We previously considered (some) methods of how to generate random variates other than the uniform given we can generate other random variates (such as the uniform).

One such technique was the acceptance-rejection method.

Today I will give a more in-depth treatment of this method.
Intro

• Finding an explicit formula for $F^{-1}(r)$ for the c.d.f. of a r.v. we wish to generate $F(x) = P(X \leq x)$ is not always possible.

• Even if we can, that may not be the most efficient method for generating a r.v. distributed according to $F$.

• Let us assume the continuous case and that $X$ has c.d.f. $F$ and p.d.f. $f$.

• I’ll give the discrete case later, which is very similar.
Basic Idea

- The basic idea is to find an alternative probability distribution $G$, with density $g(x)$, from which we can easily simulate (e.g., inverse-transform etc.).
- However, we’ll also want $g(x)$ ‘close’ to $f(x)$.
- In particular, we assume that $f(x)/g(x)$ is bounded by a constant $c \geq 1$.
- Hence $\sup_x \{f(x)/g(x)\} \leq c$.
- In practice we want $c$ as close to 1 as possible.
**Algorithm**

- Generate a r.v. $Y$ according to $G$ (remember we assumed this was easy).
- Generate $R \sim U(0, 1)$ independent of $Y$.
- If $R \leq f(Y)/cg(Y)$, then set $X = Y$, i.e., ‘accept’; otherwise start again i.e., ‘reject’.
Points of Note

- $f(Y)$ and $g(Y)$ are random variables, hence so is $f(Y)/cg(Y)$ and this ratio is independent of $R$.

- The ratio is bounded between 0 and 1; $0 < f(Y)/cg(Y) \leq 1$.

- The number of times $N$ that steps 1 and 2 need to be called (e.g., the number of iterations needed to successfully generate $X$) is itself a r.v. and has a geometric distribution with ‘success’ probability $p = P(R \leq f(Y)/cg(Y))$.

- Hence $P(N = n) = (1 - p)^{n-1}p$, $n \geq 1$, and on average the number of iterations required is given by $\mathbb{E}[N] = 1/p$.

- Ultimately we obtain $X$ as having conditional distribution of a $Y$ given that the event \{R $\leq f(Y)/cg(Y)$\} occurs.
Efficiency

- A direct calculation that the probability of ‘success’ $p = 1/c$ by first conditioning on $Y$.
- $P(R \leq f(Y)/cg(Y)|Y = y) = f(y)/cg(y)$.
- Thus unconditioning, and recalling that $Y$ has density $g(y)$ yields:
  \[
  p = \int_{-\infty}^{\infty} \frac{f(y)}{cg(y)} g(y) dy = \frac{1}{c} \int_{-\infty}^{\infty} f(y) dy = \frac{1}{c}
  \]
- Thus $\mathbb{E}[N] = c$, the bounding constant, and we now see why it is desirable to chose the alternative density $g$ so as to minimise this constant $c = \sup_x \{f(x)/g(x)\}$.
- Of course the optimal function would be $g(x) = f(x)$, but the whole point is that we are finding a $g$ which is ‘easy’ to simulate from because it is ‘difficult’ to simulate from $f$.
- In short, it is a bit of an art to find the appropriate $g$. 
Proof

- We need to show that the conditional distribution of $Y$ given that $R \leq f(Y)/cg(Y)$ is indeed $F$, i.e.:

$$P(Y \leq y | R \leq f(Y)/cg(Y)) = F(y)$$

- Let $B = \{R \leq f(Y)/cg(Y)\}$ and $A = \{Y \leq y\}$, and recall that $P(B) = p = 1/c$.

- Bayes’ Theorem then states $P(A|B) = P(B|A)P(A)/P(B)$.

- Hence we have:

$$P(Y \leq y | R \leq f(Y)/cg(Y)) = P(R \leq f(Y)/cg(Y) | Y \leq y) \frac{G(y)}{1/c} = \frac{F(y)}{cG(y)} \frac{G(y)}{1/c} = F(y)$$
• Why does $P(R \leq f(Y)/cg(Y)|Y \leq y) = \frac{F(y)}{cG(y)}$?

• $P(R \leq f(Y)/cg(Y)|Y \leq y) = \frac{P(R \leq f(Y)/cg(Y), Y \leq y)}{G(y)}$

• \[= \int_{-\infty}^{y} \frac{P(R \leq f(Y)/cg(Y)|Y = w)}{G(y)} g(w) \, dw\]

• \[= \frac{1}{G(y)} \int_{-\infty}^{y} \frac{f(w)}{cg(w)} g(w) \, dw\]

• \[= \frac{1}{cG(y)} \int_{-\infty}^{y} f(w) \, dw\]

• \[= \frac{F(y)}{cG(y)}\]
Sampling from a Normal Distribution

• Say we want to sample $X \sim \mathcal{N}(\mu, \sigma^2)$.
• We can express this as $X = \sigma Z + \mu$, where $Z \sim \mathcal{N}(0, 1)$.
• So sufficient to be able to sample $Z \sim \mathcal{N}(0, 1)$.
• Also, if we can generate $|Z|$, then by symmetry we can obtain $Z$ by independently generating a r.v. $S$ (for sign) that is ±1 with equal probability and setting $Z = S|Z|$.
• In other words, we generate Uniform $R$ and set $Z = |Z|$ if $R \leq 0.5$ and $Z = -|Z|$ if $R > 0.5$. 
Sampling from a Normal Distribution

- $|Z|$ is non-negative and has density $f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}$ for $x \geq 0$.

- For our alternative we can choose $g(x) = e^{-x}$ for $x \geq 0$ which is the Exponential distribution with rate $\lambda = 1$ (something we already know how to easily simulate from using the inverse-transform method).

- Note that $h(x) = f(x)/g(x) = e^{x-x^2/2}\sqrt{2/\pi}$.

- We can find the maximum of this by solving $h'(x) = 0$ (which must occur at that value of $x$ which maximizes the exponent $x - x^2/2$, namely at $x = 1$.)

- Thus $c = \sqrt{2e/\pi} \approx 1.32$.

- Hence $f(y)/cg(y) = e^{-(y-1)^2/2}$. 

Sampling from a Normal Distribution

• To generate $Z$ then:

• Generate $Y$ with an exponential distribution with parameter $\lambda = 1$ (that is generate $R$ and set $Y = -\log(R)$).

• Generate another $R$.

• If $R \leq e^{-(Y-1)^2/2}$, set $|Z| = Y$, otherwise go back to beginning.

• Generate $R$ and set $Z = |Z|$ if $R \leq 0.5$ and $Z = -|Z|$ if $R > 0.5$. 
Sampling from a Normal Distribution

• Note that in the ‘acceptance’ step $R \leq e^{-(Y-1)^2/2}$ if and only if $-\log(R) \geq (Y-1)^2/2$ and since $-\log(R)$ is Exponential with parameter $\lambda = 1$, we can simplify the algorithm to:

• Generate two independent exponentials at rate 1; $Y_1 = -\log(R_1)$ and $Y_2 = -\log(R_2)$.

• If $Y_2 \geq (Y_1 - 1)^2/2$ set $|Z| = Y_1$, otherwise go back to beginning.

• Generate $R$ and set $Z = |Z|$ if $R \leq 0.5$ otherwise $Z = -|Z|$. 
Sampling from a Normal Distribution

• As a nice afterthought, note that by the memoryless property of the Exponential distribution, \( P(T > s + t | T > s) = P(T > t) \), the amount by which \( Y_2 \) exceeds \( (Y_1 - 1)^2/2 \) in the ‘acceptance’ step, namely \( Y = Y_2 - (Y_1 - 1)^2/2 \), is itself exponentially distributed with rate 1 and is independent of \( Y_1 \).

• Thus, for free, we get back an independent exponential which could be used as one of the two needed in step 1 if we wanted to generate another independent \( X \sim \mathcal{N}(0, 1) \).

• So, after repeated use of the algorithm, the expected number of uniforms required to generate one \( Z \) is \( (2c + 1) - 1 = 2c \approx 2.64 \).

• Also, we might ask if we could improve the algorithm for \( Z \) by changing the rate of the exponential; that is, by using an exponential density \( g(x) = \lambda e^{-\lambda x} \) for some \( \lambda \neq 1 \)?

• The answer is no as elementary calculus shows \( \lambda = 1 \) minimizes \( c \).
Discrete Case

- The discrete case is analogous to the continuous case.
- Suppose we want to generate an $X$ that is a discrete r.v. with p.m.f. $p(k) = P(X = k)$.
- Further suppose that we can already easily generate a discrete r.v. with p.m.f. $q(k) = P(Y = k)$ so that $\sup_k \{p(k)/q(k)\} \leq c < \infty$.
- Then the following algorithm yields the required $X$:
  - Generate a r.v. $Y$ distributed as $q(k)$.
  - Generate $R$ (independent from $Y$).
  - If $R \leq p(Y)/cq(Y)$ then set $X = Y$ otherwise start over.