A Brief Summary of Statistics Course

统计学课程知识总结

Vincent

2021 年 9 月 27 日

目录

1 概率论部分

Cover：Basic axioms, random events, σ-field; random variable/vector and their properties, some special distributions; $E\,\&\,\sigma^2\,\&\,cov$ and their properties; probability-generating/moment-generating/characteristic function; weak/strong law of large number, central limit thm.; intro. to multivariate normal distribution.

1.1 Some Important Distributions

X	$p_X(k)//f_X(x)$	E	σ^2	PGF	MGF
$B(p)$		p	pq		$q+pe^s$
$B(n,p)$	$C_n^k p^k(1-p)^{n-k}$	np	npq	$(q+ps)^n$	$(q+pe^s)^n$
$G(p)$	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{ps}{1-qs}$	$\frac{pe^s}{1-qe^s}$
$H(n,M,N)$	$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$	$n\frac{M}{N}$	$\frac{nM(N-n)(N-M)}{N^2(n-1)}$		
$P(\lambda)$	$\frac{\lambda^k}{k!}e^{-\lambda}$	λ	λ	$e^{\lambda(s-1)}$	$e^{\lambda(e^s-1)}$
$U(a,b)$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$		$\frac{e^{sb}-e^{sa}}{(b-a)s}$
$N(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2		$e^{\frac{\sigma^2 s^2}{2}+\mu s}$
$\epsilon(\lambda)$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$		$\frac{\lambda}{\lambda-s}$
$\Gamma(\alpha,\lambda)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$		
$B(\alpha,\beta)$	$\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$		
χ_n^2	$\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}$	n	$2n$		
t_ν	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})(1+\frac{x^2}{\nu})^{-\frac{\nu+1}{2}}}$	0	$\frac{\nu}{\nu-2}$		
$F(m,n)$	$\frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\frac{m^{\frac{m}{2}}x^{\frac{n}{2}}x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}$	$\frac{n}{n-2}$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$		

More Properties of χ^2, t, F see section **??**.

Definition of PGF, MGF, CF see section **??**.

1.2 Probability and Probability Model

What is **Probability**? A 'belief' in the chance of an event occurring.

1.2.1 Sample and σ-Field

Def. sample space Ω: The set of all possible outcomes of one particular experiment.

Def. $\mathscr{F}$ a σ-field(or a σ-algebra) as a collection of some subsets of Ω **if**

- $\Omega \in \mathscr{F}$

- if $A \in \mathscr{F}$,then $A^C \in \mathscr{F}$
- if $A_n \in \mathscr{F}$, then $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{F}$

And $(\Omega, \mathscr{F})$ is a measurable space.

1.2.2 Axioms of Probability

P is probability measure (or probability function) defined on $(\Omega, \mathscr{F})$, satisfying

- Nonnegativity

$$P(A) \geq 0 \qquad \forall A \in \Omega$$

- Normalization

$$P(\Omega) = 1$$

- Countable Additivity

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots \quad (A_i \parallel A_j \quad \forall i \neq j)$$

Then $(\Omega, \mathscr{F}, P)$ is probability space.

Properties of Probability:

- Monotonicity

$$P(A) \leq P(B) \quad \text{for } A \subset B$$

- Finite Subadditivity (Boole Inequality)

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

- Inclusion-Exclusion Formula

$$\begin{aligned} P(\Cup_{i=1}^{n} A_i) = & \sum_{1\leq i \leq n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ & + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \cdots \\ & + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n) \end{aligned}$$

- Borel-Cantelli Lemma

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(\lim_{n\to\infty} \sup A_n) = 0$$

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\lim_{n\to\infty} \sup A_n) = 1 \quad \text{if } A_i \text{ independent}$$

1.2.3 Conditional Probability

Def. **Conditional Probability** of B given A:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

(Actually a change of σ-field from Ω to B)

Application of conditional probability:

- Multiplication Formula

$$P(\Cap_{i=1}^{n} A_i) = P(A_1) \prod_{i=2}^{n} P(A_i | A_1 \cap A_2 \cap \cdots \cap A_{i-1})$$

- Total Probability Thm

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i)$$

  where $\{A_i\}$ is a partition of Ω.

- Bayes's Rule

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

  where $\{A_i\}$ is a partition of Ω.

- Statistically Independence

$$P(A \cap B) = P(A)P(B), \text{ for } A \parallel B$$

1.3 Properties of Random Variable and Vector

1.3.1 Random Variable

Def. Random Variable: a **function** X defined on sample space Ω, mapping from Ω to some $\mathscr{X} \in \mathbb{R}$.

Then def. Cumulative Distribution Function (CDF).

$$F_X(x) = P(X \leq x)$$

For Discrete case, consider CDF as right-continuity.

- PMF:

$$p_X(x) = F_X(x^+) - F_X(x^-)$$

  PDF

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$$

- Indicator function:

$$I_{x \in A}(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

- Convolution

- $W = X + Y$

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w-x)\mathrm{d}x$$

- $V = X - Y$

$$f_V(v) = \int_{-\infty}^{\infty} f_X(x) f_Y(x-v)\mathrm{d}x$$

- $Z = XY$

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(\frac{z}{x})\mathrm{d}x$$

- Order Statistics

  Def $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ as order statistics of $\vec{X}$

$$g_{X_{(i)}} = n! \prod f(x_i) \qquad \text{for}\, x_1 < x_2 \cdots < x_n$$

  PDF of $X_{(k)}$

$$g_k(x_k) = nC_{n-1}^{k-1}[F(x_k)]^{k-1}[1-F(x_k)]^{n-k} f(x_k)$$

- p-fractile

$$\xi_p = F^{-1}(p) = \inf\{x|F(x) \geq p\}$$

1.3.2 Random Vector

A general case of random variable.

n-dimension Random Vector $\vec{X} = (X_1, X_2, \ldots, X_n)$ defined on $(\Omega, \mathscr{F}, P)$.

CDF $F(x_1, \ldots, x_n)$ defined on $\mathbb{R}^n$:

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

Joint PDF of random vector:

$$f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$$

k-dimensional Marginal Distribution: For $1 \leq k < n$ and index set $S_k = \{i_1 \ldots, i_k\}$, distribution of $\vec{X} = (X_{i_1}, X_{i_2}, \ldots, X_{i_k})$

$$F_{S_k}(x_{i_1}, X_{i_2} \leq x_{i_2} \ldots, x_{i_k}) = P(X_{i_1} \leq x_{i_1}, \ldots, X_{i_k} \leq x_{i_k}; X_{i_{k+1}}, \ldots, X_{i_n} \leq \infty)$$

Marginal distribution:

$$g_{S_k}(x_{i_1}, \ldots, x_{i_k}) = \int_{\mathbb{R}^{n-k}} f(x_1, \ldots, x_n)\mathrm{d}x_{i_{k+1}} \ldots \mathrm{d}x_{j_n} = \frac{\partial^{n-k} F(x_1, \ldots, x_n)}{\partial x_{i_{k+1}} \ldots \partial x_{i_n}}$$

Δ **Function of r.v.**

For $\vec{X} = (X_1, X_2, \cdots, X_n)$ with PDF $f(\vec{X})$ and define

$$\vec{Y} = (Y_1, Y_2, \cdots, Y_n) = (y_1(\vec{X}), y_2(\vec{X}), \cdots, y_n(\vec{X}))$$

with inverse mapping

$$\vec{X} = (X_1, X_2, \cdots, X_n) = (x_1(\vec{Y}), x_2(\vec{Y}), \cdots, x_n(\vec{Y}))$$

then

$$g(\vec{Y}) = f(x_1(\vec{Y}), x_2(\vec{Y}), \cdots, x_n(\vec{Y})) \left|\frac{\partial \vec{X}}{\partial \vec{Y}}\right| I_{D_Y}$$

(Intuitively: $g(\vec{Y})\mathrm{d}\vec{Y} = \mathrm{d}P = f(\vec{X})\mathrm{d}\vec{X}$)

1.4 Properties of E , σ^2 and cov

Expectation and Variance of common distributions see sec.**??**.

1.4.1 Expection

Expectation of r.v. $g(X)$ def.:

$$E[g(X)] = \begin{cases} \int_\Omega g(x) f_X(x)\mathrm{d}x = \int_\Omega g(x)\mathrm{d}F(x) \\ \sum_\Omega g(X) f_X(x) \end{cases}$$

Properties of expectation $E(\cdot)$:

- Linearity of Expectation

$$E(aX + bY) = aE(X) + bE(Y)$$

- Conditional Expectation

$$E(X|A) = \frac{E(XI_A)}{P(A)}$$

  Note: if take A as Y is also a r.v. then

$$m(Y) = E(X|Y) = \int x f_{X|Y}(x)\mathrm{d}x$$

  is actually a function of Y

- Law of Total Expectation

$$E\{E[g(X)|Y]\} = E[g(X)]$$

- r.v.& Event

$$P(A|X) = E(I_A|X) \Rightarrow E[P(A|X)] = E(I_A) = P(A)$$

-

$$E[h(Y)g(X)|Y] = h(Y)E[g(X)|Y]$$

1.4.2 Variance

Variance of r.v. X:

$$var(X) = E[(X - E(X))^2] = E(X^2) - (E(X))^2$$

(sometimes denoted as σ_X^2.)

Properties:

- Linear combination of Variance

$$var(aX + b) = a^2 var(X)$$

- Conditional Variance

$$var(X|Y) = E[X - E(X|Y)]^2|Y$$

- Law of Total Variance

$$var(X) = E[var(X|Y)] + var[E(X|Y)]$$

Standard Deviation def. as :

$$\sigma_X = \sqrt{var(X)}$$

Then can construct **Standardization** of r.v.

$$Y + \frac{X - E(X)}{\sqrt{var(X)}}$$

1.4.3 Covariance and Correlation

Covariance of r.v. X and Y:

$$cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$$

And (Pearson's) Correlation Coefficient

$$\rho_{X,Y} = corr(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}$$

Remark: correlation $\nRightarrow$ cause and effect.

Properties:

- Bilinear of Covariance

$$cov(X + Y, Z) = cov(X, Z) + cov(Y, Z)$$
$$cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$$

- Variance and Covariance

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

- Covariance Matrix

  Def $\Sigma = E[(X-\mu)^T(X-\mu)] = \{\sigma_{ij}\}$

$$\Sigma = \begin{pmatrix} var(X_1) & cov(X_1,X_2) & \dots & cov(X_1,X_n) \\ cov(X_2,X_1) & var(X_2) & \dots & cov(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_n,X_1) & cov(X_n,X_2) & \dots & var(X_n) \end{pmatrix} \tag{1.1}$$

Attachment: Independence:

$$X_i||X_j \Rightarrow \begin{cases} f(x_1,x_2,\cdots,x_n) = \prod f(x_i) \\ F(x_1,x_2,\cdots,x_n) = \prod F(x_i) \\ E(\prod X_i) = \prod E(X_i) \\ var(\sum X_i) = \sum var(X_i) \end{cases}$$

1.5 PGF, MGF and C.F

Generating Function: Representation of P in function space. $P \Leftrightarrow$ Generating Function.

1.5.1 Probability Generating Function

PGF: used for non-negative, integer X

$$g(s) = E(s^X) = \sum_{j=0}^{\infty} s^j P(X=j), s \in [-1,1]$$

Properties

- $P(X=k) = \dfrac{g^{(k)}(0)}{k!}$
- $E(X) = g^{(1)}(1)$
- $var(X) = g^{(2)}(1) + g^{(1)}(1) - [g^{(1)}(1)]^2$
- For $X_1, X_2, \cdots, X_n$ independent with $g_i(s) = E(s^{X_i})$, $Y = \sum_{i=1}^{n} X_i$, then

$$g_Y(s) = \prod_{i=1}^{n} g_i(s), s \in [-1,1]$$

- For X_i i.i.d with $\psi(s) = E(s^{X_i})$, Y with $G(s) = E(s^Y)$, $W = X_1 + X_2 + \cdots + X_Y$,then

$$g_W(s) = G[\psi(s)]$$

- 2-Dimensional PGF of (X,Y)

$$g(s,t) = E(s^X t^Y) = \sum_{i=o}^{\infty}\sum_{j=0}^{\infty} P_{(X,Y)}(X=i,Y=j)s^i t^j,\ s,t \in [-1,1]$$

1.5.2 Moment Generating Function

MGF:

$$M_X(s) = E(e^{sX}) = \begin{cases} \sum_j e^{sx} P(X = x_j) \\ \int_{-\infty}^{\infty} e^{sx} f_X(x) \mathrm{d}x \end{cases}$$

Properties

- MGF of $Y = aX + b$: $M_Y(s) = e^{sb} M(sa)$
- $E(X^k) = M^{(k)}(0)$
- $P(X = 0) = \lim_{s \to -\infty} M(s)$
- For $X_1, X_2, \cdots, X_n$ independent with $M_{X_i}(s) = E(e^{sX_i})$, $Y = \sum_{i=1}^{n} X_i$, then

$$M_Y(s) = \prod_{i=1}^{n} M_{X_i}(s)$$

1.5.3 Characteristic Function

C.F is actually the Fourier Transform of f.

$$\phi(t) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f_X(x) \mathrm{d}x$$

Properties

- if $E(|X|^k) < \infty$,then

$$\phi^{(k)}(t) = i^k E(X^k e^{itX}) \qquad \phi^{(k)}(0) = i^k E(X^k)$$

- For $X_1, X_2, \cdots, X_n$ independent with $\phi_{X_i}(t) = E(e^{itX_i})$, $Y = \sum_{i=1}^{n} X_i$, then

$$\phi_Y(t) = \prod_{i=1}^{n} \phi_{X_i}(t)$$

- Inverse (Fourier) Transform

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) \mathrm{d}t$$

1.6 Convergence and Limit Distribution

1.6.1 Convergence Mode

$$\begin{cases} \text{Convergence in Distribution} & X_n \xrightarrow{\mathscr{L}} X : \lim_{n \to \infty} F_n(x) = F(x) \\ \text{Convergence in Probability} & X_n \xrightarrow{p} X : \lim_{n \to \infty} P(|X_n - X) \geq \varepsilon) = 0\,, \forall \varepsilon > 0 \\ \text{Almost Sure Convergence} & X_n \xrightarrow{\text{a.s.}} X : P(\lim_{n \to \infty} X_n = X) = 1 \\ L_p \text{ Convergence} & X_n \xrightarrow{L_p} X : \lim_{n \to \infty} E(|X_n - X|^p) = 0 \end{cases}$$

Relations between convergence:

$$
\begin{tikzcd}
X_n \xrightarrow{\text{a.s.}} X \arrow[dr] & & \\
& X_n \xrightarrow{p} X \arrow[r] & X_n \xrightarrow{\mathscr{L}} X \\
X_n \xrightarrow{L_p} X \arrow[ur] & &
\end{tikzcd}
$$

Useful Thm.:

- Continuous Mapping Thm.: For continuous function $g(\cdot)$
    1. $X_n \xrightarrow{\text{a.s.}} X \Rightarrow g(X_n) \xrightarrow{\text{a.s.}} g(X)$
    2. $X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$
    3. $X_n \xrightarrow{\mathscr{L}} X \Rightarrow g(X_n) \xrightarrow{\mathscr{L}} g(X)$
- Slutsky's Thm.: For $X_n \xrightarrow{\mathscr{L}} X, Y_n \xrightarrow{p} c$
    1. $X_n + Y_n \xrightarrow{\mathscr{L}} X + c$
    2. $X_n Y_n \xrightarrow{\mathscr{L}} cX$
    3. $X_n / Y_n \xrightarrow{\mathscr{L}} X/c$
- Continuity Thm.

$$\lim_{n\to\infty} \phi_n(t) = \varphi(t) \Leftrightarrow X_n \xrightarrow{\mathscr{L}} X$$

1.6.2 Law of Large Number & Central Limit Theorem

- WLLN

$$\frac{1}{n}\sum X_i \xrightarrow{p} E(X_1)$$

- SLLN

$$\frac{1}{n}\sum X_i \xrightarrow{\text{a.s.}} C$$

- CLT

$$\frac{1}{\sigma\sqrt{n}}\sum(X_k - \mu) \xrightarrow{\mathscr{L}} N(0,1)$$

- de Moivre-Laplace Thm.

$$P(k \le S_n \le m) \approx \Phi(\frac{m + 0.5 - np}{\sqrt{npq}}) - \Phi(\frac{k - 0.5 - np}{\sqrt{npq}})$$

- Stirling Eqa

$$\frac{\lambda^k}{k!}e^{-\lambda} \approx \frac{1}{\sqrt{\lambda}\sqrt{2\pi}}e^{-\frac{(k-\lambda)^2}{2\lambda}} \xrightarrow[\lambda=n]{k=n} n! \approx \sqrt{2\pi n}(\frac{n}{e})^n$$

1.7 Inequalities

- Cauchy-Schwarz Inequality

$$|E(XY)| \le \sqrt{E(X^2)E(Y^2)}$$

- Bonferroni Inequality

$$P(\bigcup_{i=1}^{n} A_i) \ge \sum_{1\le i\le n} P(A_i) + \sum_{1\le i<j\le n} P(A_i \cap A_j)$$

- Markov Inequality

$$P(|X| \ge \epsilon) \le \frac{E(|X|^\alpha)}{\epsilon^\alpha}$$

- Chebyshev Inequality

$$P(|X - E(X)| \ge \epsilon) \le \frac{var(X)}{\epsilon^2}$$

- Jensen Inequality: For convex function $g(x)$:

$$E[g(X)] \ge g(E(X))$$

1.8 Multivariate Normal Distribution

For $X_1, X_2, \cdots, X_n$ independent and $X_k \sim N(\mu_k, \sigma_k^2)$, $k = 1, \cdots, n$, $T = \sum_{k=1}^{n} c_k X_k$, ($c_k$ const), then

$$T \sim N(\sum_{k=1}^{n} c_k\mu_k, \sum_{k=1}^{n} c_k^2\sigma_k^2)$$

Deduction in some special cases:

- Given $\mu_1 = \mu_2 = \cdots = \mu_n = \mu$, $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$, i.e. X_k i.i.d., then

$$T \sim N(\mu \sum_{k=1}^{n} c_k, \sigma^2 \sum_{k=1}^{n} c_k^2)$$

- Further take $c_1 = c_2 = \cdots = c_n = \frac{1}{n}$, i.e. $T = \sum_{k=1}^{n} X_k/n = \bar{X}$, then

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

1.8.1 Linear Transform

First consider $\epsilon_1, \epsilon_2, \cdots, \epsilon_m$ i.i.d. $\sim N(0, 1)$, $n \times 1$ const column vector $\vec{\mu}$, $n \times m$ const matrix $\mathbf{B} = \{b_{ij}\}$, def.$X_i = \sum_{j=1}^{m} b_{ij}\epsilon_j$, i.e.

$$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{pmatrix}$$

We have: $\vec{X} \sim N(\vec{\mu}, \Sigma)$, where Σ, as defined in eqa.**??** is

$$\Sigma = \mathbf{B}\mathbf{B}^T = \begin{pmatrix} var(X_1) & cov(X_1, X_2) & \dots & cov(X_1, X_n) \\ cov(X_2, X_1) & var(X_2) & \dots & cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_n, X_1) & cov(X_n, X_2) & \dots & var(X_n) \end{pmatrix} = \{\sigma_{ij}\}$$

Furthur Consider $\vec{Y} = (Y_1, \cdots, Y_n)^T$, $n \times n$ const square matrix $\mathbf{A} = \{a_{ij}\}$ and def. $\vec{Y} = \mathbf{A}\vec{X}$ i.e.

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

Then $\vec{Y} \sim N(\mathbf{A}\vec{\mu}, \mathbf{A}\Sigma\mathbf{A}^T)$

Special case: $X_1, \cdots, X_n$ i.i.d. $\sim N(\mu, \sigma^2)$, $\vec{X} = (X_1, \cdots, X_n)^T$,

$$\begin{aligned} E(Y_i) =& \mu \sum_{k=1}^n a_{ik} \\ var(Y_i) =& \sigma^2 \sum_{k=1}^n a_{ik}^2 \\ cov(Y_i, Y_j) =& \sigma^2 \sum_{k=1}^n a_{ik} a_{jk} \end{aligned}$$

Specially when $\mathbf{A} = \{a_{ij}\}$ orthonormal, we have $Y_1, \cdots, Y_n$ independent

$$Y_i \sim N(\mu \sum_{k=1}^n a_{ik}, \sigma^2)$$

1.8.2 Distributions of Function of Normal Variable: χ^2, t & F

Consider $X_1, X_2, \dots, X_n$ i.i.d. $\sim N(0,1)$; $Y, Y_1, Y_2, \dots, Y_m$ i.i.d. $\sim N(0,1)$

- χ^2 Distribution: Def. χ^2 distribution with degree of freedom n:

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

  PDF of χ_n^2:

$$g_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2} e^{-x/2} I_{x>0}$$

  Properties

  - E and var of $\xi \sim \chi_n^2$

$$E(\xi) = n \qquad var(\xi) = 2n$$

- For independent $\xi_i \sim \chi^2_{n_i}$, $i = 1, 2, \ldots, k$:

$$\xi_0 = \sum_{i=1}^{k} \xi_i \sim \chi^2_{n_1+\ldots+n_k}$$

- Denoted as $\Gamma(\alpha, \lambda)$:

$$\xi = \sum_{i=1}^{n} X_i \sim \Gamma(\frac{n}{2}, \frac{1}{2}) = \chi^2_n$$

- t Distribution: Def. t distribution with degree of freedom n:

$$T = \frac{Y}{\sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}}} = \frac{Y}{\sqrt{\frac{\xi}{n}}} \sim t_n$$

(Usually take ν instead of n)

PDF of t_ν:

$$t_\nu(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Denote: Upper α-fractile of t_ν, satisfies $P(T \geq c) = \alpha$:

$$c = t_{\nu,\alpha}$$

(Similar for χ^2_n and $F_{m,n}$ etc.)

- F Distribution: Def. F distribution with degree of freedom m and n:

$$F = \frac{\sum_{i=1}^{m} Y_i}{\sum_{i=1}^{n} X_i} =\sim F_{m,n}$$

PDF of $F_{m,n}$:

$$f_{m,n}(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1} (n + mx)^{-\frac{m+n}{2}} I_{x>0}$$

Properties

- If $Z \sim F_{m,n}$, then $\frac{1}{Z} \sim F_{n,m}$.
- If $T \sim t_n$, then $T^2 \sim F_{1,n}$
- $F_{m,n,1-\alpha} = \frac{1}{F_{n,m,\alpha}}$

- Some useful Lemma (uesd in statistic inference):

- For $X_1, X_2, \ldots, X_n$ independent with $X_i \sim N(\mu_i, \sigma_i^2)$, then

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2_n$$

- For $X_1, X_2, \ldots, X_n$ i.i.d.$\sim N(\mu, \sigma^2)$, then

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$$

  For $X_1, X_2, \ldots, X_m$ i.i.d.$\sim N(\mu_1, \sigma^2)$, $Y_1, Y_2, \ldots, Y_n$ i.i.d.$\sim N(\mu_2, \sigma^2)$, denote $S_\omega^2 = \dfrac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$, then

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_\omega} \cdot \sqrt{\frac{mn}{m+n}} \sim t_{m+n-2}$$

- For $X_1, X_2, \ldots, X_m$ i.i.d.$\sim N(\mu, \sigma^2)$, $Y_1, Y_2, \ldots, Y_n$ i.i.d.$\sim N(\mu_2, \sigma^2)$, then

$$T = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F_{m-1,n-1}$$

- For $X_1, X_2, \ldots, X_n$ i.i.d. $\sim \epsilon(\lambda)$, then

$$2\lambda n\bar{X} = 2\lambda\sum_{i=1}^{n} X_i \sim \chi^2_{2n}$$

  Remark: for $X_i \sim \epsilon(\lambda) = \Gamma(1, \lambda) \Rightarrow 2\lambda\sum_{i=1}^n X_i \sim \Gamma(n, 1/2) = \chi^2_{2n}$.

2 统计推断部分

Statistical Inference: use sample to estimate population.

Two main tasks of Statistical Inference:

- Parameter Estimation
  - Point Estimation: **??**
  - Interval Estimation: **??**
- Hypothesis Testing: **??**

2.1 Statistical Model and Statistics

Random sample comes from population X. In parametric model case, we have population distribution family:

$$\mathscr{F} = \{f(x;\vec{\theta})|\vec{\theta} \in \Theta\}$$

where parameter $\vec{\theta}$ reflect some quantities of population (e.g. mean, variance, etc.), each $\vec{\theta}$ corresponds to a distribution of population X.

Sample space: Def. as $\mathscr{X} = \{\{x_1, x_2, \dots, x_n\}, \forall x_i\}$, then $\{X_i\} \in \mathscr{X}$ is random sample from population $X \sim f(x;\vec{\theta})$.

2.1.1 Statistics

Statistic(s): function of random sample $\vec{T}(X_1, X_2, \dots, X_n)$, **but not a function of parameter**.

Some useful statistics, e.g.

- Sample mean (Consider X_i i.i.d.)

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

- Sample variance

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

- Sample moments
  - Origin moment

$$a_{n,k} = \frac{1}{n}\sum_{i=1}^{k} X_i^k \qquad k = 1, 2, 3, \dots$$

  - Center moment

$$m_{n,k} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k \qquad k = 2, 3, 4, \dots$$

- Order statistics

$$(X_{(1)}, X_{(2)}, \dots, X_{(n)}), \text{ for } X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$$

- Sample p-fractile

$$m_p = X_{(m)}, \quad m = [(n+1)p]$$

- Sample coefficient of variation

$$\hat{\nu} = \frac{S}{\bar{X}}$$

- Skewness and Kurtosis

$$\hat{\beta}_1 = \frac{m_{n,3}}{m_{n,2}^{3/2}} \qquad \hat{\beta}_2 = \frac{m_{n,4}}{m_{n,2}^2}$$

Properties

Statistic T is a function of random sample $\{X_i\}$, thus has distribution (say $g_T(t)$) called **Sampling Distribution**.

For X_i i.i.d. from $X \sim f(x)$ with population mean μ and variance σ^2

- Calculation of S^2

$$(n-1)S^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

- E and var of $\bar{X}$ and S^2

$$E(\bar{X}) = \mu \qquad var(\bar{X}) = \frac{\sigma^2}{n} \qquad E(S^2) = \sigma^2$$

Further if X_i i.i.d. from $X \sim N(\mu, \sigma^2)$ where μ and σ^2 unknown.

- Independence of $\bar{X}$ and S^2

$$\bar{X} \text{ and } S^2 \text{independent}$$

- Distribution of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

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

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

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

2.1.2 Exponential Family

Def. $\mathscr{F} = \{f(x;\vec{\theta}|\theta \in \Theta)\}$ is **Exponential Family** if $f(x;\vec{\theta})$ has the form as

$$f(x;\vec{\theta}) = C(\theta)h(x)\exp\left[\sum_{i=1}^{k} Q_i(\theta)T_i(x)\right] \quad \vec{\theta} \in \Theta$$

Canonical Form: Take $Q_i(\theta) = \varphi_i$, then $\vec{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_k)$ =$(Q_1(\theta), Q_2(\theta), \ldots, Q_k(\theta))$ is a transform from Θ to Θ^*, s.t. $\mathscr{F}$ has canonical form, i.e.

$$f(x;\vec{\varphi}) = C^*(\vec{\varphi})h(x)\exp\left[\sum_{i=1}^{k} \varphi T_i(x)\right] \quad \vec{\varphi} \in \Theta^*$$

Θ^* is canonical parameter space.

- Why we need exponential family? Have some nice properties.

2.1.3 Sufficient and Complete Statistics

- A **Sufficient Statistic** $T(\vec{X})$ for $\vec{\theta}$ contains all the information of sample when infer $\vec{\theta}$, i.e.

$$f(\vec{X}; T(\vec{X}) = f(\vec{X}; T(\vec{X}), \vec{\theta})$$

  Properties

  - **Factorization Thm.** $T(\vec{X})$ is sufficient **if and only if** $f_{\vec{X}}(\vec{x}; \vec{\theta}) = f(\vec{x}; \vec{\theta})$ can be written as

$$f(\vec{x}; \vec{\theta}) = g[t(\vec{x}); \vec{\theta}]h(\vec{x})$$

  - If $T(\vec{X})$ sufficient, then $T'(\vec{X}) = g[T(\vec{X})]$ also.(require g single-valued and invertible)
  - If $T(\vec{X})$ sufficient, then (T, T_1) also.
  - Minimal sufficient statistic $T_\theta(\vec{X})$ satisfies

$$\forall \text{sufficient statistic } S, \exists\, q_S(\cdot), \text{ s.t.} T_\theta = q_S(S)$$

    A minimal sufficient statistic not always exists.

    Sufficient & Complete $\Rightarrow$ Minimal sufficient.

  - Usually dimension of $\vec{T_\theta}$ and $\vec{\theta}$ equals.

  Sufficient statistic is not unique.

- A **Complete Statistic** $T(\vec{X})$ for $\vec{\theta}$ satisfies

$$\forall \vec{\theta} \in \Theta\,;\, \forall \varphi \text{ satisfies } E[\varphi(T(\vec{X}))] = 0, \text{ we have } P[\varphi(T) = 0; \vec{\theta}] = 1$$

  Explanation: $T \sim g_T(t)$. Rewrite as

$$\int \varphi(t) g_T(t)\,\mathrm{d}t = 0\ \forall \vec{\theta} \Rightarrow \varphi(T) = 0 \text{ a.s.}$$

  i.e. span$\{g_T(t); \forall \vec{\theta}\}$ is a complete sapce. Or to say that $\nexists$ none-zero $\varphi(t)$ so that $E(\varphi(T)) = 0$ (unbiased estimation)

$$\varphi(T) \neq 0\ \forall \vec{\theta} \Rightarrow E[\varphi(T(\vec{X}))] \neq 0$$

  So make sure the uniqueness of unbiased estimation of $\hat{\theta}$ using T.

  Properties

  - If $T(\vec{X})$ complete, then $T'(\vec{X}) = g[T(\vec{X})]$ also.(require g measurable)
  - A complete statistic not always exists.

- An **Ancillary Statistic** $S(\vec{X})$ is a statistic whose distribution does not depend on $\vec{\theta}$

  Basu Thm: $\vec{X} = (X_1, X_2, \ldots, X_n)$ is sample from $\mathscr{F} = \{f(x;\theta), \theta \in \Theta\}$. $T(\vec{X})$ is a complete and minimal sufficient statistic, $S(\vec{X})$ is ancillary statistic, then $S(\vec{X}) \parallel T(\vec{X})$.

- Exponential family: For $\vec{X} = (X_1, X_2, \ldots, X_n)$ from exponential family with canonical form, i.e.

$$f(\vec{x};\vec{\theta}) = C(\vec{\theta})h(\vec{x})\exp\left[\sum_{i=1}^{k}\theta_i T_i(\vec{x})\right], \quad \vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta$$

  Then if $\Theta \in \mathbb{R}^k$ interior point exists, then $T(\vec{X}) = (T_1(\vec{X}), T_2(\vec{X}), \ldots, T_k(\vec{X}))$ is sufficient & complete statistic.

2.2 Point Estimation

For parametric distribution family $\mathscr{F} = \{f(x, \vec{\theta}), \vec{\theta} \in \Theta\}$, random sample $\vec{X} = (X_1, X_2, \ldots, X_n)$ from $\mathscr{F}$. $g(\vec{\theta})$ is a function defined on Θ.

Mission: use sample $\{X_i\}$ to estimate $g(\vec{\theta})$, called **Parameter Estimation**.

$$\text{Parameter Estimation} \begin{cases} \text{Point Estimation} & \surd \\ \text{Interval Estimation} \end{cases}$$

Point estimation: when estimating θ or $g(\theta)$, denote the estimator (defined on sample space $\mathscr{X}$) as

$$\hat{\theta}(\vec{X}) \qquad \hat{g}(\vec{X})$$

Estimator is a statistic, with sampling distribution.

2.2.1 Optimal Criterion

Some nice properties of estimators (that we expect)

- Unbiasedness

$$E(\hat{\theta}) = \theta \quad \text{or} \quad E(\hat{g}(\vec{X})) = g(\theta)$$

  Otherwise, say $\hat{\theta}$ or $\hat{g}$ biased. Def. **Bias**: $E(\hat{\theta}) - \theta$

  Asymptotically unbiasedness

$$\lim_{n\to\infty} E(\hat{g}(\vec{X})) = g(\theta)$$

- Efficiency: say $\hat{g}_1(\vec{X})$ is more efficient than $\hat{g}_2(\vec{X})$, if

$$var(\hat{g}_1) \leq var(\hat{g}_2) \quad \forall \theta \in \Theta$$

- Mean Squared Error (MSE)

$$\text{MSE} = E[(\hat{\theta}-\theta)^2] = var(\hat{\theta}) + [Bias(\hat{\theta})]^2$$

For unbiased estimator, i.e. $Bias(\hat{\theta}) = 0$, we have

$$\text{MSE} = E[(\hat{\theta}-\theta)^2] = var(\hat{\theta})$$

- Consistency

$$\lim_{n\to\infty} P(|\hat{g}_n(\vec{X}) - g(\theta)| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0$$

- Asymptotic Normality

2.2.2 Method of Moments

Review: Population moments & Sample moments

$$\alpha_k = E(X^k) \qquad \mu_k = E[(X-E(X))^k]$$
$$a_{n,k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad m_{n,k} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

Property: $a_{n,k}$ is the unbiased estimator of α_k.(while $m_{n,k}$ unually biased for μ_k)

For sample $\vec{X} = (X_1, X_2, \ldots, X_n)$ from $\mathscr{F} = \{f(x;\theta, \theta \in \Theta)\}$, unknown parameter (or its function) $g(\theta)$ can be written as

$$g(\theta) = G(\alpha_1, \alpha_2, \ldots, \alpha_k; \mu_2, \mu_3, \ldots, \mu_l)$$

Then its **Moment Estimate** $\hat{g}(\vec{X})$ is

$$\hat{g}(\vec{X}) = G(a_{n,1}, a_{n,2}, \ldots, a_{n,k}; m_{n,2}, m_{n,3}, \ldots, m_{n,l})$$

Example: coefficient of variance & skewness

$$\hat{\nu} = \frac{S}{\bar{X}} \quad \hat{\beta}_1 = \frac{m_{n,3}}{m_{n,2^{3/2}}} = \sqrt{n}\frac{\sum_{i=1}^{n}(X_i-\bar{X})^3}{[\sum_{i=1}^{n}(X_i-\bar{X})^2]^{\frac{3}{2}}}$$

Note:

- G may not have explicit expression.
- Moment estimate may not be unique.
- If $G = \sum_{i=1}^{k} c_i\alpha_i$ (linear combination of α, without μ), then $\hat{g}(\vec{X}) = \sum_{i=1}^{k} c_i a_{n,i}$ unbiased.

  Usually $\hat{g}(\vec{X})$ is asymptotically unbiased.
- For small sample, not so accurate.
- May not contain all the information about $\vec{\theta}$, i.e. may not be sufficient statistic.
- Do not require a statistic model.

2.2.3 Maximum Likelihood Estimation

For sample $\vec{X} = (X_1, X_2, \ldots, X_n)$ with distribution $f(\vec{x};\vec{\theta})$ from $\mathscr{F} = \{f(x;\vec{\theta}), \vec{\theta} \in \Theta\}$, def. **Likelihood Function** $L(\vec{\theta};\vec{x})$, defined on Θ (as a function of $\vec{\theta}$)

$$L(\vec{\theta};\vec{x}) = f(\vec{x};\vec{\theta}) \qquad \vec{\theta} \in \Theta, \vec{x} \in \mathscr{X}$$

Also def. log-likelihood function $l(\vec{\theta};\vec{x}) = \ln L(\vec{\theta};\vec{x})$.

If estimator $\hat{\theta} = \hat{\theta}(\vec{X})$ satisfies

$$L(\hat{\theta};\vec{x}) = \sup_{\vec{\theta}\in\Theta} L(\vec{\theta};\vec{x}), \quad \vec{x} \in \mathscr{X}$$

Or equivalently take $l(\vec{\theta};\vec{x})$ instead of $L(\vec{\theta};\vec{x})$.

Then $\hat{\theta} = \hat{\theta}(\vec{X})$ is a **Maximum Likelihood Estimate**(MLE) of $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$

How to identify MLE?

- Differentiation: Fermat Lemma

$$\left.\frac{\partial L}{\partial \theta_i}\right|_{\vec{\theta}=\hat{\theta}} = 0 \qquad \left.\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right|_{\vec{\theta}=\hat{\theta}} \text{negative definite} \qquad \forall i, j = 1, 2, \ldots, k$$

- Graphing method.
- Numerically compute maximum.

Note: Some properties of MLE

- (Depend on the case, not always) unbiased.
- Invariance of MLE: If $\hat{\theta}$ is MLE of $\vec{\theta}$, invertible function $g(\vec{\theta})$, then $g(\hat{\theta})$ is MLE of $g(\vec{\theta})$.
- MLE and Sufficiency: $T = T(X_1, X_2, \ldots, X_n)$ is a sufficient statistic of $\vec{\theta}$, if MLE of $\vec{\theta}$ exists, say $\hat{\theta}$, then $\hat{\theta}$ is a function of T, i.e.

$$\hat{\theta} = \hat{\theta}(\vec{X}) = \hat{\theta}^*(T(\vec{X}))$$

- Asymptotic Normality:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma_\theta^2), \quad \sigma_\theta^2 = \frac{1}{E_\theta[\frac{\partial}{\partial\theta}\ln f(\vec{X};\theta)]^2}$$

  i.e.

$$\hat{\theta}_n \xrightarrow{d} N(\theta, \frac{\sigma_\theta^2}{n})$$

Comparison: MoM and MLE

- MoM do not require statistic model; MLE need to know PDF.
- MoM is more robust than MLE.

MLE in Exponential Family:

For sample $\vec{X}=(X_1,X_2,\dots,X_n)$ from canonical exponential family $\mathscr{F}=\{f(x;\vec{\theta}),\vec{\theta}\in\Theta\}$

$$f(x;\vec{\theta})=C(\vec{\theta})h(x)\exp\left[\sum_{i=1}^{k}\theta_i T_i(x)\right]\qquad \vec{\theta}=(\theta_1,\dots,\theta_k)\in\Theta$$

Likelihood function $L(\vec{\theta},\vec{x})=\prod_{j=i}^{n}f(x_j;\theta)$ and log-likelihood function $l(\vec{\theta},\vec{x})$

$$L(\vec{\theta},\vec{x})=C^n(\vec{\theta})\prod_{j=1}^{n}h(x_j)\exp\left[\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}T_i(x_j)\right]$$
$$l(\vec{\theta},\vec{x})=n\ln C(\vec{\theta})+\sum_{j=1}^{n}\ln h(x_j)+\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}T_i(x_j)$$

Solution of MLE: (Require $\hat{\theta}\in\Theta$)

$$\frac{n}{C(\vec{\theta})}\frac{\partial C(\vec{\theta})}{\partial\theta_i}\bigg|_{\vec{\theta}=\hat{\theta}}=-\sum_{j=1}^{n}T_i(x_j),\quad i=1,2,\dots,k$$

2.2.4 Uniformly Minimum Variance Unbiased Estimate

MSE: For $\hat{g}(\vec{X})$ is estimate of $g(\vec{\theta})$,then MSE

$$\text{MSE}(\hat{g}(\vec{X}))=E[(\hat{g}(\vec{X})-g(\vec{\theta}))^2]=var(\hat{g})+[Bias(\hat{g})]^2$$

- Unbiased estimator (i.e. $Bias(\hat{g})=0$) not unique; not always exist.

Now only consider unbiased estimators of $g(\vec{\theta})$ exists, say $\hat{g}(\vec{X})$, then

$$\text{MSE}(\hat{g}(\vec{X}))=var(\hat{g}(\vec{X}))$$

If $\forall$ unbiased estimate $\hat{g}\prime(\vec{X})$, $\hat{g}$ satisfies

$$var[\hat{g}(\vec{X})]\le var[\hat{g}\prime(\vec{X})]$$

Then $\hat{g}(\vec{X})$ is **Uniformly Minimum Variance Unbiased Estimate(UMVUE)** of $g(\vec{\theta})$

How to determine UMVUE? (Not an easy task)

- Zero Unbiased Estimate Method
- Sufficient and Complete Statistic Method
- Cramer-Rao Inequality

1. **Zero Unbiased Estimate Method**

   Let $\hat{g}(\vec{X})$ be an unbiased estimate with $var(\hat{g})<\infty$. If $\forall$ $E(\hat{l}(\vec{X}))=0$, $\hat{g}$ holds that

   $$cov(\hat{g},\hat{l})=E(\hat{g}\cdot\hat{l})=0,\quad\forall\theta\in\Theta$$

   Then $\hat{g}$ is a UMVUE of $g(\vec{\theta})$ (sufficient & necessary).

2. **Sufficient and Complete Statistic Method**

   For $T(\vec{X})$ sufficient statistic, $\hat{g}(\vec{X})$ unbiased estimate of $g(\vec{\theta})$, then

   $$h(T) = E(\hat{g}(\vec{X})|T)$$

   is an unbiased estimate of $g(\vec{\theta})$ and $var(h(T)) \leq var(\hat{g})$.

   Remark:

   - A method to improve estimator.
   - A UMVUE has to be a function of sufficient statistic.

   Lehmann-Scheffé Thm.: For $\vec{X} = (X_1, X_2, \ldots, X_n)$ from population $X \sim \mathscr{F} = \{f(x, \vec{\theta}, \theta \in \vec{\Theta})\}$. $T(\vec{X})$ sufficient and complete, and $\hat{g}(T(\vec{X}))$ be an unbiased estimator, then $\hat{g}(T(\vec{X}))$ is the unique UMVUE.

   Can be used to construct UMVUE: given $T(\vec{X})$ sufficient and complete and some unbiased estimator $\hat{g}\prime(\vec{\theta})$ then

   $$\hat{g}(T) = E(\hat{g}\prime|T)$$

   is the unique UMVUE.

3. Cramer-Rao Inequality

   Core idea: determine a lower bound of $var(\hat{g})$.

   Consider $\vec{\theta} = \theta$ (One dimension parameter); For $\{X_i\}$ i.i.d. $f(x, \theta)$: def.

   - **Score function**: Reflects the steepness of likelihood function f.

     $$S(\vec{x}; \theta) = \frac{\partial \ln f(\vec{x}; \theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i; \theta)}{\partial \theta}$$

     $$E[S(\vec{X}; \theta)] = 0$$

   - **Fisher Information**: Variance of $S(\vec{x}; \theta)$, reflects the accuracy to conduct estimation, i.e. reflects information of statistic model.

     $$I(\theta) = E\left[\left(\frac{\partial \ln f(\vec{x}; \theta)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln f(\vec{x}; \theta)}{\partial \theta^2}\right]$$

   Consider $\mathscr{F}$ satisfies some regularity conditions(in most cases, regularity conditions do hold), then the lower bound of $var(\hat{g})$ satisfies **Cramer-Rao Inequality**:

   $$var(\hat{g}(\vec{X})) \geq \frac{[g'(\theta)]^2}{nI(\theta)}$$

   Special case: $g(\theta) = \theta$ then

   $$var(\hat{\theta}) \geq \frac{1}{nI(\theta)}$$

   note:

- C-R Inequality determine a lower bound, not the infimum(i.e. UMVUE$\nRightarrow var(\hat{g}(\vec{X})) = \frac{[g'(\theta)]^2}{nI(\theta)}$).
- Take '=': Only some cases in Exponential family.
- **Efficiency**: How good the estimator is.

$$e_{\hat{g}(\vec{X})}(\theta) = \frac{[g'(\theta)]^2/(nI(\theta))}{var(\hat{g}(\vec{X}))}$$

4. **Multi-Dimensional Cramer-Rao Inequality**

ReDef. Fisher Information:

$$\mathbf{I}(\vec{\theta}) = \{I_{ij}(\vec{\theta})\} = \{E\left[\left(\frac{\partial \ln f(\vec{x};\theta)}{\partial \theta_i}\right)\left(\frac{\partial \ln f(\vec{x};\theta)}{\partial \theta_j}\right)\right]\}$$

Then covariance matrix $\Sigma(\vec{\theta})$ satisfies **Cramer-Rao Inequality**

$$\Sigma(\vec{\theta}) \geq (n\mathbf{I}(\vec{\theta}))^{-1}$$

Note: '$\geq$' holds for all diagonal elements, i.e.

$$var(\hat{\theta}_i) \geq \frac{I^*_{ii}(\vec{\theta})}{n}, \quad \forall\, i = 1, 2, \dots, k$$

2.2.5 MoM and MLE in Linear Regression

- Linear Regression Model(1-dimension case):

$$y_i = \beta_0 + \beta_1 x_0 + \epsilon_i$$

where β_0, β_1 are regression coefficient, and ϵ_i are unknown random **error**. Assume:

$$\begin{aligned} &\epsilon_i \text{ are i.i.d.} \\ &E(\epsilon_i|x_i) = 0 \\ &var(\epsilon_i) = \sigma^2 \end{aligned}$$

Mission: use data $\{(x_i, y_i)\}$ to estimate β_0, β_1(i.e. regression line), and error ϵ_i.

1. OLS: Take β_0, β_1 so that MSE min, i.e.

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Solution:

$$\begin{cases} \hat{\beta}_0 &= \bar{y} - \beta_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \{\sum_{i=1}^n (x_i - \bar{x})^2\}^{-1}\{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})\} \end{cases}$$

So get regression line:$y = \hat{\beta}_0 + \hat{\beta}_1 x$

Def. Residuals

$$\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

Residuals can be used to estimate ϵ_i: $E[(\epsilon_i)^2] = \sigma^2$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

2. MoM: Consider r.v. $\epsilon \sim f(\varepsilon; x, y, \beta_0, \beta_1)$, sample $\{\epsilon_i | \epsilon_i = y_i - \beta_0 - \beta_1 x_i\}$, then obviously

$$\bar{\epsilon} = \bar{y} - \beta_0 - \beta_1 \bar{x}$$

Take moment estimate of ϵ, we have

$$E(\epsilon_i) = 0 \qquad E(\epsilon_i x_i) = 0(\text{note that})E(\epsilon|x) = 0$$

$$\text{i.e.} \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0 \\ \frac{1}{n}\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) = 0 \end{cases}$$

Solution:

$$\begin{cases} \hat{\beta}_0 &= \bar{y} - \beta_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \end{cases}$$

(Same as OLS)

Moment estimate of σ^2

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

3. MLE: Assume $\epsilon_i \sim N(0, \sigma^2)$, then $y_i | x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$. Get likelihood function:

$$L(\beta_0, \beta_1, \sigma^2; x_1, \ldots, x_n, y_1, \ldots, y_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)}{2\sigma^2}\right]$$

Take differentiation, also get the same result.

- Linear Regression Model(Multi-dimension case):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i$$

Denote: $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$, $\vec{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})$, then for each i: $y_i = x_i^T \beta + \epsilon_i$

Further denote: Matrix form:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} = X\vec{\beta} + \vec{\epsilon}$$

Basic Assumptions: Gauss-Markov Assumptions

- OLS unbiased

$$E(\epsilon_i|x_i) = 0 \qquad E(y_i|x_i) = x_i^T\beta$$

- Homogeneity of ϵ_i

$$var(\epsilon_i) = \sigma^2$$

- Independent of ϵ

- (For MLE) ϵ_i i.i.d. $\sim N(0, \sigma^2)$

Residuals:

$$\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - x_i^T\beta$$

Def. Residual Sum of Squares (RSS)

$$\text{RSS} = \sum_{i=1}^{n} \hat{\epsilon}_i^2 = \sum_{i=1}^{n} (y_i - x_i^T\beta)^2$$

Estimator exists and unique:($\hat{\sigma}^2$ after bias correction)

$$\begin{aligned}
\hat{\beta} &= (X^TX)^{-1}(X^TY) \\
\hat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T\hat{\beta})^2 \\
\hat{\sigma}^2 &= \frac{1}{n-p-1}\sum_{i=1}^{n}(y_i - x_i^T\hat{\beta})^2
\end{aligned}$$

2.2.6 Kernel Density Estimate

Given random sample $\{X_i\}$. Def. Empirical Distribution Function

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i) \tag{2.1}$$

Problem: Overfitting when getting $\hat{f}$. Solution: Using **Kernel Estimate**, replace $I_{(-\infty,x]}(\cdot)$ with Kernel function $K(\cdot)$, then

$$\hat{f}_n(x) = \frac{F_n(x+h_n) - F - n(x-h_n)}{2h_n} = \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{x - X_i}{h_n})$$

where h_n is **bandwidth**. Take proper kernel function K to get estimate of f.

Can be considered as a convolution of sample $\{X_i\}$ and kernel function K.

Useful Kernel Functions:

- $K(x) = \dfrac{1}{2} I_{[-\frac{1}{2},\frac{1}{2}]}$

- $K(x) = (1 - |x|)I_{[-1,1]}$

- $K(x) = \dfrac{1}{2\pi} e^{-\frac{x^2}{2}}$

- $K(x) = \dfrac{1}{\pi(1+x^2)}$
- $K(x) = \dfrac{1}{2\pi}\text{sinc}^2(\dfrac{x}{2})$

2.3 Interval Estimation

$$\text{Parameter Estimation}\begin{cases}\text{Point Estimation} \\ \text{Interval Estimation} & \surd\end{cases}$$

Interval Estimation: to estimate $g(\vec{\theta})$, give **two** estimators $\hat{g}_1(\vec{X})$, $\hat{g}_2(\vec{X})$ defined on $\mathscr{X}$ as the two ends of interval (i.e. give an interval $[\hat{g}_1(\vec{X}),\ \hat{g}_2(\vec{X})]$), then random interval $[\hat{g}_1(\vec{X}),\ \hat{g}_2(\vec{X})]$ is an **Interval Estimation** of $g(\theta)$.

2.3.1 Confidence Interval

How to judge an interval estimation?

- Reliability
$$P(g(\theta) \in [\hat{g}_1, \hat{g}_2])$$
- Precision
$$E(\hat{g}_2 - \hat{g}_1)$$

Trade off: (in most cases)

Given a level of reliability, find an interval with the highest precision above the level

- For a given $0 < \alpha < 1$, if
$$P(\hat{g}_1 \le g(\vec{\theta}) \le \hat{g}_2) \ge 1 - \alpha$$

then $[\hat{g}_1, \hat{g}_2]$ is a **Confidence Interval** for $g(\vec{\theta})$, with **Confidence Level** $1-\alpha$.

Confidence Coefficient:
$$\inf_{\forall\theta\in\Theta} P(\vec{\theta} \in \Theta)$$

Other cases:

- **Confidence Limit**:Upper/Lower Confidence Limit
$$P(g \le \hat{g}_U) \ge 1 - \alpha$$
$$P(\hat{g}_L \le \theta) \ge 1 - \alpha$$
- **Confidence Region**: For high dimensional parameters $\vec{g} = (g_1, g_2, \ldots, g_k)$
$$P(\vec{g} \in S(\vec{X})) \ge 1 - \alpha \quad \forall\vec{\theta} \in \Theta$$

Mission: Determine $\hat{g}_1, \hat{g}_2$.

2.3.2 Pivot Variable Method

Idea: Based on point estimation, construct a new variable and thus find the interval estimation.

Def. **Pivot Variable** T, satisfies:

- Expression of T contains θ (thus T is not a statistic).
- Distribution of T independent of θ.

In different cases, construct different pivot variable, usually base on sufficient statistics and transform.

Knowing a proper pivot variable $T = T(\hat{\varphi}, g(\theta)) \sim f$, ($f$ is some distribution independent of $\vec{\theta}$), $\hat{\varphi}$ is a sufficient statistic), then we can take T satisfies:

$$P(f_{1-\frac{\alpha}{2}} \leq T \leq f_{\frac{\alpha}{2}}) = 1-\alpha$$

Construct the inverse mapping of $T = T(\hat{\varphi}, g(\theta)) \rightleftarrows g(\theta) = T^{-1}(T, \hat{\varphi})$, we get

$$P[T^{-1}(f_{1-\frac{\alpha}{2}}, \hat{\varphi}) \leq \hat{g} \leq T^{-1}(f_{\frac{\alpha}{2}}, \hat{\varphi})] = 1-\alpha$$

Thus get a confidence interval for θ with confidence coefficient $1-\alpha$.

2.3.3 Confidence Interval for Common Distributions

Some important properties of χ^2, t and F see section **??**.

1. Single normal population: $\vec{X} = \{X_1, X_2, \ldots, X_n\} \in \mathscr{X}$ i.i.d from Normal Distribution population $N(\mu, \sigma^2)$. Denote sample mean and sample variance:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad S_\mu = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2, \text{(for } \mu \text{ known)}$$

Estimating $\mu \,\&\, \sigma^2$: construction of pivot variable under different circumstances:

Estimation	Pivot Variable	Confidence Interval
σ^2 known, estimate μ	$T = \dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$	$\left[\bar{X} - \dfrac{\sigma}{\sqrt{n}}N_{\frac{\alpha}{2}}, \bar{X} + \dfrac{\sigma}{\sqrt{n}}N_{\frac{\alpha}{2}}\right]$
σ^2 unknown, estimate μ	$T = \dfrac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$	$\left[\bar{X} - \dfrac{S}{\sqrt{n}}t_{n-1,\frac{\alpha}{2}}, \bar{X} + \dfrac{S}{\sqrt{n}}t_{n-1,\frac{\alpha}{2}}\right]$
μ known, estimate σ^2	$T = \dfrac{nS_\mu^2}{\sigma^2} \sim \chi_n^2$	$\left[\dfrac{nS_\mu^2}{\chi^2_{n,\frac{\alpha}{2}}}, \dfrac{nS_\mu^2}{\chi^2_{n,1-\frac{\alpha}{2}}}\right]$
μ unknown, estimate σ^2	$T = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$	$\left[\dfrac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}, \dfrac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right]$

2. Double normal population: $\vec{X} = \{X_1, X_2, \ldots, X_m\}$ i.i.d. from $N(\mu_1, \sigma_1^2)$; $\vec{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ i.i.d. from $N(\mu_2, \sigma_2^2)$

   Denote sample mean, sample variance and pooled sample variance:

$$
\begin{aligned}
&\bar{X} = \frac{1}{m}\sum_{i=1}^{n} X_i \qquad && S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2 \qquad && S_{\mu_1}^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i - \mu_1)^2, (\mu_1 \text{ known})\\
&\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i && S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 && S_{\mu_2}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_2)^2, (\mu_2 \text{ known})\\
& && S_\omega^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2} &&
\end{aligned}
$$

   Estimating $\mu_1 - \mu_2$:

   When $\sigma_1^2 \neq \sigma_2^2$ unknown, estimate $\mu_1 - \mu_2$: Behrens-Fisher Problem, remain unsolved, but can deal with simplified cases.

Estimation	Pivot Variable	Confidence Interval
σ_1^2 & σ_2^2 known, estimate $\mu_1 - \mu_2$	$T = \dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} \sim N(0,1)$	$\left[\bar{X} - \bar{Y} - N_{\frac{\alpha}{2}}\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}, \bar{X} - \bar{Y} + N_{\frac{\alpha}{2}}\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}\right]$
$\sigma_1^2 = \sigma_2^2$ unknown, estimate $\mu_1 - \mu_2$	$T = \dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_\omega} \sim t_{m+n-2}$	$\left[\bar{X} - \bar{Y} - S_\omega t_{m+n-2,\frac{\alpha}{2}}\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}, \bar{X} - \bar{Y} + S_\omega t_{m+n-2,\frac{\alpha}{2}}\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}\right]$
Welch's t-Interval (when m, n large enough)	$T = \dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_X^2}{m} + \dfrac{S_Y^2}{n}}} \xrightarrow{\mathscr{L}} N(0,1)$	$\left[\bar{X} - \bar{Y} - N_{\frac{\alpha}{2}}\sqrt{\dfrac{S_1^2}{m} + \dfrac{S_2^2}{n}}, \bar{X} - \bar{Y} + N_{\frac{\alpha}{2}}\sqrt{\dfrac{S_1^2}{m} + \dfrac{S_2^2}{n}}\right]$

   Estimating $\dfrac{\sigma_1^2}{\sigma_2^2}$:

Estimation	Pivot Variable	Confidence Interval
μ_1, μ_2 known, estimate $\dfrac{\sigma_1^2}{\sigma_2^2}$	$T = \dfrac{S_{\mu_2}^2}{S_{\mu_1}^2}\dfrac{\sigma_1^2}{\sigma_2^2} \sim F_{n,m}$	$\left[\dfrac{S_{\mu_1}^2}{S_{\mu_2}^2}\dfrac{1}{F_{m,n,\frac{\alpha}{2}}}, \dfrac{S_{\mu_1}^2}{S_{\mu_2}^2}\dfrac{1}{F_{m,n,1-\frac{\alpha}{2}}}\right]$ or $\left[\dfrac{S_{\mu_1}^2}{S_{\mu_2}^2}F_{m,n,\frac{\alpha}{2}}, \dfrac{S_{\mu_1}^2}{S_{\mu_2}^2}F_{n,m,\frac{\alpha}{2}}\right]$
μ_1, μ_2 unknown, estimate $\dfrac{\sigma_1^2}{\sigma_2^2}$	$T = \dfrac{S_Y^2}{S_X^2}\dfrac{\sigma_1^2}{\sigma_2^2} \sim F_{n-1,m-1}$	$\left[\dfrac{S_X^2}{S_Y^2}\dfrac{1}{F_{m-1,n-1,\frac{\alpha}{2}}}, \dfrac{S_X^2}{S_Y^2}\dfrac{1}{F_{m-1,n-1,1-\frac{\alpha}{2}}}\right]$ or $\left[\dfrac{S_X^2}{S_Y^2}\dfrac{1}{F_{m-1,n-1,\frac{\alpha}{2}}}, \dfrac{S_X^2}{S_Y^2}F_{n-1,m-1,\frac{\alpha}{2}}\right]$

3. Non-normal population:

Estimation	Pivot Variable	Confidence Interval
Uniform Distribution: $\vec{X}$ i.i.d. from $U(0,\theta)$	$T=\frac{X_{(n)}}{\theta}\sim U(0,1)$	$\left[X_{(n)},\frac{X_{(n)}}{\sqrt[n]{\alpha}}\right]$
Exponential Distribution: $\vec{X}$ i.i.d. from $\epsilon(\lambda)$	$T=2n\lambda\bar{X}\sim\chi^2_{2n}$	$\left[\frac{\chi^2_{2n,1-\frac{\alpha}{2}}}{2n\bar{X}},\frac{\chi^2_{2n,\frac{\alpha}{2}}}{2n\bar{X}}\right]$
Bernoulli Distribution: $\vec{X}$ i.i.d. from $B(1,\theta)$	$T=\frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\bar{X}(1-\bar{X})}}\xrightarrow{\mathscr{L}}N(0,1)$	$\left[\bar{X}-N_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}},\bar{X}+N_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]$
Poisson Distribution: $\vec{X}$ i.i.d. from $P(\lambda)$	$T=\frac{\sqrt{n}(\bar{X}-\lambda)}{\sqrt{\bar{X}}}\xrightarrow{\mathscr{L}}N(0,1)$	$\left[\bar{X}-N_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}},\bar{X}+N_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}\right]$

4. General Case: Use asymptotic normality of MLE to construct CLT for large sample. MLE of θ satisfies:

$$\sqrt{n}(\hat{\theta}^*-\theta)\xrightarrow{\mathscr{L}}N(0,\frac{1}{I(\theta)})$$

where $\hat{\theta}^*$ is MLE of θ. Replace $\frac{1}{I(\theta)}$ by $\sigma^2(\hat{\theta}^*)$, then

$$T=\frac{\sqrt{n}(\hat{\theta}^*-\theta)}{\sigma(\hat{\theta}^*)}\xrightarrow{\mathscr{L}}N(0,1)$$

confidence interval:

$$\left[\hat{\theta}^*-\frac{N_{\frac{\alpha}{2}}}{\sqrt{n}}\sigma(\hat{\theta}^*),\hat{\theta}^*+\frac{N_{\frac{\alpha}{2}}}{\sqrt{n}}\sigma(\hat{\theta}^*)\right]$$

2.3.4 Fisher Fiducial Argument*

Idea: When sample is known, we can get '**Fiducial Probability**' of θ, thus can find an interval estimation based on fiducial distribution.(Similar to the idea of MLE)

Remark: Fiducial probability (denoted as $\tilde{P}(\theta)$) is 'probability of parameter', in the case that sample is known. **Fiducial probability is different from Probability**.

Thus get

$$\tilde{P}(\hat{g}_1\leq g(\theta)\leq\hat{g}_2)=1-\alpha$$

2.4 Hypothesis Testing

Hypothesis is a statement about the characteristic of population, e.g. distribution form, parameters, etc.

Mission: Use sample to test the hypothesis, i.e. judge whether population has some characteristic.

2.4.1 Basic Concepts

Parametric hypothesis testing.

For random sample $\vec{X} = (X_1, X_2, \ldots, X_n) \in \mathscr{X}$ i.i.d. from $\mathscr{F} = \{f(x;\theta); \theta \in \Theta\}$

- Null Hypothesis H_0 & Alternative Hypothesis H_1: Wonder whether a statement is true. Def. **Null Hypothesis**: $H_0 : \theta \in \Theta_0 \subset \Theta$, **a statement that we try to reject based on sample**; $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$ is **Alternative Hypothesis**.

  Thus Hypothesis Testing:

  $$H_0 : \theta \in \Theta_0 \longleftrightarrow H_1 : \theta \in \Theta_1$$

- Rejection Region R & Acceptance Region R^C: Judge whether to reject H_0 from sample, Def. **Rejection Region**:

  $$R \subset \mathscr{X}: \text{ reject } H_0 \text{ if } \vec{X} \in R$$

  Acceptance Region: accept H_0 if $\vec{X} \in R^C$

- Test Function: Describe how to make a decision.

  - Continuous Case:

    $$\varphi(\vec{X}) = \begin{cases} 1, & \vec{X} \in R \\ 0, & \vec{X} \in R^C \end{cases}$$

    i.e. $R = \{\vec{X} : \varphi(\vec{X}) = 1\}$. Where R to be determined.

  - Discrete Case: Randomized Test Function

    $$\varphi(\vec{X}) = \begin{cases} 1, & \vec{X} \in R - \partial R \\ r, & \vec{X} \in \partial R \\ 0, & \vec{X} \in R^C \end{cases}$$

    Where R and r to be determined.

- Type I Error & Type II Error: Sample is random, possible to make a wrong judge.

  - Type I Error (弃真): H_0 is true but sample falls in R, thus H_0 is rejected.

    $$P(\text{type I error}) = P(\vec{X} \in R | H_0) = \alpha(\theta)$$

  - Type II Error (取伪): H_0 is wrong but sample falls in R^C, thus H_0 is accepted.

    $$P(\text{type II error}) = P(\vec{X} \notin R | H_1) = \beta(\theta)$$

  Impossible to make probability of Type I & II Error small simultaneously, how to pick a proper test $\varphi(\vec{x})$?

  Neyman-Pearson Principle: First control $\alpha \leq \alpha_0$, then take $\min \beta$.

  How to determine α_0? Depend on problem.[1]

[1] In most cases, take $\alpha_0 = 0.05$.

		Judgement	
		Accept H_0	Reject H_0
Real Case	H_0	$\surd$	Type I Error
	H_1	Type II Error	$\surd$

- p-value: probability to get larger bias than observed $\vec{x}_0$ under H_0.

  For reject region $R = \{\vec{X}|T(\vec{X}) \geq C\}$, p-value:

  $$p(\vec{x}) = P[T(\vec{X}) \geq t(\vec{x}_0)|H_0]$$

  Remark: Under H_0, the probability to get a worse result than $\vec{x}_0$.

  Rule: Reject H_0 if $p(\vec{x}_0) \leq \alpha_0$

- Power Function: (when H_0 is given), probability to reject H_0 by sampling.

  $$\pi(\theta) = \begin{cases} P(\text{type I error}), & \theta \in \Theta_0 \\ 1 - P(\text{type II error}), & \theta \in \Theta_1 \end{cases} = \begin{cases} \alpha(\theta), & \theta \in \Theta_0 \\ 1 - \beta(\theta), & \theta \in \Theta_1 \end{cases}$$

  Express as test function:

  $$\pi(\theta) = E[\varphi(\vec{X})|\theta]$$

  A nice test: $\pi(\theta)$ small under H_0, large under H_1.

Δ **General Steps of Hypothesis Testing:**

1. Propose H_0 & H_1.
2. Determine R (usually in the form of a statistic, e.g. $R = \{\vec{X} : T(\vec{X}) \geq c\}$).
3. Select a proper α (to determine c).
4. Sampling, get sample (as well as $t(\vec{x})$), compare with R and determine whether to reject/accept H_0

2.4.2 Hypothesis Testing of Common Distributions

For some common distribution populations, determine rejection region R under certain H_0 with confidence coefficient α.

Definition of necessary statistics see section **??**.

1. Single normal population:

Condition	H_0	H_1	Testing Statistic T	Rejection Region R
σ^2 known, test μ	$\mu = \mu_0$ $\mu \le \mu_0$ $\mu \ge \mu_0$	$\mu \ne \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$T = \dfrac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0,1)$	$\|T\| > N_{\frac{\alpha}{2}}$ $T > N_\alpha$ $T < -N_\alpha$
σ^2 unknown, test μ	$\mu = \mu_0$ $\mu \le \mu_0$ $\mu \ge \mu_0$	$\mu \ne \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$T = \dfrac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{n-1}$	$\|T\| > t_{n-1,\frac{\alpha}{2}}$ $T > t_{n-1,\alpha}$ $T < -t_{n-1,\alpha}$
μ known, test σ^2	$\sigma^2 = \sigma_0^2$ $\sigma^2 \le \sigma_0^2$ $\sigma^2 \ge \sigma_0^2$	$\sigma^2 \ne \sigma_0^2$ $\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$	$T = \dfrac{nS_\mu^2}{\sigma_0^2} \sim \chi_n^2$	$T < \chi^2_{n,1-\frac{\alpha}{2}} \cup T > \chi^2_{n,\frac{\alpha}{2}}$ $T > \chi^2_{n,\alpha}$ $T < \chi^2_{n,1-\alpha}$
μ unknown, test σ^2	$\sigma^2 = \sigma_0^2$ $\sigma^2 \le \sigma_0^2$ $\sigma^2 \ge \sigma_0^2$	$\sigma^2 \ne \sigma_0^2$ $\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$	$T = \dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2$	$T < \chi^2_{n-1,1-\frac{\alpha}{2}} \cup T > \chi^2_{n-1,\frac{\alpha}{2}}$ $T > \chi^2_{n-1,\alpha}$ $T < \chi^2_{n-1,1-\alpha}$

2. Double normal population:

Condition	H_0	H_1	Testing Statistic T	Rejection Region R
σ_1^2, σ_2^2 known, test $\mu_1 - \mu_2$	$\mu_1 - \mu_2 = \mu_0$ $\mu_1 - \mu_2 \le \mu_0$ $\mu_1 - \mu_2 \ge \mu_0$	$\mu_1 - \mu_2 \ne \mu_0$ $\mu_1 - \mu_2 > \mu_0$ $\mu_1 - \mu_2 < \mu_0$	$T = \dfrac{\bar{X} - \bar{Y} - \mu_0}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} \sim N(0,1)$	$\|T\| > N_{\frac{\alpha}{2}}$ $T > N_\alpha$ $T < -N_\alpha$
σ_1^2, σ_2^2 unknown, test $\mu_1 - \mu_2$	$\mu_1 - \mu_2 = \mu_0$ $\mu_1 - \mu_2 \le \mu_0$ $\mu_1 - \mu_2 \ge \mu_0$	$\mu_1 - \mu_2 \ne \mu_0$ $\mu_1 - \mu_2 > \mu_0$ $\mu_1 - \mu_2 < \mu_0$	$T = \dfrac{\bar{X} - \bar{Y} - \mu_0}{S_\omega}\sqrt{\dfrac{mn}{m+n}}$ $\sim t_{m+n-2}$	$\|T\| > t_{m+n-2,\frac{\alpha}{2}}$ $T > t_{m+n-2,\alpha}$ $T < -t_{m+n-2,\alpha}$
μ_1, μ_2 known, test $\dfrac{\sigma_1^2}{\sigma_2^2}$	$\sigma_1^2 = \sigma_2^2$ $\sigma_1^2 \ge \sigma_2^2$ $\sigma_1^2 \le \sigma_2^2$	$\sigma_1^2 \ne \sigma_2^2$ $\sigma_1^2 < \sigma_2^2$ $\sigma_1^2 > \sigma_2^2$	$T = \dfrac{S_{\mu_2}^2}{S_{\mu_1}^2} \sim F_{n,m}$	$T < F_{n,m,1-\frac{\alpha}{2}}$ $\cup T > F_{n,m,\frac{\alpha}{2}}$ $T > F_{n,m,\alpha}$ $T < F_{n,m,1-\alpha}$
μ_1, μ_2 unknown, test $\dfrac{\sigma_1^2}{\sigma_2^2}$	$\sigma_1^2 = \sigma_2^2$ $\sigma_1^2 \ge \sigma_2^2$ $\sigma_1^2 \le \sigma_2^2$	$\sigma_1^2 \ne \sigma_2^2$ $\sigma_1^2 < \sigma_2^2$ $\sigma_1^2 > \sigma_2^2$	$T = \dfrac{S_2^2}{S_2^2} \sim F_{n-1,m-1}$	$T < F_{n-1,m-1,1-\frac{\alpha}{2}}$ $\cup T > F_{n-1,m-1,\frac{\alpha}{2}}$ $T > F_{n-1,m-1,\alpha}$ $T < F_{n-1,m-1,1-\alpha}$

3. None normal population:

Condition	H_0	H_1	Testing Statistic T	Rejection Region R
$\vec{X}$ from $B(1,p)$, test p	$p=p_0$	$p\neq p_0$	$T=\dfrac{\sqrt{n}(\bar{X}-p_0)}{\sqrt{p_0(1-p_0)}}\xrightarrow{\mathscr{L}}N(0,1)$	$\lvert T\rvert>N_{\frac{\alpha}{2}}$
$\vec{X}$ from $P(\lambda)$, test λ	$\lambda=\lambda_0$	$\lambda\neq\lambda_0$	$T=\dfrac{\sqrt{n}(\bar{X}-\lambda_0)}{\sqrt{\lambda_0}}\xrightarrow{\mathscr{L}}N(0,1)$	$\lvert T\rvert>N_{\frac{\alpha}{2}}$

2.4.3 Likelihood Ratio Test

Idea: To test $H_0:\theta\in\Theta_0\longleftrightarrow H_1:\theta\in\Theta_1$ known $\vec{x}$, examine the likelihood function $L(\theta;\vec{x})$ and **compare** $L_{\theta\in\Theta_0}$ and $L_{\theta\in\Theta}$ to see the likelihood that H_0 is true.

Def. **Likelihood Ratio** (LR):

$$\lambda(\vec{x})=\frac{\sup\limits_{\theta\in\Theta_0}L(\theta;\vec{x})}{\sup\limits_{\theta\in\Theta}L(\theta;\vec{x})}$$

Reject H_0 if $\lambda(\vec{x})<\lambda_0$. Or equivalently

Reject H_0 if $-2\ln\lambda(\vec{x})>C(=-2\ln\lambda_0)$.

where λ_0 (or equivalently $C=-2\ln\lambda_0$) satisfies:

$$E_{\Theta_0}[\varphi(\vec{X})]\leq\alpha,\quad\forall\theta\in\Theta_0$$

LR and sufficient statistic: $\lambda(\vec{x})$ can be expressed as $\lambda(\vec{x})=\lambda^*(T(\vec{x}))$, where $T(\vec{X})$ is sufficient statistic.

- Limiting Distribution of LR: Wilks' Thm.

If $\dim\Theta=k>\dim\operatorname{span}\{\Theta_0\}=s$[2], then under $H_0:\theta\in\Theta_0$:

$$\Lambda_{\theta\in\Theta_0}(\vec{x})=-2\ln\lambda(\vec{x})\xrightarrow{\mathscr{L}}\chi^2_{k-s}$$

2.4.4 Uniformly Most Powerful Test

Idea: Neyman-Pearson Principle: control α, find $\min\beta$. i.e. control α, find $\max\pi(\theta)$

Def. **Uniformly Most Powerful Test** (UMP) φ_{UMP} with level of significance α satisfies

$$\pi_{\text{UMP}}(\theta)\geq\pi(\theta),\ \forall\theta\in\Theta_1$$

Neyman-Pearson Lemma: For $\vec{X}=(X_1,X_2,\dots,X_n)$ i.i.d. from $f(\vec{x};\theta)$.

Test hypothesis $H_0:\theta=\theta_0\longleftrightarrow H_1:\theta=\theta_1$. Def. test function φ as:

$$\varphi(\vec{x})=\begin{cases}1, & \dfrac{f(\vec{x};\theta_1)}{f(\vec{x};\theta_0)}>C\\ r, & \dfrac{f(\vec{x};\theta_1)}{f(\vec{x};\theta_0)}=C\\ 0, & \dfrac{f(\vec{x};\theta_1)}{f(\vec{x};\theta_0)}<C\end{cases}\tag{2.2}$$

Then there exists C and r such that

[2]Here 'dimension' refers to 'degree of freedom'.

- $E[\varphi(\vec{x})|\theta_0] = P(\frac{f(\vec{x};\theta_1)}{f(\vec{x};\theta_0)} > C) + rP(\frac{f(\vec{x};\theta_1)}{f(\vec{x};\theta_0)} = C) = \alpha$
- This φ is UMP of level of significance α

Actually kind of 1-dimensional case of LRT.

Note: UMT exist for **simple** H_0, H_1, otherwise may not exist.

UMP and sufficient statistics: Test function $\varphi(\vec{X}$ given by eqa.**??** is function of sufficient statistics $T(\vec{X})$, i.e. $\varphi(\vec{X}) = \varphi^*(T(\vec{X}))$.

UMP and Exponential Family: For sample $\vec{X} = (X_1, X_2, \dots, X_n)$ from exponential family:

$$f(\vec{x};\theta) = C(\theta)h(\vec{x})\exp\{Q(\theta)T(\vec{x})\}$$

Test single hypothesis $H_0 : \theta = \theta_0 \longleftrightarrow H_1 : \theta = \theta_1$. If

- θ_0 is inner point of Θ
- $Q(\theta)$ monotone increase with θ

Then UMP exists, in the form of:

$$\varphi(\vec{x}) = \begin{cases} 1, & T(\vec{x}) > C \\ r, & T(\vec{x}) = C \\ 0, & T(\vec{x}) < C \end{cases} \tag{2.3}$$

where C and r satisfies $E[\varphi(\vec{x})|\theta_0] = \alpha$.

Note: or take $Q(\theta)$ mono decreased, then in eqa.**??**, take opposite inequality operators.

Δ **General Steps of UMP**:

1. Find a point $\theta_0 \in \Theta_0$ and a point $\theta_1 \in \Theta_1$. (Note: **one** point)
2. Construct test function in the form of eqa.**??**, use $E[\varphi(\vec{x})|\theta_0] = \alpha$ to determine C and r.
3. Get R and $\varphi(\vec{x})$.
4. If φ does **not** depend on θ_1, then H_1 can be generalized to $H_1 : \theta \in \Theta_1$.
5. If φ satisfies $E_{\theta\in\Theta_0}(\varphi) \leq \alpha$, then H_0 an be generalized to $H_0 : \theta \in \Theta_0$.

2.4.5 Duality of Hypothesis Testing and Interval Estimation

- Thm.: $\forall\theta_0 \in \Theta$ there exists hypothesis testing $H_0 : \theta = \theta_0 \longleftrightarrow H_1 : \theta \neq \theta_0$ of level α with rejection region R_{θ_0}. Then

  $$C(\vec{X}) = \{\theta : \vec{X} \in R_\theta^C\}$$

  is a $1-\alpha$ confidence region for θ

- Thm.: $C(\vec{X})$ is a $1-\alpha$ confidence region for θ. Then $\forall\theta_0 \in C(\vec{X})$, the rejection region of hypothesis testing $H_0 : \theta = \theta_0 \longleftrightarrow H_1 : \theta \neq \theta_0$ of level α satisfies

$$R_{\theta_0}^C = \{\vec{X} : \theta_0 \in C(\vec{X})\}$$

Idea:

$$H_0 : \theta = \theta_0 \longleftrightarrow H_1 : \theta \neq \theta_0$$

$$\updownarrow$$

$$P(R^C(\vec{X})|H_0) = P(R^C(\vec{X})|\theta_0) = 1-\alpha$$

$$\updownarrow$$

$$\text{Confidence Interval: } \theta_0 \in R^C(\vec{X})$$

Similar for Confidence Limit and One-Sided Testing.

2.4.6 Introduction to Non-Parametric Hypothesis Testing

Motivation: Usually distribution form unknown, cannot use parametric hypothesis testing.

Useful Method:

- Sign Test: Used for paired comparison $\vec{X} = (X_1, X_2, \ldots, X_n$, $\vec{Y} = (Y_1, Y_2, \ldots, Y_n)$.

  Take $Z_i = Y_i - X_i$ i.i.d., denote $E(Z) = \mu$. Test $H_0 : \mu = 0 \longleftrightarrow H_1 : \mu \neq 0$.

  Denote $n_+ = \#(\text{positive } Z_i)$ and $n_- = \#(\text{negative } Z_i)$, $n_0 = n_+ + n_-$. Then $n_+ \sim B(n_0, \theta)$, test $H_0 : \theta = \dfrac{1}{2} \longleftrightarrow H_1 : \theta \neq \dfrac{1}{2}$

  Then use Binomial Testing or large sample CLT Normal Testing.

  Remark:

  – Also can test $H_0 : \theta \leq \dfrac{1}{2} \longleftrightarrow H_1 : \theta > \dfrac{1}{2}$

  – Drawback: ignores magnitudes.

- Wilcoxon Signed Rank Sum Test: Improvement of Sign Test. Base on order statistics.

  Order Statistics of Z_i: $Z_{(1)} < Z_{(2)} < \ldots < Z_{(n)}$, where each $Z_{(j)}$ corresponds to some Z_i, denote as $Z_i = Z_{(R_i)}$, then R_i is the rank of Z_i.[3]

  Def. $\vec{R} = (R_1, R_2, \ldots, R_n)$ is **Rank Statistics** of $(Z_1, Z_2, \ldots, Z_n)$

  Def. **Sum of Wilcoxon Signed Rank**:

$$W^+ = \sum_{i=1}^{n_0} R_i I_{Z_i > 0}$$

[3] If some $X_i, X_j, \ldots$ equal, then take same rank $R = \text{mean}\{R_i, R_j, \ldots\}$.

Distribution of W^+ is complex. E and var of W^+ under H_0:

$$E(W^+) = \frac{n_0(n_0+1)}{4} \qquad var(W^+) = \frac{n_0(n_0+1)(2n_0+1)}{24}$$

Usually consider large sample CLT, construct normal approximation:

$$T = \frac{W^+ - E(W^+)}{\sqrt{var(W^+)}} \xrightarrow{\mathscr{L}} N(0,1)$$

Rejection Region: $R = \{|T| > N_{\frac{\alpha}{2}}\}$

- Wilcoxon Two-Sample Rank Sum Test: Used for two independent sample comparison.

  Assume $\vec{X} = (X_1, \ldots, X_m)$ i.i.d. $\sim f(x)$; $\vec{Y} = (Y_1, \ldots, Y_n)$ i.i.d. $\sim f(x-\theta)$, test $H_0 : \theta = 0 \longleftrightarrow H_1 : \theta \neq 0$.

  Rank X_i and Y_i as:

  $$Z_1 \leq Z_2 \leq \ldots \leq Z_{m+n}$$

  in which denote rank of Y_i as R_i, and def. **Wilcoxon two-sample rank sum**:

  $$W = \sum_{i=1}^{n} R_i$$

  E and var of W under H_0:

  $$E(W) = \frac{n(m+n+1)}{2} \qquad var(W) = \frac{mn(n+m+1)}{12}$$

  Use large sample approximation, construct CLT:

  $$T = \frac{W - E(W)}{\sqrt{var(W)}} \xrightarrow{\mathscr{L}} N(0,1)$$

- Goodness-of-Fit Test: For $\vec{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. from some certain population X. Test $H_0 : X \sim F(x)$.

  where F is theoretical distribution, can be either parametric or non-parametric.

  Idea: Define some *quantity* $D = D(X_1, \ldots, X_n; F)$ to measure the difference between F and sample. And def. *Goodness-of-fit* when observed value of D (say d_0) is given:

  $$p(d_0) = P(D \geq d_0 | H_0)$$

  Goodness-of-Fit Test: Reject H_0 if $p(d_0) < \alpha$.

  Pearson χ^2 Test: Usually used for discrete case.

  Test $H_0 : P(X_i = a_i) = p_i$, $i = 1, 2, \ldots, r$. Denote $\#(X_j = a_i) = \nu_i$, take D as:

  $$K_n = K_n(X_1, \ldots, X_n; F) = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i} \tag{2.4}$$

Pearson Thm.: For K_n defined as eqa.**??**, then under H_0:

$$K_n \xrightarrow{\mathscr{L}} \chi^2_{r-1-s}$$

Here s is number of unknown parameter, $r-1-s$ is the degree of freedom.

Note:

- a_i must **not** depend on sample.
- For continuous case, construct division:

  $$\mathbb{R} \to (-\infty, a_1, a_2, \ldots, a_{r-1}, \infty = a_r)$$

  and test $H_0 : P(X \in I_j) = p_j$

  Criterion: Pick proper interval so that np_i and ν_i both ≥ 5.

- Contingency Table Independence & Homogeneity Test

  - Independence Test:

    Test a two-parameter sample and to see whether these two parameters(features) are independent.

    Denote $Z = (X, Y)$ are some 'level' of sample, n_{ij} is number of sample with level (i, j)

    Contingency Table:

X \ Y	1	…	j	…	s	$\sum$
1	n_{11}	…	n_{1j}	…	n_{1s}	$n_{1\cdot}$
⋮	⋮	⋱	⋮	⋱	⋮	⋮
i	n_{i1}	…	n_{ij}	…	n_{is}	$n_{i\cdot}$
⋮	⋮	⋱	⋮	⋱	⋮	⋮
r	n_{r1}	…	n_{rj}	…	n_{rs}	$n_{r\cdot}$
$\sum$	$n_{\cdot 1}$	…	$n_{\cdot j}$	…	$n_{\cdot s}$	n

    Test $H_0 : X \,\&\, Y$ are independent. i.e. $H_0 : P(X = i, Y = j) = P(X = i)P(Y = j) = p_{i\cdot}p_{\cdot j}$.

    Construct χ^2 test statistic:

$$K_n = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{[n_{ij} - n(\frac{n_{i\cdot}}{n})(\frac{n_{\cdot j}}{n})]^2}{n(\frac{n_{i\cdot}}{n})(\frac{n_{\cdot j}}{n})} = n\left(\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - 1\right) \tag{2.5}$$

    Then under H_0, $K_n \xrightarrow{\mathscr{L}} \chi^2_{rs-1-(r+s-2)} = \chi^2_{(r-1)(s-1)}$

    Reject H_0 if $p(k_0) = P(K_n \geq k_0) < \alpha$

  - Homogeneity Test:

    Test R groups of sample with category rank, to see whether these groups has similar rank distribution.

Group \ Category	Category 1	$\dots$	Category j	$\dots$	Category C	$\sum$
Group 1	n_{11}	$\dots$	n_{1j}	$\dots$	n_{1C}	$n_{1\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Group i	n_{i1}	$\dots$	n_{ij}	$\dots$	n_{iC}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Group R	n_{R1}	$\dots$	n_{Rj}	$\dots$	n_{RC}	$n_{R\cdot}$
$\sum$	$n_{\cdot 1}$	$\dots$	$n_{\cdot j}$	$\dots$	$n_{\cdot C}$	n

Denote $P(\text{Category } j|\text{Group } i) = p_{ij}$. Test $H_0 : p_{ij} = p_j, \forall 1 \leq i \leq R$.

Construct χ^2 test statistic:

$$D = \sum_{i=1}^{R}\sum_{j=1}^{C} \frac{[n_{ij} - n(\frac{n_{i\cdot}}{n})(\frac{n_{\cdot j}}{n})]^2}{n(\frac{n_{i\cdot}}{n})(\frac{n_{\cdot j}}{n})} = n\left(\sum_{i=1}^{R}\sum_{j=1}^{C}\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - 1\right) \tag{2.6}$$

Then under H_0, $D \xrightarrow{\mathscr{L}} \chi^2_{R(C-1)-(C-1)} = \chi^2_{(R-1)(C-1)}$

- Test of Normality: normality is a good & useful assumption.

  For $\vec{Y} = (Y_1, Y_2, \dots, Y_n)$,

  Test H_0 : exists $\mu \,\&\, \sigma^2$ such that Y_i i.i.d. $\sim N(\mu, \sigma^2)$.

  - Kolmogorov-Smirnov Test: Assume $\vec{X}$ form population CDF $F(x)$, test $H_0 : F(x) = F_0(x)$(where can take $F_0 = \Phi$ or some other known CDF).

    use $F_n(x)$ (as defined in eqa.**??**) as approx. to $F(x)$, test

    $$D_n = \sum_{-\infty<x<+\infty} |F_n(x) - F_0(x)|$$

    Reject H_0 if $D_n > c$

    or use goodness-of-fit: denote observed value of D_n as d_n. Reject H_0 if

    $$p(d_n) = P(D_n > d_n|H_0) < \alpha$$

  - Shapiro-Wilk Test:

    Test H_0 : exists $\mu \,\&\, \sigma^2$ such that X_i i.i.d. $\sim N(\mu, \sigma^2)$.

    Denote $Y_{(i)} = \dfrac{X_{(i) - \mu}}{\sigma}$, $m_i = E(Y_{(i)})$

    Under H_0, $(X_{(i)}, m_i)$ falls close to straight line. Test Statistic: Correlation

    $$R^2 = \frac{\left(\sum_{i=0}^{n}(X_{(i)} {-} \bar{X})(m_i - \bar{m})\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(m_i - \bar{m})^2}$$

    Reject H_0 if $R^2 < c$

    Shapiro-Wilk correction:

    $$W = \frac{\left(\sum_{i=1}^{[n/2]} a_i(X_{(n+1-i)} - X_{(i)})\right)^2}{\sum_{i=1}^{n}(X_{(i)} - \bar{X})^2}$$

- Summary: Useful Non-Parameter Hypothesis Testing.

- Non-Parameter Hypothesis Testing
  - One Population Sample
    - χ^2Test
    - Binomial Test
    - One-Sample K-S Test
    - Wilcoxon Sign Test
    - Runs Test
  - Two Population Sample
    - Independent Sample
      - Mann-Whitney Test
      - K-S Test
      - Wald-Wolfowitz Test
      - Moses Test of Extreme Reactions
    - Relative Sample
      - Sign Test
      - McNemar Test
      - Wilcoxon Rank Sum Test
      - Marginal Homogeneity Test
  - Multi-Population Sample
    - Independent Sample
      - Median Test
      - K-W One-Way ANOVA Test
      - Jonckheere-Terpstra Test
    - Relative Sample
      - Friedman Rank Sum Test
      - Kendall's Coefficient of Concordance Test
      - Cochran Q Test