Model Answers Homework 3 Stat 310

Total points: 28 (plus 7 more bonus)

1. (8pts total, 2pts each). For the following random variables, identify the distribution including parameters that most closely matches the situation. What assumptions do you have to make?

a. (2pt) The metrorail comes every 10 minutes on average. Let $X_i$ be the time until the next train comes

EXAMPLE ANSWER: $X_1 \sim$ exponential$(\beta = 10)$

½ point for distribution

Also accept: normal$(\mu = 10$
uniform$(0, \geq 10)$
negative binomial$(r = 1, p = \frac{1}{10})$
ANYTHING

½ point for fitting the parameters in a way that corresponds to 10 minute interarrivals given the choice of distribution above

Like an exponential r.v., the waiting time is continuous, not discrete. The waiting time can not be negative, and the average waiting time will be 10 minutes.

½ point for explaining why the distribution is like the situation

However, we have to assume that waiting times are memoryless, (i.e, the amount of time left to wait is independent of the amount of time we have already weighted). This doesn't seem likely.

½ point for pointing out a necessary assumption or identifying how the distribution is different than the situation.

**b.** (2pts) A couple is trying to get pregnant. P(pregnancy) for 1 trial is 2.2%, $X_2$ is the number of times they try before success

Answer:

$X_2 \sim$ Negative Binomial $(r=1, p=0.022)$ $\quad$ $+\frac{1}{2}$ distribution

also accept geometric $(p=0.022)$ $\quad$ $+\frac{1}{2}$ parameter values

Reasons: The negative binomial describes the number of times it takes to get $r$ successes when each success has chance $p$ of occurring. $\quad$ $+\frac{1}{2}$ for reasons

Assumptions: but we must assume the chance of success in each trial is fixed and independent of the other trials. This wouldn't be the case since fertility fluctuates over time. $\quad$ $+\frac{1}{2}$ for assumption or caveat

c. (2pts) The patriots and Giants are playing in the Super Bowl
$X_3$ = the Patriots score minus the giants score.

ANSWER: No distribution that we know could describe this. Each team's score will be some weird sum of 7, 6, 3, 2, and 1's

Its much better to recognize what you can't do, then to mistakenly try to do it anyways.

+2 pts for realizing that none of the distributions they know can accurately model this.

(otherwise award partial points per the scheme established in a and b)

d. (2pts) $X_4 = 1$ if a coin flip is heads, 0 otherwise

Answer $X_4 \sim \text{Bernoulli}(0.5)$

+½ distribution

+½ parameter

A coin flip only has two outcomes, just like a bernoulli trial. The probability of heads on a fair coin is 0.5

+½ reasons

We have to assume the coin is fair. (i.e, that the true probability is whatever we decided it was). We should also assume that the coin can not land on its edge

+½ assumptions

**e.** (2pts) $X_5$ is your grade when we calculate it by using a spinner labelled from 3 to 19.

ANSWER $X_5 \sim$ (discrete) uniform $(3, 19)$

+½ dist.
+½ parameter values
+½ reason

reason - each integer on the interval between 3 and 19 may come up with equal probability

Assumption - We should assume that each integer has equal area (arc length) on the spinner or that Hadley doesn't spin it in a systematic way

+½ assumptions/critique

**f.** (2pts) The Auckland volcanoes erupt at a rate of 11 every 100,000 years. $X_6$ is the number of eruptions in the next 100 years.

ANSWER $X_6 \sim$ poisson $\left(\lambda = \frac{11}{1000} = 0.011\right)$

+½ distribution
+½ parameter values

reason: $X_6$ is a count (of eruptions during a 100 year period). Hence we can model it with a poisson distribution where $\lambda$ = the average/expected number of eruptions during 100 years. If the yearly rate is $\frac{11}{100000}$ eruptions, then the rate per century will be $\frac{11}{100000} \times 100 = 0.011$

+½ reason

Assumption: we must assume that each eruption is an independent event. In reality, eruptions in a small area, such as auckland, may be correlated

+½ assumption

**2.** (8pts total: 2 each). Compute $C$ so each is a pmf.

**a.** $f_1(x) = C_1 \cdot \log(x) \quad x = 1, 2, \ldots 5$

ANSWER: To be a pmf, a discrete function must sum to 1. This is because something must happen. Its also an out come of the law of total probability. So we answer 2, by solving for $C$ that causes $f(x)$ to sum to 1 over the given range

$$\sum_{x=1}^{5} C_1 \log(x) = 1$$

+1 pt for suming to 1

$$C_1 \sum_{x=1}^{5} \log(x) = 1$$

+1 pt for using correct range

$$C_1 = \frac{1}{\sum_{x=1}^{5} \log(x)}$$

$-\frac{1}{2}$ for a minor mistake

$$\boxed{C_1 = 0.209}$$

**b.** $f_2(x) = C_2 x^2 \quad x = 1, 2, \ldots n$

$$C_2 \sum_{x=1}^{n} x^2 = 1$$

-1 pt for summing to 1

+1 pt for using correct range (note $n \neq \infty$)

$$C_2 = \frac{1}{\sum_{x=1}^{n} x^2}$$

$$C_2 = \frac{6}{n(n+1)(2n+1)}$$ (using wolfram alpha)

doesn't matter where they stopped simplifying

c. $f_3(x) = \frac{C_3}{x^3}$ $x = 1, 2, \ldots$

$$C_3 \sum_{x=1}^{\infty} \frac{1}{x^3} = 1$$

+1 Summing to 1

+1 range $= 1 \to \infty$

$$C_3 = \frac{1}{\sum_{x=1}^{\infty} \frac{1}{x^3}}$$

$\approx 0.832$ (via Wolfram Alpha)

d. $f_4(x) = \frac{C_4}{x^2}$, $x = 1, 2, \ldots$

$$C_4 \sum_{x=1}^{\infty} \frac{1}{x^2} = 1$$

+1 Summing to 1

+1 range $1 \to \infty$

$$C_4 = \frac{1}{\sum_{x=1}^{\infty} \frac{1}{x^2}}$$

$= \frac{6}{\pi^2}$ (by Wolfram Alpha)

$\approx 0.608$

3. (8 pts) Calculate the expected values of the random variables with the above pmfs

ANSWER: Recall that $E(x) = \sum_x x\,p(x)$ where $p(x)$ is the pmf of $x$ as long as $\sum_x |x|\,p(x) < \infty$. So

a. $E(X_1) = \sum_{x=1}^{5} 0.209\, x \log(x)$ +1 correct use of definition of $E(x)$

$\approx 3.82$ (WA) +1 Correct range

b. $E(X_2) = \sum_{x=1}^{n} x \frac{6}{n(n+1)(2n+1)} x^2$ +1 correct use of definition

$= \frac{6}{n(n+1)(2n+1)} \sum_{x=1}^{n} x^3$ +1 correct range

$= \frac{3n(n+1)}{2(2n+1)}$

Note that when $n=\infty$ $\sum_x |x|\,p(x) = \infty$, in which case the expectation does not exist

c. $E(X_3) = \sum_{x=1}^{\infty} x \cdot \frac{1}{\sum_{x=1}^{\infty} \frac{1}{x^3}} \frac{1}{x^3}$ +1 for correct use of definition

$\approx \sum_{x=1}^{\infty} 0.832 \frac{1}{x^2}$ +1 for correct range

$= 1.37$ (WA)

**d.** $E(X_4) = \sum_{x=1}^{\infty} x \frac{6}{\pi^2} \frac{1}{x^2}$

$= \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{1}{x}$

$= \infty$

The sum does not converge, so $E(X_4)$ does not exist (or alternatively, $E(X_4) = \infty$)

**4.** (2 pts Bonus) How can you turn an arbitrary function $f$ into a pmf? What restrictions must you put on it?

ANSWER

1.) $f$ must be $\geq 0$ for all $x$, so we must adjust it to be so if it is not eg. $g(x) = e^{f(x)}$ or $g(x) = f(x) - \min_{x \in S} f(x)$ +1 point

2.) $\sum_x f = 1$ or $\int_x f \, dx = 1$, i.e. $f$ must sum to 1 if it is discrete, or integrate to 1 if it is continuous. We can make this so by multiplying $f$ by a suitable normalizing constant, $C$, provided that the sum/integral is finite. +1 point

**5.** (5 points) Attend Rob Tibshirani's lecture and write 2 points about what you learned and how it relates to stat 310.

Give 5 points for a response that shows students thought about the topic and put some effort into spotting connections to 310. Its okay if they didn't fully understand Tibshirani's lecture.

3 points for a mediocre response that does not show much thought.

1 point for a less than two paragraphs. ~~short response that does the bare minimum of what is aske~~