Lecture 6 :

- Recap (the method of moments):

$$X_1, \dots, X_n \overset{iid}{\sim} \underbrace{f(x \mid \theta_1, \dots, \theta_m)}_{\text{PDF that generates the data}}$$

- If $\mu_k = E[X^k]$

$$\boxed{\hat{\mu}_k(X_1, \dots, X_n) = \frac{\sum_{i=1}^{n} X_i^k}{n}}$$

unbiased estimator for μ_k

- So far we know how to estimate the moments of the underlying distribution from data.

- $X_1, \dots, X_n \overset{iid}{\sim} f(x \mid \theta_1, \dots, \theta_m)$

- $\mu_1 = E[X] = \int_{-\infty}^{\infty} x f(x \mid \theta_1, \dots, \theta_m)\, dx$

$$= g_1(\theta_1, \dots, \theta_m)$$

- $\mu_2 = E[X^2] = \int x^2 f(x \mid \theta_1, \dots, \theta_m)\, dx$

$$= g_2(\theta_1, \dots, \theta_m)$$

$\vdots$

- $\mu_i = E[X^i] = g_i(\theta_1, \dots, \theta_m)$

$\vdots$

- $\mu_m = E[X^m] = g_m(\theta_1, \dots, \theta_m)$

We want to learn the values of $\theta_1, \dots, \theta_m$.

$$\begin{cases} \mu_1 = g_1(\theta_1, \dots, \theta_m) \\ \mu_2 = g_2(\theta_1, \dots, \theta_m) \\ \vdots \\ \mu_m = g_m(\theta_1, \dots, \theta_m) \end{cases} \xrightarrow{\text{inversion}} \begin{cases} \theta_1 = h_1(\mu_1, \dots, \mu_m) \\ \vdots \\ \theta_m = h_m(\mu_1, \dots, \mu_m) \end{cases} \quad (*)$$

Since we know how ~~tto~~ to estimate μ_k's from data, i.e.

$$\hat{\mu}_k = \frac{\sum_{i=1}^{n} x_i^k}{n}$$

then we can just plug these estimates into the set of equations $(*)$ to find estimates of the parameters θ_k's.

Example:

$- \; x_1, \dots, x_n \sim N(\mu, \sigma^2) =$

$$= f(x \mid \underset{\theta_1}{\mu}, \underset{\theta_2}{\sigma^2}) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2}}$$

$$\begin{cases} \mu_1 = E[X] = \mu \triangleq \theta_1 \\ \mu_2 = E[X^2] = \mu^2 + \sigma^2 \triangleq \theta_1^2 + \theta_2 \end{cases}$$

$$\left(\begin{aligned} & \mathrm{Var}(X) \\ & = E[X^2] - (E[X])^2 \\ & = \mu_2 - \mu_1^2 \end{aligned} \right)$$

$$\left(\begin{aligned} & \sigma^2 = \mu_2 - \mu^2 \\ & \Rightarrow \mu_2 = \sigma^2 + \mu^2 \end{aligned} \right)$$

$$\begin{cases} \mu_1 = \underbrace{\theta_1}_{g_1(\theta_1, \theta_2)} \\ \mu_2 = \underbrace{\theta_1^2 + \theta_2}_{g_2(\theta_1, \theta_2)} \end{cases}$$

$$\Downarrow$$

$$\theta_1 = \mu_1$$

$$\theta_2 = \mu_2 - \theta_1^2 = \mu_2 - \mu_1^2$$

$$\begin{cases} \theta_1 = \mu_1 \to h_1(\mu_1, \mu_2) \\ \theta_2 = \mu_2 - \mu_1^2 \to h_2(\mu_1, \mu_2) \end{cases}$$

$$\begin{cases} \theta_1 = \mu_1 \\ \theta_2 = \mu_2 - \mu_1^2 \end{cases}$$

$$\boxed{\begin{aligned} \hat{\mu}_1 &= \bar{X} \\ \hat{\mu}_2 &= \frac{X_1^2 + \cdots + X_n^2}{n} \end{aligned}}$$

$X_1, \ldots, X_n \sim f(x \mid \mu, \sigma^2) = N(\underbrace{\mu}_{\theta_1}, \underbrace{\sigma^2}_{\theta_2})$

$$\hat{\mu} \stackrel{\Delta}{=} \hat{\theta}_1 = \bar{X}$$

$$\hat{\sigma}^2 = \hat{\theta}_2 = \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{X_1^2 + \cdots + X_n^2}{n} - \bar{X}^2$$

$$= \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

Examples

$X_1, \ldots, X_n \sim \text{gamma}(\alpha, \beta)$

$$\begin{cases} \mu_1 = \underbrace{\alpha/\beta}_{g_1(\alpha,\beta)} \\ \mu_2 = \underbrace{\dfrac{\alpha(\alpha+1)}{\beta^2}}_{g_2(\alpha,\beta)} \end{cases}$$

$$\underbrace{\begin{cases} \mu_1 = \alpha/\beta \\ \mu_2 = \dfrac{\alpha^2}{\beta^2} + \dfrac{\alpha}{\beta^2} = \mu_1^2 + \dfrac{\alpha}{\beta^2} \end{cases}}$$

$$\Downarrow$$

$$\begin{cases} \dfrac{\alpha}{\beta^2} = \mu_2 - \mu_1^2 \\ \dfrac{\alpha}{\beta} = \mu_1 \end{cases} \Rightarrow \frac{1}{\beta} = \frac{\mu_2 - \mu_1^2}{\mu_1}$$

$$\Rightarrow \boxed{\beta = \frac{\mu_1}{\mu_2 - \mu_1^2}}$$

$$\mu_1 = \alpha/\beta \Rightarrow \alpha = \mu_1 \beta \Rightarrow \alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$$

$$\begin{cases} \alpha = \dfrac{\mu_1^2}{\mu_2 - \mu_1^2} \longrightarrow h_1(\mu_1, \mu_2) \\[2ex] \beta = \dfrac{\mu_1}{\mu_2 - \mu_1^2} \longrightarrow h_2(\mu_1, \mu_2) \end{cases}$$

$$\begin{cases} \hat{\mu}_1 = \bar{X} \\[2ex] \hat{\mu}_2 = \dfrac{X_1^2 + \cdots + X_n^2}{n} \end{cases}$$

$$\Rightarrow \begin{cases} \hat{\alpha} = \dfrac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2} \\[2ex] \hat{\beta} = \dfrac{\hat{\mu}_1}{\hat{\mu}_2 - \hat{\mu}_1^2} \end{cases}$$

The method of Maximum Likelihood (ML)

$$X_1, \dots, X_n \overset{iid}{\sim} f(x \mid \underbrace{\theta_1, \dots, \theta_m}_{\theta_0})$$

The main idea behind the method of Maximum Likelihood (ML) is to formulate and analyse the so-called likelihood function;

Given the sample data $x_1, \dots, x_m$ the likelihood function $lik(\theta)$ is defined as:

$$lik(\theta) = f(\overbrace{x_1, \dots, x_n}^{\text{we observe } x_1, \dots, x_n \text{ as our sample data}} \mid \theta)$$

$$\overset{X_i = iid}{=} \overline{\prod_{i=1}^{n} f(x_i \mid \theta)}$$

$$f(x_1, \ldots, x_n \mid \theta) = ?$$

$$X_i \sim f(x \mid \theta) \longrightarrow \underbrace{f(x_i \mid \theta)}$$

the likelihood of the event $X_i = x_i$ assuming the PDF is $f(x \mid \theta)$

Example:

$$X_i \sim N(\mu, \sigma^2)$$

$$f_{X_i}(x_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Assume we observe: $x_i = 3.17$

note that: $$f(3.17 \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(3.17-\mu)^2}{2\sigma^2}}$$

Consider two cases for the parameters:

(1) $\mu = 0, \sigma = 1 \longrightarrow f(\overset{\text{data}}{3.17} \mid \mu=0, \sigma=1) = 0.5$

(2) $\mu = 0, \sigma = 2 \longrightarrow f(3.17 \mid \mu=0, \sigma=2) = 0.25$

since $0.5 > 0.25$, Case (1) is more likely to have 3.17 as outcome compared to (2).

$X_1 \sim N(\mu, \sigma^2) \longrightarrow f(x_1 | \mu, \sigma^2)$

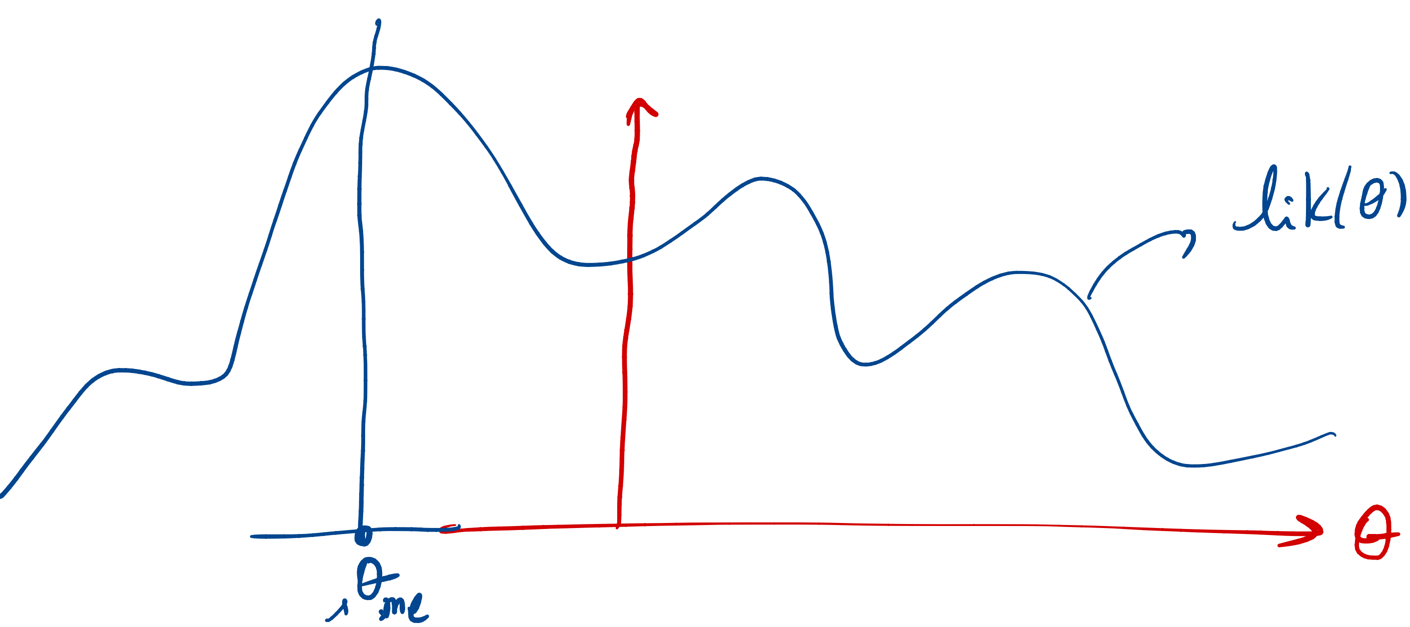
$X_2 \sim N(\mu, \sigma^2) \longrightarrow f(x_2 | \mu, \sigma^2)$

$\vdots \qquad\qquad\qquad \vdots$

$X_n \sim N(\mu, \sigma^2) \longrightarrow f(x_n | \mu, \sigma^2)$

$$f(x_1, \ldots, x_n | \mu, \sigma^2) = f(x_1 | \mu, \sigma^2) \times f(x_2 | \mu, \sigma^2) \times \cdots \times f(x_n | \mu, \sigma^2)$$

$$= \underbrace{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}}}_{lik(\mu, \sigma^2)}$$



$$\theta_{ml} = \underset{\theta}{\operatorname{argmax}}\ lik(\theta) = \underset{\theta}{\operatorname{argmax}}\ f(x_1, \ldots, x_n \mid \theta)$$

$$= \underset{\theta}{\operatorname{argmax}}\ \prod_{i=1}^{n} f(x_i \mid \theta)$$

Compared to all the other choices of θ, θ_{ml} has the largest likelihood in generating the sample data $x_1, \ldots, x_n$.

$$\boxed{f = PDF \overset{\text{continuous}}{=} \text{derivative of CDF}}$$

Using $\log(\cdot)$ without loss of generality: $g > 0$

$$\underset{\theta}{\operatorname{argmax}}\ g(\theta)$$

$$= \underset{\theta}{\operatorname{argmax}}\ \log(g(\theta))$$

because $\log(\cdot)$ is a strictly increasing function

$$\theta_{ml} = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} f(x_i|\theta)$$

$$= \underset{\theta}{\operatorname{argmax}} \log\left(\prod_{i=1}^{n} f(x_i|\theta)\right)$$

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log(f(x_i|\theta))$$

Example: $X_1, \dots, X_n \sim \text{exponential}(\underline{\lambda})$

$$f(x|\lambda) = \lambda e^{-\lambda x} \, \underline{\mathbb{1}\{x \geq 0\}}$$

$$\mathbb{1}\{x \geq 0\} = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Data: $x_1, \dots, x_n \longrightarrow$ find $\hat{\lambda}_{ml}$?

$$\hat{\lambda}_{ml} = \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \log(f(x_i|\lambda))$$

$$= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \log(\lambda e^{-\lambda x_i})$$

$$= \underset{\lambda}{\operatorname{argmax}} \left\{ \underbrace{n \log \lambda - \lambda \sum_{i=1}^{n} x_i} \right\}$$

take the derivative and set it to zero

$$\frac{d}{d\lambda}\left(n \log \lambda - \lambda \sum_{i=1}^{n} x_i \right) = 0$$

$$= n/\lambda - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\lambda}_{ml} = \frac{n}{\sum_{i=1}^{n} x_i}$$

$$= \frac{1}{\bar{X}}$$

$$\Rightarrow \boxed{\hat{\lambda}_{ml} = \frac{1}{\bar{X}}}$$

is $\hat{\lambda}_{ml}$ biased / unbiased?

$$\left\{ \begin{array}{l} \hat{\lambda}_{ml} = \dfrac{1}{\bar{X}} \\ E[\bar{X}] = \dfrac{1}{\lambda} \end{array} \right.$$

(we know that the mean of the exponential distribution, $X \sim \text{exponential}(\lambda)$ is $E[X] = \frac{1}{\lambda}$)

$$\boxed{\begin{array}{c} \text{if } E[X] = \alpha \\ \not\Downarrow \\ E[g(X)] = g(\alpha) \end{array}}$$