Slide 6: Estimation Methods

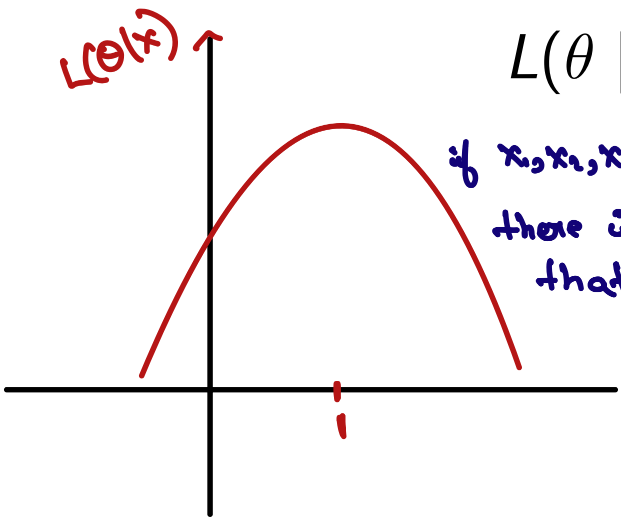
STATS 511: Statistical Inference

Kean Ming Tan

Some Definitions

Likelihood Function: Let $X_1, \ldots, X_n$ be a random sample from f_θ. The likelihood function is defined as

$x_1, x_2, x_3 \sim N(u, 1)$



$$L(\theta \mid \mathbf{x}) = f_n(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta).$$

if x_1, x_2, x_3 are close to 1
there is high chance
that $\mu = 1$

Maximum Likelihood Estimator: For each possible observed vector $\mathbf{x} = (x_1, \ldots, x_n)$, let $\hat{\theta} = \delta(\mathbf{x})$ denote a value of $\theta \in \Theta$ for which the likelihood function $L(\theta \mid \mathbf{x})$ is maximized. The value $\hat{\theta} = \delta(\mathbf{x})$ is called a **maximum likelihood estimator** of θ.

Test for a Disease

Suppose that a medical test is 90 percent reliable in the following sense: If a person has the disease, there is a probability of 0.9 that the test will give a positive response; whereas, if a person does not have the disease, there is a probability of only 0.1 that the test will give a positive response. Let X be the results of the test, i.e.,

$$X = \begin{cases} 1 & \text{if the test is positive} \\ 0 & \text{if the test is negative} \end{cases}$$

Let the parameter space $\Theta = \{0.1, 0.9\}$, where $\theta = 0.1$ means that the person tested does not have the disease, and $\theta = 0.9$ means that the person has the disease. So in this case, X is Bernoulli with parameter θ. What is the MLE of θ?

Test for a Disease MLE

Example: Picking Balls

Suppose that we have a box with a certain number of black and white balls. Denote the proportion of black balls by p. Suppose we sample 4 balls (with replacement) and we obtain the sequence of balls: black, white, black, black. What is the MLE for p?

$X=$ # of Black BAll's out of 4 BAll's

$X \sim \text{Binomial}(n=4, p)$

$\mathbb{P}(X=3) = \binom{4}{3} p^3 (1-p)$

Likelihood function: $L(p|x) = \binom{n}{x} p^x (1-p)^{n-x}$

MLE: $\hat{p} = \underset{p}{\operatorname{argmax}}\, L(p|x)$

$= \underset{p}{\operatorname{argmax}}\, \log L(p|x)$

$\ell(p) = \log \binom{n}{x} + x \log p + (n-x) \log(1-p)$

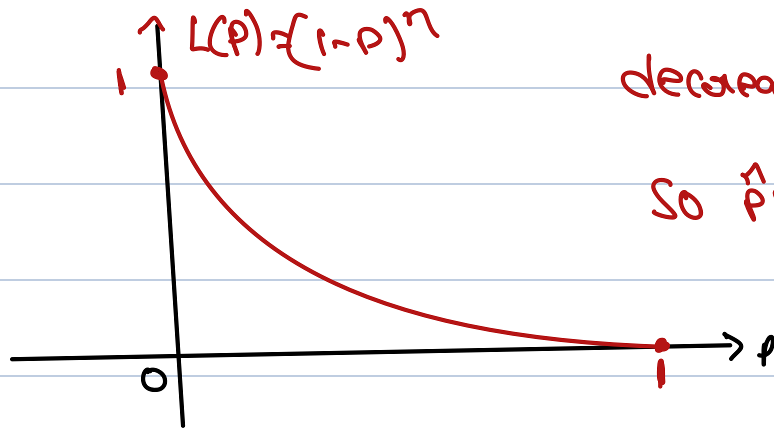
$$\frac{dl}{dp} = \frac{x}{p} - \frac{n-x}{1-p} = 0 \implies \hat{p} = \frac{x}{n}$$

what if $x = 0$?

$$l(p) = \log\binom{n}{0} + n\log(1-p)$$

$$\frac{dl}{dp} = \frac{-n}{1-p}$$ we cannot solve.

when $x=0$ $L(p) = (1-p)^n$



decreasing function of p

So $\hat{p} = 0$ is the MLE

Same logic for $x=n$ case.

Example: Sampling from a Normal Distribution

Suppose that the random variables $X_1, \ldots, X_n$ are sampled from the Normal distribution with unknown mean μ and unknown variance σ^2. Therefore, in this case, our parameter of interest is $\theta = (\mu, \sigma^2) \in \Theta$ where $\Theta = \mathbb{R} \times \mathbb{R}^+$. Suppose that we have the observed values $x_1, \ldots, x_n$. What is the MLE of θ?

$x_1, x_2, \ldots, x_n \sim N(\mu, \sigma^2)$

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$\Rightarrow \quad l(\theta|x) = \log L(\theta|x) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial l(\theta|x)}{\partial \mu} = -\frac{1}{2\sigma^2} \cdot 2\sum_{i=1}^{n}(x_i-\mu) = 0$$

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

$$\Rightarrow \quad \mu = \frac{1}{n} \sum x_i = \bar{x}$$

$$\Rightarrow \quad \boxed{\hat{\mu} = \bar{x}}$$

$$\frac{\partial l}{\partial \sigma^2} = \frac{-n}{2} \cdot \frac{1}{2\pi\sigma^2} \cdot 2\pi - \frac{1}{2} \sum (x_i - \mu)^2 \cdot \frac{-2}{\sigma^4}$$

$$\Rightarrow \quad \frac{-n}{2\sigma^2} + \frac{1}{\sigma^4} \sum (x_i - \mu)^2$$

$$\Rightarrow \quad \tilde{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

MLE for an Exponential Family

Suppose that the random variables $X_1, \ldots, X_n$ are sampled from an exponential family with density function, i.e., the density function can be rewritten as

$$f(\mathbf{x} \mid \boldsymbol{\theta}) = h(\mathbf{x})c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(\mathbf{x})\right)$$

Assume that this exponential family is full rank. How do we find the MLE for general exponential family distirbution?

Sampling from a Uniform Distribution

Suppose that the random variables $X_1, \ldots, X_n$ are sampled from the Uniform distribution on the interval $[0, \theta]$, where $\theta \in \Theta$ is unknown and that $\Theta \in \mathbb{R}^+$. What is the MLE of θ?

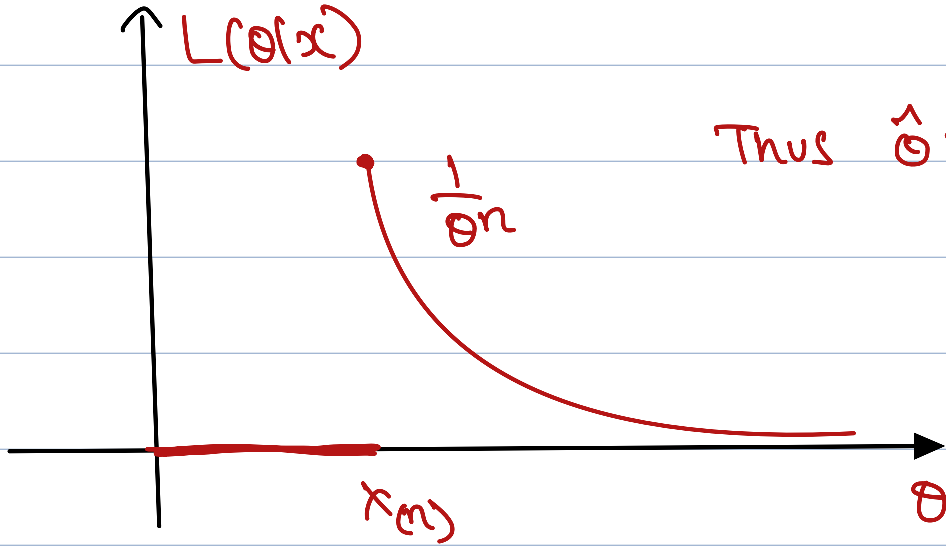
$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta \\ 0 & \text{o.w} \end{cases}$$

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } 0 \le x_i \le \theta \ \forall i \\ 0 & \text{o.w} \end{cases}$$

$$\ell(\theta|x) = -n \log \theta \Rightarrow \frac{d\ell}{d\theta} = \frac{-n}{\theta} \quad (\theta \text{ cannot be } 0)$$

so we cannot use traditional MLE

$$L(\theta|x) = \frac{1}{\theta^n} I(x_{(n)} \leq \theta)$$



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

Example: MLE with Constrained Parameter Space

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. When the parameters are not constrained, we have

$$\hat{\mu} = \bar{X}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

What if the parameters are constrained, such as $\mu \geq 0$.

$$l(\theta) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2 - \frac{n}{2}\log(2\pi)$$

for μ, $\quad \underset{\mu}{\text{maximize}} \; -\frac{1}{2\sigma^2}\sum(x_i-\mu)^2$

$$\Downarrow$$

$\underset{\mu}{\text{maximize}} \; -\frac{1}{2\sigma^2}\left(\sum x_i^2 + n\mu^2 - 2\mu\bar{x}\right)$

$$\underset{\mu}{\text{maximize}} \quad -\frac{\sum x_i^2}{2\sigma^2} + \frac{2\mu\bar{x} - n\mu^2}{2\sigma^2}$$

$$\Rightarrow \quad \underset{\mu}{\text{maximize}} \quad -n\mu^2 + 2n\mu\bar{x}$$

if $\bar{x} > 0$ then $\tilde{\mu} = \bar{x}$

if $\bar{x} < 0$ then $\underset{\mu}{\text{maximize}} \quad \underbrace{-n\mu^2 + 2n\mu\bar{x}}_{\text{decreasing in } \mu}$

So, therefore the max occurs at $\hat{\mu} = 0$

Thus, $\hat{\mu} = \begin{cases} \bar{x} & \text{if } \bar{x} \geq 0 \\ 0 & \text{if } \bar{x} < 0 \end{cases}$

Limitations of MLE I

Nonexistence of an MLE: Suppose that the random variables $X_1, \ldots, X_n$ are sampled from the Uniform distribution on the interval $(0, \theta)$, where $\theta \in \Theta$ is unknown and that $\Theta \in \mathbb{R}^+$. We write the density function as

$$f(x \mid \theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

What is the MLE of θ?

we know that $X_{(n)}$ is MLE if $x_1, x_2, \ldots, x_n \sim \text{unif}(0, \theta]$

Pick $\hat{\theta} = X_{(n)}$, as our "MLE" of θ for unif$(0,\theta)$

in this case $0 < x_1, x_2, \ldots, x_n < \theta$

Pick $\tilde{\theta} = x_{(n)} + \varepsilon$ as my MLE,

But I can pick $\hat{\theta} = x_{(n)} + \frac{\varepsilon}{2}$ has higher $L(\theta)$

$\Rightarrow$ there is no MLE for this.

Limitations of MLE II

Non-uniqueness of an MLE: Suppose that the random variables $X_1, \ldots, X_n$ are sampled from the Uniform distribution on the interval $[\theta, \theta+1]$, where $\theta \in \Theta$ is unknown and that $\Theta \in \mathbb{R}$. What is the MLE of θ? $x_1, x_2, \ldots, x_n \sim \text{unif}[\theta, \theta+1]$

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \begin{cases} 1 & \text{if } \theta \le x_i \le \theta+1 \quad \forall i \\ 0 & \text{o.w} \end{cases}$$

$$= I(x_{(1)} \geq \theta) \cdot I(x_{(n)} \le \theta+1)$$

$$= I(x_{(n)}-1 \le \theta \le x_{(1)})$$

$$L(\theta) = \begin{cases} 1 & \text{if } x_{(n)}-1 \le \theta \le x_{(1)} \\ 0 & \text{o.w} \end{cases}$$

MLE is any value of θ b/w $x_{(n)}-1$ and $x_{(1)}$

$\Rightarrow$ MLE is not unique.

Invariance Property of MLE

Theorem: Let $\hat{\theta}$ be the MLE of θ and let $g(\theta)$ be an arbitrary function of the parameter. The MLE of $g(\theta)$ is $g(\hat{\theta})$.

Examples:

- Let $\hat{\mu}$ be the MLE of μ. Then $\sqrt{\hat{\mu}}$ is an MLE of $\sqrt{\mu}$.
- Let $\hat{\mu}$ be the MLE of μ. Then μ^2 is an MLE of μ^2.

Example

Suppose that $X_1, \ldots, X_n$ form a random sample from the normal distribution with both the mean μ and the variance σ^2 are unknown. What is the MLE of the standard deviation σ and $E(X^2)$?

$E[X^2] = \sigma^2 + \mu^2$

Numerical Computation

Sometimes, there are no closed form solution for the MLE. Here is an example: Suppose that $X_1, \ldots, X_n$ form a sample of Gamma distribution for which pdf is as follows:

$$f(x \mid \alpha) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \qquad \text{for} \qquad x > 0.$$

What is the MLE of α?

Newton's Method

Newton's Method: Let $f(\theta)$ be a real-valued function of a real variable, and suppose that we wish to solve the equation $f(\theta) = 0$. Let θ_0 be an initial guess of the solution. Newton's method replaces the initial guess with the updated guess

$$\theta_1 = \theta_0 - \frac{f(\theta_0)}{f'(\theta_0)}.$$

This process is done until the solution is stabilize.

Summary of MLE

The maximum likelihood estimate of a parameter θ is that value of θ that provides the largest value of the likelihood function $L(\theta \mid \mathbf{x})$ for fixed data x.

Up Next - Bayes Estimator