Assignment 7 - DATA 605

John Mazon

```r
require(tidyverse, quietly = TRUE, warn.conflicts = FALSE)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## Warning: package 'dplyr' was built under R version 4.0.3

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
theme_set(theme_bw())
```

1. Let $X_1, X_2, \ldots, X_n$ be n mutually independent random variables, each of which is uniformly distributed on the integers from 1 to k. Let Y denote the minimum of the X_i. Find the distribution of Y.

Because we are examining the distribution of the minimum value, we can obtain the probability of value Y where $Y \in 1, 2, \ldots, k$ by differencing the likelihood that all $X - i$ are equal to or greater than Y and the likelihood that all X_i are greater than Y:

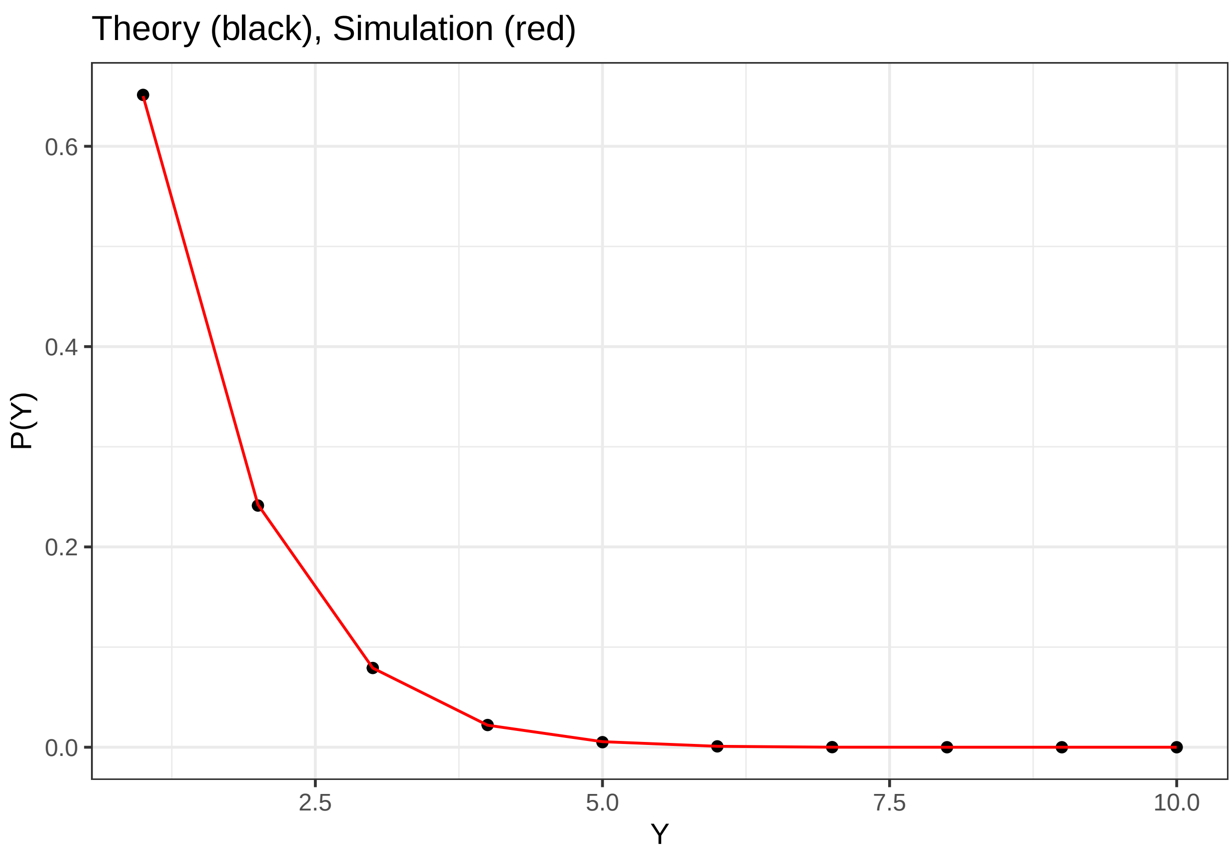
$$P(Y) = \left(\frac{k - Y + 1}{k}\right)^n - \left(\frac{k - Y}{k}\right)^n$$

As a check, I ran a simulation sampling 1e5 Y values for $n = k = 10$.

```r
n <- 10
k <- 10
f <- rep(0, k)
g <- rep(0, k)

for (j in 1:k) {
  f[j] <- ((k + 1 - j) / k)^n - ((k - j) / k)^n
}
y <- plyr::count(replicate(100000, min(sample(1:k, n, replace = TRUE))))
y$freq <- y$freq / 100000
S <- left_join(tibble(x = 1:10, P = f), y, by = "x") %>% replace_na(list(x = 0, P = 0, freq = 0))

ggplot(S, aes(x = x, y = P)) + geom_point() + geom_line(aes(y = freq), color = "red") +
  labs(x = "Y", y = "P(Y)", title = "Theory (black), Simulation (red)")
```



2. Your organization owns a copier (future lawyers, etc.) or MRI (future doctors). This machine has a manufacturer's expected lifetime of 10 years. This means that we expect one failure every ten years. (Include the probability statements and R Code for each part.).

a. What is the probability that the machine will fail after 8 years?. Provide also the expected value and standard deviation. Model as a geometric. (Hint: the probability is equivalent to not failing during the first 8 years.)

Modeling the process as a geometrically distributed random variable, we first must specify the probability of failure in a year, here $p = 1/10$. The probability of no failures in 8 years is then $P(X \geq 9) = (1-p)^8 = 0.43$ where X is the year in which the device fails.

In general, the geometric distribution follows:

$$P(X = k) = (1-p)^{k-1}p$$

for failure in year X. The mean and standard deviation are given by:

$$\mu = \frac{1}{p}$$

$$\sigma = \left(\frac{1-p}{p^2}\right)^{1/2}$$

For the problem at hand, the mean 10 years and the standard deviation is 9.49 years.

b. What is the probability that the machine will fail after 8 years?. Provide also the expected value and standard deviation. Model as an exponential. The general form of the exponential distribution is:

$$f(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0$$

which has mean and standard deviation:

$$\mu = \frac{1}{\lambda}$$
$$\sigma = \frac{1}{\lambda}$$

Because we have an expected failure age of 10 years, $\lambda = 1/10$, hence the likelihood of failure after 8 years is:

$$P(x > 8) = 1 - P(x \leq 8) = e^{-\frac{8}{10}} = 0.45$$

The mean and standard deviation are both 10 years.

c. What is the probability that the machine will fail after 8 years?. Provide also the expected value and standard deviation. Model as a binomial. (Hint: 0 success in 8 years). The general form of the binomial distribution is:

$$\begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x}$$

where x is the number of successes in n trials with p being the probability of success in a single trial.

If we treat a "success" in the binomial sense ("success" and "not success") as machine failure, then the probability of "success" in a year is $p = 1/10$. The likelihood of no "successes" in 8 years is:

$$\begin{pmatrix} n \\ 0 \end{pmatrix} p^0 (1-p)^8 = (1-p)^8 = 0.43$$

To obtain the mean and standard deviation we have to treat this as a negative binomial, which in this case (counting years to first failure) simplifies to the geometric distribution. Hence,the mean 10 years and the standard deviation is 9.49 years.

d. What is the probability that the machine will fail after 8 years?. Provide also the expected value and standard deviation. Model as a Poisson. The general form of the Poisson distribution is:

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where k is the number of events and λ is the rate parameter (expected number of events in unit time).

To model the probability of failure after 8 years as a Poisson, we assume a failure rate of $\lambda = 8/10$ Here the 8 is the number of years without a failure and 10 is the expected lifetime. Then the probability of 0 failures in an 8-year window is:

$$P(0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-8/10} = 0.45$$

The mean and standard deviation of failure age come from the exponential distribution which govern interval times between Poisson events, hence the mean and standard deviation are both 10 years.