Department of Statistics, The Chinese University of Hong Kong
STAT5010 Advanced Statistical Inference (Term 1, 2022–23)

Assignment 1 · due on 2 October 2023
Please submit your answers in `.pdf` *format via Blackboard.*

1. Show that if X is a continuous random variable, then

$$\min_a E|X-a| = E|X-m|,$$

where m is the median of X.

2. Let X and Y be independent standard normal random variables.

   (a) Show that $\frac{X}{X+Y}$ has a Cauchy distribution.

   (b) Find the distribution of $X/|Y|$.

3. Suppose that $X_1, \ldots, X_n$ are independent random variables and that for $i = 1, \ldots, n$, X_i has a Poisson distribution with mean $\lambda_i = \exp(\alpha + \beta t_i)$, where $t_1, \ldots t_n$ are observed constants and α and β are unknown parameters. Show that the joint distributions for $X_1, \ldots, X_n$ form a two-parameter exponential family and identify the statistics T_1 and T_2.

4. Let $X_1, \ldots, X_n$ be independent random variables, and let α_i and $t_i, i = 1, \ldots, n$, be known constants. Suppose $X_i \sim \Gamma(\alpha_i, \lambda_i^{-1})$ with $\lambda_i = \theta_1 + \theta_2 t_i, i = 1, \ldots, n$, where θ_1 and θ_2 are unknown parameters. Show that the joint distributions form a two-parameter exponential family. Identify the statistic T and give its mean and covariance matrix. Note that if $X \sim \Gamma(\alpha, \lambda^{-1})$, the corresponding density is given by $f_X(x; \alpha, \lambda) = \lambda^\alpha x^{\alpha-1} e^{-\lambda x}/\Gamma(\alpha)$ for $x > 0$.

5. Random variables $X_1, X_2, \ldots$ are called "m-dependent" if X_i and X_j are independent whenever $|i-j| \geq m$. Suppose $X_1, X_2, \ldots$ are m-dependent, with $E(X_j) = \mu$ and $Var(X_j) = \sigma^2 < \infty$ for $j \geq 1$. Show that $\bar{X}_n \xrightarrow{p} \mu$ as $n \to \infty$, where $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$.

6. Let $X_1, \cdots, X_n$ be independently distributed with exponential density

$$\frac{1}{2\theta} \exp(-x/2\theta) I(x \geq 0) \ ,$$

and let the ordered X's be denoted by $Y_1 \leq Y_2 \leq \cdots \leq Y_n$. It is assumed that Y_1 becomes available first, then Y_2, and so on, and that observation is continued until Y_r has been observed. This might arise, for example, in life testing where each X measures the length of life of, say, an electron tube, and n tubes are being tested simultaneously. Another application is to the disintegration of radioactive material, where n is the number of atoms, and observation is continued until r α-particles have been emitted.

(a) Show that the joint distribution of $Y_1 \leq Y_2 \leq \cdots \leq Y_r$ has density

$$\frac{1}{(2\theta)^r} \frac{n!}{(n-r)!} \exp\left\{ -\frac{\sum_{i=1}^r y_i + (n-r) y_r}{2\theta} \right\} , \quad 0 \leq y_1 \leq \cdots \leq y_r .$$

(b) Argue that the distribution of $\{\sum_{i=1}^r y_i + (n-r)y_r\}/\theta$ is χ^2 with $2r$ degrees of freedom.

(c) Let $Y_1, Y_2, \ldots$ denote the time required until the first, second, $\ldots$ event occurs in a Poisson process with parameter $1/2\theta'$. Prove that $Z_1 = Y_1/\theta'$, $Z_2 = (Y_2 - Y_1)/\theta'$, $Z_3 = (Y_3 - Y_2)/\theta', \cdots$ are independently distributed as χ^2 with 2 degrees of freedom, and the joint density of $Y_1, \ldots, Y_r$ has the density

$$\frac{1}{(2\theta')^r} \exp\left(-\frac{y_r}{2\theta'} \right) , \quad 0 \leq y_1 \leq \cdots \leq y_r .$$

The distribution of Y_r/θ' is again χ^2 with $2r$ degrees of freedom.

7. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim N(t_i\theta, 1)$, where $t_1, \ldots, t_n$ are a sequence of known constants (not all zero).

   (a) Show that the least squares estimator

   $$\hat{\theta} = \frac{\sum_{i=1}^n t_i X_i}{\sum_{i=1}^n t_i^2}$$

   is complete sufficient for the family of joint distributions.

   (b) Show that $\hat{\theta}$ and $\sum_{i=1}^n (X_i - t_i\hat{\theta})^2$ are independent.

8. Suppose X has a geometric distribution with success probability $\theta \in (0, 1)$, Y has a geometric distribution with success probability $2\theta - \theta^2$, and X and Y are independent. Find a minimal sufficient statistic T for the family of joint distributions. Is T complete?

- End -

1. Show that if X is a continuous random variable, then

$$\min_a E|X-a| = E|X-m|,$$

where m is the median of X.

Let $L(m) = E(|X-m|) = \int_{-\infty}^{+\infty} |x-m| f(x)\,dx = \int_{-\infty}^{m} (m-x) f(x)\,dx + \int_{m}^{+\infty} (x-m) f(x)\,dx$, we want to find the minimum of $L(m)$

$$\frac{\partial L}{\partial m} = (m-x)f(x)\big|_{x=m} + \int_{-\infty}^{m} f(x)\,dx + (x-c)\,f(x)\big|_{x=c} - \int_{c}^{+\infty} f(x)\,dx$$

$$= \int_{-\infty}^{m} f(x)\,dx - \int_{m}^{+\infty} f(x)\,dx$$

when $\frac{\partial L}{\partial m} = 0$, i.e. $\int_{-\infty}^{m} f(x)\,dx = \int_{m}^{+\infty} f(x)\,dx$.

$P(X \le m) = P(X > m)$ (*) (Since X is continuous)

$L(m) = E(|X-m|)$ reaches minimum.

(*) means m is the median of X.

2. Let X and Y be independent standard normal random variables.

(a) Show that $\frac{X}{X+Y}$ has a Cauchy distribution.

(b) Find the distribution of $X/|Y|$.

(a) Cauchy pdf: $f(x) = \frac{1}{\pi b} \frac{1}{1+\left(\frac{x-a}{b}\right)^2}$.

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right)$$

Let $U = X+Y$, $V = \frac{X}{X+Y}$.

$$f_{U,V}(u,v) = \frac{u}{2\pi} \exp\left(-\frac{u^2v^2 + (u-uv)^2}{2}\right)$$

$$= \frac{u}{2\pi} \exp\left(-u^2\left(v^2 - v + \frac{1}{2}\right)\right).$$

The marginal V has

$$f_V(v) = \int_0^{\infty} \frac{u}{\pi} e^{-u^2(v^2-v+\frac{1}{2})}\,du$$

Let $s = u^2$, then $du = \frac{1}{2\sqrt{s}}$

$$f_V(v) = \int_0^\infty \frac{1}{2\pi} \exp\left(-s\left(v^2 - v + \tfrac{1}{2}\right)\right) ds$$

$$= \frac{1}{2\pi} \left(\frac{1}{v^2 - v + \frac{1}{2}} \right)$$

$$= \left(\frac{1}{2\pi}\right) \frac{4}{1 + \left(\frac{v - \frac{1}{2}}{\frac{1}{2}}\right)^2}$$, which implies $a = b = \frac{1}{2}$.

(b) Let $Z = \frac{X}{|Y|}$, Find $F_Z(z) =$

$$P(Z \le z) = P(X \le z|Y|) = \frac{1}{2\pi} \int_{-\infty}^{z|Y|} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} dx\, dy$$

$$\xrightarrow{x = |y|u} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{z} |y| e^{-\frac{y^2(1+u^2)}{2}} du\, dy$$

$$f_Z(z) = F'(Z \le z) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |y| e^{-\frac{y^2(1+z^2)}{2}} dy$$

$$= \frac{1}{\pi} \frac{1}{(1+z^2)}$$. Hence $b = 1$, $a = 0$.

$Z = \frac{X}{|Y|}$ is Cauchy distributed with $a = 0$, $b = 1$.

3. Suppose that $X_1, \ldots, X_n$ are independent random variables and that for $i = 1, \ldots, n$, X_i has a Poisson distribution with mean $\lambda_i = \exp(\alpha + \beta t_i)$, where $t_1, \ldots t_n$ are observed constants and α and β are unknown parameters. Show that the joint distributions for $X_1, \ldots, X_n$ form a two-parameter exponential family and identify the statistics T_1 and T_2.

Poisson pdf: $f_X(x_i) = \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!} = \frac{\exp(\alpha x_i + \beta x_i t_i - \lambda_i)}{x_i!}$

$f_{X_1 X_2 \ldots X_n}(X_1, X_2, \ldots, X_n) =$

$$\exp\left(\alpha \sum_i X_i + \beta \sum_i X_i t_i - \sum_i \exp(\alpha + \beta t_i)\right) \frac{1}{X_i!}$$

Hence $T_1 = \sum_{i=1}^n X_i$, $T_2 = \sum_{i=1}^n t_i X_i$.

4. Let $X_1, \ldots, X_n$ be independent random variables, and let α_i and $t_i, i = 1, \ldots, n$, be known constants. Suppose $X_i \sim \Gamma(\alpha_i, \lambda_i^{-1})$ with $\lambda_i = \theta_1 + \theta_2 t_i, i = 1, \ldots, n$, where θ_1 and θ_2 are unknown parameters. Show that the joint distributions form a two-parameter exponential family. Identify the statistic T and give its mean and covariance matrix. Note that if $X \sim \Gamma(\alpha, \lambda^{-1})$, the corresponding density is given by $f_X(x; \alpha, \lambda) = \lambda^\alpha x^{\alpha-1} e^{-\lambda x}/\Gamma(\alpha)$ for $x > 0$.

$$f(X_i) = (\theta_1 + \theta_2 t_i)^{\alpha_i} X_i^{(\alpha_i - 1)} \exp(-(\theta_1 + \theta_2 t_i) X_i) / \Gamma(\alpha_i)$$

$$= \exp(\theta_1 \alpha_i + \theta_2 \alpha_i t_i + (\alpha_i - 1) \ln X_i - (\theta_1 + \theta_2 t_i) X_i - \ln(\Gamma(\alpha_i))$$

$$f(X_1, X_2, \ldots, X_n) = \exp\left(\sum_{i=1}^{n} (\alpha_i \ln X_i - (\theta_1 + \theta_2 t_i) X_i - B(\theta)\right) h(x)$$

Hence, $T = \begin{pmatrix} t_1(X) \\ t_2(X) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \alpha_i \ln X_i \\ \sum_{i=1}^{n} t_i X_i \end{pmatrix}$.

5. Random variables $X_1, X_2, \ldots$ are called "m-dependent" if X_i and X_j are independent whenever $|i-j| \geq m$. Suppose $X_1, X_2, \ldots$ are m-dependent, with $E(X_j) = \mu$ and $Var(X_j) = \sigma^2 < \infty$ for $j \geq 1$. Show that $\bar{X}_n \xrightarrow{p} \mu$ as $n \to \infty$, where $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$.

Let $Y_n^{(1)} = \dfrac{X_1 + X_{m+1} + X_{2m+1} + \cdots + X_{(n-1)m+1}}{n}$

$\vdots$

$Y_n^{(m)} = \dfrac{X_m + X_{2m} + X_{3m} + \cdots + X_{nm}}{n}$

The we have $Y_n^{(1)} \xrightarrow{P} \mu$, $Y_n^{(2)} \xrightarrow{P} \mu$, $\ldots$, $Y_n^{(m)} \xrightarrow{P} \mu$ as $n \to \infty$.

Then $\bar{X}_n = \dfrac{Y_n^{(1)} + Y_n^{(2)} + \cdots + Y_n^{(m)}}{m} \xrightarrow{P} \mu$ as $n \to \infty$

by continuous mapping theorem.

6. Let $X_1, \cdots, X_n$ be independently distributed with exponential density

$$\frac{1}{2\theta}\exp(-x/2\theta)I(x \geq 0)\ ,$$

and let the ordered X's be denoted by $Y_1 \leq Y_2 \leq \cdots \leq Y_n$. It is assumed that Y_1 becomes available first, then Y_2, and so on, and that observation is continued until Y_r has been observed. This might arise, for example, in life testing where each X measures the length of life of, say, an electron tube, and n tubes are being tested simultaneously. Another application is to the disintegration of radioactive material, where n is the number of atoms, and observation is continued until r α-particles have been emitted.

(a) Show that the joint distribution of $Y_1 \leq Y_2 \leq \cdots \leq Y_r$ has density

$$\frac{1}{(2\theta)^r}\frac{n!}{(n-r)!}\exp\left\{-\frac{\sum_{i=1}^r y_i + (n-r)y_r}{2\theta}\right\}\ , \qquad 0 \leq y_1 \leq \cdots \leq y_r\ .$$

(b) Argue that the distribution of $\{\sum_{i=1}^r y_i + (n-r)y_r\}/\theta$ is χ^2 with $2r$ degrees of freedom.

(c) Let $Y_1, Y_2, \ldots$ denote the time required until the first, second, ... event occurs in a Poisson process with parameter $1/2\theta'$. Prove that $Z_1 = Y_1/\theta'$, $Z_2 = (Y_2 - Y_1)/\theta'$, $Z_3 = (Y_3 - Y_2)/\theta', \cdots$ are independently distributed as χ^2 with 2 degrees of freedom, and the joint density of $Y_1, \ldots, Y_r$ has the density

$$\frac{1}{(2\theta')^r}\exp\left(-\frac{y_r}{2\theta'}\right)\ , \qquad 0 \leq y_1 \leq \cdots \leq y_r\ .$$

The distribution of Y_r/θ' is again χ^2 with $2r$ degrees of freedom.

(a) $f_{Y_1, Y_2, \ldots, Y_n}(y_1, y_2, \ldots, y_n)$

$= n! \frac{1}{(2\theta)^n} \exp\left(-\frac{\sum_{i=1}^n y_i}{2\theta}\right) \mathbb{1}\{0 \leq y_1 \leq \cdots \leq y_n\}.$

If $0 \leq y_1 \leq \cdots \leq y_r$, then

$f_{Y_1, Y_2, \ldots, Y_r}(y_1, \ldots, y_r) = n! \frac{1}{(2\theta)^n} \int_{y_r}^{\infty} \cdots \int_{y_n}^{\infty} \exp\left(-\frac{\sum_{i=1}^n y_i}{2\theta}\right) dy_n \ldots dy_{r+1}$

$= n! \frac{1}{(2\theta)^{n-1}} \int_{y_r}^{\infty} \cdots \int_{y_{n-2}}^{\infty} \exp\left(-\frac{\sum_{i=1}^{n-1} y_i + 2y_{n-1}}{2\theta}\right) dy_{r+1} \cdots dy_{n-1}$

$\vdots$

$= \frac{1}{(2\theta)^r} \frac{n!}{(n-r)!} \exp\left(-\frac{\sum_{i=1}^r y_i + (n-r)y_r}{2\theta}\right)$

(b) Let $f: \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_r \end{pmatrix} \longrightarrow \begin{pmatrix} Y_1 \\ Y_2 - Y_1 \\ \vdots \\ Y_r - Y_{r-1} \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_r \end{pmatrix}$

$$\nabla f^{-1} = \begin{pmatrix} 1 & & & \\ 1 & 1 & & 0 \\ \vdots & & \ddots & \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

Then we have

$$f_{u_1,u_2,\dots,u_r}(u_1,u_2,\dots,u_r) = f_{Y_1,\dots,Y_r}(u_1, u_1+u_2, \dots, \sum_{i=1}^r u_i)$$

$$= \frac{1}{(2\theta)^r} \frac{n!}{(n-r)!} \exp\left(-\frac{nu_1+(n-1)u_2+\dots+(n-r+1)u_r}{2\theta}\right)$$

$$= \prod_{i=1}^r \left(\frac{(n+1-i)}{2\theta} \exp\left(-\frac{(n+1-i)u_i}{2\theta}\right)\right)$$

Hence we have $u_i \sim \exp\left(\frac{2\theta}{n-i+1}\right)$.

$$\frac{(n-i+1)u_i}{\theta} \sim \chi^2(2) \quad i = 1, 2, \dots, r.$$

$$\frac{\sum_{i=1}^r Y_i + (n-r)Y_r}{\theta} = \sum_{i=1}^r \frac{(n-i+1)u_i}{\theta} \sim \chi^2(2r).$$

(c) $P(Z_1 \geq z) = P(Y_1 \geq z\theta') = e^{-\frac{z}{2}} \sim \chi^2(2)$.

$$P(Z_2 \geq z_2, Z_1 \geq z_1) = \int_{z_1}^{\infty} P(Z_1 = z) P(Z_2 \geq z_2 \mid Z_1 = z) dz$$

$$= \int_{z_1}^{\infty} P(Y_2 - Y_1 \geq z_2\theta' \mid Y_1 = z_0\theta') P(Z_1 = z) dz$$

$$= e^{-\frac{z_2}{2}} e^{-\frac{z_1}{2}}$$

Hence we have $z_2 \sim \chi^2(2)$

We can show that $z_i \sim \chi^2(2)$, $i = 1, 2, \dots, r$

using how we prove above.

We have $\quad p_{z_1, \dots, z_r} = \frac{1}{2^r} \exp\left(-\frac{\sum_{i=1}^{r} z_i}{2}\right)$

$$p_{Y_1, \dots, Y_r}(y_1, \dots, y_r) = \frac{1}{(2\theta)^r} \exp\left(-\frac{y_r}{2\theta}\right), \quad 0 \le y_1 \le \dots \le y_r.$$

7. Let $X_1, \dots, X_n$ be independent, with $X_i \sim N(t_i\theta, 1)$, where $t_1, \dots, t_n$ are a sequence of known constants (not all zero).

(a) Show that the least squares estimator

$$\hat{\theta} = \frac{\sum_{i=1}^n t_i X_i}{\sum_{i=1}^n t_i^2}$$

is complete sufficient for the family of joint distributions.

(b) Show that $\hat{\theta}$ and $\sum_{i=1}^n (X_i - t_i\hat{\theta})^2$ are independent.

(a). $P_{X_1, X_2, \dots, X_n}(X_1, \dots, X_n) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\sum_{i=1}^n \frac{(X_i - t_i\theta)^2}{2}\right)$

$= \frac{1}{(\sqrt{2\pi})^n} \exp\left(\theta \sum_{i=1}^n t_i X_i - \sum_{i=1}^n \frac{t_i^2\theta^2}{2} - \sum_{i=1}^n \frac{X_i^2}{2}\right)$

Since $P_{X_1, \dots, X_n}$ forms a full rank exponential family.

$T = \sum_{i=1}^n t_i X_i$ is complete sufficient.

i.e. $\hat{\theta}$ is complete sufficient.

(b). Let $Z_i = X_i - t_i\theta$, $i = 1, 2, \dots, n$.

$\hat{\theta} = \frac{\sum_{i=1}^n t_i X_i}{\sum_{i=1}^n t_i^2} = \frac{\sum_{i=1}^n t_i (Z_i + t_i\theta)}{\sum_{i=1}^n t_i^2} = \theta + \frac{\sum_{i=1}^n t_i Z_i}{\sum_{i=1}^n t_i^2}$

$X_i - t_i\hat{\theta} = Z_i - t_i \frac{\sum_{i=1}^n t_i Z_i}{\sum_{i=1}^n t_i^2}$

Since $Z_i \overset{iid}{\sim} N(0,1)$, we have

$\sum_{i=1}^n (X_i - t_i\hat{\theta})^2$ is ancillary.

Then by Basu's theorem, $\hat{\theta} \perp\!\!\!\perp \sum_{i=1}^n (X_i - t_i\hat{\theta})^2$.

8. Suppose X has a geometric distribution with success probability $\theta \in (0,1)$, Y has a geometric distribution with success probability $2\theta - \theta^2$, and X and Y are independent. Find a minimal sufficient statistic T for the family of joint distributions. Is T complete?

$$P_X(x) = (1-\theta)^x \theta, \quad P_Y(y) = (2\theta - \theta^2)(1+\theta^2 - 2\theta)^y$$

$$P_{X,Y}(x,y) = \theta(1-\theta)^x (2\theta - \theta^2)(\theta - 1)^{2y}$$

$$= \theta^2 (1-\theta)^{x+2y} (2-\theta)$$

$$= \exp\left((x+2y)\ln(1-\theta) + 2\ln\theta + \ln(2-\theta) \right)$$

$x, y = 1, 2, 3, \dots,$

Since $P_{X,Y}$ forms a full rank exponential family, $T = X + 2Y$ is complete sufficient.