Lecture #9

Point Estimation

BMIR Lecture Series on Probability and Statistics

Ching-Han Hsu, Ph.D.
Department of Biomedical Engineering
and Environmental Sciences
National Tsing Hua University

Introduction

Statistical Inferences:

- We want some methods to make decisions or to draw conclusions about a **population**.
- We need **samples from population** and utilize the information within.
- The methods can be divided into two major areas: **parameter estimation** and **hypothesis testing**.

What is statistics?

- Statistic is a function of observations or random samples,
- Statistic itself is also a random variable.
- The probability distribution of a statistic is called a **sampling distribution**.

Point Estimation

If X is a random variable with probability function $f(x)$, characterized by the unknown parameter θ, and if $X_1, X_2, \ldots, X_n$ is a **random sample** of size n, the statistics $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$ is called a **point estimator** of θ.

- $\hat{\Theta}$ is a function of random variables $X_1, X_2, \ldots, X_n$.
- $\hat{\Theta}$ is a random variable, too.
- When the sample is selected, i.e., a set of observations $x_1, x_2, \ldots, x_n$ is available, $\hat{\Theta}$ takes on a particular numerical value $\hat{\theta}$, called the **point estimate** of θ.

Question: Can you distinguish the difference among estimation, estimator, and estimate?

Point Estimation: Example

For example, suppose that X is normally distributed with an unknown mean μ.

- The sample mean $\bar{X}$ is a point estimator of the unknown parameter mean μ.
- Is $\hat{\mu} = \bar{X}$?

Similarly, if the population variance σ^2 is unknown.

- The sample variance S^2 is a point estimator of the unknown parameter σ^2.
- Is $\hat{\sigma}^2 = s^2$?

Statistical Inference

Statistical inference is concerned with the making decisions about population based on the information contained in a random sample from that population.

- The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) that X_i's are independent random variables, and (b) every X_i has the same probability distribution.
- A **statistics** is any function of the observation in a random sample.
- The probability distribution of a statistics is called a **sample distribution**.

Statistical Inference

For example the sample mean

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

- $\bar{X} = \bar{X}(X_1, X_2, \ldots, X_n)$ is a function of $X_1, X_2, \ldots, X_n$.
- The probability distribution of $\bar{X}$ is called **sampling distribution of mean**.
- The sampling distribution of a statistics depends on the population distribution, sample size, and the method of sample selection.

Sample Mean of Normal Random Sample

Suppose that a random sample of size n is drawn from a normal population with mean μ and variance σ^2. Each observation in this sample, say $X_1, X_2, \ldots, X_n$, is a normally and independently distributed random variable with mean μ and variance σ^2. The sample mean

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

is a normal distribution $\sim N\left(\mu, \frac{\sigma^2}{n}\right)$ with mean

$$\mu_{\bar{X}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

and variance

$$\sigma^2_{\bar{X}} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Unbiased Estimators

Theorem

The point estimator $\hat{\Theta}$ *is an* **unbiased estimator** *for the parameter* θ *if*

$$E(\hat{\Theta}) = \theta. \tag{1}$$

If the estimator is biased, then the difference

$$b = E(\hat{\Theta}) - \theta. \tag{2}$$

is called the **bias** *of the estimator* $\hat{\Theta}$.

Ching-Han Hsu, Ph.D.

Estimators: Sample Mean and Variance

Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the distribution represented by X with mean μ and variance σ^2. Show that the sample mean

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

and sample variance

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

are unbiased estimators of μ and σ^2.

Sample Mean is Unbiased

The sample mean is defined as:

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

$$\begin{aligned} E(\bar{X}) &= E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\ &= \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right] \\ &= \frac{1}{n}\underbrace{(\mu + \mu + \cdots + \mu)}_{n} \\ &= \mu \end{aligned}$$

The sample mean is an unbiased estimator of μ.

Sample Variance is Unbiased

The sample variance is defined as:

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

$$\begin{aligned}
E(\bar{X}^2) &= E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right]^2 = \mu^2 + \frac{\sigma^2}{n} \\
E(X_i\bar{X}) &= E\left[\frac{X_1X_i + \cdots + X_iX_i + \cdots + X_nX_i}{n}\right] \\
&= \frac{1}{n}\left[E(X_1X_i) + \cdots + E(X_iX_i) + \cdots + E(X_nX_i)\right] \\
&= \frac{1}{n}(n\mu^2 + \sigma^2) = \mu^2 + \frac{\sigma^2}{n}
\end{aligned}$$

Sample Variance is Unbiased

$$
\begin{aligned}
E(S^2) &= E\left[\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}\right] \\
&= \frac{1}{n-1} E \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 = \frac{1}{n-1} \sum_{i=1}^{n} E\left(X_i - \bar{X}\right)^2 \\
&= \frac{1}{n-1} \sum_{i=1}^{n} E\left(X_i^2 - 2X_i\bar{X} + \bar{X}^2\right) \\
&= \frac{1}{n-1} \sum_{i=1}^{n} \left(\mu^2 + \sigma^2 - 2\mu^2 - 2\frac{\sigma^2}{n} + \mu^2 + \frac{\sigma^2}{n}\right) \\
&= \frac{1}{n-1} \sum_{i=1}^{n} \frac{n-1}{n}\sigma^2 \\
&= \sigma^2
\end{aligned}
$$

The sample variance is an unbiased estimator of σ^2.

Minimal Variance Principle of Estimator

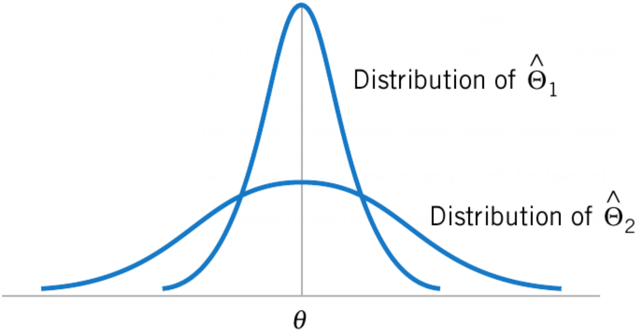

Figure 1: The sampling distribution of two unbiased estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$.

Standard Error

Definition

If we consider all unbiased estimator of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, the sample mean $\bar{X}$ is the MVUE for μ.

Definition

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation, given by $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those vales into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error** denoted by $\hat{\sigma}_{\hat{\Theta}}$, or $s_{\hat{\Theta}}$.

Standard Error: Example

Example

An article described a new method of measuring the thermal conductivity of Armco iron. Using a temperature of 100^oF and a power input of 550 watts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-^{o}F) were obtained:

41.60	41.48	42.34	41.95	41.86
42.18	41.72	42.26	41.81	42.04

- A point estimate of the mean thermal conductivity at 100^oF and 550 watts is the sample mean

$$\bar{X} = 41.924 \text{(Btu/hr-ft-}^o\text{F)}$$

Standard Error: Example

- The standard error of the sample mean is $\sigma_{\bar{X}} = \sigma / n$.
- Since σ is unknown, we replace σ by the sample deviation $s = 0.284$.
- The corresponding the estimated standard error of $\bar{X}$ is

$$\hat{\sigma}_{\hat{X}} = \frac{s}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898.$$

Mean Squared Error of Estimator

Theorem

The **mean squared error** *of an estimator* $\hat{\Theta}$ *of the parameter* θ *is defined as*

$$MSE(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2 = V(\hat{\Theta}) + (bias)^2 \qquad (3)$$

The MSE can be rewritten as

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

Mean Squared Error of Estimator

Ching-Han Hsu, Ph.D.

- Let $\hat{\Theta}_1$ and $\hat{\Theta}_2$ be two estimator of the parameter θ.
- Let $\text{MSE}(\hat{\Theta}_1)$ and $\text{MSE}(\hat{\Theta}_2)$ be the mean squared errors of $\hat{\Theta}_1$ and $\hat{\Theta}_2$.
- The **relative efficiency** of $\hat{\Theta}_1$ and $\hat{\Theta}_2$ is defined as

$$\frac{\text{MSE}(\hat{\Theta}_1)}{\text{MSE}(\hat{\Theta}_2)} \tag{4}$$

- If relative efficiency is less than 1, we would conclude that $\hat{\Theta}_1$ is a more efficient estimator of θ than $\hat{\Theta}_2$, in the sense that $\hat{\Theta}_1$ has a smaller mean squared error.

Mean Squared Error of Estimator

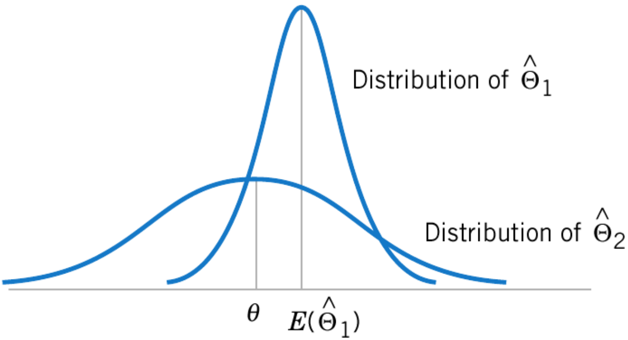

Figure 2: A biased estimator $\hat{\Theta}_1$ that has smaller variance than the unbiased estimator $\hat{\Theta}_2$.

Moments

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability distribution $f(x)$ (discrete or continuous). The kth **population moment** (or **distribution moment**) is $E(X^k), k = 1, 2, \ldots$. The corresponding kth **sample moment** is

$$E(X^k) \approx \frac{1}{n} \sum_{i=1}^{n} X_i^k, k = 1, 2, \ldots$$

Example

The first population moment is $E(X) = \mu$, and the first sample moment is $\frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$. The sample mean is the **moment estimator** of the population mean.

Moment Estimators

Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample from a probability function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The moment estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

Moment Estimators: Exponential Distribution

Example (Exponential Distribution)

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of an exponential distribution with parameter λ:

$$f(x) = \lambda e^{-\lambda x}, \quad 0 \leq x < \infty. \tag{5}$$

- λ is the only parameter and $E(X) = \mu = \frac{1}{\lambda}$.
- When $E(X) = \bar{X}$, this results in $\frac{1}{\lambda} = \bar{X}$.
- Therefore, $\hat{\lambda} = \frac{1}{\bar{X}}$ is the moment estimator of λ.
- Consider the failure rate of a part, we have collected the following failure time: $x_1 = 11.96, x_2 = 5.03, x_3 = 67.40, x_4 = 16.07, x_5 = 31.50, x_6 = 7.73, x_7 = 11.10, x_8 = 22.38$. Then, $\bar{x} = 21.65$ and $\hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{21.65} = 0.0462$.

Moment Estimators: Gaussian Distribution

Example (Gaussian Distribution)

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with parameters μ and σ^2. For the normal distribution $E(X) = \mu$ and $E(X^2) = \mu^2 + \sigma^2$. Equating $E(X) = \bar{X}$ and $E(X^2) = \frac{1}{n}\sum X_i^2$ gives

$$\mu = \bar{X}, \quad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

- The moment estimator of μ is $\hat{\mu} = \bar{X}$.
- The moment estimator of σ^2 is

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

This is a biased estimator of σ^2.

Ching-Han Hsu, Ph.D.

Moment Estimators: Gamma Distribution

Example (Gamma Distribution)

Suppose that Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a gamma distribution with parameters γ and λ:

$$f(x;\gamma,\lambda) = \begin{cases} \frac{\lambda^\gamma x^{\gamma-1} e^{-\lambda x}}{\Gamma(\gamma)} & 0 < x < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

For the gamma distribution $E(X) = \frac{\gamma}{\lambda}$ and $E(X^2) = \frac{\gamma(\gamma+1)}{\lambda^2}$. The moment estimators are found by solving

$$\frac{\gamma}{\lambda} = \bar{X}, \quad \frac{\gamma(\gamma+1)}{\lambda^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Moment Estimators: Gamma Distribution

- The resulting moment estimators for γ and λ are:

$$\hat{\gamma} = \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

$$\hat{\lambda} = \frac{\bar{X}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

- Consider the failure rate of a part, we have collected the following failure time: $x_1 = 11.96, x_2 = 5.03, x_3 = 67.40, x_4 = 16.07, x_5 = 31.50, x_6 = 7.73, x_7 = 11.10, x_8 = 22.38$. We have $\bar{x} = 21.65$ and $\sum_{i=1}^{8} x_i^2 = 6639.40$. Then

$$\hat{\gamma} = \frac{21.65^2}{\frac{1}{8} \cdot 6639.40 - 21.65^2} = 1.29$$

$$\hat{\lambda} = \frac{21.65}{\frac{1}{8} \cdot 6639.40 - 21.65^2} = 0.0598$$

Moment Estimators: Gamma Distribution

- When $\gamma = 1$, the gamma distribution reduces to the exponential distribution.
- $\hat{\gamma} = 1.29$ is slightly greater than 1, it is quite possible that either gamma ($\hat{\lambda} = 0.0598$) or exponential ($\hat{\lambda} = 0.0462$) distributions would provide a reasonable model for the data.

Maximum Likelihood Estimator

Definition

Suppose that X is a random variable with probability distribution $f(x;\theta)$, where θ is the single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size n. Then the **likelihood function** of the sample is

$$L(\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdots f(x_n;\theta) \tag{6}$$

The **maximum likelihood estimator** (MLE) of θ is the value of θ that maximizes the likelihood function $L(\theta)$

$$\hat{\theta} = \arg\max_{\theta} L(\theta) \tag{7}$$

Maximum Likelihood Estimator

- Note that the likelihood function is a function of only the unknown parameter θ.
- For any $\phi \neq \hat{\theta}$, $L(\phi) < L(\hat{\theta})$.
- If $x_1 < x_2$, then $\log x_1 < \log x_2$. This is known as the monotonically increasing property of the $\log$ function.
- If $\hat{\theta}$ is a maximizer of $L(\theta)$, then $\hat{\theta}$ is a maximizer of $\log L(\theta)$.
- The function $l(\theta) = \log L(\theta)$ is called as the **logarithmic likelihood function**.
- For a discrete distribution, the likelihood function of the sample $L(\theta)$ is simply the probability:

$$
\begin{aligned}
& P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n; \theta) \\
= \; & P(X_1 = x_1; \theta) \cdot P(X_2 = x_2; \theta) \cdots P(X_n = x_n; \theta) \\
= \; & \prod_{i=1}^{n} P(X_i = x_i; \theta)
\end{aligned}
$$

MLE: Bernoulli Distribution

Example

Let X be a Bernoulli random variable. The probability mass function is

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

where p is the parameter to be estimated. The likelihood function of a random sample of size n is

$$\begin{aligned} L(p) &= P(X_1 = x_1;\theta) \cdot P(X_2 = x_2;\theta) \cdots P(X_n = x_n;\theta) \\ &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

Show that the MLE of p is $\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i$.

MLE: Bernoulli Distribution

- The corresponding logarithmic likelihood function is

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

- To find the maximizer, we take the first derivative of $l(p)$ w.r.t. p:

$$\frac{dl(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{(n - \sum_{i=1}^{n} x_i)}{1-p}$$

- Then, equating this to zero and solving for p:

$$\begin{aligned} \frac{\sum_{i=1}^{n} x_i}{p} &= \frac{(n - \sum_{i=1}^{n} x_i)}{1-p} \\ (1-p)\sum_{i=1}^{n} x_i &= np - p\sum_{i=1}^{n} x_i \\ \hat{p} &= \frac{\sum_{i=1}^{n} x_i}{n} \end{aligned}$$

MLE: Exponential Distribution

Example

Let X be exponentially distributed with parameter λ. The likelihood function of a random sample of size n is

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

Find the MLE of λ.

- The log likelihood of $L(\lambda)$ is

$$l(\lambda) = \log L(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$$

- Take the derivative of $l(\lambda)$:

$$\frac{dl(\lambda)}{d\lambda} = \frac{d \log L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

MLE: Exponential Distribution

- Now equating this to zero, we have

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\frac{\sum_{i=1}^{n} x_i}{n}} = \frac{1}{\bar{X}}$$

- The ML estimator of λ is the reciprocal of the sample mean.
- This is the same as the moment estimator.

MLE: Exponential Distribution

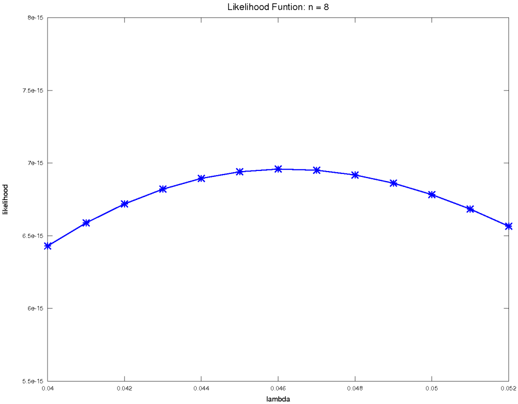

Figure 3: Likelihood function for the exponential distribution, using the failure time data: $x_1 = 11.96, x_2 = 5.03, x_3 = 67.40, x_4 = 16.07, x_5 = 31.50, x_6 = 7.73, x_7 = 11.10, x_8 = 22.38.$

MLE: Exponential Distribution

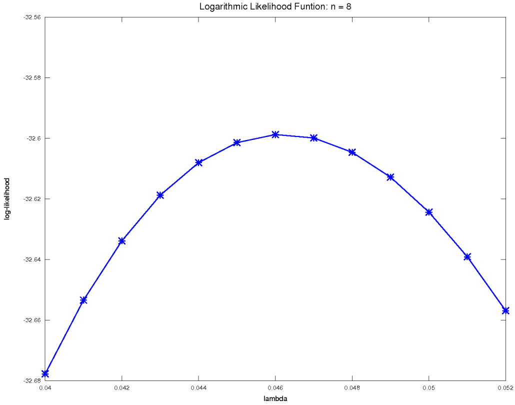

Figure 4: Logarithmic likelihood function for the exponential distribution, using the failure time data: $x_1 = 11.96, x_2 = 5.03, x_3 = 67.40, x_4 = 16.07, x_5 = 31.50, x_6 = 7.73, x_7 = 11.10, x_8 = 22.38.$

MLE: Multiple Parameters

- The MLE can be used in situations where there are several unknown parameters, say $\theta_1, \theta_2, \ldots, \theta_k$.
- The likelihood function is a function of k unknown parameters.
- The ML estimators $\{\hat{\Theta}_i\}$ would be found by equating the k partial derivatives of likelihood function to zero:

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

and then solving the resulting system of equations.

MLE: Normal Distribution with Unknown μ and σ^2

Example

Let X be normally distributed with mean μ and variance σ^2, where μ and σ^2 are unknown. The likelihood function of a random sample of size n is

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

Find the ML estimators of μ and σ^2.

- The corresponding logarithmic likelihood function is

$$\begin{aligned} l(\mu, \sigma^2) &= \log L(\mu, \sigma^2) \\ &= -\frac{n}{2} \log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \end{aligned}$$

MLE: Normal Distribution with Unknown μ and σ^2

- Taking partial derivatives w.r.t. μ and σ^2:

$$\frac{\partial \log L(\mu, \sigma^2)}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{\partial \log L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

- The ML estimators of μ and σ^2 are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

Properties of MLE

Theorem

Under very general conditions, when the sample size n is large and if $\hat{\Theta}$ is the maximum likelihood estimator of the parameter θ,

1. *$\hat{\Theta}$ is an approximately unbiased estimator for θ, i.e., $E(\hat{\Theta}) \approx \theta$,*
2. *the variance of $\hat{\Theta}$ us nearly small as the variance that could be obtained with any other estimator, and*
3. *$\hat{\Theta}$ has an approximate normal distribution.*

- Properties (1) and (2) state that the ML estimator is approximately an MVUE.
- To use ML estimation, remember that the distribution of the population must be either known or assumed.

Properties of MLE: Asymptotic

- Consider the ML estimator of σ^2, the variance of the normal distribution. We have

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

- $\hat{\sigma}^2$ is a biased estimator of σ^2. The bias is

$$E(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}$$

- The bias is negative so that $\hat{\sigma}^2$ tends to **underestimate** σ^2.
- As $n \to \infty$, $\hat{\sigma}^2$ asymptotically converges to σ^2. Then, $\hat{\sigma}^2$ is an **asymptotically unbiased estimator** of σ^2.