Review Lab 6 again!

One desired property of an LCG is a period of m. (i.e. we want our LCG to have the longest possible period of m).

Proposition: The LCG will have a full period m if and only if:

1. c and m are relatively prime, i.e. gcd (c, m) = 1
   greatest common divisor.

2. a-1 is divisible by all prime factors of m.

3. a-1 is a multiple of 4 if m is a multiple of 4.

proof: See Knuth, The Art of Computer Programming, Vol 2. §3.3.

Ex: $LCG(m, a, c, x_0, n) = LCG(256, 137, 123, 13, 256)$
has a full period of $m = 256$.
check that conditions 1, 2, and 3 above are met.
[see Lab 6]

Ex:
$LCG(m, a, C, x_0, n) = LCG(256, 135, 123, 13, 256)$

- does not have a full period of $m = 256$?
- Which condition of the above proposition is being violated?

[see Lab 6].

LCGs can be very bad (Recall RANDU from Lab 6). They are fast and good LCGs can be used for simple statistical simulation problems.

Ex. (bad LCG).

RANDU $\leftarrow$ LCG(2147483648, 65539, 0, 1, n)

Ex (decent LCG).

GlibcGCC $\leftarrow$ LCG(2^{32}, 1103515245, 12345, 13, n)

Problem with any LCG:

If an LCG is used to choose points in an n-dimensional space, the points will lie on at most $m^{1/n}$ hyper-planes.

We will use a more sophisticated pseudo-random number generator called The Mersenne Twister for our statistical simulation purposes. It is a variant of the recursive equations known as twisted generalized feedback shift-register. see Makoto Matsumoto and Takuji Nishimura, "Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator, ACM Transactions on Modeling and Computer Simulation, Vol 8, No. 1, Jan 1998, Pg. 3-30.

It has a period of $2^{19937}-1 \approx 10^{6000}$ and is currently used widely by researchers interested in statistical simulation. We will use it too.

See Lab 7 for an introduction.

Common Random Variables & their Simulation.

We can simulate or generate samples from or produce realisations of other random variables by making the following two assumptions:

1. IID samples from Uniform(0,1) RV can be generated (we have a good pseudorandom number generator).
2. real arithmetic can be performed exactly in a computer.

Inversion Sampler for Continuous RVs.

Proposition:
Let $F(x) := \int_{-\infty}^{x} f(y)dy : \mathbb{R} \to [0,1]$ be a continuous distribution function (D.F.) with density f and let its inverse

$$F^{[-1]}(u) := \inf\{x : F(x) = u\} : [0,1] \to \mathbb{R},$$

then $F^{[-1]}(U)$ has D.F. F provided $U \sim \text{Uniform}[0,1]$

Note: Infimum of a set A, denoted by $\inf(A)$ is the greatest lower bound of every element in A.

Proof:

$$P(F^{[-1]}(u) \leq x) \overset{\text{by defn.}}{=} P(\inf\{y : F(y) = u\} \leq x)$$

$$= P(u \leq F(x))$$

$$= F(x) \qquad \text{for all } x \in \mathbb{R}.$$

Now, let us summarize the inversion sampler as an Algorithm.

Algorithm for Inversion Sampler

input:
- PRNG for Uniform$(0,1)$ samples.
- $F^{[-1]}(u)$ as a procedure

output: $x \sim X$ with D.F. F.

$u \leftarrow$ Uniform$(0,1)$

$x \leftarrow F^{[-1]}(u)$

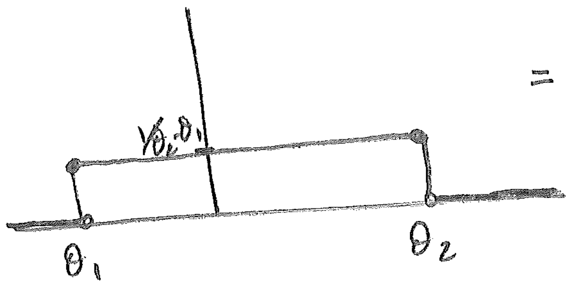
return x.

Model Uniform(θ_1, θ_2) RV.

Given two real parameters θ_1, θ_2 such that $\theta_1 \le \theta_2$, the PDF of the Uniform(θ_1,θ_2) RV is:

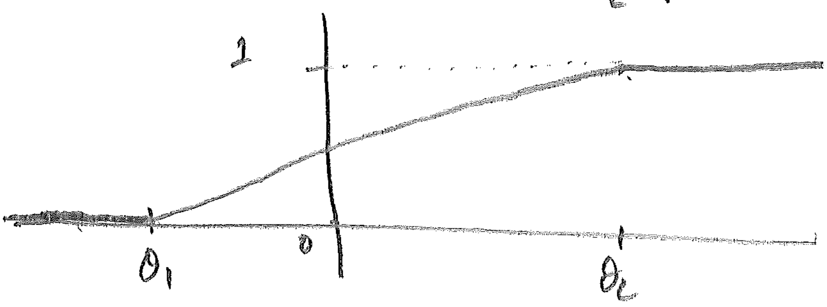
$$f(x;\theta_1,\theta_2) = \begin{cases} \dfrac{1}{\theta_2-\theta_1} & \text{if } \theta_1 \le x \le \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

$$= \mathbb{1}_{[\theta_1,\theta_2]}(x) \cdot \frac{1}{\theta_2-\theta_1}$$



and its DF given by $F(x;\theta_1,\theta_2) = \int_{-\infty}^{x} f(y;\theta_1,\theta_2)\,dy$ is:

$$F(x;\theta_1,\theta_2) = \begin{cases} 0 & \text{if } x < \theta_1 \\ \dfrac{x-\theta_1}{\theta_2-\theta_1} & \text{if } \theta_1 \le x \le \theta_2 \\ 1 & \text{if } x \ge \theta_2 \end{cases}$$



Simulation from $X \sim \text{Uniform}(\theta_1, \theta_2)$ can be achieved using the inversion sampler, provided we get an expression for $F^{[-1]}(u)$ that can be implemented as a procedure.

We need $F^{[-1]}(u)$ ...

We can get it by solving for x in terms of $u = F(x; \theta_1, \theta_2)$

$$\Leftrightarrow \quad u = \frac{x - \theta_1}{\theta_2 - \theta_1}, \quad \text{since } u \in [0,1]$$

$$\Leftrightarrow \quad (\theta_2 - \theta_1)u = x - \theta_1$$

$$\Leftrightarrow \quad x = (\theta_2 - \theta_1)u + \theta_1$$

$$\Leftrightarrow \quad F^{[-1]}(u; \theta_1, \theta_2) = \underbrace{(\theta_2 - \theta_1)u}_{\text{rescale}} + \underbrace{\theta_1}_{\text{translate}}$$

Algorithm:

input:
- $u \sim \text{Uniform}(0,1)$
- $F^{[-1]}(u)$.
- θ_1, θ_2

output: $x \sim \text{Uniform}(\theta_1, \theta_2)$

$u \longleftarrow \text{Uniform}(0,1)$

$x \longleftarrow F^{[-1]}(u) = ((\theta_2 - \theta_1) * u) + \theta_1$

return x.

Model Exponential (λ)

For a given real parameter $\lambda > 0$, an Exponential (λ) RV X has the PDF f and D.F. F:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \qquad F(x;\lambda) = 1 - e^{-\lambda x}$$

Mean and Variance of Exponential(λ) RV X.

$$E(X) = \int_{-\infty}^{\infty} x f(x;\lambda)\, dx = \int_{0}^{\infty} x \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}$$

Exercise: show this!

$$V(X) = \left(\frac{1}{\lambda}\right)^2$$

Exercise: Show that $F^{[-1]}(u;\lambda) = -\frac{1}{\lambda} \ln(1-u)$

$\ln = \log_e$ is the Natural logarithm.
In SAGE $\log$ is the natural logarithm.

Here is the Algorithm to simulate from Exponential (λ) RV.

Algorithm:

input:
- $u \sim$ Uniform$(0,1)$ from a PRNG.
- λ parameter

output:

$x \sim$ Exponential (λ)

$u \leftarrow$ Uniform$(0,1)$

$x \leftarrow -\frac{1}{\lambda} * \ln(1-u)$

Inversion Sampler for continuous RVs.

Recall we saw how to simulate from Uniform(θ_1, θ_2) and Exponential(λ) RVs.
Let us see another simulation next.

Model Cauchy RV.

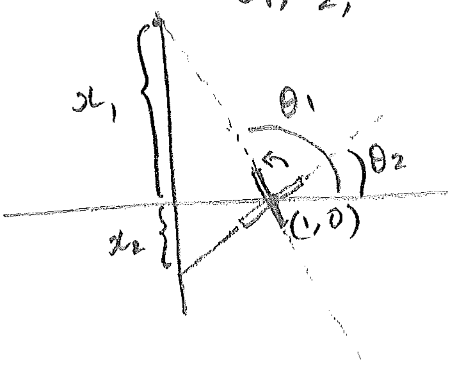
The density of the standard Cauchy RV X is:

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty$$

and its DF is:

$$F(x) = \frac{1}{\pi} \tan^{[-1]}(x) + \frac{1}{2}$$

$\theta_1, \theta_2, \ldots \overset{IID}{\sim}$ Uniform$(0, 2\pi)$



Try to construct by imagining a randomly spun double light saber centered at $(1,0)$ and noting its point of intersection with y-axis.

Recall that we need an expression for $F^{[-1]}(u)$ if we want to simulate from the standard Cauchy RV X.
This is done as follows: (Recall Uniform(θ_1, θ_2) ...).

1) Replace $F(x)$ by u in the expression for $F(x)$:

$$u = \frac{1}{\pi} \tan^{[-1]}(x) + \frac{1}{2} \quad (*)$$

2) Solve for x, i.e. we want $x = \cdots$ from $(*)$

$$\frac{1}{\pi} \tan^{[-1]}(x) + \frac{1}{2} = u \iff \frac{1}{\pi} \tan^{[-1]}(x) = u - \frac{1}{2}$$

$$\Leftrightarrow \quad \tan^{[-1]}(x) = \left(u - \tfrac{1}{2}\right)\pi$$

$$\Leftrightarrow \quad \tan\left(\tan^{[-1]}(x)\right) = \tan\left(\left(u - \tfrac{1}{2}\right)\pi\right)$$

$$\Leftrightarrow \quad x = \tan\left(\left(u - \tfrac{1}{2}\right)\pi\right)$$

Algorithm.

input: $u \sim \text{Uniform}(0,1)$.

output: $x \sim$ Cauchy.

$u \leftarrow \text{Uniform}(0,1)$

$x \leftarrow \tan\left(\left(u - \tfrac{1}{2}\right)\pi\right)$.

Note: Due to the periodicity of tan, we may use $x \leftarrow \tan(\pi u)$ instead.

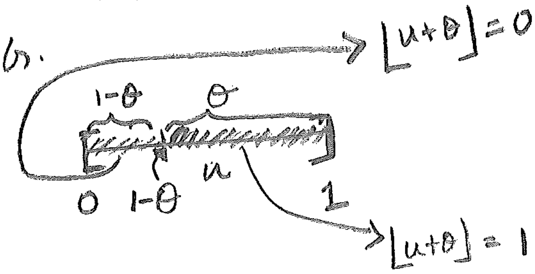
Gaussian or Normal RV (you have not seen this RV yet) is not amenable to such a simple inversion sampler.

—x—

Inversion Sampler for Discrete RVs.

We want to simulate from Bernoulli(θ) RV by transforming samples from Uniform$(0,1)$ RV.

Picture of Algorithm



Algorithm

input: $u \sim \text{Uniform}(0,1)$ PRNGs.
θ, the parameter.

$u \leftarrow \text{Uniform}(0,1)$

$x \leftarrow \lfloor u + \theta \rfloor$

return x

eg:

$\theta = \frac{1}{4}$, $u = \frac{1}{2}$: $\lfloor \frac{1}{4} + \frac{1}{2} \rfloor = \lfloor \frac{3}{4} \rfloor = 0$

but, $\theta = \frac{3}{4}$, $u = \frac{1}{2}$: $\lfloor \frac{3}{4} + \frac{1}{2} \rfloor = \lfloor 1\frac{1}{4} \rfloor = 1$