Mat 354

Homework 11

Kenny Roffo

Due November 4, 2015

1. **A-maze-ing**
In a lab, a strain of mice are selectively bred for intelligence. And...it turns out that the resulting mice are smarter than (other) mice. These mice are timed going through a maze to reach a reward of food. The time (in seconds) required for any mouse is a random variable Y with a density function given by

$$f(y) = \begin{cases} \frac{2b^2}{y^3} & b \leq y \\ 0 & b \not\leq y \end{cases}$$

where b is the minimum time needed to traverse the maze.

a) Find $F(y)$. Define it over the entire set of real numbers.

$F(y)$ is found by integrating the density function. We see

$$\begin{aligned} \int_b^y \frac{2b^2}{y^3} dy &= \int_b^y 2b^2 y^{-3} dy \\ &= \left[-b^2 y^{-2}\right]_b^y \\ &= 1 - b^2 y^{-2} \end{aligned}$$

Thus we have

$$F(y) = \begin{cases} 0 & y < b \\ 1 - \frac{b^2}{y^2} & b \leq y \end{cases}$$

b) If $c > 0$ is a positive constant, determine $P(Y > b + c)$.
No words necessary:

$$\begin{aligned} P(Y > b + c) &= 1 - P(Y < b + c) \\ &= 1 - F(b + c) \\ &= 1 - (1 - \frac{b^2}{(b + c)^2} \\ &= \frac{b^2}{(b + c)^2} \end{aligned}$$

c) If $d > c > 0$ are positive constants, find $P(Y > b + d|Y > b + c)$.
Given $Y > b + c$ we can treat $b + c$ as the minimum. Since $d > c$, we know $d - c > 0$, so we this probability simplifies greatly. Just note that treating $b + c$ as the minimum means we must use $b + c$ wherever b occurrs in the distribution function. That being said, we can begin the calculation:

$$\begin{aligned} P(Y > b + d|Y > b + c) &= P(Y > (b + c) + (d - c)) \\ &= 1 - F(d - c) \\ &= 1 - (1 - \frac{(b + c)^2}{((b + c) + (d - c))^2} \\ &= \frac{(b + c)^2}{(b + d)^2} \end{aligned}$$

d) Determine the 2.5^{th} and 97.5^{th} percentiles of the times. There is a 95% probability that a randomly selected mouses time falls between these two values. (In other words: 95% of these mice have times that fall between these two values.)
For the 2.5^{th} percentile we have

$$\begin{aligned} & F(y) = 0.025 \\ \Longrightarrow\ & 1 - \frac{b^2}{y^2} = 0.025 \\ \Longrightarrow\ & 0.975 = \frac{b^2}{y^2} \\ \Longrightarrow\ & y^2 = \frac{b^2}{0.975} \\ \Longrightarrow\ & y = \frac{b}{\sqrt{0.975}} \end{aligned}$$

and for the 97.5^{th} we've got

$$\begin{aligned}
&F(y) = 0.975 \\
\Longrightarrow\ & 1 - \frac{b^2}{y^2} = 0.975 \\
\Longrightarrow\ & 0.025 = \frac{b^2}{y^2} \\
\Longrightarrow\ & y^2 = \frac{b^2}{0.025} \\
\Longrightarrow\ & y = \frac{b}{\sqrt{0.025}}
\end{aligned}$$

Thus the percentiles are given by

$$\phi_{2.5} = \frac{b}{\sqrt{0.975}} \qquad \text{and} \qquad \phi_{97.5} = \frac{b}{\sqrt{0.025}}$$

e) State a general expression for finding the p^{th} quantile ($100p^{th}$ percentile) of Y; a formula in terms of p.

That's easy:

$$\phi_{100p} = \frac{b}{\sqrt{1-p}}$$

f) Which mouse is more likely to take more than twice the minimum time: the intelligent mice of this exercise, or the dumb mice of Exercise 4.15? Produce probabilities supporting your response.

By the same process as in part a we get that the distribution function for the mice in Exercise 4.15 is

$$F(y) = 1 - \frac{b}{y}$$

and also we see that $P(Y > 2b) = 1 - F(2b)$. Pluggin in for the distribution functions for this exercise and 4.15 respectively, we see

$$1 - \frac{b^2}{(2b)^2} = \frac{3}{4} \qquad \text{and} \qquad 1 - \frac{b}{2b} = \frac{1}{2}$$

Thus the probability is higher that the mice from this exercise will take at least twice as long as the minimum time.

2. **Max Again**

A random variable Y has distribution function

$$F(y) = \begin{cases} 0 & y < 0 \\ \frac{y(y+1)}{110} & 0 \leq y \leq 10 \\ 1 & 10 < y \end{cases}$$

This was covered in class. The density and some expectations were obtained:

$$E(Y) = 215/33 = 6.5152 \qquad \text{and} \qquad V(Y) = 6575/1089 = 6.0376$$

Imagine 5 independent runs of the experiment that results in this variable. Let M denote the maximum result for the 5 runs. Determine the following.

a) The distribution function for M: $F_M(m) = P_M(M \leq m)$ (Nothing has changed from earlier problems of this sort. If the maximum is no greater than m, what can you say about each of the 5 outcomes (Ys)? Use the independence to obtain an expression for this probability.)

The probability that the maximum of the five trials is a number m, $P(M \leq m)$, is the probability that each trial yields an outcome $\leq m$. Since all trials are independent, this implies

$$F_M(m) = P_M(M \leq m) = P(Y \leq m)^5 = F(m)^5 = \left(\frac{y(y+1)}{110}\right)^5$$

Note that this is only for choices of m between 0 and 10. Thus truly, we have

$$F_M(m) = \begin{cases} 0 & m < 0 \\ \left(\frac{y(y+1)}{110}\right)^5 & 0 \leq m \leq 10 \\ 1 & 10 < m \end{cases}$$

b) The density for M.

The density is found by simply taking the derivative of the distribution, and slapping 0 on the ends past the bounds of possible values:

$$\begin{aligned} f_M(m) &= \frac{\mathrm{d}}{\mathrm{d}m}\left[\left(\frac{m(m+1)}{110}\right)^5\right] \\ &= 5\left(\frac{m(m+1)}{110}\right)^4\left(\frac{2m+1}{110}\right) \end{aligned}$$

So

$$f_M(m) = \begin{cases} 0 & m < 0 \\ 5\left(\frac{m(m+1)}{110}\right)^4\left(\frac{2m+1}{110}\right) & 0 \leq m \leq 10 \\ 0 & 10 < m \end{cases}$$

c) The expected value of M. (Youll probably want to offload the integration to technology. Its easy but tedious.) Clearly state the integral; tell what tools you used to evaluate it; and then present the solution.

The expected value is given by

$$E(M) = \int_{-\infty}^{\infty} m f_M(m) dm$$

Thus we have

$$\begin{aligned}
E(M) &= \int_{-\infty}^{0} 0 dm + \int_{0}^{10} m \cdot 5 \left(\frac{m(m+1)}{110}\right)^4 \left(\frac{2m+1}{110}\right) dm + \int_{10}^{\infty} 0 dm \\
&= 0 + \int_{0}^{10} m \cdot 5 \left(\frac{m(m+1)}{110}\right)^4 \left(\frac{2m+1}{110}\right) dm + 0 \\
&= 9.0479
\end{aligned}$$

The calculation of this integral was done using Wolfram Alpha.

d) The standard deviation of M.
We find $E(M^2) = 82.619$ by the same process as part c. Now we find the variance

$$V(M) = E(M^2) - E(M)^2 = 0.7545$$

and take the square root for the standard deviation

$$\sigma(M) = \sqrt{V(M)} = 0.8686$$

e) The 5^{th} percentile of the distribution of M.

$$\begin{aligned}
F(m) = 0.05 &= \left(\frac{m(m+1)}{110}\right)^5 \\
\implies 60.42 &= m(m+1) \\
\implies m^2 + m - 60.42 &= 0 \\
\implies m &\in \{-8.2892, 7.2892\}
\end{aligned}$$

Since if m were negative then 0% of the data would be below it, m must be 7.2892. Thus the 5^{th} percentile of the distribution is 7.2892.

f) (Extra) The obvious follow up is to ask for parts a e when there are n (not necessarily 5) independent runs. And then look at what happens as $n \to \infty$.

Since this is extra I won't go into any gory details, but basically, as n grows larger we see

a) - The likelihood that all trials are below a value decreases, thus as n gets really big, the distribution function changes to stay close to 0 until it gets close to the maximum m value, which is 10.

b) - The density function behaves similaryly to the distribution function.

c) - The expected value approaches 10.

d) The standard deviation approaches 0.

e) The 5^{th} percentile approaches 10.