Assignment 8: Discrete random variables
MACSS 33000 1
Due Friday, September 6

1 PMF vs CMF

Consider the following function: $f(x) = \frac{1}{8}$. Find the pmf and cmf of the function and provide them in a table below.

Solution

Considering the function $f(x)$ for discrete random variable X, we may generate a pmf as the probability that X lies somewhere on the function. Since this function is uniform at $\frac{1}{8}$, we have the pmf

$$\Pr(X = x) = p_X(x) = \frac{1}{8} \text{ such that } x \in \{0, 1, ..., 7\}.$$

The range was found by solving for m such that

$$\sum_{x=0}^{m} p_X(x) = 1.$$

Using the pmf to generate the cmf, we have

$$\Pr(X \le x) = F(x) = \frac{1}{8}x + \frac{1}{8} \text{ such that } x \in \{0, 1, ..., 7\}$$

$X = 1$	$p_X(x)$	$F(x)$
$X = 0$	$\frac{1}{8}$	$\frac{1}{8}$
$X = 1$	$\frac{1}{8}$	$\frac{1}{4}$
$X = 2$	$\frac{1}{8}$	$\frac{3}{8}$
$\vdots$	$\vdots$	$\vdots$
$X = 7$	$\frac{1}{8}$	1

2 Conversion of temperatures

A city's temperature is modeled as a random variable with mean and standard deviation equal to 10 degrees Celsius. A day is described as "normal" if the temperature during that day ranges within one standard deviation from the mean. What would be the temperature range for a normal day if temperature were expressed in degrees Fahrenheit?

Solution

Temperature range in this case is the mean plus (and minus) one standard deviation.

$$\text{Temp} = 10 \deg C \pm 10 \deg C = (0 \deg C, 20 \deg C) = (35 \deg F, 68 \deg F)$$

3 Getting a traffic ticket

You drive to work 5 days a week for a full year (50 weeks), and with probability $p = 0.02$ you get a traffic ticket on any given day, independent of other days. Let X be the total number of tickets you get in the year.

1. What is the probability that the number of tickets you get is exactly equal to the expected value of X?

   (a) The random variable X counts the total number of tickets you get in the year, so we have
   $$X \sim \text{binomial}(N, p) = \text{binomial}(250, 0.02)$$

   (b) Constructing the pmf, we have
   $$p_X(k) = \binom{250}{k}(0.02)^k(0.98)^{250-k} \text{ such that } k \in \{0, ..., 250\}$$

   (c) Computing the expected value of X using the pmf,
   $$E[X] = \sum_{i=1}^{N} E[Y_i]$$
   $$= Np$$
   $$= 250(0.02)$$
   $$= 5.$$

   (d) Now finding the probability that the number of tickets you get is exactly equal to the expected value of X:
   $$\Pr(X = 5) = p_X(5)$$
   $$= \binom{250}{5}(0.02)^5(0.98)^{245}$$
   $$\approx 0.18$$

2. Calculate approximately the probability in (1) using a Poisson approximation. We are given that
   $$e^{-\lambda}\frac{\lambda^k}{k!} \approx \binom{n}{k}\pi^k(1 - \pi)^{n-k} \text{ where } \lambda = np.$$

   Using this, we have
   $$p_X(5) = e^{-5}\frac{5^5}{5!} \approx 0.18.$$

4   The Unbirthday Song

"The Unbirthday Song" from "Alice in Wonderland" can be sung to an individual on any day it is not that person's birthday with p = 364/365. You decide to sing this song to N random people until you encounter someone who's birthday is today. At this point your singing streak ends. Find the PMF, the expected value, and the variance of the number of people needed before you encounter a person whose birthday is today.

Solution

Let X be the random variable that counts the number of people encountered before finding someone whose birthday is today. This implies $X \sim \text{geometric}(\frac{364}{365})$. Generating the pmf for this:

$$f_X(k) = (1 - p)^k p \text{ for } p = \frac{364}{365}, \ k = 1, 2, ...$$

Finding the expected value,

$$E[X] = \sum_{i=1}^{\infty} E[Y - i] \text{ for } Y_i \sim \text{bernoulli}(p) = \frac{1}{p} = \frac{365}{364}.$$

Finding the variance,

$$Var[X] = \frac{1-p}{p^2} = \frac{1}{365} \cdot \frac{365^2}{364^2} = \frac{365}{364^2}$$

5 Properties of variance

For a discrete random variable X, show that:

$$Var(X) = E[X^2] - E[X]^2$$

Proof. Assume that the mean of X is finite and exists. Let $E[X] = \mu_x$ as the mean distribution of X. We may then use $r(X) = (X - \mu_X)^2$ as a measure of variability in the distribution of X. Now starting with the definition of the variance,

$$\begin{aligned}
Var(X) &= E[(X - \mu_X)^2] \\
&= E[X^2 - 2X\mu_X + \mu_X^2] \\
&= E[X^2] - 2\mu_X E[X] + \mu_X^2 \\
&= E[X^2] - 2E[X]^2 + E[X]^2 \qquad \text{Using } E[X] = \mu_x \\
&= E[X^2] - E[X]^2
\end{aligned}$$

6 Calculate an exact probability

The random variable X has a probability density function:

$$f(x; \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}$$

for $x = 0, 1, 2, ...$ (i.e. X has a Poisson distribution with parameter λ). In a lengthy manuscript, it is discovered that only 13.5 percent of the pages contain no typing errors.

If we assume that the number of errors per page is a random variable with a Poisson distribution, find the percentage of pages with exactly one error.

Solution.

$$\begin{aligned}
\Pr(X = x; \lambda) &= e^{-\lambda} \frac{\lambda^x}{x!} \\
\Pr(X = 0; \lambda) &= e^{-\lambda} \frac{\lambda^0}{0!} = 0.135 \\
e^{-\lambda} &= 0.135 \\
\lambda &= -\log(0.135) \approx 0.87
\end{aligned}$$

Now calculating the percentage of pages with exactly one error:

$$\begin{aligned}
\Pr(X = 1; \lambda) &= e^{-\lambda} \lambda \\
&= -0.135 \log(0.135) \\
&\approx 0.12
\end{aligned}$$

7   Obtaining requests for information

X is a discrete random variable. It takes the value of the number of days required for a governmental agency to respond to a request for information. X is distributed according to the following PMF:

$$f(x) = e^{-4} \frac{4^x}{x!} \text{ for } X \in \{0, 1, 2, ...\}$$

1. Given this information, what is the probability of a response from the agency in 3 days or less?

   Solution. Generating the cmf of this pmf, we have

   $$\Pr(X \leq x) = F_X(x) = e^{-\lambda} \sum_{x=0}^{3} \frac{4^x}{x!} = e^{-4} \left[1 + \frac{4}{1} + \frac{4^2}{2} + \frac{4^3}{6} \right] \approx 0.43$$

2. What is the probability the agency response takes more than 10 but less than 13 days?

   Solution.

   $$F_X(12) - F_X(10) = \sum_{x \in \{11, 12\}} f(x) = e^{-4} \left[\frac{4^{11}}{11!} + \frac{4^{12}}{12!} \right] \approx 0.0026$$

3. What is the probability the agency response takes more than 5 days?

   Solution.
   $$1 - F_X(5) = 1 - e^{-4} \left(1 + \frac{4}{1} + \frac{4^2}{2} + \frac{4^3}{6} + \frac{4^4}{4!} + \frac{4^5}{5!} \right) \approx 0.21$$

4. Suppose using X you generate a new variable, Responsive. Responsive equals 1 if an agency responds in 5 days or less and 0 otherwise. What is the expected value of Responsive?

   Solution. Let p be the probability an agency responds in 5 days or less. We may then specify the random variable X as

   $$X \sim \text{bernoulli}(p) \text{ such that } X = \begin{cases} 0 & 1 - p \\ 1 & p \end{cases}$$

   where $p = F_X(5) \approx 0.79$. Calculating the expected value of the Bernoulli random variable, we have $E[X] = p \approx 0.79$.

5. What is the variance of Responsive?

   Solution. Taking the variance of a Bernoulli random variable, we have $Var[X] = p(1 - p) \approx 0.17$.

8   Modeling electoral outcomes

Suppose we've developed a model predicting the outcome of the upcoming midterm elections in a state with 4 Congressional districts. In each district there are two candidates, a Republican and a Democrat. We have reason to believe the following PMF describes the distribution of potential

election results where $K \in \{0, 1, 2, 3, 4\}$ and is the number of seats won by Republican candidates in the upcoming election.

$$\Pr(K = k|\theta) = \binom{4}{\theta} \theta^k (1 - \theta)^{4-k}$$

Based on polling information, we think the appropriate value for θ is 0.55.7

1. What's the expected number of seats Republicans will win in the upcoming election?

   Solution. Using the expected value of Binomial distributed random variables, $E[K] = N\theta = 4(0.55) = 2.2$.

2. Given this PMF, what's the probability that no Republican legislators win in the upcoming election?

   Solution.
   $$\Pr(K = 0|\theta) = \binom{4}{0.55}(0.45)^4 \approx 0.102.$$

3. What's the probability that Republican legislators win a majority of the seats in this state?

   Solution.

   $$\Pr(\text{Republican majority}) = F_K(4) - F_K(2)$$
   $$= \sum_{k \in \{3,4\}} \binom{4}{\theta} \theta^k (1 - \theta)^{4-k}$$
   $$= \binom{4}{\theta} \theta^3 (1 - \theta) + \binom{4}{\theta} \theta^4$$
   $$\approx 0.41.$$

4. A prominent political pundit declares they are certain that Republicans will win a majority of seats in the next election and offers the following bet. If Republicans win a majority of the seats, we must pay the pundit \$15.00. If Republican's fail to win a majority of states, we will win \$20.00. Based on our model, should we take this bet? Hint: Think of the betting outcomes as a random variable. Find the expected value of this random variable.

   Solution. Let $X = \{-15, 20\}$ be a binary random variable that is -15 if the Republicans win a majority of the seats with probability p, and 20 otherwise. To use p, we use question 3 where $p \approx 0.41$. Computing the expected value of this bet to judge if we should take it, we have

   $$E[X] = \sum_{i \in \Omega} p_i X_i = -15p + 20(1 - p) = 5.52.$$

   Since our expected payout for this bet is \$5.52, we should take this bet if these rules represent all associated costs of participating.

5. Suppose we are offered a second bet with a more complicated structure. In this case we'll receive \$100 if the Republicans win a majority, \$50 if neither party wins a majority and we'll have to pay \$200 if the Democrats win a majority. Should we take this bet?

5

Solution. Similarly, let X be a random variable such that

$$X = \begin{cases} 100 & p \\ 50 & q \\ -200 & 1 - p - q \\ 0 & \text{Otherwise} \end{cases}$$

that is, \$100 if the Republicans win a majority with probability p, \$50 if neither party wins a majority with probability q, and -\$200 if the Democrats win a majority with probability $1-p-q$. Recall $p \approx 0.41$ from the previous question. Finding q, we have

$$q = \Pr(K = 2|\theta) = \binom{4}{\theta}\theta^2(1-\theta)^2 \approx 0.15.$$

Computing the expected value of this bet to judge if we should take it, we have

$$E[X] = \sum_{i \in \Omega} p_i X_i = 100p + 50q - 200(1 - p - q) = -37.75.$$

Since our expected payout for this bet is $-\$37.75$, we should not take this bet.