Imperial College
London

M2S1

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2019

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

Probability and Statistics 2

Date: Wednesday 22 May 2019

Time: 10.00 - 12.00

Time Allowed: 2 Hours

This paper has 4 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.

All required additional material will be provided.

- 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.
- Calculators may not be used.

1. Let $X_1, X_2, \ldots X_n \sim N(\mu, \sigma^2)$ be a random sample.

(a) Write down the likelihood function $\Pr(x_1, \ldots, x_n | \mu, \sigma^2)$.

(b) Determine the maximum likelihood estimators of μ and σ^2, being sure to verify that they are indeed maxima.

(c) Obtain the bias and variance of the estimators determined above, stating clearly any general results that you use.

Consider now the case in which $\mu = 0$ is known, and let $T = \sum_{i=1}^{n} X_i^2$.

(d) Stating any general results used, determine the distribution of T and find the constant α such that αT is an unbiased estimator of σ^2.

(e) Show that $U = \frac{\sqrt{T}}{\beta}$ is an unbiased estimator of σ, where

$$\beta = \frac{\sqrt{2}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

2. An urn contains N balls in total, r of which are red. Balls are removed from the urn at random. Define the indicator variable Z_i to be 1 if the ith ball removed is red, and 0 otherwise. When n balls have been removed, let S_n be the number of removed balls that are red.

(a) If n balls are removed *with* replacement, write down the probability distribution of each of the random variables Z_i. Use your answer to find the mean and variance of Z_i, and hence determine the mean and variance of S_n.

(b) If $n < r$ balls are removed *without* replacement, show that the variables Z_i are identically distributed but not independent, and hence determine the mean and variance of S_n.

(c) Suppose instead X balls are removed *with* replacement, where X is a Poisson distributed random variable with mean n. Find the mean and variance of S_X.

3. (a) Consider a probability space $(\Omega, \mathcal{F}, \Pr)$.

   (i) What is meant by saying that $\mathcal{F}$ is a sigma algebra?

   (ii) What is meant by saying that $\Pr$ is a probability function?

   (iii) Stating clearly the properties of $\mathcal{F}$ and $\Pr$ you use, show that the collection

$$\mathcal{F}_0 = \{A \in \mathcal{F} | \Pr(A) \in \{0,1\}\}$$

   is a sigma algebra.

(b) Let $Z_0, Z_1, Z_2, Z_3 \sim N(0,1)$ be independent normally distributed variables, and for $i = 1,2,3$, define

$$X_i = Z_i + \beta Z_{i-1}, \qquad Y_i = Z_i^2, \text{ and } \qquad Y = Y_1^2 + Y_2^2 + Y_3^3.$$

Stating clearly any general results that you use, determine

   (i) the joint distribution of (X_1, X_2, X_3),

   (ii) the joint distribution of $(\frac{Y_1}{Y}, \frac{Y_2}{Y}, \frac{Y_3}{Y})$.

4. (a) Define what it means for a sequence $Z_1, Z_2, \ldots$ of random variables to converge in probability to a random variable Z.

(b) Show that for any random variable Z and any non-negative function $g : \mathbf{R} \to \mathbf{R}$ for which the expectation of $g(Z)$ exists,

$$\Pr(g(Z) > r) \leq \frac{\mathrm{E}(g(Z))}{r},$$

and deduce that if Z is a random variable with expectation μ and variance σ^2,

$$\Pr(|Z - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

Let $X_1, X_2, \ldots X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples from a $\textsc{Poisson}(\lambda)$ distribution. For any $0 < p < 1$, define $\widehat{\lambda_p} = p\bar{X} + (1-p)\bar{Y}$, where $\bar{X}$ and $\bar{Y}$ denote the respective sample means.

(c) Show that $\widehat{\lambda_p}$ is an unbiased estimator of λ.

(d) Show that $\widehat{\lambda_p}$ defined above converges in probability to λ as both $m \to \infty$ and $n \to \infty$.

(e) Find the value p_0 of p for which $\widehat{\lambda_p}$ has minimal variance, and determine the variance of $\widehat{\lambda_{p0}}$.

(f) Define what it means for a sequence $Z_1, Z_2, \ldots$ of random variables to converge in distribution to a random variable Z.

(g) State the central limit theorem, and use it to give a transformation of $\widehat{\lambda_{p0}}$ (depending on λ) that converges in distribution to a standard normal random variable as the total sample size becomes large.

DISCRETE DISTRIBUTIONS

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

The PDF of the *multivariate normal distribution* is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{K/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\Big\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big\},$$

for $\boldsymbol{x} \in \mathbb{R}^K$ with $\boldsymbol{\Sigma}$ a $(K \times K)$ variance-covariance matrix and $\boldsymbol{\mu}$ a $(K \times 1)$ mean vector.

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

$$f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) \qquad F_Y(y) = F_X\left(\frac{y-\mu}{\sigma}\right)$$

$$M_Y(t) = e^{\mu t}M_X(\sigma t) \qquad \mathrm{E}[Y] = \mu + \sigma\mathrm{E}[X] \qquad \mathrm{Var}[Y] = \sigma^2\mathrm{Var}[X]$$

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

CONTINUOUS DISTRIBUTIONS

		parameters	*pdf*	*cdf*	E[X]	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	$e^{\{\mu t+\sigma^2t^2/2\}}$
$Student(\nu)$	$\mathbb{R}$	$\nu\in\mathbb{R}^+$	$\dfrac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left\{1+\frac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		0 (if $\nu>1$)	$\dfrac{\nu}{\nu-2}$ (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 SOLUTIONS

1. (a) [2 marks]

$$L(\mu,\sigma^2) = \Pr(x_1,\ldots,x_n|\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$
$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right).$$

[Seen]

(b) [6 marks] First obtain the log-likelihood

$$l(\mu,\sigma^2) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}.$$

Now differentiate partially with respect to μ and σ^2

$$\frac{\partial l(\mu,\sigma^2)}{\partial\mu} = \frac{\sum_{i=1}^{n} 2(x_i-\mu)}{2\sigma^2}.$$
$$\frac{\partial l(\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2(\sigma^2)^2}.$$

Setting the two equations above simultaneously to zero gives the estimates

$$\widehat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}.$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}.$$

To confirm that these estimates do indeed maximize the likelihood, examine the second derivative of the log-likelihood evaluated at the estimate.

$$\left.\frac{\partial^2 l(\mu,\sigma^2)}{\partial\mu^2}\right|_{\widehat{\mu},\widehat{\sigma^2}} = -\frac{n}{\sigma^2} < 0.$$
$$\left.\frac{\partial l(\mu,\sigma^2)}{\partial(\sigma^2)^2}\right|_{\widehat{\mu},\widehat{\sigma^2}} = \frac{n}{2(\sigma^2)^2} - \left.\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{(\sigma^2)^3}\right|_{\widehat{\mu},\widehat{\sigma^2}} = -\frac{n}{2(\widehat{\sigma^2})^2} < 0.$$
$$\left.\frac{\partial l(\mu,\sigma^2)}{\partial\mu\partial\sigma^2}\right|_{\widehat{\mu},\widehat{\sigma^2}} = -\left.\frac{\sum_{i=1}^{n}(x_i-\mu)}{(\sigma^2)^2}\right|_{\widehat{\mu},\widehat{\sigma^2}} = 0.$$

The hessian of the log likelihood is immediately seen to be negative definite, so these estimates do indeed maximize the likelihood.

The estimators are

$$\widehat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}, \qquad \widehat{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}.$$

[Seen]

(c) [5 marks]$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, so the bias and variance for $\widehat{\mu}$ are 0 and $\frac{\sigma^2}{n}$, respectively.

$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$, which is a gamma distribution with shape parameter $\frac{n-1}{2}$ and rate parameter $\frac{1}{2}$. This has mean $n-1$ and variance $2(n-1)$. Hence the bias of $\widehat{\sigma^2}$ is

$$\frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n},$$

and its variance is

$$2(n-1)\left(\frac{\sigma^2}{n}\right)^2 = \frac{2\sigma^4}{n}\left(1 - \frac{1}{n}\right).$$

[Seen]

(d) [3 marks] $\frac{T}{\sigma^2} \sim \chi^2(n)$, which is a gamma distribution with shape parameter $\frac{n}{2}$ and rate parameter $\frac{1}{2}$, hence, since gamma distributions form a scale family, T has a gamma distribution with shape parameter $\frac{n}{2}$ and rate parameter $\frac{1}{2\sigma^2}$. Hence $\mathrm{E}(T) = n\sigma^2$, and so $\frac{1}{n}T$ is an unbiased estimator of σ^2, i.e. $\alpha = \frac{1}{n}$.

[Seen Similar]

(e) [4 marks]

$$\begin{aligned}
\mathrm{E}(\sqrt{T}) &= \int_0^\infty t^{\frac{1}{2}} \frac{t^{\frac{n}{2}-1}}{(2\sigma^2)^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{t}{2\sigma^2}}\, dt \\
&= \frac{1}{(2\sigma^2)^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_0^\infty t^{\frac{n+1}{2}-1} e^{-\frac{t}{2\sigma^2}}\, dt \\
&= \frac{(2\sigma^2)^{\frac{n+1}{2}}\Gamma(\frac{n+1}{2})}{(2\sigma^2)^{\frac{n}{2}}\Gamma(\frac{n}{2})} = \sqrt{2}\sigma \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}
\end{aligned}$$

Using the fact that the integrand is an un-normalized gamma density. Hence taking $U = \frac{\sqrt{T}}{\beta}$ as given in the question gives an unbiased estimator.

[Unseen]

2. (a) [6 marks] As sampling occurs with replacement, the random variables Z_i are independent, and

$$\Pr(Z_i = k) = \begin{cases} \frac{r}{N} & k = 1 \\ \frac{N-r}{N} & k = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Hence

$$\mathrm{E}(Z_i) = 1 \times \frac{r}{N} + 0 \times \frac{N-r}{N} = \frac{r}{N}.$$

Linearity of expectation then gives

$$\mathrm{E}(S_n) = \sum_{i=1}^{n} \mathrm{E}(Z_i) = \frac{nr}{N}.$$

For the variance, similarly

$$\mathrm{E}(Z_i^2) = 1 \times \frac{r}{N} + 0 \times \frac{N-r}{N} = \frac{r}{N}.$$

Hence,

$$\mathrm{Var}(Z_i) = \mathrm{E}(Z_i^2) - \mathrm{E}(Z_i)^2 = \frac{r(N-r)}{N^2}.$$

Using the independence of the Z_i,

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(Z_i) = \frac{nr(N-r)}{N^2}.$$

[Seen]

(b) [9 marks] Each of the $\binom{N}{n}$ configurations of removed balls is equally likely, and there is a one-to-one correspondence between configurations with a red ball in position i and those with a red ball in position j. Hence $\Pr(Z_i = 1) = \frac{r}{N} = \Pr(Z_j = 1)$.

For $i \neq j$, the variables Z_i and Z_j are not independent since

$$\Pr(Z_i = 1, Z_j = 1) = \frac{r}{N}\frac{(r-1)}{N-1} \neq \Pr(Z_i = 1)\Pr(Z_j = 1).$$

To determine the mean of S_n, note that as in the previous part

$$\mathrm{E}(Z_i) = 1 \times \frac{r}{N} + 0 \times \frac{N-r}{N} = \frac{r}{N},$$

so again by linearity of expectation

$$\mathrm{E}(S_n) = \sum_{i=1}^{n} \mathrm{E}(Z_i) = \frac{nr}{N}.$$

To determine the variance of S_n, first note that

$$\mathrm{Cov}(Z_i, Z_j) = \mathrm{E}(Z_i Z_j) - \mathrm{E}(Z_i)\mathrm{E}(Z_j) = \frac{r}{N}\frac{(r-1)}{N-1} - \frac{r^2}{N^2},$$

Then

$$\begin{aligned}
\mathrm{Var}(S_n) = \mathrm{Var}\left(\sum_{i=1}^{n} Z_i\right) &= \sum_{i=1}^{n} \mathrm{Var}(Z_i) + 2\sum_{i<j} \mathrm{Cov}(Z_i, Z_j) \\
&= n\frac{r}{N}\frac{N-r}{N} + n(n-1)\left(\frac{r}{N}\frac{(r-1)}{N-1} - \frac{r^2}{N^2}\right) \\
&= n\frac{r}{N}\left[\frac{N-r}{N} + (n-1)\left(\frac{r-1}{N-1} - \frac{r}{N}\right)\right] \\
&= n\frac{r}{N}\left[\frac{N-r}{N} + (n-1)\frac{r-N}{N(N-1)}\right] \\
&= n\frac{r}{N}\frac{N-r}{N}\left[1 - \frac{n-1}{N-1}\right] = n\frac{r}{N}\frac{(N-r)}{N}\frac{(N-n)}{N-1}.
\end{aligned}$$

[Seen] (Optional question on problem sheet)

(c) [5 marks] Using the law of iterated expectation, and part (a),

$$\mathrm{E}(S) = \mathrm{E}\left(\mathrm{E}(S|X)\right) = \mathrm{E}\left(\frac{Xr}{N}\right) = \frac{nr}{N}.$$

Now, using the law of total variance,

$$\begin{aligned}
\mathrm{Var}(S) &= \mathrm{E}\left(\mathrm{Var}\left(S|X\right)\right) + \mathrm{Var}\left(\mathrm{E}\left(S|X\right)\right) \\
&= \mathrm{E}\left(\frac{Xr(N-r)}{N^2}\right) + \mathrm{Var}\left(\frac{Xr}{N}\right) \\
&= \frac{nr(N-r)}{N^2} + \frac{nr^2}{N^2} = \frac{nr}{N}.
\end{aligned}$$

Alternatively, show $S \sim \textsc{Poisson}\left(\frac{nr}{N}\right)$ by

$$\begin{aligned}
\Pr(S = s) = \sum_{k=0}^{\infty} \Pr(S = s|X = k)\Pr(X = k) &= \sum_{k=s}^{\infty} \binom{k}{s}\left(\frac{r}{N}\right)^s \left(1 - \frac{r}{N}\right)^{k-s} e^{-n}\frac{n^k}{k!} \\
&= n^s \left(\frac{r}{N}\right)^s \frac{e^{-n}}{s!} \sum_{k=s}^{\infty} \frac{1}{(k-s)!}\left(1 - \frac{r}{N}\right)^{k-s} n^{k-s} \\
&= \left(\frac{nr}{N}\right)^s \frac{e^{-\frac{nr}{N}}}{s!},
\end{aligned}$$

so $\mathrm{E}(S) = \mathrm{Var}(S) = \frac{nr}{N}$.
[Seen Similar]

3. (a) (i) [2 marks] $\mathcal{F}$ is a sigma algebra if

- $\emptyset \in \mathcal{F}$
- $A^c \in \mathcal{F}$ whenever $A \in \mathcal{F}$
- $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$ whenever $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}$.

[Seen]

(ii) [2 marks] $\Pr : \mathcal{F} \to [0,1]$ is said to be a probability function if

- $\Pr(A) \geq 0$ for all $A \in \mathcal{F}$,
- $\Pr(\Omega) = 1$,
- whenever $A_1, A_2, \ldots$ is a sequence of pairwise disjoint sets in $\mathcal{F}$,

$$\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i).$$

[Seen]

(iii) [8 marks]

Note first that since $\mathcal{F}$ is a sigma algebra, all sets considered lie in the domain of the probability function $\Pr$. Need to show that the three conditions hold.

- Since $\Pr(\emptyset) = 0$, certainly $\emptyset \in \mathcal{F}_0$.
- If $A \in \mathcal{F}_0$ is such that $\Pr(A) = 1$, then $\Pr(A^c) = 1 - \Pr(A) = 0$ so $A^c \in \mathcal{F}_0$; symmetrically, if $\Pr(A) = 0$, $\Pr(A^c) = 1$, so again $A^c \in \mathcal{F}_0$.
- Suppose $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}_0$. Consider two cases: either for each i, $\Pr(A_i) = 0$. or there exists an index j such that $\Pr(A_j) = 1$.

  In the former case, for any two sets A and B with $\Pr(A) = \Pr(B) = 0$, $A \cap B \subset A$, so $0 \leq \Pr(A \cap B) \leq \Pr(A) = 0$.
  Then we see that

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) = 0.$$

  By induction, we see that for any n,

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = 0.$$

  By the continuity property of the probability function $\Pr$,

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \Pr\left(\bigcup_{i=1}^{n} A_i\right) = 0.$$

  so $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_0$.
  In the latter case, if (say) $\Pr(A_j) = 1$, then

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \Pr(A_j) + \Pr\left(\bigcup_{i\neq j} A_i\right) - \Pr\left(A_j \cap \bigcup_{i\neq j} A_i\right).$$

  Since $A_j \cap \bigcup_{i\neq j} A_i \subset \bigcup_{i\neq j} A_i$, $\Pr(A_j \cap \bigcup_{i\neq j} A_i) \leq \Pr(\bigcup_{i\neq j} A_i)$. Substituting this inequality into the equation above then gives

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) \geq \Pr(A_j) = 1.$$

Hence, since $\Pr$ is a probability function,

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = 1,$$

and so $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_0$.

[Unseen]

(b) (i) [6 marks] As a linear transformation of the standard normal variables (Z_0, Z_1, Z_2, Z_3), (X_1, X_2, X_3) is multivariate normal.

Note that for each i, using the linearity of expectation,

$$\mathrm{E}(X_i) = \mathrm{E}(Z_i) + \beta\mathrm{E}(Z_{i-1}) = 0.$$

Further, by independence of the Z_i,

$$\mathrm{Var}(X_i) = \mathrm{Var}(Z_i) + \beta^2\mathrm{Var}(Z_{i-1}) = 1 + \beta^2.$$

To determine covariances,

$$\mathrm{Cov}(X_1, X_2) = \mathrm{E}(X_1X_2) - \mathrm{E}(X_1)\mathrm{E}(X_2) = \mathrm{E}(X_1X_2) = \mathrm{E}((Z_1 + \beta Z_0)(Z_2 + \beta Z_1)).$$

By linearity of expectation, this is

$$\mathrm{E}(Z_1Z_2) + \beta\mathrm{E}(Z_1^2) + \beta\mathrm{E}(Z_0Z_2) + \beta^2\mathrm{E}(Z_0Z_1) = 0 + \beta + 0 + 0 = \beta,$$

as if $i \neq j$, Z_i and Z_j are independent, and so $\mathrm{E}(Z_iZ_j) = \mathrm{E}(Z_i)\mathrm{E}(Z_j) = 0$.

By symmetry, we also have

$$\mathrm{Cov}(X_2, X_3) = \beta.$$

Finally,

$$\mathrm{Cov}(X_1, X_3) = \mathrm{E}(X_1X_3) - \mathrm{E}(X_1)\mathrm{E}(X_3) = \mathrm{E}(X_1X_3) = \mathrm{E}((Z_1 + \beta Z_0)(Z_3 + \beta Z_2)),$$

and so $\mathrm{Cov}(X_1, X_3) = 0$, since no Z_i term appears twice in the above. Hence the required distribution is

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathrm{N}_3 \left[\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1+\beta^2 & \beta & 0 \\ \beta & 1+\beta^2 & \beta \\ 0 & \beta & 1+\beta^2 \end{pmatrix}\right].$$

[Seen Method]

(ii) [2 marks] For each i, $Y_i \sim \chi^2(1)$, which is $\Gamma(\frac{1}{2}, \frac{1}{2})$ (using the rate parameterization).

A general result is that if for $i = 1, 2, \ldots k$, $U_i \sim \Gamma(\alpha_i, \beta)$ independent, then

$$\frac{1}{\sum_{i=1}^{k} U_i}(U_1, U_2, \ldots U_k) \sim \textsc{Dirichlet}(\alpha_1, \alpha_2, \ldots \alpha_k),$$

hence $(\frac{Y_1}{Y}, \frac{Y_2}{Y}, \frac{Y_3}{Y}) \sim \textsc{Dirichlet}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

4. (a) [1 mark] A sequence of random variables $Z_1, Z_2, \ldots$ converges in probability to the random variable Z if for all $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(|Z_n - Z| > \epsilon) = 0.$$

[Seen]

(b) [3 marks]

Since $g(z) \geq 0$,

$$\mathrm{E}\left(g(Z)\right) = \int_{-\infty}^{\infty} g(z) f_Z(z)\, dz \geq \int_{z:g(z)>r} g(z) f_Z(z)\, dz$$

and since on this region of integration, $g(z) > r$, we have

$$\mathrm{E}\left(g(Z)\right) \geq \int_{z:g(z)>r} r f_Z(z)\, dz = r \Pr(g(Z) > r),$$

which establishes the required inequality.

Now, applying this result to the particular non-negative function $g(Z) = (Z-\mu)^2$,

$$\Pr(|Z-\mu| > \epsilon) = \Pr\left((Z-\mu)^2 > \epsilon^2\right) \leq \frac{\mathrm{E}\left((Z-\mu)^2\right)}{\epsilon^2}.$$

Noting that $\mathrm{E}\left((Z-\mu)^2\right) = \mathrm{Var}(Z) = \sigma^2$ gives

$$\Pr(|Z-\mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2},$$

as required.

[Seen Similar]

(c) [2 marks] By linearity of expectation,

$$\mathrm{E}(\widehat{\lambda_p}) = p\mathrm{E}(\bar{X}) + (1-p)\mathrm{E}(\bar{Y}).$$

Now $E(\bar{X}) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{E}(X_i) = \frac{m\lambda}{m} = \lambda$, and similarly $\mathrm{E}(\bar{Y}) = \lambda$.

Hence

$$\mathrm{E}(\widehat{\lambda_p}) = p\lambda + (1-p)\lambda = \lambda,$$

so $\widehat{\lambda}_p$ is unbiased for p.

[Seen Method]

(d) [4 marks]

$$\Pr(|\widehat{\lambda}_p - \lambda| > \epsilon) \leq \frac{\mathrm{Var}(\widehat{\lambda}_p)}{\epsilon^2} = \frac{1}{\epsilon^2}\left(p^2\mathrm{Var}(\bar{X}) + (1-p)^2\mathrm{Var}(\bar{Y})\right) = \frac{1}{\epsilon^2}\left(p^2\frac{\lambda}{m} + (1-p)^2\frac{\lambda}{n}\right),$$

using $\mathrm{Var}(\bar{X}) = \frac{1}{m^2}\sum_{i=1}^{m} \mathrm{Var}(X_i) = \frac{\lambda}{m}$, and similarly for $\mathrm{Var}(\bar{Y})$.

Hence as $m, n \to \infty$, $\Pr(|\widehat{\lambda}_p - \lambda| > \epsilon) \to 0$, and $\widehat{\lambda}_p$ converges in probability to λ.

[Seen Method]

(e) [4 marks] As determined above,

$$\text{Var}(\widehat{\lambda}_p) = p^2\text{Var}(\bar{X}) + (1-p)^2\text{Var}(\bar{Y}) = p^2\frac{\lambda}{m} + (1-p)^2\frac{\lambda}{n},$$

so seek to minimize

$$p^2\frac{1}{m} + (1-p)^2\frac{1}{n}$$

as a function of p. This expression is a quadratic in p with a unique minimum, which can be found by setting the derivative to zero.

$$2p\frac{1}{m} - 2(1-p)\frac{1}{n} = 0,$$

so that $pn = (1-p)m$ and on solving, $p_0 = \frac{m}{m+n}$. This then gives

$$\text{Var}(\widehat{\lambda}_{p_0}) = \frac{m}{(m+n)^2}\lambda + \frac{n}{(m+n)^2}\lambda = \frac{\lambda}{m+n}.$$

[Unseen]

(f) [2 marks] A sequence $Z_1, Z_2, \ldots$ converges in distribution to a random variable Z if

$$\lim_{n\to\infty} F_{Z_n}(x) = F_Z(x)$$

at all points of continuity of F_Z.

[Seen]

(g) [4 marks] Let $Z_1, Z_2, \ldots$ be a sequence of independent, identically distributed random variables with $\text{E}(Z_i) = \mu$ and $\text{Var}(Z_i) = \sigma^2 < \infty$. Then defining $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$, the rescaled sequence $\sqrt{n}\frac{(\bar{Z}_n - \mu)}{\sigma}$ converges in distribution to a standard normal variable, i.e.

$$\lim_{n\to\infty} \Pr\left(\sqrt{n}\frac{(\bar{Z}_n - \mu)}{\sigma} \leq z\right) = \Phi(z).$$

Here, for $p_0 = \frac{m}{m+n}$,

$$\widehat{\lambda}_{p_0} = \frac{m}{m+n}\bar{X} + \frac{n}{m+n}\bar{Y} = \frac{\sum_{i=1}^m X_i + \sum_{j=1}^n Y_j}{m+n}$$

is just the sample mean of $m+n$ independent, identically distributed random variables.

Hence, by the central limit theorem,

$$\lim_{m+n\to\infty} \Pr\left(\sqrt{\frac{m+n}{\lambda}}(\lambda_{p_0} - \lambda) \leq z\right) = \Phi(z),$$

hence $\sqrt{\frac{m+n}{\lambda}}(\lambda_{p_0} - \lambda)$ converges in distribution to a standard normal variable.

[Unseen] (theorem is familiar, this example is unseen)