Mathematical Statistics and Data Analysis

Lecture 8: Parameter Estimation

Lyu Ni

DaSE@ECNU
(lni@dase.ecnu.edu.cn)

October 28, 2019

Outlines

Reading Material

Textbook:

- Rice: Chapter 8;
- Mao: Chapter 6;

Point Estimation

Example

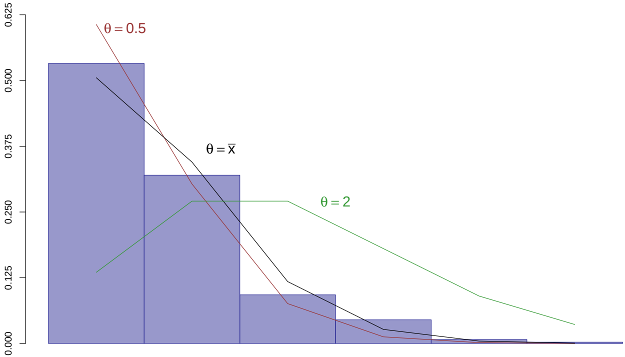
On the Error of Counting with a Haemacytometer(1907) by Student.

- The famous statistician William Gosset, who worked for Guinness brewery, took measure of the number of yeast cells per square in a hemocytometer. The count of yeast cells could be model with a probability distribution known as 'Poisson distribution' $P(\theta)$.
- This distribution $P(\theta)$ has an unknown parameter θ.
- The data is shown as follows:

Containing	0	1	2	3	4	5
Actual	213	128	37	18	3	1

- Problem: What is a guess of θ?

Example (Con'd)

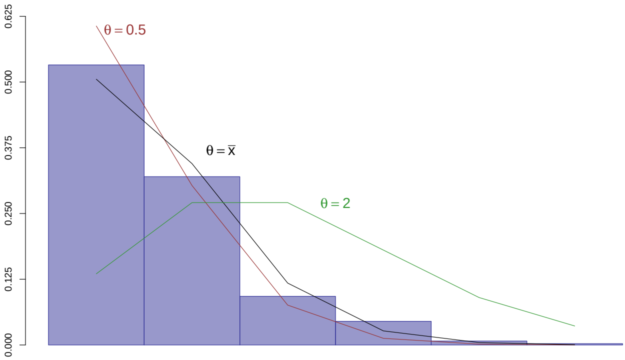


Definition

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from a population with unknown parameter θ. The statistic $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$ is called an **point estimate** of θ.

Point Estimation

Example (Con'd)



Definition

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from a population with unknown parameter θ. The statistic $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$ is called an **point estimate** of θ.

Method of Moments

The kth moment of a random variable X is defined as

$$\mu_k = E(X^k).$$

Suppose that $x_1, x_2, \cdots, x_n$ is a sample. The kth sample moment is defined as

$$a_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

Then, we can view a_k as an estimate of μ_k, and thus let $\hat{\mu}_k = a_k$.

Idea

The method of moments estimates parameters by finding expressions for them in terms of the lowest moments and then substitution sample moments into the expressions.

Method of Moments

- The p.d.f. or p.m.f. of the population is $f(x : \theta_1, \cdots, \theta_k)$;
- $(\theta_1, \cdots, \theta_k) \in \Theta$ is an unknown parameter vector;
- Θ is a parameter space.
- Suppose that the ith moment μ_i exists, $i = 1, 2, \cdots, k$;
- The parameters $\theta_1, \cdots, \theta_k$ can be written as the functions of $\mu_1, \cdots, \mu_k$, that is $\theta_j = \theta_j(\mu_1, \cdots, \mu_k)$;
- The method of moments estimates of θ_j is

$$\hat{\theta}_j = \theta_j(\hat{\mu}_1, \cdots, \hat{\mu}_k), j = 1, \cdots, k$$

- Furthermore, if $\eta = g(\theta_1, \cdots, \theta_k)$ is to be estimated, the method of moment estimate of η is

$$\hat{\eta} = g(\hat{\theta}_1, \cdots, \hat{\theta}_k)$$

Method of Moments

Example: Exponential Distribution

The p.d.f. of an exponential distribution is

$$f(x;\lambda) = \lambda e^{-\lambda x}, x > 0$$

and $x_1, x_2, \cdots, x_n$ is a sample.

- Consider $k = 1$. Since $EX = 1/\lambda$, i.e. $\lambda = 1/EX$, then the method of moment estimate of λ is

$$\hat{\lambda} = 1/\bar{x};$$

- Consider $k = 2$. Since $Var(X) = 1/\lambda^2$, i.e. $\lambda = 1/\sqrt{Var(X)}$, then the moment of method estimate of λ is

$$\hat{\lambda} = 1/s.$$

Method of Moments

Remark

- The method of moment estimate is straight forward.
- The method of moment estimate is **not unique**.
- Problem: Which one is better?

Rule of thumb

The sample moments used in the method of moment should be as **low** as possible.

Method of Moments

Example: Poisson Distribution

The p.d.f. of a Poisson distribution is

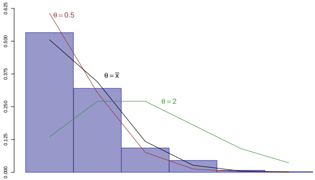
$$f(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}, x = 0, 1, 2, \cdots$$

and $x_1, x_2, \cdots, x_n$ is a sample. Since $E(X) = \lambda$, the method of moment estimate of λ is

$$\lambda = \bar{x}$$

The data are shown as follows:

Containing	0	1	2	3	4	5
Actual	213	128	37	18	3	1



Method of Moments

Example: Uniform Distribution

The p.d.f. of a uniform distribution is

$$f(x;\lambda) = \frac{1}{b-a} I_{(a,b)}(x)$$

with two unknown parameter a and b. Suppose that $x_1, x_2, \cdots, x_n$ is a sample. Since

$$E(X) = \frac{a+b}{2} \quad \text{and} \quad Var(X) = \frac{(b-a)^2}{12},$$

it is obvious that $a = EX - \sqrt{3Var(X)}$ and $b = EX + \sqrt{3Var(X)}$. Thus, the method of moment estimates of a and b are

$$\hat{a} = \bar{x} - \sqrt{3}s \quad \text{and} \quad \hat{b} = \bar{x} + \sqrt{3}s.$$

Method of Maximum Likelihood

Example One

Suppose that it is difficult to distinguish two urns from the appearance. Urn A contains 99 white balls and 1 black ball while Urn B contains 1 white ball and 99 black balls. Here we randomly select an urn and then take a ball. If this ball is a white ball, which urn do you select?

Solution: Let the event

$$A = \{\text{A white ball is taken}\}.$$

- If Urn A is chosen, the probability $P(A) = 0.99$.
- If Urn B is chosen, the probability $P(A) = 0.01$

If A occurs and then we may think that it is likely that this white ball is taken out of Urn A.

Method of Maximum Likelihood

Example Two

We flip a coin and use a random variable X to represent the result. If it heads up, then $X = 1$; otherwise, $X = 0$. Then, X is distributed as a Bernoulli distribution $B(p)$ with a unknown parameter p.

Suppose that $x_1, x_2, \cdots, x_n$ is a sample. The joint p.m.f. of $(x_1, x_2, \cdots, x_n)$ is

$$f(x_1, x_2, \cdots, x_n; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

Since p is unknown, this function could be thought to be a likelihood function of p, denoted as $L(p)$. That is,

$$L(p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}, p \in (0, 1).$$

Method of Maximum Likelihood

Example Two (Con'd)

- How to determine p?
- We would like to choose p so that the probability is as large as possible. Equivalently,

$$\hat{p} = \arg\max_{p} L(p)$$

Then,

$$\frac{\partial \ln L(p)}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0.$$

Thus, the maximum likelihood estimate of p is

$$\hat{p} = \hat{p}(x_1, x_2, \cdots, x_n) = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}.$$

Method of Maximum Likelihood

Definition

Suppose that the p.m.f. or p.d.f. of the population is $p(x;\theta), \theta \in \Theta$, where θ is a unknown parameter (vector) and Θ is the parameter space. Let $x_1, x_2, \cdots, x_n$ be a sample. The joint p.m.f. or p.d.f. of $x_1, x_2, \cdots, x_n$ could be thought to be a function of θ, denoted as $L(\theta; x_1, \cdots, x_n)$ or $L(\theta)$.

- This function $L(\theta)$ is called as the **likelihood function**.
- A statistic $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$ is called **maximum likelihood estimate (MLE)** if this statistic $\hat{\theta}$ satisfies

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$$

Method of Maximum Likelihood

Example: Normal Distribution

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from a normal distribution $N(\mu, \sigma^2)$, where $\theta = (\mu, \sigma^2)$ is a two-dimensional parameter vector. The likelihood function is

$$
\begin{aligned}
L(\mu, \sigma^2) &= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\} \right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\},
\end{aligned}
$$

and its log-likelihood function is

$$
l(\mu, \sigma^2) = \ln L(\mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 - \frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi).
$$

Method of Maximum Likelihood

Example: Normal Distribution (Con'd)

The partials with respect to μ and σ^2 are

$$\begin{aligned}
\frac{\partial(-l)}{\partial\mu} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \\
\frac{\partial(-l)}{\partial\sigma^2} &= -\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2+\frac{n}{2\sigma^2}.
\end{aligned}$$

Setting the first partial equal to zero and solving for the MLE, we obtain

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x} \quad \text{and} \quad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})=s_*^2.$$

Method of Maximum Likelihood

Example: Normal Distribution (Con'd)

The second-order partial deviates are, respectively,

$$\frac{\partial^2(-l)}{\partial\mu^2} = \frac{n}{\sigma^2} \quad \text{and} \quad \frac{\partial^2(-l)}{\partial(\sigma^2)^2} = \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{2\sigma^4}$$

$$\frac{\partial^2 l}{\partial(\sigma^2)\partial\mu} = \frac{\partial^2 l}{\partial\mu\partial(\sigma^2)} = \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu).$$

It is easy to verify the matrix is negative definite since

$$\frac{\partial l}{\partial\mu^2}\Big|_{\mu=\bar{x},\sigma^2=s_*^2} = \frac{n}{s_*^2} > 0$$

$$\left(\frac{\partial l}{\partial\mu^2}\cdot\frac{\partial l}{\partial(\sigma^2)^2} - \left(\frac{\partial l}{\partial(\sigma^2)\partial\mu}\right)^2\right)\Big|_{\mu=\bar{x},\sigma^2=s_*^2} = \frac{n^2}{2s_*^6} > 0$$

Method of Maximum Likelihood

Example: Uniform Distribution

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from a uniform distribution $U(0, \theta)$. Find the maximum likelihood estimate of θ.

Solution: The likelihood function is

$$L(\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{\{0<x_i \le \theta\}} = \frac{1}{\theta^n} I_{\{0<x_{(n)} \le \theta\}}$$

To maximize the likelihood,

- let $I_{\{x_{(n)} \le \theta\}}$ be 1;
- let $1/\theta^n$ be as large as possible.

Since $\frac{1}{\theta^n}$ is decreasing in θ, the maximum likelihood estimate of θ is

$$\hat{\theta} = x_{(n)}$$

Method of Maximum Likelihood

Theorem: Invariance Property

If $\hat{\theta}$ is the MLE of θ, then for any function of $g(\theta)$, the MLE of $g(\theta)$ is $g(\hat{\theta})$.

Example: Normal Distribution (Revisit)

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from $N(\mu, \sigma^2)$. The MLE of μ and σ^2 are respectively

$$\hat{\mu} = \bar{x} \quad \text{and} \quad \hat{\sigma^2} = s_*^2.$$

From the invariance property, find the MLE:

- The standard deviation σ;
- The probability $P(X < 3) = \Phi\left(\frac{3-\mu}{\sigma}\right)$.;
- The 90% quantile $x_{0.90} = \mu + \sigma u_{0.90}$, where $u_{0.90}$ is the 90% quantile of a standard normal r.v.

Method of Maximum Likelihood

Theorem: Invariance Property

If $\hat{\theta}$ is the MLE of θ, then for any function of $g(\theta)$, the MLE of $g(\theta)$ is $g(\hat{\theta})$.

Example: Normal Distribution (Revisit)

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from $N(\mu, \sigma^2)$. The MLE of μ and σ^2 are respectively

$$\hat{\mu} = \bar{x} \quad \text{and} \quad \hat{\sigma^2} = s_*^2.$$

From the invariance property, we have

- The MLE of σ is $\hat{\sigma} = s_*$;
- The MLE of $P(X < 3)$ is $\Phi\left(\frac{3-\bar{x}}{s_*}\right)$;
- The MLE of the 90% quantile $x_{0.90}$ is $\bar{x} + s_* u_{0.90}$.