Discrete random variables

1 Same birthdays

You go to a party with 500 guests. What is the probability that exactly one other guest has the same birthday as you? Calculate this exactly and also approximately by using the Poisson PMF. (For simplicity, exclude birthdays on February 29.)

Solution: The number of guests that have the same birthday as you is binomial with $p = 1/365$ and $n = 499$. Thus the probability that exactly one other guest has the same birthday is:

$$\begin{aligned}\binom{n}{k} p^k (1-p)^{n-k} &= \binom{499}{1} \frac{1}{365} \left(\frac{364}{365}\right)^{498} \\ &\approx 0.3486\end{aligned}$$

Let $\lambda = np = 499/365 \approx 1.367$. The Poisson approximation is $e^{-\lambda}\lambda = e^{-1.367} \times 1.367 \approx 0.3483$, which closely agrees with the correct probability based on the binomial.

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: If X is the temperature in Celsius, the temperature in Farenheit is $Y = 32 + 9X/5$. Therefore,

$$\mathrm{E}[Y] = 32 + \frac{9\mathrm{E}[X]}{5} = 32 + \frac{9 \times 10}{5} = 32 + 18 = 50$$

Also

$$\mathrm{Var}(Y) = \left(\frac{9}{5}\right)^2 \mathrm{Var}(X)$$

where $\mathrm{Var}(X)$, the square of the given standard deviation of X, is equal to 100. Thus, the standard deviation of Y is $(9/5) \times 10 = 18$. Hence a normal day in Farenheit is one for which the temperature is in the range $[32, 68]$.

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.

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

The PMF of X is the binomial PMF with parameters $p = 0.02$ and $n = 250$. The mean is $\mathrm{E}[X] = np = 250 \times 0.02 = 5$. The desired probability is

$$\Pr(X = 5) = \binom{250}{5}(0.02)^5(0.98)^{245} = 0.1773$$

b. Calculate approximately the probability in (a) using a Poisson approximation.

The Poisson approximation has parameter $\lambda = np = 5$, so the probability in (a) is approximated by

$$e^{-\lambda}\frac{\lambda^5}{5!} = 0.1755$$

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: The number of people P needed before you encounter a person whose birthday is today is a geometric random variable with parameter $p = \frac{1}{365}$. Thus, the PMF is

$$\begin{aligned}\Pr(P = k) &= (1-p)^k p, \quad k = 1, 2, \ldots \\ &= \left(1 - \frac{1}{365}\right)^k \times \frac{1}{365}\end{aligned}$$

The expected value is

$$\begin{aligned}\mathrm{E}[P] &= \frac{1-p}{p} \\ &= \frac{1 - 1/365}{1/365} \\ &= 364\end{aligned}$$

The variance is

$$\begin{aligned}\mathrm{Var}(P) &= \frac{1-p}{p^2} \\ &= \frac{1 - 1/365}{(1/365)^2} \\ &= \frac{364/365}{(1/365)^2} \\ &= 132860\end{aligned}$$

5 Properties of variance

For a discrete random variable X, show that:

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$

Solution: A useful property to know here is that $\mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X]$.

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}[(X - \mathrm{E}[X])^2] \\
&= \mathrm{E}[X^2 - (2\mathrm{E}[X] \times X) + \mathrm{E}[X]^2] \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[\mathrm{E}[X] \times X] + \mathrm{E}[\mathrm{E}[X]^2] \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[X]\mathrm{E}[X] + \mathrm{E}[X]^2 \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[X]^2 + \mathrm{E}[X]^2 \\
&= \mathrm{E}[X^2] - \mathrm{E}[X]^2
\end{aligned}$$

6 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:

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

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

$$\Pr(X \leq 3) \approx .43$$

$$\begin{aligned}
&\Pr(X \leq 3) = \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) \\
&\Pr(X = 0) = e^{-4}\frac{4^0}{0!} = 0.01831564 \\
&\Pr(X = 1) = e^{-4}\frac{4^1}{1!} = 0.07326256 \\
&\Pr(X = 2) = e^{-4}\frac{4^2}{2!} = 0.1465251 \\
&\Pr(X = 3) = e^{-4}\frac{4^3}{3!} = 0.1953668 \\
&\Pr(X \leq 3) = 0.01831564 + 0.07326256 + 0.1465251 + 0.1953668 = 0.4334701
\end{aligned}$$

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

$$\Pr(10 < X < 13) \approx .003$$

$$\begin{aligned}
\Pr(10 < X < 13) &= \Pr(X = 11) + \Pr(X = 12) \\
\Pr(X = 11) &= e^{-4}\frac{4^{11}}{11!} = 0.001924537 \\
\Pr(X = 12) &= e^{-4}\frac{4^{12}}{12!} = 0.0006415123 \\
\Pr(10 < X < 13) &= 0.001924537 + 0.0006415123 = 0.002566049
\end{aligned}$$

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

$$\Pr(X > 5) = 1 - \Pr(X \leq 5) \approx 0.22$$

We already know $\Pr(X \leq 3) = .43$ so now we only need to find $\Pr(X = 4)$ and $\Pr(X = 5)$.

$$\begin{aligned}
&\Pr(X \leq 3) = 0.4334701 \\
&\Pr(X = 4) = e^{-4}\frac{4^4}{4!} = 0.1953668 \\
&\Pr(X = 5) = e^{-4}\frac{4^5}{5!} = 0.1562935 \\
&\Pr(X \leq 5) = 0.4334701 + 0.1953668 + 0.1562935 = 0.7851304 \\
&\Pr(X > 5) = 1 - 0.7851304 = 0.2148696
\end{aligned}$$

d. 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**?

Since **Responsive** is dichotomous, E[Responsive] = Pr(Responsive = 1) = $\Pr(X \leq 5)$. From the previous question we know that $\Pr(5 < X) = 0.2148696$ and $\Pr(X \leq 5) = 0.7851304$. So E[Responsive] = 0.7851304.

e. What is the variance of **Responsive**?

Responsive is distributed Bernoulli. Var(Bernoulli) = $p(1-p)$. In this case $p = 0.7851304$. So Var(Responsive) = $(0.7851304 \times (1 - 0.7851304)) = 0.1687007$.

7 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.

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

Based on polling information, we think the appropriate value for θ is 0.55.

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

The election outcomes are distributed according to a binomial distribution so we find the expected value given our two parameters, n and θ.

$$\begin{aligned}
\text{E[Binomial]} &= n \times \theta \\
&= 4 \times .55 \\
&= 2.2
\end{aligned}$$

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

$$\Pr(k = 0) \approx 0.04100625$$

$$
\begin{aligned}
\Pr(k=0) &= \binom{4}{0}.55^0(1-.55)^4 \\
&= 1 \times 1 \times 0.04100625 \\
&= 0.04100625
\end{aligned}
$$

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

$$\Pr(2 < k) \approx .39$$

$$
\begin{aligned}
&\Pr(k=3) = \binom{4}{3}.55^3(1-.55)^1 = 4 \times 0.07486875 = 0.299475 \\
&\Pr(k=4) = \binom{4}{4}.55^4(1-.55)^0 = 1 \times 0.09150625 = 0.09150625 \\
&\Pr(2 < k) = 0.299475 + 0.09150625 = 0.3909813
\end{aligned}
$$

d. 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.**

We can should take this bet as we will net an expected return of \$6.32 from it.

We can think of *bet* as a variable with the associated values:

$$
\text{bet} = \begin{cases} -15 & \text{if Republican majority;} \\ 20 & \text{otherwise.} \end{cases}
$$

$$
\begin{aligned}
\text{E[bet]} &= \text{E[bet|Republican Majority]} + \text{E[bet|Otherwise]} \\
&= -15 \times \Pr(\text{Republican Majority}) + 20 \times \Pr(\text{Otherwise}) \\
&= -15 * (0.3909813) + 20 * (1 - 0.3909813) \\
&= -5.864719 + 12.18037 \\
&\approx 6.3
\end{aligned}
$$

e. 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?

We should take this bet as well since we will net an expected \$9.18.

We can think of bet_2 as a variable that takes the following values:

$$
\text{bet}_2 = \begin{cases} 100 & \text{if Republican majority;} \\ 50 & \text{if tie;} \\ -200 & \text{otherwise.} \end{cases}
$$

$$
\begin{aligned}
\Pr(\text{Rep Majority}) &= 0.3909813 \\
\Pr(\text{Tie}) &= \binom{4}{2}.55^2(1-.55)^2 = 6 \times 0.06125625 = 0.3675375 \\
\Pr(\text{Dem Majority}) &= 1 - \Pr(\text{Rep Majority}) - \Pr(\text{Tie}) \\
&= 1 - 0.3909813 - 0.3675375 \\
&= 0.2414812 \\
\text{E}[\text{bet}_2] &= \text{E}[\text{bet}_2|\text{Rep Majority}] + \text{E}[\text{bet}_2|\text{Tie}] + \text{E}[\text{bet}_2|\text{Otherwise}] \\
&= 100 \times \Pr(\text{Rep Majority}) + 50 \times \Pr(\text{Tie}) \\
&\quad + -200 \times \Pr(\text{Republicans don't win majority}) \\
&= 100 \times 0.3909813 + 50 \times 0.3675375 - 200 \times 0.2414812 \\
&\approx 9.2
\end{aligned}
$$

8 Calculate an exact probability

The random variable X has a probability density function:

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

for $x = 0, 1, 2, \ldots$, (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: With a Poisson variable, the pdf supplies probability of a given number (count) of events, i.e. $\Pr(X = x)$. From the prompt, we know that $\Pr(X = 0) = .135$. Using this, we can solve for λ, which is the average rate of "success" in a given interval, (i.e. the average number of times an event occurs in a given interval). Here, the event is an error occurring and the interval is the page.

$$
\begin{aligned}
f(x;\lambda) &= \frac{\lambda^x \exp(-\lambda)}{x!} \\
f(x=0;\lambda) = \frac{\lambda^0 \exp(-\lambda)}{0!} &= .135 \\
\frac{\exp(-\lambda)}{1} &= .135 \\
\frac{1}{\exp(\lambda)} &= .135 \\
\exp(\lambda) &= \frac{1}{.135} \\
\exp(\lambda) &= 7.407 \\
\lambda &= 2.002
\end{aligned}
$$

Now that we know the value of λ, we know that on average, a page contains 2.002 errors. (Note: If we want to change the interval to say, 10 pages, we just adjust λ accordingly. If there are 2.002 errors per page, there are 2.002*10=20.02 errors per 10 pages.) Now we can solve for $\Pr(X = 1)$.

$$f(x=1;\lambda) = \frac{2.002^1 \exp(-2.002)}{1!} = .27$$