Proba2 IIND-2027

Complementaria I

Iván Andrés Trujilllo Abella

ai-page.readthedocs.ios

ivantrujillo1229@gmail.com

Dataset

The following dataset click here contain the number of patients that arrives to an emergency room in holidays during a specific interval hours during several years, specifically the number of patients per hour, with this information compute:

- The probability that arrives exactly two patients in one hour $P(x = 2)$
- The probability that arrives at least 5 patients in one hour $P(x \geq 5)$

Dataset

The following dataset click here contain the number of patients that arrives to an emergency room in holidays during a specific interval hours during several years, specifically the number of patients per hour, with this information compute:

- The probability that arrives exactly two patients in one hour $P(x=2)$
- The probability that arrives at least 5 patients in one hour $P(x \geq 5)$

...

The population parameter is $\lambda = 4$

Solution

...

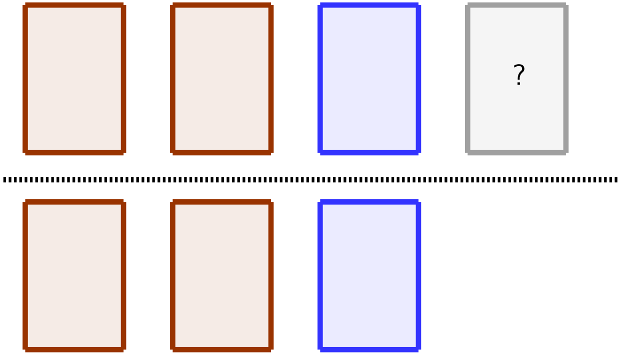
click here

why?

λ as sample mean!

It is important the following.

Freund ilustration



Estimate Θ by maximun likelihood

$$p(k, M, n, N) = \frac{\binom{n}{k}\binom{M-n}{N-k}}{\binom{M}{N}} \tag{1}$$

The **Method of maximun likelihood** consist in estimate the parameter Θ the number of *red cards* that maximize the probability of see the **data** (three red cards and a blue car), in this case Θ could be 2 or 3.

$$\frac{\binom{3}{2}\binom{1}{1}}{\binom{4}{3}} > \frac{\binom{2}{2}\binom{2}{1}}{\binom{4}{3}} \tag{2}$$

Maximun Likelihood Estimation (MLE)

Suppose that you data it is generated by a theoretical distribution, the inverse problem is determine the most probable parameter that generate the data.

More formal statement

...

We need figure out the unknown value Θ, then we collect a random sample of independent and identically distributed data points $x_1, ..., x_n$, namely we have a random sample of a distribution function $f(x \mid \Theta)$.

More formal statement

...

We need figure out the unknown value Θ, then we collect a random sample of independent and identically distributed data points $x_1, ..., x_n$, namely we have a random sample of a distribution function $f(x \mid \Theta)$.

likelihood function

Now, that we have collect data (held constant), find out which is the most likely value of Θ given the data!

$$L(\Theta | x_1, ..., x_n) \tag{3}$$

Likelihood function

definition

Likelihood function is the probability of get the observed data if we use Θ^* as the parameter generator of the data, modeled under $f()$.

$$\begin{aligned} f(x_1, ..., x_n \mid \Theta) &= f(x_1 \mid \Theta), ..., f(x_n \mid \Theta) \\ &= \prod_{i=1}^{n} f(x_i \mid \Theta) \end{aligned} \tag{4}$$

Now we need optimize the expression, for practical reasons is used the natural logarithm (**log likelihood function**), and the find it value is $\hat{\Theta}_{MLE}$ is our maximum likelihood estimator.

Poisson - example

...

for a set of random data points $X(x_1, ..., x_n)$ in which this collection (vector) is independent and identically distributed coming from a Poisson distribution ($X \overset{iid}{\sim} poisson(\lambda)$) then our parameter of interest is λ

poisson maximun likelihood function

$$\begin{aligned} L(\lambda \mid X) &= \prod_{i=1}^{n} f(x_i \mid \lambda) \\ &= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \end{aligned} \tag{5}$$

Poisson

...

$$\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!} \tag{6}$$

if we uses natural log then:

$$\ln(L(\lambda \mid X) = -n\lambda + log(\lambda)\sum x_i - \sum log(x_i) \tag{7}$$

...

$$\begin{gathered} \frac{d \ln(L(\lambda \mid X))}{d\lambda} = 0 \\ \lambda = \frac{1}{N}\sum x_i \end{gathered} \tag{8}$$

Therefore our parameter is estimated with the sample mean.

Exponential

Related with Poisson, model the time between two events.

Uniform