You may use the following facts on this Theory I question set.

- If a random vector
$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{pmatrix}, \quad \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right)$$
has a multivariate normal distribution, then the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{X}$ is normal with mean vector $\boldsymbol{\mu}_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X)$ and covariance matrix $\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$.
- A chi-squared distribution with ν degrees of freedom has mean ν and variance 2ν.

Part I

Let X_1, X_2 be iid normal $N(0, \theta)$ random variables with mean 0 and variance $\theta > 0$. Define

$$U = \frac{(X_1 + X_2)^2}{2}, \quad V = \frac{(X_1 - X_2)^2}{2}, \quad W = X_1 X_2, \quad S_1 = \text{sign}(X_1 + X_2), \quad S_2 = \text{sign}(X_1 - X_2),$$

where, for $x \in \mathbb{R}$, $\text{sign}(x) = 1$ if $x > 0$; $\text{sign}(x) = -1$ if $x < 0$; and $\text{sign}(x) = 0$ if $x = 0$.

1. Carefully show that $\text{E}_\theta|W| = 2\theta/\pi$.
2. Determine the joint pdf of $(X_1, |W|)$.
3. Considering both V and $|W|$ as estimators of θ, which is the better estimator in terms of MSE?
4. Carefully derive the joint pmf of (S_1, S_2), stating any standard results that you use.
5. If a random variable Y has moment generating function $M_Y(t) = 4^{-1}(1+e^{2t})+2^{-1}e^t$, $t \in \mathbb{R}$, then determine all possible paired constants (a, b) (for $a, b \in \mathbb{R}$) so that $aY + b$ has the same distribution as $S_1 + S_2$.
6. Show that the random vector (U, V) is independent of the random vector (S_1, S_2).
7. Prove that U and V are independent, justifying any steps that you use.
8. Identify the distribution of each of the following random variables, stating any standard results used:
   - **a)** $(S_1 + S_2 + 2)/2$
   - **b)** $(U + V)/\theta$
   - **c)** $S_1 S_2 \sqrt{U/V}$
   - **d)** $F_1(X_2)$, where $F_1(x) = P_\theta(X_1 \leq x)$, $x \in \mathbb{R}$.
9. Determine the conditional joint distribution of (X_1, X_2) given (U, V), expressing your answer as either a joint pdf or joint pmf.
Note: $X_1 + X_2 = S_1\sqrt{2U}$ and $X_1 - X_2 = S_2\sqrt{2V}$

10. Show that the conditional expectation holds: $\mathrm{E}_\theta[W|V] = [\theta - V]/2$, stating any standard results used.
Note: U, V, W are linearly related.

11. Based on Questions 9 and 10, explain why V cannot be sufficient for θ but (U, V) must be sufficient for θ.

Part II

Let (X_1, X_2) be a normal random vector with $\mathrm{E}X_1 = \mathrm{E}X_2 = 0$, $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \theta > 0$, and $\mathrm{corr}(X_1, X_2) = \rho \in (-1, 1)$ is the correlation between X_1, X_2.

12. If B is a Bernoulli(0.3) random variable, independent of (X_1, X_2), find the pdf of $Y = 2X_1 - X_2 + BX_1 + (1 - B)X_2$.

13. Prove that $\mathrm{corr}(X_1^2, X_2^2) = \rho^2$ is the correlation between X_1^2 and X_2^2.

14. If $L(\rho, \theta)$ denotes the likelihood function based on X_1, X_2, prove that the profile likelihood function

$$L(\rho) \equiv \sup_\theta L(\rho, \theta) = \frac{\sqrt{1-\rho^2}}{2\pi(X_1^2 + X_2^2 - 2\rho X_1 X_2)} \exp\left\{-\frac{1}{2}\right\} \quad \text{for each } \rho.$$

15. Show that the likelihood ratio test of $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$ has the general form

$$\begin{cases} 1 & \text{if } |\hat{p}| > c \\ 0 & \text{if } |\hat{p}| \leq c \end{cases}$$

for some c, where $\hat{p} = 2X_1X_2/(X_1^2 + X_2^2)$.

16. Assuming that θ and ρ are unknown but that one can simulate a sequence $U_1, U_2, \ldots$ of iid Uniform$(0, 1)$ random variables, state an algorithm for numerically calibrating (or approximating) "c" so that the likelihood ratio test in Question 15 has size 10%.

1. By independence $\mathrm{E}|W| = \mathrm{E}|X_1| \cdot \mathrm{E}|X_2| = \theta[\mathrm{E}|Z|]^2$ where $Z = X_1/\sqrt{\theta}$ is standard normal. Then,

$$\mathrm{E}|Z| = \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 2\int_0^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \big[- e^{-z^2/2}\big|_0^{\infty} \big] = \sqrt{\frac{2}{\pi}},$$

so that $\mathrm{E}|W| = 2\theta/\pi$.

2. For $(Y_1, Y_2) = (X_1, |W|)$, it holds that $Y_1 = X_1$ and X_2 has two possibilities Y_2/Y_1 or $-Y_2/Y_1$ when $Y_2 > 0$. Adding the density contributions over the support $y_1 \neq 0 \in \mathbb{R}$ and $y_2 > 0$ (can assume $y_1 = 0$ too without changes),

$$\begin{aligned} & f_{Y_1,Y_2}(y_1, y_2) \\ = & \left| \det \begin{pmatrix} 1 & 0 \\ y_2/y_1^2 & -1/y_1 \end{pmatrix} \right| f_{X_1,X_2}(y_1, -y_2/y_1) + \left| \det \begin{pmatrix} 1 & 0 \\ -y_2/y_1^2 & 1/y_1 \end{pmatrix} \right| f_{X_1,X_2}(y_1, y_2/y_1) \\ = & 2|1/y_1| \frac{1}{\sqrt{2\pi\theta}} e^{-y_1^2/(2\theta)} \frac{1}{\sqrt{2\pi\theta}} e^{-(y_2/y_1)^2/(2\theta)}. \end{aligned}$$

Hence the joint pdf is $f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{|y_1|\pi\theta} e^{-[y_1^2 + (y_2/y_1)^2]/(2\theta)}$ for $y_1 \in \mathbb{R}$ and $y_2 > 0$ (and otherwise, $f_{Y_1,Y_2}(y_1, y_2) = 0$).

3. Note $(X_1 - X_2)/\sqrt{2\theta} \sim N(0, 1)$ so, in squaring, $V/\theta \sim \chi_1^2$, implying $\mathrm{E}(V) = \theta$ and $\mathrm{MSE}(V) = \mathrm{E}(V - \theta)^2 = \mathrm{Var}(V) = 2\theta^2$. Using Question 1, $\mathrm{E}X_i^2 = \theta$, and independence,

$$\mathrm{MSE}(|W|) = \mathrm{E}(|W| - \theta)^2 = \mathrm{E}|W|^2 - 2\theta\mathrm{E}|W| + \theta^2 = \mathrm{E}X_1^2\mathrm{E}X_2^2 - 4\pi^{-1}\theta^2 + \theta^2 = 2\theta^2(1 - \pi^{-1}2).$$

Hence, $\mathrm{MSE}(V) > \mathrm{MSE}(|W|)$ for any $\theta > 0$ and $|W|$ is MSE-better than V.

4. Note that $(X_1 + X_2, X_1 - X_2)$ is bivariate normal with mean $(0, 0)$ and variance $2I_2$, where I_2 is the 2×2 identity matrix. Hence, by normality, $X_1 + X_2$ and $X_1 - X_2$ must be independent (i.e., $\mathrm{cov}(X_1 + X_2, X_1 - X_2) = 0$) and, consequently, so are $S_1 = \mathrm{sign}(X_1 + X_2)$ and $S_2 = \mathrm{sign}(X_1 - X_2)$ as functions of $X_1 + X_2$ and $X_1 - X_2$, respectively. As $P(S_1 = 1) = P(X_1 + X_2 > 0) = 1/2 = 1 - P(S_1 = -1)$ and $P(S_2 = 1) = P(X_1 - X_2 > 0) = 1/2 = 1 - P(S_2 = -1)$, the joint pmf of (S_1, S_2) is given by

$$f(s_1, s_2) = P(S_1 = s_1, S_2 = s_2) = \begin{cases} 1/4 & |s_1| = |s_2| = 1 \\ 0 & \text{otherwise} \end{cases}$$

5. By its moment generating function, $P(Y = 0) = 1/4 = P(Y = 2)$ and $P(Y = 1) = 1/2$ (i.e., $P(Y = y)$ is the weight on e^{yt} in $M_Y(t)$) so that $aY + b$ assumes the values

$$b, a2 + b \text{ (each with prob } 1/4) \text{ and } a + b \text{ (with prob } 1/2).$$

By independence of the binary (1 or -1) distributions of S_1 and S_2 from Question 4, the distribution of $S_1 + S_2$ is given by $P(S_1 + S_2 = -2) = 1/4 = P(S_1 + S_2 = 2)$ and $P(S_1 + S_2 = 0) = 1/2$. To match the distributions of $S_1 + S_2$ and $aY + b$ starting from the $1/2$ probability, one requires $a + b = 0$ or $a = -b$ so that $aY + b$ assumes the values

$$b, -b \text{ (each with prob } 1/4) \text{ and } 0 \text{ (with prob } 1/2).$$

Now b may be 2 or -2. Hence, either $a = 2, b = -2$ or $a = -2, b = 2$.

6. By symmetry of the normal distribution, $([X_1+X_2],[X_1-X_2]) \stackrel{d}{=} (a[X_1+X_2], b[X_1-X_2])$ holds for any $a,b \in \{\pm 1\}$, where $\stackrel{d}{=}$ denotes equality in distribution. Hence, for any $u,v \in \mathbb{R}^2$,

$$\begin{aligned} & P\left((X_1+X_2)^2/2 \le u, (X_1-X_2)^2/2 \le v, \text{sign}(X_1+X_2)=1, \text{sign}(X_1-X_2)=1\right) \\ = \; & P\left(a^2(X_1+X_2)^2/2 \le u, b^2(X_1-X_2)^2/2 \le v, \text{sign}[a(X_1+X_2)]=1, \text{sign}[b(X_1-X_2)]=1\right) \end{aligned}$$

holds for any $a,b \in \{\pm 1\}$. Also, the random variables $U = (X_1+X_2)^2/2 = a^2(X_1+X_2)^2/2$ and $V = (X_1-X_2)^2/2 = b^2(X_1-X_2)^2/2$ are unchanged in form when substituting $(a[X_1+X_2], b[X_1-X_2])$ for $([X_1+X_2],[X_1-X_2])$. Consequently,

$$\begin{aligned} & P\left(U \le u, V \le v, \text{sign}(X_1+X_2)=1, \text{sign}(X_1-X_2)=1\right) \\ = \; & P\left(U \le u, V \le v, \text{sign}(X_1+X_2)=a, \text{sign}(X_1-X_2)=b\right) \end{aligned}$$

holds for any $a,b \in \{\pm 1\}$ so that

$$\begin{aligned} & P\left(U \le u, V \le v, \text{sign}(X_1+X_2)=a, \text{sign}(X_1-X_2)=b\right) \\ = \; & \frac{1}{4} P\left(U \le u, V \le v\right) \\ = \; & P(S_1=a, S_2=b) P\left(U \le u, V \le v\right) \end{aligned}$$

holds for any $a,b \in \{\pm 1\}$. In other words,

$$P\left(U \le u, V \le v, S_1 \le s_1, S_2 \le s_2\right) = P\left(U \le u, V \le v\right) P\left(S_1 \le s_1, S_2 \le s_2\right)$$

for any $u,v,s_1,s_2 \in \mathbb{R}$ so that (U,V) and (S_1,S_2) are independent.

7. (X_1+X_2, X_1-X_2) is bivariate normal with mean $(0,0)$ and variance $2I_2$ so that X_1+X_2 and X_1-X_2 must be independent and, consequently, so are $U=(X_1+X_2)^2/2$ and $V=(X_1-X_2)^2/2$.

8. **a**) $(S_1+1)/2$ and $(S_2+1)/2$ are independent Bernoulli$(1/2)$ random variables (i.e., each assuming 0 or 1 with equal probability) so that their sum $(S_1+S_2+2)/2$ is Binomial$(2,1/2)$; see also Question 4.

b) U/θ and V/θ are iid χ^2_1 random variables so that $(U+V)/\theta$ is χ^2_2 distributed; see also Question 7.

c) U/V is $F_{1,1}$-distributed (i.e., ratio of independent χ^2_1 variables) so that $\sqrt{U/V} \stackrel{d}{=} |T_1|$, where T_1 is t-distributed with 1 degree of freedom. Also, $\sqrt{U/V}$ is independent of S_1S_2, where the later is -1 or 1 with probability $1/2$. Hence, $S_1S_2\sqrt{U/V} \stackrel{d}{=} T_1$ by symmetry of the t-distribution.

d) $F_1(X_2)$ is uniform$(0,1)$ by the probability integral transform.

9. Note that, by the hint, $(X_1,X_2) = ([S_1\sqrt{2U}+S_2\sqrt{2V}]/2, [S_1\sqrt{2U}-S_2\sqrt{2V}]/2)$. The distribution vector of (S_1,S_2) is uniform on $\{(1,1),(-1,-1),(-1,1),(1,-1)\}$, which is the same as the distribution of (S_1,S_2) conditional on (U,V), as (S_1,S_2) and (U,V) are independent. So the distribution of (X_1,X_2) conditional on (U,V) is discrete and given by

$$\begin{aligned} & P_\theta(X_1=x_1, X_2=x_2|(U,V)) \\ = \; & \begin{cases} 1/4 & (x_1,x_2) = (a+b,a-b), (-a+b,-a-b), (a-b,a+b), (-a-b,-a+b) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

for $a=\sqrt{2U}/2$ and $b=\sqrt{2V}/2$ above.

10. Note that $W = X_1X_2 = [(X_1+X_2)^2 - (X_1-X_2)^2]/4 = [U-V]/2$. U is independent of V and U/θ has a χ^2_1 distribution (i.e., $\mathrm{E}_\theta(U/\theta)=1$) so that

$$\mathrm{E}_\theta[W|V] = \mathrm{E}_\theta[(U-V)/2|V] = \frac{1}{2}\left(\mathrm{E}_\theta U - V\right) = \frac{1}{2}\left(\theta - V\right).$$

11. By Question 9, (U,V) must be sufficient for θ by definition as the distribution of (X_1,X_2) given (U,V) does not depend on θ. On the other hand, the conditional expectation of the statistic W given V does depend on θ in Question 10, which could not happen if V were indeed sufficient for θ.

12. By independence, the cdf of $Y = 2X_1 - X_2 + BX_1 + (1-B)X_2$ may be written as

$$\begin{aligned} F_Y(y) = P(Y \le y) &= P(Y \le y|B=1)P(B=1) + P(Y \le y|B=0)P(B=0) \\ &= P(3X_1 - X_2 \le y)(0.3) + P(2X_1 \le y)(0.7) \\ &= \Phi\left(\frac{y}{\sqrt{\theta(10-6\rho)}}\right)(0.3) + \Phi\left(\frac{y}{\sqrt{\theta 4}}\right)(0.7) \end{aligned}$$

for $y \in \mathbb{R}$. Hence, the pdf is

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{0.3}{\sqrt{2\pi\theta(10-6\rho)}}e^{-y^2/[2\theta(10-6\rho)]} + \frac{0.7}{\sqrt{8\pi\theta}}e^{-y^2/[8\theta]}, \quad y \in \mathbb{R}.$$

13. Note, for either $i=1$ or 2, it holds that $X_i^2/\theta \sim \chi^2_1$ so that $\mathrm{E}X_i^2 = \theta$, $2\theta^2 = \mathrm{Var}(X_i^2) = \mathrm{E}X_i^4 - [\mathrm{E}X_i^2]^2$ and $\mathrm{E}X_i^4 = 3\theta^2$. Then,

$$\begin{aligned} \mathrm{cov}(X_1^2, X_2^2) &= \mathrm{E}(X_1^2X_2^2) - \mathrm{E}[X_1^2]\cdot\mathrm{E}[X_2^2] \\ &= \mathrm{E}(\mathrm{E}[X_1^2X_2^2|X_1]) - \theta^2 \\ &= \mathrm{E}[X_1^2\mathrm{E}[X_2^2|X_1]] - \theta^2 \\ &= \mathrm{E}[\rho^2X_1^4 + \theta(1-\rho^2)X_1^2] - \theta^2 \\ &= 3\rho^2\theta^2 + \theta^2(1-\rho^2) - \theta^2 \\ &= 2\theta^2\rho^2 \end{aligned}$$

using the conditional variance $\mathrm{Var}(X_2|X_1) = \theta(1-\rho^2)$ and conditional mean $\mathrm{E}[X_2|X_1] = \rho X_1$ under normality to determine

$$\mathrm{E}[X_2^2|X_1] = \mathrm{Var}(X_2|X_1) + \{\mathrm{E}[X_2|X_1]\}^2 = \theta(1-\rho^2) + \{\rho X_1\}^2.$$

Hence, the correlation is

$$\mathrm{corr}(X_1^2, X_2) = \frac{\mathrm{cov}(X_1^2, X_2^2)}{\sqrt{\mathrm{Var}(X_1^2)\mathrm{Var}(X_2^2)}} = \rho^2.$$

14. The joint likelihood function for (θ, ρ) is

$$\begin{aligned} L(\rho,\theta) &= f(X_1|\theta)f(X_2|X_1,\theta,\rho) \\ &= \frac{1}{\sqrt{2\pi\theta}}\exp\left\{-\frac{1}{2\theta}X_1^2\right\}\frac{1}{\sqrt{2\pi\theta(1-\rho^2)}}\exp\left\{-\frac{1}{2\theta(1-\rho^2)}(X_2-\rho X_1)^2\right\} \\ &= \frac{1}{2\pi\theta}\frac{1}{\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\theta(1-\rho^2)}\left((1-\rho^2)X_1^2+(X_2-\rho X_1^2)\right)\right\} \\ &= \frac{1}{2\pi\theta}\frac{1}{\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\theta}\frac{X_1^2+X_2^2-2\rho X_1X_2}{1-\rho^2}\right\} \end{aligned}$$

For a given $\rho \in (-1,1)$, maximize $L(\rho,\theta)$ over θ by solving

$$0 = \frac{\partial \log L(\rho,\theta)}{\partial\theta} = -\frac{1}{\theta} + \frac{1}{\theta^2}\frac{X_1^2+X_2^2-2\rho X_1X_2}{1-\rho^2}$$

so that $\hat{\theta}_\rho = [X_1^2+X_2^2-2\rho X_1X_2]/[1-\rho^2]$ is the maximizer and the profile likelihood function for ρ is

$$L(\rho) = \max_{\theta>0} L(\rho,\theta) = L(\rho,\hat{\theta}_\rho) = \frac{\sqrt{1-\rho^2}}{2\pi(X_1^2+X_2^2-2\rho X_1X_2)}\exp\left\{-\frac{1}{2}\right\}$$

15. The maximizer of the profile likelihood ratio $L(\rho)$ is given by the solution to

$$0 = \frac{\partial \log L(\rho)}{\partial\rho} = \frac{1}{2}\frac{-2\rho}{1-\rho^2} + \frac{2X_1X_2}{X_1^2+X_2^2-2\rho X_1X_2}$$

or $\hat{\rho} = 2X_1X_2/(X_1^2+X_2^2)$. The likelihood ratio statistic for testing $H_0: \rho = 0$ is then

$$\lambda \equiv \frac{L(0)}{L(\hat{\rho})} = \sqrt{1-\hat{p}^2}.$$

One rejects $H_0: \rho = 0$ (at a level $\alpha \in (0,1)$) if $\lambda < c_1$ for some $c_1 \in (0,1) \Longleftrightarrow \hat{p}^2 > 1-c_1^2 \Longleftrightarrow |\hat{p}| > c \equiv \sqrt{1-c_1^2}$.

16. One needs to find c so that $\sup_{\theta>0,\rho=0} P_{\theta,\rho}(|\hat{\rho}| > c) = 0.1$. Note that, under $H_0: \rho = 0$, the test statistic

$$\hat{\rho} = 2X_1X_2/(X_1^2+X_2^2) = 2Z_1Z_2/(Z_1^2+Z_2^2)$$

where $Z_1 = X_1/\sqrt{\theta}$ and $Z_2 = X_2/\sqrt{\theta}$ are iid $N(0,1)$ variables (where independent follows from $\rho = 0$). Hence, the distribution of $\hat{\rho}$ is free of θ under $H_o: \rho = 0$. One algorithm for numerically calibrating the test could be as follows:

- Set number of iterations M, e.g., $M = 10000$
- For $i = 1,\ldots,M$, generate iid $(U_{1i}, U_{2i}$ as uniform$(0,1)$, define $Z_{1i} = \Phi^{-1}(U_{1i})$ and $Z_{2i} = \Phi^{-1}(U_{2i})$ (using the standard normal cdf) and set $\hat{\rho}_i^* = 2Z_{1i}Z_{2i}/(Z_{1i}^2+Z_{2i}^2)$.
- Define c as the 90th sample percentile of $|\hat{\rho}_1^*|,\ldots,|\hat{\rho}_M^*|$.

Let $X_1,\ldots,X_n$ be iid $\mathrm{U}(\theta_1,\theta_2)$ random variables for $-\infty<\theta_1<\theta_2<\infty$. Use the notations $X_{(1)}$ and $X_{(n)}$ for respectively the smallest and largest order statistics of the sample.

1. Show that $\left(X_{(1)},X_{(n)}\right)$ is sufficient for $\boldsymbol{\theta}\equiv(\theta_1,\theta_2)$.

It can be shown that $\left(X_{(1)},X_{(n)}\right)$ is also a complete statistic. You may use this fact without proof in what follows if it is of help.

2. Derive the MLE for $\boldsymbol{\theta}$.

Define $\hat{\eta}\equiv\frac{1}{2}\left(X_{(1)}+X_{(n)}\right)$ and $\eta\equiv\frac{1}{2}(\theta_1+\theta_2)$.

3. Show that $\hat{\eta}$ is an unbiased estimator of η.

4. Is $\hat{\eta}$ a Minimum Variance Unbiased Estimator of η? Justify your answer.

5. Derive a Method of Moments estimator of η. (Call it $\hat{\eta}_{\mathrm{MM}}$.)

6. Derive the risk of $\hat{\eta}_{\mathrm{MM}}$ as an estimator of η under squared error loss.

7. Show that there is a positive number C so that under squared error loss, the risk of $\hat{\eta}$ as an estimator of η is no larger than Cn^{-2} for any $n\geq 2$. (You need not evaluate/provide a numerical value for C, only show that it exists.)

8. On the basis of your answer to Problem **6** and what is to be shown in Problem **7**, how do you expect the performances of $\hat{\eta}$ and $\hat{\eta}_{\mathrm{MM}}$ to compare for large n? Explain.

Now consider inference for θ_2 assuming that $\theta_1=0$, based on its MLE $\hat{\theta}_2=X_{(n)}$.

9. Derive the limiting distribution of $n\left(\theta_2-X_{(n)}\right)/\theta_2$ as $n\to\infty$.

10. Use your answer to Problem **9** and derive an approximately level $(1-\alpha)$ two-sided confidence interval for θ_2.

11. Derive an approximately $(1-\alpha)$ level two-sided confidence interval for θ_2 based on the large sample approximate normality of a method of moments estimator for θ_2. Compare the length of this interval to that from Problem **10** for large n.

12. Based on your answer to Problem **9**, formulate an approximately level α test of $\mathrm{H}_0{:}\theta_2 = \theta^*$ versus $\mathrm{H}_\mathrm{a}{:}\theta_2 \neq \theta^*$ (for a $\theta^* > 0$). Confirm the consistency of your test.

PhD Exam Major Theory II Key 2016

Let $X_1, \cdots, X_n$ be IID uniform(θ_1, θ_2) random variables with parameters $\theta_1, \theta_2 \in R$ and $\theta_1 < \theta_2$, and $X_{(1)}$ and $X_{(n)}$ be the smallest and the largest order statistics of the sample.
(1). By the factorization of the likelihood function,

$$f(X_1, \cdots, X_n|\theta) = (\theta_2 - \theta_1)^{-n} I(X_{(1)} > \theta_1) I(X_{(n)} < \theta_2).$$

It is readily shown that $(X_{(1)}, X_{(n)})$ is a sufficient statistic for $\theta = (\theta_1, \theta_2)$.
(2). The MLE for θ is the suff. stat $(X_{(1)}, X_{(n)})$.
(3). Detailed derivations by first obtaining the distributions of the two extreme order statistics leads to

$$\begin{aligned} E(X_{(n)}) &= n \int x f(x) F^{n-1}(x) dx = n(\theta_2 - \theta_1)^{-n} \int_{\theta_1}^{\theta_2} x(x - \theta_1)^{n-1} dx \\ &= \theta_1 + \frac{n}{n+1}(\theta_2 - \theta_1), \end{aligned}$$

Similarly, $E(X_{(1)}) = \frac{n}{n+1}\theta_1 - \left(\frac{n}{n+1} - 1\right)\theta_2$. This leads to that $\hat{\eta}$ is an unbiased estimator of η.
(4). Yes, $\hat{\eta}$ is a Minimum Variance Unbiased Estimator of θ due to the sufficiency, unbiasedness and the completeness of $(X_{(1)}, X_{(n)})$.
(5). As $E(X_i) = \eta$, a MM estimator of η is $\hat{\eta}_{MM} = \bar{X}_n$.
(6). As $\hat{\eta}_{MM}$ is unbiased, it's risk of estimation for η is $Var(\hat{\eta}) = \sigma^2/n = (\theta_2 - \theta_1)^2/(12n)$.
(7). As $\hat{\eta}$ is also unbiased, the risk is the variance as well. It can be shown that

$$\begin{aligned} E(X_{(n)}^2) &= n \int x^2 f(x) F^{n-1}(x) dx = n(\theta_2 - \theta_1)^{-n} \int_{\theta_1}^{\theta_2} x^2 (x - \theta_1)^{n-1} dx \\ &= \frac{n}{n+2}(\theta_2 - \theta_1)^2 + 2\theta_1 \frac{n}{n+1}(\theta_2 - \theta_1) + \theta_1^2, \end{aligned}$$

Hence, $Var(X_{(n)}) = \frac{n}{(n+1)^2(n+2)}(\theta_2 - \theta_1)^2 \le C_1 n^{-2}$ for some positive constant C_1. Similarly, it may be shown that $Var(X_1) \le C_2 n^{-2}$. According to Schwarz inequality, $|Cov(X_{(1)}, X_{(n)})| = \sqrt{Var(X_{(n)}) Var(X_1)} \le \sqrt{C_1 C_2} n^{-2}$. Hence, $Var(\hat{\eta}) \le (C_1 + 2\sqrt{C_1 C_2} + C_2) n^{-2}) = C n^{-2}$ for $C = (\sqrt{C_1} + \sqrt{C_2})^2$.
(8). From the results in (6) and (7), $Var(\hat{\eta}) = O(n^{-2})$ is a smaller order than that of the MM estimator, which is of order n^{-1}. This confirms that the $\hat{\eta}$ is the Mimimum Variance Unbiased estimator.

(9).

$$\begin{aligned} & P\{n(\theta_2 - X_{(n)})/\theta_2 \le x\} = P\left(X_{(n)} \ge \theta_2 - \frac{x\theta_2}{n}\right) = 1 - F^n(-\frac{x\theta_2}{n} + \theta_2) \\ = & \ 1 - (1 - \frac{x}{n})^n \to 1 - e^{-x}. \end{aligned}$$

Hence, the asymptotic distribution of $n(\theta_2 - X_{(n)})/\theta_2$ is the standard exponential.
(10). Let $\xi_{\alpha/2}$ and $\xi_{1-\alpha/2}$ be respectively the $\alpha/2$ and $1-\alpha/2$ quantiles of the unit exp. distribution. Then, an approximate $1-\alpha$ confidence interval for θ_2 can be obtained by

$$\xi_{\alpha/2} < n(\theta_2 - X_{(n)})/\theta_2 < \xi_{1-\alpha/2}$$

which is equivalent to

$$\frac{X_{(n)}}{1 - n^{-1}\xi_{\alpha/2}} < \theta_2 < \frac{X_{(n)}}{1 - n^{-1}\xi_{1-\alpha/2}}.$$

(11). The confidence interval is $\bar{X}_n \pm z_{1-\alpha/2}\frac{s}{\sqrt{n}}$, where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $N(0,1)$, and s is the sample standard deviation. The length of the confidence interval is $2z_{1-\alpha/2}\frac{s}{\sqrt{n}}$, which is at the order of $n^{1/2}$. In contrast the length of the CI in (10) is at the order of n^{-1}, which is one order of magnitude smaller.
(12). The test statistic is $T_n(\theta_{20}) =: n(1 - X_{(n)}/\theta^*)$. We reject H_0 if either $T_n > \xi_{1-\alpha/2}$ or $T_n < \xi_{\alpha/2}$. The asymptotic distribution in (9) ensures its size converges to α as $n \to \infty$. The power of the test under H_1 is

$$\begin{aligned}
& P\left(n\frac{\theta_2 - X_{(n)} + \theta^* - \theta_2)}{\theta_2} > \xi_{1-\alpha/2}\theta^*/\theta_2\right) \\
+ \quad & P\left(n\frac{\theta_2 - X_{(n)} + \theta^* - \theta_2)}{\theta_2} < \xi_{\alpha/2}\theta^*/\theta_2\right) \\
= \quad & P\left(T(\theta_2) > n\frac{(\theta - \theta^*)}{\theta_2} + \xi_{1-\alpha/2}\theta^*/\theta_2\right) \\
+ \quad & P\left(T(\theta_2) < n\frac{(\theta - \theta^*)}{\theta_2} + \xi_{\alpha/2}\theta^*/\theta_2\right).
\end{aligned}$$

It can be argued that under H_1 one of the probability on the RHS approaches to 1 as $n \to \infty$, which confirms the consistency of the test.

Part I

Let Ω be a non-empty set. Suppose $\mathcal{C}$ is a finite partition of Ω, that is, for some $k < \infty$,

$$\mathcal{C} \equiv \{A_1, \ldots, A_k\}, \quad \Omega = \cup_{i=1}^k A_i, \quad A_i \cap A_j = \emptyset, \text{ for } i \neq j.$$

1. Show that the minimal algebra $\mathcal{A}(\mathcal{C})$ generated by $\mathcal{C}$ is the class of unions of subfamilies of $\mathcal{C}$, that is, show that

$$\mathcal{A}(\mathcal{C}) = \{\cup_{j \in I} A_j : I \subset \{1, 2, \ldots, k\}\}.$$

(This includes the empty set $I = \emptyset$ also.)

2. What is the σ-algebra generated by the partition $\mathcal{C}$?

3. Let $\mathcal{C}' \equiv \{A_1, A_2, \ldots\}$ be a countably infinite partition of Ω, that is,

$$\Omega = \cup_{i=1}^\infty A_i, \quad A_i \cap A_j = \phi, \text{ for } i \neq j.$$

Find the σ-algebra generated by the partition $\mathcal{C}'$.

Part II

Let Ω be a non-empty set. Let $\mathcal{F}$ be the collection of all subsets such that either A or A^c is finite, that is, $\mathcal{F} \equiv \{A \subset \Omega : \text{either } A \text{ or } A^c \text{ is finite}\}$.

4. Show that $\mathcal{F}$ is an algebra.

For $E \in \mathcal{F}$, define the set function P by

$$P(E) = \begin{cases} 0, & \text{if } E \text{ is finite,} \\ 1, & \text{if } E^c \text{ is finite.} \end{cases}$$

5. If Ω is countably infinite, show that P is finitely additive on $\mathcal{F}$.

6. Find an example of an algebra that is not a σ-algebra.

(Hint: You may consider $\mathcal{F}$ corresponding to a countably infinite Ω.)

Part III

Suppose the sequence of random variables $\{X_n, n \geq 1\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ converges in probability to 0, that is, $X_n \xrightarrow{P} 0$ as $n \to \infty$.

7. Show by an example that $\{X_n\}$ need not converge to 0 almost surely.

8. Show that there is a subsequence $\{X_{n_k}\}$ that converges to 0 almost surely as $k \to \infty$.

Part IV

Let $\{X_n, n \geq 1\}$ be iid random variables on a probability space $(\Omega, \mathcal{F}, P)$. Let $S_n = \sum_{j=1}^n X_j$ for $n \geq 1$ and $p \in (0, \infty)$ be a constant.

9. Assume $E|X_1|^p = \infty$. Show that

$$\limsup_{n\to\infty} \frac{|S_n|}{n^{1/p}} = \limsup_{n\to\infty} \frac{|X_n|}{n^{1/p}} = \infty \text{ almost surely.}$$

10. Assume $E|X_1|^p < \infty$ for all $0 < p < 2$. In addition for cases where $1 \leq p < 2$, assume $EX_1 = 0$. Show that for $p \in (0, 2)$

$$\lim_{n\to\infty} \frac{S_n}{n^{1/p}} = \lim_{n\to\infty} \frac{X_n}{n^{1/p}} = 0 \text{ almost surely.}$$

Part V

11. Let $S_n = \sum_{j=1}^n X_j$ for $n \geq 1$, where $\{X_n, n \geq 1\}$ is a sequence of independent mean-zero random variables. Let $\{\lambda_n, n \geq 1\}$ be a bounded sequence of positive constants. Suppose that

$$E|X_j|^{2+\lambda_n} < \infty, \ \ 1 \leq j \leq n, \ n \geq 1,$$

and

$$\lim_{n\to\infty} \frac{\sum_{j=1}^n E|X_j|^{2+\lambda_n}}{v_n^{2+\lambda_n}} = 0,$$

where $v_n^2 = \sum_{j=1}^n EX_j^2$, for $n \geq 1$.

Prove that

$$\frac{S_n}{v_n} \xrightarrow{d} Z,$$

as $n \to \infty$ where $Z \sim N(0, 1)$.

Part VI

Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables with

$$P(X_n = 1) = \frac{1}{2}\left(1 - \frac{1}{n^2}\right) = P(X_n = -1), \ \ P(X_n = n) = \frac{1}{2n^2} = P(X_n = -n),$$

for all $n \geq 1$. Let $S_n = \sum_{j=1}^n X_j, n \geq 1$.

12. Show that $\text{Var}(S_n/\sqrt{n}) \to 2$ as $n \to \infty$.

Next, for $n \geq 1$ let

$$Y_n = \begin{cases} 1, & \text{if } X_n = 1 \text{ or } n, \\ -1, & \text{if } X_n = -1 \text{ or } -n. \end{cases}$$

Let $T_n = \sum_{j=1}^{n} Y_j, n \geq 1$. Let $Z \sim N(0,1)$.

13. Show that $T_n/\sqrt{n} \xrightarrow{d} Z$ as $n \to \infty$.

14. Show that $\sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty$.

15. Using the results in problems **13** and **14**, show that $S_n/\sqrt{n} \xrightarrow{d} Z$ as $n \to \infty$.

1. Let

$$\mathcal{A} = \{\cup_{j \in I} A_j : I \subset \{1, 2, \ldots, k\}\}.$$

We first show that $\mathcal{A}$ is an algebra. To do this, note that $\Omega \in \mathcal{A}$ by taking $I = \{1, 2, \ldots, k\}$. If $A = \cup_{j \in I} A_j \in \mathcal{A}$ then $A^c = \cup_{j \in I^c} A_j \in \mathcal{A}$. If $A^i = \cup_{j \in I_i} A_j \in \mathcal{A}, i = 1, 2$ then $A^1 \cup A^2 = \cup_{j \in I_1 \cup I_2} A_j \in \mathcal{A}$.

So $\mathcal{A}$ is an algebra and $\mathcal{C} \subset \mathcal{A}$. Thus by minimality we have $\mathcal{A}(\mathcal{C}) \subset \mathcal{A}$. But since $A_j \in \mathcal{C} \subset \mathcal{A}(\mathcal{C})$, we have $\mathcal{A} \subset \mathcal{A}(\mathcal{C})$. Thus $\mathcal{A} = \mathcal{A}(\mathcal{C})$

2. We have $\sigma(\mathcal{C}) = \mathcal{A}(\mathcal{C})$.

3. Using the argument in 1., we have $\sigma(\mathcal{C}') = \{\cup_{j \in I} A_j : I \subset \{1, 2, \ldots, \}\}$.

4. Since $\Omega^c = \phi$ is finite, $\Omega \in \mathcal{F}$. If $A \in \mathcal{F}$, then either A or A^c is finite. Therefore, $A^c \in \mathcal{F}$ since either $A = (A^c)^c$ or A^c is finite. Finally, suppose $A_i \in \mathcal{F}, i = 1, 2$. If one of A_1, A_2 is finite, then $A_1 \cap A_2$ is finite and hence $A_1 \cap A_2 \in \mathcal{F}$. If neither set is finite, then A_1^c and A_2^c are finite, so $A_1^c \cup A_2^c$ is finite. Therefore $(A_1^c \cup A_2^c)^c = A_1 \cap A_2 \in \mathcal{F}$.

5. Let $A_1, \ldots, A_k \in \mathcal{F}, A_i \cap A_j = \phi$, for $i \neq j$. At most one of these sets can be infinite, since if A_1 and A_2 are both infinite, $A_1 \cap A_2 = \phi$, then A_1^c, A_2^c are finite, which implies $A_1^c \cup A_2^c$ is finite. So $(A_1^c \cup A_2^c)^c$ is infinite and in $\mathcal{F}$. However, we also have $(A_1^c \cup A_2^c)^c = A_1 \cap A_2 = \phi$, which gives a contradiction.

If none of $A_1, A_2, \ldots, A_k$ is infinite, then

$$P(\cup_{i=1}^k A_i) = 0 = \sum_{i=1}^k P(A_i).$$

If exactly one is infinite, then $\cup_{i=1}^k A_i$ is infinite and

$$P(\cup_{i=1}^k A_i) = 1 = \sum_{i=1}^k P(A_i),$$

since the right hand side is a sum of $(k-1)$ zeros and one 1.

6. Let $\Omega = \mathbb{N}$, the set of natural numbers. The set $2\mathbb{N}$ consisting of all even positive integers is a countable union of sets in $\mathcal{F}$, but is not in $\mathcal{F}$.

7. Consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$, where λ is the Lebesgue measure. Consider the sequence of random variables

$$X_1 = I_{[0,1]}, X_2 = I_{[0,\frac{1}{2}]}, X_3 = I_{(\frac{1}{2},1]}, X_4 = I_{[0,\frac{1}{3}]}, X_5 = I_{(\frac{1}{3},\frac{2}{3}]}, \ldots$$

Then $X_n \xrightarrow{P} 0$ as $n \to \infty$. But, $X_n \overset{a.s.}{\not\to} 0$.

8. Choose $Y_k = X_{n_k}$ such that
$$P(|Y_k| > \frac{1}{k}) < \frac{1}{2^k}.$$
Then given $\epsilon > 0$,
$$\sum_k P(|Y_k| > \epsilon) < \infty.$$
Hence by Borel-Cantelli,
$$P(|Y_k| > \epsilon \ i.o.) = 0,$$
implying $X_{n_k} \stackrel{a.s.}{\to} 0$ as $k \to \infty$.

9. Since $E|X_1|^p = \infty$, for all $M > 0$,
$$\sum_{n=1}^{\infty} P\Big(\frac{|X_1|}{M} > n^{1/p}\Big) = \infty,$$
which implies, as X_i's are iid
$$\sum_{n=1}^{\infty} P\Big(\frac{|X_n|}{M} > n^{1/p}\Big) = \infty.$$
By Borel-Cantelli Theorem,
$$P\Big(\frac{|X_n|}{n^{1/p}} > M \ i.o. \ (n)\Big) = 1,$$
implying
$$P\Big(\limsup \frac{|X_n|}{n^{1/p}} > M\Big) = 1.$$
Thus
$$P\Big(\limsup \frac{|X_n|}{n^{1/p}} = \infty\Big) = P\Big(\cap_{M=1}^{\infty}\Big[\limsup \frac{|X_n|}{n^{1/p}} > M\Big]\Big) = \lim_{M\to\infty} P\Big(\limsup \frac{|X_n|}{n^{1/p}} > M\Big) = 1.$$
Since
$$\frac{|X_n|}{n^{1/p}} = \frac{|S_n - S_{n-1}|}{n^{1/p}} \le \frac{|S_n|}{n^{1/p}} + \frac{|S_{n-1}|}{(n-1)^{1/p}}\Big(\frac{n-1}{n}\Big)^{1/p},$$
$$\limsup \frac{|X_n|}{n^{1/p}} \le 2 \limsup \frac{|S_n|}{n^{1/p}}$$
Thus
$$P\Big(\limsup \frac{|S_n|}{n^{1/p}} = \infty\Big) = 1.$$

10. Since $E|X_1|^p < \infty$, for all $M > 0$,

$$\sum_{n=1}^{\infty} P\Big(\frac{|X_1|}{M} > n^{1/p}\Big) < \infty,$$

which implies, as X_i's are iid

$$\sum_{n=1}^{\infty} P\Big(\frac{|X_n|}{M} > n^{1/p}\Big) < \infty.$$

By Borel-Cantelli Theorem,

$$P\Big(\frac{|X_n|}{n^{1/p}} > M \ i.o.\ (n)\Big) = 0,$$

implying

$$P\Big(\limsup \frac{|X_n|}{n^{1/p}} \leq M\Big) = 1.$$

Thus

$$P\Big(\limsup \frac{|X_n|}{n^{1/p}} = 0\Big) = 1.$$

By Marcinkiewz-Zygmund SLLN,

$$\frac{|S_n|}{n^{1/p}} \to 0, \quad \text{almost certainly}(P).$$

11. Note that $\{X_n\}$ satisfies the Lindberg condition as for $\epsilon > 0$

$$\frac{\sum_{j=1}^n E|X_j|^2 I(|X_j| > \epsilon v_n)}{v_n^2} \leq \frac{1}{v_n^2}\sum_{j=1}^n E|X_j|^2\Big(\frac{|X_j|}{\epsilon v_n}\Big)^{\lambda_n} I(|X_j| > \epsilon v_n) \leq \frac{\sum_{j=1}^n E|X_j|^{2+\lambda_n}}{\epsilon^{\lambda_n} v_n^{2+\lambda_n}} \to 0.$$

Thus

$$\frac{S_n}{v_n} \xrightarrow{d} N(0,1).$$

12. We have $EX_n = 0$,

$$EX_n^2 = 1 - \frac{1}{n^2} + \frac{n^2}{n^2} = 2 - \frac{1}{n^2} \to 2,$$

and X_n's are independent. Thus, by Cesaro mean summability theorem,

$$Var\Big(\frac{S_n}{\sqrt{n}}\Big) = \frac{1}{n}\sum_{k=1}^n \Big(2 - \frac{1}{k^2}\Big) \to 2.$$

13. Since Y_n's are iid Bernoulli random variables with

$$P(Y_n = 1) = \frac{1}{2} = P(Y_n = -1),$$

we have $T_n/\sqrt{n} \to N(0,1)$ as $n \to \infty$.

14. We have

$$\sum_{n=1}^{\infty} P(X_n \neq Y_n) = \sum_{n=1}^{\infty} P(X_n = \pm n) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

15. Since $\{X_n\}$ and $\{Y_n\}$ are tail equivalent, from **13** it follows that $S_n/\sqrt{n} \to N(0,1)$ as $n \to \infty$.

Part I

During World War II, the Allies wanted an accurate estimation of the number of German Panther tanks. Assume there are a total of n German Panther tanks. Shortly before D-Day, rumors indicated that a large numbers of Panthers were being used. To ascertain if this was true the Allies attempted to estimate the number of tanks being produced based on the serial numbers on k captured or destroyed tanks. The numbers used were gearbox numbers $Y_1, Y_2, \dots, Y_k \in \{1,2,\dots,n\}$. Note that n and k are positive integers.

1. Assume that each gearbox serial number is equally likely to be observed. Show that the MLE of n is given by $\hat{n} = \max(Y_1, \dots, Y_k)$.

2. Find the cumulative distribution function (cdf) of $\hat{n}$.

3. What is the probability mass function (pmf) of $\hat{n}$?

4. Show that $\hat{n}$ is a biased estimator of n.

5. Provide an unbiased estimator of n that is a function of $\hat{n}$.

6. Find the variance for the estimator from Problem **5**.

   Hint:

$$\sum_{r=k}^{n} r\binom{r}{k} = \frac{(n+k+kn)\Gamma(n+2)}{\Gamma(n+3)\Gamma(n-k+1)}$$

7. Show that the estimator established in Problem 5 is a uniformly minimum variance unbiased estimator (UMVUE) of n.

Part II

Let $X_1, X_2, \dots, X_n$ be independent and identically distributed (iid) random variables with density function $f(\cdot)$. Let $Y_n = \max(X_1, \dots, X_n)$ and $Z_n = \min(X_1, \dots, X_n)$.

8. Calculate the (marginal) density functions (pdfs) of Y_n and Z_n .

9. Calculate the joint density function (pdf) of Y_n and Z_n

10. Let

$$f(x) = \begin{cases} -\frac{1}{\theta}, & \theta \le x \le 0 \\ 0, & \text{otherwise} \end{cases}$$

   a. Calculate the Method of Moments estimator for θ.
   b. Calculate the Maximum Likelihood Estimator (MLE) for θ.
   c. Derive an exact $100(1-\alpha)\%$ confidence interval for θ based on the MLE (DO NOT use large sample ML theory).
   d. Suggest an unbiased estimator of θ that is a function of the MLE.

11. Let

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & \text{otherwise} \end{cases}$$

a. Calculate the density function of the sample range $R_n = Y_n - Z_n$.
b. Derive an exact $100(1-\alpha)\%$ confidence interval for θ based on the sample range R_n.

Part I

1. Note that there are $\binom{n}{k}$ sets of k distinct numbers which are subsets of $\{1,2,\dots,n\}$. Since we can assume that each gearbox serial is equally likely to be observed, we have that the probability mass function (as a function of n) is given by

$$p_Y(y;n) = \begin{cases} \dfrac{1}{\binom{n}{k}}, & \text{if } \{Y_1, Y_2, \dots, Y_k\} \in \{1,2,\dots,n\} \\ 0, & \text{otherwise.} \end{cases}$$

The total number of tanks must be greater than or equal to the largest observed serial number, $M_k = \max(Y_1, Y_2, \dots, Y_k)$. Therefore, we can write the pmf as follows

$$p_Y(y;n) = \begin{cases} \dfrac{1}{\binom{n}{k}}, & \text{if } n \geq M_k = \max(Y_1, Y_2, \dots, Y_k) \\ 0, & \text{otherwise.} \end{cases}$$

It is immediately clear that the MLE is given by $\hat{n} = M_k = \max(Y_1, Y_2, \dots, Y_n)$. Figure 1 illustrates this graphically.

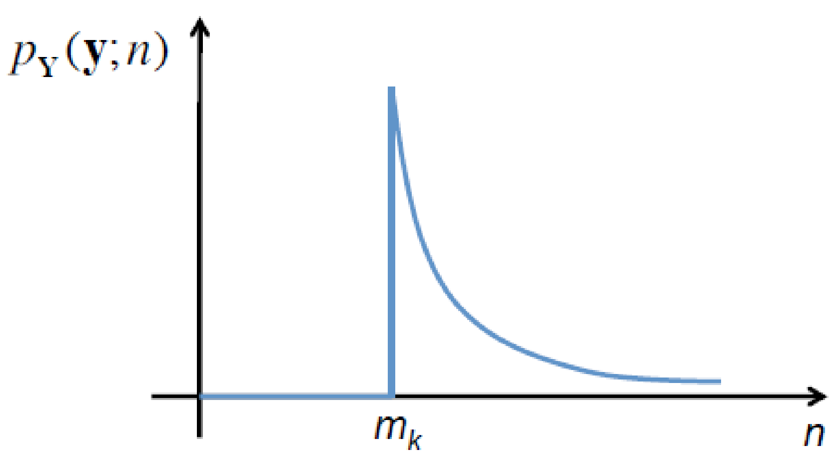

Figure 1. Graphical illustration of the pmf vs. the number of tanks n.

2. We already know the PMF of the gearbox serial numbers. The cumulative distribution function of the maximum of the gearbox numbers is given by

$$F_{M_k}(r) = P(M_k \le r) = P(Y_1 \le r, \dots, Y_k \le r) = \frac{1}{\binom{n}{k}} \times No.\, subsets\ of\ \{1,2,\dots,r\} of size\ k$$

$$= \begin{cases} 0, \ \text{if}\, r \le k-1 \\ \frac{\binom{r}{k}}{\binom{n}{k}}, \ \text{if}\, r = k, k+1, \dots, n \\ 1, \ \text{if}\, r \ge n \end{cases}$$

3. The PMF of the MLE of n is given by

$$p_{M_k}(r) \ = \ P(M_k = r) = P(M_k \le r) - P(M_k \le r-1) = F_{M_k}(r) - F_{M_k}(r-1).$$

Hence, we have

$$p_{M_k}(r) = \begin{cases} 0, & \text{if}\, r \le k-1 \text{ and } r \ \ge \ n+1 \\ \frac{\binom{r}{k} - \binom{r-1}{k}}{\binom{n}{k}} = \frac{\binom{r-1}{k-1}}{\binom{n}{k}}, & \text{if}\, r = k, k+1, \dots, n \end{cases}$$

4. Since p_{M_k} is a pmf, we have that

$$1 = \sum_{r=k}^{n} p_{M_k}(r) = \sum_{r=k}^{n} \frac{\binom{r-1}{k-1}}{\binom{n}{k}} = \frac{1}{\binom{n}{k}} \sum_{r=k}^{n} \binom{r-1}{k-1}$$

Hence

$$\sum_{r=k}^{n} \binom{r-1}{k-1} = \binom{n}{k}$$

Let's calculate the expected value of the MLE estimator

$$E(M_k) = \sum_{r=k}^{n} r \frac{\binom{r-1}{k-1}}{\binom{n}{k}} = \sum_{r=k}^{n} r \frac{\frac{(r-1)!}{(k-1)!\,(r-k)!}}{\binom{n}{k}} = \sum_{r=k}^{n} \frac{\frac{k\,r!}{k!\,(r-k)!}}{\binom{n}{k}} = \frac{k}{\binom{n}{k}} \sum_{r=k}^{n} \binom{r}{k}$$

Using the previous identity, we can rewrite the above equation as

$$E(M_k) = \frac{k}{\binom{n}{k}} \binom{n+1}{k+1} = \frac{k(n+1)}{k+1}$$

The bias of the MLE is then given by

$$bias(M_k) = E(M_k) - n = \frac{k(n+1)}{k+1} - n = \frac{k-n}{k+1}$$

5. A bias corrected estimator is

$$\widehat{N_k} = \frac{k+1}{k} M_k - 1,$$

Indeed

$$E(\widehat{N_k}) = \frac{k+1}{k} E(M_k) - 1 = \frac{k+1}{k} \frac{k(n+1)}{k+1} - 1 = n$$

6.

$$
\begin{aligned}
Var(\widehat{N_k}) &= \frac{(k+1)^2}{k^2} Var(M_k) \\
&= \frac{(k+1)^2}{k^2} \left[E(M_k^2) - E^2(M_k) \right] \\
&= \frac{(k+1)^2}{k^2} \left[\sum_{r=k}^{n} r^2 \frac{\binom{r-1}{k-1}}{\binom{n}{k}} - \frac{k^2(n+1)^2}{(k+1)^2} \right] \\
&= \frac{(k+1)^2}{k^2} \left[\frac{k}{\binom{n}{k}} \sum_{r=k}^{n} r \binom{r}{k} - \frac{k^2(n+1)^2}{(k+1)^2} \right]
\end{aligned}
$$

$$= \frac{(k+1)^2}{k^2} \left[\frac{k(n-k)!\,k!}{n!} \frac{(n+k+kn)\Gamma(n+2)}{\Gamma(n+3)\Gamma(n-k+1)} - \frac{k^2(n+1)^2}{(k+1)^2} \right]$$

Using $\Gamma(n+1) = n\Gamma(n)$ and for any natural number $\Gamma(n+1) = n!$ yields

$$Var(\widehat{N_k}) = \frac{1}{k}\frac{(n-k)(n+1)}{k+2}$$

7. To show that this is the UMVUE, we need to
   a. first show that the sample maximum is a sufficient statistic for the population maximum

      This example is described by saying that a sample of k observations is obtained from a uniform distribution on the integers $1,2,\dots,n$ i.e., $U[0,\theta]$ and $\theta = \binom{n}{k}$, with the problem being to estimate the unknown maximum n. Consider the joint probability density function of the serial numbers $Y_1, Y_2, \dots, Y_n$. Since the observations are assumed to be independent, the PDF can be written as a product of individual densities (exclude the outcomes in which a serial number occurs twice in the sample):

$$f_Y(y_1,\dots,y_n) = \frac{1}{\theta} I_{\{0\le y_1\le\theta\}} \times \cdots \times \frac{1}{\theta} I_{\{0\le y_n\le\theta\}}$$

$$= \frac{1}{\theta^n} I_{\{0\le \min\limits_{i=1,\dots,n}(y_i)\}} I_{\{\max\limits_{i=1,\dots,n}(y_i)\le\theta\}}$$

      where $I_{\{\cdot\}}$ Is the indicator function. According to the Factorization Theorem of Fisher and Neyman we have that, with $h(y_1,\dots,y_n) = I_{\{0\le \min\limits_{i=1,\dots,n}(y_i)\}}$ and the rest of the expression is a function of θ and consequently, the sample maximum is a sufficient statistic.

   b. Second, show that the family of densities is complete. We need to show to show that for every function $u(x)$ not depending on θ, it holds that $E_\theta\big(u(T)\big) = 0$ for all θ implies $P_\theta(u(T)=0) = 1$ for all θ. The family of densities is $\{g(y;\theta)|\theta > 0\}$, where

$$g(y;\theta) = \begin{cases} \frac{1}{\theta} \text{ if } 0 \le y \le \theta \\ 0 \text{ if otherwise} \end{cases}$$

$$E_\theta\big(u(T)\big) = 0 \text{ for all } \theta > 0$$

$$\Leftrightarrow \int_0^\theta u(y)\tfrac{1}{\theta}\,dy = 0 \text{ for all } \theta > 0$$

      Hence $u(y) = 0$ for all $y > 0$ and consequently this family of densities is complete.

   c. Since $\frac{k+1}{k}M_k - 1$ is a complete sufficient statistic for n (see part a and b), unbiased and has finite variance, Lehmann–Scheffé theorem states that the estimator $\frac{k+1}{k}M_k - 1$, is the UMVUE estimator and is unique.

Part II

8. Density function of Y_n

$$P(Y_n \le x) = P(X_1 \le x, \dots X_n \le x)$$
$$= P(X_1 \le x) \cdots P(X_n \le x)$$
$$= [F(x)]^n$$

And consequently: $f_{Y_n}(x) = n[F(x)]^{n-1} f(x)$

Density function of Z_n

$$P(Z_n \le x) = 1 - P(Z_n \ge x)$$
$$= 1 - P(Z_1 \ge x) \cdots P(Z_n \ge x)$$
$$= 1 - [1 - F(x)]^n$$

And $f_{Z_n}(x) = n[1 - F(x)]^{n-1} f(x)$

9. $P(Y_n \le x, Z_n \le y) = P(Y_n \le y) - P(Y_n \le y, Z_n > z)$

$$= P(Y_n \le y) - P(z < X_1, \dots, X_n \le y)$$
$$= \begin{cases} [F(y)]^n & \text{if } z > y \\ [F(y)]^n - [F(y) - F(z)]^n & \text{if } z \le y \end{cases}$$

And

$$f_{Y_n, Z_n}(y, z) = \begin{cases} 0, & z > y \\ n(n-1)[F(y) - F(z)]^{n-2} f(y) f(z), & z \le y \end{cases}$$

10. If $X \sim U[\theta, 0]$ then $f(x) = \begin{cases} \frac{1}{|\theta|}, & \theta \le x \le 0 \\ 0, & \text{otherwise} \end{cases}$

a. The method of moments estimator is given by noticing that $(X) = \frac{\theta}{2}$. Using the Law of large numbers yields $\hat{\theta}_{MoM} = \frac{2}{n} \sum_{i=1}^{n} X_i$

b. The log likelihood function is given by

$$l(\theta) = \begin{cases} \prod_{i=1}^{n} \frac{1}{|\theta|}, & \theta \le \min(X_1, \dots, X_n) \text{ and } \max(X_1, \dots, X_n) \le 0 \\ 0, & \text{otherwise} \end{cases}$$

Then, the MLE is given by $\hat{\theta}_{MLE} = \min(X_1, \dots, X_n)$.

c. First, determine the distribution of the minimum. Let M denote the minimum of the sample, then

$$P(M \le x) = 1 - [1 - F(x)]^n$$

For $X \sim U[\theta, 0]$, the distribution of the minimum is given by

$$P(M \le x) = 1 - \left[1 - \frac{x - \theta}{-\theta}\right]^n = 1 - \frac{x^n}{\theta^n}$$

Consider the following probability and for $X \sim U[\theta, 0]$ (remember $\theta < 0$)

$$P\left(\frac{M}{\theta} \le x\right) = P(M \ge \theta x) = 1 - P(M \le \theta x) = 1 - 1 + \frac{\theta^n x^n}{\theta^n} = x^n$$

for $\theta \le x \le 0$, and 0 otherwise . Now, we find the find an "a" and a "b" such that

$$P\left(a \le \frac{M}{\theta} \le b\right) = P\left(\frac{M}{a} \le \theta \le \frac{M}{b}\right) = b^n - a^n = 1 - \alpha$$

Since $\theta \le M$, we can set $b = 1$ and consequently $a = \sqrt[n]{\alpha}$.

d. The density of the minimum for $X \sim U[\theta, 0]$ is given by $f_M(x) = -n\frac{x^{n-1}}{\theta^n}$ and hence

$$E(M) = \int_{\theta}^{0} -n\frac{x^n}{\theta^n} dx = \frac{n}{\theta^n}\frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta$$

And therefore a bias corrected ML estimator is

$$\hat{\theta}_{MLE,corrected} = \frac{n+1}{n}\min(X_1, \dots, X_n)$$

11. If $X \sim U[0,\theta]$ then $f(x) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & \text{otherwise} \end{cases}$ and $F(x) = \begin{cases} 0, & x \le 0 \\ \frac{x}{\theta}, & 0 \le x \le \theta \\ 1, & x \ge \theta \end{cases}$

a. Therefore, the joint density of the minimum and maximum for $U[0,\theta]$ is given by

$$f_{Y_n,Z_n}(y,z) = \begin{cases} \frac{n(n-1)}{\theta^n}(y-z)^{n-2}, & 0 \le z \le y \le \theta \\ 0, & \text{otherwise} \end{cases}$$

In order to calculate the density function of the sample range, define the following relation

$$\begin{cases} S_n = Y_n \\ R_n = Y_n - Z_n \end{cases}$$

The transformation and its inverse are then given by

$$\begin{cases} s = y \\ r = y - z \end{cases} \quad \Leftrightarrow \quad \begin{cases} y = s \\ z = s - r \end{cases}$$

The determinant of the Jacobian J is given by

$$\det J = \det \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} = -1.$$

The joint density of S_n and R_n is given by

$$f_{S_n,R_n}(s,r) = f_{Y_n,Z_n}(s, s-r) \cdot 1 \quad \text{if } 0 \le r \le s \le \theta.$$

The marginal density of R_n is

$$f_{R_n}(r) = \begin{cases} \int_r^\theta f_{S_n,R_n}(s,r)\,ds, & 0 \le r \le \theta \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{n(n-1)}{\theta^n} r^{n-2}(\theta - r), & 0 \le r \le \theta \\ 0, & \text{otherwise} \end{cases}$$

b. We consider the following ratio

$$\frac{n(\theta - R_n)}{\theta}$$

and determine its limit distribution. Then, we have that

$$P\left(\frac{n(\theta - R_n)}{\theta} \le x\right) = P\left(R_n \ge \theta\left(1 - \frac{x}{n}\right)\right)$$

$$= \int_{\theta(1-\frac{x}{n})}^{\theta} \frac{n(n-1)}{\theta^n} r^{n-2}(\theta - r)dr \ \text{ if } 0 \le x \le n$$

$$= 1 - n\left(1 - \frac{x}{n}\right)^{n-1} + (n-1)\left(1 - \frac{x}{n}\right)^{n-1}$$

$$= 1 - \left(1 - \frac{x}{n}\right)^{n-1} + \frac{n-1}{n} x \left(1 - \frac{x}{n}\right)^{n-1}$$

For $n \to \infty$ we have

$$P\left(\frac{n(\theta - R_n)}{\theta} \le x\right) = 1 - e^{-x} - xe^{-x}$$

which is the distribution of a Gamma with parameters 2 and 1. Therefore, an exact $100(1-\alpha)\%$ confidence interval for θ based on the sample range R_n is given by

$$P\left(\Gamma_{(2,1);\alpha/2} \le \frac{n(\theta - R_n)}{\theta} \le \Gamma_{(2,1);1-\alpha/2}\right) = 1 - \alpha$$

$$P\left(\frac{\Gamma_{(2,1);\alpha/2}}{n} \le 1 - \frac{R_n}{\theta} \le \frac{\Gamma_{(2,1);1-\alpha/2}}{n}\right) = 1 - \alpha$$

$$P\left(\frac{\Gamma_{(2,1);\alpha/2} - n}{n} \le -\frac{R_n}{\theta} \le \frac{\Gamma_{(2,1);1-\alpha/2} - n}{n}\right) = 1 - \alpha$$

$$P\left(\frac{nR_n}{n - \Gamma_{(2,1);\alpha/2}} \le \theta \le \frac{nR_n}{n - \Gamma_{(2,1);1-\alpha/2}}\right) = 1 - \alpha$$