Statistics
Semester 4

Contents

Chapter 1

Revision of Probability

I'm simply gonna list rules.

$$\begin{aligned}
\mathbb{E}(X) = \mu &= \sum_{i\in\Omega} X_i \Pr(X_i) \\
\mathbb{E}(g(X)) &= \sum_{i\in\Omega} g(X_i)\Pr(X_i) \\
\mathbb{E}(aX+b) &= a\mathbb{E}(X)+b \\
\mathbb{E}(X+Y) &= \mathbb{E}(X)+\mathbb{E}(Y) \quad \text{if both variables are independent}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \sigma^2 = \mathbb{E}(X^2)-\mu^2 \\
\mathrm{Var}(aX+bY) &= a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)+2ab\,\mathrm{cov}(X,Y)
\end{aligned}$$

where

$$\mathrm{cov}(X,Y) = \mathbb{E}(XY)-\mathbb{E}(X)\cdot\mathbb{E}(Y).$$

1.1 Discrete Distributions

1. **Uniform discrete law**

$$\begin{aligned}
&X(\Omega) = \{1,2,3,\ldots,n\} \\
&\Pr(X=k) = \frac{1}{n} \quad \forall k = 1,2,3,\ldots,n \\
&\begin{cases}\mathbb{E}(X) = \frac{n+1}{2} \\ \mathrm{Var}(X) = \frac{n^2-1}{12}\end{cases}
\end{aligned}$$

2. **Bernoulli law of parameters** p $(0<p<1)$

$$\begin{aligned}
&X \sim \mathrm{B}(p) \\
&X(\Omega) = \{0,1\} \\
&\Pr(X=1) = p \quad \Pr(X=0) = 1-p \\
&\begin{cases}\mathbb{E}(X) = p \\ \mathrm{Var}(X) = p(1-p)\end{cases}
\end{aligned}$$

3. **Binomial law of parameters n and p**

$$\begin{aligned}
&X \sim \mathrm{Bin}(n,p)\\
&X(\Omega) = \{1,2,\ldots,n\}\\
&\Pr(X=1) = C_n^k p^k q^{n-k} \quad \forall k \in \{0,1,2,\ldots,n\}\\
&\begin{cases} \mathbb{E}(X) = np \\ \mathrm{Var}\,(\mathrm{X}) = np(1-p) \end{cases}
\end{aligned}$$

4. **Hypergeometric law**

$$\begin{aligned}
&X \sim \mathcal{H}(N,n,p)\\
&X(\Omega) = [\max\{0, n-N+M\}, \min\{M,n\}]\\
&\Pr(X=k) = \frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n} \quad \forall k \in X(\Omega)\\
&\begin{cases} \mathbb{E}(X) = np \\ \mathrm{Var}\,(X) = np(1-p)\left(\frac{N-n}{N-1}\right) \end{cases}
\end{aligned}$$

5. **Geometric law**

$$\begin{aligned}
&\mathrm{X} \sim \mathrm{G}(p)\\
&X(\Omega) = \mathbb{N}^*\\
&\Pr(\mathrm{X}=k) = p(1-p)^{k-1} \quad \forall k \in \mathbb{N}^*\\
&\begin{cases} \mathbb{E}(\mathrm{X}) = \frac{1}{p} \\ \mathrm{Var}\,(X) = \frac{1-p}{p^2} \end{cases}
\end{aligned}$$

6. **Poisson's law of parameter λ ($\lambda \in \mathbb{R}_+^*$)**

$$\begin{aligned}
&X \sim \mathcal{P}(\lambda)\\
&X(\Omega) = \mathbb{N}\\
&\Pr(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}\\
&\begin{cases} \mathbb{E}(\mathrm{X}) = \lambda \\ \mathrm{Var}\,(\mathrm{X}) = \lambda \end{cases}
\end{aligned}$$

1.2 Continuous Distributions

1. **Uniform law**

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{else} \end{cases}$$

$$\begin{cases} \mathbb{E}(x) = \frac{a+b}{2} \\ \mathrm{Var}\,(x) = \frac{(b-a)^2}{12} \end{cases}$$

2. **Exponential law**

$$\begin{aligned}
&x \sim \xi(\lambda)\\
&f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}\\
&\begin{cases} \mathbb{E}(x) = \frac{1}{\lambda} \\ \mathrm{Var}\,(x) = \frac{1}{\lambda^2} \end{cases}
\end{aligned}$$

3. **Normal law**

$$x \sim \mathcal{N}(\mu, \sigma)$$
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$\begin{cases} \mathbb{E}(x) = \mu \\ \mathrm{Var}\,(x) = \sigma^2 \end{cases}$$

For $\mathcal{N}(0,1)$

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x.$$
$$\pi(z) = \Phi(z) - 0.5 = \int_{0}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x.$$

1.3 Convergence

Theorem 1.3.1 Chebyshev's inequality

Let X be a random variable of expectation $\mathbb{E}(X)$ and variance $\mathrm{Var}\,(X)$. Then $\forall \varepsilon$

$$\Pr\left(|X - \mathbb{E}(X)| \geqslant \varepsilon\right) \leqslant \frac{\mathrm{Var}\,(X)}{\varepsilon^2}.$$

it can also be stated as

$$\Pr\left(|X - \mathbb{E}(X)| < \varepsilon\right) \geqslant 1 - \frac{\mathrm{Var}\,(X)}{\varepsilon^2}.$$

We say a sequence of random variables X_n converges to a $(X_n) \xrightarrow{\Pr} a$ if $\forall \varepsilon$

$$\lim_{n\to+\infty} \Pr\left(|X_n - a| > \varepsilon\right) = 0.$$

or

$$\lim_{n\to+\infty} \Pr\left(|X_n - a| \leqslant \varepsilon\right) = 1.$$

Theorem 1.3.2 Weak law of large numbers

Consider a random variable (X_n) of mean μ and variance σ^2. Consider the random variable $\tilde{X}_n = \frac{X_1+X_2+\cdots+X_n}{n}$. It can be shown that $\tilde{X}_n$ converges to μ meaning $\forall \varepsilon$

$$\lim_{n\to+\infty} \Pr\left(|\tilde{X}_n - \mu| > \varepsilon\right) = 0.$$

1.4 Approximations

Theorem 1.4.1 Binomial by a Poisson

$$\mathrm{Bin}(n,p) \sim \mathcal{P}(np) \quad \text{if } \begin{cases} n \geqslant 30 \\ p \leqslant 0.1 \\ np < 15 \end{cases}.$$

Theorem 1.4.2 Hypergeometric by a Binomial

$$\mathcal{H}(N,n,p) \sim \mathrm{Bin}(n,p) \quad \text{if } n \leqslant 0.05N.$$

Theorem 1.4.3 De Moivre–Laplace theorem

$$\mathrm{Bin}(n,p) \sim \mathcal{N}\left(np, \sqrt{np(1-p)}\right) \quad \text{if } \begin{cases} n \geqslant 30 \\ np \geqslant 5 \\ n(1-p) \geqslant 5 \end{cases}.$$

In this case the event $X = k$ can be replaced by $k - 0.5 < X < l + 0.5$

Theorem 1.4.4 Central limit theorem

Let (X_n) be a sequence of independent random variables following the same law of expectation μ and of standard deviation σ. Let $S_n = \sum_{i=1}^{n} X_i$ and $S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. It can be shown that S_n^* converges in law to $\mathcal{N}(0,1)$.

$$\mathbb{E}(S_n) = n\mu$$
$$\mathrm{Var}\,(S_n) = n\sigma^2$$

1.5 Further laws

Theorem 1.5.1 Chi square law

Let $X_1, X_2, \ldots, X_n$ be n independent random variables following the standard normal law $\mathcal{N}(0,1)$. Let $Y = {X_1}^2 + {X_2}^2 + \cdots + {X_n}^2$. We say that Y follows a chi-square law with n degrees of freedom. $Y \sim {\chi_n}^2$.

$$\mathbb{E}(Y) = n$$
$$\mathrm{Var}\,(Y) = 2n$$

It can be shown that the density function of Y is

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}.$$

where Γ is the gamma function

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} \mathrm{d}t \quad \forall x > 0.$$

Theorem 1.5.2 Student law(t-distribution)

Let X, Z be two independent random random variables such that $X \sim \mathcal{N}(0,1)$ and $Z \sim {\chi_n}^2$. Hence the random variable

$$T = \frac{X}{\sqrt{\frac{Z}{n}}}.$$

is said to be following a student law. $T \sim \mathcal{T}_n$

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

Chapter 2

Estimators

Let θ be a certain characteristic of a population P of N individuals, for exmaple letting θ be the expectation of a certain random variable X defined over the population. We take a sample of size $n < N$ of the population to estimate the value of θ.

Let Y_n be a function of the random variables $X_1, X_2, \ldots, X_n$. Y_n is called an estimator of θ if

$$\lim_{n\to+\infty} \mathbb{E}(Y_n) = \theta.$$

a consistent estimator if

$$\lim_{n\to+\infty} \mathrm{Var}\,(Y_n) = 0.$$

and an unbiased estimator if

$$\mathbb{E}(Y_n) = \theta \quad \forall n \in \mathbb{N}^*.$$

the value y_n of Y_n computed from any observed sample is called point estimation of θ

2.1 Point estimation of the mean

Let X be a random variable defined over the population P of the expected value μ and standard deviation σ. Consider a sample $(X_1, X_2, \ldots, X_n)$ of size n, randomly selected from P such that X_i are independent and follow the same law.

Consider the statistic $\bar{X}_n = \frac{X_1+X_2+\cdots+X_n}{n}$, it is a random variable whose distribution is called the sample distribution of the mean.

$$\mathbb{E}(\bar{X}_n) = \mu$$
$$\mathrm{Var}\left(\bar{X}_n\right) = \frac{\sigma^2}{n}$$

Since $\mathrm{Var}\left(\bar{X}_n\right) \xrightarrow[n\to+\infty]{} 0$ then $\bar{X}_n$ is a consistent unbiased estimator of the mean μ.

Note:-

The standard deviation of $\bar{X}_n$ is called standard error of the mean

$$\sigma(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}.$$

Due to the central limit theorem, as the sample size gets larger and larger $\bar{X}_n$ approaches a normal distribution $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

2.2 Point estimator of the variance

2.2.1 Suppose μ is unknown

Consider the random variable S^2 (estimator of σ^2)

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

The expectation of S^2 can be proved to be

$$\mathbb{E}(S^2) = \frac{n-1}{n}\sigma^2.$$

Since $\mathbb{E}(S^2) \xrightarrow[n\to+\infty]{} \sigma^2$ then S^2 is a biased estimator of σ^2.

Consider the random variable S'^2

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

Since $\mathbb{E}(S'^2) = \sigma^2$ then S'^2 is an unbiased estimator of σ^2.

Hence σ can be estimated by

$$S' = \sqrt{\frac{n}{n-1}}S.$$

and

$$\sigma(\bar{X}_n) = \frac{S}{\sqrt{n-1}}.$$

where

σ^2 variance of the population.

S^2 variance of the sample.

$\sigma^2(\bar{X}_n)$ variance of the distribution of the sample mean.

S'^2 corrected variance of the sample.

Note:-

It is better to estimate σ^2 using S'^2 than S^2 since S^2 is a biased estimator. However, if n (sample size) is big enough $\left(\frac{n}{n-1} \approx 1\right)$, then σ^2 can be estimated by S^2

2.2.2 Suppose μ is known

Consider the random variable Z^2 (not the variance of the sample)

$$Z^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Since $\mathbb{E}(Z^2) = \sigma^2$ then Z^2 is an unbiased estimator of σ^2 thus the value $z^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$ is a point estimation of the variance σ^2 of the population.

Note:-

If $n > 0.05N$ and if the sample is selected without replacement then the value of the variance changes to become

$$\mathrm{Var}\left(\bar{X}_n\right) = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}.$$

and the standard error becomes

$$\sigma(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

If the variance of the population is not known then we can use use S^2 or Z^2 to estimate $\mathrm{Var}\left(\bar{X}_n\right)$

$$\mathrm{Var}\left(\bar{X}_n\right) = \left(\frac{N-n}{N-1}\right)\frac{S^2}{n-1}.$$

and the standard error with

$$\sigma(\bar{X}_n) = \frac{S}{\sqrt{n-1}}\sqrt{\frac{N-n}{N-1}}.$$

2.3 Point estimation of a proportion (percentage)

Consider a population P of individuals with a proportion p if individuals having a certain characteristic θ. Let $(a_1, a_2, \ldots, a_n)$ be a sample randomly selected P. We define for each individual a_i the Bernoulli random variable X_i as follows

$$\begin{cases} X_i = 1 & \text{if } a_i \text{ has the characteristic } \theta \text{ with probability } p \\ X_i = 0 & \text{else} \end{cases}.$$

Let $Y_n = \frac{X_1+X_2+\cdots+X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Y_n is the random variable giving the proportion of individuals of the sample that have the characteristic θ.

$$\begin{aligned} \Pr(X_i = 1) &= \frac{\text{number of individuals of the population having } \theta}{\text{total number of individuals}} = p \\ \Pr(X_i = 0) &= 1 - p \end{aligned}$$

Thus $X_1 + X_2 + \cdots + X_n \sim \mathrm{Bin}(n, p)$

$$\begin{aligned} \mathbb{E}(X_1 + X_2 + \cdots + X_n) &= np \\ \mathrm{Var}(X_1 + X_2 + \cdots + X_n) &= np(1-p) \end{aligned}$$

$$\begin{aligned} \mathbb{E}(Y_n) &= p \\ \mathrm{Var}(Y_n) &= \frac{p(1-p)}{n} \end{aligned}$$

Hence Y_n is a consistent unbiased estimator of p. Therefore any observed value y_n of Y_n is a point estimator of P, meaning p is estimated by the proportion of the sample.

2.4 Confidence interval

2.4.1 Confidence interval for the mean

1. **Suppose that $n \geqslant 30$, the population is normally distributed, and σ is known**
   Let X be a random variable defined over a population P of mean $\mathbb{E}(X) = \mu$ and of variance $\mathrm{Var}(X) = \sigma^2$.

   Here we consider that $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Hence $\frac{\bar{X}_n - \mu}{\sigma_{\bar{X}_n}} \sim \mathcal{N}(0,1)$.

   Given the probability γ (level of confidence), we can find t such that

$$\begin{aligned} \Pr\left(-t \leqslant \frac{\bar{X}_n - \mu}{\sigma_{\bar{X}_n}} \leqslant t\right) &= \gamma \\ \Pr\left(\bar{X}_n - t\sigma_{\bar{X}_n} \leqslant \mu \leqslant \bar{X}_n + t\sigma_{\bar{X}_n}\right) &= \gamma \end{aligned}$$

where $\pi(t) = \frac{\gamma}{2}$. Knowing γ we get t. Therefore a $\gamma\%$ confidence interval for the mean μ is given by

$$\mathrm{IC}_\gamma(\mu) = [\bar{x}_n - t\sigma_{\bar{X}_n}, \bar{x}_n + t\sigma_{\bar{X}_n}].$$

where

$$\sigma_{\bar{X}_n} = \begin{cases} \frac{\sigma}{\sqrt{n}} & \text{if } \sigma \text{ is known} \\ \frac{S}{\sqrt{n-1}} & \text{if } \sigma \text{ is unknown (estimated by } S' = \sqrt{\frac{n}{n-1}}S) \end{cases}.$$

2. **Suppose that $n < 30$, the population is normally distributed, and σ is unknown**:

   Using the table of student distributed knowing γ, we determine t such that

   $$\Pr\left(\bar{X}_n - t\frac{S}{\sqrt{n-1}} \leqslant \mu \leqslant \bar{X}_n + t\frac{S}{\sqrt{n-1}}\right) = \gamma.$$

   hence the confidence inteval for the mean μ is

   $$\mathrm{IC}_\gamma(\mu) = \left[\bar{X}_n - t\frac{S}{\sqrt{n-1}}, \bar{X}_n + t\frac{S}{\sqrt{n-1}}\right].$$

Theorem 2.4.1

(a) $\bar{X}_n$ and S^2 are two independent random variance.

(b) The random variable $n\frac{S^2}{\sigma^2}$ follows a chi-square law with $n-1$ degrees of freedom.

Theorem 2.4.2
The random variable

$$\tilde{T} = \frac{\bar{X}_n - \mu}{\frac{S'}{\sqrt{n}}} = \frac{\bar{X}_n - \mu}{\frac{S}{\sqrt{n-1}}}.$$

follows a student law (t-distribution) with $n-1$ degrees of freedom

3. **Suppose that $n < 30$, the population is not normally distributed**:
   In this case we cannot use the normal distributed nor the student distribution. However we can use Chebyshev's inequality.

   $$\Pr\left(|\bar{X}_n - \mu| \leqslant \varepsilon\right) \geqslant 1 - \frac{\sigma^2_{\bar{X}_n}}{\varepsilon^2}.$$

   Take $\varepsilon = t\sigma_{\bar{X}_n}$

   $$\Pr\left(\bar{X}_n - t\sigma_{\bar{X}_n} \leqslant \mu \leqslant \bar{X}_n + t\sigma_{\bar{X}_n}\right) \geqslant 1 - \frac{1}{t^2}.$$

   Then we set $1 - \frac{1}{t^2}$ equal to γ solve for t and find the interval as follows

   $$\mathrm{IC}_\gamma = [\bar{x}_n - t\sigma_{\bar{X}_n}, \bar{x}_n + t\sigma_{\bar{X}_n}].$$

Note:-

- if σ is known then $\sigma_{\bar{X}_n} = \frac{\sigma}{\sqrt{n}}$
- if σ is unknown then we replace $\sigma_{\bar{X}_n}$ by its point estimator $\frac{S'}{\sqrt{n}} = \frac{S}{\sqrt{n-1}}$

2.4.2 Confidence interval for a proportion (precentage)

same setup as last time. If we assume this time that $\text{Bin}(n,p) \approx \mathcal{N}\left(np, \sqrt{np(1-p)}\right)$ if $(n \geqslant 30,\ np, n(1-p) \geqslant 5)$ then we can say $Y_n \sim \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$. Knowing γ we can determine t such that

$$\Pr\left(-t \leqslant \frac{Y_n - p}{\sigma_{Y_n}} \leqslant t\right) = \gamma.$$

The confidence interval becomes

$$[y_n - t\sigma_{Y_n}, y_n + t\sigma_{Y_n}].$$

where $\sigma_{Y_n} = \sqrt{\frac{p(1-p)}{n}}$ estimated by

$$\sqrt{\frac{n}{n-1}}\sqrt{\frac{y_n(1-y_n)}{n}} = \sqrt{\frac{y_n(1-y_n)}{n-1}}.$$

Therefore the confidence interval becomes

$$\text{IC}_\gamma(p) = \left[y_n - t\sqrt{\frac{y_n(1-y_n)}{n-1}}, y_n + t\sqrt{\frac{y_n(1-y_n)}{n-1}}\right].$$

Note:-

If $n \geqslant 100$ then $\frac{n}{n-1} \approx 1$, then the confidence interval is

$$\left[y_n - t\sqrt{\frac{y_n(1-y_n)}{n}}, y_n + t\sqrt{\frac{y_n(1-y_n)}{n}}\right].$$

Note:-

If the sample is selected without replace and if $n > 0.05N$ then we shall put a correcting factor $\frac{N-n}{N-1}$ to $\sigma_{Y_n} = \sqrt{\frac{p(1-p)}{n}}$, thus the confidence interval for proportion p becomes

$$\left[y_n - t\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{y_n(1-y_n)}{n-1}}, y_n + t\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{y_n(1-y_n)}{n-1}}\right].$$

2.4.3 Confidence interval for the variance

Assume $X \sim \mathcal{N}(\mu, \sigma)$ and $X_1, X_2, \ldots, X_n$ n independent random variables and identically distributed as X. We set the variables

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

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

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

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

we have

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$
$$n\frac{S^2}{\sigma^2} \sim {\chi_{n-1}}^2$$
$$n\frac{Z^2}{\sigma^2} \sim {\chi_n}^2$$

1. Suppose μ is unknown

   Since $n\frac{S^2}{\sigma^2} \sim {\chi_{n-1}}^2$, then we determine the values $v_{\alpha/2}$ and $v_{1-\alpha/2}$ from the chi-square table such that

   $$\Pr\left(v_{\alpha/2} \leqslant \frac{nS^2}{\sigma^2} \leqslant v_{1-\alpha/2}\right) = \gamma = 1 - \alpha.$$

   therefore a confidence interval of level γ (risk α) is given by

   $$IC_\gamma(\sigma^2) = \left[\frac{nS^2}{v_{1-\alpha/2}}, \frac{nS^2}{v_{\alpha/2}}\right] = \left[\frac{(n-1)S'^2}{v_{1-\alpha/2}}, \frac{(n-1)S'^2}{v_{\alpha/2}}\right].$$

2. Suppose μ is known

   From the chi-square table, we determine the values of the quantities $v_{\alpha/2}$ and $v_{1-\alpha/2}$ for the law ${\chi_n}^2$ such that

   $$\Pr\left(v_{\alpha/2} \leqslant \frac{nZ^2}{\sigma^2} \leqslant v_{1-\alpha/2}\right) = \gamma.$$

   Therefore the confidence interval of level γ is given by

   $$\mathrm{IC}_\gamma(\sigma^2) = \left[\frac{nz^2}{v_{1-\alpha/2}}, \frac{nz^2}{v_{\alpha/2}}\right].$$

2.5 Mean Squared Error

Consider the estimator $y_n = f(x_1, x_2, \ldots, x_n)$ of θ. The bias of y_n relative to θ is

$$\mathrm{Bias}(y_n) = \mathbb{E}(y_n) - \theta.$$

The mean squared error of y_n with respect to θ is

$$\mathrm{MSE}(y_n) = \mathbb{E}\left[(y_n - \theta)^2\right].$$

It can also be shown that

$$\mathrm{MSE}(y_n) = \mathrm{Var}(y_n) + \mathrm{Bias}(y_n)^2.$$

If y_n is an unbiased estimator of θ then $\mathrm{Bias}(y_n) = \mathbb{E}(y_n) - \theta = 0 \Rightarrow$

$$\mathrm{MSE}(y_n) = \mathrm{Var}(y_n).$$

Assume y_n and z_n are two estimators of the same parameter θ. We say y_n is more efficient than z_n if

$$\mathrm{MSE}(y_n) < \mathrm{MSE}(z_n).$$

Assume y_n and z_n are two unbiased estimators of θ, then y_n is more efficient then z_n if and only if $\mathrm{Var}(y_n) < \mathrm{Var}(z_n)$

Chapter 3

Hypothesis Testing

3.1 Introduction

In hypothesis testing, we have a null-hypothesis H_0 on a sample space Ω and an alternative hypothesis H_1 on Ω. We want to test H_0 against H_1.

To do so we consider a random sample size n and calculate the probability that H_0 is within a certain *significane level* and deduce if we can accpet H_0.
As an example consider $H_0 = \mu \neq m$, then H_1 is

$$\left.\begin{aligned} H_1 &= \mu > m \\ H_1 &= \mu < m \end{aligned}\right\} \text{One sided test}$$
$$\left.H_1 = \mu \neq m\right\} \text{Two sided test}$$

Type I error

$$\alpha = \Pr(\text{Type I error}) = \Pr(\text{Reject } H_0 | H_0 \text{ is true}).$$

Type II error

$$\beta = \Pr(\text{Type II error}) = \Pr(\text{Accept } H_0 | H_0 \text{ is false}).$$

Power of the test

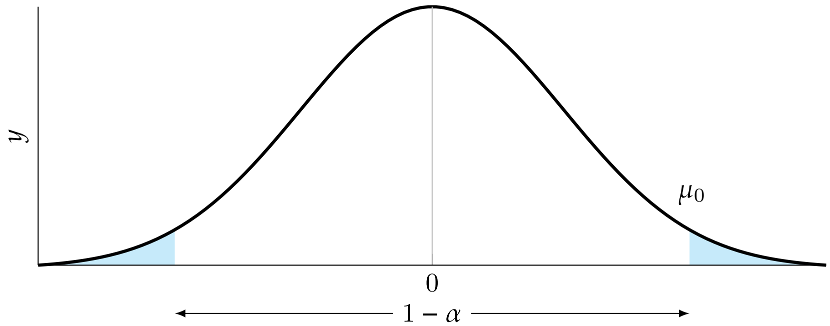
$$\pi = \begin{cases} \alpha & \text{if } H_0 \text{ is true} \\ 1-\beta & \text{if } H_1 \text{ is true} \end{cases}$$

3.2 Comparison between a mean and a reference value

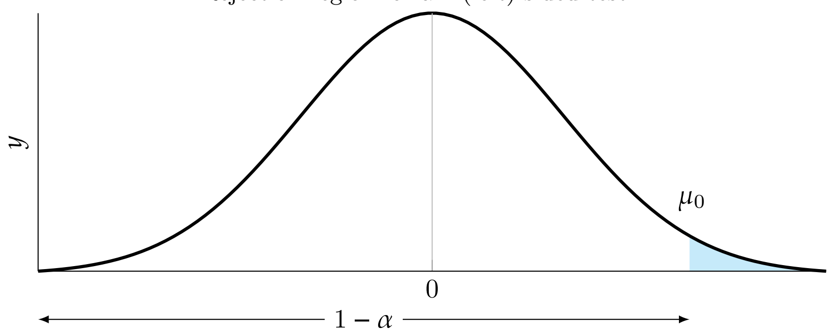
3.2.1 Two sided test

$$H_0 : \mu_0 = \mu$$
$$H_1 : \mu_0 \neq \mu$$

	H_0 is true	H_0 is false
Accept H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision



Rejection region for a 1 (left) sided test



Rejection region for a 2 sided test

- Assume σ is known. Test statistic $T = \frac{\bar{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$.

  Rule of rejection

$$|T| > t \quad \begin{cases} \Pr(|T| > t) = \alpha \\ T \sim \mathcal{N}(0,1) \end{cases} .$$

- Assume σ is unknown. Test statistic $T = \frac{\bar{x}_n - \mu_0}{\frac{S'}{\sqrt{n}}} = \frac{\bar{x}_n - \mu_0}{\frac{S}{\sqrt{n-1}}}$.

  Rule of rejection

$$|T| > t \quad \begin{cases} \Pr(|T| > t) = \alpha \\ T \sim \mathcal{T}_{n-1} \end{cases} .$$

3.2.2 One sided test

Left sided test

$$H_0 : \mu_0 = \mu$$
$$H_1 : \mu_0 < \mu$$

- Assume σ is known. Test statistic $T = \frac{\bar{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$.

  Rule of rejection

$$T > t \quad \begin{cases} \Pr(T > t) = \alpha \\ T \sim \mathcal{N}(0,1) \end{cases} .$$

- Assume σ is unknown. Test statistic $T = \frac{\bar{x}_n - \mu_0}{\frac{S'}{\sqrt{n}}} = \frac{\bar{x}_n - \mu_0}{\frac{S}{\sqrt{n-1}}}$.

Rule of rejection

$$T > t \quad \begin{cases} \Pr(T > t) = \alpha \\ T \sim \mathcal{T}_{n-1} \end{cases} .$$

Right sided test

Same as left sided test but with $T < -t$.

3.3 Comparison between a proportion and a reference value

3.3.1 Two sided test

$$\begin{aligned} H_0 &: p_0 = p \\ H_1 &: p_0 \neq p \end{aligned}$$

Test statistic $X \sim \text{Bin}(n, p)$.

Rule of rejection

$$\begin{cases} \Pr\left(X > b_{n,p_0,1-\alpha/2}\right) = \Pr\left(X > b_{n,p_0,\alpha/2}\right) = \frac{\alpha}{2} \\ X \sim \text{Bin}(n, p_0) \end{cases} .$$

Acceptance region $[b_{n,p_0,\alpha/2}, b_{n,p_0,1-\alpha/2}]$.

$$\Pr\left(X \in [b_{n,p_0,\alpha/2}, b_{n,p_0,1-\alpha/2}]\right) = 1 - \alpha.$$

However if $n \geqslant 30$ we can use the normal approximation.
Test statistic $T = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$.

Rule of rejection

$$|T| > t \quad \begin{cases} \Pr(|T| > t) = \alpha \\ T \sim \mathcal{N}(0, 1) \end{cases} .$$

3.3.2 One sided test

$$\begin{aligned} H_0 &: p_0 = p \\ H_1 &: p_0 < p \end{aligned}$$

Test statistic $X \sim \text{Bin}(n, p)$.

Rule of rejection

$$\Pr\left(X \geqslant b_{n,p_0,1-\alpha}\right) = \alpha.$$

Acceptance region $[-\infty, b_{n,p_0,1-\alpha}]$.

Normal approximation Test statistic $T = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$.

Rule of rejection

$$T > t \quad \begin{cases} \Pr(T > t) = \alpha \\ T \sim \mathcal{N}(0, 1) \end{cases} .$$

3.4 Comparison between a variance and a reference value

3.4.1 Two sided test

$$H_0 : \sigma_0^2 = \sigma^2$$
$$H_1 : \sigma_0^2 \neq \sigma^2$$

Test statistic

$$T = \begin{cases} n\frac{Z^2}{\sigma_0^2} \sim {\chi_n}^2 & \text{if } \mu \text{ is known} \\ n\frac{S^2}{\sigma_0^2} \sim {\chi_{n-1}}^2 & \text{if } \mu \text{ is unknown} \end{cases}.$$

We reject H_0 if $T \notin [v_{\alpha/2}, v_{1-\alpha/2}]$.

$$\Pr\left(T \notin [v_{\alpha/2}, v_{1-\alpha/2}]\right) = \frac{\alpha}{2}.$$

3.4.2 One sided test

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 > \sigma_0^2$$

Test statistic

$$T = \begin{cases} n\frac{Z^2}{\sigma_0^2} \sim {\chi_n}^2 & \text{if } \mu \text{ is known} \\ n\frac{S^2}{\sigma_0^2} \sim {\chi_{n-1}}^2 & \text{if } \mu \text{ is unknown} \end{cases}.$$

We reject H_0 if $T > v_{1-\alpha}$.

$$\Pr\left(T > v_{1-\alpha}\right) = \alpha.$$

3.5 Critical Probability

$$\mathrm{P_c} = \begin{cases} \Pr\left(|T| \geqslant |t_0| / \theta = \theta_0\right) & \text{for one sided tests} \\ \Pr\left(T \geqslant t_0 / \theta = \theta_0\right) & \text{for } H_1 : \theta > \theta_0 \\ \Pr\left(T \leqslant t_0 / \theta = \theta_0\right) & \text{for } H_1 : \theta < \theta_0 \end{cases}.$$

3.6 Comparison between two means

3.6.1 Two sided test

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Test statistic $T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1^2} + \frac{\sigma_2^2}{n_2^2}}}$.

Rule of rejection

$$\begin{cases} \Pr\left(|T| > t\right) = \alpha \\ T \sim \mathcal{N}(0,1) \end{cases}.$$

Note:-

$$\bar{X}_1 - \bar{X}_2 \sim \mathcal{N}\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$\mathbb{E}(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$$

$$\text{Var}\left(\bar{X}_1 - \bar{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

3.6.2 One sided test

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

Test statistic same as before.

Rule of rejection

$$\begin{cases} \Pr(T > t) = \alpha \\ T \sim \mathcal{N}(0,1) \end{cases}.$$

Note:-

If σ_1 and σ_2 are unknown, and $\sigma_1 = \sigma_2$ we perform all the same preceding tests but this a new test statistic

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathcal{T}_{n_1+n_2-2}.$$

if $\sigma_1 \neq \sigma_2$ we use the preceding tests only if n_1 and n_2 are sufficiently large (> 30).

3.7 Comparison between 2 proportions

Consider the following 2 random variables

$$P = \frac{X+Y}{n_1+n_2} \quad S_d^2 = P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

3.7.1 Two sided test

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2$$

Test statistic $T = \frac{\frac{X}{n_1} - \frac{Y}{n_2}}{S_d}$.

Rule of rejection

$$\begin{cases} \Pr(|T| > t) = \alpha \\ T \sim \mathcal{N}(0,1) \end{cases}.$$

3.7.2 One sided test

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 > p_2$$

Same job as before.

3.8 Comparison between two variances

3.8.1 Two sided test

$$H_0 : {\sigma_1}^2 = {\sigma_2}^2$$
$$H_1 : {\sigma_1}^2 \neq {\sigma_2}^2$$

Test statistic $F = \frac{S_1'^2}{S_2'^2}$.

$$F \sim \mathcal{F}_{n_1-1,n_2-1}.$$

where $\mathcal{F}$ is the Fisher distribution.

Rule of rejection

$$\begin{cases} \Pr\left(F < f_{n_1-1,n_2-1,1-\alpha/2}\right) = \Pr\left(F > f_{n_1-1,n_2-1,\alpha/2}\right) = \frac{\alpha}{2} \\ F \sim \mathcal{F}_{n_1-1,n_2-1} \end{cases}.$$

3.8.2 One sided test

$$H_0 : {\sigma_1}^2 = {\sigma_2}^2$$
$$H_1 : {\sigma_1}^2 < {\sigma_2}^2$$

Rule of rejection

$$\begin{cases} \Pr\left(F > f_{n_1-1,n_2-1,1-\alpha}\right) = \frac{\alpha}{2} \\ F \sim \mathcal{F}_{n_1-1,n_2-1} \end{cases}.$$

Note:-

Chi-squared test

$$H_0: \text{ the population follows law M}$$
$$H_1: \text{ the population does not follow law M}$$

Test statistic

$$Y = \sum_{i=1}^{k} \frac{(n_i - n_{p_i})^2}{n_{p_i}} \sim {\chi^2}_{k-1}.$$

Rule of rejection

$$\begin{cases} \Pr\left(Y > t\right) = \alpha \\ Y \sim {\chi^2}_{k-1} \end{cases}.$$

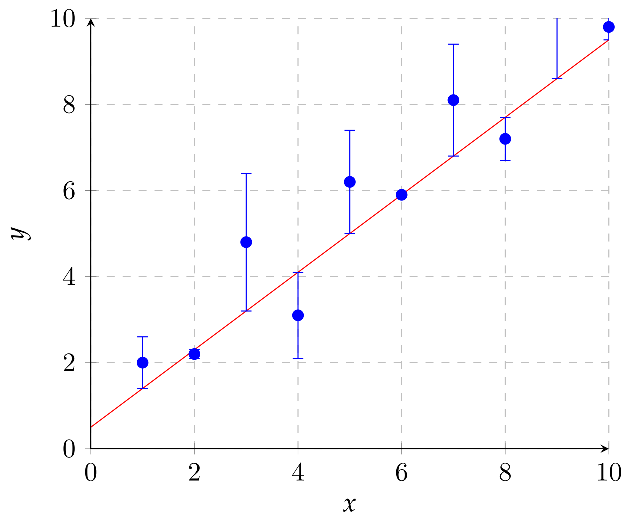
Chapter 4

Regression

$$\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

4.1 Linear Regression

The objective of linear regression is to find the best linear approximation $D : \alpha x + \beta$ of a set of data points.



The least squares method consists in minimizing the sum $S(\alpha, \beta)$ of the squares of the residuals.

$$S(\alpha,\beta) = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}(y_j - \alpha x_i - \beta)^2.$$

It can shown that the couple (a, b) that minimizes S is given by

$$a = \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(X)}$$
$$b = \mathbb{E}(Y) - a\mathbb{E}(X)$$

$$S_{\min} = \mathrm{Var}(Y)\cdot\left[1 - \frac{\mathrm{cov}(X,Y)^2}{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}\right].$$

Similarly, we can construct a regression line $D' : a'y + b'$ of y on x where a' and b' are given by

$$a' = \frac{\operatorname{cov}(X,Y)}{\operatorname{Var}(Y)}$$
$$b' = \mathbb{E}(X) - a'\mathbb{E}(Y)$$

4.2 Coefficient of linear correlation

$$\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}.$$

Some properties of $\rho(X,Y)$

$$\rho(X,Y)^2 \leqslant 1$$
$$\rho(X,Y)^2 = a \cdot a'$$
$$\rho(X,Y) = \begin{cases} \sqrt{a \cdot a'} & \text{if } a, a' > 0\,(\operatorname{cov}(X,Y) > 0) \\ -\sqrt{a \cdot a'} & \text{if } a, a' < 0\,(\operatorname{cov}(X,Y) < 0) \\ 0 & \text{if } a = a' = 0\,(\operatorname{cov}(X,Y) = 0) \end{cases}$$
$$\rho(X,Y) \begin{cases} > 0 & \text{if } X \text{ and } Y \text{ are positively correlated} \\ < 0 & \text{if } X \text{ and } Y \text{ are negatively correlated} \\ = 0 & \text{if } X \text{ and } Y \text{ are not correlated} \end{cases}$$
$$\rho(X,Y)^2 = 1 \text{ if and only if } X \text{ and } Y \text{ are perfectly correlated}$$

4.3 Residual and non-residual variance

Consider a double statistical distribution (n_{ij}, x_i, y_j) and the regression line $D : y = ax + b$ where

$$a = \frac{\operatorname{cov}(X,Y)}{\operatorname{Var}(X)}$$
$$b = \mathbb{E}(Y) - a\mathbb{E}(X)$$

The *residual variance* is defined as

$$V_R = S_{\min} = \operatorname{Var}(Y) \cdot \left[1 - \rho(X,Y)^2\right].$$

The *non-residual variance* is defined as

$$V_E = \frac{1}{n}\sum_{ij} n_{ij}(\mathbb{E}(Y) - ax_i - b)^2.$$

$$V_E = a^2\operatorname{Var}(X) = \rho^2 \cdot \operatorname{Var}(Y)$$
$$\operatorname{Var}(Y) = V_R + V_E$$