BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2016

This paper is also taken for the relevant examination for the Associateship.

M2S1

PROBABILITY AND STATISTICS II

Date: date May 2016 Time: time

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used.

Formula sheets are included on pages 5 and 6.

1. Suppose (Z_1, Z_2) is a bivariate normal random variable such that $\mathrm{E}(Z_1) = \mathrm{E}(Z_2) = 0$, $\mathrm{Var}(Z_1) = \mathrm{Var}(Z_2) = 1$ and $\mathrm{Correlation}(Z_1, Z_2) = 0$.

Recall that a χ^2_ν distribution (i.e, a chi squared distribution with ν degrees of freedom) is equivalent to a $\textsc{Gamma}\left(\frac{\nu}{2}, \frac{1}{2}\right)$ distribution.

(a) Are Z_1 and Z_2 independent? Justify your answer using the definition of independence.

(b) Identify the distributions of the following random variables.

i)
$$X_1 = Z_1^2$$

ii)
$$X_2 = Z_1^2 + Z_2^2$$

iii)
$$X_3 = \lambda(Z_1^2 + Z_2^2), \text{ where } \lambda \text{ is a positive constant}$$

iv)
$$X_4 = \begin{cases} 1 & \text{if } Z_1 > 1 \\ 0 & \text{otherwise} \end{cases}$$

(c) Identify the joint distribution of $X_5 = Z_1$ and $X_6 = 2Z_1 - Z_2$.
Are X_5 and X_6 independent? Justify your answer.

(d) Using the distribution of X_3, derive a pivot and a 95% equal-tailed confidence interval for λ. Hint: Your formula for the confidence interval should involve notation for the quantiles of an appropriate distribution.

(e) Consider the random variable

$$X_7 \sim \begin{cases} \textsc{Gamma}(\alpha_1, \beta) & \text{if } X_4 = 1 \\ \textsc{Gamma}(\alpha_2, \beta) & \text{otherwise} \end{cases},$$

where α_1, α_2, and β are positive parameters. Using the parameterization of the gamma distribution given in the formula sheet, derive $\mathrm{E}(X_7)$ and $\mathrm{Var}(X_7)$.

2. (a) State the definition of convergence in distribution, $V_n \xrightarrow{\mathcal{D}} V$.

State the definition of convergence in probability, $V_n \xrightarrow{\mathcal{P}} V$.

Suppose $(Z_1, \ldots, Z_n)$ is a random sample from a distribution with CDF $F_Z(z)$.

Show that the CDF of $U_n = \min(Z_1, \ldots, Z_n)$ is given by $F_{U_n}(u) = 1 - (1 - F_Z(u))^n$.

For the remainder of this problem suppose $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ are random samples from the distributions with probability density functions given by

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda} \text{ for } x \geq 0 \qquad \text{and} \quad f_Y(y) = \frac{1}{\xi} e^{-y/\xi} \text{ for } y \geq 0,$$

respectively, where λ and ξ are positive. (These are EXPONENTIAL distributions, but their parameters are the reciprocal of what is given in the attached formula sheet.) Also let $M_n = \min(X_1, \ldots, X_n)$ and assume that $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ are independent.

(b) Derive the maximum likelihood estimator $\widehat{\Lambda}_n$ of λ.

State the sampling distribution of $\widehat{\Lambda}_n$. Justify your answer using results from the notes.

(c) Derive the (named) sampling distribution of M_n and its parameter(s).

Show that the limiting distribution of M_n is degenerate at 0.

(d) Let $\widehat{\Xi}_n$ be the maximum likelihood estimator of ξ. The formula and sampling distribution of $\widehat{\Xi}_n$ follow directly from part (b), with $(X_1, \ldots, X_n)$ replaced with $(Y_1, \ldots, Y_n)$. Using a theorem from the notes derive the asymptotic (large n) distributions of

i) $\widehat{\Lambda}_n$

ii) $\widehat{\Lambda}_n - \widehat{\Xi}_n$.

3. Suppose that $(X_1, \ldots, X_n)$ is a random sample from a normal distribution with mean μ and variance σ^2. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, and $T_n = (n-1)S_n^2/\sigma^2$. Recall that $T_n \sim \chi^2_{n-1} = \text{GAMMA}\left(\frac{n-1}{2}, \frac{1}{2}\right)$.

(a) Starting with the known distribution of T_n, derive the probability density function of S_n^2. What named distribution does S_n^2 follow? What are its parameter(s)?

(b) Using your answer from part (a), derive the bias and mean square error of S_n^2 as an estimator of σ^2.

(c) Show that S_n^2 is a consistent estimator of σ^2. That is, show that $S_n^2 \xrightarrow{\mathcal{P}} \sigma^2$ as $n \to \infty$.

(d) Let $S_n = \sqrt{S_n^2}$ and derive $\mathrm{E}(S_n)$.

Give an expression for the bias of S_n as an estimator of σ and evaluate your expression for $n = 2$.

(e) Show that $\mathrm{E}(S_n) \leq \sigma$ for any n.

.

4. Consider the probability density function

$$f_Y(y) = -\alpha^2 \, y^{\alpha-1} \log(y) \quad \text{for} \quad 0 < y < 1, \tag{1}$$

where α is a positive parameter.

(a) Derive $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$.

(b) Suppose $(Y_1, \ldots, Y_n)$ is a random sample, such that each Y_i has the probability density function given in Equation (1). Based on this random sample, derive the method of moments estimator of α.

Now suppose that $X_1 \sim \mathrm{B{\small ETA}}(\alpha, 1)$, that $X_2 \sim \mathrm{B{\small ETA}}(\alpha, 1)$, and that X_1 and X_2 are independent. Let $U = X_1$ and $V = X_1 X_2$.

(c) Derive the joint probability density function of U and V.
Are U and V independent? Justify your answer.

(d) Derive the marginal probability density function of V.
Using your answer to part (a), derive the Covariance of U and V.

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	$\theta(1-\theta)$	$1-\theta+\theta e^t$
$Binomial(n,\theta)$	$\{0,1,...,n\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$		$n\theta$	$n\theta(1-\theta)$	$\left(1-\theta+\theta e^t\right)^n$
$Poisson(\lambda)$	$\{0,1,2,...\}$	$\lambda \in \mathbb{R}^+$	$\dfrac{e^{-\lambda}\lambda^x}{x!}$		λ	λ	$\exp\left\{\lambda\left(e^t-1\right)\right\}$
$Geometric(\theta)$	$\{1,2,...\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\dfrac{1}{\theta}$	$\dfrac{(1-\theta)}{\theta^2}$	$\dfrac{\theta e^t}{1-e^t(1-\theta)}$
$NegBinomial(n,\theta)$	$\{n,n+1,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$		$\dfrac{n}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta e^t}{1-e^t(1-\theta)}\right)^n$
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 *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*	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 — May 2016 Exam — Solution

1. (a) The joint PDF of Z_1 and Z_2 is given by

$$\begin{aligned} f_{Z_1,Z_2}(z_1,z_2) &= \frac{1}{2\pi\sqrt{(1-\rho^2)}}\exp\left\{-\frac{z_1^2-2\rho z_1 z_2+z_2^2}{2(1-\rho^2)}\right\} \quad \text{for} \quad -\infty < z_1, z_2 < \infty \\ &= \frac{1}{2\pi}\exp\left\{-\frac{z_1^2+z_2^2}{2}\right\} \\ &= f_{Z_1}(z_1)f_{Z_2}(z_2). \end{aligned}$$

Since the joint PDF factors and the joint support is the cross product of the individual supports, Z_1 and Z_2 are independent random variables.

(b) The distributions are:

(i) The square of standard normal random variables is χ^2_1, thus $X_1 \sim \chi^2_1$.

(ii) The sum of independent χ^2 random variables is also χ^2, with the degrees of freedom summing, thus $X_2 \sim \chi^2_2$.

(iii) Since χ^2_2 is equivalent to $\textsc{Gamma}\left(1, \frac{1}{2}\right)$, we have $Z_1^2 + Z_2^2 \sim \textsc{Gamma}\left(1, \frac{1}{2}\right)$ and using the scale parameter of the gamma distribution, $X_3 = \lambda(Z_2^2 + Z_3^2) \sim \textsc{Gamma}\left(1, \frac{1}{2\lambda}\right)$.

(iv) $X_4 \sim \textsc{Bernoulli}(\pi)$, where $\pi = \Pr(Z_1 > 1) = 0.16$.

(c) We can write

$$\begin{pmatrix} X_5 \\ X_6 \end{pmatrix} = \boldsymbol{M}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \quad \text{where} \quad \boldsymbol{M} = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}$$

Thus,

$$\begin{pmatrix} X_5 \\ X_6 \end{pmatrix} \sim N_2\left(\boldsymbol{M0}, \boldsymbol{MIM}^\top\right),$$

where $\mathbf{0}$ is column vector of zeros and $\boldsymbol{I}$ is a the identity matrix, i.e.,

$$\begin{pmatrix} X_5 \\ X_6 \end{pmatrix} \sim N_2\left(\mathbf{0}, \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}\right).$$

If X_5 and X_6 were independent there correlation would be zero. Since $\mathrm{Corr}(X_5, X_6) = 2/\sqrt{5} > 0$ they are not independent.

(d) The random variable $X_2 = X_3/\lambda \sim \textsc{Gamma}\left(1, \frac{1}{2}\right)$ is pivot for λ because (i) its distribution depends on no unknown parameters and (ii) it depends on λ. We can derive a 95% equal-tailed confidence interval by setting $g_{0.025}$ and $g_{0.975}$ so that $F_{X_2}(g_{0.025}) = 0.025$ and $F_{X_2}(g_{0.975}) = 0.975$ so that

$$\Pr\left(g_{0.025} \le \frac{X_3}{\lambda} \le g_{0.975}\right) = 95\%,$$

so that

$$\Pr\left(\frac{1}{g_{0.975}} \le \frac{\lambda}{X_3} \le \frac{1}{g_{0.025}}\right) = 95\%,$$

and

$$\Pr\left(\frac{X_3}{g_{0.975}} \le \lambda \le \frac{X_3}{g_{0.025}}\right) = 95\%,$$

i.e., $(X_3/g_{0.975}, X_3/g_{0.025})$ is a 95% equal-tailed CI for λ.

(e) $\mathrm{E}(X_7) = \mathrm{E}\left[\mathrm{E}(X_7|X_4)\right] = \mathrm{E}\left(\frac{\alpha_1}{\beta}X_4 + \frac{\alpha_2}{\beta}(1-X_4)\right) = \frac{0.16\alpha_1+0.84\alpha_2}{\beta}$.

$\mathrm{Var}(X_7) = \mathrm{E}\left[\mathrm{Var}(X_7|X_4)\right] + \mathrm{Var}\left[\mathrm{E}(X_7|X_4)\right] = \mathrm{E}\left(\frac{\alpha_1}{\beta^2}X_4 + \frac{\alpha_2}{\beta^2}(1-X_4)\right) + \mathrm{Var}\left(\frac{\alpha_1}{\beta}X_4 + \frac{\alpha_2}{\beta}(1-X_4)\right)$

$= \frac{0.16\alpha_1+0.84\alpha_2}{\beta^2} + \mathrm{Var}\left(\frac{\alpha_1-\alpha_2}{\beta}X_4 + \frac{\alpha_2}{\beta}\right) = \frac{0.84\alpha_1+0.16\alpha_2}{\beta^2} + (0.16)(0.84)\frac{(\alpha_1-\alpha_2)^2}{\beta^2}.$

2. (a) We say $V_n \xrightarrow{\mathcal{D}} V$ if $\lim_{n\to\infty} F_{V_n}(v) = F_V(v)$ at all points of continuity of $F_V(v)$.

We say $V_n \xrightarrow{\mathcal{P}} V$ if for all $\epsilon > 0$, $\lim_{n\to\infty} \Pr(|V_n - V| < \epsilon) = 1$ (or equivalently, $\lim_{n\to\infty} \Pr(|V_n - V| \ge \epsilon) = 0$).

From first principles,

$$\begin{aligned}\Pr(U_n > u) &= \Pr\Big(\min\{Z_1,\dots,Z_n\} > u\Big) = \Pr\Big(Z_1 > u,\dots,Z_n > u\Big)\\ &= \prod_{i=1}^n \Pr(Z_i > u) = \prod_{i=1}^n \{1 - F_Z(u)\} = \{1-F_Z(u)\}^n,\end{aligned}$$

so $F_{U_n}(u) = 1 - \Pr(U_n > u) = 1 - \{1 - F_Z(u)\}^n$.

(b) The Likelihood function is given by

$$L(\lambda|x) = \frac{1}{\lambda^n}\exp\left\{-\frac{1}{\lambda}\sum_{i=1}^n x_i\right\}, \quad \text{for} \quad \lambda > 0$$

where $x = (x_1,\dots,x_n)$ and thus the loglikelihood is given by

$$\ell(\lambda|x) = -n\log(\lambda) - \frac{1}{\lambda}\sum_{i=1}^n x_i, \quad \text{for} \quad \lambda > 0.$$

Taking the first derivative,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\ell(\lambda|x) = -\frac{n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^n x_i = 0,$$

and solving over positive λ gives the preliminary estimate, $\hat\lambda_n = \frac{1}{n}\sum_{i=1}^n x_i$. Finally, the second derivate test yields

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}\lambda}^2 \ell(\lambda|x)\right|_{\lambda=\hat\lambda} = \left.\frac{n}{\lambda^2} - \frac{2}{\lambda^3}\sum_{i=1}^n x_i\right|_{\lambda=\hat\lambda} = -\frac{n}{\hat\lambda_n^2} < 0,$$

so that the preliminary estimate does indeed maximize the likelihood function, and the maximum likelihood estimator is $\hat\Lambda_n = \frac{1}{n}\sum_{i=1}^n X_i$.

To derive the sampling distribution of $\hat\Lambda_n$, note that $X_i \overset{\text{iid}}{\sim} \textsc{Gamma}(1, 1/\lambda)$ so that $\sum_{i=1}^n X_i \overset{\text{iid}}{\sim} \textsc{Gamma}(n, 1/\lambda)$ and $\hat\Lambda_n = \frac{1}{n}\sum_{i=1}^n X_i \overset{\text{iid}}{\sim} \textsc{Gamma}(n, n/\lambda)$.

(c) By part (a), the CDF of M_n is $F_{M_n}(m) = 1 - (1 - F_X(m))^n$, where $F_X(m) = \int_0^m f_X(x)\mathrm{d}x = 1 - e^{m/\lambda}$. Thus $F_{M_n}(m) = 1 - e^{mn/\lambda}$ for $m \ge 0$ and $F_{M_n}(m) = 0$ for $m < 0$ and $M_n \sim \textsc{Exponential}(n/\lambda)$.

Because

$$\lim_{n\to\infty} F_{M_n}(m) = \begin{cases} 1 & \text{if } m > 0\\ 0 & \text{if } m < 0\end{cases}$$

$M_n \xrightarrow{\mathcal{D}} 0$, i.e., the distribution of M_n is degenerate at zero.

(d) By the central limit theorem,

$$\frac{\frac{1}{n}\sum_{i=1}^n X_i - \mathrm{E}(X_i)}{\sqrt{\mathrm{Var}(X_i)/n}} = \frac{\hat\Lambda_n - \lambda}{\sqrt{\lambda^2/n}} \xrightarrow{\mathcal{D}} Z \sim N(0,1)$$

and so $\hat\Lambda_n \overset{\text{approx}}{\sim} N(\lambda, \lambda^2/n)$.

Similarly, $\hat\Xi_n \overset{\text{approx}}{\sim} N(\xi, \xi^2/n)$ and $\hat\Lambda_n$ and $\hat\Xi_n$ are independent. Thus, because the sum of independent normal random variables is normal, $(\hat\Lambda_n - \hat\Xi_n) \overset{\text{approx}}{\sim} N\Big(\lambda - \xi, (\lambda^2 + \xi^2)/n\Big)$.

3. (a) The Jacobian of the transformation $S_n^2 = \sigma^2 T_n/(n-1)$ is $|J| = (n-1)/\sigma^2$ so that

$$\begin{aligned}f_{S_n^2}(s) = f_{T_n}\left(\frac{(n-1)s}{\sigma^2}\right)|J| &= \frac{n-1}{\sigma^2 2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{(n-1)s}{\sigma^2}\right)^{\frac{n-1}{2}-1}\exp\left(-\frac{(n-1)s}{2\sigma^2}\right)\\ &= \left(\frac{n-1}{2\sigma^2}\right)^{(n-1)/2}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)}\, s^{\frac{n-1}{2}-1}\exp\left(-\frac{(n-1)s}{2\sigma^2}\right),\end{aligned}$$

for $s > 0$. That is, $S_n^2 \sim \textsc{Gamma}\left(\frac{n-1}{2}, \frac{n-1}{2\sigma^2}\right)$.

(A simpler derivation of the PDF of S_n^2 involves identifying the scale parameter of the gamma distribution and adjusting it appropriately.)

(b) As an estimator of σ^2, $\text{bias}(S_n^2) = \text{E}(S_n^2) - \sigma^2$. From part (a), $\text{E}(S_n) = \frac{n-1}{2} / \frac{n-1}{2\sigma^2} = \sigma^2$, so that $\text{bias}(S_n^2) = \sigma^2 - \sigma^2 = 0$.

As an estimator of σ^2,

$$\text{MSE}(S_n^2) = \text{Var}(S_n^2) + \Big(\text{bias}(S_n^2)\Big)^2 = \text{Var}(S_n^2) = \frac{n-1}{2}\left(\frac{n-1}{2\sigma^2}\right)^{-2} = \frac{2\sigma^4}{n-1}.$$

(c) By Chebychev's inequality, for any non-negative function g and $r > 0$, $\Pr\Big(g(S_n^2) \geq r\Big) \leq \frac{1}{r}\text{E}\Big(g(S_n^2)\Big)$. Setting $g(S_n^2) = (S_n^2 - \sigma^2)^2 \geq 0$, we have

$$\Pr\Big((S_n^2 - \sigma^2)^2 \geq t^2\Big) \leq \frac{1}{t^2}\text{Var}(S_n^2) = \frac{2\sigma^4}{t^2(n-1)}$$

and

$$\lim_{n\to\infty} \Pr\Big(|S_n^2 - \sigma^2| \geq t\Big) \leq \lim_{n\to\infty} \frac{2\sigma^4}{t^2(n-1)} = 0,$$

for all $t > 0$, i.e., $S_n^2 \xrightarrow{\mathcal{P}} \sigma^2$.

(d)

$$\begin{aligned}
\text{E}(S_n) &= \int_0^\infty \sqrt{s} f_{S_n^2}(s)\text{d}s = \int_0^\infty \sqrt{s}\left(\frac{n-1}{2\sigma^2}\right)^{(n-1)/2} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} s^{\frac{n-1}{2}-1} \exp\left(-\frac{(n-1)s}{2\sigma^2}\right)\text{d}s \\
&= \left(\frac{n-1}{2\sigma^2}\right)^{-1/2} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty \left(\frac{n-1}{2\sigma^2}\right)^{n/2} \frac{1}{\Gamma\left(\frac{n}{2}\right)} s^{\frac{n}{2}-1} \exp\left(-\frac{(n-1)s}{2\sigma^2}\right)\text{d}s \\
&= \sigma\sqrt{\frac{2}{n-1}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.
\end{aligned}$$

Thus, the bias of S_n as an estimator or σ is

$$\text{bias}(S_n) = \text{E}(S_n) - \sigma = \sigma\left(\sqrt{\frac{2}{n-1}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} - 1\right),$$

which for $n = 2$ is $\text{bias}(S_2) = \sigma\left(\sqrt{\frac{2}{\pi}} - 1\right) \approx \sigma/5$.

(e) This can be shown using Jensen's inequality. Let $g(x) = \sqrt{x}$ for $x > 0$, so that $\frac{\text{d}^2}{\text{d}x^2}g(x) = -x^{-3/2}/4 < 0$ and g is a concave function. Thus, by Jensen's inequality,

$$\text{E}(S) = \text{E}\left(g(S_n^2)\right) \leq g\left(\text{E}(S_n^2)\right) = g(\sigma^2) = \sigma.$$

4. (a)

$$\begin{aligned}
\text{E}(Y) = \int_0^1 y f_Y(y)\text{d}y &= -\int_0^1 \alpha^2\, y^{(\alpha+1)-1} \log(y)\,\text{d}y \\
&= \frac{\alpha^2}{(\alpha+1)^2} \int_0^1 -(\alpha+1)^2 y^{(\alpha+1)-1} \log(y)\,\text{d}y = \left(1 + \frac{1}{\alpha}\right)^{-2},
\end{aligned}$$

where the second integral is one because it is a density (of the form as $f_Y(y)$ with α replaced with $\alpha + 1$) integrated over its support.

Now, $\text{Var}(Y) = \text{E}(Y^2) - [\text{E}(Y)]^2$ and

$$\text{E}(Y^2) = -\int_0^1 \alpha^2\, y^{(\alpha+2)-1} \log(y)\text{d}y = \left(\frac{\alpha}{\alpha+2}\right)^2 \int_0^1 -(\alpha+2)^2\, y^{(\alpha+2)-1} \log(y)\text{d}y = \left(1 + \frac{2}{\alpha}\right)^{-2}.$$

Finally, $\text{Var}(Y) = \text{E}(Y^2) - [E(Y)]^2 = \left(1 + \frac{2}{\alpha}\right)^{-2} - \left(1 + \frac{1}{\alpha}\right)^{-4}$.

(b) Setting $\frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y} = \mathrm{E}(Y) = \left(1+\frac{1}{\alpha}\right)^{-2}$ and solving for α,

$$\begin{aligned}
\left(1+\frac{1}{\alpha}\right)\sqrt{\bar{Y}} &= 1 \\
\frac{\sqrt{\bar{Y}}}{\alpha} &= 1-\sqrt{\bar{Y}} \\
\frac{\sqrt{\bar{Y}}}{1-\sqrt{\bar{Y}}} &= \alpha
\end{aligned}$$

gives $\hat{\alpha}_{\mathrm{MoM}} = \sqrt{\bar{Y}}/(1-\sqrt{\bar{Y}})$.

(c) First we note the joint support of U and V is define by $\Pr(0 < V < U < 1) = 1$. Second we can invert the bivariate transformation with $X_1 = U$ and $X_2 = V/U$, which has Jacobian

$$J = \begin{vmatrix} 1 & 0 \\ - & 1/u \end{vmatrix} = 1/u$$

. Thus, the joint PDF of U and V is

$$f_{UV}(u,v) = f_{X_1X_2}(u, v/u)|J| = \alpha^2 u^{\alpha-1}\left(\frac{v}{u}\right)^{\alpha-1}\frac{1}{u} = \alpha^2\,\frac{v^{\alpha-1}}{u} \quad \text{for} \quad 0 < v < u < 1.$$

Although the joint PDF of U and V appears to factor, the random variables are not independent because their joint support is not the cross product of their marginal supports. (The marginal supports of both U and V are both the unit interval, but $\Pr(V \le U) = 1$.)

(d) The marginal PDF of V is

$$f_V(v) = \int_v^1 f_{UV}(u,v)\mathrm{d}u = \int_v^1 \alpha^2\,\frac{v^{\alpha-1}}{u}\mathrm{d}u = \alpha^2 v^{\alpha-1}\Big[\log(u)\Big]_v^1 = -\alpha^2 v^{\alpha-1}\log(v) \ \text{ for } 0 < v < 1.$$

$\mathrm{Cov}(U,V) = \mathrm{E}(UV) - \mathrm{E}(U)\mathrm{E}(V)$. Since $U = X_1 \sim \textsc{Beta}(\alpha, 1)$, we know $\mathrm{E}(U) = \frac{\alpha}{\alpha+1}$. Because the PDF of V derived in part (d) is the same as the PDF of Y given in part (a), we know $\mathrm{E}(V) = (1+\frac{1}{\alpha})^{-2} = (\frac{\alpha}{\alpha+1})^2$. Finally,

$$\begin{aligned}
\mathrm{E}(UV) &= \int_0^1\int_0^u uv f_{UV}(u,v)\mathrm{d}v\mathrm{d}u = \int_0^1\int_0^u uv\alpha^2\,\frac{v^{\alpha-1}}{u}\mathrm{d}v\mathrm{d}u = \alpha^2\int_0^1\int_0^u v^{\alpha}\mathrm{d}v\mathrm{d}u \\
&= \alpha^2\int_0^1\left[\frac{v^{\alpha+1}}{\alpha+1}\right]_0^u \mathrm{d}u = \frac{\alpha^2}{\alpha+1}\int_0^1 u^{\alpha+1}\ \mathrm{d}u = \frac{\alpha^2}{\alpha+1}\left[\frac{u^{\alpha+2}}{\alpha+2}\right]_0^1 = \frac{\alpha^2}{(\alpha+1)(\alpha+2)}
\end{aligned}$$

and

$$\mathrm{Cov}(U,V) = \mathrm{E}(UV) - \mathrm{E}(U)\mathrm{E}(V) = \frac{\alpha^2}{(\alpha+1)(\alpha+2)} - \left(\frac{\alpha}{\alpha+1}\right)^3.$$