Agenda: Random Variables

- Conditional Expectation
- Conditional distribution and Monte Carlo: Rejection sampling
The conditional expectation of $Y$ given $X = x$ is the mean of the conditional distribution of $Y$ given $X = x$.

For example, in the discrete case we have

$$E[Y|X = x] = \sum_y y P_{Y|X}(y|x)$$

and more generally the conditional expectation of $h(Y)$ given $X = x$ is

$$E[h(Y)|X = x] = \sum_y h(y) P_{Y|X}(y|x).$$
Example

- \( X \sim \text{Binomial}(n, p) \).

- Set \( m \leq n \) and let \( Y \) be the number of successes in the first \( m \) trials.

What are the conditional distribution and mean of \( Y \) given \( X = x \)?

- Conditional distribution:

\[
P(Y = y | X = x) = \binom{m}{y} \binom{n-m}{x-y} \binom{n}{x},
\]

- Conditional mean:

\[
Y = I_1 + \ldots + I_m
\]

and

\[
E(Y | X = x) = E(I_1 | X = x) + \ldots + E(I_m | X = x) = P(I_1 | X = x) + \ldots + P(I_m | X = x) = \frac{x}{n} + \ldots + \frac{x}{n} = \frac{m}{n}x
\]
Conditional expectation as a random variable

- Assuming that the conditional expectation of $Y$ given $X = x$ exists for every $x$, it is a well defined function of $X$ and, hence, a random variable. (The expectation of $Y$ given $X = x$ is a function depending on $x$: $E[Y|X = x] = g(x)$. $E[Y|X] = g(X)$ is r.v.)

- Example:

$$E(Y|X) = \frac{m}{n} X.$$

- This random variable has an expectation, and a variance (proviso absolute convergence, etc.) Its expectation is

$$E[E(Y|X)]$$
**Iterated Expectation**

**Theorem**

\[ E(Y) = E[E(Y|X)] \]

**Interpretation:** the expectation of \( Y \) can be calculated by first conditioning on \( X \), finding \( E(Y|X) \) and then averaging this quantity with over \( X \):

\[
E[Y] = \sum_x E[Y|X = x]P_X(x)
\]

where

\[
E[Y|X = x] = \sum_y yP_{Y|X}(y|x).
\]

**Example:** \( E(Y|X) = \frac{m}{n}X \) and \( E(X) = np \) give

\[
E(Y) = \frac{m}{n}E(X) = mp.
\]
Proof

We need to show

\[ E(Y) = \sum_x E(Y | X = x) P(X = x) \]

where

\[ E(Y | X = x) = \sum_y y P_{Y|X}(y | x) . \]

We have

\[
\sum_x E(Y | X = x) P_X(x) = \sum_y y \sum_x P_{Y|X}(y | x) P_X(x) \\
= \sum_y y P_Y(y) = E(Y).
\]
Consider sums of the type

\[ T = \sum_{i=1}^{N} X_i, \]

where

- \( N \) is a r.v. with finite expectation \( E(N) < \infty \),
- the \( X_i \)'s are independent of \( N \), with common mean \( E[X_i] = E[X], \forall i \).

Such sums arise in a variety of applications

- Insurance companies might receive \( N \) claims in a given period of time and the amounts of the individual claims may be modeled as r.v.'s \( X_1, X_2, \ldots \).
- \( N \) is the number of jobs in a single server queue and \( X_i \) the service time for the \( i \)th job. \( T \) is the time to serve all the jobs in the queue.
\[ E(T) = \sum_{\text{all } n} E(T|N = n)P(N = n) \]

and

\[ E(T|N = n) = E\left(\sum_{i=1}^{n} X_i\right) = nE(X). \]

i.e.

\[ E(T) = \sum_{n} nE(X)P(N = n) = E(N) \cdot E(X). \]

In other words, the average time to complete \( n \) jobs when \( n \) is random is the average value of \( N \) times the average time amount to complete a job.
Density function \( f \) we wish to sample from.

- \( f \) is nonzero on an interval \([a, b]\) and zero outside the interval (\( a \) and \( b \) may be infinite).
- Let \( M \) be a function such that \( M(x) \geq f(x) \) on \([a, b]\) and let

\[
m(x) = \frac{M(x)}{\int_a^b M(x) dx}
\]

The idea is to choose \( M \) so that it is easy to generate random variables from \( m \). If \([a, b]\) is finite, \( m \) can be chosen to be the uniform distribution.
Step 1: Generate \( T \) with the density \( m \).

Step 2: Generate \( U \), uniform on \([0, 1]\) and independent of \( T \).

If \( M(T) \times U \leq f(T) \) \( \rightarrow \) accept, \( X = T \).

Otherwise \( \rightarrow \) reject, goto step 1.
Why does this work?

To show

\[ P(X \in A) = \int_A f(t) \, dt \]

\[ P(X \in A) = P(T \in A \mid \text{Accept}) = \frac{P(T \in A \text{ and Accept})}{P(\text{Accept})} \]

Condition on \( T = t \) (\( I = \int_a^b M(t) \, dt \))

\[ P(T \in A \text{ and Accept}) = \int_a^b P(T \in A \text{ and Accept} \mid T = t) f_T(t) \, dt \]

\[ = \int_a^b P(U \leq f(t)/M(t) \text{ and } t \in A) m(t) \, dt \]

\[ = \int_A \frac{f(t)}{M(t)} m(t) \, dt = \frac{1}{I} \int_A f(t) \, dt \]

Similarly

\[ P(\text{Accept}) = \int_a^b P(\text{Accept} \mid T = t) m(t) \, dt = \int_a^b \frac{f(t)}{M(t)} m(t) \, dt = \frac{1}{I}. \]
**Remark**: High efficiency if algorithm accepts with high probability, i.e. $M$ close to $f$. 
Suppose we want to sample from a density whose graph is shown below.

Figure 1: Density function
In this case we let $M(T)$ be the maximum of $f$ over the interval $[0, 1]$, namely

$$M(x) = \max(f), \quad 0 \leq x \leq 1$$

so that $m$ is the uniform density over the interval $[0, 1]$. 
x <- 0:100
M <- max(f(x))
Routine for sampling once from the density $f$

OK <- 0
while(OK<1)
{
    # Generate T
    T <- runif(1, min = 0, max = 1)
    # Generate U
    U <- runif(1, min = 0, max = 1)
    if(M*U <= f(T))
    {
        OK <- 1
        RN <- T
    }
}
This routine will sample $n$ iid samples from the density $f$

```r
sample.from.density<-function(n)
{
    RN <- NULL
    for(i in 1:n)
    {
        OK <- 0
        while(OK<1)
        {
            T <- runif(1,min = 0, max = 1)
            U <- runif(1,min = 0, max = 1)
            if(U <= f(T))
            {
                OK <- 1
                RN <- c(RN,T)
            }
        }
    }
    return(RN)
}
```
Figure 2: Histogram of the Sampled Data