HL Paper 3

The weight of tea in *Supermug* tea bags has a normal distribution with mean 4.2 g and standard deviation 0.15 g. The weight of tea in *Megamug* tea bags has a normal distribution with mean 5.6 g and standard deviation 0.17 g.

a. Find the probability that a randomly chosen *Supermug* tea bag contains more than 3.9 g of tea. [2]

b. Find the probability that, of two randomly chosen *Megamug* tea bags, one contains more than 5.4 g of tea and one contains less than 5.4 g of tea. [4]

c. Find the probability that five randomly chosen *Supermug* tea bags contain a total of less than 20.5 g of tea. [4]

d. Find the probability that the total weight of tea in seven randomly chosen *Supermug* tea bags is more than the total weight in five randomly chosen *Megamug* tea bags. [5]

Two species of plant, A and B, are identical in appearance though it is known that the mean length of leaves from a plant of species A is 5.2 cm, whereas the mean length of leaves from a plant of species B is 4.6 cm. Both lengths can be modelled by normal distributions with standard deviation 1.2 cm.

In order to test whether a particular plant is from species A or species B, 16 leaves are collected at random from the plant. The length, x, of each leaf is measured and the mean length evaluated. A one-tailed test of the sample mean, $\bar{X}$, is then performed at the 5% level, with the hypotheses: $H_0 : \mu = 5.2$ and $H_1 : \mu < 5.2$.

a. Find the critical region for this test. [3]

c. It is now known that in the area in which the plant was found 90% of all the plants are of species A and 10% are of species B. [2]

Find the probability that $\bar{X}$ will fall within the critical region of the test.

d. If, having done the test, the sample mean is found to lie within the critical region, find the probability that the leaves came from a plant of species A. [3]

The random variable $X \sim \text{Po}(m)$. Given that $\mathrm{P}(X = k - 1) = \mathrm{P}(X = k + 1)$, where k is a positive integer,

a. show that $m^2 = k(k+1)$; [2]

b. hence show that the mode of X is k . [6]

A traffic radar records the speed, v kilometres per hour (km h^{-1}), of cars on a section of a road.

The following table shows a summary of the results for a random sample of 1000 cars whose speeds were recorded on a given day.

Speed	Number of cars
$50 \le v < 60$	5
$60 \le v < 70$	13
$70 \le v < 80$	52
$80 \le v < 90$	68
$90 \le v < 100$	98
$100 \le v < 110$	105
$110 \le v < 120$	289
$120 \le v < 130$	142
$130 \le v < 140$	197
$140 \le v < 150$	31

a. Using the data in the table, [4]

(i) show that an estimate of the mean speed of the sample is $113.21\ \text{km h}^{-1}$;

(ii) find an estimate of the variance of the speed of the cars on this section of the road.

b. Find the 95% confidence interval, I, for the mean speed. [2]

c. Let J be the 90% confidence interval for the mean speed. [2]

Without calculating J, explain why $J \subset I$.

The random variable X has the distribution $\mathrm{B}(n, p)$.

(a) (i) Show that $\frac{\mathrm{P}(X=x)}{\mathrm{P}(X=x-1)} = \frac{(n-x+1)p}{x(1-p)}$.

(ii) Deduce that if $\mathrm{P}(X = x) > \mathrm{P}(X = x - 1)$ then $x < (n + 1)p$.

(iii) Hence, determine the value of x which maximizes $\mathrm{P}(X = x)$ when $(n + 1)p$ is not an integer.

(b) Given that $n = 19$, find the set of values of p for which X has a unique mode of 13.

Each week the management of a football club recorded the number of injuries suffered by their playing staff in that week. The results for a 52-week period were as follows:

Number of injuries per week	0	1	2	3	4	5	6
Number of weeks	6	14	15	9	5	2	1

a. Calculate the mean and variance of the number of injuries per week. [2]

b. Explain why these values provide supporting evidence for using a Poisson distribution model. [1]

If X is a random variable that follows a Poisson distribution with mean $\lambda > 0$ then the probability generating function of X is $G(t) = e^{\lambda(t-1)}$.

a. (i) Prove that $\mathrm{E}(X) = \lambda$. [6]

(ii) Prove that $\mathrm{Var}(X) = \lambda$.

b. Y is a random variable, independent of X, that also follows a Poisson distribution with mean λ. [3]

If $S = 2X - Y$ find

(i) $\mathrm{E}(S)$;

(ii) $\mathrm{Var}(S)$.

c. Let $T = \frac{Y}{2} + \frac{Y}{2}$. [3]

(i) Show that T is an unbiased estimator for λ.

(ii) Show that T is a more efficient unbiased estimator of λ than S.

d. Could either S or T model a Poisson distribution? Justify your answer. [1]

e. By consideration of the probability generating function, $G_{X+Y}(t)$, of $X + Y$, prove that $X + Y$ follows a Poisson distribution with mean 2λ. [3]

f. Find [2]

(i) $G_{X+Y}(1)$;

(ii) $G_{X+Y}(-1)$.

g. Hence find the probability that $X + Y$ is an even number. [3]

Engine oil is sold in cans of two capacities, large and small. The amount, in millilitres, in each can, is normally distributed according to Large $\sim \mathrm{N}(5000,\ 40)$ and Small $\sim \mathrm{N}(1000,\ 25)$.

a. A large can is selected at random. Find the probability that the can contains at least 4995 millilitres of oil. [2]

b. A large can and a small can are selected at random. Find the probability that the large can contains at least 30 milliliters more than five times the amount contained in the small can. [6]

c. A large can and five small cans are selected at random. Find the probability that the large can contains at least 30 milliliters less than the total amount contained in the small cans. [5]

When Andrew throws a dart at a target, the probability that he hits it is $\frac{1}{3}$; when Bill throws a dart at the target, the probability that he hits the it is $\frac{1}{4}$. Successive throws are independent. One evening, they throw darts at the target alternately, starting with Andrew, and stopping as soon as one of their darts hits the target. Let X denote the total number of darts thrown.

a. Write down the value of $\mathrm{P}(X=1)$ and show that $\mathrm{P}(X=2)=\frac{1}{6}$. [2]

b. Show that the probability generating function for X is given by [6]

$$G(t)=\frac{2t+t^2}{6-3t^2}.$$

c. Hence determine $\mathrm{E}(X)$. [4]

The weights of the oranges produced by a farm may be assumed to be normally distributed with mean 205 grams and standard deviation 10 grams.

a. Find the probability that a randomly chosen orange weighs more than 200 grams. [2]

b. Five of these oranges are selected at random to be put into a bag. Find the probability that the combined weight of the five oranges is less than 1 kilogram. [4]

c. The farm also produces lemons whose weights may be assumed to be normally distributed with mean 75 grams and standard deviation 3 grams. Find the probability that the weight of a randomly chosen orange is more than three times the weight of a randomly chosen lemon. [5]

---

a. Determine the probability generating function for $X \sim \mathrm{B}(1,\ p)$. [4]

b. Explain why the probability generating function for $\mathrm{B}(n,\ p)$ is a polynomial of degree n. [2]

c. Two independent random variables X_1 and X_2 are such that $X_1 \sim \mathrm{B}(1,\ p_1)$ and $X_2 \sim \mathrm{B}(1,\ p_2)$. Prove that if X_1+X_2 has a binomial distribution then $p_1=p_2$. [5]

The random variable X is assumed to have probability density function f, where

$$f(x)=\begin{cases}\frac{x}{18}, & 0 \leqslant x \leqslant 6\\ 0, & \text{otherwise.}\end{cases}$$

Show that if the assumption is correct, then

$$\mathrm{P}(a \leqslant X \leqslant b)=\frac{b^2-a^2}{36},\ \text{for } 0 \leqslant a \leqslant b \leqslant 6.$$

The discrete random variable X has the following probability distribution, where $0 < \theta < \frac{1}{3}$.

x	1	2	3
$P(X = x)$	θ	2θ	$1-3\theta$

a. Determine $\mathrm{E}(X)$ and show that $\mathrm{Var}(X) = 6\theta - 16\theta^2$. [4]

b. In order to estimate θ, a random sample of n observations is obtained from the distribution of X. [10]

(i) Given that $\bar{X}$ denotes the mean of this sample, show that

$$\hat{\theta}_1 = \frac{3 - \bar{X}}{4}$$

is an unbiased estimator for θ and write down an expression for the variance of $\hat{\theta}_1$ in terms of n and θ.

(ii) Let Y denote the number of observations that are equal to 1 in the sample. Show that Y has the binomial distribution $\mathrm{B}(n, \theta)$ and deduce that $\hat{\theta}_2 = \frac{Y}{n}$ is another unbiased estimator for θ. Obtain an expression for the variance of $\hat{\theta}_2$.

(iii) Show that $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$ and state, with a reason, which is the more efficient estimator, $\hat{\theta}_1$ or $\hat{\theta}_2$.

Bill also has a box with 10 biscuits in it. 4 biscuits are chocolate and 6 are plain. Bill takes a biscuit from his box at random, looks at it and replaces it in the box. He repeats this process until he has looked at 5 biscuits in total. Let B be the number of chocolate biscuits that Bill takes and looks at.

d. State the distribution of B . [1]

e. Find $P(B = 3)$. [2]

f. Find $P(B = 5)$. [2]

The random variable X has probability distribution Po(8).

a. (i) Find $\mathrm{P}(X = 6)$. [5]

(ii) Find $\mathrm{P}(X = 6 | 5 \leqslant X \leqslant 8)$.

b. $\bar{X}$ denotes the sample mean of $n > 1$ independent observations from X. [3]

(i) Write down $\mathrm{E}(\bar{X})$ and $\mathrm{Var}(\bar{X})$.

(ii) Hence, give a reason why $\bar{X}$ is not a Poisson distribution.

c. A random sample of 40 observations is taken from the distribution for X. [6]

(i) Find $\mathrm{P}(7.1 < \bar{X} < 8.5)$.

(ii) Given that $\mathrm{P}\left(\left|\bar{X} - 8\right| \leqslant k\right) = 0.95$, find the value of k.

The continuous random variable X has probability density function f given by

$$f(x)=\begin{cases} 2x, & 0 \leqslant x \leqslant 0.5, \\ \frac{4}{3}-\frac{2}{3}x, & 0.5 \leqslant x \leqslant 2 \\ 0, & \text{otherwise.} \end{cases}$$

a. Sketch the function f and show that the lower quartile is 0.5. [3]

b. (i) Determine $\mathrm{E}(X)$. [4]

(ii) Determine $\mathrm{E}(X^2)$.

c. Two independent observations are made from X and the values are added. [5]

The resulting random variable is denoted Y.

(i) Determine $\mathrm{E}(Y-2X)$.

(ii) Determine $\mathrm{Var}\,(Y-2X)$.

d. (i) Find the cumulative distribution function for X. [7]

(ii) Hence, or otherwise, find the median of the distribution.

A random variable X has probability density function

$$f(x)=\begin{cases} 0 & x<0 \\ \frac{1}{2} & 0 \leq x<1 \\ \frac{1}{4} & 1 \leq x<3 \\ 0 & x \geq 3 \end{cases}$$

a. Sketch the graph of $y=f(x)$. [1]

b. Find the cumulative distribution function for X. [5]

c. Find the interquartile range for X. [3]