Lecture 19

2024 Spring

July 18, 2024

Last Lecture

- Estimation and Method of Moments

MM: Step 1: $\begin{cases} \mu_1 = \mu_1(\theta_1 \cdots, \theta_K) \\ \vdots \\ \mu_K = \mu_K(\theta_1 \cdots \theta_K) \end{cases}$

Step 2: $\begin{cases} \theta_1 = \theta_1(\mu_1 \cdots \mu_K) \\ \vdots \\ \theta_K = \theta_K(\mu_1 \cdots \mu_K) \end{cases}$

Step 3. replace $\mu_1 \cdots \mu_K$ with $\hat{\mu}_1 \cdots \hat{\mu}_K$

$$\boxed{\hat{\mu}_j = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i^j}$$

This Lecture

- Maximum Likelihood Method (most commonly used)

Maximum Likelihood Method

Likelihood Function

1. The likelihood function is the joint pdf/pmf evaluated at the realized sample.

$X_1 \cdots X_n$ — sample.

$f(x_1, \cdots, x_n)$ — joint pdf/pmf

After observing the realized sample.

$X_1 = x_1, \cdots X_n = x_n$

we plug in the realized values into the joint pdf/pmf.

$f_n(x_1, x_2, \cdots x_n; \theta)$

evaluate at $(x_1 \cdots x_n)$

parameters that determine the pdf/pmf.

2. Suppose $X_1, \ldots, X_n$ are iid from $f(x; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is unknown. Then, the likelihood function is

$$f_n(x_1, \cdots x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = \begin{cases} \text{discrete case: } \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} Pr(X_i = x_i; \theta) & \text{pmf} \\ \text{continuous case: } \prod_{i=1}^{n} f(x_i; \theta) & \text{pdf.} \end{cases}$$

3. We use $L(\boldsymbol{\theta}; x_1, \ldots, x_n)$ or $L(\boldsymbol{\theta})$ to denote the likelihood function.

Likelihood function is a function of parameter(s) θ.

$L(\theta) = L(\theta; x_1, \cdots x_n)$

Comments:

1. The likelihood function measures how likely we observe the observed data given $\boldsymbol{\theta}$.

2. Smaller $L(\boldsymbol{\theta})$ means $\boldsymbol{\theta}$ is less likely to generate the observed data.

3. Larger $L(\boldsymbol{\theta})$ means $\boldsymbol{\theta}$ is more likely to generate the observed data.

Example: Suppose we have two types of coins; the first is a fair coin, and the second is an unfair coin with the probability of heads being 80%. We were given a coin (not knowing it was fair or unfair), tossed it five times, and got five heads and 0 tails.

"fair / unfair coin" $\longrightarrow$ parameter space $\Theta = \{0.5, 0.8\}$, $p \in \Theta$

"toss the coin five times" $\longrightarrow$ random sample $\{X_1 \cdots X_5\}$

"got five heads" $\longrightarrow$ realized sample $\{X_1 = H, \cdots, X_5 = H\}$

Likelihood function $\binom{5}{x} p^x (1-p)^{n-x}$, based $x=5$. $L(p) = \binom{5}{5} \cdot p^5 = p^5$

$L(0.5) = (0.5)^5 < L(0.8) = (0.8)^5$

$\Rightarrow$ our estimate is $p = 0.8$, unfair coin.

Idea of Maximum Likelihood Method Choose the $\boldsymbol{\theta}$ to maximize the likelihood function $L(\boldsymbol{\theta})$.

1. ML estimate:

$$\hat{\theta} = \hat{\theta}(\underbrace{x_1, \cdots, x_n}_{\text{realized sample}}) = \underset{\theta \in \underbrace{\Theta}_{\text{parameter space}}}{\operatorname{argmax}} L(\theta)$$

2. ML estimator:

$$\hat{\theta} = \hat{\theta}(\underbrace{X_1, \cdots X_n}_{\text{random sample}}) \rightarrow \text{is random variable.}$$

3. Log-likelihood function:

$L(\theta)$ is usually $\prod_{i=1}^{n} f(x_i|\theta)$.

maximizing product is not easy.

$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i|\theta)$.

summation is easier to maximize.

$\log(\cdot)$ is monotone. $\Rightarrow$

$$\hat{\theta} = \hat{\theta}(x_1, \cdots x_n) = \underbrace{\underset{\theta \in \Theta}{\operatorname{argmax}} L(\theta) = \underset{\theta \in \Theta}{\operatorname{argmax}} \ell(\theta)}_{\text{does not change the ML estimate after the log-transformation.}}$$

4. If $\hat{\theta}$ is the ML estimator for θ.
   the $\tau(\hat{\theta})$ is the ML estimator for $\tau(\theta)$

   ↳ $\tau(\cdot)$ is known function

Examples: $X_1 \cdots X_n$ iid $f(x;\theta)$. Find the ML estimator for θ

1. Poisson distribution

$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$, $\ell(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$, $f(x;\theta) = \frac{\theta^x}{x!} e^{-\theta}$

$\Rightarrow \ell(\theta) = \sum_{i=1}^{n} x_i \log\theta - n\theta - \sum_{i=1}^{n} \log(x_i!)$

$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - n = 0$. Solving θ from it. $\Rightarrow \hat{\theta} = \frac{\sum_{i=1}^{n} X_i}{n}$

MM estimator is the same as ML estimator.

2. $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta > 0$.

$L(\theta) = \prod_{i=1}^{n} \theta \cdot x_i^{\theta-1}$.

$\ell(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta) = n\log\theta + (\theta-1)\cdot\sum_{i=1}^{n} \log x_i$

$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i = 0$. $\Rightarrow \hat{\theta}_{ML} = -\frac{n}{\sum_{i=1}^{n} \log X_i}$.

$\hat{\theta}_{MM} = \frac{\bar{X}_n}{1-\bar{X}_n} \neq \hat{\theta}_{ML}$

3. Normal distribution Norm(μ, σ^2)

Likelihood

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i; \mu, \sigma^2)$$

log-likelihood

$$\ell(\mu, \sigma^2) = \log L(\mu, \sigma^2) = \sum_{i=1}^{n} \left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} - \frac{n}{2} \log(2\pi\sigma^2) \right\}$$

$$\begin{cases} \dfrac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = \dfrac{\sum_{i=1}^{n}(x_i - \mu)}{\sigma^2} = 0 & \cdots\cdots \quad (1) \\[2ex] \dfrac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} = \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^4} - \dfrac{n}{2} \dfrac{1}{\sigma^2} = 0 & \cdots\cdots \quad (2) \end{cases}$$

usually, one of the two equation is easy to solve.

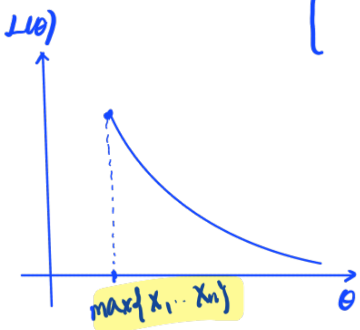
From (1), $\hat{\mu} = \overline{X}_n$, plug in $\hat{\mu}$ in (2) to replace μ.

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$ Same as MM estimator.

4. Unif$(0, \theta)$ The support depends on parameters, don't rush to take derivatives.

$$f(x) = \begin{cases} \frac{1}{\theta} & , x \in (0, \theta] \\ 0 & , \text{otherwise}. \end{cases}$$

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) = \begin{cases} \left(\frac{1}{\theta}\right)^n & , 0 < x_1, \cdots, x_n \leq \theta \quad \theta \geq \max\{X_1 \cdots X_n\} \\ 0 & , \text{ow}. \end{cases}$$



$\hat{\theta} = \max\{X_1 \cdots X_n\}$ is the ML estimator.

5. Let $X_1, \ldots, X_n$ be iid. a random sample from the inverse Gaussian distribution with pdf

$$f(x|\mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad 0 < x < \infty.$$

The log-likelihood is

$$\ell(\lambda, \mu) = \frac{n}{2}\log(\lambda) + n\lambda/\mu - \frac{n}{2}\log(2\pi) - \frac{3}{2}\sum_{i=1}^n \log x_i - \frac{\lambda}{2\mu^2}\sum_{i=1}^n x_i - \frac{\lambda}{2}\sum_{i=1}^n x_i^{-1}.$$

$$\begin{cases} \dfrac{\partial \ell(\lambda,\mu)}{\partial \lambda} = \dfrac{n}{2\lambda} + \dfrac{n}{\mu} - \dfrac{\sum_{i=1}^n x_i}{2\mu^2} - \dfrac{1}{2}\sum_{i=1}^n \dfrac{1}{x_i} = 0 \quad \cdots \quad (1) \\ \dfrac{\partial \ell(\lambda,\mu)}{\partial \mu} = -\dfrac{n\lambda}{\mu^2} + \dfrac{\lambda}{\mu^3}\sum_{i=1}^n x_i = 0 \quad \cdots\cdots \quad (2) \end{cases}$$

By (2). $\Rightarrow \hat{\mu} = \overline{X}_n$, Plug in $\hat{\mu}$ into (1) to replace μ.

$$\Rightarrow \hat{\lambda} = \left(\frac{1}{n}\sum_{i=1}^n X_i^{-1} - \frac{1}{\overline{X}_n}\right)^{-1}$$

Properties of ML Estimator

(1) ML estimator is random variable.

(2) We consider a scalar. θ.

(3) We only consider the case where the support of $x_1 \cdots x_n$ does not depend on θ.

---

1. Score function.

$$S(\theta) = S(\theta; x_1 \cdots x_n) = \frac{d\,l(\theta)}{d\theta} = \frac{d}{d\theta} \log L(\theta).$$

when the support of $x_1 \cdots x_n$ does not depend on θ.

$$S(\hat{\theta}) = 0.$$

ML estimate.

2. Information function.

$$I(\theta) = I(\theta; x_1 \cdots x_n) = -\frac{d\,S(\theta)}{d\theta} = -\frac{d^2 l(\theta)}{d\theta^2} = -\frac{d^2}{d\theta^2} \log L(\theta).$$

both depends on realized sample. $x_1 \cdots x_n$.

3. Fisher Information (Matrix) / Expected information.

$I(\theta; x_1 \cdots x_n)$ is based on realized sample $x_1 \cdots x_n$

Replace $x_1 \cdots x_n$ with $X_1, \cdots X_n$, $I(\theta; X_1, \cdots X_n)$

The FI is defined as.

$$J(\theta) = E\{I(\theta; X_1, \cdots X_n)\}$$

(mean of the information function).

If $X_1 \cdots X_n$ are iid. with $f(x;\theta)$.

$$J(\theta) = E\left\{\sum_{i=1}^{n} -\frac{d^2 \log f(X_i;\theta)}{d\theta^2}\right\} = \sum_{i=1}^{n} -E\left\{\frac{d^2 \log f(X_i;\theta)}{d\theta^2}\right\}$$

$$= -n\,E\left\{\frac{d^2 \log f(X_1;\theta)}{d\theta^2}\right\}$$