THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE EXAMINATION

NEW MODULAR SCHEME

introduced from the examinations in 2007

MODULE 5

SPECIMEN PAPER A

AND SOLUTIONS

The time for the examination is 1½ hours. The paper contains four questions, of which candidates are to attempt **three**. Each question carries 20 marks. An indicative mark scheme is shown within the questions, by giving an outline of the marks available for each part-question. The pass mark for the paper as a whole is 50%.

The solutions should not be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids. For this reason, they do not carry mark schemes. Please note that in many cases there are valid alternative methods and that, in cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of the questions and solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of the questions or solutions.

Note. In accordance with the convention used in all the Society's examination papers, the notation log denotes logarithm to base *e*. Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

1. (i) The continuous random variable X is distributed with probability density function $f(x)$ given by

$$f(x) = \alpha(1-x)^{\alpha-1}, \quad 0 < x < 1, \quad \alpha > 0.$$

Find the cumulative distribution function of X, $F(x)$ say, and hence obtain the median of X. Also sketch the graphs of $f(x)$ and $F(x)$ for the case $\alpha = 3$.

(8)

(ii) A random sample $x_1, x_2, \ldots, x_n$ is taken from this distribution with a view to estimating the unknown parameter α. Write down the likelihood function of these data, $L(x_1, x_2, \ldots, x_n \mid \alpha)$ say, and show that the maximum likelihood estimate (MLE) of α is given by

$$\hat{\alpha} = \frac{-n}{\sum_{i=1}^{n} \log(1-x_i)}.$$

(4)

Also obtain $\dfrac{d^2 \log L(x_1, x_2, \ldots, x_n \mid \alpha)}{d\alpha^2}$ as a function of α and n. Assuming that $\hat{\alpha}$ is approximately Normally distributed with mean α and variance $-1\Big/\left(\dfrac{d^2 \log L(x_1, x_2, \ldots, x_n \mid \alpha)}{d\alpha^2}\right)$, deduce an approximate 90% confidence interval for α. Evaluate this interval for the sample 0.12, 0.43, 0.07, 0.87, 0.29.

(8)

2. The random variable X has the geometric probability mass function (pmf) given by $f(x)$, where

$$f(x) = (1-p)^x p, \quad x = 0, 1, 2, ..., \quad 0 < p < 1 .$$

(i) Sketch the graph of $f(x)$ for the case $p = 1/3$, for $0 \le x \le 5$.

(5)

(ii) Show that the probability generating function (pgf) of X is given by $G(s)$ where

$$G(s) = \frac{p}{1-(1-p)s}, \quad |s| < \frac{1}{1-p} ,$$

and hence or otherwise obtain the mean and variance of X.

(9)

(iii) Suppose that Y and Z are two independent discrete random variables taking values $r = 0, 1, 2, 3, \ldots$, and that their pgfs are G_Y and G_Z. Show that the pgf of the sum $Y + Z$ is $G_Y G_Z$. Hence find the pgf of $X_1 + X_2$ where X_1 and X_2 are independent and both have the geometric pmf given above.

(6)

3. (i) Discuss briefly the importance (or otherwise) of each of the following properties of a statistic that is to be used as a point estimator of a parameter θ. Illustrate your answers by simple examples of two possible estimators of θ, one which has the property and one which does not; for example in (*A*) give one biased statistic and one unbiased statistic for the same problem.

(*A*) Unbiasedness.

(*B*) Consistency.

(*C*) Efficiency.

(8)

(ii) The random variable X follows the continuous uniform distribution over the range $0 \le x \le \theta$. $X_1, X_2, \ldots, X_n$ represents a random sample of observations from this distribution

(*A*) Show that the maximum likelihood estimator of θ is $\hat{\theta} = X_{\text{max}}$, the largest of the sample values.

(3)

(*B*) Show that the maximum likelihood estimator $\hat{\theta}$ is biased but the adjusted estimator $\{(n + 1)/n\}X_{\text{max}}$ is unbiased. You may **use** the fact that the probability density function of X_{max} is given by $f(x)$ where

$$f(x) = nx^{n-1}/\theta^n, \qquad \text{for } 0 \le x \le \theta.$$

(4)

You are **given** that the method of moments estimator of θ is $2\overline{X}$ (where $\overline{X}$ is the sample mean), that this estimator is unbiased and that its variance is $\theta^2/(3n)$. You are also **given** that the variance of the adjusted estimator in part (ii)(*B*) is $\theta^2/\{n(n+2)\}$.

(*C*) Six values selected at random from the distribution of X are 1.5, 0.7, 1.2, 2.8, 0.5, 1.1. Calculate the value of the method of moments estimator and the value of the adjusted maximum likelihood estimator $\{(n + 1)/n\}X_{\text{max}}$. Compare these estimators as point estimators of θ, basing your discussion on the properties listed in part (i).

(5)

4. The table below shows the joint distribution of two random variables, X and Y.

		Values of Y			
		1	2	3	4
Values of X	1	$6c$	$3c$	$2c$	$4c$
	2	$4c$	$2c$	$4c$	0
	3	$2c$	c	0	$2c$

(i) Find c.

(2)

(ii) Calculate the marginal distributions of X and Y.

(5)

(iii) Calculate $E(X)$ and $\text{Var}(X)$, and show that the covariance $\text{Cov}(X, Y) = 0$.

(6)

(iv) State, with a reason, whether or not X and Y are independent.

(2)

(v) The random variables U and V are defined by

$$U = 1 \text{ if } X = 1 \text{ or } 3, \quad U = 0 \text{ if } X = 2,$$

$$V = 1 \text{ if } Y = 1 \text{ or } 3, \quad V = 0 \text{ if } Y = 2 \text{ or } 4.$$

Tabulate the joint distribution of U and V and state with a reason whether or not U and V are independent.

(5)

SOLUTIONS

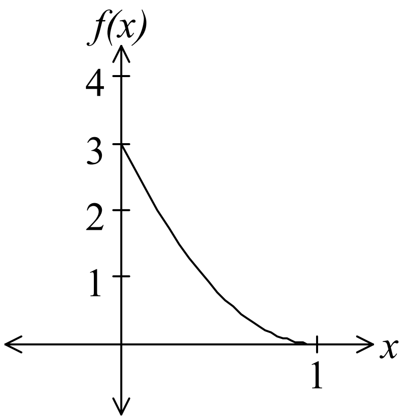
Question 1

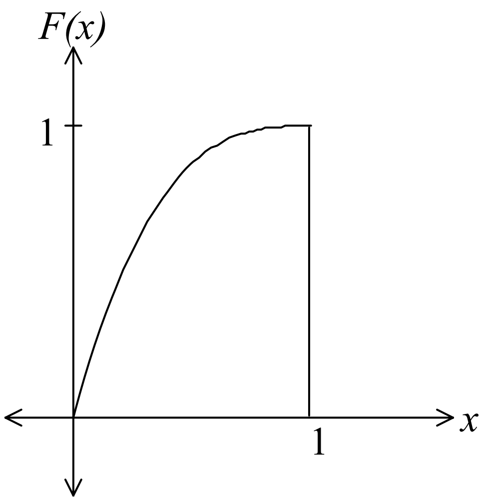
$$f(x) = \alpha(1-x)^{\alpha-1}, \qquad 0 < x < 1, \qquad \alpha > 0.$$

(i) $F(x) = \int_0^x \alpha(1-u)^{\alpha-1}\,du = \left[-(1-u)^{\alpha}\right]_0^x = 1-(1-x)^{\alpha}$ (for $0 < x < 1$ and $\alpha > 0$).

The median m is given by $F(m) = ½$, so we have $1 - (1 - m)^{\alpha} = ½$ or $(1 - m)^{\alpha} = ½$, so that $1 - m = 2^{-1/\alpha}$, i.e. $m = 1 - 2^{-1/\alpha}$.

When $\alpha = 3$, $f(x) = 3(1-x)^2$ and $F(x) = 1-(1-x)^3$ (in [0, 1]).





The solution to part (ii) is on the next page

(ii) $L=\prod_{i=1}^{n}\left[\alpha(1-x_i)^{\alpha-1}\right]=\alpha^n\prod_{i=1}^{n}(1-x_i)^{\alpha-1}$.

Hence $\log L = n\log\alpha+(\alpha-1)\sum_{i=1}^{n}\log(1-x_i)$.

$\therefore \dfrac{d\log L}{d\alpha}=\dfrac{n}{\alpha}+\sum_{i=1}^{n}\log(1-x_i)$ which on setting equal to zero gives that the maximum likelihood estimate is $\hat{\alpha}=\dfrac{-n}{\sum_{i=1}^{n}\log(1-x_i)}$, as required. [Consideration of $\dfrac{d^2\log L}{d\alpha^2}$ (see below) confirms that this is a maximum.]

$\dfrac{d^2\log L}{d\alpha^2}=-\dfrac{n}{\alpha^2}$. Hence, using the result quoted in the question, $\hat{\alpha}$ is approximately Normally distributed with mean α and variance $\dfrac{\alpha^2}{n}$. We estimate the variance by $\dfrac{\hat{\alpha}^2}{n}$, so that we have $\hat{\alpha}\sim \mathrm{N}\left(\alpha,\ \dfrac{\hat{\alpha}^2}{n}\right)$, approximately.

Hence an approximate 90% confidence interval is given by

$$0.90\approx P\left(-1.645<\frac{\hat{\alpha}-\alpha}{\hat{\alpha}/\sqrt{n}}<1.645\right),$$

leading to the interval $\left(\hat{\alpha}-\dfrac{1.645\hat{\alpha}}{\sqrt{n}}\ ,\quad \hat{\alpha}+\dfrac{1.645\hat{\alpha}}{\sqrt{n}}\right)$.

For the given sample, we have n = 5 and the values of $1-x_i$ are 0.88, 0.57, 0.93, 0.13 and 0.71. Therefore

$$\Sigma\log(1-x_i) = -0.1278 - 0.5621 - 0.0726 - 2.0402 - 0.3425 = -3.1452$$

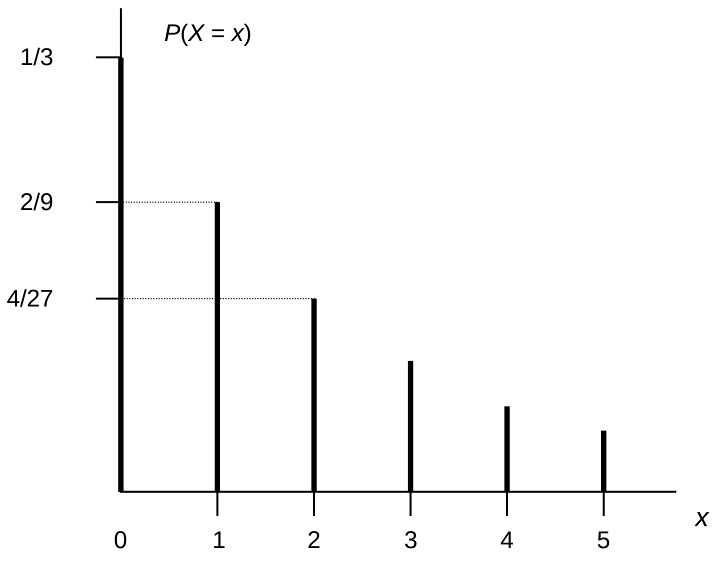
giving $\hat{\alpha}=\dfrac{5}{3.1452}=1.5897$.

Also, $\dfrac{1.645\hat{\alpha}}{\sqrt{n}}=\dfrac{1.645\times 1.5897}{\sqrt{5}}=1.1695$, so the confidence interval is (0.420, 2.759).

<u>Question 2</u>

Probability mass function: $f(x) = (1-p)^x p, \quad x = 0, 1, 2, \ldots, \quad 0 < p < 1$.

(i)



(ii) Probability generating function $G(s)$ is

$$G(s) = E\left[s^X\right] = \sum_{x=0}^{\infty} s^x (1-p)^x p = p\sum_{x=0}^{\infty} \left\{(1-p)s\right\}^x = \frac{p}{1-(1-p)s}$$

(requires $|s| < 1/(1-p)$ for convergence).

The mean is given by $E[X] = G'(1)$. We have $G'(s) = p\frac{1-p}{\left\{1-(1-p)s\right\}^2}$ and inserting $s = 1$ gives $G'(1) = \frac{p(1-p)}{p^2}$, i.e. the mean is $\frac{1-p}{p}$.

The variance is given by $\text{Var}(X) = G''(1) + \text{mean} - \text{mean}^2$. $G''(s) = \frac{2p(1-p)^2}{\left\{1-(1-p)s\right\}^3}$, so that $G''(1) = \frac{2(1-p)^2}{p^2}$. Thus the variance is given by $\frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \left(\frac{1-p}{p}\right)^2$ $= \frac{1}{p^2}\left\{(1-p)^2 + p(1-p)\right\} = \frac{1-p}{p^2}$.

The solution to part (iii) is on the next page

(iii) Let $U = Y + Z$. For U to take value r ($r = 0, 1, 2, 3, \ldots$), we need

$Y = 0$ and $Z = r$
or $Y = 1$ and $Z = r - 1$
or $Y = 2$ and $Z = r - 2$
…
or $Y = r - 1$ and $Z = 1$
or $Y = r$ and $Z = 0$.

Let $p_i = P(Y = i)$ and $\pi_i = P(Z = i)$. Then, from the above,

$$P(U = r) = p_0\pi_r + p_1\pi_{r-1} + p_2\pi_{r-2} + \ldots + p_{r-1}\pi_1 + p_r\pi_0 \ .$$

Now,

$$G_Y = E\left(s^Y\right) = p_0 s^0 + p_1 s^1 + p_2 s^2 + \ldots + p_{r-1} s^{r-1} + p_r s^r + \ldots$$

and

$$G_Z = E\left(s^Z\right) = \pi_0 s^0 + \pi_1 s^1 + \pi_2 s^2 + \ldots + \pi_{r-1} s^{r-1} + \pi_r s^r + \ldots \ .$$

Thus the coefficient of s^r in $G_Y G_Z$ is $p_0\pi_r + p_1\pi_{r-1} + p_2\pi_{r-2} + \ldots + p_{r-1}\pi_1 + p_r\pi_0$ which equals $P(U = r)$ as required.

Alternatively, argue that because Y and Z are independent we have

$$G_{Y+Z} = E\left(s^{Y+Z}\right) = E\left(s^Y\right)E\left(s^Z\right) = G_Y G_Z \ .$$

Hence, for independent random variables X_1 and X_2 having the given geometric distribution, we immediately have

$$G_{X_1+X_2} = \left(G_X\right)^2 = \frac{p^2}{\left(1-(1-p)s\right)^2} \ .$$

Question 3

This solution continues on the next page

Part (i)

Suppose that a random variable *X* has a probability mass or density function which depends on a single parameter θ. An estimator *T* is a function of the set of sample values $X_1, X_2, \ldots, X_n$ forming a random sample from *X*. If several samples (of the same size *n*) are taken, *T* will take a different numerical value for each sample. This *sampling distribution* of *T* can often be found, or at least the mean and variance of it, and these generally form the basis for examining the properties of estimators.

(*A*) An *unbiased* estimator is one whose sampling distribution has mean θ, i.e. $E(T) = \theta$ over repeated sampling. For example, a sample mean is unbiased for estimating the population mean. In principle, unbiased estimators are very useful, because when only one sample is available the value obtained for the estimator gives a good, easily understood idea of the true value of θ (though there is no guarantee that the value of the estimator will be anywhere near the true value of θ – see (*B*) and (*C*) below). In some problems it may be easier or more natural to construct an estimator that is known to be biased but can be made into an unbiased estimator by a simple transformation. An example is when estimating the variance σ^2 of a population whose mean is not known. Dividing the sample sum of squares by *n* gives a biased estimator, but using $(n - 1)$ instead of *n* removes the bias.

(*B*) Precision is important as well as bias. An estimator ought to have greater precision for larger sample size: the larger sample should provide more information. An estimator is *consistent* if the probability of it differing from the parameter being estimated by more than ε, a very small quantity, approaches 0 as sample size $\to \infty$. It is however easier to use a criterion based on variance: if the variance of the sampling distribution $\to 0$ as sample size $\to \infty$, the estimator is consistent. [Some care is needed in using this criterion for biased estimators, in case the estimator is "homing in" on the wrong place. Provided any bias itself $\to 0$ as the sample size $\to \infty$, the criterion is satisfactory.] Almost all common estimators are *consistent* in this sense. For example, for estimating a population mean, the sample mean is unbiased and its variance is σ^2/n, so the sample mean is a consistent estimator of the population mean. One case where this does not happen is the Cauchy distribution (i.e. the *t* distribution with one degree of freedom). This is a useful "heavy-tailed" distribution, but its mean and variance do not exist (neither do any other moments) because of the "weight" of the tails. However, the median or the mode both indicate the "centre" of the distribution, and it is natural to suggest that this might be estimated by the mean of a sample. But, for this distribution, the pdf of the sample mean is the same as that for a single observation; therefore no advantage arises through increasing sample size, and in this case the sample mean is not a consistent estimator of the "centre".

(*C*) If there is a choice of two estimators for the same purpose, and assuming they are both unbiased and consistent, the one with the smaller variance will be preferred. It gives more information per item sampled; somewhat less formally, it is "more likely" to give a value "nearer" to the true value of the parameter θ. The estimator with the smaller variance is said to be more *efficient*. For example the parameter μ in a Normal distribution (i.e. the population mean) may be estimated by the sample median, as an alternative to the sample mean. Both the sample mean and the sample median are unbiased and consistent. But the sample median has a larger variance for a given sample size (it can be shown to be $\pi\sigma^2/(2n)$) and so it is less efficient. The relative efficiency is measured by the inverse ratio of the variances of the two estimators. [In fact, the sample mean is fully efficient, as it has the least possible variance for this estimation problem, whereas the sample median is not fully efficient (these concepts are explored in detail in the Statistical Theory and Methods section of the

Graduate Diploma examination, which also studies the extension to biased estimators using mean square error).]

Part (ii)

The probability density function of X, $g(x)$, is $g(x) = \begin{cases} \dfrac{1}{\theta}, & 0 \le x \le \theta. \\ 0, & \text{otherwise.} \end{cases}$

(A) The likelihood function is $(1/\theta)^n$, i.e. θ^{-n}. By considering the graph of this, we see that it is a decreasing function of θ. Thus it attains its maximum value at the smallest possible value of θ, which is clearly $X_{\max}$ as θ cannot be less than the largest observed value.

(B) $E\left(\hat{\theta}\right) = E\left(X_{\max}\right) = \int_0^\theta x f(x) dx$ [$f(x)$ is as given in the question]

$= \int_0^\theta \frac{n x^n}{\theta^n} dx = \frac{n\theta}{n+1}$, so the maximum likelihood estimator is biased.

It is immediate that the adjusted estimator $\{(n + 1)/n\}X_{\max}$ is unbiased.

(C) The values of the estimators are: moments 2.6 (note that this is an *impossible* value!); adjusted maximum likelihood $(7/6) \times 2.8 = 3.27$.

Unbiasedness

We already have that both estimators are unbiased.

Consistency

The variance of the method of moments estimator is given in the question: $\theta^2/(3n)$. This tends to zero as $n \to \infty$ and so this estimator is consistent (note that it is unbiased, so this criterion based on variance can be used directly).

The variance of the adjusted maximum likelihood estimator is also given in the question: $\theta^2/\{n(n + 2)\}$. This $\to 0$ as $n \to \infty$, so this also is a consistent estimator. (Note. It is also the case that the maximum likelihood estimator itself is consistent. Its bias $\to 0$ as $n \to \infty$ and so does its variance.)

Efficiency

We have $\dfrac{\text{Var}\left(\text{method of moments estimator}\right)}{\text{Var}\left(\text{adjusted estimator}\right)} = \dfrac{\dfrac{\theta^2}{3n}}{\dfrac{\theta^2}{n(n+2)}} = \dfrac{n+2}{3}$

which is > 1 for all n (> 1). Thus the adjusted estimator is more efficient than the method of moments estimator.

Question 4

(i) The sum of all 12 table entries is $30c$. These probabilities must add up to 1, so $c = 1/30$.

(ii) The marginal distributions are given by the row and column totals.

Hence: $P(X = 1) = 15c = 1/2$; $P(X = 2) = 10c = 1/3$; $P(X = 3) = 5c = 1/6$.

Similarly: $P(Y = 1) = 12c = 2/5$; $P(Y = 2) = 6c = 1/5$; $P(Y = 3) = 6c = 1/5$; $P(Y = 4) = 6c = 1/5$.

(iii) $$E(X) = \left(1\times\frac{1}{2}\right) + \left(2\times\frac{1}{3}\right) + \left(3\times\frac{1}{6}\right) = \frac{1}{2}+\frac{2}{3}+\frac{1}{2} = \frac{5}{3}.$$

$$E(X^2) = \left(1\times\frac{1}{2}\right) + \left(4\times\frac{1}{3}\right) + \left(9\times\frac{1}{6}\right) = \frac{1}{2}+\frac{4}{3}+\frac{3}{2} = \frac{10}{3}.$$

$$\therefore \text{Var}(X) = \frac{10}{3} - \left(\frac{5}{3}\right)^2 = \frac{5}{9}.$$

We also need $E(Y)$ later: $E(Y) = \frac{2}{5}+\frac{2}{5}+\frac{3}{5}+\frac{4}{5} = \frac{11}{5}$.

Distribution of XY:

Values of xy	1	2	3	4	6	12
Probability	$6c$	$7c$	$4c$	$6c$	$5c$	$2c$

[$c = 1/30$, see above]

$$E(XY) = \left(1\times\frac{6}{30}\right) + \left(2\times\frac{14}{30}\right) + \left(3\times\frac{4}{30}\right) + \left(4\times\frac{6}{30}\right) + \left(6\times\frac{5}{30}\right) + \left(12\times\frac{2}{30}\right) = \frac{110}{30} = \frac{11}{3}$$

Also we have $E(X)E(Y) = \frac{5}{3}\times\frac{11}{5} = \frac{11}{3}$.

$$\therefore \text{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0.$$

(iv) X and Y are not independent [even though $\text{Cov}(X, Y) = 0$ and even though some cells have $P(X = x, Y = y) = P(X = x).P(Y = y)$]. For example, we have $P(X = 1, Y = 4) = 2/15$, but $P(X = 1).P(Y = 4) = 1/10$.

Solution continued on next page

(v) $U = 1$ if $X = 1$ or 3 $\quad$ $U = 0$ if $X = 2$

$V = 1$ if $Y = 1$ or 3 $\quad$ $V = 0$ if $Y = 2$ or 4

Table of joint distribution of U and V, with margins.

		Values of V: 0	Values of V: 1	
Values of U	0	$2c = 1/15$	$8c = 4/15$	$10c = 1/3$
	1	$10c = 1/3$	$10c = 1/3$	$20c = 2/3$
		$12c = 2/5$	$18c = 3/5$	

Consider for example the cell with $(U, V) = (0, 0)$. The cell probability is 1/15 but the product of the marginal probabilities is 2/15. So U and V are not independent.