Imperial College
London

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May – June 2015

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

Probability & Statistics II

Date: Monday, 18 May 2015. 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 5 & 6.

- 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) Suppose $(X_1, \ldots, X_n)$ are independent and identically distributed normal random variables, that is $X_i \overset{\text{iid}}{\sim} \text{Norm}(\mu, \sigma^2)$ for $i = 1, \ldots, n$. Assume that both μ and σ^2 are unknown.

Give a pivot for μ and state the distribution of the pivot.
Derive a $100(1-\alpha)\%$ confidence interval for μ.

Identify the following distributions:

(i) The distribution of

$$V_1 = \left(\frac{X_1 - \mu}{\sigma}\right)^2$$

(ii) The distribution of

$$V_2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$$

(iii) The distribution of

$$V_3 = \sum_{i=1}^{n} (X_i - \mu)^2$$

(b) Suppose X_1 and X_2 are independent random variables, such that $X_i \overset{\text{indep}}{\sim} \text{Norm}(0, \sigma_i^2)$.

Identify the distribution of $V_4 = \dfrac{X_1/\sigma_1}{|X_2/\sigma_2|}$.
Give an expression for the probability density function of $V_5 = X_1/|X_2|$.

(c) Suppose each $\boldsymbol{X}_i = (X_{1i}, X_{2i})^{\mathrm{T}}$ is a (2×1) random vector with

$$\boldsymbol{X}_i \overset{\text{indep}}{\sim} \mathrm{N}_2(\boldsymbol{\mu}, \Sigma), \quad \text{for} \quad i = 1, \ldots n,$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)^{\mathrm{T}}$ is a (2×1) mean vector and $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ is a (2×2) variance-covariance matrix. Let Φ be the cumulative distribution function of the standard normal distribution and M be the number of the $\boldsymbol{X}_i$ among $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ for which X_{1i} is greater than X_{2i}.

Give an expression for the joint probability density function of $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$.
Using a theorem from the notes, derive the distribution of $U_i = X_{1i} - X_{2i}$.
Derive $\pi = \Pr(X_{1i} > X_{2i})$. You may express your answer in terms of Φ.
What is the distribution of M? Justify your answer.

2. This question compares frequency and Bayesian estimators in a particular Poisson model.

(a) Suppose $X|\lambda \sim \text{POISSON}(\lambda)$ and $Y|\lambda,\xi \sim \text{POISSON}(\lambda\xi)$, with X and Y independent.
Derive the maximum likelihood estimates of λ and ξ.
For simplicity you may ignore the possibility of $x = 0$ and/or $y = 0$.

For the remainder of this problem, suppose $X_i \overset{\text{iid}}{\sim} \text{POISSON}(\lambda)$, for $i = 1, \ldots, n$.

(b) For a Bayesian analysis, assign the prior distribution $\Lambda \sim \text{GAMMA}(\alpha, \beta)$, where the gamma distribution is parameterized as in the formula sheet.
Derive the posterior distribution for Λ given $X_1, \ldots, X_n$.
Give an expression for the posterior mean of Λ. Denote this estimate of λ by $\widehat{\lambda}_1$ and the corresponding estimator of λ by $\widehat{\Lambda}_1$.
Derive the marginal mean and variance of X_i. (Do not condition on $\Lambda = \lambda$.)
Derive the correlation between X_i and X_j, for $i \neq j$. (Do not condition on $\Lambda = \lambda$.)

(c) In this part you will compare the frequency properties of two estimators of λ, specifically $\widehat{\Lambda}_1$ derived in part (b) and $\widehat{\Lambda}_2 = \frac{1}{n}\sum_{i=1}^n X_i$. You should treat $X_1, \ldots, X_n$ as random variables and treat λ as a constant. (This does not preclude the use of the estimator, $\widehat{\Lambda}_1$, derived in part (b).)
Derive the mean square error of $\widehat{\Lambda}_2$.
Derive the mean square error of $\widehat{\Lambda}_1$. (For simplicity, set $\alpha = 0$ and $\beta = 1$.)
For what values of λ does $\widehat{\Lambda}_1$ have a smaller mean square error than $\widehat{\Lambda}_2$?

3. Suppose $U_1, U_2, \ldots$ is a sequence of independent random variables that are uniformly distributed on $(0, 1)$. That is, $U_i \overset{\text{iid}}{\sim} \text{UNIFORM}(0, 1)$ for $i = 1, 2, \ldots$. For $n = 1, 2, \ldots$ and for some $\lambda > 0$, define $X_n = \prod_{i=1}^n U_i^\lambda$ and $Y_n = -\sum_{i=1}^n \log(U_i^\lambda) = -\log(X_n)$.

(a) Derive the probability density function of Y_1.
Use a result from the notes to obtain the distribution of Y_n.
Construct a Normal Approximation for the distribution of Y_n.

(b) Derive the probability density function of X_n.
Derive $\text{E}(X_n)$.

(c) Prove that $\Pr(X_n \leq x) \geq 1 - (1 - \sqrt[\lambda]{x})^n$ for $0 \leq x \leq 1$.
Prove that $X_n \xrightarrow{\mathcal{D}} 0$.

4. Suppose that I have a 10p and a 20p coin and that the coins are weighted so that when flipped the 10p coin has probability θ of coming up heads and the 20p coin has probability π of coming up heads. In each of n trials I flip both coins. Let X be the number among the n trials in which the 10p coin comes up heads and let Y be the number among the n trials in which *both* coins come up heads. You may assume (i) the trials are mutually independent, (ii) the outcomes of the two flips in each trial are independent, (iii) n is fixed in advance, and (iv) θ and π do not vary among the n trials.

(a) State the marginal distribution of X.
State the marginal distribution of Y.
State the conditional distribution of Y given X.
Derive the joint probability mass function of X and Y.
Are X and Y independent? Justify your answer.

(b) Derive the conditional probability mass function of X given Y.
Derive the conditional expectation of X given Y.

(c) Derive method of moments estimators of θ and π by setting up a system of two equations, one involving the first (marginal) moment of X and the other involving the first (marginal) moment of Y. Denote your estimators by $\widehat{\Theta}_{\text{MoM}}$ and $\widehat{\Pi}_{\text{MoM}}$.
For what values of X and Y are $\widehat{\Theta}_{\text{MoM}}$ and $\widehat{\Pi}_{\text{MoM}}$ defined?
Is the methods of moments estimator of θ biased? Prove your answer.
Consider the estimator of π given by

$$\widehat{\Pi} = \begin{cases} 0 & \text{if } X = 0 \\ \widehat{\Pi}_{\text{MoM}} & \text{otherwise} \end{cases}.$$

Is $\widehat{\Pi}$ an unbiased estimator or π? Justify your answer.

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	$\theta(1-\theta)$	$1-\theta+\theta e^t$
$Binomial(n,\theta)$	$\{0,1,...,n\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$		$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta e^t)^n$
$Poisson(\lambda)$	$\{0,1,2,...\}$	$\lambda \in \mathbb{R}^+$	$\frac{e^{-\lambda}\lambda^x}{x!}$		λ	λ	$\exp\{\lambda(e^t-1)\}$
$Geometric(\theta)$	$\{1,2,...\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\frac{1}{\theta}$	$\frac{(1-\theta)}{\theta^2}$	$\frac{\theta e^t}{1-e^t(1-\theta)}$
$NegBinomial(n,\theta)$	$\{n,n+1,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$		$\frac{n}{\theta}$	$\frac{n(1-\theta)}{\theta^2}$	$\left(\frac{\theta e^t}{1-e^t(1-\theta)}\right)^n$
or	$\{0,1,2,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n+x-1}{x}\theta^n(1-\theta)^x$		$\frac{n(1-\theta)}{\theta}$	$\frac{n(1-\theta)}{\theta^2}$	$\left(\frac{\theta}{1-e^t(1-\theta)}\right)^n$

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

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

CONTINUOUS DISTRIBUTIONS

		parameters	*pdf*	*cdf*	$\mathrm{E}[X]$	$\mathrm{Var}[X]$	*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 — May 2015 Exam — Solution

1. (a) *[Seen]* Let $T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}}$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$.
$T \sim t_{n-1}$ is a pivot for μ.

Let F_T be the cumulative distribution function of the t_{n-1} distribution and $F_T(t_{\alpha/2}) = 1 - \dfrac{\alpha}{2}$, then

$$\Pr\left(-t_{\alpha/2} \;\le\; \frac{\bar{X}-\mu}{S/\sqrt{n}} \;\le\; t_{\alpha/2}\right) = 100(1-\alpha)\%,$$

so that

$$\Pr\left(-\frac{t_{\alpha/2}S}{\sqrt{n}} - \bar{X} \;\le\; -\mu \;\le\; \frac{t_{\alpha/2}S}{\sqrt{n}} - \bar{X}\right) = 100(1-\alpha)\%,$$

and

$$\Pr\left(\bar{X} - \frac{t_{\alpha/2}S}{\sqrt{n}} \;\le\; \mu \;\le\; \bar{X} + \frac{t_{\alpha/2}S}{\sqrt{n}}\right) = 100(1-\alpha)\%,$$

i.e., $\bar{X} \pm \dfrac{S}{\sqrt{n}} t_{\alpha/2}$ is a $100(1-\alpha)\%$ CI for μ.

The distributions are:

(i) *[Seen]* V_1 is the square of a standard normal, thus $V_1 \sim \chi^2_1$, or $V_1 \sim \text{GAMMA}\left(\frac{1}{2}, \frac{1}{2}\right)$.

(ii) *[Seen]* V_2 is the sum of n independent χ^2_1 random variables, thus $V_2 \sim \chi^2_n$, or $V_2 \sim \text{GAMMA}\left(\frac{n}{2}, \frac{1}{2}\right)$.

(iii) *[Seen Similar]* $V_3 = \sigma^2 V_2$ and identifying the scale parameter of the gamma distribution, $V_3 \sim \text{GAMMA}\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right)$.

[8 marks: One mark each for stating (i) the pivot, (ii) its distribution, and the distributions of (iii) V_1, (iv) V_2, and (v) V_3. Three marks for deriving the confidence interval.]

(b) *[Seen]* V_4 is a standard normal random variable divided by an independent $\sqrt{\chi^2_\nu/\nu}$, with $\nu = 1$, thus $V_4 \sim t_1$, or $V_4 \sim \text{CAUCHY}$.

[Seen Similar] Because $V_5 = \sigma_1 V_4/\sigma_2$ we can express its density function by identifying σ_1/σ_2 as a scale parameter,

$$f_{V_5}(v) = \frac{\sigma_2}{\sigma_1} f_{V_4}\left(\frac{\sigma_2 v}{\sigma_1}\right) = \frac{\sigma_2}{\sigma_1 \pi \left(1 + \left(\frac{\sigma_2}{\sigma_1} v\right)^2\right)} \quad \text{for} \quad -\infty < v < +\infty$$

[3 marks: One mark each for (i) stating the distribution of V_4, (ii) implicitly using the Cauchy density, and (iii) giving the final pdf of V_5.]

(c) *[Seen]* The joint density function can be written in either of two ways: (i) in scalar notation,

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \left(\frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}}\right)^n \exp\left\{-\frac{1}{2(1-\rho^2)}\sum_{i=1}^{n}\left[\left(\frac{x_{1i}-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_{1i}-\mu_1}{\sigma_1}\right)\left(\frac{x_{2i}-\mu_2}{\sigma_2}\right) + \left(\frac{x_{2i}-\mu_2}{\sigma_2}\right)^2\right]\right\},$$

where $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ and $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ with $\boldsymbol{x}_i = (x_{1i}, x_{2i})^{\mathrm{T}}$
or (ii) more succinctly in matrix notation

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \left(\frac{1}{2\pi|\Sigma|^{1/2}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right\},$$

for (in both cases) $-\infty < x_{1i}, x_{2i} < +\infty$.

[Seen Similar] We know that $B\boldsymbol{X}_i \sim \mathrm{N}(B\boldsymbol{\mu}, B\Sigma B^{\mathrm{T}})$. Setting the first row of B to $(1, -1)$, this implies that $U_i = X_{1i} - X_{2i} \sim \text{Norm}(\mu_1 - \mu_2,\ \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)$.

[Seen Similar] Letting Z be a standard normal random variable,

$$\begin{aligned}\pi &= \Pr(X_{1i} > X_{2i}) = \Pr(U_i > 0) = \Pr\left(Z > \frac{0 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}}\right)\\ &= 1 - \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}}\right) = \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}}\right).\end{aligned}$$

(Either of the last two expressions is acceptable.)

[Seen Similar] Because

(i) there is a fixed number of X_i, i.e., n;
(ii) for each i, either X_{i1} is greater than X_{i2} or it is not;
(iii) the probability that X_{i1} is greater than X_{i2} is the same for each i; and
(iv) the occurrences of X_{i1} being greater than X_{i2} are independent among the X_i

we know $M \sim \text{BINOMIAL}(n, \pi)$.

[9 marks: Two marks each for (i) the joint pdf of X and (ii) deriving π. One mark each for (i) justifying that the distribution of U is normal, (ii) computing its mean, (iii) computing its variance, (iv) stating the distribution of M, and (v) justifying the distribution of M.]

2. (a) *[Seen]* The likelihood function is $L(\lambda, \xi \mid x, y) = \dfrac{e^{-\lambda}\lambda^x}{x!}\dfrac{e^{-\lambda\xi}(\lambda\xi)^y}{y!}$ and the loglikelihood function is $\ell(\lambda, \xi \mid x, y) = -\lambda(1+\xi) + (x+y)\log(\lambda) + y\log(\xi)$.

To find a candidate value for the maximum likelihood estimate, set

$$\frac{\partial \ell}{\partial \lambda} = -(1+\xi) + \frac{x+y}{\lambda} = 0 \quad \text{and} \quad \frac{\partial \ell}{\partial \xi} = -\lambda + \frac{y}{\xi} = 0.$$

Solving these equations yields candidate estimates $\hat{\lambda}_{\text{MLE}} = x$ and $\hat{\xi}_{\text{MLE}} = y/x$ (since $x > 0$).

To verify that the candidate estimates indeed maximize the log likelihood function, we note

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{x+y}{\lambda^2}, \quad \frac{\partial^2 \ell}{\partial \xi^2} = -\frac{y}{\xi^2}, \quad \text{and} \quad \frac{\partial^2 \ell}{\partial \lambda \cdot \partial \xi} = -1.$$

Evaluating at the candidate estimates,

$$\left.\frac{\partial^2 \ell}{\partial \lambda^2}\right|_{\hat{\lambda},\hat{\xi}} = -\frac{x+y}{x^2} \le 0, \quad \left.\frac{\partial^2 \ell}{\partial \xi^2}\right|_{\hat{\lambda},\hat{\xi}} = -\frac{x^2}{y} \le 0, \quad \text{and} \quad \left.\frac{\partial^2 \ell}{\partial \lambda \cdot \partial \xi}\right|_{\hat{\lambda},\hat{\xi}} = -1.$$

Because the inequalities are strict when $x, y > 0$, this verifies two of the three conditions for the two-dimensional second derivative test. The third is

$$\left[\frac{\partial^2 \ell}{\partial \lambda^2} \cdot \frac{\partial^2 \ell}{\partial \xi^2} - \left(\frac{\partial^2 \ell}{\partial \lambda \cdot \partial \xi}\right)^2\right]_{\lambda=\hat{\lambda},\xi=\hat{\xi}} = \frac{x+y}{y} - 1 = \frac{x}{y} \ge 0.$$

(If $y = 0$, $\partial\ell/\partial\xi$ is everywhere negative and the estimate of ξ is zero.) Thus, the maximum likelihood estimates are $\hat{\lambda}_{\text{MLE}} = x$ and $\hat{\xi}_{\text{MLE}} = x/y$ if $x > 0$. (The estimate of ξ is not defined if $x = 0$.)

[7 marks: One mark each for (i) the likelihood function, (ii) the loglikelihood function, (iii)-(iv) each of the derivatives of the loglikelihood function, (iv) the candidate estimates, (vi), the matrix of partial derivatives (vii) checking the signs of the diagonal terms and the determinant of the matrix of partial derivatives.]

(b) *[Seen]* Setting $\boldsymbol{X} = (X_1, \ldots, X_n)$, the posterior distribution is

$$f_{\Lambda|\boldsymbol{X}}(\lambda) \propto \prod_{i=1}^{n} f_{X_i|\Lambda}(x_i|\lambda) f_\Lambda(\lambda) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda} \propto \lambda^{\sum_{i=1}^n x_i+\alpha-1}e^{-(\beta+n)\lambda},$$

i.e., $\Lambda|X_1, \ldots, X_n \sim \text{GAMMA}(\alpha + \sum_{i=1}^n X_i, \beta + n)$.

The posterior expectation is $\hat{\lambda}_1 = (\alpha + \sum_{i=1}^n x_i)/(\beta + n)$. The corresponding estimator is $\widehat{\Lambda}_1 = (\alpha + \sum_{i=1}^n X_i)/(\beta + n)$.

The marginal moments of X_i can be derived as *[Seen]*

$$\text{E}(X_i) = \text{E}\Big[\text{E}(X_i|\Lambda)\Big] = E(\Lambda) = \alpha/\beta,$$

$$\text{Var}(X_i) = \text{E}\Big[\text{Var}(X_i|\Lambda)\Big] + \text{Var}\Big[\text{E}(X_i|\Lambda)\Big] = E(\Lambda) + \text{Var}(\Lambda) = \alpha(\beta+1)/\beta^2,$$

[Unseen]

$$\mathrm{E}(X_iX_j) = \mathrm{E}\Big[\mathrm{E}(X_iX_j|\Lambda)\Big] = \mathrm{E}(\Lambda^2) = \mathrm{Var}(\Lambda) + [\mathrm{E}(\Lambda)]^2 = \alpha(\alpha+1)/\beta^2.$$

and

$$\mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_iX_j)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}} = \frac{\mathrm{E}(X_iX_j) - \mathrm{E}(X_i)\mathrm{E}(X_j)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}} = \frac{\alpha(\alpha+1)-\alpha^2}{\alpha(\beta+1)} = \frac{1}{\beta+1}.$$

[7 marks: One mark each for (i) the general formula for a posterior distribution, (ii) the correct gamma posterior distribution, (iii) the posterior expectation, (iv) $\mathrm{E}(X_i)$*, (v)* $\mathrm{Var}(X_i)$*, (vi)* $\mathrm{E}(X_iX_j)$*, and (vii)* $\mathrm{Corr}(X_i, X_j)$*.]*

(c) *[Seen Similar]* The mean square error of $\widehat{\Lambda}_2$ can be computed as

$$\mathrm{mse}(\widehat{\Lambda}_2) = \Big(\mathrm{bias}(\widehat{\Lambda}_2)\Big)^2 + \mathrm{Var}(\widehat{\Lambda}_2) = \Big(\mathrm{E}(\widehat{\Lambda}_2) - \lambda\Big)^2 + \mathrm{Var}(\widehat{\Lambda}_2) = \mathrm{Var}(\widehat{\Lambda}_2) = \lambda/n.$$

[Unseen] Setting $\alpha = 0$ and $\beta = 1$, the mean square error of $\widehat{\Lambda}_1$ can be computed as

$$\mathrm{mse}(\widehat{\Lambda}_1) = \Big(\mathrm{E}(\widehat{\Lambda}_1) - \lambda\Big)^2 + \mathrm{Var}(\widehat{\Lambda}_1) = \left(\frac{n\lambda}{n+1} - \lambda\right)^2 + \mathrm{Var}\left(\frac{\sum_{i=1}^n X_i}{n+1}\right) = \frac{(\lambda^2+n\lambda)}{(n+1)^2} = \lambda\frac{(n+\lambda)}{(n+1)^2}.$$

[Unseen] For $\mathrm{mse}(\widehat{\Lambda}_1) < \mathrm{mse}(\widehat{\Lambda}_2)$, we need $(n+1)^2/(n+\lambda) > n$, i.e., $\lambda < 2 + \frac{1}{n}$.

[6 marks: 2 marks each for (i) mse of $\widehat{\Lambda}_2$*, (ii)* $\widehat{\Lambda}_1$*, and (iii) range of* λ *with* $\mathrm{mse}(\widehat{\Lambda}_1) < \mathrm{mse}(\widehat{\Lambda}_2)$*.]*

3. (a) *[Seen Similar]* Inverting the transformation, $Y_1 = -\log(U_1^\lambda)$, yields $U_1 = \exp(-Y_1/\lambda)$ and the Jacobian, $J = -\frac{1}{\lambda}\exp(-y/\lambda)$. Thus the density of Y_1 is

$$f_{Y_1}(y) = f_{U_1}(e^{-y/\lambda})|J| = \frac{1}{\lambda}\exp(-y/\lambda) \quad \text{for} \quad 0 < y < +\infty.$$

[Seen] Recognizing this probability density function, $Y_1 \sim \textsc{Exponential}(1/\lambda)$. Since the U_i are independent and identically distributed, we have $-\log(U_i^\lambda) \overset{\text{iid}}{\sim} \textsc{Exponential}(1/\lambda)$ and thus $Y_n = -\sum_{i=1}^n \log(U_i^\lambda) \sim \textsc{Gamma}(n, 1/\lambda)$.

[Seen Similar] For the normal approximation we use the fact that $Y_n = \sum_{i=1}^n V_i$ where $V_i \sim \textsc{Exponential}(1/\lambda)$ and apply the Central Limit Theorem with $\mu = \mathrm{E}(V_i) = \lambda$ and $\sigma^2 = \mathrm{Var}(V_i) = \lambda^2$. This yields,

$$Z_n = \frac{\sum_{i=1}^n V_i - n\lambda}{\lambda\sqrt{n}} \xrightarrow{\mathcal{D}} \mathrm{Norm}(0,1) \;\text{ and hence for large } n,\; Y_n \overset{\text{approx}}{\sim} \mathrm{Norm}\left(n\lambda, n\lambda^2\right).$$

[7 marks: Two marks for the change of variable calculations and one mark each for (i) the final pdf of Y_1*, (ii) stating the distribution of* Y_n*, (iii) justification for stated distribution of* Y_n*, (iv) correctly applying the central limit theorem, and (v) the final normal approximation for* Y_n*.]*

(b) *[Seen Similar]* We know $X_n = \exp(-Y_n)$ and that $Y_n \sim \textsc{Gamma}(n, 1/\lambda)$. Inverting this transformation, we have $Y_n = -\log(X_n)$ and Jacobian, $J = -1/x$ so that

$$f_{X_n}(x) = f_{Y_n}(-\log x)|J| = \frac{1}{\lambda^n\Gamma(n)}(-\log x)^{n-1}x^{\frac{1}{\lambda}-1} \;\text{ for }\; 0 < x < 1,$$

and *[Seen Similar]*

$$\begin{aligned}\mathrm{E}(X_n) = \int_0^1 v f_V(v)\mathrm{d}v &= -\int_0^1 \frac{1}{\lambda^n\Gamma(n)}(-\log x)^{n-1}x^{(\frac{1}{\lambda}+1)-1}\mathrm{d}x \\ &= \frac{(\frac{1}{\lambda}+1)^{-n}}{\lambda^n}\int_0^1 \frac{(\frac{1}{\lambda}+1)^n}{\Gamma(n)}(-\log x)^{n-1}x^{(\frac{1}{\lambda}+1)-1}\mathrm{d}x = \frac{1}{(\lambda+1)^n},\end{aligned}$$

where the second integral is one because it is a density (of the form as $f_{X_n}(x)$) integrated over its support.

[6 marks: Two marks for the change of variable calculations. One mark each for (i) the final pdf of X_n*, (ii) setting up the integral for* $\mathrm{E}(X_n)$ *including the range of integration, (iii) recognizing the pdf in the integrand, and (iv) the final expression for* $\mathrm{E}(X_n)$*.]*

(c) *[Unseen]* By construction $X_n \leq \min_{i=1}^n U_i^\lambda$. Thus, $U_i^\lambda \leq x$ for any i in $(1, 2, \ldots, n)$ implies that $X_n \leq x$ and

$$\Pr(X_n \leq x) \geq \Pr\left(\bigcup_{i=1}^n \{U_i^\lambda \leq x\}\right) = 1 - \Pr\left(\bigcap_{i=1}^n \{U_i \geq \sqrt[\lambda]{x}\}\right) = 1 - \left(1 - \sqrt[\lambda]{x}\right)^n,$$

where the inequality follows from a basic property of probability (i.e., if A implies B, then $\Pr(A) \leq \Pr(B)$), the first equality follows from De Morgan's Law, and the second equality holds for any $0 \leq x \leq 1$.

[Unseen] For any n and $0 \leq x \leq 1$, we have $1 - \left(1 - \sqrt[\lambda]{x}\right)^n \leq \Pr(X_n \leq x) \leq 1$. Taking the limit as $n \to \infty$ yields $\lim_{n\to\infty} \Pr(X_n \leq x) = 1$ for $0 \leq x \leq 1$. Whereas $\Pr(X_n \leq x) = 0$ for $x < 0$ and $\Pr(X_n \leq x) = 1$ for any $x > 1$ we have that $X_n \xrightarrow{\mathcal{D}} 0$ (i.e., X_n is degenerate at zero).

[7 marks: Two marks for the general outline of the first proof; one mark for the detailed justification for the first proof; two marks for convergence of $\Pr(X_n \leq x)$ *for* $0 \leq x \leq 1$*; one mark for convergence for other* x*; and one mark for establishing convergence in distribution.]*

4. (a) *[Seen]* $X \sim$ BINOMIAL(n, θ); *[Seen Similar]* $Y \sim$ BINOMIAL$(n, \theta\pi)$; and *[Unseen]* $Y \mid X = x \sim$ BINOMIAL(x, π).

[Seen Similar] The joint probability mass function is given by

$$f_{XY}(x,y) = \binom{n}{x}\theta^x(1-\theta)^{n-x}\binom{x}{y}\pi^y(1-\pi)^{x-y} = \frac{n!}{(n-x)!(x-y)!y!}(1-\theta)^{n-x}(\theta-\theta\pi)^{x-y}(\theta\pi)^y$$

for integer x and y such that $0 \leq y \leq x \leq n$.

[Seen Similar] X and Y are not independent because their joint support is not the cross product of their marginal supports and their joint mass function does not factor.

[7 marks: One mark each for (i) the marginal distribution of X*, (ii) the marginal distribution of* Y*, (iii) the conditional distribution of* Y *given* X*, (iv) the joint mass function, (v) its support, (vi) noting that* X *and* Y *are not independent, and (vii) justification this.]*

(b) *[Seen Similar]* Since $f_{X|Y}(x|y) = f_{XY}(x,y)/f_Y(y)$, the conditional probability mass function can be written

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{n!}{(n-x)!(x-y)!y!}\frac{y!(n-y)!}{n!}\frac{(1-\theta)^{n-x}(\theta-\theta\pi)^{x-y}(\theta\pi)^y}{(\theta\pi)^y(1-\theta\pi)^{n-y}} \\ &= \frac{(n-y)!}{(n-x)!(x-y)!}\left(\frac{1-\theta}{1-\theta\pi}\right)^{n-x}\left(\frac{\theta-\theta\pi}{1-\theta\pi}\right)^{x-y} \quad \text{for } y \leq x \leq n, \end{aligned}$$

[Seen Similar] Alternatively, since $f_{X|Y}(x|y) \propto f_{XY}(x,y)$, the conditional mass function can be written

$$f_{X|Y}(x|y) = \frac{k}{(n-x)!(x-y)!}(1-\theta)^{n-x}(\theta-\theta\pi)^{x-y} \quad \text{for } y \leq x \leq n,$$

for some positive k. Letting $z = x - y$,

$$\begin{aligned} \frac{1}{k} &= \sum_{z=0}^{n-y}\frac{1}{(n-y-z)!z!}(1-\theta)^{n-y-z}(\theta-\theta\pi)^z \\ &= \frac{(1-\theta\pi)^{n-y}}{(n-y)!}\sum_{z=0}^{n-y}\frac{(n-y)!}{(n-y-z)!z!}\left(\frac{1-\theta}{1-\theta\pi}\right)^{n-y-z}\left(\frac{\theta-\theta\pi}{1-\theta\pi}\right)^z = \frac{(1-\theta\pi)^{n-y}}{(n-y)!} \end{aligned}$$

and

$$f_{X|Y}(x|y) = \frac{(n-y)!}{(n-x)!(x-y)!}\left(\frac{1-\theta}{1-\theta\pi}\right)^{n-x}\left(\frac{\theta-\theta\pi}{1-\theta\pi}\right)^{x-y} \quad \text{for } y \leq x \leq n.$$

[Unseen] Noting that $X|Y \sim Y +$ BINOMIAL$\left(n - Y, \dfrac{\theta-\theta\pi}{1-\theta\pi}\right)$, we have $\mathrm{E}(X|Y) = Y + \dfrac{(n-Y)(\theta-\theta\pi)}{1-\theta\pi}$.

[Seen Similar] Alternatively, the conditional expectation may be computed directly, again letting $z = x - y$.

$$\begin{aligned}
\mathrm{E}(X|Y=y) &= \sum_{z=0}^{n-y} \frac{(n-y)!(z+y)}{(n-y-z)!z!}\left(\frac{1-\theta}{1-\theta\pi}\right)^{n-y-z}\left(\frac{\theta-\theta\pi}{1-\theta\pi}\right)^{z} \\
&= y\sum_{z=0}^{n-y} \frac{(n-y)!}{(n-y-z)!z!}\left(\frac{1-\theta}{1-\theta\pi}\right)^{n-y-z}\left(\frac{\theta-\theta\pi}{1-\theta\pi}\right)^{z} + \sum_{z=0}^{n-y} \frac{(n-y)!z}{(n-y-z)!z!}\left(\frac{1-\theta}{1-\theta\pi}\right)^{n-y-z}\left(\frac{\theta-\theta\pi}{1-\theta\pi}\right)^{z} \\
&= y + \frac{(n-y)(\theta-\theta\pi)}{1-\theta\pi}.
\end{aligned}$$

[6 marks: One mark each for (i) knowing a method for computing a conditional distribution, (ii) the correctly normalized probability mass function, (iii) the support of the mass function, and (iv) correct formulae for the conditional expectation. Two marks for deriving the conditional expectation.]

(c) *[Unseen]* From the marginal distributions of X and Y, we know $\mathrm{E}(X) = n\theta$ and $\mathrm{E}(Y) = n\theta\pi$, so we can set up a system of two equations: (i) $x = n\theta$ and (ii) $y = n\theta\pi$. Solving yields the method of moments estimates, $\hat\theta = x/n$ and $\hat\pi = y/x$; although $\hat\theta$ is always defined, $\hat\pi$ only exists if $x > 0$. Hence the method of moments estimators are $\widehat{\Theta}_{\mathrm{MoM}} = X/n$ and $\widehat{\Pi}_{\mathrm{MoM}} = Y/X$, if $X > 0$.

[Seen Similar] The estimator, $\widehat{\Theta}_{\mathrm{MoM}}$ is unbiased for θ because $\mathrm{E}(\widehat{\Theta}_{\mathrm{MoM}}) = \mathrm{E}(X/n) = n\theta/n = \theta$.

[Unseen] The expectation of $\widehat{\Pi}$ is

$$\begin{aligned}
\mathrm{E}(\widehat{\Pi}) = \mathrm{E}\Big[\mathrm{E}(\widehat{\Pi}|X)\Big] &= \mathrm{E}\left[\mathrm{E}\left(0I\{X=0\} + \frac{Y}{X}I\{X>0\} \;\Big|\; X\right)\right] = \mathrm{E}\left(\frac{X\pi}{X}I\{X>0\}\right) \\
&= \pi\Big[\Pr(X>0)\Big] < \pi.
\end{aligned}$$

Thus $\widehat{\Pi}$ is a biased estimator of π.

[7 marks: One mark each for (i) the system of equations, (ii) the method of moments estimators, (iii) noting that the estimate (or estimator) of π does not exist if $x = 0$ (or $X = 0$), (iv) arguing that $\widehat{\Theta}_{\mathrm{MoM}}$ is unbiased, and (v) stating that $\widehat{\Pi}$ is biased. Two marks for the expectation of $\widehat{\Pi}$.]