All questions may be attempted but only marks obtained on the best **four** *solutions will count.*
The use of an electronic calculator **is** *permitted in this examination.*

New Cambridge Statistical Tables are provided.

1. (a) Let Ω be a sample space, and let E and F be any two events (i.e. $E \subseteq \Omega$, $F \subseteq \Omega$). In addition, let P be a probability function mapping $P : \Omega \to \Re$. State the three axioms that P must satisfy.

   (b) Consider an experiment with sample space Ω. Give definitions of:

   (i) Random variable.

   (ii) Continuous random variable.

   In doing so, explain clearly any notation or concepts used in your definitions.

   (c) Let X be a random variable with distribution function $F(\cdot)$. Using the axioms of probability, show that if $a < b$ then $P(a < X \leqslant b) = F(b) - F(a)$. Furthermore, deduce that $F(\cdot)$ is a non-decreasing function.

   (d) Let $\{E_j\}$ be a partition of Ω, and F be any event with $P(F) > 0$. State and prove Bayes' Theorem explaining clearly any properties used in your proof.
   [You may assume that for any collection $\{E_j\}$ of mutually disjoint events $P(\bigcup_j E_j = \sum_j E_j)$]

2. (a) State the conditions required for a function f to be the probability density function of a continuous random variable.

   (b) State the conditions required for a function F to be the distribution function of a continuous random variable.

   (c) A continuous random variable X has probability density function

$$f(x) = abx^{b-1}e^{-ax^b} \quad (x > 0)$$

   for some parameters $a > 0$, $b > 0$. Show that the corresponding distribution function is

$$F(x) = 1 - e^{-ax^b} \quad (x > 0).$$

   Also, show that $Y = X^b$ has an exponential distribution. What is the parameter of this distribution?

3. (a) Write down the probability density function for a uniform distribution with parameters 0 and 1, and find the expression for the corresponding distribution function.

   (b) Suppose U has a uniform distribution with parameters 0 and 1, and let $X = -\log(U^2)$. Show that X has an exponential distribution, provide the parameter of this distribution, and derive its moment generating function (MGF).

   (c) Suppose that $U_1, \ldots, U_n$ are independent random variables, each distributed as $U(0,1)$. Write down an expression for the MGF of $S = -\sum_{i=1}^{n} \log(U_i^2)$.

   (d) It can be shown that the MGF of a chi-squared distribution with m degrees of freedom is $M(t) = (1-2t)^{-m/2}$. Deduce that in part (c) S has a chi-squared distribution and provide the degrees of freedom of this distribution.

   (e) Use the tables provided to evaluate $P\left(\prod_{i=1}^{6} U_i > 0.1\right)$, where $U_1, \ldots, U_6$ are independent random variables each distributed as $U(0,1)$.

   (You may use, without proof, the result that the MGF of a sum of independent random variables is the product of their individual MGFs).

4. (a) Let X be a Poisson random variable with mean m. Show that $E\left[e^{aX}\right] = \exp\left[m\left(e^a - 1\right)\right]$, for any constant a.

   (b) Animals are prone to parasite infection. Suppose the number of parasites living on each animal follows a Poisson distribution with mean μ. If an animal has one or more parasites, it is said to be 'infected'. A farmer wants to know the proportion, p, of infected animals in the flock. He takes n animals from the flock and counts the number of parasites on each of them. In doing so, he chooses the animals in such a way that the counts can be regarded as independent.

      (i) Let $X_1, \ldots, X_n$ denote the parasite counts, and let $S_n = \sum_{i=1}^{n} X_i$. The farmer notices that $p = 1 - e^{-\mu}$ and that $n^{-1}S_n$ is an unbiased estimator of μ. He therefore proposes to use $T = 1 - \exp\left[-n^{-1}S_n\right]$ as an estimator of p. By stating the distribution of S_n and applying the result from part (a), or otherwise, show that T is a biased estimator of p. Also, show that as $n \to \infty$, $E(T) \to p$.

      (ii) Let Y be the total number of infected animals in the sample. State the distribution of Y, and show that Y/n is an unbiased estimator of p. State the standard error of this estimator (in terms of p).

      (iii) If you were asked to choose between the estimators T and Y/n, what considerations would influence your decision? If necessary, state clearly any additional calculations that would be required (but do not attempt to carry out any such calculations).

5. (a) Let $\{X_1, \ldots, X_m\}$ be a collection of random variables, each distributed as $N(\mu_1, \sigma_1^2)$. Let $\{Y_1, \ldots, Y_n\}$ be another such collection, independent of the $\{X_i\}$ and distributed as $N(\mu_2, \sigma_2^2)$. Let $\overline{X}$ and $\overline{Y}$ denote the sample means of the $\{X_i\}$ and $\{Y_i\}$ respectively.

   (i) State the distribution of $\overline{X}$.

   (ii) State the distribution of $\overline{X} - \overline{Y}$.

   (iii) Define a new random variable Z, which is a transformation of $\overline{X} - \overline{Y}$, so that $Z \sim N(0, 1)$. (You should give an expression for Z in terms of $\overline{X} - \overline{Y}$, μ_1, μ_2, σ_1^2 and σ_2^2.)

   (b) In a company, a study was conducted to examine workers' performance before and after a training course. Performance was measured in a sample of 11 workers before the training, and again in an independent sample of 9 workers afterwards. Some summary statistics from the two samples were as follows:

	Before training	After training
Sample mean	0.71	1.06
Sample variance	0.039	0.023

   Assume that the recorded performances in the two samples follow normal distributions.

   (i) Test, at the 95% level and using a 2-tailed test, the hypothesis that the underlying variances in the two groups are the same.

   (ii) Test, at the 99% level and using a 2-tailed test, the hypothesis that the underlying means in the two groups are the same. For this test you should assume that the variances in the two groups are equal, regardless of the result from part (i).

   (iii) Explain clearly what the test results tell you about the effect of the training course.

6. A sack of vegetables is weighed twice, on separate machines. Each machine is subject to random measurement error which is normally distributed with mean zero. Errors from the two machines are independent and have standard deviations σ_1 and σ_2, respectively. Let X_1 and X_2 denote the measurements from the two machines. It is proposed to combine the two measurements using a weighted average

$$\overline{X} = aX_1 + bX_2 \ .$$

(a) Show that for the expected value of $\overline{X}$ to be equal to the true weight of the sack, we need $b = 1 - a$.

(b) For the case $b = 1 - a$, find an expression for the variance of $\overline{X}$ in terms of a, σ_1 and σ_2. Deduce that the value of a that minimises the variance of $\overline{X}$ in this case is $\sigma_2^2/\left(\sigma_1^2 + \sigma_2^2\right)$. Explain why it is desirable that $\overline{X}$ should have a small variance.

(c) Suppose that the actual weight of a sack of vegetables is 50kg, and that the two machines have error standard deviations of 2kg and 1kg respectively. If the measurements are combined according to the scheme above so as to minimise the variance of the combined measurement, then

(i) State the distribution of $\overline{X}$ and its parameter values.

(ii) Find the probability that the measured weight will be under 48kg.

(iii) Find the probability that $\overline{X}$ will be within 1kg of the correct weight.