Qualify Exam 2015

Mingwei Fei

December 17, 2022

1 MVN

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

1.1 Quadratic Form $Y^T AY$ and Chi-square distribution

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

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

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

1.1.1 Question

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

(i) *Normal distribution vs. Chi-square:*

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

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

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

(ii) *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 $Q_i \sim \chi^2_{k_i}(\lambda_i)$ for $i = 1, \ldots, n$, be independent. Then, $Q = \sum_{i=1}^n Q_i$ is a noncentral $\chi^2_k(\lambda)$, 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 δ_i and d_i in the non-central chi-square distribution.

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

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

(iii) Here A matrix is not necessarily inverse of Σ, it could be any symmetric matrix. So this is a general case of linear combination of non-central chi-square.

$$\begin{aligned}
\Sigma &= QQ^T \\
Y^T AY &= (Q^T Y)^T diag\{\lambda_1, ..., \lambda_k\}(Q^T Y) \\
Q^T Y &= \Sigma^{-1/2} Y \sim N(\mu, 1) \\
A &= \Sigma^{-1/2} diag\{\lambda_1, ..., \lambda_k\}\Sigma^{-1/2} \\
A^T &= A, \qquad \text{A is symmetric}
\end{aligned}$$

1.1.2 Proof

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

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

$$
\begin{aligned}
\Sigma &= QQ^T, \qquad \text{by semi-definite matrix} \\
Q^{-1}Y &= (Z_i), \qquad Z_i \sim N(\mu_i, I) \\
A &= Q\Lambda Q^T, \qquad \Lambda = diag\{\lambda_1, ...\lambda_k\} \\
Y^TAY &= Y^TQ\Lambda Q^TY = (Q^TY)^T diag\{\lambda_1, ...\lambda_k\}(Q^TY) \\
&= \sum_{i=1}^{k} \lambda_i Z_i^2 =\sim \sum_{i=1}^{k} \lambda_i \chi^2(d_i, \delta_i), \qquad \delta_i = \mu_i^2 \\
Y^TAY &= \sum_{i=1}^{k} \lambda_i W_i, \qquad W_i \sim \chi^2_{d_i}(\delta_i)
\end{aligned}
$$

λ_i is the eigenvalue of matrix A , d_i is the number of same eigenvalue λ_i.

W_i is the non-central chi-square distribution with noncentrality parameter $\delta_i = \sum_{i=1}^{d_i} \mu_i^2$.

1.2 MGF of Y^TAY

Use part (a) to derive the moment generating function of Y^TAY . Let m(t) denote the moment generating function. Show that m(t) exits 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.

1.2.1 Proof

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

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

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 $(Z+\mu)^2$

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

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

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

1.3 $A=\Sigma^{-1}$

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

1.3.1 Question

1.3.2 Proof

From part (a) that

$$\begin{aligned}
A = Q\Lambda Q^T, &\qquad \Sigma = QQ^T \\
(A\Sigma)^2 = A\Sigma A\Sigma &= [Q\Lambda Q^T QQ^T][Q\Lambda Q^T QQ^T] = Q\Lambda^2 Q^T \\
tr((A\Sigma)^2) = tr(A\Sigma), &\qquad \lambda_i^2=\lambda_i, \qquad \lambda_i = 1,0
\end{aligned}$$

As $r=\sum_{i=1}^k \lambda_i$ is the rank of A, then we have

$$A\Sigma = diagBlk\{I_{r\times r}, \quad 0_{n-r\times n-r}\}$$

Then Y^TAY is the sum of r chi-square $\chi^2(1,\delta_i)$

$$Y^T AY = (Q^T Y)^T I_{r\times r}(Q^T Y) = \chi^(r, \delta)$$
$$\delta = \sum_{i=1}^{r} \mu_i^2$$

The degrees of freedom is r, the non-centrality parameter is $\delta = \sum_{i=1}^{r} \mu_i^2$.

1.4 $YTAY$ Distribution

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

1.4.1 Questions

Need to link the piece of information together. In order to have noncentral chi-square,

$$A = PP^T, \qquad \text{Symmetric matrix}$$
$$Y^T AY = (P^T Y)^T (P^T Y) \sim \chi^2(r, \delta), \qquad P^T Y \sim N(\mu, I)$$
$$P^T Y = \Sigma^{-1/2} Y \sim N(\mu, I)$$

Idempotent is $A^2 = A, A^T = A$. To prove if and only if, we need to demonstrate both way. And often times we need to show contradiction.

From part(c), we already show that when A is idempotent, the $Y^T AY$ has a non-central chi-squared distribution.

$A\Sigma = (A\Sigma)^2$, $A\Sigma$ is idempotent.

1.4.2 Proof

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

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(w_i t)$$
$$= \prod_{i=1}^{n} (1 - 2w_j t)^{-1/2} exp\left(\frac{\lambda w_j t}{1 - 2w_j t}\right)$$

Then we can see that the shape parameter is $\frac{1}{2w_i}$. If we want to have a non-central chi-square distribution for Y, then all w_j need to be the same.

$$M_Y(t) = \prod_{i=1}^{n} (1 - 2w_i t)^{-1/2} exp\left(\frac{\lambda_i w_i t}{1 - 2w_i t}\right)$$
$$= (1 - 2wt)^{-n/2} exp\left(\frac{\sum_{i=1}^{n} \lambda_i wt}{1 - 2wt}\right)$$

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.

The other way is also proved from part (c).

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 $H = X(X^TX)^{-1}X^T$ and let h_{ii} denote the ith 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 $A_i = \frac{\hat{\epsilon}_i}{\sigma^2(1-h_{ii})}$, and $B_i = \frac{(n-p)\hat{\sigma}^2}{\sigma^2} - \frac{\hat{\epsilon}_i^2}{\sigma^2(1-h_{ii})}$.

2.1 Show that $B \sim \chi^2_{n-p-1}$

2.1.1 Question

In order to prove $B \sim \chi^2_{n-p-1}$, 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,

$$(h_{i1}, h_{12}, ...h_{1n})^T(h_{i1}, h_{12}, ...h_{1n}) = h_{i1}^2 + h_{i2}^2 + .. + h_{in}^2 = h_{i1}$$

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 ith 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 ϵ_i follows the same distribution

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

We also have correlation

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

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 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

$$
\begin{aligned}
e &= (I - H)Y \\
e_i &= -h_{i1}Y_1 - h_{i2}Y_2 - ... + (1 - h_{ii})Y_i - .. - h_{in}Y_n = c^T Y \\
c^T &= (-h_{i1}, -h_{i2}, ..(1 - h_{ii}), .. - h_{in}) \\
c^T c &= \sum_{j=1}^{n} h_{ij}^2 + (1 - 2h_{ii}) = (1 - h_{ii}) \\
SS(e_i) &= \frac{e_i^2}{(1 - h_{ii})}
\end{aligned}
$$

2.2 Show that A_i and B_i are independent

2.2.1 Question

How to prove independence? We have shown in part (a), there are chi-square distribution, which we just need to show that the two statistics are ancillary.

One statistics is a part of the other statistics, so 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

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

Then we can use MGF properties to prove

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

2.2.2 Proof

2.3 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 Question

To derive the exact the distribution, we can start from get the distribution of r_i. It is a normal distribution from part(a), then we can transform the distribution.

2.4 Outlier Model 1

Suppose we suspect that the ith 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 ith position and zeros elsewhere. Derive the maximum likelihood estimate of ϕ.

2.4.1 Question

We need to under the outlier model is introducing a new parameter ϕ and we need to estimate the parameter with the existing parameter β.

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

2.5 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 Question

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.

2.6 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 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 $\lambda_i, i = 1, ..m$, 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.