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.

Solution

$\Rightarrow$ Sample space ,this will be the set of all the possible outcomes which is the same to the

$\Rightarrow$ eventspace in this case

$\Rightarrow S = \{GG, BG, BR, RB, \ldots\}$,this will have a total of 9 possible outcomes

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

Solution

$$\Rightarrow P(BG) = P(B).\,P(G), = P(B) = \frac{3}{9}, P(G) = \frac{4}{9} \Rightarrow \frac{3}{9} * \frac{4}{9} = \frac{4}{27}$$

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 $P(R)$.

Solution

$\Rightarrow$ From sample space,$S = \{BG, GR, GB, GG, BR, BB, RB, RR, RG\}, \Rightarrow P(R) = \dfrac{5}{9}$

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.

Solution

$\Rightarrow$ The probability space for this will be reduced at each event when a ball is drawn and no replacement

$\Rightarrow$ unlike in the first case when the probability space remains constant

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 $P(G_2|G_1)$.

Solution

$$\Rightarrow P(G_2|R_1) = \frac{4}{8} = \frac{1}{2}$$

$$\Rightarrow P(G_2|B_1) = \frac{4}{8} = \frac{1}{2}$$

$$\Rightarrow P(G_2|G_1) = \frac{3}{8}$$

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

Solution

$$\Rightarrow P(G_2) = P(R_1G_2) + P(B_1G_2) + P(G_1G_2) = (\frac{1}{2} * \frac{2}{9}) + (\frac{1}{2} * \frac{3}{9} + (\frac{3}{8} * \frac{4}{9})) = \frac{4}{9}$$

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

Solution

$$\Rightarrow P(B_1|G_2) = \frac{P(G_2B_1)}{P(G_2)} = (\frac{4}{9} * \frac{3}{8})/\frac{4}{9} = \frac{3}{8}$$

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 $P(X_5 = 2)$.

Solution

$$\Rightarrow P(X_5 = 2) = \binom{5}{2}(0.5)^2(0.5)^2 = 0.3125$$

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 $P(X_k < \frac{k}{2})$.

Solution

$$\Rightarrow P(X_k > \frac{k}{2}) = 1 - P(X_k \leq \frac{k}{2})$$

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.

Solution

$$\Rightarrow P(Y = k) = (1 - P)^{k-1}P \,, \forall_k = 1, 2, 3, 4, \ldots$$

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:

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

Solution

$$\Rightarrow \text{From the given case E(Y) becomes } E(Y) = \frac{1}{p} \Rightarrow \frac{1}{(1-p)^{k-1}p}$$

$$\Rightarrow \text{Thus for k=1,we get :}\frac{1}{p(1-p)}$$

$$\Rightarrow \text{using the given relation we get that } : \frac{1}{p(1-p)} / \frac{1}{(1-p)^2} = \frac{(1-p)^2}{p(1-p)} = \frac{1 - 2p + 2p^2}{p(1-p)}, \text{with a=2}$$

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:

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

Solution

$\Rightarrow$ Given that Z=X+Y,we have for;

$$\Rightarrow P(X + Y = z) = \int_{x,y} f_X(x) f_Y(y) dy dx$$

$$\Rightarrow \int_y f_Y(y) dy \int_x f_X(x) dx = \int_x f_X(x) \left[\int_y f_Y(y) dy \right] dx$$

$\Rightarrow$ This simplifies into : $\Rightarrow \int_{x,y} f_X(x)[1 - F_Y(z - x)] dx = \int_{x,y} f_X(x) f_Y(z - x) dx$

$\Rightarrow$ Therefore we have: $\Rightarrow \sum_x f_X(x) f_Y(z - x) = \sum_x f_{XY}(x, z - x)$

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

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

Solution

$\Rightarrow$ From (i) ,we let P=Y ,thus we obtain $= \int_x f_X(x) f_P(z - x) dx = \int_y f_P(y) f_X(z - y) dp$

$$\Rightarrow \int_x f_X(x) f_Y(z - x) dx = \int_y f_X(z - y) f_Y(y) dy = \sum_x f_X(x) f_Y(z - x) = \sum_y f_Y(y) f_X(z - y)$$

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 $\lambda_X + \lambda_Y$.

Solution

$$\Rightarrow P(Z=z) = P(X+Y=z) = \sum P(X=x, Y=y)$$

$$\Rightarrow \sum P(X=x)P(Y=y) = \sum \frac{e^{-\Lambda_X}\Lambda_X^x}{x!} . \frac{e^{-\Lambda_Y}\Lambda_Y^y}{y!}$$

$$\Rightarrow e^{-(\Lambda_X+\Lambda_Y)} \sum \frac{\Lambda_X^x \Lambda_Y^y}{x!y!}$$

$$\Rightarrow P(Z=x) = e^{-(\Lambda_X+\Lambda_Y)} \frac{(\Lambda_X+\Lambda_Y)^2}{z!}$$

$\Rightarrow$ Thus Z has 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.

Solution

$\Rightarrow$ From Z=X+Y, we have that , $\dfrac{P(X=k, Z=n)}{P(Z=n)} = \dfrac{P(X+Y=n, X=k)}{P(Z=n)}$

$$\Rightarrow \frac{P(X+Y=n, X=k)}{P(Z=n)} = \frac{P(Y=n-k)P(X=k)}{P(Z=n)}$$

$\Rightarrow$ We have shown that Z is poison distribution with $\Lambda_X + \Lambda_Y$

$$\Rightarrow \text{Thus we have: } \Rightarrow \frac{e^{-\Lambda_X} e^{\frac{\Lambda_X^k}{k!}} e^{-\Lambda_Y} e^{\frac{\Lambda_Y^{n-\kappa}}{(n-k)!}}}{e^{-(\Lambda_X+\Lambda_Y)} \frac{(\Lambda_X+\Lambda_Y)^n}{n!}}$$

$$\Rightarrow \binom{n}{k} \left(\frac{\Lambda_X}{\Lambda_X+\Lambda_Y}\right)^k \left(\frac{\Lambda_Y}{\Lambda_X+\Lambda_Y}\right)^{n-k}$$

$\Rightarrow$ Hence is binomial distribution with parameters, n and $\dfrac{\Lambda_X}{\Lambda_X+\Lambda_Y}$

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.

Solution

=Population will be the number of all participants which is 100

=Sample is the number of the participants among the 100 who will participate which is equal to the sample size of 100 in this case

ii) Identify a reasonable parameter and parameter space.

=The grandmasters represent the parameter in this case, and parameter space accorded to the computer

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

=From total scored computer points 60 with 2 chess players an estimate will be 60/2 which is 30

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.

=Null hypothesis -the Alice's claim that the new version will perform better than the grandmaster than the previous version

=Alternative hypothesis-the Bob's claim that the new version will not have an average score of 0.7 points against the grandmaster

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

=From Alice's claim, the rejection zone will be the case when the new version of computer won't perform better than the previous version hence can be tested using the old computer version claims to obtain the result

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

=The level of significance in Bob's claim will be tested on how area under the curve from the obtained claim report is to the left or right of the testing statistics

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.

=With the random selection excluding the best 10%, it implies that for Alice's claim of new version able to outperform the previous version may not be actually the case since the participants part of the participants are not included hence claim does not hold

=As for Bob's claim, the new average will depict a different value with this exclusion hence to tell how the grandmaster will average against the new version will not hold

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.

Solution

$$\Rightarrow P(Z) = \int_0^1 f_z(z)dz \text{ ,and } \Rightarrow P(\Theta) = \int_0^1 f_\Theta(\Theta)d\Theta$$

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:

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

Solution

$$\Rightarrow P(Z,\Theta) = \int_0^1 \int_0^{\text{II}} f_z(z)f_\Theta(\Theta)dzd\Theta \Longleftrightarrow P(Z,\Theta) = \int_0^1 \int_0^{\text{II}} f_{\Theta z}(\Theta, z)d\Theta dz$$

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

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

Solution

$$\Rightarrow P(Z,\Theta) = \int_0^1 \int_0^{\text{II}} f_{\Theta z}(\Theta, z)d\Theta dz, \text{ intersection will occur at } (1-z) \le \frac{1}{2}sin\Theta$$

$$\Rightarrow P(Z \leq \frac{1}{2}sin\Theta) = \int_0^1 \int_0^{\Pi} (z - \frac{1}{2}sin\Theta)dzd\Theta = \int_0^{\Pi} z^2 - \frac{1}{2}sin\Theta z|d\Theta$$

$$\Rightarrow [\frac{\Theta}{z}/\frac{\Pi\Theta}{2z}] \text{ interval } 0 \longrightarrow \Pi, \Longrightarrow \frac{\Theta}{z} * \frac{2z}{\Pi\Theta} = \frac{2}{\Pi}$$

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.

Solution

=Based on the law of large numbers the statistician can repeat this same experiment a large number of times each time getting a different number of results coming closer to the desired value phi, which becomes stable for extended large number of trials

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

Solution

=As n approaches infinity, the distribution of X converges to a certain random variable i.e.

$$\Rightarrow X_n \longrightarrow x$$

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.

=To verify that the density function is a valid density function, then first for every exp(-x^2) must be positive which is the case since we have a positive domain as given ,similarly the area beneath this function will be equal to integrating the function over the given interval resulting to root phi

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

Solution

$\Rightarrow$ We have from $\displaystyle\int_{-\infty}^{\infty} e^{-x^2}dx,$ that it's probability density function will be given by:

$$\Rightarrow \Theta(x) = \frac{1}{\sqrt{2\Pi}}e^{-x^2} \text{ ,thus E(X) will be equal to :}$$

$$\Rightarrow E(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\Pi}}e^{-x^2}dx = -\frac{1}{\sqrt{2\Pi}}e^{-x^2} \text{ ,interval } -\infty \longrightarrow +\infty = 0$$

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

Solution

$$\Rightarrow Var(X) = E(X^2) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2}dx$$

$\Rightarrow$ Upon integration by parts we have:

$$\Rightarrow \frac{1}{\sqrt{2\Pi}}\left(-xe^{-x^2}, interval, -\infty \longrightarrow +\infty\right) - -\frac{1}{\sqrt{2\Pi}}\int_{-\infty}^{\infty} e^{-x^2}dx = 0 - -\frac{1}{2} = \frac{1}{2}$$

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 $\rho = \mathrm{cov}(X,Y)$.

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

Solution

$\Rightarrow$ This is because when dividing the product of ρ_x and ρ_y, i.e. $\rho_x\,\rho_y$,from the given function f(x,y)

$\Rightarrow$ above for the standard deviation ,it implies that the correlation will be bounded between $\lfloor$-1,1$\rfloor$,

$\Rightarrow$ thus for 2 joint independent variables X and Y will have zero covariance and zero corrleation

$\Rightarrow$ which will be equal

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

Solution

$\Rightarrow$ We have that, $\rho = corr(X,Y) = \dfrac{cov(X,Y)}{\sigma(x)\sigma(y)}$

$\Rightarrow$ Thus given $\rho = 0, \Rightarrow corr(X,Y) = cov(X,Y) = 0.\sigma(x)\sigma(y)$

$\Rightarrow cov(X,Y) = 0, and, corr(X,Y) = 0$

$\Rightarrow$ Thus we know that if both the covariance and correlation for variable X and Y are zero,

$\Rightarrow$ then the two variables are independent implying that the case described is true

THE END