Problem Sheet 3

Question 1

a)

i)

The sample space is the set of possible outcomes from undertaking in an experiment. In this case we draw one ball, observe the colour and then return it to the bag. There are three possible coloured balls that can be drawn. Hence, for one draw, there are three elements in the sample space: [R, B, G]. The second draw has the same sample space as the first draw, as the ball is returned to the bag. Therefore, the sample space for consecutive draws is all the combinations between the sample space of the two draws, when looks at as individual events: [(R, G), (R, B), (R, R), (B, R), (B, G), (B, B), (G, R), (G, B), (G, G)].

The event space represents all the subsets in the sample space that satisfies a condition. An example element of this sample space is “when a red ball is drawn first”. The sample space of this subset is: [(R, R), (R, B), (R, G)]

ii)

iii)

Instead of looking to find all the combinations P(R) and adding them. We can find:

As the draws are independent due to the ball being replaced, we can say:

Where:

iv)

Since the ball is returned in the first experiment, the probability space is uniform across both draws:

In the second experiment, the probability space of the first draw is the same as the first experiment. However, since the ball is not replaced in the second experiment, the probability space of the second draw changes to the below:

v)

vi)

vii)

b)

i)

Given:

The number of ways 2 heads can come up in 5 flips:

The probability of 2 heads coming up in 5 flips:

ii)

When p = 0.5 and even, the binomial distribution is symmetrical. That is, the probability of events occurring above the mean is equal to the probability of events below the mean.

A graph with a line

Description automatically generatediii)

We can see that as k grows, the probability of Y tends to 0. Hence, this is a geometric probability distribution.

The probability mass function for a geometric distribution describes the number of trials required to get the first successful outcome. Hence:

iv)

The expected value of a random variable X, is defined as the weighted average of all values of X. The formula for the mean of a geometric distribution:

Given:

We then substitute the following for p, into the E(X) formula:

Given the definition of E(Y) of a geometric distribution below:

Taking *p* as *a*, and taking our definition off above, we can say:

c)

i)

Since:

Therefore:

We can use the law of total probability to say that the: is the sum of all the probabilities, that both the discrete random variables X and Y take of values of *x* and *y* respectively, such that the sum of *x* and *y* is *z*. This can be shows as:

Taking the definition of the joint probability mass function, we can therefore replace:

ii)

Since X and Y are now independent, we cannot use the concept of a join probability mass function. We need to look at the random variable’s individual probability mass functions (PMF). We then take the product of the PMF of X and Y.

Hence, given:

We can say:

Given:

And:

Hence:

iii)

The probability mass function of a Poisson distribution of a discrete random variable X, is given by:

Using this formula, we can state the PMF for both of our discrete random variables as:

Now given that:

And:

And now given, that X and Y are independent events, we can state:

The PMF of Z is therefore given by:

Using the formular for a binomial expansion:

We can use the above to simplify our expression to:

iv)

We can X + Y to Z, hence we need to find:

Given the conditional probability formula, we can state:

We know that:

And:

Using the conditional probability formula, we can now say:

Question 2

a)

i)

The population is all the chess players in the world that have a FIDE recognised title of a chess grandmaster.

The sample is 100, as 100 FIDE chess grandmasters are selected out of all the FIDE chess grandmasters in the world.

The sample size is 100, as there will be 100 games of chess. One for each chess grandmaster.

ii)

A reasonable parameter for this experiment is the win rate of the computer against the FIDE chess grandmasters. The parameter space would be [0, 1]. 0 being that the computer losing all games and 1 meaning the computer wins all games, and a spectrum in between.

iii)

An estimate of this parameter is 0.6. Since the computer achieved 60 points out of a possible 100 in the experiment.

This is an unbiased estimate as the sample of FIDE grandmasters were chosen at complete random from the population of FIDE grandmasters.

iv)

Alice : The new version of the chess computer will score <= 60 points.

Alice The new version of the chess computer will score > 60 points.

Bob The new version of the chess computer will score < 70 points.

Bob The new version of the chess computer will score >= 70 points

v)

In Alices case, we will re-run the experiment with the new chess computer. The aim for allice is to reject the . For this to be the case, the parameter, which is the chess computer win rate, must be > 0.6. Which represents more than 60 points scored. Hence the critical value is 0.6. If the parameter of the new test exceeds 0.6, the is rejected and Alices is accepted.

vi)

A critical *p* value will first need to be determined. This states that, if the *p* value of the sample test is less than the critical *p* value, then we deem the sample to be statistically significant and we reject the . A common *p* value is 0.05, we will assume Bob uses this. The standard deviation of the sample of games would need to be calculated. After this, a Z test can be performed:

Once the *z* value for the sample experiment is calculated. A *p* value for the sample can be found using the statistical *z-table*. This provides a *p* value for each corresponding *z* value, and vice versa. For example:

A screenshot of a graph

Description automatically generated

If the *p* value for the sample is < the critical *p* value chosen at 0.05, then the Is rejected and we can say that: The new version of the chess computer will score >= 70 points

vii)

Given the second experiment exclude the best 10% of grandmasters from the population and therefore the sample, we can say that the second experiment was easier for the chess computer. As a result, we expect a higher win rate for the chess computer, hence, our parameter has a bias above 0.6, before accounting for the new chess computer.

The Type I error of this experiment decreases since the probability to incorrectly reject the goes down as this experiment is easier for the computer, adding a positive win rate bias.

The Type II error of this experiment increases. This is due to the parameter bias to the upside. The probability of not rejecting a goes up as the experiment is easier than the first.

b)

i)

ii)

Since and are independent. We can state the below for the joint PDF:

iii)

iv)

The Law of Large Numbers states that as the random sample size increases, the closer the sample mean will get to the real mean, or the expected value. In this case, we are referring to the drop of the needle as the experiment. Hence, as the number of drops increase, the average result of the drops should tend to the true probability of the needle crossing the line.

Since:

We can say that:

We can say that statistician can get closer to estimating by increasing the number of trials.

v)

As the number of the drops of the needle tends to infinity, the probability that the line is intersected tends to:

The Central Limit Theorem states that if we take a large enough random sample from any population, no matter what its distribution, the sample mean will be normally distributed.

Hence, since this holds:

The statistician can be confident that as n grows, the accuracy of grows as the distribution of the drops of the needle becomes more and more bell shaped. The variance the distribution drops as a result.

c)

i)

We can verify that is a valid probability density function by showing that all values are not negative and that the integral is equal to 1.

The probability density function of a normal distribution is given by:

We can say that f(x) is always positive as is a positive constant and .

Given

Substituting these values into the probability density function for a normal distribution:

Since:

This shows the integral of the probability density function is in fact 1. Confirming the second condition for validity.

ii)

Given that the random variable X is normally distributed with . We can say that the expected value of the random variable X, given by E(X), is 0. We can say this as we know the mean of the distribution is 0. We can say this as and E(X) = 0 is a fundamental property of a normal distribution.

iii)

Var(X) = , we can solve for to find the variance. Given the normal probability density function is valid, we know that:

iv)

Given the below formular for the correlation between the random variables X and Y:

And that X and Y are normally distributed with variances of 1, given by:

We can simplify the correlation formular to:

Hence, we have shown that the correlation between X and Y, equals the covariance between X and Y. This is due to the unique nature of normal distributions with variances of 1.

v)

When p = 0:

When the random variables are independent then the join probability density function is the product of the individual probability density functions:

References

<https://math.libretexts.org/Bookshelves/Algebra/Algebra_and_Trigonometry_1e_(OpenStax)/13%3A_Sequences_Probability_and_Counting_Theory/13.06%3A_Binomial_Theorem>

<https://ocw.mit.edu/courses/6-436j-fundamentals-of-probability-fall-2018/491fa11552c94413806e5a09a9d6c237_MIT6_436JF18_lec05.pdf>

<https://www.scribbr.com/statistics/poisson-distribution/>

<https://www.scribbr.com/statistics/null-and-alternative-hypotheses/>

<https://www.youtube.com/watch?v=szUH1rzwbAw>

<https://www.statlect.com/fundamentals-of-probability/legitimate-probability-density-functions>

<https://en.wikipedia.org/wiki/Normal_distribution>