Homework 3
Math 183, UCSD, Fall 2017
Due on Friday 27th, October 12:50pm
Staple pages together

Exercise 1

Lynndyl is a pretty small city in Utah, with only 106 inhabitants. Seventy-one percent of the registered voters favor their incumbent mayor in her bid for re-election. If only twenty-four citizens go to the polls:

1. What is the probability that the race ends in a tie?

\[ P(X = 12) = \binom{24}{12} (0.71)^{12} (0.29)^{12} \approx 0.016 \]

2. What number of voters do you expect to have voted for her challenger?

Let \( Y \) be an r.v. for the no. of people who voted for her challenger. Here \( P(\text{prob}) = 1 - 0.71 = 0.29. \)

\( Y \sim \text{Binom}(24, 0.29). \) As \( Y \) also follows Binom dist

\[ E(Y) = nP = 24 \times 0.29 = 6.96 \]

6.96 voters are expected to vote for her challenger.
Exercise 2

1. Do you think the following set of data is likely to have been generated by a random variable \( X \) that has the geometric distribution \( P(X = k) = \left( \frac{4}{5} \right)^{k-1} \times \frac{1}{5} \), \( (k = 1, 2 \ldots) \)? Explain.

\[
\begin{align*}
2 & \ 8 & 1 & 2 & 2 & 5 & 1 & 2 & 8 & 3 \\
5 & 4 & 2 & 4 & 7 & 2 & 2 & 8 & 4 & 7 \\
2 & 6 & 2 & 3 & 5 & 1 & 3 & 3 & 2 & 5 \\
4 & 2 & 2 & 3 & 6 & 3 & 6 & 4 & 9 & 3 \\
3 & 7 & 5 & 1 & 3 & 4 & 3 & 4 & 7 & 6
\end{align*}
\]

You may arrive at multiple conclusions here. All sensible solutions are correct.

\[\underline{\text{Ans 1:}}\] mean of all elements = 3.94. X is following geom dist where mean \( E(X) = 1/p \). \( E(Y_5) = 5 \).

\[\underline{\text{Ans 2:}}\] (YES)

Mean of the geom \( Y_5 \) is 5 and \( SD(X) \approx 4.41 \).

\[\therefore\] All the values in the data lie within \( E(X) \pm SD(X) \). We can say that the data has been generated by the dist.

2. With your cellphone, transmission errors occur with a rate of approximately 2.5 per ten seconds. What is the probability that more than two errors will occur during the next half-minute?

Since we know the average behavior here and are looking to explore a specific case we can use Poisson model here.

Avg. number of errors per half minute \( \lambda = 2.5 \times 3 = 7.5 \).

Let \( X \) be r.v. for the no. of errors per half minute.

We need to calculate \( \Pr(X > 2) \).

\[
\begin{align*}
P(0) &= \left( \frac{e^{-7.5}}{7.5!} \right)^0 = 0.00005303044 \\
P(1) &= \left( \frac{e^{-7.5}}{7.5!} \right)^1 = 0.062412412 \\
P(2) &= \left( \frac{e^{-7.5}}{7.5!} \right)^2 = 0.155555555
\end{align*}
\]

\[
P(X > 2) = 1 - (P(0) + P(1) + P(2)) = 0.980
\]
Exercise 3

The Biology department bought a brand new spectrophotometer. This device was sold with a service contract that ensures 4 free breakdown repairs from a technician. A breakdown of such a device occurs, on average, three times a year. These breakdowns are independent from each other.

1. Write explicitly a model and a random variable that describes the time elapsed until the first intervention of a technician.

Here we are modeling the time we have to wait for an event to occur for the first time. We know the frequency of this event to occur i.e., \( \lambda = 3 \) times/year.

Let \( X \) be the r.v. for the time elapsed for the first event. \( X \sim \text{Exp}(\lambda) \). The density of this model is given by

\[
    f(x) = \begin{cases} 
        3e^{-3x} & x > 0 \\
        0 & \text{otherwise} 
    \end{cases}
\]

2. Write explicitly a random variable describing the time for the service contract to be fulfilled.

Here, we use the exponential model with \( \lambda = 3/4 \) because the service contract gets fulfilled after \( 4 \) breakdowns. Since, the expected time here is \( 4 \) times the expected value of \( X \) in \( \text{Exp}(\lambda) \).

Let \( Y \) be the r.v., then

\( Y \sim \text{Exp}(3/4) \)

3. What is the average time it will take for the service contract to be fulfilled?

\[
    E(Y) = \frac{1}{\lambda} = \frac{1}{3/4} \text{ years}.
\]

So, it takes \( 4/3 \) years on avg to fulfill the service contract.
Exercise 4

For an adult at rest, the time between two consecutive eye blinks (inter-blinking time), expressed in seconds, is well-modeled by a continuous random variable $X$ having density

$$f(x) = \begin{cases} \frac{c}{5}x^3 & 0 \leq x \leq 5 \\ 0 & \text{otherwise,} \end{cases}$$

where $c$ is a positive constant.

1. What value must $c$ be if $f(x)$ is to be a valid density function?

For $f(x)$ to be a valid density function, the integral equation should be satisfied:

$$\int_{-\infty}^{\infty} f(x) \, dx = 1$$

$$\int_{0}^{5} c \cdot x^3 \, dx = \left[ \frac{c}{4} x^4 \right]_0^5 = \frac{625c}{4} = 1$$

Thus, $c = \frac{4}{625}$

2. What is the probability someone does not blink for more than 3 seconds?

$$P(X > 3) = \int_{3}^{\infty} f(x) \, dx = \int_{3}^{5} \frac{c}{4} x^3 \, dx = \left[ \frac{c}{4} \cdot \frac{x^4}{4} \right]_3^5 = \frac{625c}{16} - \frac{81c}{16} = \frac{544c}{16} = 10.8704$$

3. What is the probability someone’s inter-blinking time is exactly 1.67 seconds?

$$P(X = 1.67) = 0$$ Since, area under the curve at that point is zero.

$$\int_{1.67}^{1.67} f(x) \, dx = 0$$
4. What is the expected time between two consecutive blinks?

\[
E(x) = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{\infty} \frac{\frac{4}{625}}{x^5} x^5 \, dx = \frac{4}{625} \times \frac{x^5}{5} \bigg|_0^5 \\
= 4 \text{ seconds}
\]

5. What is the standard deviation of this duration?

\[
\text{Var}(x) = E(x^2) - [E(x)]^2 = \int_{-\infty}^{\infty} x^2 f(x) \, dx - (4)^2.
\]

\[
= \int_{0}^{\infty} \frac{\frac{4}{625}}{x^5} x^5 \, dx = \left[ -\frac{4}{625} \times \frac{x^6}{6} \right]_0^5 - 16
\]

\[
= \frac{0}{625} \times 6 - 16 = 0.6667
\]

\[
\text{SD}(x) = \sqrt{\text{Var}(x)} \approx 0.816 \text{ seconds}
\]

6. During a medical experiment, a scientist tells you she observed an inter-blinking time among the shortest 5%. Find the longest inter-blinking time she could have observed.

Here 5% is the area under the density function curve from 0 to + to where + is the longest inter-blinking time she could have observed. 

\[
\int_{0}^{\text{+}} \frac{\frac{4}{625}}{x^5} x^5 \, dx = 0.05 \implies \left[ \frac{4}{625} \times \frac{x^4}{4} \right]_0^t = 0.05
\]

\[
t^4 = 0.05 \times 625
\]

\[
t \approx 2.364 \text{ seconds}
\]
Exercise 5

The average daily temperature in October in Boston is 61°F with a standard deviation of 5°F. These temperatures closely follow a normal distribution.

1. What is the probability of observing a 67°F temperature or higher in Boston during a randomly chosen day in October?

   \[
   \text{Let } X \text{ be the r.v. for daily temp in October.} \\
   X \sim N(61, 5), \text{ we need to find } P(X \geq 67) \\
   \text{Writing } Z = \frac{X - \mu}{\sigma} = \frac{67 - 61}{5} = 1.2. \\
   P(X \geq 67) = 1 - P(Z < 1.2) = 1 - 0.8849 \quad \text{(as per 2-tailed)} \\
   = 0.1151 \\
   \text{(alternatively, we could do } \text{pnorm}(67, \text{mean}=61, \text{sd}=5, \text{lower.tail}=F))
   \]

2. How warm are the hottest 20% of the days (days with highest average high temperature) during October in Boston?

   \[
   \text{Here, for the same } X \text{ we are trying to find } t \text{ such that } P(X > t) = 0.2. \\
   \text{. e. } P(X < t) = 1 - 0.8 = 0.2 \\
   Z = 0.84 \quad \text{(as per normal lookup of 2-tailed)} \\
   t = 61.2 \quad \text{(calc., } \text{pnorm}(0.2, \text{mean}=61, \text{sd}=5, \text{lower.tail}=F))
   \]

   We use the following equation to convert °F (Fahrenheit) to °C (Celsius):

   \[
   C = \frac{5}{9} (F - 32).
   \]

4. Write the probability model for the distribution of temperature in °C in October in Boston.

   \[
   \text{we are creating a new random variable } Y \text{ with} \\
   \text{mean } = \frac{5}{9} (61 - 32) \approx 16.111 \\
   \text{std. dev. } = \frac{5}{9} \times 5 \approx 2.778. \\
   Y \sim N(16.111, 2.778)
   \]
5. What is the probability of observing a 20°C (which roughly corresponds to 68°F) temperature or higher in October in Boston? Calculate using the °C model from question 4.

\[ Y \sim N(16.111, 2.778) \text{. we need to find } P(Y \geq 20) \]

\[ Z = \frac{20 - 16.111}{2.778} \approx 1.4 \]

\[ P(Y \geq 20) = P(Z \geq 1.4) = 1 - P(Z < 1.4) \]

\[ = 1 - 0.9192 \text{ (on } Z \text{-table lookup)} \]

\[ = 0.0808 \]

Alternatively, we can use R:

\[ \text{pnorm}(20, \text{mean} = 16.111, \text{sd} = 2.778, \text{lower.tail} = \text{F}) \]

6. Estimate the IQR of the temperatures (in °C) in October in Boston.

\[ \text{IQR} = Q_3 - Q_1 \]

given \( Y \sim N(16.111, 2.778) \).

\( Q_3 \) is supposed to have 25% of the cases greater than \( \mu \) and \( Q_1 \) is supposed to have 25% of the cases less than \( \mu \) or 75% of 1.

\[ : Q_3 = \text{qnorm}(0.75, \text{mean} = 16.111, \text{sd} = 2.778, \text{lower.tail} = \text{F}) \]

\[ \approx 17.985 \]

\[ Q_1 = \text{qnorm}(0.25, \text{mean} = 16.111, \text{sd} = 2.778) \]

\[ \approx 14.237 \]

\[ \text{IQR} = 14.237 \text{ to } 17.985 \]

(PS: You can notice this using Z-tables too.)