STEP 3 – Combinations and Functions of Random Variables (Sheet 2)

Independent random variables

Understand and use the idea of independent random variables.

Algebra of expectation

Know, understand and use the algebra of expectation:

$\mathrm{E}(aX + bY + c) = a\mathrm{E}(X) + b\mathrm{E}(Y) + c$

$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$

$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

and for independent random variables:

$\mathrm{Var}(aX \pm bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$

Use cumulative distribution functions to calculate the probability density function of a related random variable; for example, X^2 from X.

Knowledge of generating functions will not be required.

Q1, (STEP III, 2005, Q12)

Five independent timers time a runner as she runs four laps of a track. Four of the timers measure the individual lap times, the results of the measurements being the random variables T_1 to T_4, each of which has variance σ^2 and expectation equal to the true time for the lap. The fifth timer measures the total time for the race, the result of the measurement being the random variable T which has variance σ^2 and expectation equal to the true race time (which is equal to the sum of the four true lap times).

Find a random variable X of the form $aT + b(T_1 + T_2 + T_3 + T_4)$, where a and b are constants independent of the true lap times, with the properties that:

(i) whatever the true lap times, the expectation of X is equal to the true race time;

(ii) the variance of X is as small as possible.

Find also a random variable Y of the form $cT + d(T_1 + T_2 + T_3 + T_4)$, where c and d are constants independent of the true lap times, with the property that, whatever the true lap times, the expectation of Y^2 is equal to σ^2.

In one particular race, T takes the value 220 seconds and $(T_1 + T_2 + T_3 + T_4)$ takes the value 220.5 seconds. Use the random variables X and Y to estimate an interval in which true race time lies.

Q2, (STEP III, 2006, Q14)

For any random variables X_1 and X_2, state the relationship between $\mathrm{E}(aX_1+bX_2)$ and $\mathrm{E}(X_1)$ and $\mathrm{E}(X_2)$, where a and b are constants. If X_1 and X_2 are independent, state the relationship between $\mathrm{E}(X_1X_2)$ and $\mathrm{E}(X_1)$ and $\mathrm{E}(X_2)$.

An industrial process produces rectangular plates. The length and the breadth of the plates are modelled by independent random variables X_1 and X_2 with non-zero means μ_1 and μ_2 and non-zero standard deviations σ_1 and σ_2, respectively. Using the results in the paragraph above, and without quoting a formula for $\mathrm{Var}\,(aX_1+bX_2)$, find the means and standard deviations of the perimeter P and area A of the plates. Show that P and A are not independent.

The random variable Z is defined by $Z = P - \alpha A$, where α is a constant. Show that Z and A are not independent if

$$\alpha \neq \frac{2(\mu_1\sigma_2^2 + \mu_2\sigma_1^2)}{\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2}\,.$$

Given that X_1 and X_2 can each take values 1 and 3 only, and that they each take these values with probability $\frac{1}{2}$, show that Z and A are not independent for any value of α.

Q3, (STEP III, 2007, Q13)

A frog jumps towards a large pond. Each jump takes the frog either 1 m or 2 m nearer to the pond. The probability of a 1 m jump is p and the probability of a 2 m jump is q, where $p+q=1$, the occurence of long and short jumps being independent.

(i) Let $p_n(j)$ be the probability that the frog, starting at a point $(n-\frac{1}{2})$ m away from the edge of the pond, lands in the pond for the first time on its jth jump. Show that $p_2(2) = p$.

(ii) Let u_n be the expected number of jumps, starting at a point $(n-\frac{1}{2})$ m away from the edge of the pond, required to land in the pond for the first time. Write down the value of u_1. By finding first the relevant values of $p_n(m)$, calculate u_2 and show that $u_3 = 3 - 2q + q^2$.

(iii) Given that u_n can be expressed in the form $u_n = A(-q)^{n-1} + B + Cn$, where A, B and C are constants (independent of n), show that $C = (1+q)^{-1}$ and find A and B in terms of q. Hence show that, for large n, $u_n \approx \dfrac{n}{p+2q}$ and explain carefully why this result is to be expected.

Q4, (STEP III, 2008, Q13)

A box contains n pieces of string, each of which has two ends. I select two string ends at random and tie them together. This creates either a ring (if the two ends are from the same string) or a longer piece of string. I repeat the process of tying together string ends chosen at random until there are none left.

Find the expected number of rings created at the first step and hence obtain an expression for the expected number of rings created by the end of the process. Find also an expression for the variance of the number of rings created.

Given that $\ln 20 \approx 3$ and that $1 + \frac{1}{2} + \cdots + \frac{1}{n} \approx \ln n$ for large n, determine approximately the expected number of rings created in the case $n = 40\,000$.

Q5, (STEP III, 2009, Q12)

(i) Albert tosses a fair coin k times, where k is a given positive integer. The number of heads he gets is X_1. He then tosses the coin X_1 times, getting X_2 heads. He then tosses the coin X_2 times, getting X_3 heads. The random variables X_4, X_5, … are defined similarly. Write down $\mathrm{E}(X_1)$.

By considering $\mathrm{E}(X_2 \mid X_1 = x_1)$, or otherwise, show that $\mathrm{E}(X_2) = \frac{1}{4}k$.

Find $\displaystyle\sum_{i=1}^{\infty} \mathrm{E}(X_i)$.

(ii) Bertha has k fair coins. She tosses the first coin until she gets a tail. The number of heads she gets before the first tail is Y_1. She then tosses the second coin until she gets a tail and the number of heads she gets with this coin before the first tail is Y_2. The random variables $Y_3, Y_4, \ldots, Y_k$ are defined similarly, and $Y = \displaystyle\sum_{i=1}^{k} Y_i$.

Obtain the probability generating function of Y, and use it to find $\mathrm{E}(Y)$, $\mathrm{Var}\,(Y)$ and $\mathrm{P}(Y = r)$.

Q6, (STEP III, 2013, Q12)

A list consists only of letters A and B arranged in a row. In the list, there are a letter As and b letter Bs, where $a \geqslant 2$ and $b \geqslant 2$, and $a + b = n$. Each possible ordering of the letters is equally probable. The random variable X_1 is defined by

$$X_1 = \begin{cases} 1 & \text{if the first letter in the row is } A; \\ 0 & \text{otherwise.} \end{cases}$$

The random variables X_k $(2 \leqslant k \leqslant n)$ are defined by

$$X_k = \begin{cases} 1 & \text{if the } (k-1)\text{th letter is } B \text{ and the } k\text{th is } A; \\ 0 & \text{otherwise.} \end{cases}$$

The random variable S is defined by $S = \sum\limits_{i=1}^{n} X_i$.

(i) Find expressions for $\mathrm{E}(X_i)$, distinguishing between the cases $i = 1$ and $i \neq 1$, and show that $\mathrm{E}(S) = \dfrac{a(b+1)}{n}$.

(ii) Show that:

(a) for $j \geqslant 3$, $\mathrm{E}(X_1 X_j) = \dfrac{a(a-1)b}{n(n-1)(n-2)}$;

(b) $\sum\limits_{i=2}^{n-2} \Big(\sum\limits_{j=i+2}^{n} \mathrm{E}(X_i X_j) \Big) = \dfrac{a(a-1)b(b-1)}{2n(n-1)}$;

(c) $\mathrm{Var}\,(S) = \dfrac{a(a-1)b(b+1)}{n^2(n-1)}$.