MA930 Data Analysis. Class test (2019)

The exam comprises two pages with a total of five questions.
Full marks are given for correct answers to each of the five questions.
Note: calculators are neither required nor allowed.

Q1. Characteristic function, moments and samples from a uniform distribution
Consider a random number X with a probability density that is uniform between $-a$ and a, but zero elsewhere.

(a) Show that the characteristic function of this distribution is

$$\phi_x(t) = \frac{\sin(at)}{at}. \tag{1}$$

(b) What value do all the odd moments of this distribution take?

(c) The sample mean is $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$. Its characteristic function to order t^4 can be written

$$\phi_{\bar{X}}(t) = \exp\left(-\frac{t^2\kappa_2}{2!} + \frac{t^4\kappa_4}{4!} + O(t^6)\right) \tag{2}$$

where κ_2 and κ_4 are the 2nd and 4th-order cumulants of the sample mean. Using the relation between the characteristic functions of the random variable and its sample mean, derive the form for the 4th-order cumulant κ_4 it terms of a and n, and show that it is negative.

Total marks: 6

Q2. Sample variance for normally distributed random numbers

A sample $\{x_1, \cdots x_n\}$ of n random numbers is drawn from a normal distribution with population mean μ_p and variance σ_p^2.

(a) Consider the sample mean $\overline{x}$ and variance estimator σ_b^2

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \quad \text{and} \quad \sigma_b^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \overline{x}^2.$$

Demonstrate that $\overline{x}$ provides an unbiased estimation of the population mean μ_p whereas σ_b^2 is a biased estimator of σ_p^2. Provide an equation for the unbiased variance estimator.

(b) Write down an equation for the likelihood $\mathcal{L}(\mu, \sigma^2)$ of the data $\{x_1 \cdots x_n\}$ having been generated by a normal distribution with mean μ and variance σ^2.

(c) Integrate $\mathcal{L}(\mu, \sigma^2)$ over μ to show that the maximum-likelihood estimation from the σ^2 marginal distribution gives an unbiased estimator for the variance.
HINT: Express $\sum_{k=1}^{n}(x_k - \mu)^2$ in terms of n, μ, $\overline{x}$ and σ_b^2.

Total marks: 7

Q3. Integral identity for the expectation
Let $C(x)$ be the cumulative distribution for a positive random variable X. Prove that

$$\langle X \rangle = \int_0^\infty dx (1 - C(x)). \tag{3}$$

Total marks: 2

Q4. Frequentist statistics for a negative binomial distribution
Consider Bernoulli random numbers with a probability of success p that are drawn until there are a total of r failures. The random variable X, defined as the number of successes drawn before stopping, has the following distribution

$$P(X = k) = \frac{(k + r - 1)!}{k!(r - 1)!}(1 - p)^r p^k \quad \text{where } k \text{ is a non-negative integer.} \tag{4}$$

(a) Provide an argument for why the distribution takes the form of equation 4 Knowledge of the standard binomial distribution can be assumed.

(b) An experiment is designed in which a coin is flipped multiple times until a total of two tails are thrown. Assuming the coin is fair ($p = 0.5$) what is the probability for stopping with k heads having been thrown? Verify that $P(0) = 1/4$.

(c) You now perform an experiment on a coin which you suspect is biased towards heads and get three heads before the second tail is thrown. Derive the p-value for this result and show it is not sigficant at the 5% level.

Total marks: 4

Q5. Autoregressive models
A first-order autoregressive model, with $|\phi| < 1$, takes the form

$$u_t = \phi u_{t-1} + \epsilon_t$$

where ϵ_t are independent random numbers drawn from a standard normal distribution.

(a) Consider first an initial condition $u_0 = 0$. Show that the mean $\langle u_t \rangle = 0$ for all $t > 0$. Show also that the variance grows to saturation with the following form

$$\langle u_t^2 \rangle = \frac{1 - \phi^{2t}}{1 - \phi^2}. \tag{5}$$

(b) Now consider a case where the process has been going on since long in the past so that any transients have decayed away. Derive a general solution for u_t. Evaluate the variance of your solution and confirm it agrees with the $t \to \infty$ limit of equation 5.

(c) The process x_t obeys a second-order autoregressive model $x_t = 2\phi x_{t-1} - \phi^2 x_{t-2} + \epsilon_t$. By using the substitution $u_t = x_t - \phi x_{t-1}$ prove that the solution is

$$x_t = \sum_{k=0}^{\infty} \epsilon_{t-k} \phi^k (k + 1) \tag{6}$$

for the case where the transients have died away.

Total marks: 6

EXAM END