

QUALIFY EXAM

UNC

## Summarize

*Author:*  
Mingwei Fei

*Special colaborator:*

Jan, 2023



# Contents

<b>1</b>	<b>2015 Section I</b>	<b>1</b>
1.1	Problem 1- Distribution . . . . .	1
1.1.1	a. Compute Expectation . . . . .	1
1.1.2	b. MLE . . . . .	1
1.1.3	c. Probability . . . . .	2
1.1.4	d. Big O . . . . .	4
1.1.5	e. Prove convergence . . . . .	6
1.2	Problem 2 - Hypothesis Test . . . . .	6
1.2.1	LRT . . . . .	6
1.3	Problem . . . . .	7
1.3.1	a. Show that $B \sim \chi^2_{n-p-1}$ . . . . .	7
1.3.2	b. Show $A_i$ and $B_i$ are independent . . . . .	10
1.3.3	c. Derive exact distribution . . . . .	12
1.3.4	d. Outlier Model 1 . . . . .	14
1.3.5	d. Outlier Model 2 . . . . .	15
1.3.6	e. Cook's Distance . . . . .	17
1.4	Problem 3 - Expectation in case control study . . . . .	18
1.4.1	a. Expectation . . . . .	18
<b>2</b>	<b>2015 Section II</b>	<b>21</b>
2.1	Problem 1- MVN and Chi-square . . . . .	21
2.1.1	a. Quadratic Form $Y^T A Y$ and Chi-square distribution . . . . .	21
2.1.2	b. MGF of $Y^T A Y$ . . . . .	26
2.1.3	c. Trace and Rank . . . . .	28
2.1.4	d. Distribution . . . . .	30
2.2	Problem 2 - Linear Model . . . . .	31
2.2.1	a. Show that $B \sim \chi^2_{n-p-1}$ . . . . .	31
2.2.2	b. Show $A_i$ and $B_i$ are independent . . . . .	34
2.2.3	c. Derive exact distribution . . . . .	36
2.2.4	d. Outlier Model 1 . . . . .	37
2.2.5	d. Outlier Model 2 . . . . .	39
2.2.6	e. Cook's Distance . . . . .	40
2.3	Problem 3 - Expectation in case control study . . . . .	42

2.3.1	a. Expectation . . . . .	42
<b>3</b>	<b>2016</b>	<b>43</b>
3.1	Problem 1- Distribution . . . . .	43
3.1.1	a. Compute Expectation, Variance, UMVUE . . . . .	43
3.1.2	b. MLE . . . . .	44
3.1.3	c. Probability . . . . .	45
3.1.4	d. Big O . . . . .	47
3.1.5	e. Prove convergence . . . . .	48
3.2	Problem 2 - Hypothesis Test . . . . .	49
3.2.1	LRT . . . . .	49
3.3	Problem . . . . .	49
3.3.1	a. Show that $B \sim \chi^2_{n-p-1}$ . . . . .	50
3.3.2	b. Show $A_i$ and $B_i$ are independent . . . . .	53
3.3.3	c. Derive exact distribution . . . . .	55
3.3.4	d. Outlier Model 1 . . . . .	56
3.3.5	d. Outlier Model 2 . . . . .	58
3.3.6	e. Cook's Distance . . . . .	59
3.4	Problem 3 - Expectation in case control study . . . . .	61
3.4.1	a. Expectation . . . . .	61

# Chapter 1

## 2015 Section I

### 1.1 Problem 1- Distribution

Let  $X_1, X_2, \dots, X_n$  be an i.i.d. sample from the density

$$f(x) = \alpha(x - \mu)^{\alpha-1} 1\{\mu \leq x \leq \mu + 1\}, 0 < \alpha < \infty, -\infty < \mu < \infty$$

#### 1.1.1 a. Compute Expectation

$$\begin{aligned} E[(X_1 - \mu)^{-r}] &= \int_{\mu}^{\mu+1} (x - \mu)^{-r} \alpha(x - \mu)^{\alpha-1} dx \\ &= \int_{\mu}^{\mu+1} \alpha(x - \mu)^{\alpha-r-1} dx \\ &= \frac{\alpha}{\alpha - r} (x - \mu)^{\alpha-r} \Big|_{\mu}^{\mu+1} \\ &= \frac{\alpha}{\alpha - r} \end{aligned}$$

In order that  $E[(X_1 - \mu)^{-r}] > 0$ , then  $r < \alpha$ . If  $\alpha - r < 0$ , then the integral will be  $\infty$ , then the expectation is not bounded.

The integral of exponential variable is bounded only when index  $> 0$  and base within  $(0, 1)$ .

#### 1.1.2 b. MLE

Assume  $\mu$  is known, show MLE of  $\alpha$ .

$$\begin{aligned}
\log f(x) &= \log \alpha + (\alpha - 1) \log(x - \mu) I(\mu \leq x \leq \mu + 1) \\
\ln(\alpha) &= \sum_{i=1}^n \log \alpha + (\alpha - 1) \log(x_i - \mu) I(\mu \leq x_i \leq \mu + 1) \\
&= n \log \alpha + (\alpha - 1) \sum_{i=1}^n \log(x_i - \mu) I(X_{(1)} > \mu, X_{(n)} < \mu + 1) \\
\frac{d \ln}{d \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \log(x_i - \mu) = 0 \\
\alpha &= - \frac{n}{\sum_{i=1}^n \log(x_i - \mu)}
\end{aligned}$$

Show that

$$\sqrt{n}(\tilde{\alpha}_n - \alpha) \xrightarrow{d} N(0, \alpha^2)$$

We need to get the Fisher Information  $I(\alpha)$ , then use the central limit theorem.  
The Fisher information

$$\begin{aligned}
I(\alpha) &= -E\left[\frac{\partial^2 \ln}{\partial \alpha \partial \alpha}\right] \\
\frac{\partial^2 \ln}{\partial \alpha \partial \alpha} &= -\frac{1}{\alpha^2} \\
I(\alpha) &= E\frac{1}{\alpha^2} = \frac{1}{\alpha^2}
\end{aligned}$$

By central limit theorem, the variance is  $I(\alpha)^{-1}$ .

### 1.1.3 c. Probability

Define  $\tilde{\mu}_n = X_{(1)}$ ,  $\hat{\mu}_n = X_n - 1$ ,  $Y_n = n^{1/\alpha}(\tilde{\mu}_n - \mu)$ ,  $Z_n = n(\mu - \hat{\mu}_n)$ , and show that for all  $0 < y, z < \infty$ ,  $Pr(Y_n > y, Z_n > z) \rightarrow e^{-y^\alpha - \alpha z}$ , as  $n \rightarrow \infty$ , and that  $Y_n, Z_n \geq 0$  almost surely for all  $n \geq 1$ .

#### Preparation

I had the temptation to derive each individual probability for  $X_{(1)}, X_{(n)}$ , but  $Pr(Y_n > y, Z_n > z)$  could not be written as the product of  $Y_n, Z_n$  as they are not independent.

Keep in mind that, the only way to construct the joint distribution is to find the independent individual distribution first, and then start from there. This is always the trick in dealing with distribution. We can see from the derivation of  $X_{(1)}, X_{(n)}$  is that to use all the  $X_i$  in forming the cumulative probability.

And statistics is all about probability and distribution.

Need to review the proof for converge in probability

Proof

$$\begin{aligned}
P(Y_n > y, Z_n > z) &= P(n^{1/\alpha}(\tilde{\mu}_n - \mu) > y, n(\mu - \hat{\mu}_n) > z) \\
&= P(n^{1/\alpha}(X_{(1)} - \mu) > y, n(\mu - X_{(n)} + 1) > z) \\
&= P(X_{(1)} > n^{-1/\alpha}y + \mu, X_{(n)} < \mu - \frac{z}{n} + 1) \\
&= \prod_{i=1}^n P(n^{-1/\alpha}y + \mu < X_i < \mu - \frac{z}{n} + 1) \\
&= \left[ (1 - z/n)^\alpha - (yn^{-1/\alpha})^\alpha \right]^n
\end{aligned}$$

We know that

$$\begin{aligned}
\left(1 - \frac{z}{n}\right)^{\alpha n} &\rightarrow e^{-\alpha z} \\
\left(1 - \frac{y^\alpha}{n}\right)^n &\rightarrow e^{-y^\alpha}
\end{aligned}$$

To find the limit of above  $P(Y_n > y, Z_n > z)$ , we will need to change the sum term into product term.

$$\begin{aligned}
(1 - z/n)^\alpha &= \left(1 - \frac{z}{n}\right)^{\alpha/n} \\
&\rightarrow (e^{-z})^{\alpha/n} = e^{-\alpha z/n}
\end{aligned}$$

So we have

$$\begin{aligned}
\left[ (1 - z/n)^\alpha - (yn^{-1/\alpha})^\alpha \right]^n &= \left[ e^{-\alpha z/n} - \frac{y^\alpha}{n} e^{\frac{\alpha z}{n}} e^{-\frac{\alpha z}{n}} \right]^n \\
&= e^{-z\alpha} \left[ 1 - \frac{y^\alpha}{n} e^{\frac{\alpha z}{n}} \right]^n
\end{aligned}$$

Since  $e^{\alpha z/n} \rightarrow 1, \alpha z/n \rightarrow 0$ .

$$\begin{aligned}
&= e^{-z\alpha} \left[ 1 - \frac{y^\alpha}{n} \right]^n \\
&= e^{-z\alpha} e^{-y^\alpha} = e^{-z\alpha - y^\alpha}
\end{aligned}$$

To see if  $Y_n, Z_n$  almost surely  $\rightarrow 0$ , is to see every each observation  $\rightarrow 0$ .

Prove  $\hat{\mu}_n \xrightarrow{p} \mu$  let  $\epsilon_i > 0$

$$\begin{aligned}
P(|\hat{\mu}_n - \mu| > \epsilon) &= P(|X_{(n)} - 1 - \mu| > \epsilon) \\
&= P(|X_{(n)} - (\mu + 1)| > \epsilon) \\
&= P(X_{(n)} - (\mu + 1) > \epsilon) + P(X_{(n)} - (\mu + 1) < -\epsilon) \\
&= P(X_{(n)} > \epsilon + (\mu + 1)) + P(X_{(n)} < (\mu + 1) - \epsilon) \\
&= 0 + \int_{\mu}^{\mu+1-\epsilon} n\alpha(x - \mu)^{n\alpha-1} dx \\
&= (x - \mu)^{n\alpha} \Big|_{\mu}^{\mu+1-\epsilon} \\
&= (1 - \epsilon)^{n\alpha} \xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

Here we need to find the distribution of  $X_{(n)}$ .

$$\begin{aligned}
F(X_{(n)} < x) &= P(X_1 < x, X_2 < x, \dots, X_n < x) = [P(X_i < x)]^n, \quad X_i \text{ are i.i.d.} \\
P(X_i < x) &= \int_{\mu}^x \alpha(x - \mu)^{\alpha-1} dx \\
&= (x - \mu)^{\alpha}
\end{aligned}$$

Then the distribution of max observation

$$\begin{aligned}
F(X_{(n)} < x) &= (x - \mu)^{n\alpha} \\
P(X_{(n)} < (\mu + 1) - \epsilon) &= ((\mu + 1) - \epsilon - \mu)^{n\alpha} = (1 - \epsilon)^{n\alpha}
\end{aligned}$$

Because  $\hat{\mu}_n \xrightarrow{p} \mu$ , then  $|\hat{\mu}_n - \mu|^r \xrightarrow{p} 0$ . So  $\hat{\mu}_n - \mu = O_p(1)$ .

$$\begin{aligned}
|\hat{\mu}_n - \mu| \text{sign}(\hat{\mu}_n - \mu) &= op(1) \\
|\hat{\mu}_n - \mu| &= op(\text{sign}(\hat{\mu}_n - \mu) \cdot 1) = op(1)
\end{aligned}$$

By LLN,

$$\frac{1}{n} \sum (X_i - \mu)^{-r} \xrightarrow{p} E[(X_i - \mu)^{-r}] < M$$

#### 1.1.4 d. Big O

Let  $\hat{\alpha}_n = \left[ -n^{-1} \sum_{i=1}^n \log(X_i - \hat{\mu}_n) \right]^{-1}$ , and show that for any  $0 < s < \alpha$ ,  $\hat{\alpha}_n - \tilde{\alpha}_n = O_p(n^{-1 \wedge s})$ , where  $a \wedge b$  denote the minimum of a and b.



Hint: we can use the fact that  $0 < r \leq 1, 0 < C_r \leq \infty, \log(1 + \Delta) \leq C_r \Delta^r$ , for all  $0 \leq \Delta < \infty$ . Using the fact, show that

$$0 \leq \frac{1}{\tilde{\alpha}_n} - \frac{1}{\hat{\alpha}_n} \leq C_{s \wedge 1} |\hat{\mu}_n - \mu|^{s \wedge 1} n^{-1} \sum_{i=1}^n (X_i - \mu)^{-s \wedge 1}$$

and then complete the proof.

### Preparation

We can use the hint and try to get the  $\hat{\alpha} - \tilde{\alpha}_n$ . The Big-O shows that bounded and the algebraic of Big-O are also Big-O.

### Proof

$$\begin{aligned} \frac{1}{\tilde{\alpha}_n} - \frac{1}{\hat{\alpha}_n} &= -n^{-1} \sum_{i=1}^n \log(X_i - \mu) + n^{-1} \sum_{i=1}^n \log(X_i - \hat{\mu}_n) \\ \log(1 + \Delta) &\leq C_r \Delta^r, \\ &= \frac{1}{n} \sum_{i=1}^n \log(X_i - X_{(n)} + 1) - \log(X_i - \mu) \\ &= \frac{1}{n} \sum_{i=1}^n \log\left(1 + \frac{\mu - X_{(n)} + 1}{X_i - \mu}\right) \\ &\leq \frac{1}{n} \sum_{i=1}^n C_r \left[\frac{\mu - X_{(n)} + 1}{X_i - \mu}\right]^r \end{aligned}$$

Furthermore,

$$\begin{aligned} \frac{\hat{\alpha}_n - \tilde{\alpha}_n}{\hat{\alpha}_n \tilde{\alpha}_n} &\leq \frac{1}{n} \sum_{i=1}^n C_r \left[\frac{\mu - X_{(n)} + 1}{X_i - \mu}\right]^r \\ \hat{\alpha}_n - \tilde{\alpha}_n &\leq \frac{1}{n} \sum_{i=1}^n C_r \left[\frac{\mu - X_{(n)} + 1}{X_i - \mu}\right]^r \hat{\alpha}_n \tilde{\alpha}_n \\ &\leq \frac{1}{n} \sum_{i=1}^n C_r \left[\frac{\mu - X_{(n)} + 1}{X_i - \mu}\right]^r \left[-n^{-1} \sum_{i=1}^n \log(X_i - \hat{\mu}_n)\right]^{-1} \left[-n^{-1} \sum_{i=1}^n \log(X_i - \mu_n)\right]^{-1} \\ &\leq C_r |\mu - X_{(n)} + 1|^r \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{X_i - \mu}\right]^r \left[-n^{-1} \sum_{i=1}^n \log(X_i - \hat{\mu}_n)\right]^{-1} \left[-n^{-1} \sum_{i=1}^n \log(X_i - \mu_n)\right]^{-1} \end{aligned}$$

We will look at each term,

$$\begin{aligned} |\mu - X_{(n)} + 1| &\xrightarrow{p} 0, & P(|\mu - X_{(n)} + 1| > \epsilon) &= 0 \\ |\mu - X_{(n)} + 1|^r &= O_p(1) \\ \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{X_i - \mu} \right]^r &= E \left[ \frac{1}{X_i - \mu} \right]^r < M, & LLN \end{aligned}$$

### 1.1.5 e. Prove convergence

Show that for any  $1/2 < \alpha < \infty$ ,  $\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{d} N(0, \alpha^2)$ .

#### Preparation

This looks like CLT, but actually it is not. I will need to show the cumulative distribution of  $\sqrt{n}(\hat{\alpha}_n - \alpha)$  is  $N(0, \alpha^2)$ . This is very tedious and not doable.

What we can do is to use the Slutsky's theorem, that the combination of the random variables converge to a combination of distributions. So we need to get familiar with the theorems and understand how to use those theorem.

That's how we study the theorem, not only understand what it is, but also how to use it.

#### Proof

## 1.2 Problem 2 - Hypothesis Test

Suppose that the distribution of a discrete random variable  $X$  is given below

X	-2	-1	0	1	2
p(x)	$\theta_1(1 - \theta_2)$	$(\frac{1}{2} - \alpha)\frac{1-\theta_1}{1-\alpha}$	$\alpha\frac{1-\theta_1}{1-\alpha}$	$(\frac{1}{2} - \alpha)\frac{1-\theta_1}{1-\alpha}$	$\theta_1\theta_2$

where  $0 < \theta_1 < \alpha < 1/2$ ,  $\alpha$  is known,  $0 < \theta_2 < 1$ , and both  $\theta_1, \theta_2$  unknown. Suppose we would like to test hypothesis

$$H_0 : \theta_1 = \alpha, \theta_2 = 1/2$$

$$H_1 : \theta_1 < \alpha, \theta_2 \neq 1/2$$

### 1.2.1 LRT

Derive the  $\alpha$  level likelihood ratio test (LRT) for  $H_0$  versus  $H_1$  and obtain its power function.

Preparation

Proof

$$LRT = \frac{\sup_{\theta \in \Theta_0} l(\theta_0)}{\sup_{\theta \in \Theta} l(\theta)}$$

$$=$$

### 1.3 Problem

Consider the linear model  $Y = X\beta + \epsilon$ , where  $\epsilon \sim N_n(0, \sigma^2 I)$ ,  $X$  is  $n \times p$  of rank  $p$ ,  $\beta$  is  $p \times 1$ , and  $(\beta, \sigma^2)$  are unknown. Define  $H = X(X^T X)^{-1} X^T$  and let  $h_{ii}$  denote the  $i$ th diagonal element of  $H$ . Further let  $\hat{\sigma}^2$  denote the usual unbiased estimator of  $\sigma^2$  for the linear regression model and let  $\hat{\epsilon}_i$  denote the ordinary residual. Let

$$A_i = \frac{\hat{\epsilon}_i^2}{\sigma^2(1 - h_{ii})}$$

$$B_i = \frac{(n - p)\hat{\sigma}^2}{\sigma^2} - \frac{\hat{\epsilon}_i^2}{\sigma^2(1 - h_{ii})}$$

1.3.1 a. Show that  $B \sim \chi_{n-p-1}^2$

Preparation

In order to prove  $B \sim \chi_{n-p-1}^2$ , we need to show that it is a sum of  $n-p-1$  normal distribution square (by definition of chi-square distribution).

We can write  $\sigma^2, \hat{\sigma}^2$  the quadratic form in the sum of squares of normal distribution. But the fraction between two quadratic forms is what distribution?

Note that the division is not F-distribution, as it didn't divided by degrees of freedom.

$\hat{\sigma}^2$  is the usual unbiased estimator of  $\sigma^2$ , we need to know that if unbiased, then the degrees of freedom is  $n-p$ .

Another thing need to pay attention that,  $\sigma^2$  is not a distribution, it is fixed and it is the variance of  $Y_i$  just unknown.

Variance of  $Y_i$  is also the variance of  $\epsilon$ , but we need to know the difference and how to use them.

$$\frac{(n - p)\hat{\sigma}^2}{\sigma^2} = \frac{Y^T(I - H)Y}{\sigma^2} = \frac{[(I - H)Y]^T[(I - H)Y]}{\sigma^2}$$

$$= \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{Var(y_i)}$$

So is  $n-p$  sum of normal distribution square. Then try to link the second term

$$\frac{\hat{\epsilon}_i^2}{\sigma^2(1 - h_{ii})} = \frac{\hat{\epsilon}_i^2}{V(e_i)}$$

This question requires understanding the matrix  $H$ , or the relationship between o.p.o matrix and scalar form. And the characteristic of  $H$  matrix itself  $H^2 = H, H^T = H$ .

If we write that in scalar form,

$$(h_{i1}, h_{i2} \dots h_{in}) \begin{pmatrix} h_{i1} \\ h_{i2} \\ \dots \\ h_{in} \end{pmatrix} = h_{i1}^2 + h_{i2}^2 + \dots + h_{in}^2 = h_{ii}$$

Variance, Covariance of  $e$

$$\begin{aligned} V(e) &= E[(e - E(e_i))(e - E(e_i))^T] = (I - H)E(\epsilon\epsilon^T)(I - H)^T \\ &= (I - H)I\sigma^2(I - H)^T \\ &= (I - H)(I - H)^T I\sigma^2 = (I - H)\sigma^2 \end{aligned}$$

Need to know that  $V(e_i)$  is given by the  $i$ th diagonal element  $1 - h_{ii}$  and  $Cov(e_i, e_j)$  is given by the  $(i, j)$ th element of  $-h_{ij}$  of the matrix  $(I - H)\sigma^2$ .

We also need to know that both  $Y_i$  and  $\epsilon_i$  follows the same distribution

$$\begin{aligned} e - E(e) &= (I - H)Y = (Y - X\beta) = (I - H)\epsilon \\ E(\epsilon\epsilon^T) &= V(\epsilon) = I\sigma^2, \quad E(\epsilon) = 0 \end{aligned}$$

We also have correlation

$$\rho_{ij} = \frac{Cov(e_i, e_j)}{\sqrt{V(e_i)V(e_j)}} = -\frac{h_{ij}}{(1 - h_{ii})(1 - h_{jj})}$$

$$SS(b) = SS(parameter) = b^T X^T Y = \hat{Y}^T Y = Y^T H^T Y = Y^T H Y = Y^T H^2 Y = \hat{Y}^T \hat{Y}$$

The average  $V(\hat{Y}_i)$  to all data points is

$$\begin{aligned} \sum_{i=1}^n \frac{V(\hat{Y}_i)}{n} &= \frac{trace(H\sigma^2)}{n} = \frac{p\sigma^2}{n} \\ \hat{Y}_i &= h_{ii}Y_i + \sum_{j \neq i} h_{ij}Y_j \end{aligned}$$

Studentized residual

$$V(e_i) = (1 - h_{ii})\sigma^2$$

where  $\sigma^2$  is estimated by  $s^2$

$$s^2 = \frac{e^T e}{n - p} = \frac{\sum_{i=1}^n e_i^2}{n - p}$$

Sum of Squares attributable to  $e_i$

$$e = (I - H)Y$$

$$e_i = -h_{i1}Y_1 - h_{i2}Y_2 - \dots + (1 - h_{ii})Y_i - \dots - h_{in}Y_n = c^T Y$$

$$Var(e_i) = c^T Var(Y) c = \sigma^2 c^T c$$

$$c^T = (-h_{i1}, -h_{i2}, \dots, (1 - h_{ii}), \dots, -h_{in}), \quad (I - H)I - H = I - H$$

$$c^T c = \sum_{j=1}^n h_{ij}^2 + (1 - 2h_{ii}) = (1 - h_{ii})$$

$$Var(e_i) = \sigma^2(1 - h_{ii})$$

$$SS(e_i) = e_i^2 = \left( (I - H)y \right)^T (I - H)y = y^T (I - H)y$$

Need to learn to write in matrix form.

Proof

Write  $B_i$  in matrix form

$$\begin{aligned} \frac{(n - p)\hat{\sigma}^2}{\sigma^2} &= \frac{Y^T(I - H)Y}{\sigma^2} = \frac{[(I - H)Y]^T[(I - H)Y]}{\sigma^2} \\ \frac{\hat{e}_i^2}{\sigma^2(1 - h_{ii})} &= \frac{[e_i^T(I - H)Y]^T[e_i^T(I - H)Y]/(1 - h_{ii})}{\sigma^2} \end{aligned}$$

where  $e_i = (0, \dots, 0, 1, 0, \dots, 0)$  Then

$$\frac{[e_i^T(I - H)Y]^T[e_i^T(I - H)Y]/(1 - h_{ii})}{\sigma^2} = \frac{[Y^T(I - H)e_i e_i^T(I - H)Y]/(1 - h_{ii})}{\sigma^2}$$

Then

$$\begin{aligned} B_i &= \frac{[Y^T(I - H)Y]}{\sigma^2} - \frac{[Y^T(I - H)e_i e_i^T(I - H)Y]/(1 - h_{ii})}{\sigma^2} \\ &= \frac{Y^T}{\sigma} \left[ (I - H) - (I - H) \frac{e_i e_i^T}{1 - h_{ii}} (I - H) \right] \frac{Y}{\sigma} \end{aligned}$$

$$\frac{Y}{\sigma} \sim N_n(0, I).$$

To show that  $C = \left[ (I - H) - (I - H) \frac{e_i e_i^T}{1 - h_{ii}} (I - H) \right]$  is a projection matrix, we need to show  $C^2 = C, C^T = C$ .

Here  $(I - H)(I - H) = I - H, e_i^T (I - H) e_i = 1 - h_{ii}$

Therefore,  $B_i \sim \chi^2(r)$ , where  $r = \text{rank}[C]$ . For projection matrix, the rank is also the trace.

$$\begin{aligned}
 \text{rank} \left[ (I - H) - (I - H) \frac{e_i e_i^T}{1 - h_{ii}} (I - H) \right] &= \text{tr} \left[ (I - H) - (I - H) \frac{e_i e_i^T}{1 - h_{ii}} (I - H) \right] \\
 &= n - p - \frac{1}{1 - h_{ii}} \text{tr}[(I - H) e_i e_i^T (I - H)] \\
 &= n - p - \frac{1}{1 - h_{ii}} \text{tr}[e_i^T (I - H) (I - H) e_i] \\
 &= n - p - \frac{1}{1 - h_{ii}} \text{tr}[e_i^T (I - H) e_i] \\
 &= n - p - \frac{1 - h_{ii}}{1 - h_{ii}} \\
 &= n - p - 1
 \end{aligned}$$

So  $B_i \sim \chi^2(n - p - 1)$ .

### 1.3.2 b. Show $A_i$ and $B_i$ are independent

#### Preparation

Note that  $A_i$  and  $B_i$  are quadratic forms, which we need to show that their projection matrix product is 0, which will need to use linear algebra.

Another method is that, there are chi-square distribution, which we just need to show that the two statistics are ancillary.

Traditional method is to derive the distributions, and if we could write the distribution of  $A_i + B_i$  as the product of distribution of  $A_i$  and  $B_i$ , then we can prove the independence.

One more question, are we able to use the moment generating function to do this? It is always easier to prove in MGF.

First, we need to write the MGF of  $A_i, B_i$

$$\begin{aligned}
 M_{A_i} &= E[\exp[A_i t]] = \int \exp[A_i t] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
 M_{B_i} &= E[\exp[B_i t]] = \int \exp[B_i t] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx
 \end{aligned}$$

Then we can use MGF properties to prove

$$\begin{aligned}
W_i &= A_i + B_i \\
M_{W_i} &= E[\exp(W_i t)] = E[\exp((A_i + B_i)t)] \\
&= E[\exp(A_i t) \exp(B_i t)], \quad \text{all based on normal distribution} \\
&= E[\exp(A_i t)] E[\exp(B_i t)] = M_{A_i} M_{B_i}
\end{aligned}$$

Also we need to notice that, this is a quadratic form chi-square distribution.

(a) Quadratic form theorem 1:

Let the  $n \times 1$  vector  $y \sim N(0, I)$ , let  $A$  be an  $n \times n$  idempotent matrix of rank  $m$ , let  $B$  be an  $n \times n$  idempotent matrix of rank  $s$ , and suppose  $BA = 0$ . Then  $y' Ay$  and  $y' By$  are independently distributed  $\chi^2$  variables.

**Why need to be idempotent matrix?** It is because for idempotent matrix  $A^2 = A$ , then the quadratic form could be written as  $y' Ay = (Ay)^T (Ay)$ , then further we can use the product of two vectors = 0 to show independence.

The proof is try write the two quadratic form in partition, so that show that they dependent on different vectors, further they are independent. Need to compare with the proof in 762 slides.

(b) Quadratic form theorem 2 (Craig's Theorem): If  $y \sim N(\mu, \Omega)$  where  $\Omega$  is positive definite, then  $q_1 = y' Ay$  and  $q_2 = y' By$  are independently distributed if  $A\Omega B = 0$

This is just a generalization of above theorem. Since  $\Omega$  is a covariance matrix of full rank it is positive definite and can be factored as  $\Omega = TT'$ . Therefore the condition  $A\Omega B = 0$  can be written  $ATT'B = 0$ .

$$\begin{aligned}
T' ATT' BT &= 0 \\
C &= T' AT, \quad K = T' BT \\
CK &= (T' AT)(T' BT) = T' \Omega BT = T' 0T = 0
\end{aligned}$$

In the 762 slides, it used the covariance matrix = 0 to prove independence in MVN.

$$\begin{aligned}
Y' AY &= Y' RR' Y = (R' Y)' (R' Y) \\
Y' BY &= Y' SS' Y = (S' Y)' (S' Y)
\end{aligned}$$

Thus  $Y' AY$  and  $Y' BY$  are independent if  $R' Y$  and  $S' Y$  are independent.  $R' Y$  and  $S' Y$  are independent if and only if

$$\begin{aligned}
Cov(R'Y, S'Y) &= 0 \\
&= R'\Sigma S = 0 \\
&= R'WW'S = 0 \\
&= C(Q'S) \perp C(Q'R)
\end{aligned}$$

Proof

$$\begin{aligned}
A_i &= \frac{[Y^T(I-H)e_ie_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2} \\
B_i &= \frac{[Y^T\left[(I-H) - (I-H)\frac{e_ie_i^T}{(1-h_{ii})}(I-H)\right]Y]}{\sigma^2}
\end{aligned}$$

We would like to show the product of two projection matrix are 0, by using the above theorem

$$\begin{aligned}
M_{A_i}M_{B_i} &= (I-H)\frac{e_ie_i^T}{1-h_{ii}}(I-H)\left[(I-H) - (I-H)\frac{e_ie_i^T}{(1-h_{ii})}(I-H)\right] \\
&= 0 \\
e_i^T(I-H)e_i &= 1-h_{ii}
\end{aligned}$$

### 1.3.3 c. Derive exact distribution

Let

$$r_i = \frac{\hat{\epsilon}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$$

Using part(a) and (b), derive the exact distribution of  $r_i^2/(n-p)$ .

Preparation

This problem needs observation of the  $r_i$  and connection between  $A_i$  and  $B_i$ . Also use the results from part (a) and (b) that,  $A_i$  and  $B_i$  are independent chi-square.

$$\frac{r_i^2}{(n-p)} = \frac{A_i}{A_i + B_i}$$

The transformation of distribution, first we can use the distribution theory that the sum of independent chi-square distribution is also a chi-square. The first method we need to try is to use MGF, then try to use the distribution transformation.



if we use distribution transformation, then we need to use two independent distributions

$$w = \frac{A_i}{A_i + B_i}$$

$$B_i = \frac{1-w}{w}A_i$$

replace  $B_i$  by  $w$ , then integrate out of  $A_i$ , then we will get the distribution of  $w$ .

the chi-square distribution  $X \sim \chi^2(k)$  is a special case of the gamma distribution, in that  $X \sim \Gamma(\frac{k}{2}, \frac{1}{2})$  using the rate parameterization of the gamma distribution.

**Proof**

We need to know the relationship between Gamma distribution and Beta distribution.  
As  $A_i \perp B_i$

$$\frac{A_i}{A_i + B_i} \sim \text{Beta}\left(\frac{1}{2}, \frac{n-p-1}{2}\right)$$

Show the distribution transformation. Assume that  $X \sim \text{Gamma}(\alpha, 1), Y \sim \text{Gamma}(\beta, 1)$ , and  $X$  and  $Y$  are independent. then find the distribution of  $\frac{X}{X+Y}$ .

$$f_{X,Y}(x,y) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-(x+y)}$$

let  $w = \frac{x}{x+y}, u = x$ , then we have

$$x = y, \quad y = \frac{(1-w)u}{w}$$

$$J = \left| \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial w} \end{pmatrix} \right|$$

$$= \left| \begin{pmatrix} 1 & 0 \\ \frac{(1-w)}{w} & -\frac{u}{w^2} \end{pmatrix} \right| = \frac{u}{w^2}$$

Then integrate out  $u$

$$f(w) = \int_0^\infty f(w, u) du$$

$$= \int_0^\infty \frac{1}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} \left(\frac{(1-w)u}{w}\right)^{\beta-1} e^{-\frac{u}{w}} \frac{u}{w^2} du$$

$$\sim \text{Beta}(\alpha, \beta)$$

### 1.3.4 d. Outlier Model 1

Suppose we suspect that the  $i$ th case is an outlier and we consider the mean shift outlier model  $Y = X\beta + d_i\phi + \epsilon$ , where  $\epsilon \sim N_n(0, \sigma^2 I)$ ,  $\phi$  is an unknown scalar, and  $d_i$  is an  $n \times 1$  vector with a 1 in the  $i$ th position and zeros elsewhere. Derive the maximum likelihood estimate of  $\phi$ .

#### Preparation

We can construct the model by introducing the unknown parameter  $\phi$ , then use the MLE to estimate.

The way to get MLE is to construct likelihood function first,  $Y$  is a normal distribution, and only  $i$ th case is an outlier, while all other cases are not. Solve the problem similar as the random model.

As the  $Y$  are independent variables,  $Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n \sim N(X_k\beta, \sigma^2)$ , while  $Y_i \sim N(X_i\beta + \phi, \sigma^2)$ .

If using scalar form, the  $\beta, \phi$  are correlated with each other, which we won't be able to solve.

$$\begin{aligned} f_k(\beta, \phi) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_k - x_k\beta)^2}{2\sigma^2}\right) \\ f_i(\beta, \phi) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - x_i\beta - \phi)^2}{2\sigma^2}\right) \\ \log f_k(\beta, \phi) &= -\frac{(y_k - x_k\beta)^2}{2\sigma^2} \\ \log f_i(\beta, \phi) &= -\frac{(y_i - x_i\beta - \phi)^2}{2\sigma^2} \\ \ln(\beta, \phi) &= \sum_{k=1, k \neq i}^n -\frac{(y_k - x_k\beta)^2}{2\sigma^2} - \frac{(y_i - x_i\beta - \phi)^2}{2\sigma^2} \end{aligned}$$

The MLE of  $\phi$

$$\begin{aligned} \frac{\partial \ln}{\partial \phi} &= \frac{(y_i - x_i\beta - \phi)}{\sigma^2} = 0 \\ \phi &= y_i - x_i\hat{\beta} \end{aligned}$$

where  $\hat{\beta}$  is the MLE of  $\beta$ .

So we need to use the linear algebra form in solve this kind of question. Need to notice that the linear regression estimate in matrix form all use the linear algebra.

#### Proof

The maximum likelihood function in linear algebra form

$$\begin{aligned}
L(\beta, \phi) &\propto \exp\left(-\frac{(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi)}{2\sigma^2}\right) \\
l(\beta, \phi) &\propto \left(-\frac{(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi)}{2\sigma^2}\right) \\
&= -\frac{1}{2\sigma^2} \left[ (Y - X\beta)^T(Y - X\beta) - 2(Y - X\beta)^T d_i\phi + d_i^T d_i\phi^2 \right]
\end{aligned}$$

The MLE

$$\begin{aligned}
\frac{\partial l}{\partial \phi} &= \left(-\frac{(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi)}{2\sigma^2}\right) \\
&\rightarrow -2(Y - X\beta)^T d_i + d_i^T d_i(2\phi) = 0 \\
\phi &= (Y - X\beta)^T d_i
\end{aligned}$$

MLE of  $\beta$

$$\begin{aligned}
X\beta &= HY \\
\hat{\beta} &= (X^T X)^{-1} X^T Y
\end{aligned}$$

There is no need to get the estimate of  $\beta$ , we only need to estimate  $(Y - X\beta)^T$ . So  $\hat{\phi} = (I - H)Y^T d_i$

### 1.3.5 d. Outlier Model 2

Suppose we wish to test  $H_0 : \phi = 0$ . Derive the test statistic for this hypothesis and derive its exact distribution under  $H_0$ .

Preparation

The hypothesis test always comes with the estimate. There are several hypothesis test method we can do, wald test, score test and likelihood ratio test. The score test is generally the way as it is easy to get the estimate, score function and fisher information under  $H_0$ .

We can use  $A_i$  and  $B_i$  are independent and they are quadratic form, so we can have F-test and try to find the connection between the  $\hat{\phi}$  and F-test.

Proof

To test  $\phi = 0$ , consider wald test

$$W_n = (\hat{\phi} - 0)^T I_n(\phi)(\hat{\phi} - 0)$$

Find the fisher information  $I_n(\phi)$

$$\begin{aligned}
 I_n(\phi) &= nI(\phi) \\
 I(\phi) &= -E\left[\frac{\partial^2 l}{\partial \phi \partial \phi}\right] \\
 &= \frac{1}{\sigma^2} \\
 I_n(\phi) &= \frac{n}{\sigma^2}
 \end{aligned}$$

We need to find the MLE of  $\sigma^2$ .

$$\begin{aligned}
 \frac{\partial^2 l}{\partial \sigma^2} &= \frac{1}{2\sigma^4}(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi) + \frac{\partial}{\partial \sigma^2} \log[(\sigma^2)^{-1/2}] \\
 \hat{\sigma}^2 &= \|Y - X\beta - d_i\phi\| \\
 &= \|(I - H)Y - d_i\hat{\phi}\|
 \end{aligned}$$

So the wald statistics

$$W_n = d_i^T Y(I - H) \left( \frac{n}{\|(I - H)Y - d_i\hat{\phi}\|} \right) (I - H)Y^T d_i$$

Under  $H_0$ , it is a  $\chi^2(1)$ .

The second method is to use F-test, as  $A_i$  and  $B_i$  are independent quadratic form

$$\begin{aligned}
 A_i &= \frac{[Y^T(I - H)e_i e_i^T(I - H)Y]/(1 - h_{ii})}{\sigma^2} = \frac{[\hat{\phi}^2]/(1 - h_{ii})}{\sigma^2} \\
 B_i &= \frac{[(I - H)Y]^T[(I - H)Y] - [\hat{\phi}^2]/(1 - h_{ii})}{\sigma^2}
 \end{aligned}$$

$$\begin{aligned}
 F &= \frac{A_i}{B_i/(n - p - 1)} \\
 &= \frac{\hat{\phi}^2/(1 - h_{ii})}{[(I - H)Y]^T[(I - H)Y] - [\hat{\phi}^2]/(1 - h_{ii})/(n - p - 1)}
 \end{aligned}$$

We can perform F-test by setting up the rejection region.

### 1.3.6 e. Cook's Distance

Let  $I = \{1, \dots, m\}$  be the subset of the first  $m$  cases in the dataset. Let  $D_I$  denote the Cook's distance based on simultaneously deleting  $m$  cases from the dataset, which is given by

$$D_I = \frac{(\hat{\beta} - \hat{\beta}_I)^T (X^T X)(\hat{\beta} - \hat{\beta}_I)}{p\hat{\sigma}^2}$$

where  $\hat{\beta}^I$  denotes the least squares estimate of  $\beta$  with the cases deleted from set  $I$  and  $\hat{\beta}$  denotes the estimate of  $\beta$  based on the full data. Show that  $D_I$  can be written as

$$D_I = \frac{1}{p} \sum_{i=1}^m h_i^2 \left( \frac{\lambda_i}{1 - \lambda_i} \right)$$

where the  $\lambda_i, i = 1, \dots, m$ , are the eigenvalues of the matrix  $P_I = X_I(X_I^T X_I)^{-1}X_I^T$  based on a spectral decomposition of  $P_I$ .

#### Preparation

The single deletion case

$$\hat{\beta}_{(1)} = \hat{\beta} - \frac{(X'X)^{-1}x_i^T \hat{\epsilon}_i}{1 - h_{ii}}$$

This method is used in Bayesian method, get marginal distribution from joint distribution, expectation of conditional distribution.

The probability involved in retrospective study sampling method

The likelihood function for the retrospective study is the conditional density of  $X = x$  given  $Y$  and  $S = 1$ , denoted by  $p(x|Y, S = 1)$ . To avoid biased sampling, we assume that the inclusion of a subject is independent of  $x$  given  $Y$ , that is,

$$P(S = 1|X, Y) = P(S = 1|Y)$$

Under unbiased sampling, we have

$$p(x|Y, S = 1) = \frac{P(S = 1|x, Y)p(x|Y)}{P(S = 1|Y)} = p(x|Y)$$

Furthermore,  $p(x|Y, S = 1)$  can be expressed as

$$p(x|Y, S = 1) = \frac{P(Y|S = 1, x)p(x|Y)P(x|S = 1)}{P(Y|S = 1)} = p(x|Y)$$

and  $p(Y|x, S = 1)$  is given by

$$\begin{aligned}
p(Y|x, S=1) &= \frac{p(X|Y, S=1)p(Y, S=1)}{p(X, S=1)} = p(Y|x) \frac{p(S=1|Y)}{p(Y|S=1)} \\
&= \frac{p(X, Y, S=1)}{p(X, S=1)} = \frac{p(Y, S=1|x)p(x)}{p(S=1|x)p(x)} \\
&= \frac{p(S=1|x, Y)p(Y)}{p(S=1|x)} = \frac{p(S=1|Y)p(Y|x)}{p(S=1|x)}
\end{aligned}$$

Finally,  $p(x|Y, S=1)$  can be expressed as

$$p(x|Y, S=1) = \frac{P(Y|x)p(x|S=1)P(S=1|Y)}{P(Y|S=1)p(S=1|x)}$$

Why do we write in this form? As it is how we select the patients in the case-control study, and it reveals the relationship between prospective and retrospective probability.

The ratio  $p(x|S=1)/p(S=1|x)$  only depends on  $x$  and  $P(S=1|Y)$  and  $P(Y|S=1)$  are determined by the sampling scheme.

We can construct the likelihood function for the retrospective study

Let  $H_1(Y) = P(S=1|Y)/P(Y|S=1)$  and  $H_2(x) = p(x|S=1)/p(S=1|x)$

$$\prod_{i=1}^n p(x_i|Y_i, S_i=1) = \prod_{i=1}^n P(y_i|x_i, \beta) \prod_{i=1}^n [H_1(y_i)H_2(x_i)]$$

Proof

## 1.4 Problem 3 - Expectation in case control study

### 1.4.1 a. Expectation

Consider the conditional mean of  $Y$  given  $(X, D)$ , denoted by  $\tilde{\mu}(X, D) = E[Y|X, D]$ . Show that

$$\begin{aligned}
\mu(X) &= \tilde{\mu}(X, 1)Pr(D=1|X) + \tilde{\mu}(X, 0)Pr(D=0|X) \\
\tilde{\mu}(X, D) &= E[Y|X, D, S=1] = \mu(X) + \{D - Pr(D=1|X)\}\gamma(X)
\end{aligned}$$

Preparation

Find the expectations in the case-control study from 762 slides.

### Proof

The expectation of conditional distribution

$$\begin{aligned} E(Y|X) &= E[E[Y|X, D]|X] \\ &= E[Y|X, D = 1]Pr(D = 1|X) + E[Y|X, D = 0]Pr(D = 0|X) \\ &= \tilde{\mu}(X, 1)Pr(D = 1|X) + \tilde{\mu}(X, 0)Pr(D = 0|X) \end{aligned}$$

In terms of  $\tilde{\mu}(X, D) = E[Y|X, D, S = 1]$ , here we know that all the patients in the case-control study with  $S = 1$ , so  $\tilde{\mu}(X, D) = E[Y|X, D, S = 1] == E[Y|X, D]$ .

$$E(Y|X) = \tilde{\mu}(X, 1)Pr(D = 1|X) + \tilde{\mu}(X, 0)Pr(D = 0|X)$$





## Chapter 2

## 2015 Section II

### 2.1 Problem 1- MVN and Chi-square

Suppose that  $Y \sim N(\mu, \Sigma)$  where  $\Sigma$  is symmetric and full rank, Let  $A$  be a symmetric matrix.

#### 2.1.1 a. Quadratic Form $Y^T A Y$ and Chi-square distribution

Show that the quadratic form  $Y^T A Y$  can be represented as

$$Y^T A Y = \sum_{i=1}^k \lambda_i W_i$$

where the  $W_i$ 's are independently distributed as noncentral chi-squared variables with  $d_i$  degrees of freedom and noncentrality parameter  $\delta_i$ , that is,  $W_i \sim \chi_{d_i}^2(\delta_i)$ ,  $i = 1, 2, \dots, k$ . Indicate what  $\lambda_i, d_i, \delta_i$  are equal to.

#### Question

- (i) *Quadratic form of normal random variable with a specific variance:*

Assume  $Y$  is a normal random variable with mean 0 and a specific variance is

$$\begin{aligned}\Sigma &= [\Sigma^{1/2}]^T [\Sigma^{1/2}], && \text{Cholesky decomposition} \\ Y &= [\Sigma^{1/2}] Z, && Z \sim N(0, 1) \\ Y_i^T Y_i &= \left([\Sigma^{1/2}] Z\right)^T [\Sigma^{1/2}] Z = Z^T \Sigma Z = \sigma^2 \chi^2\end{aligned}$$

- (ii) non-centrality chi-square distribution

And if assume  $Y$  the mean is not 0,  $Y \sim N(\mu, 1)$ , then we have

$$\begin{aligned} Y_i^T Y_i &= \sigma^2 \chi^2(\delta) \\ \delta &= \mu_i^2 \end{aligned}$$

The mean ( $\mu$ ) of the chi-square distribution is its degrees of freedom,  $k$ . Because the chi-square distribution is right-skewed, the mean is greater than the median and mode. The variance of the chi-square distribution is  $2k$ .

Suppose  $Y_{n \times n}$ , and  $\Sigma$  is full rank. so  $\Sigma$  is  $n \times n$  dimension matrix,

We can transform  $Y_i$  into  $N(\mu, 1)$  distribution, so that the quadratic form will be a non-central chi-square distribution.

If  $Z_1, \dots, Z_k$  are independent, standard normal random variables, then the sum of their squares is chi-square distribution,

$$\begin{aligned} Q &= Z_i^2 \sim \chi^2(k) \\ p(k) &= \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} \exp(-\frac{x}{2}) \end{aligned}$$

*Non-central Chi-square:*

Here the  $k$  is unknown, we need to show that the sum of non-central chi-square distribution is also a non-central chi-square distribution with distribution transformation. The distribution transformation generally use Moment Generating Function.

**Lemma:**

Let  $Q_i \sim \chi_{k_i}^2(\lambda_i)$  for  $i = 1, \dots, n$ , be independent. Then,  $Q = \sum_{i=1}^n Q_i$  is a noncentral  $\chi_k^2(\lambda)$ , where  $k = \sum_{i=1}^n k_i$  and  $\lambda = \sum_{i=1}^n \lambda_i$ .

Chi-square distribution and non-central chi-square distribution are totally different. I need to understand the components  $\delta_i$  and  $d_i$  in the non-central chi-square distribution.

The non-central chi-square distribution: Let  $(X_1, X_2, \dots, X_i, \dots, X_k)$  be  $k$  independent, normally distributed random variables with means  $\mu_i$  and unit variances. Then the random variable

$$Q = \sum_{i=1}^k X_i^2 \sim \chi^2(k, \lambda), \quad \lambda = \sum_{i=1}^k \mu_i^2$$

(iii) Theorem using spectral decomposition

For any  $k \times k$  symmetric matrix  $A$  the quadratic form defined by  $A$  can be written using its spectral decomposition as

$$Q(x) = x'Ax = \sum_{i=1}^k \lambda_i \|q_i^T x\|^2$$

where the eigendecomposition of  $A$  is  $Q'\Lambda Q$  with  $\Lambda$  diagonal with diagonal elements  $\lambda_i, i = 1, \dots, k$ ,  $Q$  is the orthogonal matrix with the eigenvectors,  $q_i, i = 1, \dots, k$  as its columns.

The proof: For any  $x \in R^k$  let  $y = Q^T x = Q^{-1}x$  Then

$$\begin{aligned} Q(x) &= \text{tr}(x'Ax) = \text{tr}(x'Q'\Lambda Qx) = \text{tr}(\Lambda Qx(Qx)^T) \\ &= \text{tr}(\Lambda y'y) = \sum_{i=1}^k y_i^2 = \sum_{i=1}^k \lambda_i \|q_i^T x\|^2 \end{aligned}$$

This proof uses a common trick of expressing the scalar  $Q(x)$  as the trace of a  $1 \times 1$  matrix so we can reverse the order of matrix multiplications.

So in the general case of linear combination of non-central chi-square. We will need to take care of the  $\Sigma$ , assume  $Z \sim N(\mu, I)$

$$\begin{aligned} Z &= \Sigma^{-1/2}Y \\ Q(Y) &= Y^T A Y = (\Sigma^{1/2}Z)^T A (\Sigma^{1/2}Z) = Z^T (\Sigma^{1/2})' A (\Sigma^{1/2}) Z \\ &= \sum_{i=1}^k \lambda_i \|q_i^T Z\|^2 \\ W_i &= \|q_i^T Z\|^2 \\ &\sim \chi^2(d_i, \delta_i) \end{aligned}$$

$\Sigma^{1/2}$  is symmetric as well, so  $\Sigma^{1/2}A\Sigma^{1/2}$  is also symmetric. Symmetric matrix could be spectral decomposed.

I need to understand the product of orthogonal matrix and other matrix, orthogonal matrix is idempotent, which the eigenvalues are either 0 or 1.

(iv) The quadratic form question often occurs in the multivariate normal distribution.

Quadratic form theorem 1: If  $y \sim N(\mu_y, \Sigma_y)$ , then

$$z = AY = N(\mu_z = A\mu_y, \Sigma_z = A\Sigma_y A')$$

where  $A$  is a matrix of constants.

Proof:

$$\begin{aligned}
E[z] &= E[AY] = AE[Y] = A\mu_y \\
Var[z] &= E[(z - E[z])(z - E[z])'] = E[(AY - A\mu_y)(AY - A\mu_y)'] \\
&= E[A(Y - \mu_y)(Y - \mu_y)'A'] = AE[(Y - \mu_y)(Y - \mu_y)']A' \\
&= A\Sigma_y A'
\end{aligned}$$

Quadratic form theorem 2: quadratic form of  $n$  standard normal is chi-square with  $n$  degrees of freedom. Here we need to pay attention to the  $n$  degrees of freedom of chi-square distribution.

Quadratic form theorem 3: If  $y \sim N(0, \sigma^2 I)$  and  $M$  is a symmetric idempotent matrix of rank  $m$  then

$$\frac{y'My}{\sigma^2} = \chi^2(tr M)$$

Proof: Since  $M$  is symmetric it can be diagonalized with an orthogonal matrix  $Q$ .

$$Q'MQ = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \lambda_n \end{pmatrix}$$

Furthermore, since  $M$  is idempotent all these roots are either zero or 1. Thus we can choose  $Q$  so that  $\Lambda$  look like

$$Q'MQ = \Lambda = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

The dimension of the identity matrix will be equal to the rank of  $M$ , since the number of non-zero roots is the rank of the matrix. Since the sum of the roots is equal to the trace, the dimension is also equal to the trace of  $M$ . Now let  $v = Q'y$ , compute the moments of

$$\begin{aligned}
v &= Q'y \\
E[v] &= Q'E[y] = 0 \\
Var[v] &= Q'Var[y]Q = Q'\sigma^2 I Q = \sigma^2 Q'Q = \sigma^2 I, \quad Q \text{ is orthogonal} \\
v &\sim N(0, \sigma^2 I)
\end{aligned}$$

Now consider the distribution of  $y'My$  using the transformation  $v$ . Since  $Q$  is orthogonal, its inverse is equal to its transpose. This means that  $y = (Q')^{-1}v = Qv$ . Now write the quadratic form as follows

$$\begin{aligned}
\frac{y'My}{\sigma^2} &= \frac{v'Q'MQv}{\sigma^2} \\
&= \frac{1}{\sigma^2} v' \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} v \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{trM} v_i^2 \\
&= \sum_{i=1}^{trM} \frac{v_i^2}{\sigma^2}
\end{aligned}$$

This is the sum of squares of  $(tr M)$  standard normal variables and so is a  $\chi^2$  variable with  $trM$  degrees of freedom.

Proof

$$Y^T A Y = \sum_{i=1}^k \lambda_i W_i$$

where  $W_i$  are independently distributed as noncentral chi-squared variables with  $d_i$  degrees of freedom and noncentrality parameter  $\delta_i$ , that is,  $W_i \sim \chi_{d_i}^2(\delta_i), i = 1, 2, \dots, k$ . Indicate what  $\lambda_i, d_i, \delta_i$  are equal to.

$$\begin{aligned}
\Sigma &= \Sigma^{1/2} \Sigma^{1/2T}, \quad \text{by semi-definite matrix} \\
\Sigma^{1/2} &= (\Sigma^{1/2})^T \\
Y^T A Y &= Y^T \Sigma^{-1/2} \Sigma^{1/2} A \Sigma^{1/2} \Sigma^{-1/2} Y \\
Z &= \Sigma^{-1/2} Y \sim N(\Sigma^{-1/2} \mu, \Sigma^{-1/2} \Sigma \Sigma^{-1/2}) = N(\Sigma^{-1/2} \mu, I) \\
\Sigma^{1/2} A \Sigma^{1/2} &= P \Lambda P', \quad \text{semi-definition matrix, symmetric, spectral decomposition}
\end{aligned}$$

By now, we have constructed quadratic form into  $Z^T B Z$ , which we can recognize that we will use the spectral decomposition

$$\begin{aligned}
Z &= \Sigma^{-1/2} Y \\
B &= \Sigma^{1/2} A \Sigma^{1/2} = P \Lambda P', \quad P \text{ is orthogonal matrix} \\
Z^T B Z &= tr(Z^T B Z) = tr(Z^T P \Lambda P' Z) = tr(\Lambda [P' Z][P' Z]^T)
\end{aligned}$$

We can see that  $\Lambda$  is the matrix of eigenvalues of  $B = \Sigma^{1/2} A \Sigma^{1/2} = A \Sigma$  which we can prove later.

The trace of  $[P'Z][P'Z]^T$  is actually  $W_i = \|P'_i Z\|$ , and we can see that it is a non-centrality chi-square distribution with noncentrality parameter  $\delta_i$ . The degrees of freedom of chi-square distribution is the number of sum of chi-squares. We can see that, in multivariate normal distribution setting, there is only one variable  $Z$ , so the degrees of freedom for  $W_i$  is 1.

$$P'_i Z = N(P_i^T \Sigma^{-1/2} \mu, P_i^T I P_i) = N(P_i^T \Sigma^{-1/2} \mu, I), \quad P_i \text{ is orthogonal vector}$$

$$\|P'_i Z\|^2 \sim \chi^2(1, \delta_i), \quad \delta_i = \frac{[P_i^T \Sigma^{-1/2} \mu]^2}{2}$$

Need to be familiar with non-centrality chi-square, the non-central parameter is half of the mean square.

To prove eigenvalues of  $\lambda_1, \lambda_2, \dots, \lambda_n$  of  $\Sigma^{1/2} A \Sigma^{1/2}$  are also eigenvalues of  $A \Sigma$ .

$$\begin{aligned} & \left| \Sigma^{1/2} A \Sigma^{1/2} - \lambda I \right| = 0 \\ & \rightarrow \left| \Sigma^{-1/2} \Sigma^{1/2} A \Sigma^{1/2} \Sigma^{1/2} - \Sigma^{-1/2} \lambda I \Sigma^{1/2} \right| = 0 \\ & \rightarrow \left| A \Sigma - \lambda I \right| = 0 \end{aligned}$$

### 2.1.2 b. MGF of $Y^T A Y$

Use part (a) to derive the moment generating function of  $Y^T A Y$ . Let  $m(t)$  denote the moment generating function. Show that  $m(t)$  exists in a small neighborhood of  $t = 0$ , say,  $|t| < t_0$  for some positive constant  $t_0$ . Find the maximal value of  $t_0$  i.i.d chi-square distribution sum MGF.

#### Preparation

In part (a), we already derive that  $Y' A Y = \sum \lambda_i W_i$ , while  $W_i$  is a one degree of freedom of chi-square distribution with non-centrality parameter. So the MGF could be directly used for the sum of chi-square distribution.

I may need to remember the standard form of MGF of chi-square, or I need to derive that from scratch.

first, get the MGF for non-centrality chi-square, then use the transformation to put  $\lambda_i$ .

Another thing needs to pay attention is that,  $Y$  is multivariate normal, not univariate normal. The MGF derivation is different from univariate normal variable.

The MGF for univariate normal:

$$\begin{aligned}
M(t) &= \prod_{i=1}^k M_i(t) \\
p(x_i) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(X - \mu_i)^2}{2}\right) \\
M_i(t) &= E[\exp(x_i^2 t)] = \frac{1}{\sqrt{2\pi}} \int \exp\left[-\frac{(1-2t)x^2 - 2\mu x + \mu_i^2}{2}\right] dx \\
&= \frac{1}{\sqrt{2\pi}} \int \exp\left[-\frac{x^2 - 2\mu_i/(1-2t)x + \mu_i^2/(1-2t)^2 - \mu_i^2/(1-2t)^2 + \mu_i^2/(1-2t)}{2((1-2t)^{-1})}\right] dx \\
&= \exp\left[\frac{\mu_i^2/(1-2t)^2 - \mu_i^2/(1-2t)}{2((1-2t)^{-1})}\right] (1-2t)^{-1/2} \int \frac{1}{\sqrt{2\pi(1-2t)^{-1}}} \exp\left[-\frac{[x - \mu_i/(1-2t)]^2}{2((1-2t)^{-1})}\right] dx \\
&= (1-2t)^{-1/2} \exp\left[\frac{\mu_i^2 t}{(1-2t)}\right] \\
M(t) &= \prod_{i=1}^k (1-2t)^{-1/2} \frac{\mu_i^2 t}{(1-2t)} = (1-2t)^{-k/2} \exp\left[\frac{\sum_i^k \mu_i^2 t}{(1-2t)}\right]
\end{aligned}$$

In which,  $(1-2t) > 0$ ,  $t < 1/2$ . We can see that the product of non-centrality chi-square distributions is also a non-central chi-square distribution.

$$M_Z(t) = E[\exp(t^T Z)] = e^{\frac{1}{2}t^T t}$$

while for the transformation

$$\begin{aligned}
P_i' Z &= N(P_i^T \Sigma^{-1/2} \mu, I) \\
\|P_i' Z\|^2 &\sim \chi^2(1, \delta_i), \quad \delta_i = \frac{[P_i^T \Sigma^{-1/2} \mu]^2}{2}
\end{aligned}$$

Then the MGF

$$\begin{aligned}
X_i &= Z + P_i^T \Sigma^{-1/2} \mu \\
M_{X_i}(t) &= E[\exp(t^T X_i)] = \exp(t^T [P_i^T \Sigma^{-1/2} \mu]) e^{\frac{1}{2}t^T t}
\end{aligned}$$

**Proof**

However, in part (a), we broke down the vector into sum of each univariate chi-square distribution. In that sense, it is still a univariate chi-square MGF question.

Another method is to let  $Z \sim N(0, 1)$ , then  $(Z + \mu)^2$  has a noncentral chi-square distribution with one degree of freedom, the MGF of  $(Z + \mu)^2$

$$\begin{aligned} E[\exp(t(Z + \mu)^2)] &= \frac{1}{\sqrt{2\pi}} \int \exp(t(Z + \mu)^2) \exp(-\frac{Z^2}{2}) dZ \\ &= \frac{1}{\sqrt{2\pi}} \int \exp[-\frac{(1-2t)Z^2 - 2\mu Z + \mu^2}{2}] dZ \\ &= (1-2t)^{-1/2} \exp[\frac{\mu^2 t}{(1-2t)}] \\ M_{W_i}(t) &= (1-2t)^{-1/2} \exp[\frac{2\delta_i t}{(1-2t)}], \quad \delta_i = \frac{1}{2}\mu_i^2 \end{aligned}$$

We also note that, chi-square with 1 degree of freedom is a gamma distribution  $\chi^2(1, 0) = \text{Gamma}(1/2, 2)$ .

As  $Y^T A Y = \sum_{i=1}^n \lambda_i W_i$ , then

$$\begin{aligned} M(t) &= E\left[\exp\left(\sum_{i=1}^n \lambda_i W_i\right)t\right] = \prod_{i=1}^k M_i(t) \\ &= \prod_i^k (1-2\lambda_i t)^{-1/2} \exp[\frac{2\delta_i \lambda_i t}{(1-2\lambda_i t)}] \end{aligned}$$

As we see that  $1 - 2\lambda_i t > 0$ , then we have  $t < \frac{1}{2\lambda_i}$ .

By definition, a non-central chi-square random variable  $\chi_{n,\lambda}^2$  with  $n$  df and parameters  $\lambda = \sum_i^n \mu_i^2$  is the sum of  $n$  **independent** normal variables  $X_i = Z_i + \mu_i, i = 1, 2, \dots, n$ .

$$\begin{aligned} \chi_{n,\lambda}^2 &= \sum_i^n X_i^2 = \sum_i^n (Z_i + \mu_i)^2 \\ M(t) &= \prod_{i=1}^k M_i(t) = \prod_i^k (1-2t)^{-1/2} \exp[\frac{\mu_i^2 t}{(1-2t)}] \\ &= (1-2t)^{-k/2} \exp[\frac{\sum_i^n \mu_i^2 t}{(1-2t)}] = (1-2t)^{-n/2} \exp[\frac{\lambda t}{(1-2t)}] \end{aligned}$$

### 2.1.3 c. Trace and Rank

Use part (a) to show that  $\text{tr}[(A\Sigma)^2] = \text{tr}(A\Sigma) = r$ , where  $r$  is the rank of  $A$ , then  $Y^T A Y$  has a chi-squared distribution. Determine its degrees of freedom and noncentrality parameter.

#### Preparation

From part (a), we see that  $Y^T A Y$  is sum of chi-square distributions. In order to show it is a chi-square, then the  $\lambda_i = 1, 0$ . Well,  $\text{tr}[(A\Sigma)^2] = \text{tr}(A\Sigma) = r$  is similar to the idempotent matrix definition, and  $\lambda_i$  is actually the eigenvalue of  $A\Sigma$ .



But how can we show the idempotency of  $A\Sigma$ ? Rank is the sum of eigenvalues, the link is how to go from the rank to show that the eigenvalues are exactly 0 or 1.

Because  $\Sigma$  is full rank, so  $R(A\Sigma) = R(A)$  as the column space won't change when it is the right product of a full rank matrix. So the number of non-zero eigenvalues is  $r$ .

$$\sum_{i=1}^k \lambda_i = \sum_{i=1}^k I(\lambda_i \neq 0)$$

$$I(\lambda_i \neq 0) = \begin{cases} 1, & \lambda_i \neq 0 \\ 0, & \lambda_i = 0 \end{cases}$$

So we can see that  $\lambda_i = I(\lambda_i \neq 0) = 1$  when  $\lambda_i \neq 0$ .

Use the spectral decomposition of  $A\Sigma = P\Lambda P'$ .

$$\begin{aligned} A\Sigma &= P\Lambda P' \\ (A\Sigma)^2 &= P\Lambda^2 P' \\ \text{tr}((A\Sigma)^2) &= \text{tr}(A\Sigma) \\ &\rightarrow \text{tr}(P\Lambda P') = \text{tr}(P\Lambda^2 P') \\ &\rightarrow \text{tr}(\Lambda^2) = \text{tr}(\Lambda), \\ &\rightarrow \sum_{i=1}^k \lambda_i^2 = \sum_{i=1}^k \lambda_i \\ &\rightarrow \sum_{i=1}^k \lambda_i(\lambda_i - 1) = 0 \\ &\rightarrow \lambda_i = 0, 1 \end{aligned}$$

**Proof**

From part (a) that

$$\begin{aligned} \text{tr}(A\Sigma) &= \text{tr}(\Sigma^{1/2} A \Sigma^{1/2}) \\ \text{tr}([\Sigma^{1/2} A \Sigma^{1/2}]^2) &= \text{tr}([\Sigma^{1/2} A \Sigma^{1/2}][\Sigma^{1/2} A \Sigma^{1/2}]) \\ &= \text{tr}([\Sigma^{1/2} A \Sigma A \Sigma^{1/2}]) \\ &= \text{tr}([A \Sigma A \Sigma^{1/2} \Sigma^{1/2}]) \\ &= \text{tr}([A \Sigma A \Sigma]) \\ &= \text{tr}([(A\Sigma)^2]) \end{aligned}$$

If  $\lambda_i = 0, 1$ , then  $Y^T A Y$  is the sum of  $r$  chi-square  $\chi^2(1, \delta_i)$ , which is also a chi-square distribution.

$$Y^T AY = \sum_{i=1}^r W_i \sim \chi^2(r, \delta), \quad \delta = 2 \sum_{i=1}^r \mu_i^2$$

#### 2.1.4 d. Distribution

Show that  $Y^T AY$  has a noncentral chi-squared distribution if and only if  $A\Sigma$  is idempotent.

##### Preparation

This problem needs to prove "if and only if", that From part(c), we have shown that if  $\lambda_i = 0, 1$  which is idempotent will lead to  $Y^T AY$  as a noncentral chi-squared distribution.

While if  $Y^T AY$  is a noncentral chi-squared distribution, then  $\lambda_i = 0, 1$  then it is a idempotent.

Idempotent is  $A^2 = A, A^T = A$ . So if  $A\Sigma = (A\Sigma)^2$ ,  $A\Sigma$  is idempotent.

we have the MGF of linear combination of non-central chi-square distribution Y

$$\begin{aligned} M(t) &= \prod_{i=1}^n M_{W_i}(\lambda_i t) \\ &= \prod_{i=1}^n (1 - 2\lambda_i t)^{-1/2} \exp\left(\frac{2\delta_i \lambda_i t}{1 - 2\lambda_i t}\right) \end{aligned}$$

Then we can see that the shape parameter is  $\frac{1}{2\lambda_i}$ . If we want to have a non-central chi-square distribution for Y, then all  $\lambda_i$  need to be the same, 1. Otherwise it won't be a chi-square.

##### Proof

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^n (1 - 2\lambda_i t)^{-1/2} \exp\left(\frac{2\delta_i \lambda_i t}{1 - 2\lambda_i t}\right) \\ &= (1 - 2t)^{-n/2} \exp\left(\frac{\sum_{i=1}^n 2\delta_i t}{1 - 2t}\right) \end{aligned}$$

And for chi-square distribution, the shape parameter has to be 1/2, so the  $w_i = 1$ . So we prove that if Y is a non-central chi-square distribution, A has to have the eigenvalues either 1 or 0.

When the eigenvalues are only 0, 1, then  $A\Sigma$  is a projection matrix, then it is idempotent.

## 2.2 Problem 2 - Linear Model

Consider the linear model  $Y = X\beta + \epsilon$ , where  $\epsilon \sim N_n(0, \sigma^2 I)$ ,  $X$  is  $n \times p$  of rank  $p$ ,  $\beta$  is  $p \times 1$ , and  $(\beta, \sigma^2)$  are unknown. Define  $H = X(X^T X)^{-1} X^T$  and let  $h_{ii}$  denote the  $i$ th diagonal element of  $H$ . Further let  $\hat{\sigma}^2$  denote the usual unbiased estimator of  $\sigma^2$  for the linear regression model and let  $\hat{\epsilon}_i$  denote the ordinary residual. Let

$$A_i = \frac{\hat{\epsilon}_i^2}{\sigma^2(1 - h_{ii})}$$

$$B_i = \frac{(n - p)\hat{\sigma}^2}{\sigma^2} - \frac{\hat{\epsilon}_i^2}{\sigma^2(1 - h_{ii})}$$

2.2.1 a. Show that  $B \sim \chi_{n-p-1}^2$

Preparation

In order to prove  $B \sim \chi_{n-p-1}^2$ , we need to show that it is a sum of  $n-p-1$  normal distribution square (by definition of chi-square distribution).

We can write  $\sigma^2, \hat{\sigma}^2$  the quadratic form in the sum of squares of normal distribution. But the fraction between two quadratic forms is what distribution?

Note that the division is not F-distribution, as it didn't divided by degrees of freedom.

$\hat{\sigma}^2$  is the usual unbiased estimator of  $\sigma^2$ , we need to know that if unbiased, then the degrees of freedom is  $n-p$ .

Another thing need to pay attention that,  $\sigma^2$  is not a distribution, it is fixed and it is the variance of  $Y_i$  just unknown.

Variance of  $Y_i$  is also the variance of  $\epsilon$ , but we need to know the difference and how to use them.

$$\begin{aligned} \frac{(n - p)\hat{\sigma}^2}{\sigma^2} &= \frac{Y^T(I - H)Y}{\sigma^2} = \frac{[(I - H)Y]^T[(I - H)Y]}{\sigma^2} \\ &= \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\text{Var}(y_i)} \end{aligned}$$

So is  $n-p$  sum of normal distribution square. Then try to link the second term

$$\frac{\hat{\epsilon}_i^2}{\sigma^2(1 - h_{ii})} = \frac{\hat{\epsilon}_i^2}{V(e_i)}$$

This question requires understanding the matrix  $H$ , or the relationship between o.p.o matrix and scalar form. And the characteristic of  $H$  matrix itself  $H^2 = H, H^T = H$ .

If we write that in scalar form,

$$(h_{i1}, h_{i2} \dots h_{in}) \begin{pmatrix} h_{i1} \\ h_{i2} \\ \dots \\ h_{in} \end{pmatrix} = h_{i1}^2 + h_{i2}^2 + \dots + h_{in}^2 = h_{ii}$$

Variance, Covariance of  $e$

$$\begin{aligned} V(e) &= E[(e - E(e_i))(e - E(e_i))^T] = (I - H)E(\epsilon\epsilon^T)(I - H)^T \\ &= (I - H)I\sigma^2(I - H)^T \\ &= (I - H)(I - H)^T I\sigma^2 = (I - H)\sigma^2 \end{aligned}$$

Need to know that  $V(e_i)$  is given by the  $i$ th diagonal element  $1 - h_{ii}$  and  $Cov(e_i, e_j)$  is given by the  $(i, j)$ th element of  $-h_{ij}$  of the matrix  $(I - H)\sigma^2$ .

We also need to know that both  $Y_i$  and  $\epsilon_i$  follows the same distribution

$$\begin{aligned} e - E(e) &= (I - H)Y = (Y - X\beta) = (I - H)\epsilon \\ E(\epsilon\epsilon^T) &= V(\epsilon) = I\sigma^2, \quad E(\epsilon) = 0 \end{aligned}$$

We also have correlation

$$\rho_{ij} = \frac{Cov(e_i, e_j)}{\sqrt{V(e_i)V(e_j)}} = -\frac{h_{ij}}{(1 - h_{ii})(1 - h_{jj})}$$

$$SS(b) = SS(parameter) = b^T X^T Y = \hat{Y}^T Y = Y^T H^T Y = Y^T H Y = Y^T H^2 Y = \hat{Y}^T \hat{Y}$$

The average  $V(\hat{Y}_i)$  to all data points is

$$\begin{aligned} \sum_{i=1}^n \frac{V(\hat{Y}_i)}{n} &= \frac{trace(H\sigma^2)}{n} = \frac{p\sigma^2}{n} \\ \hat{Y}_i &= h_{ii}Y_i + \sum_{j \neq i} h_{ij}Y_j \end{aligned}$$

Studentized residual

$$V(e_i) = (1 - h_{ii})\sigma^2$$

where  $\sigma^2$  is estimated by  $s^2$

$$s^2 = \frac{e^T e}{n - p} = \frac{\sum_{i=1}^n e_i^2}{n - p}$$

Sum of Squares attributable to  $e_i$

$$\begin{aligned}
e &= (I - H)Y \\
e_i &= -h_{i1}Y_1 - h_{i2}Y_2 - \dots + (1 - h_{ii})Y_i - \dots - h_{in}Y_n = c^T Y \\
\text{Var}(e_i) &= c^T \text{Var}(Y) c = \sigma^2 c^T c \\
c^T &= (-h_{i1}, -h_{i2}, \dots, (1 - h_{ii}), \dots, -h_{in}), \quad (I - H)I - H = I - H \\
c^T c &= \sum_{j=1}^n h_{ij}^2 + (1 - 2h_{ii}) = (1 - h_{ii}) \\
\text{Var}(e_i) &= \sigma^2(1 - h_{ii}) \\
SS(e_i) &= e_i^2 = \left( (I - H)y \right)^T (I - H)y = y^T (I - H)y
\end{aligned}$$

Need to learn to write in matrix form.

Proof

Write  $B_i$  in matrix form

$$\begin{aligned}
\frac{(n-p)\hat{\sigma}^2}{\sigma^2} &= \frac{Y^T(I-H)Y}{\sigma^2} = \frac{[(I-H)Y]^T[(I-H)Y]}{\sigma^2} \\
\frac{\hat{e}_i^2}{\sigma^2(1-h_{ii})} &= \frac{[e_i^T(I-H)Y]^T[e_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2}
\end{aligned}$$

where  $e_i = (0, \dots, 0, 1, 0, \dots, 0)$  Then

$$\frac{[e_i^T(I-H)Y]^T[e_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2} = \frac{[Y^T(I-H)e_ie_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2}$$

Then

$$\begin{aligned}
B_i &= \frac{[Y^T(I-H)Y]}{\sigma^2} - \frac{[Y^T(I-H)e_ie_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2} \\
&= \frac{Y^T}{\sigma} \left[ (I-H) - (I-H) \frac{e_ie_i^T}{1-h_{ii}} (I-H) \right] \frac{Y}{\sigma}
\end{aligned}$$

$\frac{Y}{\sigma} \sim N_n(0, I)$ .

To show that  $C = \left[ (I-H) - (I-H) \frac{e_ie_i^T}{1-h_{ii}} (I-H) \right]$  is a projection matrix, we need to show  $C^2 = C, C^T = C$ .

Here  $(I-H)(I-H) = I-H, e_i^T(I-H)e_i = 1 - h_{ii}$

Therefore,  $B_i \sim \chi^2(r)$ , where  $r = \text{rank}[C]$ . For projection matrix, the rank is also the trace.

$$\begin{aligned}
\text{rank}\left[(I - H) - (I - H)\frac{e_i e_i^T}{1 - h_{ii}}(I - H)\right] &= \text{tr}\left[(I - H) - (I - H)\frac{e_i e_i^T}{1 - h_{ii}}(I - H)\right] \\
&= n - p - \frac{1}{1 - h_{ii}}\text{tr}[(I - H)e_i e_i^T(I - H)] \\
&= n - p - \frac{1}{1 - h_{ii}}\text{tr}[e_i^T(I - H)(I - H)e_i] \\
&= n - p - \frac{1}{1 - h_{ii}}\text{tr}[e_i^T(I - H)e_i] \\
&= n - p - \frac{1 - h_{ii}}{1 - h_{ii}} \\
&= n - p - 1
\end{aligned}$$

So  $B_i \sim \chi^2(n - p - 1)$ .

## 2.2.2 b. Show $A_i$ and $B_i$ are independent

### Preparation

Note that  $A_i$  and  $B_i$  are quadratic forms, which we need to show that their projection matrix product is 0, which will need to use linear algebra.

Another method is that, there are chi-square distribution, which we just need to show that the two statistics are ancillary.

Traditional method is to derive the distributions, and if we could write the distribution of  $A_i + B_i$  as the product of distribution of  $A_i$  and  $B_i$ , then we can prove the independence.

One more question, are we able to use the moment generating function to do this? It is always easier to prove in MGF.

First, we need to write the MGF of  $A_i$ ,  $B_i$

$$\begin{aligned}
M_{A_i} &= E[\exp[A_i t]] = \int \exp[A_i t] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
M_{B_i} &= E[\exp[B_i t]] = \int \exp[B_i t] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx
\end{aligned}$$

Then we can use MGF properties to prove

$$\begin{aligned}
W_i &= A_i + B_i \\
M_{W_i} &= E[\exp(W_i t)] = E[\exp[(A_i + B_i)t]] \\
&= E[\exp[A_i t] \exp[B_i t]], \quad \text{all based on normal distribution} \\
&= E[\exp[A_i t]] E[\exp[B_i t]] = M_{A_i} M_{B_i}
\end{aligned}$$

Also we need to notice that, this is a quadratic form chi-square distribution.

(a) Quadratic form theorem 1:

Let the  $n \times 1$  vector  $y \sim N(0, I)$ , let  $A$  be an  $n \times n$  idempotent matrix of rank  $m$ , let  $B$  be an  $n \times n$  idempotent matrix of rank  $s$ , and suppose  $BA = 0$ . Then  $y'Ay$  and  $y'By$  are independently distributed  $\chi^2$  variables.

**Why need to be idempotent matrix?** It is because for idempotent matrix  $A^2 = A$ , then the quadratic form could be written as  $y'Ay = (Ay)'(Ay)$ , then further we can use the product of two vectors = 0 to show independence.

The proof is try write the two quadratic form in partition, so that show that they dependent on different vectors, further they are independent. Need to compare with the proof in 762 slides.

(b) Quadratic form theorem 2 (Craig's Theorem): If  $y \sim N(\mu, \Omega)$  where  $\Omega$  is positive definite, then  $q_1 = y'Ay$  and  $q_2 = y'By$  are independently distributed if  $A\Omega B = 0$

This is just a generalization of above theorem. Since  $\Omega$  is a covariance matrix of full rank it is positive definite and can be factored as  $\Omega = TT'$ . Therefore the condition  $A\Omega B = 0$  can be written  $ATT'B = 0$ .

$$T'ATT'BT = 0$$

$$C = T'AT, \quad K = T'BT$$

$$CK = (T'AT)(T'BT) = T'\Omega BT = T'0T = 0$$

In the 762 slides, it used the covariance matrix = 0 to prove independence in MVN.

$$Y'AY = Y'RR'Y = (R'Y)'(R'Y)$$

$$Y'BY = Y'SS'Y = (S'Y)'(S'Y)$$

Thus  $Y'AY$  and  $Y'BY$  are independent if  $R'Y$  and  $S'Y$  are independent.  $R'Y$  and  $S'Y$  are independent if and only if

$$\begin{aligned} Cov(R'Y, S'Y) &= 0 \\ &= R'\Sigma S = 0 \\ &= R'WW'S = 0 \\ &= C(Q'S) \perp C(Q'R) \end{aligned}$$

Proof

$$\begin{aligned} A_i &= \frac{[Y^T(I-H)e_ie_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2} \\ B_i &= \frac{[Y^T[(I-H) - (I-H)\frac{e_ie_i^T}{(1-h_{ii})}(I-H)]Y]}{\sigma^2} \end{aligned}$$

We would like to show the product of two projection matrix are 0, by using the above theorem

$$\begin{aligned} M_{A_i} M_{B_i} &= (I - H) \frac{e_i e_i^T}{1 - h_{ii}} (I - H) \left[ (I - H) - (I - H) \frac{e_i e_i^T}{(1 - h_{ii})} (I - H) \right] \\ &= 0 \\ e_i^T (I - H) e_i &= 1 - h_{ii} \end{aligned}$$

### 2.2.3 c. Derive exact distribution

Let

$$r_i = \frac{\hat{\epsilon}_i}{\hat{\sigma} \sqrt{1 - h_{ii}}}$$

Using part(a) and (b), derive the exact distribution of  $r_i^2/(n - p)$ .

#### Preparation

This problem needs observation of the  $r_i$  and connection between  $A_i$  and  $B_i$ . Also use the results from part (a) and (b) that,  $A_i$  and  $B_i$  are independent chi-square.

$$\frac{r_i^2}{(n - p)} = \frac{A_i}{A_i + B_i}$$

The transformation of distribution, first we can use the distribution theory that the sum of independent chi-square distribution is also a chi-square. The first method we need to try is to use MGF, then try to use the distribution transformation.

if we use distribution transformation, then we need to use two independent distributions

$$\begin{aligned} w &= \frac{A_i}{A_i + B_i} \\ B_i &= \frac{1 - w}{w} A_i \end{aligned}$$

replace  $B_i$  by w, then integrate out of  $A_i$ , then we will get the distribution of  $w$ .

the chi-square distribution  $X \sim \chi^2(k)$  is a special case of the gamma distribution, in that  $X \sim \Gamma(\frac{k}{2}, \frac{1}{2})$  using the rate parameterization of the gamma distribution.

#### Proof

We need to know the relationship between Gamma distribution and Beta distribution. As  $A_i \perp B_i$

$$\frac{A_i}{A_i + B_i} \sim \text{Beta}\left(\frac{1}{2}, \frac{n - p - 1}{2}\right)$$



Show the distribution transformation. Assume that  $X \sim \text{Gamma}(\alpha, 1)$ ,  $Y \sim \text{Gamma}(\beta, 1)$ , and  $X$  and  $Y$  are independent. then find the distribution of  $\frac{X}{X+Y}$ .

$$f_{X,Y}(x, y) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-(x+y)}$$

let  $w = \frac{x}{x+y}$ ,  $u = x$ , then we have

$$\begin{aligned} x &= y, & y &= \frac{(1-w)u}{w} \\ J &= \left| \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial w} \end{pmatrix} \right| \\ &= \left| \begin{pmatrix} 1 & 0 \\ \frac{(1-w)}{w} & -\frac{u}{w^2} \end{pmatrix} \right| = \frac{u}{w^2} \end{aligned}$$

Then integrate out  $u$

$$\begin{aligned} f(w) &= \int_0^\infty f(w, u) du \\ &= \int_0^\infty \frac{1}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} \left(\frac{(1-w)u}{w}\right)^{\beta-1} e^{-\frac{u}{w}} \frac{u}{w^2} du \\ &\sim \text{Beta}(\alpha, \beta) \end{aligned}$$

#### 2.2.4 d. Outlier Model 1

Suppose we suspect that the  $i$ th case is an outlier and we consider the mean shift outlier model  $Y = X\beta + d_i\phi + \epsilon$ , where  $\epsilon \sim N_n(0, \sigma^2 I)$ ,  $\phi$  is an unknown scalar, and  $d_i$  is an  $n \times 1$  vector with a 1 in the  $i$ th position and zeros elsewhere. Derive the maximum likelihood estimate of  $\phi$ .

##### Preparation

We can construct the model by introducing the unknown parameter  $\phi$ , then use the MLE to estimate.

The way to get MLE is to construct likelihood function first,  $Y$  is a normal distribution, and only  $i$ th case is an outlier, while all other cases are not. Solve the problem similar as the random model.

As the  $Y$  are independent variables,  $Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n \sim N(X_k\beta, \sigma^2)$ , while  $Y_i \sim N(X_i\beta + \phi, \sigma^2)$ .

If using scalar form, the  $\beta, \phi$  are correlated with each other, which we won't be able

to solve.

$$\begin{aligned}
f_k(\beta, \phi) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_k - x_k\beta)^2}{2\sigma^2}\right) \\
f_i(\beta, \phi) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - x_i\beta - \phi)^2}{2\sigma^2}\right) \\
\log f_k(\beta, \phi) &= -\frac{(y_k - x_k\beta)^2}{2\sigma^2} \\
\log f_i(\beta, \phi) &= -\frac{(y_i - x_i\beta - \phi)^2}{2\sigma^2} \\
\ln(\beta, \phi) &= \sum_{k=1, k \neq i}^n -\frac{(y_k - x_k\beta)^2}{2\sigma^2} - \frac{(y_i - x_i\beta - \phi)^2}{2\sigma^2}
\end{aligned}$$

The MLE of  $\phi$

$$\begin{aligned}
\frac{\partial \ln}{\partial \phi} &= \frac{(y_i - x_i\beta - \phi)}{\sigma^2} = 0 \\
\phi &= y_i - x_i\hat{\beta}
\end{aligned}$$

where  $\hat{\beta}$  is the MLE of  $\beta$ .

So we need to use the linear algebra form in solve this kind of question. Need to notice that the linear regression estimate in matrix form all use the linear algebra.

**Proof**

The maximum likelihood function in linear algebra form

$$\begin{aligned}
L(\beta, \phi) &\propto \exp\left(-\frac{(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi)}{2\sigma^2}\right) \\
l(\beta, \phi) &\propto \left(-\frac{(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi)}{2\sigma^2}\right) \\
&= -\frac{1}{2\sigma^2} \left[ (Y - X\beta)^T(Y - X\beta) - 2(Y - X\beta)^T d_i\phi + d_i^T d_i\phi^2 \right]
\end{aligned}$$

The MLE

$$\begin{aligned}
\frac{\partial l}{\partial \phi} &= \left(-\frac{(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi)}{2\sigma^2}\right) \\
&\rightarrow -2(Y - X\beta)^T d_i + d_i^T d_i(2\phi) = 0 \\
\phi &= (Y - X\beta)^T d_i
\end{aligned}$$

MLE of  $\beta$

$$X\beta = HY$$
$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

There is no need to get the estimate of  $\beta$ , we only need to estimate  $(Y - X\beta)^T$ . So  $\hat{\phi} = (I - H)Y^T d_i$

#### 2.2.5 d. Outlier Model 2

Suppose we wish to test  $H_0 : \phi = 0$ . Derive the test statistic for this hypothesis and derive its exact distribution under  $H_0$ .

Preparation

The hypothesis test always comes with the estimate. There are several hypothesis test method we can do, wald test, score test and likelihood ratio test. The score test is generally the way as it is easy to get the estimate, score function and fisher information under  $H_0$ .

We can use  $A_i$  and  $B_i$  are independent and they are quadratic form, so we can have F -test and try to find the connection between the  $\hat{\phi}$  and F-test.

Proof

To test  $\phi = 0$ , consider wald test

$$W_n = (\hat{\phi} - 0)^T I_n(\phi) (\hat{\phi} - 0)$$

Find the fisher information  $I_n(\phi)$

$$I_n(\phi) = nI(\phi)$$
$$I(\phi) = -E\left[\frac{\partial^2 l}{\partial \phi \partial \phi}\right]$$
$$= \frac{1}{\sigma^2}$$
$$I_n(\phi) = \frac{n}{\sigma^2}$$

We need to find the MLE of  $\sigma^2$ .

$$\frac{\partial^2 l}{\partial \sigma^2} = \frac{1}{2\sigma^4} (Y - X\beta - d_i \phi)^T (Y - X\beta - d_i \phi) + \frac{\partial}{\partial \sigma^2} \log[(\sigma^2)^{-1/2}]$$
$$\hat{\sigma}^2 = \|Y - X\beta - d_i \phi\|^2$$
$$= \|(I - H)Y - d_i \hat{\phi}\|^2$$

So the wald statistics

$$W_n = d_i^T Y(I - H) \left( \frac{n}{\|(I - H)Y - d_i \hat{\phi}\|} \right) (I - H)Y^T d_i$$

Under  $H_0$ , it is a  $\chi^2(1)$ .

The second method is to use F-test, as  $A_i$  and  $B_i$  are independent quadratic form

$$A_i = \frac{[Y^T(I - H)e_i e_i^T(I - H)Y]/(1 - h_{ii})}{\sigma^2} = \frac{[\hat{\phi}^2]/(1 - h_{ii})}{\sigma^2}$$

$$B_i = \frac{[(I - H)Y]^T[(I - H)Y] - [\hat{\phi}^2]/(1 - h_{ii})}{\sigma^2}$$

$$F = \frac{A_i}{B_i/(n - p - 1)}$$

$$= \frac{\hat{\phi}^2/(1 - h_{ii})}{[(I - H)Y]^T[(I - H)Y] - [\hat{\phi}^2]/(1 - h_{ii})/(n - p - 1)}$$

We can perform F-test by setting up the rejection region.

### 2.2.6 e. Cook's Distance

Let  $I = \{1, ..m\}$  be the subset of the first m cases in the dataset. Let  $D_I$  denote the Cook's distance based on simultaneously deleting m cases from the dataset, which is given by

$$D_I = \frac{(\hat{\beta} - \hat{\beta}_I)^T (X^T X)(\hat{\beta} - \hat{\beta}_I)}{p \hat{\sigma}^2}$$

where  $\hat{\beta}^I$  denotes the least squares estimate of  $\beta$  with the cases deleted from set I and  $\hat{\beta}$  denotes the estimate of  $\beta$  based on the full data. Show that  $D_I$  can be written as

$$D_I = \frac{1}{p} \sum_{i=1}^m h_i^2 \left( \frac{\lambda_i}{1 - \lambda_i} \right)$$

where the  $\lambda_i, i = 1, ..m$ , are the eigenvalues of the matrix  $P_I = X_I(X^T X)^{-1}X_I^T$  based on a spectral decomposition of  $P_I$ .

## Preparation

The single deletion case

$$\hat{\beta}_{(1)} = \hat{\beta} - \frac{(X'X)^{-1}x_i^T\hat{\epsilon}_i}{1 - h_{ii}}$$

This method is used in Bayesian method, get marginal distribution from joint distribution, expectation of conditional distribution.

The probability involved in retrospective study sampling method

The likelihood function for the retrospective study is the conditional density of  $X = x$  given  $Y$  and  $S = 1$ , denoted by  $p(x|Y, S = 1)$ . To avoid biased sampling, we assume that the inclusion of a subject is independent of  $x$  given  $Y$ , that is,

$$P(S = 1|X, Y) = P(S = 1|Y)$$

Under unbiased sampling, we have

$$p(x|Y, S = 1) = \frac{P(S = 1|x, Y)p(x|Y)}{P(S = 1|Y)} = p(x|Y)$$

Furthermore,  $p(x|Y, S = 1)$  can be expressed as

$$p(x|Y, S = 1) = \frac{P(Y|S = 1, x)p(x|Y)P(x|S = 1)}{P(Y|S = 1)} = p(x|Y)$$

and  $p(Y|x, S = 1)$  is given by

$$\begin{aligned} p(Y|x, S = 1) &= \frac{p(X|Y, S = 1)p(Y, S = 1)}{p(X, S = 1)} = p(Y|x) \frac{p(S = 1|Y)}{p(Y|S = 1)} \\ &= \frac{p(X, Y, S = 1)}{p(X, S = 1)} = \frac{p(Y, S = 1|x)p(x)}{p(S = 1|x)p(x)} \\ &= \frac{p(S = 1|x, Y)p(Y)}{p(S = 1|x)} = \frac{p(S = 1|Y)p(Y|x)}{p(S = 1|x)} \end{aligned}$$

Finally,  $p(x|Y, S = 1)$  can be expressed as

$$p(x|Y, S = 1) = \frac{P(Y|x)p(x|S = 1)P(S = 1|Y)}{P(Y|S = 1)p(S = 1|x)}$$

Why do we write in this form? As it is how we select the patients in the case-control study, and it reveals the relationship between prospective and retrospective probability.

The ratio  $p(x|S = 1)/p(S = 1|x)$  only depends on  $x$  and  $P(S = 1|Y)$  and  $P(Y|S = 1)$  are determined by the sampling scheme.

We can construct the likelihood function for the retrospective study

Let  $H_1(Y) = P(S = 1|Y)/P(Y|S = 1)$  and  $H_2(x) = p(x|S = 1)/p(S = 1|x)$

$$\prod_{i=1}^n p(x_i|Y_i, S_i = 1) = \prod_{i=1}^n P(y_i|x_i, \beta) \prod_{i=1}^n [H_1(y_i)H_2(x_i)]$$

Proof

## 2.3 Problem 3 - Expectation in case control study

### 2.3.1 a. Expectation

Consider the conditional mean of  $Y$  given  $(X, D)$ , denoted by  $\tilde{\mu}(X, D) = E[Y|X, D]$ . Show that

$$\begin{aligned}\mu(X) &= \tilde{\mu}(X, 1)Pr(D = 1|X) + \tilde{\mu}(X, 0)Pr(D = 0|X) \\ \tilde{\mu}(X, D) &= E[Y|X, D, S = 1] = \mu(X) + \{D - Pr(D = 1|X)\}\gamma(X)\end{aligned}$$

Preparation

Find the expectations in the case-control study from 762 slides.

Proof

The expectation of conditional distribution

$$\begin{aligned}E(Y|X) &= E[E[Y|X, D]|X] \\ &= E[Y|X, D = 1]Pr(D = 1|X) + E[Y|X, D = 0]Pr(D = 0|X) \\ &= \tilde{\mu}(X, 1)Pr(D = 1|X) + \tilde{\mu}(X, 0)Pr(D = 0|X)\end{aligned}$$

In terms of  $\tilde{\mu}(X, D) = E[Y|X, D, S = 1]$ , here we know that all the patients in the case-control study with  $S = 1$ , so  $\tilde{\mu}(X, D) = E[Y|X, D, S = 1] = E[Y|X, D]$ .

$$E(Y|X) = \tilde{\mu}(X, 1)Pr(D = 1|X) + \tilde{\mu}(X, 0)Pr(D = 0|X)$$

## Chapter 3

2016

### 3.1 Problem 1- Distribution

The first problem always is distribution, especially conditional distribution under different parameters. Then we need to get the expectation, variance, UMVUE, etc.

Let  $X_1, X_2, \dots, X_n$  be an i.i.d. sample from the density

$$f(X) = \begin{cases} 0 & \text{with probability } p \\ U[0, \theta] & \text{with probability } 1-p \end{cases}$$

First, assume  $p$  is known constant in  $(0,1)$ , and that  $\theta > 0$  is the only parameter of interest.

#### 3.1.1 a. Compute Expectation, Variance, UMVUE

Based on only one observation  $X_1$ , find all the unbiased estimator for  $\theta$  and calculate the variances. Does the UMVUE exists for  $\theta$ ?

BE AWARE: 1. only one observation  $X_1$ , but it does not mean the unbiased estimator is  $X_1$ . It could be the function of  $X_1$ , so we denote  $T(X_1)$  as the assumed unbiased estimator.

2. The conditional distribution of  $f(X)$ ,  $f(X = 0|Z = 1) = p$ ,  $f(X = U(0, \theta)|Z = 0) = 1-p$ . When there is situation that two distributions involve, we need to consider conditional distribution.

3.  $E[T(X_1)] = E[E[T(X_1)|Z]] = E[T(0)|Z = 1]p(Z = 1) + E[T(X_1)|Z = 0]p(Z = 0)$

4.  $Var[T(X_1)] = E[Var[T(X_1)|Z]] + Var[E(T(X_1)|Z)] = E[T(0)|Z = 1]p(Z = 1) + E[T(X_1)|Z = 0]p(Z = 0)$

$$\begin{aligned}
E[(X_1 - \mu)^{-r}] &= \int_{\mu}^{\mu+1} (x - \mu)^{-r} \alpha (x - \mu)^{\alpha-1} dx \\
&= \int_{\mu}^{\mu+1} \alpha (x - \mu)^{\alpha-r-1} dx \\
&= \frac{\alpha}{\alpha - r} (x - \mu)^{\alpha-r} \Big|_{\mu}^{\mu+1} \\
&= \frac{\alpha}{\alpha - r}
\end{aligned}$$

In order that  $E[(X_1 - \mu)^{-r}] > 0$ , then  $r < \alpha$ . If  $\alpha - r < 0$ , then the integral will be  $\infty$ , then the expectation is not bounded.

The integral of exponential variable is bounded only when index  $> 0$  and base within  $(0,1)$ .

### 3.1.2 b. MLE

Assume  $\mu$  is known, show MLE of  $\alpha$ .

$$\begin{aligned}
\log f(x) &= \log \alpha + (\alpha - 1) \log(x - \mu) I(\mu \leq x \leq \mu + 1) \\
\ln(\alpha) &= \sum_{i=1}^n \log \alpha + (\alpha - 1) \log(x_i - \mu) I(\mu \leq x_i \leq \mu + 1) \\
&= n \log \alpha + (\alpha - 1) \sum_{i=1}^n \log(x_i - \mu) I(X_{(1)} > \mu, X_{(n)} < \mu + 1) \\
\frac{d \ln}{d \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \log(x_i - \mu) = 0 \\
\alpha &= -\frac{n}{\sum_{i=1}^n \log(x_i - \mu)}
\end{aligned}$$

Show that

$$\sqrt{n}(\tilde{\alpha}_n - \alpha) \xrightarrow{d} N(0, \alpha^2)$$

We need to get the Fisher Information  $I(\alpha)$ , then use the central limit theorem.



The Fisher information

$$\begin{aligned}
I(\alpha) &= -E\left[\frac{\partial^2 l_n}{\partial \alpha \partial \alpha}\right] \\
\frac{\partial^2 l_n}{\partial \alpha \partial \alpha} &= -\frac{1}{\alpha^2} \\
I(\alpha) &= E\frac{1}{\alpha^2} = \frac{1}{\alpha^2}
\end{aligned}$$

By central limit theorem, the variance is  $I(\alpha)^{-1}$ .

### 3.1.3 c. Probability

Define  $\tilde{\mu}_n = X_{(1)}$ ,  $\hat{\mu}_n = X_n - 1$ ,  $Y_n = n^{1/\alpha}(\tilde{\mu}_n - \mu)$ ,  $Z_n = n(\mu - \hat{\mu}_n)$ , and show that for all  $0 < y, z < \infty$ ,  $Pr(Y_n > y, Z_n > z) \rightarrow e^{-y^\alpha - \alpha z}$ , as  $n \rightarrow \infty$ , and that  $Y_n, Z_n \geq 0$  almost surely for all  $n \geq 1$ .

#### Preparation

I had the temptation to derive each individual probability for  $X_{(1)}, X_{(n)}$ , but  $Pr(Y_n > y, Z_n > z)$  could not be written as the product of  $Y_n, Z_n$  as they are not independent.

Keep in mind that, the only way to construct the joint distribution is to find the independent individual distribution first, and then start from there. This is always the trick in dealing with distribution. We can see from the derivation of  $X_{(1)}, X_{(n)}$  is that to use all the  $X_i$  in forming the cumulative probability.

And statistics is all about probability and distribution.

Need to review the proof for converge in probability

#### Proof

$$\begin{aligned}
P(Y_n > y, Z_n > z) &= P(n^{1/\alpha}(\tilde{\mu}_n - \mu) > y, n(\mu - \hat{\mu}_n) > z) \\
&= P(n^{1/\alpha}(X_{(1)} - \mu) > y, n(\mu - X_{(n)} + 1) > z) \\
&= P(X_{(1)} > n^{-1/\alpha}y + \mu, X_{(n)} < \mu - \frac{z}{n} + 1) \\
&= \prod_{i=1}^n P(n^{-1/\alpha}y + \mu < X_i < \mu - \frac{z}{n} + 1) \\
&= \left[ (1 - z/n)^\alpha - (yn^{-1/\alpha})^\alpha \right]^n
\end{aligned}$$

We know that

$$\begin{aligned}\left(1 - \frac{z}{n}\right)^{\alpha n} &\rightarrow e^{-\alpha z} \\ \left(1 - \frac{y^\alpha}{n}\right)^n &\rightarrow e^{-y^\alpha}\end{aligned}$$

To find the limit of above  $P(Y_n > y, Z_n > z)$ , we will need to change the sum term into product term.

$$\begin{aligned}(1 - z/n)^\alpha &= \left(1 - \frac{z}{n}\right)^{\alpha/n} \\ &\rightarrow (e^{-z})^{\alpha/n} = e^{-\alpha z/n}\end{aligned}$$

So we have

$$\begin{aligned}\left[(1 - z/n)^\alpha - (y n^{-1/\alpha})^\alpha\right]^n &= \left[e^{-z\alpha/n} - \frac{y^\alpha}{n} e^{\frac{\alpha z}{n}} e^{-\frac{\alpha z}{n}}\right]^n \\ &= e^{-z\alpha} \left[1 - \frac{y^\alpha}{n} e^{\frac{\alpha z}{n}}\right]^n\end{aligned}$$

Since  $e^{\alpha z/n} \rightarrow 1, \alpha z/n \rightarrow 0$ .

$$\begin{aligned}&= e^{-z\alpha} \left[1 - \frac{y^\alpha}{n}\right]^n \\ &= e^{-z\alpha} e^{-y^\alpha} = e^{-z\alpha - y^\alpha}\end{aligned}$$

To see if  $Y_n, Z_n$  almost surely  $\rightarrow 0$ , is to see every each observation  $\rightarrow 0$ .  
Prove  $\hat{\mu}_n \xrightarrow{P} \mu$  let  $\epsilon_i > 0$

$$\begin{aligned}P(|\hat{\mu}_n - \mu| > \epsilon) &= P(|X_{(n)} - 1 - \mu| > \epsilon) \\ &= P(|X_{(n)} - (\mu + 1)| > \epsilon) \\ &= P(X_{(n)} - (\mu + 1) > \epsilon) + P(X_{(n)} - (\mu + 1) < -\epsilon) \\ &= P(X_{(n)} > \epsilon + (\mu + 1)) + P(X_{(n)} < (\mu + 1) - \epsilon) \\ &= 0 + \int_{\mu}^{\mu+1-\epsilon} n\alpha(x - \mu)^{n\alpha-1} dx \\ &= (x - \mu)^{n\alpha} \Big|_{\mu}^{\mu+1-\epsilon} \\ &= (1 - \epsilon)^{n\alpha} \xrightarrow{n \rightarrow \infty} 0\end{aligned}$$

Here we need to find the distribution of  $X_{(n)}$ .

$$F(X_{(n)} < x) = P(X_1 < x, X_2 < x, \dots, X_n < x) = [P(X_i < x)]^n, \quad X_i \text{ are i.i.d.}$$

$$\begin{aligned} P(X_i < x) &= \int_{\mu}^x \alpha(x - \mu)^{\alpha-1} dx \\ &= (x - \mu)^{\alpha} \end{aligned}$$

Then the distribution of max observation

$$\begin{aligned} F(X_{(n)} < x) &= (x - \mu)^{n\alpha} \\ P(X_{(n)} < (\mu + 1) - \epsilon) &= ((\mu + 1) - \epsilon - \mu)^{n\alpha} = (1 - \epsilon)^{n\alpha} \end{aligned}$$

Because  $\hat{\mu}_n \xrightarrow{p} \mu$ , then  $|\hat{\mu}_n - \mu|^r \xrightarrow{p} 0$ . So  $\hat{\mu}_n - \mu = O_p(1)$ .

$$\begin{aligned} |\hat{\mu}_n - \mu| \text{sign}(\hat{\mu}_n - \mu) &= o_p(1) \\ |\hat{\mu}_n - \mu| &= o_p(\text{sign}(\hat{\mu}_n - \mu)) = o_p(1) \end{aligned}$$

By LLN,

$$\frac{1}{n} \sum (X_i - \mu)^{-r} \xrightarrow{p} E[(X_i - \mu)^{-r}] < M$$

### 3.1.4 d. Big O

Let  $\hat{\alpha}_n = \left[ -n^{-1} \sum_{i=1}^n \log(X_i - \hat{\mu}_n) \right]^{-1}$ , and show that for any  $0 < s < \alpha$ ,  $\hat{\alpha}_n - \tilde{\alpha}_n = O_p(n^{-1 \wedge s})$ , where  $a \wedge b$  denote the minimum of a and b.

Hint: we can use the fact that  $0 < r \leq 1, 0 < C_r \leq \infty, \log(1 + \Delta) \leq C_r \Delta^r$ , for all  $0 \leq \Delta < \infty$ . Using the fact, show that

$$0 \leq \frac{1}{\tilde{\alpha}_n} - \frac{1}{\hat{\alpha}_n} \leq C_{s \wedge 1} |\hat{\mu}_n - \mu|^{s \wedge 1} n^{-1} \sum_{i=1}^n (X_i - \mu)^{-s \wedge 1}$$

and then complete the proof.

### Preparation

We can use the hint and try to get the  $\hat{\alpha} - \tilde{\alpha}_n$ . The Big-O shows that bounded and the algebraic of Big-O are also Big-O.

Proof

$$\begin{aligned}
\frac{1}{\tilde{\alpha}_n} - \frac{1}{\hat{\alpha}_n} &= -n^{-1} \sum_{i=1}^n \log(X_i - \mu) + n^{-1} \sum_{i=1}^n \log(X_i - \hat{\mu}_n) \\
\log(1 + \Delta) &\leq C_r \Delta^r, \\
&= \frac{1}{n} \sum_{i=1}^n \log(X_i - X_{(n)} + 1) - \log(X_i - \mu) \\
&= \frac{1}{n} \sum_{i=1}^n \log\left(1 + \frac{\mu - X_{(n)} + 1}{X_i - \mu}\right) \\
&\leq \frac{1}{n} \sum_{i=1}^n C_r \left[\frac{\mu - X_{(n)} + 1}{X_i - \mu}\right]^r
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\frac{\hat{\alpha}_n - \tilde{\alpha}_n}{\hat{\alpha}_n \tilde{\alpha}_n} &\leq \frac{1}{n} \sum_{i=1}^n C_r \left[\frac{\mu - X_{(n)} + 1}{X_i - \mu}\right]^r \\
\hat{\alpha}_n - \tilde{\alpha}_n &\leq \frac{1}{n} \sum_{i=1}^n C_r \left[\frac{\mu - X_{(n)} + 1}{X_i - \mu}\right]^r \hat{\alpha}_n \tilde{\alpha}_n \\
&\leq \frac{1}{n} \sum_{i=1}^n C_r \left[\frac{\mu - X_{(n)} + 1}{X_i - \mu}\right]^r \left[-n^{-1} \sum_{i=1}^n \log(X_i - \hat{\mu}_n)\right]^{-1} \left[-n^{-1} \sum_{i=1}^n \log(X_i - \mu_n)\right]^{-1} \\
&\leq C_r |\mu - X_{(n)} + 1|^r \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{X_i - \mu}\right]^r \left[-n^{-1} \sum_{i=1}^n \log(X_i - \hat{\mu}_n)\right]^{-1} \left[-n^{-1} \sum_{i=1}^n \log(X_i - \mu_n)\right]^{-1}
\end{aligned}$$

We will look at each term,

$$\begin{aligned}
|\mu - X_{(n)} + 1| &\xrightarrow{p} 0, \quad P(|\mu - X_{(n)} + 1| > \epsilon) = 0 \\
|\mu - X_{(n)} + 1|^r &= O_p(1) \\
\frac{1}{n} \sum_{i=1}^n \left[\frac{1}{X_i - \mu}\right]^r &= E\left[\frac{1}{X_i - \mu}\right]^r < M, \quad LLN
\end{aligned}$$

### 3.1.5 e. Prove convergence

Show that for any  $1/2 < \alpha < \infty$ ,  $\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{d} N(0, \alpha^2)$ .

Preparation

This looks like CLT, but actually it is not. I will need to show the cumulative distribution of  $\sqrt{n}(\hat{\alpha}_n - \alpha)$  is  $N(0, \alpha^2)$ . This is very tedious and not doable.

What we can do is to use the Slutsky's theorem, that the combination of the random variables converge to a combination of distributions. So we need to get familiar with the theorems and understand how to use those theorems.

That's how we study the theorem, not only understand what it is, but also how to use it.

Proof

## 3.2 Problem 2 - Hypothesis Test

Suppose that the distribution of a discrete random variable  $X$  is given below

$X$	-2	-1	0	1	2
$p(x)$	$\theta_1(1 - \theta_2)$	$(\frac{1}{2} - \alpha)\frac{1-\theta_1}{1-\alpha}$	$\alpha\frac{1-\theta_1}{1-\alpha}$	$(\frac{1}{2} - \alpha)\frac{1-\theta_1}{1-\alpha}$	$\theta_1\theta_2$

where  $0 < \theta_1 < \alpha < 1/2$ ,  $\alpha$  is known,  $0 < \theta_2 < 1$ , and both  $\theta_1, \theta_2$  unknown. Suppose we would like to test hypothesis

$$H_0 : \theta_1 = \alpha, \theta_2 = 1/2$$

$$H_1 : \theta_1 < \alpha, \theta_2 \neq 1/2$$

### 3.2.1 LRT

Derive the  $\alpha$  level likelihood ratio test (LRT) for  $H_0$  versus  $H_1$  and obtain its power function.

Preparation

Proof

$$LRT = \frac{\sup_{\theta \in \Theta_0} l(\theta)}{\sup_{\theta \in \Theta} l(\theta)}$$

$$=$$

## 3.3 Problem

Consider the linear model  $Y = X\beta + \epsilon$ , where  $\epsilon \sim N_n(0, \sigma^2 I)$ ,  $X$  is  $n \times p$  of rank  $p$ ,  $\beta$  is  $p \times 1$ , and  $(\beta, \sigma^2)$  are unknown. Define  $H = X(X^T X)^{-1} X^T$  and let  $h_{ii}$  denote the  $i$ th

diagonal element of H. Further let  $\hat{\sigma}^2$  denote the usual unbiased estimator of  $\sigma^2$  for the linear regression model and let  $\hat{\epsilon}_i$  denote the ordinary residual. Let

$$A_i = \frac{\hat{\epsilon}_i^2}{\sigma^2(1 - h_{ii})}$$

$$B_i = \frac{(n - p)\hat{\sigma}^2}{\sigma^2} - \frac{\hat{\epsilon}_i^2}{\sigma^2(1 - h_{ii})}$$

3.3.1 a. Show that  $B \sim \chi_{n-p-1}^2$

Preparation

In order to prove  $B \sim \chi_{n-p-1}^2$ , we need to show that it is a sum of n-p-1 normal distribution square (by definition of chi-square distribution).

We can write  $\sigma^2, \hat{\sigma}^2$  the quadratic form in the sum of squares of normal distribution. But the fraction between two quadratic forms is what distribution?

Note that the division is not F-distribution, as it didn't divided by degrees of freedom.

$\hat{\sigma}^2$  is the usual unbiased estimator of  $\sigma^2$ , we need to know that if unbiased, then the degrees of freedom is n-p.

Another thing need to pay attention that,  $\sigma^2$  is not a distribution, it is fixed and it is the variance of  $Y_i$  just unknown.

Variance of  $Y_i$  is also the variance of  $\epsilon$ , but we need to know the difference and how to use them.

$$\frac{(n - p)\hat{\sigma}^2}{\sigma^2} = \frac{Y^T(I - H)Y}{\sigma^2} = \frac{[(I - H)Y]^T[(I - H)Y]}{\sigma^2}$$

$$= \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{Var(y_i)}$$

So is n-p sum of normal distribution square. Then try to link the second term

$$\frac{\hat{\epsilon}_i^2}{\sigma^2(1 - h_{ii})} = \frac{\hat{\epsilon}_i^2}{V(e_i)}$$

This question requires understanding the matrix H, or the relationship between o.p.o matrix and scalar form. And the characteristic of H matrix itself  $H^2 = H, H^T = H$ .

If we write that in scalar form,

$$(h_{i1}, h_{i2} \dots h_{in}) \begin{pmatrix} h_{i1} \\ h_{i2} \\ \dots \\ h_{in} \end{pmatrix} = h_{i1}^2 + h_{i2}^2 + \dots + h_{in}^2 = h_{ii}$$

Variance, Covariance of  $e$

$$\begin{aligned} V(e) &= E[(e - E(e_i))(e - E(e_i))^T] = (I - H)E(\epsilon\epsilon^T)(I - H)^T \\ &= (I - H)I\sigma^2(I - H)^T \\ &= (I - H)(I - H)^T I\sigma^2 = (I - H)\sigma^2 \end{aligned}$$

Need to know that  $V(e_i)$  is given by the  $i$ th diagonal element  $1 - h_{ii}$  and  $Cov(e_i, e_j)$  is given by the  $(i, j)$ th element of  $-h_{ij}$  of the matrix  $(I - H)\sigma^2$ .

We also need to know that both  $Y_i$  and  $\epsilon_i$  follows the same distribution

$$\begin{aligned} e - E(e) &= (I - H)Y = (Y - X\beta) = (I - H)\epsilon \\ E(\epsilon\epsilon^T) &= V(\epsilon) = I\sigma^2, \quad E(\epsilon) = 0 \end{aligned}$$

We also have correlation

$$\begin{aligned} \rho_{ij} &= \frac{Cov(e_i, e_j)}{\sqrt{V(e_i)V(e_j)}} = -\frac{h_{ij}}{(1 - h_{ii})(1 - h_{jj})} \\ SS(b) &= SS(parameter) = b^T X^T Y = \hat{Y}^T Y = Y^T H^T Y = Y^T H Y = Y^T H^2 Y = \hat{Y}^T \hat{Y} \end{aligned}$$

The average  $V(\hat{Y}_i)$  to all data points is

$$\begin{aligned} \sum_{i=1}^n \frac{V(\hat{Y}_i)}{n} &= \frac{trace(H\sigma^2)}{n} = \frac{p\sigma^2}{n} \\ \hat{Y}_i &= h_{ii}Y_i + \sum_{j \neq i} h_{ij}Y_j \end{aligned}$$

Studentized residual

$$V(e_i) = (1 - h_{ii})\sigma^2$$

where  $\sigma^2$  is estimated by  $s^2$

$$s^2 = \frac{e^T e}{n - p} = \frac{\sum_{i=1}^n e_i^2}{n - p}$$

Sum of Squares attributable to  $e_i$

$$\begin{aligned}
e &= (I - H)Y \\
e_i &= -h_{i1}Y_1 - h_{i2}Y_2 - \dots + (1 - h_{ii})Y_i - \dots - h_{in}Y_n = c^T Y \\
\text{Var}(e_i) &= c^T \text{Var}(Y) c = \sigma^2 c^T c \\
c^T &= (-h_{i1}, -h_{i2}, \dots, (1 - h_{ii}), \dots, -h_{in}), \quad (I - H)I - H = I - H \\
c^T c &= \sum_{j=1}^n h_{ij}^2 + (1 - 2h_{ii}) = (1 - h_{ii}) \\
\text{Var}(e_i) &= \sigma^2(1 - h_{ii}) \\
SS(e_i) &= e_i^2 = \left( (I - H)y \right)^T (I - H)y = y^T (I - H)y
\end{aligned}$$

Need to learn to write in matrix form.

**Proof**

Write  $B_i$  in matrix form

$$\begin{aligned}
\frac{(n-p)\hat{\sigma}^2}{\sigma^2} &= \frac{Y^T(I-H)Y}{\sigma^2} = \frac{[(I-H)Y]^T[(I-H)Y]}{\sigma^2} \\
\frac{\hat{e}_i^2}{\sigma^2(1-h_{ii})} &= \frac{[e_i^T(I-H)Y]^T[e_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2}
\end{aligned}$$

where  $e_i = (0, \dots, 0, 1, 0, \dots, 0)$  Then

$$\frac{[e_i^T(I-H)Y]^T[e_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2} = \frac{[Y^T(I-H)e_i e_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2}$$

Then

$$\begin{aligned}
B_i &= \frac{[Y^T(I-H)Y]}{\sigma^2} - \frac{[Y^T(I-H)e_i e_i^T(I-H)Y]/(1-h_{ii})}{\sigma^2} \\
&= \frac{Y^T}{\sigma} \left[ (I-H) - (I-H) \frac{e_i e_i^T}{1-h_{ii}} (I-H) \right] \frac{Y}{\sigma}
\end{aligned}$$

$\frac{Y}{\sigma} \sim N_n(0, I)$ .

To show that  $C = \left[ (I-H) - (I-H) \frac{e_i e_i^T}{1-h_{ii}} (I-H) \right]$  is a projection matrix, we need to show  $C^2 = C, C^T = C$ .

Here  $(I-H)(I-H) = I-H, e_i^T(I-H)e_i = 1-h_{ii}$

Therefore,  $B_i \sim \chi^2(r)$ , where  $r = \text{rank}[C]$ . For projection matrix, the rank is also the trace.



$$\begin{aligned}
\text{rank}\left[(I - H) - (I - H)\frac{e_i e_i^T}{1 - h_{ii}}(I - H)\right] &= \text{tr}\left[(I - H) - (I - H)\frac{e_i e_i^T}{1 - h_{ii}}(I - H)\right] \\
&= n - p - \frac{1}{1 - h_{ii}}\text{tr}[(I - H)e_i e_i^T(I - H)] \\
&= n - p - \frac{1}{1 - h_{ii}}\text{tr}[e_i^T(I - H)(I - H)e_i] \\
&= n - p - \frac{1}{1 - h_{ii}}\text{tr}[e_i^T(I - H)e_i] \\
&= n - p - \frac{1 - h_{ii}}{1 - h_{ii}} \\
&= n - p - 1
\end{aligned}$$

So  $B_i \sim \chi^2(n - p - 1)$ .

### 3.3.2 b. Show $A_i$ and $B_i$ are independent

#### Preparation

Note that  $A_i$  and  $B_i$  are quadratic forms, which we need to show that their projection matrix product is 0, which will need to use linear algebra.

Another method is that, there are chi-square distribution, which we just need to show that the two statistics are ancillary.

Traditional method is to derive the distributions, and if we could write the distribution of  $A_i + B_i$  as the product of distribution of  $A_i$  and  $B_i$ , then we can prove the independence.

One more question, are we able to use the moment generating function to do this? It is always easier to prove in MGF.

First, we need to write the MGF of  $A_i$ ,  $B_i$

$$\begin{aligned}
M_{A_i} &= E[\exp[A_i t]] = \int \exp[A_i t] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
M_{B_i} &= E[\exp[B_i t]] = \int \exp[B_i t] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx
\end{aligned}$$

Then we can use MGF properties to prove

$$\begin{aligned}
W_i &= A_i + B_i \\
M_{W_i} &= E[\exp(W_i t)] = E[\exp[(A_i + B_i)t]] \\
&= E[\exp[A_i t] \exp[B_i t]], \quad \text{all based on normal distribution} \\
&= E[\exp[A_i t]] E[\exp[B_i t]] = M_{A_i} M_{B_i}
\end{aligned}$$

Also we need to notice that, this is a quadratic form chi-square distribution.

(a) Quadratic form theorem 1:

Let the  $n \times 1$  vector  $y \sim N(0, I)$ , let  $A$  be an  $n \times n$  idempotent matrix of rank  $m$ , let  $B$  be an  $n \times n$  idempotent matrix of rank  $s$ , and suppose  $BA = 0$ . Then  $y'Ay$  and  $y'By$  are independently distributed  $\chi^2$  variables.

**Why need to be idempotent matrix?** It is because for idempotent matrix  $A^2 = A$ , then the quadratic form could be written as  $y'Ay = (Ay)'(Ay)$ , then further we can use the product of two vectors  $= 0$  to show independence.

The proof is try write the two quadratic form in partition, so that show that they dependent on different vectors, further they are independent. Need to compare with the proof in 762 slides.

(b) Quadratic form theorem 2 (Craig's Theorem): If  $y \sim N(\mu, \Omega)$  where  $\Omega$  is positive definite, then  $q_1 = y'Ay$  and  $q_2 = y'By$  are independently distributed if  $A\Omega B = 0$

This is just a generalization of above theorem. Since  $\Omega$  is a covariance matrix of full rank it is positive definite and can be factored as  $\Omega = TT'$ . Therefore the condition  $A\Omega B = 0$  can be written  $ATT'B = 0$ .

$$T'ATT'BT = 0$$

$$C = T'AT, \quad K = T'BT$$

$$CK = (T'AT)(T'BT) = T'\Omega BT = T'0T = 0$$

In the 762 slides, it used the covariance matrix  $= 0$  to prove independence in MVN.

$$Y'AY = Y'RR'Y = (R'Y)'(R'Y)$$

$$Y'BY = Y'SS'Y = (S'Y)'(S'Y)$$

Thus  $Y'AY$  and  $Y'BY$  are independent if  $R'Y$  and  $S'Y$  are independent.  $R'Y$  and  $S'Y$  are independent if and only if

$$Cov(R'Y, S'Y) = 0$$

$$= R'\Sigma S = 0$$

$$= R'WW'S = 0$$

$$= C(Q'S) \perp C(Q'R)$$

Proof

$$A_i = \frac{[Y^T(I - H)e_ie_i^T(I - H)Y]/(1 - h_{ii})}{\sigma^2}$$

$$B_i = \frac{[Y^T[(I - H) - (I - H)\frac{e_ie_i^T}{(1 - h_{ii})}(I - H)]Y]}{\sigma^2}$$

We would like to show the product of two projection matrix are 0, by using the above theorem

$$\begin{aligned} M_{A_i} M_{B_i} &= (I - H) \frac{e_i e_i^T}{1 - h_{ii}} (I - H) \left[ (I - H) - (I - H) \frac{e_i e_i^T}{(1 - h_{ii})} (I - H) \right] \\ &= 0 \\ e_i^T (I - H) e_i &= 1 - h_{ii} \end{aligned}$$

### 3.3.3 c. Derive exact distribution

Let

$$r_i = \frac{\hat{\epsilon}_i}{\hat{\sigma} \sqrt{1 - h_{ii}}}$$

Using part(a) and (b), derive the exact distribution of  $r_i^2/(n - p)$ .

#### Preparation

This problem needs observation of the  $r_i$  and connection between  $A_i$  and  $B_i$ . Also use the results from part (a) and (b) that,  $A_i$  and  $B_i$  are independent chi-square.

$$\frac{r_i^2}{(n - p)} = \frac{A_i}{A_i + B_i}$$

The transformation of distribution, first we can use the distribution theory that the sum of independent chi-square distribution is also a chi-square. The first method we need to try is to use MGF, then try to use the distribution transformation.

if we use distribution transformation, then we need to use two independent distributions

$$\begin{aligned} w &= \frac{A_i}{A_i + B_i} \\ B_i &= \frac{1 - w}{w} A_i \end{aligned}$$

replace  $B_i$  by w, then integrate out of  $A_i$ , then we will get the distribution of  $w$ .

the chi-square distribution  $X \sim \chi^2(k)$  is a special case of the gamma distribution, in that  $X \sim \Gamma(\frac{k}{2}, \frac{1}{2})$  using the rate parameterization of the gamma distribution.

#### Proof

We need to know the relationship between Gamma distribution and Beta distribution. As  $A_i \perp B_i$

$$\frac{A_i}{A_i + B_i} \sim \text{Beta}\left(\frac{1}{2}, \frac{n - p - 1}{2}\right)$$

Show the distribution transformation. Assume that  $X \sim \text{Gamma}(\alpha, 1)$ ,  $Y \sim \text{Gamma}(\beta, 1)$ , and  $X$  and  $Y$  are independent. then find the distribution of  $\frac{X}{X+Y}$ .

$$f_{X,Y}(x, y) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-(x+y)}$$

let  $w = \frac{x}{x+y}$ ,  $u = x$ , then we have

$$\begin{aligned} x &= y, & y &= \frac{(1-w)u}{w} \\ J &= \left| \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial w} \end{pmatrix} \right| \\ &= \left| \begin{pmatrix} 1 & 0 \\ \frac{(1-w)}{w} & -\frac{u}{w^2} \end{pmatrix} \right| = \frac{u}{w^2} \end{aligned}$$

Then integrate out  $u$

$$\begin{aligned} f(w) &= \int_0^\infty f(w, u) du \\ &= \int_0^\infty \frac{1}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} \left(\frac{(1-w)u}{w}\right)^{\beta-1} e^{-\frac{u}{w}} \frac{u}{w^2} du \\ &\sim \text{Beta}(\alpha, \beta) \end{aligned}$$

### 3.3.4 d. Outlier Model 1

Suppose we suspect that the  $i$ th case is an outlier and we consider the mean shift outlier model  $Y = X\beta + d_i\phi + \epsilon$ , where  $\epsilon \sim N_n(0, \sigma^2 I)$ ,  $\phi$  is an unknown scalar, and  $d_i$  is an  $n \times 1$  vector with a 1 in the  $i$ th position and zeros elsewhere. Derive the maximum likelihood estimate of  $\phi$ .

#### Preparation

We can construct the model by introducing the unknown parameter  $\phi$ , then use the MLE to estimate.

The way to get MLE is to construct likelihood function first,  $Y$  is a normal distribution, and only  $i$ th case is an outlier, while all other cases are not. Solve the problem similar as the random model.

As the  $Y$  are independent variables,  $Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n \sim N(X_k\beta, \sigma^2)$ , while  $Y_i \sim N(X_i\beta + \phi, \sigma^2)$ .

If using scalar form, the  $\beta, \phi$  are correlated with each other, which we won't be able

to solve.

$$\begin{aligned}
f_k(\beta, \phi) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_k - x_k\beta)^2}{2\sigma^2}\right) \\
f_i(\beta, \phi) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - x_i\beta - \phi)^2}{2\sigma^2}\right) \\
\log f_k(\beta, \phi) &= -\frac{(y_k - x_k\beta)^2}{2\sigma^2} \\
\log f_i(\beta, \phi) &= -\frac{(y_i - x_i\beta - \phi)^2}{2\sigma^2} \\
\ln(\beta, \phi) &= \sum_{k=1, k \neq i}^n -\frac{(y_k - x_k\beta)^2}{2\sigma^2} - \frac{(y_i - x_i\beta - \phi)^2}{2\sigma^2}
\end{aligned}$$

The MLE of  $\phi$

$$\begin{aligned}
\frac{\partial \ln}{\partial \phi} &= \frac{(y_i - x_i\beta - \phi)}{\sigma^2} = 0 \\
\phi &= y_i - x_i\hat{\beta}
\end{aligned}$$

where  $\hat{\beta}$  is the MLE of  $\beta$ .

So we need to use the linear algebra form in solve this kind of question. Need to notice that the linear regression estimate in matrix form all use the linear algebra.

**Proof**

The maximum likelihood function in linear algebra form

$$\begin{aligned}
L(\beta, \phi) &\propto \exp\left(-\frac{(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi)}{2\sigma^2}\right) \\
l(\beta, \phi) &\propto \left(-\frac{(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi)}{2\sigma^2}\right) \\
&= -\frac{1}{2\sigma^2} \left[ (Y - X\beta)^T(Y - X\beta) - 2(Y - X\beta)^T d_i\phi + d_i^T d_i\phi^2 \right]
\end{aligned}$$

The MLE

$$\begin{aligned}
\frac{\partial l}{\partial \phi} &= \left(-\frac{(Y - X\beta - d_i\phi)^T(Y - X\beta - d_i\phi)}{2\sigma^2}\right) \\
&\rightarrow -2(Y - X\beta)^T d_i + d_i^T d_i(2\phi) = 0 \\
\phi &= (Y - X\beta)^T d_i
\end{aligned}$$

MLE of  $\beta$

$$\begin{aligned} X\beta &= HY \\ \hat{\beta} &= (X^T X)^{-1} X^T Y \end{aligned}$$

There is no need to get the estimate of  $\beta$ , we only need to estimate  $(Y - X\beta)^T$ . So  $\hat{\phi} = (I - H)Y^T d_i$

### 3.3.5 d. Outlier Model 2

Suppose we wish to test  $H_0 : \phi = 0$ . Derive the test statistic for this hypothesis and derive its exact distribution under  $H_0$ .

Preparation

The hypothesis test always comes with the estimate. There are several hypothesis test method we can do, wald test, score test and likelihood ratio test. The score test is generally the way as it is easy to get the estimate, score function and fisher information under  $H_0$ .

We can use  $A_i$  and  $B_i$  are independent and they are quadratic form, so we can have F -test and try to find the connection between the  $\hat{\phi}$  and F-test.

Proof

To test  $\phi = 0$ , consider wald test

$$W_n = (\hat{\phi} - 0)^T I_n(\phi) (\hat{\phi} - 0)$$

Find the fisher information  $I_n(\phi)$

$$\begin{aligned} I_n(\phi) &= nI(\phi) \\ I(\phi) &= -E\left[\frac{\partial^2 l}{\partial \phi \partial \phi}\right] \\ &= \frac{1}{\sigma^2} \\ I_n(\phi) &= \frac{n}{\sigma^2} \end{aligned}$$

We need to find the MLE of  $\sigma^2$ .

$$\begin{aligned} \frac{\partial^2 l}{\partial \sigma^2} &= \frac{1}{2\sigma^4} (Y - X\beta - d_i \phi)^T (Y - X\beta - d_i \phi) + \frac{\partial}{\partial \sigma^2} \log[(\sigma^2)^{-1/2}] \\ \hat{\sigma}^2 &= \|Y - X\beta - d_i \phi\| \\ &= \|(I - H)Y - d_i \hat{\phi}\| \end{aligned}$$

So the wald statistics

$$W_n = d_i^T Y(I - H) \left( \frac{n}{\|(I - H)Y - d_i \hat{\phi}\|} \right) (I - H)Y^T d_i$$

Under  $H_0$ , it is a  $\chi^2(1)$ .

The second method is to use F-test, as  $A_i$  and  $B_i$  are independent quadratic form

$$A_i = \frac{[Y^T(I - H)e_i e_i^T(I - H)Y]/(1 - h_{ii})}{\sigma^2} = \frac{[\hat{\phi}^2]/(1 - h_{ii})}{\sigma^2}$$

$$B_i = \frac{[(I - H)Y]^T[(I - H)Y] - [\hat{\phi}^2]/(1 - h_{ii})}{\sigma^2}$$

$$F = \frac{A_i}{B_i/(n - p - 1)}$$

$$= \frac{\hat{\phi}^2/(1 - h_{ii})}{[(I - H)Y]^T[(I - H)Y] - [\hat{\phi}^2]/(1 - h_{ii})/(n - p - 1)}$$

We can perform F-test by setting up the rejection region.

### 3.3.6 e. Cook's Distance

Let  $I = \{1, ..m\}$  be the subset of the first m cases in the dataset. Let  $D_I$  denote the Cook's distance based on simultaneously deleting m cases from the dataset, which is given by

$$D_I = \frac{(\hat{\beta} - \hat{\beta}_I)^T (X^T X)(\hat{\beta} - \hat{\beta}_I)}{p \hat{\sigma}^2}$$

where  $\hat{\beta}^I$  denotes the least squares estimate of  $\beta$  with the cases deleted from set I and  $\hat{\beta}$  denotes the estimate of  $\beta$  based on the full data. Show that  $D_I$  can be written as

$$D_I = \frac{1}{p} \sum_{i=1}^m h_i^2 \left( \frac{\lambda_i}{1 - \lambda_i} \right)$$

where the  $\lambda_i, i = 1, ..m$ , are the eigenvalues of the matrix  $P_I = X_I(X^T X)^{-1}X_I^T$  based on a spectral decomposition of  $P_I$ .

## Preparation

The single deletion case

$$\hat{\beta}_{(1)} = \hat{\beta} - \frac{(X'X)^{-1}x_i^T\hat{\epsilon}_i}{1 - h_{ii}}$$

This method is used in Bayesian method, get marginal distribution from joint distribution, expectation of conditional distribution.

The probability involved in retrospective study sampling method

The likelihood function for the retrospective study is the conditional density of  $X = x$  given  $Y$  and  $S = 1$ , denoted by  $p(x|Y, S = 1)$ . To avoid biased sampling, we assume that the inclusion of a subject is independent of  $x$  given  $Y$ , that is,

$$P(S = 1|X, Y) = P(S = 1|Y)$$

Under unbiased sampling, we have

$$p(x|Y, S = 1) = \frac{P(S = 1|x, Y)p(x|Y)}{P(S = 1|Y)} = p(x|Y)$$

Furthermore,  $p(x|Y, S = 1)$  can be expressed as

$$p(x|Y, S = 1) = \frac{P(Y|S = 1, x)p(x|Y)P(x|S = 1)}{P(Y|S = 1)} = p(x|Y)$$

and  $p(Y|x, S = 1)$  is given by

$$\begin{aligned} p(Y|x, S = 1) &= \frac{p(X|Y, S = 1)p(Y, S = 1)}{p(X, S = 1)} = p(Y|x) \frac{p(S = 1|Y)}{p(Y|S = 1)} \\ &= \frac{p(X, Y, S = 1)}{p(X, S = 1)} = \frac{p(Y, S = 1|x)p(x)}{p(S = 1|x)p(x)} \\ &= \frac{p(S = 1|x, Y)p(Y)}{p(S = 1|x)} = \frac{p(S = 1|Y)p(Y|x)}{p(S = 1|x)} \end{aligned}$$

Finally,  $p(x|Y, S = 1)$  can be expressed as

$$p(x|Y, S = 1) = \frac{P(Y|x)p(x|S = 1)P(S = 1|Y)}{P(Y|S = 1)p(S = 1|x)}$$

Why do we write in this form? As it is how we select the patients in the case-control study, and it reveals the relationship between prospective and retrospective probability.



The ratio  $p(x|S = 1)/p(S = 1|x)$  only depends on  $x$  and  $P(S = 1|Y)$  and  $P(Y|S = 1)$  are determined by the sampling scheme.

We can construct the likelihood function for the retrospective study

Let  $H_1(Y) = P(S = 1|Y)/P(Y|S = 1)$  and  $H_2(x) = p(x|S = 1)/p(S = 1|x)$

$$\prod_{i=1}^n p(x_i|Y_i, S_i = 1) = \prod_{i=1}^n P(y_i|x_i, \beta) \prod_{i=1}^n [H_1(y_i)H_2(x_i)]$$

Proof

### 3.4 Problem 3 - Expectation in case control study

#### 3.4.1 a. Expectation

Consider the conditional mean of  $Y$  given  $(X, D)$ , denoted by  $\tilde{\mu}(X, D) = E[Y|X, D]$ . Show that

$$\begin{aligned}\mu(X) &= \tilde{\mu}(X, 1)Pr(D = 1|X) + \tilde{\mu}(X, 0)Pr(D = 0|X) \\ \tilde{\mu}(X, D) &= E[Y|X, D, S = 1] = \mu(X) + \{D - Pr(D = 1|X)\}\gamma(X)\end{aligned}$$

Preparation

Find the expectations in the case-control study from 762 slides.

Proof

The expectation of conditional distribution

$$\begin{aligned}E(Y|X) &= E[E[Y|X, D]|X] \\ &= E[Y|X, D = 1]Pr(D = 1|X) + E[Y|X, D = 0]Pr(D = 0|X) \\ &= \tilde{\mu}(X, 1)Pr(D = 1|X) + \tilde{\mu}(X, 0)Pr(D = 0|X)\end{aligned}$$

In terms of  $\tilde{\mu}(X, D) = E[Y|X, D, S = 1]$ , here we know that all the patients in the case-control study with  $S = 1$ , so  $\tilde{\mu}(X, D) = E[Y|X, D, S = 1] = E[Y|X, D]$ .

$$E(Y|X) = \tilde{\mu}(X, 1)Pr(D = 1|X) + \tilde{\mu}(X, 0)Pr(D = 0|X)$$

