DATA605 Homework 7

Keith Colella

2023-10-16

```
library(tidyverse)
set.seed(42)
```

Question 1 - Chapter 7.2 Q#11

A company buys 100 lightbulbs, each of which has an exponential lifetime of 1000 hours. What is the expected time for the first of these bulbs to burn out? (See Exercise 10.)

Response

We can treat the lifetime for each of the 100 lightbulbs as a separate random variable ($X_1, X_2, \ldots, X_100$). We then need to find the distribution of the minimum (M) lifetime, representing the first of the bulbs to burn out. Finally, we'll calculate the mean of that distribution to obtain the expected time for the first bulb to die.

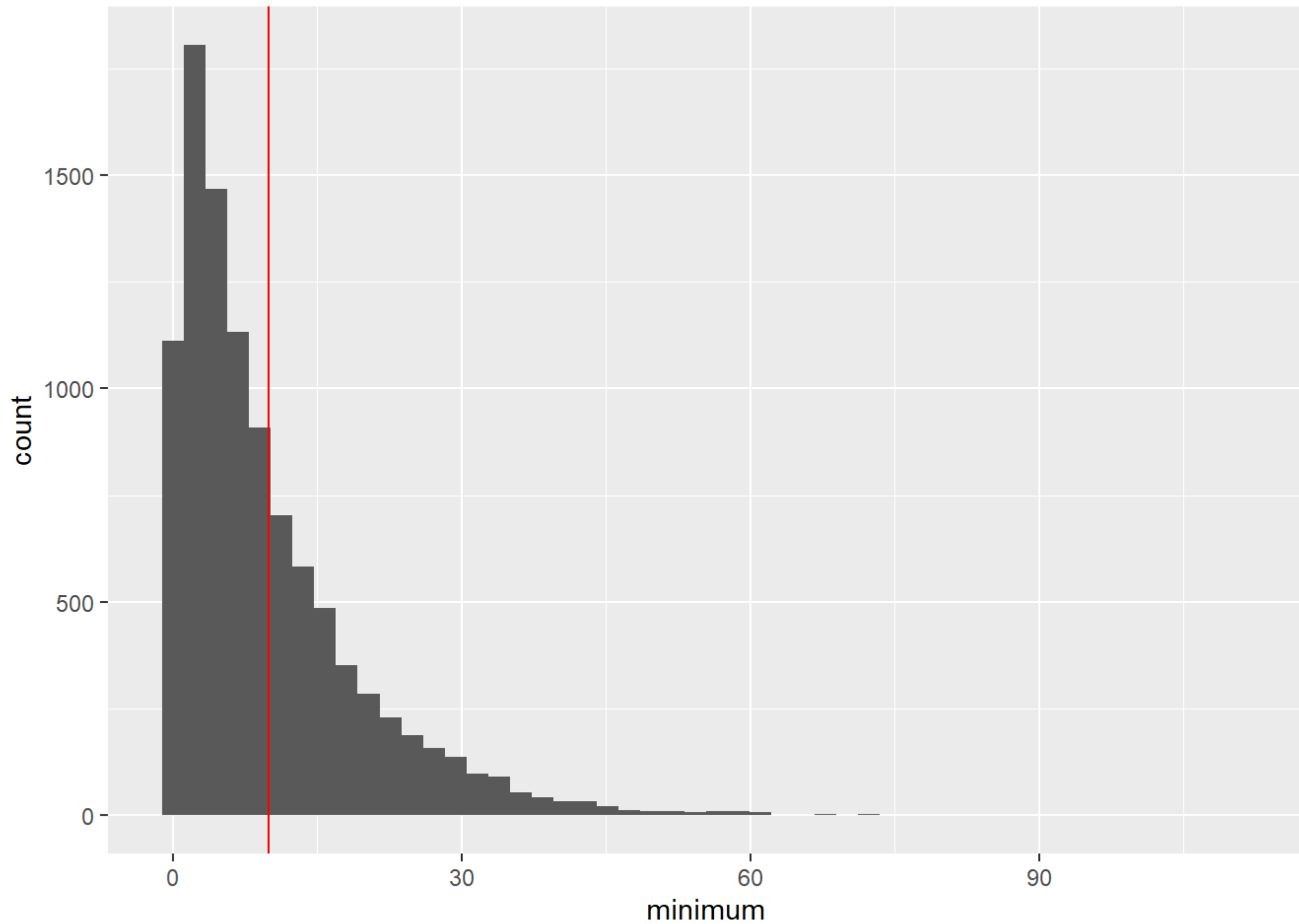
The key insight for this problem is provided by Q#10 in this same chapter, which notes that "the density for M is exponential with mean μ/n." We know the rate parameter (λ) for the distribution of bulb failures is 1/1000, so the mean μ is $1/\lambda = 1000$. Here we are concerned with 100 lightbults, so n = 100. So, to give us the mean of M, we divide μ (1000) by n (100), and get an expected value of 10 hours.

We can try and simulate this result, as well. We define an exponential distribution with our rate parameter λ and take 100 random samples, highlighting the minimum. We perform this 10,000 times, then aggregate the minimums from each iteration. We can plot that distribution and highlight the mean, which should align to our expected value of 10.

```
lambda <- 1 / 1000

n <- 100

results <- data.frame()

for (i in 1:10000) {
  minimum <- min(rexp(n, rate = lambda))
  result <- list(iteration = 0, minimum = minimum)
  results <- rbind(results, result)
}

expected_value <- mean(results$minimum)

cat('Expected value of simulated first dead bulb:', round(expected_value,4))
```

```
## Expected value of simulated first dead bulb: 9.9009
```

```
results %>%
  ggplot(aes(minimum)) +
  geom_histogram(bins = 50) +
  geom_vline(xintercept = expected_value, color = 'red')
```



The mean of our simulated distribution is very close to our expected value 10!

Question 2

Assume that X_1 and X_2 are independent random variables, each having an exponential density with parameter . Show that $Z = X1 - X2$ has density:

$$f_Z(z) = (1/2)\lambda e^{-\lambda|z|}$$

Response

Example 7.4 shows us that the **sum** of two independent random variables $Z = X_1 + X_2$ has density

$$F_Z(z) = \lambda_2 z e^{-\lambda z}$$

We can follow this example, but adapt it for $Z = X_1 - X_2$ (i.e. subtraction rather than additional). Whereas the convolution of the sum of two random variables involves the integral of $f_{X_1}(z - y)$ and $f_{X_2}(y)$ and dy, here we can rearrange our variables in terms of x, giving us three key terms: $f_X(x)$ and $f_Y(x - z)$ (and dx). Our integral is as follows.

$$\int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(x-z) dx$$

Because both variables follow an exponential distribution, their PDFs are defined as $\lambda e^{-\lambda x}$, and they are lower bound by 0. So we solve our integral as follows.

$$\int_{0}^{z} \lambda e^{-\lambda x} \lambda e^{-\lambda(x-z)} dx$$

The remainder of the solution requires integration by parts. I don't quite have the skills yet to fully break that down. But online tools help reduce this to two cases, where z > 0 and z < 0. These reduce to

$$\frac{1}{2} \lambda e^{\lambda z}$$

and

$$\frac{1}{2} \lambda e^{-\lambda z}$$

which can be expressed together as

$$\frac{1}{2} \lambda e^{-\lambda |z|}$$

Question 3

Let X be a continuous random variable with mean μ = 10 and variance σ^2 = 100/3. Using Chebyshev's Inequality, find an upper bound for the following probabilities.

a. $P(|X-10| \geq 2)$
b. $P(|X-10| \geq 5)$
c. $P(|X-10| \geq 9)$
d. $P(|X-10| \geq 20)$

Response

Chebyshev's Inequality is defined as follows.

$$P(|X-\mu| \geq k\sigma) \leq 1/k^2$$

So, for (a), $k\sigma = 2$. Given that $\sigma^2 = 100/3$, we know that $\sigma = \sqrt{100/3}$, so k is equal to $2/\sqrt{100/3}$. We can plug this into the inequality as follows.

$$\frac{1}{(2/\sqrt{100/3})^2}$$

This comes out to ~8.333, but because this is a probability, we limit the result to 1. So, $P(|X-10| \geq 2) \leq 1$, i.e. the probability that x deviates from the mean by more than 2 units is, at most, 1.

We can define a function to apply this same procedure for the other probabilities.

```r
cheb_ineq <- function(sigma, difference) {
  k <- difference / sigma
  upper_bound <- 1 / k^2
  upper_bound <- ifelse(upper_bound > 1, 1, upper_bound)
  return(upper_bound)
}

sigma <- sqrt(100/3)

a <- cheb_ineq(sigma, difference = 2)
b <- cheb_ineq(sigma, difference = 5)
c <- cheb_ineq(sigma, difference = 9)
d <- cheb_ineq(sigma, difference = 20)

cat(
  '(a) Upper bound of P(|X − 10| ≥ 2): ',round(a,4),'\n',
  '(b) Upper bound of P(|X − 10| ≥ 5): ',round(b,4),'\n',
  '(c) Upper bound of P(|X − 10| ≥ 9): ',round(c,4),'\n',
  '(d) Upper bound of P(|X − 10| ≥ 20): ',round(d,4),'\n',
  sep = ''
)
```

```
## (a) Upper bound of P(|X − 10| ≥ 2): 1
## (b) Upper bound of P(|X − 10| ≥ 5): 1
## (c) Upper bound of P(|X − 10| ≥ 9): 0.4115
## (d) Upper bound of P(|X − 10| ≥ 20): 0.0833
```