Chapter 6. Point Estimation

Math 3670 Summer 2024

Georgia Institute of Technology

Section 1.
Some General Concepts of Point Estimation

Population
(distribution (PDF, PMF)
with parameters)

$x_1 \quad x_2 \quad \cdots \quad x_n$: data

$\underbrace{X_1, X_2, \ldots, X_n}_{\text{sample}}$ i.i.d

$u(X_1, \ldots, X_n)$

Ex: $\frac{1}{n}\sum X_i$, $\max\{X_i\}$, $\sum_i X_i^2$

Goal: Estimate "Parameters"

Statistics and Point Estimators

Definition

A statistic is any quantity whose value can be calculated from sample data.

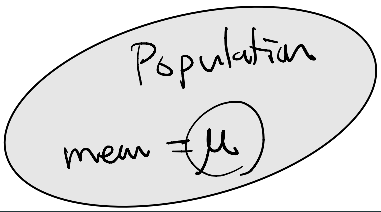
Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result.

Therefore, a statistic is a random variable and will be denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value of the statistic.

A point estimate of a parameter θ is a single number that can be regarded as a sensible value for θ.

A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data.

The selected statistic is called the point estimator of θ



$x_1, \dots, x_{10}$

$X_1, \dots, X_n$

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$: a statistic

Statistics and Point Estimators

Example

Let μ (a parameter) denote the true average breaking strength of wire connections used in bonding semiconductor wafers.

A random sample of $n = 10$ connections might be made, and the breaking strength of each one determined, resulting in observed strengths $x_1, x_2, \cdots, x_{10}$.

The sample mean breaking strength $\overline{x}$ could then be used to draw a conclusion about the value of μ.

Point estimator for μ

$= \overline{X}$

$\hat{\mu} = \frac{1}{n}(x_1 + \dots + x_{10})$

Statistics and Point Estimators

Example

An automobile manufacturer has developed a new type of bumper, which is supposed to absorb impacts with less damage than previous bumpers. The manufacturer has used this bumper in a sequence of 25 controlled crashes against a wall, each at 10 mph, using one of its compact car models.

Let X be the number of crashes that result in no visible damage to the automobile. The parameter to be estimated is

$$p = \mathbb{P}(\text{no damage in a single crash}).$$

If X is observed to be $x = 15$, then

$$\hat{p} = \qquad\qquad\qquad\qquad .$$

Population p
25 trials
# of crashes

Sample
$X_1, \dots, X_{25}$

X = # of crashes

$$\hat{p} = \frac{X}{25}$$

Statistics and Point Estimators

Example

The article "Is a Normal Distribution the Most Appropriate Statistical Distribution for Volumetric Properties in Asphalt Mixtures?" reported the following observations on $X =$ voids filled with asphalt (%) for 52 specimens of a certain type of hot-mix asphalt:

74.33	71.07	73.82	77.42	79.35	82.27	77.75	78.65	77.19
74.69	77.25	74.84	60.90	60.75	74.09	65.36	67.84	69.97
68.83	75.09	62.54	67.47	72.00	66.51	68.21	64.46	64.34
64.93	67.33	66.08	67.31	74.87	69.40	70.83	81.73	82.50
79.87	81.96	79.51	84.12	80.61	79.89	79.70	78.74	77.28
79.97	75.09	74.38	77.67	83.73	80.39	76.90		

Possible estimators for σ^2 are

$$\hat{\mu} = \frac{1}{n} \sum X_i$$

$$\hat{\sigma}^2 = \begin{cases} \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \\ \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \end{cases}$$

$$E\left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \right] = \frac{n-1}{n} \sigma^2$$

$$E\left[\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \right] = \sigma^2$$

$$X = \begin{cases} 1, & \text{with prob. } \frac{3}{4}, \quad \frac{1}{3} \\ 0 & \frac{1}{4}, \end{cases}$$

Statistics and Point Estimators

For a parameter θ, there are many different possible estimators.

Among them, which one would be best? How can we choose the best possible estimator for θ?

$\hat{\mu} = \frac{1}{n} \Sigma x_i$: random variable.

μ : a number

$$E[\hat{\mu}] = E[\frac{1}{n} \Sigma x_i] = \mu$$

Unbiased Estimators

θ : parameter

Definition

A point estimator $\widehat{\theta}$ is said to be an unbiased estimator of θ if

statistic

$$\mathbb{E}[\widehat{\theta}] = \theta$$

for every possible value of θ.

If $\widehat{\theta}$ is not unbiased, the difference

$$\mathbb{E}[\widehat{\theta}] - \theta$$

is called the bias of $\widehat{\theta}$.

Principle of Unbiased Estimation: When choosing among several different estimators of u, select one that is unbiased.

Unbiased Estimators

Example

known

Let $X \sim \text{Bin}(n, p)$.

Find an unbiased estimator of p.

"sample proportion."

$\hat{p} = \frac{X}{n}$: unbiased estimator.

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{E[X]}{n} = \frac{np}{n} = p.$$

Unbiased Estimators

Example

Let $X \sim$ Unif$(0, \theta)$.

Find an unbiased estimator of θ.

Population

Unif$(0, \theta)$

θ

$X_1, \ldots, X_n$

$[0, \theta]$

$$\hat{\theta}_1 = \underbrace{\max\{X_1, \ldots, X_n\}}_{=Y}$$

PDF of this?

CDF

$$E[\hat{\theta}_1] = \frac{\theta}{2}$$

$$P(Y \leq y) = P(X_1 \leq y, X_2 \leq y, \ldots X_n \leq y)$$

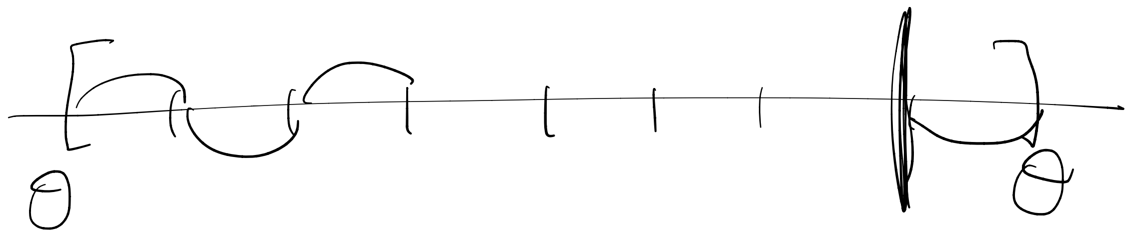
$$= P(X_1 \leq y) \cdots P(X_n \leq y) = \left(\frac{y}{\theta}\right)^n$$

$$f_Y(y) = n \cdot \theta^{-n} y^{n-1}$$

$$E[Y] = \int_0^\theta y \cdot f_Y(y)\, dy = \int_0^\theta n \cdot \theta^{-n}\, y^n\, dy$$
$$= \frac{n}{n+1} \cdot \theta^{-n} \cdot \theta^{n+1}$$
$$= \frac{n}{n+1} \cdot \theta$$

$$E[\hat\theta_1] = \frac{n}{n+1}\theta$$

$$\hat\theta_2 = \frac{n+1}{n} \max\{X_1, \ldots, X_n\} \leftarrow \text{unbiased}$$



Recall

A point estimator

Population
Distribution
Parameters

$\longrightarrow$ Sample

$\underbrace{X_1, X_2, \cdots, X_n}_{\text{RVs}}$

A parameter $\theta \leftarrow$ To estimate θ,

$\hat{\theta} = $ a function of $\underbrace{X_1, X_2, \cdots, X_n}_{\text{RVs}}$

point estimator

Point estimate for θ $= \hat{\theta}(\underbrace{x_1, \cdots, x_n}_{\text{data}})$

Example

Uniform RV on $[0, \theta] \leftarrow$ Population

$\downarrow$

$X_1, X_2, \cdots, X_n$

$\hat{\theta} =$ Best possible estimator for θ?

$\hat{\theta}_1 = \max\{X_1, X_2, \cdots, X_n\} \leftarrow$ RV

$\theta \leftarrow$ a number

Comparing these two. How?

$E[\hat{\theta}_1] = \theta \leftarrow \hat{\theta}_1$ is unbiased.

$\hat{\theta}_1 = \max\{X_1, \cdots, X_n\}$ $\quad X_1, \cdots, X_n$: i.i.d. Unif $(0, \theta)$

$$E[\hat{\theta}_1] = \int x \cdot \underbrace{f(x)}_{\text{PDF of } \hat{\theta}_1} dx$$

Conti. RV

CDF of $\hat{\theta}_1 = F(y) = P(\hat{\theta}_1 \le y)$

$$= P(\max\{X_1, \cdots, X_n\} \le y)$$

$$= P(X_1 \le y, X_2 \le y, \cdots, X_n \le y)$$

$$\underset{X_1, \cdots, X_n \text{ indep.}}{=} P(X_1 \le y)\, P(X_2 \le y) \cdots P(X_n \le y)$$

$$= \left(\frac{y}{\theta}\right)^n = \theta^{-n} \cdot y^n, \quad 0 \le y \le \theta$$

$$f(y) = n\,\theta^{-n}\, y^{n-1}.$$

$$E[\hat{\theta}_1] = \int_0^\theta x \cdot n\,\theta^{-n}\, x^{n-1}\, dx = n\,\theta^{-n} \frac{1}{n+1} \theta^{n+1}$$

$$= \frac{n}{n+1} \cdot \theta$$

$\hat{\theta}_1$ is NOT unbiased (or biased)

$$E\Big[\underbrace{\frac{n+1}{n} \cdot \hat{\theta}_1}_{= \hat{\theta}_2}\Big] = \theta$$

$$\hat{\theta}_2 = \frac{n+1}{n} \max\{X_1, \cdots, X_n\}$$

is an unbiased estimator for θ.

Q: Another unbiased estimator for θ?

$\text{Unif}(0, \theta) \longrightarrow X_1, \ldots, X_n$

$\hat{\theta}_3$ = a function of $X_1, \ldots, X_n$

$E[\hat{\theta}_3] = \theta$?

To estimate the mean, a natural choice for point estimator is "sample mean"?

$\underline{\underline{\mu}}$ $\qquad$ $\underline{\bar{X} = \frac{1}{n}\sum_i X_i}$

$\uparrow$ unbiased for μ

$E[\bar{X}] = \mu = \frac{\theta}{2}$

$E[\underline{2\bar{X}}] = \theta$

$\hat{\theta}_3 = 2\bar{X}$ unbiased estimator for θ.

Q: Which one is better? $\hat{\theta}_2, \hat{\theta}_3$

RVs: $\hat{\theta}_2$ $\qquad$ $\hat{\theta}_3$

$E[\hat{\theta}_2] = E[\hat{\theta}_3] = \theta$

$\theta \leftarrow$ a number

$$X = \begin{cases} 20, & \text{w.p. } \frac{1}{2} \\ 0, & \text{w.p. } \frac{1}{2} \end{cases} \qquad Y = \begin{cases} 40, & \text{with p. } \frac{1}{4} \\ 0, & \text{with p. } \frac{3}{4} \end{cases}$$

10

Unbiased Estimators

Proposition

If $X_1, X_2, \cdots, X_n$ is a random sample from a distribution with mean μ and variance σ^2, then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{(Sample mean)}$$
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad \text{(Sample variance)}$$

are unbiased estimators of μ and σ^2.

Why

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[X_i]}_{\mu} = \frac{\cancel{n}\cdot\mu}{\cancel{n}}$$

$$= \mu$$

$$E\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]$$
$$= E\left[\sum_{i=1}^{n} \left(X_i^2 - 2X_i \cdot \bar{X} + (\bar{X})^2\right)\right]$$
$$= E\left[\sum_{i=1}^{n} X_i^2\right] - 2E\left[\underbrace{\left(\sum_{i=1}^{n} X_i\right)\cdot \bar{X}}_{= n\cdot\bar{X}^2}\right] + E\left[\underbrace{\sum_{i=1}^{n}(\bar{X})^2}_{= n\cdot\bar{X}^2}\right]$$
$$= E\left[\sum_{i=1}^{n} X_i^2\right] - nE[(\bar{X})^2]$$

$$\mathbb{E}[\overline{X}^2] = (\underline{\mathbb{E}[\overline{X}]})^2 + \underline{Var(\overline{X})}$$

$$Var\left(\frac{1}{n}(X_1 + \cdots + X_n)\right)$$

$$= \frac{1}{n^2} Var(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2}(Var(X_1) + \cdots + Var(X_n))$$

$$= \mu^2 + \frac{\sigma^2}{n}$$

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right]$$

$$= \sum_{i=1}^{n} \underline{\mathbb{E}[X_i^2]} = \sum_{i=1}^{n} \left((\mathbb{E}[X_i])^2 + Var(X_i)\right)$$

$$= \sum_{i=1}^{n} (\mu^2 + \sigma^2) = n \cdot (\mu^2 + \sigma^2)$$

$$\mathbb{E}\left[\sum_{i=1}^{n} (X_i - \overline{X})^2\right] = n(\cancel{\mu^2} + \sigma^2) - n \cdot (\cancel{\mu^2} + \frac{\sigma^2}{n})$$

$$= (n-1)\sigma^2$$

$$\mathbb{E}\Big[\underbrace{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}_{= S^2}\Big] = \sigma^2$$

Estimators with Minimum Variance

Principle of Minimum Variance Unbiased Estimation

Among all estimators of θ that are unbiased, choose the one that has minimum variance.

The resulting $\widehat{\theta}$ is called the minimum variance unbiased estimator (MVUE) of θ.

Estimators with Minimum Variance

Example

Let $X_1, \cdots, X_n$ be a random sample from a uniform distribution on $[0, \theta]$. Consider

$$\widehat{\theta}_1 = \frac{n+1}{n} \max X_i$$
$$\widehat{\theta}_2 = 2\overline{X}. \quad = \frac{2}{n} \cdot \sum_{i=1}^{n} X_i$$

Are they unbiased? Both.

Find the variances of $\widehat{\theta}_1$ and $\widehat{\theta}_2$

$$Var(\hat{\theta}_2) = Var(2 \cdot \overline{X}) = 2^2 \cdot Var(\overline{X}) = 4 \cdot \frac{\sigma^2}{n}$$

$$\sigma^2 = \text{Variance of Unif } [0, \theta] = \frac{\theta^2}{12}$$

$$Var(\hat{\theta}_2) = \frac{\theta^2}{3n}$$

$$Var(\hat{\theta}_1) = Var\Big(\underline{\frac{n+1}{n}} \underbrace{\max\{X_1, \cdots, X_n\}}_{=Y}\Big)$$

$$= \underline{\frac{(n+1)^2}{n^2}} \cdot \underline{Var(Y)}$$

$$\text{PDF of } Y = f(y) = n \cdot \theta^{-n} y^{n-1}, \quad 0 \leq y \leq \theta$$

$$Var(Y) = E[Y^2] - (E[Y])^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2$$

$$E[Y^2] = \int y^2 \cdot f(y)\, dy = \int_0^\theta y^2 \cdot n\, \theta^{-n}\, y^{n-1}\, dy$$

$$= \frac{n}{n+2}\, \theta^2$$

$$E[Y] = \frac{n}{n+1}\, \theta$$

$$Var(Y) = \frac{n}{(n+1)^2 \cdot (n+2)} \cdot \theta^2$$

$$Var(\hat{\theta}_1) = \frac{\cancel{(n+1)^3}}{n^{\cancel{2}}} \cdot \frac{\cancel{n}}{\cancel{(n+1)^2}(n+2)} \cdot \theta^2 = \frac{\theta^2}{n(n+2)}$$

better

$$\frac{\theta^2}{3n} \geq \frac{\theta^2}{n(n+2)}$$

$$3n \leq n(n+2)$$

$$3 \leq n+2 \qquad n \geq 1$$

Estimators with Minimum Variance

Theorem

Let $X_1, \cdots, X_n$ be a random sample from a normal distribution with parameters μ and σ^2. Then the estimator $\overline{X}$ is the MVUE for μ.

That is, $Var(\overline{X}) = \frac{\sigma^2}{n} \leq Var(\hat{\mu})$

for every unbiased estimator $\hat{\mu}$ for μ

Exercise

(6.1-8) In a random sample of 80 components of a certain type, 12 are found to be defective.

1. Give a point estimate of the proportion of all such components that are not defective.
2. ~~2.~~ A system is to be constructed by randomly selecting two of these components and connecting them in series, as shown here.

The series connection implies that the system will function if and only if neither component is defective (i.e., both components work properly).
Estimate the proportion of all such systems that work properly.

1. A point estimator

Population

Ber(p)

$X_1, \dots, X_{80}$ i.i.d. Ber(p)

D N D N

$p = \mathbb{P}($ Component is Not Defective $)$

$\hat{p} = \dfrac{X_1 + \dots + X_{80}}{80} = \bar{X}$ ← unbiased estimator.

Point estimate $= \dfrac{68}{80}$ for p.

Section 2.
Methods of Point Estimation

The Method of Moments

Definition

Let $X_1, \cdots, X_n$ be a random sample from a PMF or PDF $f(x)$.

For $k = 1, 2, \cdots$, the k-th population moment, or k-th moment of the distribution $f(x)$, is $\mathbb{E}[X^k]$.

The k-th sample moment is $\frac{1}{n}(X_1^k + X_2^k + \cdots + X_n^k) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

Ex 2^{nd} moment $= \mathbb{E}[X^2] = \begin{cases} \int x^2 f(x)\,dx \\ \sum x^2 p(x) \end{cases}$

2^{nd} Sample moment $= \frac{1}{n}(X_1^2 + \cdots + X_n^2)$

The Method of Moments

Definition

Let $X_1, \cdots, X_n$ be a random sample from a distribution with PMF or PDF $f(x; \theta_1, \cdots, \theta_m)$, where $\theta_1, \cdots, \theta_m$ are parameters whose values are unknown.

Then the moment estimators

$$\widehat{\theta}_1, \cdots, \widehat{\theta}_m$$

are obtained by equating the first m sample moments to the corresponding first m population moments and solving for $\theta_1, \cdots, \theta_m$.

Population

$\theta_1, \theta_2, \cdots, \theta_m$ $\longrightarrow$ $X_1, \cdots, X_n$

In terms of $\theta_1, \cdots, \theta_m$

$$= E[X^1] = \frac{1}{n}(X_1 + \cdots + X_n)$$

$$= E[X^2] = \frac{1}{n}(X_1^2 + \cdots + X_n^2)$$

$$= E[X^3] = \frac{1}{n}(X_1^3 + \cdots + X_n^3)$$

$$\vdots$$

The Method of Moments

Example

Let $X_1, \cdots, X_n$ represent a random sample of service times of n customers at a certain facility, where the underlying distribution is assumed exponential with parameter λ.

Find the moment estimator for λ.

Estimate 1 parameter

1^{st} moment = 1^{st} sample moment

$$\frac{1}{\lambda} = \mathbb{E}[X] = \frac{1}{n}(X_1 + \cdots + X_n) = \overline{X}$$

Solve for parameter

$$\lambda = \frac{1}{\overline{X}}$$

Define $\hat{\lambda} = \frac{1}{\overline{X}}$, " ← moment estimator for λ.

The Method of Moments

$E[X] = \alpha\beta \quad Var(X) = \alpha\beta^2$

$\uparrow$

Gamma(α, β)

Example

Let $X_1, \cdots, X_n$ be a random sample of size n from a Gamma distribution.

Find the moment estimators for α, β.

Estimate 2 parameters.

$$\begin{cases} 1^{st} \text{ moment} = 1^{st} \text{ sample moment} \\ 2^{nd} \text{ moment} = 2^{nd} \text{ sample moment} \end{cases}$$

$$\begin{cases} E[X] = \frac{1}{n}\sum_1^n X_i = \overline{X} \\ E[X^2] = \frac{1}{n}\sum_1^n X_i^2 \end{cases}$$

$Var(X) + (E[X])^2 =$

$\alpha \cdot \beta^2 + (\alpha\beta)^2$

$$\begin{cases} \overline{X} = \alpha\beta \\ \frac{1}{n}\sum_1^n X_i^2 = \underline{\alpha\beta^2(\alpha+1)} \end{cases} \Rightarrow \begin{array}{l} \alpha\beta = \overline{X} \\ \hline \alpha\beta^2 = \frac{1}{n}\sum X_i^2 - (\overline{X})^2 \end{array}$$

$= \alpha\beta^2 + (\alpha\beta)^2$

$\alpha = \frac{\overline{X}}{\beta}$

Solve for α, β

$$\beta = \frac{\alpha\beta^2}{\alpha\beta} = \frac{\frac{1}{n}\sum X_i^2 - (\overline{X})^2}{\overline{X}}, \quad \alpha = \frac{(\overline{X})^2}{\frac{1}{n}\sum X_i^2 - (\overline{X})^2}$$

$$\hat{\alpha} = \frac{(\bar{X})^2}{\frac{1}{n}\sum X_i^2 - (\bar{X})^2}, \quad \hat{\beta} = \frac{\frac{1}{n}\sum X_i^2 - (\bar{X})^2}{\bar{X}}.$$

Moment estimators for α, β.

Maximum Likelihood Estimation

Definition

Let $X_1, \cdots, X_n$ have joint PMF or PDF

$$f(x_1, x_2, \cdots, x_n; \theta_1, \cdots, \theta_m)$$

where the parameters $\theta_1, \cdots, \theta_m$ have unknown values.

When $x_1, \cdots, x_n$ are the observed sample values and f is regarded as a function of $\theta_1, \cdots, \theta_m$, it is called the likelihood function.

The maximum likelihood estimates (MLE) $\widehat{\theta}_1, \cdots, \widehat{\theta}_m$ are those values of the θ_i's that maximize the likelihood function

When the X_i's are substituted in place of the x_i's, the maximum likelihood estimators result.

Maximum Likelihood Estimation

IDEA : $\begin{cases} \text{Likelihood function} = \text{PDF or PMF} \\ \text{Maximize} \quad \text{in terms of parameters} \end{cases}$

Example

Let $X_1, \cdots, X_n$ be a random sample from Bernoulli distribution.

Find the Likelihood function and the MLE for p.

PMF of Ber(p) ?

$$f(x) = \begin{cases} p & , x = 1 \\ 1-p & , x = 0 \end{cases} = p^x \cdot (1-p)^{1-x}$$

① Likelihood function = joint PMF for $X_1, \cdots, X_n$

$$= f(x_1, \cdots, x_n ; p)$$

$$= p^{x_1} (1-p)^{1-x_1} p^{x_2} (1-p)^{1-x_2} \cdots p^{x_n} \cdot (1-p)^{1-x_n}$$

$$= p^{x_1 + x_2 + \cdots + x_n} (1-p)^{n - \sum x_i}$$

$$L(p) = p^{\sum x_i} (1-p)^{n - \sum x_i} \qquad 0 \leq p \leq 1$$

assume $x_1, \cdots, x_n$ given

② Find maximizer.

$$F(p) = \log L(p) = \left(\sum_1^t x_i\right) \cdot \log p + \left(n - \sum_1^t x_i\right) \cdot \log(1-p)$$

$$F'(p) = \left(\sum_1^t x_i\right) \cdot \frac{1}{p} + \left(n - \sum_1^t x_i\right) \cdot \frac{1}{1-p} \cdot (-1)$$

$$= 0$$

$$\left(\sum_1^t x_i\right) \cdot \frac{1}{p} = \left(n - \sum_1^t x_i\right) \cdot \frac{1}{1-p}$$

$$(1 - \cancel{p})\left(\sum x_i\right) = \left(n - \cancel{\sum_1^t x_i}\right) \cdot p$$

$$\sum_1^t x_i = n \cdot p$$

Solve for p : $\quad p = \frac{1}{n} \sum_1^t x_i$

$$\hat{p}_{MLE} = \frac{1}{n} \sum_1^t x_i = \overline{X} .$$

Maximum Likelihood Estimation

Example $Exp(\lambda)$

Let $X_1, \cdots, X_n$ be a random sample from exponential distribution.

Find the Likelihood function and the MLE for λ.

① Likelihood function = joint PDF for $X_1, \cdots, X_n$

$$= \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \cdots \lambda e^{-\lambda x_n}$$

$$L(\lambda) = \lambda^n e^{-\lambda \sum_1^n x_i}, \quad \lambda > 0$$

② $F(\lambda) = \log L(\lambda)$

$$= n \cdot \log \lambda - \lambda \cdot \sum_1^n x_i$$

$$F'(\lambda) = \frac{n}{\lambda} - \sum_1^n x_i = 0$$

$$\lambda = \frac{n}{\sum x_i}$$

$$\hat{\lambda}_{MLE} = \frac{n}{\sum x_i} = \frac{1}{\overline{X}}.$$

MLE (Maximum Likelihood Estimate)

Likelihood Function = Joint PDF or PMF.

$\underline{X_1, \cdots, X_n}$

$$f(\underbrace{x_1, x_2, \cdots, x_n}_{\text{fixed}} ; \underbrace{\theta_1, \theta_2, \cdots, \theta_m}_{\text{function of parameters}}) = L(\theta_1, \cdots, \theta_m)$$

Maximize $L(\theta_1, \cdots, \theta_m)$

WANT to Find $\underline{\hat{\theta}_1, \cdots, \hat{\theta}_m}$ which

maximize $L(\theta_1, \cdots, \theta_m)$

MLEs for $\theta_1, \cdots, \theta_m$.

Maximum Likelihood Estimation

Example

Let $X_1, \cdots, X_n$ be a random sample from exponential distribution.

Find the Likelihood function and the MLE for λ.

$$f(x_1, \cdots, x_n; \lambda) = (\lambda e^{-\lambda x_1}) \cdot (\lambda \cdot e^{-\lambda x_2}) \cdots (\lambda e^{-\lambda x_n})$$

$$L(\lambda) = \lambda^n \cdot e^{-\lambda(x_1 + \cdots + x_n)} \quad \text{: Likelihood function}$$

$$F(\lambda) = \log L(\lambda) = \log\left(\lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}\right)$$

$$= \log(\lambda^n) + \log\left(e^{-\lambda \sum_i^n x_i}\right)$$

$$= n \cdot \log \lambda - \lambda \sum_i^n x_i .$$

$$F'(\lambda) = n \cdot \frac{1}{\lambda} - \sum_i^n x_i = 0$$

$$\lambda = \frac{n}{\sum_i^n x_i} = \frac{1}{\bar{x}}$$

$$\hat{\lambda}_{MLE} = \frac{1}{\bar{x}} .$$

Maximum Likelihood Estimation

Example

Let $X_1, \cdots, X_n$ be a random sample from normal distribution.

Find the Likelihood function and the MLEs for μ, σ^2.

$$f(x_1, \cdots, x_n ; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \right) \cdots \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \right)$$

$$L(\mu, \sigma^2) = (2\pi)^{-\frac{n}{2}} \cdot (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2}$$

$$F(\mu, \sigma^2) = \log L(\mu, \sigma^2)$$

$$= \log (2\pi)^{-\frac{n}{2}} - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2$$

$\mu = \theta_1$, $\sigma^2 = \theta_2$

$$F(\theta_1, \theta_2) = \log (2\pi)^{-\frac{n}{2}} - \frac{n}{2} \log \theta_2 - \frac{1}{2\theta_2} \sum_1^n (x_i - \theta_1)^2$$

$$\frac{\partial F}{\partial \theta_1} = 0 \quad , \quad \frac{\partial F}{\partial \theta_2} = 0$$

$$\frac{\partial F}{\partial \theta_1} = -\frac{1}{\cancel{2}\theta_2} \cdot \sum_{1} \cdot \cancel{2}(x_i - \theta_1) \cdot (-1)$$

$$= \frac{1}{\theta_2} \sum_{i=1}^{n} (x_i - \theta_1) = 0 \Rightarrow \theta_1 = \overline{x}.$$

$$\frac{\partial F}{\partial \theta_2} = -\frac{n}{2} \cdot \frac{1}{\theta_2} + \frac{1}{2\theta_2^2} \sum_{1} (x_i - \theta_1)^2 = 0$$

$$\frac{n}{2\theta_2} = \frac{1}{2\theta_2^2} \sum_{1} (x_i - \theta_1)^2$$

$$\theta_2 = \frac{1}{n} \sum_{1} (x_i - \theta_1)^2$$

$$\hat{\theta}_1 = \hat{\mu} = \overline{x}$$

$$\hat{\theta}_2 = \widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Exercise

(6.2-23) Two different computer systems are monitored for a total of n weeks.

Let X_i denote the number of breakdowns of the first system during the i-th week, and suppose the X_i's are independent and drawn from a Poisson distribution with parameter μ_1.

Similarly, let Y_i denote the number of breakdowns of the second system during the i-th week, and assume independence with each Y_i Poisson with parameter μ_2.

Derive the MLE's of μ_1, μ_2, and $\mu_1 - \mu_2$

$$X_1, \dots, X_n \sim \text{Pois}(\mu_1)$$

$$P(x_1, \dots, x_n) = \left(e^{-\mu_1} \cdot \frac{\mu_1^{x_1}}{x_1!}\right) \cdots \left(e^{-\mu_1} \frac{\mu_1^{x_n}}{x_n!}\right)$$

$$L(\mu_1) = \frac{e^{-n \cdot \mu_1} \; \mu_1^{\sum_i x_i}}{x_1! \cdots x_n!}$$

$$F(\mu_1) = \log L(\mu_1) = -n \cdot \mu_1 + \left(\sum_i x_i\right) \cdot \log \mu_1 - \log(x_1! \cdots x_n!)$$

$$F'(\mu_1) = -n + \frac{\sum_i x_i}{\mu_1} = 0 \qquad \mu_1 = \overline{X}$$

$$\hat{\mu}_1 = \overline{X} \qquad \hat{\mu}_2 = \overline{Y} \qquad \widehat{\mu_1 - \mu_2} = \hat{\mu}_1 - \hat{\mu}_2 = \overline{X} - \overline{Y}$$

The Invariance Principle

Let $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ be the mle's of the parameters $\theta_1, \theta_2, \ldots, \theta_m$. Then the mle of any function $h(\theta_1, \theta_2, \ldots, \theta_m)$ of these parameters is the function $h(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m)$ of the mle's.