EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

(formerly the Examinations of the Institute of Statisticians)

HIGHER CERTIFICATE IN STATISTICS, 1996

Paper I : Statistical Theory

Time Allowed: Three Hours

Candidates should answer **FIVE** *questions.*

All questions carry equal marks.

Graph paper and Official tables are provided.

Candidates may use silent, cordless, non-programmable electronic calculators.

Where a calculator is used the **method** *of calculation should be stated in full.*

Note that $\binom{n}{r}$ *is the same as* ${}^nC_r$ *and that* ln *stands for* $\log_e$.

1. The random variable X has the distribution

$$P(X = x) = \frac{1}{5},\ x = 1, 2, 3, 4, 5:$$

$$= 0 \qquad \text{otherwise}$$

(i) Draw a graph of this distribution and find E(X) and V(X).

(ii) The random variables X_1, X_2 and X_3 are independently distributed, each with the same distribution as X.

(a) Find $P(X_1 = X_2)$ and $P(X_1 = X_2 = X_3)$.

(b) By using the fact that $P(X_1 < X_2) = P(X_2 < X_1)$, or otherwise, show that $P(X_1 < X_2) = \frac{2}{5}$.

(c) By considering $P(X_1 < x_2) \,.\, P(X_3 < x_2)$ for all possible values x_2 of X_2, or otherwise, show that $P(X_1 < X_2 \text{ and } X_2 > X_3) = \frac{6}{25}$.

Turn over

2. (i) In answering a question on a multiple-choice test, a student either knows the answer or guesses. Let p be the probability that the student knows the answer, and assume that a student who guesses picked an answer at random from the m multiple choice answers given. Show that the conditional probability that a student knew the answer to a question, given that he or she answered it correctly, is

$$\frac{mp}{1+(m-1)p}.$$

If $p=\frac{1}{2}$, how many multiple choice answers for a question should the examiner include to be at least 90% certain that a student who answers correctly actually knows the answer?

(ii) One of the games at Cobley Fair involves throwing three balls at a target. The player wins if he or she scores at least two consecutive hits. There is a rule that the player must change hands at each throw, so that there are two possible strategies: throwing with the right, left and right hand, and throwing with the left, right and left hand. Suppose that for a right-handed person the probability of scoring a hit is 0.8 with the right hand but only 0.4 with the left.

Write down the possible outcomes and their probabilities using each strategy, determine which is the better strategy for a right-handed player and calculate the expected number of hits under each strategy. Comment on your answers.

3. The random variable X has a Binomial distribution so that

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x},\ x=0,1,\dots,n;$$
$$= 0 \text{ otherwise}$$

for some positive integer n and probability p, $0<p<1$.

(a) Find $\mathrm{E}(X)$.

(b) Show that

$$\frac{P(X=x+1)}{P(X=x)}=\frac{(n-x)p}{(x+1)(1-p)}\quad,\quad x=0,\ 1,\ \dots,\ n-1,$$

and hence or otherwise show that, if $P(X=k)=P(X=k+1)$ for some integer k, then $(n+1)p$ must be an integer. If $(n+1)p$ is not an integer, show that the mode, m, of the distribution is given by $[(n+1)p]$, the integer part of $(n+1)p$. Under what condition will the mean equal the mode? Give examples of values of n and p for a binomial distribution for which (i) the mean is greater than the mode, (ii) the mode is greater than the mean.

(Question continued on next page)

(c) Suppose that, in a large population, 20% favours having a single European currency. Find the probability that, in a random sample of 25 people, at most 2 people share this view.

4. A newsboy buys a stock of newspapers at 20p each to sell at 30p each. The demand for papers is a random variable D such that

$$P(D = 28) = P(D = 32) = 0.1$$
$$P(D = 29) = P(D = 31) = 0.2$$
$$P(D = 30) = 0.4$$

Show that if the newsboy's stock is s and the demand is d, the gain in pence, G, is given by

$$G(s,\ d) = 30d - 20s, \quad \text{for } d = 28,\ \ldots,\ s$$
$$= 10s, \quad \text{for } d = s,\ s+1,\ \ldots\ 32$$

and hence obtain the missing entries in the following table of $G(s,\ d)$:

		Demand d				
		28	29	30	31	32
Stock s	28	280	.....	.....	.....	.....
	29	.....	.....	290	.....	.....
	30	240	.....	.....	300	.....
	31	.....	.....	.....	.....	310
	32	.....	230	.....	.....	.....

Calculate the newsboy's expected gain for the cases $s = 28,\ 29,\ \ldots,\ 32$ and find the variance of his gain when $s = 29$ and $s = 30$. Comment on your results and give your recommendation as to the newsboy's best choice of s.

Turn over

5. The random variable X has the Poisson probability mass function

$$p(X = x) = e^{-\lambda}.\frac{\lambda^x}{x!} \ , \qquad x = 0,1,2, \ \ldots., \ \lambda > 0;$$
$$= 0 \quad \text{otherwise.}$$

Prove that E(X) = V(X) = λ .

Sketch the graph of this function in the cases $\lambda = \frac{1}{2}$, and $\lambda = 2$.

A random sample $x_1, \ \ldots., \ x_n$ is available from this distribution: obtain the maximum likelihood estimate of λ, $\hat{\lambda}_{ml}$ say, and state whether or not it is unbiased. Use the fact that $\mathrm{V}(X) = \lambda$ to write down a second unbiased estimate of λ, $\hat{\lambda}$ say, and state with a reason which of $\hat{\lambda}_{ml}$ and $\hat{\lambda}$ you would prefer.

In a particular case the sample

2 2 3 3 3 4 5 6 35

was obtained: calculate $\hat{\lambda}_{ml}$. Comment on this result, given that the sample variance is 112. It is later found that the observation of 35 was wrongly recorded: it should have been 8. Calculate $\hat{\lambda}_{ml}$ and s^2 for the corrected data and comment briefly.

6. In a certain country, household incomes X are distributed according to the Pareto density given by

$$f(x) = \frac{\alpha c^{\alpha}}{x^{\alpha+1}}, \quad x \geq c;$$

$$= 0 \quad \text{otherwise,}$$

c and α being positive constants.

(a) Calculate the cumulative distribution function of X, $F(x)$ say, and draw the graphs of $f(x)$ and $F(x)$.

Show that the median and quartiles of this distribution are given by

$$Q_1 = 1.10064c$$
$$M = 1.25992c$$
$$Q_3 = 1.58740c$$

to 5 decimal places, in the case $\alpha = 3$.

(Question continued on next page)

(b) Consider the case $\alpha = 3$. Let a random variable Y be Normally distributed, with the same median and inter-quartile range as X. Show that $\mathrm{V}(Y) = 0.1302c^2$, approximately. Compare $P(Y < c)$ with $P(X < c)$, and also compare $P(Y > 2c)$ with $P(X > 2c)$, and comment on your results.

7. The random variable X follows the exponential distribution with probability density function (pdf)

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0,\ \lambda > 0;$$
$$= 0 \quad \text{otherwise,}$$

and the random variable Y is defined as $X + c$, where c is a positive constant.

(a) (i) On the same graph, sketch the pdfs of X and Y.

(ii) Find the mean and variance of X and, by using the fact that $Y - c = X$, or otherwise, deduce the mean and variance of Y.

(b) (i) Find $P(X > x)$.

(ii) The independent random variables $X_1, \ldots, X_n$ have the same distribution as X. Show that

$$P(X_1 > x,\ X_2 > x,\ \ldots,\ X_n > x) = e^{-n\lambda x},\ x \geq 0,$$

and hence or otherwise show that

$$P(Z \leq x) = 1 - e^{-n\lambda x}, x \geq 0,$$

where $Z = \min(X_1, \ldots, X_n)$ refers to the minimum of $X_1, \ldots, X_n$.

Deduce the pdf of Z and write down the pdf of W, where $W = \min(Y_1, \ldots, Y_n)$ and $Y_i = X_i + c$, $i = 1, \ldots, n$. Comment on your results.

Turn over

8. Define the product-moment correlation ρ between two random variables X and Y, and write down the formula for the sample product-moment correlation coefficient for the paired data $(x_i,\ y_i)$, $i=1,\ \ldots,$ n. Sketch scatterplot graphs to illustrate the cases when

(a) X and Y are closely correlated;

(b) X and Y are weakly correlated;

(c) X and Y are uncorrelated but not independent.

Each candidate for a mathematics examination takes two papers, A and B, and his overall percentage mark is the mean of his percentage marks, X and Y say, on the two papers. Taking $V(X)=144$ and $V(Y)=81$, calculate the standard deviation of the overall percentage marks

(i) assuming that X and Y are independent,

(ii) assuming instead that $\rho(X,\ Y)=\frac{11}{24}$.

Which of assumptions (i) and (ii) do you think is more realistic, and why?