Statistics and Linear Models (exam scope)

Lecture 1
Definition of: likelihood function, log-likelihood function, maximum likelihood estimator.

Lecture 2
Definition of: Fisher information.
Knowing and understanding of: Theorem 2, Corollary 1, Theorem 1.

Lecture 3
Definition of: efficiency of an estimator, asymptotic efficiency of an estimator.
Knowing and understanding of: Corollary 1, Theorem 2.

Lecture 4
Definition of: minimum variance unbiased estimator, sufficient statistic.
Knowing and understanding of: factorization theorem.

Lecture 5
Knowing and understanding of: Rao-Blackwell theorem, Theorem 2.

Lecture 6
Definition of: complete statistic, regular exponential class (positive and negative example).
Knowing and understanding of: Lehmann-Scheffé theorem, Theorem 2.
III Theory of Testing Statistical Hypotheses, Section 1: Definitions 3-5.

Lecture 7
Section 1: Definitions 6-10. Section 2: Definition 1, Theorem 1, and Corollary 1.

Lectures 8
Definition of: family with monotone likelihood ratio, likelihood ratio test.
Knowing and understanding of Karlin-Rubin Theorem.

Lecture 9
Definition of: (asymptotic) likelihood ratio test, Rao score test, Wald test.
Knowing and understanding of Theorem 1.

Lecture 10
Definition of Fisher information matrix.

Lectures 11 and 12
Definition of models and related statistics from Remarks 1-3.
Knowing and understanding of Theorem 1 (distribution of quadratic forms).
Definition of a one-way ANOVA model and the solution (a form of the test statistic).

Lecture 13
Definition of the Linear Model. Knowing and understanding of Theorem 2.

Lecture 14
Knowing and understanding of: Theorem 3 and 4, the form of $\hat{b}$, and Theorem 1 (Section 4.2).

I Maximum Likelihood Methods

Suppose that $X_1, \dots, X_n$ are i.i.d. random variables with common pdf $f(x, \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \geq 1$.

1. Maximum Likelihood Estimation

Definition 1

The function $L : \Theta \to \mathbb{R}$ of the form

$$L(\theta) = L(\theta, \underline{x}) = \prod_{i=1}^{n} f(x_i, \theta), \quad \theta \in \Theta,$$

where $\underline{x} = (x_1, \dots, x_n)$ is called the likelihood function.

The function $\ell : \Theta \to \mathbb{R}$ of the form

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i, \theta), \quad \theta \in \Theta$$

is called the log-likelihood.

Example 1

Let $X_1, \dots, X_n$ denote a random sample from the distribution with pmf

$$p(x) = \begin{cases} \theta^x (1-\theta)^{1-x}, & x = 0, 1, \\ 0, & \text{elsewhere}, \end{cases}$$

where $0 < \theta < 1$. We have

$$P(\underline{X} = \underline{x}) = P((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}.$$

Thus

$$L(\theta) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}, \quad \theta \in (0, 1).$$

<u>Problem</u>: What value of θ maximize the probability $L(\theta)$ of obtaining ~~this particular~~ the observed sample $x_1, \dots, x_n$? Would it be a good estimate of θ?

We have

$$\ell(\theta) = \log L(\theta) = \left(\sum_{i=1}^{n} x_i\right) \log \theta + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-\theta),$$

$$\frac{d\ell(\theta)}{d\theta} = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} = 0,$$

$$(1-\theta)\sum_{i=1}^{n} x_i - \theta\left(n - \sum x_i\right) = \sum x_i - n\theta = 0 \Rightarrow \hat{\theta} = \frac{1}{n}\sum x_i$$

$$\frac{d^2\ell(\theta)}{d\theta^2} = -\frac{\sum x_i}{\theta^2} - \frac{n - \sum x_i}{(1-\theta)^2} < 0$$

The statistic

$$\hat{\theta} = \bar{X}$$

is called the maximum likelihood estimator of θ.

Let θ_0 denote the true value of θ.

<u>Assumptions (Regularity Conditions)</u>

(R0): The pdfs are distinct, i.e., $\theta \neq \theta' \Rightarrow f(x_i; \theta) \neq f(x_i; \theta')$

(R1): The pdfs have common support for all θ.

(R2) $\theta_0 \in \operatorname{int} \Theta$.

<u>Theorem 1</u>

Under assumptions (R0) - (R1)

$$\lim_{n\to\infty} P_{\theta_0}\left[L(\theta_0, \underline{X}) > L(\theta, \underline{X})\right] = 1 \quad \text{for all } \theta \neq \theta_0.$$

Remark 1

Asymptotically, the likelihood function is maximized at the true value θ_0.

Definition 2

We say that $\hat{\theta} = \hat{\theta}(\underline{X})$ is a maximum likelihood estimator (mle) of θ if

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\, L(\theta, \underline{X}).$$

In other words

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta, \underline{X})$$

~~Remark 2~~

~~The mle can not exist or~~

Example 2

$X_1, \dots, X_n$, i.i.d. $X_i \sim f(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \mathbb{1}_{(0,+\infty)}(x)$

$$L(\theta) = \frac{1}{\theta^n} e^{-\frac{\sum_i x_i}{\theta}}$$

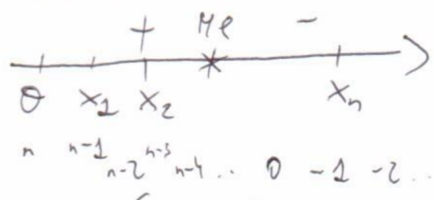
$$\ell(\theta) = -n \log \theta - \frac{1}{\theta} \sum_i x_i$$

$$\frac{d\ell(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{\sum_i x_i}{\theta^2} = 0 \iff n\theta = \sum_i x_i \iff \hat{\theta} = \bar{X}$$

$$\frac{d^2\ell(\theta)}{d\theta^2} = \frac{n}{\theta^2} - \frac{2\sum_i x_i}{\theta^3} = \frac{1}{\theta^3}\left(n\theta - 2\sum_i x_i\right)\Big|_{\theta = \bar{x}} = \frac{1}{\bar{x}^3}\left(n\bar{x} - 2n\bar{x}\right) < 0$$

Example 3

$X_1, \dots, X_n$ i.i.d. $X_i \sim f(x, \theta) = \frac{1}{2} e^{-|x-\theta|}$, $x, \theta \in \mathbb{R}$



$$L(\theta) = \left(\tfrac{1}{2}\right)^n e^{-\sum_i |x_i - \theta|}$$

$$\ell(\theta) = -n \log 2 - \sum_i |x_i - \theta|$$

$$\ell'(\theta) = \sum_i \mathrm{sgn}(x_i - \theta) = 0 \implies \hat{\theta} = \mathrm{med}\{X_1, \dots, X_n\}$$

Example 4

$X_1, \dots, X_n$ i.i.d. $X_i \sim \mathcal{U}(0,\theta)$, $f(x,\theta) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x)$

$$L(\theta) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} \mathbb{1}_{[0,\theta]}(x_i) = \left(\frac{1}{\theta}\right)^n \mathbb{1}_{[0,+\infty)}(x_{(1)}) \mathbb{1}_{(-\infty,\theta]}(x_{(n)})$$

$\forall_i \quad 0 \leq x_i \leq \theta$

$\hat{\theta} = X_{(n)} \qquad 0 \leq x_{(1)}$ & $x_{(n)} \leq \theta$

Theorem 2

Let $X_1, \dots, X_n$ be i.i.d. with the pdf $f(x,\theta)$, $\theta \in \Theta$. For a specified function $g: \Theta \to \mathbb{R}$, let $\eta = g(\theta)$ be a parameter of interest. Suppose $\hat{\theta}$ is the mle of θ. Then $g(\hat{\theta})$ is the mle of $\eta = g(\theta)$.

2. Rao-Cramér Lower Bound and Efficiency

Let X be a random variable with pdf $f(x,\theta)$, $\theta \in \Theta$, where Θ is a open set.

Assumptions (Additional Regularity Conditions)

(R3) The pdf $f(x,\theta)$ is twice differentiable as a function of θ.

(R4) The integral $\int_{\mathbb{R}} f(x,\theta)\,dx$ can be differentiated twice under the ~~sig~~ integral sign as a function of θ.

We have

$$1 = \int_{-\infty}^{+\infty} f(x,\theta)\,dx \quad \Big/ \frac{\partial}{\partial \theta}$$

$$0 = \int_{-\infty}^{+\infty} \frac{\partial f(x,\theta)}{\partial \theta}\,dx.$$

Equivalently,

$$0=\int_{-\infty}^{+\infty}\frac{\frac{\partial f(x,\theta)}{\partial\theta}}{f(x,\theta)}f(x,\theta)\,dx=\int_{-\infty}^{+\infty}\frac{\partial \log f(x,\theta)}{\partial\theta}f(x,\theta)\,dx. \quad (*)$$

Thus
$$\mathbb{E}\left[\frac{\partial \log f(X,\theta)}{\partial\theta}\right]=0.$$

Furthermore, differentiate one more $(*)$, we obtain

$$0=\int_{-\infty}^{+\infty}\frac{\partial^2 \log f(x,\theta)}{\partial\theta^2}f(x,\theta)\,dx+\int_{-\infty}^{+\infty}\frac{\partial \log f(x,\theta)}{\partial\theta}\underbrace{\frac{\partial \log f(x,\theta)}{\partial\theta}f(x,\theta)}_{\frac{\partial f(x,\theta)}{\partial\theta}}dx.$$

Therefore
$$-\int_{-\infty}^{+\infty}\frac{\partial^2 \log f(x,\theta)}{\partial\theta^2}f(x,\theta)\,dx=\mathbb{E}\left[\frac{\partial \log f(X,\theta)}{\partial\theta}\right]^2.$$

Definition 1

The number
$$I(\theta)=\mathbb{E}\left[\left(\frac{\partial \log f(X,\theta)}{\partial\theta}\right)^2\right]$$
is called the Fisher information.

Corollary 1

Under the assumptions (R0) - (R4)
$$I(\theta)=-\int_{-\infty}^{\infty}\frac{\partial^2 \log f(x,\theta)}{\partial\theta^2}f(x,\theta)\,dx=Var\left[\frac{\partial \log f(X,\theta)}{\partial\theta}\right].$$

Example 1

$X\sim b(1,\theta)$, $f(x,\theta)=\theta^x(1-\theta)^{1-x}$

$$\log f(x,\theta)=x\log\theta+(1-x)\log(1-\theta)$$

$$\frac{\partial \log f(x,\theta)}{\partial\theta}=\frac{x}{\theta}-\frac{1-x}{1-\theta}$$

$$\frac{\partial^2 \log f(x,\theta)}{\partial\theta^2}=-\frac{x}{\theta^2}-\frac{1-x}{(1-\theta)^2}$$

Therefore,

$$I(\theta) = -\mathbb{E}\left[-\frac{X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\right] = \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}$$

Example 2:

$X_1, \dots, X_n$ i.i.d. such that

$$X_i = \theta + e_i, \quad i = 1, \dots, n, \qquad \text{(location model)}$$

where $e_1, \dots, e_n$ are i.i.d ~~with~~ $e_i \sim f(x)$.

Then $X_i \sim f_X(x, \theta) = f(x-\theta)$

Assume that f satisfies the regularity conditions. Then

$$I(\theta) = \int_{-\infty}^{+\infty} \left(\frac{f'(x-\theta)}{f(x-\theta)}\right)^2 f(x-\theta)dx = \begin{Bmatrix} z = x-\theta \\ dz = dx \end{Bmatrix} = \int_{-\infty}^{+\infty} \left(\frac{f'(z)}{f(z)}\right)^2 f(z)\,dz.$$

Hence, in the location model, the Fisher information does not depend on θ.

Suppose that X_i has the Laplace distribution, i.e., $f(x,\theta) = \frac{1}{2}e^{-|x_i-\theta|}$

Since

$$X_i = \theta + e_i,$$

$$e_i \sim f(z_i) = \frac{1}{2}e^{-|z_i|}.$$

Furthermore,

$$f'(z) = -\frac{1}{2}e^{-|z|}\,\mathrm{sgn}(z).$$

Therefore,

$$I(\theta) = \int_{-\infty}^{+\infty} \left(\frac{f'(z)}{f(z)}\right)^2 f(z)dz = \int_{-\infty}^{+\infty} f(z)dz = 1$$

Remark 2:

If $X_1, \dots, X_n$ are i.i.d, $X_i \sim f(x,\theta)$ and $I(\theta)$ is the Fisher information of X_1, then $nI(\theta)$ is the Fisher information of the sample.

Proof

$$\mathrm{Var}\left(\frac{\partial \log L(\theta, \underline{X})}{\partial \theta}\right) = \mathrm{Var}\left(\sum_{i=1}^{n} \frac{\partial \log f(X_i, \theta)}{\partial \theta}\right) = \sum_{i=1}^{n} \mathrm{Var}\left(\frac{\partial \log f(X_i, \theta)}{\partial \theta}\right) = nI(\theta)$$

Theorem 1 (Cramér - Rao inequality)

Let $X_1, \ldots, X_n$ be i.i.d. with pdf $f(x, \theta)$, $\theta \in \Theta$.
Assume that the regularity conditions (R0) - (R4) hold.
Let $Y = u(X_1, \ldots, X_n)$ be a statistic with mean $\mathbb{E}Y = \mathbb{E}[u(X_1, \ldots, X_n)] = k(\theta)$. Then

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

Proof

I) (Consider the) continuous case

1. We have $k(\theta) = \mathbb{E}Y = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} u(x_1, \ldots, x_n) f(x_1, \theta) \cdots f(x_n, \theta)\, dx_1 \ldots dx_n$

2. By the above

$$k'(\theta) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} u(x_1, \ldots, x_n) \left[\sum_{i=1}^{n} \frac{1}{f(x_i, \theta)} \frac{\partial f(x_i, \theta)}{\partial \theta}\right] f(x_1, \theta) \cdots f(x_n, \theta)\, dx_1 \ldots dx_n$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} u(x_1, \ldots, x_n) \left[\sum_{i=1}^{n} \frac{\partial \log f(x_i, \theta)}{\partial \theta}\right] f(x_1, \theta) \cdots f(x_n, \theta)\, dx_1 \ldots dx_n$$

3. Define random variable $Z = \sum_{i=1}^{n} \frac{\partial \log f(X_i, \theta)}{\partial \theta}$

4. Then $\mathbb{E}Z = 0$, $\mathrm{Var}\, Z = nI(\theta)$

5. Moreover, $k'(\theta) = \mathbb{E}[Y \cdot Z] = \mathbb{E}Y \cdot \mathbb{E}Z + \rho \sigma_Y \sigma_Z \overset{(4)}{=} \rho \sigma_Y \sqrt{nI(\theta)}$,
where $\rho = \mathrm{corr}(Y, Z)$

6. Thus

$$\rho = \frac{k'(\theta)}{\sigma_Y \sqrt{nI(\theta)}}.$$

7. Since $\rho^2 \leq 1$, we have

$$\frac{[k'(\theta)]^2}{\sigma_Y^2 nI(\theta)} \leq 1 \iff \mathrm{Var}\, Y \geq \frac{[k'(\theta)]^2}{nI(\theta)}$$

Corollary 1

Under the assumptions of Theorem 1, if $Y = u(X_1, \dots, X_n)$ is an unbiased estimator of θ ($k(\theta) = \theta$), then

$$\operatorname{Var} Y \geqslant \frac{1}{nI(\theta)}.$$

Example 3

$X_1, \dots, X_n$ i.i.d $X_1 \sim b(1, \theta)$, $\frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}$

MLE $\hat{\theta} = \bar{X}$, $E\bar{X} = \theta$, $\operatorname{Var} \bar{X} = \frac{\theta(1-\theta)}{n}$

The variance of $\bar{X}$ attains the Cramér-Rao lower bound.

Definition 3

Under the assumptions (R0) - (R4), if $Y = u(X_1, \dots, X_n)$ is an unbiased estimator of a parameter θ, the number

$$e_Y = \frac{\frac{1}{nI(\theta)}}{\operatorname{Var} Y} = \frac{1}{nI(\theta) \operatorname{Var} Y} \qquad e_Y \in [0, 1]$$

is called the efficiency of that estimator.

If $e_Y = 1$ it is said that the estimator Y is efficient.

Example 4

$X_1, \dots, X_n$ i.i.d $X_i \sim \operatorname{Pois}(\theta)$, $\theta > 0$, MLE $\hat{\theta} = \bar{X}$, $E\bar{X} = \theta$, $\operatorname{Var} \bar{X} = \frac{\theta}{n}$

$P(X=x) = \frac{\theta^x e^{-\theta}}{x!}$

We have

$$\frac{\partial \log f(x, \theta)}{\partial \theta} = \frac{\partial}{\partial \theta}(x \log \theta - \theta - \log x!) = \frac{x}{\theta} - 1$$

$$I(\theta) = E\left[\frac{\partial \log f(X, \theta)}{\partial \theta}\right]^2 = E\left[\frac{X}{\theta} - 1\right]^2 = \frac{1}{\theta^2} E[X - \theta]^2 = \frac{1}{\theta^2} \sigma_X^2 = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

$$e_Y = \frac{1}{\left(n \frac{1}{\theta}\right) \frac{\theta}{n}} = 1$$

$Y = \bar{X}$ - efficient estimator of θ.

Example 5

$X_1, \dots, X_n$ i.i.d. $X_i \sim Beta(\theta, 1)$, $f(x, \theta) = \theta x^{\theta-1} \mathbb{1}_{(0,1)}(x)$, $\theta > 0$

$$\log f(x,\theta) = \log\theta + (\theta-1)\log x$$

$$\frac{\partial \log f(x,\theta)}{\partial\theta} = \frac{1}{\theta} + \log x$$

x^{θ}

$$\frac{\partial^2 \log f(x,\theta)}{\partial\theta^2} = -\frac{1}{\theta^2}$$

Thus $I(\theta) = \frac{1}{\theta^2}$.

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i,\theta) = n\log\theta + (\theta-1)\sum_{i=1}^n \log x_i$$

$$\ell'(\theta) = \frac{n}{\theta} + \sum_{i=1}^n \log(x_i) = 0 \Rightarrow \hat{\theta} = -\frac{n}{\sum_{i=1}^n \log X_i}$$

MLE of θ.

$$\ell''(\theta) = -\frac{n}{\theta^2} < 0$$

Let $Y_i = -\log X_i$, $i = 1, \dots, n$.

$$F_Y(x) = P(Y \le x) = P(-\log X \le x) = P(X \ge e^{-x}) =$$

$$1 - (e^{-x})^{\theta} = 1 - e^{-x\theta} \qquad Exp\left(\frac{1}{\theta}\right) \overset{d}{=} \Gamma\left(1, \frac{1}{\theta}\right)$$

$$W = \sum_{i=1}^n Y_i = -\sum_{i=1}^n \log X_i \sim \Gamma\left(n, \frac{1}{\theta}\right)$$

Fact

$$EW^k = \frac{(n+k-1)!}{\theta^k (n-1)!} \quad \text{for } k > -n.$$

Then,

$$E\hat{\theta} = n E[W^{-1}] = n\frac{(n-2)!}{\theta^{-1}(n-1)!} = \theta\frac{n}{n-1}$$

$$E\hat{\theta}^2 = n^2 E[W^{-2}] = n^2\frac{(n-3)!}{\theta^{-2}(n-1)!} = \theta^2\frac{n^2}{(n-1)(n-2)}$$

As a result,

$$\operatorname{Var}\hat{\theta} = \mathbb{E}\hat{\theta}^2 - (\mathbb{E}\hat{\theta})^2 = \theta^2 \frac{n^2}{(n-1)(n-2)} - \theta^2 \frac{n^2}{(n-1)^2} = \theta^2 \frac{n^2(n-1) - n^2(n-2)}{(n-1)^2(n-2)} =$$

(10)

$$\theta^2 \frac{n^2}{(n-1)^2(n-2)},$$

and

$$e_{\hat{\theta}} = \frac{1}{nI(\theta)\operatorname{Var}\hat{\theta}} = \frac{1}{n \cdot \frac{1}{\theta^2} \cdot \theta^2 \frac{n^2}{(n-1)^2(n-2)}} = \frac{(n-1)^2(n-2)}{n^3} < 1$$

but $\frac{(n-1)^2(n-2)}{n^3} \to 1$

$\hat{\theta}$ is not efficient, but is asymptotically efficient.

Assumption (Additional Regularity Conditions)

(R5) The pdf $f(x,\theta)$ is three times differentiable as a function of θ. Further, for all $\theta \in \Theta$, there exists a constant c and a function $M(x)$ such that

$$\left| \frac{\partial^3 \log f(x,\theta)}{\partial \theta^3} \right| \leq M(x) \quad \text{and} \quad \mathbb{E}_{\theta_0}[M(X)] < +\infty$$

for all $\theta_0 - c < \theta < \theta_0 + c$ and all x in the support of X.

Theorem 2

Assume that $X_1, \dots, X_n$ are i.i.d with pdf $f(x,\theta_0)$, for $\theta_0 \in \Theta$ such that the regularity conditions (R0)-(R5) are satisfied. Suppose that Fisher information satisfies $0 < I(\theta_0) < +\infty$. Then any consistent sequence $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ of solutions of the equation $\frac{dL(\theta)}{d\theta} = \frac{dL(\theta, \underline{x}_n)}{d\theta} = 0$ satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{I(\theta_0)}\right).$$

Definition 4

Let $X_1, \dots, X_n$ be i.i.d with the pdf $f(x, \theta)$. Suppose $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, \dots, X_n)$ is an estimator of θ_0 such that $\sqrt{n}\left(\hat{\theta}_{1n} - \theta_0\right) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2_{\hat{\theta}_1}\right)$.

(i) The number

$$e\left(\hat{\theta}_1\right) = \frac{\frac{1}{I(\theta_0)}}{\sigma^2_{\hat{\theta}_1}}$$

is called the asymptotic efficiency of $\hat{\theta}_{1n}$.

(ii) If $e\left(\hat{\theta}_1\right) = 1$, it is said that $\hat{\theta}_{1n}$ is asymptotically efficient.

(iii) Suppose $\hat{\theta}_{2n} = \hat{\theta}_{2n}(X_1, \dots, X_n)$ is an estimate of θ_0 such that $\sqrt{n}\left(\hat{\theta}_{2n} - \theta_0\right) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2_{\hat{\theta}_2}\right)$.

The number

$$e\left(\hat{\theta}_1, \hat{\theta}_2\right) = \frac{\sigma^2_{\hat{\theta}_2}}{\sigma^2_{\hat{\theta}_1}}$$

is called the asymptotic relative efficiency of $\hat{\theta}_{1n}$ with respect to $\hat{\theta}_{2n}$.

Example 6

$X_i = \theta + e_i$, $i = 1, \dots, n$, $e_1, \dots, e_n$ i.i.d

(i) $e_1 \sim$ Laplace distribution $\frac{1}{2}e^{-|x|}$

MLE of θ is $\hat{\theta}_{1n} = Me\{X_1, \dots, X_n\} = Q_2$, $I(\theta_0) = 1$.

Also $\sqrt{n}\left(\hat{\theta}_{1n} - \theta_0\right) \xrightarrow{\mathcal{D}} N(0, 1)$.

Let $\hat{\theta}_{2n} = \bar{X}_n$
CLT implies

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

where $\sigma^2 = \operatorname{Var} X_1 = \operatorname{Var}(e_1 + \theta) = \operatorname{Var} e_1 = E e_1^2 = \int_{-\infty}^{+\infty} z^2 \frac{1}{2} e^{-|z|} dz$

$$= \int_0^{\infty} z^2 e^{-z} dz = \Gamma(3) = 2.$$

Thus, $e(Q_2, \bar{X}) = \frac{2}{1} = 2.$

The sample median is twice as efficient as the sample mean (asymptotically).

(ii) $e_i \sim N(0,1)$.

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \tfrac{\pi}{2}), \qquad \frac{\pi}{2} = \frac{1}{[2f(0)]^2}$$

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \sim N(0,1),$$

$$e(Me, \bar{X}) = \frac{1}{\frac{\pi}{2}} = \frac{2}{\pi} \approx 0.636 \approx \frac{1}{1.57}$$

$\bar{X}$ is 1.57 times more efficient than Q_2.

<u>Corollary 3</u>

Under the assumptions of Theorem 2, suppose $g(x)$ is a continuous function of x which is differentiable at θ_0 such that $g'(\theta_0) \neq 0$. Then

$$\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta_0)\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{[g'(\theta_0)]^2}{I(\theta_0)}\right).$$

<u>3. Numerical finding of MLEs (Newton's method)</u>

$\hat{\theta}^{(0)}$ - initial guess, $\hat{\theta}^{(1)} = \hat{\theta}^{(0)} - \frac{l'(\hat{\theta}^{(0)})}{l''(\hat{\theta}^{(0)})}$, and so on ~~etc~~

<u>Example 1</u>

$X_1, \dots, X_n$ i.i.d. $f(x, \theta) = \frac{\exp\{-(x-\theta)\}}{[1 + \exp\{-(x-\theta)\}]^2}$ $\quad x \in \mathbb{R}$, $\theta \in \mathbb{R}$

<u>No explicit form</u>

II Sufficiency

1. Measures of Quality of Estimators

Suppose $f(x,\theta)$, $\theta \in \Theta$ is the "density" of a variable X. Consider a point estimator $Y_n = u(X_1,..,X_n)$ based on a sample $X_1,..,X_n$.

Definition 1

For a given integer n, $Y = u(X_1,..,X_n)$ is called a minimum variance unbiased estimator, (MVUE), of the parameter θ, if Y is unbiased and its variance is smaller than or equal to the variance of every other unbiased estimate of θ.

Example 1

$X_1,..,X_9$ i.i.d. $X_1 \sim N(\theta, \sigma^2)$.

X_1 - unbiased estimator of θ, $Var\, X_1 = \sigma^2$

$\overline{X}$ - —"— —"—, $Var\, \overline{X} = \frac{\sigma^2}{n}$

Problem: Is there ~~any another~~ an unbiased estimate of θ with variance smaller than $\frac{\sigma^2}{n}$?

2. Sufficient Statistics

Suppose that $X_1,..,X_n$ are i.i.d. random variables with the density $f(x,\theta)$, $\theta \in \Theta$, ~~and~~ Let $Y_1 = u_1(X_1,..,X_n)$ be a statistic.

Example 1

$X_1, \dots, X_n$ i.i.d.

$$X_i \sim f(x,\theta) = \theta^x (1-\theta)^{1-x}, \; x = 0,1, \; \theta \in (0,1).$$

Then,

$$Y_1 = \sum_{i=1}^n X_i \sim f_{Y_1}(y_1,\theta) = \binom{n}{y_1} \theta^{y_1} (1-\theta)^{n-y_1}, \; y_1 = 0,1,\dots,n.$$

We will find the conditional probability

$$P(X_1 = x_1, \dots, X_n = x_n \mid Y_1 = y_1) = P(A \mid B),$$

say, where $y_1 = 0,1,\dots,n$.

(i) If $\sum_{i=1}^n x_i \neq y_1$, then $P(A|B) = 0$ because $A \cap B = \emptyset$.

(ii) If $\sum_{i=1}^n x_i = y_1$, then $A \subset B$, ~~and~~ Thus $A \cap B = A$ & $P(A|B) = \frac{P(A)}{P(B)} =$

$$\frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}}{\binom{n}{y_1} \theta^{y_1} (1-\theta)^{n-y_1}} = \frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}}{\binom{n}{\sum_i x_i} \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}} = \frac{1}{\binom{n}{\sum_i x_i}}$$ and

it does not depend on θ.

In general, let $f_{Y_1}(y_1,\theta)$ be the "density" of the statistic $Y_1 = u_1(X_1, \dots, X_n)$. The conditional probability

$$X_1 = x_1, \dots, X_n = x_n \mid Y_1 = y_1$$

equals

$$\frac{f(x_1,\theta) \cdot \ldots \cdot f(x_n,\theta)}{f_{Y_1}(u_1(X_1,\dots,X_n),\theta)}.$$

Definition 1

The statistic Y_1 is called a sufficient statistic for the parameter θ if and only if

$$\frac{f(x_1,\theta)\cdot\ldots\cdot f(x_n,\theta)}{f_{Y_1}\left(u_1(x_1,\ldots,x_n),\theta\right)} = H(x_1,\ldots,x_n),$$

where $H(x_1,\ldots,x_n)$ does not depend upon θ.

Example 2

$X_1,\ldots,X_n$ i.i.d $X_1 \sim \Gamma(2,\theta)$ $\quad f(x,\theta) = \dfrac{x e^{-\frac{x}{\theta}}}{\Gamma(2)\,\theta^2}\,\mathbb{1}_{(0,+\infty)}(x)$

$Y_1 = \sum_{i=1}^n X_i \sim \Gamma(2n,\theta)$ $\quad f_{Y_1}(y_1,\theta) = \dfrac{1}{\Gamma(2n)\,\theta^{2n}}\,y_1^{2n-1} e^{-y_1/\theta}\,\mathbb{1}_{(0,+\infty)}(y_1)$

We have

$$\frac{\prod_{i=1}^n f(x_i,\theta)}{f_{Y_1}(y_1,\theta)} = \frac{\frac{x_1 e^{-\frac{x_1}{\theta}}}{\Gamma(2)\,\theta^2}\cdots\frac{x_n e^{-\frac{x_n}{\theta}}}{\Gamma(2)\,\theta^2}}{\frac{1}{\Gamma(2n)\,\theta^{2n}}\,y_1^{2n-1}\,e^{-y_1/\theta}} = \frac{\frac{1}{\theta^{2n}}\left(\prod_{i=1}^n x_i\right) e^{-\sum_{i=1}^n x_i/\theta}}{\frac{1}{\Gamma(2n)\,\theta^{2n}}\left(\sum_{i=1}^n x_i\right)^{2n-1} e^{-\sum_{i=1}^n x_i/\theta}} =$$

$$\frac{\Gamma(2n)\left(\prod_{i=1}^n x_i\right)}{\left(\sum_{i=1}^n x_i\right)^{2n-1}} \quad \text{does not depend on } \theta.$$

Y_1 - sufficient statistic for θ.

Example 3

$Y_1 \leqslant \ldots \leqslant Y_n$ - the order statistics of a random sample of size n from the distribution with pdf

$$f(x,\theta) = e^{-(x-\theta)}\,\mathbb{1}_{(\theta,+\infty)}(x).$$

The pdf of $Y_1 = \min\{X_1, \dots, X_n\}$ is

$$f_{Y_1}(y_1, \theta) = n e^{-n(y_1 - \theta)} \mathbb{1}_{(\theta, \infty)}(y_1).$$

We have

$$\forall_i \; \theta < x_i < +\infty$$

$$\frac{\prod_{i=1}^{n} f(x_i, \theta)}{f_{Y_1}(y_1, \theta)} = \frac{\prod_{i=1}^{n} e^{-(x_i - \theta)} \mathbb{1}_{(\theta, +\infty)}(x_i)}{n e^{-n(y_1 - \theta)} \mathbb{1}_{(\theta, \infty)}(y_1)} =$$

$$\frac{e^{-\sum_{i=1}^{n} x_i + n\theta} \mathbb{1}_{(\theta, +\infty)}(x_{(1)}) \mathbb{1}_{(-\infty, +\infty)}(x_{(n)})}{n e^{-n x_{(1)} + n\theta} \mathbb{1}_{(\theta, +\infty)}(x_{(1)})} =$$

$$\frac{e^{-\sum_{i=1}^{n} x_i}}{n e^{-n x_{(1)}}} \quad \text{does not depend on } \theta.$$

Y_1 - sufficient statistic for θ.

<u>Theorem 1 (factorization theorem)</u>

Let $X_1, \dots, X_n$ denote a random sample from a distribution $f(x, \theta)$, $\theta \in \Theta$. The statistic $Y_1 = u_1(X_1, \dots, X_n)$ is a sufficient statistic for θ if and only if there are two nonnegative functions k_1 and k_2, such that

$$\prod_{i=1}^{n} f(x_i, \theta) = k_1\left(u_1(x_1, \dots, x_n), \theta\right) \cdot k_2(x_1, \dots, x_n),$$

where $k_2(x_1, \dots, x_n)$ does not depend upon θ.

Example 4

$X_1, \dots, X_n$ i.i.d $X_1 \sim N(\theta, \sigma^2)$, σ^2 - known

Let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. We have

$$\sum_{i=1}^{n}(x_i-\theta)^2 = \sum_{i=1}^{n}\left[(x_i-\bar{x})+(\bar{x}-\theta)\right]^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\theta)^2$$

because

$$2\sum_{i=1}^{n}(x_i-\bar{x})(\bar{x}-\theta) = 2(\bar{x}-\theta)\sum_{i=1}^{n}(x_i-\bar{x}) = 0.$$

The joint density of $X_1, \dots, X_n$ has the form

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left[-\sum_{i=1}^{n}(x_i-\theta)^2/2\sigma^2\right] = \cancel{\exp}$$

$$\exp\left[-n(\bar{x}-\theta)^2/2\sigma^2\right]\left\{\frac{\exp\left[-\sum_{i=1}^{n}(x_i-\bar{x})^2/2\sigma^2\right]}{\left(\sqrt{2\pi}\,\sigma\right)^n}\right\}$$

Thus, $\bar{X}$ is ~~the~~ a sufficient statistic for the parameter θ.

Example 5

$X_1, \dots, X_n$ i.i.d $X_i \sim f(x,\theta) = \theta x^{\theta-1}\mathbb{1}_{(0,1)}(x)$, $\theta > 0$.

The joint density of $X_1, \dots, X_n$

$$\theta^n\left(\prod_{i=1}^{n} x_i\right)^{\theta-1} = \underbrace{\left[\theta^n\left(\prod_{i=1}^{n} x_i\right)^{\theta}\right]}_{k_1(u_1(x_1,\dots,x_n),\,\theta)} \underbrace{\frac{1}{\prod_{i=1}^{n} x_i}}_{k_2(x_1,\dots,x_n)}$$

$\prod_{i=1}^{n} X_i$ - sufficient statistic for θ.

3. Properties of a Sufficient Statistic

Suppose $X_1, \dots, X_n$ is a random sample with a "density" $f(x, \theta)$, $\theta \in \Theta$

Remark 1

A sufficient statistic is not unique

Proof

Let $Y_1 = u_1(X_1, \dots, X_n)$ be a sufficient statistic for θ.
Let $Y_2 = g(Y_1)$, where $g: \mathbb{R} \xrightarrow{1-1} \mathbb{R}$. Then,

$$\prod_{i=1}^{n} f(x_i, \theta) = k_1[u_1(x_1, \dots, x_n), \theta] \cdot k_2(x_1, \dots, x_n)$$
$$= k_1(y_1, \theta) k_2(\underline{x}) = k_1(g^{-1}(y_2), \theta) k_2(\underline{x}).$$

By the factorization theorem Y_2 is also a sufficient statistic.

Lemma 1

If X_1 and X_2 are random variables such that ~~$Var X_1$~~ the variance of X_2 exist, then

$$\mathbb{E}X_2 = \mathbb{E}[\mathbb{E}[X_2 | X_1]] \quad \text{and} \quad Var X_2 \geqslant Var[\mathbb{E}[X_2 | X_1]].$$

Y_1 - sufficient statistic for θ

Y_2 - unbiased estimator of θ

$$\mathbb{E}[Y_2 | Y_1] = \varphi(Y_1),$$
$$\theta = \mathbb{E}[Y_2] = \mathbb{E}[\varphi(Y_1)],$$
$$Var Y_2 \geqslant Var[\varphi(Y_1)].$$

Theorem 1 (Rao-Blackwell)

Let $X_1, \dots, X_n$ be a random sample with the "density" $f(x;\theta)$, $\theta \in \Theta$. Let $Y_1 = u_1(X_1, \dots, X_n)$ be a sufficient statistic for θ, and let $Y_2 = u_2(X_1, \dots, X_n)$, not a function of Y_1, be an unbiased estimator of θ. Then $E[Y_2 | Y_1] = \varphi(Y_1)$ (a function of the sufficient statistic) is an unbiased estimator of θ, and its variance is smaller than ~~be equal to~~ the variance of Y_2.

ok

Theorem 2

Let $X_1, \dots, X_n$ be a random sample with $f(x;\theta)$, $\theta \in \Theta$. If a sufficient statistic $Y_1 = u_1(X_1, \dots, X_n)$ for θ exists and a maximum likelihood estimator $\hat{\theta}$ of $\hat{\theta}$ exists and is unique, then $\hat{\theta}$ is a function of Y_1.

Proof

1. $f_{Y_1}(y_1;\theta)$ - density of Y_1.
2. We have
$$L(\theta) = L(\theta, x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i;\theta) = f_{Y_1}(y_1;\theta) \cdot H(x_1, \dots, x_n) =$$
$$f_{Y_1}[u_1(x_1, \dots, x_n);\theta] \cdot H(x_1, \dots, x_n),$$ where $H(x_1, \dots, x_n)$ does not depend on θ.
3. Thus L and f_{Y_1} as ~~a~~ functions of θ are maximized simultaneously.
4. Since $\hat{\theta}$ is unique, it also maximized f_{Y_2} and must

be a function of $u_1(x_1, \dots, x_n)$.

Example 1

$X_1, \dots, X_n$ i.i.d $\quad X_i \sim f(x, \theta) = \theta e^{-\theta x} \mathbb{1}_{(0, +\infty)}(x)$, $\theta > 0$

We want to find the MVUE of θ.

$$L(\theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

$$Y_1 = \sum_{i=1}^n X_i \text{ - sufficient statistic}$$

$$\ell(\theta) = \log L(\theta) = n \log \theta - \theta \sum_{i=1}^n x_i$$

$$\ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^n x_i = 0 \Rightarrow \theta = \frac{1}{\bar{x}}$$

$$\ell''(\theta) = -\frac{n}{\theta^2} < 0 \Rightarrow \hat{\theta} = \frac{1}{\bar{X}} \text{ MLE of } \theta.$$

Y_2 - asymptotically unbiased

$X_1 \sim \Gamma(1, \frac{1}{\theta})$, $Y_1 \sim \Gamma(n, \frac{1}{\theta})$, and

$$\mathbb{E}\left[\frac{1}{\bar{X}}\right] = n \mathbb{E}\left[\frac{1}{Y_1}\right] = n \int_0^\infty \frac{1}{x} \frac{\theta^n}{\Gamma(n)} x^{n-1} e^{-\theta x} dx =$$

$$n \int_0^\infty \frac{\theta^n}{\Gamma(n)} x^{n-2} e^{-\theta x} dx = n \frac{\Gamma(n-1)\, \theta}{\Gamma(n)} \int_0^\infty \frac{\theta^{n-1}}{\Gamma(n-1)} x^{n-2} e^{-\theta x} dx =$$

$$\frac{n}{n-1} \theta.$$

Thus $\dfrac{n-1}{\sum_{i=1}^n X_i}$ is the MVUE of the parameter θ.

Remark 2.

Y_2 - estimate of θ, $EY_2 = \theta$, Y_1 - sufficient statistic

$\varphi(Y_1) = \mathbb{E}[Y_2|Y_1]$, $Var\,\varphi(Y_1) \leq Var\,Y_2$.

Y_3 - estimate of θ, $EY_3 = \theta$, Y_3 - is not a suff. stat.

$\psi(Y_3) = \mathbb{E}[\varphi(Y_1)|Y_3]$, $\mathbb{E}\psi(Y_3) = \theta$ and $Var\,\psi(Y_3) \leq Var\,\varphi(Y_1)$

Since Y_3 is not a sufficient statistic, the conditional distribution of Y_1 given Y_3 depends upon θ. Thus $\psi(Y_3)$ is not a statistic. because $\psi(Y_3)$ depends on θ.

Example 2.

X_1, X_2, X_3 i.i.d. $Exp(\theta)$, $\theta > 0$.

$$(X_1, X_2, X_3) \sim \frac{1}{\theta^3} e^{-(x_1+x_2+x_3)/\theta}, \quad x_i > 0, \; i = 1, 2, 3.$$

The factorization theorem implies that

$Y_1 = X_1 + X_2 + X_3$ is a sufficient statistic for θ.

$E Y_1 = \mathbb{E}(X_1 + X_2 + X_3) = 3\theta$. Thus $\mathbb{E}\left[\frac{Y_1}{3}\right] = \theta$. $\bar{X} = \varphi(Y_1) = \frac{Y_1}{3}$

Let $Y_2 = X_2 + X_3$, $Y_3 = X_3$. The one-to-one transformation

$$x_1 = y_1 - y_2, \quad x_2 = y_2 - y_3, \quad x_3 = y_3,$$

has Jacobian

$$J = \begin{vmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{vmatrix} = 1,$$

and the joint distribution of (Y_1, Y_2, Y_3) is

$$g(y_1, y_2, y_3) = \frac{1}{\theta^3} e^{-y_1/\theta} \mathbb{1}(0 < y_3 < y_2 < y_1 < +\infty)$$

The joint distribution of (Y_1, Y_3)

$$g_{13}(y_1, y_3, \theta) = \int_{y_3}^{y_1} g(y_1, y_2, y_3)\, dy_2 = \int_{y_3}^{y_1} \left(\frac{1}{\theta}\right)^3 e^{-y_1/\theta}\, dy_2 =$$

$$= \frac{1}{\theta^3} e^{-y_1/\theta}(y_1 - y_3), \quad 0 < y_3 < y_1 < +\infty.$$

Obviously, $g_3(y_3, \theta) = \frac{1}{\theta} e^{-y_3/\theta}, \quad 0 < y_3 < +\infty.$

The condition distribution of Y_1 given $Y_3 = y_3$, is

$$g_{1|3}(y_1|y_3) = \frac{g_{13}(y_1, y_3, \theta)}{g_3(y_3, \theta)} = \frac{\frac{1}{\theta^3} e^{-y_1/\theta}(y_1 - y_3)}{\frac{1}{\theta} e^{-y_3/\theta}} =$$

$$\frac{1}{\theta^2}(y_1 - y_3) e^{-(y_1 - y_3)/\theta}, \quad 0 < y_3 < y_1 < +\infty$$

We have

$$\mathbb{E}\left[\frac{Y_1}{3} \Big| Y_3\right] = \mathbb{E}\left[\frac{Y_1 - Y_3}{3} \Big| Y_3\right] + \mathbb{E}\left[\frac{Y_3}{3} \Big| Y_3\right] =$$

$$\frac{1}{3}\int_{y_3}^{+\infty} \frac{1}{\theta^2}(y_1 - y_3)^2 e^{-(y_1 - y_3)/\theta}\, dy_1 + \frac{Y_3}{3} = \left\{\begin{array}{l} z = y_1 - y_3 \\ dz = dy_1 \end{array}\right\} =$$

$$\frac{1}{3}\int_0^{+\infty} \frac{1}{\theta^2} z^2 e^{-z/\theta}\, dz + \frac{Y_3}{3} = \frac{1}{3}\int_0^{+\infty} \frac{1}{\theta^2} \frac{\Gamma(3)\theta^3}{\underbrace{\Gamma(3)\theta^3} _{\Gamma(3,\theta)}} z^2 e^{-z/\theta}\, dz + \frac{Y_3}{3} =$$

$$= \frac{1}{3}\Gamma(3)\theta + \frac{Y_3}{3} = \frac{2}{3}\theta + \frac{Y_3}{3}.$$

Put $\psi(Y_3) = \mathbb{E}\left[\frac{Y_1}{3} \Big| Y_3\right] = \frac{2}{3}\theta + \frac{Y_3}{3}$.

$\mathbb{E}\,\psi(Y_3) = \theta$ and $Var\,\psi(Y_3) \leq Var\left(\frac{Y_1}{3}\right)$,

but $\psi(Y_3)$ is not a statistic. (depends upon the parameter θ)

4. Complete statistics and uniqueness of MVUEs.

Let $X_1, \dots, X_n$ be a random sample from the Poisson distribution

$$f(x, \theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0, 1, 2, \dots, \quad \theta > 0.$$

Recall that $Y_1 = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ and

$$g_1(y_1, \theta) = \frac{(n\theta)^{y_1} e^{-n\theta}}{y_1!}, \quad y_1 = 0, 1, \dots .$$

Consider the family $\{ g_1(y_1, \theta) : \theta > 0 \}$.

Suppose that the function $u(Y_1)$ of Y_1 is such that $E_\theta[u(Y_1)] = 0$ for every $\theta > 0$. We shall show that then

$$u(y_1) = 0 \quad \text{for } y_1 = 0, 1, 2, \dots$$

We have, for all $\theta > 0$,

$$0 = E[u(Y_1)] = \sum_{y_1 = 0}^{\infty} u(y_1) \frac{(n\theta)^{y_1} e^{-n\theta}}{y_1!} =$$

$$e^{-n\theta} \left[u(0) + u(1) \frac{n\theta}{1!} + u(2) \frac{(n\theta)^2}{2!} + \dots \right].$$

Since $e^{-n\theta} > 0$, ~~the~~ we have

$$u(0) + u(1) \frac{n\theta}{1!} + u(2) \frac{(n\theta)^2}{2!} + \dots = 0 \quad (\text{polynomial } (\theta))$$

(Polynomial vanishes, then coefficients as well)

$$u(0) = 0, \quad u(1) \frac{n\theta}{1!} = 0, \quad \frac{n^2\theta^2}{2!} u(2) = 0, \dots$$

Thereby

$$u(0) = 0, \quad u(1) = 0, \quad u(2) = 0, \dots .$$

Definition 1

Let Z be a random variable with "the density" from the family $\{h(z,\theta) : \theta \in \Theta\}$.

If

$$\bigwedge_{\theta \in \Theta} \bigwedge_{u: \mathbb{R} \to \mathbb{R}} \mathbb{E}_\theta[u(Z)] = 0 \implies u(z) \equiv 0 \text{ almost surely},$$

then ~~the family~~ $\{h(z,\theta) : \theta \in \Theta\}$ is called a complete family while Z is called a complete statistic.

Let $X_1, \dots, X_n$ be a sample with $f(x,\theta), \theta \in \Theta$.
Let $Y_1 = u_1(X_1, \dots, X_n)$ be a sufficient statistic for θ, while $f_{Y_1}(y_1,\theta)$ is the density of Y_1. If Y_2 is ~~un~~ an unbiased estimator of θ which is not a function of Y_1, then $\varphi(Y_1) = \mathbb{E}[Y_2 | Y_1]$ is also the unbiased estimator of θ. Suppose, there is another function $\cancel{\varphi}$ ψ of Y_1 such that $\mathbb{E}[\psi(Y_1)] = \theta$ for any $\theta \in \Theta$.

Thus

$$\mathbb{E}[\varphi(Y_1) - \psi(Y_1)] = 0 \quad \text{for} \quad \theta \in \Theta.$$

If the ~~sa~~ family $\{f_{Y_1}(y_1,\theta) : \theta \in \Theta\}$ is complete, $\varphi(y_1) - \psi(y_1) = 0$ except on a set of points that has probability zero. Equivalently $\varphi(y_1) = \psi(y_1)$ a.s.
Therefore, $\varphi(Y_1)$ is the unique function of Y_1 such that $\mathbb{E}[\varphi(Y_1)] = \theta$. As a result, the Rao-Blackwell

implies that $\varphi(Y_1)$ is uniquely determined MVUE of θ.

Theorem 1 (Lehmann-Scheffé)

Let $X_1, \ldots, X_n$ be a sample for $f(x, \theta)$, $\theta \in \Theta$.
Let $Y_1 = u_1(X_1, \ldots, X_n)$ be a sufficient statistic for θ, and let the family $\{f_{Y_1}(y_1, \theta) : \theta \in \Theta\}$ be complete.
If there is a function φ of Y_1 that is $\varphi(Y_1)$, an unbiased estimator of θ, then ~~this function~~ $\varphi(Y_1)$ is uniquely determined MVUE of θ.

5. Exponential Class of Distributions

Consider a family of distributions $\{f(x, \theta) : \theta \in \Theta\}$, where $\Theta = \{\theta : \gamma < \theta < \delta\}$ while γ and δ are known constants (they may be $\pm\infty$), and

$$f(x, \theta) = \begin{cases} \exp[p(\theta)K(x) + S(x) + q(\theta)], & x \in S, \\ 0, & \text{elsewhere}, \end{cases}$$

where S is the support of X.

Definition 1

It is said that $f(x, \theta)$ is a member of the regular exponential class if

(i) S does not depend upon θ,

(ii) $p(\theta)$ is a nontrivial continuous function of θ

(iii) - if X is continuous, $K'(x) \not\equiv 0$ and $S(x)$ is a continuous function of $x \in S$
- if X is discrete, $K(x)$ is a nontrivial function of $x \in S$ 〃 〃

Example 1

(i) The family $\{f(x|\theta): 0<\theta<+\infty\}$, where

$$f(x|\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{x^2}{2\theta}} = \exp\left[-\frac{1}{2\theta}x^2 - \log\sqrt{2\pi\theta}\right], \quad x \in \mathbb{R}$$

is a regular exponential class of the continuous type.

(ii) The family $\{f(x,\theta): 0<\theta<+\infty\}$, where

$$f(x,\theta) = \frac{1}{\theta}\mathbb{1}_{(0,\theta)}(x) = \exp\{-\log\theta\}\mathbb{1}_{(0,\theta)}(x)$$

is not a regular exponential class.

Let $X_1,..,X_n$ denote a random sample from a distribution being a regular exponential class. The joint distribution has the form

$$\exp\left[p(\theta)\sum_{i=1}^{n}K(x_i) + \sum_{i=1}^{n}S(x_i) + nq(\theta)\right] \quad \text{for } x_i \in S, \; i=1,..,n.$$

Equivalently

$$\exp\left[p(\theta)\sum_{i=1}^{n}K(x_i) + nq(\theta)\right] \cdot \exp\left[\sum_{i=1}^{n}S(x_i)\right].$$

The factorization theorem implies that $Y_1 = \sum_{i=1}^{n}K(X_i)$ is a sufficient statistic for θ.

Theorem 2.

Let $f(x|\theta)$, $\gamma<\theta<\delta$ be the distribution of a random variable X being a member of a regular exponential class. If $X_1,..,X_n$ is a random sample from $f(x,\theta)$, the statistic $Y_1 = \sum_{i=1}^{n}K(X_i)$ is a complete sufficient stat for θ.

Corollary 1

If $X_1, \dots, X_n$, a random sample, comes from a regular exponential class and φ is a function of $Y_1 = \sum_{i=1}^{n} K(X_i)$ such that $\mathbb{E}[\varphi(Y_1)] = \theta$, then $\varphi(Y_1)$ is a uniquely determined MVUE of the parameter θ.

Example 2

$X_1, \dots, X_n$ i.i.d $X_1 \sim f(x, \theta) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}$, $x \in \mathbb{R}$, $\theta \in \mathbb{R}$, $\sigma^2 > 0$.

Therefore

$$f(x, \theta) = \exp\left\{\frac{\theta}{\sigma^2} x - \frac{x^2}{2\sigma^2} - \log \sqrt{2\pi}\,\sigma - \frac{\theta^2}{2\sigma^2}\right\}$$

is a member of REC. Specifically,

$$p(\theta) = \frac{\theta}{\sigma^2}, \; K(x) = x, \; S(x) = \frac{-x^2}{2\sigma^2} - \log\sqrt{2\pi\sigma^2}, \; q(\theta) = -\frac{\theta^2}{2\sigma^2}.$$

Thus $Y_1 = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic for θ.

Since $\varphi(Y_1) = \frac{Y_1}{n} = \overline{X}$ is an unbiased estimator of θ, $\varphi(Y_1)$ is a uniquely determined MVUE of θ.

$\overline{X}$ is also a complete sufficient statistic for θ because $Y_1 \xrightarrow{1-1} \overline{X}$.

Example 3

$X \sim \mathrm{Pois}(\theta)$, $\theta \in (0, +\infty)$, $S = \{0, 1, 2, \dots\}$

$$f(x, \theta) = e^{-\theta} \frac{\theta^x}{x!} = \exp\left\{(\log\theta)x + \log\left(\frac{1}{x!}\right) + (-\theta)\right\} \quad \text{REC}$$

$p(\theta) = \log\theta$, $K(x) = x$, $S(x) = \log\left(\frac{1}{x!}\right)$, $q(\theta) = -\theta$.

$Y_1 = \sum_{i=1}^{n} X_i$ - complete and sufficient statistic for θ

$\mathbb{E}\, Y_1 = n\theta$ $\quad \varphi(Y_1) = \overline{X}$ - UD MVUE for θ.

III Theory of Testing Statistical Hypotheses

1. Introduction

Let X be a random variable with the distribution $f(x,\theta)$, $\theta \in \Theta$. Let Θ_0 and Θ_1 be $\underbrace{\text{such that}}_{\text{subsets of } \Theta}$

$$\Theta_0 \cup \Theta_1 = \Theta \quad \text{and} \quad \Theta_0 \cap \Theta_1 = \emptyset .$$

Definition 1

Supposition $\theta \in \Theta_0$ is called the null hypothesis and is denoted by $H_0: \theta \in \Theta_0$, while supposition $\theta \in \Theta_1$ is called the alternative hypothesis and is denoted by $H_1: \theta \in \Theta_1$.

Definition 2

The testing formulation

$$H_0: \theta \in \Theta_0$$

against

$$H_1: \theta \in \Theta_1$$

is called the testing problem. Cheking statistal hypotheses is called testing (verifying) hypotheses.

Definition 3

If $\# \Theta_0 = 1$ $\left(\# \Theta_1 = 1\right)$ the hypothesis H_0 (H_1) is called simple. Otherwise, it is said that the hypothesis H_0 (H_1) is composite.

Let $X_1, \dots, X_n$ be a sample with $f(x, \theta)$. Consider the testing problem

$$H_0: \theta \in \Theta_0,$$
$$H_1: \theta \in \Theta_1.$$

Let $\mathcal{X} = \{X_1(\omega), \dots, X_n(\omega) : \omega \in \Omega\}$ be the sample space.

Definition 4

The statistic $T = T(X_1, \dots, X_n)$ allowing one to assert in the above problem is called the test statistic.

Definition 5

The set $C = \{\underline{x} : \underline{x} = (x_1, \dots, x_n), \underline{x} \in \mathcal{X}\}$ such that for $\underline{x} \in C$, $T(\underline{x})$ leads to rejection of the null hypothesis is called the critical region.

Remark 1

The critical region C of the form

(i) $\{\underline{x} : T(\underline{x}) > c_1\}$ for some $c_1 \in R$ is called the right-tailed critical region.

(ii) $\{\underline{x} : T(\underline{x}) < c_2\}$ for some $c_2 \in \mathbb{R}$ is called the left-tailed critical region.

(iii) $\{\underline{x} : T(\underline{x}) > c_3\} \cup \{\underline{x} : T(\underline{x}) < c_4\}$ for some $c_3, c_4 \in \mathbb{R}$ is called two-tailed critical region

(iv) In general the critical region (i) or (ii) is called a one-tailed critical region.

Definition 6

An error relying on rejection of a true null hypothesis H_0 is called the Type I error (the error of the first kind).

An error relying on acceptance of a false null hypothesis H_0 is called the Type II error (the error of the second kind).

Illustration

		Decision	
		H_0	H_1
Truth	H_0	X	Type I error
	H_1	Type II error	X

Definition 7

Let C be a critical region. The measurable function of the form $\mathbb{1}_C(\underline{x})$ is called a (non-randomized) test of the hypothesis H_0 against the alternative H_1 and is denoted by $\varphi(\underline{x})$ or φ, for short.

Definition 8

A number $\alpha \in (0,1)$ is called the significance level.

Remark 2

Usually, $\alpha = 0.01$, $\alpha = 0.05$, $\alpha = 0.1$.

Definition 9

Let $\alpha \in (0,1)$. It is said ~~(called)~~ that the test φ is at the significance level α, if (and only if)

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\varphi(\underline{X})] = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\underline{X} \in C) \leq \alpha .$$

If

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\varphi(\underline{X})] = \alpha,$$

it is said that the test φ has the size α.

Definition 10

The function $\gamma : \Theta \to [0,1]$ defined as follows:

$\gamma(\theta) = P_\theta(\underline{X} \in C) = \mathbb{E}_\theta[\varphi(\underline{X})]$ for $\theta \in \Theta$

is called ~~the~~ power function of the test φ.

The number $\gamma(\theta)$ for $\theta \in \Theta_1$ is called the power of the test φ under the alternative θ.

Remark 3

Statistical tests are constructed in such a manner in order to minimize the probability of making the Type II error under ~~give~~ fixed probability of making the Type I error equals α.

2. Neyman - Pearson Lemma

Definition 1

It is said that the test φ_0 is ~~the~~ uniformly most powerful at the significance level α, if for any another test φ at the same significance level

$$\mathbb{E}_\theta[\varphi(\underline{X})] \leq \mathbb{E}_\theta[\varphi_0(\underline{X})] \quad \text{for } \cancel{\text{each}}\ \text{any } \theta \in \Theta_1.$$

Theorem 1 (Neyman Pearson Lemma)

Let $X_1, \ldots, X_n$ be a sample with $f(x, \theta)$.

Consider the testing problem

$$H_0: \theta \in \Theta_0,$$
$$H_1: \theta = \theta_1,$$

and the α-size φ_0 test of the form

$$\varphi_0(\underline{x}) = \begin{cases} 1, & \text{if } \prod_{i=1}^n f(x_i, \theta_0) < k \prod_{i=1}^n f(x_i, \theta_1), \\ \gamma, & \text{if } \qquad = \quad , \\ 0, & \text{if } \qquad > \quad , \end{cases}$$

where the constants k and γ ~~are~~ satisfy the condition $\mathbb{E}_{\theta_0}[\varphi(\underline{X})] = \alpha$. Then, φ_0 is the UMP ~~test~~ in the problem (H_0, H_1).

Corollary 1

Under the conditions of Theorem 1, $\gamma_{\varphi_0}(\theta_1) \geq \alpha$.

Example 1

$X_1, \dots, X_n$ i.i.d $X_i \sim f(x, \theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta)^2}{2}\right\}$, $x \in \mathbb{R}$.

We verify

$$H_0: \theta = 0,$$
$$H_1: \theta = 1.$$

We have

$$\frac{L(1, \underline{x})}{L(0, \underline{x})} = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^n (x_i - 1)^2}{2}\right)}{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2}\right)} = \exp\left(\sum_{i=1}^n x_i - \frac{n}{2}\right) > k_1.$$

Thus

$$\sum_{i=1}^n x_i - \frac{n}{2} > \log k_1 \iff \sum_{i=1}^n x_i > \log k_1 + \frac{n}{2} = k.$$

Thereby, the critical region of the UMP test has the form

$$C = \{\underline{x} : \sum_{i=1}^n x_i > k\},$$

while the constant k satifies the condition $P_0(\underline{X} \in C) = \alpha$

So, $P_0\left(\sum_{i=1}^n X_i > k\right) = \alpha$. Since, under H_0, $\sum_{i=1}^n X_i \sim N(0, n)$, we have $P_0\left(\sum_{i=1}^n X_i > k\right) = P_0\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}} > \frac{k}{\sqrt{n}}\right) =$

$1 - \Phi\left(\frac{k}{\sqrt{n}}\right) = \alpha$. As a result, $\frac{k}{\sqrt{n}} = \Phi^{-1}(1-\alpha)$ and $k = \sqrt{n}\,\Phi^{-1}(1-\alpha)$.

On the other hand, under H_1, $\sum_{i=1}^n X_i \sim N(n, n)$, and

$$\gamma(\theta_1) = \gamma(1) = P_1\left(\sum_{i=1}^n X_i > k\right) = P_1\left(\frac{\sum_{i=1}^n X_i - n}{\sqrt{n}} > \frac{k-n}{\sqrt{n}}\right) =$$

$$1 - \Phi\left(\frac{k-n}{\sqrt{n}}\right) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \sqrt{n}\right).$$

3. The UMP for a composite ~~hyp~~ alternative

Example 1

$X_1, \dots, X_n$ i.i.d , $X_1 \sim N(0, \theta)$, $\theta > 0$.

We test

$$H_0: \theta = \theta_0,$$
$$A: \theta > \theta_0.$$

We find the UMP α-level test.

We have

$$L(\theta) = L(\theta, \underline{x}) = \left(\frac{1}{2\pi\theta}\right)^{n/2} \exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n x_i^2\right\}.$$

Let $\theta_1 > \theta_0$, and $k_1 > 0$. Then,

$$\frac{L(\theta_1)}{L(\theta_0)} \geq k_1 \iff \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left\{-\frac{\theta_0 - \theta_1}{2\theta_0\theta_1}\sum_{i=1}^n x_i^2\right\} \geq k_1 \iff$$

$$\sum_{i=1}^n x_i^2 \geq k$$

The critical region has a form

$$C = \left\{(x_1, \dots, x_n) : \sum_{i=1}^n x_i^2 \geq k\right\}$$

and correspods the the UMP test in the problem

$$H_0: \theta = \theta_0,$$
$$H_1: \theta = \theta_1,$$

where the constant k satisfies the condition $P_0\left(\sum_{i=1}^n X_i^2 \geq k\right) = \alpha$.

Since $\sum_{i=1}^n X_i^2/\theta \sim \chi^2_n$, we have

$$P_0\left(\sum_{i=1}^n X_i^2 \geq k\right) = P_0\left(\sum_{i=1}^n X_i^2/\theta_0 \geq \frac{k}{\theta_0}\right) = 1 - F_{\chi^2(n)}\left(\frac{k}{\theta_0}\right).$$

As a result,
$$\frac{k}{\theta_0} = q_{\chi^2(n)}(0.95)$$

Since the above holds for any $\theta_1 > \theta_0$, the UMP test in the problem (H_0, A) has the form $\square_C$.

Example 2

$X_1, \dots, X_n$ i.i.d, $X_i \sim N(\theta, 1)$

We verify

$$H_0: \theta = \theta_0,$$
$$H_1: \theta \neq \theta_0.$$

It will be shown that there is no UMP test in the above problem. We have $\Theta = \mathbb{R}$. Let $\theta_1 \in \Theta$ and $\theta_1 \neq \theta_0$. Consider

$$\frac{L(\theta_1)}{L(\theta_0)} = \frac{\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta_1)^2\right]}{\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta_0)^2\right]} \geq k.$$

Equivalently,

$$\exp\left[(\theta_2 - \theta_0)\sum_{i=1}^{n} x_i - \frac{n}{2}(\theta_2^2 - \theta_0^2)\right] \geq k$$

or

$$(\theta_1 - \theta_0)\sum x_i \geq \frac{n}{2}(\theta_1^2 - \theta_0^2) + \log k,$$

so,

$$\sum_{i=1}^{n} x_i \geq \frac{n}{2}(\theta_2^2 + \theta_0^2) + \frac{\log k}{\theta_1 - \theta_0} \quad \text{provided that } \theta_1 > \theta_0$$

or

$$\sum_{i=1}^{n} x_i \leq \frac{n}{2}(\theta_1 + \theta_0) + \frac{\log k}{\theta_1 - \theta_0} \quad \text{provided that } \theta_1 > \theta_0.$$

The first expression defines the critical region of the UMP test in the problem of testing $H_0: \theta = \theta_0$ against $H_1^+: \theta = \theta_1$ provided that $\theta_1 > \theta_0$, while the second expression defines the critical region of the UMP test ... $H_0: \theta = \theta_0$ ~ $H_1^-: \theta = \theta_1$ in $\theta_1 < \theta_0$.

The UMP test does not exist because these regions are different.

Remark 1

Under the assumptions of Example 2, there are UMP tests in the problems

$$H_0: \theta = \theta_0, \quad H_1^+: \theta > \theta_0, \qquad \text{and} \qquad H_0: \theta = \theta_0, \quad H_1^-: \theta < \theta_0.$$

Let $X_1, \dots, X_n$ be ~~th~~ sample from $f(x, \theta)$, $\theta \in \Theta$, and $Y = u(X_1, \dots, X_n)$ a sufficient statistic for θ.

Factorization theorem implies

$L(\theta, \underline{x}) = k_1[u(\underline{x}), \theta] \cdot k_2(\underline{x})$, where $k(\underline{x})$ does not depend on θ. Thereby, for $\theta_0, \theta_1 \in \Theta$

$$\frac{L(\theta_1, \underline{x})}{L(\theta_0, \underline{x})} = \frac{k_1[u(\underline{x}), \theta_1]}{k_1[u(\underline{x}), \theta_0]}$$

depends on $x_1, \dots, x_n$ only by $u(\underline{x})$.

Corollary 1

The UMP test is a function of a sufficient statistic.

Definition 1

It is said that the family of distributions ~~likelihood~~ $\{L(\theta, x): \theta \in \Theta\}$ has monotone likelihood ratio with respect to a statistic $Y = u(\underline{X})$ if for all $\theta_0, \theta_1 \in \Theta$ such that $\theta_0 < \theta_1$ the ratio

$$\frac{L(\theta_1, \underline{x})}{L(\theta_0, \underline{x})}$$

is a nondecreasing function of $y = u(\underline{x})$.

Theorem 1 (Karlin-Rubin) (continuous version)

Consider the testing problem

$$H_0 : \theta \leq \theta_0,$$
$$H_1 : \theta > \theta_0.$$

If the family of distributions $\{L(\theta, \underline{x}) : \theta \in \Theta\}$ has monotone likelihood ratio with respect to a statistic $Y = u(\underline{X})$, then, in the problem (H_0, H_1), there is UMP test with the critical region

$$C = \{\underline{x} : u(\underline{x}) > k\},$$

where the constant satisfies the condition $P_{\theta_0}(Y > k) = P_{\theta_0}(u(\underline{X}) > k) = \alpha$.

Example 3

$X_1, \dots, X_n$ i.i.d $X_i \sim b(1, p)$, $p = \theta$, $0 < \theta < 1$. Let $\theta_1 > \theta_0$ and consider the ratio

$$\frac{L(\theta_1, \underline{x})}{L(\theta_0, \underline{x})} = \frac{\theta_1^{\sum_{i=1}^n x_i}(1-\theta_1)^{n-\sum x_i}}{\theta_0^{\sum x_i}(1-\theta_0)^{n-\sum x_i}} = \left[\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right]^{\sum_{i=1}^n x_i}\left(\frac{1-\theta_1}{1-\theta_0}\right)^n$$

which is an increasing function of $y = u(\underline{x}) = \sum_{i=1}^n x_i$.

As a result, the family of distributions $\{L(\theta, \underline{x}) : \theta \in (0,1)\}$ has monotone likelihood ratio with respect to $Y = \sum_{i=1}^n X_i$.

Consider the testing problem

$$H_0 : \theta \leq \theta_0$$

against

$$H_1 : \theta > \theta_0.$$

The test Δ_C, where $C = \{\underline{x} : \sum_{i=1}^n x_i \geq k\}$, is the UMP test in the problem (H_0, H_1) at the significance level α if $P_{\theta_0}(\sum X_i \geq k) = \alpha$.

Let $X_1, \dots, X_n$ be the sample from $f(x,\theta), \theta \in \Theta$,
where $f(x,\theta) = \exp[p(\theta)K(x) + S(x) + q(\theta)]$, $x \in \mathcal{S}$, while
$\mathcal{S}$ is the support of X_1 independent from θ.

Assume that $p(\theta)$ is an increasing function of θ.

Then, for $\theta_1 > \theta_0$, the ratio

$$\frac{L(\theta_1)}{L(\theta_0)} = \exp\left\{[p(\theta_1) - p(\theta_0)]\sum_{i=1}^{n} K(x_i) + n[q(\theta_1) - q(\theta_0)]\right\}$$

is a nondecreasing function of $Y = \sum_{i=1}^{n} K(X_i)$.

Thus, in the testing problem

$H_0: \theta \leq \theta_0$,

$H_1: \theta > \theta_0$,

there exists the UMP test of the form $\mathbb{1}_C$,
where $C = \{\underline{x} : \sum_{i=1}^{n} K(x_i) \geq k\}$ while k satisfies
the condition $\alpha = P_{\theta_0}(\underline{X} \in C)$

4. Likelihood Ratio Tests

Let $X_1, \dots, X_n$ be a sample with $f(x,\theta), \theta \in \Theta$.
We test

$H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$,

where θ_0 is a fixed constant.

Let $\hat{\theta}$ be MLE of the parameter θ.

Def 1

The test with a critical region $C = \{\Lambda \leq c\}$, where

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}$$

is called the likelihood ratio test in the problem (H_0, H_1). If $P_{\theta_0}(\Lambda \leq c) = \alpha$, the test has a size α.

Example 1

$X_1, \ldots, X_n$ i.i.d $X_i \sim f(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x, \theta > 0$.

Then
$$L(\theta) = \theta^{-n} \exp\{-\tfrac{n}{\theta}\bar{X}\}.$$

Furthermore, $\hat{\theta} = \bar{X}$ is the MLE of θ. We have

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = \frac{\theta_0^{-n} \exp\{-\frac{n}{\theta_0}\hat{\theta}\}}{\hat{\theta}^{-n} \exp\{-\frac{n}{\hat{\theta}}\hat{\theta}\}} = \left(\frac{\hat{\theta}}{\theta_0}\right)^n \exp\{-\tfrac{n}{\theta_0}\hat{\theta} + n\}.$$

A critical region has the form

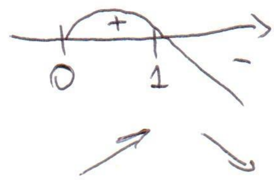
$$\Lambda \leq c_\alpha,$$

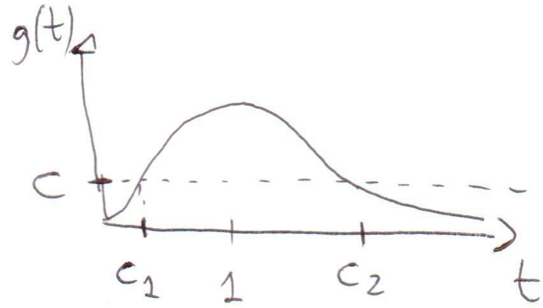
where $\Lambda = g(t) = t^n e^{-nt}$, while $t = \frac{\bar{X}}{\theta_0}$.

Moreover, $g'(t) = n t^{n-1} e^{-nt} + t^n e^{-nt}(-n) = n e^{-nt}(t^{n-1} - t^n) = 0$

$n e^{-t} \cdot t^{n-1}(1-t)$

$\Leftrightarrow t = 1, \; t = 0$





Thereby,

$\Lambda \leq c \Leftrightarrow g(t) \leq c \Leftrightarrow t \leq c_1$ or $t \geq c_2$

So,

$$\frac{\bar{X}}{\theta_0} \leq c_1 \text{ or } \frac{\bar{X}}{\theta_0} \geq c_2.$$

Under H_0, the statistic $\frac{2}{\theta_0}\sum_{i=1}^n X_i \sim \chi^2(2n)$. As a result, the critical region of the α-size LRT has the form

$$C = \{\underline{x} : \tfrac{2}{\theta_0}\sum_{i=1}^n x_i \leq q_{\chi^2_{2n}}(\tfrac{\alpha}{2})\} \cup \{\underline{x} : \tfrac{2}{\theta_0}\sum_{i=1}^n x_i \geq q_{\chi^2_{2n}}(1-\tfrac{\alpha}{2})\},$$

where $q_{\chi^2_{2n}}(\alpha)$ is the α-quantile of the chi-square distr. with 2 d.of.f.

Example 2.

$X_1, \dots, X_n$ i.i.d, $X_1 \sim N(\theta, \sigma^2)$, $\theta \in \mathbb{R}$, $\sigma^2 > 0$ and known.

We verify

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta \neq \theta_0 ,$$

where θ_0 is fixed. We have

$$L(\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta)^2\right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\} \exp\left\{-\frac{1}{2\sigma^2} n(\bar{x} - \theta)^2\right\}$$

Furthermore, $\hat{\theta} = \bar{X}$ is the MLE of θ. Thus,

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = \exp\left\{-\frac{1}{2\sigma^2} n(\bar{X} - \theta_0)^2\right\},$$

and $\Lambda \leqslant c$ is equivalent to $-2\log\Lambda \geqslant -2\log c$.

Under H_0, $-2\log\Lambda = \left(\frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim \chi^2(1)$.

We reject H_0 in favour of H_1, if

$$-2\log\Lambda \geqslant q_{\chi^2(1)}(1-\alpha).$$

Theorem 1

Let $X_1, \dots, X_n$ be a sample with $f(x, \theta)$, $\theta_0 \in \Theta \subseteq \mathbb{R}$ satisfying the regularity conditions (R0) - (R5).

Under $H_0 : \theta = \theta_0$,

$$-2\log\Lambda \xrightarrow{\mathcal{D}} \chi^2(1).$$

Remark 1

If there is a problem with finding an exact form of the statistic Λ, we can apply the test based on the statistic $\chi^2_L = -2\log\Lambda$ at the asymptotic significance level α rejecting H_0 in favour of H_1 when

$$\chi^2_L \geq q_{\chi^2(1)}(1-\alpha).$$

Definition 2

In the testing problem $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$ the test based on the statistic

$$\chi^2_W = \left\{\sqrt{nI(\hat{\theta})}\,(\hat{\theta} - \theta_0)\right\}^2$$

is called the Wald test. We reject H_0, at the asymptotic significance level α, when

$$\chi^2_W \geq q_{\chi^2(1)}(1-\alpha).$$

Definition 3

In the problem of verifying $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$ the test based on the statistic

$$\chi^2_R = \left\{\frac{\ell'(\theta_0)}{\sqrt{nI(\theta_0)}}\right\}^2$$

is called the Rao-score test. We reject H_0, at the asymptotic significance level α, when

$$\chi^2_R \geq q_{\chi^2(1)}(1-\alpha).$$

Example 3

$X_1, \dots, X_n$ i.i.d , $X_i \sim B(\theta, 1)$

We test

$$H_0 : \theta = 1 \quad \text{against} \quad H_1 : \theta \neq 1.$$

Under H_0, $X_i \sim \mathcal{U}(0,1)$,

Moreover, $\hat{\theta} = \dfrac{-n}{\sum_{i=1}^{n} \log X_i}$ - ~~EN~~ MLE of θ.

We have,

$$f(x,\theta) = \frac{\Gamma(\theta+1)}{\Gamma(\theta)\Gamma(1)} x^{\theta-1}(1-x)^{1-1} = \theta x^{\theta-1} \mathbb{1}_{(0,1)}(x)$$

$$L(\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1}, \quad L(1) = 1$$

$$L(\hat{\theta}) = \left(\frac{-n}{\sum \log X_i}\right)^n \left(\prod_{i=1}^{n} x_i\right)^{-\frac{n}{\sum \log X_i} - 1} =$$

$$n^n \left(-\sum \log X_i\right)^{-n} \exp\left(\log \left(\prod_{i=1}^{n} x_i\right)^{-\frac{n}{\sum \log X_i} - 1}\right) =$$

$$n^n \left(-\sum \log X_i\right)^{-n} \exp\left(\left[\frac{n}{\sum \log X_i} - 1\right] \log \prod_{i=1}^{n} X_i\right) =$$

$$\exp(n \log n) \left(-\sum \log X_i\right)^{-n} \exp\left[-n - \sum \log X_i\right] =$$

$$\left(-\sum_{i=1}^{n} \log X_i\right)^{-n} \exp\left(-\sum_{i=1}^{n} \log X_i\right) \exp\left[n(\log n - 1)\right]$$

Thus

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = \frac{1}{L(\hat{\theta})}. \quad \text{Therefore,}$$

$$\chi^2_L = -2 \log \Lambda = -2\left\{-n \log\left(-\sum_{i=1}^{n} \log X_i\right) - \sum_{i=1}^{n} \log X_i + n(\log n - 1)\right\}$$

Recall that $I(\theta) = \theta^{-2}$. As a result

$$\chi_W^2 = \left\{\sqrt{nI(\hat\theta)}\,(\hat\theta - \theta_0)\right\}^2 = \left\{\sqrt{\frac{n}{\hat\theta^2}}\,(\hat\theta - 1)\right\}^2 = n\left(1 - \frac{1}{\hat\theta}\right)^2 =$$

$$= n\left(1 + \frac{\sum_{i=1}^n \log X_i}{n}\right)^2.$$

Furthermore,

$$\ell'(1) = \ell'(\theta_0) = \sum_{i=1}^n \frac{\partial \log f(X_i,\theta)}{\partial\theta}\Big|_{\theta=\theta_0} = \sum_{i=1}^n \frac{\partial \log\left(\theta X_i^{\theta-1}\right)}{\partial\theta}\Big|_{\theta=\theta_0}$$

$$= \sum_{i=1}^n \frac{\partial\left[\log\theta + (\theta-1)\log X_i\right]}{\partial\theta}\Big|_{\theta=\theta_0} = \sum_{i=1}^n \left(1 + \log X_i\right) = n + \sum_{i=1}^n \log X_i$$

Finally,

$$\chi_R^2 = \left\{\frac{\ell'(\theta_0)}{\sqrt{nI(\theta_0)}}\right\}^2 = \left(\frac{\sum \log X_i + n}{\sqrt{n}}\right)^2 = n\left(1 + \frac{\sum_{i=1}^n \log X_i}{n}\right)^2$$

Example 4

Consider the shift model

$$X_i = \theta + e_i, \quad i = 1,\dots,n,$$

where $e_i \sim f(x) = \frac{1}{2}e^{-|x|}$.

We test

$$H_0: \theta = \theta_0,$$
$$H_1: \theta \neq \theta_0.$$

MLE of θ is $\hat\theta = \text{med}\{X_1,\dots,X_n\}$, $\quad X_i \sim f(x,\theta) = \frac{1}{2}\exp\{-|x-\theta|\}$

$$L(\theta_0) = 2^{-n}\exp\left\{-\sum_{i=1}^n |X_i - \theta_0|\right\}$$

$$L(\hat\theta) = 2^{-n}\exp\left\{-\sum_{i=1}^n |X_i - \hat\theta|\right\}$$

So

$$-2\log\Lambda = -2\log\frac{L(\theta_0)}{L(\hat\theta)} = 2\left[\sum_{i=1}^n |X_i - \theta_0| - \sum_{i=1}^n |X_i - \hat\theta|\right]$$

We reject H_0, at the asymptotic significance level α

when $$-2\log\Lambda \geq q_{\chi^2(1)}(1-\alpha).$$

Since $I(\theta) = 1$, $\chi^2_W = \{\sqrt{n}(\hat{\theta} - \theta_0)\}^2$.

Furthermore,

$$\frac{\partial \log f(x_i - \theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\left[\log\tfrac{1}{2} - |x_i - \theta|\right] = \operatorname{sgn}(x_i - \theta)$$

Finally,

$$\chi^2_R = \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\operatorname{sgn}(X_i - \theta_0)\right\}^2.$$

5. Likelihood Ratio Tests: multidimensional case

Let $X_1,\dots,X_n$ be a sample from $f(x,\underline{\theta})$, $\underline{\theta} = (\theta_1,\dots,\theta_p) \in \Theta \subset R^p$, $p>1$

The likelihood function has a form

$$L(\underline{\theta}) = \prod_{i=1}^{n} f(x_i,\underline{\theta}),$$

the log-likelihood

$$\ell(\underline{\theta}) = \log L(\underline{\theta}) = \sum_{i=1}^{n}\log f(x_i,\underline{\theta}).$$

Let $\underline{\theta}_0$ be ~~the~~ three value of the parameter θ.

Impose (additional) regularity conditions

(R6) There is an open subset $\Theta_0 \subseteq \Theta$, such that $\underline{\theta}_0 \in \Theta_0$ and all third partial derivatives of $f(x,\underline{\theta})$ exist for all $\underline{\theta} \in \Theta_0$.

(R7) $$E_{\underline{\theta}}\left[\frac{\partial}{\partial\theta_j}\log f(x,\underline{\theta})\right] = 0 \quad \text{for } j = 1,\dots,p.$$

$$I_{jk}(\underline{\theta}) = \operatorname{Cov}\left(\frac{\partial\log f(X,\underline{\theta})}{\partial\theta_j}, \frac{\partial\log f(X,\underline{\theta})}{\partial\theta_k}\right) =$$

$$= -E_{\underline{\theta}}\left[\frac{\partial^2\log f(X,\underline{\theta})}{\partial\theta_j\partial\theta_k}\right] \quad \text{for } j,k = 1,\dots,p.$$

(R8) For all $\underline{\theta} \in \Theta_0$,

$$I(\underline{\theta}) = \left[I_{jk}(\underline{\theta}) \right]_{j,k=1}^{p}$$

is positive definite.

(R9) There exist ~~functio~~ functions $M_{jkl}(x)$, such that

$$\left| \frac{\partial^3 \log f(x, \underline{\theta})}{\partial \theta_j \, \partial \theta_k \, \partial \theta_l} \right| \leq M_{jkl}(x) \quad \text{for all } \underline{\theta} \in \Theta_0$$

and

$$\mathbb{E}_{\underline{\theta}_0} \left[M_{jkl}(X) \right] < +\infty \quad \text{for all } j,k,l = 1, \ldots, p.$$

Definition 1

The quantity

$$\hat{\underline{\theta}} = \underset{\underline{\theta} \in \Theta}{\operatorname{argmax}} \, L(\underline{\theta})$$

is called the maximum likelihood estimator of the parameter $\underline{\theta}$.

Remark 1

If $\hat{\underline{\theta}}$ is the MLE of $\underline{\theta}$, then $g(\hat{\underline{\theta}})$ is the MLE of $g(\underline{\theta})$

Example 1

$X_1, \ldots, X_n$ i.i.d $X_1 \sim N(\mu, \sigma^2)$. Let $\underline{\theta} = (\mu, \sigma^2)$

and $\Theta = \mathbb{R} \times (0, +\infty)$. We have

$$L(\underline{\theta}) = L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2} \right\}$$

$$= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\},$$

$$l(\underline{\theta}) = \log L(\underline{\theta}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Moreover,

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

$$\frac{\partial l(\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Leftrightarrow \; -\frac{n}{2\sigma^4}\left[\sigma^2 - \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right] = 0$$

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

$$\frac{\partial^2 l(\mu,\sigma^2)}{\partial\mu^2} = -\frac{n}{\sigma^2}\bigg|_{\sigma=\hat{\sigma}} = -\frac{n}{\hat{\sigma}^2}$$

$$\frac{\partial^2 l(\mu,\sigma^2)}{\partial\mu\,\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu) = \frac{\partial^2 l(\mu,\sigma^2)}{\partial\sigma^2\,\partial\mu}\bigg|_{\substack{\mu=\hat{\mu}\\ \sigma^2=\hat{\sigma}^2}} = 0$$

$$\frac{\partial^2 l(\mu,\sigma^2)}{\partial^2\sigma^4} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2\bigg|_{\substack{\mu=\hat{\mu}\\ \sigma^2=\hat{\sigma}^2}} =$$

$$= \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6}\, n\hat{\sigma}^2 = \frac{n}{2\hat{\sigma}^4} - \frac{n}{\hat{\sigma}^4} = -\frac{n}{2\hat{\sigma}^4}.$$

To sum up,

$$\frac{\partial^2 l(\mu,\sigma^2)}{\partial\cdot\,\partial\cdot} = \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{bmatrix} \qquad -\frac{n}{\hat{\sigma}^2} < 0 \;\&\; \underset{\displaystyle \frac{n^2}{2\hat{\sigma}^6}}{\underset{\shortparallel}{\det|\;|}} > 0$$

Indeed, $\cancel{\hat{\mu}=\bar{X}}$

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

is the MLE of $\underline{\theta} = (\mu, \sigma^2)$.

Under H_0, the statistic T has the t-Student distribution with $n-1$ degrees of freedom. Taking $c'' = q_{t(n-1)}\left(1-\frac{\alpha}{2}\right)$, we obtain the α-size test.

Remark 1

If σ^2 is known, the LRT statistic has the form

$$U = \frac{\sqrt{n}\,(\bar{X}-\mu_0)}{\sigma}.$$

Under H_0, $U \sim \mathcal{N}(0,1)$. We reject H_0 when $|U| > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$.

Remark 2

Let $X_1, \ldots, X_m$ be a sample from $\mathcal{N}(\mu_1, \sigma_1^2)$ distribution, and $Y_1, \ldots, Y_n$ be an independent from X_i's sample from $\mathcal{N}(\mu_2, \sigma_2^2)$ distribution. We test

$$H_0: \sigma_1^2 = \sigma_2^2,$$
$$H_1: \sigma_1^2 \neq \sigma_2^2.$$

(i) If μ_1, μ_2 are not known, the LRT statistic has the form

$$F = \frac{\frac{1}{m-1}\sum_{i=1}^{m}(X_i-\bar{X})^2}{\frac{1}{n-1}\sum_{j=1}^{n}(Y_j-\bar{Y})^2}.$$

Under H_0, $F \sim F(m-1, n-1)$. We reject H_0 when

$$F < q_{F(m-1,n-1)}\left(\frac{\alpha}{2}\right) \quad \text{or} \quad F > q_{F(m-1,n-1)}\left(1-\frac{\alpha}{2}\right).$$

(ii) If μ_1, μ_2 are known, the LRT statistic has the form

$$\cancel{F =}\quad F = \frac{\frac{1}{m}\sum_{i=1}^{m}(X_i-\mu_1)^2}{\frac{1}{n}\sum_{j=1}^{n}(Y_j-\mu_2)^2}.$$

Under H_0, $F \sim F(m,n)$. We reject H_0 when

$$F < q_{F(m,n)}\left(\frac{\alpha}{2}\right) \quad \text{or} \quad F > q_{F(m,n)}\left(1-\frac{\alpha}{2}\right).$$

Remark 3

Let $X_1, \ldots, X_m$ be a sample from $N(\mu_1, \sigma^2)$ distribution.
Let $Y_1, \ldots, Y_n$ be — " — — " — $N(\mu_2, \sigma^2)$ — " —.
~~Both~~ The samples are independent, while σ^2 is unknown.

We verify

$$H_0: \mu_1 = \mu_2,$$
$$H_1: \mu \neq \mu_2.$$

The LRT statistic has the form

$$T = \sqrt{\frac{mn}{m+n}} \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{m+n-2}\left[\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{j=1}^{n}(Y_j - \bar{Y})^2\right]}}.$$

Under H_0, ~~the~~ the T statistic has the t-Student distribution with $m+n-2$ degrees of freedom. We reject H_0 when

$$|T| > q_{t(m+n-2)}\left(1-\frac{\alpha}{2}\right).$$

IV Inferences about Normal Models

1. Quadratic Forms

Definition 1

A homogeneous polynomial of degree 2 in n variables is called a quadratic form in those variables. If both the variables and the coefficients are real, the form is called a real quadratic form.

Example 1

$X_1^2 + X_1X_2 + X_2^2$ is a quadratic form in the two variables X_1 and X_2

$X_1^2 + X_2^2 + X_3^2 - 2X_1X_2$ is a quadratic form in the three variables X_1, X_2, and X_3.

$(X_1-1)^2 + (X_2-2)^2 = X_1^2 + X_2^2 - 2X_1 - 4X_2 + 5$ is not quadratic form in X_1 and X_2, although it is a quadratic form in the variables $X_1 - 1$ and $X_2 - 2$.

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}\left(X_i - \frac{X_1 + \ldots + X_n}{n}\right)^2 =$$

$$\frac{n-1}{n}\left(X_1^2 + X_2^2 + \ldots + X_n^2\right) - \frac{2}{n}\left(X_1X_2 + \ldots + X_1X_n + \ldots + X_{n-1}X_n\right)$$

is a quadratic form in the n variables $X_1, X_2, \ldots, X_n$.

Theorem 1

Let $Q = Q_1 + Q_2 + \ldots + Q_{k-1} + Q_k$, where $Q, Q_1, \ldots, Q_k$ are $k+1$ random variables that are real quadratic forms in n independent random variables which are normally distributed with common mean and variance $\mu^{=0}$ and σ^2, respectively. Let $Q/\sigma^2, Q_1/\sigma^2, \ldots, Q_{k-1}/\sigma^2$ have chi-square distributions with degrees of freedom $r, r_1, \ldots, r_{k-1}$, respectively. Let Q_k be nonnegative. Then:

a) $Q_1, \ldots, Q_k$ are independent, and hence

b) Q_k/σ^2 has a chi-square distribution with $r - (r_1 + \ldots + r_{k-1}) = r_k$ degrees of freedom.

Proof (later)

Example 2

Let the random variable X have a distribution that is $N(\mu, \sigma^2)$. Let a and b denote positive integers greater than 1 and let $n = ab$. Consider a random sample of size n from this normal distribution

$$
\begin{array}{l}
X_{11}, X_{12}, \ldots, X_{1j}, \ldots, X_{1b} \\
X_{21}, X_{22}, \ldots, X_{2j}, \ldots, X_{2b} \\
\vdots \\
X_{i1}, X_{i2}, \ldots, X_{ij}, \ldots, X_{ib} \\
\\
X_{a1}, X_{a2}, \ldots, X_{aj}, \ldots, X_{ab}.
\end{array}
$$

We now define $a+b+1$ statistics. They are

$$\overline{X}_{\cdot\cdot} = \frac{X_{11} + \ldots + X_{1b} + \ldots X_{a1} + \ldots + X_{ab}}{ab} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} X_{ij}}{ab},$$

$$\overline{X}_{i\cdot} = \ldots = \frac{\sum_{j=1}^{b} X_{ij}}{b}, \quad i = 1, 2, \ldots, a,$$

and

$$\overline{X}_{\cdot j} = \ldots = \frac{\sum_{i=1}^{a} X_{ij}}{a}, \quad j = 1, 2, \ldots, b.$$

25.02.14

3

Consider the variance S^2 of the random sample of size $n = ab$. We have the algebraic identity

$$(ab-1)S^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X}_{..})^2$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\left[(X_{ij} - \overline{X}_{i.}) + (\overline{X}_{i.} - \overline{X}_{..})\right]^2$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X}_{i.})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{i.} - \overline{X}_{..})^2$$
$$+ 2\sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X}_{i.})(\overline{X}_{i.} - \overline{X}_{..}).$$

Furthermore,

$$2\sum_{i=1}^{a}\left[(\overline{X}_{i.} - \overline{X}_{..})\sum_{j=1}^{b}(X_{ij} - \overline{X}_{i.})\right] = 2\sum_{i=1}^{a}\left[(\overline{X}_{i.} - \overline{X}_{..})(b\overline{X}_{i.} - b\overline{X}_{i.})\right] = 0,$$

and

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{i.} - \overline{X}_{..})^2 = b\sum_{i=1}^{a}(\overline{X}_{i.} - \overline{X}_{..})^2.$$

Thus

$$(ab-1)S^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X}_{i.})^2 + b\sum_{i=1}^{a}(\overline{X}_{i.} - \overline{X}_{..})^2,$$

or, for brevity,

$$Q = Q_1 + Q_2.$$

Since $X_{ij} \sim N(\mu, \sigma^2)$, we have $(ab-1)S^2/\sigma^2 = Q/\sigma^2 \sim \chi^2(ab-1)$. For each fixed value of i, $\sum_{j=1}^{b}(X_{ij} - \overline{X}_{i.})^2/\sigma^2 \sim \chi^2(b-1)$ and $\sum_{j=1}^{b}(X_{1j} - \overline{X}_{1.})^2/\sigma^2, \dots, \sum_{j=1}^{b}(X_{aj} - \overline{X}_{a.})/\sigma^2$ are independent. Therefore,

$$\sum_{i=1}^{a}\left[\sum_{j=1}^{b}(X_{ij} - \overline{X}_{i.})^2/\sigma^2\right] = Q_1/\sigma^2 \sim \chi^2(a[b-1]).$$

Now $Q_2 = b\sum_{i=1}^{a}(\overline{X}_{i.} - \overline{X}_{..})^2 \geq 0$. Theorem 1 implies that Q_1 and Q_2 are independent and $\frac{Q_2}{\sigma^2} \sim \chi^2(\cancel{a(b-1)}\; ab-1-a(b-1) = a-1)$

Similarly, replacing $X_{ij} - \bar{X}_{..}$ by $(X_{ij} - \bar{X}_{.j}) + (\bar{X}_{.j} - \bar{X}_{..})$, in $(ab-1)S^2$,

25.02.14

4

we ~~obtain~~ get

$$(ab-1)S^2 = \sum_{j=1}^{b}\sum_{i=1}^{a}(X_{ij} - \bar{X}_{.j})^2 + a\sum_{j=1}^{b}(\bar{X}_{.j} - \bar{X}_{..})^2,$$

or, for ~~brevity~~ short,

$$Q = Q_3 + Q_4.$$

As a result $Q_3/\sigma^2 \sim \chi^2(b[a-1])$ and $Q_4/\sigma^2 \sim \chi^2(b-1)$ and are independent.

In like manner, in $(ab-1)S^2$, replacing $X_{ij} - \bar{X}_{..}$ by $(\bar{X}_{i.} - \bar{X}_{..}) + (\bar{X}_{.j} - \bar{X}_{..}) + (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})$, we ~~get~~ obtain

$$(ab-1)S^2 = b\sum_{i=1}^{a}(\bar{X}_{i.} - \bar{X}_{..})^2 + a\sum_{j=1}^{b}(\bar{X}_{.j} - \bar{X}_{..})^2 + \sum_{j=1}^{b}\sum_{i=1}^{a}(X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2,$$

or, for clarity,

$$Q = Q_2 + Q_4 + Q_5.$$

Moreover,

$$\frac{Q}{\sigma^2} \sim \chi^2(ab-1), \quad \frac{Q_2}{\sigma^2} \sim \chi^2(a-1), \quad \frac{Q_4}{\sigma^2} \sim \chi^2(b-1).$$

Since $Q_5 \geq 0$, the theorem asserts that, Q_2, Q_4, and Q_5 are independent and that $\frac{Q_5}{\sigma^2} \sim \chi^2([a-1][b-1])$.

Finally,

$$\frac{\frac{Q_4}{\sigma^2(b-1)}}{\frac{Q_3}{\sigma^2 b(a-1)}} \sim F(b-1, a-1), \text{ and}$$

$$\frac{\frac{Q_4}{\sigma^2(b-1)}}{\frac{Q_5}{\sigma^2(a-1)(b-1)}} \sim F(b-1, [a-1][b-1]).$$

2. One-way ANOVA

Consider

$$X_{11}, \ldots, X_{1b};$$
$$\vdots \quad \ddots \quad \vdots$$
$$X_{a1}, \ldots, X_{ab},$$

~~a ra~~ independent indentically distributed (iid) random variables, where $X_{ij} \sim N(\mu_j, \sigma^2)$, $i = 1, .., a$, $j = 1, .., b$, and all parameters are unknown.

The appropriate model for the observations is as follows

$$X_{ij} = \mu_j + e_{ij}; \quad i = 1, .., a, \; j = 1, .., b,$$

where e_{ij} are iid $N(0, \sigma^2)$.

Suppose that it is desired to test the composite hypothesis

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_b = \mu,$$

μ unspecified, against

$$H_1: \sim H_0.$$

A likelihood ratio test will be used.

Remark 1.

The problem is often summarized that we have one factor at b levels. The model is called a one-way model. As we will see, the likelihood ratio test can be thought of in terms of estimates of variance. Hence, this is an example of an analysis of variance (ANOVA). In short, we say that this example is a one-way ANOVA problem.

The total parameter space is

$$\Omega = \{(\mu_1, \mu_2, \dots, \mu_b, \sigma^2) : -\infty < \mu_j < \infty,\ 0 < \sigma^2 < +\infty\}$$

and

$$\omega = \{(\mu_1, \dots, \mu_b, \sigma^2) : -\infty < \mu_1 = \mu_2 = \dots = \mu_b = \mu < \infty,\ 0 < \sigma^2 < +\infty\}.$$

The likelihood functions, denoted by $L(\omega)$ and $L(\Omega)$ are, respectively

$$L(\omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{ab/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\mu)^2\right]$$

and

$$L(\Omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{ab/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\mu_j)^2\right].$$

Now

$$\frac{\partial \log L(\omega)}{\partial \mu} = \sigma^{-2}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\mu)$$

and

$$\frac{\partial \log L(\omega)}{\partial (\sigma^2)} = -\frac{ab}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\mu)^2.$$

Solving $\dfrac{\partial \log L(\omega)}{\partial \mu} = 0$ and $\dfrac{\partial \log L(\omega)}{\partial (\sigma^2)} = 0$, we obtain

$$\hat{\mu} = \bar{x}_{..} = \frac{1}{ab}\sum_{j=1}^{b}\sum_{i=1}^{a} x_{ij},$$

$$\hat{\sigma}_0^2 = v = \frac{1}{ab}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_{..})^2,$$

Sufficient condition!

and these values maximize $L(\omega)$.

Furthermore,

$$\frac{\partial \log L(\Omega)}{\partial \mu_j} = \sigma^{-2} \sum_{i=1}^{a} (x_{ij} - \mu_j), \quad j = 1, 2, \dots, b,$$

and

$$\frac{\partial \log L(\Omega)}{\partial (\sigma^2)} = -\frac{ab}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \mu_j)^2$$

~~Then~~ As a result

$$\hat{\mu}_j = \bar{x}_{.j} = \frac{1}{a} \sum_{i=1}^{a} x_{ij}, \quad j = 1, 2, \dots, b$$

sufficient condition!

$$\hat{\sigma}_1^2 = w = \frac{1}{ab} \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{.j})^2$$

maximize $L(\Omega)$. These maxima are, respectively,

$$L(\hat{\omega}) = \left[\frac{ab}{2\pi \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2} \right]^{ab/2} \exp\left[-\frac{ab \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2}{2 \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2} \right]$$

$$= \left[\frac{ab}{2\pi \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2} \right]^{ab/2} \exp\left[-\frac{ab}{2} \right]$$

and

$$L(\hat{\Omega}) = \left[\frac{ab}{2\pi \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{.j})^2} \right]^{ab/2} \exp\left[-\frac{ab}{2} \right].$$

Finally,

$$\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left[\frac{\sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{.j})^2}{\sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2} \right]^{ab/2} := \left[\frac{Q_3}{Q} \right]^{\frac{ab}{2}}.$$

We reject the hypothesis H_0 if $\Lambda \leqslant \lambda_0$.

We find λ_0.

We have

$$\frac{Q_3}{Q} = \frac{Q_3}{Q_3 + Q_4} = \frac{1}{1 + \frac{Q_4}{Q_3}}.$$

~~The significance level of the test of H_0 is~~ Therefore,

$$\alpha = P_{H_0}\left[\frac{1}{1 + Q_4/Q_3} \leq \lambda_0^{\frac{2}{ab}}\right] \overset{!}{=} P_{H_0}\left[\frac{\frac{Q_4}{b-1}}{\frac{Q_3}{b(a-1)}} \geq c\right],$$

where

$$c = \frac{b(a-1)}{b-1}\left(\lambda_0^{-2/ab} - 1\right).$$

But

$$F = \frac{\frac{Q_4}{\sigma^2(b-1)}}{\frac{Q_3}{\sigma^2 b(a-1)}} = \frac{\frac{Q_4}{b-1}}{\frac{Q_3}{b(a-1)}}$$

has an F-distribution with $b-1$ and $b(a-1)$ degrees of freedom. As a result, ~~The constant c is so selected as to yield the desired value of α i.e~~

$$c = q_{F(b-1,\, b(a-1))}(1-\alpha).$$

Remark 2

The samples may be of different sizes, for instance,

$a_1, a_2, \ldots, a_b$

5. The Analysis of Variance

Recall that the one-way analysis of variance (ANOVA) problem ~~considered~~ concerned one factor at b levels. Now, we have two factors A and B with levels a and b, respectively.

Let X_{ij}, $i=1,..,a$ and $j=1,..,b$ denote the response for Factor A at level i and Factor B at level j.

Denote the total sample size by $n=ab$. We shall assume that the X_{ij}'s are iid $N(\mu_{ij}, \sigma^2)$.

The mean μ_{ij} is often referred to as the mean of the (i,j)th cell.

First, we will consider the <u>additive model</u>, where

$$\mu_{ij} = \bar{\mu} + (\bar{\mu}_{i.} - \bar{\mu}) + (\bar{\mu}_{.j} - \bar{\mu});$$

that is, the mean in the (i,j)th cell is due to additive effects of the levels, i of Factor A and j of Factor B, over the average (constant) $\bar{\mu}$.

Let $\alpha_i = \bar{\mu}_{i\cdot} - \bar{\mu}$, $i = 1, \dots, a$; $\beta_j = \bar{\mu}_{\cdot j} - \bar{\mu}$, $j = 1, \dots, b$; and $\mu = \bar{\mu}$. Then

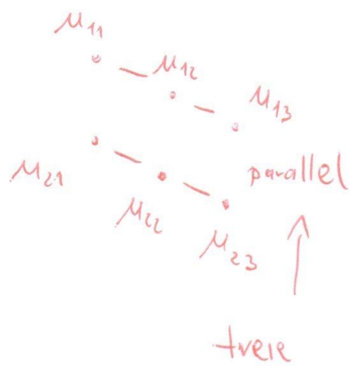
$$\mu_{ij} = \mu + \alpha_i + \beta_j ,$$

where $\sum_{i=1}^{a} \alpha_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$. We refer to this model as ~~being~~ a <u>two-way</u> ANOVA model.

<u>Example 1 (Mean profile plots)</u>

For $a = 2$, $b = 3$, $\mu = 5$, $\alpha_1 = 1$, $\alpha_2 = -1$, $\beta_1 = 1$, $\beta_2 = 0$, $\beta_3 = -1$, we have

		Factor B		
		1	2	3
Factor A	1	$\mu_{11} = 7$	$\mu_{12} = 6$	$\mu_{13} = 5$
	2	$\mu_{21} = 5$	$\mu_{22} = 4$	$\mu_{23} = 3$



If $\beta_1 = \beta_2 = \beta_3 = 0$, we have

		Factor B		
		1	2	3
Factor A	1	$\mu_{11} = 6$	$\mu_{12} = 6$	$\mu_{13} = 6$
	2	$\mu_{21} = 4$	$\mu_{22} = 4$	$\mu_{23} = 4$

The hypotheses of interest are,

$$H_{0A}: \alpha_1 = \dots = \alpha_a = 0 \quad \text{versus} \quad H_{1A}: \alpha_i \neq 0, \text{ for some } i,$$

and

$$H_{0B}: \beta_1 = \dots = \beta_b = 0 \quad \text{versus} \quad H_{1B}: \beta_j \neq 0, \text{ for some } j.$$

If H_{0A} (H_{0B}) is true then the mean of the (i,j)th cell does not depend on the level of A (B). [cf. example, case 2]

We call these hypotheses <u>main effect</u> hypotheses.

The likelihood ratio test for H_{0B} versus H_{1B}

Recall that $Q = Q_4 + Q_3$. That is

$$\underbrace{(ab-1)S^2}_{\substack{\text{total sum}\\ \text{of squares}\\ [TSS]\\ [SST]}} = \underbrace{\sum_{j=1}^{b}\sum_{i=1}^{a}(\bar{X}_{.j}-\bar{X}_{..})^2}_{\substack{\text{sum of squares}\\ \text{among columns}\\ \text{means}}} + \underbrace{\sum_{j=1}^{b}\sum_{i=1}^{a}(X_{ij}-\bar{X}_{.j})^2}_{\substack{\text{sum of squares}\\ \text{within columns}}}$$

is decomposed into

needless

We have shown it on exercises. (499)

Recall that $Q = Q_2 + Q_4 + Q_5$. That is

$$\underbrace{(ab-1)s^2}_{TSS} = \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{i.}-\bar{X}_{..})^2}_{\substack{\text{sum of squares}\\ \text{among rows}}} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{.j}-\bar{X}_{..})^2}_{\substack{\text{sum of square}\\ \text{among columns}}} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij}-\bar{X}_{i.}-\bar{X}_{.j}+\bar{X}_{..})^2}_{\substack{\text{sum of squares}\\ \text{the remains}}}.$$

is decomposed

$$\frac{Q}{\sigma^2} \sim \chi^2(ab-1)$$
$$\frac{Q_2}{\sigma^2} \sim \chi^2(a-1)$$
$$\frac{Q_4}{\sigma^2} \sim \chi^2(b-1)$$
$$\frac{Q_5}{\sigma^2} \sim \chi^2([a-1][b-1])$$

(LRT for H_{0B} against H_{1B})

The ~~A~~ test statistic ~~has a form~~

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

has, under H_{0B}, an F-distribution with $b-1$ and $(a-1)(b-1)$ degrees of freedom.

The hypothesis H_{0B} is rejected if $F \geq c$, where $\alpha = P_{H_{0B}}(F \geq c)$.

We shall compute ~~a the~~ the distribution of F under ~~the~~ the alternatives.

We have

$$E[X_{ij}] = \mu + \alpha_i + \beta_j$$

$$E[\bar{X}_{i.}] = \mu + \alpha_i, \quad E[\bar{X}_{.j}] = \mu + \beta_j, \text{ and } E[\bar{X}_{..}] = \mu.$$

The noncentrality parameter Q_4/σ^2 is

$$\frac{a}{\sigma^2}\sum_{j=1}^{b}(\mu+\beta_j-\mu)^2 = \frac{a}{\sigma^2}\sum_{j=1}^{b}\beta_j^2$$

and that of Q_5/σ^2 is

$$\frac{1}{\sigma^2}\sum^{b}\sum^{a}(\mu+\alpha_i+\beta_j-\mu-\alpha_i-\mu-\beta_j+\mu)^2 = 0$$

Thus, if the hypothesis H_{0B} is true, F has a noncentral F-distribution with $b-1$ and $(a-1)(b-1)$ degrees of freedom and noncentrality parameter $\frac{a}{\sigma^2}\sum_{j=1}^{b}\beta_j^2$.

The likelihood ratio test for H_{0A} versus H_{1A}

The test statistic

$$F = \frac{\frac{Q_2}{a-1}}{\frac{Q_5}{[a-1][b-1]}}$$

has, under H_{0A}, an F-distribution with $a-1$ and $(a-1)(b-1)$ degrees of freedom. The hypothesis H_{0A} is rejected if $F \geq c$, where $\alpha = P_{H_{0A}}(F \geq c)$.

If the hypothesis H_{1A} is true, F has a noncentral F-distribution with $a-1$ and $(a-1)(b-1)$ degrees of freedom and noncentrality parameter $\frac{b}{\sigma^2}\sum_{i=1}^{a}\alpha_i^2$.

Remark 1

The above analysis-of-variance problem is usally refered to as a two-way classification with one observation per cell.

Let X_{ijk}, $i=1,..,a$; $j=1,..,b$; $k=1,..,c$, denote $n=abc$ random variables which are independent and have normal distributions with common, but unknown, variance σ^2. Denote the mean of each X_{ijk}, $k=1,..,c$ by μ_{ij}.

Consider the parameters,

$$\gamma_{ij} = \mu_{ij} - \{\mu + (\bar{\mu}_{i.} - \mu) + (\bar{\mu}_{.j} - \mu)\}, \quad i=1,..,a, \; j=1,..,b.$$
$$= \mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \mu$$

These parameters are called interaction parameters. ~~and~~ [They reflect the specific contribution to the cell mean over and above the additive model]

Recall that ~~Let~~ $\alpha_i = \bar{\mu}_{i\cdot} - \bar{\mu}$, $i = 1, \dots, a$; $\beta_j = \bar{\mu}_{\cdot j} - \bar{\mu}$, $j = 1, \dots, b$, and $\mu = \bar{\mu}$.

Then

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij},$$

where $\sum_{i=1}^{a} \alpha_i = 0$, $\sum_{j=1}^{b} \beta_j = 0$, and $\sum_{i=1}^{a} \gamma_{ij} = \sum_{j=1}^{b} \gamma_{ij} = 0$.

This model is called a <u>two-way model with interactions</u>.

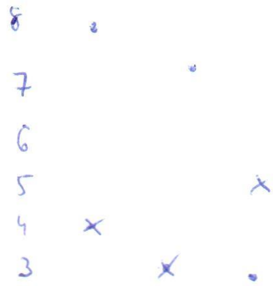
<u>Example 2</u>

For $a = 2$, $b = 3$, $\mu = 5$, $\alpha_1 = 1$, $\alpha_2 = -1$, $\beta_1 = 1$, $\beta_2 = 0$, $\beta_3 = -1$,

$\gamma_{11} = 1$, $\gamma_{12} = 1$, $\gamma_{13} = -2$, $\gamma_{21} = -1$, $\gamma_{22} = -1$, and $\gamma_{23} = 2$,

we have

		Factor B		
		1	2	3
Factor A	1	$\mu_{11} = 8$	$\mu_{12} = 7$	$\mu_{13} = 3$
	2	$\mu_{21} = 4$	$\mu_{22} = 3$	$\mu_{23} = 5$



First, we consider the testing problem

$$H_{0AB}: \gamma_{ij} = 0 \text{ for all } i,j \quad \text{versus} \quad H_{1AB}: \gamma_{ij} \neq 0, \text{ for some } i,j.$$

We have

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(X_{ijk} - \bar{X}_{\cdots})^2 = \overbrace{bc\sum_{i=1}^{a}(\bar{X}_{i\cdot\cdot} - \bar{X}_{\cdots})^2}^{\text{row differences}} + \overbrace{ac\sum_{j=1}^{b}(\bar{X}_{\cdot j\cdot} - \bar{X}_{\cdots})^2}^{\text{column differences}} +$$

$$+ \; c\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X}_{\cdots})^2 + \quad \Big\} \text{ interaction}$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(X_{ijk} - \bar{X}_{ij\cdot})^2 \quad \Big\} \text{ within cells}$$

$a-1$, $b-1$, $(a-1)(b-1)$ $\quad ab(c-1)$

$a - 1, b - 1,$ $ab - \cancel{b} - \cancel{a} + 1$ $\quad abc - ab$

$ab - 1$

$abc - 1 - (ab - 1)$

$abc - ab = ab(c-1)$

The test statistic

$$F = \frac{\dfrac{c\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X}_{\cdots}\right)^2}{(a-1)(b-1)}}{\dfrac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(X_{ijk} - \bar{X}_{ij\cdot}\right)^2}{ab(c-1)}}$$

has, under H_{0AB}, an F-distribution with $(a-1)(b-1)$ and $ab(c-1)$ degrees of freedom.

If H_{0AB} is accepted, ~~we~~ then, we test $\qquad c\sum^{b}\sum^{a}\frac{\gamma_{ij}^2}{\sigma^2}$

(i) $H_{0A}: \alpha_1 = \ldots = \alpha_a = 0$ versus $H_{1A}: \alpha_i \neq 0$, for some i,

on the basis of

$$F = \frac{bc\sum_{i=1}^{a}\left(\bar{X}_{i\cdot\cdot} - \bar{X}_{\cdots}\right)^2 / [a-1]}{\dfrac{\sum\sum\sum\left(X_{ijk} - \bar{X}_{ij\cdot}\right)^2}{ab(c-1)}},$$

which, under H_{0A}, has an F-distribution with $(a-1)$ and $ab(c-1)$ degrees of freedom.

(ii) $H_{0B}: \beta_1 = \ldots = \beta_b = 0$ versus $H_{1B}: \beta_j \neq 0$, for some j,

on the basis of

$$F = \frac{ac\sum_{j=1}^{b}\left(\bar{X}_{\cdot j\cdot} - \bar{X}_{\cdots}\right)^2 / [b-1]}{\sum\sum\sum\left(X_{ijk} - \bar{X}_{ij\cdot}\right)^2 / [ab(c-1)]},$$

which, under H_{0B}, has an F-distribution with $b-1$ and $ab(c-1)$ degrees of freedom.

$$(a-1)+(b-1)+(a-1)(b-1)+ab(c-1) = \cancel{a}-1\cancel{+b}\cancel{-1}+\cancel{ab}-\cancel{b}-\cancel{a}+\cancel{1}+abc-\cancel{ab}$$
$$abc-1$$

6.05.14

4. LS Estimation for Linear Model

We have p predictors $x_1, \dots, x_p$ and a response variable Y. We consider the model of the form

$$Y = h(x_1, \dots, x_p) + \varepsilon,$$

where ε is a random variable (a random error), and h is a specified function. We will restrict our attention to the case where h is linear in the β-coefficients. Our data consists of n vectors of the form $(Y_i, x_{i1}, \dots, x_{ip})$, for $i = 1, \dots, n$. We will center the x's, i.e. $x_{c_{ij}} = x_{ij} - \bar{x}_j$, where $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$. The linear model is

$$Y_i = \alpha + x_{ci1}\beta_1 + x_{ci2}\beta_2 + \dots + x_{cip}\beta_p + \varepsilon_i \quad , i = 1, \dots, n,$$

where $\alpha, \beta_1, \dots, \beta_p$ are unknown parameters (regression coefficients). We assume that the random errors $\varepsilon_1, \dots, \varepsilon_n$ are iid.

The matrix formulation of the model is as follows.

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \alpha + \begin{bmatrix} x_{c11} & \dots & x_{c1p} \\ \vdots & & \\ x_{cn1} & \dots & x_{cnp} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

or equivalently as

$$\underline{Y} = \mathbb{1}\alpha + \underline{X}_c \underline{\beta} + \underline{\varepsilon},$$

or in a more compact form

$$\underline{Y} = \underline{X}\,\underline{b} + \underline{\varepsilon}, \text{ where } \underline{X} = [\mathbb{1}\ \underline{X}_c] \text{ and } \underline{b} = (\alpha, \underline{\beta}')'.$$

We will assume that the $n \times (p+1)$ matrix $\underline{X}$ has full column rank $p+1$.

Let V be the space spanned by the columns of $\underline{X}$.

Then V is $(p+1)$-dimensional vector space of $\mathbb{R}^n$.

Put $\underline{\eta} = \underline{X}\,\underline{b}$. Then

$$\underline{Y} = \underline{\eta} + \underline{\varepsilon}, \quad \text{for } \underline{\eta} \in V.$$

Reading: Except for random error, $\underline{Y}$ would lie in the subspace V. So, to estimate η, find a vector in V which lies "closest" to Y (in terms of a given norm)

4.1 Least Squares

The LS estimator of η has a form

$$\hat{\eta} = \arg\min_{\eta \in V} \| Y - \eta \|^2, \quad \text{where } \|v\|^2 = \sum_{i=1}^{n} v_i^2 .$$

Let $V^\perp$ be the subspace which consists of all vectors in $\mathbb{R}^n$ which are orthogonal to all vectors in V, that is,

$$V^\perp = \{ w \in \mathbb{R}^n : w'v = 0, \text{ for all } v \in V \}.$$

The dimension of $V^\perp$ is $n-(p+1)$.

Definition 1

Let v be a vector in $\mathbb{R}^n$ and let V be a subspace of $\mathbb{R}^n$. We say that $\hat{v}$ is the projection of v onto V if

(i) $\hat{v} \in V$,

(ii) $v - \hat{v} \in V^\perp$.

Theorem 1

Projections are unique.

Proof

1. Let $\hat{v}_1$ and $\hat{v}_2$ be projections of v onto V.
2. Since V is a subspace, from (i) $\hat{v}_1 - \hat{v}_2 \in V$.
3. But $\hat{v}_1 - \hat{v}_2 = (v - \hat{v}_2) - (v - \hat{v}_1) \in V^\perp$.
4. Thus $\|\hat{v}_1 - \hat{v}_2\|^2 = 0$.
5. Finally $\hat{v}_1 = \hat{v}_2$. □

The columns of $\underline{X}$ form a basis for the subspace V. Therefore, we will say that $\underline{X}$ is a basis matrix for V and that $\underline{X}$ has full column rank, which implies that $(X'X)^{-1}$ exists.

Theorem 2

Let $\underline{X}$ be a basis matrix for a subspace V, let $\underline{H} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$, and let $\underline{v}$ be a vector in $\mathbb{R}^n$. Then the projection of $\underline{v}$ onto V is $\underline{Hv}$

Proof (we will check (i) & (ii))

1. We have $\underline{H}\underline{v} = \underline{X}\{(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{v}\} \in V$ and (i) follows.

2. Let $\underline{u} \in V$.

3. Because $\underline{X}$ is a basis matrix for V, we have $\underline{u} = \underline{X}\underline{c}$ ~~for~~, where $\underline{c} \in \mathbb{R}^{p+1}$.

4. Then
$$(\underline{v} - \underline{H}\underline{v})'\underline{u} = \underline{v}'(\underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{X}\underline{c} = \underline{v}'(\underline{X} - \underline{X})\underline{c} = 0$$

5. (ii) is satisfied.

~~Conclusion~~ Corollary 1

The projection matrix $\underline{H}$ is idempotent (i.e $\underline{H}^2 = \underline{H}$) and symmetric. All eigenvalues of $\underline{H}$ are either 0 or 1 and the rank of $\underline{H}$ is equal to its trace. The matrix $\underline{I} - \underline{H}$ is the projection matrix onto $V^\perp$.

Theorem 3 Consider the model $\underline{Y} = \eta + \varepsilon$, for $\eta \in V$. Let $\underline{H}$ be the projection matrix onto V. Let $\hat{\eta} = \underline{H}\underline{Y} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$. Then $\hat{\eta}$ is the LS solution.

Proof

1. Let $\eta \in V$.

2. Then $\underline{H}\underline{Y} - \eta \in V$.

3. But $(\underline{I} - \underline{H})\underline{Y} \in V^\perp$

4. Therefore,
$$\|\underline{Y} - \eta\|^2 = \|\underline{Y} - \underline{H}\underline{Y} + \underline{H}\underline{Y} - \eta\|^2 = \|(\underline{I} - \underline{H})\underline{Y} + (\underline{H}\underline{Y} - \eta)\|^2 = \underbrace{\|(\underline{I}-\underline{H})\underline{Y}\|^2}_{\text{does not depend on } \eta} + \|\underline{H}\underline{Y} - \eta\|^2$$

5. We minimize the left-side by taking $\eta = \underline{H}\underline{Y}$.

6. Hence, the LS solution is the projection $\underline{H}\underline{Y}$.

7. Uniqueness - exercise. [~~ex 16~~]

As a result, the LS estimate $\hat{\underline{b}}$ of $\underline{b}$ must satisfy

$$X\hat{b} = \underline{H}\underline{Y} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y} \quad / \cdot X'$$

$$X'X\hat{b} = X'Y$$ - estimating equation (normal equation) for the multiple regression model.

$(X'X)^{-1}/$
$$\hat{b} = (X'X)^{-1}X'Y.$$

6.05.14

The estimate $\hat{Y} = X\hat{b}$ is called the <u>fitted</u> or <u>predicted</u> value of $\underline{Y}$.
The <u>residual</u> or <u>estimate</u> of the error vector is given by $\hat{\varepsilon} = \underline{Y} - \underline{\hat{Y}}$.
Since $\hat{Y} \in V$ and $\hat{\varepsilon} \in V^{\perp}$, we have $\underline{\hat{Y}} \perp \hat{\varepsilon}$.

<u>Theorem</u> 4

Consider the model $\underline{Y} = \underline{X}\,\underline{b} + \underline{\varepsilon}$ and assume that $\varepsilon_1, .., \varepsilon_n$ are iid. and that $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \sigma^2 < \infty$. Then

a) $E(\hat{b}) = \underline{b}$ and $\mathrm{Cov}(\hat{b}) = \sigma^2 (X'X)^{-1}$.

b) $E(\hat{Y}) = \underline{X}\,\underline{b}$ and $\mathrm{Cov}(\hat{Y}) = \sigma^2 H$

~~c) $E(\hat{\varepsilon}) = \underline{0}$ and $\mathrm{Cov}(\hat{\varepsilon}) = \sigma^2(I-H)$~~ $\hat{\varepsilon}$ <u>ok</u>

~~c~~d) $E(\hat{\sigma}^2) = \sigma^2$, where $\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n} \hat{\varepsilon}_i^2$

<u>Proof</u>

1. We have

$$\hat{b} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(Xb+\varepsilon) = b + (X'X)^{-1}X'\varepsilon$$
$$\hat{Y} = X\hat{b} = Xb + H\varepsilon$$
$$\hat{\varepsilon} = Y - \hat{Y} = Xb + \varepsilon - Xb - H\varepsilon = (I-H)\varepsilon.$$

2. Since $E(\varepsilon) = 0$ and $\mathrm{Cov}(\varepsilon) = \sigma^2 I$, Then

$$E\hat{b} = b, \quad \mathrm{Cov}(\hat{b}) = (X'X)^{-1}X'\sigma^2 I \cdot X \cdot (X'X)^{-1} = \sigma^2 (X'X)^{-1},$$
$$E\hat{Y} = Xb, \quad \mathrm{Cov}(\hat{Y}) = H\,\sigma^2 I\, H' = \sigma^2 H \quad \left(\underset{\text{symetry}}{H' = H}\right)$$

correct but obvious

?? ~~$E\hat{\eta} = EHY = EH(\hat{\varepsilon} + \hat{Y}) = H(I-H)E\varepsilon + E\hat{Y}$ $HE(Xb+\varepsilon) = HXb = X(X'X)^{-1}X'Xb = Xb$~~

?? ~~$\mathrm{Cov}(\hat{\eta}) = \mathrm{Cov}(HY) = H\,\mathrm{Cov}\,Y\,H' = H(\mathrm{Cov}(Y - \hat{Y} + \hat{Y}))H' = H(\mathrm{Cov}\,\hat{\varepsilon} + \mathrm{Cov}\,\hat{Y})H' =$~~

~~$H(I-H)\sigma^2 I (I-H)'H' + H\sigma^2 H H' =$~~

~~$\mathrm{Cov}(\hat{\eta}) = \mathrm{Cov}(HY) = H\,\mathrm{Cov}(Y)\,H' = H\sigma^2 I H = H\sigma^2$~~ $H\sigma^2$

$E\hat{\varepsilon} = 0$, $\mathrm{Cov}(\hat{\varepsilon}) = \sigma^2(I-H)$

3. We have

$$(n-p-1)\hat{\sigma}^2 = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{\varepsilon}^T\hat{\varepsilon} = \varepsilon^T(I-H)(I-H)\varepsilon = \varepsilon^T(I-H)\varepsilon.$$

4. Hence

$$E[(n-p-1)\hat{\sigma}^2] = E[\varepsilon^T(I-H)\varepsilon] = \cancel{E[tr(\varepsilon^T(I-H)\varepsilon)] = E[tr((I-H)\varepsilon\varepsilon^T)]} \overset{ok}{=}$$
$$tr[(I-H)E[\varepsilon\varepsilon^T]] = tr[(I-H)\sigma^2 I] = tr[(I-H)\sigma^2] = (n-p-1)\sigma^2.$$

$X \sim \mu, \Sigma$
$E[X'AX] = tr(\overset{I-H}{A}\Sigma) + \mu' A \mu = tr((I-H)\sigma^2) = \mathrm{rank}(\ )\sigma^2$

4.2 Basics of LS Inference under Normal Errors

6.05.14

Theorem 1

Consider the model $\underline{Y} = \underline{X}\underline{b} + \underline{\varepsilon}$ and ~~asume~~ assume that ~~$\varepsilon_1, \dots, \varepsilon_n$~~ ε has a $N_n(0, \sigma^2 I)$ distribution. Then the LS estimators satisfy the following:

(a) $\hat{b}$ has a $N(\underline{b}, \sigma^2 (X'X)^{-1})$ distribution.

(b) $\hat{Y}$ has a $N(\underline{X} b, \sigma^2 H)$ distribution.

(c) $\hat{\varepsilon}$ has a $N(0, \sigma^2 (I - H))$ — " — .

(d) $(n-p-1)\hat{\sigma}^2 / \sigma^2$ has a $\chi^2(n-p-1)$ distribution.

(e) $\hat{Y}$ and $\hat{\varepsilon}$ are independent.

(f) $\hat{b}$ and $\hat{\sigma}^2$ are independent.

Proof

(a, b, c) obvious.

d) We have $\dfrac{(n-p-1)\hat{\sigma}^2}{\sigma^2} = \sigma^{-2} \varepsilon'(I - H)\varepsilon$ and $I - H$ is idempotent of rank $n-p-1$.

e) Since

$$\begin{bmatrix} \hat{Y} \\ \hat{\varepsilon} \end{bmatrix} = \begin{bmatrix} H \\ I - H \end{bmatrix} \varepsilon + \begin{bmatrix} X b \\ 0 \end{bmatrix}$$

Then $\hat{Y}$ and $\hat{\varepsilon}$ have a jointly normal distribution, while their covariance matrix has a form

$$\begin{bmatrix} H \\ 1 - H \end{bmatrix} \sigma^2 I \begin{bmatrix} H \\ 1 - H \end{bmatrix}' = \sigma^2 \begin{bmatrix} H & 0 \\ 0 & I - H \end{bmatrix}.$$

f) Since $\hat{Y} = X\hat{b}$ and $\hat{\sigma}^2 = \frac{1}{(n-p-1)} \hat{\varepsilon}^T \hat{\varepsilon}$, e) entails f).

652

Corollary 1

Consider the model $\underline{Y} = \underline{X}\underline{b} + \underline{\varepsilon}$ and asume that ε has a $N_n(0, \sigma^2 I)$ distribution. Then the random variables

$$T_j \quad \cancel{t_j} = \frac{\hat{\beta}_j - \beta}{\hat{\sigma}\sqrt{(X_c' X_c)^{-1}_{jj}}}, \qquad j = 1, \dots, p,$$

where $(X_c' X_c)^{-1}_{jj}$ is the jth diagonal entry of $(X_c' X_c)^{-1}$ and X_c is the centered design matrix, have t-distributions with $n-p-1$ degrees of freedom.

Therefore, a level α test for the hypotheses

$$H_0 : \beta_j = 0 \quad \text{versus} \quad H_1 : \beta_j \neq 0 \qquad j = 1, \dots, p$$

is given by

$$\text{reject } H_0 \text{ if } |T_j| = \frac{|\hat{\beta}_j|}{\hat{\sigma}\sqrt{(X_c' X_c)^{-1}_{jj}}} > t_{1-\alpha/2, n-p-1},$$

where $t_{1-\alpha/2, n-p-1}$ is the $(1-\alpha/2)$-quantile of the t-distribution

6. Tests of General Linear Hypotheses

Consider the model

$$\underline{Y} = \underline{X}\,\underline{b} + \underline{\varepsilon},$$

where $\underline{X}$ is an $n \times (p+1)$ design matrix, $\underline{b} = (\alpha, \underline{\beta}')'$. The above model we will call the full model.

We test a general linear hypothesis

$$H_0 : \underline{A}\,\underline{b} = \underline{0},$$

against the alternative

$$H_1 : \underline{A}\,\underline{b} \neq 0,$$

where $\underline{A}$ is a $q \times (p+1)$ specified matrix of full row rank $q < p+1$.

So, the rows of $\underline{A}$ provide the linear constraints.

Example 1

1) Suppose we are predicting Y based on a second degree polynomial model in x_1 and x_2, i.e.,

$$E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 . \quad (*)$$

Suppose our null hypothesis is that the first-order terms suffice to predict Y.

The corresponding matrix $\underline{A}$ is

$$\underline{A} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

because, under H_0, $E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2$

2) Suppose for the model $(*)$, we think the slope parameters of x_1 and x_2 are the same. Then the null hypothesis can be expressed with the matrix

$$\underline{A} = [\,0 \;\; 1 \;\; -1 \;\; 0 \;\; 0 \;\; 0\,].$$

Let V_F, (where F stands for the full model), denote the column space of $\underline{X}$.

For the hypothesis H_0, the reduced model is the full model subject to H_0, i.e., the subspace given by

$$V_R = \{ \underline{v} \in V_F : \underline{v} = \underline{X}\,\underline{b} \text{ and } \underline{A}\underline{b} = 0 \}.$$

We will show (later) that $\dim(V_r) = (p+1) - q$.

Suppose we have a norm $\|\cdot\|$ for fitting models.

Let $$\hat{\eta}_F = \underset{\underline{\eta} \in V_F}{\operatorname{argmin}} \|\underline{Y} - \underline{\eta}\|.$$

Then, the distance between $\underline{Y}$ and the subspace V_F is

$$d_F = d(\underline{Y}, V_F) = \| Y - \hat{\eta}_F \| .$$

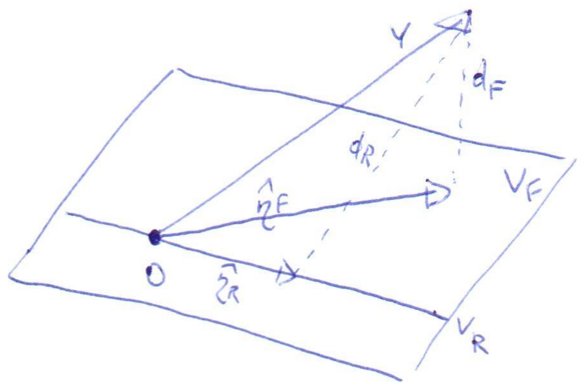
Let

$$\hat{\eta}_R = \arg\min_{\eta \in V_R} \| Y - \eta \|$$

and let

$d_R = d(\underline{Y}, V_R)$ denote the distance between $\underline{Y}$ and the subspace V_R.

We have $d(\underline{Y}, V_R) \geq d(\underline{Y}, V_F)$ (minimum over a larger set can be smaller)



An intuitive test statistic has a form

$$RD_{\|\cdot\|} = d(\underline{Y}, V_R) - d(\underline{Y}, V_F).$$

Small values of $RD_{\|\cdot\|}$ indicate ~~that~~ H_0 while large values indicate that H_1 is true. Therefore, we will reject H_0 in favour of H_1 if $RD_{\|\cdot\|} \geq c$.

We will find c.

Assume that $\|\cdot\|$ is the Euclidean norm. Let $\underline{H}_F$ and $\underline{H}_R$ denote the projection matrices onto the subspaces V_F and V_R, respectively. Then

$$d^2_{LS}(\underline{Y}, V_F) = \| \underline{Y} - \underline{H}_F \underline{Y} \|^2_{LS} = \underline{Y}'(I - \underline{H}_F)\underline{Y}$$

$$d^2_{LS}(\underline{Y}, V_R) = \| \underline{Y} - \underline{H}_R \underline{Y} \|^2_{LS} = \underline{Y}'(I - \underline{H}_R)\underline{Y}.$$

Therefore

$$RD_{LS} = d^2_{LS}(\underline{Y}, V_R) - d^2_{LS}(\underline{Y}, V_F) = \underline{Y}'(\underline{H}_F - \underline{H}_R)\underline{Y}.$$

6.1 Distribution Theory for the LS Test for Normal Errors

Assume that $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I)$.

Definition 1

Let V_1 and V_2 be two subspaces of $\mathbb{R}^n$ and assume that $V_1 \subset V_2$.
Then the space V_2 mod V_1 ~~is def~~ has a form

$$V_2 | V_1 = \{ \underline{v} \in V_2 : \underline{v} \perp \underline{w}, \text{ for all } \underline{w} \in V_1 \}.$$

Lemma 1

The matrix $\underline{H}_F - \underline{H}_R$ is the projection matrix onto the space $V_F | V_R$.

Proof

1. Let $\underline{U}_R$ be an o.n. basis matrix for V_R ~~and let~~
2. Let $[\underline{U}_R : \underline{U}_2]$ be an extention of $\underline{U}_R$ to an o.n. basis matrix of V_F.
3. Then $\underline{U}_2$ is a basis matrix for $V_F | V_R$ and $\underline{U}_2 \underline{U}_2'$ is the projection matrix onto $V_F | V_R$.

$$\underline{U}_2 (\underbrace{\underline{U}_2' U}_{I_{p+1-q}})^{-1} U_2'$$

4. Also, $\underline{H}_R = \underline{U}_R \underline{U}_R'$ and

$$\underline{H}_F = [\underline{U}_R : \underline{U}_2][\underline{U}_R : \underline{U}_2]' = \underline{U}_R \underline{U}_R' + \underline{U}_2 \underline{U}_2' = \underline{H}_R + \underline{U}_2 \underline{U}_2'. \quad \square$$

Lemma 2

Let $\underset{n \times q}{\underline{C}} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{A}'$. Then $\underline{C}$ is a basis matrix for $V_F | V_R$. Further, the dimension of $V_F | V_R$ is q and V_R is $p+1-q$.

Proof (Ex)

$$C'C = A(X'X)^{-1}\cancel{X'X(X'X)^{-1}}A'$$

upisać

As a result $\underline{H}_F - \underline{H}_R = C(C'C)^{-1}C'$. Then

$$RD_{LS} = \underline{Y}'(\underline{H}_F - \underline{H}_R)\underline{Y} = Y' X(X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}A\underbrace{(X'X)^{-1}X'Y}_{\hat{b}_{LS}}$$

$$= (A\hat{b}_{LS})'[A(X'X)^{-1}A']^{-1}A\hat{b}_{LS}.$$

So, the standardised thest statistic has a form

$$F_{LS} = \frac{\frac{1}{q}(A\hat{b}_{LS})'[A(X'X)^{-1}A']^{-1}A\hat{b}_{LS}}{\hat{\sigma}^2}, \qquad LRT$$

where $\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n} \hat{e}_i^2$.

Theorem 1 Under the model $\underline{Y} = \underline{X}\underline{b} + \underline{\varepsilon}$ and the assumption that $\underline{\varepsilon}$ ~~has~~ has a $N_n(\underline{0}, \sigma^2 I)$ distribution, the statistic F_{LS} has an F-distribution with q and $n-p-1$ degrees of freedom and noncentrality parameter $\theta = (Ab)'[A(X'X)^{-1}A']^{-1}Ab / \sigma^2$.

Proof

1. ~~We have~~ $(n-p-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-p-1)$.
2. $\hat{\sigma}^2$ is independent of $\hat{\underline{b}}_{LS}$
3. Hence numerator and denominator of F_{LS} are independent.
4. $\hat{\underline{b}}_{LS} \sim N_{p+1}(\underline{b}, \sigma^2(\underline{X}'\underline{X})^{-1})$
5. Hence $\underline{A}\hat{\underline{b}}_{LS} \sim N_q(\underline{A}\underline{b}, \sigma^2\underline{A}(\underline{X}'\underline{X})^{-1}\underline{A}')$

and ~~numer~~ numerator $\times \frac{q}{\sigma^2} \sim \chi^2(q, \theta)$ $\square$.

Corollary 1

We reject H_0 in favour of H_1 if $F_{LS} \geqslant F_{1-\alpha, q, n-p-1}$

"now you can find c" ~~as~~ ~~α~~