

# Poisson processes

Josemari Sarasola

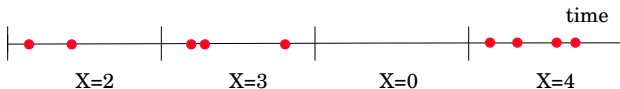
Statistics for Business

Gizapedia



# Poisson distribution

Poisson distribution gives the probability of occurring  $x$  punctual events (failures, arrival of customers, accidents) in a period of time (minute, hour, day) .



# Poisson distribution

These conditions must be held to use the Poisson distribution:

- punctual events must take place absolutely randomly, and hence are independent;
- punctual events happen at a constant  $\lambda$  rate in each period of time.

## Mass function

$$P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!}; \quad x = 0, 1, 2, \dots$$

## Expression, expectation and variance

$$X \sim P(\lambda) \begin{cases} \mu = \lambda \\ \sigma^2 = \lambda \end{cases}$$

It has only 1 parameter:  $\lambda$ .

## Poisson distribution as limit of the binomial distribution

We can approximate binomial distribution by means of the Poisson distribution, when  $n$  is big and  $p$  small:

$$B(n, p) \rightarrow P(\lambda = np)$$

It's a good approximation when  $n \geq 20$  and  $p \leq 0.05$ , and very good when  $n \geq 100$  and  $np \leq 10$ .

# Poisson distribution

## Return periods

- 6 magnitude earthquakes take place once in 1000 years in Iberian peninsula Iberiar Penintsulan. So, return period for such earthquakes is 1000 years.
- Fatal road accidents occur one in 15 days in Gipzukoa. So, return periods for such accidents is 15 days.

### Return period ( $T$ )

It's the time needed for  $\lambda = 1$ . So, if *lambda* for  $t$  time is  $\lambda_t$ ,

$$T = \frac{t}{\lambda_t}.$$

Example: if we have 4 failures every 20 months, return period is  $20/4=5$ .

## R software commands

- Customers arrive at a rate of 4 per hour randomly.  $X$ : number of customers in an hour.
- $P[X = 2]$   
`dpois(2,4)`
- $P[X \leq 2]$   
`ppois(2,4)`
- $P[X > 6]$   
`ppois(6,4,lower.tail=FALSE)`

# Exponential distribution

If the number of events is random, as in the Poisson distribution, time between events will be also random. If the number of events follows the Poisson distribution, time between consecutive events follows this distribution, named the exponential distribution:

## Density function and distribution function

$$f(x) = \lambda e^{-\lambda x}; x > 0$$

$$F(x) = 1 - e^{-\lambda x}; x > 0$$

## Expected value and variance

$$X \sim \text{Exp}(\lambda) : \mu = \frac{1}{\lambda}; \sigma^2 = \frac{1}{\lambda^2}$$

## Time units in exponential distributions

- The expected value is obvious: if we have  $\lambda = 4$  per hour, average time between events will be  $1/\lambda = 0.25$  hours.
- Pay attention to this coincidence: if  $\lambda$  is per hour,  $x$  data and results about the exponential will be also in hours; if  $\lambda$  is per minute,  $x$  data and results will be in minutes.



## Memorylessness

If  $t$  time has passed, probability of supplementary time being  $x$  or more (so, in total,  $t + x$ ) and probability of passing  $x$  time units from the beginning are the same:

$$P(X > t + x | X > t) = P(X > x)$$

This property is linked with the randomness of the process.

## As model for durations and lifespans

- The exponential distribution is at its origin used for time between events, but its shape is useful also to model durations and lifespans.
- But for that we should take into account that memorylessness is always a property for the exponential distribution; for lifespans that means that there is no decay or deterioration over time.

## As model for durations and lifespans

- So, lifespans for a new car and a 10 years old car will be the same in the exponential distribution. That means that in the exponential distribution failure rate remains constant.
- In practice, objects don't behave at a constant failure rate. Actually, the failure rate increases over time (a 10 years old car has a bigger failure rate than a new car). For increasing failure rates, we should use the *Weibull distribution*.

## As model for durations and lifespans

- Sometimes, failure rate is decreasing at the beginning, then it remains constant and at the ending it's increasing.
- For example, persons have a decreasing failure rate when they are children. Infants have a critical health, but over time they get better and better. When we are young, our failure rate (tendency to illness) remains constant, and finally when we become older, the failure rate increases dramatically.
- For those changing failure rates, the best distribution about lifespans is *Gompertz distribution*.

# Gamma distribution

As the number of events over a period of time following a Poisson distribution is random, **the time till the  $k$ -th event** will also be random. The distribution of that time is mathematically fixed and has its own name: the Gamma distribution. It has 2 parameters,  $\lambda$  (the mean number of events, as usual), and  $k$ , that is the forthcoming  $k$ -th event. Briefly we write it ( $\Gamma$ , gamma capital letter):

$$X \sim \Gamma(k, \lambda)$$

When we have  $k = 1$ , the Gamma distribution is the same as the exponential distribution, logically.

Likewise,  $\Gamma(k, \lambda)$  distribution is the sum of  $k$   $Exp(\lambda)$  independent distributions. Thus, we may calculate easily the mean and variance for the Gamma distribution (remember the mean and variances will be the sum of means and variances of exponential distributions).

# Gamma distribution

## Mean and variance

Thus, as Gamma distribution is the sum of  $k$  exponential distributions:

$$X \sim \Gamma(k, \lambda) : \mu = \frac{k}{\lambda}; \sigma^2 = \frac{k}{\lambda^2}$$

# Gamma distribution

## Calculation of probabilities

Calculating probabilities from the density or distribution functions is rather complex. Look at the density function of the Gamma distribution:

$$f(x; k, \lambda) = \frac{x^{k-1} e^{-\lambda x}}{\frac{1}{\lambda^k} \Gamma(k)} \quad x > 0; k, \lambda > 0$$

$\Gamma(k)$  being a special integral function.

We can calculate probabilities more easily thinking about number of events instead of time between  $k$  events.

E.g., having  $X \sim \Gamma(k = 3, \lambda = 0.3)$ , how do we calculate  $P[X > 10]$ ?

The condition to have more than 10 minutes till the 3rd event is having less than 3 events in 10 minutes. For a period of 10 minutes,  $\lambda = 0.3 \times 10 = 3$ . Thus:

$$P[X > 10] = \frac{e^{-3} 3^0}{0!} + \frac{e^{-3} 3^1}{1!} + \frac{e^{-3} 3^2}{2!}$$