Queen Mary
University of London

BSc and MSci Examination

Thursday 10th May 2012 10:00 - 12:30

PHY328 Statistical Data Analysis Duration: 2 hours 30 minutes

YOU ARE NOT PERMITTED TO READ THE CONTENTS OF THIS QUESTION PAPER UNTIL INSTRUCTED TO DO SO BY AN INVIGILATOR.

Answer ALL questions from Section A and TWO questions from Section B

ONLY NON-PROGRAMMABLE CALCULATORS ARE PERMITTED IN THIS EXAMINATION. PLEASE STATE ON YOUR ANSWER BOOK THE NAME AND TYPE OF MACHINE USED.

COMPLETE ALL ROUGH WORKINGS IN THE ANSWER BOOK AND CROSS THROUGH ANY WORK WHICH IS NOT TO BE ASSESSED.

IMPORTANT NOTE:
THE *A*CADEMIC REGULATIONS STATE THAT POSSESSION OF UNAUTHORISED MATERIAL AT ANY TIME WHEN A STUDENT IS UNDER EXAMINATION CONDITIONS IS AN ASSESSMENT OFFENCE AND CAN LEAD TO EXPULSION FROM QMUL.

PLEASE CHECK NOW TO ENSURE YOU DO NOT HAVE ANY NOTES, MOBILE PHONES OR UNATHORISED ELECTRONIC DEVICES ON YOUR PERSON. IF YOU HAVE ANY THEN PLEASE RAISE YOUR HAND AND GIVE THEM TO AN INVIGILATOR IMMEDIATELY. PLEASE BE AWARE THAT IF YOU ARE FOUND TO HAVE HIDDEN UNAUTHORISED MATERIAL ELSEWHERE, INCLUDING TOILETS AND CLOAKROOMS IT WILL BE TREATED AS BEING FOUND IN YOUR POSSESSION. UNAUTHORISED MATERIAL FOUND ON YOUR MOBILE PHONE OR OTHER ELECTRONIC DEVICE WILL BE CONSIDERED THE SAME AS BEING IN POSSESSION OF PAPER NOTES. MOBILE PHONES CAUSING A DISRUPTION IS ALSO AN ASSESSMENT OFFENCE.

EXAM PAPERS CANNOT BE REMOVED FROM THE EXAM ROOM.

Examiners:
Dr A. Bevan and Dr. J. Wilson

SECTION A. Answer all questions.

A1

(a) Write down the conditions on the probability $\mathcal{P}$ of some quantity.

(b) Write down the total probability $\mathcal{P}$ for two independent events $\mathcal{P}_1$ and $\mathcal{P}_2$ to both occur.

(c) Write down the total probability $\mathcal{P}$ for either of two independent events $\mathcal{P}_1$ or $\mathcal{P}_2$ to occur.

(d) Write down Bayes theorem and explain each term in full.

(e) Write down the frequentist definition of probability in terms of the number of interesting events occurring n and the total number of events N in a given sample.

(f) Write down the normalisation condition for a probability density function $\mathcal{P}(x)$, and the relationship between $\mathcal{P}(x)$ and the corresponding likelihood function $\mathcal{L}(x)$.

[10]

A2

(a) Write down the equations used for computing the arithmetic mean, standard deviation (with unbiased variance) and the related quantities: the Pearson skew, covariance and the Pearson correlation coefficient.

(b) Write down the formulae used to compute the expectation value and variance of some quantity x described by some probability density function $\mathcal{P}$.

(c) Write down the mean and variance for (i) a Binomial and (ii) a Poisson distribution.

(d) Write down in full the definition of median and modal value.

[10]

A3

(a) Write down the equation corresponding to a χ^2 sum and explain each of the terms.

(b) Explain the term numbers of degrees of freedom ν and give the value of ν for a data sample of N events.

(c) Explain fully how the gradient descent method works in the context of determining an optimal set of parameters $\underline{p}$ for some test statistic T (for example a χ^2).

[10]

A4

(a) For a Gaussian distribution, what are the two-tailed probabilities of obtaining the result of some measurement within (i) 1σ, (ii) 2σ, and (iii) 3σ of the mean value.

(b) Consider the following measurements and state if these are consistent with each other (i) $x_1 = 1.0 \pm 0.5$ and $x_2 = 0.0 \pm 0.5$, (ii) $x_1 = 1.0 \pm 0.5$ and $x_2 = 0.3 \pm 0.1$, and (iii) $x_1 = 2.0 \pm 0.5$ and $x_2 = 0.0 \pm 0.1$.

(c) Write down the general combination of errors formula used to determine the variance for some function $f(x, y)$ given the variance on x and y.

(d) Explain fully what statistical and systematic uncertainties are, and give examples.

(e) What is a *blind analysis* technique, and when might it be useful?

[10]

A5

(a) Write down the form for an extended likelihood function $\mathcal{L}$ comprising of two components, A and B, where A corresponds to signal and B corresponds to background. Explain the meaning of each term.

(b) For a Gaussian likelihood function $G(x, \mu, \sigma)$, determine the change in $-\ln \mathcal{L}$ from the minimum value that corresponds to a change in x of $\pm 3\sigma$.

(c) Describe a perceptron and list four possible activation functions.

[10]

SECTION B. Answer two of the three questions in this section.

B1

a) Fully describe the process of testing a null hypothesis and subsequently placing a confidence level on the conclusions drawn. **[5]**

b) Using Bayes theorem determine the probability that it will rain tomorrow given that rain is forecast, and that 90% of the time when it rains, rain has been correctly forecast. In contrast when no rain is forecast it rains only 5% of the time, and it rains a total of 50 days in the year. **[5]**

c) Two points A and B are separated by a known distance L. A third point C is measured to be L_1 m from A and L_2 m from B. Describe in full how you would use Bayes theorem to estimate the position of C relative to point A, sketching the expected contributions from each measurement, and the combined constraint obtained from both of the measurements. Is the solution you expect to obtain unique? **[15]**

B2

a) Given two measurements of the same observable $x_1 = 1.0 \pm 1.5$ and $x_2 = 2.0 \pm 0.7$, estimate the mean value, and uncertainty, on the average of the observable using a χ^2 scan. Compare the result you obtain with that obtained when computing a weighted average of the two results directly. **[10]**

b) Briefly describe how you would measure the detector efficiency of a particle detector, and compute the uncertainty on that quantity. **[5]**

c) Assuming that $y = ax^2 + b$ demonstrate how the method of least squares to determine the values of a and b given a set of measurement points x_i and y_i. Assume that the uncertainties on x_i, $\sigma(x_i)$, are all the same. **[10]**

B3

a) Write down the functional form of a Fisher linear discriminant, explaining the significance of each element, and fully explain how the coefficients α_i are determined given two samples of data A (signal) and B (background). **[5]**

b) Compute the coefficients of a Fisher discriminant to optimally separate the samples A and B where

$$\mu_A = \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}, \tag{1}$$

and the covariance matrix for events of type A is

$$V_A = \begin{pmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{pmatrix},$$

for sample A, and:

$$\mu_B = \begin{pmatrix} 0.0 \\ 0.0 \end{pmatrix}, \tag{2}$$

and the covariance matrix for events of type B is

$$V_B = \begin{pmatrix} 2.0 & 0.2 \\ 0.2 & 2.0 \end{pmatrix},$$

for sample B. **[10]**

c) Fully describe how to construct a Multi-Layer Perceptron from individual perceptrons, and how weight parameters can be determined using back-propagation, and how these may subsequently be validated. **[10]**