MTH211A: Theory of Statistics

Mid-semester Examination

SOLUTION SET

Time: 120 minutes **Total marks:** 40

Name: ______________________ **Roll number:** ____________

1. *Answer all questions.*
2. *All notations used are as discussed in class.*

Q.1 Let $X_1, \cdots, X_n$ be a random sample from $\texttt{normal}(\mu_x, \sigma^2)$, $Y_1, \cdots, Y_m$ be a random sample from $\texttt{normal}(\mu_y, \sigma^2)$, and $\{X_1, \cdots, X_n, Y_1, \cdots, Y_m\}$ are mutually independent. Find the distribution of

$$T = \{(\bar{X}_n - \mu_x) - (\bar{Y}_m - \mu_y)\} \Big/ \sqrt{\frac{nS_x^2 + mS_y^2}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)}.$$

[6]

As $X_1, \dots, X_n \overset{IID}{\sim} N(\mu_x, \sigma^2)$, we have

$$\left.\begin{array}{ll} \text{(i)} & \bar{X}_n \sim N(\mu_x, \sigma^2/n) \\ \text{and (ii)} & \dfrac{n S_x^2}{\sigma^2} \sim \chi^2_{(n-1)} \end{array}\right\} \text{ind.}$$

Similarly,

$$\left.\begin{array}{ll} \text{(iii)} & \bar{Y}_n \sim N(\mu_y, \sigma^2/m) \\ \text{(iv)} & \dfrac{m S_Y^2}{\sigma^2} \sim \chi^2_{(m-1)} \end{array}\right\} \text{ind}$$

Further, as $\underset{\sim}{X} \perp\!\!\!\perp \underset{\sim}{Y}$, we have $\bar{X}_n \perp\!\!\!\perp \bar{Y}_n$ and $S_x^2 \perp\!\!\!\perp S_Y^2$.

Thus, $(\bar{X}_n - \bar{Y}_n) \sim N\left(\mu_x - \mu_y, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right)$

$$\Rightarrow Z = \left[\frac{(\bar{X}_n - \mu_x) - (\bar{Y}_n - \mu_y)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}}\right] \sim N(0,1) \quad \text{— (v)}$$

$$\text{— (vi)}$$

and $W = \dfrac{nS_x^2 + mS_Y^2}{\sigma^2} \sim \chi^2_{(m+n-2)}$

Finally, observe that as $\bar{X}_n \perp\!\!\!\perp (\bar{Y}_n, S_Y^2)$ ~~and~~ $\bar{X}_n \perp\!\!\!\perp S_x^2$ and $\bar{X}_n \perp\!\!\!\perp (S_x^2, S_Y^2, \bar{Y}_n)$ and similarly, $S_x^2 \perp\!\!\!\perp (\bar{Y}_n, S_Y^2)$, we have $(\bar{X}_n, \bar{Y}_n) \perp\!\!\!\perp (S_x^2, S_Y^2)$.
$\bar{Y}_n \perp\!\!\!\perp (S_x^2, S_Y^2, \bar{X}_n)$, and

Thus, (v) and (vi) are independently distributed.

Now, $\dfrac{Z}{\sqrt{W/(m+n-2)}} \sim t_{(m+n-2)}$ distributn.. $\Rightarrow T \sim t_{(m+n-2)}$

x —— x

Q.2 Consider a trivariate discrete random variable $\mathbf{W} = (X, Y, Z)'$ with pmf

$$f_{\mathbf{W}}(\mathbf{w}) = f_{\mathbf{W}}\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = p_1^x p_2^y p_3^z, \text{ with } x, y, z \in \{0,1\},\ x+y+z=1,\ p_1, p_2, p_3 > 0, \text{ and } p_1+p_2+p_3 = 1.$$

(a) Show that the marginal distributions of X, Y and Z are Bernoulli. Find the parameters.

(b) Based on a random sample $\mathbf{w}_1, \ldots, \mathbf{w}_n$, find the expectation of the trace of the sample variance covariance matrix $S_n = n^{-1}\sum_{i=1}^{n}(\mathbf{w}_i - \bar{\mathbf{w}})(\mathbf{w}_i - \bar{\mathbf{w}})^\top$. [3+5]

(a) $P(X=1) = P(X=1, Y=0, Z=0) = p_1$ as $x+y+z=1$

$\Rightarrow P(X=0) = 1-p_1 \Rightarrow X \sim \text{Bern}(p_1)$

Similarly, $Z \sim \text{Bern}(p_3)$ and $Y \sim \text{Bern}(p_2)$.

(b) Variance - covariance matrix

$$S_n = \begin{bmatrix} S_X^2 & S_{XY} & S_{XZ} \\ S_{XY} & S_Y^2 & S_{YZ} \\ S_{XZ} & S_{YZ} & S_Z^2 \end{bmatrix}$$

where S_X^2 = sample variance of $X_1, \ldots, X_n$.

S_{XY} = sample covariance of $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$,

and so on.

Now, $\operatorname{tr}[E(S_n)]$

$= E[\operatorname{tr}(S_n)] = E[S_X^2 + S_Y^2 + S_Z^2] = \frac{n}{(n-1)}\{\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2\}$,

where σ_X^2 = population var. of X $= E[(X - E(X))^2] = p_1(1-p_1)$.

Thus, $\operatorname{tr}[E(S_n)] = \left(\frac{n}{n-1}\right)\{p_1(1-p_1) + p_2(1-p_2) + p_3(1-p_3)\}$.

$= \left(\frac{n}{n-1}\right)\{1 - (p_1^2 + p_2^2 + p_3^2)\}$

x ———— x

Q.3 Let $X_1, \cdots, X_n$ be a random sample from the location-scale exponential family with pdf

$$f_X(x) = \theta_2^{-1} \exp\{-(x-\theta_1)/\theta_2\} \quad \text{with } x > \theta_1, \text{ and } \theta_1, \theta_2 > 0.$$

Find the conditional distribution of $X_{(n)}$ given $X_{(1)}$. [6]

The joint distributn. of $X_{(1)}, X_{(n)}$ is given by:

$$f_{S,T}(s,t) = n(n-1)\left[F_x(t) - F_x(s)\right]^{n-2} f_x(s)\, f_x(t),$$

$$\theta_1 < s < t < \infty .$$

where $S = X_{(1)}$, $T = X_{(n)}$,

F_x and f_x are the CDF and PDF of X_1.

Marginal distributn. of $X_{(1)}$ is:

$$f_S(s) = n\, f_x(s)\left[1 - F_x(s)\right]^{n-1} \; ; \; s > \theta_1$$

Thus, the conditional distn. of $X_{(n)} = T$ given $X_{(1)} = S = s$ is:

$$f_{T|S=s}(t) = \frac{f_{S,T}(s,t)}{f_S(s)} = \frac{(n-1)\left[F_x(t) - F_x(s)\right]^{n-2} f_x(t)}{\left[1 - F_x(s)\right]^{n-1}}, \quad -(1)$$

$t > s$.

Now, $$F_x(t) = \int_{\theta_1}^{t} \frac{e^{\theta_1/\theta_2}}{\theta_2} e^{-x/\theta_2}\, dx = e^{\theta_1/\theta_2}\left\{e^{-\theta_1/\theta_2} - e^{-t/\theta_2}\right\}.$$

$$= 1 - \alpha e^{-t/\theta_2} \quad \text{where } \alpha = e^{\theta_1/\theta_2}$$

So,

$$f_{T|(S=s)} = (n-1)\left[\alpha\left\{e^{-s/\theta_2} - e^{-t/\theta_2}\right\}\right]^{n-2} \alpha^{-(n-1)} e^{(n-1)s/\theta_2} \frac{\alpha}{\theta_2} e^{-t/\theta_2}$$

$t > s$.

$$= \frac{(n-1)}{\theta_2} e^{-t/\theta_2} e^{s/\theta_2}\left\{1 - e^{-(t-s)/\theta_2}\right\}^{(n-2)} \quad t > s.$$

$$= \frac{(n-1)}{\theta_2} e^{-(t-s)/\theta_2}\left\{1 - e^{-(t-s)/\theta_2}\right\}^{(n-2)} \; ; \; t > s.$$

x ———— x

Q.4 Based on a random samples $X_1, \cdots, X_n$, find a minimal sufficient statistic for the parameter $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ of the inverse-Gaussian distribution with pdf

$$f_X(x) = \sqrt{\theta_1}\exp\left\{-\theta_1(x-\theta_2)^2/(2\theta_2 x)\right\}/\sqrt{2\pi x^3}, \quad \text{with} \quad x > 0, \theta_1 > 0 \quad \text{and} \quad \theta_2 \in \mathbb{R}.$$

[4]

To find the minimal sufficient statistic, we consider the ratio of joint pdf of $X_1, \dots, X_n$, for 2 different realizations $\underset{\sim}{x}$ and $\underset{\sim}{y}$:

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{f_{\underset{\sim}{X}}(\underset{\sim}{x})}{f_{\underset{\sim}{Y}}(\underset{\sim}{y})} = \left(\frac{\prod_{i=1}^n x_i}{\prod_{i=1}^n y_i}\right)^{-3/2} \exp\left\{-\theta_1 \sum_{i=1}^n \frac{(x_i-\theta_2)^2}{2\theta_2 x_i} + \frac{\theta_1}{2\theta_2}\sum_{i=1}^n \frac{(y_i-\theta_2)^2}{y_i}\right\}$$

$$= \left(\frac{\prod_{i=1}^n x_i}{\prod_{i=1}^n y_i}\right)^{-3/2} \exp\left\{-\frac{\theta_1}{2\theta_2}\left[\sum_{i=1}^n x_i - \sum_{i=1}^n y_i + \theta_2^2\left(\sum_{i=1}^n \frac{1}{x_i} - \sum_{i=1}^n \frac{1}{y_i}\right)\right]\right\}$$

Observe that, the ratio $\lambda_{\underset{\sim}{\theta}}(\underset{\sim}{x}, \underset{\sim}{y})$ will be a constant function $\left(\prod y_i / \prod x_i\right)^{3/2}$ w.r.t. $\underset{\sim}{\theta} = (\theta_1, \theta_2)'$ iff

$$\sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad \text{and} \quad \sum_{i=1}^n \frac{1}{x_i} = \sum_{i=1}^n \frac{1}{y_i}.$$

So, $\underset{\sim}{T}(\underset{\sim}{X}) = \begin{bmatrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n \frac{1}{X_i} \end{bmatrix}$ is a minimal sufficient statistic.

X ——————————— X

Q.5 Let $X_i \overset{\text{iid}}{\sim}$ normal (μ, σ^2), for $i = 1, \ldots, n$. Define $U = (a_1 X_1 + \cdots + a_n X_n)$ and $V = (b_1 X_1 + \cdots + b_n X_n)$. Find conditions on $\mathbf{a} = (a_1, \ldots, a_n)'$ and $\mathbf{b} = (b_1, \ldots, b_n)'$ such that U and V are independently distributed. [4]

Claim: $\begin{pmatrix} U \\ V \end{pmatrix} \sim$ Bivariate normal distn..

To see this, we take $\underset{\sim}{c} = (c_1 \; c_2)'$ and observe that

$$c_1 U + c_2 V = (a_1 c_1 + b_1 c_2) X_1 + \cdots + (a_n c_1 + b_n c_2) X_n$$

$$= d_1 X_1 + \cdots + d_n X_n.$$

$$\sim N\left(\mu \sum_{i=1}^n d_i, \; \sigma^2 \sum_{i=1}^n d_i^2\right),$$

where $\sum_{i=1}^n d_i = c_1 \sum_{i=1}^n a_i + c_2 \sum_{i=1}^n b_i = \underset{\sim}{c}' \begin{pmatrix} \Sigma a_i \\ \Sigma b_i \end{pmatrix}$.

$$\Sigma d_i^2 = c_1^2 \sum_{i=1}^n a_i^2 + c_2^2 \sum_{i=1}^n b_i^2 + 2 c_1 c_2 \sum_{i=1}^n a_i b_i$$

$$= \underset{\sim}{c}' \begin{pmatrix} \sum_{i=1}^n a_i^2 & \sum_{i=1}^n a_i b_i \\ \sum_{i=1}^n a_i b_i & \sum_{i=1}^n b_i^2 \end{pmatrix} \underset{\sim}{c}$$

$$\therefore \begin{bmatrix} U \\ V \end{bmatrix} \sim N_2\left(\begin{pmatrix} \Sigma a_i \\ \Sigma b_i \end{pmatrix}, \begin{pmatrix} \|\underset{\sim}{a}\|^2 & \langle \underset{\sim}{a}, \underset{\sim}{b} \rangle \\ \langle \underset{\sim}{a}, \underset{\sim}{b} \rangle & \|\underset{\sim}{b}\|^2 \end{pmatrix} \right)$$

As, for bivariate normal distn.s, Covariance $= 0 \iff$ independence,

$$\operatorname{cov}(U, V) = \langle \underset{\sim}{a}, \underset{\sim}{b} \rangle = 0 \iff U \perp\!\!\!\perp V.$$

x —— x

Q.6 Let $X_1, \cdots, X_n$ be a random sample from $\texttt{Gamma}(\alpha, \beta)$ distribution. In each of the following cases, indicate with proper justification, if the statistic is (i) not sufficient, (ii) sufficient but not minimal sufficient, or (iii) minimal sufficient. [3 + 3 + 4]

(a) $T_1 = (\log(\bar{X}), n^{-1}\sum_{i=1}^{n}\log(X_i))'$. [Minimal suffient]

A ~~The~~ minimal sufficient statistic for $\underset{\sim}{\theta} = (\alpha, \beta)'$ is obtained by taking the ratio:

$$\frac{f_{\underset{\sim}{X}}(\underset{\sim}{x})}{f_{\underset{\sim}{X}}(\underset{\sim}{y})} = \left(\frac{\prod x_i}{\prod y_i}\right)^{\alpha-1} \exp\left\{-\beta\left(\sum x_i - \sum y_i\right)\right\} = \lambda_{\underset{\sim}{\theta}}(\underset{\sim}{x}, \underset{\sim}{y}).$$

$\lambda_{\underset{\sim}{\theta}}(\underset{\sim}{x}, \underset{\sim}{y})$ is a constant function of $\underset{\sim}{\theta}$, if $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ and $\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} y_i$.

So, $\underset{\sim}{T}(\underset{\sim}{x}) = \left[\prod_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i\right]$ is minimal sufficient.

Observe that $\underset{\sim}{T_1}(\underset{\sim}{x})$ is a one-one function $\underset{\sim}{h}$ of $\underset{\sim}{T}$, where $\underset{\sim}{h}\left(\binom{x}{y}\right) = \left(\log(y/n), \log(x)/n\right)$. So, $\underset{\sim}{T_1}$ is also minimal sufficient.

(b) $T_2 = (Y_{(1)}, \ldots, Y_{(n)})'$, where $Y_i = |\log(X_i)|$, for $i = 1, \ldots, n$.

[Not sufficient]

Proof: If not, i.e., if $\underset{\sim}{T_2}$ is sufficient, then $\underset{\sim}{T}$ must be a function of $\underset{\sim}{T_2}$. Then for two different realizations $\underset{\sim}{x}$ and $\underset{\sim}{y}$ of $\{X_1, \ldots, X_n\}$, $\underset{\sim}{T_2}(\underset{\sim}{x}) = \underset{\sim}{T_2}(\underset{\sim}{y}) \Rightarrow \underset{\sim}{T}(\underset{\sim}{x}) = \underset{\sim}{T}(\underset{\sim}{y})$.

Now, let $\underset{\sim}{x} = (x_1, \ldots, x_n)$ and $\underset{\sim}{y} = (1/x_1, x_2, \ldots, x_n)$,

then $T_2(\underset{\sim}{x}) = T_2(\underset{\sim}{y})$ but $T(\underset{\sim}{x}) \neq T(\underset{\sim}{y})$.

Hence $\underset{\sim}{T_2}$ is not sufficient.

(c) $T_3 = (\sum_{i=1}^{n}\min\{X_i, 1\}, \sum_{i=1}^{n}\max\{X_i, 1\}, \prod_{i=1}^{n}\min\{X_i, 1\}, \prod_{i=1}^{n}\max\{X_i, 1\})'$.

[Sufficient, but not minimal sufficient]

Note that $\sum_{i=1}^{n}\min\{x_i, 1\} + \sum_{i=1}^{n}\max\{x_i, 1\} = \sum_{i=1}^{n} x_i + n$,

as $\min\{x_i, 1\} + \max\{x_i, 1\} = (x_i + 1)$ for all $i = 1, \ldots, n$.

Further, $\left(\prod_{i=1}^{n}\min\{x_i, 1\}\right) \times \left(\prod_{i=1}^{n}\max\{x_i, 1\}\right) = \prod_{i=1}^{n} x_i$.

So, $\underset{\sim}{T}$ is a function $\underset{\sim}{h}$ of $\underset{\sim}{T_3}$ where $\underset{\sim}{h}\left(\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}\right) = \begin{pmatrix} x+y-n \\ z.w \end{pmatrix}$.

Observe that $\underset{\sim}{h}$ is not a bijective fn. (for e.g., for $(x, y, z/c, wc)$ and (x, y, z, w), we have same functional value).

So, $\underset{\sim}{T_3}$ is sufficient, but not minimal sufficient.

Q.7 Suppose $X_1, \cdots, X_n$ are n iid observations from the pdf $f(x \mid \theta) = \theta x^{\theta-1} \exp\{-x^\theta\}$, with $x > 0$, and $\theta > 0$. Define $Y_i = \log X_i$, for $i = 1, \ldots, n$. Show that $Y_{(n)}/Y_{(1)}$ is ancillary for θ. [4]

Let $Y_1 = \log X_1$, i.e., $X_1 = \exp(Y_1)$ and $J_{xy} = \exp(Y_1)$.

Then $f_{Y_1}(y) = \theta \left(e^y\right)^{\theta-1} e^{-e^{y\theta}} e^y; \quad y \in \mathbb{R}.$

$= \theta e^{\theta y - y - e^{y\theta} + y}; \quad y \in \mathbb{R}.$

$= \theta \exp\{\theta y - e^{y\theta}\}; \quad y \in \mathbb{R}.$

Define $Z = \theta Y$ then

$f_Z(z) = \theta \exp\{z - e^z\} \frac{1}{\theta} \quad ; z \in \mathbb{R}$

$= \exp\{z - e^z\}$ for $z \in \mathbb{R}$.

Observe that the distribution of Z does not depend on θ.

So, Y belongs to the scale family.

$T(\underset{\sim}{Y}) = \frac{Y_{(n)}}{Y_{(1)}} = T(c\underset{\sim}{Y}) = \frac{\max\{cY_1, \cdots, cY_n\}}{\min\{cY_1, \cdots, cY_n\}}$.

So, T is ancillary for θ.

Q.8 The random variable X takes the values $\{-1, 0, 1\}$ according to the distribution $\{F_p : 0 < p < 1/3\}$ such that

$$X = \begin{cases} -1 & \text{with probability } p, \\ 0 & \text{with probability } 2p, \\ 1 & \text{with probability } 1-3p. \end{cases}$$

[3+3]

(a) Is the family of distribution $\{F_p; 0 < p < 1/3\}$ complete?

(b) Let X_1, X_2 be a random sample from this distribution. Is the statistic $T = |X_1 + X_2|$ complete?

(a) $E_p[g(x)] = 0 \Rightarrow g(-1)p + g(0)\cdot 2p + g(1)(1-3p) = 0$

$\Rightarrow p[g(-1) + 2g(0) - 3g(1)] + g(1) = 0$ for all $p \in (0, 1/3)$

So, $g(1) \equiv 0$

and $g(-1) + 2g(0) \equiv 0$.

Now, consider the function g: $g(0) = 1$, $g(-1) = -2$ and $g(1) = 0$. Then $E_p[g(x)] = 0$ for all $p \in (0, 1/3)$, but $P_x[g(x) = 0] = (1-3p) \neq 1$. So, X is not complete.

(b) First we find the distribution of T. T takes 3 possible values,

$P(T=0) = P(x_1=0, x_2=0) + P(x_1=1, x_2=-1) + P(x_1=-1, x_2=1)$
$= 4p^2 + 2p(1-3p)$

Similarly, $P(T=1) = \cancel{2\times 2p\times p + 2\times 2p\times(1-3p)}\ 4p^2 + 2(1-3p)\times 2p$
$= 4p^2 + 4p(1-3p)$

and $P(T=2) = p^2 + (1-3p)^2$.

Now $E_p(g(T)) = 0$ for all $p \in (0, 1/3)$

$\Rightarrow g(0)\{4p^2 + 2p - 6p^2\} + g(1)\{4p - 8p^2\} + g(2)\{10p^2 - 6p + 1\} = 0 \quad \forall p \in (0, 1/3)$

$\Rightarrow 2p^2\{-g(0) - 4g(1) + 5g(2)\} + 2p\{g(0) + 2g(1) - 3g(2)\} + g(2) = 0 \quad \forall p \in (0, 1/3)$

$\Rightarrow g(2) = 0$ and $-g(0) = 4g(1)$ and $2g(1) = -g(0)$,

which implies $g(0) = g(1) = g(2) = 0$. Thus, T is complete.

x———x