Contents

0.1 Poisson probability distribution

A discrete random variable that is often used is one which estimates the number of occurrences over a specified time period or space.

(remark : a specified space can be a specified length , a specified area, or a specified volume.)

If the following two properties are satisfied, the number of occurrences is a random variable described by the Poisson probability distribution

Properties

(1) The probability of an occurrence is the same for any two intervals of equal length.

(2) The occurrence or non-occurrence in any interval is independent of the occurrence or non-occurrence in any other interval.

1 Poisson Formulae

Here the Poisson Mean is the expected number of occurence per unit period. It is usually denoted as either m or λ

$$P(X = k) = \frac{m^{`} \times e^{-m}\lambda}{k!}$$

Given that there is on average 2 occurrences per hour, what is the probability of no occurrences in the next hour?
i.e. Compute $P(X = 0)$ given that $m = 2$

$$P(X = 0) = \frac{2^0 e^{-2}}{0!}$$

- $2^0 = 1$
- $0! = 1$

The equation reduces to

$$P(X = 0) = e^{-2} = 0.1353$$

The Poisson Distribution

(Part E - Probability Distributions)

- The Poisson Distribution is a statistical distribution showing the frequency probability of specific events when the average probability of a single occurrence is known.
- The Poisson distribution is a discrete probability distribution.
- Application of the Poisson distribution enables managers to introduce optimal scheduling systems.
- For example, if the average number of people that rent movies on a friday night at a single video store location is 400, a Poisson distribution can answer such questions as, "What is the probability that more than 600 people will rent movies?".
- One of the most famous historical practical uses of the Poisson distribution was estimating the annual number of Prussian cavalry soldiers killed due to horse-kicks.
- Other modern examples include estimating the number of car crashes in a city of a given size; in physiology, this distribution is often used to calculate the probabilistic frequencies of different types of neurotransmitter secretions.

Question:
Suppose that random variable X follows a Poisson distribution with rate parameter `L`.
If we increase the value of `L`, which of the following is true?

Options:

1. The spread increases but the center remains unchanged.
2. Both the spread and the center increase.
3. The center increases but the spread decreases.
4. The spread increases but the center decreases.

Comments:

- center - i.e. the measures of centrality, such as mean and median.
- spread - i.e. measures of disperion, such as variance and range.

Exercise

- Generate 100 random numbers from the Poisson distribution - specifying a value for `lambda` (i.e. what the rate parameter is called when using `R`) .
- Compute the mean and variance for this set of numbers.
- Repeat the process a few times, each time increasing the value of lambda.

```
#generate three data sets
X1 <- rpois(100, lambda= 4)
X2 <- rpois(100, lambda= 8)
X3 <- rpois(100, lambda= 18)
```

```
#Now get the mean and variance for each data set
mean(X1);var(X1)
mean(X2);var(X2)
mean(X3);var(X3)
```

2 Notation for Poisson Distribution

A discrete random variable X is said to follow a Poisson distribution with parameter m, written $X \sim \text{Po}(m)$, if it has probability distribution

$$P(X = k) = e^{-m}\frac{m^k}{k!}$$

where

- $k = 0, 1, 2, \ldots$
- $m > 0$.

The Poisson probability function is given by

- f(x) the probability of x occurrences in an interval.
- λ is the expected value of the mean number of occurrences in any interval. (We often call this the Poisson mean)
- e=2.71828284

2.1 Binomial Distribution: Worked Example

Suppose a gambler is playing a simple coin flip game. The gambler does not know that the coin has been tampered with such that the probability of a Head is 47%.

Suppose the gamble plays this coin flip game nine times. What is the probability that he wins precisely 3 times.

3 Characteristics of a Poisson Experiment

A Poisson experiment is a statistical experiment that has the following properties:

- The experiment results in outcomes that can be classified as successes or failures.
- The average number of successes (m) that occurs in a specified region is known.
- The probability that a success will occur is proportional to the size of the ***region***.
- The probability that a success will occur in an extremely small region is virtually zero.
- The `pois` family of functions are used to compute probabilities and quantiles.

Note that the specified region could take many forms. For instance, it could be a length, an area, a volume, a period of time, etc.

3.1 The Poisson Random Variable

- A Poisson random variable is the number of successes that result from a Poisson experiment.
- The probability distribution of a Poisson random variable is called a Poisson distribution.
- This distribution describes the number of occurrences in a unit period (or space)
- The expected number of occurrences is m.
- `R` refers to the mean number of occurrences as `lambda` rather than `m`.

The probability that there will be k occurrences in a unit time period is denoted $P(X = k)$, and is computed as below. Remark: This is known as the probability density function. The corresponding `R` command is `dpois()`.

$$P(X = k) = \frac{m^k e^{-m}}{k!}$$

3.2 Example

What is the probability of one occurrences in the next hour?
i.e. Compute $P(X = 1)$ given that $m = 2$

$$P(X = 1) = \frac{2^1 e^{-2}}{1!}$$

- $2^1 = 2$
- $1! = 1$

The equation reduces to

$$P(X = 1) = 2 \times e^{-2} = 0.2706$$

3.3 Poisson probability distribution

A discrete random variable that is often used is one which estimates the number of occurrences over a specified time period or space.

(remark : a specified space can be a specified length , a specified area, or a specified volume.)

If the following two properties are satisfied, the number of occurrences is a random variable described by the Poisson probability distribution

Properties

1. The probability of an occurrence is the same for any two intervals of equal length.

2. The occurrence or non-occurrence in any interval is independent of the occurrence or non-occurrence in any other interval.

The Poisson probability function is given by

- f(x) the probability of x occurrences in an interval.
- λ is the expected value of the mean number of occurrences in any interval. (We often call this the Poisson mean)
- e=2.71828284

3.4 Poisson Approximation of the Binomial Probability Distribution

The Poisson distribution can be used as an approximation of the binomial probability distribution when p, the probability of success is small and n, the number of trials is large. We set (other notation) and use the Poisson tables.

As a rule of thumb, the approximation will be good wherever both and

4 Poisson Distribution (Example)

- Suppose that electricity power failures occur according to a Poisson distribution with an average of 2 outages every twenty weeks.
- Calculate the probability that there will not be more than one power outage during a particular week.

Solution:

- The average number of failures per week is: $m = 2/20 = 0.10$
- "Not more than one power outage" means we need to compute and add the probabilities for "0 outages" plus "1 outage".

Recall:

$$P(X = k) = e^{-m}\frac{m^k}{k!}$$

- $P(X = 0)$

$$P(X = 0) = e^{-0.10}\frac{0.10^0}{0!} = e^{-0.10} = 0.9048$$

- $P(X = 1)$

$$P(X = 1) = e^{-0.10}\frac{0.10^1}{1!} = e^{-0.10} \times 0.1 = 0.0905$$

- $P(X \leq 1)$

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.9048 + 0.0905 = 0.995$$

- Probability Density Function $P(X = k)$
  - For a given poisson mean m, which in `R` is specified as `lambda`
  - `dpois(k,lambda = ...)`
- Cumulative Density Function $P(X \leq k)$
  - `ppois(k,lambda = ...)`

From before: $P(X = 0)$ given than the mean number of occurrences is 2.

```
> dpois(0,lambda=2)
[1] 0.1353353
> dpois(1,lambda=2)
[1] 0.2706706
> dpois(2,lambda=2)
[1] 0.2706706
```

Compute the cumulative distribution functions for the values $k = \{0, 1, 2\}$, given that the mean number of occurrences is 2

```
> ppois(0,lambda=2)
[1] 0.1353353
> ppois(1,lambda=2)
[1] 0.4060058
> ppois(2,lambda=2)
[1] 0.6766764
```

- The Poisson distribution can sometimes be used to approximate the binomial distribution
- When the number of observations n is large, and the success probability p is small, the $\text{Bin}(n, p)$ distribution approaches the Poisson distribution with the parameter given by $m = np$.
- This is useful since the computations involved in calculating binomial probabilities are greatly reduced.
- As a rule of thumb, n should be greater than 50 with p very small, such that np should be less than 5.
- If the value of p is very high, the definition of what constitutes a “success” or “failure” can be switched.

4.1 Poisson Approximation: Example

Suppose we sample 1000 items from a production line that is producing, on average, 0.1% defective components.

Using the binomial distribution, the probability of exactly 3 defective items in our sample is

$$P(X=3) = {}^{1000}C_3 \times (0.001)^3 \times 0.999^{997}$$

Lets compute each of the component terms individually.

- ${}^{1000}C_3$

$${}^{1000}C_3 = \frac{1000 \times 999 \times 998}{3 \times 2 \times 1} = 166,167,000$$

- 0.001^3

$$0.001^3 = 0.000000001$$

- 0.999^{997}

$$0.999^{997} = 0.36880$$

Multiply these three values to compute the binomial probability

$$P(X=3) = 0.06128$$

- Lets use the Poisson distribution to approximate a solution.
- First check that $n \geq 50$ and $np < 5$ (Yes to both).
- We choose as our parameter value $m = np = 0.001 \times 1000 = 1$

$$P(X=3) = e^{-1}\frac{1^3}{3!} = \frac{e^{-1}}{6} = \frac{0.36787}{6} = 0.06131$$

- Compare this answer with the Binomial probability $P(X=3) = 0.06128$.
- Very good approximation, with much less computation effort.

```
> # Poisson Mean m = 1000 * 0.001 = 1
> dbinom(3,size=1000,prob=0.001)
[1] 0.06128251
>
> dpois(3,lambda=1)
[1] 0.06131324
```

- Probability Density Function
- Cumulative Density Function

If X is a continuous random variable then we can say that the probability of obtaining a **precise** value x is infinitely small, i.e. close to zero.

$$P(X = x) \approx 0$$

Consequently, for continuous random variables (only), $P(X \leq x)$ and $P(X < x)$ can be used interchangeably.

$$P(X \leq x) \approx P(X < x)$$

A random variable X is called a continuous uniform random variable over the interval (a, b) if it's probability density function is given by

$$f_X(x) = \frac{1}{b-a} \qquad \text{when } a \leq x \leq b$$

The corresponding cumulative density function is

$$F_x(x) = \frac{x-a}{b-a} \qquad \text{when } a \leq x \leq b$$

The mean of the continuous uniform distribution is

$$E(X) = \frac{a+b}{2}$$

$$V(X) = \frac{(b-a)^2}{12}$$

The most interesting property of the exponential distribution is the ***memoryless*** property. By this , we mean that if the lifetime of a component is exponentially distributed, then an item which has been in use for some time is a good as a brand new item with regards to the likelihood of failure.

The exponential distribution is the only distribution that has this property.

A pair of dice is thrown. Let X denote the minimum of the two numbers which occur. Find the distributions and expected value of X.

A fair coin is tossed four times. Let X denote the longest string of heads. Find the distribution and expectation of X.

A fair coin is tossed until a head or five tails occurs. Find the expected number E of tosses of the coin.

The coin is tossed three times. Let X denote the number of heads that appear.

- (a) Find the distribution f of X.
- (b) Find the expectation E(X).

- Now consider an experiment with only two outcomes. Independent repeated trials of such an experiment are called Bernoulli trials, named after the Swiss mathematician Jacob Bernoulli (16541705).
- The term ***independent trials*** means that the outcome of any trial does not depend on the previous outcomes (such as tossing a coin).
- We will call one of the outcomes the "success" and the other outcome the "failure".
- Let p denote the probability of success in a Bernoulli trial, and so $q = 1-p$ is the probability of failure. A binomial experiment consists of a fixed number of Bernoulli trials.
- A binomial experiment with n trials and probability p of success will be denoted by

$$B(n, p)$$

- a probability mass function (pmf) is a function that gives the probability that a discrete random variable is exactly equal to some value.
- The probability mass function is often the primary means of defining a discrete probability distribution

4.2 Poisson

M=15 (1/2 hour or 30 minutes)
5 minute period m=2.5
X : No of arrivals
P(X=0) when M = 2.5

$$\begin{aligned} P(X=0) &= 1 - P(X \geq 1)(Complement) \\ &= 1 - 0.9179 \\ &= 0.0821 \end{aligned}$$

4.3 Example

Thirty-eight students took the test. The X-axis shows various intervals of scores (the interval labeled 35 includes any score from 32.5 to 37.5). The Y-axis shows the number of students scoring in the interval or below the interval.

cumulative frequency distributionA can show either the actual frequencies at or below each interval (as shown here) or the percentage of the scores at or below each interval. The plot can be a histogram as shown here or a polygon.

5 Poisson Distribution

A researcher takes a random sample of 500 urban residents and finds that 122 have fibre-optic broadband access. Calculate a 90the true percentage of residents who have fibre-optic broadband access.

The following table gives the results of operations in a hospital according to the complexity of the operation.

Let A be the event that an operation is simple and B be the event that an operation is successful. Calculate P r(B), P r(A—B), P r(A—BC), P r(B—A) and P r(B A).

Past experience shows that there, on average, are 2 traffic accidents on a particular stretch of road every week.
What is the probability of:

- Four accidents during a randomly selected week?
- No accidents during a randomly selected week?
- The Poisson mean $\lambda = 2$ per week.
- (Unit period is 1 week for both questions)
- We use this following formula
$$P(X = k) = \frac{e^{-\lambda} \times \lambda^k}{k!}$$
- Using our value for the Poisson mean
$$P(X = k) = \frac{e^{-2} \times 2^k}{k!}.$$

Probability of four accidents during a randomly selected week : $P(X = 4)$.

$$P(X = 4) = \frac{e^{-2} \times 2^4}{4!}$$

Probability of no accidents during a randomly selected week : $P(X = 0)$.

$$P(X = 0) = \frac{e^{-2} \times 2^0}{0!}$$

What is the expected value and standard deviation of the distribution?