Statistics 2 - Problem Sheet 1

Dom Hutchinson

January 31, 2020

Question - 1.

Question 1 a)
If $\mathbf{Y}$ and $\mathbf{X}$ are random vectors st $\mathbf{Y} = C\mathbf{X}$ where C is a matrix of fixed coefficients, show that if Σ_X and Σ_Y are the covariance matrices for $\mathbf{X}$ and $\mathbf{Y}$ respecitvely then

$$\Sigma_y = C\Sigma_X C^T$$

Answer 1 a)

$$\begin{array}{rcl}
\mu_y &=& \mathbb{E}(Y) \\
&=& \mathbb{E}(CX) \\
&=& C\mathbb{E}(X) \\
&=& C\mu_X \\
\Sigma_y &=& \mathbb{E}[(Y-\mu_Y)(Y-\mu_Y)^T] \\
&=& \mathbb{E}[(CX - C\mu_X)(CX - C\mu_X)^T] \\
&=& \mathbb{E}[C(X-\mu_X)(X-\mu_X)^T C^T] \\
&=& C\mathbb{E}[(X-\mu_X)(X-\mu_X)^T]C^T \\
&=& C\Sigma_X C^T
\end{array}$$

Question 1 b)
Consider a multivariate normal random vector $\mathbf{X} \sim \text{Normal}(\boldsymbol{\mu}_X, \Sigma_X)$ and suppose that the covariance matrix can be decomposed $\Sigma_X = CC^T$ (THis can always be done for a full rank covariance matrix using a Choleski decomposition). Show that $\Sigma_X^{-1} = C^{-T}C^{-1}$ and that $\mathbf{Y} = C^{-1}(\mathbf{X} - \boldsymbol{\mu}_X) \sim \text{Normal}(\mathbf{0}, I)$.

Answer 1 b)

$$\begin{array}{rrcl}
& \Sigma_X \Sigma_X^{-1} &=& I \\
\implies & CC^T\Sigma_X^{-1} &=& 1 \\
\implies & C^T\Sigma_X^{-1} &=& C^{-1} \\
\implies & \Sigma_X^{-1} &=& C^{-T}C^{-1}
\end{array}$$

$$\begin{array}{rrcl}
\text{Note that} & \mathbf{Y} &=& C^{-1}(\mathbf{X} - \boldsymbol{\mu}_X) \\
& &=& C^{-1}\mathbf{X} - C^{-1}\boldsymbol{\mu}_X \\
\text{We have} & \mathcal{M}_X(\mathbf{t}) &=& \exp\left\{\mathbf{t}^T\boldsymbol{\mu}_X + \frac{1}{2}\mathbf{t}^T\Sigma_X\mathbf{t}\right\} \\
& &=& \exp\left\{\mathbf{t}^T\boldsymbol{\mu}_X + \frac{1}{2}\mathbf{t}^T CC^T\mathbf{t}\right\} \\
\implies & \mathcal{M}_Y(\mathbf{t}) &=& \exp\{-\mathbf{t}^T C^{-1}\boldsymbol{\mu}_X\}\mathcal{M}_X(C^{-t}\mathbf{t}) \\
& &=& \exp\{-\mathbf{t}^T C^{-1}\boldsymbol{\mu}_X\}\exp\left\{\mathbf{t}^T C^{-1}\boldsymbol{\mu}_X + \frac{1}{2}\mathbf{t}^T C^{-1}CC^TC^{-T}\mathbf{t}\right\} \\
& &=& \exp\{\frac{1}{2}\mathbf{t}^T\mathbf{t}\} \\
& &=& \mathcal{M}_Z(\mathbf{t}) \text{ where } \mathbf{Z} \sim \text{Normal}(\mathbf{0}, I)
\end{array}$$

Since $\mathbf{Y}$ has the same moment generating function as $\mathbf{Z} \sim \text{Normal}(\mathbf{0}, I)$ they have the same distribution.

Question 1 c)
Assuming that $\mathbf{X} \sim \text{Normal}(\boldsymbol{\mu}_X, \Sigma_X)$ show that

$$(\mathbf{X} - \boldsymbol{\mu}_X)^T \Sigma_X^{-1} (\mathbf{X} - \boldsymbol{\mu}_X) = \mathbf{Y}^T \mathbf{Y} \text{ where } \mathbf{Y} \sim \text{Normal}(\mathbf{0}, I)$$

Answer 1 c)
It is reasonable to assume that Σ_X is full rank and thus $\exists\ C$ st $\Sigma_X = CC^T$.
Thus

$$\begin{array}{rcl} \mathbf{Y} & = & C^{-1}(\mathbf{X} - \boldsymbol{\mu}_X) \\ (\mathbf{X} - \boldsymbol{\mu}_X)^T \Sigma_X^{-1} (\mathbf{X} - \boldsymbol{\mu}_X) & = & (\mathbf{X} - \boldsymbol{\mu}_X)^T (CC^T)^{-1} (\mathbf{X} - \boldsymbol{\mu}_X) \\ & = & (\mathbf{X} - \boldsymbol{\mu}_X)^T C^{-T} C^{-1} (\mathbf{X} - \boldsymbol{\mu}_X) \\ & = & \mathbf{Y}^T \mathbf{Y} \text{ as required} \end{array}$$

Question 1 d)
If $Z_i \overset{\text{iid}}{\sim} \text{Normal}(0, 1)$ random variables then

$$\sum_{i=1}^n Z_i^2 \sim \chi_n^2$$

What is the distribution of

$$(\mathbf{X} - \boldsymbol{\mu}_X)^T \Sigma_X^{-1} (\mathbf{X} - \boldsymbol{\mu}_X)$$

if $\mathbf{X} \sim \text{Normal}(\boldsymbol{\mu}_X, \Sigma_X)$?

Answer 1 e) -

$$\begin{array}{rcl} (\mathbf{X} - \boldsymbol{\mu}_X)^T \Sigma_X^{-1} (\mathbf{X} - \boldsymbol{\mu}_X) & = & \mathbf{Y}^T \mathbf{Y} \\ & = & \displaystyle\sum_{i=1}^n y_i^2 \text{ where } y_i \sim \text{Norma}(0, 1) \\ & \sim & \chi_n^2 \end{array}$$

Question - 2.

Question 2 a)
Define $\mathbf{y} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$ and $B = \begin{pmatrix} -1 & 2 & -1 \\ 2 & -3 & 0 \end{pmatrix}$.

Answer 2 a)
Find $B^T \mathbf{y}$.

$$\begin{pmatrix} -1 & 2 & -1 \\ 2 & -3 & 0 \end{pmatrix}^T \begin{pmatrix} 1 \\ -3 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -3 \end{pmatrix} = \begin{pmatrix} -1 - 7 \\ 2 + 9 \\ -1 + 0 \end{pmatrix} = \begin{pmatrix} -7 \\ 11 \\ -1 \end{pmatrix}$$

Question 2 b)
Let A be a full rank 3×3 matrix and B be a full rank 5×3 matrix.
State the dimensions of the following, if they exist. For those that do not exist, state why in a single sentence.

Answer 2 b)
Let $M \in \mathbb{R}(a, b)$ & $N \in \mathbb{R}(c, d)$ with $a, b, c, d \in \mathbb{N}$.

MN is a valid matrix multiplication iff $b = c$. If $b = c$ then $(MN) \in \mathbb{R}(a, d)$.
$M + N$ is a valid matrix addition iff $a = c$ **and** $b = d$. If these critera are fulfilled then $(M + N) \in \mathbb{R}(a, b)$.

i) $A^{-1}B^T$
$A^{-1} \in \mathbb{R}(3 \times 3),\ B^T \in \mathbb{R}(3 \times 5) \implies A^{-1}B^T \in \mathbb{R}(3 \times 5)$

ii) $A^{-1}B$
$A^{-1} \in \mathbb{R}(3 \times 3),\ B \in \mathbb{R}(5 \times 3) \implies$ these matrices are incompatible for multiplication.

iii) $B^{-1}A$
$B^{-1} \in \mathbb{R}(3 \times 5),\ A \in \mathbb{R}(3 \times 3) \implies$ these matrices are incompatible for multiplication.

iv) BA
$B \in \mathbb{R}(5 \times 3),\ A \in \mathbb{R}(3 \times 3) \implies BA \in \mathbb{R}(5 \times 3)$

v) $B^{-1}A^T$
$B^{-1} \in \mathbb{R}(3 \times 5),\ A^T \in \mathbb{R}(3 \times 3) \implies$ these matrices are incompatible for multiplication.

vi) BA^{-1}
$B \in \mathbb{R}(5 \times 3),\ A^{-1} \in \mathbb{R}(3 \times 3) \implies BA^{-1} \in \mathbb{R}(5 \times 3)$.

vii) $(BA)^{-1}$
We know that $BA \in \mathbb{R}(5 \times 3) \implies (BA)^{-1} \in \mathbb{R}(3 \times 5)$.

viii) B^TA
$B^T \in \mathbb{R}(3 \times 5),\ A \in \mathbb{R}(3 \times 3) \implies$ these matrices are incompatible for multiplication.

ix) $B + A$
$B \in \mathbb{R}(3 \times 5),\ A \in \mathbb{R}(3 \times 3)$. Matrices must have the exact same dimensions in order to be added together, thus this in an illegal equation.

x) $B + A^T$
$B \in \mathbb{R}(3 \times 5),\ A \in \mathbb{R}(3 \times 3)$. Matrices must have the exact same dimensions in order to be added together, thus this in an illegal equation.

Question - 3.

The *Exponential Distribution* is often a reasonable model of the times between random events. Suppose then, that $x_1, \dots, x_n$ are observations of times between hardware faults on a computer network, and it is reasonable to treat the faults as independent. To plan for fault tolerance the network managers need a reasonable model for the fault occurence rate. The *pdf* of an *Exponential Distribution* is

$$f(x) = \mathbb{1}\{x \geq 0\}\lambda e^{-\lambda x}$$

where λ is a positive parameter.
The variance of an exponential random variable is λ^{-2}.

Question 3 a) - Let $X \sim \text{Exponential}(\lambda)$. Find $\mathbb{E}(X)$.

Answer 3 a)

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f_X(t)dt \\ &= \int_{-\infty}^{\infty} \mathbb{1}\{t \geq 0\} x\lambda e^{-\lambda t}dt \\ &= \int_{0}^{\infty} x\lambda e^{-\lambda t}dt \\ &= \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}dx \text{ integration by parts} \\ &= \left[-xe^{-\lambda x} - \tfrac{1}{\lambda}e^{-\lambda x}\right]_0^\infty \\ &= [0-0] - \left[0 - \tfrac{1}{\lambda}\right] \\ &= \frac{1}{\lambda}\end{aligned}$$

Question 3 b) - Hence, suggest an estimator, $\hat{\lambda}$, for λ.

Answer 3 b)
Since $\mathbb{E}(X) = 1/\lambda \implies \lambda = 1/\mathbb{E}(X)$ and $\bar{X} \to_{\mathbb{P}} \mathbb{E}(X)$.
Thus I suggest

$$\hat{\lambda} = \frac{1}{\bar{X}} \text{ where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Question 3 c) - What is the variance of $\hat{\lambda}^{-1}$?

Answer 3 c)

$$\begin{aligned}\text{var}\left(\hat{\lambda}^{-1}\right) &= \text{var}(\bar{X}) \\ &= \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{n}{n^2}\text{var}(X_1) \\ &= \frac{1}{n\lambda^2}\end{aligned}$$

Question 3 d)
Let $\bar{x} = \frac{1}{n}\sum x_i$.
Find a first order *Taylor Expansion* of $\hat{\lambda}$ about $\mathbb{E}(\bar{x})$, considering $\hat{\lambda}$ as a function of $\bar{x}$.

Answer 3 d)
We have

$$\hat{\lambda}(\bar{x}) = \frac{1}{\bar{x}} \implies \frac{d}{d\bar{x}}\hat{\lambda}(\bar{x}) = -\frac{1}{\bar{x}^2}$$

A *First Order Taylor Expansion* give us the following approximation

$$f(x) \approx f(a) + f'(a)(x-a) \text{ for } a \in \mathbb{R}$$

Define $f(x) = \hat{\lambda}(x)$ and $a = \mathbb{E}(\bar{x}) \in \mathbb{R}$. Then

$$\begin{aligned}\hat{\lambda}(\bar{x}) &\approx \hat{\lambda}(\mathbb{E}(\bar{x})) + [\bar{x} - \mathbb{E}(\bar{x})]\tfrac{d}{d\bar{x}}\hat{\lambda}(\bar{x}) \\ &= \frac{1}{\mathbb{E}(\bar{x})} - \frac{\bar{x} - \mathbb{E}(\bar{x})}{\mathbb{E}(\bar{x})^2} \\ &= \frac{\bar{x}}{\mathbb{E}(\bar{x})^2} \\ &= \lambda^2\bar{x}\end{aligned}$$

Question 3 e)
Hence find an approximation for the variance of $\hat{\lambda}$, in terms of n and $\bar{x}$. This use of Taylor expansions to computer approximate variances via linearisation is known as the Δ-*method* in statistics.

Answer 3 e)
By definition

$$\text{var}(\hat{\lambda}) = \mathbb{E}[(\hat{\lambda} - \mathbb{E}(\hat{\lambda}))^2] = \mathbb{E}(\hat{\lambda}^2) - \mathbb{E}(\hat{\lambda})^2$$

Note that

$$\begin{array}{rcl} \mathbb{E}(\hat{\lambda}) & \approx & \mathbb{E}\left(\lambda^2 \bar{X}\right) \\ & = & \lambda^2 \mathbb{E}(\bar{X}) \\ & = & \lambda^2 \cdot \frac{1}{\lambda} \\ & = & \lambda \\ \mathbb{E}(\hat{\lambda}^2) & \approx & \mathbb{E}\left[\left(\lambda^2 \bar{x}\right)^2\right] \\ & = & \mathbb{E}[\lambda^4 \bar{X}^2] \\ & = & \lambda^4 \mathbb{E}[\bar{X}^2] \\ & = & \lambda^4[\text{var}(\bar{X}) + \mathbb{E}(\bar{X})^2] \\ & = & \lambda^4\left[\frac{1}{n}\text{var}(X_1) + \left(\frac{1}{\lambda}\right)^2\right] \\ & = & \lambda^4\left[\frac{1}{n\lambda^2} + \frac{1}{\lambda^2}\right] \\ & = & \lambda^2\left[\frac{1}{n} + 1\right] \\ \implies \text{var}(\hat{\lambda}) & = & \lambda^2\left[\frac{1}{n} + 1\right] - \lambda^2 \\ & = & \dfrac{\lambda^2}{n} \end{array}$$

Question - 4.

Consider again the setup from the previous question, but now taking a Bayesian approach. This means taht we need to augment our model with a prior distribution for the parameter, $\lambda \sim \Gamma(\alpha, \theta)$. So the prior *pdf* of λ is

$$f(\lambda) = \frac{\lambda^{\alpha-1} e^{-\lambda/\theta}}{\theta^\alpha \Gamma(\alpha)}$$

with $\mathbb{E}(\lambda) = \alpha\theta$ and $\text{var}(\lambda) = \alpha\theta^2$.

Question 4 a) - Write down the *pdf* for the joint distribution of the data $x_1, x_2, \dots$ given λ.

Answer 4 a)

$$\begin{array}{rcl} f_n(\mathbf{x}; \lambda) & = & \displaystyle\prod_{i=1}^n f(x_i; \lambda) \\ & = & \displaystyle\prod_{i=1}^n \mathbb{1}\{x_i \geq 0\} \lambda e^{-\lambda x_i} \\ & = & \mathbb{1}\{\text{all } x \geq 0\} \lambda^n e^{-\lambda n \bar{X}} \end{array}$$

Question 4 b) - By considering the joint distribution of λ and $\mathbf{x}$, indentify the posterior distribution of λ given $\mathbf{x}$.

Answer 4 b)

$$\begin{aligned} f(\lambda;\mathbf{x}) &\propto f(\mathbf{x};\lambda)f(\lambda) \\ &= \lambda^n e^{-\lambda n\bar{x}} \cdot \frac{\lambda^{\alpha-1}e^{-\lambda/\theta}}{\theta^\alpha\Gamma(\alpha)} \\ &\propto \lambda^{n+\alpha-1}e^{-\lambda(n\bar{x}+1/\theta)} \\ &= \lambda^{n+\alpha-1}e^{-\lambda\frac{n\bar{x}\theta+1}{\theta}} \\ &\sim \Gamma\left(n+\alpha, \frac{\theta}{n\bar{x}\theta+1}\right) \end{aligned}$$

Question 4 c) - What are the posterior expection and variance of λ?

Answer 4 c)

$$\mathbb{E}[\lambda;\mathbf{x}] = \frac{\theta(n+\alpha-1)}{n\bar{x}\theta+1} \quad \text{var}(\lambda;\mathbf{x}) = \frac{\theta^2(n+\alpha-1)}{(n\bar{x}\theta+1)^2}$$

Question 4 d)
Consider the situation in which $n \to \infty$.
What happens to the Bayesian and frequentist inferences aboute λ in this case?

Answer 4 d)

$$\mathbb{E}[\lambda;\mathbf{x}] \overset{n\to\infty}{\longrightarrow} \frac{1}{\bar{x}} = \hat{\lambda}_{\text{Frequentist}}$$