STAT 370: Theoretical Statistics - Problem Set 1

Group 5: Bonnie Lin, Daniel Njoo

January 25, 2018

```
library(dplyr)
set.seed(370)
numsim <- 10000
```

Question 3

A machine produces bearings with mean diameter 3.00 inches and standard deviation 0.01 inches. Bearings with diameters in excess of 3.02 inches or less than 2.98 inches will be deemed to fail to meet quality specifications. Assume that the underlying process is normally distributed. Approximately what fraction of this machine's production will fail to meet specifications? Simulate from a normal distribution to check your analytic answer. Then assume that the underlying process is Gamma distributed with the same mean and sd. Simulate from an appropriate Gamma distribution to check your analytic answer. How important is the normality assumption?

Analytical Answer

Assuming Normal distribution:

Let b represent a bearing produced by the machine. Then:

$P(b > 3.02 \text{ or } b < 2.98)$

$=(1 - P(b \leq 3.02)) + P(b \geq 2.98)$ — since it's a continuous probability distribution

=(1-pnorm(3.02,3,0.01))+pnorm(2.98,3,0.01)

=0.0455003

–

Assuming Gamma distribution

Let b represent a bearing produced by the machine. Then:

$E[X] = \frac{\alpha}{\beta}$ and $V[X]\frac{\alpha}{\beta^2}$, so solving for α and β, we find $\alpha = 90000$ and $\beta = 30000$, so

$(1 - P(b \leq 3.02)) + P(b \geq 2.98)$

=(1-pgamma(3.02,90000,30000))+pnorm(2.98,90000,30000)

=0.0242801

Empirical Answer

Question 5

A personnel manager for a certain industry has records of the number of employees absent per day. The average number absent is 5.5, and the standard deviation is 2.5. Because there are many days with zero, one or two absent and only a few with more than 10 absent, the frequency distribution is highly skewed. The manager wants to publish an interval in which at least 75% of these values lie. Use the result from the previous problem to find such an interval. Generate data from a distribution that matches these parameters and compare your result to the true value.

Analytical Answer

Given Chebyshev's Inequality, we know that $P(|X - \mu| \geq k\sigma) \geq 1 - \frac{1}{k^2}$, and that $1 - \frac{1}{k^2} = 0.75$ so $k = 2$.

So the interval is $\mu \pm 2\sigma = (0, 10.6)$

Empirical Answer

```
lambda <- 1/5.5
cdf <- pexp(1:16,lambda)
pdf <- c()

for (i in 1:15) {
  pdf[i] <- cdf[i+1]-cdf[i]
}

absent_vec <- sample(1:15, 15, replace=T, prob=pdf); absent_vec
```

```
##  [1]  3  1  2 13  7  1  3  3  5  6  6 10  7 12  8
```

Question 9

Shuffle an ordinary deck of 52 playing cards containing four aces. Then turn up cards from the top until the first ace appears. On the average, how many cards are required to produce the first ace?

Analytical Answer

4 Aces can divide a deck into 5 portions. Each of these portions would on average have $\frac{48}{5} = 9.6$ cards in them.

Therefore the average number of cards required to draw the first ace would be $9.6 + 1 = 10.6$, in order to include the drawing of the ace itself.

Empirical Answer

```
question9 <- function() {

  deck <- c(rep("Ace", 4), rep("no", 48))       # create deck with 4 Aces, and 48 other
  shuffled.deck <- shuffle(deck)                # shuffle
  locations <- grep("Ace", shuffled.deck)       # find locations
  num.before <- min(locations)                  # find location of first Ace
  return(num.before)

}

num.cards.for.ace <- do(numsim) * question9()   # repeat
```

```r
num.cards.for.ace %>%
  unlist() %>%
  mean()
```

```
## [1] 10.7175
```