STAT 5430
Lecture 03, M, Jan 27

- Homework 1 is assigned in Canvas (submit/due by next Monday, Feb 3, by midnight)

practice on point estimation (method of moments & likeihood estimation)

- Office hours to be announced
  Mine: FM, 12-1 PM & by appointment
  TA (Min-Yi): WR 11-12 in Snedecor 2404

Point Estimation

Background, continued

Definitions:

1. A (Borel measurable) function $\gamma : \Theta \to \mathbb{R}^d$, some $1 \le d < \infty$, is called a **parametric function**.

   $\gamma(\theta)$

   θ is a parameter

2. If a statistic $T = h(X_1, \dots, X_n)$ is used to estimate $\gamma(\theta)$, then T is called an **estimator of** $\gamma(\theta)$; and the observed value $t = h(x_1, \dots, x_n)$ is called an **estimate of** $\gamma(\theta)$.

   realized value of estimator T

Example: X_1, X_2, X_3 iid $N(\mu, \sigma^2)$

$\gamma(\mu) = \mu^2$ ← parametric function

$T = h(X_1, X_2, X_3) = (\bar{X}_3)^2$ ← estimator of $\gamma(\mu)$

Suppose $x_1 = 1, x_2 = 2, x_3 = 3$ are observed, then

$t = \left(\frac{1+2+3}{3}\right)^2 = 2^2 = 4$ is an estimate of $\gamma(\mu)$ (observed value of $T = (\bar{X}_3)^2$)

Some General Approaches to Point Estimation

(how to get statistics or estimators)

✓ I. Method of Moments

✓ II. Maximum Likelihood (popular)

III. Bayes Estimators (popular)

We'll next discuss I. & II., and return to Bayes estimators at a later point.

Point Estimation

Method of Moments Estimation (MOM estimation)

Definition: Let $X_1, \ldots, X_n$ be a r.s. from pdf/pmf $f(x|\underbrace{\theta_1, \ldots, \theta_k}_{k \text{ parameters}})$. Then,

pop. distribution

(a) $\mathrm{E}\{(X_1)^j\} \equiv \mu_j(\theta_1, \ldots, \theta_k)$ is the j**th population moment**, $j = 1, 2, \ldots$

parametric functions

e.g. $X_1 \sim N(\mu, \sigma^2)$, $E(X_1) = \mu$
$E(X_1^2) = Var(X_1) + (EX_1)^2$
$= \sigma^2 + \mu^2$

(b) $\mu_j' \equiv \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i)^j$ is the j**th sample moment**, $j = 1, 2, \ldots$

statistic
$j = 1, 2, 3, \ldots$

estimators based on $X_1, \ldots, X_n$

(c) The method of moments estimators (MMEs), say $\tilde{\theta}_1, \ldots, \tilde{\theta}_k$, of $\theta_1, \ldots, \theta_k$ are defined as the solution to

k parameters
$\Rightarrow$ k equations

$$\left.\begin{array}{rcl} \mu_1(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) & = & \mu_1' \\ \vdots & \vdots & \vdots \\ \mu_k(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) & = & \mu_k' \end{array}\right\} (*)$$

pick $\tilde{\theta}_1, \ldots, \tilde{\theta}_k$ so that pop./model moments match the sample moments

(d) The system of equations $(*)$ is called the method of moments equations (MMEquations).

Point Estimation

Method of Moments Estimation, cont'd

Example: Let $X_1, \ldots, X_n$ be a random sample from a Beta(α, β) distribution, $\alpha > 0, \beta > 0$. Find the MMEs of α, β.

Solution: $\theta_1 = \alpha$, $\theta_2 = \beta$

Then, $\mu_1(\theta_1, \theta_2) = EX_1 = \dfrac{\theta_1}{\theta_1 + \theta_2}$

& $\mu_2(\theta_1, \theta_2) = E(X_1^2) = \dfrac{(\theta_1 + 1)\theta_1}{(\theta_1 + \theta_2 + 1)(\theta_1 + \theta_2)}$

Hence, MME equations

$$\mu_1(\tilde\theta_1, \tilde\theta_2) = \frac{\tilde\theta_1}{\tilde\theta_1 + \tilde\theta_2} = \mu_1' = \bar X_1$$

$$\mu_2(\tilde\theta_1, \tilde\theta_2) = \frac{(\tilde\theta_1 + 1)\tilde\theta_1}{\underbrace{(\tilde\theta_1 + \tilde\theta_2 + 1)}_{\tilde\theta_1/\mu_1'}\underbrace{(\tilde\theta_1 + \tilde\theta_2)}_{\mu_1'}} = \mu_2' = \frac{1}{n}\sum_{i=1}^n X_i^2$$

Solve

$$\tilde\theta_1 = \frac{\bar X_n \left[\sum_{i=1}^n X_i(1 - X_i)\right]}{\sum_{i=1}^n (X_i - \bar X_n)^2}$$

$\leftarrow$ using $\mu_2' - (\mu_1')^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar X_n)^2$

$$\tilde\theta_2 = \left(\frac{1 - \mu_1'}{\mu_1'}\right)\tilde\theta_1$$

Point Estimation

Remarks on Method of Moments Estimators (MMEs)

1. Method of Moments doesn't work if there are not enough population moments.

e.g. Cauchy distribution $f(x|\theta) = \frac{1}{\pi(1+(x-\theta)^2)}$, $x \in \mathbb{R}$

Cauchy mean doesn't exist! $E|X_1| = +\infty = \int_{-\infty}^{\infty} |x| f(x|\theta)\,dx$ $\quad \theta \in \mathbb{R}$

2. MMEquations can have no or multiple solutions!

e.g. $X_1, \dots, X_n$ iid (discrete)

$P(X_1=x)$	θ	$1-2\theta$	θ
x	1	2	3

where $0 \le \theta \le \frac{1}{2}$.

try: . .

$\bar{X}_n = \mu_1' = \mu_1(\tilde{\theta})$
$= EX_1 = 2$ MME doesn't work here

Definition: For a parametric function $\gamma(\theta_1, \dots, \theta_k)$, we <u>define</u> the MME $\tilde{\gamma}(\theta_1, \dots, \theta_k)$ of $\gamma(\theta_1, \dots, \theta_k)$ as estimator of

$$\gamma(\theta_1, \dots, \theta_k) \to \tilde{\gamma}(\theta_1, \dots, \theta_k) = \gamma(\tilde{\theta}_1, \dots, \tilde{\theta}_k),$$

where $\underbrace{\tilde{\theta}_1, \dots, \tilde{\theta}_k}_{\text{parameter estimators}}$ are MMEs of $\theta_1, \dots, \theta_k$.

Example: Let $X_1, \dots, X_n$ be iid $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$. Find the MME of $\sin(\mu^2)$.

Solution: $\mu_1(\mu, \sigma^2) = EX_1 = \mu$, $\mu_2(\mu, \sigma^2) = EX_1^2 = \sigma^2 + \mu^2$

So the MMEs are $\tilde{\mu} = \mu_1(\tilde{\mu}, \tilde{\sigma}^2) = \mu_1' = \bar{X}_n$

"and" & $(\tilde{\mu})^2 + \tilde{\sigma}^2 = \mu_2(\tilde{\mu}, \tilde{\sigma}^2) = \mu_2' = \frac{1}{n}\sum_{i=1}^n X_i^2$

& Find $\tilde{\mu} = \bar{X}_n$ & $\tilde{\sigma}^2 = \mu_2' - (\mu_1')^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

So, the MME of $\gamma(\mu, \sigma^2) = \sin(\mu^2)$ is $\gamma(\tilde{\mu}, \tilde{\sigma}^2) = \sin((\bar{X}_n)^2)$.

Point Estimation

Maximum Likelihood Estimation

Definition: Let $f(x_1, \dots, x_n|\theta)$ be the joint pdf/pmf of $(X_1, \dots, X_n)$. Then,

$$L(\theta) = f(\underbrace{x_1, \dots, x_n}_{\text{data fixed}}|\theta), \quad \theta \in \Theta$$

← Joint "probability" of data values, treated as a function of θ.

[as a function of θ, given $(x_1, \dots, x_n)$] is called the **likelihood function.**

Note:

1. If $X_1, \dots, X_n$ are iid with common pdf/pmf $f(x|\theta)$, then

$$L(\theta) = \underbrace{f(x_1, \dots, x_n|\theta)}_{\text{joint}} = \prod_{i=1}^{n} \underbrace{f(x_i|\theta)}_{\text{marginals}}$$

2. If $X_1, \dots, X_n$ are discrete r.v.'s, then

$$L(\theta) = f(x_1, \dots, x_n|\theta) = P(X_1 = x_1, \dots, X_n = x_n|\theta)$$

Definition: Let $(X_1, \dots, X_n)$ have point pdf/pmf $f(x_1, \dots, x_n|\theta)$, $\theta \in \Theta$. Then, for a given set of observations $(x_1, \dots, x_n)$, the **maximum likelihood estimate** (MLE) of θ is a point $\hat{\theta}$ in Θ, say $\hat{\theta} = h(x_1, \dots, x_n)$, such that

$$f(x_1, \dots, x_n|\hat{\theta}) = \max_{\theta \in \Theta} f(x_1, \dots, x_n|\theta) = \max_{\theta \in \Theta} L(\theta)$$

And the **maximum likelihood estimator** (MLE) of θ is defined as $\hat{\theta} = h(X_1, \dots, X_n)$.

So, MLE $\hat{\theta} = h(X_1, \dots, X_n)$ & $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$

parameter space

Example/Discussion:

$\Theta = [0,1]$

Suppose X_1, X_2, X_3 iid Bernoulli(θ) so

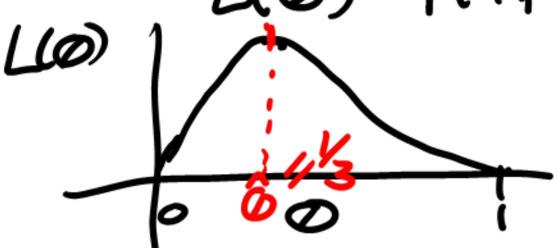
$$X_i = \begin{cases} 1 & \text{w.p. } \theta \\ 0 & \text{w.p. } 1-\theta \end{cases}$$

where $0 \le \theta \le 1$

$\theta \in \Theta$ Observe $X_1 = 0, X_2 = 1, X_3 = 0$

Estimate θ

$$L(\theta) = P(X_1 = 0, X_2 = 1, X_3 = 0|\theta) = \prod_{i=1}^{3} f(x_i, \theta) = \theta(1-\theta)^2, \quad 0 \le \theta \le 1.$$



$\hat{\theta} = 1/3$ (pick θ value for which the data "$X_1=0, X_2=1, X_3=0$" seem most plausible or have highest likelihood)

Point Estimation

Finding Maximum Likelihood Estimators (MLEs)

Finding the MLE $\hat{\theta}$ requires *maximizing* the likelihood $L(\theta)$ function *over the parameter space* $\theta \in \Theta$. There are several potential ways to achieve this.

1. If $L(\theta)$ is smooth (i.e., differentiable) in θ (which happens often), consider using calculus to maximize $L(\theta)$.
2. If $L(\theta)$ is *not* smooth, need to think more carefully about how to maximize $L(\theta)$ over Θ for the specific model at hand.
3. Often times in practice, $L(\theta)$ is maximized numerically using some computing.
4. Maximizing $\log L(\theta)$ is equivalent to maximizing $L(\theta)$ & can be easier.
5. In particular, if $X_1, \ldots, X_n$ are iid with common pdf/pmf $f(x|\theta)$ where the support $\{x : f(x|\theta) > 0\}$ changes with θ, then using *indicator functions* to write $f(x|\theta)$ and $L(\theta)$ can help in maximization.

Using Calculus to Determine the MLE

If the likelihood function $L(\theta) = f(x_1, \ldots, x_n|\theta)$ is differentiable, it can often be maximized over Θ using calculus.

Assume $\Theta \subset \mathbb{R}$ is open and that $L(\theta)$ is twice differentiable on Θ. Then,

$$\hat{\theta} \text{ maximizes } L(\theta) \iff \left.\frac{dL(\theta)}{d\theta}\right|_{\hat{\theta}} = 0 \quad \text{and} \quad \left.\frac{d^2L(\theta)}{d\theta^2}\right|_{\hat{\theta}} < 0.$$

Since $\log(\cdot)$ is an increasing function, $\hat{\theta}$ maximizes $L(\theta) \iff \hat{\theta}$ maximizes $\log L(\theta)$. Hence,

$$\hat{\theta} \text{ is an MLE if } \quad \left.\frac{d\log L(\theta)}{d\theta}\right|_{\hat{\theta}} = 0 \quad \text{and} \quad \left.\frac{d^2\log L(\theta)}{d\theta^2}\right|_{\hat{\theta}} < 0.$$