STAT 5430: Summary to date
Where we have been
& where we are headed

- Completed: Introduction to Statistical Inference
  - definitions/notation
  - random samples for inference about parametric population distributions

- Next: Point Estimation
  - Defining statistics & point estimators
  - Some strategies for point estimation
    - Method of Moments Estimation (MME)
    - Maximum Likelihood Estimation (MLE)

Point Estimation

Background

Definition: Let $X_1, \ldots, X_n$ be a random sample. A (Borel measurable) function of the random sample, say $T = h(X_1, \ldots, X_n)$, is called a **statistic** or an **estimator**.

Examples:

Definition: The probability distribution of a statistic T is called the **sampling distribution of** T.

Example: Suppose $X_1, \ldots, X_n$ is a r.s. from $N(\mu, \sigma^2)$.

Point Estimation

Background, continued

Definitions:

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

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

Example:

Some General Approaches to Point Estimation

I. Method of Moments

II. Maximum Likelihood

III. Bayes Estimators

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

Point Estimation

Method of Moments Estimation

Definition: Let $X_1, \ldots, X_n$ be a r.s. from pdf/pmf $f(x|\theta_1, \ldots, \theta_k)$. Then,

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

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

(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

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

(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.

Point Estimation

Remarks on Method of Moments Estimators (MMEs)

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

2. MMEquations can have no or multiple solutions!

Definition: For a parametric function $\gamma(\theta_1, \ldots, \theta_k)$, we define the MME $\tilde{\gamma}(\theta_1, \ldots, \theta_k)$ of $\gamma(\theta_1, \ldots, \theta_k)$ as

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

where $\tilde{\theta}_1, \ldots, \tilde{\theta}_k$ are MMEs of $\theta_1, \ldots, \theta_k$.

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

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,

$$L(\theta) = f(x_1, \ldots, x_n|\theta), \quad \theta \in \Theta$$

[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

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

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

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

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

Example/Discussion:

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,

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

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

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

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.

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

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

Find the MLE of θ.

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,

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

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

Point Estimation

Finding Maximum Likelihood Estimators (MLEs)/Multiparameter Case

Suppose $X_1, X_2, \ldots, X_n$ have joint pmf/pdf $f(x_1, x_2, \ldots, x_n | \underset{\sim}{\theta})$ where $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)' \in \Theta \subset \mathbb{R}^k$ (i.e., k parameters).

Want to find MLEs $\hat{\underset{\sim}{\theta}} \equiv (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)'$ of $\underset{\sim}{\theta}$, which solve

$$L(\hat{\underset{\sim}{\theta}}) = \max_{\underset{\sim}{\theta} \in \Theta} L(\underset{\sim}{\theta}), \quad \text{where} \quad L(\underset{\sim}{\theta}) \equiv f(x_1, x_2, \ldots, x_n | \underset{\sim}{\theta})$$

.

Result: If $\Theta \subset \mathbb{R}^k$ is open and $L(\underset{\sim}{\theta}) \equiv f(x_1, x_2, \ldots, x_n | \underset{\sim}{\theta})$ has 2nd order partial derivatives on Θ, then $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are MLEs of $\theta_1, \theta_2, \ldots, \theta_k$ provided

1. for each $i = 1, \ldots, k$
$$\frac{\partial \log L(\underset{\sim}{\theta})}{\partial \theta_i}\Big|_{\hat{\underset{\sim}{\theta}}} = 0;$$

2. denote the $k \times k$ Hessian matrix at $\hat{\underset{\sim}{\theta}}$ as
$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & h_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k1} & h_{k2} & \cdots & h_{kk} \end{pmatrix}$$
where
$$h_{ij} = \frac{\partial^2 \log L(\underset{\sim}{\theta})}{\partial \theta_i \partial \theta_j}\Big|_{\hat{\underset{\sim}{\theta}}} \quad \text{for } i, j = 1, \ldots k,$$
and let
$$\Delta_i = \det \begin{pmatrix} h_{11} & \cdots & h_{1i} \\ \vdots & \ddots & \vdots \\ h_{i1} & \cdots & h_{ii} \end{pmatrix} \quad \text{for } i = 1, \ldots k,$$
be the determinant of the $i \times i$ submatrix of H consisting of the first i rows. Then, we need $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 > 0, \cdots$ and so on. (Must compute k determinants $\Delta_1, \Delta_2, \ldots, \Delta_k$ to see if they alternate in positive/negative.)

Point Estimation

Finding Maximum Likelihood Estimators/Example in Multiparameter Case

Example: Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$. Find MLEs of $\mu \& \sigma^2$.

Solution: Write $\theta_1 \equiv \mu$ and $\theta_2 \equiv \sigma^2$ and

$$\begin{aligned}
L(\theta_1, \theta_2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_i-\theta_1)^2}{2\theta_2}} = (2\pi)^{-n/2} \theta_2^{-n/2} e^{-\sum_{i=1}^{n} \frac{(x_i-\theta_1)^2}{2\theta_2}} \\
\log L(\theta_1, \theta_2) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\theta_2 - \sum_{i=1}^{n} \frac{(x_i-\theta_1)^2}{2\theta_2}
\end{aligned}$$

Then, setting

$$\frac{\partial \log L(\underset{\sim}{\theta})}{\partial \theta_i}\Big|_{\underset{\sim}{\hat{\theta}}} = 0 \quad \text{for } i = 1, 2,$$

we see that the MLEs $\underset{\sim}{\hat{\theta}} \equiv (\hat{\theta}_1, \hat{\theta}_2)'$ satisfy

$$\sum_{i=1}^{n} \frac{(x_i - \hat{\theta}_1)}{\hat{\theta}_2} = 0 \qquad \& \qquad -\frac{n}{2\hat{\theta}_2} - \sum_{i=1}^{n} \frac{(x_i-\hat{\theta}_1)^2}{2(\hat{\theta}_2)^2} = 0,$$

implying that

$$\hat{\theta}_1 = \bar{x}_n = \sum_{i=1}^{n} x_i/n, \qquad \hat{\theta}_2 = \sum_{i=1}^{n}(x_i - \hat{\theta}_1)^2/n = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2/n,$$

though need to check 2nd partials conditions too. Note

$$\begin{aligned}
\frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial \theta_1^2} &= \frac{\partial}{\partial \theta_1}\left[\frac{\partial \log L(\theta_1,\theta_2)}{\partial \theta_1}\right] = -\frac{n}{\theta_2}, \\
\frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial \theta_2^2} &= \frac{\partial}{\partial \theta_2}\left[\frac{\partial \log L(\theta_1,\theta_2)}{\partial \theta_2}\right] = \frac{n}{2(\theta_2)^2} - \frac{\sum_{i=1}^{n}(x_i-\theta_1)^2}{(\theta_2)^3} \\
\frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial \theta_2 \partial \theta_1} &= \frac{\partial}{\partial \theta_2}\left[\frac{\partial \log L(\theta_1,\theta_2)}{\partial \theta_1}\right] = -\frac{\sum_{i=1}^{n}(x_i-\theta_1)}{(\theta_2)^2} = \frac{\partial}{\partial \theta_1}\left[\frac{\partial \log L(\theta_1,\theta_2)}{\partial \theta_2}\right] = \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial \theta_1 \partial \theta_2}
\end{aligned}$$

and hence

$$h_{11} \equiv \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial \theta_1^2}\Big|_{\underset{\sim}{\hat{\theta}}} = -\frac{n}{\hat{\theta}_2}, \quad h_{22} \equiv \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial \theta_2^2}\Big|_{\underset{\sim}{\hat{\theta}}} = \frac{n}{2(\hat{\theta}_2)^2} - \frac{\sum_{i=1}^{n}(x_i-\hat{\theta}_1)^2}{(\hat{\theta}_2)^3} = -\frac{n}{2(\hat{\theta}_2)^2}$$

$$h_{12} \equiv \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial \theta_1 \partial \theta_2}\Big|_{\underset{\sim}{\hat{\theta}}} = -\frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)}{(\hat{\theta}_2)^2} = 0 = \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial \theta_2 \partial \theta_1}\Big|_{\underset{\sim}{\hat{\theta}}} \equiv h_{21}$$

$$H \equiv \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} -n/\hat{\theta}_2 & 0 \\ 0 & -n/(2\hat{\theta}_2^2) \end{bmatrix} \quad \Rightarrow \quad \Delta_1 \equiv -\frac{n}{\hat{\theta}_2} < 0 \ \& \ \Delta_2 \equiv \det(H) = \frac{n^2}{2\hat{\theta}_2^3} > 0$$

Point Estimation

Maximum Likelihood Estimators (MLEs) of Parametric Functions

Definition: For a parametric function $\gamma(\theta_1, \theta_2, \ldots, \theta_k)$, we define $\gamma(\hat{\theta}_1, \ldots, \hat{\theta}_k)$ as the MLE of $\gamma(\theta_1, \theta_2, \ldots, \theta_k)$, where $\hat{\theta}_1, \ldots, \hat{\theta}_k$ are the MLEs of $\theta_1, \ldots, \theta_k$.

Last Example: Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$. Find the MLE of $\log(\mathrm{E}X_1^2) = \log(\mu^2 + \sigma^2)$.