UNIVERSITY OF LONDON

MSci EXAMINATION 2006

For Internal Students of

Royal Holloway

DO NOT TURN OVER UNTIL TOLD TO BEGIN

PH4515B: COMPUTING AND STATISTICAL DATA ANALYSIS

Time Allowed: **TWO AND A HALF** hours

Answer ***THREE QUESTIONS*** only

No credit will be given for attempting any further questions

Approximate part-marks for questions are given in the right-hand margin

Only CASIO fx85WA Calculators or CASIO fx85MS Calculators are permitted

PH4515B/12

2005-06

GENERAL PHYSICAL CONSTANTS

Permeability of vacuum	μ_0	=	$4\pi \times 10^{-7}$	$H\ m^{-1}$
Permittivity of vacuum	ε_0	=	8.85×10^{-12}	$F\ m^{-1}$
	$1/4\pi\varepsilon_0$	=	9.0×10^{9}	$m\ F^{-1}$
Speed of light in vacuum	c	=	3.00×10^{8}	$m\ s^{-1}$
Elementary charge	e	=	1.60×10^{-19}	C
Electron (rest) mass	m_e	=	9.11×10^{-31}	kg
Unified atomic mass constant	m_u	=	1.66×10^{-27}	kg
Proton rest mass	m_p	=	1.67×10^{-27}	kg
Neutron rest mass	m_n	=	1.67×10^{-27}	kg
Ratio of electronic charge to mass	e/m_e	=	1.76×10^{11}	$C\ kg^{-1}$
Planck constant	h	=	6.63×10^{-34}	J s
	$\hbar = h/2\pi$	=	1.05×10^{-34}	J s
Boltzmann constant	k	=	1.38×10^{-23}	$J\ K^{-1}$
Stefan-Boltzmann constant	σ	=	5.67×10^{-8}	$W\ m^{-2}\ K^{-4}$
Gas constant	R	=	8.31	$J\ mol^{-1}\ K^{-1}$
Avogadro constant	N_A	=	6.02×10^{23}	mol^{-1}
Gravitational constant	G	=	6.67×10^{-11}	$N\ m^2\ kg^{-2}$
Acceleration due to gravity	g	=	9.81	$m\ s^{-2}$
Volume of one mole of an ideal gas at STP		=	2.24×10^{-2}	m^3
One standard atmosphere	P_0	=	1.01×10^{5}	$N\ m^{-2}$

MATHEMATICAL CONSTANTS

$e \cong 2.718$ $\pi \cong 3.142$ $\log_e 10 \cong 2.303$

PART MARKS

1. Consider the cumulative distribution function

$$F(x;\theta) = 1 - \left(1 + \frac{x}{\theta}\right) e^{-x/\theta},$$

where $0 \le x < \infty$ and $\theta > 0$. The mean and variance of x are $E[x] = 2\theta$ and $V[x] = 2\theta^2$.

(a) Show that the corresponding probability distribution function is

$$f(x;\theta) = \frac{x}{\theta^2} e^{-x/\theta}.$$

[3]

(b) Consider n independent values sampled from this distribution: $x_1, \ldots, x_n$.

Show that the log-likelihood function is of the form

$$\ln L(\theta) = -2n \ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i + const.$$

Show that the Maximum Likelihood estimator for θ is $\hat{\theta} = \bar{x}/2$, where $\bar{x}$ is the sample mean. **[5]**

(c) Show that $\hat{\theta}$ is an unbiased estimator. (Use the properties of the distribution given above.) **[4]**

(d) Find the variance $V[\hat{\theta}]$. **[4]**

(e) Find the minimum variance bound MVB and thus show explicitly that the estimator $\hat{\theta}$ is efficient. **[4]**

2. Consider the joint probability density function (pdf) for the continuous random variables *x* and *y*,

$$f(x,y;\theta)=\frac{1}{\theta^2}e^{-(x+y)/\theta}\,,\quad \text{where both } x\geq 0 \text{ and } y\geq 0\,.$$

(a) Find the conditional pdf $h(x\,|\,y;\theta)$. **[3]**

(b) Are *x* and *y* independent? Justify your answer. **[2]**

(c) Let $u=x+y$ and $v=x-y$. Show that the joint pdf $g(u,v)$ is

$$g(u,v)=\frac{1}{2\theta^2}e^{-u/\theta}$$

for a particular region of the (*u*,*v*) plane and zero elsewhere. Find the region of the (*u*,*v*) plane where the formula above holds. **[6]**

(d) Show that the marginal pdf of $v=x-y$ is

$$g_v(v)=\frac{1}{2\theta}\exp\left[-\frac{|v|}{\theta}\right].$$

[3]

(e) Suppose we have a sample of independent (x,y) pairs, i.e. $(x_1,y_1),\ldots,(x_n,y_n)$.

Write down the log-likelihood function for θ.

Find the Maximum Likelihood estimator for θ.

Describe briefly and illustrate with a simple sketch the graphical method for determining the variance of the estimator $\hat{\theta}$. **[6]**

PART MARKS

3. (a) Suppose a number of events of two different types, 'signal' and 'background', are modelled as Poisson variables n_s and n_b, with mean values *s* and *b*, respectively. Suppose the value of *b* is known, that we only observe the sum $n = n_s + n_b$, and we want to estimate *s*.

Given a single observation of *n*, show that the log-likelihood function can be written

$$\ln L(s) = n \ln(s+b) - s + const.$$ **[4]**

(b) As an estimator of *s* take $\hat{s} = n - b$. Is this the Maximum Likelihood (ML) estimator?

Find the expectation value and variance of $\hat{s}$. Show that the relative statistical error $\sigma_{\hat{s}} / s$ is

$$\frac{\sigma_{\hat{s}}}{s} = \frac{\sqrt{s+b}}{s}.$$ **[6]**

(c) Describe briefly the concept of a *P*-value.

Suppose *b* is known and we observe n_{obs}. Describe how one would go about computing a *P*-value to provide evidence that *s* is non-zero. **[5]**

(d) Now suppose there is no background, i.e. $b = 0$, and we observe $n = 0$. Find an upper limit on *s* at 90% confidence level and evaluate it numerically. **[5]**

4. Two measured values, y_1 and y_2, are both assumed to be independent and to follow a Gaussian distribution with a common mean, $E[y_1] = E[y_2] = \lambda$, and standard deviations σ_1 and σ_2, respectively.

(a) Show that the Least-Squares estimator for λ is

$$\hat{\lambda} = \frac{\sigma_2^2 y_1 + \sigma_1^2 y_2}{\sigma_1^2 + \sigma_2^2}.$$ **[5]**

(b) Find the bias and variance of $\hat{\lambda}$. **[4]**

(c) Suppose one were to assume an estimator of the form

$$\hat{\lambda} = w y_1 + (1 - w) y_2$$

for some appropriately chosen constant *w*. Show that the value of *w* that minimizes the variance of this estimator gives the same $\hat{\lambda}$ as the LS estimator from (a). **[5]**

(d) Suppose now that the values y_1 and y_2 have each been determined from two data sets containing *n* and *m* independent observations, respectively, of a variable *x*,

$$y_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{and} \qquad y_2 = \frac{1}{m}\sum_{j=1}^{m} x_j \ ,$$

and that *c* of these *x* values are common between the two sets. Suppose the variance of *x* is $V[x] = \sigma^2$ and its expectation value is $E[x] = \lambda$.

Show that the covariance of y_1 and y_2 is $\mathrm{cov}[y_1, y_2] = \dfrac{c\sigma^2}{nm}$.

Explain briefly how such a non-zero covariance would be taken into account in estimating λ (you do not have to actually find the estimator). **[6]**

PART MARKS

5. Under two hypotheses H_0 and H_1 the continuous random variable x follows the probability density functions (pdfs)

$$f(x|H_0) = (1+\theta)x^{\theta},$$

$$f(x|H_1) = (1+\theta)(1-x)^{\theta},$$

where $0 \le x \le 1$ and $\theta > 0$.

(a) First consider only hypothesis H_0. Describe how to generate random values that follow $f(x|H_0)$ using both the transformation method and the acceptance-rejection method. Assume one has a means to generate values r uniformly distributed between 0 and 1. For the transformation method, find an appropriate transformation $x(r)$. **[8]**

(b) We observe values of x that follow either H_0 or H_1 with probabilities π_0 and $\pi_1 = 1-\pi_0$, respectively. Suppose we want to select values that correspond to H_0 by requiring $x > x_c$ for an appropriately chose value x_c.

Find the selection efficiencies for both hypotheses, i.e., the probabilities $P(x > x_c | H_0)$ and $P(x > x_c | H_1)$ as a function of the parameters given.

Find the purity, i.e., the probability $P(H_0 | x > x_c)$. **[7]**

(c) Suppose now that the parameter θ is not known in advance but has been estimated to be $\hat{\theta}$ using an estimator with variance $V[\hat{\theta}]$. Using error propagation, write down an expression for the standard deviation of $\hat{\alpha} = \ln \hat{\theta}$.

Suppose that both θ and π_0 are estimated with estimators $\hat{\theta}$ and $\hat{\pi}_0$, which have variances $V[\hat{\theta}]$ and $V[\hat{\pi}_0]$ and a covariance $\mathrm{cov}[\hat{\theta}, \hat{\pi}_0]$. Using error propagation, write down an expression for the variance of $\hat{\beta} = \hat{\pi}_0 \ln \hat{\theta}$. **[5]**

END