HERIOT-WATT UNIVERSITY

F71SB2 Statistics II

Wednesday, 15 March 2006, 13.30 – 15.30

Attempt ALL 8 questions.
A total of 100 marks is available.

Approved electronic calculators may be used

1. A group of 8 students consists of 4 sets of twins. Suppose that 3 students are selected at random without replacement.

   (a) Find the probability that the selection does not contain a pair of twins. [5]

   (b) Hence, or otherwise, find the probability that the selection contains one pair of twins. [3]

2. (a) Define the cumulative distribution function (*cdf*) of a random variable X, $F_X(x)$, by expressing it as a probability and state the properties it must satisfy. [5]

   (b) Let X be a random variable with range $S_x = \{1, 2, 3, 4\}$. For each of the following determine whether the given values can serve as the values of the *cdf* of X, giving reasons for your answers:

   i) $F_X(1) = 0.3, F_X(2) = 0.5, F_X(3) = 0.8, F_X(4) = 1.2$; [2]

   ii) $F_X(1) = 0.5, F_X(2) = 0.4, F_X(3) = 0.7, F_X(4) = 1.0$. [2]

   (c) Consider a discrete random variable X with the following *cdf*:

$$F_X(x) = \begin{cases} 0, & \text{for } x < 1 \\ \frac{1}{3}, & \text{for } 1 \leq x < 4 \\ \frac{1}{2}, & \text{for } 4 \leq x < 6 \\ \frac{5}{6}, & \text{for } 6 \leq x < 10 \\ 1, & \text{for } x \geq 10. \end{cases}$$

   i) Draw the graph of the *cdf* of X. [3]

   ii) Calculate $P(2 < X \leq 6)$. [3]

   iii) Calculate $P(X = 4)$. [3]

CONTINUED/

3. For a random sample of 100 motorists, the number of times each required to sit the driving test was noted. The results were as follows.

Number of attempts (x) :	1	2	3	4	≥ 5
Number of motorists (f_x) :	42	27	20	11	0

It is assumed that these data are a random sample from a Geometric(p) distribution for some value of p. (Hint: If $X \sim$ Geometric(p) then $P(X = x) = (1 - p)^{x-1}p$ and $E(X) = p^{-1}$.)

(a) Apply the method-of-moments to the above data set to estimate the unknown parameter, p. [4]

(b) Compute the expected numbers in each category for a sample of size 100 using your estimated value of p. [6]

(c) Compare the expected numbers with the actual numbers given above. Do you think the geometric model fits these data well? [2]

4. Let X be a continuous random variable whose probability density function (*pdf*) is

$$f(x) = \begin{cases} \theta x^{\theta-1}, & \text{for } 0 < x < 1; \\ 0, & \text{elsewhere} \end{cases}$$

where $\theta > 0$ is a parameter of the distribution.

(a) By evaluating an appropriate integral, show that $E(X) = \frac{\theta}{\theta+1}$. [4]

(b) Let the values in a random sample of size 5 from this distribution be

0.41 0.84 0.89 0.94 0.98.

Calculate $\hat{\theta}$, the method-of-moments estimate of θ, from these data. [4]

CONTINUED/

5. (a) Let X be a random variable whose distribution is $N(\mu, \sigma^2)$, with probability density function (*pdf*) given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x, \mu < \infty, \quad \sigma > 0.$$

Derive the *pdf* of the random variable $Z = \frac{X-\mu}{\sigma}$, and identify fully its distribution. [6]

(b) Suppose that $X \sim N(80, 16)$. Use statistical tables to calculate the following probabilities:

i) $P(X > 85)$; [3]

ii) $P(77 < X < 81)$. [5]

6. The inside-leg measurements (in inches) of 30 randomly selected male students are given below (in ascending order):

25.1	25.5	27.2	27.9	28.8	28.9	29.3	29.5	29.9	30.1
30.3	30.6	30.6	31.1	31.5	31.7	31.7	31.9	32.2	32.4
32.6	32.9	33.2	33.6	33.8	34.2	34.5	34.9	35.3	36.3

(a) Calculate the 5-point summary for these data. [5]

(b) Construct a stem-and-leaf diagram for the data. Does the plot support the suggestion that the distribution of inside-leg measurements is Normal? [6]

7. Suppose that the time, T, until you receive the next call on your mobile phone, measured in hours, follows an Exp(0.5) distribution.

(a) Calculate the probability that you will not receive a phone call for at least the next two hours, assuming that you have not received a call in the last one hour. [4]

(b) Let X be the random variable representing the number of phone calls that you receive in one hour. Identify fully the distribution of X and hence, or otherwise, find the probability that you will not receive a call in the next one hour. [4]

(c) Let $T_1, T_2, \ldots, T_{25}$, be independent random variables from an Exp(0.5) distribution. Use the Central Limit Theorem to calculate the approximate probability $P(\bar{T} > 1)$, where $\bar{T} = \frac{1}{25}\sum_{i=1}^{25} T_i$. (Hint: If $X \sim \text{Exp}(\lambda)$, then $E(X) = \lambda^{-1}$ and $\text{var}(X) = \lambda^{-2}$.) [6]

CONTINUED/

8. The population distribution of the length (measured in microns) of a certain kind of bacteria (A) is known to be Normal with unknown mean and variance. Eight bacteria are selected at random and their lengths measured. The resulting data are:

$$6.3 \quad 7.3 \quad 6.6 \quad 6.8 \quad 8.0 \quad 7.6 \quad 7.1 \quad 7.0.$$

For these data $\sum x_i = 56.7$ and $\sum x_i^2 = 403.95$.

(a) Evaluate the sample mean and sample variance for these data. [3]

(b) Use the sample mean and sample variance, evaluated in (a), to calculate a 95% confidence interval for the unknown mean, μ_A, of the population. [6]

(c) Explain briefly the meaning of an observed '95% confidence interval' for an unknown parameter μ. [3]

(d) Now suppose that a sample of a different kind of bacteria (B) is taken, and a 95% confidence interval for the difference $\mu_A - \mu_B$ in the mean length of bacteria A and B is calculated as $(0.42, 1.58)$. Comment on any difference in the mean length of the two kinds of bacteria. [3]

[Grand total: 100]

END OF PAPER