Homework 3

Stat 345 - Spring 2020

Name: ________________________________

Problem 1

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

$$p(x) = \frac{3x+1}{22} \qquad x = 0, 1, 2, 3$$

a) Verify that this pmf is a valid probability mass function.

The pmf is valid if

$$\sum_{i=0}^{3} P(X = x) = 1$$

$$\frac{3 \cdot 0 + 1}{22} + \frac{3 \cdot 1 + 1}{22} + \frac{3 \cdot 2 + 1}{22} + \frac{3 \cdot 3 + 1}{22} = \frac{1 + 4 + 7 + 10}{22} = 1$$

b) Find $P(1 \leq X < 3)$ using pmf.

$$P(1 \leq X < 3) = P(X = 1) + P(X = 2) = \frac{3 \cdot 1 + 1}{22} + \frac{3 \cdot 2 + 1}{22} = \frac{11}{22} = 0.5$$

c) Determine the cumulative distribution function (cdf) of X.

$$F(x) = P(X \leq x) = \sum_{y=0}^{x} P(X = y) = \sum_{y=0}^{x} \frac{3y+1}{22}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{3 \cdot 0+1}{22} = \frac{1}{22} & 0 \leq x < 1 \\ \frac{1}{22} + \frac{3 \cdot 1+1}{22} = \frac{5}{22} & 1 \leq x < 2 \\ \frac{5}{22} + \frac{3 \cdot 2+1}{22} = \frac{12}{22} & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

d) Determine expected value and variance for X.

$$E(X) = \sum_{i=0}^{3} xP(X = x) = 0 \cdot \frac{3 \cdot 0 + 1}{22} + 1 \cdot \frac{3 \cdot 1 + 1}{22} + 2 \cdot \frac{3 \cdot 2 + 1}{22} + 3 \cdot \frac{3 \cdot 3 + 1}{22} = \frac{4 + 14 + 30}{22} = \frac{24}{11} = 2.18$$

$$E(X^2) = \sum_{i=0}^{3} x^2 P(X = x) = 0^2 \cdot \frac{3 \cdot 0 + 1}{22} + 1^2 \cdot \frac{3 \cdot 1 + 1}{22} + 2^2 \cdot \frac{3 \cdot 2 + 1}{22} + 3^2 \cdot \frac{3 \cdot 3 + 1}{22} = \frac{4 + 28 + 90}{22} = \frac{122}{22} = 5.55$$

$$Var(X) = E(X^2) - (E(X))^2 = 5.55 - (2.18)^2 = 0.8$$

Problem 2

Let X be a random variable with probability density function

$$f(x) = c(1 - x^2) \qquad -1 < x < 1$$

and $f(x) = 0$ otherwise

a) What is the value of c?

The pdf is valid if

$$\int_{-1}^{1} c(1 - x^2)dx = 1$$

$$1 = \int_{-1}^{1} c(1 - x^2)dx = c(x - \frac{x^3}{3})\Big|_{-1}^{1} = c((1 - 1/3) - (-1 + 1/3)) = \frac{4}{3}c$$

Therefore, $c = \frac{3}{4}$ and pdf is $f(x) = \frac{3}{4}(1 - x^2)$.

b) What is the cumulative distribution function of X?

$$F(x) = P(X \leq x) = \int_{-1}^{x} \frac{3}{4}(1 - y^2)dy = \frac{3}{4}(y - \frac{y^3}{3})\Big|_{-1}^{x} = \frac{3}{4}(x - x^3/3 + 2/3)$$

$F(x) = 0$, if $x < -1$ and $F(x) = 1$, if $x \geq 1$.

c) Find $P(-0.5 \leq X < 0.2)$.

Using cdf:

$$P(-0.5 \leq X < 0.2) = P(X < 0.2) - P(X < -0.5) = F(0.2) - P(-0.5) =$$

$$= \frac{3}{4}(0.2 - 0.2^3/3 + 2/3) - \frac{3}{4}(-0.5 - (-0.5)^3/3 + 2/3) = 0.49$$

Using pdf:

$$P(-0.5 \leq X < 0.2) = \int_{-0.5}^{0.2} \frac{3}{4}(1 - x^2)dx = 0.49$$

.

d) Determine expected value and variance for X.

$$E(X) = \int_{-1}^{1} x \cdot \frac{3}{4}(1 - x^2)dx = \frac{3}{4}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Big|_{-1}^{1} = 0$$

$$E(X^2) = \int_{-1}^{1} x^2 \cdot \frac{3}{4}(1 - x^2)dx = \frac{3}{4}\left(\frac{x^3}{3} - \frac{x^5}{5}\right)\Big|_{-1}^{1} = 0.2$$

$$Var(X) = E(X^2) - (E(X))^2 = 0.2 - 0^2 = 0.2$$

Problem 3

Suppose there are 3.4 millions flights every month worldwide. The probability that the commercial airplane will crash is 10^{-6}. What is the probability that there will be

Since each flight has a small probability of crashing it seems reasonable to suppose that the number of crashes (X) is approximately Poisson distributed with rate $\lambda = np = 3.4 \cdot 10^6 \cdot 10^{-6} = 3.4$.

a) at least 2 such accidents in the next month?

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-3.4}\frac{3.4^0}{0!} - e^{-3.4}\frac{3.4^1}{1!} = 1 - e^{-3.4} - 3.4e^{-3.4} = 0.853$$

b) at most 1 accident in the next month?

$$P(X \leq 1) = P(X = 0) + P(X = 1) = e^{-3.4}\frac{3.4^0}{0!} + e^{-3.4}\frac{3.4^1}{1!} = 0.147$$

$dpois(0, 3.4) + dpois(1, 3.4) = 0.147$ Note that the probability can be computed as $P(A) = 1 - P(A^c)$:

$$P(X \leq 1) = 1 - P(X \geq 2) = 0.147$$

Problem 4

If the probability of hitting a target is 0.2, and 11 shots are fired independently. Define the random variable X and its distribution.

$X \sim Binomial(n = 11, p = 0.2)$, here $p = 0.2$ is the probability of success (hitting) and n is the size. The pmf is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

where k is the number of successes.

a) What is the probability of the target being hit at least 3 times?

$$P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2) = 1 - \binom{11}{0} 0.2^0 (1 - 0.2)^{11-0} - \binom{11}{1} 0.2^1 (1 - 0.2)^{11-1} -$$

$$-\binom{11}{2} 0.2^2 (1 - 0.2)^{11-2} = 0.383$$

$1 - (dbinom(0, 11, 0.2) + dbinom(1, 11, 0.2) + dbinom(2, 11, 0.2)) = 0.383$

b) What is the probability of scoring more hits than misses?

$$P(X \geq 6) = \sum_{k=6}^{11} \binom{11}{k} 0.2^k (1 - 0.2)^{11-k} = 0.012$$

$dbinom(6, 11, 0.2) + dbinom(7, 11, 0.2) + dbinom(8, 11, 0.2) + dbinom(9, 11, 0.2) + dbinom(10, 11, 0.2) + dbinom(11, 11, 0.2) = 0.012$

c) What is the expected value and variance of the number of hits in 11 shots?

$$E(X) = np = (11)(0.2) = 2.2$$
$$Var(X) = np(1-p) = (11)(0.2)(1-0.2) = 1.76$$

d) What is the conditional probability that the target is hit at least 3 times, given that it is hit at least once?

Using Bayes' rule,

$$P(X \geq 3 | X \geq 1) = \frac{P(X \geq 3 \cap X \geq 1)}{P(X \geq 1)} = \frac{P(X \geq 3)}{1 - P(X = 0)} = \frac{0.383}{1 - \binom{11}{0} 0.2^0 (1-0.2)^{11-0}} = \frac{0.383}{0.914} = 0.419$$

e) Suppose the person has to pay \$5 to enter the shooting range and he gets \$3 dollars for each hit. Find the expectation and the variance of the profit.

Denote the profit by $Y = 3X - 5$.

$$E(Y) = E(3X - 5) = 3E(X) - 5 = 3(2.2) - 5 = 1.6$$
$$Var(Y) = Var(3X - 5) = 3^2 Var(X) - 0 = 9(1.76) = 15.84$$

Problem 5

The life of a semiconductor laser at a constant power is normally distributed with a mean of 7000 hours and a standard deviation of 600 hours.

Denote the life of the laser by X, then $X \sim N(7000, 600^2)$ and $Z = \frac{X-7000}{600} \sim N(0,1)$, $\Phi(z) = P(Z \leq z)$.

a) What is the probability that a laser fails before 5000 hours?

$$P(X < 5000) = P\left(\frac{X - 7000}{600} < \frac{5000 - 7000}{600}\right) = P(Z < -3.33) = \Phi(-3.33) = 0.0004$$

$pnorm(-3.33) = 0.0004$

b) What is the life in hours that 90% of the lasers exceed?

Find x such that

$$P(X > x) = 0.90$$

$$P(X > x) = P\left(\frac{X - 7000}{600} > \frac{x - 7000}{600}\right) = P\left(Z > \frac{x - 7000}{600}\right) = 1 - \Phi\left(\frac{x - 7000}{600}\right) = 0.90$$

$$\Phi\left(\frac{x - 7000}{600}\right) = 0.10$$

Since $\Phi(-1.28) = 0.10$ using standard normal table (also $pnorm(-1.28) = 0.10$), then

$$\frac{x - 7000}{600} = -1.28$$

and $x = 6232$ hours.

c) If four lasers are used in a product and they are assumed to fail independently, what is the probability that all four are still operating after 7000 hours?

The standard normal density curve is symmetric with the center at $\mu = 0$.

$(P(X > 7000))^4 = P(Z > 0)^4 = (\frac{1}{2})^4 = 0.0625$

d) *(bonus (5 pts))* Given that a laser already lasts 5000 hours, what is the probability that it lasts at least another 3000 hours?

Using Bayes' rule,

$$P(X > 3000 + 5000|X > 5000) = \frac{P(X > 8000 \cap X > 5000)}{P(X > 5000)} = \frac{P(X > 8000)}{P(X > 5000)} = \frac{1 - P(Z < \frac{8000-7000}{600})}{1 - P(Z < \frac{5000-7000}{600})} =$$

$$= \frac{1 - \Phi(1.67)}{1 - \Phi(-3.33)} = \frac{0.04745968}{0.9995658} = 0.047$$

Problem 6

Insects are expected to be attracted to the roses. A commercial insecticide is advertised as being 98% efective. Suppose 2000 insects infest a rose garden where the insecticide has been applied, and let X=number of surviving insects. Evaluate the probability that fewer than 35 insects survive.

The size is $n = 2000$, the probability of surviving is $p = 1 - 0.98 = 0.02$.

a) Write the expression for the probability using Binomial distribution that fewer than 35 insects survive. Do not calculate.

$$P(X < 35) = \sum_{k=0}^{34} \binom{2000}{k} (0.02)^k (0.98)^{2000-k}$$

If you want to calculate this in R (not required):

```
res = 0
f <- function(x){ choose(2000, x)*(0.02)^x*(0.98)^(2000-x) }
for(x in 0:34){ res = res + f(x) }
res
[1] 0.1913021
```

b) Write the expression for the probability using the Poisson distribution that fewer than 35 insects survive. Do not calculate.

The rate $\lambda = np = 2000(0.02) = 40$

$$P(X < 35) = \sum_{k=0}^{34} e^{-40} \frac{40^k}{k!}$$

If you want to calculate this in R (not required):

```
res = 0
f <- function(x){ exp(-40)*40^x/factorial(x) }
for(x in 0:34){ res = res + f(x) }
res
[1] 0.1938755
```

c) Use the normal approximation to the probability that fewer than 35 insects survive. Calculate the probability.

We can approximate either Binomial or Poisson with the Normal distribution. We will add the continuity correction to approximate a discrete distribution by the continuous. Let's approximate Binomial by Normal ($np(1-p)$ is large enough):

$$P(X < 35) = P(X < 35.5) = P\left(\frac{X - np}{\sqrt{np(1-p)}} < \frac{35.5 - np}{\sqrt{np(1-p)}}\right) = P\left(Z < \frac{35.5 - 40}{\sqrt{40(0.98)}}\right) =$$

$$= P(Z < -0.719) = 0.236$$

Now let's approximate Poisson by Normal (λ is large enough):

$$P(X < 35) = P(X < 35.5) = P\left(\frac{X - \lambda}{\sqrt{\lambda}} < \frac{35.5 - \lambda}{\sqrt{\lambda}}\right) = P\left(Z < \frac{35.5 - 40}{\sqrt{40}}\right) =$$

$$= P(Z < -0.712) = 0.238$$

. The approximated probability is close to the true one of 0.191.

Problem 7

Imagine that subway trains in New York City always arrive exactly on time and run every day 24 hours a day, with the time between successive trains fixed at 12 minutes. Bob arrives at the train stop at a uniformly random time on a certain day (the time that Bob arrives is independent of the train arrival process).

a) What is the distribution of how long Bob has to wait for the next train? What is the average time that he has to wait?

The distribution is Uniform on (0, 12), so the mean is 6 minutes.

b) Given that the train has not yet arrived after 7 minutes, what is the probability that Bob will have to wait at least 2 more minutes?

Let T be the waiting time. Then by applying Bayes' rule,

$$P(T \geq 2 + 7 | T > 7) = \frac{P(T \geq 9 \cap T > 6)}{P(T > 7)} = \frac{P(T \geq 9)}{P(T > 7)} = \frac{1/12}{5/12} = \frac{1}{5}$$

c) Bob moves to Los Angeles where subway system is less organized. Now, when any train arrives, the time until the next train arrives is an Exponential random variable with mean 12 minutes. Bob arrives at the train stop at a random time. What is the distribution of his waiting time for the next train? What is the average time that Bob has to wait?

By the memoryless property, the distribution is Exponential with parameter 1/12 regardless of when Bob arrives; how much longer the next train will take to arrive is independent of how long ago the previous one arrived. The average time that Bob has to wait is 12 minutes.

d) When Bob complains to a friend how much worse transportation is in Los Angeles, the friend says that Bob arrives at a uniform instant between the previous train arrival and the next train arrival. The average length of that interval between trains is 12 minutes, but since Bob is equally likely to arrive at any time in that interval, his average waiting time is only 6 minutes. Explain what is wrong with his friend's reasoning.

Bob's friend is replacing a random variable (the time between trains) by its expectation (12 minutes), therefore he is ignoring the variability in interarrival times. The average length of a time interval between two trains is 12 minutes, but Bob is not equally likely to arrive at any of these intervals: Bob is more likely to arrive during a long interval between buses than to arrive during a short interval between buses. For example, if one interval between buses is 50 minutes and another interval is 5 minutes, then Bob is 10 times more likely to arrive during the 50-minute interval. This phenomenon is known as length-biased sampling.

e) *(bonus (5 pts))* Tired of the poor public transportation in Los Angeles, Bob bought the car. Eventually, in a few years Bob moves back to New York City and wants to sell his car. He decides to sell it to the first person to offer at least \$13,000 for it. Assume that the offers are independent Exponential random variables with mean \$10,000. Find the expected number of offers Bob will have.

The offers on the car are independent, each distributed as $X_i \sim Exponential(10^{-4})$. He decides to sell it to the first person to offer at least \$10,000 for the car. This process can be modelled using $Geometric(p)$ distribution, because Bob accepting offers until the first success ($> 13,000$). The probability of success is $p = P(X_i \geq 13,000) = 1 - P(X_i < 13,000) = 1 - F(13,000) = 1 - (1 - e^{10^{-4}}) = e^{13,000 \cdot 10^{-4}} = e^{-1.3}$. The expected value for Geometric(p) r.v. is $1/p$ that leads to Bob's expected number of offers to get the first successful offer is $e^{1.3} = 3.67$.

Problem 8

A truncated discrete distribution is one in which a particular class cannot be observed and is eliminated from the sample space. In particular, X has range $0, 1, 2, ...$ and 0 class cannot be observed (as is usually the case), the zero-truncated random variable T has pmf

$$P(T = x) = \frac{P(X = x)}{P(X > 0)}, \quad x = 1, 2, ...$$

a) Find pmf of the zero-truncated Poisson random variable T if $X \sim Poisson(\lambda)$

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, ...$$

$$P(X > 0) = 1 - P(X = 0) = 1 - e^{-\lambda}\frac{\lambda^0}{0!} = 1 - e^{-\lambda}$$

Therefore for $t = 1, 2, ...$

$$P(T = x) = \frac{P(X = x)}{P(X > 0)} = \frac{e^{-\lambda}\lambda^x}{x!(1 - e^{-\lambda})}, \quad x = 1, 2, ...$$

b) *(bonus (5 pts))* Find the expected value of the zero-truncated Poisson random variable T.

The expected value for any discrete zero-truncated random variable is

$$E(T) = \sum_{x=1}^{\infty} xP(T = x) = \sum_{x=1}^{\infty} x\frac{P(T = x)}{P(X > 0)} = \frac{1}{P(X > 0)}\sum_{x=1}^{\infty} xP(T = x) =$$

$$\frac{1}{P(X>0)}\sum_{x=0}^{\infty} xP(T=x) = \frac{E(X)}{P(X>0)}$$

If $X \sim Poisson(\lambda)$ with $E(X) = \lambda$,

$$E(T) = \frac{\lambda}{1-e^{-\lambda}}$$

c) Find pmf of the zero-truncated Negative Binomial random variable T if $X \sim NegativeBinomial(r,p)$

$$P(X=x) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0,1,2,...$$

$$P(X>0) = 1 - P(X=0) = 1 - \binom{r+0-1}{0} p^r (1-p)^0 = 1 - p^r$$

Therefore for $t = 1,2,...$

$$P(T=x) = \frac{P(X=x)}{P(X>0)} = \frac{\binom{r+x-1}{x} p^r (1-p)^x}{1-p^r}, \quad x = 1,2,...$$

d) *(bonus (5 pts))* If we let $r \to 0$ in c), we get an interesting distribution, called the *logarithmic series distribution*. A random variable Y has a logarithmic series distribution with parameter p if

$$P(Y=y) = \frac{-(1-p)^y}{ylog(p)}, \quad y = 1,2,..., \quad 0 < p < 1$$

Find the expected value of Y.

Note that the sum of geometric series $\sum_{y=0}^{\infty}(1-p)^y = \frac{1}{1-(1-p)} = \frac{1}{p}, \quad |p| < 1.$

$$E(Y) = \sum_{y=1}^{\infty} \frac{-y(1-p)^y}{ylog(p)} = -\sum_{y=1}^{\infty} \frac{(1-p)^y}{log(p)} = -\frac{1}{log(p)}\sum_{y=1}^{\infty}(1-p)^y = -\frac{1}{log(p)}\left(\sum_{y=0}^{\infty}(1-p)^y - 1\right) = -\frac{1}{log(p)}\left(\frac{1}{p} - 1\right)$$

Interesting facts: *Zero-truncated discrete distributions* (there are many of them) can model the count data when the number of instances or individuals falling into the zero-category class cannot be determined. For example, zero-truncated discrete distributions can model the length of hospital stay in days (the minimum of at least one day is required) or model the numbers of traffic violation for drivers during a certain period (there will be no record of those who have received no tickets). The *logarithmic series distribution* has proven useful in modeling how rare a species is relative to other species in a certain location. It is also can be used to model the distribution of numbers of items of a product purchased by a customer during a certain time period.