Chapter 2

A few exact sampling techniques

The basic building block of all of the methods that we will discuss are routines like `random()` in the $C/C++$ programming language which, with each call, returns an integer in the range $[0, M]$ where M is a very large positive integer. The best of these routines generate a deterministic periodic sequence of integers that, for most practical purposes, is indistinguishable from a random independent sequence of integers chosen uniformly (each one as likely as any other). Random number generators are interesting in their own right but we will not discuss them here.

Assuming that your random number generator (we'll assume it's `random()`) is actually producing an independent sequence of uniformly chosen integers, one can easily construct a very good approximation of a sequence of independent $\mathcal{U}(0, 1)$ random variables by the transformation

$$U = \frac{\texttt{random}()}{M}$$

with appropriate modifications if outcomes of $U = 0$ or $U = 1$ are problematic. In the next few sections we'll introduce techniques that can be used to transform samples from one distribution (e.g. $\mathcal{U}(0, 1)$) that can be easily generated into samples from a more complicated distribution.

2.1 Inversion

Suppose that our goal is to generate a sample from the distribution of a random variable X for which we have the function

$$F(x) = \mathbf{P}\left[X \leq x\right].$$

The function F is called the probability distribution function for X. This function is increasing and right continuous with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Exercise 14. *Use the requirements of a probability measure from Chapter 1 to show that F is increasing, right continuous, and satisfies $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.*

If F is differentiable then X has density $\pi = F'$.

Exercise 15. *Suppose that $X \in \mathbb{R}$ has density π and distribution function F. Use the definition of the Riemann integral to show that, for any continuous, bounded function f,*

$$\mathbf{E}\left[f(X)\right] = \int f(x)F'(x)dx,$$

i.e. $\pi = F'$.

In fact, this function completely characterizes the statistical behavior of a random variable. Let

$$F^{\dagger}(u) = \inf\left\{x : F(x) \geq u\right\}.$$

If F happens to be one-to-one then $F^{\dagger}$ is just the usual inverse of F. Now notice that

$$\left\{u : F^{\dagger}(u) \leq y\right\} = \left\{u : u \leq F(y)\right\}$$

since on the one hand if $F(y) \geq u$ then $F^{\dagger}(u) \leq y$ by the definition of $F^{\dagger}$ and on the other hand note that by right continuity of F and the definition of $F^{\dagger}$, $F(F^{\dagger}(u)) \geq u$, so that since F is increasing $F^{\dagger}(u) \leq y$ implies that $u \leq F(y)$.

Therefore, if $U \sim \mathcal{U}(0,1)$, the probability distribution function of $Y = F^\dagger(U)$ is

$$\begin{aligned}\mathbf{P}\left[Y \le y\right] &= \mathbf{P}\left[F^\dagger(U) \le y\right] \\ &= \mathbf{P}\left[U \le F(y)\right] \\ &= F(y),\end{aligned}$$

i.e. we have succeeded in generating a sample with the distribution of X.

Example 4. *Suppose we want to generate a single index Y from the set $\{0, 1, \ldots, n-1\}$ so that $\mathbf{P}\left[Y = i\right] = p_i$ where $p_i \ge 0$ and $\sum_{0=1}^{n-1} p_i = 1$. The probability distribution function for Y is*

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ s(j) & \text{if } x \in [j, j+1) \text{ for some } j \in \{0, 1, \ldots, n-2\} \\ 1 & \text{if } x \ge n-1 \end{cases}$$

where $s(j) = \sum_{i=0}^{j} p_i$. We also find that for $u \in [0,1)$,

$$F^\dagger(u) = \begin{cases} 0 & \text{if } u \in [0, s(0)) \\ j & \text{if } u \in [s(j-1), s(j)) \text{ for some } j \in \{1, 2, \ldots, n-1\}. \end{cases}$$

That $Y = F^\dagger(U)$ for $U \sim \mathcal{U}(0,1)$ has the correct distribution is intuitively clear since the length of each interval $[s(j-1), s(j))$ is p_j.

Example 5. *Suppose our goal is to sample an exponential random variable with parameter λ, i.e.*

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0. \end{cases}$$

We have just shown that if $U \sim \mathcal{U}(0,1)$ then

$$Y = F^\dagger(U) = -\frac{1}{\lambda}\log(1-U)$$

is distributed according to exp(λ).

Exercise 16. *Consider the distribution on $(0,1)$ with*

$$\pi(x) = \frac{1}{2\sqrt{x}}.$$

Write a code that uses inversion to generate samples from π using samples from $\mathcal{U}(0,1)$ and assemble a histogram of the output of your scheme. Produce a QQ plot to graphically compare your samples to π.

2.2 Change of variables

In the last section we showed that a particular change of variables could always be used (assuming you can evaluate F^{-1}) to generate samples with a desired distribution. Let's now think slightly more generally about what happens to a continuous distribution under a change of variables. Let's see what happens when we apply some smooth invertible function φ to the variables sampled from a density $\tilde{\pi}$. Its easiest to consider what happens to the moments of $\tilde{\pi}$:

$$\int f(\varphi(y))\tilde{\pi}(y)dy.$$

Make the change of variables $x = \varphi(y)$. In 1D we obtain

$$\int f(x)|\varphi'(\varphi^{-1}(x))|^{-1}\tilde{\pi}(\varphi^{-1}(x))dx.$$

The variable $X = \varphi(Y)$ then has density

$$\pi(x) = |\varphi'(\varphi^{-1}(x))|^{-1}\tilde{\pi}(\varphi^{-1}(x)).$$

In higher dimensions the formula becomes

$$\pi(x) = \frac{\tilde{\pi}(\varphi^{-1}(x))}{|\det\left(D\varphi(\varphi^{-1}(x)\right)|}$$

with

$$(D\varphi)_{ij} = \frac{\partial \varphi_i}{\partial y_j}.$$

Thinking of the above calculations in the opposite direction, suppose the goal is to generate samples from π but we are unable to do that directly. Suppose that for some smooth, invertible change of variables φ we are able to efficiently generate samples $Y^{(k)}$ from $\tilde{\pi}(y) = |\det\left(D\varphi(y)\right)|\,\pi(\varphi(y))$. Then by taking $X^{(k)} = \varphi^{-1}(Y^{(k)})$ we obtain samples from π.

Example 6. *Box-Muller Suppose that u_0 and u_1 are independent $\mathcal{U}(0,1)$ random variables. Define the function $\varphi : [0,1]^2 \to \mathbb{R}^2$ by*

$$\varphi_1(u_0,u_1) = \sqrt{-2\log u_0}\cos(2\pi u_1), \qquad \varphi_2(u_0,u_1) = \sqrt{-2\log u_0}\sin(2\pi u_1).$$

The Jacobian of this transformation is the determinant of the matrix

$$\begin{bmatrix} -\frac{1}{\sqrt{-2\log u_0 u_0}}\cos(2\pi u_1) & -2\pi\sqrt{-2\log u_0}\sin(2\pi u_1) \\ -\frac{1}{\sqrt{-2\log u_0 u_0}}\sin(2\pi u_1) & -2\pi\sqrt{-2\log u_0}\cos(2\pi u_1) \end{bmatrix}$$

which can easily be computed to obtain $|det(D\varphi)| = \frac{2\pi}{u_0}$. *In terms of the variables* $(y_0, y_1) = \varphi(u_0, u_1)$, u_0 *can be written*

$$u_0 = e^{-\frac{y_0^2+y_1^2}{2}}.$$

The density of $X = \varphi(U)$ *is therefore*

$$\pi(x) = \frac{1}{2\pi}e^{-\frac{x_0^2+x_1^2}{2}}$$

(the π *on the right hand side of the last display is the area of the unit disk and not the density* $\pi(x)$*).*

Exercise 17. *Write a routine that generates two independent Gaussian random variables. Use your code to generate many samples and produce a (2 dimensional) histogram of the results. Produce a QQ plot to graphically compare the distribution of the samples you generate to the standard normal distribution.*

Exercise 18. *Use a change of variables similar to the one you used for the Gaussian to generate a uniformly distributed sample on the unit disk given two independent samples from* $\mathcal{U}(0,1)$ *and produce a (2 dimensional) histogram to verify your code.*

2.3 Rejection

For the algorithm described in this subsection we again assume that our goal is to draw samples from the density π. Assume that we can draw samples from $\tilde{\pi}$ instead and that for some constant $K \geq 1$,

$$\pi \leq K\tilde{\pi}. \tag{2.1}$$

We generate a sample $X \sim \pi$ by generating pairs $(Y^{(k)}, U^{(k)})$ of independent random variables with $Y^{(k)} \sim \tilde{\pi}$ and $U^{(k)} \sim \mathcal{U}(0,1)$ until index τ when

$$U^{(\tau)} \leq \frac{\pi(Y^{(\tau)})}{K\tilde{\pi}(Y^{(\tau)})}$$

at which point we set $X = Y^{(\tau)}$.

The first question one must ask is whether this scheme returns a sample in finite time, i.e. is $\tau < \infty$? We will return to this question in a moment. For now, we assume that $\mathbf{P}[\tau < \infty] = 1$. Having made that assumption, breaking up the event $\{Y^{(\tau)} \in dx\}$ according to the value of τ and plugging in the definition of τ yields the expansion

$$\begin{aligned}
\mathbf{P}\left[Y^{(\tau)} \in dx\right] &= \mathbf{P}\left[Y^{(\tau)} \in dx,\, \tau < \infty\right] = \sum_{k=1}^{\infty} \mathbf{P}\left[Y^{(k)} \in dx,\, \tau = k\right] \\
&= \sum_{i=1}^{\infty} \mathbf{P}\left[Y^{(k)} \in dx,\, U^{(k)} \leq \frac{\pi(Y^{(k)})}{K\tilde{\pi}(Y^{(k)})},\, U^{(\ell)} > \frac{\pi(Y^{(\ell)})}{K\tilde{\pi}(Y^{(\ell)})}\, \forall\, \ell < k\right]
\end{aligned}$$

Using the fact that the different samples are independent we can factor the last expression to obtain

$$\begin{aligned}
\mathbf{P}\left[Y^{(\tau)} \in dx\right] = \sum_{k=1}^{\infty} \mathbf{P}\left[Y^{(k)} \in dx,\, U^{(k)} \leq \frac{\pi(Y^{(k)})}{K\tilde{\pi}(Y^{(k)})}\right] \\
\times \mathbf{P}\left[U^{(1)} > \frac{\pi(Y^{(1)})}{K\tilde{\pi}(Y^{(1)})}\right]^{k-1}
\end{aligned}$$

Finally, appealing to the relation

$$\begin{aligned}
\mathbf{P}\left[Y^{(k)} \in dx,\, U^{(k)} \leq \frac{\pi(Y^{(k)})}{K\tilde{\pi}(Y^{(k)})}\right] = \mathbf{P}\left[U^{(k)} \leq \frac{\pi(Y^{(k)})}{K\tilde{\pi}(Y^{(k)})} \,\middle|\, Y^{(k)} \in dx\right] \\
\times \mathbf{P}\left[Y^{(k)} \in dx\right]
\end{aligned}$$

and the uniform distribution of the $U^{(\ell)}$ variables we conclude that

$$\begin{aligned}
\lim_{|dx| \to 0} \frac{\mathbf{P}\left[Y^{(\tau)} \in dx\right]}{|dx|} &= \sum_{k=1}^{\infty} \left(1 - \int \frac{\pi(y)}{K\tilde{\pi}(y)} \tilde{\pi}(y) dy\right)^{k-1} \frac{\pi(x)}{K\tilde{\pi}(x)} \tilde{\pi}(x) \\
&= \frac{\pi(x)}{K} \sum_{k=0}^{\infty} \left(1 - \frac{1}{K}\right)^{k} \\
&= \pi(x)
\end{aligned}$$

Note that to use this algorithm you need to relate some distribution that you can easily sample (say a Gaussian) to the distribution that you'd actually like to sample, by a bound of the form (2.1). In all but some simple cases this is not possible. Nonetheless, rejection sampling can be a useful component of more complicated sampling schemes.

We return now to considering the cost of this algorithm. As mentioned above, one must first address the question of whether or not $\tau < \infty$. Fortunately, again decomposing the event $\tau < \infty$ by the possible values taken by τ and using the independence of the variables, we find that

$$\begin{aligned}\mathbf{P}\left(\tau < \infty\right) &= \sum_{k=1}^{\infty} \mathbf{P}\left(\tau = k\right) \\ &= \frac{1}{K} \sum_{k=1}^{\infty} \left(1 - \frac{1}{K}\right)^{k-1} \\ &= 1\end{aligned}$$

so the algorithm will at least exit properly with probability one (in fact we already assumed this with our previous calculation).

While returning a sample from π in finite time is of course a very basic requirement, it is hardly enough. We would also like to know how much computational effort will be expended to generate that sample. The cost of the scheme is characterized by the expectation of τ. We can compute the the mean time to exit as

$$\begin{aligned}\mathbf{E}\left[\tau\right] = \mathbf{E}\left[\sum_{k=1}^{\infty} \mathbf{1}_{\{\tau \leq k\}}\right] &= \sum_{k=1}^{\infty} \mathbf{E}\left[\mathbf{1}_{\{\tau \leq k\}}\right] = \sum_{k=1} \mathbf{P}\left(\tau \geq k\right) \\ &= \sum_{k=1}^{\infty} \left(1 - \frac{1}{K}\right)^{k-1} \\ &= K.\end{aligned}$$

Clearly, if the K chosen is not optimal, or if the best possible K so that (2.1) holds is very large, the rejection algorithm will be very costly.

Exercise 19. *Write a routine to generate a single sample from the uniform measure on the unit disk from two independent samples from $\mathcal{U}(0,1)$ using* *rejection. Make sure you clearly identify the target density* π*, the reference density* $\tilde{\pi}$*, and your choice of* K *(which should be justified). There's a natural choice of* K *for this problem that allows you to apply the sampling algorithm without having to know the area of the unit disk in advance... what is it? Verify your code by producing a histogram. Compare (numerically) the cost of this approach with the more direct approach in Exercise 18 by comparing, e.g. the expected number of* $\mathcal{U}(0, 1)$ *variables required per sample from the unit disk and the expected wall clock time per sample from the unit disk.*

2.4 bibliography