General Methods for Random Variate Generation

We now assume that we have a stream of independent uniforms and wish to generate from more general distributions.

Inversion

If $X \sim F_X$ (continuous) $\Rightarrow U = F_X(X) \sim \mathrm{U}(0,1)$.

Proof (sketch)

$$\begin{aligned} F_U(u) &= \Pr(U \leq u) = \Pr(F_X(X) \leq u) \\ &= \Pr(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u \quad \text{c.d.f. of a uniform.} \end{aligned}$$

This result is known as the probability integral transform, and allows us to transform any continuous random variable into a uniform random variable...**and vice versa**! We therefore have the following method of generating $X_i \sim F_X$:

1. generate $U_i \sim \mathrm{U}(0,1)$;
2. set $X_i := F_X^{-1}(U_i)$.

Inversion

Note that a key requirement above is the existence of $F_X^{-1}(\cdot)$.
What if F_X is not continuous?

For arbitrary random variables (i.e not necessarily continuous), we can use the generalized inverse distribution function,

$$F_X^{-}(u) = \min\{x : F_X(x) \geq u\},$$

and replace step 2 above with "Set $X_i = F_X^{-}(U_i)$".

Note also that, for an independent sample $X_1, \ldots, X_n$, we simply start with an independent sample $U_1, \ldots, U_n$.

Examples

Want $X \sim \text{Bernoulli}(p)$, *i.e.*,

$$X = \begin{cases} 1 & \text{w.p.} \quad p \\ 0 & \text{w.p.} \quad 1-p \end{cases}$$

i.e. $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$.

Algorithm (Bernoulli)

1. Generate $U = u \sim U(0,1)$.
2. If $u > 1 - p$, set $X = 1$,
   else set $X = 0$.
   Then $X \sim \text{Bernoulli}(p)$.

Examples

Want $X \sim \exp(\lambda)$. So,

$$F_X(x) = 1 - e^{-\lambda x}.$$

Setting $U = 1 - e^{-\lambda X} \Rightarrow X = -\frac{1}{\lambda}\log(1-U)$,

$$\Rightarrow F_X^{-1}(U) = -\lambda^{-1}\log(1-U).$$

Algorithm (Exponential)

1. Generate $U = u \sim U(0,1) \Rightarrow 1 - U \sim U(0,1)$.
2. Set $X = -\lambda^{-1}\log(u)$. Then $X \sim \exp(\lambda)$.

Examples

Want $X \sim$ Cauchy, i.e.

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,\mathrm{d}y = \frac{1}{2} + \pi^{-1}\tan^{-1}x$$

$$\Rightarrow F_X^{-1}(U) = \tan\left[\pi\left(U - \frac{1}{2}\right)\right].$$

Algorithm (Cauchy)

1. Generate $U \sim U(0,1)$.
2. Set $X = \tan\left[\pi\left(U - \frac{1}{2}\right)\right]$.

Examples

Want to generate $X \sim N(0,1)$. Note that

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\,\mathrm{d}y$$

But $F_X^{-1} = ?$. This can only be solved numerically.

<u>This is a severe limitation on the inversion method.</u>

Rejection

Suppose we wish to generate X with a known pdf $f_X(\cdot)$, but we cannot use inversion.

If it is possible to generate a different RV Y from a distribution with known pdf g_Y (using e.g. inversion), this may be enough, as long as:

- the support of g_Y encompasses that of f_X, i.e. $f_X(x) > 0 \Rightarrow g_Y(x) > 0$,
- and there exists $M > 0$ such that $\forall x$ s.t. $f_X(x) > 0$,

$$\frac{f_X(x)}{g_Y(x)} \leq M < \infty.$$

Algorithm Outline (rationale later):

1. Generate Y from g_Y.
2. For some function $h(\cdot)$ with values in [0,1], given $Y = y$, set $X = y$ with probability $h(y)$, otherwise return to 1.

What is the distribution of the accepted RV X?

We seek $\mathrm{P}[Y \leq x | U \leq h(Y)]$.

$$\begin{aligned}
\mathrm{P}[(Y \leq x) \cap (U \leq h(Y))] &= \int_{y=-\infty}^{x} \int_{u=0}^{h(y)} f_{Y,U}(y,u) \mathrm{d}u\mathrm{d}y \\
&= \int_{y=-\infty}^{x} \int_{u=0}^{h(y)} g_Y(y) \mathrm{d}u\mathrm{d}y \\
&= \int_{y=-\infty}^{x} g_Y(y) h(y) \mathrm{d}y
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{P}[U \leq h(Y)] &= \int_{y=-\infty}^{\infty} \int_{u=0}^{h(y)} F_{Y,U}(y,u) \mathrm{d}u\mathrm{d}y \\
&= \int_{y=-\infty}^{\infty} \int_{u=0}^{h(y)} g_Y(y) \mathrm{d}u\mathrm{d}y \\
&= \int_{y=-\infty}^{\infty} g_Y(y) h(y) \mathrm{d}y
\end{aligned}$$

What is the distribution of the accepted RV X?

What if we choose

$$h(y) = \frac{f_X(y)}{Mg_Y(y)}?$$

We have

$$\begin{aligned}
\mathrm{P}[Y \le x | U \le h(Y)] &= \frac{\int_{y=-\infty}^{x} g_Y(y)h(y)\mathrm{d}y}{\int_{y=-\infty}^{\infty} g_Y(y)h(y)\mathrm{d}y} = \frac{\int_{y=-\infty}^{x} (f_X(y)/M)\mathrm{d}y}{\int_{y=-\infty}^{\infty} (f_X(y)/M)\mathrm{d}y} \\
&= \int_{y=-\infty}^{x} f_X(y)\mathrm{d}y = F_X(x)
\end{aligned}$$

and so pdf of accepted X's is $f_X(x)$ - exactly what we want!

Rejection Sampling Algorithm

1. Generate $Y = y \sim g(.)$.
2. Generate $U = u \sim U(0, 1)$.
3. If $u \leq \frac{f(y)}{Mg(y)}$ set $X = y$.
4. Otherwise GOTO 1.

Here $M = \sup_x \frac{f(x)}{g(x)}$.

- Mg is an "envelope" for f.
- The condition $u \leq \frac{f(y)}{Mg(y)}$ implies that $Mg(y)u \leq f(y)$
- Now, $Mg(y)u$ is a point "at random" below $Mg(y)$ – we accept this point if it lies below $f(y)$.

How many "goes" to accept a value?

Let Z = number of rejections before accepting X.
Let $p = \mathrm{P}(\text{accept } X \text{ at each attempt})$,
$\Pr(Z = k) = p(1-p)^k$, $k = 0, 1, 2, \ldots$, *i.e.* $Z \sim \text{Geometric}(p)$.

$$\Rightarrow \mathrm{E}[\text{number of rejections per accepted } X \text{ variate}] = \mathrm{E}[Z] = \frac{1-p}{p}.$$

$$\begin{aligned} p = \mathrm{P}(\text{accept } X \text{ at each attempt}) &= \int_{-\infty}^{\infty} h(y) g(y)\, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} \frac{f(y)}{M g(y)} g(y)\, \mathrm{d}y \\ &= \frac{1}{M} \int_{-\infty}^{\infty} f(y)\, \mathrm{d}y = \frac{1}{M} \end{aligned}$$

So $\mathrm{E}[Z] = \frac{1-p}{p} = M - 1$ Alternatively: M is the expected number of trials per accepted X.

How to choose g such that the procedure is efficient?

So, ideally, we want to have M small. This can be achieved through prudent choice of g: the more the shape of g mimics the shape of f, the more efficient the procedure will be.

- ⊕ Can almost always find a suitable g, except perhaps if f is unbounded or if tails of f are heavy – in both cases, we can't always find $g(\cdot)$ such that $\frac{f(x)}{g(x)} < M \forall x$.
- ⊖ In order for the procedure to be efficient:
  - (i) we need to understand the shape of f_X in detail, to choose g_Y to mimic f_X;
  - (ii) it should be easy to simulate from g_Y.

  Typically, sampling distributions $g_Y(\cdot)$ for which exact sampling is straightforward require large M. TRADE OFF.

Extension; discard complicated normalisation constants

Suppose our target pdf $f_X(\cdot)$ can be written

$$f_X(x) = \frac{f_X^*(x)}{\int f_X^*(y)dy}.$$

If we choose $h(y) = \frac{f^*(y)}{Mg(y)}$, we can proceed almost exactly as before, noting only that

$$\mathrm{P}[\text{accept } X] = \int_{-\infty}^{\infty} h(y)g(y)\,\mathrm{d}y = \int \frac{f^*(y)g(y)}{Mg(y)}\,\mathrm{d}y = \frac{\int f^*(y)\,\mathrm{d}y}{M}.$$

Now test $u \le \frac{f^*(y)}{Mg(y)}$ where $y \sim g(\cdot)$ and $M = \sup_x \frac{f^*(x)}{g(x)}$.
Hence, we only need to know f "up to proportionality". Also only need to know the form of g up to proportionality: if $g(x) \propto g^*(x)$, work as before, with $M = \sup_x \frac{f^*(x)}{g^*(x)}$, and by testing $u \le \frac{f^*(y)}{Mg^*(y)}$.

Example: Generate from a normal using Cauchy as g

$$f^*(x) = e^{-x^2/2}, \quad g^*(x) = (1+x^2)^{-1}.$$

Want, $M = \sup_x \frac{f^*(x)}{g^*(x)}$.

Let $y = \log\left(\frac{f^*(x)}{g^*(x)}\right) = \log(f^*(x)) - \log(g^*(x)) = -\frac{x^2}{2} + \log(1+x^2)$.

Maximize y:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = x\left(\frac{2}{1+x^2} - 1\right) = 0 \Rightarrow x = 0 \quad \text{or} \quad x = \pm 1.$$

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = -1 + \frac{2-2x^2}{(1+x^2)^2} \begin{cases} x = 0 \Rightarrow \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} > 0 \Rightarrow \text{min} \\ x = \pm 1 \Rightarrow \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} < 0 \Rightarrow \text{max}\,. \end{cases}$$

So set

$$M = \sup_x \frac{f^*(x)}{g^*(x)} = \frac{f^*(1)}{g^*(1)} = 2e^{-1/2}.$$

NB: $x = -1$ gives the same M, as f^* and g^* are quadratic in x.

Example: Generate from a normal using Cauchy as g

Algorithm

1. Generate $U_1 = u_1 \sim U(0,1)$
2. Set $y = \tan\left[\pi\left(u_1 - \frac{1}{2}\right)\right]$, (so y is Cauchy – inversion).
3. Generate $U_2 = u_2 \sim U(0,1)$.
4. If $u_2 \le \frac{f^*(y)}{Mg^*(y)} = \frac{e^{-y^2/2}(1+y^2)}{2e^{-1/2}}$ set $X = y$, then $X \sim \mathrm{N}(0,1)$.
5. Otherwise GOTO 1.

Example: Generate from a normal using Cauchy as g

Acceptance Probability

$$
\begin{aligned}
\mathrm{P}(\text{accept } X) &= \int_{-\infty}^{\infty} \frac{f^*(y)}{Mg^*(y)} g(y)\,\mathrm{d}y, \\
&= \frac{1}{M} \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{(1+y^2)^{-1}} (\pi(1+y^2))^{-1}\,\mathrm{d}y, \\
&= \frac{e^{1/2}}{2} \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\pi}\,\mathrm{d}y, \\
&= \frac{e^{1/2}}{2\pi}\sqrt{2\pi} = \frac{e^{1/2}}{\sqrt{2\pi}} \approx 0.657.
\end{aligned}
$$

Example: Want $X \sim \text{Beta}(\alpha, \beta)$

$$
\begin{aligned}
f(x) &= \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{o/w} \end{cases} \\
&\propto \underbrace{x^{\alpha-1}(1-x)^{\beta-1}}_{f^*}
\end{aligned}
$$

Example of Implementation

(a) Work with f^*. Clearly easier than working with f

(b) "Lazy" choice of g? $U(0,1)$.

(c) Identify $M = \sup_x \frac{f^*}{g}$.

$$
\frac{f^*(x)}{g(x)} = x^{\alpha-1}(1-x)^{\beta-1}
$$

Setting $\frac{\mathrm{d}}{\mathrm{d}x} = 0$, gives maximum when $x = \frac{\alpha-1}{\alpha+\beta-2}$ - exercise

$$
\Rightarrow M = \frac{(\alpha-1)^{\alpha-1}(\beta-1)^{\beta-1}}{(\alpha+\beta-2)^{\alpha+\beta-2}}
$$

Example: Sampling from Bayesian posterior densities

$$\begin{aligned} f(\theta|D) &= \frac{p(\theta)f(D|\theta)}{\int p(\theta)f(D|\theta)} \\ &\propto p(\theta)f(D|\theta). \end{aligned}$$

e.g. Let θ = parameter $\quad p(\theta) =$ 'prior distribution'

$$f(D|\theta) = l(\theta) = \text{'likelihood'}.$$

Then the 'posterior' $f(\theta|D) \equiv f(\theta) \propto l(\theta)p(\theta)$, *i.e.* $f^*(\theta) = l(\theta)p(\theta)$.

Example: Sampling from Bayesian posterior densities

Use prior as our distribution to simulate from (i.e. $g \equiv p$).
We require

$$M = \sup_\theta \frac{f^*(\theta)}{p(\theta)} = \sup_\theta l(\theta) = l(\hat{\theta}),$$

where $\hat{\theta}$ = MLE. I.e. M is equal to the maximised likelihood.

Algorithm

1. Generate $U = u \sim U(0,1)$ and $\theta \sim p(x)$.
2. If
$$u \leq \frac{f^*(\theta)}{Mg(\theta)} = \frac{l(\theta)p(\theta)}{l(\hat{\theta})p(\theta)} = \frac{l(\theta)}{l(\hat{\theta})},$$
accept θ, then $\theta \sim f(\theta|D)$.
3. Otherwise GOTO 1.