Emory University

MATH 362 Mathematical Statistics II

Learning Notes

Jiuru Lyu

January 23, 2024

Contents

1 Estimation

1.1 Introduction

Definition 1.1.1 (Model). A *model* is a distribution with certain parameters.

Example 1.1.2 The normal distribution: $N(\mu, \sigma^2)$.

Definition 1.1.3 (Population). The *population* is all the objects in the experiment.

Definition 1.1.4 (Data, Sample, and Random Sample). *Data* refers to observed value from sample. The *sample* is a subset of the population. A *random sample* is a sequence of independent, identical $(i.i.d.)$ random variables.

Definition 1.1.5 (Statistics). *Statistics* refers to a function of the random sample.

Example 1.1.6 The sample mean is a function of the sample:

$$\overline{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n).$$

Example 1.1.7 Central Limit Theorem

We randomly toss $n = 200$ fair coins on the table. Calculate, using the central limit theorem, the probability that at least 110 coins have turned on the same side.

$$\overline{X} = \frac{X_1 + \cdots + X_{200}}{200} \quad \overset{\text{CLT}}{\sim} \quad N\left(\mu, \sigma^2\right),$$

where

$$\mu = \mathbf{E}\left(\overline{X}\right) = \frac{\sum_{i=1}^{200} \mathbf{E}(X_i)}{200},$$

$$\sigma^2 = \mathbf{Var}\left(\overline{X}\right) = \mathbf{Var}\left(\frac{X_1 + \cdots + X_{200}}{200}\right) = \frac{\sum_{i=1}^{200} \mathbf{Var}(X_i)}{200^2}.$$

Definition 1.1.8 (Statistical Inference). The process of *statistical inference* is defined to be the process of using data from a sample to gain information about the population.

Example 1.1.9 Goals in statistical inference

1. **Definition 1.1.10 (Estimation).** To obtain values of the parameters from the data.
2. **Definition 1.1.11 (Hypothesis Testing).** To test a conjecture about the parameters.
3. **Definition 1.1.12 (Goodness of Fit).** How well does the data fit a given distribution.
4. Linear Regression

1.2 The Method of Maximum Likelihood and the Method of Moments

Example 1.2.1 Given an unfair coin, or p-coin, such that

$$X = \begin{cases} 1 & \text{head with probability } p, \\ 0 & \text{tail with probability } 1-p. \end{cases}$$

How can we determine the value p?

Solution 1.

1. Try to flip the coin several times, say, three times. Suppose we get HHT.
2. Draw a conclusion from the experiment.

Key idea: The choice of the parameter p should be the value that maximizes the probability of the sample.

$$\mathbf{P}(X_1 = 1, X_2 = 1, X_3 = 0) = \mathbf{P}(X_1 = 1)\mathbf{P}(X_2 = 1)\mathbf{P}(X_3 = 0) = p^2(1-p) := f(p).$$

Solving the optimization problem $\max_{p>0} f(p)$, we find it is most likely that $p = \frac{2}{3}$. This method is called the *likelihood maximization method.* □

Definition 1.2.2 (Likelihood Function). For a random sample of size n from the discrete (or continuous) pdf $p_X(k;\theta)$ (or $f_Y(y;\theta)$), the *likelihood function*, $L(\theta)$, is the product of the pdf evaluated at $X_i = k_i$ (or $Y_i = y_i$). That is,

$$L(\theta) := \prod_{i=1}^{n} p_X(k_i;\theta) \quad \text{or} \quad L(\theta) := \prod_{i=1}^{n} f_Y(y_i;\theta).$$

Definition 1.2.3 (Maximum Likelihood Estimate). Let $L(\theta)$ be as defined in Definition 1.2.2. If θ_e is a value of the parameter such that $L(\theta_e) \geq L(\theta)$ for all possible values of θ, then we call θ_e the *maximum likelihood estimate* for θ.

Theorem 1.2.4 The Method of Maximum Likelihood
Given random samples $X_1, \ldots, X_N$ and a density function $p_X(x)$ (or $f_X(x)$), then we have the likelihood function defined as

$$\begin{aligned} L(\theta) = p_X(X;\theta) &= \mathbf{P}(X_1, X_2, \ldots, X_N) \\ &= \mathbf{P}(X_1)\mathbf{P}(X_2)\cdots\mathbf{P}(X_N) && [independent] \\ &= \prod_{i=1}^{N} p_X(X_i;\theta) && [identical] \end{aligned}$$

Then, the maximum likelihood estimate for θ is given by

$$\theta^* = \arg\max_\theta L(\theta),$$

where

$$L\Big(\arg\max_\theta L(\theta)\Big) = L^*(\theta) = \max_\theta L(\theta).$$

Example 1.2.5 Consider the Poisson distribution $X = 0, 1, \ldots,$ with $\lambda > 0$. Then, the pdf is given by

$$p_X(k,\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots$$

Given data $k_1, \ldots, k_n$, we have the likelihood function

$$L(\lambda) = \prod_{i=1}^{n} p_X(X = k;\lambda) = \prod_{i=1}^{n} e^{-\lambda} = \frac{\lambda^{k_i}}{k_i!} = e^{-n\lambda}\frac{\lambda^{\sum k_i}}{k_1!\cdots k_n!}$$

Then, to find the maximum likelihood estimate of λ, we need to $\max_\lambda L(\lambda)$. That is to solve $\dfrac{\partial L(\lambda)}{\partial \lambda} = 0$ and $\dfrac{\partial^2 L(\lambda)}{\partial \lambda^2} < 0$.