Chapter 1: Sampling Distributions and Statistical Inference

1 Some general concepts

Random sample - If $X_1, \ldots, X_n$ be independently and identically distributed random variables each having distribution function F, then $(X_1, \ldots, X_n)$ constitutes a random sample drawn from a theoretical distribution F.

Sample space - Let $(X_1, \ldots, X_n)$ be a random sample drawn from a theoretical distribution having distribution function F. Suppose $(x_1, \ldots, x_n)$ be a realization on $(X_1, \ldots, X_n)$. Then $(x_1, \ldots, x_n)$ is known as a sample point as it specifies a point in an n-dimensional space. Clearly, a sample point may vary from one occasion to the other. The totality of all such sample points constitute the sample space $\mathbb{X}$.

Statistic - Let $(X_1, \ldots, X_n)$ be a random sample drawn from a theoretical distribution having distribution function F_θ (F is characterized by a parameter θ). Consider $T(X_1, \ldots, X_n)$ such that $T: \mathbb{R}^n \rightarrow \mathbb{R}^k, k \geq 1$, and T is independent of any unknown θ. Then T is said to be a statistic.

In other words, a statistic is a function of sample observations which is independent of any unknown parameter. Here, if $k=1$, we get real valued statistics, e.g., $\bar{X}$, the sample mean; S^2, the sample variance, etc. For $k>1$, we get vector valued statistics, e.g., $\left(\bar{X}, S^2\right)$, $\left(\sum X_i, \sum X_i^2\right)$, etc.

1.1 Sampling from a finite population

Suppose we have a population consisting of a finite number, say N, of identifiable units, namely $T_1, \ldots, T_N$. A sample $\mathbb{S}$ is a non-empty collection of units from $\{T_1, \ldots, T_N\}$ with or without replacement, that is, a sample of size n is $\left(T_{i_1}, \ldots, T_{i_n}\right), i_j \in \{1, \ldots, N\} \; \forall \, j$. Let $t_1, \ldots, t_N$ be the variate values (values corresponding to some variable of interest) of $T_1, \ldots, T_N$ respectively. Consider the variate values for the sample units to be $X_1, \ldots, X_n$.

If the sample units are drawn one by one without replacement (WOR), then the sampling distribution will be

$$p(s) = 1 \Big/ \binom{N}{n}, \; s \in \mathbb{S}.$$

Consider the sample mean $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Let the observed value of $\bar{X}$ corresponding to s be $\bar{X}(s)$. Clearly, $\bar{X}(s)$ occurs with probability $p(s)$. The distribution of $\bar{X}$ will be called the sampling distribution of $\bar{X}$.

Standard error of a statistic is the standard deviation of the sampling distribution of that statistic.

In sampling WOR from a finite population, the independence condition in the definition of a random sample is not satisfied.

Here,

$$\begin{aligned}
P\left(X_1 = t_1\right) &= \frac{1}{N}, \\
P\left(X_2 = t_2 \mid X_1 = t_1\right) &= \frac{1}{N-1}, \\
P\left(X_2 = t_1 \mid X_1 = t_1\right) &= 0.
\end{aligned}$$

Thus the pmf of X_2 depends on X_1, violating the notion of independence. However,

$$\begin{aligned}
P\left(X_2 = t_2\right) &= \sum_{j=1}^N P\left(X_2 = t_2 \mid X_1 = t_j\right) P\left(X_1 = t_j\right) \\
&= \sum_{j \neq 2} P\left(X_2 = t_2 \mid X_1 = t_j\right) P\left(X_1 = t_j\right) \\
&= \frac{1}{N} \sum_{j \neq 2} P\left(X_2 = t_2 \mid X_1 = t_j\right) \\
&= \frac{1}{N}(N-1) \frac{1}{N-1} = \frac{1}{N}.
\end{aligned}$$

Hence, $X_1 \overset{d}{=} X_2$ (identically distributed). Similarly we can show that $X_1, \ldots, X_n$ all have the same distribution, but they are not independent. Random variables which are identically distributed but not necessarily independent are called exchangeable random variables.

1.2 Expectation and standard error of sample mean

Suppose a sample of size n is drawn from a finite population of size N. Let the variate values be $t_1, \ldots, t_N$. The population mean is

$$\mu = \frac{1}{N} \sum_{j=1}^{N} t_j,$$

and the population variance is

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} (t_j - \mu)^2 .$$

Let us denote the sample by $X_1, \ldots, X_n$. The sample mean is $\bar{X}$. Then, we have,

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{j=1}^{n} X_i\right) = \frac{1}{n} \sum_{j=1}^{n} E(X_i) . \tag{1}$$

$$\begin{aligned}
\operatorname{Var}(\bar{X}) &= E\{\bar{X} - E(\bar{X})\}^2 \\
&= \frac{1}{n^2} E\left\{\sum_{i=1}^{n} (X_i - E(X_i))\right\}^2 \\
&= \frac{1}{n^2} \sum_{i=1}^{n} E\{X_i - E(X_i)\}^2 + \frac{1}{n^2} \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} E\{X_i - E(X_i)\}\{X_j - E(X_j)\} \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \operatorname{Var}(X_i) + \frac{1}{n^2} \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} \operatorname{Cov}(X_i, X_j) .
\end{aligned} \tag{2}$$

Case I. Sampling with replacement (WR).

Note that,

$$P(X_i = t_j) = \frac{1}{N}, \forall i = 1, \ldots, n; j = 1, \ldots, N.$$

Hence,

$$E(X_i) = \sum_{j=1}^{N} t_j P(X_i = t_j) = \frac{1}{N} \sum_{j=1}^{N} t_j = \mu, \forall i = 1, \ldots, n.$$

And,

$$\begin{aligned}
\operatorname{Var}(X_i) = E(X_i - \mu)^2 &= \frac{1}{N} \sum_{j=1}^{N} (t_j - \mu)^2 \cdot P(X_i = t_j) \\
&= \frac{1}{N} \sum_{j=1}^{N} (t_j - \mu)^2 = \sigma^2
\end{aligned}$$

Also,

$$
\begin{aligned}
\operatorname{Cov}(X_i, X_j) &= E(X_i-\mu)(X_j-\mu) \\
&= \sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(t_{j_1}-\mu)(t_{j_2}-\mu)P(X_i=t_{j_1}, X_j=t_{j_2}) \\
&= \frac{1}{N^2}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(t_{j_1}-\mu)(t_{j_2}-\mu) \\
&= \frac{1}{N^2}\sum_{j_1=1}^{N}(t_{j_1}-\mu)\sum_{j_2=1}^{N}(t_{j_2}-\mu) \\
&= 0.
\end{aligned}
$$

Thus, using Eqns. (1) and (2), we get,

$$
\begin{aligned}
E(\bar{X}) &= \frac{1}{n}\sum_{i=1}^{n}\mu = \mu. \\
\operatorname{Var}(\bar{X}) &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}
$$

Standard error of $\bar{X}$ is $\frac{\sigma}{\sqrt{n}}$.

Case II. Sampling without replacement (WOR).

As in Case I, $E(X_i)=\mu$ and $\operatorname{Var}(X_i)=\sigma^2$, since distribution of $X_1,\ldots,X_n$ are identical. The covariance term needs special attention. For $i \neq j$,

$$
\begin{aligned}
P(X_i=t_{j_1}, X_j=t_{j_2}) &= P(X_i=t_{j_1} \mid X_j=t_{j_2})P(X_j=t_{j_2}) \\
&= \begin{cases} \frac{1}{N-1}\frac{1}{N}, j_1 \neq j_2 \\ 0, j_1 = j_2 \end{cases}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\operatorname{Cov}(X_i, X_j) &= \sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(t_{j_1}-\mu)(t_{j_2}-\mu)P(X_i=t_{j_1}, X_j=t_{j_2}) \\
&= \sum_{\substack{j_1=1 \\ j_1\neq j_2}}^{N}\sum_{j_2=1}^{N}(t_{j_1}-\mu)(t_{j_2}-\mu)\frac{1}{N(N-1)} \\
&= \frac{1}{N(N-1)}\sum_{j_1=1}^{N}(t_{j_1}-\mu)\left\{\sum_{j_2=1}^{N}(t_{j_2}-\mu)-(t_{j_1}-\mu)\right\} \\
&= \frac{1}{N(N-1)}\left\{\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(t_{j_1}-\mu)(t_{j_2}-\mu)-\sum_{j_1=1}^{N}(t_{j_1}-\mu)^2\right\} \\
&= \frac{1}{N(N-1)}\left(0-N\sigma^2\right) = \frac{-\sigma^2}{N-1}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E(\bar{X}) &= \mu \\
\operatorname{Var}(\bar{X}) &= \frac{1}{n^2}\left(n\sigma^2 + n(n-1)\frac{-\sigma^2}{N-1}\right) \\
&= \frac{N-n}{N-1}\frac{\sigma^2}{n}
\end{aligned}
$$

Standard error of $\bar{X}$ is $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$. The factor $\frac{N-n}{N-1}$ is called the finite population correction factor. As $N \to \infty, n$ fixed, the factor goes to unity.

1.3 Exact sampling distributions

1.3.1 Chi-squared distribution

Suppose

$$X_1, \ldots, X_n \overset{iid}{\sim} N(0,1)$$

Then, $\sum_{i=1}^{n} X_i^2$ is said to follow a chi-squared distribution with n degrees of freedom (df). We write,

$$Y = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2.$$

If $Y \sim \chi_n^2$, then the pdf of Y is given by

$$f(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, y > 0.$$

Also,

$$E(Y) = n, \ \text{Var}(Y) = 2n.$$

Remark 1.1. *1. Let $X \sim N\left(\mu, \sigma^2\right)$. Then,*

$$\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_1^2$$

2. If $X_1, \ldots, X_n \overset{iid}{\sim} N\left(\mu, \sigma^2\right)$, then,

(a)

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2.$$

(b)

$$\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2 \sim \chi_1^2.$$

1.3.2 Student's t-distribution.

Let $X \sim N(0,1), Y \sim \chi_n^2$, and X, Y are independent. Then, the statistic $T = \frac{X}{\sqrt{Y/n}}$ is said to follow a Student's t-distribution with n df, and we write,

$$T \sim t_n$$

The pdf of T is given by

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

If $T \sim t_n$, then, $E(T) = 0$, for $n > 1$, otherwise undefined.

$$\text{Var}(T) = \begin{cases} \frac{n}{n-2}, n > 2 \\ \infty, 1 < n \leq 2 \\ \text{undefined}, n \leq 1 \end{cases}$$

1.3.3 F-distribution

Let

$$X \sim \chi_p^2, Y \sim \chi_q^2,$$

and X, Y are independent. Then, the statistic

$$F = \frac{X/p}{Y/q}$$

is said to follow an F-distribution with (p, q) df. We write,

$$F \sim F_{p,q}.$$

The pdf of F is given by

$$f(x)=\frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)}\left(\frac{p}{q}\right)^{p/2}x^{p/2-1}\left(1+\frac{p}{q}x\right)^{-\frac{p+q}{2}},x>0$$

If $X\sim F_{p,q}$, then,

$$E(X)=\frac{q}{q-2},q>2.$$

$$\operatorname{Var}(X)=\frac{2q^2(p+q-2)}{p(q-2)^2(q-4)},q>4.$$

1.4 Sampling distribution of $\bar{X}$ and S^2 in case of Normal population

Let

$$X_1,\ldots,X_n\overset{IID}{\sim}N\left(\mu,\sigma^2\right).$$

Define

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

Then,

$$\bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right),$$

$$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{n-1},$$

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

1.5 Convergence in Probability Theory

A sequence of random variables $X_1,X_2,\ldots$, converges in probability to a random variable X if, for every $\epsilon>0$,

$$\lim_{n\to\infty}P\left(|X_n-X|\geq\epsilon\right)=0.$$

Frequently, statisticians are concerned with situations in which the limiting random variable is a constant and the random variables in the sequence are sample means (of some sort). This leads to an important result as below.

1.5.1 Weak Law of Large Numbers (WLLN)

Let $X_1,X_2,\ldots$, be iid random variables with $E\left(X_i\right)=\mu$ and $\operatorname{Var}\left(X_i\right)=\sigma^2<\infty$. Define $\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$. Then, for every $\epsilon>0$,

$$\lim_{n\to\infty}P\left(\left|\bar{X}_n-\mu\right|<\epsilon\right)=1$$

that is, $\bar{X}_n$ converges in probability to μ.

The proof is a straighforward application of Chebychev's inequality.

We have, for every $\epsilon>0$,

$$P\left(\left|\bar{X}_n-\mu\right|\geq\epsilon\right)\leq\frac{\operatorname{Var}\left(\bar{X}_n\right)}{\epsilon^2}=\frac{\sigma^2}{n\epsilon^2}.$$

Hence,

$$P\left(\left|\bar{X}_n-\mu\right|<\epsilon\right)=1-P\left(\left|\bar{X}_n-\mu\right|\geq\epsilon\right)\geq1-\frac{\sigma^2}{n\epsilon^2}\to1,$$

as $n\to\infty$. The WLLN quite elegantly states that, under general conditions, the sample mean approaches the population mean as $n\to\infty$.

1.5.2 Central Limit Theorem (CLT)

DeMoivre-Laplace CLT: Let $\{X_n\}$ be a sequence of iid Bernoulli random variables with $P(X_n = 1) = p = 1 - P(X_n = 0)$, $0 < p < 1$. Define the partial sums $S_n = \sum_{i=1}^{n} X_i$. Then,

$$\frac{S_n - np}{\sqrt{np(1-p)}} \sim N(0,1), \quad \text{asymptotically.}$$

Lindeberg-Levy CLT: Let $\{X_n\}$ be a sequence of iid random variables with $\text{Var}(X_n) = \sigma^2 < \infty$, and common mean μ. Let $S_n = \sum_{i=1}^{n} X_i$, $\bar{X}_n = S_n/n, n = 1, 2, \ldots$ Then, for every $x \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) = \lim_{n\to\infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le x\right) = \Phi(x),$$

where $\Phi(\cdot)$ is the CDF of a $N(0,1)$ distribution.

Remark 1.2.

- *Asymptotic distribution of* $\bar{X}_n$ *is given by*
$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$
- *Lindeberg Levy CLT directly implies DeMoivre-Laplace CLT (Note that* $E(X_i) = p, \text{Var}(X_i) = p(1-p)$*)*
- *CLT is concerned with the convergence of random variables to normality under certain general conditions. Initially, the theorems have been stated by several statisticians for a sequence of independent random variables.*
- *Whatever the exact distribution of* S_n *may be, discrete or continuous, the limiting distribution is continuous.*

2 Point Estimation

In many scenarios, finding a natural estimator for a parameter of interest is a trivial problem, and can be found using intuitive reasoning. However, in many complex scenarios, such natural estimators are not available easily, for which we need to resort to specific techniques of finding estimators. We shall discuss two such important methods in this course.

2.1 Method of Moments Estimator

Consider a random sample $X_1, \ldots, X_n$ from a population with distribution function F_θ, where $\theta = (\theta_1, \ldots, \theta_k)$. Method of moments estimators are found by equating the first k sample moments to the corresponding population moments, and solving the system of equations simultaneously. More precisely, define

$$\begin{aligned}
m_1' &= \frac{1}{n}\sum_{i=1}^{n} X_i,\ \mu_1' = E(X), \\
m_2' &= \frac{1}{n}\sum_{i=1}^{n} X_i^2,\ \mu_2' = E\left(X^2\right), \\
&\vdots \\
m_k' &= \frac{1}{n}\sum_{i=1}^{n} X_i^k,\ \mu_k' = E\left(X^k\right).
\end{aligned}$$

The population moments are ideally functions of the population parameters, so that we can explicitly express them as

$$\mu_r' = \mu_r'(\theta_1, \ldots, \theta_k), r = 1, \ldots, k.$$

The Method of Moments estimator (MOME) $\hat{\theta} = \left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right)$ is given by the solution to the system of equations

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

Example 2.1.1. *(Sampling from Normal). Let* $X_1, \ldots, X_n \overset{iid}{\sim} N\left(\mu, \sigma^2\right)$*. The parameter is given by* $\theta = (\theta_1, \theta_2)$*, where* $\theta_1 = \mu$ *and* $\theta_2 = \sigma^2$*. We have,* $m_1' = \bar{X}$ *and* $m_2' = \sum_{i=1}^n X_i^2/n$*. Also,* $\mu_1' = \mu$ *and* $\mu_2' = \sigma^2 + \mu^2$*. Thus, the system of equations for obtaining the MOME is:*

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

Solving for μ *and* σ^2 *gives the MOME as:*

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

Example 2.1.2. *(Sampling from Uniform). Let* $X_1, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$*. The system of equations for obtaining the MOME is:*

$$\bar{X} = \frac{\theta}{2},$$

so as to get the MOME of θ *as* $\hat{\theta} = 2\bar{X}$*.*

2.1.1 Exercises

1. Consider a random sample of size n from the Exponential distribution with rate λ. Find the MOME of λ.
2. Consider a random sample of size n from the Gamma(α, β) distribution. Find the MOME of (α, β)

2.2 Maximum Likelihood Estimator

We shall start with the concept of a likelihood function, which forms the core of a maximum likelihood estimator.

Definition 2.1. *(Likelihood function). Let* $f\left(x_1, \ldots, x_n \mid \theta\right)$ *denote the joint pmf or pdf of a random sample* $X_1, \ldots, X_n$*. Then, given that* $\left(X_1, \ldots, X_n\right) = \left(x_1, \ldots, x_n\right)$ *is observed, the likelihood function of* θ *is given by*

$$L\left(\theta; x_1, \ldots, x_n\right) = f\left(x_1, \ldots, x_n \mid \theta\right)$$

If the pmf/pdf of the underlying random variable is given by $f(x \mid \theta)$, then the joint likelihood is given by

$$f\left(x_1, \ldots, x_n \mid \theta\right) = \prod_{i=1}^n f\left(x_i \mid \theta\right)$$

so that the likelihood function of θ can be expressed as

$$L\left(\theta; x_1, \ldots, x_n\right) = \prod_{i=1}^n f\left(x_i \mid \theta\right)$$

Example 2.2.1. *Let* $X \overset{iid}{\sim} \text{Bin}(10, p),\ 0 < p < 1$*. Then for the observation* $x = 2$*, the likelihood function of* p *is given by*

$$f(p; x = 2) = \begin{pmatrix} 10 \\ 2 \end{pmatrix} p^2(1-p)^8, 0 < p < 1.$$

This is a polynomial function in p *of degree 10.*

Example 2.2.2. *Let* $X_1, \ldots, X_n \overset{iid}{\sim} N\left(\mu, \sigma^2\right)$*. Then, for observations* $X_1 = x_1, \ldots, X_n = x_n$*, the likelihood function of* $\left(\mu, \sigma^2\right)$ *is given by*

$$L\left(\mu, \sigma^2; x_1, \ldots, x_n\right) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left(x_i - \mu\right)^2\right\}, \mu \in \mathbb{R}, \sigma > 0$$

Example 2.2.3. *Consider a random sample $X_1, \ldots, X_n$ from the Uniform $(0, \theta)$ distribution. The likelihood function of θ is given by*

$$L(\theta; x_1, \ldots, x_n) = \left(\frac{1}{\theta}\right)^n, \theta > \max(x_1, \ldots, x_n)$$

Note the domain of the parameter θ in the above example. The pdf of an Uniform $(0, \theta)$ distribution is given by

$$f(x \mid \theta) = \frac{1}{\theta},\ 0 < x < \theta,$$

so that while defining the likelihood as a function of θ, one has to take care of the constraint that the joint pdf is positive only when θ is greater than each of the observations $x_1, \ldots, x_n$. This immediately reduces to the condition that θ must exceed the maximum of the observations.

Definition 2.2. *(Maximum Likelihood Estimator (MLE)). For each sample point $x_1, \ldots, x_n$, let $\hat{\theta}(x_1, \ldots, x_n)$ be a parameter value at which the likelihood function $L(\theta; x_1, \ldots, x_n)$ attains its maximum as a function of θ, with $x_1, \ldots, x_n$ held fixed. A maximum likelihood estimator (MLE) of the parameter θ based on a random sample $X_1, \ldots, X_n$ is then given by $\hat{\theta}(X_1, \ldots, X_n)$*

Intuitively, the MLE is a reasonable choice as a point estimator in the sense that it gives the parameter point expressed as a function of the sample at which the outcomes (observations) are most likely. The general strategy for finding an MLE is to differentiate the likelihood function with respect to the parameter(s), and find the point where global maximum is attained. However, there are many situations where resorting to differentiation will not be possible. We shall explore these different scenarios below.

In many scenarios, we will see that working with the logarithm of the likelihood function is easier. The logarithm of the likelihood, also called the log-likelihood is monotone increasing function of the likelihood, and hence the maximizer of the latter will be exactly equal to that of the former.

Example 2.2.4. *(Normal MLE; known variance). Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma_0^2), \mu \in \mathbb{R}, \sigma_0 > 0$, where σ_0 is known, but μ is unknown. We need to find the MLE of μ based on the above random sample. Let us first write the likelihood function of μ.*

$$\begin{aligned} L(\mu; x_1, \ldots, x_n) &= \prod_{i=1}^{n} \left[\frac{1}{\sigma_0\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_0^2}(x_i - \mu)^2\right\}\right] \\ &= \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}. \end{aligned}$$

Notice that it will be easier for us to differentiate the log-likelihood instead of working with the likelihood itself. The log-likelihood is given by

$$l(\mu; x_1, \ldots, x_n) = -\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu)^2 + \text{ constant },$$

where the constant involves terms free of μ. Differentiating both sides of the above equation and equating it to zero, we get,

$$2\sum_{i=1}^{n}(x_i - \mu) = 0,$$

so as to get the solution $\hat{\mu} = \bar{X}$. Note that

$$\frac{d^2}{d\mu^2} l(\mu; x_1, \ldots, x_n) = \frac{d}{d\mu}\left\{2\sum_{i=1}^{n}(X_i - \mu)\right\} = -2n < 0,$$

so that the solution gives the global maximum. Hence the MLE of μ is $\bar{X}$.

Example 2.2.5. *(Bernoulli MLE). Consider a random sample $X_1, \ldots, X_n$ from a Bernoulli distribution unknown probability of success p, where $0 \le p \le 1$. The likelihood function is given by*

$$\begin{aligned} L(p; x_1, \ldots, x_n) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

The log-likelihood is given by,

$$l(p; x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \ln(p) + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p).$$

Differentiating both sides wrt p, we get,

$$\frac{d}{dp} l(p; x_1, \ldots, x_n) = \frac{1}{p} \sum_{i=1}^{n} x_i - \frac{1}{1-p} \left(n - \sum_{i=1}^{n} x_i\right).$$

Equating the above to zero gives the solution

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

Hence the MLE is given by the sample proportion of success. [Check the double-derivative]

Example 2.2.6. *(Uniform MLE). Let $X_1, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$. The likelihood function is given by*

$$L(\theta; x_1, \ldots, x_n) = \left(\frac{1}{\theta}\right)^n, \theta > \max(x_1, \ldots, x_n).$$

Note that the above likelihood function is positive only in the range $(\max(x_1, \ldots, x_n), \infty)$ and zero otherwise. Also, the function is monotone decreasing in θ, so that the supremum is attained at the point where θ attains its infimum, so that the MLE is given by

$$\hat{\theta} = \max(X_1, \ldots, X_n).$$

2.2.1 Exercises

1. Consider sampling n random variables from a $N(\mu, \sigma^2)$ distribution with both the parameters unknown. Find the MLE of μ and σ^2.

2. Let $X_1, \ldots, X_n \overset{iid}{\sim} U(\theta, \theta + 1)$. Find the MLE of θ.

2.2.2 Invariance property of MLE

Theorem 2.1. *(Invariance property of MLE). Let $\hat{\theta}$ be the MLE of the parameter θ. Then, for any function $g(\theta)$, the MLE of $g(\theta)$ is given by $g(\hat{\theta})$.*

Example 2.2.7. *Consider sampling from an exponential distribution with rate λ. We need to find the MLE of the mean parameter μ, given by $\mu = 1/\lambda$. Let us find the MLE of λ first. The likelihood functio based on a random sample of size n is given by*

$$L(\lambda; x_1, \ldots, x_n) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right), \lambda > 0.$$

The log-likelihood function is

$$l(\lambda; x_1, \ldots, x_n) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i.$$

Differentiating both sides and equating to zero, we get,

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0,$$

so as to obtain the MLE of λ as $\hat{\lambda} = n / \left(\sum_{i=1}^{n} X_i\right)$. [Check the double-derivative condition].
Thus, by the invariance property of MLE, the MLE of $\mu = 1/\lambda$ is given by

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \bar{X}.$$

3 Confidence Intervals

A confidence interval is the mean of the estimate $\mp$ the variance of the estimate. This is the range of values that we expect to include our estimate with a predefined probability:

3.1 Computing normal confidence interval:

From the standard normal table, we have the following:

$$P(-1.96 < Z < 1.96) = P(Z < 1.96) - P(Z < -1.96) = 0.9750 - 0.0250 = 0.95.$$

i.e., there exists 95% probability that a standard normal variable will fall between -1.96 to 1.96.
Now, if $X_i \sim N(\mu, \sigma^2); i = 1, 2, \ldots, n$

$$\begin{aligned}
&Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \\
&\therefore P(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96) = 0.95 \\
&\Rightarrow P(-1.96\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96\frac{\sigma}{\sqrt{n}}) = 0.95 \\
&\Rightarrow P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95.
\end{aligned}$$

Hence, $(\bar{X} \mp 1.96\frac{\sigma}{\sqrt{n}})$ is the confidence interval for μ.

4 Testing of Hypothesis

4.1 Testing univariate sample

Tests for Normal mean

Random sample:

$$X_1, \ldots, X_n \overset{iid}{\sim} N\left(\mu, \sigma^2\right).$$

Case I: Population variance σ^2 known.

H_0	H_1	Test Statistic	Rejection rule
$\mu \leq \mu_0$	$\mu > \mu_0$	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	obsvd. $(Z) > z_\alpha$
$\mu \geq \mu_0$	$\mu < \mu_0$	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	obsvd. $(Z) < -z_\alpha$
$\mu = \mu_0$	$\mu \neq \mu_0$	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	obsvd. $(\lvert Z\rvert) > z_{\alpha/2}$

Table 1: ***Testing univariate samples: variance σ^2 known***

Case II: Population variance σ^2 unknown.

Tests for proportion: large samples

Random sample:

$$X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}(p)$$

Test statistic:

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

Sample size n is large.

H_0	H_1	Test Statistic	Rejection rule
$\mu \leq \mu_0$	$\mu > \mu_0$	$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$	obsvd. $(T) > t_{\alpha,n-1}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$	obsvd. $(T) < -t_{\alpha,n-1}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$	obsvd. $(\|T\|) > t_{\alpha/2,n-1}$

Table 2: ***Testing univariate samples: variance σ^2 unknown***

H_0	H_1	Test Statistic	Rejection rule
$p \leq p_0$	$p > p_0$	$Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$	obsvd. $(Z) > z_\alpha$
$p \geq p_0$	$p < p_0$	$Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$	obsvd. $(Z) < -z_\alpha$
$p = p_0$	$p \neq p_0$	$Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$	obsvd. $(\|Z\|) > z_{\alpha/2}$

Table 3: ***Testing univariate Bernoulli samples***

4.2 Testing two independent samples

Testing Normal means

Random sample:

$$X_1, \ldots, X_n \overset{iid}{\sim} N\left(\mu_X, \sigma_X^2\right)$$
$$Y_1, \ldots, Y_m \overset{iid}{\sim} N\left(\mu_Y, \sigma_Y^2\right)$$

Case I: σ_X^2, σ_Y^2 are known.

H_0	H_1	Test Statistic	Rejection rule
$\mu_X - \mu_Y \leq \mu_0$	$\mu_X - \mu_Y > \mu_0$	$Z = \frac{\bar{X}-\bar{Y}-\mu_0}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$	obsvd. $(Z) > z_\alpha$
$\mu_X - \mu_Y \geq \mu_0$	$\mu_X - \mu_Y < \mu_0$	$Z = \frac{\bar{X}-\bar{Y}-\mu_0}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$	obsvd. $(Z) < -z_\alpha$
$\mu_X - \mu_Y = \mu_0$	$\mu_X - \mu_Y \neq \mu_0$	$Z = \frac{\bar{X}-\bar{Y}-\mu_0}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$	obsvd. $(\|Z\|) > z_{\alpha/2}$

Table 4: ***Testing two independent univariate samples: variances known.***

Case II: σ_X^2, σ_Y^2 are unknown.
Here, the degrees of freedom of the t-distribution is given by

$$\nu = \frac{\left(S_X^2/n + S_Y^2/m\right)^2}{\frac{\left(S_X^2/n\right)^2}{n-1} + \frac{\left(S_Y^2/m\right)^2}{m-1}}$$

This is known as Satterthwaite approximation formula.

Case III: σ_X^2, σ_Y^2 are unknown but equal.
S_p^2 is the pooled variance, given by

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}.$$

Testing two proportions from independent samples: large samples

H_0	H_1	Test Statistic	Rejection rule
$\mu_X - \mu_Y \leq \mu_0$	$\mu_X - \mu_Y > \mu_0$	$T = \frac{\bar{X}-\bar{Y}-\mu_0}{\sqrt{S_X^2/n+S_Y^2/m}}$	obsvd. $(T) > t_{\alpha,\nu}$
$\mu_X - \mu_Y \geq \mu_0$	$\mu_X - \mu_Y < \mu_0$	$T = \frac{\bar{X}-\bar{Y}-\mu_0}{\sqrt{S_X^2/n+S_Y^2/m}}$	obsvd. $(T) < -t_{\alpha,\nu}$
$\mu_X - \mu_Y = \mu_0$	$\mu_X - \mu_Y \neq \mu_0$	$T = \frac{\bar{X}-\bar{Y}-\mu_0}{\sqrt{S_X^2/n+S_Y^2/m}}$	obsvd. $(\|T\|) > t_{\alpha/2,\nu}$

Table 5: ***Testing two independent univariate samples: variances unknown***

H_0	H_1	Test Statistic	Rejection rule
$\mu_X - \mu_Y \leq \mu_0$	$\mu_X - \mu_Y > \mu_0$	$T = \frac{\bar{X}-\bar{Y}-\mu_0}{\sqrt{S_p^2(1/n+1/m)}}$	obsvd. $(T) > t_{\alpha,n+m-2}$
$\mu_X - \mu_Y \geq \mu_0$	$\mu_X - \mu_Y < \mu_0$	$T = \frac{\bar{X}-\bar{Y}-\mu_0}{\sqrt{S_p^2(1/n+1/m)}}$	obsvd. $(T) < -t_{\alpha,n+m-2}$
$\mu_X - \mu_Y = \mu_0$	$\mu_X - \mu_Y \neq \mu_0$	$T = \frac{\bar{X}-\bar{Y}-\mu_0}{\sqrt{S_p^2(1/n+1/m)}}$	obsvd. $(\|T\|) > t_{\alpha/2,n+m-2}$

Table 6: ***Testing two independent univariate samples: variances unknown but equal***

Random sample:

$$X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}\,(p_1)$$
$$Y_1, \ldots, Y_m \overset{iid}{\sim} \text{Ber}\,(p_2)$$

Test statistic:

$$\hat{p}_1 - \hat{p}_2,$$

where

$$\hat{p}_1 = \frac{\sum_{i=1}^n X_i}{n}, \hat{p}_2 = \frac{\sum_{i=1}^m Y_i}{m}.$$

Sample sizes n and m are large.

H_0	H_1	Test Statistic	Rejection rule
$p_1 - p_2 \leq p_0$	$p_1 - p_2 > p_0$	$Z = \frac{\hat{p}_1-\hat{p}_2-p_0}{\sqrt{\hat{p}_1\hat{q}_1/n+\hat{p}_2\hat{q}_2/m}}$	obsvd. $(Z) > z_\alpha$
$p_1 - p_2 \leq 0$	$p_1 - p_2 > 0$	$Z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n+1/m)}}$	obsvd. $(Z) > z_\alpha$
$p_1 - p_2 \geq p_0$	$p_1 - p_2 < p_0$	$Z = \frac{\hat{p}_1-\hat{p}_2-p_0}{\sqrt{\hat{p}_1\hat{q}_1/n+\hat{p}_2\hat{q}_2/m}}$	obsvd. $(Z) < -z_\alpha$
$p_1 - p_2 \geq 0$	$p_1 - p_2 < 0$	$Z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n+1/m)}}$	obsvd. $(Z) < -z_\alpha$
$p_1 - p_2 = p_0$	$p_1 - p_2 \neq p_0$	$Z = \frac{\hat{p}_1-\hat{p}_2-p_0}{\sqrt{\hat{p}_1\hat{q}_1/n+\hat{p}_2\hat{q}_2/m}}$	obsvd. $(\|Z\|) > z_{\alpha/2}$
$p_1 - p_2 = 0$	$p_1 - p_2 \neq 0$	$Z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n+1/m)}}$	obsvd. $(\|Z\|) > z_{\alpha/2}$

Table 7: ***Testing proportions from two independent populations. Note the difference in the test statistic for*** $p_0 \neq 0$***.***

Here $\hat{p}$ is the pooled estimator of proportion, given by

$$\hat{p} = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m} = \frac{\sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i}{n+m}.$$

4.3 Testing means of paired Normal data

Random sample: $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} N_2\left(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho\right)$.
Define $Z_i = X_i - Y_i,\ i = 1, \ldots, n$. Then, $\bar{Z} = \bar{X} - \bar{Y}$. Denote expecation and variance of Z_i's as

$$\mu_Z = \mu_X - \mu_Y, \quad \sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y.$$

Note that, $Z_1, \ldots, Z_n \overset{iid}{\sim} N\left(\mu_Z, \sigma_Z^2\right)$
Case I: $\sigma_X^2, \sigma_Y^2, \rho$ are known.

H_0	H_1	Test Statistic	Rejection rule
$\mu_Z \leq \mu_0$	$\mu_Z > \mu_0$	$Z = \frac{\bar{Z}-\mu_0}{\sigma_Z/\sqrt{n}}$	obsvd. $(Z) > z_\alpha$
$\mu_Z \geq \mu_0$	$\mu_Z < \mu_0$	$Z = \frac{\bar{Z}-\mu_0}{\sigma_Z/\sqrt{n}}$	obsvd. $(Z) < -z_\alpha$
$\mu_Z = \mu_0$	$\mu_Z \neq \mu_0$	$Z = \frac{\bar{Z}-\mu_0}{\sigma_Z/\sqrt{n}}$	obsvd. $(\lvert Z\rvert) > z_{\alpha/2}$

Table 8: ***Testing means of bivariate samples: all other parameters known***

Case II: $\sigma_X^2, \sigma_Y^2, \rho$ are unknown.

H_0	H_1	Test Statistic	Rejection rule
$\mu_Z \leq \mu_0$	$\mu_Z > \mu_0$	$T = \frac{\bar{Z}-\mu_0}{S_Z/\sqrt{n}}$	obsvd. $(T) > t_{\alpha,n-1}$
$\mu_Z \geq \mu_0$	$\mu_Z < \mu_0$	$T = \frac{\bar{Z}-\mu_0}{S_Z/\sqrt{n}}$	obsvd. $(T) < -t_{\alpha,n-1}$
$\mu_Z = \mu_0$	$\mu_Z \neq \mu_0$	$T = \frac{\bar{Z}-\mu_0}{S_Z/\sqrt{n}}$	obsvd. $(\lvert T\rvert) > t_{\alpha/2,n-1}$

Table 9: ***Testing means of bivariate samples: all other parameters unknown***