Homework 3 Solutions

1. Chapter 8 Exercise 1 In the network intrusion example in Section 8.4, suppose X is not normally distributed, but instead has a uniform distribution on (450,550). Find $\Pr(X \geq 535)$ in this case. $\Pr(X \geq 535) = 1 - \Pr(X < 535)$

```
1 - punif(535, 450, 550)
```

```
## [1] 0.15
```

2. Suppose that a packaging service charges a specific rate for packages based on their weight. For packages weighing between 0 and 10 lbs, the service charges $5. For packages weighing between 10 and 20 lbs, the service charges $10. And for packages weighing between 20 and 30 lbs, the service charges $20.

   a. Suppose that 50% of the packages are between 0 and 10lbs, 30% between 10 and 20 lbs, and 20% between 20 and 30 lbs. What is the average charge for packages? The average charge is the weighted sum of the charges, $0.5 \cdot \$5 + 0.3 \cdot \$10 + 0.2 \cdot \$20$.

```
0.5*5 + 0.3*10 + 0.2*20
```

```
## [1] 9.5
```

   b. Suppose that within each category, the distribution of weights is uniformly distributed. The cost per lb is $0.20. What is the average cost of a package? We can compute the average weight in lbs and then multiply by the cost per lb to obtain the average cost (remember that $\mathrm{E}(c \cdot X) = c \cdot \mathrm{E}(X)$). The average weight for the small packages is 5 lbs, for medium packages it's 15 lbs, and for large packages it's 25 lbs. The weighted average is therefore

```
0.5*5 + 0.3*15 + 0.2*25
```

```
## [1] 12
```

and multiply by 0.2 gives 2.4. 0.32

   c. What is the probability that a package weighs more than 22 lbs conditional on knowing that it weighs more than 16lbs? Let X be the weight of the package. $\Pr(X > 22|X > 16) = \Pr(X > 22)/\Pr(X > 16)$. The numerator is equal to $0.2 \cdot (30 - 22)/(30 - 20) = 0.16$. $\Pr(X > 16) = \Pr(16 < X < 20) + \Pr(20 \leq X < 30) = 0.3 \cdot (20 - 16)/(20 - 10) + 0.2 \cdot (30 - 20)/(30 - 20) = 0.32$. Therefore $\Pr(X > 22|X > 16) = 0.16/0.32 = 1/2$.

3. Suppose that you are running a web server and the average number of hits per hour is 20.

   a. For a random hit, what is the probability the next hit comes in less than a minute? The time until the next hit is an exponential random variable with rate parameter 20 hits / 60 minutes = 1/3. $\Pr(X < 1) =$

```
pexp(1, rate = 1/3)
```

```
## [1] 0.2834687
```

   b. For 10 consecutive hits, what is probability that the minimum time between hits is less than a minute? Because of the ambiguity of question, I want to allow them to use either 10 or 9 independent random variables. I'm gonna go with 9. We want to find the probability that the minimum of 9 independent exponential(1/3) variables is less than 1. The minimum is an exponential random variable with rate $9 \cdot 1/3 = 3$ and the probability it is less than 1 is $1 - e^{-3} =$

```r
1 - exp(-3)
```

```
## [1] 0.9502129
```

c. What is the average time until the 10th hit? The average time till the 10th hit is the expected value of the sum of 10 exponential random variables, each with with mean $(1/3)^{-1} = 3$ and therefore the average time is 30. Another way to explain this is that the sum is a Gamma$(k = 10, \theta = 3)$ and the mean is $10 \cdot 3 = 30$.

d. What is the probability that you will get more than 20 hits in a half-hour? The number of hits in a half-hour is Poisson random variable with mean $\lambda = 10$ expected hits on average per half-hour. Therefore the probability that there are more than 20 hits in a half hour is equal to

```
r  1 - ppois(20, 10)
## [1] 0.001588261
```

4. Now suppose that the time between hits (measured in hours) is gamma distribution with scale parameter $\theta = 0.1$ and shape parameter 0.05.

a. For a random hit, what is the probability that the next hit comes in less than a minute?

```r
pgamma(1, shape = 0.05, scale = 0.1)
```

```
## [1] 0.9999998
```

b. For 10 consecutive hits, what is the probability that the minimum time between hits is less than a minute? Again, I will accept either 10 or 9 but I will do the problem with 9. Let $X_1, X_2, \ldots, X_9$ be iid Gamma$(k = 0.1, \theta = 0.05)$ random variables. $\Pr(\min(X_1, \ldots, X_9) < 1) = 1 - \Pr(\min(X_1, \ldots, X_9) \geq 1) = 1 - \Pr(X_1 > 1 \cap X_2 > 1 \cap \cdots \cap X_9 > 1) = 1 - \Pr(X > 1)^9$

```
r  1 - (1 - pgamma(1, shape = 0.05, scale = 0.1))^9
## [1] 1
```

c. What is the average time until the 10th hit? Let X_i be the time between the $(i-1)$th hit and the ith hit. Then the time till the 10th hit is $\sum_{i=1}^{10} X_i$ with X_i iid Gamma$(k = 0.1, \theta = 0.05)$. The sum is Gamma$(k = 10 \cdot 0.1, \theta = 0.05)$ with mean equal to $10 \cdot 0.1 \cdot 0.05 = 0.05$.

5. A civil engineer is collecting data on a certain road. She needs to have data on 25 trucks. Cars pass at an average of 90 per hour, and trucks pass at an average of 10 per hour.

a. What is the average amount of time that she will have to wait to have data on 25 trucks? Trucks pass at 10 per hour, so we should expect to wait 2.5 hours for 25 trucks.

b. In this time (from part a), what is the expected number of cars that will pass? The expected number of cars that pass in 2.5 hours is $90 \cdot 2.5 = 225$.

c. What is the probability she will observed more than 300 cars in this time? Let X be the number of cars that pass in 2.5 hours. X is a Poisson$(\lambda = 225)$ random variable. The probability X is greater than 300 is equal to

```
r   1 - ppois(300, 225)
## [1] 8.104138e-07
```

6. Suppose that a lightbulb has a mean lifetime of 50,000 hours.

a. If we've ran the lightbulb for 20,000 hours, what is the probability that it will last for 50,000 hours? We will model the lifetime of a lightbulb X as an exponential random variable with parameter $\lambda = 1/50000$ hours. $\Pr(X > 50000|X > 20000) = \Pr(X > 30000)$ by the memoryless property and $\Pr(X > 30000) = e^{-30000/50000}$

```
exp(-3/5)
```

```
## [1] 0.5488116
```

b. Suppose we have 5 lightbulbs in our house. What is the probability that all lightbulbs will last 50,000 hours?
Since the lightbulbs are independent, the probability all will last more than 50000 hours is the product of the probabilities that each of them last more than 50000 hours which equals

```
(exp(-1))^5
```

```
## [1] 0.006737947
```

c. Suppose we have 5 lightbulbs in our house. One goes out after 20,000 hours. What is the probability that at least one more lightbulb goes out in the next 1,000 hours? If one goes out then there are 4 light bulb remaining. We want to find the probability that at least one light bulb goes out in the next 1,000 hours, which is one minus the probability that all last longer than 1,000 hours due to the memoryless property of the exponential distribution. Using the logic above we can compute this as

```
1 - (exp(-1/50))^4
```

```
## [1] 0.07688365
```

7. a. Suppose that X is a random variable with pdf f $(x) = cxe^{-x}$ for $0 < x < 1$. Find the value of c. To find the answer note that in order for a pdf to be valid it must be non-negative (which is trivially satisfied here) and it must integrate to 1. So we set $\int_0^1 cxe^{-x}dx = 1$ and note that we can take the c outside the integral. Integration by parts leads to $\int_0^1 xe^{-x}dx = 1 - 2e^{-1}$ and therefore $c = 1/(1 - 2e^{-1})$.
   b. Suppose that X is a random variable with pdf $f(x) = c \cdot x \cdot exp(x)$ for $0 < x < 1$. Find the value of c. We do the same thing, noting that $\int_0^1 xe^x = 1$ and therefore $c = 1$.

8. Suppose that the percentage of impurities in a batch of chemical follows a beta distribution with shape parameters $\alpha = 2$ and $\beta = 10$.

   a. What is the probability that a random batch has more than 10% impurities? The probability that a random batch has more than 10% impurities is

```
1 - pbeta(0.1, 2, 10)
```

```
## [1] 0.6973569
```

b. Suppose that you can measure the impurities and classify the product into three grades, A with between 0 and 10% impurities, B with between 10 and 25% impurities, and C with more than 25% impurities. You can sell grade A for \$100 per batch, B for \$40 per batch, and C for \$10 per batch. What is the average price of a batch? What we need to do is find out what percentage of the batches correspond to each category and then add up the weighted sum. That is the average price = $\Pr(\text{grade A}) \cdot 100 + \Pr(\text{grade B}) \cdot 40 + \Pr(\text{grade C}) \cdot 10$.

r   100*pbeta(0.1, 2, 10) + 40*(pbeta(0.4, 2, 10) - pbeta(0.1, 2, 10)) + 10*(1 - pbeta(0.4, 2, 10))

```
## [1] 57.25159
```