

Qualify Exam 2015

Mingwei Fei

January 21, 2023

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

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

1.2 b. MLE

Assume μ is known, show MLE of α .

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

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 $I(\alpha)^{-1}$.

1.3 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 $n \geq 1$.

1.3.1 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

1.3.2 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

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

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

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

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 $\hat{\mu}_n - \mu = O_p(1)$.

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

By LLN,

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

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

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

and then complete the proof.

1.4.1 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.

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

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,

$$\begin{aligned} |\mu - X_{(n)} + 1| &\xrightarrow{P} 0, \quad P(|\mu - X_{(n)} + 1| > \epsilon) = 0 \\ F_{X_{(n)}}(x) &= (x - \mu)^{n\alpha} \Big|_{\mu+1-\epsilon}^{\mu+1+\epsilon} \end{aligned}$$

1.4.2 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 $(Z + \mu)^2$

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

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 $X_i = Z_i + \mu_i, i = 1, 2, ..n$.

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

1.5 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.

1.5.1 Preparation

From part (a), we see that $Y^T AY$ is sum of chi-square distributions. In order to show it is a chi-square, then the $\lambda_i = 1, 0$. Well, $tr[(A\Sigma)^2] = tr(A\Sigma) = r$ is similar to the idempotent matrix definition, and λ_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 Σ 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 $A\Sigma = P\Lambda P'$.

$$\begin{aligned} A\Sigma &= P\Lambda P' \\ (A\Sigma)^2 &= P\Lambda^2 P^T \\ tr((A\Sigma)^2) &= tr(A\Sigma) \\ &\rightarrow tr(P\Lambda P') = tr(P\Lambda^2 P^T) \\ &\rightarrow tr(\Lambda^2) = tr(\Lambda), \\ &\rightarrow \sum_{i=1}^k \lambda_i^2 = \sum_{i=1}^k \lambda_i \\ &\rightarrow \sum_{i=1}^k \lambda_i(\lambda_i - 1) = 0 \\ &\rightarrow \lambda_i = 0, 1 \end{aligned}$$

1.5.2 Proof

From part (a) that

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

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

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

1.6 d. Distribution

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

1.6.1 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 λ_i need to be the same, 1. Otherwise it won't be a chi-square.

1.6.2 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 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 , β is $p \times 1$, and (β, σ^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 σ^2 for the linear regression model and let $\hat{\epsilon}_i$ denote the ordinary residual. Let

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

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

2.1.1 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 σ^2 , we need to know that if unbiased, then the degrees of freedom is $n-p$.

Another thing need to pay attention that, σ^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 ϵ , but we need to know the difference and how to use them.

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

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 $-(I-H)\sigma^2$.

We also need to know that both Y_i and ϵ_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

$$\sum_{i=1}^n \frac{V(\hat{Y}_i)}{n} = \frac{\text{trace}(H\sigma^2)}{n} = \frac{p\sigma^2}{n}$$

$$\hat{Y}_i = h_{ii}Y_i + \sum_{j \neq i} h_{ij}Y_j$$

Studentized residual

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

where σ^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

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

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

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

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

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

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

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

Need to learn to write in matrix form.

2.1.2 Proof

Write B_i in matrix form

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

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

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

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

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

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 $C^2 = C, C^T = C$.

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

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

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

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

2.2 b. Show A_i and B_i are independent

2.2.1 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

$$M_{A_i} = E[\exp[A_i t]] = \int \exp[A_i t] \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2}) dx$$

$$M_{B_i} = E[\exp[B_i t]] = \int \exp[B_i t] \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2}) dx$$

Then we can use MGF properties to prove

$$W_i = A_i + B_i$$

$$M_{W_i} = E[\exp(W_i t)] = E[\exp[(A_i + B_i)t]]$$

$$= E[\exp[A_i t] \exp[B_i t]], \quad \text{all based on normal distribution}$$

$$= E[\exp[A_i t]] E[\exp[B_i t]] = M_{A_i} M_{B_i}$$

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 χ^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 Ω 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 Ω 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 $ATT'B = 0$.

$$T' ATT' BT = 0$$

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

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

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

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

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

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

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

2.2.2 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

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

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 $r_i^2/(n-p)$.

2.3.1 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.

2.3.2 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

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

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

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

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 $\epsilon \sim N_n(0, \sigma^2 I)$, ϕ 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 ϕ .

2.4.1 Preparation

We can construct the model by introducing the unknown parameter ϕ , 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 β, ϕ 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 ϕ

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

where $\hat{\beta}$ is the MLE of β .

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.

2.4.2 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 β

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

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

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 .

2.5.1 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.

2.5.2 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 σ^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.

2.6 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 β with the cases deleted from set I and $\hat{\beta}$ denotes the estimate of β 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 .

2.6.1 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 $H_2(x) = p(x|S=1)/p(S=1|x)$

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

2.6.2 Proof

3 Problem 3 - Expectation in case control study

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

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

3.1.1 Preparation

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

3.1.2 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 $\tilde{\mu}(X, D) = E[Y|X, D, S = 1] = E[Y|X, D]$.

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