U20 Math 585 - Stochastic Processes

Homework 1

Solution

Instruction: Please type or write your answers clearly and show your work. You are encouraged to use the Rmarkdown version of this assignment as a template to submit your work. Unless stated otherwise, all programming references in the assignment will be in `R`. For this assignment, problems roughly covers content from the first 3 lectures.

1. Let X, Y be discrete random variables, show that

$$Var(X) = \mathbf{E}[Var(X|Y)] + Var(\mathbf{E}[X|Y])$$

Proof Note that

$$\begin{aligned}
\mathbf{E}[\mathbf{Var}(\mathbf{X}|\mathbf{Y})] &= \mathbf{E}[\mathbf{E}[X^{|}Y] - (\mathbf{E}[X|Y])^2] \\
&= \underbrace{\mathbf{E}[\mathbf{E}[\mathbf{X^2}|\mathbf{Y}]]}_{\mathbf{E}[X^2]} - \mathbf{E}[(\mathbf{E}[X|Y])^2] \\
\mathbf{Var}(\mathbf{E}[X|Y]) &= \mathbf{E}[(\mathbf{E}[X|Y])^2] - (\underbrace{\mathbf{E}[\mathbf{E}[X|Y]}_{(\mathbf{E}[X])^2})^2
\end{aligned}$$

Since $Var(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$, we have $\mathbf{E}[\mathbf{Var}(\mathbf{X}|\mathbf{Y})] + \mathbf{Var}(\mathbf{E}[X|Y]) = \mathbf{Var}(X)$

2. The time until a bus arrives has an exponential distribution with mean 30 minutes.

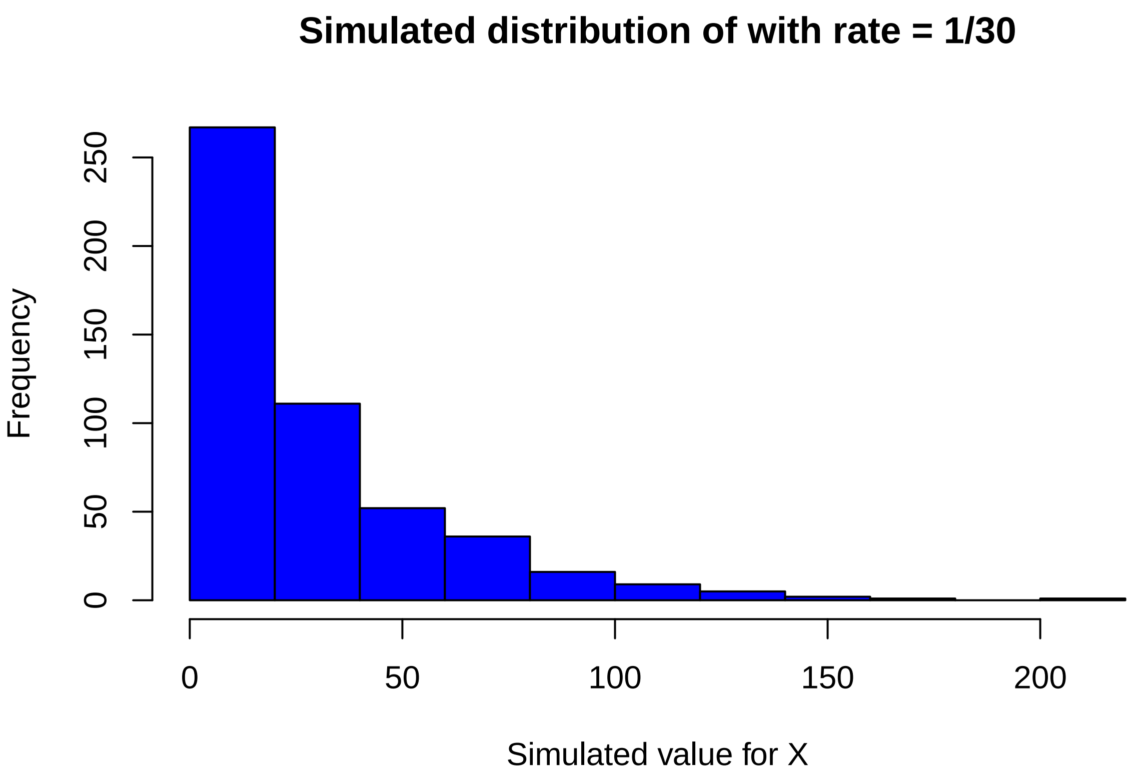
a) Use the command `rexp()` to simulate the probability that the bus arrives in the first 20 minutes.

b) Use the command `pexp()` to compare the exact probability.

c) Let X be the time of until a bus arrives. Note that if the mean arrival time is 30 minutes, that implies the rate $\lambda = 1/30$. Suppoose we generate $n = 500$ observations, we'll have a distribution that looks like the following:

```r
n <- 500
X <- rexp(n,rate = 1/30)

hist(X,col = 'blue',
     main = "Simulated distribution of with rate = 1/30",
     xlab = 'Simulated value for X',
     ylab = 'Frequency')
```



The probability $P(X \leq 20)$ is the proportion of simulated random variables that are less than 20, which in this case can be coded as `sum(ifelse(X<=20,1,0))/n`, and for this set of observations, the answer is about 0.534.

b) The exact probability can be computed using `pexp(20,1/30)`, giving 0.4865829.

We can also dive deeper in the comparison and investigate the difference between simulated and theoretical probability values using a series of increasing sample sizes, as we generate X using `rexp()` and compare $P(X < 20)$ obtained using the simulated random variables and `pexp()`.

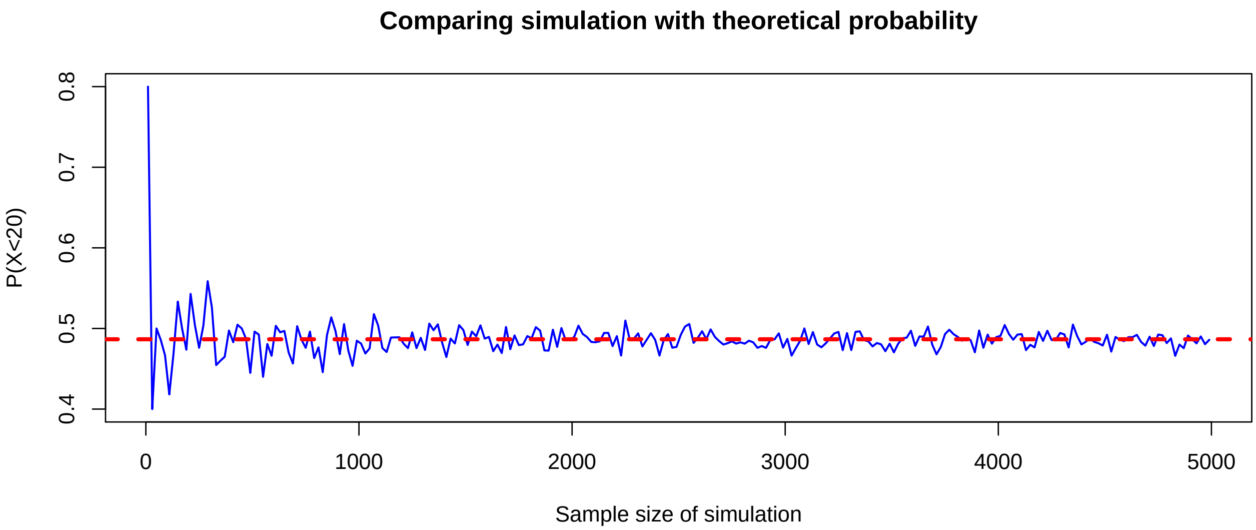
```
p_sim <-function(n,rate,x){
  simVals <-rexp(n =n, rate = rate)
  print(head(simVals))
  return(sum(ifelse(simVals<=x,1,0))/n)
}


ns<-seq(10,5000,20)

probList<-lapply(ns, function(n){p_sim(n=n,rate = 1/30, x = 20)})

plot(ns,unlist(probList),
     lwd = 1.5,
     col='blue',
     type = 'l',
     ylab='P(X<20)',
     xlab='Sample size of simulation',
     main ='Comparing simulation with theoretical probability')

abline(b=0,a=pexp(20,1/30),col = 'red',lwd = 3,lty = 2)
legend(70,0.35,legend = c('Simulated','Exact'),col = c('blue','red'),lty=1:2, cex=0.8)
```



3. Let $X_0, X_1, \ldots$ be a Markov chain with state space $\{1, 2, 3\}$ and transition matrix

$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array}$$

and initial distribution $\alpha = (1/2, 0, 1/2)$. Find the following:

a) $P(X_2 = 1|X_1 = 3)$

b) $P(X_1 = 3|X_2 = 1)$ (hint: use the properties of conditional distribution)

c) $P(X_9 = 1|X_2 = 3, X_4 = 1, X_7 = 2)$

4. Consider the Markov chain $X_0, X_1, X_2 \ldots$ with state space $\{A, B, C, D, E\}$ and transition matrix

$$\begin{bmatrix} 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 2/3 & 1/3 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 0 & 2/5 \\ 3/4 & 0 & 0 & 0 & 1/4 \\ 1/2 & 1/2 & 0 & 0 & 0 \end{bmatrix}$$

a) Construct a weighted directed graph of the Markov chain using the `markovchain` package.

5. You start with five dice. Roll all the dice and put aside those dice that come up 6. Then, roll the remaining dice, putting aside those dice that come up 6. And so on. Let X_n be the number of dice that are sixes after n rolls.

a) Describe the transition matrix $\mathbf{P}$ for this Markov chain. (hint: first, define the state space S)
b) Find the probability of getting all sixes by the third play.
c) What do you expect $\mathbf{P}^{100}$ to look like? Use tetchnology to confirm your answer.