Jerry Cain
CS109

Section #3
April 20, 2021

Section #3: Discrete and Continuous Random Variables

1 Warmups

1.1 Website Visits

You have a website where only one visitor can be on the site at a time, but there is an infinite queue of visitors, so that immediately after a visitor leaves, a new visitor will come onto the website. On average, visitors leave your website after 5 minutes. Assume that the length of stay is exponentially distributed. What is the probability that a user stays more than 10 minutes, if we calculate this probability:

a. using the random variable X, defined as the length of stay of the user?

b. using the random variable Y, defined as the number of users who leave your website over a 10-minute interval?

If this problem doesn't convince you that the Poisson and Exponential RVc are coupled, then I'm not sure will! As defined above, $X \sim \text{Exp}(\lambda = \frac{1}{5})$.

$$P(X > 10) = 1 - F_X(10) = 1 - (1 - e^{-10\lambda}) = e^{-2} \approx 0.1353$$

Alternatively, we have that Y is the number of users leaving on the website in the next 10 minutes. The average number of users leaving is 2 users per 10 minutes. $Y \sim \text{Poi}(\lambda = 2)$.

$$\begin{aligned} P(Y = 0) &= \frac{2^0 e^{-2}}{0!} \\ &= e^{-2} \approx 0.1353 \end{aligned}$$

1.2 Continuous Random Variables

Let X be a continuous random variable with the following probability density function:

$$f_X(x) = \begin{cases} c(e^{x-1} + e^{-x}) & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a. Find the value of c that makes f_X a valid probability distribution.

b. What is $P(X > 0.75)$?

a. We need $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 c(e^{x-1} + e^{-x})dx$$
$$1 = c\left[e^{x-1} - e^{-x}\right]_{x=0}^{1}$$
$$1 = c(e^{1-1} - e^{-1} - (e^{0-1} - e^{-0}))$$
$$c = \frac{1}{1 - e^{-1} - (e^{-1} - 1)} = \frac{1}{2 - \frac{2}{e}}$$

b.

$$P(X > 0.75) = \int_{0.75}^{1} c(e^{x-1} + e^{-x})dx$$
$$= c\left[e^{x-1} - e^{-x}\right]_{x=0.75}^{1}$$
$$= c\left(e^{1-1} - e^{-1} - (e^{0.75-1} - e^{-0.75})\right)$$
$$= c\left(1 - e^{-1} - e^{-0.25} + e^{-0.75}\right) = \frac{1 - e^{-1} - e^{-0.25} + e^{-0.75}}{2 - \frac{2}{e}}$$

2 Problems

2.1 More Bit Strings

Once again, we're sending sending bit strings across potentially noisy communication channels, just like last week. However, this week we're identifying bit string corruptions in a slightly different way. Now, whenever we want to send n bits of information, we send an extra as the $n+1^{st}$ bit. Specifically, if the sum of the n data bits is even, the extra $n+1^{st}$ bit sent is set to 0. If the sum of the n data bits is odd, then $n+1^{st}$ bit appended is set to 1. If the recipient of the bit string adds all bits and gets an odd number, that recipient knows there's a problem and can request a repeat transmission. We'll assume that each bit is erroneously inverted with probability nonzero $p \leq 0.5$, and that all bit corruptions are independent of one another.

a. Assuming that n = 4 and p = 0.1, what is the probability the transmitted message has errors that go undetected?

b. For arbitrary n and p, what is the probability that a bit string has errors that go undetected? You may leave it as a sum of O(n) terms.

c. Simplify your answer from part b by letting $a = \sum_{\text{odd}k} \binom{n+1}{k} p^k (1-p)^{n+1-k}$ and $b = \sum_{\text{even}k} \binom{n+1}{k} p^k (1-p)^{n+1-k}$ and then considering what $a+b$ and $a-b$ equal. Leverage the fact that, in general, $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

a. Note that a bit string with an odd number of errors will be flagged as erroneous, but those bit strings with an even number of errors, regardless of n, will be taken as correct. That means the probability errors will go undetected is:

$$\binom{5}{2}(0.1)^2(0.9)^3 + \binom{5}{4}(0.1)^4(0.9)^1 = 0.07335 \tag{1}$$

b. The above generalizes to arbitrary n, so that the probability of interest is:

$$\sum_{\text{even } k\geq 2} \binom{n+1}{k} p^k (1-p)^{n+1-k} \tag{2}$$

c. If a and b are defined that way, then:

$$a + b = \sum_{\text{all} k} \binom{n+1}{k} p^k (1-p)^{n+1-k} = 1 \tag{3}$$

and

$$a - b = \sum_{\text{all} k} \binom{n+1}{k} (-p)^k (1-p)^{n+1-k} = (1-2p)^{n+1} \tag{4}$$

Solving for a, we arrive at:

$$a = \sum_{\text{even} k} \binom{n+1}{k} (-p)^k (1-p)^{n+1-k} = \frac{1+(1-2p)^{n+1}}{2} \tag{5}$$

Now, a includes a term for no errors, so we need to subtract that one term off so it doesn't contribute. That leaves us with our final answer for general n and p, which is:

$$\frac{1+(1-2p)^{n+1}}{2} - (1-p)^{n+1} \tag{6}$$

2.2 Air Quality

Throughout the United States, the Environmental Protection Agency monitors levels of PM2.5, a type of dangerous air pollution. These PM2.5 measurements can be approximately modeled by a normal distribution.

a. Let us model PM2.5 measurements with a normal distribution that has a mean of 8. If three-quarters of all measurements fall below 11.4, what is the standard deviation? Round to the nearest integer.

b. PM2.5 values above 12 can pose some health risks, especially to sensitive populations. Using the standard deviation found above, what is the probability that a randomly selected PM2.5 measurement is over 12?

c. What is the probability that a randomly selected PM2.5 measurement is between 7 and 8?

a. $\Phi(\frac{x-\mu}{\sigma}) = \Phi(\frac{11.4-8}{\sigma}) = 0.75 \implies \frac{3.4}{\sigma} \approx .68 \implies \sigma \approx 5.$

b. $P(q > 12) = 1 - P(q < 12) = 1 - \Phi(\frac{12-8}{5}) = 1 - \Phi(.8) = 1 - 0.7881 = 0.2119.$

c. $P(7 < h < 8) = P(h < 8) - P(h < 7) = \Phi(\frac{8-8}{5}) - \Phi(\frac{7-8}{5})$
$= \Phi(\frac{8-8}{5}) - \Phi(\frac{-1}{5})) = \Phi(\frac{8-8}{5}) - (1 - \Phi(\frac{1}{5}))$
$= \Phi(0) - (1 - \Phi(0.2)) = 0.5 - (1 - 0.5793) = 0.0793$