Problem Sheet 3

This problem sheet consists of two questions. Each question contains three parts.

- Parts a), worth 40% of the marks, in a question, test your knowledge of the core material, and you should aim to provide good answers to this part of all questions.
- Parts b), worth 30% of the marks, involve taking the concepts you have been taught, but applying them in ways you may not have directly been shown. You should attempt all the parts b), but you can still get a good mark without completing all of them.
- Parts c), worth 30% of the marks, are difficult questions which will test your understanding of the concepts taught in unfamiliar situations.

Unless indicated otherwise, you should show your working for all questions.

Question 1 (50%):

a) We have a bag that contains 2 red balls, 3 blue balls and 4 green balls. For our first experiment, we take a random ball from the bag, observing its colour, return the ball to the bag, and then take a second ball at random, observing its colour.

   i) Describe the sample space and event space for this experiment. You do not need to explicitly list every element of the event space.

   ii) Let BG denote the event in which the first ball drawn is blue, and the second is green. Calculate $P(BG)$.

   iii) Let R denote the event in which either the first or second ball drawn is red, including the case where both are red. Calculate $P(R)$.

   In a second experiment, we keep the same bag, but instead of returning the first ball to the bag, we do not replace it before drawing the second.

   iv) Explain how the probability space for this new experiment differs to that for the first experiment.

   v) Let R_1, B_1, G_1 be the events in which, respectively, a red, blue or green ball is drawn first, and G_2 the event in which a green ball is drawn second. Calculate $P(G_2|R_1)$, $P(G_2|B_1)$ and $P(G_2|G_1)$.

   vi) Using your answer from the previous question, calculate $P(G_2)$.

   vii) Using Bayes' Theorem, calculate $P(B_1|G_2)$.

b) Consider an infinite sequence of independent (possibly unfair) coin flips where the probability of each being heads is $p \in [0,1]$. We define the family of random variables X_k to be the number of heads after k flips.

i) Using the fact that X_n is binomially distributed with $X_n \sim B(n,p)$, calculate $P(X_5 = 2)$.

ii) Let $k > 0$ be even, and $p = \frac{1}{2}$. Calculate $P(X_k > \frac{k}{2})$. Hint: you might want to also consider $P(X_k < \frac{k}{2})$.

We now define the random variable Y as the length of the run of either all heads or all tails at the start of an infinite series of coin flips. For example $Y = 4$ if the sequence of coin flips start with HHHHT... or TTTTH....

iii) Give an expression of $P(Y = k)$ in terms of p and k.

iv) Using your answer to the previous question, show that

$$E(Y) = \frac{1 - 2p + 2p^2}{p(1-p)}$$

You will probably want to use the fact that:

$$\sum_{i=1}^{\infty} i a^{i-1} = \frac{1}{(a-1)^2}$$

c) Let X and Y be two discrete random variables. We define $Z = X + Y$, i.e. $\forall \omega \in \Omega, Z(\omega) = X(\omega) + Y(\omega)$.

i) Show that:

$$P(Z = z) = \sum_x f_{X,Y}(x, z - x)$$

ii) Now assume that X and Y are independent. Show that:

$$P(Z = z) = \sum_x f_X(x) f_Y(z - x) = \sum_y f_X(z - y) f_Y(y)$$

From now on, we assume that X and Y are independent random variables which have the Poisson distributions with parameters λ_X and λ_Y, respectively.

iii) Show that Z has the Poisson distribution, with parameter $\lambda_X + \lambda_Y$.

iv) Show that the conditional distribution of X, given $X + Y = n$, is binomial, and find its parameters.

Question 2 (50%):

a) You and your team have built a chess computer. To investigate how well it performs, you organise a tournament with 100 randomly selected grandmasters (a title given to expert chess players by the world chess organization FIDE), where your computer will play each of the grandmasters once, and it is recorded whether the game is a win for the computer, a win for the grandmaster, or a draw. For each win against a grandmaster, the computer scores 1 point, and for each draw the computer scores half a point. At the end of the tournament, the chess computer has scored 60 points.

i) Identify the population, sample and sample size.

ii) Identify a reasonable parameter and parameter space.

iii) Using the results of the tournament, give an estimate of the parameter. Explain whether this is an unbiased estimate.

Subsequently, you and your team go away to improve the chess computer. After a while, you organise another tournament, where the computer will again play 100 randomly selected grandmasters. Two members of your team make claims about the new version of the computer. Alice claims that the new version will perform better against grandmasters than the previous version. Bob claims that the new version will score an average of at least 0.7 points in games against a grandmaster.

iv) Give the null and alternative hypotheses for both Alice and Bob's claims.

v) Using the concept of the critical value, explain how Alice's claim can be tested.

vi) Using the concept of the p-value, explain how Bob's claim can be tested.

Based on the results of the second tournament, you decide there is sufficient evidence for both Alice and Bob's claims to be accepted. However, you later discover that both your tournaments occurred at the same time as the Candidates Tournament (one of the most important chess tournaments), and so, each time, your random selection of grandmasters excluded the best 10%.

vii) With reference to Type I and/or Type II errors, explain how this discovery might change your attitude to each of Alice and Bob's claims.

b) Consider a two-dimension plane in which we mark the lines $y = n$ for $n \in \mathbb{Z}$. We now randomly "drop a needle" (i.e. draw a line segment) of length 1 on the plane: its centre is given by two random co-ordinates (X, Y), and the angle is given (in radians) by a random variable Θ. In this question, we will be concerned with the probability that the needle intersects one of the lines $y = n$. For this purpose, we define the random variable Z as the distance from the needle's centre to the nearest line beneath it (i.e. $Z = Y - \lfloor Y \rfloor$, where $\lfloor Y \rfloor$ is the greatest integer not greater than Y). We assume:

- Z is uniformly distributed on $[0,1]$.
- Θ is uniformly distributed on $[0,\pi]$.
- Z and Θ are independent and jointly continuous.

i) Give the density functions of Z and Θ.

ii) Give the joint density function of Z and Θ (hint: use the fact that Z and Θ are independent).

By geometric reasoning, it can be shown that an intersection occurs if and only if:

$$(z,\theta) \in [0,1] \times [0,\pi] \text{ is such that } z \leq \frac{1}{2}\sin\theta \text{ or } 1-z \leq \frac{1}{2}\sin\theta$$

iii) By using the joint distribution function of Z and Θ, show that:

$$P(\text{The needle intersects a line}) = \frac{2}{\pi}$$

Suppose now that a statistician is able to perform this experiment n times without any bias. Each drop of the needle is described by a random variable X_i which is 1 if the needle intersects a line and 0 otherwise. For any n, we assume the random variables $X_1, \ldots, X_n$ are independent and identically distributed and that the variance of the population is $\sigma^2 < \infty$.

iv) Explain, with reference to the Law of Large Numbers, how the statistician could use this experiment to estimate the value of π with increasing accuracy.

v) Explain what happens to the distribution of $\bar{X}$ as $n \to \infty$.

c) i) Using the fact that $\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi}$, explain how we can verify that the density function for the normal distribution $N\left(0, \left(\frac{1}{\sqrt{2}}\right)^2\right)$ is a valid density function.

ii) Show that if $X \sim N\left(0, \left(\frac{1}{\sqrt{2}}\right)^2\right)$, then $E(X) = 0$.

iii) Given that $\int_{-\infty}^{\infty} x^2 e^{-x^2}dx = \frac{\sqrt{\pi}}{2}$, show that $\text{var}(X) = \frac{1}{2}$.

Let $f: \mathbb{R}^2 \to \mathbb{R}$ be given by:

$$f(x,y) = \frac{e^{\left(-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right)}}{2\pi\sqrt{(1-\rho^2)}}$$

for $-1 < \rho < 1$. This function is the joint distribution function for two normally distributed variables $X, Y \sim N(0,1)$, with $\rho = \text{cov}(X,Y)$.

iv) Explain why, in this instance, the correlation of X and $Y, \rho(X,Y)$, is equal to the covariance of X and Y, $\text{cov}(X,Y)$.

v) Show that if $\rho = 0$, then X and Y are independent.