Poisson processes

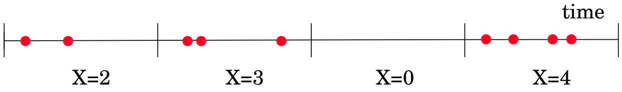
Josemari Sarasola

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 λ rate in each period of time.

Mass function

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

Expression, expectation and variance

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

In has only 1 parameter: λ.

Poisson distribution

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:

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

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

- 6 magnitude earthquakes tahe 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 λ_t, $T = \frac{t}{\lambda_t}$.

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

Poisson distribution

R software commands

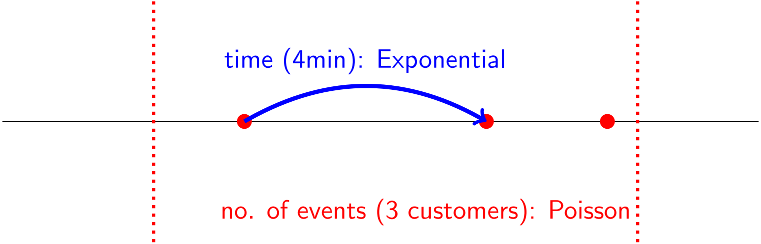
- Customers arrive at a rate of 4 per hour randomly. X: number of customers in an hour.
- $P[X = 2]$?

  ```
  dpois(2,4)
  ```
- $P[X \leq 2]$?

  ```
  ppois(2,4)
  ```
- $P[X > 6]$?

  ```
  ppois(6,4,lower.tail=FALSE)
  ```

If the number of events is random, time between events will be also random. The number of events follows the Poisson distribution, and we will say that time between consecutive events follows the exponential distribution:



Exponential distribution

Let's give more information about the exponential distribution:

Distribution function [F(x)] (and complementary)

$$F(x) = P[X < x] = 1 - e^{-\lambda x};\ x > 0$$

$$P[X > x] = e^{-\lambda x};\ x > 0$$

Expected value and variance

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

Remark: as time is a continuous variable: $P[X < x] = P[X \leq x]$

Exponential distribution

Time units in exponential distributions

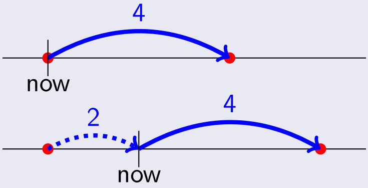
- The expected value is obvious: if we have $\lambda = 4$ per hour, average time between events will be $1/\lambda = 0.25\ hours$.
- Pay attention to this coincidence: if lambda is per hour, x data and results about the exponential will be also in hours; if lambda is per minute, x data and results will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

Exponential distribution

Memorylessness

As events occur randomly and independently, the probability of having more than 4min between consecutive events, and, known that 2min have gone, the probability of having again more than 4min between consecutive events (so, in total at least 6min) are the same:

$$P[X > 4] = P[X > 6/X > 2]$$



That is knowing that 2min have gone doesn't give us any relevant information.

Exponential distribution

Memorylessness

So, generally, if t time is gone, 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 in directly linked wih the randomness of the process. So, generally the exponential distribution tells us about the time till the next event, and not only about the time between consecutive events.

Exponential distribution

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.

Exponential distribution

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 arte. 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.

Exponential distribution

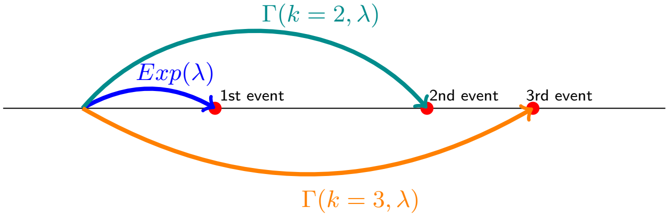
As model for durations and lifespans

- 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 illnes) 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 (remember that time till next event is $Exp(\lambda)$.
The distribution of time till k-th event is mathematically fixed and has its own name: the Gamma distribution. It has 2 parameters, λ (the mean number of events, as usual), and k, that is the forthcoming k-th event. Briefly we write it (Γ, gamma capital letter):

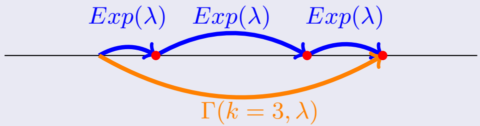
$$X \sim \Gamma(k, \lambda)$$



Gamma distribution

Link with $Exp(\lambda)$

Logically: $Exp(\lambda) \equiv \Gamma(k=1,\lambda)$. Likewise, $\Gamma(k,\lambda)$ distribution is the sum of k $Exp(\lambda)$ independent distributions.



$$Exp(\lambda) + Exp(\lambda) + Exp(\lambda) = \Gamma(k=3,\lambda)$$

Mean and variance of $\Gamma(k,\lambda)$

Thus, we may calculate easily the mean and variance for the Gamma distribution adding the means $(1/\lambda)$ and variances $(1/\lambda^2)$ of k exponential distributions.

As Gamma distribution is the sum of k exponential distributions:

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

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

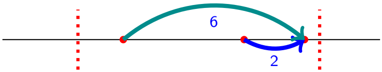
$$f(x;k,\lambda) = \frac{x^{k-1}e^{-\lambda x}}{\frac{1}{\lambda^k}\Gamma(k)} \quad x > 0; k, \lambda > 0$$

$\Gamma(k)$ 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:

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

A Poisson process (example):



Distributions related to a Poisson process (λ: mean no. of events, k: rank of event):

Name	Notation	Variable	Example
Poisson	$P(\lambda)$	No of events in a period	x=3
Exponential	$G(p)$	Time till the next event	x=2
Gamma	$\Gamma(k,\lambda)$	Time till k-th event	k=2,x=6

Name	Formula	Support
$P(\lambda)$	$P[X=x]=\frac{e^{-\lambda}\lambda^x}{x!}$	$x=0,1,2,..,\infty$
$Exp(\lambda)$	$P[X<x]=1-e^{-\lambda x}$	$x>0$
$\Gamma(k,\lambda)$	complex (solved with Poisson)	$x>0$