Stats 743b - Assignment 2

Tommy Flynn (400121496)

May 11, 2023

Question 1:

Suppose $\mathbf{C} = ((c_{ij}))_{i,j=1}^{k}$ is a symmetric non-singular product-decomposable matrix with elements of the form $c_{ij} = a_i b_j$. Then we have $\mathbf{C}^{-1}$ as a tri-diagonal matrix with elements $i \leq j$ as below.

$$c^{ij} = \begin{cases} \frac{a_2}{a_1(a_2 b_1 - a_1 b_2)} & \text{for } i = j = 1 \\ \frac{a_{i+1} b_{i-1} - a_{i-1} b_{i+1}}{(a_i b_{i-1} - a_{i-1} b_i)(a_{i+1} b_i - a_i b_{i+1})} & \text{for } 2 \leq i = j \leq k-1 \\ \frac{b_{k-1}}{b_k(a_k b_{k-1} - a_{k-1} b_k)} & \text{for } i = j = k \\ -\frac{1}{a_{i+1} b_i - a_i b_{i+1}} & \text{for } j = i+1 \text{ and } 1 \leq i \leq k-1 \\ 0 & \text{for } j > i+1 \end{cases} \tag{1}$$

To show this we check $\mathbf{C}\mathbf{C}^{-1} = \mathbf{I}$ by cases, multiplying rows of $\mathbf{C}$ by columns of $\mathbf{C}^{-1}$.

Case 1: Row 1 $\times$ Col 1

$$a_1 b_1 \cdot \frac{a_2}{a_1(a_2 b_1 - a_1 b_2)} + a_1 b_2 \cdot \frac{-1}{a_2 b_1 - a_1 b_2} = \frac{a_2 b_1 - a_1 b_2}{a_2 b_1 - a_1 b_2} = 1$$

Case 2: Row k $\times$ Col k

$$a_k b_k \cdot \frac{b_{k-1}}{b_k(a_k b_{k-1} - a_{k-1} b_k)} + a_{k-1} b_k \cdot \frac{-1}{a_k b_{k-1} - a_{k-1} b_k} = \frac{a_k b_{k-1} - a_{k-1} b_k}{a_k b_{k-1} - a_{k-1} b_k} = 1$$

Case 3: Row i $\times$ Col i, $1 < i < k$

$$a_i b_i \cdot \frac{a_{i+1} b_{i-1} - a_{i-1} b_{i+1}}{(a_i b_{i-1} - a_{i-1} b_i)(a_{i+1} b_i - a_i b_{i+1})} + a_{i-1} b_i \cdot \frac{-1}{a_i b_{i-1} - a_{i-1} b_i} + a_i b_{i+1} \cdot \frac{-1}{a_{i+1} b_i - a_i b_{i+1}}$$
$$= \frac{a_i a_{i+1} b_i b_{i-1} - a_{i-1} a_{i+1} b_i^2 + a_i a_{i-1} b_i b_{i+1} - a_i^2 b_{i-1} b_{i+1}}{a_i a_{i+1} b_i b_{i-1} - a_{i-1} a_{i+1} b_i^2 + a_i a_{i-1} b_i b_{i+1} - a_i^2 b_{i-1} b_{i+1}} = 1$$

Case 4: Row i $\times$ Col j, $1 \leq i < j < k$

$$a_i b_j \cdot \frac{a_{j+1} b_{j-1} - a_{j-1} b_{j+1}}{(a_j b_{j-1} - a_{j-1} b_j)(a_{j+1} b_j - a_j b_{j+1})} + a_i b_{j-1} \cdot \frac{-1}{a_j b_{j-1} - a_{j-1} b_j} + a_i b_{j+1} \cdot \frac{-1}{a_{j+1} b_j - a_j b_{j+1}}$$
$$= a_i \frac{a_{j+1} b_j b_{j-1} - a_{j-1} b_j b_{j+1} - a_{j+1} b_j b_{j-1} + a_j b_{j-1} b_{j+1} - a_j b_{j-1} b_{j+1} + a_{j-1} b_j b_{j+1}}{(a_j b_{j-1} - a_{j-1} b_j)(a_{j+1} b_j - a_j b_{j+1})} = 0$$

Case 5: Row i $\times$ Col k, $1 \leq i < k$

$$a_i b_k \cdot \frac{b_{k-1}}{b_k(a_k b_{k-1} - a_{k-1} b_k)} + a_i b_{k-1} \cdot \frac{-1}{a_k b_{k-1} - a_{k-1} b_k} = \frac{a_i b_{k-1} - a_i b_{k-1}}{a_k b_{k-1} - a_{k-1} b_k} = 0$$

Cases 1 to 3 imply the diagonal elements of $\mathbf{C}\mathbf{C}^{-1}$ are 1 and cases 4 and 5 imply the off diagonal elements are 0. Thus, $\mathbf{C}\mathbf{C}^{-1} = \mathbf{I}$ showing that $\mathbf{C}^{-1}$ is given by equation (1).

Question 2:

Let the location family have density function $f_X(x;\theta) = f(x-\theta)$, for $x, \theta \in \mathbb{R}$. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics obtained from a random sample of size n from this location family of distributions.

a) Let $\hat{\theta} = \vec{a}'\vec{X}$ for coefficient vector $\vec{a}$ and order statistics $\vec{X}$. Let the order statistics from the standard distribution be $\vec{Z} = \vec{X} - \theta\vec{j}$ with mean, variance, and vector of ones given by $\vec{\mu}$, $\mathbf{\Sigma}$, $\vec{j}$. Then $\mathbb{E}[\hat{\theta}] = \vec{a}'\mathbb{E}[\vec{X}] = \vec{a}'(\vec{\mu} + \theta\vec{j})$. Unbiasedness imposes the constraints $\vec{a}'\vec{j} = 1$, $\vec{a}'\vec{\mu} = 0$ with the objective function below.

$$Q(\vec{a}) = \vec{a}'\mathbf{\Sigma}\vec{a} - 2\lambda_1(\vec{a}'\vec{j} - 1) - 2\lambda_2(\vec{a}'\vec{\mu})$$

$$\implies \tfrac{\partial Q}{\partial \vec{a}} = 2\mathbf{\Sigma}\vec{a} - 2\lambda_1\vec{j} - 2\lambda_2\vec{\mu} = 0$$

$$\implies \vec{a} = \lambda_1\mathbf{\Sigma}^{-1}\vec{j} + \lambda_2\mathbf{\Sigma}^{-1}\vec{\mu}$$

$$\implies \begin{cases} \vec{a}'\vec{j} = 1 = \lambda_1\vec{j}'\mathbf{\Sigma}^{-1}\vec{j} + \lambda_2\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{j} \implies \lambda_1 V_1 + \lambda_2 V_3 = 1 \\ \vec{a}'\vec{\mu} = 0 = \lambda_1\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{j} + \lambda_2\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu} \implies \lambda_1 V_3 + \lambda_2 V_2 = 0 \end{cases}$$

$$\implies \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \tfrac{1}{V_1V_2 - V_3^2}\begin{pmatrix} V_2 & -V_3 \\ -V_3 & V_1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \tfrac{1}{V_1V_2 - V_3^2}\begin{pmatrix} V_2 \\ -V_3 \end{pmatrix}$$

$$\implies \hat{\theta} = \vec{a}'\vec{X} = \tfrac{1}{V_1V_2 - V_3^2}\left(V_2(\vec{j}'\mathbf{\Sigma}^{-1}) - V_3(\vec{\mu}'\mathbf{\Sigma}^{-1})\right)\vec{X}$$

b) The variance of $\hat{\theta}$ is given by the following:

$$\text{Var}[\hat{\theta}] = \text{Var}\left[\tfrac{1}{V_1V_2 - V_3^2}\left(V_2(\vec{j}'\mathbf{\Sigma}^{-1}) - V_3(\vec{\mu}'\mathbf{\Sigma}^{-1})\right)\vec{X}\right]$$

$$= \tfrac{1}{(V_1V_2 - V_3^2)^2}\left(V_2(\vec{j}'\mathbf{\Sigma}^{-1}) - V_3(\vec{\mu}'\mathbf{\Sigma}^{-1})\right)\mathbf{\Sigma}\left(V_2(\vec{j}'\mathbf{\Sigma}^{-1}) - V_3(\vec{\mu}'\mathbf{\Sigma}^{-1})\right)'$$

$$= \tfrac{1}{(V_1V_2 - V_3^2)^2}\left(V_2\vec{j}' - V_3\vec{\mu}'\right)\left(V_2(\mathbf{\Sigma}^{-1}\vec{j}) - V_3(\mathbf{\Sigma}^{-1}\vec{\mu})\right)$$

$$= \tfrac{1}{(V_1V_2 - V_3^2)^2}\left(V_1V_2^2 - 2V_2V_3^2 + V_2V_3^2\right)$$

$$= \tfrac{1}{(V_1V_2 - V_3^2)^2}\left(V_1V_2^2 - V_2V_3^2\right)$$

$$= \tfrac{V_2}{(V_1V_2 - V_3^2)}$$

Question 3:

Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics obtained from a random sample of size n from Uniform$(\theta, \theta+1)$.

a) We know that the order statistics from standard uniform are Beta distributed giving us the mean and covariance below.

$\mu_i = \frac{i}{n+1}$

$\sigma_{ij} = \frac{i(n-j+1)}{(n+1)^2(n+2)}, i \leq j$

We calculate $\mathbf{\Sigma}^{-1}, i \leq j$ using Question 1.

$$\frac{\sigma^{ij}}{(n+1)^2(n+2)} = \begin{cases} -\frac{1}{n+1} & \text{for } j = i+1, 1 \leq i \leq n-1 \\ \frac{2}{n+1} & \text{for } 1 \leq i = j \leq n \\ 0 & \text{for } j > i+1 \end{cases}$$

Using the above we calculate $\vec{j}'\mathbf{\Sigma}^{-1}$ by direct multiplication. All middle elements have 3 non-zero terms that perfectly cancel, leaving only the first and final terms to survive as follows $\vec{j}'\mathbf{\Sigma}^{-1} = \begin{pmatrix}\frac{1}{n+1} & 0 & \cdots & 0 & \frac{1}{n+1}\end{pmatrix}$. Similarly we have $\vec{\mu}'\mathbf{\Sigma}^{-1} = \begin{pmatrix}0 & 0 & \cdots & 0 & \frac{1}{n+1}\end{pmatrix}$. This allows us to calculate the BLUE from Question 2.

$\vec{j}'\mathbf{\Sigma}^{-1}\vec{X} = (X_{1:n} + X_{n:n})(n+1)(n+2)$

$\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{X} = X_{n:n}(n+1)(n+2)$

$V_1 = \vec{j}'\mathbf{\Sigma}^{-1}\vec{j} = 2(n+1)(n+2)$

$V_2 = \vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu} = n(n+2)$

$V_3 = \vec{j}'\mathbf{\Sigma}^{-1}\vec{\mu} = (n+1)(n+2)$

$\implies \hat{\theta} = \frac{1}{V_1V_2 - V_3^2}\left(V_2(\vec{j}'\mathbf{\Sigma}^{-1}) - V_3(\vec{\mu}'\mathbf{\Sigma}^{-1})\right)\vec{X} = \frac{nX_{1:n} - X_{n:n}}{n-1}$

The variance is $\text{Var}[\hat{\theta}] = \frac{V_2}{(V_1V_2 - V_3^2)} = \frac{n}{(n-1)(n+1)(n+2)}$

b) The likelihood function is given by $\mathcal{L}(\theta; x_i) = 1, \theta \leq x_i \leq \theta + 1 \implies x_{n:n} - 1 \leq \theta \leq x_{1:n} \implies$ the MLE is given by $\alpha(X_{n:n} - 1) + (1-\alpha)X_{1:n}, \alpha \in (0,1)$ so choose $\frac{X_{n:n} - 1 + X_{1:n}}{2}$ for unbiasedness.

c) The distribution of the max and min is given by $n(n-1)[F(x_n) - F(x_1)]^{n-2} f(x_1) f(x_n)$.

$f_{X_{1:n}, X_{n:n}}(x_1, x_n) = n(n-1)(x_n - x_1)^{n-2}, \quad \theta \leq x_1 \leq x_n \leq \theta + 1$

$\implies \mathbb{E}[\frac{X_{n:n}-1+X_{1:n}}{2}] = \frac{n(n-1)}{2}\int_\theta^{\theta+1}\int_\theta^{x_n}(x_n+x_1)(x_n-x_1)^{n-2}\,dx_1 dx_n - \frac{1}{2}$

$= \frac{n}{2}\int_\theta^{\theta+1} x_n(x_n-\theta)^{n-1} + \theta(x_n-\theta)^{n-1} + \frac{(x_n-\theta)^n}{n}\,dx_n - \frac{1}{2}$

$= \frac{1}{2}((\theta+1) - \frac{1}{n-1} + \theta + \frac{1}{n+1}) - \frac{1}{2}$

$= \theta$

Since the variance is unchanged by location shift (from standard distribution), we have the following:

$\text{Var}[\frac{X_{n:n}-1+X_{1:n}}{2}] = \frac{1}{4}\text{Var}[X_{n:n}+X_{1:n}]$

$= \frac{1}{4}(\text{Var}[X_{1:n}] + \text{Var}[X_{n:n}] + 2\text{Cov}[X_{1:n}, X_{n:n}])$

$= \frac{1}{4}(\frac{n}{(n+1)^2(n+2)} + \frac{n}{(n+1)^2(n+2)} + \frac{2}{(n+1)^2(n+2)})$

$= \frac{1}{2(n+1)(n+2)}$

$\implies \text{MSE}[\frac{X_{n:n}-1+X_{1:n}}{2}] = \frac{1}{2(n+1)(n+2)} + 0 = \frac{1}{2(n+1)(n+2)}$

d) Both estimates are unbiased and as $n \to \infty$ the MSE and variance both go to $\frac{1}{n^2} \to 0$.

Question 4:

Let $X_1, \cdots, X_n$ be a random sample of size n from a population with density function

$$f(x;\mu) = \frac{e^{-(x-\mu)}}{(1+e^{-(x-\mu)})^2}, x, \mu \in \mathbb{R}$$

and order statistics $X_{1:n} < \cdots < X_{n:n}$.

a) Consider the ratio of the likelihoods of samples of order statistics $\vec{X}$ and $\vec{Y}$:

$$\frac{n! \prod_{i=1}^n \frac{e^{-(x_i-\mu)}}{(1+e^{-(x_i-\mu)})^2}}{n! \prod_{i=1}^n \frac{e^{-(x_i-\mu)}}{(1+e^{-(x_i-\mu)})^2}} = \frac{e^{-\sum_{i=1}^n x_i}}{e^{-\sum_{i=1}^n y_i}} \cdot \left(\prod_{i=1}^n \frac{e^{-\mu} + e^{-y_i}}{e^{-\mu} + e^{-x_i}} \right)^2$$

This is free of $\mu \iff$ the vectors of order statistics agree $\vec{X} = \vec{Y}$ therefore $T(\vec{X}) = (X_{1:n}, \ldots, X_{n:n})'$ is a (minimal) sufficient statistic of μ.

b) The log-likelihood is given by $\ell(\mu; x) = n\mu - n\bar{x} - 2\sum_{i=1}^n \log(1 + e^{-(x_i-\mu)})$

$\implies \frac{\partial \ell}{\partial \mu} = n - 2\sum_{i=1}^n \frac{e^{-(x_i-\mu)}}{1+e^{-(x_i-\mu)}}$

I set up a gradient based optimization in R to maximize the likelihood above. Simulating 1000 iterations results in the following:

$\text{Bias}(\hat{\mu}) = 0.01263244, \quad \text{MSE}(\hat{\mu}) = 0.3036554.$

c) Using tables of the means and covariances I calculated the vector $\vec{a}$ for $\vec{a}'\vec{X}$ as:

$$\vec{a} = \begin{pmatrix} -0.06960859 \\ -0.009760527 \\ 0.05569379 \\ 0.1154093 \\ 0.408266 \\ 0.408266 \\ 0.1154093 \\ 0.05569379 \\ -0.009760527 \\ -0.06960859 \end{pmatrix}$$

Applying this to our simulation we get $\text{Var}[\mu_{\text{blue}}] = 0.3792105$.

d) The MSE of the MLE is slightly smaller than the variance of the BLUE.

e) The fisher information is given by:

$$\mathcal{I}_X(\mu) = -\mathbb{E}\left[\frac{\partial^2}{\partial \mu^2} \log f(X|\mu)\right] = -\mathbb{E}\left[-2\sum_{i=1}^n \frac{e^{-(x_i-\mu)}}{(1+e^{-(x_i-\mu)})^2}\right] = 2\int_{-\infty}^{\infty} \left(\frac{e^{-(x_i-\mu)}}{(1+e^{-(x_i-\mu)})^2}\right)^2 dx = \frac{1}{3}$$

$\implies \mathcal{I}_{X_1,\ldots,X_{10}}(\mu) = \frac{10}{3} = 3.33$

The MSE in part b) is about 0.3 i.e., $\text{MSE}(\hat{\mu}) \approx \frac{1}{\mathcal{I}_{X_1,\ldots,X_{10}}(\mu)} = \frac{3}{10} = 0.3$.

f) Suppose we wanted to test $H_0 : \mu = a$ vs. $H_1 : \mu \neq a$, we could use the MLE and perform the following test:

First, use the gradient based optimization and likelihood from part b) to calculate the estimated MLE $\hat{\mu}$. Then calculate the fisher information $\mathcal{I}(\hat{\mu})$ using our work from part e). Next calculate the test statistic $T = \mathcal{I}(\hat{\mu})(\hat{\mu} - a)^2$ which is asymptotically $\chi^2(1)$. Finally we would reject or accept the null-hypothesis using the critical values from the aforementioned distribution against the predetermined significance level.

Question 5:

Let the scale family have density function $f_X(x;\sigma) = \frac{1}{\theta}f(\frac{x}{\theta})$, for $x \in \mathbb{R}, \theta \in \mathbb{R}^+$. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics obtained from a random sample of size n from this scale family of distributions.

a) Let $\hat{\theta} = \vec{a}'\vec{X}$ for coefficient vector $\vec{a}$ and order statistics $\vec{X}$. Let the order statistics from the standard distribution be $\vec{Z} = \frac{\vec{X}}{\theta}$ with mean and variance $\vec{\mu}$ and $\mathbf{\Sigma}$. Then $\mathbb{E}[\hat{\theta}] = \vec{a}'\mathbb{E}[\vec{X}] = \vec{a}'\vec{\mu}\theta$. This imposes constraint $\vec{a}'\vec{\mu} = 1$ with objective function below.

$\mathcal{L}(a) = \vec{a}'\mathbf{\Sigma}\vec{a} - 2\lambda(\vec{a}'\vec{\mu} - 1)$

$\implies \frac{\partial \mathcal{L}}{\partial \vec{a}} = 0 = 2\mathbf{\Sigma}\vec{a} - 2\lambda\mu$

$\implies \vec{a} = \lambda\mathbf{\Sigma}^{-1}\vec{\mu}$

$\implies \vec{a}'\vec{\mu} = 1 = \lambda\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}$

$\implies \lambda = \frac{1}{\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}}$

Therefore, the BLUE of θ is given by $\hat{\theta} = \vec{a}'\vec{X} = \lambda\mathbf{\Sigma}^{-1}\vec{\mu}\vec{X} = \frac{\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{X}}{\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}}$

b) The variance of $\hat{\theta}$ is given by the following:

$\mathrm{Var}[\hat{\theta}] = \mathrm{Var}\left[\frac{\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{X}}{\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}}\right]$

$= \frac{\vec{\mu}'\mathbf{\Sigma}^{-1}\mathrm{Var}[\vec{X}]\mathbf{\Sigma}^{-1}\vec{\mu}}{(\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu})^2}$

$= \frac{\vec{\mu}'\mathbf{\Sigma}^{-1}\theta^2\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\vec{\mu}}{(\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu})^2}$

$= \frac{\theta^2\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}}{(\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu})^2}$

$= \frac{\theta^2}{\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}}$

Question 6:

Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics obtained from a random sample of size n from Uniform$(0, \theta)$.

a) We know that the order statistics from standard uniform are Beta distributed giving us the mean and covariance below.

$\mu_i = \frac{i}{n+1}$

$\sigma_{ij} = \frac{i(n-j+1)}{(n+1)^2(n+2)}, i \leq j$

Then we calculate $\mathbf{\Sigma}^{-1}, i \leq j$ using the relation in Question 1.

$$\frac{\sigma^{ij}}{(n+1)^2(n+2)} = \begin{cases} -\frac{1}{n+1} & \text{for } j = i+1, 1 \leq i \leq n-1 \\ \frac{2}{n+1} & \text{for } 1 \leq i = j \leq n \\ 0 & \text{for } j > i+1 \end{cases}$$

Now we calculate $\vec{\mu}'\mathbf{\Sigma}^{-1}$ by direct multiplication. The first element goes to zero, the middle elements have 3 non-zero terms that perfectly cancel, and the final term survives as follows $\vec{\mu}'\mathbf{\Sigma}^{-1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & \frac{1}{n+1} \end{pmatrix}$. This allows us to calculate the BLUE from Question 5.

$$\implies \vec{\mu}'\mathbf{\Sigma}^{-1}\vec{X} = X_{n:n}(n+1)(n+2)$$

$$\implies \vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu} = n(n+2)$$

$$\implies \hat{\theta} = \frac{\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{X}}{\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}} = \frac{n+1}{n}X_{n:n}$$

$$\text{Var}[\hat{\theta}] = \frac{\theta^2}{\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}} = \theta^2\frac{1}{n(n+2)}$$

b) In Assignment 1 and in the lectures we discovered $X_{n:n}$ is a complete and sufficient statistic for θ. Since $\hat{\theta} = \frac{n+1}{n}X_{n:n}$ is unbiased estimator of θ and a function of a complete and sufficient statistic, by the Lehmann-Scheffe theorem we know it is the uniformly minimum variance unbiased estimator of θ. Thus, $\hat{\theta}$ performs well in an even larger class of estimators.

Question 7:

Suppose we have a random sample of size n from bivariate normal:

$$\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2\left(\vec{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}\right)$$

a) Let $\vec{\theta} = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)'$ then the density of is given by

$$f_{X,Y}(x, y; \vec{\theta}) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}$$

$$= \tfrac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)} \exp\left\{-\tfrac{1}{2}(\vec{X}-\vec{\mu})'\mathbf{\Sigma}^{-1}(\vec{X}-\vec{\mu})\right\}$$

$$= \tfrac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)} \exp\left\{-\tfrac{1}{2}(\vec{X}'\mathbf{\Sigma}^{-1}\vec{X} - 2\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{X} + \vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu})\right\}$$

$$= \tfrac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)} \exp\left\{-\tfrac{1}{2}\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}\right\} \exp\left\{-\tfrac{1}{2}\mathrm{tr}(\mathbf{\Sigma}^{-1}\vec{X}\vec{X}') + \vec{\mu}'\mathbf{\Sigma}^{-1}\vec{X}\right\}$$

$$= \tfrac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)} \exp\left\{-\tfrac{1}{2}\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}\right\} \exp\left\{-\tfrac{1}{2}\mathrm{vec}(\mathbf{\Sigma}^{-1})\mathrm{vec}(\vec{X}\vec{X}') + \vec{\mu}'\mathbf{\Sigma}^{-1}\vec{X}\right\}$$

Note: $\mathbf{\Sigma}^{-1}\vec{\mu} = \begin{pmatrix} \frac{\mu_x}{\sigma_x^2(1-\rho^2)} - \frac{\mu_y\rho}{\sigma_x\sigma_y(1-\rho^2)} \\ \frac{\mu_y}{\sigma_y^2(1-\rho^2)} - \frac{\mu_x\rho}{\sigma_x\sigma_y(1-\rho^2)} \end{pmatrix}, \quad -\frac{1}{2}\mathrm{vec}(\mathbf{\Sigma}^{-1}) = \begin{pmatrix} -\frac{1}{2\sigma_x^2(1-\rho^2)} \\ \frac{\rho}{2\sigma_x\sigma_y(1-\rho^2)} \\ \frac{\rho}{2\sigma_x\sigma_y(1-\rho^2)} \\ -\frac{1}{2\sigma_y^2(1-\rho^2)} \end{pmatrix}, \quad \mathrm{vec}(\vec{X}\vec{X}') = \begin{pmatrix} x^2 \\ xy \\ xy \\ y^2 \end{pmatrix}$

Define: $h(\vec{X}) = \frac{1}{2\pi}, \quad c(\vec{\theta}) = \frac{1}{\sigma_x\sigma_y(1-\rho^2)} \exp\left\{-\frac{1}{2}\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}\right\}$

$$\implies w(\vec{\theta}) = \begin{pmatrix} w_1(\vec{\theta}) \\ w_2(\vec{\theta}) \\ w_3(\vec{\theta}) \\ w_4(\vec{\theta}) \\ w_5(\vec{\theta}) \end{pmatrix} = \begin{pmatrix} \frac{\mu_x}{\sigma_x^2(1-\rho^2)} - \frac{\mu_y\rho}{\sigma_x\sigma_y(1-\rho^2)} \\ \frac{\mu_y}{\sigma_y^2(1-\rho^2)} - \frac{\mu_x\rho}{\sigma_x\sigma_y(1-\rho^2)} \\ -\frac{1}{2\sigma_x^2(1-\rho^2)} \\ \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)} \\ -\frac{1}{2\sigma_y^2(1-\rho^2)} \end{pmatrix}, \quad t(\vec{X}) = \begin{pmatrix} t_1(\vec{X}) \\ t_2(\vec{X}) \\ t_3(\vec{X}) \\ t_4(\vec{X}) \\ t_5(\vec{X}) \end{pmatrix} = \begin{pmatrix} x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix}$$

Therefore, $f_{X,Y}(x, y; \vec{\theta}) = h(\vec{X})c(\vec{\theta})\exp\left\{\sum_{i=1}^{5} w_i(\vec{\theta})t_i(\vec{X})\right\}$ showing that the bivariate normal distribution belongs to the exponential family. Moreover, the exponential family structure is preserved for the iid sample.

b) Because we have the distribution in exponential family form we can easily extract the natural sufficient statistic of $\vec{\theta}$ as $T(\vec{X}) = \begin{pmatrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i^2 \\ \sum_{i=1}^n X_i Y_i \\ \sum_{i=1}^n Y_i^2 \end{pmatrix}$. For $T(\vec{X})$ to be a complete statistic we need $\mathcal{S} = \left\{ w(\vec{\theta}\,) : \vec{\theta} \in \Theta = \mathbb{R}^2 \times \mathbb{R}^{+2} \times (-1,1) \right\}$ to contain an open set in $\mathbb{R}^5$. Notice that $w(\vec{\theta}\,)$ is a continuously differentiable invertible map with Jacobian $J = \frac{1}{4\sigma_x^4\sigma_y^4(1-\rho^6)} \neq 0$. An application of the inverse function theorem ensures locally invertibility so it must contain an open set in $\mathbb{R}^5$. Therefore, $T(\vec{X})$ is a complete sufficient statistic for $\vec{\theta}$.

c) The natural parameters for the bivariate normal distribution are $w(\vec{\theta}\,) = \begin{pmatrix} \frac{\mu_x}{\sigma_x^2(1-\rho^2)} - \frac{\mu_y \rho}{\sigma_x \sigma_y (1-\rho^2)} \\ \\ \frac{\mu_y}{\sigma_y^2(1-\rho^2)} - \frac{\mu_x \rho}{\sigma_x \sigma_y (1-\rho^2)} \\ -\frac{1}{2\sigma_x^2(1-\rho^2)} \\ \\ \frac{\rho}{\sigma_x \sigma_y (1-\rho^2)} \\ \\ -\frac{1}{2\sigma_y^2(1-\rho^2)} \end{pmatrix}$.

Question 8:

Suppose we have bivariate normal $\vec{X} = \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2 \left(\vec{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$.

a) By our work in Question 8, we know that that this belongs to the exponential family with density below.

$$f_{X,Y}(x, y; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[x^2 - 2\rho xy + y^2 \right] \right\}$$

By substituting the knowns into our results from Question 7 (or by inspection) we get

$$w(\vec{\theta}) = \begin{pmatrix} w_1(\vec{\theta}) \\ w_2(\vec{\theta}) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2(1-\rho^2)} \\ \\ \frac{\rho}{(1-\rho^2)} \end{pmatrix}, \quad t(\vec{X}) = \begin{pmatrix} t_1(\vec{X}) \\ t_2(\vec{X}) \end{pmatrix} = \begin{pmatrix} x^2 + y^2 \\ xy \end{pmatrix}.$$

b) From a) for iid random sample we have the natural sufficient statistic of ρ as $T(\vec{X}) = \begin{pmatrix} \sum_{i=1}^n X_i^2 + Y_i^2 \\ \sum_{i=1}^n X_i Y_i \end{pmatrix}$.

Now consider below:

$$\frac{\left(\frac{1}{2\pi\sqrt{1-\rho^2}} \right)^n \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[x_1^2 - 2\rho x_1 y_1 + y_1^2 \right] \right\}}{\left(\frac{1}{2\pi\sqrt{1-\rho^2}} \right)^n \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[x_2^2 - 2\rho x_2 y_2 + y_2^2 \right] \right\}}$$

This is clearly free of $\vec{\theta} \iff T(\vec{X}_1) = T(\vec{X}_2)$ so $T(\vec{X})$ must be the minimal sufficient statistic for $\vec{\theta}$.

c) We already know that this model belongs to the exponential family from Question 7. From part a) we know that the natural parameters and sufficient statistic are of dimension $2 \subset \mathbb{R}^2$ and the parameter space for ρ is $(-1, 1) \subset \mathbb{R}$. That is, $\ell = 1 < 2 = k$. Lastly we check the covariance:

$$\text{Cov}[T(\vec{X})] = \begin{pmatrix} \text{Var}[X^2 + Y^2] & \text{Cov}[X^2 + Y^2, XY] \\ \text{Cov}[XY, X^2 + Y^2] & \text{Var}[XY] \end{pmatrix} = \begin{pmatrix} 4 & -2\rho \\ -2\rho & 1 + \rho^2 \end{pmatrix}$$

The determinant is 4 i.e. positive definite, so the model must belong to the curved 2-parameter exponential family.

d) The natural parameters are given by $\begin{pmatrix} w_1(\vec{\theta}) \\ w_2(\vec{\theta}) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2(1-\rho^2)} \\ \\ \frac{\rho}{(1-\rho^2)} \end{pmatrix}$.

Question 9:

Suppose we have $\vec{X} = (X, Y)'$, where $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p^2)$ are independent with parameter $\theta = p$ and $m, n \in \mathbb{N}, p \in (0, 1)$.

We have joint density given below.

$f_{X,Y}(x, y; \theta) = f_X(x; \theta) f_Y(y; \theta) = \binom{n}{x} p^x (1-p)^{n-x} \cdot \binom{m}{y} p^{2y} (1-p^2)^{m-y}$

$= \binom{n}{x}\binom{m}{y}(1-p)^n(1-p^2)^m \exp\left\{x \log(\frac{p}{1-p}) + y \log(\frac{p^2}{1-p^2})\right\}$

Define: $h(\vec{X}) = \binom{n}{x}\binom{m}{y}, \quad c(\theta) = (1-p)^n(1-p^2)^m$

$$w(\vec{\theta}) = \begin{pmatrix} w_1(\theta) \\ w_2(\theta) \end{pmatrix} = \begin{pmatrix} \log(\frac{p}{1-p}) \\ \\ \log(\frac{p^2}{1-p^2}) \end{pmatrix}, \quad T(\vec{X}) = \begin{pmatrix} t_1(\vec{X}\,) \\ t_2(\vec{X}\,) \end{pmatrix} = \begin{pmatrix} X \\ Y \end{pmatrix}$$

Then $f_{X,Y}(x, y; \theta) = h(\vec{X}) c(\theta) \exp\left\{\sum_{i=1}^{2} w_i(\theta) t_i(\vec{X})\right\}$. It is clear we have $\ell = 1 < 2 = k$. Now we check the covariance below.

$$\text{Cov}[T(\vec{X})] = \begin{pmatrix} \text{Var}[X] & \text{Cov}[X, Y] \\ \text{Cov}[Y, X] & \text{Var}[Y] \end{pmatrix} = \begin{pmatrix} np(1-p) & 0 \\ 0 & mp^2(1-p^2) \end{pmatrix}$$

The determinant is $\Delta = nmp^3(1-p)(1-p^2)$ which is positive for say $p = 0.5, (0.5) \approx (nm)0.047 > 0$, therefore this model belongs to the curved 2-parameter exponential family.

Question 10:

Let $X_1, \ldots, X_n$ be a random sample from the inverse Gaussian distribution with density:

$$f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \ x, \mu, \lambda \in \mathbb{R}^+$$

a) Define $h(x) = \sqrt{\frac{1}{2\pi x^3}}, \ c(\vec{\theta}) = \sqrt{\lambda} e^{\lambda/\mu}$

$w(\vec{\theta}) = \begin{pmatrix} w_1(\vec{\theta}) \\ w_2(\vec{\theta}) \end{pmatrix} = \begin{pmatrix} -\frac{\lambda}{2\mu^2} \\ \\ -\frac{\lambda}{2} \end{pmatrix}, \quad T(\vec{X}) = \begin{pmatrix} t_1(\vec{X}) \\ t_2(\vec{X}) \end{pmatrix} = \begin{pmatrix} X \\ \frac{1}{X} \end{pmatrix}$

Then $f_X(x; \vec{\theta}) = h(\vec{X})c(\theta) \exp\left\{\sum_{i=1}^2 w_i(\theta) t_i(\vec{X})\right\}$. The nautral parameters are continuously differentiable with open domain $\mathbb{R}$ so $\left\{w(\vec{\theta}) : \vec{\theta} \in \mathbb{R}^2\right\}$ must contain an open set in $\mathbb{R}^2$. That means for an iid random sample, $\sum_{i=1}^n X_i$ and $\frac{1}{\sum i=1^n X_i}$ are complete sufficient statistics. Since complete sufficient statistics are closed under bijections both $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $T = \frac{n}{\sum_{i=1}^n \frac{1}{X_i} - \frac{1}{\bar{X}}}$ are complete and sufficient statistics.

b) For this question I followed Carl J. Schwarz and M. Samanta (An Inductive Proof of the Sampling Distributions for the MLE's of the Parameters in an Inverse Gaussian Distribution 1991).

Let $n = 2$ with joint density

$$f(x_1, x_2; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x_1^3 x_2^3}} \exp\left\{-\sum_{i=1}^2 \frac{\lambda(x_i - \mu)^2}{2\mu^2 x_i}\right\}, x_1, x_2 > 0$$

and define $\bar{X} = \frac{X_1+X_2}{2}$ and $S = \sum_{i=1}^2 (\frac{1}{X_i} - \frac{1}{\bar{X}})$. Solving the inverse trnasformation leads to a quadratic term with $X_2 = 2\bar{X}(1 \pm \sqrt{1 + \frac{2}{\bar{X}S+2}})$. So suppose $X_1 < X_2$ then the transformation has Jacobian $J = \frac{2\bar{X}^{3/2}}{\sqrt{S}(\bar{X}S+2)^{3/2}}$. Along with the case $X_1 > X_2$ we get the joint pdf of $\bar{X}$ and S as:

$$f(\bar{x}, s) = \sqrt{\frac{2\lambda}{2\pi\bar{x}^3}} \exp\left\{-\frac{2\lambda(\bar{x}-\mu)^2}{2\mu^2\bar{x}}\right\} \cdot \frac{\sqrt{\lambda/2s}}{\Gamma(1/2)} \exp\left\{\frac{-\lambda s}{2}\right\}$$

This shows that $\bar{X} \sim \text{IG}(\mu, 2\lambda)$ and $\frac{n\lambda}{T} \sim \chi^2(1)$ and they are independent. They use this as the base case of their induction and go on to show the general property that $\bar{X} \sim \text{IG}(\mu, n\lambda)$ and $\frac{n\lambda}{T} \sim \chi^2(n-1)$ with independence.

Question 11:

Let $X_1, \cdots, X_n$ be a random sample of size n from Normal(μ, σ^2) with corresponding order statistics $X_{1:n} < \cdots < X_{n:}$. Let $T = \frac{X_{n:n} - X_{n-1:n}}{S}$ for sample standard deviation S.

a) Note that T is location-scale invariant since an affine transformation $x_i \to a + bx_i$ results in$\frac{a+bX_{n:n}-a-bX_{1:n}}{bS} = T$. Therefore, we have a location-scale invariant statistic for the normal distribution so it is an ancillary statistic for the parameters.

b) Since $\bar{X}$ and S are complete and sufficient, by Basu's theorem they are independent with the ancillary statistic T.

$\implies \mathbb{E}[T] = \frac{\mathbb{E}[X_{n:n}-X_{1:n}]}{\sigma\sqrt{2}\Gamma(\frac{n}{2})}\sqrt{n-1}\Gamma(\frac{n-1}{2})$

$= \frac{\mu-\mu+\sigma(\alpha_{n:n}-\alpha_{1:n})}{\sigma\sqrt{2}\Gamma(\frac{n}{2})}\sqrt{n-1}\Gamma(\frac{n-1}{2})$

$= \frac{\alpha_{n:n}-\alpha_{1:n}}{\sqrt{2}\Gamma(\frac{n}{2})}\sqrt{n-1}\Gamma(\frac{n-1}{2})$

$\implies \mathbb{E}[T^2] = \frac{\mathbb{E}[X_{n:n}^2]-2\mathbb{E}[X_{1:n}X_{n:n}]+\mathbb{E}[X_{1:n}]}{\sigma^2}$
$= \frac{1}{\sigma^2}(\alpha_{n:n}^2\sigma^2 + 2\alpha_{n:n}\sigma\mu + \mu^2 + \alpha_{1:n}^2\sigma^2 + 2\alpha_{1:n}\sigma\mu + \mu^2 - 2\alpha_{1:n}\alpha_{n:n}\sigma^2 - 2\alpha_{1:n}\sigma\mu - 2\alpha_{n:n}\sigma\mu - 2\mu^2)$

$= \alpha_{n:n}^2 - 2\alpha_{1:n}\alpha_{n:n} + \alpha_{1:n}$

$\mathrm{Var}[T] = \mathbb{E}[T^2] - \mathbb{E}[T]^2$

c) Using the linked tables and the above expressions we calculate the mean and variance for $n = 10$:

$\mathbb{E}[T] = 3.164006$.

$\mathrm{Var}[T] = 10.10632 - 3.164006^2 = 0.09538603$.

d) Using 1000 simulations, we have $\mathbb{E}[T] = 3.15519$ and $\mathrm{Var}[T] = 0.09385712$. The simulations and the table calculations match quite closely.

e) The critical region will be determined as the upper $(1-\alpha)$-quantile of the simulated test statistics. If $X_{n:n}$ is larger than this value it is considered an outlier.

d) Using 1000 simulations, the upper 5% critical value for T with $n = 10$ is $t_c = 3.704414$.

g) From the data, our test statistic is $T = 3.079277 < 3.704414 \implies 4.5747$ is not large enough to be an outlier .

Question 12:

We calculate the best linear unbiased predictor using the table values as :

$$(0.0843)*(1.7918)+(0.0921)*(2.3026)+(0.0957)*(2.7726)+(0.0986)*(3.2581)+(0.1011)*(3.5264)+(0.1036)*(3.8067)+(0.1060)*(3.9703)+(0.1085)*(4.0943)+(0.2101)*(4.2905)$$

$$= 3.467115 < 4.5747$$

The result is much closer to the mean (3.4388) than the largest censored value.