COMP312/DATA304/DATA474
Simulation & Stochastic Models

The Poisson Process: Basic Elements

Alejandro C. Frery

T1 2023

VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA

School of Mathematics and Statistics
New Zealand

What is it about?

The Poisson Process (not to be confused with the Poisson distribution) is at the core of the theory underlying, among others, queues.

We will see the two basic elements that arise from the Poisson Process, namely:

- The Exponential distribution,
- The Poisson distribution.

Did we say distribution?

Reminders:

- We don't write or plot a distribution.
- We know the distribution of random variable when we are able to compute the probability of all imaginable events.
- The distribution of a random variable X is always characterized by its cumulative distribution function $F_X(t) = \Pr(X \leq t)$, for every $t \in \mathbb{R}$.
- If the random variable is discrete, we can also characterize its distribution with its probability function $\boldsymbol{p} = (p_0, p_1, p_2, \dots)$, where $p_i = \Pr(X = i)$.
- If the random variable is continuous, we can also characterize its distribution with its probability density function f_X, which satisfies (i) $f_X \geq 0$, and (ii) $F_X(t) = \int_{-\infty}^{t} f(x)dx$.

Did we say distribution?

Reminders:

- We don't write or plot a distribution.
- We know the distribution of random variable when we are able to compute the probability of all imaginable events.
- The distribution of a random variable X is always characterized by its cumulative distribution function $F_X(t) = \Pr(X \leq t)$, for every $t \in \mathbb{R}$.
- If the random variable is discrete, we can also characterize its distribution with its probability function $\boldsymbol{p} = (p_0, p_1, p_2, \dots)$, where $p_i = \Pr(X = i)$.
- If the random variable is continuous, we can also characterize its distribution with its probability density function f_X, which satisfies (i) $f_X \geq 0$, and (ii) $F_X(t) = \int_{-\infty}^{t} f(x)dx$.

Did we say distribution?

Reminders:

- We don't write or plot a distribution.
- We know the distribution of random variable when we are able to compute the probability of all imaginable events.
- The distribution of a random variable X is always characterized by its cumulative distribution function $F_X(t) = \Pr(X \leq t)$, for every $t \in \mathbb{R}$.
- If the random variable is discrete, we can also characterize its distribution with its probability function $\boldsymbol{p} = (p_0, p_1, p_2, \dots)$, where $p_i = \Pr(X = i)$.
- If the random variable is continuous, we can also characterize its distribution with its probability density function f_X, which satisfies (i) $f_X \geq 0$, and (ii) $F_X(t) = \int_{-\infty}^{t} f(x)dx$.

Did we say distribution?

Reminders:

- We don't write or plot a distribution.
- We know the distribution of random variable when we are able to compute the probability of all imaginable events.
- The distribution of a random variable X is always characterized by its cumulative distribution function $F_X(t) = \Pr(X \leq t)$, for every $t \in \mathbb{R}$.
- If the random variable is discrete, we can also characterize its distribution with its probability function $\boldsymbol{p} = (p_0, p_1, p_2, \dots)$, where $p_i = \Pr(X = i)$.
- If the random variable is continuous, we can also characterize its distribution with its probability density function f_X, which satisfies (i) $f_X \geq 0$, and (ii) $F_X(t) = \int_{-\infty}^{t} f(x)dx$.

Did we say distribution?

Reminders:

- We don't write or plot a distribution.
- We know the distribution of random variable when we are able to compute the probability of all imaginable events.
- The distribution of a random variable X is always characterized by its cumulative distribution function $F_X(t) = \Pr(X \leq t)$, for every $t \in \mathbb{R}$.
- If the random variable is discrete, we can also characterize its distribution with its probability function $\boldsymbol{p} = (p_0, p_1, p_2, \dots)$, where $p_i = \Pr(X = i)$.
- If the random variable is continuous, we can also characterize its distribution with its probability density function f_X, which satisfies (i) $f_X \geq 0$, and (ii) $F_X(t) = \int_{-\infty}^{t} f(x)dx$.

The Exponential distribution

Definition

We say that the random variable $T\colon \Omega \to \mathbb{R}_+$ obeys the standard (unitary-mean) Exponential law if its cumulative distribution function is

$$F_T(t) = \left(1 - e^{-t}\right)\mathbb{1}_{\mathbb{R}_+}(t).$$

Consequently, the density that characterizes its distribution is

$$f_T(t) = e^{-t}\mathbb{1}_{\mathbb{R}_+}(t).$$

We denote this situation as $T \sim \operatorname{Exp}(1)$.

More generally, for every $\lambda > 0$, we say that $V = \lambda T$ follows an Exponential distribution with mean λ (or, equivalently, with rate $\theta = 1/\lambda$). We denote this situation as $V \sim \operatorname{Exp}(\lambda)$, making explicit mention that λ is the mean.

The mean and variance of $T \sim \operatorname{Exp}(\lambda)$ are, respectively, λ and λ^2.

Properties

A no, a yes, and another no:

- The Exponential distribution is not a location family: if $V \sim \text{Exp}(\lambda)$, then $V + \mu$, with $\mu \in \mathbb{R}$ is not exponentially-distributed unless $\mu = 0$.
- The Exponential distribution is a scale family: if $V \sim \text{Exp}(\lambda)$, then ηV, with $\eta > 0$ is exponentially-distributed with mean $\eta\lambda$.
- The Exponential distribution is not preserved by convolutions: if V_1 and V_2 are independent identically distributed standard Exponential random variables, then $V_1 + V_2$ is not exponentially-distributed.

The maximum likelihood estimator of λ based on the random sample $V_1, V_2, \ldots, V_n$ is $\widehat{\lambda} = n^{-1} \sum_{i=1}^{n} V_i$.

It is the Dory of the distributions (it lacks memory).

Properties

A no, a yes, and another no:

- The Exponential distribution is not a location family: if $V \sim \text{Exp}(\lambda)$, then $V + \mu$, with $\mu \in \mathbb{R}$ is not exponentially-distributed unless $\mu = 0$.
- The Exponential distribution is a scale family: if $V \sim \text{Exp}(\lambda)$, then ηV, with $\eta > 0$ is exponentially-distributed with mean $\eta\lambda$.
- The Exponential distribution is not preserved by convolutions: if V_1 and V_2 are independent identically distributed standard Exponential random variables, then $V_1 + V_2$ is not exponentially-distributed.

The maximum likelihood estimator of λ based on the random sample $V_1, V_2, \ldots, V_n$ is $\widehat{\lambda} = n^{-1} \sum_{i=1}^{n} V_i$.

It is the Dory of the distributions (it lacks memory).

Properties

A no, a yes, and another no:

- The Exponential distribution is not a location family: if $V \sim \text{Exp}(\lambda)$, then $V + \mu$, with $\mu \in \mathbb{R}$ is not exponentially-distributed unless $\mu = 0$.
- The Exponential distribution is a scale family: if $V \sim \text{Exp}(\lambda)$, then ηV, with $\eta > 0$ is exponentially-distributed with mean $\eta\lambda$.
- The Exponential distribution is not preserved by convolutions: if V_1 and V_2 are independent identically distributed standard Exponential random variables, then $V_1 + V_2$ is not exponentially-distributed.

The maximum likelihood estimator of λ based on the random sample $V_1, V_2, \dots, V_n$ is $\widehat{\lambda} = n^{-1} \sum_{i=1}^{n} V_i$.

It is the Dory of the distributions (it lacks memory).

Properties

A no, a yes, and another no:

- The Exponential distribution is not a location family: if $V \sim \text{Exp}(\lambda)$, then $V + \mu$, with $\mu \in \mathbb{R}$ is not exponentially-distributed unless $\mu = 0$.
- The Exponential distribution is a scale family: if $V \sim \text{Exp}(\lambda)$, then ηV, with $\eta > 0$ is exponentially-distributed with mean $\eta\lambda$.
- The Exponential distribution is not preserved by convolutions: if V_1 and V_2 are independent identically distributed standard Exponential random variables, then $V_1 + V_2$ is not exponentially-distributed.

The maximum likelihood estimator of λ based on the random sample $V_1, V_2, \ldots, V_n$ is $\widehat{\lambda} = n^{-1}\sum_{i=1}^{n} V_i$.

It is the Dory of the distributions (it lacks memory).

Properties

A no, a yes, and another no:

- The Exponential distribution is not a location family: if $V \sim \text{Exp}(\lambda)$, then $V + \mu$, with $\mu \in \mathbb{R}$ is not exponentially-distributed unless $\mu = 0$.
- The Exponential distribution is a scale family: if $V \sim \text{Exp}(\lambda)$, then ηV, with $\eta > 0$ is exponentially-distributed with mean $\eta\lambda$.
- The Exponential distribution is not preserved by convolutions: if V_1 and V_2 are independent identically distributed standard Exponential random variables, then $V_1 + V_2$ is not exponentially-distributed.

The maximum likelihood estimator of λ based on the random sample $V_1, V_2, \ldots, V_n$ is $\widehat{\lambda} = n^{-1} \sum_{i=1}^{n} V_i$.

It is the Dory of the distributions (it lacks memory).

Properties

A no, a yes, and another no:

- The Exponential distribution is not a location family: if $V \sim \text{Exp}(\lambda)$, then $V + \mu$, with $\mu \in \mathbb{R}$ is not exponentially-distributed unless $\mu = 0$.
- The Exponential distribution is a scale family: if $V \sim \text{Exp}(\lambda)$, then ηV, with $\eta > 0$ is exponentially-distributed with mean $\eta\lambda$.
- The Exponential distribution is not preserved by convolutions: if V_1 and V_2 are independent identically distributed standard Exponential random variables, then $V_1 + V_2$ is not exponentially-distributed.

The maximum likelihood estimator of λ based on the random sample $V_1, V_2, \ldots, V_n$ is $\widehat{\lambda} = n^{-1} \sum_{i=1}^{n} V_i$.

It is the Dory of the distributions (it lacks memory).

Assume we want to model the time until failure of a simple system, e.g., a LED bulb, that does not age with time. What we want is a random variable Y such that the fact that the bulb is functioning up to time s does not affect the probability of failure at time $s + t$.

$$\Pr(Y > s + t \mid Y > s) = \Pr(Y > t).$$

Using the definition of the conditional probability:

$$\Pr(Y > s + t \mid Y > s) = \frac{\Pr(Y > s + t \cap Y > s)}{\Pr(Y > s)},$$

but the events overlap such that

$$= \frac{\Pr(Y > s + t)}{\Pr(Y > s)}.$$

Using the definition of the Exponential distribution:

$$= \frac{e^{-(s+t)}}{e^{-s}} = e^{-t} = \Pr(Y > t),$$

which does not depend on s.

So, the Exponential distribution is a model without memory.

The books by Johnson et al. (1994, 1995) are among the most comprehensive references about continuous distributions.

The Poisson distribution

From observations to hypotheses

We are observing cars passing, customers arriving, planes departing, raindrops falling on a dry surface, failures of computers in a large data storage facility...
They all are

- discrete events
- that occur at different instants
- one event does not seem to interfere with the others.

From observations to hypotheses

We are observing cars passing, customers arriving, planes departing, raindrops falling on a dry surface, failures of computers in a large data storage facility…
They all are

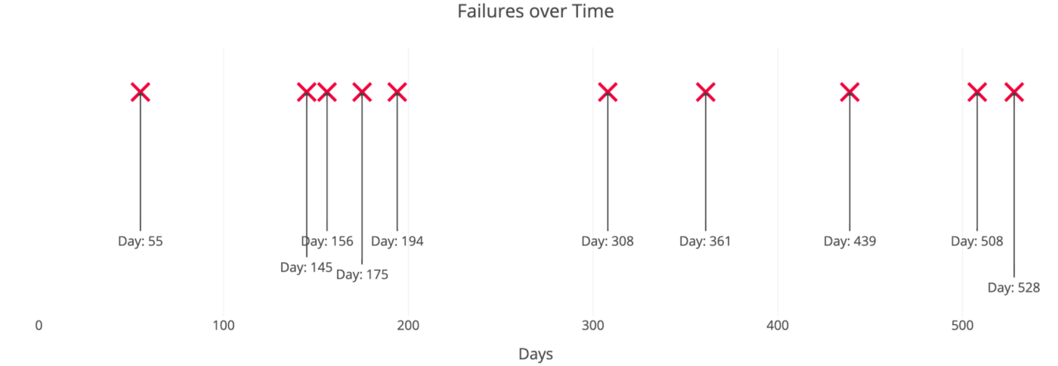
- discrete events
- that occur at different instants
- one event does not seem to interfere with the others.

From observations to hypotheses

We are observing cars passing, customers arriving, planes departing, raindrops falling on a dry surface, failures of computers in a large data storage facility…
They all are

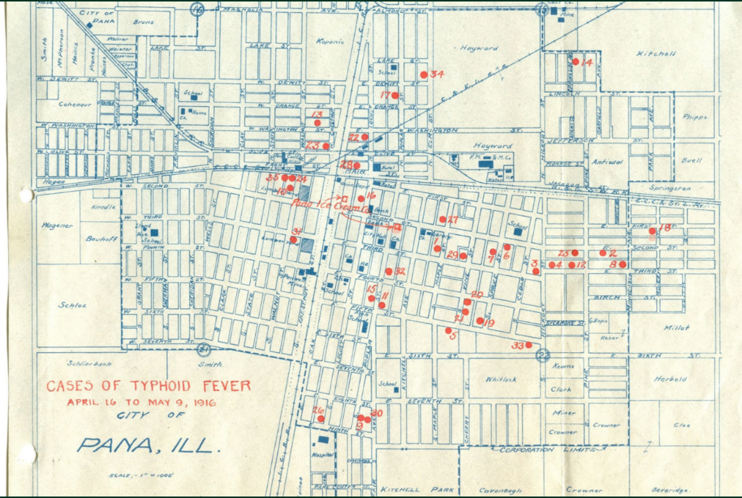
- discrete events
- that occur at different instants
- one event does not seem to interfere with the others.

From observations to hypotheses



Source: `https://towardsdatascience.com/the-poisson-distribution-and-poisson-process-explained-4e2cb17d459`

From observations to hypotheses



Source: `https://blogs.illinois.edu/view/7859/2090039869`

Translating:

- discrete events: counting process, we are looking for a discrete distribution that describes $\Pr(X_{[t_1,t_2]} = k)$ with $k = 0, 1, 2, \ldots$
- that occur at different instants, so they do not overlap: $\lim_{t_2 \to t_1} \Pr(X_{[t_1,t_2]} > 1) = 0$,
- and different events do not interfere: there is statistical independence: $\Pr(X_{[t_1,t_2]} = k \text{ and } X_{[t_3,t_4]} = \ell) = \Pr(X_{[t_1,t_2]} = k) \Pr(X_{[t_3,t_4]} = \ell)$ provided $[t_1, t_2] \cap [t_3, t_4] = \emptyset$.

From observations to hypotheses

Translating:

- discrete events: counting process, we are looking for a discrete distribution that describes $\Pr(X_{[t_1,t_2]} = k)$ with $k = 0, 1, 2, \ldots$
- that occur at different instants, so they do not overlap: $\lim_{t_2 \to t_1} \Pr(X_{[t_1,t_2]} > 1) = 0$,
- and different events do not interfere: there is statistical independence: $\Pr(X_{[t_1,t_2]} = k \text{ and } X_{[t_3,t_4]} = \ell) = \Pr(X_{[t_1,t_2]} = k) \Pr(X_{[t_3,t_4]} = \ell)$ provided $[t_1, t_2] \cap [t_3, t_4] = \emptyset$.

From observations to hypotheses

Translating:

- discrete events: counting process, we are looking for a discrete distribution that describes $\Pr(X_{[t_1,t_2]} = k)$ with $k = 0, 1, 2, \dots$
- that occur at different instants, so they do not overlap: $\lim_{t_2 \to t_1} \Pr(X_{[t_1,t_2]} > 1) = 0$,
- and different events do not interfere: there is statistical independence: $\Pr(X_{[t_1,t_2]} = k \text{ and } X_{[t_3,t_4]} = \ell) = \Pr(X_{[t_1,t_2]} = k)\Pr(X_{[t_3,t_4]} = \ell)$ provided $[t_1, t_2] \cap [t_3, t_4] = \emptyset$.

Counting events

The Bernoulli law is the simplest model for dichotomic (Yes, No) events:

$$\Pr(X = 1) = 1 - \Pr(X = 0) = p,$$

with $0 < p < 1$. Call 1 "success" and 0 "failure." Nothing to question here, right?

Observe n Bernoulli events $X_1, X_2, \dots, X_n$ and sum their results $X = \sum_{i=1}^{n} X_i$, you may have anything from 0 to n. Assume further that the n trials are independent. You may want to prove that

$$\Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} = \frac{n!}{(n-k)!k!} p^k (1 - p)^{n-k},$$

with $k = 0, 1, 2, \dots, n$.

We say that the random variable X follows a Binomial distribution with probability $p \in (0, 1)$ of success in each of the $n \geq 1$ trials. Anything to question here?

Counting events

Let us write $\lambda = np$. With this, we have

$$\Pr(X = k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Now let's assume that the probability of any event is very small, but that we make many independent observations, i.e., $p \to 0$ and $n \to \infty$, but such that $np = \lambda$. Carrying on:

$$\Pr(X = k) = \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n!}{n^k (n-k)!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

We will obtain the limit in three parts.

First part: expand numerator and denominator, check the terms that cancel out, group, and find that

$$\lim_{n\to\infty} \frac{n!}{n^k(n-k)!} = 1.$$

Second part: use the definition $e = \lim_{t\to\infty}(1+1/t)^t$, and see that

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

Third part: it is easy to see that

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1.$$

Gathering these results, we have that

$$\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

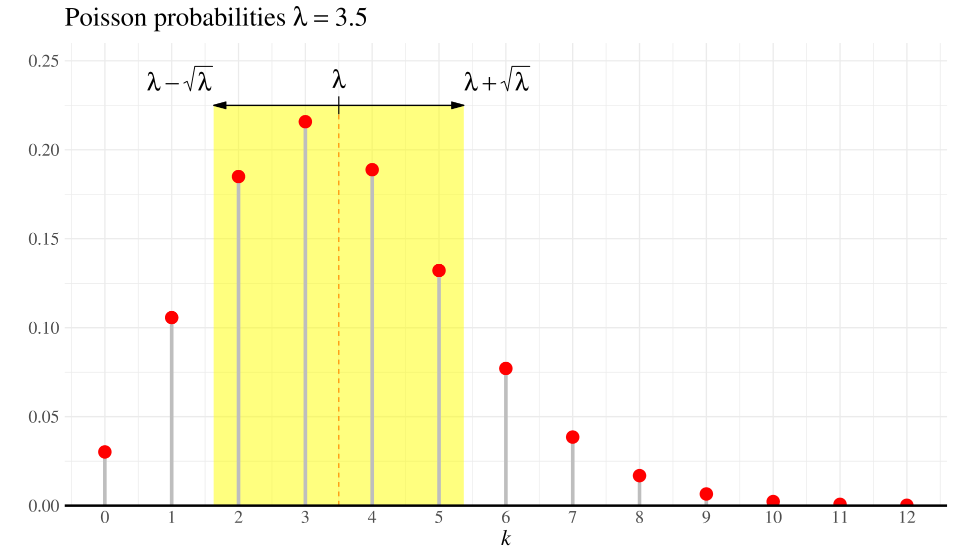
is the limit distribution of a Binomial counting process when the number of trials $n \to \infty$ but when, at the same time, the probability of success is $p \to 0$ such that $np = \lambda$.

This is the Poisson distribution, denoted as $X \sim \text{Po}(\lambda)$, with $\lambda \in \mathbb{R}_+$. At this point you should make plots of $\Pr(X = k; \lambda)$.

It is useful to remind that $\text{E}(X) = \text{Var}(X) = \lambda$.

Do you remember how to obtain the maximum likelihood estimator of λ given the random sample $X_1, X_2, \ldots, X_m$ of i.i.d. Poisson deviates?

What do you see in this plot? (From Frery 2021)



Properties

Some properties of the Poisson distribution are the following:

- It is not a location family of distributions, i.e., if $X \sim \text{Po}(\lambda)$, then $X + \mu$, with $\mu \in \mathbb{R}$, is not Poisson distributed unless $\mu = 0$.
- It is not an scale family of distributions, i.e., if $X \sim \text{Po}(\lambda)$, then ηX, with $\eta > 0$, is not Poisson distributed unless $\eta = 1$.
- The distribution of the sum of independent Poisson random variables with parameters $\lambda_1, \lambda_2, \dots, \lambda_n$ is Poisson distributed with parameter $\lambda = \sum_{i=1}^{n} \lambda_i$.

Properties

Some properties of the Poisson distribution are the following:

- It is not a location family of distributions, i.e., if $X \sim \text{Po}(\lambda)$, then $X + \mu$, with $\mu \in \mathbb{R}$, is not Poisson distributed unless $\mu = 0$.
- It is not an scale family of distributions, i.e., if $X \sim \text{Po}(\lambda)$, then ηX, with $\eta > 0$, is not Poisson distributed unless $\eta = 1$.
- The distribution of the sum of independent Poisson random variables with parameters $\lambda_1, \lambda_2, \dots, \lambda_n$ is Poisson distributed with parameter $\lambda = \sum_{i=1}^{n} \lambda_i$.

Properties

Some properties of the Poisson distribution are the following:

- It is not a location family of distributions, i.e., if $X \sim \text{Po}(\lambda)$, then $X + \mu$, with $\mu \in \mathbb{R}$, is not Poisson distributed unless $\mu = 0$.
- It is not an scale family of distributions, i.e., if $X \sim \text{Po}(\lambda)$, then ηX, with $\eta > 0$, is not Poisson distributed unless $\eta = 1$.
- The distribution of the sum of independent Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$ is Poisson distributed with parameter $\lambda = \sum_{i=1}^{n} \lambda_i$.

One of the most comprehensive references...

... about discrete distributions is the book by Johnson et al. (1993).

Tufte (2001) discusses, with plenty of examples, how to make good plots.

Tufte (2004) is also very critical about the use of "PowerPoint-like" slides.

References

Frery, A. C. (2021), Standard deviation, *in* B. S. Daya Sagar, Q. Cheng, J. McKinley & F. Agterberg, eds, 'Encyclopedia of Mathematical Geosciences', Springer International Publishing, pp. 1–3.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994), *Continuous Univariate Distributions*, Vol. 1 of *Wiley Series in Probability and Mathematical Statistics*, 2 edn, John Wiley & Sons, New York.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995), *Continuous Univariate Distributions*, Vol. 2, 2 edn, John Wiley & Sons, New York.

Johnson, N. L., Kotz, S. & Kemp, A. W. (1993), *Univariate Discrete Distributions*, Wiley Series in Probability and Mathematical Statistics, 2 edn, John Wiley & Sons, New York.

Tufte, E. R. (2001), *The Visual Display of Quantitative Information*, 2 edn, Graphics Press.

Tufte, E. R. (2004), *The Cognitive Style of PowerPoint*, Graphics Press.