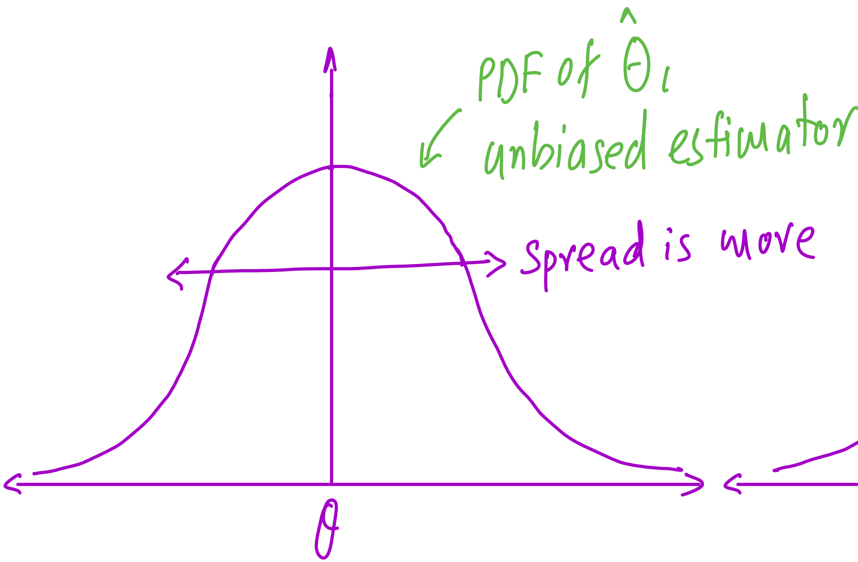
Lecture 29:
Estimation - Part II

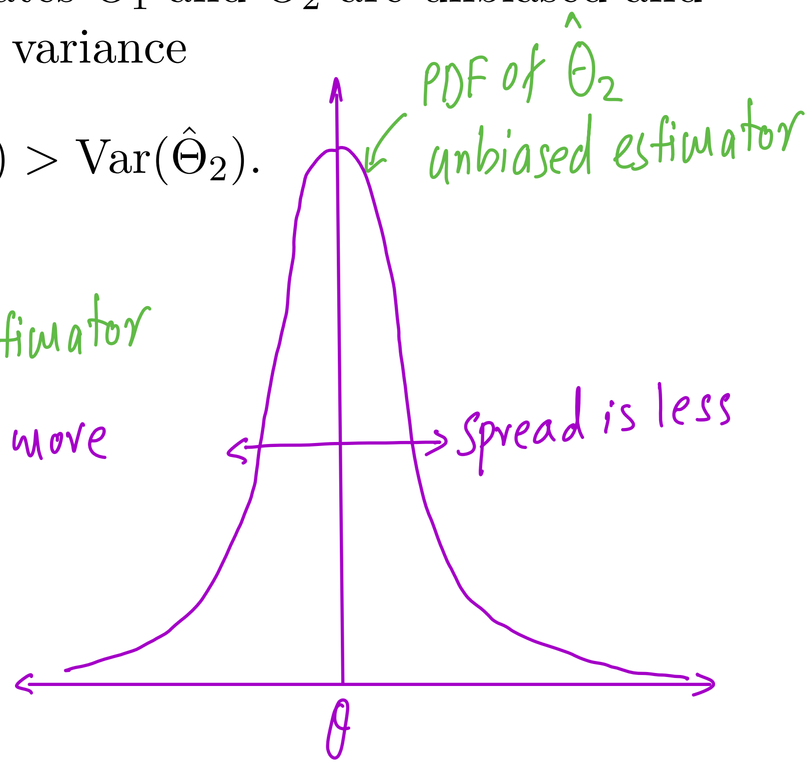
Satyajit Thakor
IIT Mandi

Point estimation

- Recall that an $\overbrace{\text{unbiased estimator}}^{E(\hat{\Theta}) = \theta}$ is more desirable in practice.
- If we have two or more unbiased estimators to choose from, which is better to choose?
- Suppose that two point estimates $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased and symmetric around θ and have variance

$$\mathrm{Var}(\hat{\Theta}_1) > \mathrm{Var}(\hat{\Theta}_2).$$



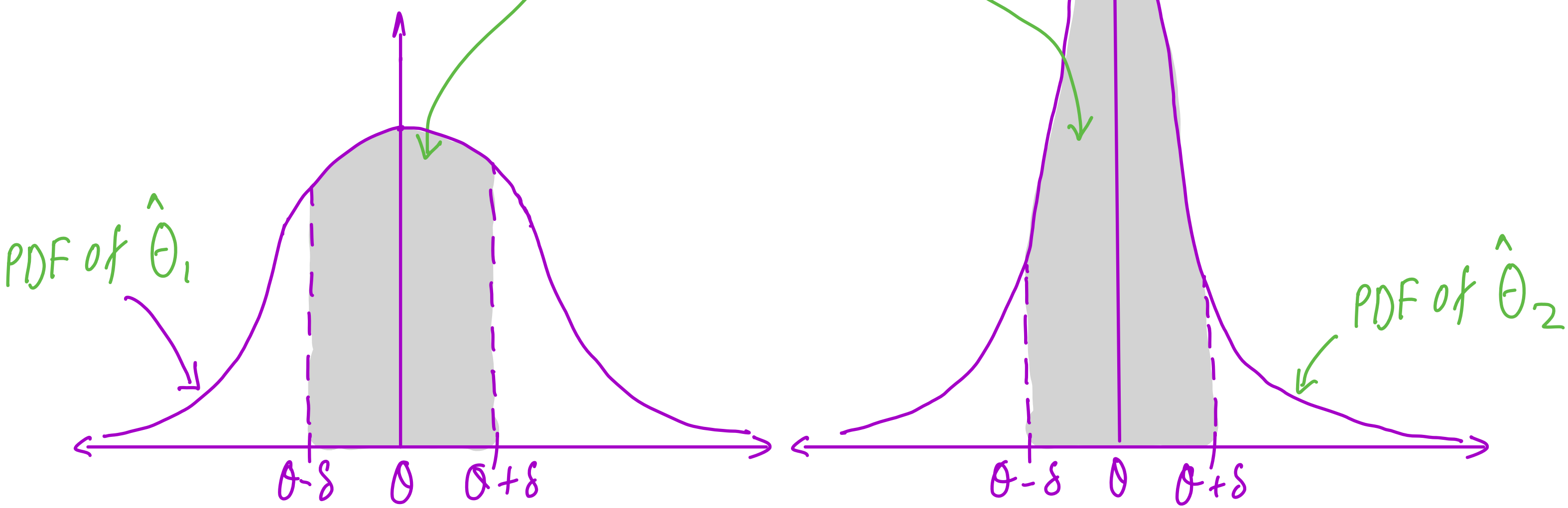


Point estimation

- Which is a better point estimate? Answer: $\hat{\Theta}_2$
- $\hat{\Theta}_2$ is better in the sense that it is likely to provide an estimate closer to the true value θ than the estimate provided by $\hat{\Theta}_1$, that is,

$$P(|\hat{\Theta}_1 - \theta| \le \delta) < P(|\hat{\Theta}_2 - \theta| \le \delta)$$

for any $\delta > 0$.



Maximum likelihood estimation

- Till now we discussed simple ways to define an estimate, e.g., sample mean, sample variance.
- Now we will study one of the most popular (and scientific) method for parameter estimation.

- Some times, it happens that you know the family distribution of the random variable, e.g., Bernoulli, binomial, normal, etc.
- But you do not know the crucial parameter. For example:
  - In Bern(p), p is not known.
  - In Binom(n, p), p is not known.
  - In Pois(λ), λ is not known.
  - In $\mathcal{N}(\mu, \sigma^2)$, μ and σ^2 are not known.

Maximum likelihood estimation

- Maximum likelihood method: Find the parameter value such that the likelihood function is maximized.
- Let $x_1, \ldots, x_n$ be a sample from some population or distribution and let θ be a parameter we want to estimate.
- The quantity

$$L(x_1, \ldots, x_n; \theta) = f(x_1, \ldots, x_n; \theta)$$

is called the likelihood function, where f is either the joint PFD or PMF with parameter θ.
- Since $X_1, \ldots, X_n$ are iid, we have

$$\begin{aligned} L(x_1, \ldots, x_n; \theta) &= f(x_1, \ldots, x_n; \theta) \\ &= f(x_1; \theta) \cdots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i, \theta) \end{aligned}$$

Maximum likelihood estimation

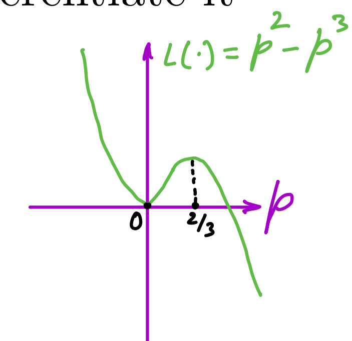
- Let's try to understand likelihood function and its maximization by example.
- We want to find the parameter p of the Bernoulli random variable "$X = 1$ if not defective and $X = 0$ otherwise".
- Assume that we have only three samples $x_1 = 1, x_2 = 1, x_3 = 0$. What is the best estimate of p given this information? $2/3$?
- The likelihood function is

$$L(x_1, x_2, x_3; p) = P(X_1 = 1; p)P(X_2 = 1; p)P(X_3 = 0; p) = p^2(1 - p)$$

- To find the maximum of the likelihood function, differentiate it and equate to zero:

differentiate: $\frac{d}{dp} L(x_1, x_2, x_3; \theta) = \frac{d}{dp}(p^2 - p^3)$

$= 2p - 3p^2$



Maximum likelihood estimation

equate to zero: $2\hat{p} - 3\hat{p}^2 = 0$ $\quad$ $\hat{p}$ is the estimate

$$\Rightarrow \quad \hat{p} = 0 \text{ or } \hat{p} = 2/3.$$

- To verify that the value indeed maximizes the likelihood function, check whether the second derivative at the estimated value is negative.

second derivative test:

$$\left.\frac{d^2}{dp^2} L(x_1, x_2, x_3; \theta)\right|_{p=\hat{p}} = 2 - 6p \Big|_{p=0} = \underbrace{2 > 0}_{\Rightarrow \hat{p}=0 \text{ does not maximize } L(\cdot)}$$

$$\left.\frac{d^2}{dp^2} L(x_1, x_2, x_3; \theta)\right|_{p=\hat{p}} = 2 - 6p \Big|_{p=2/3} = \underbrace{-2 < 0}_{\Rightarrow \hat{p}=2/3 \text{ maximizes } L(\cdot).}$$

- 2/3 is indeed a reasonable estimate of the parameter from the given sample: on average 2 are not defective out of 3.

Maximum likelihood estimation

- Basic idea of maximum likelihood estimation: the reasonable estimator of a parameter based on a sample is that parameter value that produces the largest probability of obtaining the sample.

- Definition: Given a random sample (or independent observations) $x_1, x_2, \ldots, x_n$ from a PDF/PMF $f(x; \theta)$, the maximum likelihood estimator $\hat{\theta}$ is that which maximizes the likelihood function

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

Maximum likelihood estimation

generalization of x_1, x_2, x_3.

- $\underbrace{\text{Example}}$: Given a sample $x_1, \ldots, x_n$ find maximum likelihood estimator of p for $X \sim \text{Bern}(p)$.

For a Bernoulli r.v. X_i with $P(X_i=1)=p$ and $P(X_i=0)=1-p$, we can write:

$$P(X_i=x_i) = p^{x_i}(1-p)^{1-x_i}, \text{ for } x_i = 1, 0.$$

Hence, the likelihood function is:

$$\begin{aligned} L(x_1, \ldots, x_n; p) &= \prod_{i=1}^{n} f(x_i; p) \\ &= \prod_{i=1}^{n} P(X_i = x_i; p) \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i} \end{aligned}$$

Maximum likelihood estimation

- Note that the expression is too complex to find $\hat{p}$.
- If a value maximizes a function then it also maximizes any monotically increasing function of the function.
- Hence, take log both sides:

$$\underbrace{\log L(x_1, \dots, x_n; p)}_{\text{called Log-Likelihood function}} = \sum x_i \log p + \left(n - \sum x_i\right) \log(1-p)$$

- Differentiate:

$$\Rightarrow \frac{d}{dp} \log L(x_1, \dots, x_n; p) = \frac{\sum x_i}{p} - \frac{\left(n - \sum x_i\right)}{1-p}$$

Maximum likelihood estimation

- Equate to zero:

$$\Rightarrow \frac{\sum x_i}{\hat{p}} - \frac{(n - \sum x_i)}{1-\hat{p}} = 0$$

$$\Rightarrow \sum x_i - \cancel{\sum x_i \hat{p}} = n\hat{p} - \cancel{\sum x_i \hat{p}}$$

$$\Rightarrow \hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Second derivative test:

$$\frac{d^2}{dp^2} \log L(x_1, \ldots, x_n; p)\Big|_{p=\hat{p}} = \frac{-\sum x_i}{p^2} - \frac{(n - \sum x_i)}{(1-p)^2}\Big|_{p=\hat{p}}$$

Maximum likelihood estimation

$$= -\frac{\sum x_i}{\left(\frac{\sum x_i}{n}\right)^2} - \frac{(n - \sum x_i)}{\left(\frac{n - \sum x_i}{n}\right)^2}$$

$$= -\frac{n^2}{\sum x_i} - \frac{n^2}{n - \sum x_i}$$

$$= -\left(\frac{n^2}{\sum x_i} + \frac{n^2}{n - \sum x_i}\right)$$

$$< 0. \quad \left(\because \sum_{i=1}^{n} x_i \text{ can only take values } 0, 1, \ldots, n\right)$$

$\Rightarrow \hat{p} = \frac{\sum x_i}{n}$ maximizes the log-likelihood function.

$\Rightarrow \hat{p}$ is the maximum likelihood estimate.