STAT241/251 Lecture Notes
Chapter 6 Part 5

Yew-Wei Lim

Introduction to Poisson Probability distribution

When n is large ($n \geq 20$) and p is small (such that $np < 5$),
We can use Poisson distribution to approximate the Binomial distribution (Section 6.5, page 113).

What is a Poisson Probability distribution?

A random variable X is said to have a Poisson probability distribution when its pmf is:

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

We write $X \sim Poisson(\lambda)$.

Properties:

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

The Poisson probability distribution often provides a good model for the probability distribution of the number X of rare events that occur in space, time, volume or any other dimension, where λ is the average value of X. Examples of rare events are industrial accidents, automobile accidents, flaws in wood panel, etc.

Simple Example

The number of knots in a particular type of wood has a Poisson distribution with an average of 1.5 knots in 10 cubic feet of the wood. Find the probability that a 10-cubic-foot block of the wood has at most one knot.

Let X be the number of knots in a 10-cubic-foot block of wood.

$$X \sim Poisson(1.5) \qquad \text{[in 10 cubic feet of wood]}$$

$$\begin{aligned} P(X \leq 1) =& P(X = 0) + P(X = 1) \\ =& \frac{e^{-1.5}(1.5)^0}{0!} + \frac{e^{-1.5}(1.5)^1}{1!} \\ =& 0.22313 + 0.334695 \\ \approx& 0.5578 \end{aligned}$$

Ch 6.5 Poisson approximation to the Binomial

When n is large ($n \geq 20$) and p is small (such that $np < 5$), we can use the Poisson distribution to approximate the Binomial distribution.

Example (from notes, pg 113):

On average, one per cent of 50-kg dry concrete bags are underfilled below 49.5kg. what is the probability of finding 4 or more of these underfilled bags in a lot of 200?

Solution:
Let X be the number of underfilled bags out of 200.
Then $X \sim Binomial(200, 0.01)$.
Exact answer:

$$\begin{aligned} P(X \geq 4) =& 1 - P(X \leq 3) \\ =& 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)] \\ =& 1 - [(0.99)^{200} + {}^{200}C_1(0.01)^1(0.99)^{199} + {}^{200}C_2(0.01)^2(0.99)^{198} \\ &+ {}^{200}C_3(0.01)^3(0.99)^{197}] \\ =& 0.142 \end{aligned}$$

Why not use Normal approximation to Binomial?
We cannot use Normal approximation to Binomial because $np \geq 5, nq \geq 5$ conditions are not satisfied (in this case $np = 2$ so not able to fulfill $np \geq 5$ requirement)

Poisson approximation can be used because n is large ($n = 200$) and p is small ($np = 2 < 5$).
We then use $X \overset{approx.}{\sim} Poisson(2)$
use rate $\lambda = np$ to approximate probabilistic behavior of X. The reason for choice of np is because $E(X)$ of Binomial distribution is np.

$$\begin{aligned} P(X \geq 4) =& 1 - P(X \leq 3) \\ =& 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)] \\ =& 1 - [\frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!} + \frac{e^{-2}2^3}{3!}] \\ =& 0.143 \quad \text{(very close to exact answer)} \end{aligned}$$

Ch 6.4 Poisson Process

Characteristics of a Poisson process:

- Events occur singly at random
- Whether an event occurs in a particular point in time or space is independent of what happens elsewhere
- Events have a low probability of occurrence at any given instant
- Events occur uniformly, that is, the mean number of occurrences of events in any interval (time or space) is constant.
- The average number of events per interval is proportional to the size of the interval.

If the above conditions are satisfied, the random variable X representing the number of occurrences in a given interval of time/space is said to have a Poisson distribution.

We write:

$$X \sim Poisson(\underbrace{\lambda}_{rate})$$

Pmf is :

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, 3, \ldots$$

Examples of Poisson random variables:

- Number of car accidents in a particular stretch of road in 5 hours
- Number of typing errors in a given page
- Number of customers arriving at a counter in a given time
- Number of flaws in a given length of copper wire
- Number of earthquakes in a decade

Example

Suppose, on average, 1 earthquake greater than 8.0 on the Richter scale occurs worldwide per year.

(a) What is the probability there are no earthquake greater than 8.0 in the next year? In the next three years?

Ans:

Let X be the number of occurrences of earthquake (greater than 8.0 on Richter scale) in 1 year.

$$X \sim Poisson(\lambda = 1)$$

time unit is 1 year

$$\lambda = 1 \quad \text{rate is 1 per year}$$

$$P(X = 0) = \frac{e^{-1}1^0}{0!} = e^{-1} = 0.368$$

To answer the question on probability of no earthquakes in 3 years, we need a new rate.

Let Y be the number of occurrences of earthquakes (greater than 8.0 on Richter scale) in 3 years.

$$Y \sim Poisson(\lambda = 1 \times 3)$$

$$\text{ie., } Y \sim Poisson(3)$$

$$P(Y = 0) = \frac{e^{-3}3^0}{0!} = e^{-3} = 0.0497$$

Apart from the number of occurrences, we may be interested in the time/distance/space between consecutive occurrences of the event of interest.
Apart from calculating probabilities involving counts of number of occurences, students must be comfortable calculating probabilities involving time (or distance) between consecutive occurrences of the event of interest.

If it is time between consecutive occurrences that is of interest to us, we call that waiting time. Let us represent the random variable waiting time as T. Note that waiting time T is a continuous random variable.
What is the relationship between count of number of occurrences $Y \sim Poisson(\lambda)$ and waiting time T which is continuous random variable?

An example will help. If Y is the count of the number of occurences of earthquakes in one unit of time (say year) and say earthquakes occur at a rate of λ per year, then $Y \sim Poisson(\lambda)$ and T, the waiting time in years between occurrences of earthquake is **an exponential distribution with a mean of** $\frac{1}{\lambda}$. [Proof given in a separate document].

$$T \sim Exponential(Mean = \underbrace{\frac{1}{\lambda}}_{\text{the } \lambda \text{ comes from } Y \sim Poisson(\lambda)})$$

$$\Rightarrow f_T(t) = \lambda e^{-\lambda t}, t > 0$$

$$\Rightarrow F_T(t) = 1 - e^{-\lambda t}$$ (see page 114 for derivations of these results)

Q: What is the probability you wait less than 2 years until occurrence of an earthquake over 8.0?

Recall $\lambda = 1$ per year (from earlier Poisson question)

$\Rightarrow$

$$T \sim Exponential(Mean = \frac{1}{1} = 1)$$

$$f_T(t) = e^{-t}, \quad t > 0$$

$$F_T(t) = 1 - e^{-t}$$

You wish to find $P(T < 2)$ ($P(T < 2)$ is the same as $P(T \leq 2)$ since T is continuous)

$$\begin{aligned} P(T \leq 2) =& F_T(2) \\ =& 1 - e^{-2} \\ =& 0.865 \end{aligned}$$

Note: Another solution possible if you convert time question to one involving occurrence.

$$\begin{aligned} & P(\text{ waiting time less than 2 years for next occurrence }) \\ =& 1 - P(\text{ no occurrence in next 2 years}) \\ & [X \sim Poisson(\lambda = 2)] \\ =& 1 - (\frac{e^{-2}2^0}{0!}) \\ =& 1 - e^{-2} = 0.865 \end{aligned}$$