Stat 113 (Introduction to Mathematical Statistics)

Lesson 2.5: The Poisson Distribution

Learning Outcomes

At the end of the lesson, students must be able to

1. Describe the properties of a Poisson process,
2. Derive the probability distribution of a random variable having a Poisson distribution,
3. Compute probabilities associated with a random variable with a Poisson distribution, and
4. Compute the mean and variance of a random variable with a Poisson distribution.

Introduction

The binomial distribution problems that we solved in Lesson 2.3 all had relatively small values for n, so evaluating $P(X = x) = \binom{n}{x} p^x q^{n-x}$ was not particularly difficult. But suppose $n = 1000$ and $x = 500$. Evaluating $P(X = 500)$ would be a formidable task for many scientific calculators, even today. Two hundred years ago, the prospect of doing cumbersome binomial calculations by hand was a catalyst for mathematicians to develop some easy-to-use approximations.

One of the first such approximations was the Poisson limit, which eventually gave rise to the Poisson distribution, in honor to *Simeon Denis Poisson* (1781–1840). It can be shown that as $p \to 0$

$$\begin{aligned} \lim_{x\to\infty} P(X = x) &= \lim_{x\to\infty} \binom{n}{x} p^x q^{n-x} \\ &= \frac{e^{-np}(np)^x}{x!} \end{aligned}$$

Definition:

A random variable X is said to have a Poisson distribution with parameter $\lambda = np$ if its probability mass function is given by

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

where $e \approx 2.71828$.

Remarks:

1. A Poisson-distributed random variable counts the number of occurrences of an event in a given unit of time, space, area, or distance, such as the number of earthquakes per hour, number of car accidents per day, or the number of typing errors per page of a document.

2. A Poisson-distributed random variable follows the Poisson process. A Poisson process has the following properties:

   a. the number of occurrences in non-overlapping intervals are independent random variables,
   b. the probability of an occurrence in a sufficiently short interval is proportional to the length of the interval, and
   c. the probability of 2 or more occurrences in a sufficiently short interval is zero.

Example 1:

Let X be the number of automobile accidents at a busy intersection per week. Suppose that $X \sim Pois(2)$. Calculate

a. the probability of exactly 3 accidents in a week,

b. the probability of at least one accident in a week,

c. the probability that there are at most 3 accidents in a week, and

d. the probability that there are exactly 16 accidents in a 4 week period.

SOLUTION:

a. The probability of exactly 3 accidents in a week is

$$P(X = 3) = \frac{e^{-2}2^3}{3!} \approx 0.1804$$

The above probability can also be generated using the R command **dpois(3,2)**.

b. The probability of at least one accident in a week is

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \frac{e^{-2}2^0}{0!} \\ &\approx 0.8647 \end{aligned}$$

The above probability can be generated using either **1- dpois(0, 2)** or **1- ppois(0, 2, lower.tail = TRUE)**.

c. The probability that there are at most 3 accidents in a week is given by

$$\begin{aligned} P(X \leq 3) &= P(Y = 3) + P(Y = 2) + P(Y = 1) + P(Y = 0) \\ &\approx 0.1804 + 0.2707 + 0.2707 + 0.1353 \\ &= 0.8571 \end{aligned}$$

The above answer can also be generated using the R command **ppois(3, 2, lower.tail = TRUE)**

d. On average, we expect 2 accidents per week, so in four weeks we expect $\lambda' = 2 \times 4 = 8$ accidents. Thus, the probability that there are exactly 16 accidents in a 4 week period is

$$\begin{aligned} P(X = 16) &= \frac{e^{-8}8^{16}}{16!} \\ &\approx 0.0045 \end{aligned}$$

Theorem:

If $X \sim Pois(\lambda)$, then

a. $m_X(t) = e^{\lambda(e^t - 1)}$.

b. $E(X) = V(X) = \lambda$.

Example 2:

The number of typing errors made by a typist has a Poisson distribution with an average of four errors per page. If more than four errors appear on a given page, the typist must retype the whole page. What is the probability that a randomly selected page does not need to be retyped?

SOLUTION:

Let Y be the number of typing errors per page. Then $Y \sim Pois(4)$. The probability that a randomly selected page does not need retyping is

$$\begin{aligned} P(\text{do not retype the page}) &= P(Y \leq 4) \\ &= P(Y = 4) + P(Y = 3) + P(Y = 2) + P(Y = 1) + P(Y = 0) \\ &\approx 0.1954 + 0.1954 + 0.1465 + 0.0733 + 0.0183 \\ &= 0.6289 \end{aligned}$$

The above probability can be calculated using the R command **ppois(4, 4, lower.tail = TRUE)**.

Example 3:

Customers arrive at a checkout counter in a department store according to a Poisson distribution at an average of seven per hour. During a given hour, what are the probabilities that

a) no more than three customers arrive?
b) at least two customers arrive?
c) exactly five customers arrive?