Lecture 18 More On Maximum Likelihood Estimation

BIO210 Biostatistics

Xi Chen

Spring, 2022

School of Life Sciences
Southern University of Science and Technology

南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

MLE For Parameters of Normal Distributions

Practice: Compute the MLE for μ and σ of a normal distribution based on the observation $x_1, x_2, x_3, ..., x_n$.

1. $\theta : \mu, \sigma$
2. $\boldsymbol{\Omega} : \{(\mu, \sigma) \mid \mu \in (-\infty, +\infty), \sigma \geq 0\}$
3. $f_X(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$
4. $\boldsymbol{\mathcal{L}} = f(x_1, x_2, x_3, ..., x_n; \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \cdot e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$

MLE For Parameters of Normal Distributions

$$\ell = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}$$

$$= -\frac{n}{2\sigma^2} \cdot \mu^2 + \frac{\sum_{i=1}^n x_i}{\sigma^2} \cdot \mu - \frac{\sum_{i=1}^n x_i^2}{2\sigma^2} - n \ln \sqrt{2\pi} - n \ln \sigma$$

$$= -n \ln \sigma - \left[\frac{\sum_{i=1}^n (x_i - \mu)^2}{2}\right] \cdot \sigma^{-2} - n \ln \sqrt{2\pi}$$

Let $\dfrac{\partial \ell}{\partial \mu} = 0$

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

Let $\dfrac{\partial \ell}{\partial \sigma} = 0$

$$\Rightarrow \hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

$$\Rightarrow \hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

MLE for σ^2 of Normal Distribution Is Biased

Population: $X \sim \mathcal{N}(\mu, \sigma)$; Sample with size n: $X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma)$

MLE: $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, Note: $\sigma^2 = var(X) = E[X^2] - \mu^2$

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{n-1}{n}\sigma^2$$

Bessel's Correction

Unbiased Variance Estimator

$$\hat{\sigma}^2 = \boxed{\frac{n}{n-1}} \cdot E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Bessel's Correction - Friedrich Bessel

Degree of Freedom

Loosely speaking: degree of freedom is the number of values in the final calculation of a statistic that are free to vary.

$$var(X) = E[(X - E[X])^2] = E[(X - \mu)^2]$$

Think: If you take a sample of size n to estimate the population variance, you first need to get the mean, and subtract the mean from each observation. How many values are free to vary in the following scenario:

$$(X_1 - \mu, X_2 - \mu, X_3 - \mu, ..., X_n - \mu)$$

$$(X_1 - \bar{X}, X_2 - \bar{X}, X_3 - \bar{X}, ..., X_n - \bar{X})$$

Advantages and Disadvantages of MLE

Advantages:

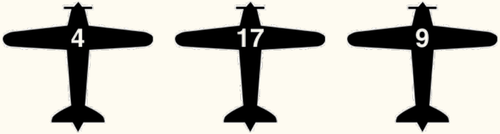
- Intuitive and straightforward to understand.
- If the model is correctly assumed, the MLE is efficient (meaning small variance or mean squared error).

Disadvantages:

- Rely on assumptions of a model (need to know the PMF/PDF).
- Sometimes difficult or impossible to solve the derivate of $\mathcal{L}$ or ℓ.

Example of The Limitation of MLE

Population size: an airline has numbered their planes $1, 2, 3, 4, \cdots, N$. You choose a simple random sample from the N planes, and it turns out as:



Question: What is the MLE for N? $i.e.$ what value should N take to make your observation most likely?

Source: Brilliant.org

Other Estimators

- Minimum-variance unbiased estimator
- Best linear unbiased estimator

Probability vs. Likelihood

$$\mathcal{L}(\theta; x_1, x_2, x_3, ..., x_n) = f(x_1, x_2, x_3, ..., x_n; \theta)$$

the likelihood of the parameter(s) θ taking certain values given that a bunch of data $x_1, x_2, ..., x_n$ are observed.

the probability mass/density of observing the data $x_1, x_2, ..., x_n$ with model parameter(s) θ.