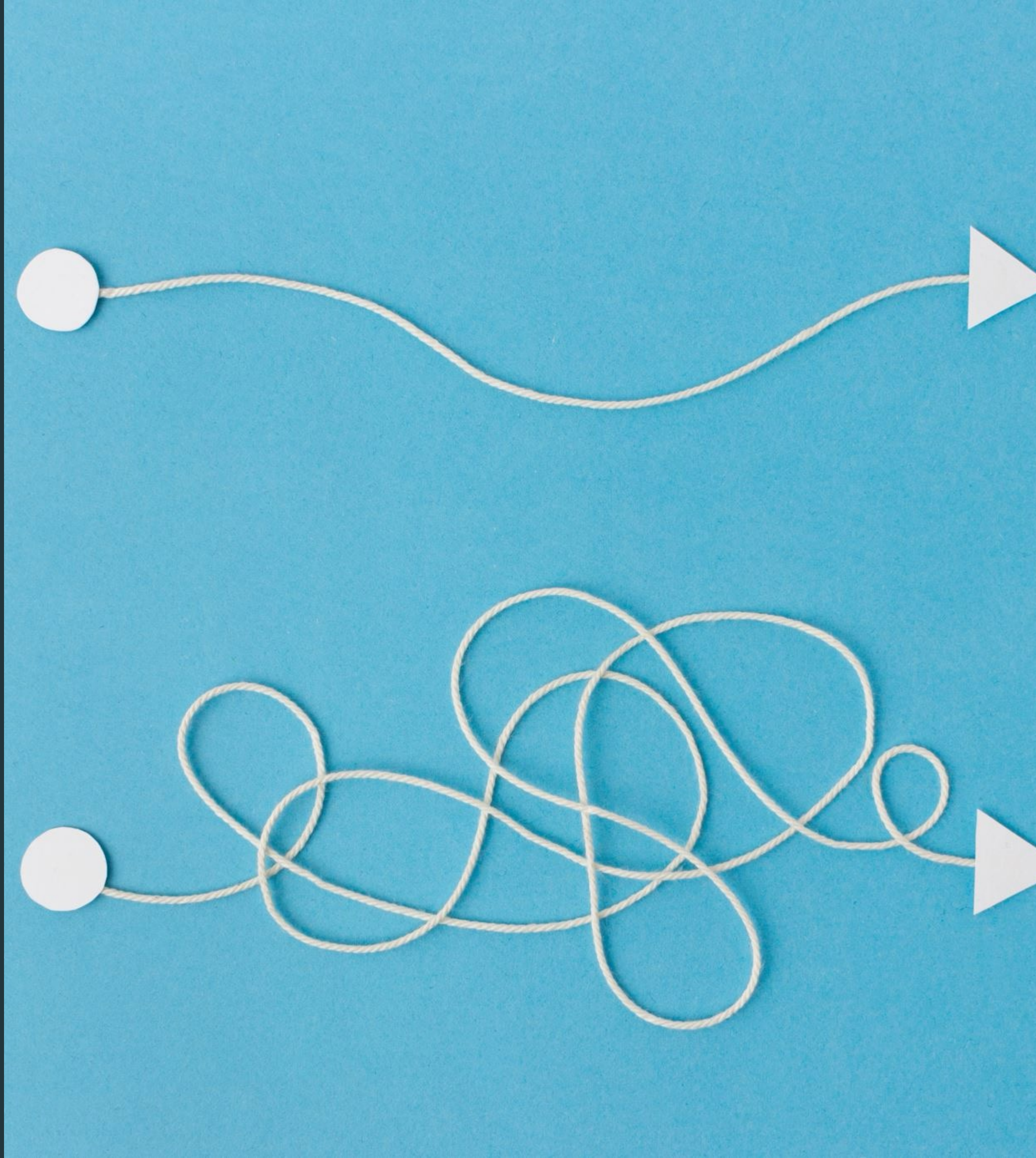
Week 2 – Lecture 1

Estimation of Parameters

and

Fitting of Probability Distributions



Objectives

- Statistics and their distributions
- Estimation of Parameters
- Unbiased Estimators
- Standard Errors of Estimators
- Bootstrap

Statistics and Their Distributions

- Consider selecting two different samples of size n from the same population distribution. The values in the second sample will virtually always differ at least a bit from those in the first sample. For example, a first sample of $n = 3$ cars of a particular type might result in fuel efficiencies $x_1 = 30.7, x_2 = 29.4, x_3 = 31.1$, whereas a second sample may give $x_1 = 28.8, x_2 = 30.0, x_3 = 32.5$. Before we obtain data, there is uncertainty about the value of each sample observation.
- Because of this uncertainty, before the data becomes available we now regard each observation as a random variable and denote the sample by $X_1, X_2, \ldots, X_n$.

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. The distribution of a statistic is called its sampling distribution.

Example: $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$, $M_2 = \frac{X_1^2+X_2^2+\cdots+X_n^2}{n}$

Statistics and Their Distributions (cont.)

Definition: The random variables $X_1, X_2, \dots, X_n$ are said to form a (simple) random sample of size *n* if

1. The X_i's are independent rv's.
2. Every X_i has the same probability distribution.

Property 1:

Let $X_1, X_2, \dots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

1. $E(\bar{X}) = \mu_{\bar{\boldsymbol{X}}} = \mu$
2. $Var(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$, and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
3. The sample total $T = X_1 + X_2 + \cdots + X_n$ has $E(T) = n\mu$, $Var(T) = n\sigma^2$ and $\sigma_T = \sqrt{n}\,\sigma$.

Statistics and Their Distributions (cont.)

Property 2:

Let $X_1, X_2, \ldots, X_n$ be a random sample **from a normal distribution** with mean value μ and standard deviation σ. Then for *any n, X* is normally distributed (with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$), as is *T* (with mean $n\mu$ and standard deviation $\sqrt{n}\,\sigma$).

The Central Limit Theorem (CLT):

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then if *n* is sufficiently large ($n > 30$), $\bar{X}$ has approximately a normal distribution with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, and T also has approximately a normal distribution with $\mu_T = n\mu$, and $\sigma_T = \sqrt{n}\,\sigma$. The larger the value of n, the better the approximation.

Estimation of Parameters

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,..., x_{10}. The sample mean breaking strength $\bar{x}$ could then be used to draw a conclusion about the value of μ. Similarly, if σ^2 is the variance of the breaking strength distribution (population variance, another parameter), the value of the sample variance s^2 can be used to infer something about σ^2.

- Statistical inference is almost always directed toward drawing some type of conclusion about one or more parameters (population characteristics).
- To do so requires that an investigator obtain sample data from each of the populations under study.
- Conclusions can then be based on the computed values of various sample quantities.

Estimation of Parameters (cont.)

Notation:

- From now on, we will use θ to denote a population parameter. θ can denote μ, or σ, or σ^2, etc, based on the context.

Definition:

- A point estimate of a parameter θ is a single number that can be regarded as a sensible value for θ.
- A point estimate of a parameter 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, denoted by $\hat{\theta}$. Note that $\hat{\theta}$ is also a random variable.

Example:

We can use the statistic $\bar{X} = \frac{\boldsymbol{X_1+X_2+\cdots+X_n}}{\boldsymbol{n}}$ to estimate the parameter μ, which is the population mean. Then $\bar{X}$ is a point estimator of μ, and we usually write $\hat{\mu} = \boldsymbol{\bar{X}}$. The statement "a point estimate of μ is 5.77" can be written concisely as $\hat{\mu} = \mathbf{5.77}$.

Estimation of Parameters (cont.)

Example:

An automobile manufacturer has developed a new type of bumper, which is sup posed 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 = the number of crashes that result in no visible damage to the automobile. The parameter to be estimated is p = the pro portion of all such crashes that result in no damage [alternatively, p = P(no damage in a single crash)]. If X is observed to be x = 15, the most reasonable estimator and estimate are

Estimator $\hat{p} = \frac{X}{n}$ Estimate $= \frac{x}{n} = \frac{15}{25} = 0.6$

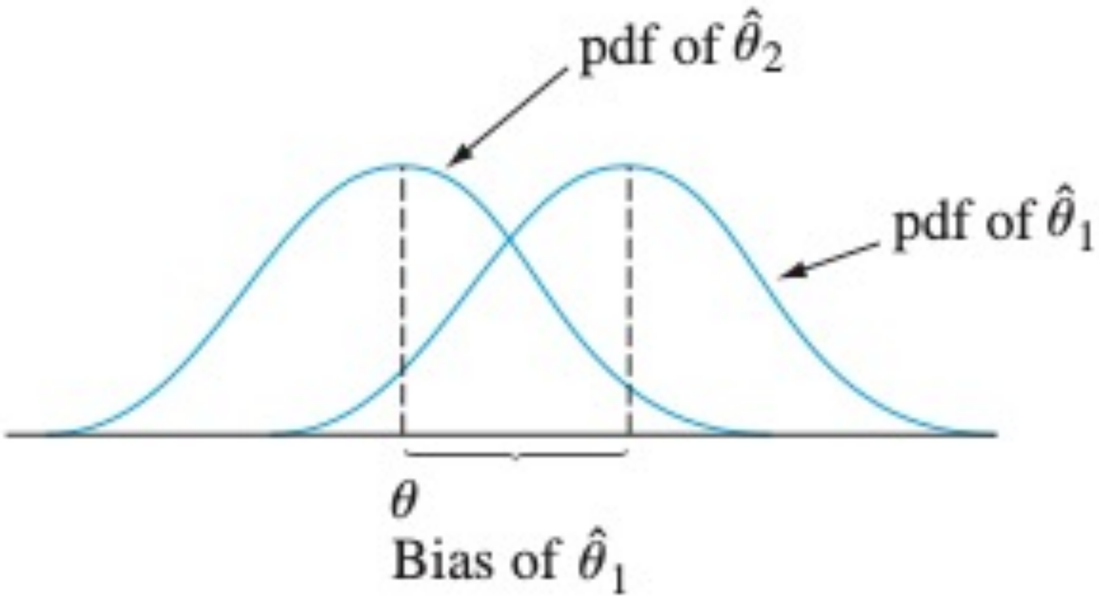
Estimation of Parameters (cont.)

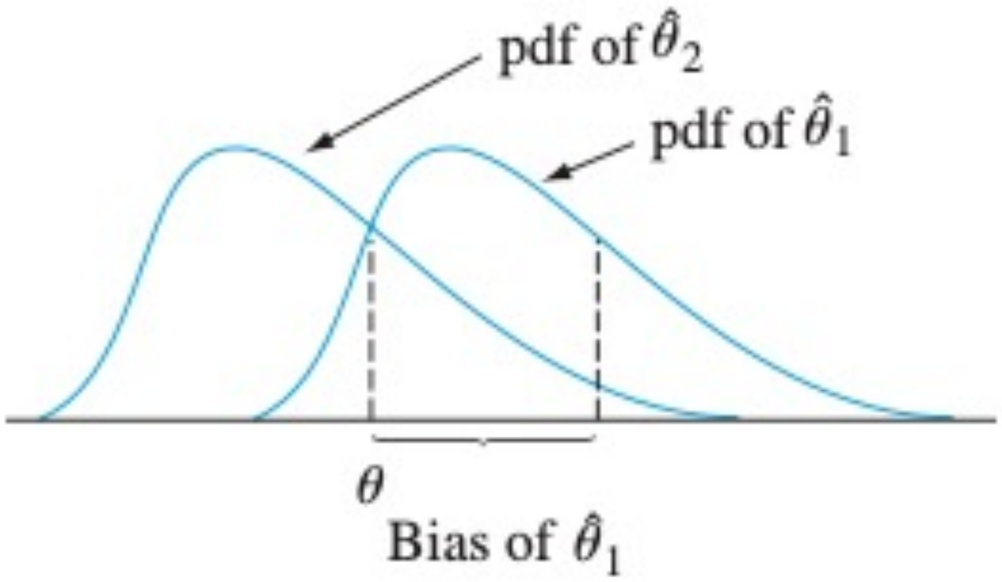
- In most problems, though, there will be more than one reasonable estimator.
- For example, if the data are (approximately) symmetric, then all of the following can be used to estimate μ.

  1) $\hat{\mu} = \bar{X}$
  2) $\hat{\mu} = \tilde{X}$
  3) $\hat{\mu} = \frac{\min X_i + \max X_i}{2}$
  4) $\hat{\mu} =$ 10% trimmed mean
  5) etc

Unbiased Estimators

Definition: A point estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if $E(\hat{\theta}) = \theta$ for every possible value of θ. If $\hat{\theta}$ is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the bias of $\hat{\theta}$.





Unbiased Estimators (cont.)

Example:

1) If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ, then $\bar{X}$ is an unbiased estimator of μ. If in addition the distribution is continuous and symmetric, then $\tilde{X}$, and any trimmed mean are also unbiased estimators of μ.

2) Let $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ and variance σ^2. Then the estimator

$$\hat{\sigma}^2 = S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

is unbiased for estimating σ^2.

Unbiased Estimators (cont.)

Choosing Estimators:

- When choosing among several different estimators of θ, select one that is unbiased.
- Among all estimators of θ that are unbiased, choose the one that has minimum variance. The resulting $\hat{\theta}$ is called the minimum variance unbiased estimator (MVUE) of θ.
- The best estimator for a parameter depends on many factors, especially the distribution that is being sampled.

Example:

Let $X_1, \dots, X_n$ be a random sample from a normal distribution with parameters μ and σ. Then the estimator $\hat{\mu} = \bar{X}$ is the MVUE for μ.

Standard Error of An Estimator

Definition: The standard error of an estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}} = \sqrt{Var(\hat{\theta})}$. It is the magnitude of a typical deviation between an estimate and the value of θ.

- If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into $\sigma_{\hat{\theta}}$ yields the estimated standard error (estimated standard deviation) of the estimator.
- The estimated standard error can be denoted either by $\hat{\sigma}_{\hat{\theta}}$ (the ˆ over σ emphasizes that $\sigma_{\hat{\theta}}$ is being estimated) or by $s_{\hat{\theta}}$.

Standard Error of An Estimator (cont.)

Example 1:

Assuming that breakdown voltage is normally distributed, $\hat{\mu} = \overline{X}$ is the best estimator of μ. If the value of σ is known to be 1.5, the standard error of $\overline{X}$ is $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 1.5/\sqrt{20} = .335$. If, as is usually the case, the value of σ is unknown, the estimate $\hat{\sigma} = s = 1.462$ is substituted into $\sigma_{\overline{X}}$ to obtain the estimated standard error $\hat{\sigma}_{\overline{X}} = s_{\overline{X}} = s/\sqrt{n} = 1.462/\sqrt{20} = .327$.

Example 2:

The standard error of $\hat{p} = X/n$ is

$$\sigma_{\hat{p}} = \sqrt{V(X/n)} = \sqrt{\frac{V(X)}{n^2}} = \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}}$$

Since p and $q = 1 - p$ are unknown (else why estimate?), we substitute $\hat{p} = x/n$ and $\hat{q} = 1 - x/n$ into $\sigma_{\hat{p}}$, yielding the estimated standard error $\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n} = \sqrt{(.6)(.4)/25} = .098$. Alternatively, since the largest value of pq is attained when $p = q = .5$, an upper bound on the standard error is $\sqrt{1/(4n)} = .10$.

Bootstrap

Suppose that the population pdf is f $(x; \theta)$, a member of a particular parametric family, and that data $x_1, x_2, \ldots, x_n$ gives $\hat{\theta} = 21.7$. We now use statistical software to obtain "bootstrap samples" from the pdf f(x; 21.7), and for each sample calculate a "bootstrap estimate" $\hat{\theta}^*$:

First bootstrap sample: $x_1^*, x_2^*, \ldots, x_n^*$; estimate $= \hat{\theta}_1^*$

Second bootstrap sample: $x_1^*, x_2^*, \ldots, x_n^*$; estimate $= \hat{\theta}_2^*$

$\vdots$

Bth bootstrap sample: $x_1^*, x_2^*, \ldots, x_n^*$; estimate $= \hat{\theta}_B^*$

$B = 100$ or 200 is often used. Now let $\bar{\theta}^* = \Sigma\hat{\theta}_i^*/B$, the sample mean of the bootstrap estimates. The **bootstrap estimate** of $\hat{\theta}$'s standard error is now just the sample standard deviation of the $\hat{\theta}_i^*$'s:

$$s_{\hat{\theta}} = \sqrt{\frac{1}{B-1}\Sigma(\hat{\theta}_i^* - \bar{\theta}^*)^2}$$

Bootstrap (cont.)

Example:

A theoretical model suggests that X, the time to breakdown of an insulating fluid between electrodes at a particular voltage, has $f(x; \lambda) = \lambda e^{-\lambda x}$, an exponential distribution. A random sample of $n = 10$ breakdown times (min) gives the following data:

41.53 18.73 2.99 30.34 12.33 117.52 73.02 223.63 4.00 26.78

Since $E(X) = 1/\lambda, E(\overline{X}) = 1/\lambda$, so a reasonable estimate of λ is $\hat{\lambda} = 1/\overline{x} = 1/55.087 = .018153$. We then used a statistical computer package to obtain $B = 100$ bootstrap samples, each of size 10, from $f(x; .018153)$. The first such sample was 41.00, 109.70, 16.78, 6.31, 6.76, 5.62, 60.96, 78.81, 192.25, 27.61, from which $\Sigma x_i^* = 545.8$ and $\hat{\lambda}_1^* = 1/54.58 = .01832$. The average of the 100 bootstrap estimates is $\overline{\lambda}^* = .02153$, and the sample standard deviation of these 100 estimates is $s_{\hat{\lambda}} = .0091$, the bootstrap estimate of $\hat{\lambda}$'s standard error. A histogram of the 100$\hat{\lambda}_i^*$'s was somewhat positively skewed, suggesting that the sampling distribution of $\hat{\lambda}$ also has this property.