STAT 5430
Lecture 04, W, Jan 29

- 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

Maximum Likelihood Estimation

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

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

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

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

Note:

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

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

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

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

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

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

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

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

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.

80%
1. If $L(\theta)$ is smooth (i.e., differentiable) in θ (which happens often), consider using calculus to maximize $L(\theta)$.

20%
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. (don't use calculus)
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.

range or what's possible for data

e.g. indicator $I(A)$ of an event A is $I(A) = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$

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,

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

Point Estimation

Finding Maximum Likelihood Estimators (MLEs)/Example using Calculus

Example: Let $X_1, \ldots, X_n$ be a random sample from a Geometric(p) distribution, $0 < p < 1$. Find the MLE of p.

Solution: $L(p) \equiv f(x_1, \ldots, x_n | p) = \prod_{i=1}^{n} f(x_i | p) = \prod_{i=1}^{n} \left[p(1-p)^{x_i - 1} \right]$ (as long as each $x_i \in \{1, 2, 3, 4, \ldots\}$)

$$= p^n (1-p)^{\sum_{i=1}^n x_i - n}$$

$$\log L(p) = n \log p + \left(\sum_{i=1}^n x_i - n\right) \log(1-p)$$

(Note: We will assume that it NOT true "$x_1 = \cdots = x_n = 1$"; that is, assume $\sum_{i=1}^n x_i > n$)

$$\left.\frac{d \log L(p)}{dp}\right|_{p=\hat{p}} = \frac{n}{\hat{p}} + \frac{\left(\sum_{i=1}^n x_i - n\right)}{(1-\hat{p})}(-1) = 0$$

ie. $n(1-\hat{p}) = \hat{p}\left(\sum_{i=1}^n x_i - n\right)$

ie. $\hat{p} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\overline{x}_n} \in (0,1)$ ($\hat{p}$ is in parameter $(0,1)$ if $\sum_{i=1}^n x_i > n$)

check: $\frac{d^2 \log L(p)}{dp^2} = \frac{-n}{p^2} - \frac{\left(\sum_{i=1}^n x_i - n\right)}{(1-p)^2} < 0$ $\left(\frac{d p^{-1}}{dp} = -1 \cdot p^{-2}\right)$

for all $p \in (0,1)$ including $\hat{p} = \frac{1}{\overline{x}_n}$

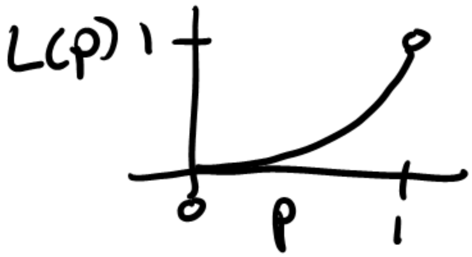
$\Rightarrow$ MLE of p is $\hat{p} = 1/\overline{X}_n$ (if $\sum_{i=1}^n x_i > n$)

Note: For discrete r.v.s (Geometric, binomial, etc), pathological cases for MLE can occur if $x_1 = \cdots = x_n = M$, where M is either minimum or maximum value of X_1

e.g. Geometric $x_1 = \cdots = x_n = 1$

$\Rightarrow L(p) = p^n$

$\Rightarrow$ on $p \in (0,1)$, $L(p)$ has no maximum here in the parameter space $(0,1)$



$\Rightarrow$ No MLE for p when $x_1 = \cdots = x_n = 1$ here

But, if $(0,1]$ is parameter space,

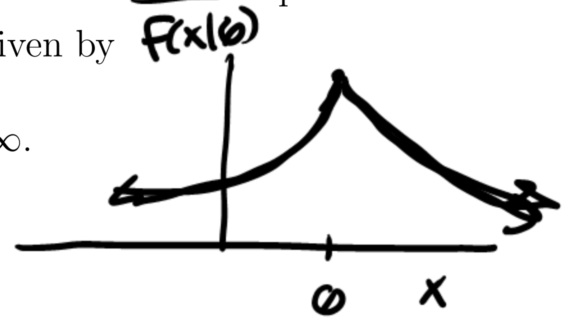
then $L(p) = p^n$ has a max at $\hat{p} = 1$

(MLE exists)

Point Estimation

Finding Maximum Likelihood Estimators (MLEs)/Examples without Calculus

Example: (Non-differentiable likelihood) Let $X_1, \ldots, X_n$ be a random sample from a Double Exponential(θ) distribution, $\theta \in \mathbb{R}$, with pdf given by



continuous

$$f(x|\theta) = \frac{1}{2}e^{-|x-\theta|}, \quad -\infty < x < \infty.$$

Find the MLE of θ.

$$L(\theta) = f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$= \left(\frac{1}{2}\right)^n e^{-\sum_{i=1}^{n} |x_i - \theta|}$$

$\hat{\theta}$ maximizes $L(\theta)$ $\Leftrightarrow$ $\hat{\theta}$ minimizes $\sum_{i=1}^{n} |x_i - \theta|$

$\Leftrightarrow$ $\hat{\theta} = \text{median}(x_1, \ldots, x_n)$

(in contrast, $\sum_{i=1}^{n} (x_i - \theta)^2$ is minimized at $\theta = \bar{x}_n$)

Example: Let $\theta \geq 1$ be an integer. Let X be a r.v. with a *discrete* uniform distribution on $\{1, \ldots, \theta\}$; that is,

distribution has support or range depending on θ

$$P(X = x|\theta) = \begin{cases} \frac{1}{\theta} & \text{for } x = 1, \ldots, \theta \\ 0 & \text{otherwise.} \end{cases} = \frac{1}{\theta} I(\text{pos integer } x \leq \text{integer } \theta)$$

If $X = 2$ is observed, what is the maximum likelihood estimate of θ?

$L(\theta) = P(X=2|\theta) = f(2|\theta)$ for $\theta = 1, 2, 3, \ldots$

$$= \frac{1}{\theta} I(2 \leq \text{integer } \theta) = \begin{cases} 0 & \text{if } \theta = 1 \\ 1/2 & \text{if } \theta = 2 \\ 1/3 & \text{if } \theta = 3 \\ \vdots & \vdots \\ 1/m & \text{if } \theta = m \end{cases}$$

$\Rightarrow$ MLE of θ is $\hat{\theta} = 2$

or $L(2) = \max_{\theta \in \{1,2,3,\ldots\}} L(\theta)$