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.

Candidates are reminded of the following:

- If $X \sim N(\mu, \sigma^2)$ then X can be written as $X = \mu + \sigma Z$, where $Z \sim N(0,1)$.
- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, and $\Gamma(1/2) = \sqrt{\pi}$.
- For $|x| < 1$ and $r > 0$, $(1-x)^{-r} = \sum_{\ell=0}^{\infty} \binom{r+\ell-1}{\ell} x^\ell$.

1. (a) Give the definition of the moment generating function, $M(t)$ say, of a random variable X. Show that $M'(0) = \mathrm{E}[X]$ and that $M''(0) = \mathrm{E}[X^2]$, where $M'(\cdot)$ and $M''(\cdot)$ denote the first and second derivatives of the moment generating function.

   (b) Show that the moment generating function of the $N(0,1)$ distribution is $M(t) = e^{t^2/2}$. Hence, or otherwise, show that the moment generating function of the $N(\mu, \sigma^2)$ distribution is $M(t) = e^{\mu t + \sigma^2 t^2/2}$.

   (c) Suppose that $X \sim N(\mu, \sigma^2)$, and that $Y = e^X$. Find the mean and variance of Y. If you wanted to choose values of μ and σ^2 such that the variance of Y was extremely small, what options would be available to you?

2. (a) A point is chosen at random from the interval (0,1), thereby dividing the interval into two parts. Let U be the length of the longer part, so that $1 - U$ is the length of the shorter part.

   (i) Name the distribution of U, giving the values of its parameters.

   (ii) Let $Y = (1 - U)/U$ be the ratio of the shorter part to the longer part. Find the probability density function of Y.

   (b) The speed of a molecule in a uniform gas in equilibrium is a random variable V with density

   $$f_V(v) = av^2 \exp(-bv^2), \quad (v > 0)\ ,$$

   where a and b are constants.

   (i) The kinetic energy of the molecule is defined as $W = \frac{mV^2}{2}$, where m is the mass of the molecule. Find the probability density function of W, and show that this corresponds to a gamma distribution. Give the parameters of this distribution in terms of b and m.

   (ii) Use this result to show that $a = 4b^{3/2}/\sqrt{\pi}$.

3. (a) $X_1, \ldots, X_n$ are mutually independent discrete random variables, each taking non-negative integer values. For $i = 1, \ldots, n$, denote by $\Pi_i(z)$ the probability generating function of X_i. Define $S = \sum_{i=1}^{n} X_i$, and let $\Pi_S(z)$ be the probability generating function of S. Show that $\Pi_S(z) = \prod_{i=1}^{n} \Pi_i(z)$. Explain your reasoning carefully.

   (b) Show that the probability generating function of a geometric distribution with parameter p is $\Pi(z) = pz\,[1-(1-p)z]^{-1}$.

   (c) Suppose that a random variable Y has a negative binomial distribution with parameters r and p.

      (i) By considering how the negative binomial distribution arises in the context of sequences of Bernoulli trials, explain carefully why Y can be regarded as a sum of independent geometrically distributed random variables.

      (ii) Write down an expression for the probability generating function of Y.

      (iii) Use your answer from part (c)(ii) to find the probability mass function of the $NB(r, p)$ distribution.

4. (a) Suppose that X_1 and X_2 are discrete random variables, with means μ_1 and μ_2 respectively.

      (i) What is meant by saying that X_1 and X_2 are independent? Define carefully any additional notation that you use.

      (ii) Show that if X_1 and X_2 are independent then $E(X_1X_2) = \mu_1\mu_2$.

   (b) X and Y are two continuous random variables with joint density

$$f(x,y) = \begin{cases} K(x+y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

      (i) Find the value of K.

      (ii) Without carrying out any calculations, state whether or not X and Y are independent. Explain your reasoning clearly.

      (iii) Find the marginal densities of X and Y.

      (iv) Find the conditional density of Y given $X = x$.

      (v) For this example, verify the identity $E(Y) = E_X\,[E(Y|X)]$. What is the name of this identity?

5. (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) Sheep are prone to parasite infection. Suppose the number of parasites living on each animal follows a Poisson distribution with mean μ. If the animal has one or more parasites, it is said to be 'infected'. A farmer wants to know the proportion, p, of her flock that is infected. She takes n animals from the flock and counts the number of parasites on each of them. She 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 estinator of μ. She 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).

6. The following data are recorded weight increases (in grams) of babies over a period of one week. The first group of 10 babies is used as control, and the second group of 14 babies has received an anti-allergic treatment.

Control:	-60	-45	-30	0	0	0	10	10	10	35	75			
Treated:	-75	-75	-50	-50	-50	-45	-30	-30	-25	-10	0	10	25	100

(a) Assume that the two groups of measurements can be regarded as independent samples from populations with means μ_1 and μ_2, and variances σ_1^2 and σ_2^2, respectively. Calculate estimates of μ_1, μ_2, σ_1^2 and σ_2^2.

(b) Test, at the 5% level, the hypothesis that the $\sigma_1^2 = \sigma_2^2$. State your conclusions clearly.

(c) Assuming that $\sigma_1^2 = \sigma_2^2$, calculate a 95% confidence interval for $\mu_1 - \mu_2$.

(d) What do these analyses tell you about the effectiveness of the new treatment? State any additional assumptions you have made during the analyses, and comment on their validity.