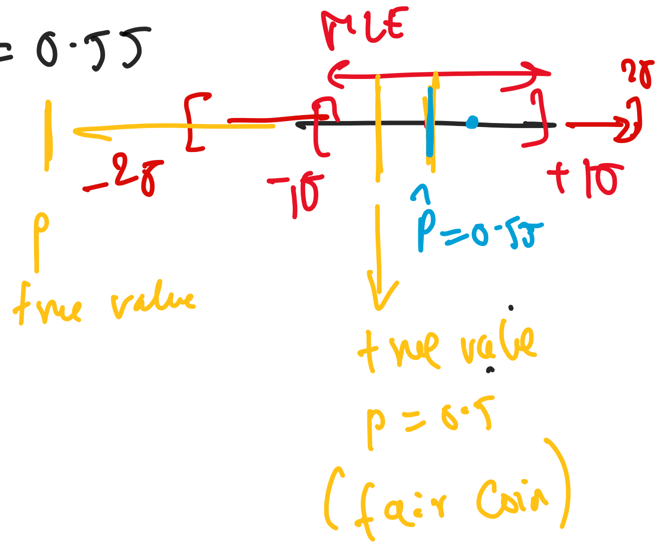
Maximum likelihood function

Thursday, 17 March 2022 11:57

In the previous example we get

$MLE \Rightarrow \hat{p} = 0.55$

$MLE \pm \sigma$



Log likelihood function:

It is often easier to work with the natural log of the likelihood function. log. likelihood function.

- Since ln(x) is an increasing function, the maxima of the likelihood and log-likelihood coincides.

previous example:

$$P(55 \text{ heads} \mid p) = \binom{100}{55} p^{55} (1-p)^{45}$$

$$\ln P(55 \text{ heads} \mid p) = \ln\left[\binom{100}{55} p^{55} (1-p)^{45}\right]$$

$$= \ln\binom{100}{55} + 55 \ln p + 45 \ln(1-p)$$

$$\frac{d}{dp} \ln P(55 \text{ heads} \mid p) = 55 \frac{d}{dp} \ln p + 45 \frac{d}{dp} \ln(1-p)$$

$$= \frac{55}{p} - \frac{45}{1-p} = 0 \Rightarrow \boxed{\hat{p} = 0.55}$$

Continuous distribution:

Normal distribution.

data : $x_1, x_2, \dots, x_n$ ← iid

drawn from a $N(\mu, \sigma^2)$ distribution

μ and σ unknown.

Q: Find the MLE of the unknown pair $(\hat{\mu}, \hat{\sigma})$

estimate we always put a hat.

$(\mu \quad \hat{\mu})$ $(\sigma, \hat{\sigma})$

true value, estimate, true, estimate

let $\underbrace{X_1, X_2, \dots, X_n}_{\text{random variables.}}$ follows a $N(\mu, \sigma^2)$ (independent)

let x_i be the value X_i takes.

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\underbrace{f(x_1 \dots x_n \mid \mu, \sigma)}_{\text{joint pdf.}} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

↑ product

$$= \left(\frac{1}{\quad}\right)^n \exp\left(-\sum^n \frac{(x_i-\mu)^2}{\quad}\right)$$ ← likelihood

$$(\sqrt{2\pi}\sigma) \exp\left(\sum_{i=1} \frac{}{2\sigma^2} \right)$$ function

log-likelihood function:

$$\ln f(x_1 \dots x_n | \mu, \sigma) = -n \ln(2\pi) - n \ln(\sigma) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial \mu} \ln f(x_1 \dots x_n | \mu, \sigma) = \sum_{i=1}^{n} \frac{2(x_i-\mu)}{2\sigma^2}(\ ) = 0$$

$$\hat{\mu} = \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\Rightarrow \hat{\mu} = \sum_{i=1}^{n} x_i - \mu \sum_{i=1}^{n} 1$$

$$\Rightarrow \sum_{i=1}^{n} x_i = n\mu \Rightarrow \boxed{\hat{\mu} = \frac{\sum^n x_i}{n} = \bar{x}}$$

$$\frac{\partial \ln f(x_1 \dots x_n | \mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \sum \frac{(x_i-\mu)^2}{\sigma^3} = 0$$

$$\Rightarrow \boxed{\hat{\sigma}^2 = \frac{\sum (x_i - \mu)^2}{n}}$$

$$\hat{\mu} = \bar{x} \Rightarrow \hat{\mu} = \bar{x}$$: Mean of the data

$$\hat{\sigma}^2 = \sum \frac{1}{n} (x_i - \bar{x})^2$$: variance of the data

(Bias and unbiased estimators)

(μ, σ) unknown $\rightarrow$ Two estimators

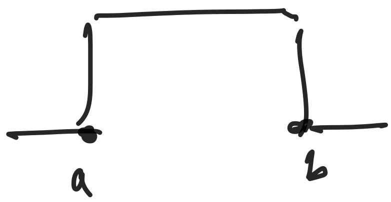
2. (μ is unknown as σ_i is known)
3. (σ_i is unknown and μ is known)

$\Big\}$ one estimator.

Ex: Uniform distribution:

data: $x_1 \dots x_n$.

distribution: $\mathcal{U}(a,b)$



Find the MLE for a & b.

The density of $\mathcal{U}(a,b)$ is $\frac{1}{b-a}$ on $[a,b]$

$$f(x_1 \dots x_n \mid a,b) = \begin{cases} \left(\frac{1}{b-a}\right)^n & \text{if all } x_i \text{ are in } [a,b] \\ 0 & \text{otherwise.} \end{cases}$$

This is maximized by taking $b-a$ as small as possible.

Only rest: $[a,b]$ must include all the data

MLE for the pair (a,b) is

$$\boxed{\hat{a} = \min(x_1 \dots x_n) \qquad \hat{b} = \max(x_1 \dots x_n)}$$

- Why we use the density to find the MLE for continuous distributions?

- Sufficient Statistics.
- Biased / unbiased estimators.