Imperial College London

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May – June 2014

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science.

Probability & Statistics II

Date: Monday, 19 May 2014. Time: 2.00pm – 4.00pm. Time allowed: 2 hours.

This paper has FOUR questions.

Candidates should start their solutions to each question in a new main answer book

Supplementary books may only be used after the relevant main book(s) are full.

Statistical tables are provided on pages 4 & 5.

- DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO.
- Affix one of the labels provided to each answer book that you use, but DO NOT USE THE LABEL WITH YOUR NAME ON IT.
- Credit will be given for all questions attempted, but extra credit will be given for complete or nearly complete answers to each question as per the table below.

Raw mark	up to 12	13	14	15	16	17	18	19	20
Extra credit	0	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	$2\frac{1}{2}$	3	$3\frac{1}{2}$	4

- Each question carries equal weight.
- Calculators may not be used.

1. (a) State the Kolmogorov Axioms and the definition of a probability function.

   (b) You have three boxes, each containing a mixture of red and blue marbles. There is a total of 100 marbles in each box, but the numbers of red marbles differ. Box 1 contains 80 red marbles; Box 2 contains 50 red marbles; and Box 3 contains 10 red marbles. You randomly choose a box and then randomly draw a single marble from the selected box. What is the probability that you choose a red marble?
   Given that you select a red marble, what is the chance that the selected box was Box 1?

   (c) Scores on a certain exam follow the normal distribution with mean equal to μ and standard deviation equal to 30.
   What is the probability that a randomly selected student scores more than one standard deviation above the mean?
   Consider a random sample of nine students. Give an expression for the probability that the median score of the nine student scores is at least one standard deviation above μ. (You needn't provide a numerical value for the probability by evaluating your expression.)

   (d) Let X_i for $i = 1, \ldots, 9$ be the scores of the nine students in the random sample in Part (c). Derive a 95% confidence interval for μ based on the nine scores. Let

$$Z = \begin{cases} 1 & \text{if your confidence interval contains the true value of } \mu, \\ 0 & \text{if your confidence interval does not contain the true value of } \mu. \end{cases}$$

   What is the variance of Z?

2. Consider a random variable X with PDF

$$f_X(x) = \begin{cases} k\frac{x}{\theta} & \text{for } 0 < x < \theta \\ k\frac{1-x}{1-\theta} & \text{for } \theta < x < 1 \,, \\ 0 & \text{elsewhere} \end{cases} \quad \text{where} \quad 0 \leq \theta \leq 1. \tag{1}$$

   (a) Sketch the PDF of X; find k; and compute $\mathrm{E}(X)$.

   (b) Suppose $X_1, \ldots, X_n$ is an iid sample each with PDF $f_X(x)$, that is $X_i \overset{\text{iid}}{\sim} f_X(x)$. Derive the method of moments estimator of θ.

   (c) Now assume $\theta = 0$. Derive the CDF of X_i.
   Using a theorem, derive the CDF of $U_n = \max(X_1, \ldots, X_n)$.

   (d) Let $V_n = \sqrt{n}(1 - U_n)$. Using the fact that $\lim_{n\to\infty} \left(1 + \frac{y}{n}\right)^n = e^y$, show $V_n \xrightarrow{D} V$ as $n \to \infty$ and derive the CDF of V.

3. (a) Let T be an arbitrary estimator of an unknown parameter α.
Define the bias of T as an estimator of α. Denote this quantity by bias(T).
What does it mean for T to be an unbiased estimator of α?
The mean square error of T as an estimator of α is $\text{mse}(T) = \text{E}[(T-\alpha)^2]$. Show that

$$\text{mse}(T) = [\text{bias}(T)]^2 + \text{Var}(T).$$

(b) Suppose that the conditional distribution of the random variable Y given X is

$$Y|X \sim \text{N}(X, 1)$$

and the that the marginal distribution of X is $X \sim \text{N}(\alpha u, v^2)$, where u an v are known constants and α is an unknown parameter.
Given an expression for the joint PDF of X and Y.
Derive the marginal PDF of Y.

(c) Now consider a random sample, $Y_i \overset{\text{indep}}{\sim} \text{N}(\alpha u_i, v^2+1)$ for $i = 1, \ldots, n$, where v and $u_1, u_2, \ldots, u_n$ are known constants and α is an unknown parameter.
Derive the maximum likelihood estimator, $\widehat{A}_{\text{MLE}}$ of α.

(d) Derive the bias and mean square error of $\widehat{A}_{\text{MLE}}$ as an estimator of α.

4. Suppose that $\boldsymbol{X}$ is an $K \times 1$ random vector that follows the multivariate normal distribution, $\boldsymbol{X} \sim \text{N}_K(\mathbf{0}, \boldsymbol{I})$, where $\mathbf{0}$ is the $K \times 1$ (mean) vector of zeros and $\boldsymbol{I}$ is the $K \times K$ identity (variance-covariance) matrix. Let $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{M}\boldsymbol{X}$, where $\boldsymbol{a}$ is a $K \times 1$ vector and $\boldsymbol{M}$ is a $K \times K$ invertible matrix. (Both $\boldsymbol{a}$ and $\boldsymbol{M}$ are known non-random constants.)

(a) Using matrix notation, give an explicit expression for the joint PDF of $\boldsymbol{X}$.
Show that $\boldsymbol{Y} \sim \text{N}_K(\boldsymbol{a}, \boldsymbol{M}\boldsymbol{M}^\top)$.

Now suppose $K = 2$, $X_i \overset{\text{iid}}{\sim} \text{N}(0, 1)$ for $i = 1, 2$, and let

$$\begin{aligned} Y_1 &= \mu_1 + \sigma_1 X_1 \\ Y_2 &= \mu_2 + \sigma_2 \rho X_1 + \sigma_2\sqrt{1-\rho^2}X_2. \end{aligned}$$

Finally, suppose that conditional distribution of U given Y_1 and Y_2 is

$$U|Y_1, Y_2 \sim \text{N}(\alpha + \beta Y_1, \gamma + \tau Y_2^2).$$

(b) Using the result in part (a) deduce the joint distribution of Y_1 and Y_2.
What named distribution is this?

(c) What is the marginal mean of U, $\text{E}(U)$?

(d) What is the marginal variance of U, $\text{Var}(U)$?

DISCRETE DISTRIBUTIONS

	range $\mathbb{X}$	*parameters*	*pmf* f_X	*cdf* F_X	$\mathsf{E}[X]$	$\mathsf{Var}[X]$	*mgf* M_X
$Bernoulli(\theta)$	$\{0,1\}$	$\theta \in (0,1)$	$\theta^x(1-\theta)^{1-x}$		θ	$\theta(1-\theta)$	$1-\theta+\theta e^t$
$Binomial(n,\theta)$	$\{0,1,\dots,n\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$		$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta e^t)^n$
$Poisson(\lambda)$	$\{0,1,2,\dots\}$	$\lambda \in \mathbb{R}^+$	$\dfrac{e^{-\lambda}\lambda^x}{x!}$		λ	λ	$\exp\{\lambda(e^t-1)\}$
$Geometric(\theta)$	$\{1,2,\dots\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\dfrac{1}{\theta}$	$\dfrac{(1-\theta)}{\theta^2}$	$\dfrac{\theta e^t}{1-e^t(1-\theta)}$
$NegBinomial(n,\theta)$	$\{n,n+1,\dots\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$		$\dfrac{n}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta e^t}{1-e^t(1-\theta)}\right)^n$
or	$\{0,1,2,\dots\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n+x-1}{x}\theta^n(1-\theta)^x$		$\dfrac{n(1-\theta)}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta}{1-e^t(1-\theta)}\right)^n$

The *gamma function* is given by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$.

The location/scale transformation $Y = \mu + \sigma X$ gives

$f_Y(y) = f_X\left(\dfrac{y-\mu}{\sigma}\right)\dfrac{1}{\sigma}$ $\quad F_Y(y) = F_X\left(\dfrac{y-\mu}{\sigma}\right)$ $\quad M_Y(t) = e^{\mu t}M_X(\sigma t)$ $\quad \mathsf{E}[Y] = \mu + \sigma\mathsf{E}[X]$ $\quad \mathsf{Var}[Y] = \sigma^2\mathsf{Var}[X]$

CONTINUOUS DISTRIBUTIONS

		parameters	*pdf*	*cdf*	$\mathsf{E}[X]$	$\mathsf{Var}[X]$	*mgf*
$Uniform(\alpha,\beta)$ (stand. model $\alpha = 0, \beta = 1$)	(α,β)	$\alpha < \beta \in \mathbb{R}$	$\dfrac{1}{\beta - \alpha}$	$\dfrac{x - \alpha}{\beta - \alpha}$	$\dfrac{(\alpha+\beta)}{2}$	$\dfrac{(\beta-\alpha)^2}{12}$	$\dfrac{e^{\beta t} - e^{\alpha t}}{t(\beta - \alpha)}$
$Exponential(\lambda)$ (stand. model $\lambda = 1$)	$\mathbb{R}^+$	$\lambda \in \mathbb{R}^+$	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\left(\dfrac{\lambda}{\lambda - t}\right)$
$Gamma(\alpha,\beta)$ (stand. model $\beta = 1$)	$\mathbb{R}^+$	$\alpha, \beta \in \mathbb{R}^+$	$\dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$		$\dfrac{\alpha}{\beta}$	$\dfrac{\alpha}{\beta^2}$	$\left(\dfrac{\beta}{\beta - t}\right)^\alpha$
$Weibull(\alpha,\beta)$ (stand. model $\beta = 1$)	$\mathbb{R}^+$	$\alpha, \beta \in \mathbb{R}^+$	$\alpha\beta x^{\alpha-1} e^{-\beta x^\alpha}$	$1 - e^{-\beta x^\alpha}$	$\dfrac{\Gamma(1 + 1/\alpha)}{\beta^{1/\alpha}}$	$\dfrac{\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2}{\beta^{2/\alpha}}$	
$Normal(\mu,\sigma^2)$ (stand. model $\mu = 0, \sigma = 1$)	$\mathbb{R}$	$\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$	$\dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$		μ	σ^2	$e^{\{\mu t + \sigma^2 t^2/2\}}$
$Student(\nu)$	$\mathbb{R}$	$\nu \in \mathbb{R}^+$	$\dfrac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left\{1 + \frac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		$0 \quad (\text{if } \nu > 1)$	$\dfrac{\nu}{\nu - 2} \quad (\text{if } \nu > 2)$	
$Pareto(\theta,\alpha)$	$\mathbb{R}^+$	$\theta, \alpha \in \mathbb{R}^+$	$\dfrac{\alpha\theta^\alpha}{(\theta + x)^{\alpha+1}}$	$1 - \left(\dfrac{\theta}{\theta + x}\right)^\alpha$	$\dfrac{\theta}{\alpha - 1}$ (if $\alpha > 1$)	$\dfrac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)}$ (if $\alpha > 2$)	
$Beta(\alpha,\beta)$	$(0,1)$	$\alpha, \beta \in \mathbb{R}^+$	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$		$\dfrac{\alpha}{\alpha + \beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	

M2S1 – Autumn, 2013 – Exam Solution

1. (a) Given a Borel Field (or sigma algebra), $\mathcal{B}$, a probability function, Pr, is a function with domain $\mathcal{B}$ that satisfies the Kolmogorov Axioms, namely

   i) $\Pr(A) \geq 0, \ \forall A \in \mathcal{B}$.

   ii) $\Pr(S) = 1$.

   iii) If $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint then $\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$.

   [5 marks: one for identifying the domain as a Borel field, one for relating the probability function to the axioms, and one for each of the three axioms.]

   (b) Let R be the event that you choose a red marble and B_i be the event that you choose Box i, for $i = 1, 2, 3$. By the law of total probability,

$$\Pr(R) = \sum_{i=1}^{3} \Pr(R|B_i)\Pr(B_i) = \frac{1}{3} \times \left(\frac{80}{100} + \frac{50}{100} + \frac{10}{100}\right) = \frac{7}{15} \approx 0.4667.$$

   By Bayes Theorem,

$$\Pr(B_1|R) = \frac{\Pr(R|B_1)\Pr(B_1)}{\Pr(R)} = \frac{\frac{80}{100} \times \frac{1}{3}}{\frac{7}{15}} = \frac{4}{7} \approx 0.5714.$$

   (In both cases, either the decimal of the ratio is acceptable.)

   [4 marks: two for each of the two calculations.]

   (c) A normal random variable takes on values within one standard deviation of its mean with probability 68%. The remaining 32% is divided evenly between the two tails, so the probability that a randomly selected students score is more than one standard deviation above μ is 16%. (Alternatively, students may report $1 - \Phi(1)$, where Φ is the standard normal CDF.)

   Let X be the number of students who score more than one standard deviation above μ. $X \sim \text{Binomial}(n = 9, p = 0.16)$ and

$$\Pr(\text{Median score is more than one std dev above } \mu) = \Pr(X \geq 5) = \sum_{k=5}^{9} \binom{9}{k} 0.16^k 0.84^{n-k}.$$

   [5 marks: two for the initial probability and three for the probability that the sample median is more then one std deviation above μ.]

   (d) Starting with the pivot, $\frac{\bar{X}-\mu}{\sigma/\sqrt{9}} \sim \mathrm{N}(0,1)$,

$$\Pr\left(-z_{0.025} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{9}} \leq z_{0.025}\right) = 95\%,$$

   where $\bar{X} = \frac{1}{9}\sum_{i=1}^{9} X_i$, $\sigma = 30$, and $\Pr(Z \leq z_{0.025}) = 0.975$ for $Z \sim \mathrm{N}(0,1)$, we have

$$\Pr\left(-\frac{\sigma\ z_{0.025}}{\sqrt{9}} - \bar{X} \leq -\mu \leq \frac{\sigma\ z_{0.025}}{\sqrt{9}} - \bar{X}\right) = 95\%,$$

   and

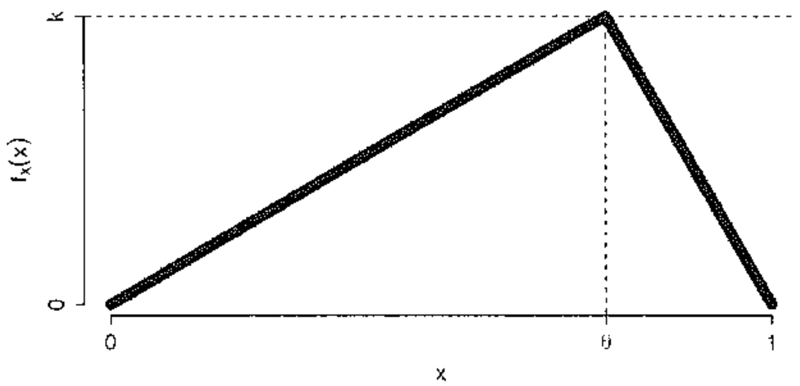
$$\Pr\left(\bar{X} - \frac{\sigma\ z_{0.025}}{\sqrt{9}} \leq \mu \leq \bar{X} + \frac{\sigma\ z_{0.025}}{\sqrt{9}}\right) = 95\%,$$

   i.e., $\bar{X} \pm \frac{\sigma}{\sqrt{9}} z_{0.025}$ is a 95% CI for μ. Substituting in $\sigma = 30$ and $z_{0.025} = 1.96$, we have $\bar{X} \pm 19.6$ is a 95% CI for μ.

   Finally, since Z is a Bernoulli random variable with probability of success equal to 0.95, $\mathrm{Var}(Z) = 0.95 \times (1 - 0.95) = \frac{19}{400} = 0.0475$. (Either the decimal of fraction form is acceptable.)

   [6 marks: two for the derivation of the CI, two for the numerical value of the CI, and two for $\mathrm{Var}(Z)$.]

2. (a) Sketch of $f_X(x)$:



The area under the density is $k/2$. Setting this area equal to one yields $k = 2$.

The expectation of X is

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^1 x f_X(x)dx = \int_0^\theta \frac{2x^2}{\theta}dx + \int_\theta^1 \frac{2(x - x^2)}{1-\theta}dx = \frac{2x^3}{3\theta}\Bigg|_0^\theta + \frac{2(\frac{x^2}{2} - \frac{x^3}{3})}{1-\theta}\Bigg|_\theta^1 \\
&= \frac{2\theta^2}{3} + \frac{1}{3(1-\theta)} - \frac{3\theta^2 - 2\theta^3}{3(1-\theta)} = \frac{1-\theta^2}{3(1-\theta)} = \frac{1+\theta}{3}.
\end{aligned}$$

[6 marks: one for sketch, two for value of k, and three for the expectation.]

(b) The methods of moments estimate is the $\hat{\Theta}_{\mathrm{MoM}} = \theta$ that solves

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} = \mathrm{E}(X) = \frac{1+\theta}{3},$$

i.e., $\hat{\Theta}_{\mathrm{MoM}} = 3\bar{X} - 1$.

[4 marks: two for setting up equation, two for estimator written in notation of a random variable ($\bar{X}$ rather than $\bar{x}$).]

(c) The CDF of X_i is

$$F_{X_i}(x) = \int_0^x f_X(t)dt = \int_0^x 2(1-t)dt = 2t - t^2\Big|_0^x = x(2-x) \quad \text{for} \quad 0 \le x \le 1.$$

The CDF of $U_n = \max(X_1, \ldots X_n)$ is

$$F_{U_n}(u) = [F_{X_i}(u)]^n = u^n(2-u)^n \quad \text{for} \quad 0 \le u \le 1.$$

[4 marks: two for each CDF.]

(d) The CDF of V_n is

$$\begin{aligned}
F_{V_n}(v) &= \Pr(V_n \le v) = \Pr\left(\sqrt{n}(1-U_n) \le v\right) = \Pr\left(U_n \ge 1 - \frac{v}{\sqrt{n}}\right) \\
&= 1 - F_{U_n}\left(1 - \frac{v}{\sqrt{n}}\right) = 1 - \left(2 - \frac{2v}{\sqrt{n}} - \left(1 - \frac{2v}{\sqrt{n}} + \frac{v^2}{n}\right)\right)^n \\
&= 1 - \left(1 - \frac{v^2}{n}\right)^n \to 1 - e^{-v^2} \quad \text{as} \quad n \to \infty, \quad \text{for all} \quad v \ge 0.
\end{aligned}$$

So $V_n \xrightarrow{D} V$ with $F_V(v) = 1 - e^{-v^2}$ for $v \ge 0$.

[6 marks: one for identifying convergence in distribution with CDFs, three for CDF of V_n, and two for limit.]

3. (a) The bias of T as an estimator of α is $\mathrm{E}(T) - \alpha$.
T is an unbiased estimator of α if $\mathrm{E}(T) = \alpha$, i.e., its bias as an estimator of α is zero.
Finally,

$$\begin{aligned}
\mathrm{mse}(T) &= \mathrm{E}\left[(T-\alpha)^2\right] \\
&= \mathrm{E}\left[\left\{T - \mathrm{E}(T) + \mathrm{E}(T) - \alpha\right\}^2\right] \\
&= \mathrm{E}\left[\left\{T - \mathrm{E}(T)\right\}^2 + \{T - \mathrm{E}(T)\}\{\mathrm{E}(T) - \alpha\} + \left\{\mathrm{E}(T) - \alpha\right\}^2\right] \\
&= \mathrm{E}\left[\left\{T - \mathrm{E}(T)\right\}^2\right] + [\mathrm{E}(T) - \alpha]\,\mathrm{E}[T - \mathrm{E}(T)] + [\mathrm{E}(T) - \alpha]^2 \\
&= \mathrm{Var}(T) + 0 + [\mathrm{bias}(T)]^2
\end{aligned}$$

[4 marks: one for each of the two definitions and two for showing mse = bias^2 + var.]

(b) The joint PDF of X and Y is given by

$$f_{XY}(x,y) = f_{Y|X}(y|x) f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-x)^2}{2}\right\} \times \frac{1}{\sqrt{2\pi v^2}} \exp\left\{-\frac{(x-\alpha u)^2}{2v^2}\right\} \tag{1}$$

for $-\infty < x, y < +\infty$. The marginal distribution of Y can be obtained by integration,

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx. \tag{2}$$

We can simplify this integral by completing the square in x in (1),

$$\begin{aligned}
f_{XY}(x,y) &= \frac{1}{2\pi v} \exp\left\{-\frac{v^2(y^2 - 2xy + x^2) + (x^2 - 2\alpha x u + \alpha^2 u^2)}{2v^2}\right\} \\
&= \frac{1}{2\pi v} \exp\left\{-\frac{1+v^2}{2v^2}\left(x - \frac{yv^2 + \alpha u}{1+v^2}\right)^2 - \frac{1}{2v^2}\left(v^2 y^2 + \alpha^2 u^2 - \frac{(yv^2 + \alpha u)^2}{1+v^2}\right)\right\} \\
&= \frac{1}{2\pi v} \exp\left\{-\frac{1+v^2}{2v^2}\left(x - \frac{yv^2 + \alpha u}{1+v^2}\right)^2 - \frac{(y-\alpha u)^2}{2(1+v^2)}\right\}.
\end{aligned}$$

Substituting this last expression into (2) yields

$$\begin{aligned}
f_Y(y) &= \frac{1}{2\pi v} \exp\left\{-\frac{(y-\alpha u)^2}{2(1+v^2)}\right\} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1+v^2}{2v^2}\left(x - \frac{yv^2+\alpha u}{1+v^2}\right)^2\right\} dx \\
&= \frac{1}{\sqrt{2\pi(1+v^2)}} \exp\left\{-\frac{(y-\alpha u)^2}{2(1+v^2)}\right\}
\end{aligned}$$

for $-\infty < y < +\infty$.
[5 marks: two marks for joint distribution, one for (2), one for completing square, and one for final expression for marginal distribution.]

(c) The likelihood function is given by $L(\alpha) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(1+v^2)}} \exp\left\{-\frac{(y_i - \alpha u_i)^2}{2(1+v^2)}\right\}$ and the loglikelihood function by

$$\ell(\alpha) = -\frac{\sum_{i=1}^{n}(y_i - \alpha u_i)^2}{2(1+v^2)}.$$

We can maximize the loglikeilhood function by setting its first derivative equal to zero and solving,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\ell(\alpha) = \frac{\sum_{i=1}^{n}(y_i - \alpha u_i)u_i}{1+v^2} = 0.$$

i.e.,

$$\sum_{i=1}^{n} y_i u_i - \alpha \sum_{i=1}^{n} u_i^2 = 0.$$

The value of α that solves this equation is $\hat{A}_{\text{MLE}} = \dfrac{\sum_{i=1}^n u_i Y_i}{\sum_{i=1}^n u_i^2}$.

Noting that

$$\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\ell(\alpha) = -\frac{\sum_{i=1}^n u_i^2}{1+v^2} < 0$$

we see that the second derivative test confirms that $\hat{A}_{\text{MLE}}$ indeed maximizes the loglikelihood and is the maximum likelihood estimator of α.

[6 marks: one mark each for (i) likelihood, (ii) loglikelihood, (iii) first derivative, (iv) MLE, (v) second derivative, and (vi) verifying maximum from second derivative.]

(d) First,

$$\mathrm{E}(\hat{A}_{\text{MLE}}) = \mathrm{E}\left(\frac{\sum_{i=1}^n u_i Y_i}{\sum_{i=1}^n u_i^2}\right) = \frac{\sum_{i=1}^n u_i \mathrm{E}(Y_i)}{\sum_{i=1}^n u_i^2} = \alpha\frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n u_i^2} = \alpha.$$

The bias of $\hat{A}_{\text{MLE}}$ as an estimator of α is zero; $\hat{A}_{\text{MLE}}$ is an unbiased estimator of α.
Second, $\text{MSE}(\hat{A}_{\text{MLE}}) = \left(\text{BIAS}(\hat{A}_{\text{MLE}})\right)^2 + \mathrm{Var}(\hat{A}_{\text{MLE}}) = \mathrm{Var}(\hat{A}_{\text{MLE}})$ and

$$\text{MSE}(\hat{A}_{\text{MLE}}) = \mathrm{Var}\left(\frac{\sum_{i=1}^n u_i Y_i}{\sum_{i=1}^n u_i^2}\right) = \frac{\sum_{i=1}^n u_i^2 \mathrm{Var}(Y_i)}{\left(\sum_{i=1}^n u_i^2\right)^2} = \frac{(1+v^2)\sum_{i=1}^n u_i^2}{\left(\sum_{i=1}^n u_i^2\right)^2} = \frac{1+v^2}{\sum_{i=1}^n u_i^2}.$$

[5 marks: two for $\mathrm{E}(\hat{A}_{\text{MLE}})$ and bias, and three for $\mathrm{Var}(\hat{A}_{\text{MLE}})$ and MSE.]

4. (a) The PDF of the multivariate normal distribution with mean zero and variance-covariance matrix equal to the identity matrix is given by $f_{\boldsymbol{X}}(x) = \frac{1}{(2\pi)^{K/2}}\exp\left\{-\boldsymbol{x}^\top\boldsymbol{x}/2\right\}$.
To derive $f_{\boldsymbol{Y}}(\boldsymbol{y})$, note that (i) $\boldsymbol{X} = \boldsymbol{M}^{-1}(\boldsymbol{Y}-\boldsymbol{a})$ and (ii) $\mathrm{d}x_i/\mathrm{d}y_j$ is the (i,j) element of $\boldsymbol{M}^{-1}$ so that the Jacobian of the transformation is $|\det(\boldsymbol{M}^{-1})|$. Thus,

$$\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= f_{\boldsymbol{X}}\left(\boldsymbol{M}^{-1}(\boldsymbol{y}-\boldsymbol{a})\right)|\det(\boldsymbol{M}^{-1})| \\
&= \frac{1}{(2\pi)^{K/2}|\boldsymbol{M}|}\exp\left\{-\frac{1}{2}\left(\boldsymbol{M}^{-1}(\boldsymbol{y}-\boldsymbol{a})\right)^\top\left(\boldsymbol{M}^{-1}(\boldsymbol{y}-\boldsymbol{a})\right)\right\} \\
&= \frac{1}{(2\pi)^{K/2}|\boldsymbol{M}|}\exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{a})^\top(\boldsymbol{M}\boldsymbol{M}^\top)^{-1}(\boldsymbol{y}-\boldsymbol{a})\right\},
\end{aligned}$$

i.e., $\boldsymbol{Y} \sim \mathrm{N}_K\left(\boldsymbol{a}, \boldsymbol{M}\boldsymbol{M}^\top\right)$.

[6 marks: two for the joint PDF of $\boldsymbol{X}$, one for $\boldsymbol{X} = \boldsymbol{M}^{-1}(\boldsymbol{Y}-\boldsymbol{a})$, one for Jacobian, and two for deriving the PDF of $\boldsymbol{Y}$.]

(b) In the notation of part (a), $\boldsymbol{X} \sim \mathrm{N}_2(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{M}\boldsymbol{X}$ with $K = 2$, $\boldsymbol{a} = (\mu_1, \mu_2)^\top$, and

$$\boldsymbol{M} = \begin{pmatrix}\sigma_1 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1-\rho^2}\end{pmatrix} \quad \text{so that} \quad \boldsymbol{M}\boldsymbol{M}^\top = \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$$

and $\boldsymbol{Y} \sim \mathrm{N}_2\left(\begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$, a bivariate normal distribution.

[6 marks: one mark each for (i) K, (ii) $\boldsymbol{a}$, (iii) $\boldsymbol{M}$, (iv) $\boldsymbol{M}\boldsymbol{M}^\top$, (v) distribution of $\boldsymbol{Y}$, and (vi) "bivariate normal".]

(c) $\mathrm{E}(U) = \mathrm{E}\Big[\mathrm{E}(U|Y_1, Y_2)\Big] = \mathrm{E}[\alpha + \beta Y_1] = \alpha + \beta\mu_1$.

[4 marks: two for law of iterated expectations and two for calculations/answer.]

(d) $\mathrm{Var}(U) = \mathrm{Var}\Big[\mathrm{E}(U|Y_1, Y_2)\Big] + \mathrm{E}\Big[\mathrm{Var}(U|Y_1, Y_2)\Big] = \mathrm{Var}[\alpha + \beta Y_1] + \mathrm{E}[\gamma + \tau Y_2^2]$
$= \beta^2\sigma_1^2 + \gamma + \tau(\sigma_2^2 + \mu_2^2)$.
[4 marks: two for law of total variance and two for calculations/answer.]