UNC BIOS 767

Summary

Creating my own project

Author:
Mingwei Fei

Special colaborator:
Bonnie

Jan, 2023

Contents

Chapter 1

ANOVA

1.1 ANOVA

1.1.1 Two Way Interaction ANOVA

Consider the two way ANOVA table with interaction, given by

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + \epsilon_{ijk},$$

where $i = 1, 2, ...a, j = 1, 2.., b$ and $k = 1, ..N$. Further suppose that the ϵ_{ijk} are i.i.d and $\epsilon_{ijk} \sim N(0, \sigma^2)$, where σ^2 is unknown. Let M_α, M_η and M_γ denote the orthogonal operations for the α, η spaces, and interaction space, respectively.

(a) Show that $\sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\gamma_{ij}$ is estimable, and find the UMVUE and the variance of the UMVUE, where the c_{ij}'s are real numbers that satisfy $\sum_{i=1}^{a} c_{ij} = \sum_{j=1}^{b} c_{ij} = 0$. Let

$$\beta = (\mu, \alpha_1, \alpha_2, ..\alpha_a, \eta_1, \eta_2, ..\eta_b, \gamma_{11}, \gamma_{12}, ...\gamma_{ab}),$$
$$X = (J_a \otimes J_b \otimes J_N, I_a \otimes J_b \otimes J_N, J_a \otimes I_b \otimes J_N, I_a \otimes I_b \otimes J_N),$$

The two way ANOVA model could be written as

$$Y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2)$$

The contrast

$$\lambda^T\beta = \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\gamma_{ij},$$

where

$$\lambda^T = (0, ...0_{a+b+1}, c_{11}, ...c_{ab}),$$

We need to find ρ such that $\lambda^T = \rho^T X$ to show estimability.
Let

$$\rho^T = \frac{1}{N}(c_{11}J_N^T, c_{12}J_N^T, ...c_{ab}J_N^T,),$$

we have

$$\rho^T X = \frac{1}{N}(c_{11}J, c_{12}J, ...c_{ab})\otimes J_N^T)(J_a\otimes J_b\otimes J_N, I_a\otimes J_b\otimes J_N, J_a\otimes I_b\otimes J_N, I_a\otimes I_b\otimes J_N),$$

$$\rho^T X = \frac{1}{N}\{[(c_{11}, ...c_{ab})(J_a \otimes J_b)] \otimes J_N^T J_N,$$

$$[(c_{11}, ...c_{ab})(I_a \otimes J_b)] \otimes J_N^T J_N,$$
$$[(c_{11}, ...c_{ab})(J_a \otimes I_b)] \otimes J_N^T J_N,$$
$$[(c_{11}, ...c_{ab})(I_a \otimes I_b)] \otimes J_N^T J_N\}$$

Since $\sum_{i=1}^{a} c_{ij} = \sum_{i=1}^{b} c_{ij} = 0$ Then

$$[(c_{11}, ...c_{ab})(J_a \otimes J_b)] = 0, [(c_{11}, ...c_{ab})(I_a \otimes I_b)] = 0, (c_{11}, ...c_{ab})(I_a \otimes J_b) = 0$$
$$[(c_{11}, ...c_{ab})(I_a \otimes I_b)] = (c_{11}, c_{12}, ...c_{ab})$$
$$\rho^T X = [0, 0_a, 0_b, (c_{11}, ...c_{ab})] = \lambda^T$$

Thus, $\lambda^T\beta = \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\gamma_{ij}$ is estimable.

The UMVUE of $\lambda^T\beta$ is $\rho^T MY$, where $M = I_a \otimes I_b \otimes P_N$, where

$$P_N = \frac{1}{N} J_N^N$$

Therefore, we have

$$\rho^T MY = \frac{1}{N}[(c_{11}, ...c_{ab})] \otimes J_N^T][(I_{ab} \otimes P_N)]Y$$
$$\rho^T MY = \frac{1}{N}[(c_{11}, ...c_{ab})] \otimes J_N^T]Y = \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\bar{Y}_{ij.}$$

The variance of UMVUE is

$$Var(\rho^T MY) = Var(\sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\bar{Y}_{ij.}) = \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}^2 Var(\bar{Y}_{ij.}) = \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}^2 \frac{\sigma^2}{N}$$

(b) Using Kronecker product and notation, derive the orthogonal projection operator for the interaction space, denoted by M_γ.
Let s be an arbitrary index. Define J_s as the $s \times 1$ vector of ones, $P_s = \frac{1}{N} J_s J_s^{'}$ and $Q_s = I_s - P_s$, where I_s is the $s \times s$ identity matrix. Thus P_s is the orthogonal projection operator onto $C(J_s)$ and Q_s is the orthogonal projection operator onto $C(J_s)^{\perp}$. Computing M_γ. The interaction space is given by $C(Q_a \otimes Q_b \otimes P_N)$. This yields

$$M_\gamma = Q_a \otimes Q_b \otimes P_N$$

Compute M_μ

$$\begin{aligned}
M_\mu &= (J_a \otimes J_b \otimes J_N)[(J_a \otimes J_b \otimes J_N)^T (J_a \otimes J_b \otimes J_N)]^{-1}(J_a \otimes J_b \otimes J_N)^T \\
&= (J_a \otimes J_b \otimes J_N)[(J_a^{'} \otimes J_b^{'} \otimes J_N^{'})(J_a \otimes J_b \otimes J_N)]^{-1}(J_a^{'} \otimes J_b^{'} \otimes J_N^{'}) \\
&= (J_a \otimes J_b \otimes J_N)[(J_a^{'}J_a \otimes J_b^{'}J_b \otimes J_N^{'}J_N)]^{-1}(J_a^{'} \otimes J_b^{'} \otimes J_N^{'}) \\
&= (J_a \otimes J_b \otimes J_N)(abN)^{-1}(J_a^{'} \otimes J_b^{'} \otimes J_N^{'}) \\
&= \frac{1}{a}J_aJ_a^{'} \otimes \frac{1}{b}J_bJ_b^{'} \otimes \frac{1}{N}J_NJ_N^{'} \\
&= P_a \otimes P_b \otimes P_N
\end{aligned}$$

Compute M_γ, the γ space is $(Q_a \otimes Q_b \otimes J_N)$, thus

$$\begin{aligned}
M_\gamma &= (Q_a \otimes Q_b \otimes J_N)[(Q_a \otimes Q_b \otimes J_N)^T(Q_a \otimes Q_b \otimes J_N)]^{-1}(Q_a \otimes Q_b \otimes J_N)^T \\
&= (Q_a \otimes Q_b \otimes J_N)[(Q_a^{'}Q_a \otimes Q_b^{'}Q_b \otimes J_N^{'}J_N)]^{-1}(Q_a^{'} \otimes Q_b^{'} \otimes J_N^{'}) \\
&= (Q_a \otimes Q_b \otimes J_N)[(Q_a^{-} \otimes Q_b^{-1} \otimes N^{-1}](Q_a^{'} \otimes Q_b^{'} \otimes J_N^{'}) \\
&= (Q_a \otimes Q_b \otimes P_N)
\end{aligned}$$

Now $M = M_\mu + M_\alpha + M_\eta + M_\gamma$, we have

$$\begin{aligned}
M &= (P_a \otimes P_b \otimes P_N) + (Q_a \otimes P_b \otimes P_N) + (P_a \otimes Q_b \otimes P_N) + (Q_a \otimes Q_b \otimes P_N) \\
M &= (P_a + Q_a) \otimes P_b \otimes P_N + (P_a + Q_a) \otimes Q_b \otimes P_N = I_a \otimes I_b \otimes P_N
\end{aligned}$$

The error space is $I - M$

$$I - M = I_a \otimes I_b \otimes I_N - I_a \otimes I_b \otimes P_N = I_a \otimes I_b \otimes Q_N$$

(c) Derive the simply possible scalar expression for $E[Y'(M_\alpha + M_\eta)Y]$.

$$E[Y'(M_\alpha + M_\eta)Y] = tr((M_\alpha + M_\eta)\Sigma) + \mu'(M_\alpha + M_\eta)\mu$$

where

$$\mu = E[Y] = \mu \otimes J_a \otimes J_b \otimes J_N + \alpha \otimes J_b \otimes J_N + J_a \otimes \eta \otimes J_N + \gamma \otimes J_N$$

$\alpha = (\alpha_1, \alpha_2, ...\alpha_a)^T, \eta = (\eta_1, \eta_2, ...\eta_b)^T, \gamma = (\gamma_{11}, ..\gamma_{ab})^T$. And $\Sigma = \sigma^2 I_{ab}$.
Therefore,

$$E[Y'(M_\alpha + M_\eta)Y] = tr((M_\alpha + M_\eta)\Sigma) + \mu'(M_\alpha + M_\eta)\mu$$

$$= tr(M_\alpha\Sigma) + tr(M_\eta\Sigma) + \mu'M_\alpha\mu + \mu'M_\eta\mu$$

$$\begin{aligned}
\mu'M_\alpha\mu = &(\mu J_a \otimes J_b \otimes J_N + \alpha \otimes J_b \otimes J_N + J_a \otimes \eta \otimes J_N + \gamma \otimes J_N)^T(Q_a \otimes P_b \otimes P_N) \\
&(\mu J_a \otimes J_b \otimes J_N + \alpha \otimes J_b \otimes J_N + J_a \otimes \eta \otimes J_N + \gamma \otimes J_N)
\end{aligned}$$

because $Q_a J_a = 0$

$$\mu' M_\alpha \mu = (\alpha \otimes J_b \otimes J_N + \gamma \otimes J_N)^T (Q_a \otimes P_b \otimes P_N)(\alpha \otimes J_b \otimes J_N + \gamma$$
$$= (\alpha \otimes J_b \otimes J_N)^T (Q_a \otimes P_b \otimes P_N)(\alpha \otimes J_b \otimes J_N) + 2(\gamma \otimes J_N$$
$$+(\gamma \otimes J_N)^T (Q_a \otimes P_b \otimes P_N)(\gamma \otimes J_N)$$
$$= (\alpha^T Q_a \alpha) \otimes (J_b^T P_b J_b) \otimes (J_N^T P_N J_N) + 2(\gamma \otimes J_N)^T (\alpha Q_a) \otimes$$
$$= (\alpha^T Q_a \alpha) \otimes (J_b^T P_b J_b) \otimes (J_N^T P_N J_N) + 2[\gamma^T (\alpha Q_a \otimes P_b J_b)] \otimes$$

break down into each term

$$\alpha' Q_a \alpha = \alpha^T [I - \frac{1}{a} J_a^a] \alpha = [(I - \frac{1}{a} J_a^a)\alpha]^T [(I - \frac{1}{a} J_a^a)\alpha]$$
$$= [(\alpha - \bar{\alpha} J_a]^T [(\alpha - \bar{\alpha} J_a] = \sum_{i=1}^{n} (\alpha_i - \bar{\alpha})^2$$
$$J_b^T P_b J_b = J_b^T \frac{1}{b} J_b^b J_b = b$$
$$J_N^T P_N J_N = N$$
$$(\alpha^T Q_a \alpha) \otimes (J_b^T P_b J_b) \otimes (J_N^T P_N J_N) = bN \sum_{i=1}^{n} (\alpha_i - \bar{\alpha})^2$$

(d) Derive the F-test for the hypothesis: $H_0 : \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} \gamma_{ij} = 4$, and state its distribution under the null and alternative hypothesis.
The orthogonal operator projection for $M\rho$ is

$$M_{MP} = (M\rho)[(M\rho)^T (M\rho)]^- (\rho' M) = (M\rho)[\rho^T M \rho]^- (\rho' M)$$

The F-test is given by:

$$F = \frac{(\rho' MY - 4)' (\rho' M \rho)^- (\rho' MY - 4)/r(M_{MP})}{MSE}$$

Also because $M\rho = \rho$,

$$\rho' MY = \rho' Y = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} \bar{Y}_{ij.}$$
$$\rho' M \rho = \rho' \rho = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{c_{ij}^2}{N}$$
$$MSE = \frac{1}{abN - ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{N} (Y_{ijk} - \bar{Y}_{ij.})^2$$

Thus,

$$F = \frac{\frac{N}{\sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}^2}(\sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\bar{Y}_{ij.} - 4)^2}{\frac{1}{abN-ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{N}(Y_{ijk} - \bar{Y}_{ij.})^2} \sim F[1, ab(N-1), \gamma]$$

Under $H_0, \gamma = 0$, and under $H_1, \gamma = \frac{(\sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}^2)N}{2\sigma^2 \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}^2}$

(e) Using only Kronecker product development and notation, obtain the simplest possible expression for $M_\alpha + M_\eta$.

Compute M_α, the α space is $(Q_a \otimes J_b \otimes J_N)$, thus

$$\begin{aligned}
M_\alpha &= (Q_a \otimes J_b \otimes J_N)[(Q_a \otimes J_b \otimes J_N)^T(Q_a \otimes J_b \otimes J_N)]^{-1}(Q_a \otimes J_b \otimes J_N)^T \\
&= (Q_a \otimes J_b \otimes J_N)[(Q_a' Q_a \otimes J_b' J_b \otimes J_N' J_N)]^{-1}(Q_a' \otimes J_b' \otimes J_N') \\
&= (Q_a \otimes J_b \otimes J_N)[(Q_a^- \otimes b^{-1} \otimes N^{-1}](Q_a' \otimes J_b' \otimes J_N') \\
&= (Q_a \otimes P_b \otimes P_N)
\end{aligned}$$

Here Q_a is symmetric, semi-definite.
Compute M_η, the η space is $(J_a \otimes Q_b \otimes J_N)$, thus

$$\begin{aligned}
M_\eta &= (J_a \otimes Q_b \otimes J_N)[(J_a \otimes Q_b \otimes J_N)^T(J_a \otimes Q_b \otimes J_N)]^{-1}(J_a \otimes Q_b \otimes J_N)^T \\
&= (J_a \otimes Q_b \otimes J_N)[(J_a' J_a \otimes Q_b' Q_b \otimes J_N' J_N)]^{-1}(J_a' \otimes Q_b' \otimes J_N') \\
&= (J_a \otimes Q_b \otimes J_N)[(a^- \otimes Q_b^{-1} \otimes N^{-1}](J_a' \otimes Q_b' \otimes J_N') \\
&= (P_a \otimes Q_b \otimes P_N) \\
M_\alpha + M_\eta &= (Q_a \otimes P_b \otimes P_N) + (P_a \otimes Q_b \otimes P_N)
\end{aligned}$$

1.1.2 ANOVA Table

Breaking a sum of squares into independent components

We consider a two way ANOVA table without interaction. The model is given by

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \epsilon_{ijk}, \qquad i = 1, ..a, j = 1, ..b, k = 1, ..N, n = abN$$

Source	DF	SS	MS
Meam	1	$Y'\left(\frac{J_n^n}{n}\right)Y$	$Y'\left(\frac{J_n^n}{n}\right)Y$
Treatment (α)	a-1	$Y'M_{M_\alpha}Y$	$\frac{Y'M_{M_\alpha}Y}{a-1}$
Treatment (η)	b-1	$Y'M_{M_\eta}Y$	$\frac{Y'M_{M_\alpha}Y}{b-1}$
Error (ϵ)	n-a-b + 1	$Y'(I-M)Y$	$\frac{Y'(I-M)Y}{n-a-b+1}$

How to understand the degrees of freedom? M_α is the orthogonal projection operator onto the column space of X_α, where X_α is the design matrix corresponding to the model

$$Y_{ijk} = \mu + \alpha_i + \epsilon_{ijk},$$

Note that, it is not the only x_α in the column of X, instead the whole design matrix has shown the treatment effect of α. Besides, even the mean of J_n^n in this design matrix, is different from the mean in the previous two-way ANOVA model.

And we can break up the α treatment sums of square into $a-1$ separate components, each having 1 degree of freedom. That is, the quadratic form $Y'M_{M_\alpha}Y$ must be decomposed into

$$Y'M_{M_\alpha}Y = \sum_{i=1}^{a-1} Y'M_iY$$

where each M_i has rank 1 and $M_iM_j = 0, i \neq j$. Thus, in terms of subspaces, we decompose $C(M_\alpha)$ into a sum of $a-1$ orthogonal subspaces each of dimension 1. Thus

$$C(M_\alpha) = C(M_1) + C(M_2) + \ldots + C(M_{a-1})$$

So $Y'M_TY$ correspond to the sums of squares for a set of orthogonal contrasts.

Find the OPO

There are two ways to find M_α. One is to find the column space, that we use the orthogonal projection operator to get it. The second is used widely in hypothesis testing, we found the M_{H_0} and M_{H_1}, then use $M_\alpha = M_{H_1} - M_{H_0}$.

And $C(M_\mu)$ and $C(M_\alpha)$ are orthogonal.

You can see that $C(1, X) = C(X)$, as the column of X add up to J_n^n. So the OPO from $C(X)$ is actually M, however $M_\alpha = M - M_\mu$.

1.1.3 Repeated ANOVA

The projection onto each component of X are independent.

1.2 linear model

Suppose that Y is a 4×1 vector with $E(Y) = \mu, \mu \in C(E)$, where E is the set $E = \{\mu : \mu' = (\beta_1 + \beta_2 - \beta_3, \beta_2 + \beta_3, -\beta_2 - \beta_3, -\beta_1 - \beta_2 + \beta_3)\}$ where the β_i are real numbers, $i = 1, 2, 3$. Further assume that $Cov(Y) = \sigma^2 I_{4\times4}$, where σ^2 is unknown.

(a) Derive $\hat{\mu}$, the ordinary least squares estimate of μ, by carrying out the appropriate projection.
E(Y) is in the column space of C(E), we need to find the o.p.o on C(E). Also

$Cov(Y) = \sigma^2 I_{4\times 4}$, we can use ordinary least squares estimator as i.i.d.

$$\mu' = (\beta_1 + \beta_2 - \beta_3, \beta_2 + \beta_3, -\beta_2 - \beta_3, -\beta_1 - \beta_2 + \beta_3) = X\beta$$
$$\beta = (\beta_1, \beta_2, \beta_3)^T$$
$$X = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$$
$$C(X) = X_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & -1 \\ -1 & -1 \end{bmatrix}$$
$$M_\mu = X_1(X_1'X_1)^{-1}X_1^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix}$$
$$\hat{\mu} = M_\mu Y = 1/2 \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} (y_1, y_2, y_3, y_4)^T = 1/2(y_1 - y_4, y_2 - y_3, y_3 - y_2, y_4 - y_1)^T$$

(b) Find the BLUE of $\beta_2 - \beta_3$ or show that it is nonestimable.

$$\lambda = (0, 1, -1)^T, \qquad \lambda^T\beta = \beta_2 - \beta_3$$
$$\lambda^T = \rho^T X = (\rho_1, \rho_2, \rho_3, \rho_4) \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & -1 & 1 \end{bmatrix} = (0, 1, -1)$$
$$\rho_1 = \rho_4, \qquad \rho_2 - \rho_3 = 1, \qquad \rho_2 - \rho_3 = -1$$

The contradict of $\rho_2 - \rho_3$ indicate that $\beta_2 - \beta_3$ is not estimable.

(c) Consider testing $H_0 : \beta_2 + \beta_3 = 0$ versus $H_1 : \beta_2 + \beta_3 \neq 0$. Let E_0 denote the set E assuming that H_0 is true. Explicitly give the sets E_0 and $E \cap E_0^\perp$.
Find the M_0, M_1 are the o.p.o onto C(X) for H_0, H_1.**Not on C(Y)**.Then the $C(M_0), C(M_1)$ and the sets E_0 and $E \cap E_0^\perp$ relationship needs attention.

$$\lambda = (0, 1, 1)^T, \qquad \lambda^T\beta = \beta_2 + \beta_3$$
$$\lambda^T = \rho^T X = (\rho_1, \rho_2, \rho_3, \rho_4) \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & -1 & 1 \end{bmatrix} = (0, 1, 1)$$
$$\rho = (1, 2, 1, 1)^T$$

$\beta_2 + \beta_3$ is estimable with one $\rho = (1, 2, 1, 1)^T$. Then we can have $H_0 : \rho^T MY = 0$ that $\rho^T M \perp C(E_0)$.

$$M_1 = (M\rho)[(M\rho)^T(M\rho)]^{-1}(M\rho)^T$$
$$M\rho = \rho_N = (0, 1, -1, 0)^T$$
$$M_1 = \rho_N[(\rho_N)^T(\rho_N)]^{-1}(\rho_N)^T = 1/2 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$= 1/2(0, 1, -1, 0)^T$$

And the complement of $M_0 = M - M_1$

$$M_0 = M - M_1 = 1/2 \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$
$$= 1/2(1, 0, 0, -1)^T$$

Also we can just look at C(Y) column space

$$E_0 = span\{(\beta_1 + 2\beta_2, 0, 0, -\beta_1 - 2\beta_2)^T\} = span\{(1, 0, 0, -1)^T\}$$

(d) Assuming normality for Y , construct the simplest possible expression for the F statistic for the hypothesis $H_0 : \mu \in E_0$ versus $H_1 : \mu \notin E_0$, where E_0 is specified in part (c), and give the distribution of the F statistic under the null and alternative hypotheses.

$$M\rho = \rho_N = (0, 1, -1, 0)^T \in M, \qquad r(\rho_N) = 1$$
$$M_\rho = \rho_N[(\rho_N)^T(\rho_N)]^{-1}(\rho_N)^T = 1/2 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$MSE = \|(I - M)Y\| = 1/2(y_1 + y_4)^2 + 1/2(y_2 + y_4)^2$$
$$F = \frac{Y^T M_\rho Y/r(\rho)}{MSE} = \frac{2(y_2 - y_3)^2}{(y_1 + y_4)^2 + (y_2 + y_4)^2} \sim F(1, 2, \gamma), \qquad r(M - M_\rho) = 1, r(I - M) = 2$$

In which, under $H_0, \gamma = 0$, and under H_1.

$$\gamma = \frac{\|(M_1)X\beta\|}{2\|(I - M)Y\|/2}$$
$$= \frac{(\beta_2 + \beta_3)^2}{\sigma^2}$$

(e) Assuming normality for Y , construct an exact 95% confidence interval for $\beta_2+\beta_3$.
From part(d), we have

$$
\begin{aligned}
\lambda' &= (0,1,1) \\
\rho &= (1,1,0,1)^T \\
\lambda^T\beta &= \rho MY = M_1Y = 1/2(y_2-y_3) \\
F &= \frac{\|\lambda^T\beta\|/r(\rho)}{\sigma^2} = \frac{\lambda'\beta[\lambda'[X'X]^{-1}\lambda]^{-1}(\lambda'\beta)^T}{\sigma^2} \sim F(1,2,\gamma) \\
[\lambda'[X'X]^{-1}\lambda]^{-1} &= 2 \\
\{\beta &: \frac{\lambda'\beta[\lambda'[X'X]^{-1}\lambda]^{-1}(\lambda'\beta)^T}{\sigma^2} \leq F(0.95,1,2)\}
\end{aligned}
$$

Chapter 2

Parameter Estimates

2.1 The Standard Exponential Distribution

The standard exponential distribution family

$$p(y|\theta) = \phi\Big[\exp\Big(y\theta - b(\theta)\Big) - c(y)\Big] - \frac{1}{2}s(y,\phi)$$

We will explore the fun characteristics of the exponential family

(i) Mean and Variance by derivatives

$$\begin{aligned} log\int p(y|\theta) = log\int \phi\Big[\exp\Big(y\theta - b(\theta)\Big) - c(y)\Big] - \frac{1}{2}s(y,\phi)dv = 0 \\ log\int \exp\{(y\theta)\}h(y)v(dy) = b(\theta) \\ \partial_\theta log\int \exp\{(y\theta)\}h(y)v(dy) = \partial_\theta b(\theta) \end{aligned}$$

To proceed we need to move the gradient past the integral sign. In general derivatives can not be moved past integral signs (both are certain kinds of limits, and sequences of limits can differ depending on the order in which the limits are taken). However it turns out that the move is justified in this case by an appeal to the dominated convergence theorem.

$$
\begin{aligned}
\partial_\theta b(\theta) &= \partial_\theta log \int \exp\{(y\theta)\}h(y)v(dy) \\
&= \frac{\int y \exp\{(y\theta)\}h(y)v(dy)}{\int \exp\{(y\theta)\}h(y)v(dy)} \\
&= \int y \exp\{y\theta - b(\theta)\}h(x)v(dy) \\
&= E[y]
\end{aligned}
$$

Also we can see that the first derivative of $b(\theta)$ is equal to the mean of the sufficient statistics. Similar for the variance.

Another proof is to use the Bartlett's identities

Suppose that differentiation and integration are exchangeable and all the necessary expectations are finite. We have the following results:

$$
\begin{aligned}
E\_\xi\Big(\partial_j l_n\Big) &= 0, \\
E_\xi\Big(\partial^2_{j,k} l_n\Big) + E_\xi\Big(\partial_j l_n \partial_k l_n\Big) &= 0
\end{aligned}
$$

By the above two equations, we can get the expectation and variance.

2.2 The Bernoulli Distribution

The standard exponential distribution family

$$
p(y|\theta) = \phi\Big[\exp\Big(y\theta - b(\theta)\Big) - c(y)\Big] - \frac{1}{2}s(y, \phi)
$$

For Bernoulli distribution,

$$
\begin{aligned}
p(x|\pi) &= \pi^x(1-\pi)^{1-x} \\
&= \exp\{\log\Big(\frac{\pi}{1-\pi}\Big)x + \log(1-\pi)\}
\end{aligned}
$$

We see that Bernoulli distribution is an exponential family distribution with

$$
\begin{aligned}
\theta &= \log\Big(\frac{\pi}{1-\pi}\Big) \\
b(\theta) &= -\log(1-\pi) = \log\Big(1+\exp(\theta)\Big)x \\
\phi &= 1
\end{aligned}
$$

2.2.1 Mean and Variance

For a univariate random variable Y, in this case, all the Y_i have the same π

$$\begin{aligned}
\frac{\partial b(\theta)}{\partial \theta} &= \frac{\exp(\theta)}{1+\exp(\theta)} = \frac{1}{1+\exp(-\theta)} = \mu = E(Y) \\
\frac{\partial^2 b(\theta)}{\partial \theta\, \partial \theta} &= \frac{\exp(\theta)}{\Big[1+\exp(\theta)\Big]^2} = \mu(1-\mu) = Var(Y)
\end{aligned}$$

In regression model, $logit(\pi) = X\beta$, which β is a vector, then we will use the chain rule. And each individual y_i has its own equation that π_i is different.

$$\begin{aligned}
\theta &= X\beta, \qquad \theta_i = x_i^T \beta \\
\partial_\beta b(\theta_i) &= \partial_{\theta_i} b(\theta_i) \partial_\beta \theta_i \\
&= \frac{\exp(\theta_i)}{1+\exp(\theta_i)} x_i = \frac{1}{1+\exp(-\theta)} x_i = \mu_i x_i \\
\partial_\beta^2 b(\theta_i) &= \frac{\exp(\theta_i)}{\Big[1+\exp(\theta_i)\Big]^2} x_i^{\otimes 2} = \mu_i(1-\mu_i) x_i^{\otimes 2}
\end{aligned}$$

And we will need to connect this with the Fisher Information or Newton-Raphson algorithm

$$\begin{aligned}
\theta_i &= k\Big(x_i^T\beta\Big) = x_i^T\beta \\
\xi &= (\beta, \phi) \\
ln(\xi) &= \sum_{i=1}^n \phi \Big[y_i k\Big(x_i^T\beta\Big) - b\Big(k\Big(x_i^T\beta\Big)\Big) - c(y_i)\Big] - \frac{1}{2} s(y_i, \phi) \\
\dot{ln}(\beta) &= \frac{\partial ln(\beta)}{\partial \beta} = \phi \sum_{i=1}^n \Big[y_i - \dot{b}\Big(k\Big(x_i^T\beta\Big)\Big)\Big] \dot{k}\Big(x_i^T\beta\Big) x_i \\
&= \sum_{i=1}^n \Big[y_i - \mu_i\Big] x_i \\
\ddot{ln}(\beta) &= \frac{\partial^2 ln(\beta)}{\partial \beta\, \partial \beta} = -\phi \sum_{i=1}^n \ddot{b}\Big(k(x_i^T\beta)\Big) \dot{k}(x_i^T\beta)^2 x_i x_i^T + \phi \sum_{i=1}^n \Big[y_i - \dot{b}(k(x_i^T\beta))\Big] \ddot{k}(x_i^T\beta) x_i x_i^T \\
&= -\sum_{i=1}^n \ddot{b}\Big(\theta_i\Big) x_i x_i^T = -\sum_{i=1}^n V(\theta_i) x_i x_i^T, \qquad \partial_\beta^2 b(\theta_i) = V(\theta_i)
\end{aligned}$$

let

$$
\begin{aligned}
V(\theta) &= diag\{V(\theta_i)\}, \qquad e_i = y_i - \mu_i \\
\sum_{i=1}^{n} V(\theta_i) x_i x_i^T &= XV(\theta)V^T \\
\mu_i &= \dot{b}(\theta_i), \qquad v_i = \ddot{b}(\theta_i) \\
\dot{\theta}_i &= \partial_\beta \theta_i = \dot{k}(x_i^T\beta)x_i, \qquad \ddot{\theta}_i = \partial_\beta^2 \theta_i = \ddot{k}(x_i^T\beta)x_i x_i^T \\
\dot{b}(\theta_i) &= \partial_\theta b(\theta)\Big|_{\theta=\theta_i}, \dot{k}(\eta) = \partial_\eta k(\eta), \ddot{k}(\eta) = \partial_\eta^2(\eta)
\end{aligned}
$$

So

$$
E\Big[-\ddot{l}n(\beta)\Big] = \phi \sum_{i=1}^{n} v_i \dot{\theta}_i^{\otimes 2}
$$

Another set is to use $E(y_i), Var(y_i)$ which is also used commonly as that are the information we generally get. It is used a lot in GEE.

$$
\begin{aligned}
&\partial_\mu \theta = \partial_\theta \mu^{-1}, \qquad \partial_\mu \mu = \partial_\theta \mu \partial_\mu \theta = 1 \\
&\partial_\theta \mu = \partial_\theta b(\theta) = \ddot{b}(\theta) \\
&\partial_\mu \theta = \Big(\partial_\theta \mu\Big)^{-1} = \ddot{b}(\theta)^{-1}
\end{aligned}
$$

Then we have the connection between the two system

$$
\begin{aligned}
\partial_\beta \theta &= \partial_\beta \mu_i \partial_{\mu_i} \theta_i = \partial_\beta \mu_i \Big[\ddot{b}(\theta_i)\Big]^{-1} \\
\partial_\beta^2 \theta_i &= \Big(\partial_{\mu_i}^2 \theta_i\Big)\Big(\partial_\beta \mu_i\Big)^{\otimes 2} + \partial_{\mu_i}\theta_i \Big(\partial_\beta^2 \mu_i\Big) \\
&= -\dddot{b}(\theta_i)\ddot{b}(\theta_i)^{-3}\Big(\partial_\beta \mu_i\Big)^{\otimes 2} + \Big[\ddot{b}(\theta_i)\Big]^{-1}\Big(\partial_\beta^2 \mu_i\Big)
\end{aligned}
$$

The generalized estimation model

$$
\begin{aligned}
V(\beta) &= \mathrm{diag}\Big(v_1(\beta), \ldots, v_n(\beta)\Big) \\
e(\beta) &= (y_1 - \mu_1(\beta), \ldots, y_n - \mu_n(\beta))' \\
D_\theta(\beta)' &= \Big(\partial_\beta \beta_1(\beta), \ldots, \partial_\beta \beta_n(\beta)\Big)_{p\times n} \\
D(\beta)^T &= \Big(\partial_\beta \mu_1(\beta), \ldots, \partial_\beta \mu_n(\beta)\Big)_{p\times n} \\
\dot{l}_n(\beta) &= \phi D_\theta(\beta)^T e(\beta) = \phi D(\beta)' V(\beta)^{-1} e(\beta) \\
E\Big[-\ddot{l}_n(\beta)\Big] &= \phi D_\theta(\beta)' V D_\theta(\beta) = \phi D(\beta)' V(\beta)^{-1} D(\beta)
\end{aligned}
$$

Chapter 3

Convergence Theorem

3.1 Measurement Theorem and Integral

3.1.1 Continuous Convergence

Definition 3.1.1. f_n converges continuously to f, written $f_n \xrightarrow{c} f$ if for any convergent sequence $x_n \to x$ we have $f_n(x_n) \to f(x)$.

We can show by triangle inequality

$$|f_n(x_n) - f(x)| \leq |f_n(x_n) - f(x_n)| + |f(x_n) - f(x)| \leq \|f_n - f\|_K + |f(x_n) - f(x)|$$

the first term on the right-hand side converges to zero by uniform convergence on compact sets and the second term on the right-hand side converges to zeros by continuity of f.

3.1.2 Convergence Mode

It is very important to understand the definition and the notation for each definition.

(i) Converge almost everywhere

A sequence X_n converges almost everywhere (a.e) to X, denoted $X_n \xrightarrow{a.e.} X$, if $X_n(w) \to X(w)$ for all $w \in \Omega - N$ where $\mu(N) = 0$. If μ is a probability, we write a.e. as a.s. (almost surely).

$$\lim_{n\to\infty} X_n = X$$
$$P(\sup_{m\geq n} \Big| X_m - X \Big| > \epsilon) \to 0$$

Remarks: Pay attention to the notation, it says that among all the observations that after X_n, the biggest difference is less than a certain value. When the $\sup\limits_{m\geq n}$ come up, it has listed almost all the observations, which is the same as almost sure.

(ii) Converges in probability A sequence X_n converges in measure to a measurable function X, denoted $X_n \xrightarrow{\mu} X$, if $\mu(|X_n - X| \geq \epsilon) \to 0$ for all $\epsilon > 0$. If μ is a probability measure, we say X_n converges in probability to X.

$$lim_{n\to\infty} P(\|X_n - X\| > \epsilon) = 0$$

(iii) Converges in L_r-distance (rth moment)

Notation: $c = (c_1, ..., c_k) \in R^k, \|c\|_r = \left(\sum_{j=1}^k |c_j|^r\right)^{1/r}, r > 0$. If $r \geq 1$, then $\|c\|_r$ is the L_r- distance between 0 and c. When $r = 2, \|c\| = \|c\|_2 = \sqrt{c^t c}$.

$$X_n \xrightarrow{L_r} X$$
$$lim_{n\to\infty} E\|X_n - X\|_r^r = 0$$

(iv) Converges in distribution Let $F, F_n, n = 1, 2, ..,$ be c.d.f.'s on R^k and $P, P_n, n = 1, 2, ..$ be their corresponding probability measures. We say that $\{F_n\}$ converges to F weakly and write $F_n \xrightarrow{w} F$ iff, for each continuity point x of F,

$$lim_{n\to\infty} F_n(x) = F(x)$$

We say that $\{X_n\}$ converges to X in distribution and write $X_n \xrightarrow{d} X$ iff $F_{X_n} \xrightarrow{w} F_X$.

Note: converges in distribution is the cumulative distribution is the same.

$\xrightarrow{a.s.}, \xrightarrow{p}, \xrightarrow{L_r}$: measures how close is between X_n and X as $n \to \infty$.

$F_{X_n} \xrightarrow{w} F_X$: F_{X_n} is close to F_X. but X_n and X may not be close, they may be on different spaces.

Example: Let $\theta_n = 1 + n^{-1}$ and X_n be an random variable having the exponential distribution $E(0, \theta_n), n = 1, 2..$ Let X be a random variable having the exponential distribution $E(0, 1)$.

For any $x > 0$, as $n \to \infty$,

$$F_{X_n}(x) = 1 - e^{-x/\theta_n} \to 1 - e^{-x} = F_X(x)$$

Since $F_{X_n}(x) = 0 = F_X(x)$ for $x \leq 0$, we have shown that $X_n \xrightarrow{d} X$.

How about $X_n \xrightarrow{p} X$?

We will need the distribution of $X_n - X$ as we need to get the probability $P(|X_n - X| > \epsilon)$.

The distribution has two cases depends on whether X_n and X are independent or not.

(i) Suppose that X_n and X are not independent, and $X_n \equiv \theta_n X$ (then X_n has the given c.d.f.).
$X_n - X = (\theta_n - 1)X = n^{-1}X$, which has the c.d.f. $(1 - e^{-nx})I_{[0,\infty)}(x)$.
Then $X_n \xrightarrow{p} X$ because, for any $\epsilon > 0$,

$$P(|X_n - X| \geq \epsilon) = e^{-n\epsilon} \to 0$$

Also, $X_n \xrightarrow{L_p} X$ for any $p > 0$, because

$$E(|X_n - X|^p) = n^{-p} E X^p \to 0$$

(ii) Suppose that X_n and X are independent random variables. Since p.d.f.'s for X_n and $-X$ are $\theta_n^{-1}e^{-x/\theta_n}I_{(0,\infty)(x)}$ and $e^x I_{(-\infty,0)}(x)$, respectively, we have let $y = X_n - X, x = X_n$, then $-X = y - X_n < 0$. In the below range, $y \in (-\infty, x)$

$$P(|X_n - X| \leq \epsilon) = \int_{-\epsilon}^{\epsilon}\int_0^{\infty} \theta_n^{-1}e^{-x/\theta_n}e^{y-x}I_{(0,\infty)}(x)I_{(-\infty,x)}(y)dxdy$$

which converges to (by the dominated convergence theorem)

$$\begin{aligned}\int_{-\epsilon}^{\epsilon}\int_0^{\infty} e^{-x}e^{y-x}I_{(0,\infty)}(x)I_{(-\infty,-x)}(y)dxdy &= 1 - e^{-\epsilon} \\ &= \int_0^{\epsilon} e^{-2x}\int_{-\epsilon}^{x} e^y dydx \\ &= \int_0^{\epsilon} e^{-x}dx \\ &= 1 - e^{-\epsilon}\end{aligned}$$

Thus, $P(|X_n - X| \leq \epsilon) \to e^{-\epsilon} > 0$ for any $\epsilon > 0$ and, therefore, $X_n \xrightarrow{p} X$ does not hold.

3.1.3 Relationship between convergence modes

(i) If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{p} X$.

Proof:

$$P(\Big|X_i - X\Big| > \epsilon) \leq P(\sup_{m\geq n}\Big|X_m - X\Big| > \epsilon) \to 0$$

(ii) If $X_n \xrightarrow{L_r} X$ for an $r > 0$, then $X_n \xrightarrow{p} X$. Consider the definition of moment convergence and probability convergence, the link that connect Expectation and Probability with inequality is Markov Inequality.

For any positive and increasing function $g(\dot{})$ and random variable Y,

$$P(\Big|Y\Big| > \epsilon) \leq E\Big[\frac{g(|Y|)}{g(\epsilon)}\Big]$$

In particular, we choose $Y = \Big|X_n - X\Big|$ and $g(y) = |y|^r$. It gives that

$$P(\Big|X_n - X\Big| > \epsilon) \leq E\Big[\frac{\Big|X_n - X\Big|^r}{\epsilon^r}\Big] \to 0$$

(iii) If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$.

Prove: need to use the definition of convergence in probability, and construct the cumulative probability $F_X(x)$.

The purpose is to induce $F_{X_n}(x)$, so that we can compare $F_{X_n}(x)$ and $F(x)$. So the $F(x)$ will be rewritten as $F_{X_n}(x)$ and a probability involves $X_n - X$.

Assume $k = 1$, let x be a continuity point of F_X and $\epsilon > 0$ be given. Then

$$\begin{aligned} F_X(x-\epsilon) &= P(X \leq x-\epsilon, X_n \leq x) + P(X \leq x-\epsilon, X_n > x) \\ &\leq P(X_n \leq x) + P(X \leq x-\epsilon, X_n > x), \qquad P(X_n \leq x) > P(X \leq x-\epsilon, X_n \leq x) \\ &\leq F_{X_n}(x) + P(|X_n - X| > \epsilon), \qquad X_n - X > x - (x-\epsilon) = \epsilon \end{aligned}$$

Letting $n \to \infty$, we obtain that

$$F_X(x-\epsilon) \leq liminf_n F_{X_n}(x)$$

Switching X_n and X in the previous argument,

$$\begin{aligned} F_X(x+\epsilon) &= P(X \leq x+\epsilon, X_n \leq x) + P(X \leq x+\epsilon, X_n > x) \\ &\geq P(X_n \leq x) + P(X \leq x+\epsilon, X_n > x) \\ &\geq F_{X_n}(x) + P(|X_n - X| > \epsilon) \end{aligned}$$

Letting $n \to \infty$, we obtain that

$$\begin{aligned} F_X(x-\epsilon) &\leq liminf_n F_{X_n}(x) \\ F_X(x+\epsilon) &\geq limsup_n F_{X_n}(x) \end{aligned}$$

Since ϵ is arbitrary and F_X is continuous at x,

$$F_X(x) = lim_{n\to\infty} F_{X_n}(x)$$

(iv) Skorohod's theorem: a conditional converse of (i)-(iii). If $X_n \xrightarrow{d} X$, then there are random vectors $Y_n, Y_n \xrightarrow{a.s.} Y$.

(v) If, for every $\epsilon > 0, \sum_{n=1}^{\infty} P(\|X_n - X\| \geq \epsilon) < \infty$, then $X_n \xrightarrow{a.s.} X$.

(vi) If $X_n \xrightarrow{p} X$, then there is a subsequence $\{X_{n_j}, j = 1, 2..\}$ such that $X_{n_j} \xrightarrow{a.s.} X$ as $j \to \infty$.

We need to show that such a sequence exists, and prove by the almost surely definition. Such a sequence generally use the 2^{-k}. Because 2^{-k} is almost surely convergence, so any sequence that is smaller than this sequence, will definitely be almost surely convergence as well.

For any $\epsilon > 0, P(\Big|X_n - X\Big| > \epsilon) \to 0$, we choose $\epsilon = 2^{-m}$ then there exists a X_{n_m} such that

$$P(\Big|X_{n_m} - X\Big| > 2^{-m}) < 2^{-m}$$

Particularly, we can choose n_m to be increasing. For the sequence $\{X_{n_m}\}$, we note that for any $\epsilon > 0$, when n_m is large,

$$P(\sup_{k \geq m}\Big|X_{n_k} - X\Big| > \epsilon) \leq \sum_{k \geq m} P(\Big|X_{n_k} - X\Big| > 2^{-k}) \leq \sum_{k \geq m} 2^{-k} \to 0$$

Thus, $X_{n_m} \xrightarrow{a.s.} X$.

Remarks: Need to pay attention to the SUP and sum of probability, it is similar to the max of the sequence. So we need to think about the all sequence observations probability.

(vii) If $X_n \xrightarrow{d} X$, and $P(X \equiv c) \equiv 1$, where $c \in R^k$ is a constant vector, then $X_n \xrightarrow{p} c$. Let $X \equiv c$.

Prove by Polya's theorem:

$$P(\Big|X - n - c\Big| > \epsilon) \leq 1 - F_n(c + \epsilon) + F_n(c - \epsilon) \to 1 - F_X(c + \epsilon) + F_X(c - \epsilon) = 0$$

Remarks: Polya's theorem is very useful when dealing with the F_n change to F.

(viii) Moment convergence: Suppose that $X_n \xrightarrow{d} X$, then for any $r > 0$,

$$\lim_{n \to \infty} E\|X_n\|_r^r = E\|X\|_r^r < \infty$$

iff $\{\|X_n\|_r^r\}$ is uniformly integrable (UI) in the sense that

$$\lim_{t \to \infty} sup E(\|X_n\|_r^r I_{\|X_n\|_r > t}) = 0$$

In particular, $X_n \xrightarrow{L_r} X$ if and only if $\{\|X_n - X\|_r^r\}$ is UI

(viii) If $X_n \xrightarrow{p} X$ and $|X_n|^r$ is uniformly integrable, then $X_n \xrightarrow{r} X$.

3.1.4 Polya's theorem

If $F_n \xrightarrow{w} F$ and F is continuous on R^k, then

$$\lim_{n\to\infty} sup_{X\in R^k}|F_n(x) - F(x)| = 0.$$

This proposition implies the following useful result: If $c_n \in R^k$ with $C_n \to C$, then

$$F_n(C_n) \to F(C)$$

3.1.5 Fatou's lemma

Given a measure space $(\Omega, \mathbf{F}, \mu)$, and a set $X \in F$, let $\{f_n\}$ be a sequence of $(F, B_{R\geq 0})$ - measurable non-negative functions: $f_n : X \to [0, +\infty]$. Define the function $f : X \to [0, +\infty]$ by setting $f(x) = \underset{n\to\infty}{liminf} f_n(x)$, for every $x \in X$. Then f is $(F, B_{R\geq 0})$ - measurable, and also

$$\int_X f d\mu \leq \underset{n\to\infty}{liminf} \int_X f_n d\mu,$$

where the integral may be infinite.

Remarks: this lemma is used a lot in expectation of sequence.

3.1.6 Big O and Little o

In calculus, two sequences of real numbers, $\{a_n\}$ and $\{b_n\}$, satisfy

(i) $a_n = O(b_n)$ iff $\|a_n\| \leq M|b_n|$ for all n and a constant $M < \infty$. Note that the equal sign does not mean equality.

(ii) $a_n = o(b_n)$ iff $a_n/b_n \to 0$ as $n \to \infty$.

Definition 3.1.2. Let $X_1, X_2, ..$ be random vectors and $Y_1, Y_2, ..$ be random variables defined on a common probability space.

(i) $X_n = O(Y_n)$ a.s. iff $P(\|X_n\| = O(|Y_n|)) = 1$

Since $a_n = O(1)$ means that $\{a_n\}$ is bounded, $\{X_n\}$ is said to be bounded in probability if $X_n = O_p(1)$. ie, O(1) - as $x \to 0$ if it is bounded on a neighborhood of zero. And we say it is o(1) as $x \to 0$ if $f(x) \to 0, x \to 0$.

$X_n = O(Y_n)$ and $Y_n = O(Z_n)$ implies $X_n = O(Z_n)$.

$X_n = O(Y_n)$ does not imply $Y_n = O_p(X_n)$.

If $X_n = O(Z_n)$, then $X_n Y_n = O_p(Y_n Z_n)$.

If $X_n = O(Z_n)$ and $Y_n = O(Z_n)$, then $X_n + Y_n = O_p(Z_n)$.

If $X_n \xrightarrow{d} X$ for a random variable X, then $X_n = O_p(1)$.

If $E(|X_n|) = O(a_n)$, then $X_n = O_p(a_n)$, where $a_n \in (0, \infty)$.

If $X_n \xrightarrow{a.s.} X$, then $sup_n|X_n| = O_p(1)$.

(ii) $X_n = o(Y_n)$ a.s. iff $X_n/Y_n \xrightarrow{a.s.} 0$

$X_n = o(Y_n)$ implies $X_n = O_p(Y_n)$.

(iii) $X_n = O_p(Y_n)$ iff, for any $\epsilon > 0$, there is a constant $C_\epsilon > 0$ such that

$$sup_n P(\|X_n\| \geq C_\epsilon(|Y_n|)) < \epsilon$$

(iv) $X_n = o_p(Y_n)$ iff $X_n/Y_n \xrightarrow{p} 0$.

3.1.7 Big O_p and Little o_p

A sequence X_n of random vectors is said to be $O_p(1)$ if it is bounded in probability (tight) and $o_p(1)$ if it converges in probability to zero. Suppose X_n and Y_n are random sequences taking values in any normed vector space, then

$$X_n = O_p(Y_n)$$
$$Pr(\|X_n\| \leq M\|Y\|) \geq 1 - \epsilon$$

Means $X_n/\|Y_n\|$ is bounded in probability

and

$$X_n = o_p(Y_n)$$
$$\frac{X_n}{\|Y_n\|} \xrightarrow{P} 0, \qquad n \to 0$$
$$Pr(\|X_n\| \geq \epsilon\|Y_n\|) \to 0$$

Means $X_n/\|Y_n\|$ converges in probability to zero.

These notations are often used when the sequence Y_n is deterministic, for example, $X_n = O_p(n^{-1/2})$.

they are also often used when the sequence Y_n is random, for example, we say two estimators $\hat{\theta}_n$ and $\tilde{\theta}_n$ of a parameter θ are asymptotically equivalent if

$$\hat{\theta}_n - \tilde{\theta}_n = o_p(\hat{\theta}_n - \theta)$$
$$\hat{\theta}_n - \tilde{\theta}_n = o_p(\tilde{\theta}_n - \theta)$$

We also use O(1), o(1) and O_p, o_p for terms in equations. For example, a function f is differentiable at x if

$$f(x+h) = f(x) + f^{'}(x)h + o(h)$$

one case of Slutsky's theorem says

$$X_n \xrightarrow{w} X \qquad \to X_n + o_p(1) \xrightarrow{w} X$$

Bibliography

[1] Armstrong, M.A. *Basic Topology.* England: Editorial Board, 2000.

[2] Coxeter, H.S.M. *Introduction to Geometry.* Toronto: John Wiley and Sons Inc, 1969.

[3] Chronister, James. *Blender Basics.* Estados Unidos: 2004.

[4] Engelking, Ryszard. *General Topology.* Berlin: Hildermann, 1989.

[5] Engelking, Ryszard y Karol Sieklucki. *Topology: A Geometric Approach.* Berlin: Hildermann, 1992.

[6] Hatcher, Allen. *Algebraic Topology.* Cambridge: Cambridge University Press, 2002.

[7] Hoffman, Kenneth y Ray Kunze. *Linear Algebra, second edition.* New Jersey: Prentice Hall. Inc, 1961.

[8] Kackzynski, T., K. Mischainkow y M. Mrozen. *Algebraic Topology: A Computational Approach.* 2000.

[9] Kosniowski, Czes. *Topología Algebraica.* Barcelona: Editorial Revert, 1986.

[10] Lefschetz, Solomon. *Algebraic Topology.* Rhode Island: American Mathematical Society, 1991.

[11] May, J. P. *A Concise Course in Algebraic Topology.* Chicago.

[12] Mill, J. Van. *Infinite-Dimensional Topology.* Amsterdam: Elsevier Science Publications B. V, 1989.

[13] Munkres, James R. *Topology.* New York: Prentice Hall Inc, 2000.

[14] Roseman, Dennis. *Elementary Topology.* New Jersey: Prentice Hall Inc, 1999.

[15] Sato, Hajime. *Algebraic Topology: An Intuituve Approach.* Rhode Island: American Mathematical Society, 1996.

[16] Zomorodian, Afra J. *Topology for Computing.* United States: Cambridge University Press, 2005.