STAT333: Statistical Methods I

Lecture Notes

Lecturer
Samuel Iddi, PhD

Department of Statistics and Actuarial Science
University of Ghana

siddi@ug.edu.gh

February 28, 2022

Learning Objectives:

- ◇ Explain the nature and objective of a statistical inquiry
- ◇ Explain the process of a statistical cycle
- ◇ State and explain the principles of estimation
- ◇ Identify the appropriate method of estimation for a given statistical problem
- ◇ Assess and choose best estimators.

Course Syllabus:

- ◇ Introduction: general context.
- ◇ Sampling distributions: sample statistics, distributions, Law of large numbers, Central limit theory.
- ◇ Sampling distributions: distributions associated with normal samples, Gamma family and associated sampling distribution
- ◇ Estimation: frequency substitution, method of moments, least squares, weighted least squares, maximum likelihood.

Course Syllabus:

- ◇ Criteria of estimation: mean absolute error, unbiased estimators, standard error, relative efficiency, absolute measure of efficiency.
- ◇ Interval estimation: error bounds, confidence intervals, pivot function
- ◇ Interval estimation: one sample and two sample problems
- ◇ Introduction to Bayesian Inference (Bayes' Theorem, priors, posterior distribution, credible intervals)
- ◇ Introduction to statistical simulations (Monte Carlo Integration techniques)

Textbook:

1. Odoom, S. I. K. (2007). Statistical Methods: basic concepts and selected applications. *Lecture Notes*. Department of Statistics, University of Ghana.

Reference:

1. Hoel, P. G., Port, S. C., and Stone, C. J. (1972). *Introduction to statistical theory*. Boston: Houghton-Mifflin.
2. Baglivo, J. A. (2005). *Mathematica Laboratories for Mathematical Statistics: Emphasizing Simulation and Computer Intensive Methods*. Philadelphia: SIAM.
3. Ramachandran, K. M. and Tsokos, C. P. (2009). *Mathematical Statistics with Applications*. London:Academic Press.
4. Wackerly, D. D., Mendenhall III, W. and Scheaffer, R. L. (2008). *Mathematical Statistics with Applications*. 7th Ed. Duxbury: Thomson Brooks/Cole.

Office hour:

Friday: 9:30am to 1:00pm or by appointment.

Phone Number: 0267603243, Office: 208

E-mail: `siddi@ug.edu.gh`; `isamuel.gh@gmail.com`
(Sending an email is the easiest way to contact me)

Personal Website: `www.samueliddi.com`

Teaching Assistants

TA: Darhlilove Botchway

Tutorial: Friday, 9:30pm-11:20pm

Phone number:0553536320

E-mail: darhlilove16@gmail.com

Grading

Exercises/assignments (10%), Pop quizzes (10%), Interim Assessment (30%) and Final Exams (50%).

Guidelines

- Homeworks should be submitted **one** week from the day assigned.
- Late submission will **not** be accepted.
- Duplicate solutions will **not** be graded.
- TA will discussion solutions during tutorial hours.

Statistical methods

- Data are produced in many areas of application.
- For example
  - Health: testing the effect of new drug
  - Politics: questioning people attitude towards a government.
  - Market research: assess consumers preference of product
- Data are usually sampled from a conceivable population usually called **target population**.

Statistical methods

- The **statistical cycle**, in any statistical investigation, involves one or more of the following stages
  - collection of data
  - data wrangling (also called data cleaning or munging)
  - data organization and presentation (using tables, graphs, etc.)$\Rightarrow$ exploratory
  - analysis and interpretation $\Rightarrow$ inferential
- The design of a study, sampling strategy, and appropriate data collection methods are essential part of a research process.

Sampling

- Practically, it may not be feasible to collect and study the entire population.
- This may be due to limitations in **time** and **resources**
- Thus we obtain a **sample**, any set of measurements or observations that forms part of (statistical) **population.**
- A statistical population is the complete set of measurements or observations that is conceivable in any given situation.
- Each distinct measurement or observation in the sample or population is called a **unit.**
- Inference regarding the population is made from the data obtained from the sample.

Statistical model

- The information or data we collect are used to construct a model.
- A model forms the bridge between the problem under investigation and statistical theory.
- **A statistical model** summarizes relevant information on the essential characteristics of the situation and of the available data.
- In other words, statistical models uses sample information to create a plan upon which the population data can be generated.

Measurement Model

- Model built on a sample of repeated measurements of a variable on an object
- No target population in the usual sense: the set of measurements obtained is not a purposefully selected sample from a population of measurements.
- Population can be visualized as the set of all values that would result if the measurement process were repeated indefinitely.
- The values obtained in the sequence of measurements will usually vary. This may be due to
  - limited precision of the measurement process
  - fluctuations in the conditions (eg. temperature, humidity)

Measurement Model

- ◇ Each observation X resulting from a measurement can be therefore be represented as

$$X = \mu + \epsilon$$

where ϵ denotes the error of measurements.

- ◇ The usual assumptions about the distribution of values ϵ are as follows:

(i) The errors are independent of one another

(ii) the chance of getting an error of a particular value, at one determination is the same as at another.

(iii) the common distribution of errors is continuous, has mean zero, and is symmetric about zero.

(iv) sometimes the common distribution of errors is assumed to be a Normal distribution with mean zero and a variance σ^2 that is unknown

Population model

- Certain inference may involve some kind of relationship between selected units and a larger population.
- A reliable way to establish such a relationship is to select the items by, for example, **simple random sampling**.
- **Population model** is a statistical model involving selection of subjects from a larger population, according to a well-defined random sampling procedure.
- Inferences made about population based on the selected sample are valid.
- An example is when we try to determine the quality of a batch of manufactured articles.

Randomization Model

- ◇ Randomization model is based on available subjects at a time rather than a random sample from a well-defined population.
- ◇ Commonly in experimental studies in medicine and social sciences where random assignment of treatement to subjects occur.
- ◇ Inferences for models based solely on randomization depend on the fact that all possible allocations of subjects to treatment and control are equally likely.
- ◇ This is the only chance consideration involvement as opposed to other models where chance elements may be introduced in random selection from a population or imposition of a probability distribution on the actual measurements taken.

Parametric Model

- A model is parametric if stringent or strong assumptions are made about the underlying distribution of the population from which the sample is taken.

Advantages

- Parametric models are sharper compared to non-parametric models
- It is easier to assess the accuracy of inferences in parametric models.
- since it has finite parameters, it is possible to estimate the distribution.

Non-Parametric Models

- Non-parametric models are based on weaker, more general assumptions as compared to parametric models.
- Methods based on them are less sharp and often less efficient.
- Accessing accuracy of inference is mostly impossible

Sampling distribution

- Two major problems of statistical inference
  - Estimation
  - Hypothesis Testing
- In **estimation**, we compute numerical characteristics from data that can be used to estimate the parameters of a model.
- From a sample, numerical characteristics are used to estimate model parameters and decide between two **hypothesis**.
- Both methods require computing numerical characteristics from sample.

Sampling distribution

- Let $\boldsymbol{x} = (x_1, x_2, x_3, \ldots, x_n)$ be a vector of sample values whose n is the sample size.
- Sample statistic: a sample characteristic that is a function of this sample vector.
- Sample statistics will be different for different samples and as such are also random.
- We can assign a probability distribution to that sample statistic that reflects this behaviour
- This is called **sampling distribution**

Sample statistics

- Examples of sample statistics:

$$
\begin{aligned}
\bar{x}_n &= \frac{1}{n}\Sigma_{i=1}^{n} x_i \longrightarrow \text{ Sample mean} \\
s_{xx} &= \Sigma_{i=1}^{n}(x_i - \bar{x})^2 \longrightarrow \text{ Sum of squares of deviations} \\
s^2 &= \frac{1}{n}\Sigma_{i=1}^{n}(x_i - \bar{x})^2 \longrightarrow \text{ Sample variance}
\end{aligned}
$$

- Let $t(X)$ or $t_n(X)$ denote a sample statistic
- The probability distribution of the sample statistic $t_n(X)$ is sampling distribution.

Some general results on sampling distributions

- Let $x_1, x_2, x_3, \ldots, x_n$ be sample of independent observations mean μ and variance σ^2 (Nature of underlying distribution is unknown).

- Theorem 1: The distribution of the sample mean, $\bar{x}$, over all possible samples of size n has mean and variance,

$$E(\bar{x}_n) = \mu \text{ and } Var(\bar{x}_n) = \frac{\sigma^2}{n}, \text{ respectively}$$

Some general results on sampling distributions

◇ <u>Theorem 2:</u> The statistic, $s_{xx} = \Sigma_{i=1}^{n}(x_i - \bar{x})^2$ has mean,

$$E(s_{xx}) = (n-1)\sigma^2$$

and variance

$$var(s_{xx}) = \frac{(n-1)^2}{n}\mu_4 - \frac{(n-1)(n-3)\sigma^4}{n}$$

where μ_4 is the fourth central moment of the parent distribution.

Some general results on sampling distributions

- Corollary: From theorem 2, it follows that, over all samples of size n, the sample variance, $s^2 = \frac{1}{n}\Sigma_{i=1}^{n}(x_i - \bar{x})^2$, has mean and variance;

$$\begin{aligned} E(s^2) &= \frac{(n-1)}{n}\sigma^2 \quad \text{and} \\ Var(s^2) &= \frac{(n-1)^2}{n^3}\mu_4 - \frac{(n-1)(n-3)}{n^3}\sigma^4 \end{aligned}$$

- The statistics, $\hat{\sigma}^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(x_i - \bar{x})^2$, is more commonly used sample variance.

$$\begin{aligned} E(\hat{\sigma}^2) &= \sigma^2 \\ Var(\hat{\sigma}^2) &= \frac{\mu_4}{n} - \frac{(n-3)}{n(n-1)}\sigma^4 \end{aligned}$$

Weak Law of Large numbers

- **Chebychev's theorem**: The probability that the random variable x can assume values which lies within (or outside) any specific distance d from it's mean is given by

$$P(|x-\mu| \geq d) \leq \frac{\sigma^2}{d^2} \quad \text{or} \quad P(|x-\mu| \leq d) \geq 1-\frac{\sigma^2}{d^2}$$

- If we apply the Chebychev's theorem to the random variable $\bar{x}_n$ which has mean μ and variance $\frac{\sigma^2}{n}$, then

$$P(|\bar{x}_n-\mu| \geq d) \leq \frac{\sigma^2}{nd^2} \quad \text{or} \quad P(|\bar{x}_n-\mu| \leq d) \geq 1-\frac{\sigma^2}{nd^2}$$

Weak Law of Large numbers

- If we let $n \to \infty$, then

$$P(|\bar{x}_n - \mu| \leq d) \to 1 \text{ as } n \to \infty$$

  no matter how large or small the value of d is.

- **Law of Large numbers**: By taking sufficiently large sample, then with probability close to 1 we can ensure that the sample mean $\bar{x}_n$ is close to the population mean, μ.

Central Limit Theorem

- Another important results of probability theorem is the CLT.
- For $\bar{x}_n$ with mean μ and variance $\frac{\sigma^2}{n}$, define the standardized sample mean

$$z_n = \frac{\bar{x}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- As $n \to \infty$, the distribution of z_n gets closer to the standard normal distribution with mean 0, and variance 1. i.e $z_n \sim \mathcal{N}(0, 1)$

Special Cases

- If $X \sim B(n, , p)$ then for large n,

$$Z_n = \frac{X - np}{\sqrt{npq}} \sim \mathcal{N}(0, 1)$$

- If $X \sim \text{Poi}(\lambda)$, then as $\lambda \to \infty$,

$$\frac{X - \lambda}{\sqrt{\lambda}} \sim \mathcal{N}(0, 1)$$

Distributions associated with normal samples

- We look at sampling distribution of statistics associated with random samples from a normal distribution.
- Suppose $X = (x_1, x_2, x_3, \ldots, x_n)$ is a sample of size n from a normally distributed population with mean μ and variance σ^2 ($X \sim \mathcal{N}(\mu, \sigma^2)$).
- Since $\bar{x}_n$ is a linear combination of x_i which are normally distributed, it follows by the **reproductive property** that

$$\bar{x}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) .$$

and

$$Z = \frac{(\bar{x}_n - \mu)\sqrt{n}}{\sigma} \sim \mathcal{N}(0, 1)$$

Linear statistics of normal sample

- A linear function of the observations in a sample is called a **linear statistics.**
- Let $X = (x_1, x_2, x_3, \ldots, x_n)$. The general form of a linear statistics is

$$L(X) = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

where $a_0, a_1, \ldots, a_n$ are constants, some of which may be zero.

Linear statistics of normal sample

◇ Theorem: If X is a random sample of size n, from $\mathcal{N}(\mu, \sigma^2)$, then the linear statistics $L(X)$ has the sampling distribution which is normal with mean and variance given respectively by

$$
\begin{aligned}
E(L) &= a_0 + \mu \sum_{i=1}^{n} a_i \\
\text{var}(L) &= \sigma^2 \sum_{i=1}^{n} a_i^2
\end{aligned}
$$

Examples

- If X has a normal distribution, $\mathcal{N}(\mu, \sigma^2)$ then $Y = a + bX$ has a normal distribution, $\mathcal{N}(a + b\mu, b^2\sigma^2)$.

- If $X_1, X_2, \ldots, X_k$ are independent normal variables with respective means $\mu_1, \mu_2, \ldots, \mu_k$ and variances, $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ then $Y = X_1 + X_2 + \cdots + X_k$ also has a normal distribution with mean and variance as follows

$$\begin{aligned} \mu &= \mu_1 + \mu_2 + \cdots + \mu_k \\ \sigma^2 &= \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2 \end{aligned}$$

Chi-square distribution

- Let $X_1, X_2, \ldots, X_n$ be n indepedent random variables each of which is standard normal $\mathcal{N}(0, 1)$
- Let $U = \sum_i^n X_i^2$
- The probability distribution of U is called the **chi-square** (χ^2) distribution with n degrees of freedom and is denoted by $\chi^2_{(n)}$
- The mean and variance of the χ^2_n are respectively

$$
\begin{aligned}
E(U) &= n \\
\text{var}(U) &= 2n
\end{aligned}
$$

Chi-square distribution

- Theorem:

  If U and V are independent random variables with distributions χ^2_m and χ^2_n, respectively, then

  $$W = U + V$$

  has a chi-square distribution with $m + n$ degrees of freedom.

- That is, $W \sim \chi^2_{m+n}$.

Sampling distribution of sum of squares of normally distributed random sample

- Let $x_1, x_2, \ldots, x_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$ and let

$$z_1 = \frac{x_1 - \mu}{\sigma}, z_2 = \frac{x_2 - \mu}{\sigma}, \ldots, z_n = \frac{x_n - \mu}{\sigma}$$

  then $z_1, z_2, \ldots, z_n$ are all standard normal random variables and are mutually independent.

- It follows from the definition of the chi-square distribution that

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

  has a χ_n^2 distribution.

Sampling distribution of sum of squares of normally distributed random sample

- If the sample observations are standardized by the sample mean, $\bar{x}$ such that

$$z_1' = \frac{x_1 - \bar{x}}{\sigma}, z_2' = \frac{x_2 - \bar{x}}{\sigma}, \ldots, z_n' = \frac{x_n - \bar{x}}{\sigma}$$

  then

  (1) $z_1', z_2', \ldots, z_n'$ are no longer standard normal although they still have a normal distribution

  (2) $z_1', z_2', \ldots, z_n'$ are no longer independent

- It implies that, the corresponding sum of squares

$$U' = \sum_{i=1}^{n} z_i^2 = \frac{1}{\sigma} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{\sigma^2} S_{xx}$$

  is **not** χ_n^2.

Sampling distribution of sum of squares of normally distributed random sample

- In fact, U' follows a chi-square distribution but the degree of freedom is not n but $(n-1)$.
- Theorem:

  For random sample of size n drawn from a normal distribution $\mathcal{N}(\mu, \sigma^2)$

  (i) S_{xx} and $\bar{x}$ are independent sample statistics.

  (ii) the distribution of $\frac{1}{\sigma^2}S_{xx}$ is χ^2_{n-1} or $S_{xx} \sim \sigma^2\chi^2_{n-1}$.

Proof of (ii)

- We can express S_{xx} as follows:

$$\begin{aligned} S_{xx} &= \Sigma(x_i - \bar{x})^2 = \Sigma[(x_i - \mu) - (\bar{x} - \mu)]^2 \\ &= \Sigma(x_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2(\bar{x} - \mu)(\Sigma x_i - n\mu) \\ &= \Sigma(x_i - \mu)^2 - n(\bar{x} - \mu)^2 \end{aligned}$$

- Dividing through by σ^2 yields

$$\frac{S_{xx}}{\sigma^2} = \frac{\Sigma(x_i - \mu)^2}{\sigma^2} - \frac{n(\bar{x} - \mu)^2}{\sigma^2}$$

Proof of (ii)

- The first term on the right hand side is χ^2_n and the second term
$$\frac{n(\bar{x}-\mu)^2}{\sigma^2} = \left(\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$$
is square of standard normal, which is χ^2_1.

- Thus, the left hand side is χ^2_{n-1}.

The $t-$ distribution

- Let Z and U be independent random variables such that

  (i) Z is $\mathcal{N}(0, 1)$ and

  (ii) U is χ^2_n

  The distribution of

  $$T = \frac{Z}{\sqrt{\frac{U}{n}}}$$

  is called **Student's $t-$ distribution** with n degrees of freedom.

- It is denoted by t_n

Properties of the $t-$ distrbituion

- $\diamond$ For all values of n, the distribution t_n is symmetric about zero.
- $\diamond$ For $n = 1$, the mean and variance are both infinite.
- $\diamond$ For $n = 2$, the mean is 0 but the variance is infinite.
- $\diamond$ For $n > 2$, the mean is 0 and variance is $\frac{n}{n-2}$
- $\diamond$ As $n \to \infty$, the distribution $t_n \to \mathcal{N}(0, 1)$

Example

- Let $x_1, x_2, \ldots, x_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$.
- We know that

  (i) $Z = \frac{(\bar{x}-\mu)}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$

  (ii) $U' = \frac{S_{xx}}{\sigma^2}$ is χ^2_{n-1}.

- Hence,

$$T = \frac{Z}{\sqrt{\frac{U'}{n-1}}} \sim t_{n-1}$$

Example

- By substitution

$$T = \frac{(\bar{x} - \mu)\sqrt{n}}{\sqrt{\frac{S_{xx}}{n-1}}} = \frac{(\bar{x} - \mu)\sqrt{n}}{\widehat{\sigma}} = \frac{(\bar{x} - \mu)}{\frac{\widehat{\sigma}}{\sqrt{n}}}$$

where $\widehat{\sigma} = \sqrt{\frac{S_{xx}}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$

- $\widehat{\sigma}$ is the sample standard deviation which is used to replace $s = \sqrt{\frac{S_{xx}}{n}}$ when the sample size is small.

The $F-$ distribution

- Let U and V be independent random variables whose distributions are, respectively, χ^2_m and χ^2_n.
- The distribution of the ratio

$$F = \frac{\frac{U}{m}}{\frac{V}{n}}$$

  is called the F- distribution with degrees of freedom or parameters (m, n).
- It is denoted by $F_{m,n}$.
- Note that in stating the parameters, m and n, the degree of freedom associated with the numerator comes first.

Example

- Let $x_1, x_2, \ldots, x_m$ be a random sample from a population, $\mathcal{N}(\mu_1, \sigma^2)$ and $y_1, y_2, \ldots, y_n$ be a second random variable from the population $\mathcal{N}(\mu_2, \sigma^2)$.

- Then

  (i) $U = \frac{S_{xx}}{\sigma^2}$ is χ^2_{m-1}

  (ii) $V = \frac{S_{yy}}{\sigma^2}$ is χ^2_{n-1}

- If the two samples are independent of each other, then

$$F = \frac{\frac{U}{m-1}}{\frac{V}{n-1}}$$

  has the distribution $F_{m-1,n-1}$.

Example

◇ By substitution

$$
\begin{aligned}
F &= \frac{S_{xx}/m-1}{S_{yy}/n-1} \\
&= \frac{\text{sample variance on the sample } (x_1, x_2, \ldots, x_m)}{\text{sample variance on the sample } (y_1, y_2, \ldots, y_n)}
\end{aligned}
$$

◇ Note that the two populations that give rise to the two samples have the same variance.

◇ The chi-square, t- and F- distributions are widely used in inference problem involving the normal population.

The Gamma family of distributions

- The Gamma family of distribution has the density function

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\,,\ x > 0$$

where α and β are parameters which are positive.

- The parameter, α determines the shape of the density curve and is called the **shape** parameter.

- $\Gamma(\alpha)$ is the gamma function defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$$

The Gamma family of distributions

- When α is a positive integer, then

$$\Gamma(\alpha) = (\alpha - 1)!$$

- The parameter, β is called the **scale** parameter and determines the sketch of the density.

- We denote the gamma distribution with shape parameter α and scale parameter β as $\Gamma(\alpha, \beta)$

Special cases

1. When $\alpha = 1$, $f(x)$ reduces to the exponential distribution with parameter β and density

$$f(x) = \beta e^{-\beta x},\ x > 0$$

   denote the exponential distribution with parameter β as $Exp(\beta)$

2. When the shape and scale parameters are $\frac{n}{2}$ and $\frac{1}{2}$ respectively,

$$\Gamma\left(\frac{n}{2}, \frac{1}{2}\right) \text{ is identical to } \chi^2_n$$

Properties

1. For $X \sim \Gamma(\alpha, \beta)$,

$$\begin{aligned} E(X) &= \frac{\alpha}{\beta} \\ Var(X) &= \frac{\alpha}{\beta^2} \end{aligned}$$

2. If $X \sim \Gamma(\alpha, \beta)$ and $k > 0$, then $Y = kX$ is distributed as $\Gamma(\alpha, \frac{1}{k}\beta)$

3. If $X_1 \sim \Gamma(\alpha_1, \beta)$ and $X_2 \sim \Gamma(\alpha_2, \beta)$, and X_1 and X_2 are independent, then $Y = X_1 + X_2$ is $\Gamma(\alpha_1 + \alpha_2, \beta)$.

Properties

This can be extended for the sum of n independent gamma variables. i.e $X_1 + X_2 + \ldots + X_n \sim \Gamma(\alpha_1 + \alpha_2 + \ldots + \alpha_n, \beta)$

4. If $X \sim \Gamma(k, \beta)$ where k is a positive integer

$$
\begin{aligned}
Y &= \beta X \sim \Gamma\left(k, \frac{1}{\beta}.\beta\right) = \Gamma(k, 1) \\
Z &= 2Y \sim \Gamma\left(k, \frac{1}{2}\right) = \chi^2_{2k}
\end{aligned}
$$

Hence if $X \sim \Gamma(k, \beta)$ and k is positive integer, then $Z = 2\beta X \sim \chi^2_{2k}$

Samples from exponential population

- Let $x_1, x_2, \ldots, x_n$ be a random sample from an exponential population, Exp(θ), if $x_1, x_2, \ldots, x_n$ are independent, then each is distributed as $\Gamma(1, \theta)$

- The statistics $\sum_{i=1}^{n} X_i \sim \Gamma(n, \theta)$.

- Also, $2\theta \sum_{i=1}^{n} x_i \sim \Gamma\left(n, \frac{1}{2}\right) = \chi^2_{2n}$.

- But $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$.

- This implies,

$$\begin{aligned} 2\theta n \bar{x}_n &\sim \chi^2_{2n} \\ \bar{x}_n &\sim \frac{1}{2n\theta} \chi^2_{2n} \end{aligned}$$

Samples from uniform distribution

- If $Y \sim U(0,1)$ then $Z = -\ln Y$ is distributed as Exp(1) with density $f(z) = e^{-z}, z > 0$.

- **Prove the above as an exercise.**

- If the random sample $y_1, y_2, \ldots, y_n$ are drawn from $U(0,1)$, then the transformed values $z_1, z_2, \ldots, z_n$ constitute a random sample from Exp(1)$\sim \Gamma(1,1)$.

- The sampling distribution of $2\sum_{i=1}^{n} z_i$ is

$$2\sum_{i=1}^{n} z_i \sim \Gamma\left(n, \frac{1}{2}\right) \sim \chi^2_{2n}.$$

- Thus, $-2\sum_i \ln y_i \sim \chi^2_{2n}$.

Samples from uniform distribution

- If $X \sim U(0, \theta)$ then $Y = \frac{X}{\theta} \sim U(0, 1)$
- **Prove the above as an exercise.**
- If $x_1, x_2, \ldots, x_n$ is a random sample drawn from $U(0, \theta)$ then

$$-2\sum_{i=1}^{n} \ln\left(\frac{x_i}{\theta}\right) \sim \chi^2_{2n}.$$

- **Proof**: Since $\frac{x_i}{\theta} \sim U(0, 1)$

$$\begin{aligned}
-\ln\left(\frac{x_i}{\theta}\right) &\sim \text{Exp}(1) = \Gamma(1, 1) \\
\Rightarrow 2\sum_{i=1}^{n}\left[-\ln\left(\frac{x_i}{\theta}\right)\right] &\sim \Gamma(n, 1/2) \\
\text{thus, } -2\sum_{i=1}^{n} \ln\left(\frac{x_i}{\theta}\right) &\sim \chi^2_{2n} \qquad \ldots(*)
\end{aligned}$$

Samples from uniform distribution

- We can use the above results to derive the sampling distribution of a statistics involving the geometric mean as follows:

$$\begin{aligned} \sum_{i=1}^{n} \ln\left(\frac{x_i}{\theta}\right) &= \sum_{i=1}^{n} [\ln x_i - \ln\theta] \\ &= \ln\left(\prod_{i=1}^{n} x_i\right) - n\ln\theta \end{aligned}$$

- but $M_G(x) = \left[\prod_{i=1}^{n} x_i\right]^{1/n} \Rightarrow [M_G(x)]^n = \prod_{i=1}^{n} x_i$

Samples from uniform distribution

- thus,

$$\sum_{i=1}^{n} \ln\left(\frac{x_i}{\theta}\right) = \ln M_G(x)^n - n\ln\theta$$
$$= n\ln M_G(x) - n\ln\theta$$

- From equation (*) above, it follows that

$$-2\sum_{i=1}^{n} \ln\left(\frac{x_i}{\theta}\right) = -2n\left[\ln M_G(x) - \ln\theta\right] \sim \chi^2_{2n}$$

Samples from uniform distribution

- Let $x_1, x_2, \ldots, x_n$ be a random sample from the distribution $U(0, \theta)$ and $X_{(n)}$ be the maximum value in the sample.

- The density function of X_n is

$$h(x) = \frac{n}{\theta} \left(\frac{x}{\theta} \right)^{n-1}, 0 \leq x \leq \theta.$$

Prove this density as an exercise.

Samples from uniform distribution

- Let $Y = -\ln\left[\frac{X_{(n)}}{\theta}\right]$, the density of Y is given by

$$g(y) = ne^{-ny}, 0 \leq y \leq \infty$$

Prove this as an exercise

- If $Y = -\ln\left[\frac{X_{(n)}}{\theta}\right] \sim \text{Exp}(n)$ or $\Gamma(1, n)$, it follows that

$$2nY = -2n\ln\left[\frac{X_{(n)}}{\theta}\right] \sim \Gamma\left(1, \frac{n}{2n}\right) = \Gamma(1, 1/2) = \chi^2_2.$$

Estimation

- **Estimation**: a process of determining unknown parameters or numerical characteristics of a population using sample data.

- When a single value is suggested for the unknown parameter, this is called a ***point estimate***.

- **Example**: finding the mean of a sample as an estimate of unknown population mean.

- When an interval is determined within which it can be established that the unknown parameter lies with some specified 'degree of confidence', this refers to as an ***Interval Estimation***.

Estimation

- Suppose the sample mean $\bar{x}$ is used as an estimate of a population mean, μ. Then $\bar{x} - \mu$ represents *error*.

- Usually the procedure or rule for computing the estimate is called an **Estimator**.

- The **estimate** is the value of the estimator for the available data.

Methods of estimation

- There are several methods for estimating a population quantity.

- Most commonly used methods are
  - substitution
  - method of moments
  - least square
  - weighted least squares
  - maximum likelihood

Method of substitution

- This method is based on the principle that the available sample is reasonably representative of the population from which it is drawn.

- For example
  - the mean of a sample will be close to the mean of the population
  - the median of the sample is expected to be close to the population median

- Approaches include: ***frequency substitution*** and ***method of moment/moment substitution***.

Frequency substitution

- Consider a random sample from a multinomial population of k classes, $A_1, A_2, \ldots, A_k$ with class probabilities, $p_1, p_2, \ldots, p_k$ which are unknown.

- Using the substitution principle, the natural estimate of the population proportion $p_j, j = 1, 2, \ldots, k$ is the proportion of individuals of class A_j in the sample.

- That is,

$$\hat{p}_j = \frac{x_j}{n}$$

  where x_j is the number of individuals in the sample who belong to class A_j and n is the total sample size.

Example

A survey carried out in a certain community gave the following returns of the employment status and marital status of a random sample of 550 adult males.

Marital status	self-employed	employed	unemployed	Total
married	92	180	34	306
unmarried	36	127	81	244
Total	128	307	115	550

◇ Denote the class probabilities by $p_1, p_2, \ldots, p_6$ as shown below

Marital status	self-employed	employed	unemployed
married	p_1	p_3	p_5
unmarried	p_2	p_4	p_6

Example

◇ By using frequency substitution, we can compute estimate of the class probabilities

Marital status	self-employed	employed	unemployed
married	$\hat{p}_1 = \frac{92}{550} = 0.167$	$\hat{p}_3 = \frac{180}{550} = 0.327$	$\hat{p}_5 = \frac{34}{550} = 0.062$
unmarried	$\hat{p}_2 = \frac{36}{550} = 0.066$	$\hat{p}_4 = \frac{127}{550} = 0.231$	$\hat{p}_6 = \frac{81}{550} = 0.147$

◇ We can also obtain estimate of the ratio of employees to self-employed persons in the population, $\frac{p_3+p_4}{p_1+p_2}$ as

$$\frac{\hat{p}_3 + \hat{p}_4}{\hat{p}_1 + \hat{p}_2} = \frac{0.327 + 0.231}{0.167 + 0.066} = 2.395$$

Method of moment/moment substitution

- Recall that, the kth (non-central) population moment is given by

$$\mu_k' = E(X^k).$$

- An appropriate estimator is the kth sample moment

$$m_k' = E(x_i^k) = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

- Similarly, an estimator for the kth central moment

$$\mu_k = E(X - \mu)^k$$

is the sample central moment

$$m^k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$$

Method of moments

- Suppose we want to estimate the unknown parameters $\theta_1, \theta_2, \ldots, \theta_k$, we can express the k population moments in terms of the k parameters as follows:

$$\begin{aligned}
\mu_1 &= h_1(\theta_1, \theta_2, \ldots, \theta_k) \\
\mu_2 &= h_2(\theta_1, \theta_2, \ldots, \theta_k) \\
\vdots & \ \ldots \ \vdots \\
\mu_k &= h_k(\theta_1, \theta_2, \ldots, \theta_k)
\end{aligned}$$

  where $h_1, h_2, \ldots, h_k$ are known functions.

- The population moments are then replaced by their corresponding sample moments $m_1, m_2, \ldots, m_k$

- Estimates of the k parameters are obtained by solving the resulting k equations.

Example

Example: Let $x_1, x_2, \ldots, x_k$ denote a random sample from the exponential distribution Exp(θ) with density

$$f(x) = \theta e^{-\theta x}, x > 0.$$

Find an estimate of the parameter θ using the method of moment approach.

Solution:

- Since we have one parameter, we compute the first non-central moment:

$$\mu_1' = E(X) = \frac{1}{\theta}$$

- Replace μ_1' by the corresponding sample moment, $\bar{x}$. Then

$$\bar{x} = \frac{1}{\hat{\theta}} \Rightarrow \hat{\theta} = \frac{1}{\bar{x}}$$

Example

Example: Consider the estimation of the parameters α and β of the gamma distribution, $\Gamma(\alpha, \beta)$. Estimate α and β using the method of moment.

Solution:

- The first two moments are $\mu = \alpha/\beta$ and $\sigma^2 = \alpha/\beta^2$.
- The moment substitution estimates for μ and σ^2 are the sample mean, $\bar{x}$ and sample variance, s^2.
- Thus,

$$\bar{x} = \frac{\hat{\alpha}}{\hat{\beta}} \tag{1}$$

$$s^2 = \frac{\hat{\alpha}}{\hat{\beta}^2} \tag{2}$$

Example

- Solving for $\hat{\alpha}$ and $\hat{\beta}$, we get

$$s^2 = \frac{\bar{x}}{\hat{\beta}} \Rightarrow \hat{\beta} = \frac{\bar{x}}{s^2}$$

- Substituting $\hat{\beta}$ into (1), we get

$$\bar{x} = \hat{\alpha} \times \frac{s^2}{\bar{x}} \Rightarrow \hat{\alpha} = \frac{\bar{x}^2}{s^2}$$

Example

Example: A certain type of electrical component is known to have a lifetime of X hours, where X is a random variable with a uniform distribution on the interval $[a, b]$. The density function of X is

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

Find an estimate for a and b using the Method of Moment.

Solution:

◇ The first two moments of X are

$$\begin{aligned} E(X) &= \frac{a+b}{2} \\ \text{var}(X) &= \frac{(b-a)^2}{12} \end{aligned}$$

Example

- Replacing $E(X)$ and $\mathrm{var}(X)$ by the corresponding sample moments, $\bar{x}$ and s^2, we have

$$\begin{aligned}\bar{x} &= \frac{\hat{a}+\hat{b}}{2} \\ s^2 &= \frac{(\hat{a}+\hat{b})^2}{12}\end{aligned}$$

- Solving for $\hat{a}$ and $\hat{b}$ yields

$$\begin{aligned}\hat{a} &= \bar{x} - s\sqrt{3} \\ \hat{b} &= \bar{x} + s\sqrt{3}\end{aligned}$$

Exercise

From the example above and using only the first 2 non-central moments, show that the moment estimates of *a* and *b* are respectively

$$\hat{a} = m'_1 - \sqrt{3(m'_2 - m'_1)}$$

and

$$\hat{b} = m'_1 + \sqrt{3(m'_2 - m'_1)}$$

where $m'_1 = \bar{x}$ and $m'_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ are the first and second non-central sample moment.

Least squares

- Commonly used to determine the best fit to data.
- Given a pair of data points $(x_i, y_i), i = 1, 2, \ldots, n$ where y_i represents the outcome and x_i the explanatory variable.
- Suppose that the line $E(y_i) = a + bx_i$ best describe the relationship between x and y.
- Given the observed data, we wish to determine the values of a and b that is a good fit to the data.
- The least-square principle regards the estimates as good if the estimates of the expected values are close to the actual observations.

Computation of least squares estimates

- A measure of closeness is the sum of squares (error associated with the line)

$$W = \sum_{i=1}^{n}(y_i - E(y_i))^2 = \sum_{i=1}^{n}(y_i - (a + bx_i))^2$$

- We desire to find the values of *a* and *b* that minimizes the error sum of squares, W.

- We learned from multivariate calculus that, to minimize a function, we find the values of the parameters by differentiating with respect to the parameters and setting the resulting equations to zero.

- These equations are called the **normal equations.**

Computation of least squares estimates

◇ Consider the linear fit example, we differentiate W with respect to a and b yields

$$\begin{aligned}
\frac{\partial W}{\partial a} &= \sum_{i=1}^{n} 2(y_i - (ax_i + b))(-x_i) \\
\frac{\partial W}{\partial b} &= \sum_{i=1}^{n} 2(yi - (ax_i + b))(1)
\end{aligned}$$

◇ Setting the two equations to 0 yields

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - (ax_i + b))x_i &= 0 \\
\sum_{i=1}^{n}(yi - (ax_i + b)) &= 0
\end{aligned}$$

Computation of least squares estimates

◇ These can be rewritten as

$$\begin{aligned} \left(\sum_i x_i^2\right) a + \left(\sum_i x_i\right) b &= \sum_i x_i y_i \\ \left(\sum_i x_i\right) a + \left(\sum_i 1\right) b &= \sum_i y_i \end{aligned}$$

◇ The values $\sum x_i, \sum y_i, \sum x_i^2$ and $\sum x_i y_i$ can be computed from sample data and these equations solved for the estimates $\hat{a}$ and $\hat{b}$.

Exercise

Show that by solving the normal equations above,

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

and

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

General case

- The above computational procedure applies to cases with $k > 0$ number of parameters.

- Suppose that $E(y_i) = h(\theta_1, \theta_2, \ldots, \theta_k)$.

- Then we define

$$W = \sum_{i=1}^{n} (y_i - h(\theta_1, \theta_2, \ldots, \theta_k))$$

- The values of $\theta_1, \theta_2, \ldots, \theta_k$ that minimizes W are obtained as the solution to the k set of simultaneous equations (normal equations):

$$\frac{\partial W}{\partial \theta_1} = 0, \frac{\partial W}{\partial \theta_2} = 0, \ldots, \frac{\partial W}{\partial \theta_k} = 0.$$

Weighted Least Squares

- The least squares method describe earlier assumes equality of variance associated with any single observation.

- In some situations in practice, this restrictive assumption may not be valid.

- The ***weighted least squares*** weights the observation proportional to the reciprocal of the error variance for that observation.

- Thus, the weighted error sum of squares

$$W = \sum_{i=1}^{n} \frac{(y_i - E(y_i))^2}{\sigma_i^2}$$

is rather minimized.

Weighted Least Squares

- ◇ The intuition behind the weight is that, if an observation has large variance, then that observation is comparatively unreliable and so should be given comparatively little weight.

- ◇ **Example**: Suppose that $E(y_i) = a + bx_i$ and $\text{var}(y_i|x_i) = \sigma_i^2$. Obtain the weighted least square estimates for a and b.

- ◇ **Solution**: Let $w_i = \frac{1}{\sigma_i^2}$ then

$$W = \sum_{i=1}^{n} w_i(y_i - (a + bx_i))^2.$$

Example

- By obtaining the normal equations and solving for a and b we get

$$\begin{aligned}\hat{a} &= \bar{y}_w - b\bar{x}_w \\ \hat{b} &= \frac{\sum w_i(x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum w_i(x_i - \bar{x}_w)^2}\end{aligned}$$

where $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$ and $\bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}$ are the weighted means.

- **Show the detailed work as an exercise.**

Maximum likelihood method

- Widely used for parametric models where the underlying distribution is known and the number of parameters is finite.

- The goal is to obtain the value of the unknown parameter of the distribution that is likely to produce the sample or that gives the highest probability to produce the sample.

- Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ denote the sample from a distribution, with density $f(x; \theta)$, which is known except for the value of the parameter θ.

- The probability of observing the sample is equal to

$$P(X_1 = x_i \cap X_2 = x_2 \cap \cdots \cap X_n = x_n).$$

Maximum likelihood method

- This joint probability function, which depends on the unknown parameter, is called the **likelihood** of the sample and is denoted

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} P(X_i = x_i) \text{ [discrete data]}$$

or

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \theta) \text{ [continuous data]}$$

where $f(x_i; \theta)$ is the pdf of X.

- The maximum likelihood estimate of θ is the value $\theta = \theta^*$ for which $L(\theta; \boldsymbol{x})$ has the largest possible value.

Maximum likelihood method

- If the parameter space is Θ the θ^* (or $\hat{\theta}$) must lie in Θ and must satisfy
$$L(\theta^*; \boldsymbol{x}) \geq L(\theta; \boldsymbol{x})$$
for all $\theta \in \Theta$.

- To compute the maximum likehood estimator (MLE), we usually maximize the $\ln L(\theta; \boldsymbol{x})$ called the **log-likelihood.**

Maximum likelihood method

- The value of θ that maximizes $\ln L$ is also the value that maximizes L.

- Thus, the value $\theta = \theta^*$ that maximizes the log-likelihood function

$$\ell(\theta; \boldsymbol{x}) = \ln L(\theta; \boldsymbol{x})$$

is the solution of the equations

$$\frac{dL}{d\theta} = 0 \text{ or } \frac{d\ell}{d\theta} = 0.$$

Example

- **Example**: The arrival of customers at a service counter is assumed to constitute a Poisson process. If the rate of this process is λ per hour, then the number of customers arriving in any period of duration T hours has a Poisson distribution with mean $T\lambda$. Let X^* customers be recorded over a 15 hour period. Find the maximum likelihood estimate of λ.

- **Solution**: Let X be the number of customers arriving in any period. The probability mass function of X is given by

$$P(X_i = x_i; \lambda) = \frac{e^{-T\lambda}(T\lambda)^{x_i}}{x_i!}$$

- The likelihood function for X^* arrivals in a 15 hour period is

$$L(\lambda) = \frac{e^{-15\lambda}(15\lambda)^{X^*}}{X^*!}.$$

Example

- The log-likelihood is

$$\ell(\lambda) = \ln L = -15\lambda + X^* \ln(15\lambda) - \ln(X^*!)$$

- The MLE of λ satisfies $\frac{d\ell}{d\lambda} = 0$, i.e.

$$\frac{d\ell}{d\lambda} = -15 + \frac{X^*}{\lambda} = 0$$

- Thus, $\lambda^* = \frac{X^*}{15}$.

- Note that since $\lambda^* = \frac{X^*}{15}$ is the only turning point, and the second derivative $\frac{d^\ell}{d\lambda^2} = -\frac{X^*}{\lambda^2}$ is negative, this point is the absolute maximum value.

Example

- **Example**: Suppose $x_1, x_2, \ldots, x_n$ are a random sample from a geometric distribution with parameter $p, 0 \leq p \leq 1$. Find the MLE, $\hat{p}$.

- **Solution**: For a geometric distribution, the pmf is given by

$$f(x; p) = p(1-p)^{x-1}, 0 \leq p \leq 1, x = 1, 2, \ldots,$$

- The likelihood is

$$L(p) = \prod_{i=1}^{n} p(1-p)^{x-1} = p^n (1-p)^{-n+\sum_{i=1}^{n} x_i}$$

- Taking the natural log of $L(p)$,

$$\ell(p) = \ln L(p) = n \ln p + (-n + \sum x_i) \ln(1-p)$$

Example

- Taking the derivative with respect to p, we get

$$\frac{d\ell}{dp} = \frac{n}{p} - \frac{-n + \sum x_i}{(1-p)}$$

- Setting $\frac{d\ell}{dp} = 0$ and solving for p we get

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}.$$

- Thus, the maximum likelihood estimator

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}$$

MLE

- In some instances, the method of derivatives cannot be used to find the MLE.
- For example, when the likelihood is not differentiable in the range space.
- In such circumstance, special structures available in the specific situation can be used to solve the problem.
- **Example**: Let $x_1, x_2, \ldots, x_n$ be a random sample from $U(0, \theta), \theta > 0$. Find the MLE of θ.

Example

◇ **Solution**: The pdf of X is given by

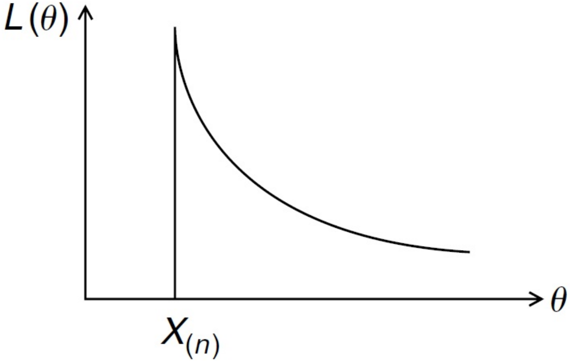
$$f(x) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

◇ The likelihood is

$$L(\theta) = \begin{cases} \frac{1}{\theta^n} & 0 \leq x_1, x_2, \ldots, x_n \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

◇ We can graph this likelihood function.

Example



- The likelihood is $\frac{1}{\theta^n}$ when $\theta \geq \max(x_i)$ and 0 for $\theta < \max(x_i)$.
- The maximum therefore occur at this point of discontinuity, $\hat{\theta} = \max(X_i) = X_{(n)}$.

MLE for more than one parameter

- The ML method is applicable for a vector of parameters, say $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$.

- The MLEs are obtained from the solutions of the systems of equations

$$\frac{\partial}{\partial \theta_i} \ln L(\theta_1, \theta_2, \ldots, \theta_k) = 0$$

  for $i = 1, 2, \ldots, k$.

- **Example**: Let $x_1, x_2, \ldots, x_n$ be a random sample from $N(\mu, \sigma^2)$ where μ and σ^2 are not known. Find the MLEs of μ and σ.

Example

◇ **Solution**: The pdf of X is given by

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

◇ The likelihood for the sample is

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

◇ The log-likelihood is

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

Example

- and

$$
\begin{aligned}
\frac{\partial \ell}{\partial \mu} &= \frac{1}{\sigma^2} \sum_i (x_i - \mu) \\
\frac{\partial \ell}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_i (x_i - \mu)^2
\end{aligned}
$$

- Setting these derivatives to zero

$$
\begin{aligned}
\sum_i (x_i - \mu) &= 0 \\
\sum_i (x_i - \mu)^2 &= n\sigma^2
\end{aligned}
$$

- Solving for μ and σ yields

$$
\mu^* = \bar{x}, \quad \text{and} \quad \sigma^* = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})}.
$$

Exercises

(1) Let $x_1, x_2, \ldots, x_n$ be a random sample from a population with gamma distribution and parameters α andβ. Find the MLEs for the unknown parameters, α and β.

[**Hint**: $\beta^* = \frac{\bar{x}}{\alpha}$ and α will be obtained from solving $-n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\ln\left(\frac{\bar{x}}{\alpha}\right) + \sum_{i=1}^{n} \ln x_i$.]

(2) The pdf of a random variable X is given by

$$f(x) = \begin{cases} \frac{2x}{\alpha^2} e^{-\frac{x^2}{\alpha^2}} & x > 0 \\ 0 & \text{Otherwise} \end{cases}$$

Using a random sample of size n, obtain the MLE α^* for α.

(3) Let $x_1, x_2, \ldots, x_n$ be a random sample from the truncated exponential distribution with pdf

$$f(x) = \begin{cases} e^{-(x-\theta)} & x \geq 0 \\ 0 & \text{Otherwise} \end{cases}$$

Show that the MLE of θ is $\min(X_i)$.

Desirable properties of point estimators

- We have looked at several methods for determining estimators of unknown population parameters.
- It is possible to find several estimators for the same parameters.
- For example, in the estimation of θ for the uniform distribution on $[0, \theta]$ given a sample of size n, we will have that
  - the moment estimator is $\hat{\theta} = 2\bar{X}$ (twice the sample mean)
  - the MLE is $\hat{\theta}^* = X_{(n)}$
- The question then arises; of which of these estimators should we use?
- We shall look at some desirable properties of estimators to answer this question.

Unbiased estimators

- It will be desirable if the expected value of the estimator is equal to the true value of the parameter.
- Such an estimator will be called an **unbiased estimator.**
- A point estimator $\hat{\theta}$ is unbiased of the parameter θ if

$$E(\hat{\theta}) = \theta$$

  for all possible values of θ.
- Otherwise, the estimator is said to be **bias** and can be computed by

$$B = E(\hat{\theta}) - \theta$$

Example

Example: Let $x_1, x_2, \ldots, x_n$ be a random sample from a Bernoulli population with parameter, p.

(1) Find the moment estimator for p.
(2) Show that the moment estimator is an unbiased estimator.

Solution: (1) The pmf is

$$f(x) = p^x(1-p)^{1-x}, x = 0, 1$$

◇ The first population moment is $\mu' = p$.

◇ Replacing with the sample moment results in the estimator, $\hat{p} = \bar{X} = \frac{\sum_i X_i}{n}$.

Example

(2) To show that $\hat{p}$ is unbiased

$$E(\hat{p}) = \frac{\sum_i E(X_i)}{n} = \frac{\sum_i p}{n} = \frac{np}{n} = p$$

◇ Thus, $\hat{p}$ is an unbiased estimator for p.

Example: Let $x_1, x_2, \ldots, x_n$ be a random sample from a population with finite mean, μ. Show that the sample mean $\bar{X}$ and $\frac{1}{3}\bar{X} + \frac{2}{3}X_1$ are both unbiased estimators of μ.

Solution:

$$\begin{aligned}
E(\bar{X}) &= E\left(\frac{\sum_i X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}.n\mu \\
E\left(\frac{1}{3}\bar{X} + \frac{2}{3}X_1\right) &= \frac{1}{3}E(\bar{X}) + \frac{2}{3}E(X_1) \\
&= \frac{1}{3}\mu + \frac{2}{3}\mu = \mu
\end{aligned}$$

Exercise

Let $\bar{x}$ denote the mean of the random sample $x_1, x_2, \ldots, x_n$ from a population with mean, μ and standard deviation, σ.

(1) Show that $\bar{x}$ (the sample mean) is an unbiased estimator for μ.

(2) Show that the sample variance $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})$ is a biased estimator for σ^2. Compute the bias term.

(3) Show that the sample variance $\hat{\sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$ is an unbiased estimator for σ^2.

Mean square error (MSE)

- As we saw in the last example, there can be more than one unbiased estimator of a parameter.
- In such situation, which of the unbiased estimators should one choose?
- Another desirable property is to choose the estimator with low variance.
- If the estimators are biased, then we desire the estimator with low bias and low variance.
- This leads to the **mean square error** of an estimator.

MSE

- The MSE of an estimator, $\hat{\theta}$ is defined as follows

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

- We can show that the MSE is a combination of bias and variance of the estimator as follows:

$$\begin{aligned} MSE(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 = E\left(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta} - \theta\right)^2 \\ &= E(\hat{\theta} - E(\hat{\theta}))^2 + (\hat{\theta} - \theta)^2 + 2E[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)] \\ &= \text{var}(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 + 0 \end{aligned}$$

- Thus,

$$MSE(\hat{\theta}) = \text{var}(\hat{\theta}) + B^2$$

where B is the bias.

MSE

- The MSE is used as the standard criterion for comparing the performance of estimators.
- Generally, Mean square error (MSE) measures, on average, how close an estimator comes to the true value of the parameter.
- An estimator with the smallest MSE is usually preferred.
- When estimators are unbiased, then their variances can be used for comparing performance.
- For an unbiased estimator $\hat{\theta}$, the square-root of the mean squared error, $\sqrt{MSE(\hat{\theta})}$, is often referred to as the **standard error**.

Example

Example: Let $x_1, x_2, \ldots, x_n$ be a random sample from the uniform distribution, $U(0, \theta)$. Find the mean squared error (MSE) of the moment estimator of θ.

Solution: We have shown previously that the moment estimator for θ is

$$\hat{\theta} = 2\bar{X}$$

◇ To find the MSE, we need to find the bias and variance of the estimator.

◇ We first determine if $\hat{\theta}$ is biased or not.

$$\begin{aligned} E(\hat{\theta}) &= E(2\bar{X}) = 2E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = 2\left(\frac{\sum_{i=1}^{n} E(X_i)}{n}\right) \\ &= 2\frac{\sum_i \theta/2}{n} = \frac{n\theta}{n} = \theta \end{aligned}$$

Example

- Thus, $\hat{\theta}$ is unbiased ($B = 0$).
- Next we find the variance of $\hat{\theta}$

$$
\begin{aligned}
\text{var}(\hat{\theta}) &= \text{var}\left(\frac{2\sum_i X_i}{n}\right) = 4\frac{\sum_i \text{var}(X_i)}{n^2} \\
&= 4\frac{n\theta^2/12}{n^2} = \frac{\theta^2}{3n}
\end{aligned}
$$

[Note that $\text{var}(X_i) = \frac{\theta^2}{12}$].

- Therefore,

$$MSE(\hat{\theta}) = \frac{\theta^2}{3n} + 0^2 = \frac{\theta^2}{3n}$$

Example

Example: Let $x_1, x_2, \ldots, x_n$ be a sample of size $n = 3$ from a distribution with unknown mean $\mu, -\infty < \mu < \infty$, where the the variance σ^2 is a known positive number. Show that both $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = \frac{2X_1+X_2+5X_3}{8}$ are unbiased estimators for μ. Compare the variances $\hat{\theta}_1$ and $\hat{\theta}_2$.

Solution:

$$
\begin{aligned}
E(\hat{\theta}_1) &= E(\bar{X}) = 1/3 \times 3\mu = \mu \\
E(\hat{\theta}_2) &= \frac{1}{8}[2E(X_1) + E(X_2) + 5E(X_3)] \\
&= \frac{1}{8}[2\mu + \mu + 5\mu] = \mu
\end{aligned}
$$

Example

$$
\begin{aligned}
\text{var}(\hat{\theta}_1) &= \text{var}\left(\frac{\sum_{i=1}^{3} X_i}{3}\right) = \frac{1}{9}\sum_i \text{var}(X_i) \\
&= \frac{1}{9} \times 3\sigma^2 = \frac{\sigma^2}{3} \\
\text{var}(\hat{\theta}_2) &= \text{var}\left(\frac{2X_1 + X_2 + 5X_3}{8}\right) \\
&= \frac{4}{64}\sigma^2 + \frac{1}{64}\sigma^2 + \frac{25}{64}\sigma^2 = \frac{30}{64}\sigma^2
\end{aligned}
$$

- Because $\text{var}(\hat{\theta}_1) < \text{var}(\hat{\theta}_2)$, $\bar{X}$ is a better estimator (since it has the smaller variance).

Example

Let $x_1, x_2, \ldots, x_n$ be a random sample from the distribution $U(0, \theta)$ and $X_{(n)}$ be the maximum value in the sample. Compute the MSE of the maximum likelihood estimator for θ, $\hat{\theta} = X_{(n)}$.

Solution: We first need to compute the bias term $B = E(\hat{\theta}) - \theta = E(X_{(n)}) - \theta$

- ◇ To find $E(X_{(n)})$ we must know the density of function of $X_{(n)}$. We have established earlier that the density function of $X_{(n)}$ is

$$h(x) = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} = \frac{n}{\theta^n}x^{n-1}, 0 \leq x \leq \theta$$

- ◇ From this, we have

$$E(X_n) = \frac{n\theta}{n+1} \text{ and } \mathrm{var}(X_{(n)}) = \frac{n\theta^2}{(n+2)(n+1)^2}$$

[Prove these as an exercise]

- Thus,

$$B = E(X_{(n)}) - \theta = \frac{n\theta}{n+1} - \theta = -\frac{\theta}{n+1}$$

- Next, we find the var$(\hat{\theta})$

$$\text{var}(\hat{\theta}) = \text{var}(X_{(n)}) = \frac{n\theta^2}{(n+2)(n+1)^2}$$

- Therefore, the MSE of $\hat{\theta}$ is

$$\begin{aligned} MSE(\hat{\theta}) &= \text{var}(\hat{\theta}) + B^2 \\ &= \frac{n\theta^2}{(n+2)(n+1)^2} + \frac{\theta^2}{(n+1)^2} = \frac{\theta^2(n+n+2)}{(n+1)^2(n+2)} \\ &= \frac{2(n+1)\theta^2}{(n+1)^2(n+2)} = \frac{2\theta^2}{(n+1)(n+2)} \end{aligned}$$

Exercise

Consider the estimation of the variance of a normal population $N(\mu, \sigma^2)$, given the random sample $x_1, x_2, \ldots, x_n$. Find the MSE of the two commonly used estimators for the population variance, σ^2. That is

$$\begin{aligned} s^2 &= \frac{1}{n} S_{xx} \\ \hat{\sigma}^2 &= \frac{1}{n-1} S_{xx} \end{aligned}$$

where $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$. Which has the smallest MSE?

Other desirable properties of point estimators

- Another desired property of an estimator is for its value to be closer to the true parameter being estimated when the sample size is large.
- This leads to the notion of **consistent estimators**, which describes the behaviour of the estimator when sample size becomes infinitely large.
- The estimator $\hat{\theta}_n$ is said to be consistent for θ if, for any $\epsilon > 0$

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1$$

or

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0.$$

- That is, the estimator has a higher probability of being close to the population value when the sample size gets large.

Consistent estimator

Theorem: Let $\hat{\theta}_n$ be an estimator of θ, and let var($\hat{\theta}_n$) be finite. If $\lim_{n\to\infty} E(\hat{\theta}_n - \theta)^2 = 0$, then $\hat{\theta}_n$ is a consistent estimator of $\hat{\theta}_n$.

Proof: Using the Chebyshev's inequality

$$P(|\hat{\theta}_n - \theta| \geq \epsilon) \leq \frac{E(\hat{\theta}_n - \theta)^2}{\epsilon^2}$$

- We have

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geq \epsilon) \leq \lim_{n\to\infty} \frac{E(\hat{\theta}_n - \theta)^2}{\epsilon^2} = 0.$$

Showing $\hat{\theta}_n$ is a consistent estimator for θ.

Consistent estimator

- Note that $E(\hat{\theta}_n - \theta)^2 = \text{var}(\hat{\theta}_n) + B(\hat{\theta}_n)^2$.
- The estimator need not be unbiased.
- An unbiased estimator $\hat{\theta}_n$ of θ is consistent if

$$\lim_{n \to \infty} \text{var}(\hat{\theta}_n) = 0.$$

[Sufficient condition].

Procedure to test for consistency

1. Check whether the estimator $\hat{\theta}_n$ is unbiased or not.
2. Calculate $\text{var}(\hat{\theta}_n)$ and $B(\hat{\theta}_n)$, the bias of $\hat{\theta}_n$
3. An unbiased estimator is consistent if $\text{var}(\hat{\theta}_n) \to 0$ as $n \to \infty$.
4. A biased estimator is consistent if both $\text{var}(\hat{\theta}_n) \to 0$ and $B(\hat{\theta}_n) \to 0$.

Example

Example: Let $x_1, x_, \ldots, x_n$ be a random sample with mean, μ and variance which is finite. Then, the sample mean $\bar{X}$ is a consistent estimator of the population mean, μ.

Solution: This can be shown in two ways:

1) Using the Chebyshev's inequality,

$$P\left(|\bar{X} - \mu| \leq \epsilon\right) \geq 1 - \frac{\sigma^2_{\bar{X}}}{\epsilon^2}$$

Taking the limit as $n \to \infty$

$$\lim_{n\to\infty} P\left(|\bar{X} - \mu| \leq \epsilon\right) \geq \lim_{n\to\infty} \left(1 - \frac{\sigma^2_{\bar{X}}}{\epsilon^2}\right) = 1.$$

Hence, $\bar{X}$ is a consistent estimator of μ.

Example

2) Since $\bar{X}$ is an unbiased estimator of μ,

$$\lim_{n\to\infty} \text{var}(\bar{X}) = \lim_{n\to\infty} \left(\frac{\sigma^2}{n}\right) = 0.$$

Thus, $\bar{X}$ is a consistent estimator of μ.

Example

Example: Let $x_1, x_2, \ldots, x_n$ be a random sample from $N(\mu, \sigma^2)$ population. Show that the maximum likelihood estimators for μ and σ^2 are consistent.

Solution:We have shown earlier that the MLE for μ is $\hat{\mu} = \bar{X}$ and σ^2 is $s^2 = \frac{1}{n}\sum_{i=1}(X_i - \bar{X})^2$.

- ◇ We have established that $\hat{\mu}$ is an unbiased estimator for μ. Thus,

$$\lim_{n\to\infty} \text{var}(\bar{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

Therefore, $\hat{\mu} = \bar{X}$ is a consistent estimator for μ.

- ◇ We have also shown earlier that $E(s^2) = \frac{n-1}{n}s^2$ and a bias, $B(s^2) = -\frac{1}{n}\sigma^2$.

Example

- Thus,

$$\lim_{n\to\infty} B(s^2) = \lim_{n\to\infty} -\frac{1}{n}\sigma^2 = 0$$

- Next we look for $\mathrm{var}(s^2)$,

$$\mathrm{var}(s^2) = \mathrm{var}\left(\frac{1}{n}S_{xx}\right) = \frac{1}{n^2}\mathrm{var}(S_{xx}) = \frac{1}{n^2}2(n-1)\sigma^4$$

[Note that $\mathrm{var}\left(\frac{S_{xx}}{\sigma^2}\right) = 2(n-1)$]

- Thus,

$$\lim_{n\to\infty} \mathrm{var}(s^2) = \lim_{n\to\infty} \frac{2(n-1)\sigma^4}{n^2} = 0.$$

Hence, s^2 is a consistent estimator of σ^2.

Efficiency

- When we have more than one unbiased estimator for a parameter θ, we have indicated that the one with the least variance is desirable.

- Another concept that compares variances of estimators is **efficiency.**

- An estimator $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if

$$MSE(\hat{\theta}_1) \leq MSE(\hat{\theta}_2).$$

- The **relative efficiency** of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is defined as

$$RE(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}.$$

Relative efficiency

- If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators for $\hat{\theta}$, the relative efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is the ratio

$$RE(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)} = \frac{\text{var}(\hat{\theta}_2)}{\text{var}(\hat{\theta}_1)}$$

- If $\text{var}(\hat{\theta}_2) > \text{var}(\hat{\theta}_1)$ or $RE(\hat{\theta}_1, \hat{\theta}_2) > 1$, then $\hat{\theta}_1$ is relatively more efficient that $\hat{\theta}_2$.

- That is to say, $\hat{\theta}_1$ has a smaller variance as compared to $\hat{\theta}_2$.

- In general, the relative efficiency depends on the parameter value θ as well as the sample size.

- In many important cases, it depends only on the sample size and can therefore be easily computed.

Example

Example: Let $x_1, x_2, \ldots, x_n, n > 3$, be a random sample from a population with true mean, μ and variance, σ^2. Consider the following estimators of μ

$$
\begin{aligned}
\hat{\theta}_1 &= \frac{1}{3}(X_1 + X_2 + X_3) \\
\hat{\theta}_2 &= \frac{1}{8}X_1 + \frac{3}{4(n-2)}(X_2 + \cdots + X_{n-1}) + \frac{1}{8}X_n \\
\hat{\theta}_3 &= \bar{X}
\end{aligned}
$$

(i) Show that each of the three estimators is unbiased.

(ii) Find $RE(\hat{\theta}_1, \hat{\theta}_2)$, $RE(\hat{\theta}_3, \hat{\theta}_1)$ and $RE(\hat{\theta}_3, \hat{\theta}_2)$. Which of these three estimators is more efficient?

Example

Solution: (a)

$$
\begin{aligned}
E(\hat{\theta}_1) &= \frac{1}{3}E(X_1 + X_2 + X_3) = \frac{3}{3}\mu = \mu \\
E(\hat{\theta}_2) &= \frac{1}{8}E(X_1) + \frac{3}{4(n-2)}E(X_2 + \cdots + X_{n-1}) + \frac{1}{8}E(X_n) \\
&= \frac{1}{8}\mu + \frac{3}{4(n-2)}(n-2)\mu + \frac{1}{8}\mu = \mu \\
E(\hat{\theta}_3) &= E(\bar{X}) = \frac{n\mu}{n} = \mu
\end{aligned}
$$

Hence $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ are unbiased estimators of μ.

Example

(b) Next we find the variances

$$
\begin{aligned}
\text{var}(\hat{\theta}_1) &= \frac{1}{3^2}\text{var}(X_1 + X_2 + X_3) = \frac{3}{9}\sigma^2 = \frac{\sigma^2}{3} \\
\text{var}(\hat{\theta}_2) &= \frac{1}{8^2}\text{var}(X_1) + \frac{3^2}{4^2(n-2)^2}\text{var}(X_2 + \cdots + X_{n-1}) \\
&\quad + \frac{1}{8^2}\text{var}(X_n) \\
&= \frac{1}{64}\sigma^2 + \frac{9}{16(n-2)^2}(n-2)\sigma^2 + \frac{1}{64}\sigma^2 = \frac{n+16}{32(n-2)}\sigma^2 \\
\text{var}(\hat{\theta}_3) &= \text{var}(\bar{X}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}
$$

Example

- Thus the relative efficiencies are

$$\begin{aligned} RE(\hat{\theta}_1, \hat{\theta}_2) &= \frac{\text{var}(\hat{\theta}_2)}{\text{var}(\hat{\theta}_1)} = \frac{\sigma^2(n+16)/2(n-2_2)}{\sigma^2/3} \\ &= \frac{3(n+16)}{32(n-2)} < 1 \text{ for } n \geq 4. \end{aligned}$$

- Similarly,

$$\begin{aligned} RE(\hat{\theta}_3, \hat{\theta}_1) &= \frac{\text{var}(\hat{\theta}_1)}{\text{var}(\hat{\theta}_3)} = \frac{\sigma^2/3}{\sigma^2/n} = \frac{n}{3} > 1 \text{ for } n \geq 4. \\ RE(\hat{\theta}_3, \hat{\theta}_2) &= \frac{\text{var}(\hat{\theta}_2)}{\text{var}(\hat{\theta}_3)} = \frac{\sigma^2(n+16)/32(n-2)}{\sigma^2/n} \\ &= \frac{n^2+16n}{32(n-2)} > 1 \text{ for } n \geq 4. \end{aligned}$$

- Thus, $\hat{\theta}_3$ is more efficient than $\hat{\theta}_2$ and $\hat{\theta}_1$.

Example

Example: In the estimation of the parameter θ of the uniform distribution on $[0, \theta]$, find the efficiency of the maximum likelihood estimator, θ^* relative to the moment estimator, $\hat{\theta}$. Which is more efficient.

Solution:

- ◇ We have earlier shown that the moment estimator, $\hat{\theta} = 2\bar{X}$ with $MSE = \sigma^2/3n$.
- ◇ Also, the MLE for θ, $\theta^* = X_{(n)} = \max(X_i)$ with MSE given by $\frac{2\theta^2}{(n+1)(n+2)}$.

Example

- Therefore,

$$
\begin{aligned}
R(\theta^*, \hat{\theta}) &= \frac{MSE(\hat{\theta})}{MSE(\theta^*)} = \frac{\theta^2}{3n} \times \frac{(n_1)(n+2)}{2\theta^2} \\
&= \frac{(n+1)(n+2)}{6n} > 1 \text{ for } n \geq 3.
\end{aligned}
$$

- Thus, the MLE for θ is more efficient than the moment estimator.

- The relative efficiency increases as the sample size grows.

Exercise

Let $x_1, x_2, \ldots, x_n, n \geq 2$ be a random sample from a normal population with mean, μ and variance, σ^2. Consider the following two estimators of σ^2

(1) $\hat{\theta}_1 = \frac{1}{n-1} S_{xx}$

(2) $\hat{\theta}_2 = \frac{1}{n} S_{xx}$.

Find $RE(\hat{\theta}_1, \hat{\theta}_2)$. Which is more efficient.

Uniformly minimum variance unbiased estimator

- There is the possibility of having one unbiased estimator that is more efficient than any other unbiased estimator.

- An unbiased estimator $\hat{\theta}$ is said to be **uniformly minimum unbiased estimator (UMVUE)** for the parameter θ_0 if, for any other unbiased estimator $\hat{\theta}$

$$\text{var}(\hat{\theta}_0) \leq \text{var}(\hat{\theta})$$

  for all possible values of θ.

- It is often not easy to find an UMVUE for a parameter.

- The **Cramer-Rao inequality (information inequality)** gives a lower bound for the variance of any unbiased estimator.

Cramer-Rao inequality

- Let $x_1, x_2, \ldots, x_n$ be a random sample from a population with pdf, $f_\theta(x)$ that depends on the parameter, θ.
- **Cramer-Rao inequality**: If $\hat{\theta}$ is an unbiased estimator of θ, then under very general condition, the following inequality is true

$$\text{var}(\hat{\theta}) \geq \frac{1}{\text{var}\left(\Phi_\theta(\boldsymbol{x})\right)}$$

where $\Phi_\theta(\boldsymbol{x}) = \frac{\partial}{\partial\theta} l(\theta; \boldsymbol{x}) = \frac{\partial}{\partial\theta} \ln(L(\theta; \boldsymbol{x}))$

- $\text{var}\left(\Phi_\theta(\boldsymbol{x})\right) = I_n(\theta)$ is called **information content.**

Cramer-Rao inequality

- Note that the function $\Phi_\theta(\boldsymbol{x})$ is a random variable.
- Its first moment, $E(\Phi_\theta(\boldsymbol{x})) = 0$
- **Proof**: Let $\Phi_\theta(\boldsymbol{x}) = \frac{\partial}{\partial\theta} l(\theta; \boldsymbol{x}) = \frac{\partial}{\partial\theta} \ln f_\theta(\boldsymbol{x}))$

$$
\begin{aligned}
E(\Phi_\theta(\boldsymbol{x})) &= \int \Phi_\theta(\boldsymbol{x}) f_\theta(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int \frac{\partial}{\partial\theta} \ln[f_\theta(\boldsymbol{x})] f_\theta(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int \frac{f'_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})} f_\theta(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int f'_\theta(\boldsymbol{x}) d\boldsymbol{x} = \int \frac{\partial}{\partial\theta} f_\theta(\boldsymbol{x}) d\boldsymbol{x} \\
&= \frac{\partial}{\partial\theta} \int f_\theta(\boldsymbol{x}) d\boldsymbol{x} = \frac{\partial}{\partial\theta}(1) = 0
\end{aligned}
$$

Cramer-Rao inequality

- Also note that $\frac{\partial^2}{\partial\theta^2}\int f_\theta(\boldsymbol{x})d\boldsymbol{x} = 0$
- The variance, $\text{var}(\Phi_\theta(\boldsymbol{x})) = E\left(\frac{\partial l}{\partial\theta}\right)^2 = -E(\frac{\partial^2 l}{\partial\theta^2})$
- **Proof**:

$$\begin{aligned}\text{var}(\Phi_\theta(\boldsymbol{x})) &= E(\Phi_\theta(\boldsymbol{x})^2) - E(\Phi_\theta(\boldsymbol{x}))^2 \\ &= \int \Phi_\theta(\boldsymbol{x})^2 f_\theta(\boldsymbol{x})d\boldsymbol{x} - 0^2 = E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right]\end{aligned}$$

[the first part of the equation to proof]

- We haven seen that $\frac{\partial l}{\partial\theta} = \frac{\partial \ln f_\theta(\boldsymbol{X})}{\partial\theta} = \left[\frac{1}{f_\theta(\boldsymbol{X})}\right] f'_\theta(\boldsymbol{x})$
- Therefore,

$$\frac{\partial^2 l}{\partial\theta^2} = \frac{f''_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})} + f'_\theta(\boldsymbol{x})\left[-\frac{1}{f_\theta(\boldsymbol{x})^2} \times f'_\theta(\boldsymbol{x})\right]$$

Cramer-Rao inequality

$$
\begin{aligned}
\frac{\partial^2 l}{\partial\theta^2} &= \frac{f''_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})} - \frac{f'_\theta(\boldsymbol{x})^2}{f_\theta(\boldsymbol{x})^2} = \frac{f''_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})} - \left[\frac{f'_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})}\right]^2 \\
&= \frac{f''_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})} - \left[\frac{\partial l}{\partial\theta}\right]^2 \\
\Rightarrow \left[\frac{\partial l}{\partial\theta}\right]^2 &= -\frac{\partial^2 l}{\partial\theta^2} + \frac{f''_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathrm{var}(\Phi_\theta(\boldsymbol{x})) &= E\left(-\frac{\partial^2 l}{\partial\theta^2} + \frac{f''_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})}\right) \\
&= -E\left(\frac{\partial^2 l}{\partial\theta^2}\right) + E\left(\frac{f''_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})}\right) \\
&= -E\left(\frac{\partial^2 l}{\partial\theta^2}\right) + \int \frac{f''_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x})} f_\theta(\boldsymbol{x})d\boldsymbol{x}
\end{aligned}
$$

Cramer-Rao inequality

$$\text{var}(\Phi_\theta(\boldsymbol{x})) \quad = \quad -E\left(\frac{\partial^2 l}{\partial\theta^2}\right) + \int f_\theta''(\boldsymbol{x})d\boldsymbol{x}$$

- We have seen earlier that $\int f_\theta''(\boldsymbol{x})d\boldsymbol{x} = \frac{\partial}{\partial\theta}\int f_\theta(\boldsymbol{x})d\boldsymbol{x} = 0$
- Thus,

$$\text{var}(\Phi_\theta(\boldsymbol{x})) \quad = \quad E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right] = -E\left(\frac{\partial^2 l}{\partial\theta^2}\right)$$

- The Cramer-Rao inequality can therefore be written us

$$\text{var}(\hat{\theta}) \geq \frac{1}{\text{var}\left(\Phi_\theta(\boldsymbol{x})\right)} = \frac{1}{E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right]} = \frac{1}{-E\left(\frac{\partial^2 l}{\partial\theta^2}\right)}$$

Cramer-Rao inequality

- For a single observation (sample size $n = 1$), the likelihood function is

$$L(\theta; x) = f_\theta(x)$$

- The function $\Phi(.)$ is written

$$\Phi_{\theta,1}(x) = \ln f_\theta(x)$$

- The information content of a single observation is therefore

$$I_1(\theta) = \text{var}(\Phi_{\theta,1}).$$

- For a sample of n independent observations, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, the likelihood function is

$$L(\theta; \boldsymbol{x}) = f_\theta(x_1) f_\theta(x_2) \ldots f_\theta(x_n)$$

Cramer-Rao inequality

◇ Thus, we write the function $\Phi(.)$ as

$$\Phi_{\theta,n} = \frac{\partial}{\partial\theta} \ln L(\theta; \boldsymbol{x}) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f_\theta(x_i) = \sum_{i=1}^{n} \Phi_{\theta,1}(x_i)$$

◇ The information content,

$$\begin{aligned} I_n(\theta) &= \text{var}\left(\Phi_{\theta,n}(\boldsymbol{x})\right) = \sum_{i=1}^{n} \text{var}(\Phi_{\theta,1}(x_i)) \\ &= n\text{var}\left[\Phi_{\theta,1}(x)\right] \\ &= nI_1(\theta) \end{aligned}$$

◇ This shows that the information content is directly proportional to the sample size.

Efficiency

Theorem: If $\hat{\theta}$ is an unbiased estimator of θ and if

$$\text{var}(\hat{\theta}) = \frac{1}{nE\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right)^2} = \frac{1}{-nE\left(\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x)\right)}$$

then $\hat{\theta}$ is a **uniformly minimum variance unbiased estimator (UMVUE)** of θ.

- $\hat{\theta}$ is also referred to as an **efficient estimator (absolute efficiency).**
- Note that, if $L(\theta)$ is the likelihood function, then $\hat{\theta}$ is efficient if

$$\text{var}(\hat{\theta}) = \frac{1}{E\left(\frac{\partial}{\partial\theta}\ln L(\theta)\right)^2} = \frac{1}{-E\left(\frac{\partial^2}{\partial\theta^2}\ln L(\theta)\right)}$$

Cramer-Rao procedure to test for efficiency

(1) For the pdf, find $\frac{\partial}{\partial\theta} \ln f_\theta(x)$ and $\frac{\partial^2}{\partial\theta^2} \ln f_\theta(x)$

(2) Calculate $\frac{1}{-nE\left(\frac{\partial^2}{\partial\theta^2} \ln f_\theta(x)\right)}$ if $f_\theta(x)$ is smooth or else calculate $\frac{1}{nE\left(\frac{\partial}{\partial\theta} \ln f_\theta(x)\right)^2}$

(3) Calculate var$(\hat{\theta})$

(4) If the results in step 2 is equal to the results of step 3, then $\hat{\theta}$ is efficient for θ.

Example

Example: Let $x_1, x_2, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ population with density function, $f(x)$. Show that $\bar{X}$ is an efficient estimator for μ.

Solution: We first calculate the Cramer-Rao lower bound.

$$\begin{aligned}
f(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} = c e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \\
\ln(f(x)) &= \ln c + \ln e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} = c' - \frac{1}{2\sigma^2}(x_i-\mu)^2 \\
\frac{\partial}{\partial\mu}\ln(f(x)) &= \frac{x_i-\mu}{\sigma^2} \\
\frac{\partial^2}{\partial\mu^2}\ln(f(x)) &= -\frac{1}{\sigma^2}
\end{aligned}$$

Example

◇ Thus

$$\frac{1}{-nE\left(\frac{\partial^2}{\partial\mu^2}\ln f_\mu(x)\right)} = \frac{1}{nE(\frac{1}{\sigma^2})} = \frac{\sigma^2}{n} = \text{var}\left(\bar{X}\right).$$

◇ Therefore, $\bar{X}$ is an efficient estimator of μ. That is $\bar{X}$ is an UMVUE of μ.

Example: Suppose $p(x)$ is the Poisson distribution with parameter λ. Show that the sample mean $\bar{X}$ is an efficient estimator of λ.

Example

Solution: The density function $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

- Finding the lower bound of the Cramer-Rao inequality:

$$\begin{aligned}
\ln p(x) &= -\lambda + x\ln\lambda - \ln x! \\
\frac{\partial}{\partial\lambda}\ln p(x) &= -1 + \frac{x}{\lambda} \\
\frac{\partial^2}{\partial\lambda^2}\ln p(x) &= -\frac{x}{\lambda^2}
\end{aligned}$$

- This implies

$$\frac{1}{-nE\left(\frac{\partial^2}{\partial\lambda^2}\ln f_\lambda(x)\right)} = \frac{1}{nE\left(\frac{X}{\lambda^2}\right)} = \frac{1}{\frac{n}{\lambda^2}E(X)} = \frac{\lambda}{n}.$$

Example

◇ Also,

$$\text{var}(\hat{\lambda}) = \text{var}\left(\frac{\sum_{i=1}^{n}}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{var}(X_i) = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$$

◇ Since

$$\text{var}(\hat{\lambda}) = \frac{1}{-nE\left(\frac{\partial^2}{\partial\lambda^2}\ln f_\lambda(x)\right)} = \frac{\lambda}{n}$$

, $\hat{\lambda} = \bar{X}$ is an efficient estimator of λ.

Example

Example: For the exponential distribution with density $f(x) = \theta e^{-\theta x}$. Calculate the information content of a random sample of size n.

Solution:

$$\begin{aligned} f(x) &= \theta e^{-\theta x} \\ \ln f(x) &= \ln\theta - \theta x \\ \frac{\partial}{\partial\theta}\ln f(x) &= \frac{1}{\theta} - x \\ \frac{\partial^2}{\partial\theta^2}\ln f(x) &= -\frac{1}{\theta^2} \end{aligned}$$

- ◇ The information content is $\therefore I_n(\theta) = nE\left(\frac{\partial^2}{\partial\theta^2}\ln f(x)\right) = nE(-\frac{1}{\theta^2}) = \frac{n}{\theta^2}$.
- ◇ The lower bound of the variance of unbiased estimators of θ, based on a random sample of size n is $\frac{1}{I_n(\theta)} = \frac{\theta^2}{n}$.

Example

Example: Let $x_1, x_2, \ldots, x_n$ be a random sample from a Bernoulli trial with probability of success, p. Show that the maximum likelihood estimator is an efficient estimator.

Solution: $f(x) = p^x(1-p)^{1-x}, x = 0, 1$

- ◇ It is easy to show that the ML estimator for p is

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

- ◇ Compute the lower bound of the Cramer-Rao inequality using the log-likelihood given by

$$l = \ln p \sum_i X_i + (n - \sum_i X_i) \ln(1-p).$$

Example

$$\begin{aligned}
\frac{\partial l}{\partial p} &= \frac{\sum_i X_i}{p} - \frac{n - \sum_i X_i}{1-p} \\
&= \frac{\sum_i X_i - p\sum_i X_i - np + p\sum_i X_i}{p(1-p)} = \frac{\sum_i X_i - np}{p(1-p)} \\
\Rightarrow \left(\frac{\partial l}{\partial p}\right)^2 &= \left(\frac{\sum_i X_i - np}{p(1-p)}\right)^2
\end{aligned}$$

- Since we have used the likelihood for the n observations, the Cramer-Rao lower bound is

$$\frac{1}{E\left(\frac{\partial l}{\partial p}\right)} = \frac{1}{E\left(\frac{\sum_i X_i - np}{p(1-p)}\right)^2} = \frac{p^2(1-p)^2}{E\left(\sum_{i=1} X_i - np\right)^2}.$$

- We can see that $E\left(\sum_{i=1} X_i - np\right)^2 = np(1-p)$ is the variance of a binomial random variable.

Example

- Thus,

$$\frac{1}{E\left(\frac{\partial l}{\partial p}\right)^2} = \frac{p^2(1-p)^2}{np(1-p)} = \frac{p(1-p)}{n}$$

- Also,

$$\begin{aligned}\text{var}(\hat{p}) &= \text{var}\left(\frac{\sum_{i=1} X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1} \text{var}(X_i) \\ &= \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}\end{aligned}$$

- Since $\text{var}(\hat{p}) = \frac{1}{E\left(\frac{\partial l}{\partial p}\right)^2} = \frac{p(1-p)}{n}$, $\hat{p}$ is an efficient estimator for p.

Exercise

A sample of size n consist of independent observations, $X_1, X_2, \ldots, X_n$ where X_i has a Poisson distribution with mean $\lambda T_i, i = 1, 2, \ldots, n$ and $T_1, T_2, \ldots, T_n$ are known constants. Show that

(i) the information lower bound to the variance of unbiased estimators of λ is

$$\frac{\lambda}{\sum_{i=1}^{n} T_i}$$

(ii) the maximum likelihood estimator

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} T_i}$$

attains the lower bound for the variance of unbiased estimators of λ.

Properties of selected estimators

Moment Estimators

- ◇ Generally consistent whether the sampled population is finite or not.
- ◇ Usually not unbiased. Example, we saw that the moment estimator s_n^2 of the population variance is in fact biased.
- ◇ Because of the consistency property, the bias is likely to be small when sample size is large.
- ◇ In some important problems, moment estimators are identical with maximum likelihood estimators.
- ◇ In such situations, they are among the most efficient estimators.
- ◇ An advantage is in its simplicity and applicability to situations where very little is known about the underlying probability model.

Properties of selected estimators

Least squares estimators

- Are unbiased estimators
- Estimators are generally linear functions of the sample observations, i.e. they are called **linear estimators**.
- They are the best linear unbiased estimators, having smaller variance than any other linear unbiased estimators.
- When samples are independently normally distributed, these estimators are identical with ML estimators.
- In such cases, they have the smaller variance (MSE) and are more efficient than any other unbiased estimators.

Properties of selected estimators

Maximum likelihood estimators

- ◇ Produces some of the best least estimators.
- ◇ Not all ML estimators are unbiased but are generally consistent.
- ◇ The ML estimator attains the information lower bound variance for unbiased estimators when the sample size is large.
- ◇ The ML estimator is asymptotically normal.

Interval estimation

- In the previous sections, we studied methods to find point estimators for unknown population parameters.
- Standard error or variances of the estimator was also found as a measure of the average estimation error involved in the use a particular estimator.
- Suppose that $\bar{X} = 10.7$ is obtained as the estimator of a population mean, μ and standard error of $\bar{X}, SE(\bar{X}) = 1.5$.
- We interpret these as:
  - the population mean is approximately 10.7
  - the method that gave this estimate is subject to an error of average magnitude 1.5.

Interval estimation

- That is, the unknown parameter is likely to lie in the interval

$$10.7 \pm 1.5$$

- The interval does not provide the degree of assurance that μ lies in this interval (No measure of reliability).
- Another type of estimation called **interval estimation** is the focus of this section.

Probability error bound and confidence intervals

◇ Recall the Chebychev's inequality

$$P(|\hat{\theta} - \theta| \leq \epsilon) \geq 1 - \frac{MSE(\hat{\theta})}{\epsilon^2}$$

◇ Suppose we take ϵ to be twice the standard error of $\hat{\theta}$, then

$$P(|\hat{\theta} - \theta| \leq 2SE(\hat{\theta})) \geq 1 - \frac{MSE(\hat{\theta})}{4SE(\hat{\theta})^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0.75$$

◇ We interpret this as:

if $\hat{\theta}$ is used as an estimator of θ, then with probability of 75% or better, the error of estimation does not exceed two standard errors.

◇ Alternatively, this means that

at least 75% of the estimates produced by $\hat{\theta}$, lies within two standard errors of the true value of θ.

Constructing a confidence interval

- The interval $\left[\hat{\theta} - 2SE(\hat{\theta}), \hat{\theta} + 2SE(\hat{\theta})\right]$ or $\hat{\theta} \pm 2SE(\hat{\theta})$.
- This interval produced from the Chebychev's inequality is inefficient since it does not use any additional information regarding the sample distribution of the statistics or estimator.
- Suppose that $t_n(\boldsymbol{x})$ is a sample statistic and a number c can be found such that

$$P(|t_n(\boldsymbol{x}) - \theta| \leq c) = \gamma$$

then with probability γ, the maximum error bound is c.

Constructing confidence interval

- Equivalently,

$$P(t_n(\boldsymbol{x}) - c \leq \theta \leq t_n(\boldsymbol{x}) + c) = \gamma.$$

- The endpoints of the interval $[t_n(\boldsymbol{x}) - c, t_n(\boldsymbol{x}) + c]$ are random variables and hence the interval is a random interval.

- If two numbers c_1 and c_2 can be found such that

$$P(c_1 \leq t_n(\boldsymbol{x}) - \theta \leq c_2) = \gamma$$

then

$$P(t_n(\boldsymbol{x}) - c_2 \leq \theta \leq t_n(\boldsymbol{x}) - c_1) = \gamma$$

Constructing confidence interval

- Thus, the interval $[t_n(\boldsymbol{x}) - c_2, t_n(\boldsymbol{x}) - c_1]$ has a probability γ, called confidence level, of including θ.

- Generally, if two sample statistics $l_1(\boldsymbol{x})$ and $l_2(\boldsymbol{x})$ can be found such that

$$P(l_1(\boldsymbol{x}) \leq \theta \leq l_2(\boldsymbol{x})) = \gamma$$

  then the interval $[l_1(\boldsymbol{x}), l_2(\boldsymbol{x})]$ is the confidence interval of θ at confidence level γ.

- The endpoints of the confidence interval are called confidence limit at confidence level γ.

Pivot method

- A general method for constructing a confidence interval is using a pivot function.

- A pivot function is a function
  - of a random sample (a statistics or an estimator), and the unknown parameter θ, where θ is the only unknown quantity.

  - that has a probability distribution that does not depend on the parameter θ (has known sampling distribution)

- Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ denote a random sample from a normal population with mean μ and variance σ^2 both unknown.

Pivot method

- Let $\bar{X}$ be the sample mean and $S_{xx} = \sum_{i=1}^{n}(X_i - \bar{X})^2$, the sum of squares
  - $T_1(\boldsymbol{x}, \mu) = \frac{\bar{X}-\mu}{\sqrt{\frac{S_{xx}}{n(n-1)}}} = \frac{(\bar{X}-\mu)/\sqrt{n}}{\sqrt{\frac{S_{xx}}{n-1}}} \sim t_{n-1}$
  - $T_2(\boldsymbol{x}, \sigma) = \frac{S_{xx}}{\sigma^2} \sim \chi^2_{n-1}$.
- Both $T_1(\boldsymbol{x}, \mu)$ and $T_2(\boldsymbol{x}, \mu)$ are pivot functions because
  - their sampling distributions are completely determined by the sample size.
  - both involve any one unknown parameter.
- The unknown parameter is one of which we seek the confidence interval.

Pivot method

- Usually, the distribution of the pivot is $Z-, t-, \chi^2-$ or $F-$ distribution.

  Example: If $x_1, x_2, \ldots, x_n$ is sampled from $N(\mu, \sigma^2)$
  - if μ is unknown and σ is known, then

    $$T(\boldsymbol{x}, \mu) = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

    is a pivot with distribution, $N(0, 1)$
  - if μ is unknown and σ unknown, then

    $$T(\boldsymbol{x}, \mu) = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

    is a pivot function with distribution, t_{n-1}

Pivot method

- ∘ If σ^2 is unknown, the pivot function

$$T(\boldsymbol{x}, \sigma^2) = \frac{S_{xx}}{\sigma^2} = \frac{(n-1)S}{\sigma^2}$$

is a pivot function, where $S = \frac{1}{n-1} S_{xx}$.

- ◇ Once an appropriate pivot function, $T(\boldsymbol{x}, \theta)$ has been determined, we can construct confidence interval for the unknown parameter as follows:

$$P(c_1 \leq T(\boldsymbol{x}, \theta) \leq c_2) = \gamma.$$

- ◇ Next we write this in the equivalent form

$$P(l_1(\boldsymbol{x}) \leq \theta \leq l_2(\boldsymbol{x})) = \gamma.$$

- ◇ The random interval $[l_1(\boldsymbol{x}), l_2(\boldsymbol{x})]$ thus has probability γ of including θ.

Example

Example; Suppose $x_1, x_2, \ldots, x_n$ are drawn from $N(\mu, 1)$. Construct a 95% confidence interval for μ.

Solution: An appropriate pivot function is

$$T(\boldsymbol{x}, \mu) = \frac{\bar{X} - \mu}{1/\sqrt{n}} = \frac{(\bar{X} - \mu)\sqrt{n}}{1} \sim N(0, 1)$$

- Thus,

$$P(c_1 \leq T(\boldsymbol{x}, \mu) \leq c_2) = 0.95$$

- Since $T(\boldsymbol{x}, \mu) \sim N(0, 1)$, we can determine the values c_1 and c_2.
- c_1 is the value of $T(\boldsymbol{x}, \mu)$ with tail area 0.025. Thus, the 2.5th percentile of the Z table is $c_1 = -1.96$.
- c_2 is the 97.5th percentile of Z i.e. $c_2 = 1.96$

Example

◇ Therefore

$$
\begin{aligned}
P(-1.96 \leq (\bar{X}-\mu)\sqrt{n} \leq 1.96) &= 0.95 \\
P\left(-\frac{1.96}{\sqrt{n}} \leq \mu - \bar{X} \leq \frac{1.96}{\sqrt{n}}\right) &= 0.95 \\
P\left(\bar{X} - \frac{1.96}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{1.96}{\sqrt{n}}\right) &= 0.95
\end{aligned}
$$

◇ Hence, the 95% confidence interval for μ is $\bar{X} \pm \frac{1.96}{\sqrt{n}}$.

Example

If the organic sulfur found in MSM tablets is normally distributed with $\sigma = 14.2mg$. Construct a 95% for μ if on average the table contains about $80.5mg$ for a randomly selected 24 MSM tablets.

Solution

$\sigma = 14.2, n = 24, \bar{X} = 80.5$, then $\frac{\sigma}{\sqrt{n}} = \frac{14.2}{\sqrt{24}} = 2.8986$,
$Z_{1-\alpha} = Z_{0.95} = 1.96$. By substituting, the 95% CI is

$$\begin{aligned}\left[\bar{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] &= [80.5 - 1.96(2.8986), 80.5 + \\ &\quad 1.9(2.8986)] \\ &= [74.8184, 86.1813]\end{aligned}$$

There is 95% chance that population mean, μ lies between approximately 74.82 and 86.18.

Exercise

Suppose the distribution of shear bond strength is known to be normally distributed with $\sigma = 4.6$. If the average strength of 13 premolar samples is 9.6MPa, construct a 99% and 90% CI for μ.

Example

Example: Let $x_1, x_2, \ldots, x_{20}$ denote a sample from a normal population with mean μ and variance σ^2. Suppose that $\bar{X} = 12.3$ and $S_{xx} = 361.28$. Find the 90% confidence interval for σ.

- **Solution**: An appropriate pivot function is

$$T = \frac{S_{xx}}{\sigma^2} \sim \chi^2_{n-1}.$$

- We find c_1 and c_2 such that

$$P(c_1 \leq T \leq c_2) = 0.90$$

- There can be several possible choices of c_1 and c_2.

- Take c_1 and c_2 to be the 5th and 95th percentile of the χ^2_{19}, then $c_1 = 10.12$ and $c_2 = 30.14$.

- Therefore

Example

- The event

$$10.12 \leq \frac{S_{xx}}{\sigma^2} \leq 30.14$$

  is equivalent to

$$\begin{aligned}
\frac{1}{30.14} \leq &\quad \frac{\sigma^2}{S_{xx}} \quad \leq \frac{1}{10.12} \\
\sqrt{\frac{S_{xx}}{30.14}} \leq &\quad \sigma \quad \leq \sqrt{\frac{S_{xx}}{10.12}}
\end{aligned}$$

- Thus

$$P\left(\sqrt{\frac{S_{xx}}{30.14}} \leq \sigma \leq \sqrt{\frac{S_{xx}}{10.12}}\right) = 0.90$$

- There require confidence interval after substituting $S_{xx} = 361.28$ is $3.46 \leq \sigma \leq 5.97$.

Exponential parameter

- We have seen examples of CIs for normal parameters.
- The pivot method also applies to other distributions.
- Suppose that $X \sim Exp(\theta)$, the ML estimator for θ is $\hat{\theta} = \frac{1}{\bar{X}}$.
- For a random sample of size n from X we can use the pivot function

$$T(\boldsymbol{x}, \theta) = 2n\theta\bar{X} \sim \chi^2_{2n}$$

  to construct the CI for θ
- That is

$$\begin{aligned} P(c_1 \leq 2n\theta\bar{X} \sim\leq c_2) &= \gamma \\ P\left(\frac{c_1}{2n\bar{X}} \leq \theta \leq \frac{c_2}{2n\bar{X}}\right) &= \gamma \end{aligned}$$

- The interval $\left[\frac{c_1}{2n\bar{X}}, \frac{c_2}{2n\bar{X}}\right]$ is therefore the confidence interval for θ at γ confidence level.

Uniform parameter

- Suppose that $X \sim U(0, \theta)$. If n samples are drawn from X, the maximum likelihood estimator for θ has been determined to be $X_{(n)} = \max(X_i)$.

- The pivot
$$T(\boldsymbol{x}, \theta) = -2n \ln\left(\frac{X_{(n)}}{\theta}\right) \sim \chi^2_2$$
can be used to derive the CI for θ.

- That is
$$\begin{aligned} P\left(c_1 \le -2n \ln\left(\frac{X_{(n)}}{\theta}\right) \le c_2\right) &= \gamma \\ P\left(X_{(n)} \exp\left(\frac{c_1}{2n}\right) \le \theta \le X_{(n)} \exp\left(\frac{c_2}{2n}\right)\right) &= \gamma \end{aligned}$$

Uniform parameter

- Thus at γ confidence level, the CI for θ is $\left[X_{(n)} \exp\left(\frac{c_1}{2n}\right), X_{(n)} \exp\left(\frac{c_2}{2n}\right)\right]$
- Also, define $GM(\boldsymbol{x}) = (x_1, x_2, \ldots, x_n)^{\frac{1}{n}}$, the geometric mean of the sample $x_1, x_2, \ldots, x_n$ from $U(0, \theta)$
- We have shown that

$$T(\boldsymbol{x}, \theta) = -2n[\ln GM(\boldsymbol{x}) - \ln \theta] \sim \chi^2_{2n}$$

- The confidence interval for θ can also be derived at the confidence level γ as

$$GM(\boldsymbol{x}) \exp\left(\frac{c_1}{2n}\right) \leq \theta \leq GM(\boldsymbol{x}) \exp\left(\frac{c_2}{2n}\right)$$

Large sample confidence intervals

- The asymptotic (large sample) properties of the ML estimators make it possible to obtain approximate confidence intervals for parameters.

- If $\hat{\theta}_n$ is the ML estimator of θ based on the sample of size n, then as $n \to \infty$

$$Z_n = \frac{\hat{\theta}_n - \theta}{\sqrt{\text{var}(\hat{\theta}_n)}}$$

  is distributed approximately as normal, $N(0, 1)$.

- Thus,

$$P\left(c_1 \leq \frac{\hat{\theta}_n - \theta}{SE(\hat{\theta}_n)} \leq c_2\right) = \gamma$$

Example

- This leads to the approximate confidence interval

$$\hat{\theta}_n - cSE(\hat{\theta}_n) \leq \theta \leq \hat{\theta}_n + cSE(\hat{\theta}_n)$$

  where c is read from the standard normal table.

- **Example**: The estimator $\bar{X}$ is normally distributed with mean μ and variance, σ^2/n, where σ is assumed to be known.

- The large sample confidence interval for μ at γ confidence level is

$$\left[\bar{X} - c\frac{\sigma}{\sqrt{n}}, \bar{X} + c\frac{\sigma}{\sqrt{n}}\right]$$

  where c can be found to be $Z_{\alpha/2}$ and $1 - \alpha = \gamma$.

- α is called the significance level.

Approximate confidence interval for proportion, p

- **Example**: Let X_n denote the number of successes in n independent binomial trials each with success probability p. The ML estimator for p is $\hat{p} = \frac{X_n}{n}$ and $\text{var}(p) = \frac{p(1-p)}{n}$.

- Thus,
$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}, \quad Z' = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$
both of which are approximately normal, $N(0, 1)$ when n is large.

Approximate confidence interval for proportion, p

- The $(1-\alpha)\%$ confidence interval for p can be obtained from

$$P\left(c \leq Z = \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq c\right) = 1-\alpha$$

- Since $Z \sim N(0,1)$, the confidence interval is

$$\hat{p} - Z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Approximate confidence interval for Poisson parameter

- **Example**: Given a random sample $x_1, x_2, \ldots, x_n$ from the Poisson population with mean θ, the ML estimator for θ is $\bar{X}_n$ and $\text{var}(\bar{X}_n) = \frac{\theta}{n}$

- Define the pivot

$$Z_n = \frac{\bar{X}_n - \theta}{\sqrt{\frac{\theta}{n}}} \sim N(0,1)$$

- Thus,

$$P\left(\left|\frac{\bar{X}_n - \theta}{\sqrt{\frac{\theta}{n}}}\right| \leq Z_{\alpha/2}\right) = 1 - \alpha$$

Example

- The event

$$\begin{aligned} \frac{(\bar{X}_n - \theta)^2}{\frac{\theta}{n}} &\leq Z^2_{\alpha/2} \\ n(\bar{X}_n - \theta)^2 &\leq Z^2_{\alpha/2}\frac{\theta}{n} \\ n\theta^2 - (2n\bar{X} + Z^2_{\alpha/2})\theta + n\bar{X}^2_n &\leq 0 \end{aligned}$$

- For a 95% confidence interval for θ, we solve the quadratic equation

$$n\theta^2 - (2k\bar{X} + 3.84)\theta + n\bar{X}_n \leq 0$$

- The solution has the form $l_1 \leq \theta \leq l_2$ and provides the required confidence interval.

Applications

Example: Two statistics professors want to estimate average scores for an elementary statistics course that has two sections. Each professor teaches one section and each section has a large number of students. A random sample of 50 scores from each section produced the following results:

(a) Section I: $\bar{x}_1 = 77.01, \hat{\sigma}_1 = 10.32$

(b) Section II: $\bar{x}_2 = 72.22, \hat{\sigma}_2 = 11.02$

Calculate 95% confidence intervals for each of these three samples

- ◇ **Solution**: Because $n = 50$ is large, we could use normal approximation. For $\alpha = 0.05$, from the normal table: $Z_{\alpha/2} = Z_{0.025} = 1.96$.

Applications

- The confidence intervals are:
  (a) We have

$$\bar{x}_1 \pm Z_{\alpha/2}\frac{\hat{\sigma}_1}{\sqrt{n}} = 77.01 \pm 1.96\frac{10.32}{\sqrt{50}}$$

  which gives a 95% confidence interval $(74.149, 79.871)$.

  (b) We compute

$$\bar{x}_2 \pm Z_{\alpha/2}\frac{\hat{\sigma}_2}{\sqrt{n}} = 72.22 \pm 1.96\frac{11.02}{\sqrt{50}}$$

  which gives the interval $(69.165, 75.275)$.

Application

Example: An auto manufacturer gives a bumper-to-bumper warranty for 3 years or 36,000 miles for its new vehicles. In a random sample of 60 of its vehicles, 20 of them needed five or more major warranty repairs within the warranty period. Estimate the true proportion of vehicles from this manufacturer that need five or more major repairs during the warranty period, with confidence coefficient 0.95. Interpret.

Application

Solution: Here we need to find a 95% confidence interval for the true proportion, p. Here, $\hat{p} = 20/60 = 1/3$. For $\alpha = 0.05, Z_{\alpha/2} = Z_{0.025} = 1.96$. Hence, a 95% confidence interval for p is

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{1}{3} \pm 1.96\sqrt{\frac{\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)}{60}}$$

which gives the confidence interval as (0.21405, 0.45262).

◇ That is, we are 95% confident that the true proportion of vehicles from this manufacturer that need five or more major repairs during the warranty period will lie in the interval $(0.21405, 0.45262)$.

Example

Let X be a random variable representing film thickness of glass onomer cement. The sample mean and sample standard deviation of 10 samples are $\bar{x} = 19.9$ and $s = 1.3$. If X is normally distributed, find the 95% CI for μ.

Solution

We have, degree of freedom $n - 1 = 10 - 1 = 9$, standard error $= \frac{s}{\sqrt{n}} = \frac{1.3}{\sqrt{10}} = 0.4111$, $t_{9,0.975} = 2.2622$. Hence, the 95% CI for μ is given by

$$\left[\bar{X} - t_{9,0.975}\frac{s}{\sqrt{n}}, \bar{X} + t_{9,0.975}\frac{s}{\sqrt{n}}\right] = [18.97, 20.83]$$

Example

Of 331 crowns studied, 291 were successful to end of study. Construct a 95% CI for the population proportion, p of successful crowns.

Solution

$n = 331$ and $X = 291$,
$\hat{p} = \frac{291}{331} = 0.879, Z_{1-\alpha/2} = Z_{0.975} = 1.96$. We obtain

$$\left(\hat{p} - Z_{1-\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + Z_{1-\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

$$= \left[0.879 - 1.96\sqrt{\frac{0.879(0.121)}{331}}, 0.879 + 1.96\sqrt{\frac{0.879(0.121)}{331}}\right]$$

$$= (0.844, 0.914)$$

The 95% CI for p is given by $(0.844, 0.914)$

Confidence interval for two population parameters

- We have studied confidence intervals of true population parameters from a **single population.**
- Often times, interest is in obtaining confidence interval for parameters of interest based on two indepedent samples from **two populations.**
- Let $x_1, x_2, \ldots, x_{n_1}$ be a sample from the population $N(\mu_1, \sigma^2)$ and $y_1, y_2, \ldots, y_{n_2}$ a sample drawn from the population, $N(\mu_2, \sigma^2)$
- The sample sizes n_1 and n_2 need not be equal.
- Note that both populations have a common variance, σ^2.

Difference in means of two normal populations

- We require a confidence interval for the difference, $\mu_2 - \mu_1$ between the two population means.

Case I: (σ^2 is known)

- We use a pivot based on the difference in sample means $\bar{y} - \bar{x}$.
- This is a natural estimator for $\mu_2 - \mu_1$.
- If the two random samples are indepedent, then

$$\bar{y} - \bar{x} \sim N\left(\mu_2 - \mu_1, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

Difference in means of two normal populations

and

$$Z = \frac{(\bar{y} - \bar{x}) - (\mu_2 - \mu_1)}{SE(\bar{y} - \bar{x})} \sim N(0, 1)$$

where $SE(\bar{y} - \bar{x}) = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

- Thus, the $(1 - \alpha)\%$ confidence interval for $\mu_2 - \mu_1$ is

$$\bar{y} - \bar{x} \pm Z_{\alpha/2}\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Difference in means of two normal populations

Case II: (σ^2 is unknown)

- When σ^2 is not known, then Z can no longer be a pivot.
- An appropriate pivot will be to replace the $SE(\bar{y} - \bar{x})$ with the corresponding sample estimate, $se(\bar{y} - \bar{x})$.
- An unbiased estimator for the common variance, σ^2 is obtained by pooling the two samples together.
- That is,

$$\hat{\sigma}^2 = \frac{S_{xx} + S_{yy}}{n_1 + n_2 - 2}$$

Difference in means of two normal populations

- The standard error of the difference in means is estimated by

$$se(\bar{y}-\bar{x}) = \hat{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{S_{xx}+S_{yy}}{n_1+n_2-2}}$$

- This leads to the pivot

$$T = \frac{(\bar{y}-\bar{x})-(\mu_2-\mu_1)}{se(\bar{y}-\bar{x})}$$

- It can be shown that the sampling distribution of T is the t-distribution with n_1+n_2-2 degrees of freedom.

- Since this distribution is symmetric about zero, the confidence interval for $\mu_2-\mu_1$ is

$$\bar{y}-\bar{x} \pm t_{n_1+n_2-2,1/2(1+\gamma)}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{S_{xx}+S_{yy}}{n_1+n_2-2}}$$

Example

Suppose that there are six establishments that gave output data under an old service conditions, and six that gave data under a new conditions, where in each case the samples at drawn at random from the entire industry. The data from the two raw samples from the two populations of new and old is as follows:

Service conditions	Output per man-hour					
Old	10.0	32.0	9.7	15.1	17.2	10.1
New	13.6	2.8	11.1	23.8	19.0	9.2

Is there a change in output as a result of the introduction of new service conditions?

- ◇ Let μ_1 be the mean output per man-hour under the old service and μ_2 the mean under the new conditions.
- ◇ A change in output is then indicated by the difference $\mu_2 - \mu_1$ of the population means.

Example

- The sample means are $\bar{x} = 15.68$ for the old conditions and $\bar{y} = 16.58$ for the new conditions.

- The corresponding sum of squares are $S_{xx} = 46.999$ and $S_{yy} = 33.768$

- The sample sizes $n_1 = 6, n_2 = 6$ and the estimate of the common variance is

$$\hat{\sigma}^2 = \frac{S_{xx} + S_{yy}}{10} = 8.077$$

- The standard error, of $\bar{y} - \bar{x}$ is

$$se(\bar{y} - \bar{x}) = \hat{\sigma}\sqrt{\frac{1}{6} + \frac{1}{6}} = 1.641$$

Example

- With the given sample sizes, the pivot function T is distributed as t_{10}
- The 95% confidence interval for

$$\mu_2 - \mu_1$$

is therefore

$$\begin{array}{ccc} \bar{y} - \bar{x} & \pm & t_{10;0.05/2} se(\bar{y} - \bar{x}) \\ 0.90 & \pm & 2.23(1.641) \end{array}$$

- That is, $\mu_2 - \mu_1 \in [-2.76, 4.56]$

Interval estimate for variance of two normal populations

- We want to determine if two populations can reasonably be assumed to have equal variance or not.
- Let the populations be $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$.
- We determine the confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$ or $\frac{\sigma_2^2}{\sigma_1^2}$.
- A natural starting point is the sample variances

$$s_1^2 = \frac{S_{xx}}{n_1 - 1} \text{ and } s_2^2 = \frac{S_{yy}}{n_2 - 1}$$

- Recall that

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$$

Interval estimate for variance of two normal populations

- The central confidence interval, at confidence level γ is given by

$$\begin{aligned} P\left(c_1 \leq \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \leq c_2\right) &= \gamma \\ P\left(c_1 \frac{s_2^2}{s_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq c_2 \frac{s_2^2}{s_1^2}\right) &= \gamma \end{aligned}$$

Example: Consider the previous example. The sample estimates of the population variances are

$$\begin{aligned} s_1^2 &= \frac{S_{xx}}{n_1 - 1} = 46.999/5 = 9.3998 \\ s_2^2 &= \frac{S_{yy}}{n_2 - 1} = 33.768/5 = 6.7536 \end{aligned}$$

Example

- The pivot for determining the confidence interval for $\frac{\sigma_2^2}{\sigma_1^2}$ is distributed as $F_{5,5}$.

- The 95% central confidence interval is given by

$$
\begin{aligned}
F_{5,5}(0.025)\frac{6.7536}{9.998} \leq & \quad \frac{\sigma_2^2}{\sigma_1^2} \quad \leq F_{5,5}(0.975)\frac{6.7536}{9.998} \\
(0.14)(0.718) \leq & \quad \frac{\sigma_2^2}{\sigma_1^2} \quad \leq (7.15)(0.718) \\
0.10 \leq & \quad \frac{\sigma_2^2}{\sigma_1^2} \quad \leq 5.13
\end{aligned}
$$

- Since this confidence interval includes the value 1, we infer that the assumption $\frac{\sigma_2^2}{\sigma_1^2} = 1$ or $\sigma_1 = \sigma_2$, is reasonable.

Difference in means of two normal populations with unequal variances

- We consider how to deal with comparison of two population means when we have unequal population variance.

- Let the two populations be $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, and let n_1 and n_2 be random samples be drawn from X and Y, respectively.

- The difference $(\bar{y} - \bar{x}) \sim N\left(\mu_2 - \mu_1, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

Case I: (σ_1^2 and σ_2^2 are known)

- The standardized difference,

$$Z = \frac{(\bar{y} - \bar{x}) - (\mu_2 - \mu_1)}{SE(\bar{y} - \bar{x})} \sim N(0, 1)$$

Difference in means of two normal populations with unequal variances

- The confidence interval, at γ confidence level, is

$$\bar{y} - \bar{x} \pm Z_{\frac{1}{2}(1+\gamma)} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Case II: (σ_1^2 and σ_2^2 are unknown)

- We will replace the denominator of Z with the estimate of standard error, $se(\bar{y} - \bar{x})$

- This leads to

$$T = \frac{(\bar{y} - \bar{x}) - (\mu_2 - \mu_1)}{se(\bar{y} - \bar{x})},$$

where $se(\bar{y} - \bar{x}) = \sqrt{\frac{S_{xx}}{n_1(n_1-1)} + \frac{S_{yy}}{n_2(n_2-1)}}$.

Difference in means of two normal populations with unequal variances

- Using the following degrees of freedom,

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}},$$

  T is approximately t_v (t-distribution with v degrees of freedom)

- The number given in this formula is always rounded down for the degrees of freedom.

Example

Assuming that two populations are normally distributed with unknown and unequal variances. Two independent samples are taken with the following summary statistics:
$n_1 = 16, \bar{x}_1 = 20.17, s_1 = 4.3$ and $n_2 = 11, \bar{x}_2 = 19.23, s_2 = 3.8$
Construct a 95% confidence interval for $\mu_1 - \mu_2$.

Solution:

- First we compute the degrees of freedom,

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} = \frac{\left(\frac{4.3^2}{16} + \frac{3.8^2}{11}\right)^2}{\frac{\left(\frac{4.3^2}{16}\right)^2}{16-1} + \frac{\left(\frac{3.8^2}{11}\right)^2}{11-1}} = 23.312.$$

Hence, $v = 23$, and $t_{0.025,23} = 2.069$.

Example

Solution:

- ◇ The 95% confidence interval for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \quad \pm \quad Z_{\alpha/2;v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(20.17 - 19.23) \quad \pm \quad (2.069)\sqrt{\frac{4.3^2}{16} + \frac{3.8^2}{11}}$$

- ◇ Which gives the 95% confidence interval as $-2.3106 < \mu_1 - \mu_2 < 4.1906$.

Large sample approximation for non-normal population

- Samples may be drawn from two population which are not necessarily normal.

- We can use the central limit theorem (CLT) to obtain an approximate confidence interval for the difference between the population means.

- Let $x_1, x_2, \ldots, x_{n_1}$ be samples drawn from a population with mean μ_1 and variance, σ_1^2 and $y_1, y_2, \ldots, y_{n_2}$ be samples drawn from a population with mean μ_2 and variance, σ_2^2.

- For large samples, the CLT guarantees that

$$\bar{x} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ and } \bar{y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

Large sample approximation for non-normal population

- Thus, the difference $\bar{y} - \bar{x}$ is approximately, $N\left(\mu_2 - \mu_1, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

- The standardized difference,

$$Z = \frac{(\bar{y} - \bar{x}) - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Because of large sample sizes, we can replace the unknown variances in the denominator of Z by their sample estimates s_1^2 and s_2^2.

- Therefore we use the pivot

$$Z' = \frac{(\bar{y} - \bar{x}) - (\mu_2 - \mu_1)}{\sqrt{\frac{S_{xx}}{n_1(n_1-1)} + \frac{S_{yy}}{n_2(n_2-1)}}} \sim N(0, 1)$$

Example

A study of two kinds of machine failures shows that 58 failures of the first kind took on the average 79.7 minutes to repair with a standard deviation of 18.4 minutes, whereas 71 failures of the second kind took on average 87.3 minutes to repair with a standard deviation of 19.5 minutes. Find a 99% confidence interval for the difference between the true average amounts of time it takes to repair failures of the two kinds of machines.

Example

Solution

- Here, $n_1 = 58, n_2 = 71, \bar{x}_1 = 79.7, s_1 = 18.4, \bar{x}_2 = 87.3,$ and $s_2 = 19.5$. Then the 99% confidence interval for $\mu_1 - \mu_2$ is given by

$$
\begin{aligned}
(\bar{x}_1 - \bar{x}_2) &\pm Z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\
(79.7 - 87.3) &\pm 2.575\sqrt{\frac{18.4^2}{58} + \frac{19.5^2}{71}}
\end{aligned}
$$

- That is, we are 99% certain that $\mu_1 - \mu_2$ is located in the interval $(-16.215, 1.0149)$.

Large sample confidence interval for binomial proportions

- We describe the procedure for large sample confidence interval for the difference of true proportions, $p_1 - p_2$, in two binomially distributed populations.

- Let p_1 be the proportion of individuals with a certain characteristics in population 1 and p_2 the proportion in group 2.

- Also, let X_1 denote the number with the characteristics of interest in n_1 independent binomial trials with probability p_1.

- When n_1 is large, X_1 is approximately, $N(n_1 p_1, n_1 p_1 q_1)$ where $q_1 = 1 - p_1$

$$\hat{p}_1 = \frac{X_1}{n_1} \sim N\left(p_1, \frac{p_1 q_1}{n_1}\right)$$

Large sample confidence interval for binomial proportions

- Similarly,

$$\hat{p}_2 = \frac{X_2}{n_1} \sim N\left(p_2, \frac{p_2 q_1}{n_2}\right)$$

  when n_2 is large.

- The difference $\hat{p}_1 - \hat{p}_2$ is a natural estimate of $p_1 - p_2$.

- For independently large sample from both populations,

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

- Thus,

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{SE(\hat{p}_1 - \hat{p}_2)}$$

Large sample confidence interval for binomial proportions

- Since we have large samples, we can replace $SE(\hat{p}_1 - \hat{p}_2)$ with the sample estimate,

$$se(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

- Hence,

$$Z' = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{se(\hat{p}_1 - \hat{p}_2)}$$

can be used as a pivot function for constructing a confidence interval for $p_1 - p_2$.

Example

Iron deficiency, the most common nutritional deficiency worldwide, has negative effects on work capacity and on motor and mental development. In a 1999-2000 survey by the National Health and Nutrition Examination Survey (NHANES), iron deficiency was detected in 58 of 573 white, non-Hispanic females (10% rounded to whole number) and 95 of 498 (19% rounded to whole number) black, non-Hispanic females. Let p_1 be the proportion of black, non-Hispanic females with iron deficiency and let p_2 be the proportion of black, non-Hispanic females with iron deficiency. Obtain a 95% confidence interval for $p_1 - p_2$.

Solution:

- ◇ Here, $n_1 = 573$ and $n_2 = 498$. Also, $\hat{p}_1 = \frac{58}{573} = 0.10122 \approx 0.1$, and $\hat{p}_2 = \frac{95}{498} = 0.1907 \approx 0.19$. For $\alpha = 1 - \gamma = 0.05, Z_{0.05} = 1.96$

Example

$$
\begin{aligned}
(\hat{p}_1 - \hat{p}_2) &\pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \\
&= (0.1 - 0.19) \pm 1.96\sqrt{\frac{(0.1)(0.9)}{573} + \frac{(0.19)(0.81)}{498}} \\
&= (-0.13232, -0.047685).
\end{aligned}
$$

◇ Here, the true difference of $p_1 - p_2$ is located in the negative portion of the real line, which tells us that the true proportion of black, non-Hispanic females with iron deficiency is larger than the proportion of white, non-Hispanic females with iron deficiency.

Good luck with your exams!!