MA930 Data Analysis. Class test (2017)

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 for the Poisson distribution
For a Poisson distribution the probability of k events (k can be zero or any positive integer) is:

$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

The parameter λ is the mean number of events.

(a) Prove that the characteristic function for the Poisson distribution is

$$\phi_\lambda(t) = e^{\lambda(e^{it}-1)}.$$

(b) Derive the variance $\langle(k-\lambda)^2\rangle$ of the Poisson distribution using its characteristic function.

Total marks: 4

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. Frequentist statistics

(a) A cheap but not fully reliable test for dust allergy is compared to a definitive test. One hundred people are tested: 90 people who genuinely have an allergy are correctly identified; 3 people who don't have allergies are also correctly identified; and the cheap test predicts that 92 have an allergy. What are the type I (false negative) and type II (false positive) error rates?

(b) You are concerned that a coin is biased towards heads and decide to do a significance test. You flip the coin ten times and get 9 heads and 1 tail. Find a sufficient approximation for the p-value that allows you to state whether or not the result is significant at the 5% or 1% levels.

Total marks: 4

Q4. Autoregressive models

A second-order autoregressive model obeys the difference equation

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$$

where ϵ_t are independent, normally distributed random numbers with zero mean and variance σ_ϵ^2, and c, ϕ_1, ϕ_2 are constants. The constants are such that a statistical steady-state exists. You can assume that any initial transients have died away.

(a) Find the mean $\langle X \rangle$ of the process. Then show that

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t, \tag{1}$$

where $x_t = X_t - \langle X \rangle$ measures the difference from the mean.

(b) Multiply both sides of equation (1) by x_{t-1} and take expectations to show that

$$\langle x^2 \rangle_1 = \frac{\phi_1 \langle x^2 \rangle}{1 - \phi_2} \tag{2}$$

where $\langle x^2 \rangle_1$ is $\langle x_t x_{t-1} \rangle$, which is the autocovariance at one time-step difference.

(c) Use equation (1) to find another equation linking $\langle x^2 \rangle$ and $\langle x^2 \rangle_1$. Use this, together with equation (2), to derive an equation for the variance $\langle x^2 \rangle$ in terms of ϕ_1, ϕ_2 and σ_ϵ^2 only.

Total marks: 6

Q5. Backpropagation in a network with one hidden layer

Consider a network for categorical classification with an input layer of size n_x with an additional bias neuron; a hidden layer of size n_h with an additional bias neuron; and one output neuron which gives the prediction p. The weights between the input and hidden layer w_{ij}, and hidden layer and output v_j, have dimensions $(n_x + 1, n_h)$ and $(n_h + 1, 1)$, respectively. There are n_s samples so that, in matrix form, the prediction of the network can be written

$$\mathbf{H} = f(\tilde{\mathbf{X}}\mathbf{w}) \quad \text{and} \quad \mathbf{P} = f(\tilde{\mathbf{H}}\mathbf{v}) \quad \text{with} \quad f(z) = \frac{1}{1 + e^{-z}},$$

and where, for example $\mathbf{H}$ is an (n_s, n_h) matrix and $\tilde{\mathbf{H}}$ is an $(n_s, n_h + 1)$ matrix for the hidden neuron layer. The vector of targets $\mathbf{T}$ is binary (i.e. elements are 0 or 1). The cost function is

$$C = -\sum_{s=1}^{n_s} \left[T_s \log(P_s) + (1 - T_s) \log(1 - P_s) \right].$$

(a) The gradient of the cost function for the w weights can be written in matrix form as

$$\frac{\partial C}{\partial \mathbf{w}} = \frac{1}{n_s} \tilde{\mathbf{X}}' \boldsymbol{\Delta}^h$$

where $\mathbf{X}'$ is the transpose of $\mathbf{X}$, which is the input matrix. By considering $\partial C / \partial w_{ij}$, derive the form for the elements Δ_{sj}^h of the matrix $\boldsymbol{\Delta}^h$ in terms of P_s, T_s, H_{sj} and v_j.

Total marks: 4

EXAM END