Mathematical Statistics

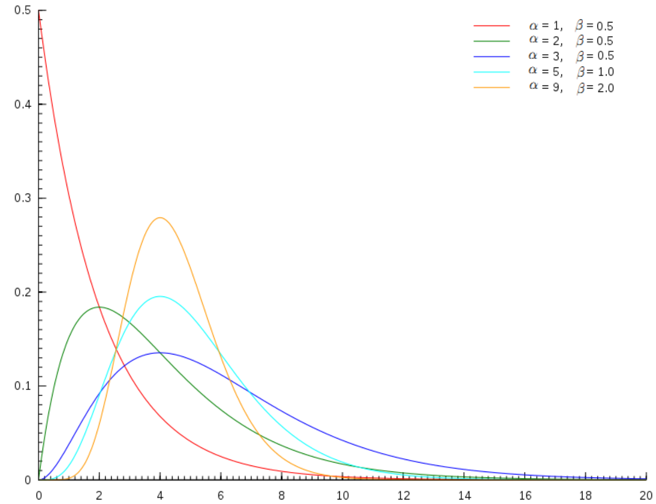
杜金鸿,15338039

2017 年 6 月 29 日

目录

1 Ditribution

1.1 Gamma Distribution

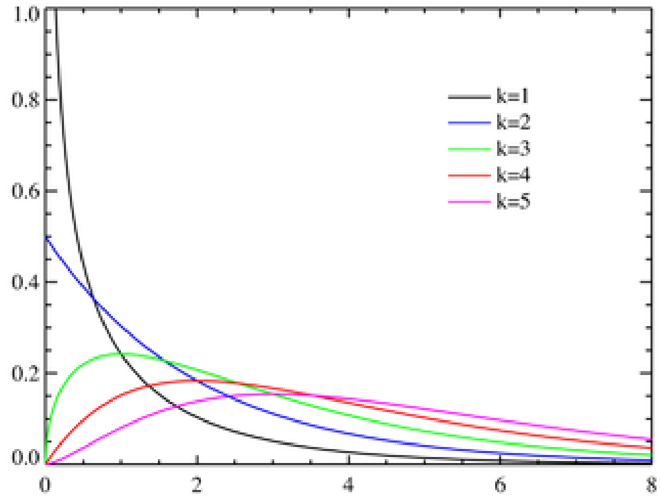


sign	$X \sim \Gamma(\alpha, \beta)$
Parameters	$\alpha > 0$ shape $\beta > 0$ rate
p.d.f	$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}, & 0 < x < \infty \\ 0, & x \leqslant 0 \end{cases}$
c.d.f	$F(x;\alpha,\beta) = \begin{cases} \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \beta x), & 0 < x < \infty \\ 0, & x \leqslant 0 \end{cases}$
Mean	$\mathrm{E}X = \frac{\alpha}{\beta}$
Variance	$\mathrm{Var}X = \frac{\alpha}{\beta^2}$
Property	$X \sim \Gamma(\alpha, \beta) \implies \frac{X}{\beta} \sim \Gamma(\alpha, 1)$ $\begin{cases} X_i \sim \Gamma(\alpha_i, \beta) \\ X_i \perp X_j \end{cases} \implies \sum_{i=1}^{n} X_i \sim \Gamma(\sum_{i=1}^{n} \alpha_i, \beta)$

$$\Gamma(s,x) = \int_x^{\infty} t^{s-1} e^{-t} \mathrm{d}t$$

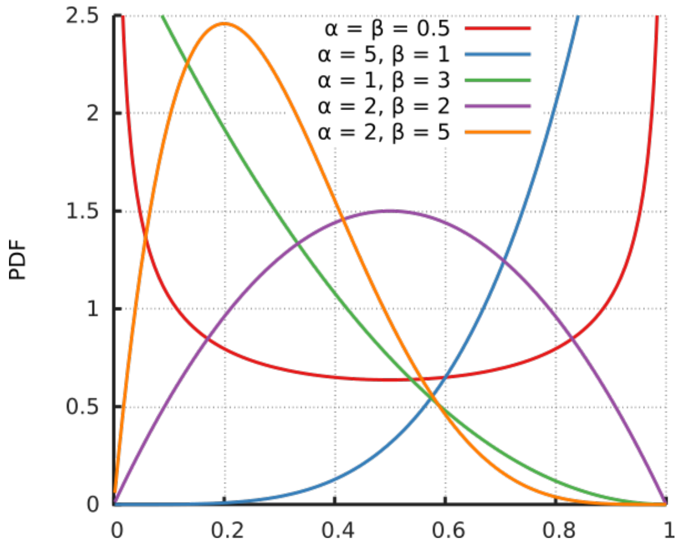
$$\gamma(s,x) = \int_0^{x} t^{s-1} e^{-t} \mathrm{d}t$$

1.2 Chi-Square Distribution



Sign	$X \sim \chi^2(k)$
Parameters	$k \in N_+$ degrees of freedom
p.d.f	$f(x;k) = \begin{cases} \dfrac{1}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, & 0 < x < \infty \\ 0, & x \leqslant 0 \end{cases}$
c.d.f	$F(x;k) = \begin{cases} \frac{1}{\Gamma(\frac{k}{2})} \gamma(\frac{k}{2}, \frac{x}{2}), & 0 < x < \infty \\ 0, & x \leqslant 0 \end{cases}$
Mean	$\mathrm{E}X = k$
Variance	$\mathrm{Var}X = 2k$
Property	$\chi^2(k) = \Gamma\left(\frac{k}{2}, 2\right)$ $Z_1, \cdots, Z_k \overset{iid}{\sim} N(0,1) \implies \sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$

1.3 Beta Distribution



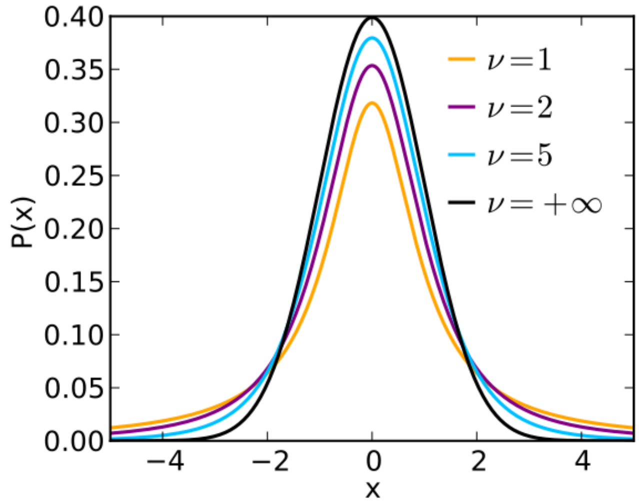
Sign	$X \sim Beta(\alpha,\beta)$
Parameters	$\alpha > 0$ shape $\beta > 0$ rate
p.d.f	$f(x;\alpha,\beta) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{Beta(\alpha,\beta)}, & 0<x<1 \\ 0, & elsewhere \end{cases}$
c.d.f	$F(x;k) = I_x(\alpha,\beta)$
Mean	$\mathrm{E}X = \frac{\alpha}{\alpha+\beta}$
Variance	$\mathrm{Var}X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Property	$\begin{cases} X_1 \sim \Gamma(\alpha,\theta) \\ X_2 \sim \Gamma(\beta,\theta) \\ X_1 \perp X_2 \end{cases} \implies \frac{X_1}{X_1+X_2} \sim Beta(\alpha,\beta)$ $Beta(1,1) = U([0,1])$

$$Beta(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$Beta(x;\alpha,\beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t$$

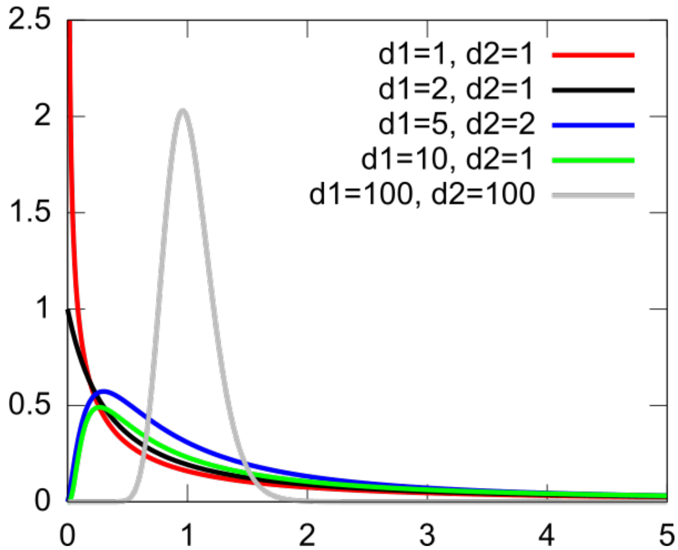
$$I_x(\alpha,\beta) = \frac{Beta(x;\alpha,\beta)}{Beta(\alpha,\beta)}$$

1.4 Student's t-Distribution



Sign	$X \sim t(\nu)$
Parameters	$\nu > 0$ degress of freedom
p.d.f	$f(x;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad -\infty < x < \infty$
c.d.f	
Mean	$\mathrm{E}X = \begin{cases} 0, & \nu > 1 \\ undefined, & elsewhere \end{cases}$
Variance	$\mathrm{Var}X = \begin{cases} \frac{\nu}{\nu-2}, & \nu > 2 \\ \infty, & 1 < \nu \leqslant 2 \\ undefined, & elsewhere \end{cases}$
Property	$\begin{cases} Z \sim N(0,1) \\ X \sim \chi^2(k) \\ Z \perp X \end{cases} \Longrightarrow \frac{Z}{\sqrt{\frac{X}{k}}} \sim t(k)$

1.5 F Distribution



Sign	$X \sim F(d_1, d_2)$
Parameters	$d_1, d_2 > 0$ degress of freedom
p.d.f	$f(x; d_1, d_2) = \frac{1}{Beta(\frac{d_1}{2}, \frac{d_2}{2})} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} \frac{x^{\frac{d_1}{2}-1}}{\left(1+\frac{d_1}{d_2}x\right)^{\frac{d_1+d_2}{2}}}$
c.d.f	
Mean	$\mathrm{E}X = \frac{d_2}{d_2-2} \qquad d_2 > 2$
Variance	$\mathrm{Var}X = \frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)} \qquad d_2 > 4$
Property	$\begin{cases} X_1 \sim \chi^2(d_1) \\ X_2 \sim \chi^2(d_2) \\ X_1 \perp X_2 \end{cases} \Longrightarrow \frac{\frac{X_1}{d_1}}{\frac{X_2}{d_2}} \sim F(d_1, d_2)$ $X \sim F(d_1, d_2) \Longrightarrow \frac{1}{X} \sim F(d_2, d_1)$

2 Convergence in Distribution

2.1 Property

1.
$$X_n \xrightarrow{D} c \qquad \Longleftrightarrow \qquad X_n \xrightarrow{P} c$$

2.
$$\begin{cases} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{cases} \qquad \Longrightarrow \qquad \begin{cases} X_n \pm Y_n \xrightarrow{P} X \pm Y \\ X_n \times Y_n \xrightarrow{P} X \times Y \end{cases}$$

3.
$$\begin{cases} X_n \xrightarrow{P} X \\ g(x) \in C(R) \end{cases} \qquad \Longrightarrow \qquad g(X_n) \xrightarrow{P} g(X)$$

4. ***Slutsky's Theorem***
$$\begin{cases} X_n \xrightarrow{D} X \\ Y_n \xrightarrow{P} c \end{cases} \qquad \Longrightarrow \qquad \frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}$$

2.2 Law of Large Numbers(LLN)

2.3 Central Limit Theorem

3 $\overline{X}$ and $\frac{nS^2}{\sigma^2}$

3.1 Definition

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

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

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

3.2 Property

1. If $EX = \mu < \infty$, then

$$\overline{X} \xrightarrow{P} \mu;$$

2. If $VarX = \sigma^2 < \infty$, then

$$S^2 \xrightarrow{P} \sigma^2, \quad S^{*2} \xrightarrow{P} \sigma^2;$$

3.

$$E\overline{X} = \mu$$

$$Var\overline{X} = \frac{\sigma^2}{n}$$

$$ES^2 = \frac{n-1}{n}\sigma^2 \qquad ES^{*2} = \sigma^2$$

$$VarS^2 = \frac{E(X-EX)^4 - \sigma^4}{n} - \frac{2[E(X-EX)^4] - 2\sigma^4}{n^2} + \frac{E(X-EX)^4 - 3\sigma^4}{n^3}$$

$$Cov(\overline{X}, S^2) = \frac{n-1}{n^2} E(X-EX)^3$$

3.3 ***Student's Theorem***

3.3.1 One Sample

$X_1, \cdots, X_n$ denote a random sample of size $n \geqslant 2$ from a distribution $N(\mu, \sigma^2)$

1. $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad U = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
2. $\overline{X} \perp S^2$
3. $\frac{nS^2}{\sigma^2} \sim \chi^2(n-1) \qquad \frac{(n-1)S^{*2}}{\sigma^2} \sim \chi^2(n-1)$
4. $\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n-1}}} \sim t(n-1) \qquad \frac{\overline{X} - \mu}{\frac{S^*}{\sqrt{n}}} \sim t(n-1)$

3.3.2 Two Samples

$$X_1, \cdots, X_{n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1^2), \quad Y_1, \cdots, Y_{n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2^2), \quad X_1, \cdots, X_{n_1}, Y_1, \cdots, Y_{n_2} \text{ are mutually independent.}$$

mixed samples variance :

$$S_w^2 = \frac{(n_1-1)S_1^{*2} + (n_2-1)S_2^{*2}}{n_1+n_2-2}$$

1. $\dfrac{\dfrac{n_1 S_1^2}{(n_1-1)\sigma_1^2}}{\dfrac{n_2 S_2^2}{(n_2-1)\sigma_2^2}} \sim F(n_1-1, n_2-1) \qquad \dfrac{\dfrac{S_1^{*2}}{\sigma_1^2}}{\dfrac{S_2^{*2}}{\sigma_2^2}} \sim F(n_1-1, n_2-1)$

2. if $\sigma_1^2 = \sigma_2^2$, then $\dfrac{(\overline{X}-\mu_1)-(\overline{Y}-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$

4 Order Statistics

4.1 Definition

$X_1, \cdots, X_n$ are i.i.d from f(x)(continuous-type), $Y_1, \cdots, Y_n$ denote $\{X_1, \cdots, X_n\}$ ranging by incresed order.

$$a < Y_1 < Y_2 < \cdots < Y_n < b$$

(We ignore the case that some Y_i $(i = 1, \cdots, n)$ are equal because its probability measure is 0)
range :

$$Y_n - Y_1$$

midrange :

$$\frac{Y_1 + Y_n}{2}$$

median(if n is odd):

$$Y_{\frac{n+1}{2}}$$

4.2 Some Special Distributions of Order Statistics

4.2.1 The joint pdf of $Y_1, \cdots, Y_n$

$$g(y_1, \cdots, y_n) = \begin{cases} n! f(y_1) \cdots, f(y_n), & a < y_1 < \cdots < y_n < b \\ 0, & elsewhere \end{cases}$$

4.2.2 The marginal pdf

$$\begin{aligned} g_1(y_1) &= \begin{cases} n! \frac{[1 - F(y_1)]^{n-1}}{(n-1)!} f(y_1), & a < y_1 < b \\ 0, & elsewhere \end{cases} \\ &= \begin{cases} n[1 - F(y_1)]^{n-1} f(y_1), & a < y_1 < b \\ 0, & elsewhere \end{cases} \\ &. \\ &. \\ &. \end{aligned}$$

$$g_k(y_k) = \begin{cases} \frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k}f(y_k), & a < y_k < b \\ 0, & elsewhere \end{cases}$$

.

.

.

$$g_n(y_n) = \begin{cases} n!\frac{[F(y_n)]^{n-1}}{(n-1)!}f(y_n), & a < y_n < b \\ 0, & elsewhere \end{cases}$$

$$= \begin{cases} n[F(y_n)]^{n-1}f(y_n), & a < y_n < b \\ 0, & elsewhere \end{cases}$$

4.2.3 The joint pdf of any 2 order statistics

$$g_{ij}(y_i, y_j) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(y_i)]^{i-1}[F(y_j)-F(y_i)]^{j-i-1}f(y_i)f(y_j), & a < y_i < y_j < b \\ 0, elsewhere \end{cases}$$

*To seek the pdf of sample range $Z_1 = Y_n - Y_1$ on specific occasion:

(1) calculate $f_{Z_1}(z_1) = g_{1n}(y_1, y_n)$, it means that on special occasion $g_{1n}(y_1, y_n)$ is a function of $z_1 = y_n - y_1$

(2) adopt the *supplementary variable technique*, let $Z_2 = Y_n$, calculate the joint pdf of Z_1 and Z_2

(3) calculate the marginal pdf of Z_1

5 Statistical Inference

5.1 Satatistic and Estimator

5.1.1 Statistic

Suppose that $X_1, X_2, \cdots, X_n$ are the observations of a random sample. Then any function $T(X_1, X_2, \cdots, X_n)$ not depend upon any unknow parameters is a random variable and a statistic.

The p.d.f of statistic may depend upon the unknown parameters.

5.1.2 Estimator

A kind of statistics like $\hat{\theta}(X_1, X_2, \cdots, X_n)$ that can be use to estimate the unknown parameters.

5.2 Unbiased Estimator

5.2.1 Definition

$$Bias = E(\hat{\theta}) - \theta = 0 \qquad \Longleftrightarrow \qquad E(\hat{\theta}) = \theta$$

5.2.2 Asymptotically Unbiased Estimator

$$\lim_{n\to\infty} Bias = \lim_{n\to\infty} E(\hat{\theta}) - \theta = 0 \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} E(\hat{\theta}) = \theta$$

* Unbiased estimator doesn't have the invariance property.

5.3 Consistent Estimator

5.3.1 Definition

$$\hat{\theta} \xrightarrow{P} \theta$$

5.3.2 Property

1. $\hat{\theta} \xrightarrow{D} \theta \qquad \Longleftrightarrow \qquad \hat{\theta} \xrightarrow{P} \theta$

2. $\begin{cases} E(\hat{\theta}) \to \theta \\ Var(\hat{\theta}) \to 0 \end{cases} \qquad \Longrightarrow \qquad E(\hat{\theta} - \theta)^2 \to 0 \qquad \Longrightarrow \qquad \hat{\theta} \xrightarrow{P} \theta$

3. *Invariance Property*

   $\begin{cases} \hat{\theta} \xrightarrow{P} \theta \\ g(x) \in C(R) \end{cases} \qquad \Longrightarrow \qquad g(\hat{\theta}) \xrightarrow{P} g(\theta)$

6 Point Estimation

6.1 Maximum Likehood Estimate

6.1.1 Definition

Likehood function

$$L(\theta;x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f(x_i;\theta) \qquad \theta=(\theta_1,\cdots,\theta_m)\in\Theta$$

Maximum Likehood Estimator (MLE) of θ is

$$\hat{\theta}=\arg\max_{\theta} L(\theta;x_1,x_2,\cdots,x_n)$$

It may be attained by the process of differentiation

$$\frac{\partial L(\theta;x_1,x_2,\cdots,x_n)}{\partial\theta}=0 \qquad \Longleftrightarrow \qquad \begin{cases} \dfrac{\partial L(\theta;x_1,x_2,\cdots,x_n)}{\partial\theta_1}=0 \\ \quad\vdots \\ \dfrac{\partial L(\theta;x_1,x_2,\cdots,x_n)}{\partial\theta_m}=0 \end{cases}$$

or

$$\frac{\partial \ln L(\theta;x_1,x_2,\cdots,x_n)}{\partial\theta}=0 \qquad \Longleftrightarrow \qquad \begin{cases} \dfrac{\partial \ln L(\theta;x_1,x_2,\cdots,x_n)}{\partial\theta_1}=0 \\ \quad\vdots \\ \dfrac{\partial \ln L(\theta;x_1,x_2,\cdots,x_n)}{\partial\theta_m}=0 \end{cases}$$

When this way can't work, it may be attained by considering the relationship of the definitional domain of $L(\theta;x_1,x_2,\cdots,x_n)$ and Θ.

6.1.2 Property

1. ***Invariance Property*** of MLE

$$\widehat{g(\theta)}=g(\hat{\theta})$$

2. Consistency

$$\hat{\theta}\xrightarrow{P}\theta$$

3. Asymptotically Sufficient

$$\frac{\dfrac{[k'(\theta)]^2}{nI(\theta)}}{Var(Y)}\to 1 \qquad (n\to\infty)$$

4. Asymptotically Normality

$$\frac{\hat{\theta}-\theta}{\sqrt{\dfrac{1}{nI(\hat{\theta})}}}\overset{\bullet}{\sim}N(0,1)$$

6.2 Method of Moments

6.2.1 Definition

j^{th} Population Moment

$$\mu_j(\theta) = E(X^j) \qquad or \qquad \widetilde{\mu}_j(\theta) = E(X - \overline{X})^j$$

j^{th} Sample Moment (***Moment Estimator***)

$$M_j(\theta) = \frac{1}{n}\sum_{i=1}^{n} X_i^j \qquad or \qquad \widetilde{M}_j(\theta) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^j$$

Supposed that j^{th} Population Moment is equal to j^{th} Sample Moment

$$\begin{cases} \mu_1(\theta) = M_1(\theta) \\ \quad\vdots \\ \mu_k(\theta) = M_k(\theta) \end{cases} \qquad or \qquad \begin{cases} \widetilde{\mu}_1(\theta) = \widetilde{M}_1(\theta) \\ \quad\vdots \\ \widetilde{\mu}_k(\theta) = \widetilde{M}_k(\theta) \end{cases}$$

here $k \in N_+$ s.t. the above equtions of θ can be solved only.

6.2.2 Property of Moment Estimator

1. unbiased
2. consistent

* Moment Estimator doesn't have the invariance property.

7 Confident Intervals

7.1 One Sample

7.1.1 μ

1. $N(\mu, \sigma^2)$ with known σ^2
   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of μ:

$$\frac{\overline{X}-\mu}{\sqrt{\dfrac{\sigma^2}{n}}} \sim N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}\right| < z_{1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

2. $N(\mu, \sigma^2)$ with unknown σ^2
   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of μ:

$$\frac{\overline{X}-\mu}{\sqrt{\dfrac{S^2}{n-1}}} \sim t(n-1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{\overline{X}-\mu}{\sqrt{\frac{S^2}{n-1}}}\right| < t_{n-1,1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

   or

$$\frac{\overline{X}-\mu}{\sqrt{\dfrac{S^{*2}}{n}}} \sim t(n-1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{\overline{X}-\mu}{\sqrt{\frac{S^{*2}}{n}}}\right| < t_{n-1,1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

3. Non-normal sample with large n and known σ^2
   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of μ:

$$\frac{\overline{X}-\mu}{\sqrt{\dfrac{\sigma^2}{n}}} \overset{\bullet}{\sim} N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

4. Non-normal sample with large n and unknown σ^2
   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of μ:

$$\frac{\overline{X}-\mu}{\sqrt{\dfrac{S^2}{n-1}}} \overset{\bullet}{\sim} N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{\overline{X}-\mu}{\sqrt{\frac{S^2}{n-1}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

   or

$$\frac{\overline{X}-\mu}{\sqrt{\dfrac{S^{*2}}{n}}} \overset{\bullet}{\sim} N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{\overline{X}-\mu}{\sqrt{\frac{S^{*2}}{n}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

7.1.2 σ^2

1. $N(\mu,\sigma^2)$ with unknown μ and σ^2

   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of σ^2:

$$\frac{nS^2}{\sigma^2} \sim \chi^2(n-1) \quad \Longrightarrow \quad P\left\{\chi^2_{n-1,\frac{\alpha}{2}} < \frac{nS^2}{\sigma^2} < \chi^2_{n-1,1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

$$\Longleftrightarrow \quad P\left\{\frac{nS^2}{\chi^2_{n-1,\frac{\alpha}{2}}} < \sigma^2 < \frac{nS^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right\} = 1-\alpha$$

   or

$$\frac{(n-1)S^{*2}}{\sigma^2} \sim \chi^2(n-1) \quad \Longrightarrow \quad P\left\{\chi^2_{n-1,\frac{\alpha}{2}} < \frac{(n-1)S^{*2}}{\sigma^2} < \chi^2_{n-1,1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

$$\Longleftrightarrow \quad P\left\{\frac{(n-1)S^{*2}}{\chi^2_{n-1,\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^{*2}}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right\} = 1-\alpha$$

2. $N(\mu,\sigma^2)$ with known μ and unknown σ^2

   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of σ^2:

$$\frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n-1) \quad \Longrightarrow \quad P\left\{\chi^2_{n-1,\frac{\alpha}{2}} < \frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} < \chi^2_{n-1,1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

7.1.3 p of $B(n,p)$

1. Method 1

   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of $p = \frac{\mu}{n}$:

$$\frac{X-np}{\sqrt{np(1-p)}} = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\bullet}{\sim} N(0,1) \quad \Longrightarrow \quad P\left\{\left|\frac{\frac{X}{n}-p}{\sqrt{\frac{np(1-p)}{n}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

   by solving the quadratic inequality of p.

2. Method 2

   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of $p = \frac{\mu}{n}$:

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \overset{\bullet}{\sim} N(0,1) \quad \Longrightarrow \quad P\left\{\left|\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

3. Method 3

   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of $p = \frac{\mu}{n}$:

$$\arcsin\sqrt{\frac{X}{n}} \overset{\bullet}{\sim} N\left(\arcsin\sqrt{p}, \frac{1}{4n}\right) \quad \Longrightarrow \quad P\left\{\left|\frac{\arcsin\sqrt{\frac{X}{n}} - \arcsin\sqrt{p}}{\sqrt{\frac{1}{4n}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

7.1.4 β of $\Gamma(\alpha,\beta)$ with known α

Given $\theta \in (0,1)$, look for the $1-\theta$ confident intervals of β:

$$\frac{2X}{\beta} \sim \Gamma(\alpha,2) = \chi^2(2\alpha) \qquad \Longrightarrow \qquad P\left\{\chi^2_{\frac{\alpha}{2}} < \frac{2X}{\beta} < \chi^2_{1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

7.2 Two Samples

7.2.1 $\mu_1 - \mu_2$

1. $N(\mu_1,\sigma_1^2), N(\mu_2,\sigma_2^2)$ with known σ_1^2,σ_2^2
   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of $\mu_1-\mu_2$:
   $$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right| < z_{1-\frac{\alpha}{2}}\right\} = 1-\alpha$$
2. $N(\mu_1,\sigma_1^2), N(\mu_2,\sigma_2^2)$ with unknown $\sigma_1^2=\sigma_2^2=\sigma^2$
   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of $\mu_1-\mu_2$:
   $$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2) \qquad \Longrightarrow \qquad P\left\{\left|\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| < t_{n_1+n_2-2,1-\frac{\alpha}{2}}\right\} = 1-\alpha$$
   where $S_w^2 = \dfrac{(n_1-1)S_1^{*2}+(n_2-1)S_2^{*2}}{n_1+n_2-2}$
3. Non-normal samples with known σ_1^2,σ_2^2
   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of $\mu_1-\mu_2$:
   $$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \stackrel{\bullet}{\sim} N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$
4. Non-normal samples with unknown σ_1^2,σ_2^2
   Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of $\mu_1-\mu_2$:
   $$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1-1}+\dfrac{S_2^2}{n_2-1}}} \stackrel{\bullet}{\sim} N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1-1}+\frac{S_2^2}{n_2-1}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$
   or
   $$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^{*2}}{n_1}+\dfrac{S_2^{*2}}{n_2}}} \stackrel{\bullet}{\sim} N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^{*2}}{n_1}+\frac{S_2^{*2}}{n_2}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

7.2.2 $\dfrac{\sigma_1^2}{\sigma_2^2}$

Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of $\dfrac{\sigma_1^2}{\sigma_2^2}$:

$$\frac{\frac{S^{*2}}{\sigma_1^2}}{\frac{S^{*2}}{\sigma_2^2}} \sim F(n_1-1, n_2-1) \qquad \Longrightarrow \qquad P\left\{F_{n_1-1,n_2-1,\frac{\alpha}{2}} < \frac{\frac{S^{*2}}{\sigma_1^2}}{\frac{S^{*2}}{\sigma_2^2}} < F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

7.2.3 $p_1 - p_2$

Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of $p_1 - p_2$:

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \dot{\sim} N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

8 Test of Statistical Hypotheses

8.1 Definition

The experimental values of $X_1, \cdots, X_n$ are $x_1, \cdots, x_n$. Let $X = (X_1, \cdots, X_n)$, $x = (x_1, \cdots, x_n)$.

$$H_0 : Null\ hypothesis \qquad x \in C : Critical\ region/Rejection\ region$$

$$H_1 : Alternative\ hypothesis \qquad x \in C^* = C^c$$

1. Kinds

$$\begin{cases} simple\ statistical\ hypothesis & \text{completely specifies the distribution, like } \theta = 75. \\ composite\ statistical\ hypothesis & \text{describle many distributions, like } \theta \geqslant 75. \end{cases}$$

$$\begin{cases} One-sided\ hypothesis \\ Two-sided\ hypothesis \end{cases}$$

2. $Power\ function$: $\quad K(\theta) = Pr(X \in C)$

   $Power$: $\quad K(\theta_0) \quad \forall \theta_0 \in \Theta$

3. $Significant\ level$: $\quad \alpha = \sup\limits_{\theta \in \Theta} Pr(X \in C; H_0) = \sup\limits_{\theta \in \Theta_0} Pr(X \in C)$

   $P-value$: $\quad p = Pr[T(X) \geqslant T(x); H_0] \quad$ when $C^* = \{x : T(x) \geqslant c\}$

   or $\quad p = Pr[T(X) \leqslant T(x); H_0] \quad$ when $C^* = \{x : T(x) \leqslant c\}$

4. Error

$$\begin{cases} Type\ I\ Error : \alpha = Pr(X \in C | \theta \in \Theta_0) \\ Type\ II\ Error : \beta = 1 - Pr(X \in C | \theta \in \Theta_1) \end{cases}$$

Table of error types		H0 True	H0 False
Decision About H0	Reject	Type I error (False Positive)	Correct inference (True Positive)
	Fail to reject	Correct inference (True Negative)	Type II error (False Negative)

8.2 One Sample

8.2.1 μ

1. $N(\mu, \sigma^2)$ with known σ^2: z test

$$\frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

when $H_0 : \mu = \mu_0$ is accepted.

2. $N(\mu, \sigma^2)$ with unknown σ^2: t test

$$\frac{\overline{X} - \mu}{\sqrt{\dfrac{S^2}{n-1}}} \sim t(n-1)$$

when $H_0 : \mu = \mu_0$ is accepted.

3. Non-normal samples with known σ^2

$$\frac{\overline{X} - \mu}{\sqrt{\dfrac{\sigma^2}{n}}} \stackrel{\bullet}{\sim} N(0,1)$$

when $H_0 : \mu = \mu_0$ is accepted.

4. Non-normal samples with known σ^2

$$\frac{\overline{X} - \mu}{\sqrt{\dfrac{S^2}{n-1}}} \stackrel{\bullet}{\sim} N(0,1)$$

or

$$\frac{\overline{X} - \mu}{\sqrt{\dfrac{S^{*2}}{n}}} \stackrel{\bullet}{\sim} N(0,1)$$

when $H_0 : \mu = \mu_0$ is accepted.

8.2.2 σ^2

8.2.3 p of $B(n,p)$

1.

$$\frac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \stackrel{\bullet}{\sim} N(0,1)$$

when $H_0 : p = p_0$ is accepted.

2.

$$\frac{\hat{p} - p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}} \stackrel{\bullet}{\sim} N(0,1)$$

when $H_0 : p = p_0$ is accepted.

8.3 Two Sample

8.3.1 $\mu_1 - \mu_2$

1. $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ with known σ_1^2, σ_2^2

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

when $H_0 : \mu_1 = \mu_2$ is accepted.

2. $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ with unknown $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

where $S_w^2 = \dfrac{(n_1-1)S_1^{*2}+(n_2-1)S_2^{*2}}{n_1+n_2-2}$, when $H_0 : \mu_1 = \mu_2$ is accepted.

3. Non-normal samples with known σ_1^2, σ_2^2

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \overset{\bullet}{\sim} N(0,1)$$

when $H_0 : \mu_1 = \mu_2$ is accepted.

4. Non-normal samples with unknown σ_1^2, σ_2^2

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1-1}+\frac{S_2^2}{n_2-1}}} \overset{\bullet}{\sim} N(0,1)$$

or

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^{*2}}{n_1}+\frac{S_2^{*2}}{n_2}}} \overset{\bullet}{\sim} N(0,1)$$

when $H_0 : \mu_1 = \mu_2$ is accepted.

8.4 Paired Design

$(X_1, Y_1), \cdots, (X_n, Y_n)$ are the paired samples.

Let

$$D_i = X_i - Y_i, \quad (i = 1, 2 \cdots, n)$$

then

$$D_1, \cdots, D_n \overset{iid}{\sim} N(\mu_1 - \mu_2, \sigma_D^2)$$

then we have paired t test

$$T = \frac{\overline{D}}{\sqrt{\frac{S_D^2}{n-1}}} \sim t(n-1)$$

or

$$T = \frac{\overline{D}}{\sqrt{\frac{S_D^{2*}}{n}}} \sim t(n-1)$$

when $H_0 : \mu_1 = \mu_2$ is accepted.

9 Chi-Square Tests

9.1 Definition

Let $X_1,\cdots,X_n \sim M(n,p_1,\cdots,p_n)$, $\sum\limits_{i=1}^{n} X_i = n$, $\sum\limits_{i=1}^{n} p_i = 1$.

$$P\{X_1 = x_1,\cdots,X_n = x_n\} = \frac{n!}{x_1!\cdots x_n!} p_1^{x_1}\cdots p_n^{x_n}$$

then

$$Q_{n-1} = \sum_{i=1}^{n} \frac{(X_i - np_i)^2}{np_i} \overset{\bullet}{\sim} \chi^2(n-1)$$

9.2 One Sample

1. Goodness-of-fitness tests

$$H_0: p_1 = p_{10},\cdots,p_n = p_{n0}$$
$$H_1: p_{i_0} \neq p_{i_0 0}$$

then

$$Q_{n-1} = \sum_{i=1}^{n} \frac{(X_i - n\hat{p}_i)^2}{n\hat{p}_i} \overset{\bullet}{\sim} \chi^2(n-1)$$

where $P\{Q_{n-1} \geqslant \chi^2_{1-\alpha}(n-1)\} = \alpha$.

9.3 Mutiple Samples

Let $X_{1j},\cdots,X_{nj} \sim M(n,p_{1j},\cdots,p_{nj})$, $\sum\limits_{i=1}^{n} X_{ij} = n$, $\sum\limits_{i=1}^{n} p_{ij} = 1$, where $j = 1,\cdots,m$.

$$H_0: p_{11} = \cdots = p_{1m},\cdots,p_{n1} = \cdots = p_{nm}$$
$$H_1: p_{i_0 j_0} \neq p_{i_0 j_0}$$

then

$$\sum_{j=1}^{m}\sum_{i=1}^{n} \frac{(X_{ij} - n_j p_{ij})^2}{n_j p_{ij}} \overset{\bullet}{\sim} \chi^2[m(n-1)]$$

Use $\dfrac{X_{i1}+\cdots+X_{im}}{n_1+\cdots+n_m} \sim p_{i1} = \cdots = p_{im}$ then

$$\sum_{j=1}^{m}\sum_{i=1}^{n} \frac{\left[X_{ij} - n_j\left(\dfrac{X_{i1}+\cdots+X_{im}}{n_1+\cdots+n_m}\right)\right]^2}{n_j\left(\dfrac{X_{i1}+\cdots+X_{im}}{n_1+\cdots+n_m}\right)} \overset{\bullet}{\sim} \chi^2[(m-1)(n-1)]$$

9.4 Contingency Table

Two factors A, B, a levels:$A_1, \cdots, A_a$, b levels:$B_1, \cdots, B_b$.
Let $p_{ij} = P(A_i \cap B_j)$
then

$$Q_{ab-1} = \sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(X_{ij} - np_{ij})^2}{np_{ij}} \overset{\bullet}{\sim} \chi^2(mn-1)$$

Let $\begin{cases} \hat{p}_{i.} = \dfrac{X_{i.}}{n}, \quad X_{i.} = \sum\limits_{j=1}^{n} X_{ij} \\ \hat{p}_{.j} = \dfrac{X_{.j}}{n}, \quad X_{.j} = \sum\limits_{i=1}^{m} X_{ij} \\ \hat{p}_{ij} = \hat{p}_{i.}\hat{p}_{.j} \end{cases}$ then

$$Q = \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{\left(X_{ij} - n\dfrac{X_{i.}}{n}\dfrac{X_{.j}}{n} \right)^2}{n\dfrac{X_{i.}}{n}\dfrac{X_{.j}}{n}} \overset{\bullet}{\sim} \chi^2[(m-1)(n-1)]$$

10 Sufficiency

Here we first discuss the measures of quality of estimators. As one of them, UMVE is always acceptable and reliable. So, to find out the UMVE, we discuss the relationship between UMVE and sufficiency. Then, to find out the uniqueness of UMVE, we introduce completeness to illustrate it. However, the completeness is not always easy to find or to proof. So, we discuss a special complete p.d.f. family, exponential class. Then we discuss the relatonship among sufficience, completeness, uniqueness and independence.

Our discussion begins from one parameter, to the function of a parameter, finally to parameters.

Although the UMVE is always useful, we need to point out that there is no one way works well for any situation. Besides, the relavant estimators or statistics may not exist.

How to find the unique UMVE? If we have the complete sufficient statistic Y_1, then we have 2 ways to find the UMVE of $g(\theta)$. First, find $u(Y_1)$, the function of Y_1 and $EY_1 = g(\theta)$. Second, find $\varphi(Y_1) = E(Y_2|Y_1)$, where Y_2 is the unbiased estimator of $g(\theta)$.

10.1 Measures of Quality of Estimators

10.1.1 Desicion Function/Desicion Rule

1. Definition

   Let Y be an estimator of θ and y be the observed value of Y, then $\hat{\theta} = \delta(y)$, the function of y, is decision function.

2. Loss Function

   i Squared-error Loss Function
   $$L(\theta,\delta) = (\theta-\delta)^2$$

   ii Absolute-error Loss Function
   $$L(\theta,\delta) = |\theta-\delta|$$

   iii Goal-post Loss Function
   $$L(\theta,\delta) = \begin{cases} 0, & |\theta-\delta| \leqslant a \\ b, & |\theta-\delta| > a \end{cases}$$

3. Risk Function

   The expectation of loss functuon,
   $$R(\theta,\delta) = E\{L[\theta,\delta(Y)]\}$$

   If the loss function is $L(\theta,\delta) = (\theta-\delta)^2$, then the Mean Squared-error($\boldsymbol{MSE}$) is:
   $$\begin{aligned} MSE &= \{E[\delta(Y)]-\theta\}^2 + Var[\delta(Y)] \\ &= \{Bias[\delta(Y)]\}^2 + Var[\delta(Y)] \end{aligned}$$

4. Criterion of Selecting Desicion Function

   a. MSE principle

   b. Minimax principle

$$\max_{\theta} R[\theta, \delta_0(y)] = \min_{\delta} \max_{\theta} R[\theta, \delta(y)]$$

   c. Likelihood principle

10.1.2 Properties of Estimator

Let $Y = u(X_1, \cdots, X_n)$ be the estimator of θ

1. Unbiased Estimator (***UE***)

$$E_\theta(Y) = \theta \qquad \forall \theta \in \Theta$$

2. Efficient Estimator (***EE***)

$$\begin{cases} E_\theta(Y_0) = E_\theta(Y) \\ Var_\theta(Y_0) \leqslant Var_\theta(Y) \end{cases} \qquad \forall \theta \in \Theta$$

3. Unbiased Minimum Variance Estimator (***UMVE***)

$$\begin{cases} E_\theta(Y_0) = \theta \\ Var_\theta(Y_0) \leqslant Var_\theta(Y) \end{cases} \qquad \forall \theta \in \Theta, \forall Y \in \{Y : E_\theta(Y) = \theta\}$$

4. Minimum Mean-squared-error Estimator (***MMSE***)

$$E_\theta[\delta_0(Y) - \theta]^2 \leqslant E_\theta[\delta(Y) - \theta]^2 \qquad \forall \theta \in \Theta, \forall \delta$$

10.2 Sufficient Statistic

10.2.1 Definition

1. One Parameter
   Let $X_1, X_2, \cdots, X_n$ be the statistics with the joint distribution $f(x_1, x_2, \cdots, x_n; \theta)$ and $Y = u(X_1, X_2, \cdots, X_n)$ be a statistic with p.d.f. $g(y; \theta)$, where $\theta \in \Theta \subset R$. Then Y is a sufficient statistic for θ if and only if

$$\frac{f(x_1, x_2 \cdots, x_n; \theta)}{g[u(x_1, x_2, \cdots, x_n); \theta]} = H(x_1, x_2, \cdots, x_n)$$

   If $X_1, X_2, \cdots, X_n$ are the observations of a random sample, that is they are i.i.d. Then it equals to

$$\frac{\prod_{i=1}^{n} f(x_i; \theta)}{g[u(x_1, x_2, \cdots, x_n); \theta]} = H(x_1, x_2, \cdots, x_n)$$

2. Parameters

Let $X_1, X_2, \cdots, X_n$ be the statistics with the joint distribution $f(x_1, x_2, \cdots, x_n; \boldsymbol{\theta})$ and $Y_i = u_i(X_1, X_2, \cdots, X_n)$ $(i = 1, 2, \cdots, m)$ be statistics with joint p.d.f. $g(y_1, y_2, \cdots, y_m; \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subset R^m$. Then $Y_1, Y_2 \cdots, Y_m$ are joint sufficient statistics for $\boldsymbol{\theta}$ if and only if

$$\frac{f(x_1, x_2 \cdots, x_n; \boldsymbol{\theta})}{g(y_1, y_2, \cdots, y_m; \boldsymbol{\theta})} = \frac{f(x_1, x_2 \cdots, x_n; \boldsymbol{\theta})}{g[u_1(x_1, x_2, \cdots, x_n), u_2(x_1, x_2, \cdots, x_n), \cdots, u_m(x_1, x_2, \cdots, x_n); \boldsymbol{\theta}]} = H(x_1, x_2, \cdots, x_n)$$

If $X_1, X_2, \cdots, X_n$ are the observations of a random sample, that is they are i.i.d. Then it equals to

$$\frac{\prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})}{g(y_1, y_2, \cdots, y_m; \boldsymbol{\theta})} = \frac{\prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})}{g[u_1(x_1, x_2, \cdots, x_n), u_2(x_1, x_2, \cdots, x_n), \cdots, u_m(x_1, x_2, \cdots, x_n); \boldsymbol{\theta}]} = H(x_1, x_2, \cdots, x_n)$$

10.2.2 *Factorization Theorem of Neyman*

1. One Parameter

Let $X_1, X_2 \cdots, X_n$ be i.i.d. with p.d.f. $f(x; \theta)$ and $Y = u(X_1, X_2, \cdots, X_n)$ be a statistic. Then Y is a sufficient statistic if and only if there are 2 functions $k_1, k_2 \geqslant 0$ s.t.

$$\prod_{i=1}^{n} f(x_i; \theta) = k_1[u(x_1, x_2, \cdots, x_n); \theta] \cdot k_2(x_1, x_2, \cdots, x_n)$$

2. Parameters

Let $X_1, X_2 \cdots, X_n$ be i.i.d. with p.d.f. $f(x; \boldsymbol{\theta})$ and $Y = u(X_1, X_2, \cdots, X_n)$ be a statistic. Then $Y_1, Y_2, \cdots, Y_m$ are joint sufficient statistics if and only if there are 2 functions $k_1, k_2 \geqslant 0$ s.t.

$$\begin{aligned}\prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) &= k_1(y_1, y_2, \cdots, y_m; \boldsymbol{\theta}) \cdot k_2(x_1, x_2, \cdots, x_n) \\ &= k_1[u_1(x_1, x_2, \cdots, x_n), u_2(x_1, x_2, \cdots, x_n), \cdots, u_m(x_1, x_2, \cdots, x_n); \boldsymbol{\theta}] \cdot k_2(x_1, x_2, \cdots, x_n)\end{aligned}$$

10.2.3 Properties

1. Sufficient Statistic & MLE

   Let $X_1, X_2 \cdots, X_n$ be i.i.d. with p.d.f. $f(x; \theta)$, $Y = u(X_1, X_2, \cdots, X_n)$ be a sufficient statistic for θ and $\hat{\theta}$ be the unique maximum likelihood estimator of θ, then $\hat{\theta}$ is a function of Y.

2. Invariance Property

   If $Y = u(X_1, X_2, \cdots, X_n)$ is a sufficient statistic for θ and $g(y)$ is one-to-one Borel function, then $g(Y)$ is also a sufficient statistic for θ.

3. Sufficient Statistic & UMVE

   Rao and Blackwell Theorem

   Let $X_1, X_2 \cdots, X_n$ be i.i.d. with p.d.f. $f(x; \theta)$, $Y_1 = u_1(X_1, X_2, \cdots, X_n)$ be a sufficient statistic for θ and

$Y_2 = u_2(X_1, X_2, \cdots, X_n)$ be an unbiased estimator of $\boldsymbol{\theta}$. Then the statistic $E(Y_2|Y_1) = \varphi(Y_1)$ is also an unbiased estimator of $\boldsymbol{\theta}$, with

$$\begin{cases} E[\varphi(Y_1)] = EY_2 = \theta \\ Var[\varphi(Y_1)] \leqslant Var(Y_2) \end{cases}$$

It means that UMVE is a function of a sufficient statistic for $\boldsymbol{\theta}$.

Corollary: Let $X_1, X_2 \cdots, X_n$ be i.i.d. with p.d.f. $f(x; \boldsymbol{\theta})$, $Y_1 = u_1(X_1, X_2, \cdots, X_n)$ be a sufficient statistic for $\boldsymbol{\theta}$ and $Y_2 = u_2(X_1, X_2, \cdots, X_n)$ be an unbiased estimator of $h(\boldsymbol{\theta})$ ($h(\cdot)$ is a Borel function). Then the statistic $E(Y_2|Y_1) = \varphi(Y_1)$ is also an unbiased estimator of $h(\boldsymbol{\theta})$, with

$$\begin{cases} E[\varphi(Y_1)] = EY_2 = h(\theta) \\ Var[\varphi(Y_1)] \leqslant Var(Y_2) \end{cases}$$

10.3 Complete family

10.3.1 Definition

1. One Parameter

   Let the random variable Z have a p.d.f. in $\{h(z; \theta) : \theta \in \Theta \subset R\}$. If $\forall u(z)$ is a Borel function, and

   $$E[u(Z)] = 0, \forall \theta \in \Theta \qquad \Longrightarrow \qquad u(z) = 0 \qquad (a.e.)$$

   then $\{h(z; \theta) : \theta \in \Theta\}$ is a complete family of p.d.f.

2. Parameters

   Let the random variables $Z_1, Z_2, \cdots, Z_m$ have a joint p.d.f. in $\{h(z_1, z_2, \cdots, z_m; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset R_m\}$. If $\forall u(z_1, z_2, \cdots, z_m)$ is a Borel function, and

   $$E[u(Z_1, Z_2, \cdots, Z_m)] = 0, \forall \boldsymbol{\theta} \in \Theta \qquad \Longrightarrow \qquad u(z_1, z_2, \cdots, z_m) = 0 \qquad (a.e.)$$

   then $\{h(z; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ is a complete family of p.d.f.

10.3.2 Properties

(1) Invariance Property

If $Y = u(X_1, X_2, \cdots, X_n)$ is a complete statistic for θ (the famlily of p.d.f. $\{g(y; \theta) : \theta \in \Theta\}$ is complete) and $g(y)$ is a Borel function. Then $g(Y)$ is also a complete statistic for $\boldsymbol{\theta}$.

(2) Complete Statistic & Unique UE

If $Y = u(X_1, X_2, \cdots, X_n)$ is a complete statistic for θ (the famlily of p.d.f. $\{g(y; \theta) : \theta \in \Theta\}$ is complete) and Y is also an unbiased estimator of $\boldsymbol{\theta}$. Then Y is the unique unbiased estimator of $\boldsymbol{\theta}$.

(3) Complete Sufficient Statistic & Unique UMVE

Theorem of Lehmann and Scheffé

Let $X_1, X_2 \cdots, X_n$ be i.i.d. with p.d.f. $f(x;\theta)$, $Y = u(X_1, X_2, \cdots, X_n)$ is a complete sufficient statistic for θ. If $g(Y)$ is an unbiased estimator of θ, then $g(Y)$ is the unique unbiased minimum variance estimator of θ.

'Unique' is in the meaning of probability. That means if Y_2 is another UMVE of θ, then $Y = Y_2$ $(a.e.)$.

Y is a complete sufficient statistic for θ means that Y is a sufficient statistic for θ and the family $\{h(y;\theta): \theta \in \Theta\}$ of p.d.f. is complete.

Corollary: Let $X_1, X_2 \cdots, X_n$ be i.i.d. with p.d.f. $f(x;\theta)$, $Y = u(X_1, X_2, \cdots, X_n)$ is a complete sufficient statistic for θ. If $g(Y)$ is an unbiased estimator of $h(\theta)$ ($h(\cdot)$ is a Borel function), then $g(Y)$ is the unique unbiased minimum variance estimator of $h(\theta)$.

10.4 The Exponential Class of p.d.f

10.4.1 Definition

1. One Parameter

Consider a family $\{f(x;\theta): \theta \in \Theta \subset R\}$ of p.d.f. where $\Theta = \{\theta : \gamma < \theta < \delta\}$ $(\gamma, \delta \in R)$, and where

$$f(x;\theta) = e^{p(\theta)K(x)+S(x)+q(\theta)} I_D(x), \qquad D \in \mathscr{B}(R)$$

If it obey these regularity conditions:

(1) D does not depend upon θ;

(2) $p(\theta)$ is a nontrival continuous function on Θ;

(3) If X is a continuous random variable: $K'(x) \not\equiv 0$ and $S(x)$ is a continuous function on D;
If X is a discrete random variable: $K(x)$ is a nontrival function on D.

then the family is an exponential class.

2. Parameters

Consider a family $\{f(x;\boldsymbol{\theta}): \boldsymbol{\theta} \in \Theta \subset R^m\}$ of p.d.f. where $\Theta = \{\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m): \gamma_i < \theta_i < \delta_i, i = 1, 2, \cdots, m\}$ $(\gamma_i, \delta_i \in R)$, and where

$$f(x;\boldsymbol{\theta}) = e^{\sum\limits_{i=1} mp_i(\boldsymbol{\theta})K_i(x)+S(x)+q(\boldsymbol{\theta})} I_D(x), \qquad D \in \mathscr{B}(R)$$

If

(1) D does not depend upon $\boldsymbol{\theta}$;

(2) $p_i(\boldsymbol{\theta})$ $\quad (i = 1, 2, \cdots, m)$ are nontrival continuous functions on Θ;

(3) If X is a continuous random variable: $K_i'(x)$ $\quad (i = 1, 2, \cdots, m)$ are continuous and linearly independent functions on D, and $S(x)$ is a continuous function on D;
If X is a discrete random variable: $K_i'(x)$ $\quad (i = 1, 2, \cdots, m)$ are nontrival and linearly independent functions on D.

then the family is an exponential class.

10.4.2 Property

Exponential Class & Complete Sufficient Statistic

1. One Parameter
   Let $X_1, X_2 \cdots, X_n$ be i.i.d. with p.d.f. $f(x;\theta)$ that is in an exponential class. Then $Y = \sum_{i=1}^{n} K(X_i)$ is a complete sufficient statistic for θ and

   (1) $g_Y(y;\theta) = R(y)e^{p(\theta)y + nq(\theta)}$;

   (2) $EY = \dfrac{\mathrm{d}[-nq(\theta)]}{\mathrm{d}[p(\theta)]} = -\dfrac{nq'(\theta)}{p'(\theta)}$;

   (3) $Var(Y) = \dfrac{\mathrm{d}^2[-nq(\theta)]}{\mathrm{d}[p(\theta)]^2} = \dfrac{n}{[p'(\theta)]^3}[p''(\theta)q'(\theta) - p'(\theta)q''(\theta)]$.

2. Parameters
   Let $X_1, X_2 \cdots, X_n$ be i.i.d. with joint p.d.f. $f(x_1, x_2, \cdots, x_n; \boldsymbol{\theta})$ that is in an exponential class. Then $Y_1 = \sum_{i=1}^{n} K_1(X_i), Y_2 = \sum_{i=1}^{n} K_2(X_i), \cdots, Y_m = \sum_{i=1}^{n} K_m(X_i)$ are joint complete sufficient statistics for $\boldsymbol{\theta}$ and

   (1) $g_{Y_1,Y_2,\cdots,Y_m}(y_1, y_2, \cdots, y_m; \boldsymbol{\theta}) = R(y_1, y_2, \cdots, y_m)e^{\sum_{i=1}^{m} p_i(\boldsymbol{\theta})y + nq(\boldsymbol{\theta})}$;

   (2) $EY_i = \dfrac{\partial[-nq(\boldsymbol{\theta})]}{\partial[p_i(\boldsymbol{\theta})]}$;

   (3) $Cov(Y_i, Y_j) = \dfrac{\partial^2[-nq(\boldsymbol{\theta})]}{\partial[p_i(\boldsymbol{\theta})]\partial[p_j(\boldsymbol{\theta})]}$.

10.5 Minimal Sufficient

10.5.1 Definition

If $Y = u(X_1, X_2, \cdots, X_n)$ is a sufficient statistic for $\boldsymbol{\theta}$ and a function of any other sufficient statistic for $\boldsymbol{\theta}$, then Y is a minimal sufficient statistic for $\boldsymbol{\theta}$.

If $(Y_1, Y_2, \cdots, Y_m) = (u_1(X_1, X_2, \cdots, X_n), u_2(X_1, X_2, \cdots, X_n), \cdots, u_m(X_1, X_2, \cdots, X_n))$ $(m \leqslant n)$ are joint sufficient statistics for $\boldsymbol{\theta}$ and a function of any other sufficient statistic for $\boldsymbol{\theta}$, then $(Y_1, Y_2, \cdots, Y_n)$ are joint minimal sufficient statistics for $\boldsymbol{\theta}$.

10.5.2 Properties

1. Invariance Property
   If $Y = u(X_1, X_2, \cdots, X_n)$ is a minimal sufficient statistic for $\boldsymbol{\theta}$ and $g(y)$ is one-to-one Borel function, then $g(Y)$ is also a minimal sufficient statistic for $\boldsymbol{\theta}$.

   Minimal sufficient statistic is not unique.

2. Maximum Likelihood Estimator, Sufficient Statistic & Minimal Sufficient Statistic
   If $Y = u(X_1, X_2, \cdots, X_n)$ is the unique maximum likelihood estimator of θ as well as a sufficient statistic for θ, then Y is also a minimal sufficient statistic for θ.

3. Complete Sufficient Statistic & Minimal Sufficient Statistic
   If $Y = u(X_1, X_2, \cdots, X_n)$ is a complete sufficient statistic for θ, then Y is also a minimal sufficient statistic for θ.

10.6 Ancillary Statistic

10.6.1 Definition

If $Y = u(X_1, X_2, \cdots, X_n)$ is a statistic whose distribution doesn't depend on θ, then Y is an ancillary statistic for θ.

10.6.2 Kinds

1. Location-Invariant Statistic
   Let the p.d.f of $X_1, X_2, \cdots, X_n$ be $f(x-\theta)$, that is, θ is a location parameter. If $Y = u(X_1, X_2, \cdots, X_n)$ satisfies
   $$u(x_1+c, x_2+c, \cdots, x_n+c) = u(x_1, x_2, \cdots, x_n)$$
   then Y is a location-invariant statistic.

   Location-invariant statistic is an ancillary for the scale parameter.

2. Scale-Invariant Statistic
   Let the p.d.f of $X_1, X_2, \cdots, X_n$ be $\frac{1}{\theta} f\left(\frac{1}{\theta} x\right) \quad (\theta > 0)$, that is, θ is a scale parameter. If $Y = u(X_1, X_2, \cdots, X_n)$ satisfies
   $$u(cx_1, cx_2, \cdots, cx_n) = u(x_1, x_2, \cdots, x_n)$$
   then Y is a scale-invariant statistic.

   Scale-invariant statistic is an ancillary for the scale parameter.

3. Location-and-Scale-Invariant Statistic
   Let the p.d.f of $X_1, X_2, \cdots, X_n$ be $\frac{1}{\theta_2} f\left(\frac{x-\theta_1}{\theta_2}\right) \quad (\theta_2 > 0)$. If $Y = u(X_1, X_2, \cdots, X_n)$ satisfies
   $$u(cx_1+d, cx_2+d, \cdots, cx_n+d) = u(x_1, x_2, \cdots, x_n)$$
   then Y is a location-and-scale-invariant statistic.

   Location-and-scale-invariant statistic is an ancillary for the location and scale parameters.

10.6.3 Properties

Complete Sufficient Statistic & Ancillary Statistic
If $Y_1 = u_1(X_1, X_2, \cdots, X_n)$ is a complete sufficient statistic for θ, $Y_2 = u_2(X_1, X_2, \cdots, X_n)$ is an ancillary statistic for θ and not a function of Y_1 alone, then Y_1 and Y_2 are independent.

11 Efficiency

11.1 Fisher Information

11.1.1 Definition

If X is a random varible with p.d.f $f(x;\theta)$, $\quad \theta \in \Theta$ which obeys these regularity conditions:

1. If $\theta \neq \theta'$, then $P\{f(X;\theta) \neq f(X;\theta')\} > 0$;
2. The support S for θ is common;
3. The support S' for $f(x;\theta)$ is common;
4. $f(x;\theta)$ is twice differentiable for θ;
5. $f(x;\theta)$ and $f'(x;\theta)$ are uniform continuous, which is $\begin{cases} \dfrac{\partial}{\partial\theta}\displaystyle\int_S f(x;\theta)\mathrm{d}x = \int_S \dfrac{\partial}{\partial\theta} f(x;\theta)\mathrm{d}x \\ \dfrac{\partial^2}{\partial\theta^2}\displaystyle\int_S f(x;\theta)\mathrm{d}x = \int_S \dfrac{\partial^2}{\partial\theta^2} f(x;\theta)\mathrm{d}x \end{cases}$

then the Fisher Information is

$$\begin{aligned} I(\theta) &= \int_R \left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\right]^2 f(x;\theta)\mathrm{d}x \\ &= -\int_R \frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} f(x;\theta)\mathrm{d}x \end{aligned}$$

or

$$\begin{aligned} I(\theta) &= E\left[\frac{\partial \ln f(X;\theta)}{\partial\theta}\right]^2 \\ &= -E\left[\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right] \end{aligned}$$

Because

$$E\left[\frac{\partial \ln f(X;\theta)}{\partial\theta}\right] = 0$$

we also have

$$I(\theta) = Var\left[\frac{\partial \ln f(X;\theta)}{\partial\theta}\right]$$

11.1.2 Fisher Information About The Sample

If the fisher information in one observation is $I(\theta)$, then the fisher information in a random sample of size n is $I_n(\theta) = nI(\theta)$.

11.2 ***Rao − Cramér Inequality***

Let $X_1, X_2 \cdots, X_n$ be i.i.d. with p.d.f. $f(x;\theta)$ $\theta \in \Theta$ that obeys the regularity conditions and $Y = u(X_1, X_2, \cdots, X_n)$ be an unbiased estimator of $k(\theta)$, then

$$Var(Y) \geqslant \frac{[k'(\theta)]^2}{nI(\theta)},$$

here $\frac{[k'(\theta)]^2}{nI(\theta)}$ is called Rao-Cramér lower bound.

11.3 Efficient Estimator

11.3.1 Definition

If $Y = u(X_1, X_2, \cdots, X_n)$ is an unbised estimator of $k(\theta)$, then the ratio

$$\frac{\frac{[k'(\theta)]^2}{nI(\theta)}}{Var(Y)}$$

is called the efficiency of Y.

If the efficiency of Y is equivalent to 1, that is

$$Var(Y) = \frac{[k'(\theta)]^2}{nI(\theta)},$$

then Y is an efficient estimator of $k(\theta)$.

11.3.2 Asymptotically Efficient

If $Y = u(X_1, X_2, \cdots, X_n)$ is an unbised estimator of $k(\theta)$ and

$$Var(Y) \longrightarrow \frac{[k'(\theta)]^2}{nI(\theta)} \qquad (n \to +\infty),$$

then Y is an asymptotically efficient estimator of $k(\theta)$.

11.4 Limiting Distribution of Maximum Likelihood Estimators

11.4.1 Properties

If $X_1, X_2, \cdots, X_n$ are random varibles with p.d.f $f(x;\theta)$, $\theta \in \Theta$ which obeys these regularity conditions:

1. If $\theta \neq \theta'$, then $P\{f(X;\theta) \neq f(X;\theta')\} > 0$;

2. The support S for θ is common;

3. The support S' for $f(x;\theta)$ is common;

4. $f(x;\theta)$ is differentiable for θ.

then

(1) $\hat{\theta}$, the limiting maximum likelihood estimator of θ exists with probability 1;

(2) $\hat{\theta} \xrightarrow{P} \theta \qquad (n \to \infty)$.

If further

5. $f(x;\theta)$ is thrice differentiable for θ, $\left|\dfrac{\partial^3 \ln f(x;\theta)}{\partial \theta^3}\right| \leqslant H(x),\ E[H(x)] < \infty$

then

$$\frac{\hat{\theta}-\theta}{\frac{1}{\sqrt{n}}} \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right) \qquad (n \to \infty)$$

or

$$\hat{\theta} \overset{\bullet}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$$

11.4.2 Confident Intervals

1. Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of θ:

$$\frac{\hat{\theta}-\theta}{\sqrt{\dfrac{1}{nI(\theta)}}} \overset{\bullet}{\sim} N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{\hat{\theta}-\theta}{\sqrt{\frac{1}{nI(\theta)}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

2. Given $\alpha \in (0,1)$, look for the $1-\alpha$ confident intervals of θ:

$$\frac{\hat{\theta}-\theta}{\sqrt{\dfrac{1}{nI(\hat{\theta})}}} \overset{\bullet}{\sim} N(0,1) \qquad \Longrightarrow \qquad P\left\{\left|\frac{\hat{\theta}-\theta}{\sqrt{\frac{1}{nI(\hat{\theta})}}}\right| < z_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

12 Theory of Statistic Tests

12.1 Best Tests

12.1.1 Definition

The best test is given by the best rejection region.

C is the best rejection region of size α for thr simple hypothesis test if for any rejection region A of size α,

$$P\{(X_1,X_2,\cdots,X_n)\in A;H_0\} = P\{(X_1,X_2,\cdots,X_n)\in C;H_0\} = \alpha$$

$$P\{(X_1,X_2,\cdots,X_n)\in C;H_1\} \geqslant P\{(X_1,X_2,\cdots,X_n)\in A;H_1\}$$

that is

$$\beta_C \leqslant \beta_A$$

12.1.2 ***Neyman − Pearson Theorem***

Let $X_1,X_2,\cdots,X_n$ be i.i.d. with p.d.f. $f(x;\theta)$ $\theta\in\Theta=\{\theta_0,\theta_1\}$. The likelihood function is

$$L(\theta|x_1,x_2,\cdots,x_n) = \prod_{i=1}^{n} f(x_i;\theta).$$

Consider the simple hypothesis

$$H_0:\theta=\theta_0 \qquad\qquad H_1:\theta=\theta_1.$$

Let

$$C = \left\{(x_1,x_2,\cdots,x_n) : \frac{L(\theta_0|x_1,x_2,\cdots,x_n)}{L(\theta_1|x_1,x_2,\cdots,x_n)} \leqslant k\right\}$$

and

$$\alpha = P\{(X_1,X_2,\cdots,X_n)\in C;H_0\},$$

then C is a best rejection region of size α for testing H_0 versus H_1.

And we have type I error

$$\alpha = P\{(X_1,X_2,\cdots,X_n)\in C;H_0\} = \int_C L(\theta_0|\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

and type II error

$$\beta = P\{(X_1,X_2,\cdots,X_n)\in C^*;H_1\} = \int_{C^*} L(\theta_1|\boldsymbol{x})\mathrm{d}\boldsymbol{x},$$

where $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$.

Furthermore, if $X_1,X_2,\cdots,X_n$ are the statistic with the joint distribution $f(x_1,x_2,\cdots,x_n|\theta)$ and the likelihood function is $L(\theta|x_1,x_2,\cdots,x_n)=f(x_1,x_2,\cdots,x_n;\theta)$, the conclusion remains true.

12.2 Uniformly Most Powerful Tests

12.2.1 Definition

The rejection region C gives a uniformly most powerful test for the simple hypothesis H_0 against composite hypothesis H_1 if C is the best rejection region for test H_0 against each simple hypothesis in H_1.

12.2.2 Monotone Likelihood Ratio (MLR)

A family of density functions $\{f_\theta(x)\}_{\theta\in\Theta}$ indexed by a parameter θ taking values in an ordered set Θ is said to have a monotone likelihood ratio in the statistic $T(x_1,x_2,\cdots,x_n)$ if for any $\theta_1 < \theta_2$, $\frac{f(x_1,x_2,\cdots,x_n;\theta_2)}{f(x_1,x_2,\cdots,x_n;\theta_1)}$ is a non-decreasing function of $T(x_1,x_2,\cdots,x_n)$, then

$$\frac{L(\theta_2;x_1,x_2,\cdots,x_n)}{L(\theta_1;x_1,x_2,\cdots,x_n)} \qquad \Longleftrightarrow \qquad T(x)\leqslant c$$

If the family has MLR in $T(x_1,x_2,\cdots,x_n)$, then

1. A UMPT of size α for
$$H_0:\theta=\theta_0 \qquad\qquad H_1:\theta>\theta_0$$
is given by $C=\{(x_1,x_2,\cdots,x_n):T(x_1,x_2,\cdots,x_n)\geqslant c\}$
2. A UMPT of size α for
$$H_0:\theta=\theta_0 \qquad\qquad H_1:\theta<\theta_0$$
is given by $C=\{(x_1,x_2,\cdots,x_n):T(x_1,x_2,\cdots,x_n)\leqslant c\}$
3. A UMPT of size α for
$$H_0:\theta\leqslant\theta_0 \qquad\qquad H_1:\theta>\theta_0$$
is given by $C=\{(x_1,x_2,\cdots,x_n):T(x_1,x_2,\cdots,x_n)\geqslant c\}$
4. A UMPT of size α for
$$H_0:\theta\geqslant\theta_0 \qquad\qquad H_1:\theta<\theta_0$$
is given by $C=\{(x_1,x_2,\cdots,x_n):T(x_1,x_2,\cdots,x_n)\leqslant c\}$

12.3 Likelihood Ratio Tests

12.3.1 Definition

Let $X_1,X_2,\cdots,X_n$ be i.i.d. with p.d.f. $f(x;\theta)$ $\theta\in\Theta=\{\theta_0,\theta_1\}$. The likelihood function is

$$L(\theta|x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n} f(x_i;\theta).$$

Consider the ratio of two likelihood functions

$$\lambda(x_1,x_2,\cdots,x_n)=\frac{\sup\limits_{\theta\in\Theta_0} L(\theta|x_1,x_2,\cdots,x_n)}{\sup\limits_{\theta\in\Theta} L(\theta|x_1,x_2,\cdots,x_n)},$$

gives the rejection region

$$C=\{(x_1,x_2,\cdots,x_n):\lambda(x_1,x_2,\cdots,x_n)\leqslant\lambda_0\},$$

where λ_0 is determined by

$$P(\lambda\leqslant\lambda_0;H_0)=\sup_{\theta\in\Theta_0} P(\lambda\leqslant\lambda_0|\theta)=\alpha.$$

The test is called likelihood ratio test.

12.3.2 λ's Limiting Distribution

$$-2\ln\lambda(X_1, X_2, \cdots, X_n) \overset{\bullet}{\sim} \chi^2(r)$$

with

$$r = dim(\Theta) - dim(\Theta_0)$$

.

13 Inferences About Normal Models

13.1 Quadratic Form

Let $Q = Q_1 + Q_2 + \cdots + Q_k$, where $Q, Q_1, \cdots, Q_k$ are $k+1$ random variables that are real quadratic forms in n independent random variables which are normally distributed with the means $\mu_1, \mu_2, \cdots, \mu_n$ and the same variance σ^2. Let $\dfrac{Q}{\sigma^2} \sim \chi^2(r), \dfrac{Q_1}{\sigma^2} \sim \chi^2(r_1), \cdots, \dfrac{Q_{k-1}}{\sigma^2} \sim \chi^2(r_{k-1})$ and Q_k be nonnegative. Then

1. $Q_1, \cdots, Q_k$ are independent;
2. $\dfrac{Q_k}{\sigma^2} \sim \chi^2(r_k),\ r = r_1 + \cdots + r_k$.

Let $\dfrac{Q}{\sigma^2} \sim \chi^2\left(r, \sum_{i=1}^{n} \mu_i\right), \dfrac{Q_1}{\sigma^2} \sim \chi^2(r_1, \mu_1), \cdots, \dfrac{Q_{k-1}}{\sigma^2} \sim \chi^2(r_{k-1}, \mu_{k-1})$ and Q_k be nonnegative. Then

1. $Q_1, \cdots, Q_k$ are independent;
2. $\dfrac{Q_k}{\sigma^2} \sim \chi^2(r_k, \mu_k),\ r = r_1 + \cdots + r_k$.

Here we only consider the former.

$$\begin{aligned}
Q &= \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X_{..}})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X_{i.}})^2 + b\sum_{i=1}^{a}(\overline{X_{i.}} - \overline{X_{..}})^2 \\
&= Q_1 + Q_2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X_{.j}})^2 + a\sum_{j=1}^{b}(\overline{X_{.j}} - \overline{X_{..}})^2 \\
&= Q_3 + Q_4 \\
&= b\sum_{i=1}^{a}(\overline{X_{i.}} - \overline{X_{..}})^2 + a\sum_{j=1}^{b}(\overline{X_{.j}} - \overline{X_{..}})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X_{i.}} - \overline{X_{.j}} + \overline{X_{..}})^2 \\
&= Q_2 + Q_4 + Q_5
\end{aligned}$$

13.2 Analysis of Variance(ANOVA)

13.2.1 One Way

Let $X_{a_ij} \overset{i.i.d.}{\sim} N(\mu_j, \sigma^2)$, then

$$\begin{aligned}
SST = Q &= \sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \overline{X_{..}})^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \overline{X_{i.}})^2 + \sum_{j=1}^{b} a_j(\overline{X_{.j}} - \overline{X_{..}})^2 \\
&= Q_3 + Q_4 \\
&= SSW + SSB
\end{aligned}$$

If the likelihood ratio λ is used to test

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \qquad\qquad H_1 : \mu_i \neq \mu_j,$$

then when H_0 is true, $\lambda \leqslant \lambda_0$ is equivalent to $F \geqslant c$, where

$$F = \frac{\dfrac{Q_4}{b-1}}{\dfrac{Q_3}{\sum\limits_{j=1}^{b} a_j - b}} \sim F\left(b-1, \sum_{j=1}^{b} a_j - b\right)$$

where $\dfrac{Q}{\sigma^2} \sim \chi^2\left(\sum\limits_{j=1}^{b} a_j - 1\right)$, $\dfrac{Q_3}{\sigma^2} \sim \chi^2\left(\sum\limits_{j=1}^{b} a_j - b\right)$, $\dfrac{Q_4}{\sigma^2} \sim \chi^2(b-1)$, however, when H_0 is false, some of these distributions will be noncentral F-distribution or noncentral χ^2-distribution.

13.2.2 Two Way

Let $X_{ij} \overset{i.i.d.}{\sim} N(\mu_{ij}, \sigma^2)$, $\mu_{ij} = \mu + \alpha_i + \beta_j$ and $\sum\limits_{i=1}^{a} \beta_j = \sum\limits_{j=1}^{b} \alpha_j = 0$, then

$$\begin{aligned}
SST = Q = \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X_{..}})^2 \\
&= b\sum_{i=1}^{a}(\overline{X_{i.}} - \overline{X_{..}})^2 + a\sum_{j=1}^{b}(\overline{X_{.j}} - \overline{X_{..}})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X_{i.}} - \overline{X_{.j}} + \overline{X_{..}})^2 \\
&= Q_2 + Q_4 + Q_5 \\
&= SSA + SSB + SSE.
\end{aligned}$$

If the likelihood ratio λ is used to test

$$H_{A0} : \mu_{1j} = \mu_{2j} = \cdots = \mu_{aj} \qquad\qquad H_{A1} : \mu_{i.} \neq \mu_{j.}$$

or

$$H_{B0} : \mu_{i1} = \mu_{i2} = \cdots = \mu_{ib} \qquad\qquad H_{B1} : \mu_{.i} \neq \mu_{j.},$$

the same as

$$H_{A0} : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0 \qquad\qquad H_{A1} : \alpha_i \neq 0$$

or

$$H_{B0} : \beta_1 = \beta_2 = \cdots = \beta_b = 0 \qquad\qquad H_{B1} : \beta_j \neq 0,$$

then when H_0 is true, $\lambda \leqslant \lambda_0$ is equivalent to $F \geqslant c$, where

$$F = \frac{\dfrac{Q_2}{a-1}}{\dfrac{Q_5}{(a-1)(b-1)}} \sim F\left(a-1, (a-1)(b-1)\right)$$

or

$$F = \frac{\dfrac{Q_4}{b-1}}{\dfrac{Q_5}{(a-1)(b-1)}} \sim F\left(b-1, (a-1)(b-1)\right),$$

$\frac{Q}{\sigma^2} \sim \chi^2(ab-1)$, $\frac{Q_2}{\sigma^2} \sim \chi^2(a-1)$, $\frac{Q_4}{\sigma^2} \sim \chi^2(b-1)$, $\frac{Q_5}{\sigma^2} \sim \chi^2((a-1)(b-1))$, however, when H_0 is false, some of these distributions will be noncentral F-distribution or noncentral χ^2-distribution.

If there is interaction between the two factors, that is $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$ and $\sum_{i=1}^{a} \beta_j = \sum_{j=1}^{b} \alpha_j = \sum_{i=1}^{a} \gamma_{ij} = \sum_{j=1}^{b} \gamma_{ij} = 0$, then

$$
\begin{aligned}
SST = Q &= \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (X_{ijk} - \overline{X_{...}})^2 \\
&= bc \sum_{i=1}^{a} (\overline{X_{i..}} - \overline{X_{...}})^2 + ac \sum_{j=1}^{b} (\overline{X_{.j.}} - \overline{X_{...}})^2 + c \sum_{i=1}^{a} \sum_{j=1}^{b} (X_{ij.} - \overline{X_{i..}} - \overline{X_{.j.}} + \overline{X_{...}})^2 \\
&\quad + \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k}^{c} (X_{ijk} - \overline{X_{ij.}})^2 \\
&= Q_2 + Q_4 + Q_5 + Q_6 \\
&= SSA + SSB + SSAB + SSE.
\end{aligned}
$$

If the likelihood ratio λ is used to test

$$H_{AB0} : \mu_{11} = \cdots = \mu_{21} = \cdots = \mu_{ab} \qquad H_{AB1} : \mu_{ij} \neq \mu_{mn}$$

the same as

$$H_{AB0} : \gamma_{11} = \cdots = \gamma_{21} = \cdots = \gamma_{ab} = 0 \qquad H_{B1} : \gamma_{ij} \neq 0,$$

then when H_0 is true, $\lambda \leqslant \lambda_0$ is equivalent to $F \geqslant c$, where

$$F = \frac{\frac{Q_5}{(a-1)(b-1)}}{\frac{Q_6}{ab(c-1)}} \sim F\left((a-1)(b-1), ab(c-1)\right),$$

$\frac{Q}{\sigma^2} \sim \chi^2(abc-1)$, $\frac{Q_2}{\sigma^2} \sim \chi^2(a-1)$, $\frac{Q_4}{\sigma^2} \sim \chi^2(b-1)$, $\frac{Q_5}{\sigma^2} \sim \chi^2((a-1)(b-1))$, $\frac{Q_6}{\sigma^2} \sim \chi^2(ab(c-1))$, however, when H_0 is false, some of these distributions will be noncentral F-distribution or noncentral χ^2-distribution.