\mathbb{Z} = \{\ldots,-2,-1,0,1,2,\ldots\}$ and at each unit of time, either increases or decreases by 1. It increases with probability p.

The reader should carefully verify that $\{S_n\}$ is a Markov chain with transition matrix

$$P(j,k) = \begin{cases} p & \text{if } k = j+1\\ 1-p & \text{if } k = j-1\,.\end{cases}$$

Example 3.4 (Ehrenfest Model of Diffusion). Suppose N molecules are contained in two containers. At each unit of time, a molecule is selected at random and moved from its container to the other. Let X_n be the number of molecules in the first container at time n.

$$\begin{aligned}P(k,k+1) &= 1-\frac{k}{N}\\ P(k,k-1) &= \frac{k}{N}\,.\end{aligned}$$

In other words,

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0\\ \frac{1}{N} & 0 & 1-\frac{1}{N} & 0 & \cdots & 0 & 0\\ 0 & \frac{2}{N} & 0 & 1-\frac{2}{N} & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{N}\\ 0 & 0 & 0 & 0 & \cdots & 1 & 0\end{bmatrix}.$$

Example 3.5 (Bernoulli-Laplace Model of Diffusion). There are two urns, each always with N particles. There are a total of N white particles, and a total of N black particles. At each unit of time, a particle is chosen from each urn and interchanged.

Let X_n be the number of white particles in the first urn. Then

$$\begin{aligned} P(k,k+1) &= \left(1-\frac{k}{N}\right)^2 \\ P(k,k-1) &= \left(\frac{k}{N}\right)^2 \\ P(k,k) &= 2\frac{k}{N}\left(1-\frac{k}{N}\right)\,. \end{aligned}$$

3.1 n-step transition probabilities

Our first goal is to show that

$$\mathbb{P}\{X_{n+m}=k \mid X_m=j\} = P^n(j,k)\,, \tag{2}$$

where $P^n(j,k)$ is the (j,k)th entry of the nth matrix power of P. Define $\boldsymbol{M}_{m,n}$ to be the matrix with (j,k) entry equal to the left-hand side of (2).

Then

$$\begin{aligned} \mathbb{P}\{X_{m+n}=k \mid X_m=j\} &= \sum_{\ell} \mathbb{P}\{X_{m+n}=k|X_m=j, X_{m+n-1}=\ell\} \\ &\qquad \times \mathbb{P}\{X_{m+n-1}=\ell \mid X_m=j\} \\ &= \sum_{\ell} P(\ell,k) M_{m,n-1}(j,\ell)\,. \end{aligned}$$

Thus

$$\boldsymbol{M}_{m,n} = \boldsymbol{M}_{m,n-1}\boldsymbol{P}\,.$$

Notice that $\boldsymbol{M}_{m,0}$ is the indentity matrix I. Thus, we can continue to get

$$\boldsymbol{M}_{m,n} = \boldsymbol{M}_{m,0}\underbrace{\boldsymbol{P}\cdots\boldsymbol{P}}_{n \text{ times.}} = \boldsymbol{P}^n\,.$$

Let μ be the row vector $(\mu(1),\ldots)$. Suppose that X_0 has the distribion

$$\mathbb{P}\{X_0=k\} = \mu(k)\,.$$

Then the distribution at time n is given by

$$\begin{aligned}\mathbb{P}\{X_n = k\} &= \sum_{\ell} \mathbb{P}\{X_n = k | X_0 = \ell\} \mathbb{P}\{X_0 = \ell\} \\ &= \sum_{\ell} \mu(\ell) P^n(\ell, k) \, .\end{aligned}$$

In other words, the distribution at time n is given by $\mu \boldsymbol{P}^n$.