F72SE2

HERIOT-WATT UNIVERSITY

SCHOOL OF MATHEMATICAL AND COMPUTER SCIENCES
ACTUARIAL MATHEMATICS AND STATISTICS

STATISTICS 5

Thursday 17 March 2005 — 4.30 p.m.– 6.30 p.m.

Attempt all FIVE questions

Approved electronic calculators may be used

1. (a) Suppose that $X_1, X_2, ..., X_n$ are independent random variables, and let $Z = X_1 + X_2 + ... + X_n$. Show carefully that

$$M_Z(\theta) = M_{X_1}(\theta)M_{X_2}(\theta)\cdots M_{X_n}(\theta)$$

where $M_Z(\theta)$ and $M_{X_i}(\theta)$ are the moment generating functions of Z and $X_i, 1 \leq i \leq n$, respectively. [3]

(b) Suppose that $X_1, X_2, ..., X_n$ are independent, identically distributed random variables with $X_i \sim Poisson(\lambda)$ for some $\lambda > 0$.

Using moment generating functions, or otherwise, show that $Z = X_1 + X_2 + ... + X_n$ has a $Poisson(\lambda n)$ distribution. (Be sure to give reasons.) [4]

(c) Suppose that $Y \sim Poisson(n)$ where $n > 0$ is very *large.*

Use part (b) above and the Central Limit Theorem to explain why $\frac{Y-n}{\sqrt{n}}$ has an approximately $N(0,1)$ distribution. [3]

(d) A speed camera is installed on a busy road. In any given week, the number of cars which are caught speeding by the camera has a $Poisson(1)$ distribution, independent of the number of cars caught in all other weeks.

i) Use the Central Limit Theorem to approximate the probability that, in total, more than 60 cars are caught speeding by the camera in one year (52 weeks). [5]

ii) Let X equal the number of weeks in one year (i.e. in 52 weeks) in which no cars are caught speeding by the camera.

Approximate $\Pr(X \leq 15)$. [5]

2. **(a)** Suppose that $X \sim Exp(2)$. Carefully show that $M_X(\theta) = \frac{2}{2-\theta}$ for $\theta < 2$, where $M_X(\theta)$ is the moment generating function of X.

Use $M_X(\theta)$ to determine $E(X^3)$. [5]

(b) Suppose that N is an integer-valued random variable with $\Pr(N = n) = (4/5)^{n-1}(1/5)$ for $n \geq 1$.

Show that

$$G_N(s) = \frac{s}{5 - 4s} \qquad \text{(for } |s| \leq 1\text{)},$$

where $G_N(s)$ is the probability generating function for N. Use $G_N(s)$ to determine $E(N)$ and $Var(N)$. [6]

(c) Suppose that $X_1, X_2, X_3, ..$ are independent, identically distributed random variables with common moment generating function $M_X(\theta)$. Also suppose that N is a non-negative integer-valued random variable (with probability generating function $G_N(s)$) and suppose that N is independent of the variables $\{X_1, X_2, ...\}$.

Let $Z = X_1 + X_2 + ... + X_N$ (where $Z = 0$ if $N = 0$).

Write down an expression for $M_Z(\theta)$, the moment generating function of Z, in terms of $M_X(\theta)$ and $G_N(\theta)$. Hence, or otherwise, show that $E(Z) = E(N)E(X)$. [4]

(d) A radio station is running a phone-in contest. The first person to telephone the station with the correct answer to a trivia question wins 2 concert tickets.

Suppose that the probability that a given caller answers the question correctly is 1/5 (independently of all other callers) and suppose that the times between consecutive calls (in minutes) to the station are i.i.d. exponential random variables with parameter $\lambda = 2$.

The DJ at the radio station starts to answer the phone at 9AM. Using moment generating functions, or otherwise, determine the distribution of the length of time that elapses until the first correct answer is received by the the station. (You may assume that the time until the DJ answers the first call has an $Exp(2)$ distribution and is independent of the times between all subsequent calls to the station.) [5]

3. Let $\underline{X} = (X_1, X_2, ..., X_n)$ denote a random sample from a $U(0, \theta)$ distribution where $\theta > 0$ is unknown.

(a) Explain why the method-of-moments estimator for θ is given by

$$\hat{\theta} = 2\bar{X}$$

where $\bar{X}$ denotes the sample mean. Show that the MSE of this estimator is given by $\frac{\theta^2}{3n}$ and explain why it is a *consistent* estimator of θ. [5]

(b) Consider the family of estimators for θ of the form $a\bar{X}$. Express the bias, variance and MSE of $a\bar{X}$ as function of a and identify the value of a which defines the *most efficient* estimator of this form. [7]

(c) Use Chebyshev's inequality to obtain an upper bound for the frequency with which the estimator $2\bar{X}$ falls outside the interval $(0.9\theta, 1.1\theta)$ when the sample size $n = 100$. [4]

(d) For realised values $\underline{x} = (x_1, ..., x_n)$ identify the likelihood $L(\theta; \underline{x})$ and sketch its graph. Identify the maximum likelihood estimate of θ. [4]

4. **(a)** Let X denote a single sample from a $Bin(n,p)$ distribution where $p \in (0,1)$ is unknown. Prove carefully that the maximum likelihood estimator for p is given by $\hat{p} = \frac{X}{n}$ [6]

(b) A class of 10 students sit an exam which they fail if they make *2 or more* errors. The distribution of the number of errors made by each student is independent of the numbers made by all other student and follows a Poisson(λ) distribution where $\lambda > 0$ is unknown.

i) Explain carefully why the probability that a student passes the exam is is given by

$$p(\lambda) = e^{-\lambda}(1 + \lambda)$$

and verify that $p(\lambda)$ decreases monotonically with λ. [4]

ii) Suppose now that 4 students out of the 10 pass the exam. Explain why the likelihood for this observation is given by

$$L(\lambda) \propto e^{-4\lambda}(1 + \lambda)^4(1 - e^{-\lambda} - \lambda e^{-\lambda})^6$$

and use the results of (a) and b(i) to show that the MLE of λ satisfies

$$5e^{-\hat{\lambda}}(1 + \hat{\lambda}) = 2.$$

Suggest how you might solve this equation. *(You may quote a result concerning invariance of maximum likelihood estimators under 1-1 transformation of parameters.)* [7]

iii) Suppose now that you had observed the number of errors made by the 10 students to be

$$0, 0, 1, 1, 2, 2, 3, 3, 3, 5.$$

Give an estimate for λ from these data and give reasons for your choice of estimate. [3]

5. (a) Explain what you understand by the *frequentist* view of probability. Give an examples of 'probabilities' that are acceptable and unacceptable to a pure frequentist. [5]

(b) Let $\underline{X}$ denote a random sample from a distribution with mean μ and variance σ^2. Determine the bias of the estimator (for σ^2) defined by

$$g(\underline{X}) = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

where $\bar{X}$ denotes the sample mean. [5]

(c) The lifetimes (in months) of a certain kind of component follow an Exp(λ) distribution where λ is unknown. Lifetimes are assumed independent between components. An engineer wishes to estimate λ and selects a random sample of 10 components and sets them in operation at $t = 0$. He observes them continuously for the next 2 months noting the precise times at which any component fails. He observes that five have failed at times $t_1, t_2, .., t_5 < 2$ and that the remaining 5 are still operational at $t = 2$. Explain carefully why the likelihood for these data is given by

$$L(\lambda) \propto \lambda^5 e^{-\lambda(\sum t_i + 10)}$$

and find the MLE of λ if $\sum t_i = 3.2$. [10]

END OF PAPER