Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2007

This paper is also taken for the relevant examination for the Associateship.

# M2S1

# PROBABILITY AND STATISTICS II

Date: Tuesday, 8th May 2007 Time: 10 am – 12 noon

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used.

Statistical tables will not be available.

A Formula Sheet is provided on pages 5-6.

1. (a) A coin shows heads with probability $p$, independently on each toss. Let $\pi_n$ be the probability that the number of heads after $n$ tosses is even. Show carefully that $\pi_{n+1} = (1-p)\pi_n + p(1-\pi_n), n \geq 1$, and hence derive $\pi_n$ explicitly. [The number 0 is even.]

(b) Explain what is meant by the *indicator function* $I_A$ of an event $A$.

Let $I_i$ be the indicator function of the event $A_i, 1 \leq i \leq n$, and let $N = \sum_1^n I_i$ be the number of values of $i$ such that $A_i$ occurs. Show that $E[N] = \sum_i p_i$, where $p_i = P(A_i)$, and find $var[N]$ in terms of the quantities $p_{ij} = P(A_i \cap A_j)$.

(c) A fair die has two green faces, two red faces and two blue faces, and the die is thrown once. Let $X = 1$ if a green face is uppermost, $X = 0$ otherwise, and let $Y = 1$ if a blue face is uppermost, $Y = 0$ otherwise.

Find $cov[X, Y]$.

2. (a) The random variable $X$ is uniformly distributed on the interval $[0, 1]$. Find the cumulative distribution function and the probability density function of $Y$, where

$$Y = \frac{2X}{1-X}.$$

(b) Let $X$ and $Y$ be independent random variables with respective density functions $f_X$ and $f_Y$.

(i) Show that $Z = Y/X$ has density function

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(xz)|x| dx.$$

(ii) Deduce that $T = \tan^{-1}(Y/X)$ is uniformly distributed on $(-\frac{\pi}{2}, \frac{\pi}{2})$ if and only if

$$\int_{-\infty}^{\infty} f_X(x) f_Y(xz)|x| dx = \frac{1}{\pi(1+z^2)}, \ z \in \mathbb{R}.$$

(iii) Verify that the condition in (ii) holds if $X$ and $Y$ both have the normal distribution with mean 0 and variance $\sigma^2 > 0$.

3. (a) The President of Statistica relaxes by fishing in the clear waters of Lake Tchebychev. The number of fish that she catches is a Poisson variable with parameter $\lambda$. The weight of each fish in Lake Tchebychev is an independent normally distributed random variable with mean $\mu$ and variance $\sigma^2$. [Since $\mu$ is much larger than $\sigma$, fish of negative weight are rare, and much prized by gourmets.] Let $Z$ be the total weight of the President's catch. Compute $E[Z]$ and $E[Z^2]$.

Show, quoting any results you need, that the probability that the President's catch weighs less than $\lambda\mu/2$ is less than $4(\mu^2+\sigma^2)/(\lambda\mu^2)$.

(b) Conditional on $Y=y$, the random variable $X$ has the $Binomial(n,y)$ distribution, and the marginal distribution of $Y$ is $Beta(\alpha,\beta)$.

What is the marginal probability mass function of $X$? What is the conditional distribution of $Y$, given $X=x$?

4. (a) Let $U$ be uniformly distributed on $[0,1]$ and let $X=\{-\frac{1}{\beta}\log_e(U)\}^{1/\alpha}$.

Find the probability density function of $X$, and calculate the $r$th moment of $X$, $E[X^r]$, $r\geq 1$.

(b) A random variable $X$ has probability density function $f_X(x)=\frac{1}{2}\exp(-|x|),\ x\in\mathbb{R}$.

Find the moment generating function of $X$, and calculate $var[X]$.

The random variables $X_1,X_2,\ldots$ are independent, identically distributed, with the same distribution as $X$. Define $S_n=n^{-1/2}\sum_{i=1}^n X_i$.

Find the moment generating function of $S_n$ and show that, as $n\to\infty$, it converges to the moment generating function of a random variable $Y$, which you should identify.

Explain briefly how the result that $S_n$ converges in distribution to $Y$ could alternatively be deduced from the Central Limit Theorem.

5. (a) Let $X_1, X_2, X_3, X_4$ be independent, identically distributed $N(0,1)$ random variables. Identify the following distributions:

(i) The distribution of
$$Y_1 = 2X_1 - 3X_2 + X_3;$$

(ii) The distribution of
$$Y_2 = X_1/X_2;$$

(iii) The distribution of
$$Y_3 = \frac{X_1^2 + X_2^2}{2X_3^2};$$

(iv) The distribution of
$$Y_4 = \frac{\sqrt{3}X_4}{\sqrt{(X_1^2 + X_2^2 + X_3^2)}};$$

(v) The *joint* distribution of
$$Y_5 = X_1 + X_2 \text{ and } Y_6 = 2X_1 + X_2.$$

(vi) The *conditional* distribution of $Y_6$, given $Y_5 = y_5$.

(b) Let $X_1, \ldots, X_n$ be independent, identically distributed $N(\mu, \sigma^2)$, where both $\mu$ and $\sigma^2$ are unknown.

State, without proof, the joint distribution of the random variables $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ and $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

Explain clearly how the joint distribution allows construction of an appropriate test statistic for testing the null hypothesis $H_0 : \mu = \mu_0$ against the alternative hypothesis $H_1 : \mu \neq \mu_0$. Describe in detail how you would carry out the test.

How would you test the hypothesis $H_0 : \sigma^2 = \sigma_0^2$ against the alternative $H_1 : \sigma^2 \neq \sigma_0^2$, if $\mu$ were *known*?

## DISCRETE DISTRIBUTIONS

| | RANGE $\mathbb{X}$ | PARAMETERS | MASS FUNCTION $f_X$ | CDF $F_X$ | $\mathsf{E}_{f_X}[X]$ | $\mathsf{Var}_{f_X}[X]$ | MGF $M_X$ |
|---|---|---|---|---|---|---|---|
| $Bernoulli(\theta)$ | $\{0,1\}$ | $\theta \in (0,1)$ | $\theta^x(1-\theta)^{1-x}$ | | $\theta$ | $\theta(1-\theta)$ | $1-\theta+\theta e^t$ |
| $Binomial(n,\theta)$ | $\{0,1,...,n\}$ | $n \in \mathbb{Z}^+, \theta \in (0,1)$ | $\binom{n}{x}\theta^x(1-\theta)^{n-x}$ | | $n\theta$ | $n\theta(1-\theta)$ | $\left(1-\theta+\theta e^t\right)^n$ |
| $Poisson(\lambda)$ | $\{0,1,2,...\}$ | $\lambda \in \mathbb{R}^+$ | $\frac{e^{-\lambda}\lambda^x}{x!}$ | | $\lambda$ | $\lambda$ | $\exp\left\{\lambda\left(e^t-1\right)\right\}$ |
| $Geometric(\theta)$ | $\{1,2,...\}$ | $\theta \in (0,1)$ | $(1-\theta)^{x-1}\theta$ | $1-(1-\theta)^x$ | $\frac{1}{\theta}$ | $\frac{(1-\theta)}{\theta^2}$ | $\frac{\theta e^t}{1-e^t(1-\theta)}$ |
| $NegBinomial(n,\theta)$ | $\{n,n+1,...\}$ | $n \in \mathbb{Z}^+, \theta \in (0,1)$ | $\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$ | | $\frac{n}{\theta}$ | $\frac{n(1-\theta)}{\theta^2}$ | $\left(\frac{\theta e^t}{1-e^t(1-\theta)}\right)^n$ |
| or | $\{0,1,2,...\}$ | $n \in \mathbb{Z}^+, \theta \in (0,1)$ | $\binom{n+x-1}{x}\theta^n(1-\theta)^x$ | | $\frac{n(1-\theta)}{\theta}$ | $\frac{n(1-\theta)}{\theta^2}$ | $\left(\frac{\theta}{1-e^t(1-\theta)}\right)^n$ |

For **CONTINUOUS** distributions (see over), define the **GAMMA FUNCTION**

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$$

and the LOCATION/SCALE transformation $Y = \mu + \sigma X$ gives

$$f_Y(y) = f_X\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} \qquad F_Y(y) = F_X\left(\frac{y-\mu}{\sigma}\right) \qquad M_Y(t) = e^{\mu t}M_X(\sigma t) \qquad \mathsf{E}_{f_Y}[Y] = \mu+\sigma\mathsf{E}_{f_X}[X] \qquad \mathsf{Var}_{f_Y}[Y] = \sigma^2\mathsf{Var}_{f_X}[X]$$

## CONTINUOUS DISTRIBUTIONS

| | $\mathbb{X}$ | PARAMS. | PDF $f_X$ | CDF $F_X$ | $\mathsf{E}_{f_X}[X]$ | $\mathsf{Var}_{f_X}[X]$ | MGF $M_X$ |
|---|---|---|---|---|---|---|---|
| $Uniform(\alpha,\beta)$ (standard model $\alpha=0,\beta=1$) | $(\alpha,\beta)$ | $\alpha<\beta\in\mathbb{R}$ | $\dfrac{1}{\beta-\alpha}$ | $\dfrac{x-\alpha}{\beta-\alpha}$ | $\dfrac{(\alpha+\beta)}{2}$ | $\dfrac{(\beta-\alpha)^2}{12}$ | $\dfrac{e^{\beta t}-e^{\alpha t}}{t(\beta-\alpha)}$ |
| $Exponential(\lambda)$ (standard model $\lambda=1$) | $\mathbb{R}^+$ | $\lambda\in\mathbb{R}^+$ | $\lambda e^{-\lambda x}$ | $1-e^{-\lambda x}$ | $\dfrac{1}{\lambda}$ | $\dfrac{1}{\lambda^2}$ | $\left(\dfrac{\lambda}{\lambda-t}\right)$ |
| $Gamma(\alpha,\beta)$ (standard model $\beta=1$) | $\mathbb{R}^+$ | $\alpha,\beta\in\mathbb{R}^+$ | $\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ | | $\dfrac{\alpha}{\beta}$ | $\dfrac{\alpha}{\beta^2}$ | $\left(\dfrac{\beta}{\beta-t}\right)^\alpha$ |
| $Weibull(\alpha,\beta)$ (standard model $\beta=1$) | $\mathbb{R}^+$ | $\alpha,\beta\in\mathbb{R}^+$ | $\alpha\beta x^{\alpha-1}e^{-\beta x^\alpha}$ | $1-e^{-\beta x^\alpha}$ | $\dfrac{\Gamma(1+1/\alpha)}{\beta^{1/\alpha}}$ | $\dfrac{\Gamma(1+2/\alpha)-\Gamma(1+1/\alpha)^2}{\beta^{2/\alpha}}$ | |
| $Normal(\mu,\sigma^2)$ (standard model $\mu=0,\sigma=1$) | $\mathbb{R}$ | $\mu\in\mathbb{R},\sigma\in\mathbb{R}^+$ | $\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$ | | $\mu$ | $\sigma^2$ | $e^{\{\mu t+\sigma^2t^2/2\}}$ |
| $Student(\nu)$ | $\mathbb{R}$ | $\nu\in\mathbb{R}^+$ | $\dfrac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left\{1+\frac{x^2}{\nu}\right\}^{(\nu+1)/2}}$ | | $0$ (if $\nu>1$) | $\dfrac{\nu}{\nu-2}$ (if $\nu>2$) | |
| $Pareto(\theta,\alpha)$ | $\mathbb{R}^+$ | $\theta,\alpha\in\mathbb{R}^+$ | $\dfrac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}}$ | $1-\left(\dfrac{\theta}{\theta+x}\right)^\alpha$ | $\dfrac{\theta}{\alpha-1}$ (if $\alpha>1$) | $\dfrac{\alpha\theta^2}{(\alpha-1)(\alpha-2)}$ (if $\alpha>2$) | |
| $Beta(\alpha,\beta)$ | $(0,1)$ | $\alpha,\beta\in\mathbb{R}^+$ | $\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ | | $\dfrac{\alpha}{\alpha+\beta}$ | $\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ | |