UNIVERSITY OF LONDON [C245 2004]

B.ENG. AND M.ENG. EXAMINATIONS 2004

For Internal Students of Imperial College London

This paper is also taken for the relevant examination for the Associateship of the City & Guilds of London Institute.

COMPUTING C245

STATISTICS

Date Friday 7th May 2004 2.30 - 4.30 pm

Answer Question 1 and THREE others

[Before starting, please make sure that the paper is complete. There should be a total of FIVE questions. Ask the invigilator for a replacement if this copy is faulty.]

1. For each part, circle the letter of the answer you believe to be correct. Marks will be subtracted for incorrect answers.

   (i) A sample of five values has a mean of 4. Four of the values are 6, 1, 5, 5. What is the missing value?

   (a) 20
   (b) 3
   (c) 4
   (d) 17
   (e) 5

   (ii) The standard deviation is measured in units which are:

   (a) The squares of the original units
   (b) The square roots of the original units
   (c) The same as the units of variance
   (d) The same as the original units
   (e) The square of the units of variance

   (iii) Two random variables, A and B, can each take values 0 or 1. Which of the following conditions necessarily implies $P(A = 0|B = 0) = P(B = 0|A = 0)$?

   (a) A and B independent
   (b) $P(A = 0 \cap B = 0) = P(A = 1 \cap B = 1)$
   (c) $P(A = 0) = P(B = 1)$
   (d) $P(A = 0 \cap B = 1) = P(A = 1 \cap B = 0)$
   (e) $P(A = 1) = P(B = 0)$

   (iv) Random variable A can take values $0, 1,$ and 2. Random variable B can take values 0 and 1.
   If $P(B = 0) = 0.4$, $P(A = 0 \cap B = 0) = 0.1$, $P(A = 1 \cap B = 0) = 0.2$ and $P(A = 2 \cap B = 1) = 0.3$, what is the conditional probability $P(B = 0|A = 2)$?

   (a) 0.25
   (b) 0.1
   (c) 0.2
   (d) 0.75
   (e) 0.4

(v) I can model the probability distribution of the number of successive heads in a sequence of tosses of a biased coin by a geometric distribution with parameter p, where $p = 0.4$ is the probability of a head. What is the probability that I will obtain a run of two or fewer heads before a tail occurs?

(a) 0.6
(b) 0.24
(c) 0.16
(d) 0.1
(e) 0.936

(vi) Counts of a particular phenomenon are known to follow a Poisson distribution. If observation shows that counts of size 0 occur 50 % of the time, what is the parameter of the distribution?

(a) 1
(b) 0.693
(c) 1.443
(d) 0.5
(e) 2

(vii) If the function $f(x) = \begin{cases} cx^3 & 0 < x < 1 \\ 0 & \text{else} \end{cases}$ is a pdf, what is the value of c?

(a) 1/4
(b) 4
(c) 1/3
(d) 3
(e) 2

(viii) What is the pdf of a random variable which has cdf $F(x) = 1 - \exp(-x^2)$, for $x > 0$?

(a) $1 - 2x\exp(-x^2)$
(b) $2x\exp(-x^2)$
(c) $\exp(-x^2)$
(d) 1
(e) Cannot be analytically determined

(ix) The values 3, 5 and 7 are known to come from an exponential distribution. What is the maximum likelihood estimator of the parameter of this distribution?

(a) 15
(b) 5
(c) 1/15
(d) 1/5
(e) 10

(x) An estimator $\hat{\theta}$ is an unbiased estimator of a parameter θ if

(a) $E(\hat{\theta}) = \theta/(n-1)$
(b) $E(\hat{\theta}) = \mu$
(c) $E(\hat{\theta}) = \theta$
(d) $\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\theta)$
(e) $\mathrm{Var}(\hat{\theta}) \to 0$ as the sample size increases

2. (i) In 1693, Sir Samuel Pepys asked Isaac Newton which of the following three events was more likely:

A: to get exactly one 6 in a throw of six fair dice;

B: to get exactly two 6's in a throw of twelve fair dice;

C: to get exactly three 6's in a throw of eighteen fair dice.

What probability distribution is an appropriate one for answering this question? Without evaluating the probabilities, write down expressions for the exact probabilities of A, B and C using this distribution.

(ii) A lecturer's assistant preparing slides for a new course occasionally makes mistakes of two types, type A and type B. Type A mistakes are errors in theory, where the assistant has misunderstood the material. Type B mistakes are simple typographical errors. The number of type A errors per slide is a random variable following a *Poisson* (λ_A) distribution. The number of type B errors per slide is a random variable following a *Poisson* (λ_B) distribution. These two random variables are independent: what happens with type A errors does not influence what happens with type B errors. Suppose that $\lambda_A = 0.5$ and $\lambda_B = 1$.

(a) Find the probability that a slide contains no errors.
(b) Find the probability that a slide contains a single error.
(c) Find the conditional probability of there being 1 type A error given that there is 1 type B error.
(d) Find the conditional probability of there being 1 type A error given that there is 1 error altogether.
(e) Find the probability that a slide contains at least one error of each type.

3. (i) Suppose that the number of calls to a file server every hour follows a Poisson distribution with parameter μ.

   (a) What is the probability that there will be no calls in an hour?

   (b) Use this information to show that the distribution of the time to the first arrival is exponential.

   (c) If X has an exponential distribution, show that

$$P(X > s + t \mid X > t) = P(X > s).$$

   (ii) (a) In order for a rocket launch system to do its job, it is necessary that component $T1$ lasts more than two hours, component $T2$ lasts more than two hours, and at least one of components $T3$ or $T4$ lasts more than two hours. If the time to failure of each component is a random variable following an exponential distribution with parameter λ, independent of the failure time of the other components, derive, in terms of λ, the probability that the job will be completed successfully.

   (b) If observation shows that, on average, the lifetimes of components of the above types are all two hours, give a numerical value for the probability that the job will be successfully completed within two hours.

4. (i) A random variable X is known to follow a distribution $f(x) = ce^{-(x-\mu)^2}$ for some constant c and unknown parameter μ.

   (a) Determine the value of c.

   (b) If a sample of values is drawn independently from this distribution, show that the maximum likelihood estimator of μ is given by the sample mean.

   (c) Using the fact that $E(x) = \mu$, show that the maximum likelihood estimator is an unbiased estimator of μ.

   (d) What would be the variance of the mean of a sample of size 10 drawn from this distribution?

   (ii) A bivariate distribution has pdf $f(x, y) = 6(x + y^2)/5$ for $0 < x,\ y < 1$.

   (a) Find the marginal distribution of x.

   (b) Find the conditional distribution of y given $x = 0.25$.

   (c) Determine whether x and y are independent.

5. (i) Define the following terms, as used in statistical hypothesis testing:
   - (a) Type 1 error
   - (b) Type 2 error
   - (c) Significance level
   - (d) Power

   (ii) The data below were collected in a classic study of the relationship between crime and drinking. Carry out an appropriate test of the hypothesis that there is no relationship between the type of crime and drinking behaviour. Describe each step of the process clearly, and state your conclusion in words.

	Arson	Rape	Violence	Stealing	Coining	Fraud
Drinker	50	88	155	379	18	63
Abstainer	43	62	110	300	14	144

   (iii) Define the terms likelihood, prior distribution, and posterior distribution, and show how they are used in Bayesian estimation.

END OF PAPER

COMP 245: Probability and Statistics for Students of Computing
2004

This sheet contains important formulae you may need in the examination. It does not contain definitions, concepts, or other material, and it does not contain simple formulae you would be expected to be able to derive or remember yourselves.

(Arithmetic) mean $\bar{x} = \frac{1}{n}\sum x_i$

Median: order the sample values $\{x_1, x_2, \ldots, x_n\}$ so that $x_{(1)}$ is the smallest, $x_{(2)}$ is the next smallest, and so on, then the median is the value $x_{(n+1)/2}$.

Quartiles: the first quartile is $x_{([n+1]/4)}$, using the same ideas as in defining the median.

Geometric mean: $x_G = \sqrt[n]{\prod x_i}$ and *Harmonic mean:* $x_H = \left(\frac{1}{n}\sum \frac{1}{x_i}\right)^{-1} = \frac{n}{\sum 1/x_i}$

Variance $s^2 = \frac{1}{(n-1)}\sum (x_i - \bar{x})^2$, *standard deviation*: $\sqrt{\text{variance}}$

Skewness: $\frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s}\right)^3$

S the set of all possible events, ϕ the empty set
Notation: $s \in S$ $\quad A \subset B$
$\phi \subset A \subset S$ for all A
$A \cup B$ (A or B) $\quad A \cap B$ (A and B) $\quad$ both commutative
$A \cap B$ is the *joint event* of A and B
A and B are *disjoint events* if $A \cap B = \phi$
Complement of A denoted by: A' or $\overline{A}$

$$P(\phi) = 0 \qquad P(S) = 1 \qquad P(A) = 1 - P(A')$$

For two disjoint events A and B (i.e. events for which $A \cap B = \phi$) $P(A \cup B) = P(A) + P(B)$

For *any* two events A and B: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Generalise: $P(\cup_i A_i) = \sum_i P(A_i) - \sum P(A_i \cap A_j) + \sum P(A_i \cap A_j \cap A_k) + \ldots\ldots$

Two events are said to be *independent* if the occurrence or non-occurrence of one is not affected by whether or not the other occurs

If two events A and B are independent, then $P(A \cap B) = P(A) \,.\, P(B)$

The probability that A will occur, *given that B has occurred* is denoted $P(A \mid B)$
If A and B are independent then $P(A \mid B) = P(A)$.

In general, $P(A \cap B) = P(A \mid B) \,.\, P(B)$ and $P(A \cap B) = P(B \mid A) \,.\, P(A)$

From this $P(A \mid B) = P(B \mid A)P(A)/P(B)$ *(Bayes theorem)*

Now $P(B) = P(B \mid A)P(A) + P(B \mid A')P(A')$ *(theorem of total probability)*

So Bayes theorem can also be written as $P(A|B) = \dfrac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')}$

The mean or *expected value* of a random variable is $E(X) = \sum_x xP(x)$, often denoted μ.

The variance of a random variable is

$V(X) = \sum_x (x-\mu)^2 P(x) = E(X^2) - E(X)^2 = E\left[(X - E(X))^2\right]$, often denoted σ^2.

The skewness of a random variable is $S(X) = \sum \left(\dfrac{x-\mu}{\sigma}\right)^3 P(x) = \dfrac{E\left[(x-\mu)^3\right]}{\sigma^3}$

$E(aX + bY) = aE(X) + bE(Y)$

$V(aX + bY) = a^2V(X) + b^2V(Y)$ if X and Y are independent

$V(aX + bY) = a^2V(X) + b^2V(Y) + 2abCov(X,Y)$, always, with $Cov(X,Y)$ the *covariance* of X and Y, defined as $Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$

More generally, the mean of the sum of a weighted combination of $X_1,\ldots,X_n$, $\sum a_i X_i$ is $\mu = \sum a_i \mu_i$ and the variance of the sum of a weighted combination of *independent* $X_1,\ldots,X_n$, $\sum a_i X_i$ is $\sigma^2 = \sum a_i^2 \sigma_i^2$. If they are not independent then

$\sigma^2 = \sum a_i^2 \sigma_i^2 + \sum_{i \neq j} a_i a_j Cov(X_i, X_j)$

The discrete uniform distribution

Let S be the set of integers from 1 to n.

$P(X = x) = 1/n$ with $\mu = \dfrac{(n+1)}{2}$ and $\sigma^2 = \dfrac{1}{12}(n^2 - 1)$

Bernoulli distribution

Let $P(E) = P(X=1) = p$ and $P(E') = P(X=0) = 1-p = q$

$P(X = x) = p^x q^{1-x}$ with $\mu = p$ $\sigma^2 = pq$

Binomial distribution

$P(X = x) = \dbinom{n}{x} p^x q^{n-x}$, Notation: $B(n,p)$ $\mu = np$ $\sigma^2 = npq$

Geometric (e.g. prob x failures before first success, parameter q)

$P(X = x) = q^x p$ $\mu = \dfrac{q}{p}$ $\sigma^2 = \dfrac{q}{p^2}$ $x = 0,1,2,3,\ldots.$

Poisson

$P(X = x) = \dfrac{e^{-\mu}\mu^x}{x!}$ Mean = variance = μ

The *probability distribution function* (or *cumulative distribution function*, the cdf) is $F(x) = P(X \le x)$. The *probability density function* or pdf is $f(x) = F'(x)$ (the derivative of *F*), so that $F(x) = \int_{-\infty}^{x} f(y)dy$

$$\mu = E(X) = \int x f(x)dx \qquad \sigma^2 = E(X^2) - E(X)^2$$

Uniform: Pdf $\quad f(x) = \begin{cases} 1/(b-a) & a \le x < b \\ 0 & otherwise \end{cases}$

The Exponential distribution: Pdf $\quad f(x) = \begin{cases} \lambda \exp(-\lambda x) & x > 0 \\ 0 & x \le 0 \end{cases}$

and cdf $\quad F(x) = 1 - \exp(-\lambda x)$ when $x > 0$

$\mu = 1/\lambda \qquad \sigma^2 = 1/\lambda^2$

The Normal distribution

$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ for $-\infty < x < \infty$

μ is the mean and σ is the standard deviation

The *standard normal distribution* has $\mu = 0$ and $\sigma = 1$: $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$

If a random variable *X* follows a $N(\mu, \sigma^2)$ distribution, then the random variable $(X - \mu)/\sigma$ follows a standard normal distribution *N*(0,1), often denoted $\phi(x)$. The cdf of the standard normal distribution is often denoted $\Phi(x)$.

The area between two points *a* and *b* under a normal curve $N(\mu, \sigma^2)$ is the same as the area under a N(0,1) curve between points $(a-\mu)/\sigma$ and $(b-\mu)/\sigma$.

Joint, marginal, and conditional densities

$$f(x,y) \qquad f_X(x) = \int f(x,y)dy \qquad f(y \mid x) = \frac{f(x,y)}{f_X(x)}$$

Given a random sample $x_1, \ldots, x_n$ from a distribution $p(x;\theta)$, the likelihood function for θ is $L(\theta) = \prod_{i=1}^{n} p(x_i;\theta)$. A 95% confidence interval for the mean μ of a distribution is approximately given by $\bar{x} \pm 1.96 \times s/\sqrt{n}$.

If *T* is a random variable denoting the lifetime of a component, with pdf $f(t)$ and cdf $F(t)$ the *survivor function* or *reliability function* is $R(t) = 1 - F(t)$ and the hazard function is $r(t) = f(t)/R(t)$. $\quad R(t) = \exp\left[-\int_0^t r(s)ds\right]$

The standard normal tables gives values of $\Phi(x) = F(x)$ for a N(0,1) distribution:

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
.0	.5	.9	.816	1.8	.964	2.8	.997
.1	.540	1.0	.841	1.9	.971	3.0	.998
.2	.579	1.1	.864	2.0	.977	3.5	.9998
.3	.618	1.2	.885	2.1	.982	1.282	.9
.4	.655	1.3	.903	2.2	.986	1.645	.95
.5	.691	1.4	.919	2.3	.989	1.96	.975
.6	.726	1.5	.933	2.4	.992	2.326	.99
.7	.758	1.6	.945	2.5	.994	2.576	.995
.8	.788	1.7	.955	2.6	.995	3.09	.999

The chi-squared table gives the values of x for which $\chi^2(k)$ has $P(X > x) = p$, where $\chi^2(k)$ is the chi-squared distribution with k degrees of freedom.

k	**.995**	**.975**	**.05**	**.025**	**.01**	***k***	**.995**	**.975**	**.05**	**.025**	**.01**
1	.000	.001	3.84	5.02	6.63	**18**	6.26	8.23	28.87	31.53	34.81
2	.010	.051	5.99	7.38	9.21	**20**	7.43	9.59	31.42	34.17	37.57
3	.072	.216	7.81	9.35	11.34	**22**	8.64	10.98	33.92	36.78	40.29
4	.207	.484	9.49	11.14	13.28	**24**	9.89	12.40	36.42	39.36	42.98
5	.412	.831	11.07	12.83	15.09	**26**	11.16	13.84	38.89	41.92	45.64
6	.676	1.24	12.59	14.45	16.81	**28**	12.46	15.31	41.34	44.46	48.28
7	.990	1.69	14.07	16.01	18.48	**30**	13.79	16.79	43.77	46.98	50.89
8	1.34	2.18	15.51	17.53	20.09	**40**	20.71	24.43	55.76	59.34	63.69
9	1.73	2.70	16.92	19.02	21.67	**50**	27.99	32.36	67.50	71.41	76.15
10	2.16	3.25	13.31	20.48	23.21	**60**	35.53	40.48	79.08	83.30	88.38
12	3.07	4.40	21.03	23.34	26.22	**70**	43.28	48.76	90.53	95.02	100.4
14	4.07	5.63	23.68	26.12	29.14	**80**	51.17	57.15	101.9	106.6	112.3
16	5.14	6.91	26.30	28.85	32.00	**100**	67.33	74.22	124.3	129.6	135.8

The Student's t table gives the values of x for which $t(v)$ has $P(|X| > x) = p$, where $t(v)$ is the Student t distribution with v degrees of freedom.

v	**.10**	**.05**	**.02**	**.01**	***v***	**.10**	**.05**	**.02**	**.01**
1	6.31	12.71	31.82	63.66	**9**	1.83	2.26	2.82	3.25
2	2.92	4.30	6.96	9.92	**10**	1.81	2.23	2.76	3.17
3	2.35	3.18	4.54	5.84	**12**	1.78	2.18	2.68	3.05
4	2.13	2.78	3.75	4.60	**15**	1.75	2.13	2.60	2.95
5	2.02	2.57	3.36	4.03	**20**	1.72	2.09	2.53	2.85
6	1.94	2.45	3.14	3.71	**25**	1.71	2.06	2.48	2.78
7	1.89	2.36	3.00	3.50	**40**	1.68	2.02	2.42	2.70
8	1.86	2.31	2.90	3.36	**∞**	1.645	1.96	2.326	2.576