Chapter 7
Statistical Estimation and Sampling Distributions

7.1 Point Estimates

7.1.1 Parameters

- Parameters
  - In statistical inference, the term **parameter** is used to denote a quantity θ, say, that determines the shape of an unknown probability distribution.

  - For example, mean, variance, or a particular quantile of the probability distribution

  - When parameters are unknown, one of the goals of statistical inference is to estimate them.

7.1.2 Statistics

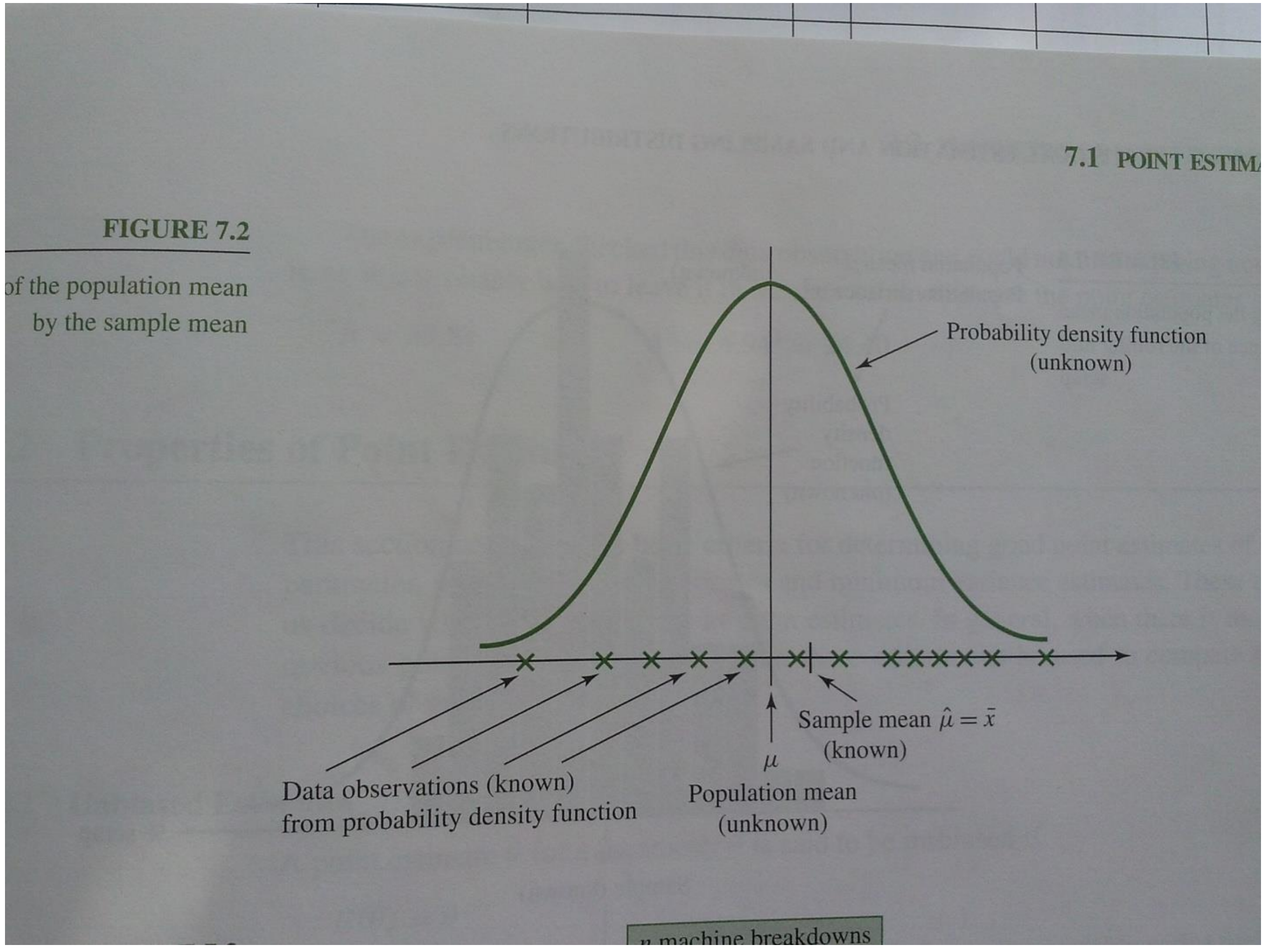
- Statistics

  - Is a function of a random sample. For example, sample mean, sample variance, or a particular sample quantile.

  - Statistics are random variables whose observed values can be calculated from a set of observed data.

7.1.3 Estimation

- Estimation
  - A procedure of “guessing” properties of the population from which data are collected.
  - A point estimate of an unknown parameter is a statistic $\hat{\theta}$ that represents a “best guess” at the value of θ.


7.1 POINT ESTIMA
FIGURE 7.2
of the population mean
by the sample mean
Probability density function
(unknown)
Sample mean $\hat{\mu} = \bar{x}$
(known)
μ
Population mean
(unknown)
Data observations (known)
from probability density function
machine breakdowns

- **Example 1** (Machine breakdowns)
  - How to estimate
    P(machine breakdown due to operator misuse) ?

- **Example 43** (Rolling mill scrap)
  - How to estimate the mean and variance of the probability distribution of % scrap?

7.2 Properties of Point Estimates

7.2.1. Unbiased Estimates

- Definitions
  - A point estimate $\hat{\theta}$ for a parameter θ is said to be unbiased if
$$E(\hat{\theta}) = \theta.$$
  - If a point estimate is not unbiased, then its bias is defined to be
$$bias(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

7.2.1. Unbiased Estimates

- Point estimate of a population mean
- Let $X_1, \cdots, X_n$ be a random sample from a distribution with mean μ. Then the sample mean $\overline{X}$ is an unbiased estimate of μ.

-

- Point estimate of a population variance

  Let $X_1, \cdots, X_n$ be a random sample from a distribution with variance σ^2.
  Then the sample variance S^2 is an unbiased estimate of σ^2.

Proof of $E(S^2) = \sigma^2$

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}((X_i - \mu) - (\bar{X} - \mu))^2$$

$$= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) + n(\bar{X} - \mu)^2$$

$$= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

$$note \;\; E(X_i) = \mu, \;\; E\left((X_i - \mu)^2\right) = Var(X_i) = \sigma^2$$

$$E(\bar{X}) = \mu, \;\; E\left((\bar{X} - \mu)^2\right) = Var(\bar{X}) = \frac{\sigma^2}{n}$$

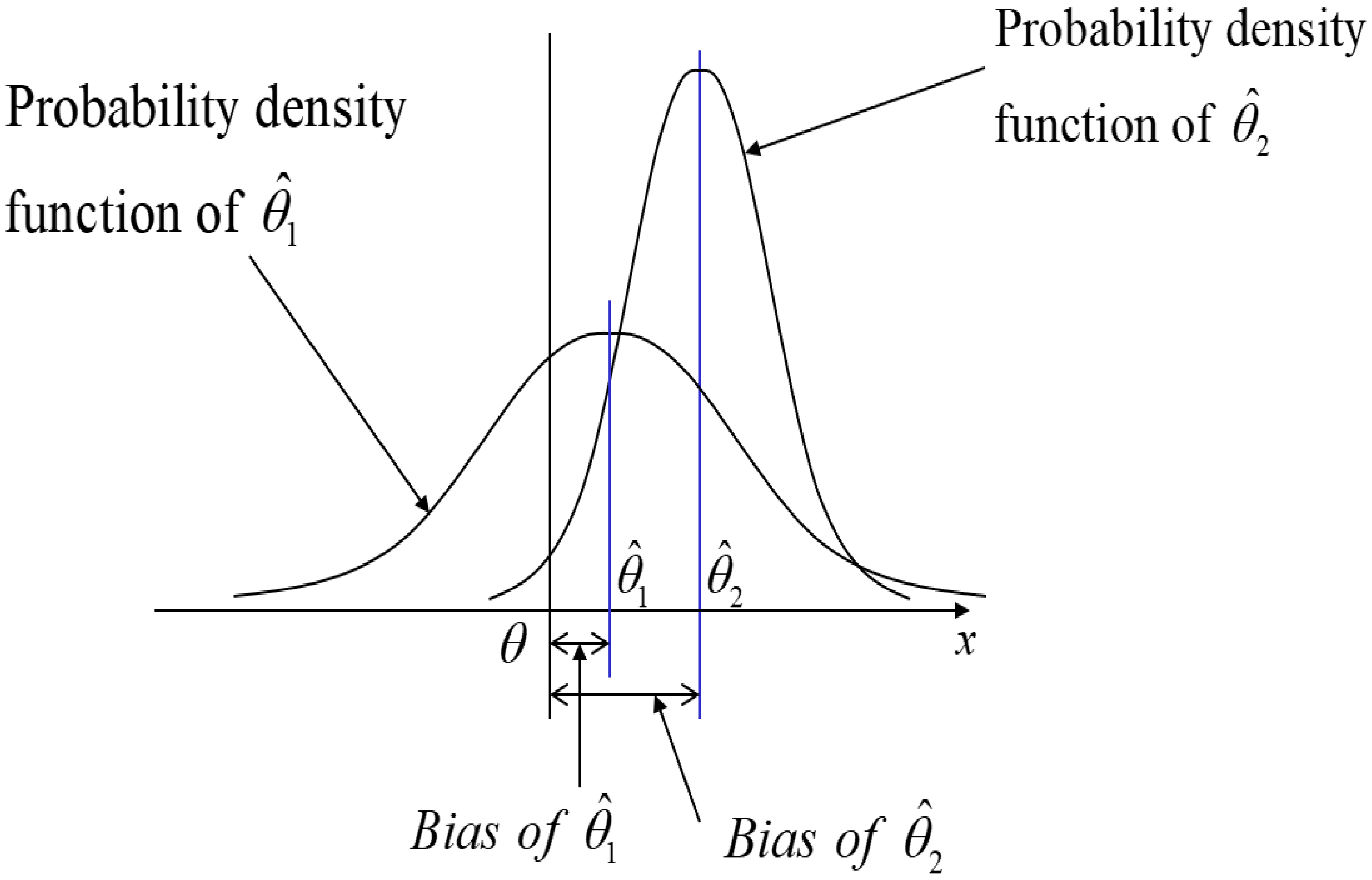
$$E(S^2) = \frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}\sigma^2 - n\left(\frac{\sigma^2}{n}\right)\right) = \sigma^2$$

7.2.2. Minimum Variance Estimates

- An unbiased point estimate whose variance is smaller than any other unbiased point estimate: **minimum variance unbiased estimate (MVUE)**
- Relative efficiency

  The relative efficiency of an unbiased point estimate $\hat{\theta}_1$ to another unbiased estimate $\hat{\theta}_2$ is $\frac{\mathrm{Var}(\hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1)}$.
- Mean squared error (MSE)
  - $\mathrm{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$.
  - $\mathrm{MSE}(\hat{\theta}) = Var(\hat{\theta}) + bias^2(\hat{\theta})$.



Probability density function of $\hat{\theta}_1$
Probability density function of $\hat{\theta}_2$
$\hat{\theta}_1$
$\hat{\theta}_2$
θ
x
Bias of $\hat{\theta}_1$
Bias of $\hat{\theta}_2$

7.3 Sampling Distribution

7.3.1 Sample Proportion

- If $X \sim B(n, p)$, then the sample proportion $\hat{p} = \frac{\mathrm{X}}{\mathrm{n}}$ has approximately the distribution $\mathrm{N}(\mathrm{p}, \frac{\mathrm{p}(1-\mathrm{p})}{\mathrm{n}})$.
- The standard error of $\hat{p}$ is defined as

$$\mathrm{s.e.}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}.$$

When n is large, $\mathrm{s.e.}(\hat{p})$ is approximated by $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

7.3.2 Sample Mean

- Distribution of Sample Mean

  Let $X_1, \cdots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then the Central Limit Theorem says that

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{for large n.}$$

- Standard error of the sample mean
  - The standard error of the sample mean is defined as $\mathrm{s.e.}(\overline{X}) = \frac{\sigma}{\sqrt{n}}$.
  - When σ is not known and n is large, the standard error is replaced with $\frac{s}{\sqrt{n}}$ as an approximate value.

7.3.3 Sample Variance

- Distribution of Sample Variance
- Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Then $\frac{(\mathrm{n}-1)\mathrm{S}^2}{\sigma^2} \sim \chi^2_{n-1}$.

- Fact 7.3.3a

  Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

  Then, $\bar{X}$ and S^2 are independent.

- t-statistic

  Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

  Then

$$T = \frac{\sqrt{n}\left(\overline{X} - \mu\right)}{S} \sim t_{n-1}$$

7.4 Constructing Parameter Estimates

7.4.1 The Method of Moments

- Method of moments point estimate(MME) for One Parameter

  If a data set of observations $x_1, \cdots, x_n$ from a probability distribution that depends on one parameter θ, then the MME $\hat{\theta}$ of θ is found by solving the equation

$$\overline{x} = E(X).$$

- Method of moments point estimates for Two Parameters

The method of moments point estimates (MME) of the two unknown parameters are found by solving

$$\bar{x} = E(X) \text{ and}$$
$$s^2 = Var(X)$$

- Examples

(1) Given a random sample of size n from $Exp'l(\lambda)$, find the MME of λ.

(2) Given a random sample of size n from $\mathrm{N}(\mu, \sigma^2)$, find the MME's of μ and σ^2.

(3) Suppose that the data values, 2.0, 2.4, 3.1, 3.9, 4.5, 4.8, 5.7, 9.9, are obtained from $\mathrm{U}(0, \theta)$. Find the MME $\hat{\theta}$ of θ.

$\overline{x} = 4.5375$ and $\mathrm{E(X)} = \frac{\theta}{2}$. So, $\hat{\theta} = 9.075$.

7.4.2 Maximum Likelihood Estimates

- Maximum Likelihood Estimate for One Parameter

Let $\mathrm{x}_1, \cdots, x_n$ be a data set observed from a distribution $\mathrm{f}(\mathrm{x};\theta)$ depending upon one unknown parameter θ.

Then the maximum likelihood estimate(MLE) $\hat{\theta}$ of the parameter is the value of θ at which the likelihood function below is maximized:

$$L(\theta) = f(x_1;\theta) \times \cdots \times f(x_n;\theta).$$

- Example

  Let $\mathrm{x}_1, \cdots, x_n$ be Bernoulli observations with parameter p.

  Then the MLE of p is $\hat{p} = \frac{\sum_i^n x_i}{n}$.

Maximum Likelihood Estimate for Two Parameters

For two unknown parameters, θ_1 and θ_2, the MLE's of them are the values of the parameters at which the likelihood function is maximized.

- Example

  Suppose we have a data set of size n from $\mathrm{N}(\mu,\ \sigma^2)$. Then the MLE's of μ and σ^2 are obtained as

$$\hat{\mu} = \overline{x} \quad \text{and} \quad \widehat{\sigma^2} = \frac{\sum_i^n (x_i - \overline{x})^2}{n}.$$

7.4.3 Examples

Example 27. Glass Sheet Flaws

- At a glass manufacturing company, 30 randomly selected sheets of glass are inspected. If the distribution of the number of flaws per sheet is Poisson with parameter λ, how should λ be estimated?

- MME

  $E(X) = \lambda$. So $\hat{\lambda}_{MM} = \bar{x}$.

- MLE

  $\hat{\lambda}_{ML} = \bar{x}$.

Example 26: Fish Tagging and Recapture

Suppose a fisherman wants to estimate the fish stock N of a lake and that 34 fish have been tagged and released back into the lake.

If, over a period of time, the fisherman catches 50 fish and 9 of them are tagged, then an intuitive estimate of the total number of fish is

$$\widehat{N} = 34 \times \frac{50}{9} \approx 189.$$

This is under the assumption that the proportion of the tagged fish is roughly equal to the proportion of the fisherman’s catch that is tagged.

- Under the assumption that all the fish are equally likely to be caught, the distribution of the number of the tagged fish X in the fisherman's catch of 50 fish is Hypergeometric with $\mathrm{r} = 34, \mathrm{n} = 50$ and N unknown.

- To find the MME of N:

  $\mathrm{E(X)} = \mathrm{n}\frac{\mathrm{r}}{\mathrm{N}} = 9.$ So $\widehat{N}_{MM} = 50 \times \frac{34}{9} = 188.89.$

- MME is computationally easy to obtain for this distribution.

- Example 36: Bee Colonies

The data of size 13 on the proportion of worker bees that leave a colony with a queen bee is the set of 0.28, 0.32, 0.09, 0.35, 0.45, 0.41, 0.06, 0.16, 0.16, 0.46, 0.35, 0.29, 0.31.

If the entomologist wishes to model this proportion with a Beta distribution, how should the parameters be estimated?

- The Beta pdf:

$$f(x;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1},\ 0 \le x \le 1,\ \ a,b > 0.$$

$$E(X) = \frac{a}{a+b}. \quad Var(X) = \frac{ab}{(a+b+1)(a+b)^2}.$$

- $\overline{x} = 0.3007.\ s^2 = 0.01966.$
- The MME's are obtained by solving

$$\frac{a}{a+b} = \overline{x} \quad \text{and} \quad \frac{ab}{(a+b+1)(a+b)^2} = s^2.$$

- $\hat{a} = 2.92, \qquad \hat{b} = 6.78.$
- The MME's are computationally easier to obtain than the MLE's.

MLE for $\mathrm{U}(0, \Theta)$

- For some distribution, the MLE may not be found by differentiation. You have to look at the curve of the likelihood function itself.

- The MLE of $\theta = max\{X_1, \cdots, X_n\}$.

Chapter Summary

7.1 Point Estimates

7.2 Properties of Point Estimates

Mean squared error

7.3 Sampling Distributions

7.4 Constructing Parameter Estimates

Method of Moments

Method of Maximum Likelihood