Section 02 - Poisson distribution

MCB 111

September 16, 2022

The number of independent events in a time interval

An example from physical science: you are watching a meteor shower from Charles River. Suppose meteors arrive independently from each other, and during a very small time interval δt minute, the probability of seeing a meteor is proportional to the length of the time interval:

$$\mathbb{P}(\text{seeing a meteor during } \delta t \text{ minute}) \approx \lambda \delta t \tag{1}$$

How many stars (X) can I see for the entire meteor shower that lasts T minutes?

What is the probability that I see no meteor at all?

$$\begin{aligned}\mathbb{P}(X = 0) &\approx (1 - \lambda\delta t)^{\frac{T}{\delta t}} \\ &\approx \exp\left(-\lambda\delta t\frac{T}{\delta t}\right) \\ &= \exp\left(-\lambda T\right)\end{aligned} \tag{2}$$

What is the probability that I see exactly one meteor?

$$\begin{aligned}\mathbb{P}(X = 1) &\approx \binom{T/\delta t}{1} (\lambda \delta t)^1 (1 - \lambda \delta t)^{\frac{T}{\delta t} - 1} \\ &\approx \frac{T}{\delta t} (\lambda \delta t) \exp\left(-\lambda \delta t \frac{T}{\delta t} - \lambda \delta t\right) \\ &\approx \lambda T \exp\left(-\lambda T\right)\end{aligned} \tag{3}$$

What is the probability that I see exactly two meteors?

$$\begin{aligned}\mathbb{P}(X=1) &\approx \binom{T/\delta t}{2} (\lambda\delta t)^2 (1-\lambda\delta t)^{\frac{T}{\delta t}-2} \\ &\approx \frac{\frac{T}{\delta t}\left(\frac{T}{\delta t}-1\right)}{2!}(\lambda\delta t)^2 \exp\left(-\lambda\delta t\frac{T}{\delta t}-2\lambda\delta t\right) \\ &\approx \frac{(\lambda T)^2}{2!}\exp\left(-\lambda T\right)\end{aligned} \tag{4}$$

What is the probability that I see exactly n meteors?

$$\begin{aligned}
\mathbb{P}(X=n) &\approx \begin{pmatrix} T/\delta t \\ n \end{pmatrix} (\lambda\delta t)^n (1-\lambda\delta t)^{\frac{T}{\delta t}-n} \\
&\approx \frac{\frac{T}{\delta t}\cdots\left(\frac{T}{\delta t}-n+1\right)}{n!}(\lambda\delta t)^n \exp\left(-\lambda\delta t\frac{T}{\delta t}-n\lambda\delta t\right) \\
&\approx \frac{(\lambda T)^n}{n!}\exp\left(-\lambda T\right)
\end{aligned} \tag{5}$$

This formula gives us the Poisson distribution with parameter "λT"

Poisson distribution's probability mass function

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} \exp(-\lambda), \quad n = 0, 1, 2, \cdots \tag{6}$$

Very useful when you are counting something that occurs independently in space/time with a fixed rate

Poisson distribution's probability mass function

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} \exp(-\lambda), \quad n = 0, 1, 2, \cdots \tag{6}$$

Very useful when you are counting something that occurs independently in space/time with a fixed rate

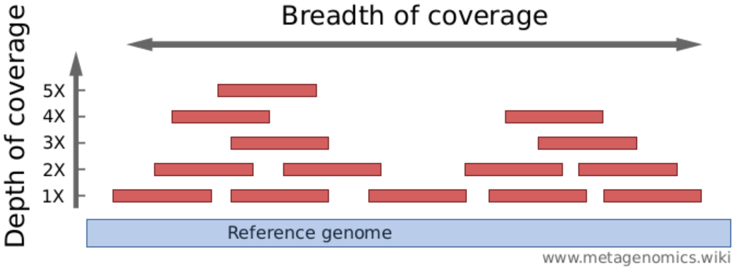
- Genome mutations

Poisson distribution's probability mass function

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} \exp(-\lambda), \quad n = 0, 1, 2, \cdots \tag{6}$$

Very useful when you are counting something that occurs independently in space/time with a fixed rate

- Shotgun sequencing: the number of reads covering a site in the genome



Poisson distribution's probability mass function

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!}\exp(-\lambda), \quad n = 0, 1, 2, \cdots \tag{6}$$

Very useful when you are counting something that occurs independently in space/time with a fixed rate

- The number of bird poops on the floor

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats (cyclic process)

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats (cyclic process)
- The number of Canada geese in Harvard Yard

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats (cyclic process)
- The number of Canada geese in Harvard Yard (they move by families)

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats (cyclic process)
- The number of Canada geese in Harvard Yard (they move by families)
- The number of redline trains arriving at Harvard Sq

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats (cyclic process)
- The number of Canada geese in Harvard Yard (they move by families)
- The number of redline trains arriving at Harvard Sq (adjacent trains repulse each other)

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats (cyclic process)
- The number of Canada geese in Harvard Yard (they move by families)
- The number of redline trains arriving at Harvard Sq (adjacent trains repulse each other)
- The number of neurons firing in a neural network

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats (cyclic process)
- The number of Canada geese in Harvard Yard (they move by families)
- The number of redline trains arriving at Harvard Sq (adjacent trains repulse each other)
- The number of neurons firing in a neural network (cooperative firing)

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats (cyclic process)
- The number of Canada geese in Harvard Yard (they move by families)
- The number of redline trains arriving at Harvard Sq (adjacent trains repulse each other)
- The number of neurons firing in a neural network (cooperative firing)
- The number of fruitflies in BioLabs

Poisson distribution cannot be used to describe events with correlation among instances

Do you think these events in a given space/time can be described by a Poisson distribution?

- The number of heartbeats (cyclic process)
- The number of Canada geese in Harvard Yard (they move by families)
- The number of redline trains arriving at Harvard Sq (adjacent trains repulse each other)
- The number of neurons firing in a neural network (cooperative firing)
- The number of fruitflies in BioLabs (Maybe, but correlation induced by experiments?)

Nice properties of the Poisson distribution

Parameter λ = Mean = Variance

$$
\begin{aligned}
\mathbb{E}(X) &= 0 + \sum_{n=1}^{\infty} n \exp(-\lambda) \frac{\lambda^n}{n!} \\
&= \exp(-\lambda) \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \\
&= \lambda \exp(-\lambda) \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \\
&= \lambda \exp(-\lambda) \sum_{n=0}^{\infty} \frac{\lambda^n}{(n)!} \\
&= \lambda \exp(-\lambda) \exp(\lambda) \\
&= \lambda
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
\mathbb{E}(X) &= 0 + \sum_{n=1}^{\infty} n \exp(-\lambda) \frac{\lambda^n}{n!} \\
&= \exp(-\lambda) \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \\
&= \lambda \exp(-\lambda) \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \\
&= \lambda \exp(-\lambda) \sum_{n=0}^{\infty} \frac{\lambda^n}{(n)!} \\
&= \lambda \exp(-\lambda) \exp(\lambda) \\
&= \lambda
\end{aligned} \tag{7}
$$

"Variance=λ" is left for you to do

The sum of independent Poisson variables is also Poisson

$$\begin{aligned} & X_1 \sim \text{Poisson}(\lambda_1) \\ & X_2 \sim \text{Poisson}(\lambda_2) \\ & \Downarrow \\ & X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2) \end{aligned} \tag{8}$$

Why?

Intuitive answer:
In the meteor shower example, if λ_1 is the rate of seeing red meteors, and λ_2 is the rate of seeing green meteors, then in a short interval δ_t, the probability of seeing either a green or a red meteor is

Intuitive answer:

In the meteor shower example, if λ_1 is the rate of seeing red meteors, and λ_2 is the rate of seeing green meteors, then in a short interval δ_t, the probability of seeing either a green or a red meteor is

$$\begin{aligned} &\mathbb{P}(\text{seeing some meteor in } \delta t \text{ minutes}) \\ &\quad = 1 - \mathbb{P}(\text{seeing nothing}) \\ &\quad = 1 - (1 - \lambda_1 \delta t)(1 - \lambda_2 \delta t) \\ &\quad = (\lambda_1 + \lambda_2)\delta t - \lambda_1 \lambda_2 (\delta t)^2 \\ &\quad \approx (\lambda_1 + \lambda_2)\delta t \end{aligned} \tag{9}$$

Intuitive answer:

In the meteor shower example, if λ_1 is the rate of seeing red meteors, and λ_2 is the rate of seeing green meteors, then in a short interval δ_t, the probability of seeing either a green or a red meteor is

$$\begin{aligned} &\mathbb{P}(\text{seeing some meteor in } \delta t \text{ minutes}) \\ &= 1 - \mathbb{P}(\text{seeing nothing}) \\ &= 1 - (1 - \lambda_1 \delta t)(1 - \lambda_2 \delta t) \\ &= (\lambda_1 + \lambda_2)\delta t - \lambda_1 \lambda_2 (\delta t)^2 \\ &\approx (\lambda_1 + \lambda_2)\delta t \end{aligned} \tag{9}$$

So you can define the rate of seeing either red or green meteors as $\lambda_1 + \lambda_2$, which defines another Poisson process (as well as a Poisson distribution).

This summation property is very useful in biology for developing null hypotheses where sub-processes of interest are independently Poisson. Some examples:

- The spike trains of some independent neurons
- Mutations at independent genomic regions
- DNA double-strand breaks across a group of cells

In the two-color meteor shower example, what is the probability that the next meteor is green?

In the two-color meteor shower example, what is the probability that the next meteor is green?

Answer: $\mathbb{P}(\text{next is green}) = \lambda_2/(\lambda_1 + \lambda_2)$

In the two-color meteor shower example, what is the probability that the next meteor is green?

Answer: $\mathbb{P}(\text{next is green}) = \lambda_2/(\lambda_1 + \lambda_2)$

In general, in a Poisson process with n types of outcomes, the probability that the next outcome is of type i is

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_n} \tag{10}$$