IPDA1005 Introduction to Probability and Data Analysis

Worksheet 7

Solution

1. A pair of dice is rolled until a sum of 8 turns up and let X be the number of rolls and Y be the number of failures.
   (a) Name the distribution of Y.
   (b) Find $P(Y = 5)$.
   (c) Find $P(X = 2)$.
   (d) Find $P(X \leq 5)$.
   (e) Find the expectation and variance of Y.
   (f) Find the expectation and variance of X.

Solution: Note that $X = Y + 1$.

(a) $Y \sim \text{Geometric}\left(\frac{5}{36}\right)$

(b) $P(Y = 5) = \left(\frac{5}{36}\right)\left(\frac{31}{36}\right)^5$

(c) $P(X = 2) = P(Y = 1) = \frac{5}{36}\frac{31}{36} = \frac{155}{1296}$

(d) $P(X \leq 5) = P(Y \leq 4) = 1 - \left(\frac{31}{36}\right)^5 = .5265$

(e) $E(Y) = q/p = \frac{31/36}{5/36} = \frac{31}{5}$
$\text{Var}(Y) = \frac{31}{36}\frac{1296}{25} = \frac{1116}{25}$

(f) $E(X) = E(Y) + 1 = \frac{36}{5}$
$\text{Var}(X) = \text{Var}(Y) = \frac{31}{36}\frac{1296}{25} = \frac{1116}{25}$

2. Each of 12 routers has been returned to a distributor because of an audible, high-pitched noise when the router is operating. Suppose that 7 of these routers have a defective circuit board and the other 5 have less serious problems. If the routers are examined in random order, let X be the number among the first 6 examined that have a defective circuit board. Calculate the following:

(a) $P(X = 5)$

(b) $P(X \leq 4)$

(c) The probability that X exceeds its mean value by more than one standard deviation.

(d) Consider a large shipment of 400 routers, of which 40 have defective circuit boards. If X is the number among 15 randomly selected routers that have defective circuit boards, write out (but don't calculate) $P(X \leq 5)$.

(e) It turns out that under certain conditions, we can use the binomial distribution to calculate approximate probabilities from the hypergeometric distribution. In particular: Let the population size, N, and number of population successes, M, get large with the ratio M/N approaching p. Then $h(x; n, M, N)$ approaches the binomial PMF $b(x; n, p)$; so for n/N small, the two are approximately equal provided that p is not too near either 0 or 1. Use this result to calculate the probability in part (d).

Solution:

From the problem description, it's clear that the random variable X has a hypergeometric distribution with $N = 12$, $M = 7$, and $n = 6$. Thus,

$$h(x; n, M, N) = \frac{\binom{7}{x}\binom{12-7}{6-x}}{\binom{12}{6}}$$

and

$$E(X) = n \cdot \frac{M}{N} = 3.5 \qquad \text{and} \qquad \text{Var}(X) = n\left(\frac{M}{N}\right)\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right) = 0.795$$

(a) $P(X = 5) = \frac{\binom{7}{5}\binom{12-7}{6-5}}{\binom{12}{6}} = 0.114$

(b) $P(X \leq 4) = 1 - P(X > 4) = 1 - [P(X = 5) + P(X = 6)]$. We can calculate $P(X = 6)$ in the same way as we calculated $P(X = 5)$ in part (a), and so

$$P(X \leq 4) = 1 - \left[\frac{\binom{7}{5}\binom{12-7}{6-5}}{\binom{12}{6}} + \frac{\binom{7}{6}\binom{12-7}{6-6}}{\binom{12}{6}}\right] = 1 - (0.114 + 0.008) = 0.879$$

(c) What we're after here is $P(X > \mu + \sigma)$. From the calculation of $E(X)$ and $\text{Var}(X)$ above, $\mu = E(X) = 3.5$, and $\sigma = \sqrt{\text{Var}(X} = 0.892$, so that

$$P(X > \mu + \sigma) = P(X > 3.5 + 0.892) = P(X > 4.392) = P(X = 5 \text{ or } 6) = 0.121$$

from Part (b).

(d) In this instance,

$$h(x;15,40,400) = \frac{\binom{40}{x}\binom{400-40}{15-x}}{\binom{400}{15}}$$

and hence

$$P(X \leq 5) = \sum_{x=0}^{5} h(x;15,40,400)$$

(e) Here, $M/N = 0.1$, so $h(x;15,40,400) \approx b(x;15,0.1)$. Using this approximation, $P(X \leq 5) \approx B(5;15,0.1) = 0.998$ using the tables you were provided during the workshop. The exact value calculated using the hypergeometric distribution is also 0.998.

3. A geologist has collected 10 specimens of basaltic rock and 10 specimens of granite. The geologist instructs a laboratory assistant to randomly select 15 of the specimens for analysis.

   (a) What is the PMF of the number of granite specimens selected for analysis?

   (b) What is the probability that all specimens of one of the two types of rock are selected for analysis?

   (c) What is the probability that the number of granite specimens selected for analysis is within 1 standard deviation of its mean value?

Solution: If X represents the number of granite specimens in the sample of 15, then possible values of X are $5, 6, 7, 8, 9, 10$ because if we had fewer than five samples, we'd need more than ten samples of basaltic rock, which isn't possible.

(a) X is hypergeometric with PMF given by

$$p(x) = h(x;15,10,20) = \frac{\binom{10}{x}\binom{10}{15-x}}{\binom{20}{15}}, \qquad x = 5, 6, 7, 8, 9, 10$$

x	5	6	7	8	9	10
$p(x)$	0.0163	0.1354	0.3483	0.3483	0.1354	0.0163

(b) If all 10 specimens of basaltic rock are selected, then $X = 5$, but if all 10 specimens of granite are selected, $X = 10$. Therefore,

$$P(X = 5 \text{ or } X = 10) = P(X = 5) + P(X = 10) = 0.0163 + 0.0163 = 0.0326$$

(c) We require $P(\mu - \sigma < X < \mu + \sigma)$. Using the expressions for the mean and variance of a hypergeometric distribution shown above, $\mu = 7.5$ and $\sigma = 0.9934$, and so

$$P(7.5 - 0.9934 < X < 7.5 + 0.9934) = P(6.5066 < X < 8.4934)$$

This quantity equals $P(X = 7) + P(X = 8) = 0.3483 + 0.3483 = 0.6966$.

4. According to climatologists, the probability of observing daily rainfall exceeding some very high threshold (mm rain) during a Perth winter is about $p = 0.005$.

   (a) Let X be the number of winters until the next exceedance. So, $X = 1$ if it happens next winter, $X = 2$ if it doesn't happen until the second winter from now, etc. What is the PMF of X?

   (b) On average, how long do we have to wait to observe a daily rainfall as extreme as this threshold? (Hint: recall the relationship between the negative binomial distribution and the geometric distribution.)

Solution:

(a) If an exceedance occurs in the X^{th} winter with probability p, then the previous $X - 1$ seasons have not experienced extreme rainfall, with probability $1 - p$ in each year. Thus,

$$P(X = x) = (1 - p)^{x-1} p = (0.995)^{x-1}(0.005), \qquad x = 1, 2, 3, \ldots.$$

(b) Intuitively, we would expect an event that occurs with probability p in any given year to occur, on average, once every $1/p$ years. Also, we have learned the expected number of trials required to obtain a success is $1/p = 1/.005 = 200$.

5. In order to estimate the number of fish in a lake, 300 fish were caught, tagged and released back into the lake. After allowing sufficient time for the fish to mix, another sample of 50 fish were caught. It was found that 6 of the second sample were tagged. Use this information to estimate the number of fish in the lake.

Solution: If N represents the number of fish in the lake, $\frac{300}{N}$, the population proportion of tagged fish, is estimated by the sample proportion $\frac{6}{50}$. Equating these two, weget the estimate of N to be $\hat{N} = \frac{(300)(50)}{6} = 2500$.

6. Let X be the number of material anomalies occurring in a particular region of an aircraft gas-turbine disk. The article "Methodology for Probabilistic Life Prediction of Multiple-Anomaly Materials" (*Amer. Inst. of Aeronautics and Astronautics J.*, 2006: 787–793) proposes a Poisson distribution for X. Assume that $\lambda = 4$.
   (a) Calculate both $P(X \leq 4)$ and $P(X < 4)$.
   (b) What is $P(4 \leq X \leq 8)$?
   (c) What is the probability that the observed number of anomalies exceeds the expected number by no more than one standard deviation?

Solution: In this case, the Poisson PMF is $p(x; \lambda) = p(x; 4) = \frac{e^{-4}4^x}{x!}$

(a) Using the tables, we can obtain $P(x \leq 4) = 0.6288$ and $P(X < 4) = P(X \leq 3) = 0.4335$.

(b) Recall that $P(a \leq X \leq b) = F(b) - F(a-1)$, where $F(x)$ is the CDF of the random variable X. Thus, $P(4 \leq X \leq 8) = F(8) - F(3) = 0.9786 - 0.4335 = 0.5451$

(c) For a Poisson distribution, $\mu = E(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$ and hence $\sigma = \lambda^{1/2}$. Thus, for this problem, $\mu = 4$ and $\sigma = 2$, and we require $P(4 < X \leq 6) = P(X = 5) + P(X = 6) = 0.2605$.

7. Let X have a Poisson distribution with parameter λ. Show that $E(X) = \lambda$ directly from the definition of expected value. [Hint: The first term in the sum equals 0, and then x can be cancelled. Now factor out λ and show that what is left sums to 1.]

Solution: We can write $E(X)$ as

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!} \\
&= 0 + \sum_{x=1}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x(x-1)!} \\
&= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \\
&= \lambda \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} \\
&= \lambda
\end{aligned}$$

where the last summation is equal to one because it represents the sum of a Poisson PMF over all its values.

8. In some applications the distribution of a discrete random variable X resembles the Poisson distribution except that zero is not a possible value of X. For example, let X be the number of tattoos that an individual wants removed when s/he arrives at a tattoo removal facility. Suppose the PMF of X is

$$p(x;\theta) = k\frac{e^{-\theta}\theta^x}{x!}, \quad x = 1, 2, 3, \ldots$$

   (a) Find a simple expression for k. [Hint: The sum of all probabilities in the Poisson PMF is equal to 1, and the PMF above must also sum to 1.]
   (b) If the mean value of X is 2.313035, find the value of θ? [Hint: you will need some trial and error.]
   (c) Calculate the value of $P(X \leq 5)$.

   [Adapted from "An Exploratory Investigation of Identity Negotiation and Tattoo Removal" (*Academy of Marketing Science Review*, Vol. 12, No. 6, 2008).]

Solution:

(a) A PMF has to sum to one over all its values, so we need to find k such that $\sum_{x=1}^{\infty} k\frac{e^{-\theta}\theta^x}{x!} = 1$. First, consider a Poisson PMF with parameter θ, which we *know* has to sum to one over all its values:

$$\begin{aligned} 1 &= \sum_{x=0}^{\infty} \frac{e^{-\theta}\theta^x}{x!} \\ &= e^{-\theta}\frac{\theta^0}{0!} + \sum_{x=1}^{\infty} \frac{e^{-\theta}\theta^x}{x!} \\ &= e^{-\theta} + \sum_{x=1}^{\infty} \frac{e^{-\theta}\theta^x}{x!} \Longrightarrow \sum_{x=1}^{\infty} \frac{e^{-\theta}\theta^x}{x!} = 1 - e^{-\theta} \end{aligned}$$

Thus,

$$\sum_{x=1}^{\infty} k\frac{e^{-\theta}\theta^x}{x!} = k\sum_{x=1}^{\infty} \frac{e^{-\theta}\theta^x}{x!} = k(1 - e^{-\theta}) \Longrightarrow k = \frac{1}{1 - e^{-\theta}}$$

(b) Again, we begin a Poisson distribution with $\mu = \theta$. We know that

$$\begin{aligned} \theta &= \sum_{x=0}^{\infty} x\frac{e^{-\theta}\theta^x}{x!} \\ &= 0 + \sum_{x=1}^{\infty} x\frac{e^{-\theta}\theta^x}{x!} \end{aligned}$$

Multiplying both sides by k yields

$$k\theta = \sum_{x=1}^{\infty} x \cdot k\frac{e^{-\theta}\theta^x}{x!}$$

The right hand side is the mean of the specified distribution, and hence $E(X) = k\theta = \frac{\theta}{1-e^{-\theta}}$, so we need to solve $\frac{\theta}{1-e^{-\theta}} = 2.313035$. This is analytically intractable, but by trying values $\theta = 0, 1, \ldots$ we find that $\theta = 2$, i.e., $\frac{2}{1-e^{-2}} = 2.313035$.

(c) We can now calculate $P(X \leq 5)$ as

$$P(X \leq 5) = \sum_{x=1}^{5} \frac{e^{-2}}{1-e^{-2}} \frac{2^x}{x!} = 0.9808$$

9. An article in the *Los Angeles Times* (Dec. 3, 1993) reports that 1 in 200 people carry the defective gene that causes inherited colon cancer. In a sample of 1000 individuals, what is the **exact** distribution of the number who carry this gene? What is the **approximate** distribution? Use this approximate distribution to calculate the approximate probability that
   (a) Betwen 5 and 8 (inclusive) people carry the gene.
   (b) At least 8 people carry the gene.

Solution: The exact distribution of X is binomial with $n = 1000$ and $p = 1/200$. We will use the Poisson approximation with $\lambda = np = 5$ because n is large and p is small, and we can use the tables to calculate the required probabilities.

(a) $P(5 \leq X \leq 8) = F(8) - F(5-1) = 0.9319 - 0.4405 = 0.4914$.

(b) $P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.8666 = 0.1334$

Note: The software `R` gives the exact probabilities to be $P(5 \leq X \leq 8) = 0.4923436$ and $P(X \geq 8) = 0.1328476$.

(*Sources: All questions adapted from Devore and Berk (2012) and Carlton and Devore (2017), except for Question 3*)

Bibliography

1. Carlton, M.A. and Devore, J.L. (2017) *Probability with Applications in Engineering, Science, and Technology*, 2nd ed. Springer: New York.

2. Devore, J.L. and Berk, K.N. (2012) *Modern Mathematical Statistics with Applications*. Springer: New York.