STAT 30100 : Mathematical Statistics-1

Winter 2020

Homework 3

Solutions by

Jinhong Du

12243476

STAT 30100, Homework 3

Problems 1-6 are based on the handout "Distribution of Quadratic Forms", and problem 7 is based on Ferguson chapters 5-7.

1. Prove Theorem 2 on the handout "Distribution of Quadratic Forms".

   Theorem 2: Let $X \sim N_p(0, \Sigma)$, $Q = X^\top AX$ (A symmetric). Then $Q \sim \chi^2_r(0)[\equiv \chi^2_r]$ iff $\Sigma(A\Sigma A - A)\Sigma = 0$ and $r = \mathrm{tr}(A\Sigma)$.

   Proof. Suppose that $\mathrm{rank}(\Sigma) = m \le p$. Since $X \sim N_p(0, \Sigma)$, there exists $B \in \mathbb{R}^{p\times m}$ and $Z \sim N_m(0, I_m)$ such that $X = BZ$ and $\Sigma = BB^\top$. Then $Q = Z^\top(B^\top AB)Z$ and $B^\top AB$ is symmetric.
   $\Longleftarrow$ Since $\Sigma(A\Sigma A - A)\Sigma = 0$, we have $B(B^\top AB)(B^\top AB)B^\top = B(B^\top AB)B^\top$. Thus, $B^\top B(B^\top AB)(B^\top AB)B^\top B = B^\top B(B^\top AB)B^\top B$. Since B has full column rank, $B^\top B \in \mathbb{R}^{m\times m}$ has full rank and invertible ($B^\top Bx = 0 \Longrightarrow x^\top B^\top Bx = \|Bx\|_2 = 0 \Longrightarrow Bx = 0 \Longrightarrow \mathrm{Null}(B) \subset \mathrm{Null}(B^\top B)$). So $(B^\top AB)(B^\top AB) = B^\top AB$, i.e., $B^\top AB$ is idempotent.
   $r = \mathrm{tr}(A\Sigma) = \mathrm{tr}(ABB^\top) = \mathrm{tr}(B^\top AB) = \sum_{i=1}^m \lambda_i(B^\top AB)$. Since $\lambda_i(B^\top AB)$ can only be 0 or 1 for the idempotent matrix $B^\top AB$, we have $r = \mathrm{rank}(B^\top AB)$.
   Therefore, by Theorem 1, we have $Q \sim \chi^2_r(0)$.
   $\Longrightarrow$ Since $Q = Z^\top(B^\top AB)Z \sim \chi^2_r$, from Theorem 1 we have $B^\top AB$ is idempotent and $\mathrm{rank}(B^\top AB) = r$.
   Since $B^\top AB$ is idempotent, we have $(B^\top AB)(B^\top AB) = B^\top AB$ and $B(B^\top AB)(B^\top AB)B^\top = B(B^\top AB)B^\top$, i.e., $\Sigma(A\Sigma A - A)\Sigma = 0$.
   Since $B^\top AB$ is idempotent, we have $r = \mathrm{rank}(B^\top AB) = \mathrm{tr}(B^\top AB) = \mathrm{tr}(ABB^\top) = \mathrm{tr}(A\Sigma)$. □

2. Extend the Fisher-Cochran Theorem to the case $\sum_{i=1}^k A_i = A$ (instead of $\sum_{i=1}^k A_i = I$), where A is idempotent but not necessarily I and not necessarily full rank. Make any other changes needed to make the theorem true and prove it.

   Let $X \sim N_n(\mu, I)$. Let $Q_i = X^\top A_i X$, $i = 1, \ldots, k$, be k quadratic forms such that $X^\top AX = \sum_{i=1}^k Q_i$, where $A = \sum_{i=1}^k A_i$ is idempotent. Let $r_i = \mathrm{rank}(A_i)$ and $r = \mathrm{rank}(A)$. Then $Q_i \sim \chi^2_{r_i}(\lambda_i)$ with $\lambda_i = \mu^\top A_i\mu$ and the Q_i's are mutually independent iff $\sum_{i=1}^k r_i = r$.

   Proof. Since $Q_i = X^\top A_i X$ ($i = 1, \ldots, k$) are quadratic forms, $A_1, \ldots, A_k$ are symmetric. So A is symmetric. Since A is idempotent, from Theorem 1 we have $X^\top AX \sim \chi^2_r(\delta)$ where $\delta = \mu^\top A\mu$.
   Let $A = U\Sigma U^\top$ be the eigen-decomposition of A, where $U \in \mathbb{R}^{n\times n}$ is orthonormal and $\Sigma \in \mathbb{R}^{n\times n}$ is diagonal with digonal entries $\sigma_1, \ldots, \sigma_n \in \{0, 1\}$. Without loss of generality, suppose that $\sigma_1 = \cdots = \sigma_r = 1$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$. Define $Y = U_r^\top X \sim N_n(U_r^\top\mu, I_r)$ where $U_r \in \mathbb{R}^{n\times r}$ is the matrix formed by concatenating first r columns of U. Then $X^\top AX = Y^\top Y$. Let $B_i = U_r^\top A_i U_r$, then $\mathrm{rank}(B_i) = \mathrm{rank}(A_i) = r_i$ and $\sum_{i=1}^k B_i = U_r^\top AU_r = I_r$.
   From the Fisher-Cochran Theorem, we have $Q_i \sim \chi^2_{r_i}(\lambda_i)$ with $\lambda_i = (U_r^\top\mu)^\top(U_r A_i U_r^\top)(U_r^\top\mu) = \mu^\top A_i\mu$ and the Q_i's are mutually independent if and only if $\sum_{i=1}^k r_i = r$.

   □

3. Prove Theorem 4 on the handout "Distribution of Quadratic Forms". You may use any of the results listed above Theorem 4 on the handout, but do not use any of the results listed below Theorem 4 on the handout.

   Theorem 4: Let $X \sim N_n(\mu, I)$. Let $Q_1 = X^\top A_1 X$, $Q_2 = X^\top A_2 X$. Suppose Q_1 and Q_2 each have a (noncentral) chi-squared distribution. Then Q_1 and Q_2 are independent iff $A_1 A_2 = 0$.

Proof. Suppose that $Q_1 \sim \chi^2_{r_1}(\delta_1)$ and $Q_2 \sim \chi^2_{r_2}(\delta_2)$. Q_1 and Q_2 each have a (noncentral) chi-squared distribution, by Theorem 1 we have A_1 and A_2 are idempotent.
$\Longrightarrow$ Since Q_1 and Q_2 are independent, by Lemma 2 we have $Q_1+Q_2 = X^\top(A_1+A_2)X \sim \chi^2_{r_1+r_2}(\delta_1+\delta_2)$. So by Theorem 1, we have $A_1 + A_2$ is idempotent. So

$$A_1 + A_2 = (A_1 + A_2)^2 = A_1^2 + A_1A_2 + A_2A_1 + A_2^2 = A_1 + A_2 + A_1A_2 + A_2A_1,$$

which implies $A_1A_2 + A_2A_1 = 0$. Since

$$A_1(A_1A_2 + A_2A_1) = A_1A_2 + A_1A_2A_1 = 0$$
$$(A_1A_2 + A_2A_1)A_1 = A_1A_2A_1 + A_2A_1 = 0,$$

by subtracting the latter from the former, we get $A_1A_2 - A_2A_1 = 0$, i.e. $A_1A_2 = A_2A_1$. Therefore, $A_1A_2 = A_2A_1 = 0$.
$\Longleftarrow$ Since

$$A_1(A_1A_2 + A_2A_1) = A_1A_2 + A_1A_2A_1 = 0$$
$$(A_1A_2 + A_2A_1)A_1 = A_1A_2A_1 + A_2A_1 = 0,$$

by subtracting the latter from the former, we get $A_1A_2 - A_2A_1 = 0$. So $A_2A_1 = A_1A_2 = 0$. Thus,

$$(A_1 + A_2)^2 = A_1^2 + A_1A_2 + A_2A_1 + A_2^2 = A_1 + A_2,$$

i.e. $A_1 + A_2$ is idempotent. Since $X^\top(A_1 + A_2)X = Q_1 + Q_2$ and $\mathrm{tr}(A_1 + A_2) = \mathrm{tr}(A_1) + \mathrm{tr}(A_2)$, by extension of Fisher-Cocharn Theorem (Problem 2), we have that Q_1 and Q_2 are independent. $\square$

4. Prove Theorem 7 on the handout "Distribution of Quadratic Forms". You may assume all the other theorems on the handout.

   Theorem 7: Let $X \sim N_n(\mu, I)$, $Q_i = X^\top A_i X$, $i = 1, \ldots, k$, $X^\top X = \sum_{i=1}^k Q_i$, and $r_i = r(A_i)$. Then the following are equivalent:

   1. The Q_i's are mutually independent.
   2. Each $Q_i \sim \chi^2_{r_i}(\lambda_i)$ with $\lambda_i = \mu^\top A_i \mu$.
   3. The A_i's are idempotent.
   4. $A_iA_j = 0$, for all $i \neq j$.
   5. $r_1 + r_2 + \cdots + r_k = n$.

Proof. ($5 \Longrightarrow 1$) By Fisher-Cochran Theorem.
($1 \Longleftrightarrow 4$) By Craig's Theorem.

Solution (cont.)

($4 \implies 3$) Since $A_iA_j = 0$, for all $i \neq j$, we have $A_i = A_iI = A_i \sum_{j=1}^k A_j = \sum_{j=1}^k A_iA_j = A_i^2$. Therefore, A_i is idempotent for all i.

($3 \iff 2$) By Theorem 1, we have for $i = 1, \ldots, k$, $Q_i \sim \chi^2_{r_i}(\lambda_i)$ with $\lambda_i = \mu^\top A_i \mu$ if and only if A_i is idempotent.

($3 \implies 5$) Since A_i's are idempotent, we have

$$\sum_{i=1}^k r_i = \sum_{i=1}^k \sum_{j=1}^n \lambda_j(A_i) = \sum_{i=1}^k \operatorname{tr}(A_i) = \operatorname{tr}\left(\sum_{i=1}^k A_i\right) = \operatorname{tr}(I) = n,$$

where $\lambda_j(A_i) \in \{0,1\}$ is the jth eigenvalue of the idempotent matrix A_i.

Now since we have proved that any statement can imply any other statement, these five statements are equivalent. $\square$

5. Let $Y \sim N_n(\mu, \sigma^2 I)$, and let A be $n \times p$ with $\operatorname{rank}(A) = p < n$. Find the distributions of $Q_1 = \sigma^{-2} Y^\top A(A^\top A)^{-1} A^\top Y$ and $Q_2 = \sigma^{-2} Y^\top Y - Q_1$, and show they are independent.

Proof. Let $H = A(A^\top A)^{-1}A^\top$, then H is symmetric. Since $H^2 = A(A^\top A)^{-1}A^\top A(A^\top A)^{-1}A^\top = A(A^\top A)^{-1}A^\top = H$, we have that H is idempotent. Also, $\operatorname{rank}(H) = \operatorname{tr}(H) = \operatorname{tr}(A(A^\top A)^{-1}A^\top) = \operatorname{tr}(A^\top A(A^\top A)^{-1}) = \operatorname{tr}(I_p) = p$, i.e., H is full rank. By Theorem 3, we have $Q_1 \sim \chi^2_p(\delta_1)$ where $\delta_1 = \sigma^{-2}\mu^\top H \mu$.

Since $(I-H)^\top = I-H$, $I-H$ is symmetric. Since $(I-H)^2 = I - 2H + H^2 = I - H$, $I-H$ is idempotent. Also, $\operatorname{rank}(I-H) = \operatorname{tr}(I-H) = \operatorname{tr}(I) - \operatorname{tr}(H) = n - p$. By Theorem 3, we have $Q_2 \sim \chi^2_{n-p}(\delta_2)$ where $\delta_1 = \sigma^{-2}\mu^\top (I-H)\mu$.

Notice that $\sigma^{-1}Y \sim N_n(\sigma^{-1}\mu, I_n)$. Since $p + (n-p) = n$ and $Y^\top(\sigma^{-2}I)Y = Q_1 + Q_2$, by Problem 2, we have Q_1 and Q_2 are independent. $\square$

6. **Simple linear regression:** Let $Y \sim N_n(\mu, \sigma^2 I)$ with $\mu_i = \alpha + \beta(x_i - \overline{x})$, $i = 1, \ldots, n$, $n > 2$, where α, β, and the x_is are constants and $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$. Let $S_X^2 = \sum_{i=1}^n (x_i - \overline{x})^2$, $S_Y^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2$, $S_{XY} = \sum_{i=1}^n (x_i - \overline{x})Y_i$, and $b = \frac{S_{XY}}{S_X^2}$. Let $Q_1 = n\overline{Y}^2 + b^2 S_X^2$, $Q_2 = S_Y^2 - b^2 S_X^2$. Show that $\sigma^{-2}Q_1$ and $\sigma^{-2}Q_2$ are independent chi-square distributed with 2 and $n-2$ d.f., respectively. [Hint: Apply problem 5.]

Proof. Let $X = \left(x_1, \ldots, x_n\right)^\top \in \mathbb{R}^n$ and $\overline{X} = \overline{x}\mathbf{1}$. Since $S_Y^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2 = Y^\top(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top)Y$ and $n\overline{Y}^2 = Y^\top\left(\frac{1}{n}\mathbf{1}\mathbf{1}^\top\right)Y$, we have $\sigma^{-2}(Q_1 + Q_2) = \sigma^{-2}(n\overline{Y}^2 + S_Y^2) = \sigma^{-2}(Y^\top Y) \sim \chi^2_n(\sigma^{-2}\mu^\top\mu)$.

Since

$$\begin{aligned} S_{XY}^2 &= Y^\top(X - \overline{X})(X - \overline{X})^\top Y \\ S_X^2 &= (X - \overline{X})^\top(X - \overline{X}), \end{aligned}$$

Solution (cont.)

we have

$$Q_1 = n\overline{Y}^2 + b^2 S_X^2 = n\overline{Y}^2 + \frac{S_{XY}^2}{S_X^2} = Y^\top \left(\frac{1}{n}\mathbf{1}\mathbf{1}^\top + (X-\overline{X})[(X-\overline{X})^\top (X-\overline{X})]^{-1}(X-\overline{X})^\top \right) Y.$$

Let $H = \frac{1}{n}\mathbf{1}\mathbf{1}^\top + (X-\overline{X})[(X-\overline{X})^\top(X-\overline{X})]^{-1}(X-\overline{X})^\top$. Since $H^\top = \frac{1}{n}\mathbf{1}\mathbf{1}^\top + (X-\overline{X})[(X-\overline{X})^\top(X-\overline{X})]^{-1}(X-\overline{X})^\top = H$, H is symmetric. Since

$$\begin{aligned} H^2 &= \frac{1}{n}\mathbf{1}\mathbf{1}^\top \cdot \frac{1}{n}\mathbf{1}\mathbf{1}^\top + \frac{1}{n}\mathbf{1}\mathbf{1}^\top (X-\overline{X})[(X-\overline{X})^\top(X-\overline{X})]^{-1}(X-\overline{X})^\top + \\ &\quad (X-\overline{X})[(X-\overline{X})^\top(X-\overline{X})]^{-1}(X-\overline{X})^\top \cdot \frac{1}{n}\mathbf{1}\mathbf{1}^\top \\ &\quad (X-\overline{X})[(X-\overline{X})^\top(X-\overline{X})]^{-1}(X-\overline{X})^\top (X-\overline{X})[(X-\overline{X})^\top(X-\overline{X})]^{-1}(X-\overline{X})^\top \\ &= \frac{1}{n^2}\mathbf{1}(\mathbf{1}^\top\mathbf{1})\mathbf{1}^\top + 0 + 0 + (X-\overline{X})[(X-\overline{X})^\top(X-\overline{X})]^{-1}(X-\overline{X})^\top \\ &= H \end{aligned}$$

by noticing that $\mathbf{1}^\top(X-\overline{X}) = (X-\overline{X})^\top\mathbf{1} = 0$, H is idempotent.

$$\begin{aligned} \operatorname{rank}(H) = \operatorname{tr}(H) &= \operatorname{tr}\left(\frac{1}{n}\mathbf{1}\mathbf{1}^\top\right) + \operatorname{tr}((X-\overline{X})[(X-\overline{X})^\top(X-\overline{X})]^{-1}(X-\overline{X})^\top) \\ &= \operatorname{tr}(\frac{1}{n}\mathbf{1}^\top\mathbf{1}) + \operatorname{tr}((X-\overline{X})^\top(X-\overline{X})[(X-\overline{X})^\top(X-\overline{X})]^{-1}) \\ &= 2. \end{aligned}$$

By Theorem 3, we have $\sigma^{-2}Q_1 \sim \chi_2^2(\delta_1)$ where $\delta_1 = \sigma^{-2}\mu^\top H\mu$. By Problem 5, we have $Q_2 = (Q_1+Q_2) - Q_1 \sim \chi_{n-2}^2(\delta_2)$ where $\delta_2 = \sigma^{-2}\mu^\top(I-H)\mu$, Q_1 and Q_2 are independent.

□

7. Let $X_1, X_2, \ldots$ be i.i.d. double exponential (Laplace) distributed random variables with density $f(x) = \frac{1}{2\tau}\exp\{-\frac{|x|}{\tau}\}$ for $-\infty < x < \infty$, where the parameter $\tau > 0$ represents the mean absolute deviation, $\tau = \mathbb{E}|X|$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n |X_i|$.

   (a) Find the joint asymptotic distribution of $\overline{X}_n$ and $\overline{Y}_n$.

Since $\frac{1}{2\tau}e^{-\frac{|x|}{\tau}}x$ and $\frac{1}{2\tau}e^{-\frac{|x|}{\tau}}x|x|$ are odd functions, we have

$$\begin{aligned} \mathbb{E}(X_i) &= \int_{\mathbb{R}} \frac{1}{2\tau}e^{-\frac{|x|}{\tau}} x \mathrm{d}x = 0 \\ \mathbb{E}(X_i \cdot |X_i|) &= \int_{\mathbb{R}} \frac{1}{2\tau}e^{-\frac{|x|}{\tau}} x|x| \mathrm{d}x = 0 \\ \operatorname{Cov}(X_i \cdot |X_i|) &= \mathbb{E}(X_i|X_i|) - \mathbb{E}(X_i)\cdot\mathbb{E}|X_i| = 0. \end{aligned}$$

Solution (cont.)

Also,

$$\begin{aligned}
\mathbb{E}(|X_i|) &= \int_{\mathbb{R}} \frac{1}{2\tau} e^{-\frac{|x|}{\tau}} |x| \mathrm{d}x \\
&= 2\int_0^{+\infty} \frac{1}{2\tau} e^{-\frac{x}{\tau}} x \mathrm{d}x \\
&= -\frac{\tau^2}{2} e^{-x}(x+1)\Big|_0^{+\infty} \\
&= \tau \\
\mathrm{Var}(X_i) &= \int_{\mathbb{R}} \frac{1}{2\tau} e^{-\frac{|x|}{\tau}} x^2 \mathrm{d}x \\
&= \int_0^{+\infty} \frac{1}{2\tau} e^{-\frac{x}{\tau}} x^2 \mathrm{d}x + \int_{-\infty}^0 \frac{1}{2\tau} e^{\frac{x}{\tau}} x^2 \mathrm{d}x \\
&= -\frac{\tau^2}{2} e^{-x}(x^2+2x+2)\Big|_0^{+\infty} + \frac{\tau^2}{2} e^{-x}(x^2+2x+2)\Big|_{-\infty}^0 \\
&= 2\tau^2 \\
\mathrm{Var}(|X_i|) &= \mathbb{E}(|X_i|^2) - [\mathbb{E}(X_i)]^2 = \mathrm{Var}(X_i) - \tau^2 = \tau^2
\end{aligned}$$

Let $Z_n = \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ and $\overline{Z} = \begin{pmatrix} \overline{X}_n \\ \overline{Y}_n \end{pmatrix}$. Then $\mu = \mathbb{E}(Z_1) = \begin{pmatrix} 0 \\ \tau \end{pmatrix}$. By central limit theorem. $\sqrt{n}(\overline{Z}_n - \mu) \xrightarrow{D} N(\mathbf{0}_{2\times 1}, \Sigma)$, where

$$\Sigma = \mathrm{Cov}(Z_1) = \begin{pmatrix} 2\tau^2 & 0 \\ 0 & \tau^2 \end{pmatrix}.$$

(b) Find the asymptotic distribution of $\frac{\overline{Y}_n - \tau}{\overline{X}_n}$.

From (a), we have $\frac{1}{\sqrt{2}\tau}\sqrt{n}\overline{X}_n \xrightarrow{D} N(0,1) \triangleq Z_1$, $\frac{1}{\tau}\sqrt{n}(\overline{Y}_n - \tau) \xrightarrow{D} N(0,1) \triangleq Z_2$ and the two asymptotic distributions are independent. Define $g : \mathbb{R}^2 \mapsto \mathbb{R}$ such that $g\left(\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)\right) = \frac{y}{x}$ for $x, y \in \mathbb{R}$. Then $\mathbb{P}(\{x \in \mathbb{R} : g\left(\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)\right) \text{ is not continuous}\}) = \mathbb{P}(\{x \in \mathbb{R} : x = 0\}) = 0$. By Slutsky Theorem, we have

$$g\left(\left(\begin{smallmatrix} \frac{1}{\sqrt{2}\tau}\sqrt{n}\overline{X}_n \\ \frac{1}{\tau}\sqrt{n}(\overline{Y}_n - \tau) \end{smallmatrix}\right)\right) \xrightarrow{D} \left(\left(\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}\right)\right),$$

i.e.,

$$\frac{1}{\sqrt{2}} \frac{\overline{Y}_n - \tau}{\overline{X}_n} \xrightarrow{D} \frac{Z_2}{Z_1} \overset{D}{=\!=} \text{Cauchy(0,1)} \triangleq C.$$

So

$$\frac{\overline{Y}_n - \tau}{\overline{X}_n} \xrightarrow{D} \sqrt{2}C.$$

Since the characteristic function of C is $\phi_C(t) = e^{-it\mu}$, we have

$$\phi_{\sqrt{2}C}(t) = \mathbb{E}e^{it\sqrt{2}C} = \mathbb{E}e^{i(\sqrt{2}t)C} = e^{-i\sqrt{2}t\mu},$$

which is the characteristic function of Cauchy$(0, \frac{1}{\sqrt{2}})$. So the asymptotic distribution of $\frac{\overline{Y}_n - \tau}{\overline{X}_n}$ is Cauchy$(0, \frac{1}{\sqrt{2}})$.