UM–SJTU JOINT INSTITUTE

VE401 / ECE4010J

PROBABILISTIC METHODS IN ENGINEERING

Lecture Notes

Li Junhao

April 20, 2023

Contents

1 Basic Probability Theories

1.1 Sample Space

Suppose that a non-empty set S is given. A σ-field $\mathcal{F}$ on S is a family of subsets of S such that
(i) $\emptyset \in \mathcal{F}$;
(ii) if $A \in \mathcal{F}$, then $S \backslash A \in \mathcal{F}$;
(iii) if $A_1, A_2, A_3, \ldots \in \mathcal{F}$ is a finite or countable sequence of subsets, then the union $\bigcup_k A_k \in \mathcal{F}$.

1.2 Independence

Random Variables X and Y are independent if

$$P[X \cap Y] = P[X]P[Y] \quad \text{or} \quad f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$$

This is equivalent to

$$P[X \mid Y] = P[X] \quad \text{if } P[Y] \neq 0$$
$$P[Y \mid X] = P[Y] \quad \text{if } P[X] \neq 0$$

Do notice that the condition $P[Y] \neq 0$ (or $P[X] \neq 0$) is important.

1.3 Expectation

$$E[X + Y] = E[X] + E[Y] \quad E[cX] = cE[X]$$

This is always true no matter whether X and Y are independent or not.
Theorem. Suppose $g(x)$ is a linear function, then

$$E[g(X)] = g(E[X])$$

If $g(x)$ is not linear, this statement is usually false.

1.4 Moment

Definition. Given a random variable X, the quantities

$$\mathrm{E}\left[X^n\right], \quad n \in \mathbb{N},$$

are known as the $\boldsymbol{n}^{\text{th}}$ (ordinary) moments of X. The quantities

$$\mathrm{E}\left[\left(\frac{X - \mu}{\sigma}\right)^n\right], \quad n = 3, 4, 5, \ldots,$$

are called the $\boldsymbol{n}^{\text{th}}$ central moments of X.

2 Discrete Random Variable Distribution

2.1 Moment-Generating Function

Definition. Let (X, f_X) be a random variable and such that the sequence of moments $\mathrm{E}\left[X^n\right], n \in \mathbb{N}$, exists. If the power series

$$m_X(t) := \mathrm{E}\left[e^{tX}\right] = \sum_{k=0}^{\infty} \frac{\mathrm{E}\left[X^k\right]}{k!} t^k \tag{1}$$

has radius of convergence $\varepsilon > 0$, the thereby defined function

$$m_X : (-\varepsilon, \varepsilon) \to \mathbb{R}$$

is called the moment-generating function for X.
And we have

$$E\left[X^k\right] = \left.\frac{d^k m_X(t)}{dt^k}\right|_{t=0} \tag{2}$$

Theorem The moment-generating function of a linear function of a ramdom variable is given by

$$M_Y(s) = E\left[e^{s(aX+b)}\right] = e^{sb} E\left[e^{saX}\right] = e^{sb} M_X(sa) \tag{3}$$

2.2 Bernoulli Distribution

$$X \sim \text{ Bernoulli } (p), \quad f_X(x) = \begin{cases} 1-p & \text{for } x = 0 \\ p & \text{for } x = 1 \end{cases}$$

With $E[X] = p$ and $\mathrm{Var}[X] = p(1-p)$.

2.3 Binomial Distribution

$$X \sim \text{ B } (p), \quad f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

With $E[X] = np$ and $\mathrm{Var}[X] = np(1-p)$.
The moment-generating function is

$$M(s) = \sum_{x=0}^{\infty} e^{sx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{\infty} \binom{n}{x} \left(e^s p\right)^x (1-p)^{n-x} = (1 - p + e^s p)^x$$

2.4 Geometric Distribution

Geometric random variable X represents the total number of trails at first success.

$$X \sim \text{ Geom } (p), \quad f_X(x) = (1-p)^{x-1} p$$

With $E[X] = \dfrac{1}{p}$ and $\mathrm{Var}[X] = \dfrac{1-p}{p^2}$
The moment-generating function is

$$M(s) = \sum_{x=1}^{\infty} e^{sx} p (1-p)^{x-1} = p e^s \sum_{x=0}^{\infty} \left(e^s (1-p)\right)^x = \frac{p e^s}{1 - e^s (1-p)}$$

2.5 Pascal Distribution

The Pascal random variable X means that the r^{th} success is obtained in the X^{th} trial.

$$f_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \tag{4}$$

With $E[X] = \dfrac{r}{p}$ and $\mathrm{Var}[X] = \dfrac{r(1-p)}{p^2}$
The moment-generating function is

$$m_X : (-\infty, -\ln q) \to \mathbb{R}, \quad m_X(t) = \frac{(pe^t)^r}{(1-qe^t)^r}, \quad q = 1-p$$

2.6 Negative Binomial

Definition. Negative Binomial is defined as

$$\binom{-r}{x} = (-1)^x \binom{r-1+x}{r-1} (r > 0) \tag{5}$$

3 Poisson Process

3.1 Precise Postulates of Rate of Arrivals

(i) The probability that exactly one arrival will occur in an interval of width Δt is

$$\lambda \cdot \Delta t + o(\Delta t).$$

(ii) The probability that exactly zero arrivals will occur in the interval is

$$1 - \lambda \cdot \Delta t + o(\Delta t).$$

(iii) The probability that two or more arrivals occur in the interval is

$$o(\Delta t).$$

3.2 Poisson Distribution

During any time interval t, the probability to have x arrivals is:

$$p(x,t) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \tag{6}$$

Definition. Let $k \in \mathbb{R}$. A random variable (X, f_X) with

$$X : S \to \mathbb{N}$$

and density function $f_X : \mathbb{N} \to \mathbb{R}$ given by

$$f_X(x) = \frac{k^x e^{-k}}{x!} \tag{7}$$

is said to follow a Poisson distribution with parameter k.
The expectation, variance and moment-generating function of Poisson distribution is

$$E[X] = k = \lambda t \quad \mathrm{Var}[X] = k = \lambda t \quad m_X(t) = e^{k\left(e^t - 1\right)}$$

3.3 The Time for the First Arrival

The time for the first arrival follows exponential distribution, i.e.

$$f_T(t) = \lambda e^{-\lambda t}, \quad E[T] = \frac{1}{\lambda}, \quad \mathrm{Var}[T] = \frac{1}{\lambda^2}$$

3.4 The Time for the k-th Arrival

The time for the k-th arrival follows

$$f_{T_k}(t) = \frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)!}, \quad E[T] = \frac{k}{\lambda}, \quad \mathrm{Var}[T] = \frac{k}{\lambda^2}$$

The time for the first arrival follows exponential distribution, i.e.

$$f_T(t) = \lambda e^{-\lambda t}, \quad E[T] = \frac{1}{\lambda}, \quad \mathrm{Var}[T] = \frac{1}{\lambda^2}$$

3.5 Approximate Binomial Distribution

In binomial distribution, if n is very large and p approaches 0, then we can use Poisson distribution to approximate binomial distribution, i.e., we take $k := np$, and yield

$$\binom{n}{m} p^m (1-p)^{n-m} \approxeq \frac{k^m}{m!} e^{-k} \tag{8}$$

4 Continuous Random Variables

4.1 Exponential Distribution

Definition. A continuous random variable (X, f_β) follows exponential distribution with parameter β if the probability density function is defined by

$$f_\beta(x) = \begin{cases} \beta e^{-\beta x}, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{9}$$

Expectation: $\mathrm{E}[X] = \frac{1}{\beta}$ Variance: $\mathrm{Var}[X] = \frac{1}{\beta^2}$ Moment-generating function:

$$M_X : (-\infty, \beta) \to \mathbb{R}, \quad M_X(t) = \frac{1}{1 - t/\beta}$$

4.2 Gamma Distribution

Definition. Let $\alpha, \beta \in \mathbb{R}, \alpha, \beta > 0$. A continuous random variable $(X, f_{\alpha,\beta})$ follows a gamma distribution with parameters α and β if the probability density function is given by

$$f_{\alpha,\beta}(x) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{10}$$

where $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z}\, \mathrm{d}z, \alpha > 0$ is the Euler gamma function. Expectation: $\mathrm{E}[X] = \frac{\alpha}{\beta}$. Variance: $\mathrm{Var}[X] = \frac{\alpha}{\beta^2}$ Moment-generating function:

$$M_X : (-\infty, \beta) \to \mathbb{R}, \quad M_X(t) = \frac{1}{(1 - t/\beta)^\alpha}$$

4.3 Relationship between Poisson Distribution, Exponential Distribution and Gamma Distribution

(1) The time needed for the next r arrivals in a Poisson process with rate λ is determined by a Gamma distribution with parameters $\alpha = r$ and $\beta = \lambda$.
(2) Sum of exponential distributed random variables with the same β follows Gamma distribution.
(3) Sum of gamma distributed random variables with the same β follows Gamma distribution (The new α will be the sum of α' s) (check the M.G.F)

4.4 Normal Distribution

Let $\mu \in \mathbb{R}, \sigma > 0$. A continuous random variable (X, f_X) with density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-((x-\mu)/\sigma)^2/2} \tag{11}$$

is said to follow a normal distribution with parameters μ and σ.
Moment-generating function:

$$M_X : \mathbb{R} \to \mathbb{R}, \quad M_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

4.4.1 Independent Normal Distribution

Theorem. Suppose $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
Corollary. Let $X_1, \ldots, X_n$ be a sample of size n from the distribution of a random variable X that follows a normal distribution with mean μ and variance σ^2. Then $\bar{X} \sim N(\mu, \frac{\sigma^2}{N})$.

4.4.2 Approximate the Binomial Distribution

$$P[X \le y] = \sum_{x=0}^{y} \binom{n}{x} p^x (1-p)^{n-x} \approx \Phi(\frac{y + \frac{1}{2} - np}{\sqrt{np(1-p)}}) \tag{12}$$

$$P[k \le X \le l] = \approx \Phi(\frac{l + \frac{1}{2} - np}{\sqrt{np(1-p)}}) - \Phi(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}) \tag{13}$$

Requirements:

$$np > 5 \text{ if } p \le \frac{1}{2} \quad \text{or} \quad n(1-p) > 5 \text{ if } p > \frac{1}{2}$$

Notice that the term 1/2 is called the **half-unit correction**.

4.4.3 Error Function

In Mathematica, the cumulative distribution function is expressed through the error function, defined as

$$\text{erf}(z) := \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt, \quad \text{erfc}(z) := 1 - \text{erf}(z)$$

4.4.4 Estimates on Variability

Theorem. Let X be normally distributed with parameters μ and σ. Then

$$\begin{aligned} P[-\sigma < X - \mu < \sigma] &= 0.68 \\ P[-2\sigma < X - \mu < 2\sigma] &= 0.95 \\ P[-3\sigma < X - \mu < 3\sigma] &= 0.997 \end{aligned}$$

Hence 68% of the values of a normal random variable lie within one standard deviation of the mean, 95% lie within two standard deviations, and 99.7% lie within three standard deviations.

4.5 The Chi Distribution

The Chi distribution has the following definition

$$\chi_n := \sqrt{\sum_{i=1}^{n} Z_i^2},$$

where Z_i follows the normal distribution. And its probability density function is

$$f_{\chi_n}(y) = F'_{\chi_n}(y) = \frac{2}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} y^{n-1} e^{-y^2/2} \tag{14}$$

Chi-squared distribution

$$\begin{aligned} f_{\chi_n^2}(y) = F'_{\chi_n^2}(y) &= \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{n/2-1}} \frac{d}{dy} \int_0^{\sqrt{y}} e^{-r^2/2} r^{n-1} dr \\ &= \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} y^{n/2-1} e^{-y/2} \end{aligned} \tag{15}$$

A chi-squared distribution with n degrees of freedom corresponds to the Gamma distribution with $\alpha = n/2$ and $\beta = 1/2$

$$\mathrm{E}\left[\chi_n^2\right] = n, \quad \mathrm{Var}\left[\chi_n^2\right] = 2n$$

Lemma. Let $\chi^2_{\gamma_1}, \ldots, \chi^2_{\gamma_n}$ be n independent random variables following chi-squared distributions with $\gamma_1, \ldots, \gamma_n$ degrees of freedom, respectively. Then

$$\chi^2_\alpha := \sum_{k=1}^{n} \chi^2_{\gamma_k} \tag{16}$$

is a chi-squared random variable with $\alpha = \sum_{k=1}^{n} \gamma_k$ degrees of freedom.

5 Multivariable Random Variables

5.1 Expectation

We define the expected value or expectation for X as the vector

$$\mathrm{E}[X] = \begin{pmatrix} \mathrm{E}\left[X_1\right] \\ \vdots \\ \mathrm{E}\left[X_n\right] \end{pmatrix}$$

And we have

$$\mathrm{E}\left[X_k\right] = \sum_{x_k} x_k f_{X_k}\left(x_k\right) = \sum_{x\in\Omega} x_k f_X(x) \tag{17}$$

Or

$$\mathrm{E}\left[X_k\right] = \int_{\mathbb{R}} x_k f_{X_k}\left(x_k\right) dx_k = \int_{\mathbb{R}^n} x_k f_X(x) dx \tag{18}$$

Theorem. Suppose $\varphi : \mathbb{R}^n \to \mathbb{R}$ is a continuous function. Then

$$\varphi \circ X : S \to \mathbb{R}$$

defines a scalar random variable. It is possible to prove that in this case,

$$\mathrm{E}[\varphi \circ X] = \sum_{x\in\Omega} \varphi(x) f_X(x), \quad \text{or} \quad \mathrm{E}[\varphi \circ X] = \int_{\mathbb{R}^n} \varphi(x) f_X(x) dx$$

For $\varphi\left(x_1, \ldots, x_n\right) = x_k$ we regain the definition of $E\left[X_k\right]$.

5.2 Correlation between Random Variables

5.2.1 Covariance

$$\mathrm{Cov}[X, Y] = \mathrm{E}\left[\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right] \tag{19}$$

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] \tag{20}$$

When using this formula, we must be careful about if X and Y are independent.
Theorem. If X and Y are independent, then $\mathrm{Cov}[X, Y] = 0$. But if $\mathrm{Cov}[X, Y] = 0$, X and Y may still be dependent.

The covariance matrix of $\boldsymbol{X}$ is given by

$$\mathrm{Var}[\mathbf{X}] = \begin{pmatrix} \mathrm{Var}\left[X_1\right] & \mathrm{Cov}\left[X_1, X_2\right] & \cdots & \mathrm{Cov}\left[X_1, X_n\right] \\ \mathrm{Cov}\left[X_1, X_2\right] & \mathrm{Var}\left[X_2\right] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{Cov}\left[X_{n-1}, X_n\right] \\ \mathrm{Cov}\left[X_1, X_n\right] & \cdots & \mathrm{Cov}\left[X_{n-1}, X_n\right] & \mathrm{Var}\left[X_n\right] \end{pmatrix} \tag{21}$$

Theorem. Suppose there's a linear transformation matrix $C \in \mathrm{Mat}(n \times n; \mathrm{R})$

$$\mathrm{Var}[C\boldsymbol{X}] = C\,\mathrm{Var}[\boldsymbol{X}]C^T \tag{22}$$

5.2.2 Correlation Coefficient

$$\rho(X, Y) = \mathrm{Cov}(\tilde{X}, \tilde{Y}) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{E[XY] - E[X]E[Y]}{\sqrt{(E[X^2] - E^2[X])(E[Y^2] - E^2[Y])}} \tag{23}$$

It can be shown that ρ_{XY} has the following properties (i) $-1 \leq \rho_{XY} \leq 1$ (ii) $|\rho_{XY}| = 1$ if and only if there exist numbers $\beta_0, \beta_1 \in \mathbb{R}, \beta_1 \neq 0$, such that

$$Y = \beta_0 + \beta_1 X$$

almost surely.
Theorem X and Y are deterministically linearly related if and only if

$$\tilde{X} + \tilde{Y} = 0 \quad \text{or} \quad \tilde{X} - \tilde{Y} = 0 \tag{24}$$

And we have

$$\mathrm{Var}[\tilde{X} + \tilde{Y}] = \mathrm{Var}[\tilde{X}] + \mathrm{Var}[\tilde{Y}] + 2\,\mathrm{Cov}[\tilde{X}, \tilde{Y}] = 2 + 2\varrho_{XY}$$

$$\mathrm{Var}[\tilde{X} - \tilde{Y}] = \mathrm{Var}[\tilde{X}] + \mathrm{Var}[\tilde{Y}] - 2\,\mathrm{Cov}[\tilde{X}, \tilde{Y}] = 2 - 2\varrho_{XY}$$

5.2.3 Fisher Transformation

$$\ln\left(\sqrt{\frac{\mathrm{Var}[\tilde{X}+\tilde{Y}]}{\mathrm{Var}[\tilde{X}-\tilde{Y}]}}\right) = \frac{1}{2}\ln\left(\frac{1+\rho_{XY}}{1-\rho_{XY}}\right) = \mathrm{Artanh}\left(\rho_{XY}\right) \in \mathbb{R} \tag{25}$$

or

$$\rho_{XY} = \tanh\left(\ln\left(\frac{\sigma_{\tilde{X}+\tilde{Y}}}{\sigma_{\tilde{X}-\tilde{Y}}}\right)\right) \tag{26}$$

If follow that if $\rho_{XY} > 0$, then X and Y are positively correlated. If $\rho_{XY} < 0$, then X and Y are negatively correlated.

5.3 The Bivariate Normal Distribution

Theorem. Let A be an invertible 2×2 matrix and define $\boldsymbol{Y} = A\boldsymbol{X}$. Then the joint density of $\boldsymbol{Y}$ is given by

$$f_{\boldsymbol{Y}}(y) = \frac{1}{2\pi\sqrt{|\det \Sigma_{\boldsymbol{Y}}|}} e^{-\frac{1}{2}\left\langle y-\mu_{\boldsymbol{Y}}, \Sigma_{\boldsymbol{Y}}^{-1}(y-\mu_{\boldsymbol{Y}})\right\rangle}$$

where $\mu_{\boldsymbol{Y}} = \mathrm{E}[\boldsymbol{Y}], \Sigma_{\boldsymbol{Y}} = \mathrm{Var}[\boldsymbol{Y}]$ and $\langle\cdot,\cdot\rangle$ denotes the euclidean scalar product in $\mathbb{R}^2$.
Corollary. It can also be expressed as

$$f_{\boldsymbol{Y}}\left(y_1, y_2\right) = \frac{1}{2\pi\sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\varrho^2}} e^{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)^2 - 2\varrho\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)\left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right) + \left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right)^2\right]}$$

where μ_{Y_i} is the mean and $\sigma^2_{Y_i}$ the variance of $Y_i, i = 1, 2$, and ϱ is the correlation of Y_1 and Y_2.

5.4 The Hypergeometric Distribution

Definition. Let $N, n, r \in \mathbb{N}\backslash\{0\}, r, n \leq N$, and $n < \min\{r, N-r\}$. A random variable (X, f_X) with

$$X : S \to \Omega = \{0, \ldots, n\}$$

and density function $f_X : \Omega \to \mathbb{R}$ given by

$$f_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \tag{27}$$

is said to have a hypergeometric distribution with parameters N, n and r.
That means, for a box containing N balls, among which are r red balls. The hypergeometirc distribution represent the probability of "exactly x red balls out of n selected".
The hypergeometric distribution takes its name from the hypergeometric identity:

$$\binom{a+b}{r} = \sum_{k=0}^{r}\binom{a}{k}\binom{b}{r-k} = \sum_{i+j=r}\binom{a}{i}\binom{b}{j} \tag{28}$$

Actually, this process can be segregated into several identity Bernoulli trails with $p = \dfrac{r}{N}$. So that

$$E[X] = \frac{nr}{N} \quad \mathrm{Var}[X] = n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}$$

Compare it with the variance of binomial distribution, the expression above differs by $\frac{N-n}{N-1}$, so when $n << N$, we can use binomial distribution to approximate the hypergeometirc distribution.

Example. A production lot of 200 units has 8 defectives. A random sample of 10 units is selected, and we want to find the probability that the random sample will contain exactly one defective. We note that the sampling fraction is $n/N = 10/200 = 0.05$, so we can use the binomial approximation. Then $p = r/N = 8/200 = 0.04$ and

$$P[X = 1] \approx \binom{10}{1}(0.04)^1(0.96)^9 = 0.277$$

6 Transformation of Random Variables

6.1 Genetic Method for Transformation of Random Variables

6.2 Monotonic and Differentiable Function

Theorem. Let X be a continuous random variable with density f_X. Let $Y = \varphi \circ X$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is strictly monotonic and differentiable. The density for Y is then given by

$$f_Y(y) = f_X\left(\varphi^{-1}(y)\right) \cdot \left|\frac{d\varphi^{-1}(y)}{dy}\right| \quad \text{for } y \in \operatorname{ran}\varphi$$

and

$$f_Y(y) = 0 \quad \text{for } y \notin \operatorname{ran}\varphi$$

6.3 Transformation of MultiVariable Random Variables

Theorem. Let $(\boldsymbol{X}, f_{\boldsymbol{X}})$ be a continuous multivariate random variable and let $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ be a differentiable, bijective map with inverse φ^{-1}. Then $\boldsymbol{Y} = \varphi \circ \boldsymbol{X}$ is a continuous multivariate random variable with density

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}} \circ \varphi^{-1}(\boldsymbol{y}) \cdot \left|\det D\varphi^{-1}(\boldsymbol{y})\right|,$$

where $D\varphi^{-1}$ is the Jacobian of φ^{-1}.

Example. Let $((X, Y), f_{XY})$ be a continuous bivariate random variable. Let $U = X/Y$. Then the density f_U of U is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv, v) \cdot |v| dv.$$

Proof. Consider the transformation $\varphi : (X, Y) \mapsto (U, V)$ where

$$\varphi(x, y) = \begin{pmatrix} x/y \\ y \end{pmatrix}.$$

Then

$$\varphi^{-1}(u, v) = \begin{pmatrix} uv \\ v \end{pmatrix}.$$

We calculate

$$D\varphi^{-1}(u, v) = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \begin{pmatrix} v & u \\ 0 & 1 \end{pmatrix}$$

so

$$\left|\det D\varphi^{-1}(u,v)\right| = |v|$$

Then

$$f_{UV}(u,v) = f_{XY}(uv,v)|v|$$

The marginal density f_U is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{UV}(u,v)dv = \int_{-\infty}^{\infty} f_{XY}(uv,v)\cdot|v|dv$$

6.4 Convolution Method

The convolution of two functions f and g is defined by

$$(f*g)(y) := \int_{-\infty}^{\infty} f(y-x)g(x)dx. \tag{29}$$

Let (X, f_X) and (Y, f_Y) be independent, continuous random variables. Then their sum $Z = X + Y$ has density $f_Z = f_X * f_Y$.

7 Reliability

7.1 Failure Density, Reliability Function and Hazard Rate

Failure Density represents the probability density of "the system failing at time t".

$$\begin{aligned} f_A(t) &= \lim_{\Delta t\to 0} \frac{P[t \le T \le t+\Delta t]}{\Delta t} \\ &= \lim_{\Delta t\to 0} \frac{F_A(t+\Delta t) - F_A(t)}{\Delta t} \end{aligned} \tag{30}$$

Reliability Function represents the probability of "the system surviving at time t".

$$R_A(t) = 1 - \int_0^t f_A(s)ds = 1 - F_A(t). \tag{31}$$

Harzard Rate represents the probability of that "the system failing at time t" under the condition of "the system surviving before t".

$$\varrho_A(t) := \lim_{\Delta t\to 0} \frac{P[t \le T \le t+\Delta t \mid t \le T]}{\Delta t} = \frac{f_A(t)}{R_A(t)} \tag{32}$$

Theorem. Let X be a random variable with failure density f, reliability function R and hazard rate ϱ. Then

$$R(t) = e^{-\int_0^t \varrho(x)dx} \tag{33}$$

7.2 The Weibull Density and Weibull Random Variable

One hazard function in widespread use is the function

$$\varrho(t) = \alpha\beta t^{\beta-1}, \quad t > 0, \quad \alpha,\beta > 0$$

The reliability function is then given by

$$R(t) = e^{-\int_0^t \alpha\beta x^{\beta-1}dx} = e^{-\alpha t^\beta} \tag{34}$$

And the failure density is given by

$$f(t) = \varrho(t)R(t) = \alpha\beta t^{\beta-1}e^{-\alpha t^\beta} \tag{35}$$

Definition. A random variable (X, f_X) is said to have a Weibull distribution with parameters α and β if its density is given by

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1}e^{-\alpha x^\beta}, & x > 0, \\ 0, & \text{otherwise}, \end{cases} \qquad \alpha, \beta > 0$$

Theorem. Let X be a Weibull random variable with parameters α and β. The mean and variance of X are given by

$$\mu = \alpha^{-1/\beta}\Gamma(1 + 1/\beta)$$

and

$$\sigma^2 = \alpha^{-2/\beta}\Gamma(1 + 2/\beta) - \mu^2$$

7.3 Series and Parallel Configuration

For systems with series configuration, the reliability function is

$$R_{\text{series}}(t) = P[\text{no component fails before } t] = \prod_{i=1}^{k} R_i(t) \tag{36}$$

For systems with parallel configuration, the reliability function is

$$\begin{aligned} R_{\text{parallel}}(t) &= 1 - P[\text{ all components fail before } t] \\ &= 1 - \prod_{i=1}^{k} (1 - R_i(t)) \end{aligned} \tag{37}$$

For a combinational system with the configuration ""$A \,||\, (B + C)$" we have

$$R_{general}(t) = 1 - (1 - R_A(t))(1 - R_{BC}(t)) = 1 - (1 - R_A(t))(1 - R_B(t)R_C(t))$$

8 Limit Theories for Probability

8.1 The Chebyshev Inequality

Let $c > 0$ be any real number. Then

$$\begin{aligned} \mathrm{E}\left[X^2\right] &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \geq \int_{|x|\geq c} x^2 f_X(x)dx \\ &\geq c^2 \int_{|x|\geq c} f_X(x)dx \\ &= c^2 \cdot P[|X| \geq c] \end{aligned}$$

More generally, for $k \in \mathbb{N}\backslash\{0\}$,

$$P[|X| \geq c] \leq \frac{\mathrm{E}\left[|X|^k\right]}{c^k} \tag{38}$$

8.2 Central Limit Theorem

Central Limit Theorem. Let (X_i) be a sequence of independent, but not necessarily identical random variables whose third moments exist and satisfy a certain technical condition. Let

$$Y_n = X_1 + \cdots + X_n$$

Then for any $z \in \mathbb{R}$,

$$P\left[\frac{Y_n - E[Y_n]}{\sqrt{\mathrm{Var}[Y_n]}} \leq z\right] \overset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx \tag{39}$$

Approximate to Normal Distribution. Suppose $S_n = X_1 + ... + X_n$, where X_i are a sequence of i.i.d. random vairables with mean value μ and variance σ^2. If n is sufficiently large, the probability $P[S_n \leq c]$ can be approximately calculated by regarding it as normal distribution:
(1) Calculate the mean value $n\mu$ and the variance $n\sigma^2$;
(2) Calculate the normalized value $z = \dfrac{c - n\mu}{\sqrt{n}\sigma}$;
(3) Calculate the approximate value $P(S_n \leq c) \approx \Phi(z)$.

8.3 The Weak Law of Large Numbers

Theorem. Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for any $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right] \overset{n\to\infty}{\longrightarrow} 0 \tag{40}$$

9 Sampling and Data Visualization

9.1 Sample Size

Sample size n should not be larget than 5% of the population size.
The large sample has not only yielded a result that is different from the true proportion (that is to be expected in statistics), it has also perturbed the distribution of the remaining population.

9.2 Quartiles

Suppose that our list of n data has been ordered from smallest to largest, so that

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$$

Then the median is given by

$$q_2 = \begin{cases} x_{(n+1)/2} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left(x_{n/2} + x_{n/2+1}\right) & \text{if } n \text{ is even} \end{cases}$$

9.3 Histograms and Category Width

The number of categories can be typically defined based on Sturges's rule, i.e.,

$$k = \lceil \log_2(n) \rceil + 1 \tag{41}$$

The precision of the data $\{x_1, \ldots, x_n\}$ is the smallest decimal place of the values x_i. The sample range is given by

$$\max_{1 \le i \le n} \{x_i\} - \min_{1 \le i \le n} \{x_i\}$$

If the number of bins k has been determined (e.g., by Sturges's rule), then the bin width is calculated as

$$h = \frac{\max\{x_i\} - \min\{x_i\}}{k} \tag{42}$$

According to Freedman and Diaconis, numerical calculations show that

$$h = \frac{2 \cdot \mathrm{IQR}}{\sqrt[3]{n}} \tag{43}$$

Where IQR is a measure of dispersion of the data, which is defined as

$$\mathrm{IQR} = q_3 - q_1 \tag{44}$$

9.3.1 Describe a Histogram

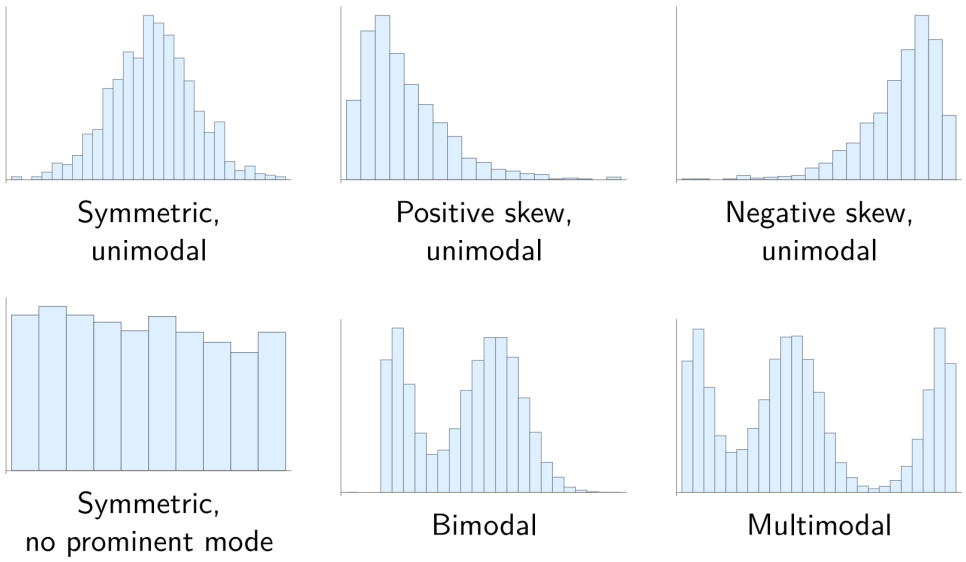


Figure 1: Different shapes of histograms.

Description:
This histogram has a unimodal shape (1/2 Mark) which is consistent with a normal distribution. It is not significantly skewed, (1/2 Mark) again consistent with a normal distribution. Therefore, there is no evidence that the data does not come from a normal distribution.

9.4 Boxplots

We define the **inner fences** f_1 and f_3 using the interquartile range as follows:

$$f_1 = q_1 - \frac{3}{2}\text{IQR}, \quad f_3 = q_3 + \frac{3}{2}\text{IQR} \tag{45}$$

The **whiskers** (lines extending to the left and right of the box) end at the adjacent values

$$a_1 = \min\{x_k : x_k \geq f_1\}, \quad a_3 = \max\{x_k : x_k \leq f_3\} \tag{46}$$

We define the **outer fences**

$$F_1 = q_1 - 3\text{IQR}, \quad F_3 = q_3 + 3\text{IQR}. \tag{47}$$

Measurements x_k that lie outside the inner fences but inside the outer fences are called **near outliers**. Those outside the outer fences are known as **far outliers**.

Boxplot of normal distribution has following traits:
(1) A symmetric median line in the middle of the box;
(2) Equally long whiskers;
(3) Very few near outliers and no far outliers.

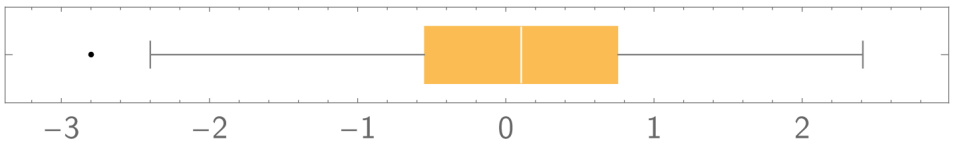


Figure 2: Boxplot of normal distribution.

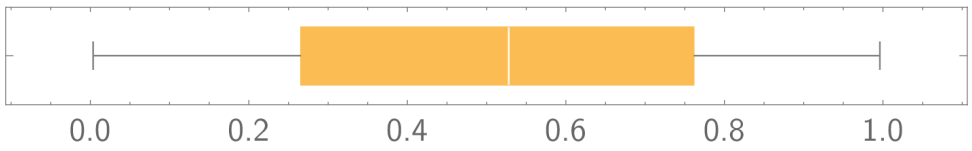


Figure 3: Boxplot of uniform distribution.

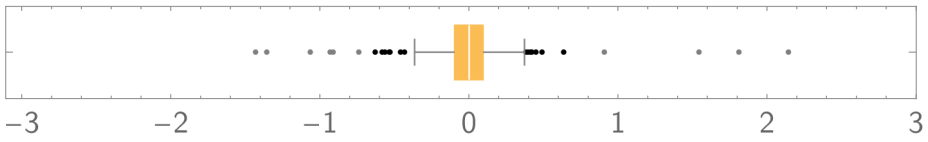


Figure 4: Boxplot of Cauchy distribution.

Description:
The whiskers are moderately asymmetric (1/2 Mark) but the median line is not too far from the center of the box, (1/2 Mark). There is no outlier, (1/2 Mark) and in summary no strong evidence that the data does not come from a normal distribution. (1/2 Mark)

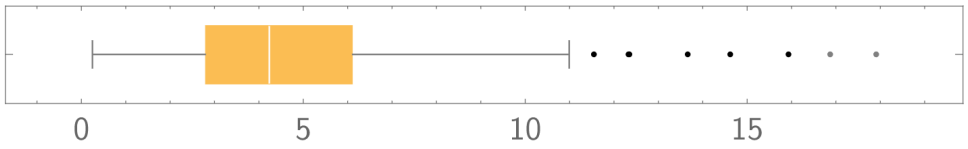


Figure 5: Boxplot of Gamma distribution.

10 Point Estimation

10.1 Sample Mean and Sample Variance

Definition. The difference

$$\theta - \mathrm{E}[\hat{\theta}]$$

is called the bias of an estimator $\hat{\theta}$ for a population parameter θ. If $\mathrm{E}[\widehat{\theta}] = \theta$, we say that $\widehat{\theta}$ is unbiased. The mean square error of $\hat{\theta}$ is defined as

$$\mathrm{MSE}[\widehat{\theta}] := \mathrm{E}\left[(\widehat{\theta} - \theta)^2\right]$$

And we have

$$\begin{aligned} \mathrm{MSE}[\widehat{\theta}] &= \mathrm{E}\left[(\widehat{\theta} - \mathrm{E}[\widehat{\theta}])^2\right] + (\theta - \mathrm{E}[\widehat{\theta}])^2 \\ &= \mathrm{Var}[\widehat{\theta}] + (\mathrm{bias})^2. \end{aligned} \tag{48}$$

Estimator of Mean Value. Let $X_1, \ldots, X_n$ be a random sample of size n from a distribution with mean μ. The sample mean $\bar{X}$ is an unbiased estimator for μ
Standard Deviation. The standard deviation of $\bar{X}$ is given by $\sqrt{\mathrm{Var}[\bar{X}]} = \sigma/\sqrt{n}$ and is called the standard error of the mean.
Unbiased Estimator of Variance.

$$S^2 := \frac{1}{n-1} \sum_{k=1}^{n} \left(X_k - \bar{X}\right)^2. \tag{49}$$

10.2 Estimator of Moments

Estimator of Moments. Given a random sample $X_1, \ldots, X_n$ of a random variable X, for any integer $k \geq 1$,

$$\widehat{\mathrm{E}\left[X^k\right]} = \frac{1}{n} \sum_{i=1}^{n} X_i^k \tag{50}$$

is an unbiased estimator for the k-th moment of X.
Usually, by calculating $E[X]$ we can get a function w.r.t θ, then we only need to solve the equation

$$f(\bar{\theta}) = \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

10.3 Method of Maximum Likelihood

Let X_θ be a random variable with parameter θ and density f_{X_θ}. Given a random sample $(X_1, \ldots, X_n)$ that yielded values $(x_1, \ldots, x_n)$ we define the likelihood function L by

$$L(\theta) = \prod_{i=1}^{n} f_{X_\theta}(x_i) \tag{51}$$

We then maximize $L(\theta)$. The location of the maximum is then chosen to be the estimator $\hat{\theta}$.

Example. Suppose it is known that X follows a Poisson distribution with parameter k and we wish to estimate k. The density for X is given by $f_k(x) = \dfrac{e^{-k}k^x}{x!}, x \in \mathbb{N}$. Given a random sample $X_1, \ldots, X_n$ the likelihood function is

$$L(k) = \prod_{i=1}^{n} f_k(x_i) = e^{-nk}\frac{k\sum x_i}{\prod x_i!}.$$

To simplify our calculations, we take the logarithm:

$$\ln L(k) = -nk + \ln k \sum_{i=1}^{n} x_i - \ln \prod x_i!$$

Maximizing $\ln L(k)$ will also maximize $L(k)$. We take the first derivative and set it equal to zero:

$$\frac{d \ln L(k)}{dk} = -n + \frac{1}{k}\sum_{i=1}^{n} x_i = 0$$

so we find

$$\widehat{k} = \bar{x}$$

10.4 Independance of Sample Mean and Sample Variance

Theorem. Let $X_1, \ldots, X_n, n \geq 2$, be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then
(i) The sample mean $\bar{X}$ is independent of the sample variance S^2,
(ii) $\bar{X}$ is normally distributed with mean μ and variance σ^2/n,
(iii) $(n-1)S^2/\sigma^2$ is chi-squared distributed with $n-1$ degrees of freedom.

The Helmert Transformation. A sample of size n taken from a normal population X with mean μ and variance σ^2 is transformed as follows:

$$\begin{aligned}
Y_1 &= \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n) \\
Y_2 &= \frac{1}{\sqrt{2}}(X_1 - X_2) \\
Y_3 &= \frac{1}{\sqrt{6}}(X_1 + X_2 - 2X_3) \\
&\vdots \\
Y_n &= \frac{1}{\sqrt{n(n-1)}}(X_1 + X_2 + \cdots + X_{n-1} - (n-1)X_n)
\end{aligned}$$

Then, $Y_1, Y_2, ..., Y_n$ are independent and normally distributed. The random variable Y_1 is normally distributed with mean $\sqrt{n}\mu$ and variance σ^2 and $Y_2, Y_3, ..., Y_n$ have mean 0 and variance σ^2. .

11 Confidence Interval

11.1 Interval Estimation for μ with Known σ^2

A 95% confidence interval does not mean that "the probability of the parameter θ locating in the interval is 95%. Instead, it means "if we samples and calculate a 95% confidence interval, then the probability that the interval contains θ is 95%". Do remember that L_1 and L_2 are functions based on our observation, which means they are random variables. So $[L_1, L_2]$ is a random interval.

Theorem. Let $X_1, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. A $100(1-\alpha)\%$ confidence interval on μ is given by

$$\bar{X} \pm \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \tag{52}$$

where $z_{\alpha/2}$ satisfies the equation

$$\alpha/2 = P\left[Z \geq z_{\alpha/2}\right] = \frac{1}{\sqrt{2\pi}} \int_{z_{\alpha/2}}^{\infty} e^{-x^2/2} dx \tag{53}$$

Theorem. Let $X_1, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2.
(i) A $100(1-\alpha)\%$ upper confidence bound on μ is given by $\bar{X} + \frac{z_\alpha \cdot \sigma}{\sqrt{n}}$.
(ii) A $100(1-\alpha)\%$ lower confidence bound on μ is given by $\bar{X} - \frac{z_\alpha \cdot \sigma}{\sqrt{n}}$.

Theorem. Let $X_1, \ldots, X_n, n \geq 2$, be i.i.d. random variables. Then if $\bar{X}$ and S^2 are independent, the $X_k, k = 1, \ldots, n$ follow a normal distribution.

11.2 Interval Estimation for σ^2 with Unknown μ

Theorem. Let $X_1, \ldots, X_n, n \geq 2$, be a random sample of size n from a normal distribution with mean μ and variance σ^2. A $100(1-\alpha)\%$ confidence interval on σ^2 is given by

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right] \tag{54}$$

Given $\alpha \in [0,1]$ and $\gamma > 0$ we define $\chi^2_{1-\alpha/2,\gamma}, \chi^2_{\alpha/2,\gamma} \in [0, \infty)$ by

$$\int_0^{\chi^2_{1-\alpha/2,\gamma}} f_{\chi^2_\gamma}(x)dx = \int_{\chi^2_{\alpha/2,\gamma}}^{\infty} f_{\chi^2_\gamma}(x)dx = \alpha/2 \tag{55}$$

Theorem. Let $X_1, \ldots, X_n, n \geq 2$, be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then with $100(1-\alpha)\%$ confidence

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

and $\left[0, \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}\right]$ is a $100(1-\alpha)\%$ upper confidence interval for σ^2. Similarly, with $100(1-\alpha)\%$ confidence

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

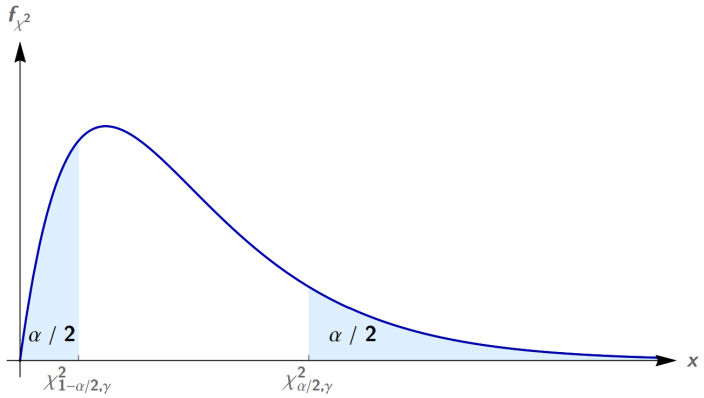


Figure 6: Chi-squared distribution in interval estimation

and $\left[\dfrac{(n-1)S^2}{\chi^2_{\alpha,n-1}}, \infty\right)$ is a $100(1-\alpha)\%$ lower confidence interval for σ^2

11.3 T-Distribution

Definition. Let Z be a standard normal variable and let χ^2_γ be an independent chi-squared random variable with γ degrees of freedom. The random variable

$$T_\gamma = \frac{Z}{\sqrt{\chi^2_\gamma/\gamma}} \tag{56}$$

is said to follow a T-distribution with γ degrees of freedom.
Theorem. The density of a T distribution with γ degrees of freedom is given by

$$f_{T_\gamma}(t) = \frac{\Gamma((\gamma+1)/2)}{\Gamma(\gamma/2)\sqrt{\pi\gamma}}\left(1+\frac{t^2}{\gamma}\right)^{-\frac{\gamma+1}{2}} \tag{57}$$

As the degree of freedom γ increases, T-distribution will approach normal distribution.
One intuitive reason that the T-distribution approaches the normal distribution as the degrees of freedom increases is that when the sample size n grows larger, the estimate for the sample variance improves, so that S^2 is likely to be closer to σ^2.

11.3.1 Interval Estimation for μ with Unknown Variance

Theorem. Let $X_1, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then a $100(1-\alpha)\%$ confidence interval on μ is given by

$$\bar{X} \pm t_{\alpha/2,n-1}S/\sqrt{n} \tag{58}$$

where $t_{\alpha/2,\gamma}$ are defined by

$$\int_{t_{\alpha/2,\gamma}}^{\infty} f_{T_\gamma}(t)dt = \alpha/2$$

Example. An article in the Journal of Testing and Evaluation presents the following 20 measurements on residual flame time (in seconds) of treated specimens of children's nightwear:

9.85	9.93	9.75	9.77	9.67	9.87	9.67	9.94	9.85	9.75
9.83	9.92	9.74	9.99	9.88	9.95	9.95	9.93	9.92	9.89

We wish to find a 95% confidence interval on the mean residual flame time. The sample mean and standard deviation are

$$\bar{x} = 9.8525, \quad s = 0.0965$$

We refer to the table for the T distribution with $20 - 1 = 19$ degrees of freedom and and $\alpha/2 = 0.025$ to obtain $t_{0.025,19} = 2.093$. Hence we are 95% certain that

$$\mu = (9.8525 \pm 0.0451)\,\text{sec}$$

Example. The proportion of color blind individuals in a population is to be estimated. Suppose the sample percentage of color blind individuals is 30%. Now we are able to get a 95% confidence interval with the true percentage deviating less than 3.1% points from the sample percentage. What would have been minimum sample size n to obtain this estimate? We know that $\bar{X} = 0.3$, sample standard deviation $S = \sqrt{0.3 * 0.7}$. So half of the length of confidence interval is

$$\frac{t_{\alpha/2,n-1} \cdot S}{\sqrt{n}} \leq 0.031$$

By using "InverseCDF" function in MMA, we can get $n \geq 842$.

12 Hypotheses Testing

Our goal is to find statistical evidence that allows us to reject the null hypothesis. The process of using statistical data to decide whether or not a hypothesis should be **rejected** is called "performing a hypothesis test".

12.1 Fisher's Null Hypothesis Test

The P-value is an upper bound of the probability of obtaining the data if H_0 is true. If D represents the statistical data. Suppose that $H_0 : \theta \leq \theta_0$

$$P[D \mid H_0] = P[D \mid \theta \leq \theta_0] \leq P[D \mid \theta = \theta_0] = P\text{-value} \tag{59}$$

and we will reject H_0 if this value is small, and we say **we reject H_0 by at the [P-value] level of significance**.

Example. We take a sample of 36 cars and find their gas mileages. We decide to base our rejection of H_0 on the sample mean. If $\mu = 26$ and $\sigma = 5$, the sample mean is normally distributed with $\mu = 26$ and standard deviation $\sigma/\sqrt{n} = 5/6$. Suppose that we find a sample mean $\bar{x} = 28.04$mpg. We now calculate the P-value of the test, i.e., the probability of obtaining this or a larger value of the sample mean if H_0 were true.

$$\begin{aligned} P[\bar{X} \geq 28.04 \mid \mu \leq 26, \sigma = 5] &\leq P[\bar{X} \geq 28.04 \mid \mu = 26, \sigma = 5] \\ &= P\left[\frac{\bar{X} - 26}{5/6} \geq \frac{28.04 - 26}{5/6}\right] \\ &= P[Z \geq 2.45] = 1 - P[Z \leq 2.45] \\ &= 1 - 0.9929 = 0.0071. \end{aligned}$$

This is the P-value of the test. Since it is very small, we decide to reject the null hypothesis at the 0.7% level of significance.

Explanation. However, if H_0 is rejected, it's still possible that H_0 is true, since we only calculate $P[D \mid H_0]$ but what's more meaningful is the probability $P[H_0 \mid D]$. Instead, what Fisher Test can tell us is that: if H_0 is true, then there's at least [P-value] probability that $\bar{X}$ is equal or larger than the sample size $\bar{x}$.

12.2 Neyman-Pearson Decision Theory

We don't need evidence that H_0 or H_1 is true. Instead, we act by assuming that H_0 or H_1 is true.
The statistical test will end with either
(1) Failing to reject H_0, therefore accepting H_0 or
(2) Rejecting H_0, thereby accepting H_1.
If we accept H_0, we do not necessarily believe H_0 to be true; we simply decide to act as if it were true. The same is the case if we decide to accept H_1; we are not necessarily convinced that H_1 is true, we merely decide to assume that it is.
We define the probability of committing a **Type I error**,

$$\begin{aligned}\alpha &:= P[\text{ Type I error }] = P[\text{ reject } H_0 \mid H_0 \text{ true }] \\ &= P[\text{ accept } H_1 \mid H_0 \text{ true }].\end{aligned}$$

The probability of committing a **Type II error** is denoted

$$\begin{aligned}\beta &:= P[\text{ Type II error }] = P[\text{ fail to reject } H_0 \mid H_1 \text{ true }] \\ &= P[\text{ accept } H_0 \mid H_1 \text{ true }].\end{aligned}$$

Related to β is the power of the test, defined as

$$\begin{aligned}\text{Power } &:= 1-\beta = P[\text{ reject } H_0 \mid H_1 \text{ true }] \\ &= P[\text{ accept } H_1 \mid H_1 \text{ true }].\end{aligned}$$

12.2.1 α and Critical Region

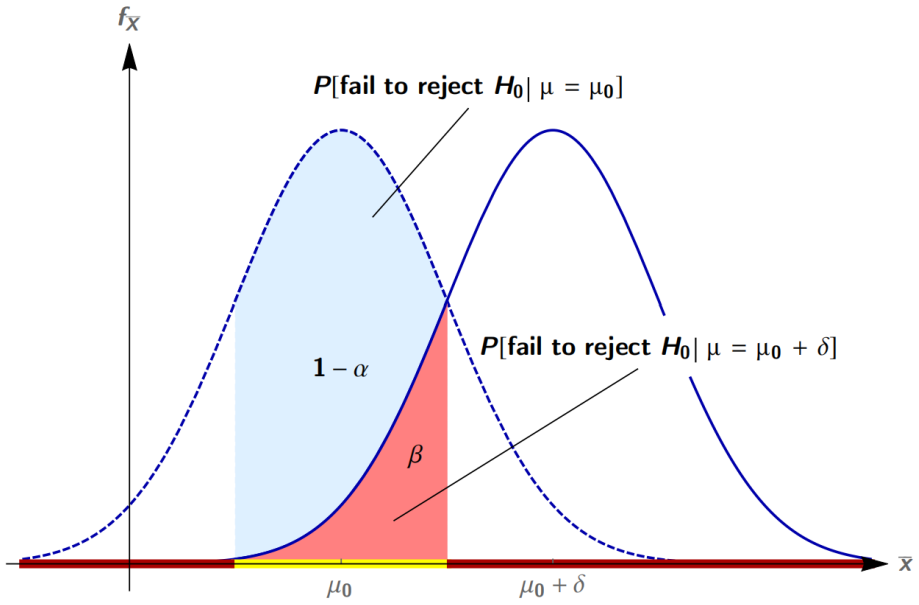
A **Critical Region** is that, if the data falls in that region, we can conclude that the type I error would be small, which means H_0 is probably wrong. **In order for the statistical procedure to be valid, the critical region must be fixed before data are obtained.** With the fixed sample size n, the critical region is given by

$$\bar{x} \notin [\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}] \tag{60}$$

If H_0 is a one-sided hypothesis, then the critical region will be

$$\bar{x} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{x} \leq \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} \tag{61}$$

In this scheme, The decision whether to reject H_0 or not is not driven by the probability of H_0 being true or not, but solely by the probability of committing an error if H0 is falsely rejected.



12.2.2 β and Sample Size

With fixed critical region, we can only adjust the sample size n to make β as small as possible.

Based on the curves in the figures, we can see β is equal to the area on the interval $[\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$. When $\mu = \mu_0 + \delta$, by normalizing it with $\frac{(t-\mu_0)\sqrt{n}}{\sigma}$, we have

$$P[\text{fail to reject } H_0 \mid \mu = \mu_0 + \delta] = \frac{1}{\sqrt{2\pi}} \int_{-z_{\alpha/2}}^{z_{\alpha/2}} e^{-(t-\delta\sqrt{n}/\sigma)^2/2} dt \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{\alpha/2}-\delta\sqrt{n}/\sigma} e^{-t^2/2} dt$$

From the figure, we know that a smaller δ will yield a larger β, and the upper bound is taken at δ_0:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{\alpha/2}-\delta\sqrt{n}/\sigma} e^{-t^2/2} dt \leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{\alpha/2}-\delta_0\sqrt{n}/\sigma} e^{-t^2/2} dt \approx \beta$$

While also we have

$$\beta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z_\beta} e^{-t^2/2} dt$$

Then according to the equations above, we get

$$n \approx \frac{\left(z_{\alpha/2} + z_\beta\right)^2 \sigma^2}{\delta_0^2} \tag{62}$$

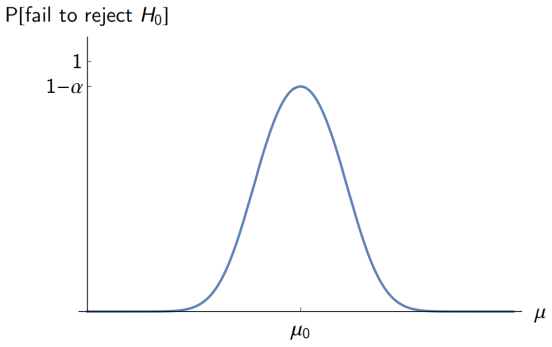
Or if H_0 is a one-sided hypothesis, we have

$$n \approx \frac{\left(z_\alpha + z_\beta\right)^2 \sigma^2}{\delta_0^2} \tag{63}$$

Remark 1. Actually, only after n is determined can we calculate the critical region.
Remark 2. In Neyman-Pearson test, we should not let H_0 and H_1 oppose to each other completely. That's because, if $\delta_0 \to 0$, it's obvious that we should have an extremely large n to acquire a small β. What's worse, when $\delta_0 = 0$, the two curves shown in the above figure will overlap and $\beta = 1 - \alpha$. That is nonsense.

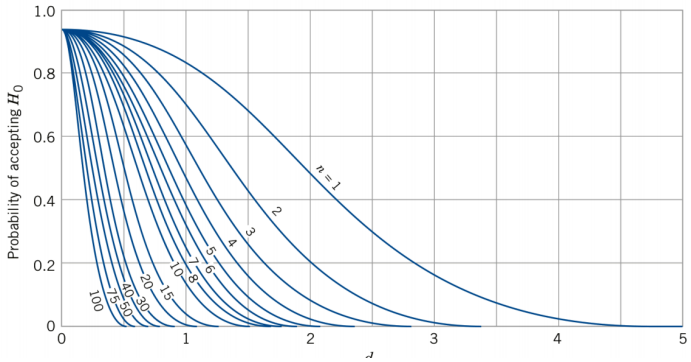
12.2.3 Operating Characteristic Curves (OC Curves)

OC curves can describe the relationship between $\beta(\mu)$ and μ. **Do notice that it's not the same as the shape of normal distribution.** For different values of α, the curve will



scale correspondingly; if the sample size n gets larger, the curve will be more narrow but have the same maximum point $(\mu_0, 1-\alpha)$.

In the second graph, d is defined as



$$d := \frac{|\mu - \mu_0|}{\sigma} \tag{64}$$

Then the OC curve can give us two aspects of information:
(1) Given the sample size n, we can easily read the value $\beta(\mu)$.
(2) Given β and $H_1 : \mu \geq \mu_0 + \delta + 0$, we can draw a horizontal line $y = \beta$ and the box $x \geq d_0 = \frac{\delta_0}{\sigma}$. Then by choosing a curve that lies under $y = \beta$ when $x = \frac{\delta_0}{\sigma}$, we can select a reasonable sample size n.

12.2.4 Steps of Neyman-Pearson Hypothesis Testing

(i) Select appropriate hypotheses H_1 and H_0 and a test statistic;
(ii) Fix α and β for the test;
(iii) Use α and β to determine the appropriate the sample size;
(iv) Use α and the sample size to determine the critical region;
(v) Obtain the sample statistic; if the test statistic falls into the critical region, reject H_0 at significance level α and accept H_1. Otherwise, accept H_0.

12.3 Null Hypothesis Significance Testing

To be probabilistically pure in the NHST sense, an experiment should be run once, and if the null hypothesis is not rejected, it should not be repeated.

13 Single Sample Test for the Mean and Variance

The following hypothesis tests are suitable if the standard deviation σ is unknown.

13.1 The T-Test for μ

Any test based on the statistic

$$T_{n-1} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \tag{65}$$

is called a T-test.
We reject at significance level α

$$H_0 : \mu = \mu_0 \text{ if } |T_{n-1}| > t_{\alpha/2,n-1}$$

$$H_0 : \mu \leq \mu_0 \text{ if } T_{n-1} > t_{\alpha,n-1}$$

$$H_0 : \mu \geq \mu_0 \text{ if } T_{n-1} < -t_{\alpha,n-1}$$

The T-test based on OC curves. Here we use use sample standard deviation s to replace σ.

$$d := \frac{|\mu - \mu_0|}{s} \tag{66}$$

13.2 The Chi-Squared Test for σ

Let $X_1, \ldots, X_n$ be a random sample of size n from a normal distribution and let S^2 denote the sample variance. Let σ^2 be the unknown population variance and σ_0^2 a null value of that variance. Then a test for the variance based on the statistic

$$\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma_0^2} \tag{67}$$

is called a chi-squared test.
We reject at significance level α

$$H_0 : \sigma = \sigma_0 \text{ if } \chi^2_{n-1} > \chi^2_{\alpha/2,n-1} \text{ or } \chi^2_{n-1} < \chi^2_{1-\alpha/2,n-1}$$

$$H_0 : \sigma \leq \sigma_0 \text{ if } \chi^2_{n-1} > \chi^2_{\alpha,n-1}$$

$$H_0 : \sigma \geq \sigma_0 \text{ if } \chi^2_{n-1} < \chi^2_{1-\alpha,n-1}$$

The Chi-squared test based on OC curves. Here we define the abscissa parameter for the two-tailed chi-squared test is

$$\lambda = \frac{\sigma}{\sigma_0} \tag{68}$$

Example. Returning to Example 19.5 , the engineers concerned are dissatisfied that H_0 was not rejected. A second test (this time of Neyman-Pearson type) is to be performed to

establish that the standard deviation is less than $\sigma_0 = 1.5$ mm.
1. If we want to preset $\alpha = 0.05$, what is the critical region for the test at a sample size $n = 20$?
2. If $n = 20$, what true value of σ is necessary so that the test will have a power of $1-\beta = 0.9$?
3. For $\alpha = 0.05$, make a statement on the sample size necessary to ensure that H_0 is rejected with 90% probability if $\sigma = 1.35$.
Solution. 1. From the table for the χ^2_{19} distribution we see that $P\left[\chi^2_{1-0.05,19} \leq 10.1\right] = 0.05$, so the critical region for the variance is

$$\frac{(n-1)s^2}{\sigma_0^2} < 10.1 \quad \Leftrightarrow \quad s^2 < \frac{2.25 \cdot 10.1}{19} = 1.20$$

i.e., $s < 1.09$.
2. For $n = 20$, the line intersects the horizontal rule $\beta = 0.1$ at $\lambda = 0.6$. This means that

$$\sigma < 0.6\sigma_0 = 0.9$$

is necessary for H_0 to be rejected 90% of the time.
3. We take $y = 1 - \beta = 0.1$ and $x = \lambda = \dfrac{\sigma}{\sigma_0} = \dfrac{1.35}{1.50} = 0.9$, then the graph shows that a sample size significantly larger than $n = 100$ would be necessary.

14 Non-Parametric Single Sample Test for the Median

A non-parametric test should follow two basic rules:
(1) Non-parametric statistics do not assume the dependence on any parameter.
(2) Distribution-free statistics do not assume that X follows any particular distribution (such as the normal distribution).

14.1 Sign Test for the Median

The median of a random variable X is defined as the value M such that

$$P[X < M] + \frac{1}{2}P[X = M] = \frac{1}{2} \quad \text{or} \quad P[X > M] + \frac{1}{2}P[X = M] = \frac{1}{2} \tag{69}$$

Given a sample $X_1, \ldots, X_n$, define

$$Q_+ = \#\left\{X_k : X_k - M_0 > 0\right\}, \quad Q_- = \#\left\{X_k : X_k - M_0 < 0\right\}$$

So $Q_+$is the number of "positive signs" and $Q_-$the number of "negative signs." We note that

$$P\left[Q_- \leq k \mid M = M_0\right] = \sum_{x=0}^{k} \binom{n}{x} \frac{1}{2^n}$$

We reject at significance level α

$$H_0 : M \leq M_0 \text{ if } P\left[Q_- \leq k \mid M = M_0\right] < \alpha$$

$$H_0 : M \geq M_0 \text{ if } P\left[Q_+ \leq k \mid M = M_0\right] < \alpha$$

Example. A certain six-sided die is suspected of being unbalanced. Based on past experience, it is suspected that the median is greater than 3.5. We decide to test the null hypothesis

$$H_0 : M \leq 3.5.$$

We note that there are 6 negative signs,

$$Q_- = 6.$$

By rolling the dice for 20 times, we then find that

$$P\left[Q_- \leq 6 \mid M = 3.5\right] = \frac{1}{2^{20}} \sum_{x=0}^{6} \binom{20}{x} = 0.0577.$$

This is the P-value of the test. It would be reasonable to decide not to reject H_0, i.e., the results do not provide convincing evidence that H_0 is false.

14.2 Rank Test

14.2.1 Symmetric Distribution

The analysis of ranks supposes that the data comes from a distribution that is symmetric about its median.
A random variable X is said to be symmetric about $a \in \mathbb{R}$ if

$$X - a \quad \text{and} \quad -(X - a)$$

have the same distribution. In terms of the density function f_X this means that

$$f_X(x - a) = f_X(a - x) \tag{70}$$

14.2.2 Rank

Several rules of giving assign a rank to each point:
(1) The signed rank is found by multiplying the rank with -1 if $X_i - M_0 < 0$ and +1 if $X_i - M_0 > 0$.
(2) The positive ranks as well as the negative ranks are summed separately, yielding two statistics $W_+$and W_-.
(3) Ties in ranks are assigned the average of their ranks.
(4) The total sum of the ranks is always $n(n+1)/2$.
Theorem For non-small sample sizes ($n \geq 10$) a normal distribution with parameters

$$E[W] = \frac{n(n+1)}{4}, \quad \text{Var}[W] = \frac{n(n+1)(2n+1)}{24} \tag{71}$$

may be used as an approximation. However, in that case the variance needs to be reduced if there are ties: for each group of t ties, the variance is reduced by $(t^3 - t)/48$.
So, suppose that there are n groups of ties, each having t_i ties, then the variance will be

$$\text{Var}[W] = \frac{n(n+1)(2n+1)}{24} - \sum_{i=1}^{n} \frac{t_i^3 - t_i}{48} \tag{72}$$

14.2.3 Wilcoxon Signed Rank Test

Wilcoxon Signed Rank Test. Let $X_1, \ldots, X_n$ be a random sample of size n from a **symmetric distribution**. Order the n absolute differences $|X_i - M|$ according to magnitude, so that $X_{R_i} - M_0$ is the R_i th smallest difference by modulus. If ties in the rank occur, the mean of the ranks is assigned to all equal values. Let

$$W_+ = \sum_{R_i>0} R_i, \qquad |W_-| = \sum_{R_i<0} |R_i| .$$

We reject at significance level α

$$H_0 : M \leq M_0 \text{ if } |W_-| \text{ is smaller than the critical value for } \alpha$$

$$H_0 : M \geq M_0 \text{ if } W_+ \text{ is smaller than the critical value for } \alpha$$

$$H_0 : M = M_0 \text{ if } W = \min(W_+, |W_-|) \text{ is smaller than the critical value for } \alpha/2$$

Example. Returning to the previous example, we want to test $H_0 : M \leq 3.5$ and have the following observations, ordered from smallest to largest:

X_i	$X_i - M_0$	R_i	X_i	$X_i - M_0$	R_i
3	-0.5	-5.5	2	-1.5	-13
3	-0.5	-5.5	5	1.5	+13
3	-0.5	-5.5	5	1.5	+13
3	-0.5	-5.5	5	1.5	+13
4	0.5	+5.5	5	1.5	+13
4	0.5	+5.5	1	-2.5	-18
4	0.5	+5.5	6	2.5	+18
4	0.5	+5.5	6	2.5	+18
4	0.5	+5.5	6	2.5	+18
4	0.5	+5.5	2	2.5	+18

We calculate the sum of the negative ranks,

$$w_- = -5.5 - 5.5 - 5.5 - 5.5 - 13 - 18 = -53.$$

Consulting a table, the critical value for $n = 20$ and $\alpha = 0.05$ is 60. For $\alpha = 0.01$ it is 43. Since $|w_-|$lies between these values, the P-value of the test is between 1% and than 5%, most likely around $2\% - 3\%$. Alternatively, we may use the normal distribution with mean $\mu = n(n+1)/4 = 105$ and variance

$$\sigma^2 = \frac{n(n+1)(2n+1)}{24} - \frac{10^3 - 10}{48} - 2 \cdot \frac{5^3 - 5}{48}.$$

Then

$$z = \frac{|w_-| - \mu}{\sigma} = -1.977$$

and we find that $P[Z < -1.977] = 0.024$. That means we can reject H_0 with the significance level 0.024.

Remark. If we yield a relatively low $P-$value, we can make following possible conclusions:
(1) The die results follow a non-symmetric distribution, or
(2) The die results follow a symmetric distribution, but the median does not follows H_0.

15 Inferences of Proportions

15.1 Parameter Estimation

It follows immediately that the following is a $100(1-\alpha)\%$ confidence interval for p :

$$\widehat{p} \pm z_{\alpha/2}\sqrt{p(1-p)/n}$$

But the interval depends on the unknown parameter p, which we are actually trying to estimate! One solution to the problem is to replace p by $\widehat{p}$, i.e., to write

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n} \tag{73}$$

Remark. The requirement is that the sample size n should be enough to make $t_{\alpha/2}$ approximate $z_{\alpha/2}$. Otherwise, we should use **binomial distribution** and calculate the confidence interval by hand rather than using normal approximation.

We may want to be able to claim that "with $xx\%$ probability, $\hat{p}$ differs from p by at most d". Then if we have done pre-sampling and a rough estimator $\hat{p}$ has been obtained, then

$$n = \frac{z_{\alpha/2}^2 \widehat{p}(1-\widehat{p})}{d^2} \tag{74}$$

Corollary. If no estimate for p is available, we can at least use that $x(1-x) < 1/4$ for all $x \in \mathbb{R}$ to deduce that

$$n = \frac{z_{\alpha/2}^2}{4d^2} \tag{75}$$

will ensure $|p - \widehat{p}| < d$ with $100(1-\alpha)\%$ confidence.

15.2 Hypothesis Testing for Proportion

Test for Proportion. Let $X_1, \ldots, X_n$ be a random sample of (large) size n from a Bernoulli distribution with parameter p and let $\widehat{p} = \bar{X}$ denote the sample mean. Then any test based on the statistic

$$Z = \frac{\widehat{p} - p_0}{\sqrt{p_0\left(1-p_0\right)/n}}$$

is called a large-sample test for proportion. We reject at significance level α

$$H_0 : p = p_0 \text{ if } |Z| > z_{\alpha/2}$$

$$H_0 : p \leq p_0 \text{ if } Z > z_{\alpha}$$

$$H_0 : p \geq p_0 \text{ if } Z < -z_{\alpha}$$

15.3 Two Proportions

An unbiased estimator for $p_1 - p_2$ is

$$\widehat{p_1 - p_2} := \widehat{p}_1 - \widehat{p}_2 = \bar{X}^{(1)} - \bar{X}^{(2)}$$

where $\bar{X}^{(1)}$ and $\bar{X}^{(2)}$ are the sample means of the respective random samples.
Since we know that

$$\widehat{p_1 - p_2} \sim N\left(p_1 - p_2, \frac{p_1\left(1-p_1\right)}{n_1} + \frac{p_2\left(1-p_2\right)}{n_2}\right)$$

We can get the confidence interval:

$$\widehat{p}_1 - \widehat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}_1\left(1-\widehat{p}_1\right)}{n_1} + \frac{\widehat{p}_2\left(1-\widehat{p}_2\right)}{n_2}} \tag{76}$$

For hypothesis testing, the statistic should be

$$Z = \frac{\widehat{p}_1 - \widehat{p}_2 - (p_1 - p_2)_0}{\sqrt{\frac{\widehat{p}_1\left(1-\widehat{p}_1\right)}{n_1} + \frac{\widehat{p}_2\left(1-\widehat{p}_2\right)}{n_2}}} \tag{77}$$

15.4 Pooled Estimator and Pooled Test

We define the pooled estimator for the proportion,

$$\widehat{p} := \frac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1 + n_2}. \tag{78}$$

And for hypothesis testing, we define the statistic

$$Z = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{79}$$

Remark 1. When comparing two proportions, we must firstly make sure that the two populations are independent.
Remark 2. We should first guarantee that the two criteria are not related with each other.
Example. Many consumers think that automobiles built on Mondays are more likely to have serious defects than those built on any other day of the week. To support this theory, a random sample of 100 cars built on Monday is selected and inspected. Of these, eight are found to have serious defects. A random sample of 200 cars produced on other days reveals 12 with serious defects. Do these data support the stated contention? We test

$$H_0 : p_1 \leq p_2.$$

where p_1 denotes the proportion of cars with serious defects produced on Mondays. Estimates for p_1 and p_2 are

$$\widehat{p}_1 = 8/100 = 0.08, \quad \widehat{p}_2 = 12/200 = 0.06$$

The pooled estimate for the common population proportion is

$$\widehat{p} = \frac{100 \cdot 0.08 + 200 \cdot 0.06}{100 + 200} = 20/300 = 0.066.$$

The observed value of the test statistic is

$$\frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.08 - 0.06}{\sqrt{0.066 \cdot 0.934\left(\frac{1}{100} + \frac{1}{200}\right)}} = 0.658.$$

Then we know that $P[Z \geq 0.658] = 0.2546$, so there is no evidence that H_0 might be false.

16 Comparison of Variances and Means

16.1 Comparison of Two Variances

16.1.1 The F-Distribution

Definition. Let $X^2_{\gamma_1}$ and $X^2_{\gamma_2}$ be independent chi-squared random variables with γ_1 and γ_2 degrees of freedom, respectively. The random variable

$$F_{\gamma_1,\gamma_2} = \frac{X^2_{\gamma_1}/\gamma_1}{X^2_{\gamma_2}/\gamma_2} \tag{80}$$

is said to follow an F-distribution with γ_1 and γ_2 degrees of freedom. And we can get that

$$P\left[F_{\gamma_1,\gamma_2} < x\right] = P\left[\frac{1}{F_{\gamma_1,\gamma_2}} > \frac{1}{x}\right] = 1 - P\left[F_{\gamma_2,\gamma_1} < \frac{1}{x}\right] \tag{81}$$

The probability density function of the $F-$distribution is

$$f_{\gamma_1,\gamma_2}(x) = \gamma_1^{\gamma_1/2}\gamma_2^{\gamma_2/2}\frac{\Gamma\left(\dfrac{\gamma_1+\gamma_2}{2}\right)}{\Gamma\left(\dfrac{\gamma_1}{2}\right)\Gamma\left(\dfrac{\gamma_2}{2}\right)}\frac{x^{\gamma_1/2-1}}{(\gamma_1 x+\gamma_2)^{(\gamma_1+\gamma_2)/2}} \tag{82}$$

Remark. As x increases, the value of Chi-squared distribution goes to 0 exponentially, but the $F-$distribution goes to 0 polynomially.

We define the critical point of the F-distribution $f_{\alpha,\gamma_1,\gamma_2}$ such that

$$P\left[F_{\gamma_1,\gamma_2} > f_{\alpha,\gamma_1,\gamma_2}\right] = \alpha$$

And from the relaion between $F_{\gamma 1,\gamma 2}$ and $F_{\gamma 2,\gamma 1}$, we have

$$f_{1-\alpha,\gamma_1,\gamma_2} = \frac{1}{f_{\alpha,\gamma_2,\gamma_1}} \tag{83}$$

16.1.2 The F-Test

Theorem. Let S_1^2 and S_2^2 be sample variances based on independent random samples of sizes n_1 and n_2 drawn from normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. If $\sigma_1^2 = \sigma_2^2$, then the statistic

$$S_1^2/S_2^2$$

follows an F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

The $F-$test is based on the statistic

$$F_{n_1-1,n_2-1} = \frac{S_1^2}{S_2^2} \tag{84}$$

We reject at significance level α

$$H_0 : \sigma_1 \le \sigma_2 \quad \text{if} \quad \frac{S_1^2}{S_2^2} > f_{\alpha,n_1-1,n_2-1}$$

$$H_0 : \sigma_1 \ge \sigma_2 \quad \text{if} \quad \frac{S_2^2}{S_1^2} > f_{\alpha,n_2-1,n_1-1}$$

$$H_0 : \sigma_1 = \sigma_2 \quad \text{if} \quad \frac{S_1^2}{S_2^2} > f_{\alpha/2,n_1-1,n_2-1} \text{ or } \frac{S_2^2}{S_1^2} > f_{\alpha/2,n_2-1,n_1-1}$$

Remark. For the F-test to be applicable, it is essential that the populations are normally distributed, and it will be more powerful if the sample sizes n_1 and n_2 are equal.

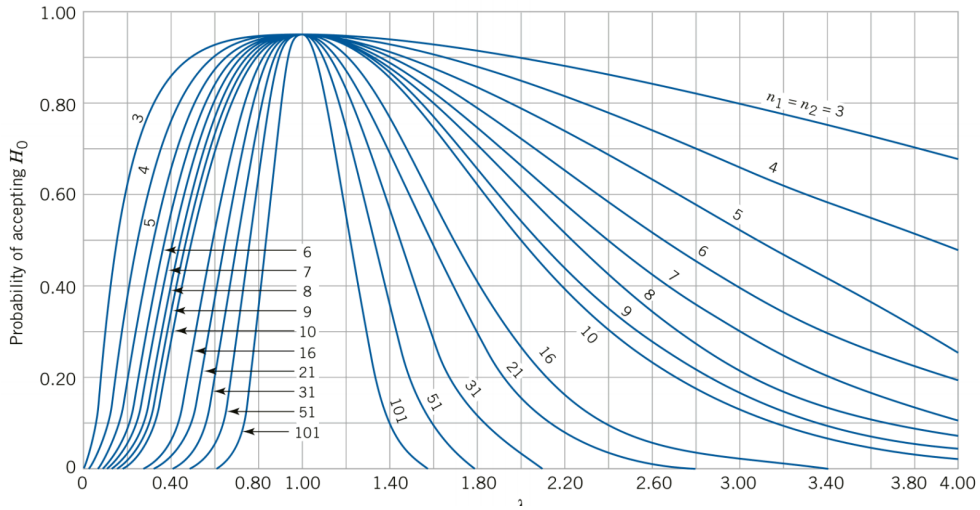


Figure 7: OC curves for F-test.

16.2 Comparison of Two Means

16.2.1 Neyman-Pearson Test with Variances Known

If we know σ_1 and σ_2, the test statistic is

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \tag{85}$$

We can also use the OC curves for the normal distribution to find power and sample size for a test. In that case, we use

$$d = \frac{|\mu_1 - \mu_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad \text{or} \quad d = \frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

for two-sided test and one-sided test respectively, with $n = n_1 = n_2$ (equal sample sizes). If $n_1 \neq n_2$, the table is used with the equivalent sample size

$$n = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}$$

Example. Continuing from Example 23.1, if H_1 is true, we want to find the sample sizes (number of days) required to detect this difference with a probability of 0.90 . We have $d = 10/\sqrt{40 + 50} = 1.05$ and using the chart for $\alpha = 0.05$ (one-sided) we find $n = n_1 = n_2 = 9$.

We may use the confidence interval

$$\mu_1 - \mu_2 = \bar{x}^{(1)} - \bar{x}^{(2)} \pm z_\alpha \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \tag{86}$$

To help checking whether or not to reject H_0.

16.2.2 Compare Two Means with Unknown but Equal Variances

We use the test statistic

$$T_{n_1+n_2-2} = \frac{Z}{\sqrt{X^2_{n_1+n_2-2}/\left(n_1+n_2-2\right)}} = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{S_p^2\left(1/n_1 + 1/n_2\right)}} \tag{87}$$

where we define the pooled estimator for the variance

$$S_p^2 = \frac{\left(n_1-1\right)S_1^2 + \left(n_2-1\right)S_2^2}{n_1+n_2-2} \tag{88}$$

We immediately obtain the following $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$,

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t_{\alpha/2,n_1+n_2-2}\sqrt{S_p^2\left(1/n_1 + 1/n_2\right)} \tag{89}$$

Theorem. Student's T-Test for Equal Variances. Suppose two random samples of sizes n_1 and n_2 from two normal distributions $N\left(\mu_1, \sigma^2\right)$ and $N\left(\mu_2, \sigma^2\right)$ are given. Denote by $\bar{X}^{(1)}$ and $\bar{X}^{(2)}$ the means of the two samples and let S_p^2 be the pooled sample variance (23.1). Let $\left(\mu_1 - \mu_2\right)_0$ be a null value for the difference $\mu_1 - \mu_2$. Then the test based on the statistic

$$T_{n_1+n_2-2} = \frac{\left(\bar{X}^{(1)} - \bar{X}^{(2)}\right) - \left(\mu_1 - \mu_2\right)_0}{\sqrt{S_p^2\left(1/n_1 + 1/n_2\right)}}$$

is called a Student's (pooled) test for equality of means. We reject at significance level α

$$H_0 : \mu_1 - \mu_2 = \left(\mu_1 - \mu_2\right)_0 \text{ if } \left|T_{n_1+n_2-2}\right| > t_{\alpha/2,n_1+n_2-2}$$
$$H_0 : \mu_1 - \mu_2 \leq \left(\mu_1 - \mu_2\right)_0 \text{ if } T_{n_1+n_2-2} > t_{\alpha,n_1+n_2-2}$$
$$H_0 : \mu_1 - \mu_2 \geq \left(\mu_1 - \mu_2\right)_0 \text{ if } T_{n_1+n_2-2} < -t_{\alpha,n_1+n_2-2}$$

In the case of equal variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and equal sample sizes $n_1 = n_2 = n$, we can use the usual OC curves for the T-test with

$$d = \frac{\left|\mu_1 - \mu_2\right|}{2S_p}$$

However, we must use the **modified sample size** $n^* = 2n - 1$ when reading the charts.
Remark. When performing pre-tests, we must **gather new data** for a comparison of means test.

16.2.3 The Welch-Satterthwaite Approximation

Welch's T-Test for Unequal Variances. Suppose two random samples of sizes n_1 and n_2 from two normal distributions $N\left(\mu_1, \sigma_1^2\right)$ and $N\left(\mu_2, \sigma_2^2\right)$ are given. Denote by $\bar{X}^{(1)}$ and $\bar{X}^{(2)}$ the means of the two samples and let γ given by (23.2). Let $\left(\mu_1 - \mu_2\right)_0$ be a null value for the difference $\mu_1 - \mu_2$. Then the test based on the statistic

$$T_\gamma = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

is called a Welch's (pooled) test for equality of means. We reject at significance level α - $H_0 : \mu_1 - \mu_2 = \left(\mu_1 - \mu_2\right)_0$ if $\left|T_\gamma\right| > t_{\alpha/2,\gamma}$, - $H_0 : \mu_1 - \mu_2 \leq \left(\mu_1 - \mu_2\right)_0$ if $T_\gamma > t_{\alpha,\gamma}$, - $H_0 : \mu_1 - \mu_2 \geq \left(\mu_1 - \mu_2\right)_0$ if $T_\gamma < -t_{\alpha,\gamma}$. where

$$\gamma = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\dfrac{\left(S_1^2/n_1\right)^2}{n_1-1} + \dfrac{\left(S_2^2/n_1\right)^2}{n_2-1}} \tag{90}$$

17 Non-Parametric Comparisons

17.1 Wilcoxon Rank-Sum Test

24.1. Wilcoxon Rank-Sum Test. Let X and Y be two random samples following some continuous distributions. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n, \boldsymbol{m} \leq \boldsymbol{n}$, be random samples from X and Y and associate the rank $R_i, i = 1, \ldots, m+n$, to the R_i th smallest among the $m+n$ total observations. If ties in the rank occur, the mean of the ranks is assigned to all equal values. Then the test based on the statistic

$$W_m := \text{ sum of the ranks of } X_1, \ldots, X_m.$$

is called the Wilcoxon rank-sum test. We reject $H_0 : P[X > Y] = 1/2$ (and similarly the analogous one-sided hypotheses) at significance level α if W_m falls into the corresponding critical region.
For large values of $m(m \geq 20), W_m$ is approximately normally distributed with

$$\mathrm{E}\left[W_m\right] = \frac{m(m+n+1)}{2}, \quad \mathrm{Var}\left[W_m\right] = \frac{mn(m+n+1)}{12}$$

If there are many ties, the variance may be corrected by taking

$$\mathrm{Var}\left[W_m\right] = \frac{mn(m+n+1)}{12 - \sum_{\text{groups}} \frac{t^3+t}{12}}$$

17.2 Paired T-Test

We will assume that X and Y follow a joint bivariate normal distribution. Then it is not hard to see that $D = X - Y$ follows a normaldistribution.
Then

$$T_{n-1} = \frac{\bar{D} - \mu_D}{\sqrt{S_D^2/n}}$$

follows a T-distribution with $n-1$ degrees of freedom. We may find confidence intervals for μ_D and conduct hypothesis tests as we would for any normally distributed random variable. AT-test for D is called a paired T-test for X and Y.

17.3 Estimate and Test for Covariance

The natural choice (method of moments!) for an estimator for the correlation coefficient is then

$$R := \hat{\rho} = \frac{\sum \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum \left(X_i - \bar{X}\right)^2}\sqrt{\sum \left(Y_i - \bar{Y}\right)^2}}$$

We can thus test $H_0 : \varrho = \varrho_0$, by using the test statistic

$$\begin{aligned} Z &= \frac{\sqrt{n-3}}{2}\left(\ln\left(\frac{1+R}{1-R}\right) - \ln\left(\frac{1+\varrho_0}{1-\varrho_0}\right)\right) \\ &= \sqrt{n-3}\left(\mathrm{Artanh}(R) - \mathrm{Artanh}\left(\varrho_0\right)\right) \end{aligned}$$

18 Categorial Data

Definition. A random vector $((X_1, \ldots, X_k), f_{X_1 X_2 \cdots X_k})$ with

$$(X_1, \ldots, X_k) : S \to \Omega = \{0, 1, 2, \ldots, n\}^k$$

and (joint) distribution function $f_{X_1 X_2 \cdots X_k} : \Omega \to \mathbb{R}$ given by

$$f_{X_1 X_2 \cdots x_k}(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \tag{91}$$

$p_1, \ldots, p_k \in (0, 1), n \in \mathbb{N} \backslash \{0\}$ is said to have a multinomial distribution with parameters n and $p_1, \ldots, p_k$.
We have

$$E[X_i] = np_i \quad \text{Var}[X_i] np_i(1 - p_i) \quad \text{Cov}[X_i, X_j] = -np_i p_j$$

18.1 The Pearson Statistic

If n is large, then the Pearson statistic

$$\sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \text{Chi}(k - 1) \tag{92}$$

where k is the number of categories.
Cochran's Rule states that for n, we should require

$$E[X_i] = np_i \geq 1 \quad \text{for all } i = 1, ..., k$$

$$E[X_i] = np_i \geq 5 \quad \text{for } 80\% \text{ of } i = 1, ..., k$$

18.2 Test for Multinomial Distribution

Pearson's Chi-squared Goodness-of-Fit Test. Let $(X_1, \ldots, X_k)$ be a sample of size n from a categorical random variable with parameters $(p_1, \ldots, p_k)$ satisfying Cochran's Rule. Let $(p_{1_0}, \ldots, p_{k_0})$ be a vector of null values. Then the test

$$H_0 : p_i = p_{i_0}, \quad i = 1, \ldots, k,$$

based on the statistic

$$X_{k-1}^2 = \sum_{i=1}^{k} \frac{(X_i - np_{i_0})^2}{np_{i_0}} \tag{93}$$

is called an chi-squared goodness-of-fit test. We reject H_0 at significance level α if

$$X_{k-1}^2 > \chi_{\alpha, k-1}^2$$

18.3 Goodness-of-Fit for a Discrete Distribution

$$\sum_{i=1}^{k} \frac{(X_i - E[X_i])^2}{E[X_i]} \sim \text{Chi}(k - 1 - m) \tag{94}$$

where m represents the number of parameters.

18.4 Goodness-of-Fit for a Continuous Distribution

For continuous distribution, we have

$$p_i = P[a_{i-1} \leq X \leq a_i] = \int_{a_{i-1}}^{a_i} f(x)dx$$

We follow the following steps when we need to test if the data corresponds with a certain continuous distribution:
(1) Calculate the estimator of the parameter (e.g. $\hat{\beta}$).
(2) Arbitrarily choose the number of categories (e.g. $k = 10$).
(3) Calculate the boundaries of each category.
(4) For each category, obtain the frequency.
(5) Use $\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \text{Chi}(k-1-m)$ to calculate Pearson statistic.

18.5 Cell Probabilities and Independence

If the row and column categorizations are independent, then it should be the case that

$$H_0 : p_{ij} = p_{i\cdot}p_{\cdot j}, \quad i = 1, \ldots, r, j = 1, \ldots, c.$$

We will therefore develop a test to determine whether there is statistical evidence that the above statement is false.
Hence, if H_0 is assumed, the expected number of elements in the (i, j) th cell is

$$E_{ij} = n \cdot \widehat{p_{ij}} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$$

We can now compare the observed frequencies O_{ij} in the (i, j) th cell to the expected frequencies E_{ij}. We will again use the Pearson statistic

$$X^2_{(r-1)(c-1)} = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

which follows a chi-squared distribution with

$$k - 1 - m = rc - 1 - (r + c - 2) = rc - r - c + 1 = (r - 1)(c - 1)$$

degrees of freedom. We reject H_0 if the value of $X^2_{(r-1)(c-1)}$ exceeds the critical value of the corresponding chi-squared distribution.

19 Simple Linear Regression

In this section, we assume that the mean $\mu_{Y|x}$ of $Y \mid x$ is given by

$$\mu_{Y|x} = \beta_0 + \beta_{1x} \quad \text{for some } \beta_0, \beta_1 \in \mathbb{R}.$$

This is called a simple linear regression model with model parameters β_0 and β_1. Another way of writing this model is

$$Y \mid x = \beta_0 + \beta_1 x + E$$

where $E[E] = 0$. Our goal is to find estimators

$$B_0 := \widehat{\beta_0} = \text{ estimator for } \beta_0, \quad b_0 = \text{ estimate for } \beta_0,$$
$$B_1 := \widehat{\beta_1} = \text{ estimator for } \beta_1, \quad b_1 = \text{ estimate for } \beta_1,$$

19.1 Least Squares Estimation

Given a sample of size n, we define the error sum of squares

$$\mathrm{SS_E} := e_1^2 + e_2^2 + \cdots + e_n^2 = \sum_{i=1}^{n} \left(y_i - (b_0 + b_1 x_i)\right)^2 .$$

Since we determine the estimators for β_0 and β_1 by minimizing this sum of squares, b_0 and b_1 are called least-squares estimates.

19.2 The Normal Equation

These are linear equations for b_0 and b_1, which may be easily solved:

$$\begin{aligned} b_1 &= \frac{n\sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ b_0 &= \frac{1}{n}\sum_{i=1}^{n} y_i - b_1 \cdot \frac{1}{n}\sum_{i=1}^{n} x_i \end{aligned} \tag{95}$$

We take the notation

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \quad S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Then the equation above can be expressed as

$$b_0 = \bar{y} - b_1\bar{x} \quad b_1 = \frac{S_{xy}}{S_{xx}} \quad \mathrm{SS_E} = S_{yy} - b_1 S_{xy} \tag{96}$$

Model Assumption.
(i) For each value of x, the random variable $Y \mid x$ follows a normal distribution with variance σ^2 and mean $\mu_{Y|x} = \beta_0 + \beta_1 x$.
(ii) The random variables $Y \mid x_1$ and $Y \mid x_2$ are independent if $x_1 \neq x_2$.
A random sample of size n consists of n pairs $(x_i, Y_i), i = 1, \ldots, n$, where the random variables $Y_i = Y \mid x_i$ are i.i.d. normal with variance σ^2 and mean $\mu_{Y|x_i} = \beta_0 + \beta_1 x_i$
Remark. We do not require that $x_i \neq x_j$. The random sample may contain more than a single measurement of $Y \mid x_i$. All the x_i are treated in the same way, e.g., when calculating $\bar{x}$.

19.3 Distribution of the Least Squares Estimators

Theorem. Given a random sample of $Y \mid x$ of size n, the statistics

$$\frac{B_1 - \beta_1}{\sigma/\sqrt{\sum (x_i - \bar{x})^2}} \quad \text{and} \quad \frac{B_0 - \beta_0}{\sigma\sqrt{\frac{\sum x_i^2}{n\sum (x_i - \bar{x})^2}}} \tag{97}$$

follow a standard normal distribution. **In particular, B_0 and B_1 are unbiased estimators**.
If $\sum (x_i - \bar{x})^2$ is small, i.e., x_i are closed to each other, then the estimator B_1 will have a large vairance, which means the slope is unstable.
If $\sum x_i^2$ is small, i.e., the points are close to the y-axis, then the estimator B_0 will have a small variance, which means the intersect is stable.

19.4 Least Square Estimator for the Variance

The variance σ^2 of $Y \mid x$ is assumed to be the same for all values of x. To estimate it, we use the error sum of squares,

$$S^2 := \frac{\mathrm{SS_E}}{n-2} = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \widehat{\mu}_{Y|x_i}\right)^2 \tag{98}$$

It turns out that this estimator is unbiased for σ^2 and in fact

$$(n-2)S^2/\sigma^2 = \frac{\mathrm{SS_E}}{\sigma^2}$$

follows a chi-squared distribution with $n-2$ degrees of freedom.

19.5 Confidence Interval

We have $100(1-\alpha)\%$ confidence intervals

$$B_1 \pm t_{\alpha/2,n-2}\frac{S}{\sqrt{S_{xx}}}, \qquad B_0 \pm t_{\alpha/2,n-2}\frac{S\sqrt{\sum x_i^2}}{\sqrt{nS_{xx}}} \tag{99}$$

for β_1 and β_0, respectively.

19.6 Hypotheses Testing for Linear Regression

We say that a regression is **significant** if there is statistical evidence that $\beta_1 \neq 0$.
Test for Significance of Regression. Let $(x_i, Y \mid x_i), i = 1, \ldots, n$ be a random sample from $Y \mid x$. We reject

$$H_0 : \beta_1 = 0$$

at significance level α if the statistic

$$T_{n-2} = \frac{B_1}{S/\sqrt{S_{xx}}}. \tag{100}$$

satisfies $|T_{n-2}| > t_{\alpha/2,n-2}$

19.7 Distribution of Estimated Mean

Since B_0 and B_1 are unbiased estimators for β_0 and β_1, we know that

$$\hat{\mu}_{Y|x} = B_0 + B_1 x$$

is also an unbiased estimator for $\mu_{Y|x}$, and

$$\frac{\hat{\mu}_{Y|x} - \mu_{Y|x}}{\sigma\sqrt{\frac{1}{n} + \frac{(x-\bar{x}^2)}{S_{xx}}}} \tag{101}$$

follows a standard normal distribution.

20 Prediction and Model Analysis

20.1 Prediction

An estimate is a statistical statement on the value of an unknown, but fixed, population parameter, while a prediction is a statistical statement on the value of an essentially random quantity. And we defined a $100(1-\alpha)\%$ prediction interval $[L_1, L_2]$ for a random variable

$$P[L_1 \leq X \leq L_2] = 1 - \alpha$$

For $Y \hat{\mid} x$, after standardizing and dividing by S/σ we obtain the T_{n-2} random variable

$$T_{n-2} = \frac{\widehat{Y \mid x} - Y \mid x}{S\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \tag{102}$$

We thus obtain the following $100(1-\alpha)\%$ prediction interval for $Y \mid x$:

$$\widehat{Y \mid x} \pm t_{\alpha/2,n-2} S\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} \tag{103}$$

20.2 Model Analysis

The total variation of the response variable

$$\mathrm{SS_T} = S_{yy} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \tag{104}$$

We will also call this the Total Sum of Squares. It represents the variation of Y regardless of any model.
And we define

$$R^2 = \frac{\mathrm{SS_T} - \mathrm{SS_E}}{\mathrm{SS_T}} = \frac{S_{xy}^2}{S_{xx} S_{yy}} \tag{105}$$

The coefficient R^2 expresses the proportion of the total variation in Y that is explained by the linear model.
By using R only, we can reexpress the statistic as

$$T_{n-2} = T_{n-2} = \frac{B_1}{S/\sqrt{S_{xx}}} = \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} \tag{106}$$

Test for Significance of Regression. Let $(x_i, Y_i), i = 1, \ldots, n$ be a random sample from $Y \mid x$. We reject

$$H_0 : \beta_1 = 0 \quad \text{(or regression not significant.)}$$

at significance level α if the statistic

$$F_{1,n-2} = (n-2)\frac{R^2}{1-R^2} = (n-2)\frac{\mathrm{SS_T} - \mathrm{SS_E}}{\mathrm{SS_E}}.$$

satisfies $F_{1,n-2} > f_{\alpha,1,n-2}$.
Test for Correlation. Let (X, Y) follow a bivariate normal distribution with correlation coefficient $\varrho \in (-1, 1)$. Let R be the estimator (24.1) for e. Then

$$H_0 : \varrho = 0$$

is rejected at significance level α if

$$\left| \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \right| > t_{\alpha/2,n-2}.$$

20.3 Lake-of-Fit Error

$\mathrm{SS_E}$ can be regared as composed of two factors: pure error $\mathrm{SS_{E;pe}}$ and the lack-of-fit $\mathrm{SS_{E;pe}}$.

$$\mathrm{SS_{E;pe}} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i) \quad \mathrm{SS_{E;lf}} = \mathrm{SS_E} - \mathrm{SS_{E;pe}} \tag{107}$$

Test for Lack of Fit. Let $x_1, \ldots, x_k$ be regressors and $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$, $i = 1, \ldots, k$, the measured responses at each of the regressors. Let $S_{E;pe}$ and $\mathrm{SS_{E;\,lf}}$ be the pure error and lack-of-fit sums of squares for a linear regression model. Then

$$H_0\text{: the linear regression model is appropriate}$$

is rejected at significance level α if the test statistic

$$F_{k-2,n-k} = \frac{\mathrm{SS_{E;If}}/(k-2)}{\mathrm{SS_{E;pe}}/(n-k)}$$

satisfies $F_{k-2,n-k} > f_{\alpha,k-2,n-k}$. In this formula, n represents the length of data, and k represents the number of groups.

21 Multiple Linear Regression

21.1 Sum of Squares Error

We have the decomposition

$$\mathrm{SS_T} = \mathrm{SS_R} + \mathrm{SS_E}$$

where
(i) $\mathrm{SS_T}$ represents the total variation of the response variable Y,
(ii) $\mathrm{SS_R}$ (called the regression sum of squares) represents the variation of the response predicted by the regression model and
(iii) $\mathrm{SS_E}$ represents the deviation of the response from the model.
Lemma. The regression sum of squares can be expressed as

$$\mathrm{SS_R} = \left\langle \boldsymbol{b}, X^T\boldsymbol{Y} \right\rangle - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2$$

In particular, in the case of the multilinear model,

$$\mathrm{SS_R} = b_0\sum_{i=1}^{n} Y_i + \sum_{j=1}^{p} b_j \sum_{i=1}^{n} x_{ji}Y_i - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2,$$

and in the polynomial model,

$$\mathrm{SS_R} = b_0\sum_{i=1}^{n} Y_i + \sum_{j=1}^{p} b_j \sum_{i=1}^{n} x_i^j Y_i - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2.$$

Analogously to (27.2), the coefficient of multiple determination,

$$R^2 = \frac{\mathrm{SS_R}}{\mathrm{SS_T}}$$

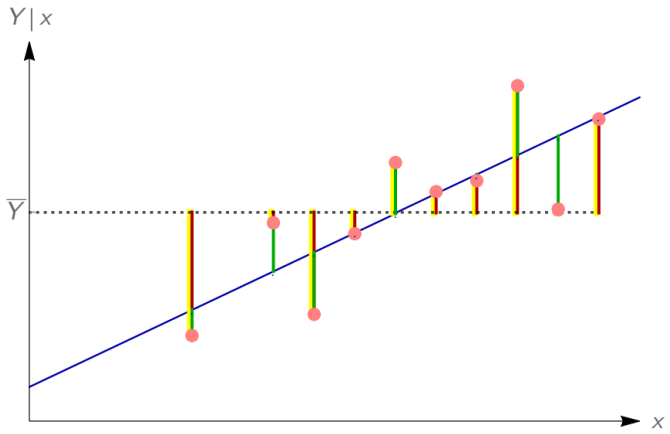


Figure 8: Yellow lines represent $\mathrm{SS_T}$, red lines represent $\mathrm{SS_E}$, and green lines represent $\mathrm{SS_R}$

gives the proportion of the response variation in Y explained by the model.

F-Test for Significance of Regression. Let $x_1, \ldots, x_p$ be the predictor variables in a multilinear model (28.1) for Y. Then

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0,$$

is rejected at significance level α if the test statistic

$$F_{p,n-p-1} = \frac{\mathrm{SS_R}/p}{\mathrm{SS_E}/(n-p-1)} = \frac{\mathrm{SS_R}/p}{S^2} = \frac{n-p-1}{p}\frac{R^2}{1-R^2}$$

satisfies $F_{p,n-p-1} > f_{\alpha,p,n-p-1}$.

21.2 Distribution of the Least-Squares Estimators

Let us write

$$\left(X^T X\right)^{-1} = \begin{pmatrix} \xi_{00} & * & \cdots & \cdots & * \\ * & \xi_{11} & \ddots & & * \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & * \\ * & \cdots & \cdots & * & \xi_{pp} \end{pmatrix}$$

where the starred values are uninteresting for us. Hence,

$$\mathrm{Var}\left[B_i\right] = \xi_{ii}\sigma^2, \quad i = 0, \ldots, p$$

We have therefore proved the following result: **Theorem**. The random vector $\boldsymbol{b}$ follows a normal distribution with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma^2\left(X^T X\right)^{-1}$. It is also possible to prove: **Theorem**. The statistic $(n-p-1)S^2/\sigma^2 = \mathrm{SS_E}/\sigma^2$ is independent of $\boldsymbol{b}$.

21.3 Confidence Intervals and Prediction Intervals

We immediately obtain the following $100(1-\alpha)\%$ confidence intervals for the model parameters:

$$\beta_j = b_j \pm t_{\alpha/2,n-p-1} S\sqrt{\xi_{jj}}, \quad j = 0, \ldots, p \tag{108}$$

Or by using the covariance matrix, we have

$$\beta_j = b_j \pm t_{\alpha/2,n-p-1}\sqrt{\text{Var}[B_j]}, \quad j = 0, \ldots, p \tag{109}$$

We have the following $100(1-\alpha)\%$ confidence interval for $\mu_{Y|x_0}$:

$$\mu_{Y|x_0} = \widehat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p-1} S\sqrt{x_0^T \left(X^T X\right)^{-1} x_0} \tag{110}$$

And $100(1-\alpha)\%$ prediction interval for $Y \mid x_0$:

$$Y \mid x_0 = \widehat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p-1} S\sqrt{1 + x_0^T \left(X^T X\right)^{-1} x_0} \tag{111}$$

Note that $p+1$ represents the number of parameters.

21.4 Model Sufficiency Test

For single parameter test

$$H_0 : \beta_j = 0$$

we use the statistic

$$T_{n-p-1} = \frac{b_j}{S\sqrt{\xi_{jj}}} = \frac{b_j}{\sqrt{\text{Var}[B_j]}} \tag{112}$$

Partial F-Test for Model Sufficiency. Let $x_1, \ldots, x_p$ be possible predictor variables for Y and (29.5) and (29.6) the full and reduced models, respectively. Then

$$H_0 : \text{the reduced model is sufficient}$$

is rejected at significance level α if the test statistic

$$F_{p-m,n-p-1} = \frac{n-p-1}{p-m} \frac{\text{SS}_{\text{E; reduced}} - \text{SS}_{\text{E;full}}}{\text{SS}_{\text{E;full}}}$$

satisfies $F_{p-m,n-p-1} > f_{\alpha,p-m,n-p-1}$.