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 \( \frac{\sigma}{\sqrt{n}} \)

**Recall:**

\[
E[\bar{X}_n] = \mu \quad \text{where} \quad \mu = E[X_i]
\]

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

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

Recall Uniform Distribution

\( U \sim U(0,1) \)

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

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

\( E[U] = \frac{1}{2} \)

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

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

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

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

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

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

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

$-\log(1-U) \overset{D}{=} -\log(U)$ \hspace{1cm} \text{same distribution}

Thus, it suffices to define $X$ s.t.

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

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

$X = -\frac{1}{\lambda} \log(U)$ \hspace{1cm} E[X] = $\frac{1}{\lambda}$ where $\lambda$ = rate parameter
Example 2: Generate a gamma \((r, \lambda)\) random variable.

Recall that if \(X_1, X_2, \ldots, X_r\) are i.i.d. exponential RVs with rate \(\lambda\), then

\[ X = X_1 + X_2 + \ldots + X_r \sim \text{gamma}(r, \lambda) \]

In R, generate an \(r\)-dim uniform random vector

\[ U = \text{runif}(r, 0, 1) \]

Then

\[ X = \left(-\frac{1}{\lambda}\right) \cdot \sum \left(\log(U)\right) \]

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

Go to R practice: inverse-transform-sims-continuous, \(r\) on website.
Rejection Method

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

For example, Normal distribution

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

Difficult to take $F^{-1}$!!

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

- Suppose $X$ is a continuous RV with PDF $f$
- Suppose $Y$ is a continuous RV with PDF $g$
  and
  $$\frac{f(y)}{g(y)} \leq c \text{ for all } y$$

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

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

Step 2: If $U \leq \frac{f(Y)}{c \cdot g(Y)}$, 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 \( f \).

---

**Application - Monte Carlo Methods**

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

1. Let \( g(x) \) be a function and compute \( \Theta \) where

\[
\Theta = \int_{0}^{1} g(x) \, dx
\]

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

\[
\Theta = E[g(U)]
\]

If \( U_1, \ldots, U_n \) are i.i.d. \( U(0,1) \) RVs, then

\( g(U_1), \ldots, g(U_n) \) are i.i.d. RVs with mean \( \Theta \).

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

**SLLN** (with probability 1)

Monte Carlo Approach: Approximate \( \Theta \) by generating large # of random numbers \( U_i \) and taking the average value of \( g(U_i) \).
2. More generally, approximate $P(X \in A)$ by simulating random variates $X_1, X_2, \ldots, X_n$ from the distribution of $X$ and computing the ratio:

$$\frac{\text{# of } X_i \text{'s } \in A}{n}$$

3. Approximate (multidim. setting) by generating $n$ $k$-dimensional vectors (i.i.d. Uniforms)

$$(U_1^1, U_2^1, \ldots, U_k^1)$$
$$(U_1^2, U_2^2, \ldots, U_k^2)$$
$${\hspace{1cm}}$$
$$(U_1^n, \ldots, U_k^n)$$

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

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