Biostat 250A HW5

Elvis Cui

Han Cui

Dept. of Biostat

UCLA

Biostat 250A Homework 5 due 11/9/20

1 Suppose Y is a p-dimensional multivariate normal distribution with mean m and covariance matrix V. Let Y be partitioned into Y_1 and Y_2 so that $Y' = (Y_1', Y_2')$ and the dimension of Y_1 and Y_2 are, respectively, p_1 and p_2 with $p_1+p_2=p$. Find the distribution of Y_1 given Y_2 if V is singular and express the conditional distribution in terms of only m, p_1 and the 4 submatrices in the partitioned matrix V and an appropriate g-inverse.

2 Do Miscellaneous Exercise 2, problem 15 on p.32 of text.

3 If m is a px1 vector of constants, $y \sim N_p(m, I)$ and A is idempotent of rank k. Write down the distribution of $(y-a)^T A(y-a)$? Justify your answer.

3a (Define a non-central chi-square distribution by its density). Let p_m be the density of a central chi-square distribution with m degrees of freedom, and for each nonnegative s, let $q_j = \exp(-s/2)\,(s/2)^j/j!$. For s=0, $q_0 = 1$ and $q_j = 0$ for $j > 0$. A random variable with density

$$h(z) = \sum_{j=0}^{\infty} q_j\, p_{m+2j}(z), \qquad z > 0$$

is said to have a non-central chi-square distribution with m degrees of freedom and non-centrality parameter s. Use the power series expansion of exp(x) to find the moment generating function of such a random variable.

3b Use (4a) to show that if X_i is chi-squared random variable with n_i degrees of freedom and non-centrality parameter c_i, i = 1, 2, and X_1, X_2 are independent, then X_1+X_2 is a chi-square random variable with n_1+n_2 degrees of freedom and non-centrality parameter c_1+c_2. Find the mean and variance of X_1+X_2.

4 Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be independent random samples from $N(\mu_1, v_1^2)$ and $N(\mu_2, v_2^2)$ respectively. Let $\bar{X}$, $\bar{Y}$, S_1^2 and S_2^2 denote the respective sample means and variances, and let c be a fixed constant. For (a)-(d), identify the distribution of each of the following statistics by finding a suitable value of k:

(a) $\sqrt{n_2}\,(\bar{Y}-c)/kv_2$; (b) $k\sqrt{n_1}\,(\bar{X}-c)/S_1$; (c) $k(X_1+X_2)/|Y_1-Y_2|$ and (d) $k\{(X_1-c)^2 + (X_2-c)^2\}/S_2^2$.

5 Let (X_j, Y_j), j = 1, 2, …, n be a random sample from the bivariate normal distribution with parameters m_1, m_2, v_1^2, v_2^2 and correlation r. If d is a fixed constant, find a constant k so that

$$T = \frac{k(\bar{X} - \bar{Y} - d)}{\{\sum_{i=1}^{n}(X_i - Y_i - \bar{X} + \bar{Y})^2\}^{1/2}}$$

has a non-central t distribution with m degrees of freedom and non-centrality parameter s. Express m and s as a function of the parameters and the constant d. What is the expectation of T?

6 Find the expectation of a non-central F-distribution with numerator and denominator degrees of freedom n and m, respectively, and non-centrality parameter Φ.

P1:

1 Suppose Y is a p-dimensional multivariate normal distribution with mean m and covariance matrix V. Let Y be partitioned into Y_1 and Y_2 so that $Y' = (Y_1', Y_2')$ and the dimension of Y_1 and Y_2 are, respectively, p_1 and p_2 with $p_1+p_2=p$. Find the distribution of Y_1 given Y_2 if V is singular and express the conditional distribution in terms of only m, p_1 and the 4 submatrices in the partitioned matrix V and an appropriate g-inverse.

Sol. $Y \sim \mathcal{N}_p(m, V)$, $m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$, $V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$

Let $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} I & -V_{12}V_{22}^{+} \\ 0 & I \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$

where V_{22}^{+} is the Moore-Penrose inverse.

Then we have

$$\mathrm{Var}\, Z = \begin{pmatrix} I & -V_{12}V_{22}^{+} \\ 0 & I \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \begin{pmatrix} I & -V_{12}V_{22}^{+} \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} V_{11} - V_{12}V_{22}^{+}V_{21} & V_{12} - V_{12}V_{22}^{+}V_{22} \\ V_{21} - V_{21}V_{22}^{+}V_{22} & V_{22} \end{pmatrix}$$

Lemma: $V_{12} - V_{12}V_{22}^{-}V_{22} = 0$ & $V_{21} - V_{22}V_{22}^{-}V_{21} = 0$. ($V_{22}^{-}$ is any g-inv)

Pf: $V \geq 0$ (maybe singular)

$\Rightarrow B^T V B \geq 0 \quad \forall B \in \mathbb{R}^{p \times k}$

Let $B^T = (\ O\ , I - V_{22}^{-}V_{22})$, then

$B^T V B = O$ & $VB = \begin{pmatrix} V_{12} - V_{12}V_{22}^{-}V_{22} \\ O \end{pmatrix}$

$\Rightarrow (B^T V^{\frac{1}{2}})(V^{\frac{1}{2}} B) = 0$ ($V \geq 0 \Rightarrow V^{\frac{1}{2}} \exists$)

$\Rightarrow V^{\frac{1}{2}} B = 0 \Rightarrow VB = 0$

$\Rightarrow V_{12} - V_{12}V_{22}^{-}V_{22} = 0$

Similarly, $B^T V = 0 \Rightarrow V_{21} - V_{22}V_{22}^{-}V_{21} = 0$

$$\Rightarrow \mathrm{Var}\, Z = \begin{pmatrix} V_{11} - V_{12}V_{22}^{+}V_{21} & O \\ O & V_{22} \end{pmatrix}$$

Thus, $Z_1 \perp\!\!\!\perp Z_2$

$\Rightarrow Z_1 \mid Z_2 = z_2 \sim \mathcal{N}_{p_1}(m_1 - V_{12}V_{22}^{+}m_2, V_{11} - V_{12}V_{22}^{+}V_{21})$

$\Rightarrow Y_1 \mid Y_2 = y_2 \sim \mathcal{N}_{p_1}(m_1 + V_{12}V_{22}^{+}(y_2 - m_2), V_{11} - V_{12}V_{22}^{+}V_{21})$

Note: in fact, V_{22}^{+} does not have to be Moore-Penrose inv., it can be the g-inv. s.t.

$$V_{22}V_{22}^{-}V_{22} = V_{22}$$
$$\&$$
$$V_{22}^{-}V_{22}V_{22}^{-} = V_{22}^{-}.$$

P_2:

15. Suppose that $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)' \sim N_4(\mathbf{0}, \mathbf{I}_4)$, and let $Q = Y_1Y_2 - Y_3Y_4$.

    (a) Prove that Q does not have a chi-square distribution.

    (b) Find the m.g.f. of Q.

Sol. $Q = Y^T A Y$ where $A = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 0 \end{pmatrix}$

Since $\begin{vmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{vmatrix} < 0$ so

A is not p.s.d. $\Rightarrow Q \not\sim \chi^2$ by the fundamental theorem.

MGF of Q:

$$\mathbb{E}e^{tQ} = \mathbb{E}e^{tY^TAY} \quad \text{(law of SILLY statisticians)}$$

$$= \int_{\mathbb{R}^4} e^{ty^TAy}\left(ke^{-\frac{1}{2}y^Ty}\right)\lambda(dy)$$

$$= \int_{\mathbb{R}^4} k\, e^{-\frac{1}{2}y^T(I-2tA)y}\lambda(dy)$$

(plug-in $k = \frac{1}{(2\pi)^2}$)

$$= \det|I-2tA|^{-\frac{1}{2}}$$

$$= \frac{1}{\sqrt{\begin{vmatrix} 1 & -t & & \\ -t & 1 & & \\ & & 1 & t \\ & & t & 1 \end{vmatrix}}} = \frac{1}{\sqrt{\begin{vmatrix} 1 & -t \\ -t & 1 \end{vmatrix}\begin{vmatrix} 1 & t \\ t & 1 \end{vmatrix}}}$$

$$= \frac{1}{1-t^2}, \quad -1 < t < 1.$$

P_3:

3a (Define a non-central chi-square distribution by its density). Let p_m be the density of a central chi-square distribution with m degrees of freedom, and for each nonnegative s, let $q_j = \exp(-s/2)\,(s/2)^j/j!$. For s=0, $q_0 = 1$ and $q_j = 0$ for $j > 0$. A random variable with density

poisson weight

$$h(z) = \sum_{j=0}^{\infty} q_j\, p_{m+2j}(z), \qquad z > 0$$

is said to have a non-central chi-square distribution with m degrees of freedom and non-centrality parameter s. Use the power series expansion of exp(x) to find the moment generating function of such a random variable.

3b Use (3a) to show that if X_i is chi-squared random variable with n_i degrees of freedom and non-centrality parameter c_i, i = 1, 2, and X_1, X_2 are independent, then X_1+X_2 is a chi-square random variable with n_1+n_2 degrees of freedom and non-centrality parameter c_1+c_2. Find the mean and variance of X_1+X_2.

Sol. (a): $Z \sim \chi^2_m(s)$

$$\mathbb{E}e^{tZ} = \int_{\mathbb{R}^+} e^{tz}\sum_{j=0}^{\infty} q_j P_{m+2j}(z)\,dz$$

(monotone convergence thm)

$$= \sum_{j=0}^{\infty} q_j \int_{\mathbb{R}^+} e^{tz} P_{m+2j}(z)\,dz$$

$$= \sum_{j=0}^{\infty} q_j \frac{1}{(1-\beta t)^{\alpha_j}}, \quad \beta = 2, \ \alpha_j = \frac{m}{2} + j$$

$$= \sum_{j=0}^{\infty} \frac{e^{-\frac{s}{2}}}{(\frac{s}{2})^j j!} \frac{1}{(1-2t)^j}(1-2t)^{-\frac{m}{2}}$$

$$= e^{\frac{st}{1-2t}}(1-2t)^{-\frac{m}{2}}, \quad t < \frac{1}{2}$$

(b): $X_i \sim \chi^2_{n_i}(c_i)$, $i = 1, 2$. $\Rightarrow$

$$\mathbb{E}e^{tX_i} = e^{\frac{c_i t}{1-2t}}(1-2t)^{-\frac{n_i}{2}}, \quad i = 1, 2. \Rightarrow$$

$$\mathbb{E}e^{t(X_1+X_2)} \overset{\text{(indept.)}}{=} \mathbb{E}e^{tX_1}\,\mathbb{E}e^{tX_2}$$

$$= e^{\frac{(c_1+c_2)t}{1-2t}}(1-2t)^{-\frac{n_1+n_2}{2}} \Rightarrow$$

$$X_1 + X_2 \sim \chi^2_{n_1+n_2}(c_1+c_2)$$

P4:

4 Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be independent random samples from $N(\mu_1, v_1^2)$ and $N(\mu_2, v_2^2)$ respectively. Let $\bar{X}$, $\bar{Y}$, S_1^2 and S_2^2 denote the respective sample means and variances, and let c be a fixed constant. For (a)-(d), identify the distribution of each of the following statistics by finding a suitable value of k:

(a) $\sqrt{n_2}(\bar{Y}-c)/kv_2$; (b) $k\sqrt{n_1}(\bar{X}-c)/S_1$; (c) $k(X_1+X_2)/|Y_1-Y_2|$ and (d) $k\{(X_1-c)^2+(X_2-c)^2\}/S_2^2$.

Sol. (a): $\bar{Y} \sim N(\mu_2, \frac{v_2^2}{n_2}) \Rightarrow$

$$\sqrt{n_2}(\bar{Y}-c) \sim N(\sqrt{n_2}(\mu_2-c), v_2^2) \Rightarrow$$

$$\frac{\sqrt{n_2}(\bar{Y}-c)}{kv_2} \sim N\left(\frac{\sqrt{n_2}(\mu_2-c)}{kv_2}, \frac{1}{k}\right).$$

with $k=\sqrt{n_2}$ we know RHS $\stackrel{d}{=} N\left(\frac{\mu_2-c}{v_2}, \frac{1}{\sqrt{n_2}}\right)$

(b): From 202A we know that

$$\frac{\bar{x}-\mu_1}{\sqrt{\frac{S_1^2}{n_1}}} \sim t_{n_1-1}(0)$$

Thus, Let $k=1$ & we know

$$\frac{\sqrt{n_1}(\bar{x}-c)}{S_1} = \frac{\sqrt{n_1}(\bar{x}-c)/v_1}{\sqrt{(n_1-1)S_1^2/v_1^2}\sqrt{\frac{1}{n_1-1}}} \quad \& \ \bar{x} \perp\!\!\!\perp S_1$$

Denominator $\sim \sqrt{\chi^2_{n_1-1}(0)/n_1-1}$

Numerator $\sim N\left(\frac{\sqrt{n_1}(\mu_1-c)}{v_1}, 1\right)$

$$\Rightarrow \frac{\sqrt{n_1}(\bar{X}-c)}{S_1} \sim t_{n_1-1}\left(\frac{\sqrt{n_1}(\mu_1-c)}{v_1}\right)$$

(c): $X_1+X_2 \sim N(\mu_1+\mu_2, 2v_1^2)$

$$Y_1-Y_2 \sim N(0, 2v_2^2)$$

$$\Rightarrow \frac{X_1+X_2}{|Y_1-Y_2|} = \frac{(X_1+X_2)/\sqrt{2v_1^2}}{\sqrt{(Y_1-Y_2)^2/2v_2^2}} \frac{v_1}{v_2} \Rightarrow$$

Numerator $\sim N((\mu_1+\mu_2)/\sqrt{2}v_1, 1)$ $\Rightarrow$

Denominator^2 $\sim \chi_1^2$ & $\perp\!\!\!\perp$

$$\frac{X_1+X_2}{|Y_1-Y_2|} \sim \frac{v_1}{v_2} t_1\left(\frac{\mu_1+\mu_2}{\sqrt{2}v_1}\right), \text{ Let } k=\frac{v_2}{v_1}$$

$$\Rightarrow \frac{v_2}{v_1}\frac{X_1+X_2}{|Y_1-Y_2|} \sim t_1\left(\frac{\mu_1+\mu_2}{\sqrt{2}v_1}\right)$$

(d): $(X_1-c)^2/v_1^2 \sim \chi_1^2((\mu_1-c)^2)$

$$(X_2-c)^2/v_1^2 \sim \chi_1^2((\mu_1-c)^2)$$

$$\frac{(n_2-1)S_2^2}{v_2^2} \sim \chi^2_{(n_2-1)}(0)$$

Let $k=\left(\frac{v_2^2}{v_1^2}\right)$, then

$$\frac{v_2^2}{v_1^2}\frac{(X_1-c)^2+(X_2-c)^2}{S_2^2} \sim F_{2,n_2-1}(2(\mu_1-c)^2)$$

P5:

5 Let (X_j, Y_j), $j = 1, 2, \ldots, n$ be a random sample from the bivariate normal distribution with parameters m_1, m_2, v_1^2, v_2^2 and correlation r. If d is a fixed constant, find a constant k so that

$$T = \frac{k(\bar{X} - \bar{Y} - d)}{\{\sum_{i=1}^{n}(X_i - Y_i - \bar{X} + \bar{Y})^2\}^{1/2}}$$

has a non-central t distribution with m degrees of freedom and non-centrality parameter s. Express m and s as a function of the parameters and the constant d. What is the expectation of T?

Sol. $(X_j, Y_j) \sim \mathcal{N}\left(\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{pmatrix} v_1^2 & r v_1 v_2 \\ r v_2 v_1 & v_2^2 \end{pmatrix}\right)$

$$\bar{X} - \bar{Y} - d \sim \mathcal{N}\left(m_1 - m_2 - d, \frac{v_1 + v_2 - 2 r v_1 v_2}{n}\right)$$

Note that $X_i - Y_i \sim \mathcal{N}(m_1 - m_2, v_1 + v_2 - 2 r v_1 v_2)$

Let $Z_i = X_i - Y_i$ & $Z_i \perp\!\!\!\perp Z_j$ $\forall\ 1 \le i \ne j \le n$.

Then $\sum_{i=1}^{n}(X_i - Y_i - \bar{X} + \bar{Y})^2$

$$= \sum_{i=1}^{n} Z_i^2 - n(\bar{Z})^2, \quad \bar{Z} = \frac{1}{n}\sum_i Z_i$$

$$= Z^T\left(I - \frac{11^T}{n}\right)Z$$

But $Z^T Z = Z^T\left(I - \frac{11^T}{n}\right)Z + Z^T \frac{11^T}{n} Z$

$\Rightarrow$ By Fisher-Cochran, (General Version. The proof is the same.)

$cZ^T\left(I - \frac{11^T}{n}\right)Z \sim \chi^2_{n-1}(\delta^2)$ where

$$\delta^2 = (m_1 - m_2)1^T\left(I - \frac{11^T}{n}\right)1(m_1 - m_2)$$

$= 0$, and $c = (v_1 + v_2 - 2 r v_1 v_2)^{-1}$

More importantly,

$$\bar{X} - \bar{Y} = \bar{Z} = \frac{1}{n}1^T Z, \ \&$$

$$\mathrm{Cov}\left(1^T Z, \left(I - \frac{11^T}{n}\right)Z\right) = 0 \Rightarrow$$

$$\bar{Z} \perp\!\!\!\perp Z^T\left(I - \frac{11^T}{n}\right)Z$$

Thus, $T = \dfrac{k(\bar{Z} - d)}{\left(Z^T\left(I - \frac{11^T}{n}\right)Z\right)^{\frac{1}{2}}}$

Let $k = \sqrt{n(n-1)}$

we have $T \sim t_{n-1}(m)$

$m = \sqrt{n}(m_1 - m_2 - d) / \sqrt{v_1 + v_2 - 2 r v_1 v_2}$.

Let $\tilde{T} = \dfrac{\mathcal{N}(m, 1)}{\sqrt{\chi^2_{n-1}(0)/(n-1)}}$ & $\mathcal{N}(m, 1) \perp\!\!\!\perp \chi^2_{n-1}(0)$

Then

$$\tilde{T} \overset{d}{=} T.$$

$$\mathbb{E}\tilde{T} = \sqrt{n-1}\, m\, \mathbb{E}\left(\left(\chi^2_{n-1}(0)\right)^{-\frac{1}{2}}\right)$$

$$= \sqrt{n-1}\, m \int \frac{1}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}}} x^{\frac{n-1}{2} - 1} e^{-\frac{x}{2}} \cdot x^{-\frac{1}{2}}\, dx$$

$$= \sqrt{\frac{n-1}{2}}\, m \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

P6:

6 Find the expectation of a non-central F-distribution with numerator and denominator degrees of freedom n and m, respectively, and non-centrality parameter Φ.

Sol. $X \sim F_{n,m}(\phi) \overset{d}{=} \frac{X_1}{X_2} \frac{m}{n}$ where

$X_1 \sim \chi^2_n(\phi)$, $X_2 \sim \chi^2_m(0)$, $X_1 \perp\!\!\!\perp X_2$

$$\Rightarrow \mathbb{E}X = \mathbb{E}X_1 \, \mathbb{E}\left(\frac{1}{X_2}\right) \frac{m}{n}$$

$$= \frac{m}{2n} \frac{\Gamma(\frac{m}{2}-1)}{\Gamma(\frac{m}{2})} \mathbb{E}X_1$$

By 3a, we have the pdf for X_1:

$$f_{X_1}(x) = \sum_{j=0}^{\infty} g_j \, p_{n+2j}(x),$$

by Monotone Convergence thm,

$$\mathbb{E}X_1 = \sum_{j=0}^{\infty} g_j (n+2j) = n + 2\sum_{j=0}^{\infty} j g_j = n+\phi$$

(or $\mathbb{E}X_1 = \mathbb{E}\,\mathrm{Tr}(Y^T Y) = \mathrm{Tr}(\mu\mu^T + I) = n+\phi$)

Also, $\frac{\Gamma(\frac{m}{2}-1)}{\Gamma(\frac{m}{2})} = \frac{2}{m-2} \Rightarrow$

$$\mathbb{E}X = \frac{m}{n} \frac{n+\phi}{m-2} \quad \square.$$