Method of Maximum Likelihood

Likelihood function of data: Joint pdf or pmf of sample data considered as a function of θ with data held fixed at the observed values $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

$$L(\theta) = L(\theta; x_1, x_2, , x_n) \underset{\text{indep.}}{=} \prod_{i=1}^{n} f_\theta(x_i)$$

joint pdf/pmf: θ is fixed, think of this a fn. of $x_1, \ldots, x_n$

Likelihood fn: $x_1, \ldots, x_n$ fixed, think of this as a fn. of θ.

- A function of θ — the data are held fixed.

Maximum likelihood estimator (MLE) of θ: The value $\hat{\theta}$ of θ that maximizes the likelihood function as a function of θ.

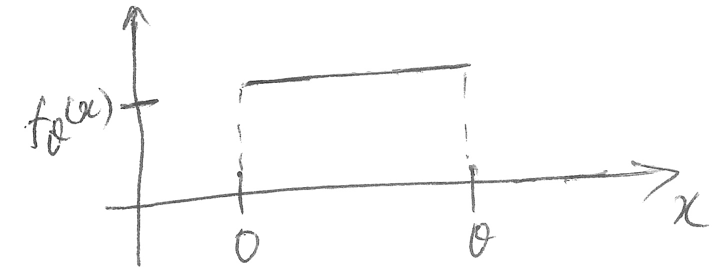
- Can think of MLE as the value of θ that is "most likely" to have led to the observed data.
- Essentially a calculus problem.

How to find MLE?

Direct approach: Directly maximize the likelihood function.

Ex: Let $X_1, X_2, \ldots, X_n$ represent a random sample from a Uniform $(0, \theta)$ distribution where $\theta > 0$. Find the MLE of θ.

Recall: $f_\theta(x) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & \text{o/w} \end{cases}$



PDF of X, which represents the population

Clearly, θ = max. in the population.

will see that: $\hat{\theta}$ = max. in the sample is MLE of θ.

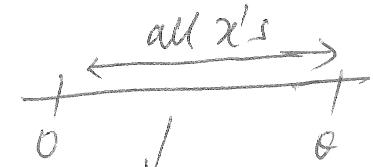
Note: $E[X] = \frac{\theta}{2} \Rightarrow \theta = 2E[X]$ — estimate this by $\bar{X}$

$\Rightarrow \hat{\theta} = 2\bar{X}$ can also be used as an estimator of θ.

Have two estimators of θ: $\boxed{\begin{aligned} \hat{\theta}_{MLE} &= X_{(n)} \\ \hat{\theta}_{MOME} &= 2\bar{X} \end{aligned}}$ which one is better?

By def: $L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \frac{1}{\theta} \cdot \frac{1}{\theta} \cdots \frac{1}{\theta}$, $0 \le x_1, x_2, \ldots x_n \le \theta$

$= \left(\frac{1}{\theta}\right)^n$

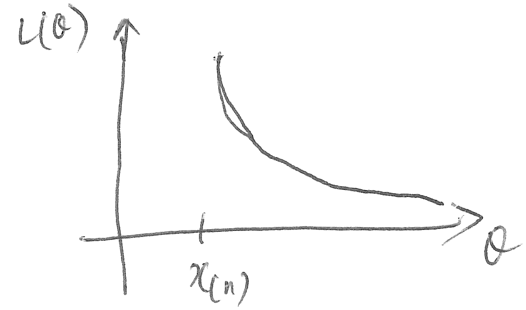


$\Rightarrow L(\theta) = \frac{1}{\theta^n}$, $x_{(n)} \le \theta$

$0 \le x_{(1)}$, $x_{(n)} \le \theta$

sample min. sample max.

always holds



$\Rightarrow$ Since $L(\theta)$ is decreasing in θ, it is maximized at the smallest value of θ, which is $x_{(n)}$

$$\Rightarrow \hat{\theta}_{MLE} = \underbrace{X_{(n)}}_{\text{sample max.}}$$

Ex: Fish length (in) example.
$n=5$ $x_1 = 12.3$, $\overset{x_2}{7.2}$, $\overset{x_3}{23.9}$, $\overset{x_4}{15.8}$, $\overset{x_5}{9.6}$

$\Rightarrow \hat{\theta}_{MLE} = X_{(n)} = 23.9$ in.

Imp note: log in this class means "natural" log.

Differentiation technique: Maximize the log-likelihood function $\log\{L(\theta)\}$ with respect to θ instead of $L(\theta)$ as the former tends to be easier. The value of θ that maximizes $L(\theta)$ also maximizes $\log\{L(\theta)\}$. (**Why?**) because $\log[L(\theta)]$ is increasing in θ.

Step 1: Set up the log-likelihood function.

$$\Rightarrow \log\{L(\theta)\} = \log\left\{\prod_{i=1}^{n} f_\theta(x_i)\right\} = \sum_{i=1}^{n} \log\{f_\theta(x_i)\}$$

Step 2: Find the *likelihood equation* by partially differentiating $\log\{L(\theta)\}$ with respect to θ and setting the derivative to equal to zero.

$$\frac{\partial}{\partial\theta} \log[L(\theta)] = 0$$

Step 3: Solve the likelihood equation for θ. The solution is MLE if it is a point of maxima (no need to verify).

solve it

Recall: Some useful properties of natural log:

- $\log(ab) = \log(a) + \log(b)$
- $\log(a^b) = b\log(a)$
- $\log(e^a) = a$

Ex: Let $X_1, X_2, \ldots, X_n$ represent a random sample from an Exponential (λ) distribution where $\lambda > 0$. Find the MLE of λ.

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

All x_i's are going to be > 0

$$L(\lambda) = \prod_{i=1}^{n} f_\lambda(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} = \lambda^n e^{-n\lambda\bar{x}}$$

Recall: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow \sum_{i=1}^{n} x_i = n\bar{x}$

$$\Rightarrow \log[L(\lambda)] = \log[\lambda^n e^{-n\lambda\bar{x}}] = n\log(\lambda) - n\lambda\bar{x}$$

$$\Rightarrow 0 = \frac{\partial}{\partial\lambda} = \frac{n}{\lambda} - n\bar{x} \Rightarrow \lambda = \frac{1}{\bar{x}} \Rightarrow \hat{\lambda}_{MLE} = \frac{1}{\bar{X}}$$

Method of moments: $d = 1$ $\quad \bar{X} = E[X] = \frac{1}{\lambda}$

$$\Rightarrow \hat{\lambda}_{MOME} = \frac{1}{\bar{X}} = \hat{\lambda}_{MLE}$$

Method of moments:

Population moments: $E[X^k]$ = k-th population moment, $k = 1, 2,$

→ $k=1$: $E[X] = \mu$ ← population

→ $k=2$: $E[X^2] = \mu^2 + \sigma^2$.

Recall: $Var[X] = E[X^2] - [E(X)]^2$

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

Sample moments:

$\frac{1}{n}\sum_{i=1}^{n} X_i^k$ = kth sample moment, $k = 1, 2, 3$

→ $k=1$: $\bar{X}$ = first sample moment

Method of moments: Let d = # unknown parameters in the model

set up d equations: $\frac{1}{n}\sum_{i=1}^{n} X_i^k = \underbrace{E[X^k]}_{\text{these involve the unknown } \theta}$, $k = 1, 2, \ldots, d$.

solve of θ; solution: $\hat{\theta}_{MOME}$

Uniform example: $d = 1$

Set up: $\bar{X} = \frac{\theta}{2} \Rightarrow \theta = \boxed{2\bar{X} = \hat{\theta}_{MOME}}$

Ex: Suppose $X_1, X_2, \ldots, X_n$ denote a random sample from a Bernoulli (p) distribution, where p is unknown. Find its MLE.

Method of moments: $d = 1$

$$\bar{X} = E[X] = p$$

$$\Rightarrow \hat{p}_{MOME} = \bar{X} = \hat{p}$$

$$f_p(x) = \begin{cases} p^x (1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

Recall: "Natural" estimator of p is $\hat{p} = \bar{X} =$ prop. of 1's in the sample

Method of ML:

$$L(p) = \prod_{i=1}^{n} f_p(x_i) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\sum_i x_i} (1-p)^{n - \sum_i x_i}$$

$$\boxed{\sum_i x_i = n\bar{x}}$$

$$= p^{n\bar{x}} (1-p)^{n - n\bar{x}}$$

$$\Rightarrow \log[L(p)] = (n\bar{x}) \log(p) + (n - n\bar{x}) \log(1-p)$$

$$\Rightarrow 0 = \frac{\partial}{\partial p} \log[L(p)] = \frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p}$$

$$\Rightarrow p = \bar{x} \quad \Rightarrow \quad \hat{p}_{MLE} = \bar{X} = \hat{p}$$

verify

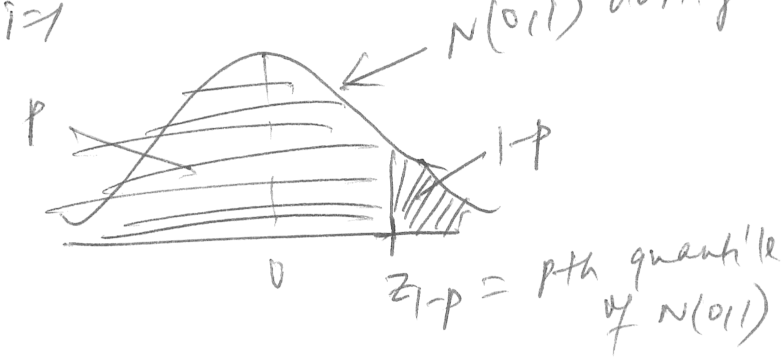
Since $p \in [0,1]$, need to check on the boundaries. Even then will see that $\hat{p}_{MLE} = \hat{p}$.

Invariance property of maximum likelihood estimators:

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

$g(\hat{\theta})$ is MLE of $g(\theta)$.

Ex: $X \sim N[\mu, \sigma^2]$ Then:
$$\left.\begin{aligned} \hat{\mu}_{MLE} &= \bar{X} \\ \hat{\sigma}^2_{MLE} &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 \end{aligned}\right\} \text{ see HW.}$$



pth quantile of X: $\mu + z_{1-p}\sigma$

$g(\mu, \sigma^2) = \longrightarrow$

Its MLE: $g(\hat{\mu}, \hat{\sigma}^2) = \hat{\mu} + z_{1-p}\hat{\sigma}$ is

the MLE of pth quantile of X.

Finding standard error (SE) of $\hat{\theta}$

Easy when: $\hat{\theta}$ is ~~~~ ~~~~ $\overline{X}$.

(a) Know: $E[\overline{X}] = \mu$, $var[\overline{X}] = \frac{\sigma^2}{n}$, $SE[\overline{X}] = \frac{\sigma}{\sqrt{n}}$

$$\Rightarrow \hat{SE} = \frac{s}{\sqrt{n}}, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

(b) $E[\hat{p}] = p$, $var[\hat{p}] = \frac{p(1-p)}{n}$ $\Rightarrow SE[\hat{p}] = \sqrt{\frac{p(1-p)}{n}}$

$$\Rightarrow \hat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

(c) Generally not easy to find ~~some~~ $\hat{SE}$'s of other estimators. — Later on will see "bootstrap" method to approximate $\hat{SE}$'s.

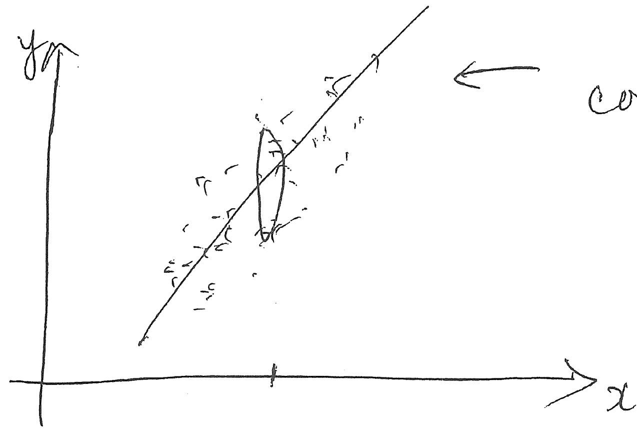
(d) If n is large, can appeal to large sample results regarding MLE to get its SE.

Detour

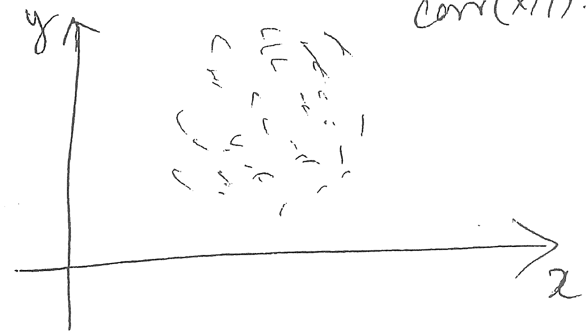
COVARIANCE & CORRELATION

- $\mathrm{COV}[X,Y] = E[(X-E(X))(Y-E(Y))]$
  $= E[XY] - E[X]\cdot E[Y]$.
- If X and Y are indep: $E[XY] = E[X]\cdot E[Y]$
  $\Rightarrow \mathrm{cov}(X,Y) = 0$.
- Converse is not true.
- $\mathrm{COV}[X,X] = \mathrm{var}[X]$.
- $\mathrm{CORR}(X,Y) = \dfrac{\mathrm{COV}(X,Y)}{SD(X)\cdot SD(Y)}$, $-1 \le \mathrm{CORR}(X,Y) \le 1$

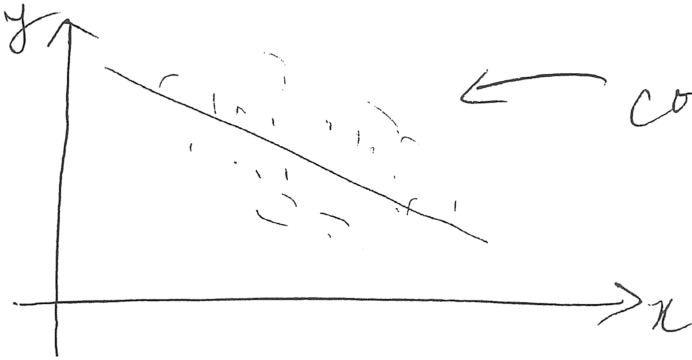
  Measures strength of linear relationship b/w X and Y
- Perfect correlation: $Y = aX+b$.
  In this case: $\mathrm{CORR}(X,Y) = \begin{cases} 1, & a>0 \\ -1, & a<0. \end{cases}$



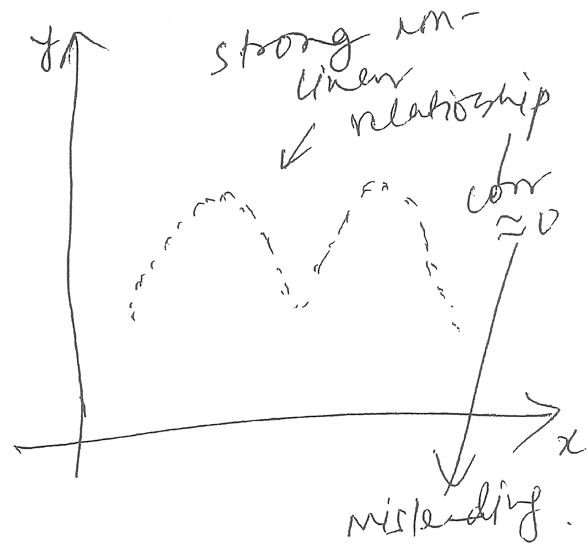
$\leftarrow$ $\mathrm{corr}(X,Y) > 0$.

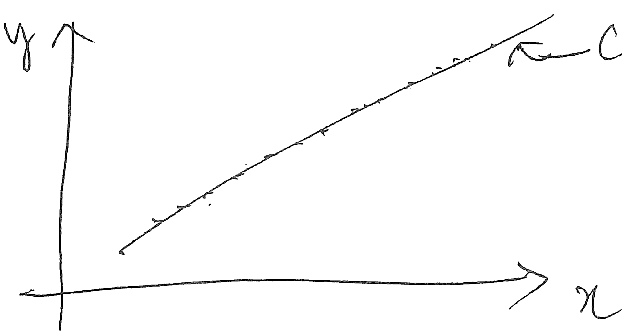


$\leftarrow$ $\mathrm{corr}(X,Y) \simeq 0$.



$\leftarrow$ $\mathrm{corr}(X,Y) < 0$





$\leftarrow$ $\mathrm{corr}(X,Y) = 1$

VARIANCE (OR COVARIANCE) MATRIX

$$\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_d \end{bmatrix}, \quad E[\underline{X}] = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_d) \end{bmatrix}$$

random vector

$$\mathrm{Var}[\underline{X}] = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_d) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_d) \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}(X_1, X_d) & \cdots & \cdots & \mathrm{Var}(X_d) \end{bmatrix}_{d \times d}$$

variances on diagonal

variance matrix
or
covariance matrix

(i,j) th element of $\mathrm{Var}[\underline{X}] = \mathrm{Cov}(X_i, X_j)$.

- Symmetric matrix.
- A positive semi definite matrix