MA930 Data Analysis. Class test (2018)

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. Random-number generator with unknown distribution
Your supervisor has asked you to make sense of some old software code. In it you found the following function that converts a random number Y drawn from the standard flat distribution between 0 and 1 to a new random number X

$$X = \sqrt{1 + 8Y} - 1. \tag{1}$$

(a) What is the range of values the random numbers generated by this formula can take?
(b) What is the distribution $f(x)$ for the random numbers X?

Total marks: 4

Q2. Characteristic functions and probability distributions

(a) Derive the characteristic function for a Binomial distribution of parameter p with n draws. (HINT: Consider the characteristic function of a sum of n Bernoulli random numbers.)

A Poisson distribution is given by $P(k) = \lambda^k e^{-\lambda}/k!$ for $k = 0, 1, \cdots$.
(b) Derive the mean for this distribution.
(c) By using characteristic functions, or otherwise, show that the Binomial distribution tends to a Poisson distribution in the limit $p \to 0$ and $n \to \infty$ such that $\lambda = np$ remains finite.

Consider a biological process where for a given event a Poisson-distributed number k of vesicles is released. Each vesicle contains a random quantity of hormone a_j that is Gaussian distributed with mean μ and standard deviation σ. Thus, for each event, the total hormone released is

$$A = \sum_{j=1}^{k} a_j \quad \text{in the event that } k > 0 \text{ or } A = 0 \text{ if } k = 0 \tag{2}$$

Note that k and the amounts a_j are all independent random variables.
(d) Derive the form of the probability density $p(A)$ in terms of a sum. You may find it useful to consider the conditional densities $p(A|k)$ given k releases.
(**Note**: be careful with how you write the contribution for $k = 0$).

Total marks: 7

Q3. Bayesian statistics

Consider a Poisson random process that is characterised by a mean parameter λ. A total of n independent samples are drawn: k_1, k_2 ... k_n.
(a) Provide the likelihood function for this data set, given a process of mean λ.
(b) What is the maximum likelihood estimator for λ?
(c) A gamma distribution $g(\lambda; \alpha, \beta) = \beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}/\Gamma(\alpha)$ provides a conjugate prior for a Poisson process. Taking account of the samples drawn, what values α', β' characterise the posterior?
(d) The mean of a gamma distribution $g(\lambda; \alpha, \beta)$ is α/β. Show that the mean calculated from the posterior distribution tends to the maximum likelihood result as $n \to \infty$ and provide the order $1/n$ correction in terms of α, β and n.

Total marks: 5

Q4. Correlated autoregressive models

Two autoregressive models obey the equations

$$\begin{aligned} X_t &= a + \phi X_{t-1} + \epsilon_t \\ Y_t &= b + \psi Y_{t-1} + \epsilon_t \end{aligned}$$

where ϵ_t are independent random numbers with zero mean and variance σ_ϵ^2. Note that both processes are driven by the same random numbers and are therefore correlated. The other quantities a, b, ϕ and ψ are constants chosen such that both processes have a statistical steady state. You can also assume these processes have been going on since infinitely long in the past.

(a) Provide formulae for both the mean and variance of X.

(b) By re-writing the difference equation in terms of $x_t = X_t - \langle X \rangle$, solve to provide the general solution for X_t in terms of a weighted sum over the history of the noise $\{\epsilon_t\}$.

(c) Show that the autocovariance (when $n > 0$) is equal to

$$\langle X_{t+n} Y_t \rangle - \langle X \rangle \langle Y \rangle \quad = \quad \phi^n \frac{\sigma_\epsilon^2}{1 - \phi\psi} \tag{3}$$

and provide the form for $n < 0$ with an explanation of how you arrived at it.

Total marks: 5

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

A1. Random-number generator with unknown distribution

(a) The range of X is from 0 to 2.

(b) It is the cumulative distribution function of X which we call $C(x)$ which we can use to map Y on to X. So this equation represents $C^{-1}(y) = x$. Hence re-arranging gives

$$y = \frac{1}{8}\left((x+1)^2 - 1\right) = C(x) \tag{4}$$

The probability density is just the derivative of this object, which gives

$$f(x) = \frac{1}{4}(x+1) \tag{5}$$

which is correctly normalised over the range $x = 0 \to 1$.

A2. Characterisic functions and probability distributions

(a) Definition of the characteristic function is $\phi_x(t)$. For a binomial distribution the answer is

$$\phi_x(t) = \Big(1 + p(e^t - 1)\Big) \tag{6}$$

(b) The mean of a Poisson is

$$\begin{aligned} \langle k \rangle &= e^{-\lambda} \sum_0^\infty k \frac{\lambda^k}{k!} && (7) \\ &= e^{-\lambda} \lambda \sum_0^\infty \frac{\lambda^{k-1}}{(k-1)!}! = \lambda && (8) \end{aligned}$$

(c) The characteristic function of a poisson is

$$\phi_x(t) = e^{\lambda(e^t - 1)} \tag{9}$$

Now we write the characteristic function of the binomial as

$$e^{n \log(1+p(e^t-1))} \simeq e^{np(e^t-1)+O(np^2)} \tag{10}$$

which yields the result as required for $np = \lambda$ being finite.

(d) We consider first the conditional density $p(A|k)$ given k releases. This is a normal distribution $g_k(A)$ with mean $k\mu$ and variance $k\sigma^2$. The total distribution is the sum of these weighted with the Poisson factor $P(k)$

$$p(A) = \sum_{k=0}^\infty g_k(A) P(k) \tag{11}$$

where it should be noted that $g_0(A)$ is a Dirac delta function.

A3. Bayesian statistics

(a) The Likelihood is

$$\mathcal{L} = \frac{\lambda^{\sum_{j=1}^{n} k_j} e^{-\lambda n}}{\prod_{j=1}^{n} k_j!} \tag{12}$$

(b) Setting $d\mathcal{L}/d\lambda = 0$ yields

$$0 = -n\mathcal{L} + \frac{\sum_{j=1}^{n} k_j}{\hat{\lambda}} \tag{13}$$

so that

$$\hat{\lambda} = \frac{1}{n}\sum_{j=1}^{n} k_j \tag{14}$$

(c) We have $\alpha' = \alpha + \sum_{j=1}^{n} k_j$ and $\beta' = \beta + n$.

(d) The mean calculated from the posterior is

$$\frac{\alpha'}{\beta'} = \frac{\alpha + \sum_{j=1}^{n} k_j}{\beta + n} \simeq \hat{\lambda} + \frac{\alpha - \beta}{n} \tag{15}$$

A4. Correlated autoregressive models

(a) The quantities required are

$$\langle X \rangle = \frac{a}{1-\phi} \quad \text{and} \quad \text{Var} = \frac{\sigma_\epsilon^2}{1-\phi^2} \tag{16}$$

(b) The equation for x_t is

$$x_t = \phi x_{t-1} + \epsilon_t \tag{17}$$

which has solution

$$x_t = \sum_{k=0}^{\infty} \epsilon_{t-k} \phi^k \tag{18}$$

(c) The solut for y_t is needed which is simply $y_t = \sum_{k=0}^{\infty} \epsilon_{t-k} \psi^k$. The autocovariance is therefore

$$\langle x_{t_n} y_t \rangle = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \phi^k \psi^j \langle \epsilon_{t+n-k} \epsilon_{t-j} \rangle = \sigma_\epsilon^2 \phi^n \sum_{j=0}^{\infty} (\phi\psi)^j \tag{19}$$

which can be summed to give the required result. The route to providing the answer for $n < 0$ is to just swap all variables corresponding to x and y to give

$$\langle x_t y_{t+n} \rangle = \psi^n \frac{\sigma_\epsilon^2}{1 - \psi\phi}. \tag{20}$$

A5. Backpropagation in a network with one hidden layer

(a) It is simplest to consider a single data point for which the cost is $c = -[t\log(p) + (1-t)\log(p)]$ and the activations are $p = f(h_j v_j)$ and $h_j = f(x_i w_{ij})$. The chain of derivatives is

$$\frac{\partial c}{\partial w_{ij}} = \frac{\partial c}{\partial p}\frac{\partial p}{\partial h_j}\frac{\partial h_j}{\partial w_{ij}} = \left[\frac{p-t}{p(1-p)}\right][p(1-p)v_j]\,[h_j(1-h_j)\tilde{x}_i] = \tilde{x}_i\,(h_j(1-h_j)v_j(p-t)) \quad (21)$$

Hence the matrix entries are

$$\Delta_{sj} = H_{sj}(1-H_{sj})v_j(P_s - T_s) \quad (22)$$

where $j = 1 \to n_h$ only.