MAST30027: Modern Applied Statistics

Week 9 Lab

1. Suppose that $X \sim \text{bin}(n, \theta)$ and $\theta \sim \text{beta}(a, b)$. That is, $X|\theta$ has the pmf

$$p_{X|\theta}(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

and θ has pdf

$$f_\theta(x) = \beta(a,b)^{-1} x^{a-1} (1-x)^{b-1}.$$

The marginal distribution of X is given by

$$p_X(x) = \int_0^1 p_{X\theta}(x) f_\theta(\theta) d\theta.$$

X is said to have a beta-binomial distribution. It is possible, but not easy, to work out p_X for a beta-binomial. However, it is easy to estimate it using simulation.

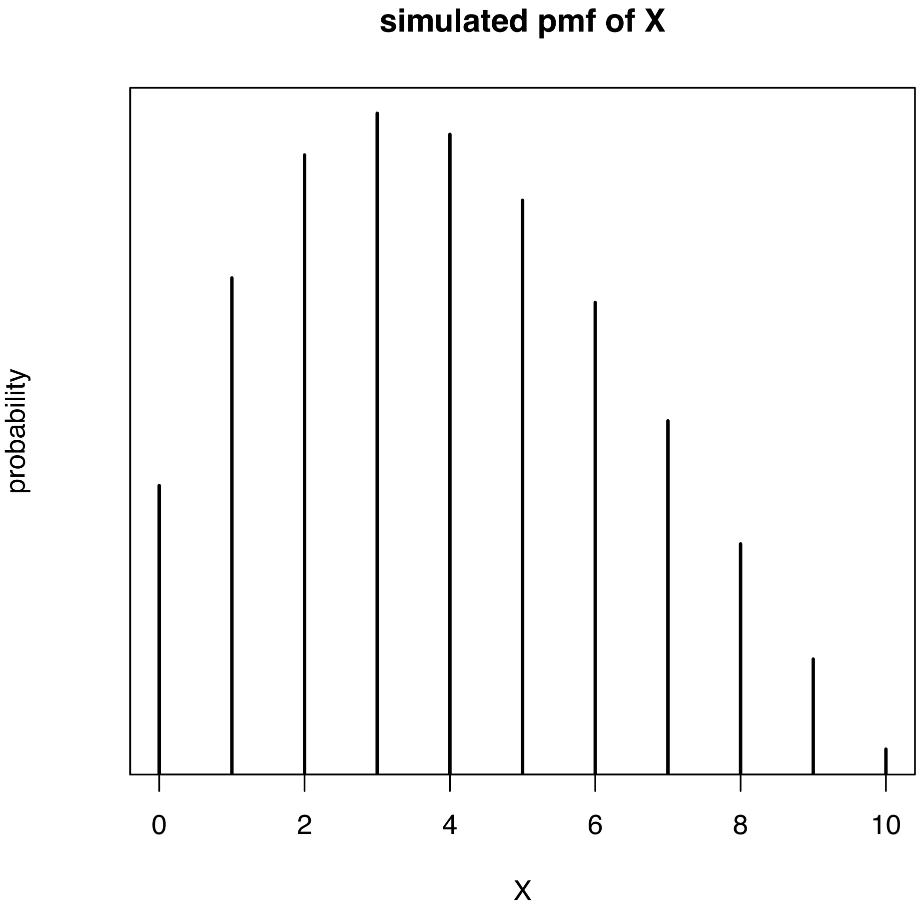
Generate a sample of size 1000,000 from a beta-binomial with $n = 10$, $a = 2$ and $b = 3$. Use it to estimate the pmf of X.

Solution:

```
> set.seed(300279)
> N <- 1e6
> tha <- rbeta(N, 2, 3)
> X <- rbinom(N, 10, tha)
> (p_X <- table(X)/N)

X
       0        1        2        3        4        5        6        7 
0.066230 0.109620 0.135315 0.144036 0.139627 0.125845 0.104476 0.079741 
       8        9       10 
0.054031 0.029951 0.011128 

> plot(sort(unique(X)), p_X, type='h',xlab="X",
+      ylab="probability", lwd=2,main="simulated pmf of X")
```



2. Suppose that Z follows a truncated exponential distribution, which has the pdf $p(z) = \frac{e^{-z}}{1-e^{-1}}$, $0 < z < 1$. Its theoretical mean and variance are known to be $E(Z) = 0.418$ and $\text{Var}(Z) = 0.079$.

   (a) Construct an rejection sampling algorithm to generate a sample of observations from the truncated exponential distribution.

   Solution: When $0 < z < 1$, $p(z) < (1 - e^{-1})^{-1} < \infty$, so we can use a rectangular envelope to contain the density. We get the following rejection algorithm:

   1° Generate $X \sim (0, 1)$.

   2° Generate $Y \sim (0, 1/(1 - e^{-1}))$ independently.

   3° If $Y \leq p(X)$ return X; otherwise go to 1°.

   (b) Write an R program to implement the algorithm in (a) and use it to generate a sample of 10000 observations. Plot a histogram of the sample. Calculate the sample mean and variance, and compare them with the theoretical mean and variance.

   Solution:

   We implement the algorithm in (a) and generate a sample of 10000 observations. We compute the sample mean and variance and compare them to the theoretical values. We also plot a histogram of the sample.

```
> trunc_exp <- function(){
+   # generate a truncated exponetial random variable using the rejection algorithm
+   x <- runif(1)
+   y <- runif(1, 0, 1/(1 - exp(-1)))
+   while (y > exp(-x)/(1 - exp(-1))){
+     x <- runif(1)
+     y <- runif(1, 0, 1/(1 - exp(-1)))
+   }
```

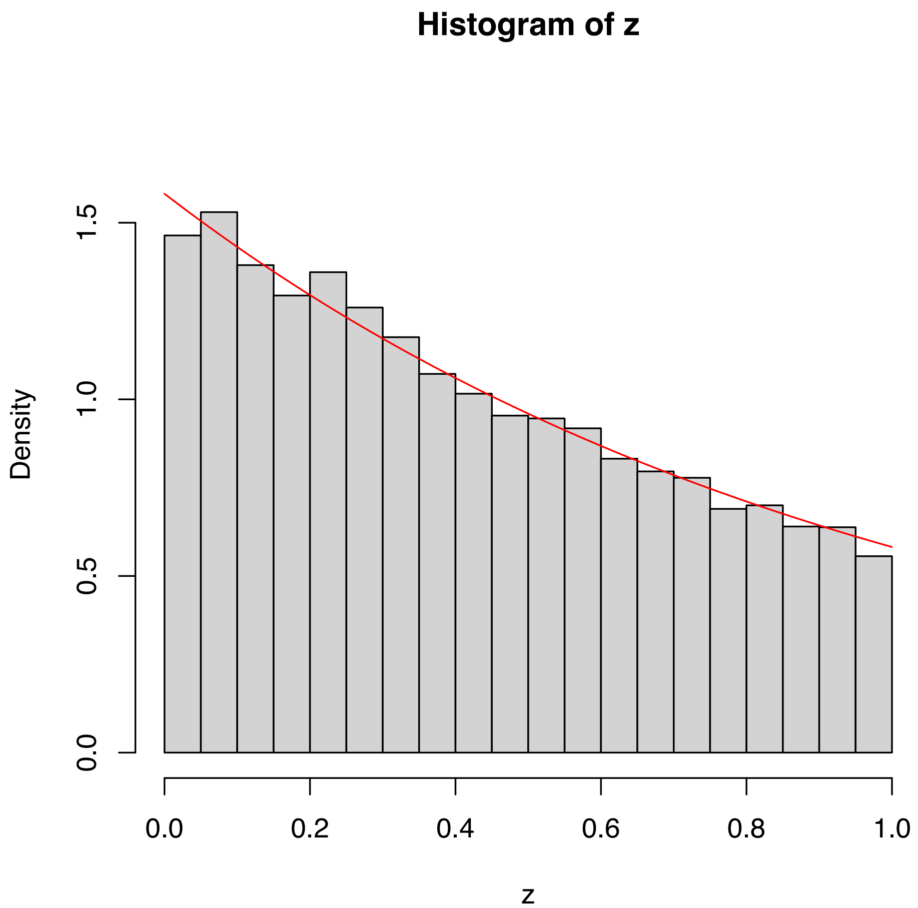
```
+   return(x)
+ }
> # generate a sample of size n
> set.seed(30027)
> n <- 10000
> z <- rep(0, n)
> for (i in 1:n) z[i] <- trunc_exp()
> # sample mean and variance
> mean(z)

[1] 0.4159879

> var(z)

[1] 0.07836399

> # histogram with density on top
> hist(z, freq=FALSE, ylim=c(0,1.8))
> curve(exp(-x)/(1 - exp(-1)), 0, 1, add=TRUE, col="red")
```



We can see that the sample mean and variance are similar to the theoretical values.

(c) Show that the following algorithm also simulates from the truncated exponential distribution.

1° Generate U from Unif(0,1);

2° If $U > e^{-1}$ then deliver $Z = -\ln(U)$; otherwise go to 1°.

Solution:

Now consider the alternative algorithm. According to the algorithm, the cdf of Z is

$$F_Z(z) \quad = \quad P(Z \le z)$$

$$
\begin{aligned}
&= P(-\ln(U) \leq z | U > e^{-1}) \\
&= P(U \geq e^{-z} | U > e^{-1}) \\
&= \frac{P(U \geq \max\{e^{-z}, e^{-1}\})}{P(U > e^{-1})} \\
&= \begin{cases} \frac{P(U \geq 1)}{1-e^{-1}} = 0 & \text{if } z \leq 0, \\ \frac{P(U \geq e^{-z})}{1-e^{-1}} = \frac{1-e^{-z}}{1-e^{-1}} & \text{if } 0 < z < 1, \\ \frac{P(U \geq e^{-1})}{1-e^{-1}} = 1 & \text{if } z \geq 1 \end{cases}
\end{aligned}
$$

Thus the pdf of Z is $F'_Z(z) = \frac{e^{-z}}{1-e^{-1}}$, $0 < z < 1$, the same as $p(z)$.