Quick note on HW 2 - CLT questions

To plot the normal density curve on top of the histogram of sample means, note that you need the std. deviation of the sample mean which is $\sigma/\sqrt{n}$

Recall: $E[\bar{X}_n] = \mu$ where $\mu = E[X_i]$

$$Var[\bar{X}_n] = \frac{\sigma^2}{n} \qquad \sigma^2 = Var[X_i]$$

$$\Rightarrow \text{std. dev. of } \bar{X}_n = \sqrt{Var(\bar{X}_n)} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Recall Uniform Distribution

$U \sim U(0,1)$

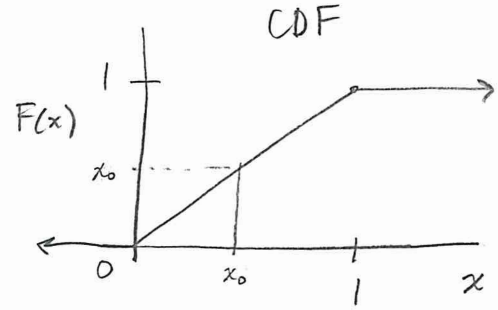
PDF of U : $f(x) = 1$ for $0 \le x \le 1$

CDF of U : $F(x) = x$ for $0 \le x \le 1$

$E[U] = \frac{1}{2}$

$Var(U) = \frac{1}{12}$

CDF



Recall Inverse Transform Method

from previous lecture : generate $U \sim U(0,1)$, then generate

$X = F^{-1}(U)$ ← RV with distribution F

Example 1: Generate an exponential random variable with rate 1 (i.e. parameter $\lambda = 1$).

Distribution function: $F(x) = 1 - e^{-\lambda x} = 1 - e^{-x}$ (for $x > 0$)

Let $x = F^{-1}(u)$. Then

$$u = F(x) = 1 - e^{-x}$$

$$\Rightarrow 1 - u = e^{-x}$$

$$\Rightarrow \log(1-u) = \log(e^{-x}) = -x$$

$$\Rightarrow x = -\log(1-u)$$

Note: log = natural logarithm in R

Thus, we can define RV X as

$$X = F^{-1}(U) = -\log(1-U)$$

However, $1-U$ is also uniformly distributed on $(0,1)$ so

$$-\log(1-U) \overset{D}{=} -\log(U)$$ ← same distribution

Thus, it suffices to define X s.t.

$$\boxed{X = -\log(U)}$$ where $U \sim Unif(0,1)$

More generally, $X \sim exp(\lambda)$ can be generated by setting

$$\boxed{X = -\frac{1}{\lambda}\log(U)}$$

$E[X] = \frac{1}{\lambda}$ ← where λ = rate parameter

Example 2: Generate a gamma(r, λ) random variable

Recall that if $X_1, X_2, \ldots, X_r$ are i.i.d. exponential RVs with rate λ, then

$$X = X_1 + X_2 + \cdots + X_r \sim \text{gamma}(r, \lambda)$$

In R, generate an r-dim uniform random vector

U = runif(r, 0, 1).

Then X = (-1/λ) · sum(log(U)).

$$\left(\text{i.e. } X = \frac{-1}{\lambda} \sum_{i=1}^{r} \log(U_i) \right)$$

Go to R practice: inverse-transform-sims-continuous.r on website

---

Rejection Method

- didn't get to this

Q. How to simulate a random variate from CDF F when there is no convenient formula for F?

For example, Normal distribution

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$$

Difficult to take F^{-1} !!

Rejection Method is based on $f = F'$ (density function) PDF

- Suppose X is a continuous RV with PDF f
- Suppose Y is a continuous RV with PDF g and

$$\frac{f(y)}{g(y)} \le c \quad \text{for all } y$$

- Suppose we know how to generate a random variate from PDF g. Then use the following steps to generate a random variate from f.

Step 1 : Generate $Y \sim g$ and and $U \sim U(0,1)$ [Y & U indep.]

Step 2 : If $U \le \frac{f(Y)}{c\, g(Y)}$, set $X = Y$. Otherwise go back to Step 1.

* Accept the generated value with a probability proportional to $\frac{f(Y)}{g(Y)}$.

Then RV X has density f.

Fun Fact: # of iterations needed to obtain X is a geometric RV with mean c.

Application – Monte Carlo Methods

* One of earliest applications of random numbers was to compute (approx.) integrals.

① Let $g(x)$ be a function & compute θ where

$$\theta = \int_0^1 g(x)\,dx$$

Note that if $U \sim U(0,1)$, then we can express θ as

$$\theta = E[g(U)]$$

If $U_1, \dots, U_n$ are i.i.d. $U(0,1)$ RVs, then $g(U_1), \dots, g(U_n)$ are i.i.d. RVs with mean θ.

SLLN $\Rightarrow$ (with probability 1)

$$\frac{1}{n}\sum_{i=1}^{n} g(U_i) \to E[g(U)] = \theta \quad \text{as } n \to \infty$$

Monte Carlo Approach: Approximate θ by generating large # of random numbers u_i & taking the average value of $g(u_i)$.

[Monte Carlo Methods Continued]

2. More generally,

approximate $P(X \in A)$ by simulating random variates $X_1, X_2, \ldots, X_n$ from the distribution of X & computing the ratio: $\boxed{\dfrac{\#\text{ of } X_i\text{'s} \in A}{n}}$

3. Approximate (multidim. setting)

$$\theta = \int_0^1 \int_0^1 \cdots \int_0^1 \underbrace{g(x_1, x_2, \ldots, x_k)}_{\text{any function}} \cdot \underset{\substack{\text{joint}\\ \text{PDF of } U_i\text{'s}}}{1} \cdot dx_1\, dx_2 \cdots dx_k$$

by generating n k-dimensional vectors (i.i.d. Uniforms)

$$(U_1^1, U_2^1, \ldots, U_k^1)$$
$$(U_1^2, U_2^2, \ldots, U_k^2)$$
$$\vdots$$
$$(U_1^n, \ldots, U_k^n)$$

where

and then since $g(U_1^i, U_2^i, \ldots, U_k^i)$ are i.i.d. with mean θ, we can estimate θ by computing

$$\frac{1}{n} \sum_{i=1}^{n} g(U_1^i, U_2^i, \ldots, U_k^i).$$