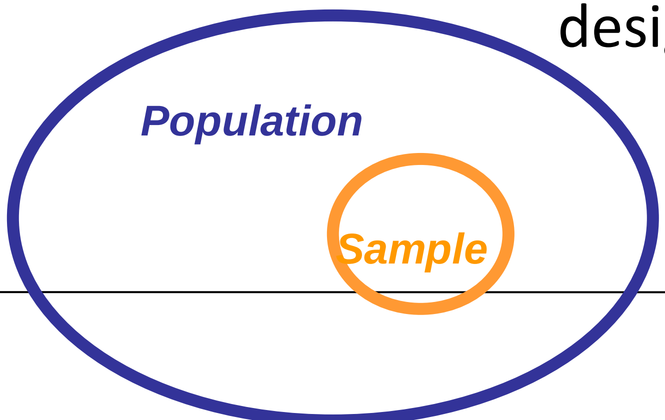
Chapter 7 Point Estimation

Population versus sample

- **Population:** The entire group of individuals in which we are interested but can't usually assess directly.
- A **parameter** is a number describing a characteristic of the population.

Sample: The part of the population we actually examine and for which we do have data.

How well the sample represents the population depends on the sample design.



Mathematically (in this course), $X_1, \ldots, X_n$ is called a random sample (in this course) if they are i.i.d. (independent and identically distributed) random variables with some distribution.

A statistic is a function of the observable random variables in the sample and known constants.

A simple example of estimator

The sample mean of the numeric data set $X_1, X_2, \ldots, X_n$, is

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

The population mean: E(X)

If we assume the components of the data vector are i.i.d., generated from some underlying distribution, the population mean is just E[X].

We can say the sample mean is an estimator of the population mean.

Roughly, an estimator is a statistic that approximates parameter of interest

1. Two general methods of finding point estimator

Method of moments estimator (MME)

Maximum Likelihood estimator (MLE)

General Description

Let $X_1, \ldots, X_n$ be iid sample from $f(x|\theta_1, \ldots, \theta_k)$.
Define

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \mu_1 = EX$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \qquad \mu_2 = EX^2$$

$$\vdots$$

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \qquad \mu_k = EX^k$$

μ_j will typically be a function of $\theta_1, \ldots, \theta_k$. The method of moments estimator is obtained by solving

$$\begin{aligned} m_1 &= \mu_1(\theta_1, \ldots, \theta_k) \\ m_2 &= \mu_2(\theta_1, \ldots, \theta_k) \\ &\vdots \\ m_k &= \mu_k(\theta_1, \ldots, \theta_k) \end{aligned}$$

Example: MME for $(X_1, \ldots, X_n) \sim Unif(0, \theta)$

Remind you:

Gamma distribution:

Definition: a random variable X is said to be gamma distributed with parameters (α, β), denoted as $X \sim Gamma(\alpha, \beta)$ if its pdf is given by

$$f(x) = \begin{cases} \frac{\beta^{\alpha} e^{-\beta x} x^{\alpha - 1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Mean and Variance:

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

Example: Given a random sample of size n from a Gamma distribution, estimate the two parameters by MME.

Maximum Likelihood Estimator (MLE)

Definition of the likelihood function For a sample $X = (X_1, \ldots, X_n)$ with joint pdf or pmf $f(X|\theta)$, the likelihood function is just the pdf or pmf, but think of it as a function of θ:

$$L(\theta|X) = f(X|\theta)$$

In this course, we have always that observations are i.i.d. Therefore, the resulting density for the samples is

$$L(\theta|X) = \prod_{i=1}^{n} f(X_i|\theta)$$

The maximum likelihood estimator (MLE) is just the maximizer $\widehat{\theta}(X)$ of the likelihood function. Often we maximize $l(\theta|X) = \log L(\theta|X)$ instead because it is usually easier.

In real life, for complicated likelihood, MLE is found by optimization software.

HHTTHT- ——

$\underbrace{\hspace{6em}}_{100 \quad 50H\ 50T}$

$\mu\ \theta = \frac{1}{2}$

$p(\text{seq} \mid \mu)$

$\mu = 0.1$

$\mu = 0.9$

Invariance Property of MLE

If $\widehat{\theta}$ is the MLE of θ, then for any function $h(\theta)$, $h(\widehat{\theta})$ is the MLE of $h(\theta)$.

Example: MLE for $(X_1, \dots, X_n) \sim N(\mu, \sigma^2)$

$$f(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$L(\theta | x_1 \dots x_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}}$$

$$\ell(\theta|\dots) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_i (x_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \ell}{\partial \mu} = -\frac{-2\sum_i (x_i-\mu)}{2\sigma^2} = \frac{\sum_i (x_i-\mu)}{\sigma^2} = 0, \quad \left(\sum_i x_i\right) = n\cdot\mu \Rightarrow \underbrace{\hat{\mu}_{MLE} = \frac{\sum x_i}{n} = \overline{X}}$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{\cancel{2}\sigma^2} + \frac{\sum (x_i-\mu)^2}{\cancel{2}\sigma^4} = 0, \quad \Leftrightarrow n\sigma^2 = \sum (x_i-\mu)^2$$

$$\Rightarrow \underbrace{\hat{\sigma}^2 = \frac{1}{n}\sum (x_i - \overline{x})^2}$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2\sum (x_i-\mu)^2}{2\sigma^3} = 0$$

sample var: $S^2 = \frac{1}{n-1}\sum (x_i - \overline{x})^2$

Example: MLE for $X_1, \ldots, X_n \sim Bin(k, p)$, k known, p unknown.

$$P(x_i \mid p) = \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i},$$

$$L(p),\quad L = \left[\prod_{i=1}^{n} \binom{k}{x_i}\right] p^{\Sigma x_i} (1-p)^{\Sigma (k-x_i)}$$

$$\ell = \log \prod_{i=1}^{n} \binom{k}{x_i} + \Sigma x_i \log p + \Sigma (k - x_i) \log (1-p)$$

$$\frac{\partial \ell}{\partial p} = \frac{\Sigma x_i}{p} - \frac{\Sigma (k - x_i)}{1-p} = 0$$

$$\hat{p} = \frac{\Sigma x_i}{nk}$$

Example: MLE for $X_1, \ldots, X_n \sim Unif(0, \theta)$.

$$f(x_i|\theta) = \frac{1}{\theta}$$

$$L(\theta) = \left(\frac{1}{\theta}\right)^n \qquad \hat{\theta} = 0 \,?$$

$$f(x_i|\theta) = \begin{cases} \frac{1}{\theta}, & \text{if } x_i < \theta \\ 0 & \text{if } x_i > \theta \end{cases}$$

$$L(\theta) = \begin{cases} \left(\frac{1}{\theta}\right)^n & \text{if all } x_i \le \theta \\ 0 & \text{o.w.} \end{cases}$$

θ should be $\max\{x_1, \ldots, x_n\}$

$$\hat{\theta} = \max\{x_1, \ldots, x_n\}$$

Example:

(a) Compute the MLE for μ given $(X_1, \ldots, X_n) \sim N(\mu, 1)$.

(b) Compute the MLE for μ^2 given $(X_1, \ldots, X_n) \sim N(\mu, 1)$.

a) $L = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\Sigma(x_i-\mu)^2}{2}}$

$l = -\frac{n}{2}\log 2\pi - \frac{\Sigma(x_i-\mu)^2}{2}$

$\frac{\partial l}{\partial \mu} = \Sigma(x_i-\mu) = 0 \Rightarrow \hat{\mu} = \bar{X}$

b) $\widehat{\mu^2} = \bar{X}^2$

2. Evaluation of Estimators

Definition

An estimator $\widehat{\theta}$ is an unbiased estimator for θ if $E(\widehat{\theta}) = \theta$ (for all θ).

$\widehat{\theta}$ is (weakly) consistent if $\widehat{\theta} \to \theta$ weakly, that is $\lim_{n\to\infty} P(|\widehat{\theta} - \theta| < c) = 1 \; \forall c > 0$.

Definition. The mean squared error (MSE) of an estimator $\widehat{\theta}$ of θ is $E[(\widehat{\theta} - \theta)^2]$.

One property of MSE. $MSE(\widehat{\theta}) = Var(\widehat{\theta}) + bias^2(\widehat{\theta})$ where the bias of an estimator is $E[\widehat{\theta}] - \theta$.

$$E(\hat{\theta}-\theta)^2 \qquad \underbrace{\mu = E\hat{\theta}}$$

$$= E(\hat{\theta}-\mu+\mu-\theta)^2$$

$$= E\left((\hat{\theta}-\mu)^2 + 2(\hat{\theta}-\mu)(\mu-\theta) + (\mu-\theta)^2\right)$$

$$= Var(\hat{\theta}) + \underbrace{(\mu-\theta)^2}_{bias^2} + 2\underbrace{E(\hat{\theta}-\mu)\boxed{(\mu-\theta)}}_{0?}$$

Example: Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$.

$E\bar{X} = \mu$

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

$ES^2 = \sigma^2$

Let $\hat{\sigma}^2 = \frac{1}{n}\sum_i (X_i - \bar{X})^2 = \frac{n-1}{n} S^2$ be the MLE.
$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$, so MLE is biased.

(FYI) All these are consistent.

$X \sim exp(\lambda)$, $EX = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$

Example: Consider a random sample of size 3 from $exp(1/\theta)$: $f(y|\theta) = (1/\theta)e^{-y/\theta}, y > 0$. $\hat{\theta}_1 = Y_1$, $\hat{\theta}_2 = (Y_1 + Y_2)/2$, $\hat{\theta}_3 = (Y_1 + 2Y_2)/3$, $\hat{\theta}_4 = \bar{Y}$ are all unbiased estmators for θ. Compare the variances of these estimators.

$EY_1 = \theta$,

$Var(\hat{\theta}_1) = Var(Y_1) = \theta^2$

$Var(\hat{\theta}_2) = Var\left(\frac{Y_1 + Y_2}{2}\right) = \frac{1}{2} Var(Y_1) = \frac{\theta^2}{2}$

$Var(\hat{\theta}_3) = \frac{1}{9} Var(Y_1) + \frac{4}{9} Var(Y_2) = \frac{5}{9}\theta^2$

$Var(\hat{\theta}_4) = \frac{\theta^2}{3}$

(FYI)

MME and MLE are not necessarily unbiased (although usually consistent)

MME and MLE are usually asymptotically normal