Stat 330: Homework #6

Due on March 5, 2014 at 3:00pm

Mr. Lanker, Section A

Josh Davis

Problem 1

Problem 4.24 from Baron.

Installation of some software package requires downloading 82 files. On the average, it takes 15 sec to download one file, with a variance of 16 sec^2. What is the probability that the software is installed in less than 20 minutes?

Solution

Given the problem description, we can let X be the random variable where $X =$ number of minutes it takes to download a file. We know that $n = 82$, $\mu = 15$, $\sigma^2 = 16$, so $\sigma = 4$. Then we also know that in order to download 82 files in 20 minutes, each file must download in 0.244 minutes (or 14.6 seconds for easier understanding).

Let $S_n = X_1 + X_2 + \cdots + X_n$. Given this, we know that using the Central Limit Theorem, we can determine that given enough samples, or n, the distribution will converge to the Standard Normal. This gives us:

$$\begin{aligned}
\Pr\{\text{installed in 20 minutes}\} &= \Pr\{S_n \leq 20\} \\
&= \Pr\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{20 - (82)(15/60)}{(4/60)\sqrt{82}}\right\} \\
&= \Pr\{z \leq -0.828\} \\
&= \Phi(-0.828) \\
&= 0.2033
\end{aligned}$$

Problem 2

Problem 4.23 from Baron.

The lifetime of a certain electronic component is a random variable with the expectation of 5000 hours and a standard deviation of 100 hours. What is the probability that the average lifetime of 400 components is less than 5012 hours?

Solution

Let $X =$ number of hours a component lasts. We know that $n = 400$, $\mu = 5000$, and $\sigma = 100$. Like before, we will be using the Central Limit Theorem, however, instead we are looking at the sample mean, not the S_n. We know that for the sum, $S_n \sim \text{Normal}(\text{mean } = n\mu, \text{var } = \sigma^2/n^2)$. Thus we want the mean of the sum which gives us $\overline{X} \sim \text{Normal}(\text{mean } = \mu, \text{var } = \sigma^2/n)$

Now we want the probability that $\overline{X}$ lasts 5012 hours. This gives us:

$$\begin{aligned}
\Pr\{\text{lasts 5012 hours}\} &= \Pr\{\overline{X} \geq 5012\} \\
&= \Pr\left\{\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n^2}} \geq \frac{5012 - (5000)}{\sqrt{100^2/400^2}}\right\} \\
&= \Pr\{Z \geq 2.4\} \\
&= 1 - \Phi(2.4) \\
&= 1 - 0.0082 \\
&= 0.9918
\end{aligned}$$

Problem 3

Problem 4.28 from Baron.

Seventy independent messages are sent form an electronic transmission center. Messages are processed sequentially, one after another. Transmission of time of each message is Exponential with parameter $\lambda = 5\text{min}^{-1}$. Find the probability that all 70 messages are transmitted in less than 12 minutes. Use the Central Limit Theorem.

Solution

Let $X \sim \text{Exponential}(\lambda)$, where $\lambda = 5 \text{ min}^{-1}$. We know that $n = 70$, $\mu = 1/\lambda$, and $\sigma^2 = 1/\lambda^2$, or $\sigma = 1/\lambda$. Once again, using the Central Limit Theorem, we want to calculate the time to transmit all n messages. Thus $S_n = X_1 + \cdots + X_n$, and letting $n \to \infty$, we get:

$$
\begin{aligned}
\Pr\{\text{transmits in 12 mins}\} &= \Pr\{S_n \leq 12\} \\
&= \Pr\left\{ \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{12 - (70)(1/5)}{(1/5)\sqrt{70}} \right\} \\
&= \Pr\{z \leq -1.20\} \\
&= \Phi(-1.20) \\
&= 0.115
\end{aligned}
$$

Problem 4

In this problem, we will esimate the probability that a Geometric random variable is greater than or equal to another Geometric random variable.

Let $X \sim \text{Geometric}(p_x = 0.2)$ and $Y \sim \text{Geometric}(p_y = 0.1)$.

Part A

Get an estimate for $\Pr(X \geq Y)$ to three decimals using a simulation study. Use the R code posted (hw6prob4.R) with this part.

Solution

Our estimate based off of running the code ran is 0.3572.

Part B

Get an exact answer for $\Pr(X \geq Y)$ by using the law of total probability and the cumulative probability function for the geometric distribution. Hint:

$$
\Pr(X \geq Y) = \sum_{k=1}^{\infty} \Pr(X \geq Y \mid Y = k) \Pr(Y = k) = \sum_{k=1}^{\infty} \Pr(X \geq k) \Pr_Y(k)
$$

using the positive integers as a partition for Y.

Solution

Knowing that X and Y are Geometric, we know that the cdf of the distribution is $F(x) = Pr(X \leq x) = 1 - (1-p)^x$.

Problem 4 continued on next page...

Given the hint, this becomes:

$$\begin{aligned}
\Pr(X \geq Y) &= \sum_{k=1}^{\infty} \Pr(X = k) \Pr(Y > k) + \Pr(X = k) \Pr_Y(k) \\
&= \sum_{k=1}^{\infty} (1 - \Pr_Y(Y > k)) \Pr(X = k) + \sum_{k=1}^{\infty} \Pr(X = k) \Pr_Y(k) \\
&= \sum_{k=1}^{\infty} (1 - (1 - (1 - p_y)^k)) \Pr_Y(k) + \sum_{k=1}^{\infty} \Pr(X = k) \Pr_Y(k) \\
&= \sum_{k=1}^{\infty} (1 - p_y)^k \Pr_Y(k) + \sum_{k=1}^{\infty} \Pr(X = k) \Pr_Y(k) \\
&= \sum_{k=1}^{\infty} (1 - p_y)^k p_x (1 - p_x)^{k-1} + \sum_{k=1}^{\infty} \Pr(X = k) p_y (1 - p_y)^{k-1} \\
&= \sum_{k=1}^{\infty} (1 - p_y)^k p_x (1 - p_x)^{k-1} + \sum_{k=1}^{\infty} p_x (1 - p_x)^{k-1} p_y (1 - p_y)^{k-1} \\
&= \sum_{k=1}^{\infty} p_x (1 - p_y)(1 - p_y)^{k-1} (1 - p_x)^{k-1} + \sum_{k=1}^{\infty} p_x p_y (1 - p_x)^{k-1} (1 - p_y)^{k-1}
\end{aligned}$$

By using the sum of a geometric series equation:

$$\sum_{k=1}^{\infty} r^{k-1} = 1 + r + r^2 + r^3 + \ldots = \frac{1}{1 - r}$$

we get:

$$\begin{aligned}
\Pr(X \geq Y) &= p_x (1 - p_y) \left[1 + (1 - p_y)^1 (1 - p_x)^1 + \ldots\right] + p_x p_y \left[1 + (1 - p_y)^1 (1 - p_x)^1 + \ldots\right] \\
&= \frac{p_x (1 - p_y)}{1 - (1 - x)(1 - y)} \frac{p_x p_y}{1 - (1 - x)(1 - y)} \\
&= \frac{p_x (1 - p_y) + p_x p_y}{1 - (1 - x)(1 - y)} \\
&= \frac{p_x (1 - p_y) + p_x p_y}{p_x + p_y - p_x p_y} \\
&= \frac{p_x - p_x p_y + p_x p_y}{p_x + p_y - p_x p_y} \\
&= \frac{p_x}{p_x + p_y - p_x p_y}
\end{aligned}$$

Substituting our values into the equation gives us:

$$\begin{aligned}
\Pr(X \geq Y) &= \frac{(0.1)}{(0.1) + (0.2) - (0.1)(0.2)} \\
&= \frac{0.1}{0.28} \\
&= 0.3571
\end{aligned}$$

Badda Bing, Badda Boom!

Problem 5

Suppose that a rabbit is trying to get to its home in the field. There are three places the rabbit can be: the field, the park, or the road. The rabbit starts in the park. Estimate the probability that the rabbit finds home before finding a car's tire.

A rabbit in the field will find its home with probability 0.40 and will otherwise return to the park. A rabbit in the park will find its way to the field with probability 0.80, and will go on the road with probability 0.20. A rabbit on the road will successfully get back to the park with probability 0.50. Otherwise, it won't be a good ending for the bunny.

How many rabbits will find their way to their home?

Solution

Running the R code provided, we get:

```
base::source("./hw6prob5.R")

## [1] "Results:  7621 rabbits found home,  2379 rabbits died.  (Longest path was of length 35)"
## [1] "Proportion of rabbits finding home:  0.7621"
```

Thus the probability was 0.7621 which means that 7621 found their way home and the rest died.

Problem 6

Draw 100,000 random values from the sum of 12 uniform random variables values minus 6.

That is, let $U_i \sim \text{Uniform}(0, 1)$ random variables for $i = 1, 2, \ldots, 12$ and let $X = U_1 + U_2 + \cdots + U_{12} - 6$ such that $\text{E}[X] = 0$ and $\text{Var}[X] = 1$ and the shape of X is approximately normal. Report the portions for each number.

How do these differ from the true probabilities $\Pr(Z < z)$ for each of these cases?

Solution

Running the R code provided, we see the simulated results, the true results, and the difference in the two values:

```
base::source("./hw6prob6.R")

## [1] "Proportion under -3: 0.00104. P(Z < -3) = 0.00135. Difference: 0.00031"
## [1] "Proportion under -2: 0.02257. P(Z < -2) = 0.02275. Difference: 0.00018"
## [1] "Proportion under -1: 0.16030. P(Z < -1) = 0.15866. Difference: 0.00164"
## [1] "Proportion under 0: 0.49865. P(Z < 0) = 0.50000. Difference: 0.00135"
## [1] "Proportion under 1: 0.83823. P(Z < 1) = 0.84134. Difference: 0.00311"
## [1] "Proportion under 2: 0.97805. P(Z < 2) = 0.97725. Difference: 0.00080"
## [1] "Proportion under 3: 0.99897. P(Z < 3) = 0.99865. Difference: 0.00032"
```

As we can see, the true values are very close to the simulated values as given by the difference for each iteration.