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 Expectation of Quadratic Form

Theorem: Let $X$ be an $n \times 1$ random vector with mean $\mu$ and covariance $\Sigma$ and let A be a symmetric matrix $n \times n$. Then, the expectation of the quadratic form

$$E[X^T AX] = \mu^T A\mu + tr(A\Sigma)$$

Proof (the proof is not easy as I thought, we need to use twice the trace property):

$$\begin{aligned}
E[X^T AX] &= E\Big[tr(X^T AX)\Big] = E\Big[tr(AXX^T)\Big] \\
&= tr\Big(AE\Big[XX^T\Big]\Big) = tr\Big[A\Big(Var(X) + E(X)E(X)^T\Big)\Big] \\
&= tr\Big[A(\Sigma + \mu\mu^T)\Big] \\
&= tr(A\Sigma) + tr(\mu^T A\mu) \\
&= tr(A\Sigma) + \mu^T A\mu
\end{aligned}$$

### 1.1.2 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 $\epsilon_{ijk}$ are i.i.d and $\epsilon_{ijk} \sim N(0, \sigma^2)$, where $\sigma^2$ is unknown. Let $M_\alpha, M_\eta$ and $M_\gamma$ denote the orthogonal operations for the $\alpha, \eta$ spaces, and interaction space, respectively.

By the previous o.p.o method we learnt, we can calculate the M using the column space of each effector. In the design matrix, the column space of A (all the columns add up to intercept) will not include the intercept component, which is exactly the Kroneck product.

Then based on that, we can further get the quadratic form of each effect, and F-test for hypothesis test.

We can summarize the subspaces and the orthogonal projection operators for the two-way ANOVA model as follows.

| Effect | Subspace | Orthogonal Projection Operator |
|---|---|---|
| $\mu$ | $C\Big(J_a \otimes J_b \otimes J_N\Big)$ | $P_a \otimes P_b \otimes P_N$ |
| $\alpha$ | $C\Big(Q_a \otimes J_b \otimes J_N\Big)$ | $Q_a \otimes P_b \otimes P_N$ |
| $\eta$ | $C\Big(J_a \otimes Q_b \otimes J_N\Big)$ | $P_a \otimes Q_b \otimes P_N$ |
| $\gamma$ | $C\Big(Q_a \otimes Q_b \otimes J_N\Big)$ | $Q_a \otimes Q_b \otimes P_N$ |
| Error | $C\Big(I - M\Big)$ | $I_a \otimes I_b \otimes Q_N$ |
| Total | | $I_a \otimes I_b \otimes I_N$ |

(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

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

The two way ANOVA model could be written as

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

The contrast is only the weighted two way interactions

$$\begin{aligned}\lambda^T\beta &= \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\gamma_{ij},\\ \lambda^T &= (0, ...0_{a+b+1}, c_{11}, ...c_{ab}),\end{aligned}$$

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

$$\begin{aligned}\rho^T &= \frac{1}{N}(c_{11}J_N^T, c_{12}J_N^T, ...c_{ab}J_N^T),\\ \rho^T X &= \frac{1}{N}(c_{11}, c_{12}, ...c_{ab}) \otimes J_N^T)(J_a \otimes J_b \otimes J_N, I_a \otimes J_b \otimes J_N, J_a \otimes I_b \otimes J_N, I_a \otimes I_b \otimes J_N),\\ \rho^T X &= \frac{1}{N}\Bigg\{\Big[(c_{11}, ...c_{ab})(J_a \otimes J_b)\Big] \otimes J_N^T J_N, \qquad \Big[(c_{11}, ..\end{aligned}$$

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

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

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

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

The variance of UMVUE is

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

(b) Using Kronecker product and notation, derive the orthogonal projection operator for the interaction space, denoted by $M_\gamma$.

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 $\mathbf{Q}_s = I_s - P_s$, where $I_s$ is the $s \times s$ identity matrix.

Thus $P_s$ is the orthogonal projection operator onto $C(J_s)$ and $\mathbf{Q}_s$ is the orthogonal projection operator onto $C(J_s)^\perp$. PAY ATTENTION THAT, WE ALWAYS WANT TO TAKE OUT THE FIRST COLUMN TO GET ORTHOGONAL MATRIX.

Computing $M_\gamma$. 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 $M_\mu$, THE FIRST COLUMN STAY AS IT IS.

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

Compute $M_\gamma$, the $\gamma$ 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 $E[Y'(M_\alpha + M_\eta)Y]$.

$$
\begin{aligned}
E[Y'(M_\alpha + M_\eta)Y] &= tr((M_\alpha + M_\eta)\Sigma) + \mu'(M_\alpha + M_\eta)\mu \\
\mu &= E[Y] = \mu \otimes J_a \otimes J_b \otimes J_N + \alpha \otimes J_b \otimes J_N + J_a \otimes \eta \otimes J_N + \gamma \otimes J_N \\
\alpha &= (\alpha_1, \alpha_2, ...\alpha_a)^T, \qquad \eta = (\eta_1, \eta_2, ...\eta_b)^T, \qquad \gamma = (\gamma_{11}, ..\gamma_{ab})^T \\
\Sigma &= \sigma^2 I_{ab} \\
E[Y'(M_\alpha + M_\eta)Y] &= tr((M_\alpha + M_\eta)\Sigma) + \mu'(M_\alpha + M_\eta)\mu \\
&= tr(M_\alpha\Sigma) + tr(M_\eta\Sigma) + \mu'M_\alpha\mu + \mu'M_\eta\mu
\end{aligned}
$$

We have $M_\alpha$ and $M_\eta$ orthogonal to $M_\mu$

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

because $Q_aJ_a = 0$

$$\begin{aligned}
\mu' M_\alpha \mu &= (\alpha \otimes J_b \otimes J_N + \gamma \otimes J_N)^T (Q_a \otimes P_b \otimes P_N)(\alpha \otimes J_b \otimes J_N + \gamma \otimes J_N) \\
&= (\alpha \otimes J_b \otimes J_N)^T (Q_a \otimes P_b \otimes P_N)(\alpha \otimes J_b \otimes J_N) + 2(\gamma \otimes J_N)^T (Q_a \otimes
\end{aligned}$$

$$+(\gamma \otimes J_N)^T (Q_a \otimes P_b \otimes P_N)(\gamma \otimes J_N)$$

$$\begin{aligned}
&= (\alpha^T Q_a \alpha) \otimes (J_b^T P_b J_b) \otimes (J_N^T P_N J_N) + 2(\gamma \otimes J_N)^T (\alpha Q_a) \otimes (P_b J_b) \\
&= (\alpha^T Q_a \alpha) \otimes (J_b^T P_b J_b) \otimes (J_N^T P_N J_N) + 2[\gamma^T (\alpha Q_a \otimes P_b J_b)] \otimes (J_N^T P
\end{aligned}$$

break down into each term

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

(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 $M\rho = \rho$,

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

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 $M_\alpha$, the $\alpha$ 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 $M_\eta$, the $\eta$ space is $(J_a \otimes Q_b \otimes J_N)$, thus

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

### 1.1.3 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 ($\alpha$) | a-1 | $Y'M_{M_\alpha}Y$ | $\frac{Y'M_{M_\alpha}Y}{a-1}$ |
| Treatment ($\eta$) | b-1 | $Y'M_{M_\eta}Y$ | $\frac{Y'M_{M_\alpha}Y}{b-1}$ |
| Error ($\epsilon$) | n-a-b + 1 | $Y'(I-M)Y$ | $\frac{Y'(I-M)Y}{n-a-b+1}$ |

How to understand the degrees of freedom? WE CAN USE THE RANK OF THE COLUMN SPACE OF X. FOR INTERCEPT, THERE IS ONLY ONE COLUMN. FOR

$\alpha$, THE RANK IS $a-1$. when we get the M rank = a, while $r(M_\mu) = 1$ and $r(M_\alpha) = r(M) - r(M_\mu) = a - 1$ because $M_\alpha$ and $M_\mu$ are orthogonal.

$M_\alpha$ is the orthogonal projection operator onto the column space of $X_\alpha$, where $X_\alpha$ is the design matrix corresponding to the model

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

I need to be aware that, we need to use Gram-Schmidt method to create two orthogonal column space $M_\mu$ and $M_\alpha$. Well, in the notes, we just took off the first column to get the $M = M_\mu + M_\alpha$. The two M are different.

And we can break up the $\alpha$ 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 $M_\alpha$. 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 $M_\alpha = M - M_\mu$.

### 1.1.4 Exercise

(i) Consider balanced two way ANOVA model with interaction $y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + \epsilon_{ijk}, i = 1, ..a, j = 1 \ldots, b, k = 1, .N$, with $\epsilon_{ijk} \sim N(0, \sigma^2)$ i.i.d. Find $E\Big[Y'\Big(\frac{1}{n}J_n^n + M_\alpha\Big)Y\Big]$ in terms of $\mu, \alpha_i, \eta_j, \gamma_{ij}$.

1. We will need to break down the quadratic form into simpler components (trace, expectations $\|MY\|$ ) for computation efficiency, after that we can calculate using matrix linear algebra. 2. The link of expectation: $E[Y] = E[\mu + \alpha_i + \eta_j + \gamma_{ij}]$, here we can write the $E(Y) = M_\mu Y = \bar{y}_{..}J_{abN}^1$, or $E(Y) = MY = \bar{y}_{i.}J_{bN}^1$. 3. The o.p.o would be the key. $M_\mu = \frac{1}{n}J_n^n$, and $M_\alpha = M - M_\mu$, so $M_\mu$ and $M_\alpha$ are orthogonal.

$$
\begin{aligned}
E\Big[Y'\Big(\frac{1}{n}J_n^n + M_\alpha\Big)Y\Big] &= E\Big[Y'MY\Big] \\
C(M) &= C\left(J_{abN} \quad Blk\Big(J_{bN}\Big)\right) = C\Big[Blk\Big(J_{bN}\Big)\Big] \\
M = X(X'X)^{-1}X' &= \begin{pmatrix} Blk\{\frac{1}{bN}J_{bN}\} & .. & 0 & 0 \\ 0 & Blk\{\frac{1}{bN}J_{bN}\} & .. & 0 \\ & .. & .. & .. \\ 0 & 0 & .. & Blk\{\frac{1}{bN}J_\}\} \end{pmatrix} \\
M_\alpha &= M - M_\mu \\
M_\mu Y_{ijk} &= Y_{..} = \mu + \alpha_{..} + \eta_{..} + \gamma_{..} \\
M_\alpha Y_{ijk} &= Y_{i.} - Y_{..} = \alpha_{i.} - \alpha_{..} + \gamma_{i.} - \gamma_{..} \\
E\Big[Y'M_\mu + M_\alpha Y\Big] &= tr\Big(\sigma^2 I_{a\times a}\Big) + E(Y)'M_\mu E(Y) + E(Y)'M_\alpha E(Y) \\
&= a\sigma^2 + bN\sum_{i=1}^{a}\left(\alpha_{i.} - \alpha_{..} + \gamma_{i.} - \gamma_{..}\right)^2 + abN\Big(\mu + \alpha_{..} + \eta_{..} + \gamma_{..}\Big)^2
\end{aligned}
$$

## 1.2 Repeated ANOVA

The model setup: assume there are 2 groups (boys and girls), and there are 10 boys, 15 girls. The measurement 4 times at age (8, 10, 12,14),

$$
\begin{aligned}
Y_i &\sim N_n(\mu_i, \Sigma) \\
\Sigma &= \sigma_b^2 J_n J_n' + \sigma_e^2 I_n
\end{aligned}
$$

The repeated ANOVA assumes the $\sigma_b^2$ is constant among the measurements of the same person. The two variance components could be written in the random effect ANOVA model.

$$
Y_{ijk} = \mu_{ij} + b_{ij} + \epsilon_{ijk}
$$

The means $\mu_{ij}$ total degrees of freedom (parameters) $(H-1)(n-1)$, and the error term degrees of freedom $(\sum_{h=1}^{H}(n_h - H)(n-1)$.

The mean squares term generally will subtract the interaction terms

| Effect | SS | DF | MS | F |
|---|---|---|---|---|
| Groups | $SS_G$ | H-1 | $MS_G = \frac{SS_G}{H-1}$ | $F = \frac{MS_G}{MS_{EU}}$ |
| Errors for groups | $SS_{GU}$ | $(\sum_{h=1}^{H} n_h - H)$ no n | $MS_{EU} = \frac{SS_{GU}}{\sum_{h=1}^{H} n_h - H)}$ | |
| Time | $SS_T$ | $n-1$ | $MS_T = \frac{SS_T}{n-1}$ | $F = \frac{MS_T}{MS_E}$ |
| Group by Time | $SS_{GT}$ | $(H-1)(n-1)$ | | |
| Errors | $SS_E$ | $(\sum_{i=1}^{H} n_h - H)(n-1)$ | $MS_E$ | |
| Total | $SS_{ToT}$ | $n\sum_{i=1}^{H} n_h - 1$ | | |

SSG and SSGT are like the one-way ANOVA, that we can get the form The interaction term, $M = M_G + M_T + M_{GT}, SS_{GT} = SSM - SS_G - SS_T$ $C(M) = C(X_{11}, ..X_{ab}) - J_n$, so $SS_M = \sum_{h=1}^{H} n_h \sum_{j=1}^{n} (\bar{Y}_{h.j} - \bar{Y}_{...})^2$ Then $M_{GT} = SS_M - SS_T - SS_G$

$$SS_G = \sum_{h=1}^{H} n_h n(\bar{Y}_{h..} - \bar{Y}_{...})^2$$

$$SS_{GU} = n \sum_{h=1}^{H} \sum_{i=1}^{n_h} (Y_{hi.} - Y_{h.})^2$$

$$SS_T = n(\sum_{h=1}^{H} n_h) \sum_{j=1}^{n} (\bar{Y}_{..j} - \bar{Y}_{...})^2$$

$$SS_{GT} = \sum_{h=1}^{H} n_h \sum_{j=1}^{n} (\bar{Y}_{h.j} - \bar{Y}_{...})^2 - SS_T - SS_G$$

$$SS_E = \sum_{h=1}^{H} n_h \sum_{j=1}^{n} (\bar{Y}_{hij} - \bar{Y}_{h.j})^2$$

$$SS_{ToT} = \sum_{h=1}^{H} n_h \sum_{j=1}^{n} (\bar{Y}_{hij} - \bar{Y}_{...})^2$$

$SS_E = SS_{ToT} - SS_G - SS_T - SS_{GT}$

## 1.3 linear model

Suppose that Y is a $4 \times 1$ 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 $\beta_i$ are real numbers, $i = 1, 2, 3$. Further assume that $Cov(Y) = \sigma^2 I_{4\times4}$, where $\sigma^2$ is unknown.

(a) Derive $\hat{\mu}$, the ordinary least squares estimate of $\mu$, 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 $Cov(Y) = \sigma^2 I_{4\times4}$, 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

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

NOTE: THE CONTRAST TEST WE GENERALLY USE

$$F - test = \frac{(\Lambda'\beta - d)^T \left(\Lambda'(X'X)\Lambda\right)^{-1} (\Lambda'\beta - d) \Big/ r(\Lambda)}{\sigma^2}$$

In this way, we don't need to calculate the $\rho, P'$ matrix.

(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 $H_1$.

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

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

$$\lambda' = (0,1,1)$$

$$\rho = (1,1,0,1)^T$$

$$\lambda^T\beta = \rho MY = M_1 Y = 1/2(y_2-y_3)$$

$$F = \frac{\|\lambda^T\beta\|/r(\rho)}{\sigma^2} = \frac{\lambda'\beta[\lambda'[X'X]^{-1}\lambda]^{-1}(\lambda'\beta)^T}{\sigma^2} \sim F(1,2,\gamma)$$

$$[\lambda'[X'X]^{-1}\lambda]^{-1} = 2$$

$$\{\beta : \frac{\lambda'\beta[\lambda'[X'X]^{-1}\lambda]^{-1}(\lambda'\beta)^T}{\sigma^2} \le F(0.95,1,2)\}$$

## 1.4 Likelihood Ratio Test

THE LIKELIHOOD FUNCTION IS THE KEY: FOR NORMAL DISTRIBUTION, WE NEED TO KNOW IF $\sigma^2$ IS KNOWN OR NOT. THE DISTRIBUTION WILL BE DIFFERENT.

SECOND, I need to get familiar with the $X\beta = MY$, especially $M$ only be calculated from column space $X$, so we don't need to rely on $\beta$ or get a MLE of $\hat{\beta}$.

### 1.4.1 Exercise

Consider the two way ANOVA with one observation per cell $Y_{ij} = \mu + \alpha_i + \eta_j + \epsilon_{ij}, i = 1, 2, j = 1, ..b, \epsilon \sim N(0, \sigma^2)$. The likelihood ratio test for

$$H_0 : \alpha_1 = \alpha_2$$
$$H_1 : \alpha_1 \neq \alpha_2$$

(a) Develop likelihood ratio test

$$\begin{aligned} LRT &= \frac{\underset{\theta\in\Theta_0}{Sup} l(\theta|y)}{\underset{\theta\in\Theta}{Sup} l(\theta|y)} \\ &= \frac{(\sigma_0^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma_0^2} Y'(I - M_0)Y\Big)}{(\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2} Y'(I - M)Y\Big)} \end{aligned}$$

Need to be aware of that, $\sigma^2 = Y'(I - M)Y$

$$\begin{aligned} LRT &= \frac{(\sigma_0^2)^{-\frac{n}{2}} \exp\Big(-\frac{n}{2}\Big)}{(\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{n}{2}\Big)} \\ &= \frac{(\sigma_0^2)^{-\frac{n}{2}}}{(\sigma^2)^{-\frac{n}{2}}} \\ &= \frac{Y'(I - M)Y}{Y'(-M_0)Y}^{\frac{n}{2}} \end{aligned}$$

The test statistics

$$LRT = \frac{Y'(I - M)Y}{Y'(-M_0)Y}^{\frac{n}{2}} < c, \qquad \text{reject H0}$$

$I - M$ and $I - M_0$ are not orthogonal, and we won't be able to get a distribution of LRT. So we generally use F-test

(b) Derive F-test

$$I - M_0 = I - M + (M - M_0)$$

Because $C(X_0) \subset C(X)$, we can prove that $(I-M)(M-M_0) = 0$, when $M_0 \subset M$.
So

$$LRT = \frac{Y'(I-M_0)Y}{Y'(I-M)Y} = \frac{Y'(I-M)Y + Y'(M-M_0)Y}{Y'(I-M)Y} = 1 + \frac{Y'(M-M_0)Y}{Y'(I-M)Y} > k, \qquad \text{reject H0}$$

$$= \frac{Y'(M-M_0)Y \Big/ r(M-M_0)}{Y'(I-M)Y \Big/ r(I-M)} > k'$$

The F-test $r(M-M_0) = 1$, and $r(I-M) = 2b - (1+b-1+1) = b-1$

$$F - test = \frac{Y'(M-M_0)Y}{Y'(I-M)Y \Big/ (b-1)} \sim F(1, b-1, \frac{\mu'(M-M_0)\mu}{2\sigma^2})$$

Pay attention that, the non-centrality is the half of the expected value under $H_1$. Under $H_0$, it is 0.

The scalar form, just need to know that

$$\begin{aligned}
M_\alpha Y &= \bar{Y}_{i.} - \bar{Y}_{..} = \mu + \alpha_i + \eta_. - \mu - \alpha_. - \eta_. = \alpha_i - \alpha_. \\
M_\mu Y &= \frac{1}{2b} J_{2b} Y_{ij} = \bar{Y}_{..} = \mu + \alpha_. + \eta_. \\
(I-M)Y &= \sum_{i=1}^{2}\sum_{j=1}^{b} Y_{ij} - (M_\mu + M_\alpha + M_\eta)Y_{ij} \\
&= \sum_{i=1}^{2}\sum_{j=1}^{b} Y_{ijk} - \bar{Y}_{..} - (\bar{Y}_{i.} - \bar{Y}_{..}) - (\bar{Y}_{.j} - \bar{Y}_{..}) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{b} Y_{ijk} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} \\
&= \mu + \alpha_i + \eta_j + \epsilon_{ij} - \Big[\mu + \alpha_. + \eta_.\Big] - \Big[\alpha_i - \alpha_.\Big] - \Big[\eta_j - \eta_.\Big] \\
&= \epsilon_{ij}, \qquad \text{due to no interaction term}
\end{aligned}$$

$M - M_0$Y could be written by breaking down into orthogonal components

$$\begin{aligned}
(M - M_0)Y &= (M_\mu + M_\alpha + M_\eta - M_\mu - M_{\alpha_0} - M_\eta)Y \\
&= (M_\alpha - M_{\alpha_0})Y
\end{aligned}$$

Under $H_0$, $M_{\alpha_0} = \frac{1}{2b}J_{2b}^{2b} - M_\mu = 0$

$$(M - M_0)Y = b\sum_{i=1}^{2}(\bar{Y}_{i.} - \bar{Y}_{..})^2 = b\sum_{i=1}^{2}(\alpha_{i.} - \alpha_{.})^2$$

$$r = \frac{b}{2\sigma^2}\sum_{i=1}^{2}(\alpha_{i.} - \alpha_{.})^2$$

## 1.5 F Test

F-test need to find two independent quadratic form, and rank. To do that, we will need to find M, directly construct the F-test from the data.

### 1.5.1 Exercise

Consider the two way ANOVA with one observation per cell $Y_{ij} = \mu + \alpha_i + \eta_j + \epsilon_{ij}, i = 1, 2, j = 1, ..b$. Consider the more general error structure

$$\begin{pmatrix} \epsilon_{1j} \\ \epsilon_{2j} \end{pmatrix} \sim N_2\Big( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \Big)$$

where $\epsilon_{1j}, \epsilon_{2j}$ are i.i.d. for $j = 1, ..b$. Define $D_j = Y_{1j} - Y_{2j}, j = 1, ..b$

The hypothesis test for

$$\begin{aligned} H_0 &: \alpha_1 = \alpha_2 \\ H_1 &: \alpha_1 \neq \alpha_2 \end{aligned}$$

(a) Develop F-test, which is only a function of the $D_j$'s.

We will need to develop the model of $D_j$, get the distribution and variance from the bi-normal of $Y_{ij}$. It is collapsed from two dimensions to one dimension, we often use delta method.

$$D_j = Y_{1j} - Y_{2j} = \alpha_1 - \alpha_2 + \epsilon_{1i} - \epsilon_{2j}$$

By Delta method,

$$\begin{aligned} Var(D_j) &= \Big(\partial_{Y_{1j}} \partial_{Y_{2j}}\Big) Var(Y_{ij}) \begin{pmatrix} \partial_{Y_{1j}} \\ \partial_{Y_{2j}} \end{pmatrix} \\ &= \Big(1 - 1\Big) \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ &= \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2 = \sigma_b^2 \end{aligned}$$

I need to notice that there are b $j's$, which this is MVN

$$\begin{aligned} D &\sim N_b\Big((\alpha_1 - \alpha_2)J_b, \sigma_b^2 I_{b\times b}\Big) \\ D &= (\alpha_1 - \alpha_2)J_b + \sigma_b^2 \epsilon, \ \epsilon = N(0, I) \end{aligned}$$

We will use F-test,

$$\begin{aligned} C(X) &=\sim J_b, \ M = \frac{1}{b} J_b^b \\ MD_j &= \bar{Y}_{1.} - \bar{Y}_{2.} \\ I - M &= I - \frac{1}{b} J_b^b, \ (I - M)D_j &= D_j - \bar{D}_j = Y_{1j} - Y_{2j} - (\bar{Y}_{1.} - \bar{Y}_{2.}) \end{aligned}$$

Then

$$F = \frac{D^T M D \Big/ r(M)}{D^T (I-M) D \Big/ r(I-M)} \quad r(M) = 1, \qquad r(I-M) = b-1$$

$$D^T M D = b(\bar{Y}_{1.} - \bar{Y}_{2.})^2$$

$$D^T (I-M) D = \sum_{i=1}^{b} (Y_{1j} - Y_{2j} - (\bar{Y}_{1.} - \bar{Y}_{2.}))^2$$

We can show that this F-test is the same as the F-test in section 1-5. There is no general condition on parameters. Whether the variance is independent or general, it does not affect the F-test. As the difference between the two observations are independent from each other.

# 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

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

$$log\int \exp\{(y\theta)\}h(y)v(dy) = b(\theta)$$

$$\partial_\theta log\int \exp\{(y\theta)\}h(y)v(dy) = \partial_\theta b(\theta)$$

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 $\pi$

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

In regression model, $logit(\pi) = X\beta$, which $\beta$ is a vector, then we will use the chain rule. And each individual $y_i$ has its own equation that $\pi_i$ 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 $\mu$ 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 $\sup_{m\geq n}$ 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 $\mu$ 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

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

(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, $P(|X_n - X| \leq \epsilon) \rightarrow e^{-\epsilon} > 0$ 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:

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

(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

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

(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

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

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

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

Switching $X_n$ and X in the previous argument,

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

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

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

Since $\epsilon$ 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 $\epsilon > 0, \sum_{n=1}^{\infty} P(\|X_n - X\| \ge \epsilon) < \infty$, 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

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

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 $\theta$ 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.