

1. You have two identical fair coins. You toss the first coin and if the output is heads then you stay with this coin and toss it again. If the the output is tails then you switch to the other coin and repeat the same process. This process can be summarized as follows:

Step 1: Select coin 1

Step 2: Toss the coin

Step 3: If result = Heads, go to step 2

Step 4: If result = Tails, switch to the other coin and go to step 2

- (a) What is the probability that after 99 tosses you end up with the same coin that you started with?

Solution: This means that you never switched the coin or switched it multiple times so that you again have coin 1.

- (b) What is the probability that after 100 tosses you end up with the same coin that you started with?

Solution:

- (c) What if instead of fair coins you have identical biased coins with probability of heads = p ($p \neq \frac{1}{2}$)?

Solution:

2. You are dealt a hand of 5 cards from a standard deck of 52 cards which contains 13 cards of each suite (hearts, diamonds, spades and clubs).

- (a) What is the probability that you get an ace, a king, a queen, a joker and a 10 of the same suite? Let us call such a hand as the King's hand.

Solution: Hint: it is very small :-)

- (b) Let n be the number of times you play this game. What is the minimum value of n so that the probability of having no King's hand in these n turns is less than $\frac{1}{e}$?

Solution: Hint: it is very large :-)

3. A spacecraft explodes while entering the earth's atmosphere and disintegrates into 10000 pieces. These pieces then fall on your town which contains 1600 houses. Each piece is equally likely to fall on every house. What is the probability that no piece falls on your house (assume you have only one house in the town and all houses are of the same size and equally spaced - for example you can assume that the houses are arranged in a 40×40 grid).

4. What's in a name?

- (a) Why is the hypergeometric distribution called so? (We understand what is geometric but what is “hyper”?)

Solution:

Hint 1: Find the ratio of $p_X(j)$ to $p_X(j - 1)$ for a hypergeometric random variable.

Hint 2: Why is the geometric distribution called the geometric distribution?

- (b) Why is the negative binomial distribution called so?

Solution:

5. Consider a binomial random variable whose distribution $p_X(x)$ is fully specified by the parameters n and p .

- (a) What is the ratio of $p_X(j)$ to $p_X(j - 1)$?

Solution:

- (b) Based on the above ratio can you find the value(s) of j for which $p_X(j)$ will be maximum ?

Solution:

6. Consider a Poisson random variable whose distribution $p_X(x)$ is fully specified by the parameter λ .

- (a) What is the ratio of $p_X(j)$ to $p_X(j - 1)$?

Solution:

- (b) Based on the above ratio can you find the value(s) of j for which $p_X(j)$ will be maximum ?

Solution:

7. For each of the following random variables show that the sum of the probabilities of all the values that the random variable can take is 1?

- (a) A negative binomial random variable whose distribution is fully specified by p (probability of success) and r (fixed number of desired successes)

Solution:

- (b) A hypergeometric random variable whose distribution is fully specified by N (number of objects in the given source), a (number of favorable objects in the source) and n (size of the sample that you want to select)

Solution:

- (c) A Poisson random variable whose distribution is fully specified by λ (i.e., arrival rate in unit time)

Solution: To prove that $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1$

8. There are 100 seats in a movie theatre. Customers can buy tickets online. Based on past data, the theatre owner knows that 5% of the people that book tickets do not show up (of course, he gets to keep the money they paid for the ticket). To make more money he decides to sell more tickets than the number of seats. For example, if he sells 102 tickets, then as long as at least 2 customers don't show up, he will be able to make extra money while not dissatisfying any customers.

- (a) If he sells 105 tickets what is the probability that no customer would be denied a seat on arrival.

Solution:

- (b) What is the maximum number of seats that he can sell so that there is at least a 90% chance that every customer will get a seat on arrival?

Solution: If you can find an analytical solution, then great! If you can't then you can write a program to find the answer. In that case, paste the code here.

- (c) Suppose he makes a profit of 5 INR for every satisfied customer and a loss of 50 INR (as penalty) for every dissatisfied customer (i.e., a customer who does not get a seat). What is his expected gain/loss if he sells 105 tickets?

Solution:

P.S.: This is what many international airlines do. They often sell more tickets than the number of available seats thereby profiting twice from the same seat!

9. In recently conducted elections, there were a total of 100 counting centres. The losing party claims that some of the counting machines were rigged by a hacker. To verify these allegations, the Election Commission decides to manually recount the votes in some

centres (obviously, manual recounting in all centres would be prohibitively expensive so it can only do so in some centres).

- (a) If 5% of the machines were rigged then in how many centres should recounting be ordered so that there is a 50% chance that rigging would be detected (i.e., in at least one of the selected centres the number of votes counted manually will not match the number of votes counted by the machine)

Solution: If you can find an analytical solution, then great! If you can't then you can write a program to find the answer. In that case, paste the code here.

- (b) If the hacker knows that the Election commission can only afford to do a recounting in 10 randomly sampled centres then what is the maximum number of machines he/she can rig so that there is less than 50% chance that the rigging will get detected.

Solution: If you can find an analytical solution, then great! If you can't then you can write a program to find the answer. In that case, paste the code here.

10. Amar and Bala are two insurance agents. Their manager has given them a list of 40 potential customers and a target of selling a total of 5 policies by the end of the day. They decide to split the list in half and each one of them talks to 20 people on the list. Amar is a better salesman and has a probability p_1 of selling a policy when he talks to customer. On the other hand, Bala has a probability p_2 ($< p_1$) of selling a policy when he talks to a customer. The customers do not know each other and hence one customer does not influence another. What is the probability that they will be able to meet their target by the end of the day? (it doesn't matter if Amar sells more policies than Bala or the other way round - the only thing that matters is that the total should be **exactly** 5).

Solution: I like this problem. So easy to state. So easy to relate to. But not easy to solve :-). The beauty is that if $p_1 = p_2$ then the problem becomes trivial but when you make $p_1 \neq p_2$ then it becomes hard! Further, imagine that instead of Amar and Bala what would happen if we had Amar, Bala and Chandu!

11. Find the expectation and variance of the following discrete random variables
- (a) A binomial random variable whose distribution is fully specified by p (probability of success) and n (number of trials)

Solution:

- (b) A negative binomial random variable whose distribution is fully specified by p (probability of success) and r (fixed number of desired successes)

Solution:

- (c) A hypergeometric random variable whose distribution is fully specified by N (number of objects in the given source), a (number of favorable objects in the source) and n (size of the sample that you want to select)

Solution:

- (d) A Poisson random variable whose distribution is fully specified by λ (i.e., arrival rate in unit time)

Solution:

12. Consider a language which has only 5 words w_1, w_2, w_3, w_4, w_5 . The way you construct a sentence in this language is by selecting one of the 5 words with probabilities p_1, p_2, p_3, p_4, p_5 respectively ($\sum_{i=1}^5 p_i = 1$). This word will be the first word in the sentence. You will then repeat the same process for the second word and continue to form a sentence of arbitrary length. As should be obvious, the i -th word in the sentence is independent of all words which appear before it (and after it). What is the expected position at which the word w_2 will appear for the first time?

Solution:

13. Two fair dice are rolled. Let X be the sum of the two numbers that show up and let Y be the difference between the two numbers that show up (number on first dice minus number on second dice).

- (a) Show that $E[XY] = E[X]E[Y]$

Solution:

- (b) Are X and Y independent? Explain your answer.

Solution:

14. The *martingale doubling system* is a betting strategy in which a player doubles his bet each time he loses. Suppose that you are playing roulette in a fair casino where the roulette contains only 36 numbers (no 0 or 00). You bet on red each time and hence your probability of winning each time is $1/2$. Assume that you enter the casino with 100 rupees, start with a 1-rupee bet and employ the martingale system. Your strategy is to

stop as soon as you have won one bet or you do not have enough money to double the previous bet.

- (a) Under what condition will you not have enough money to double your previous bet?

Solution:

- (b) What would your expected winnings be under this system? (for every 1 INR you bet you get 2 INR if you win)

Solution:

15. You have 800 rupees and you play the following game. A box contains two green hats and two red hats. You pull out the hats out one at a time without replacement until all the hats are removed. Each time you pull out a hat you bet half of your present fortune that the pulled hat will be a green hat. What is your expected final fortune?

Solution:

16. There are 6 dice. Each dice has 0 on five sides and on the 6th side it has a number between 1 and 6 such that no two dice have the same number (i.e, if dice 2 has the number 3 on the 6th side then no other dice can have the number 3 on the sixth side). All the 6 dice are rolled and let X be the sum of the numbers on the faces which show up.

- (a) Find $E[X]$ and $Var(X)$

Solution:

- (b) Suppose you are the owner of a casino which has a game involving these 6 dice. The players can bet on the sum of the numbers that will show up. If I bet on the number 21 what should the payoff be so that the bet looks as attractive as possible to me but in the long run the casino will not lose (e.g., in the game of roulette a payoff of 1:35 looks attractive while still protecting the interests of the casino).

Solution:

17. A friend invites you to play the following game. He will toss a fair coin till the first heads appears. If the first head appears on the k -th toss then he will give you 2^k rupees. His condition is that you should first pay him a 100 million rupees to get a chance to play this game. Would you be willing to pay this amount to get a chance to play this game? Explain your answer.

Solution: Hint: What do you “expect” from your friends?

18. Suppose you are playing with a deck of 20 cards which contain 10 red cards and 10 black cards. The dealer opens the cards one by one but you cannot see a card before he opens it. Before he opens a card you are supposed to guess the color of the card.
- (a) If you are guessing randomly then what is the expected number of correct guesses that you will make?

Solution:

- (b) Can you think of a better strategy than random guessing?

Solution: This is similar to (but much easier than) what skilled blackjack players do to increase their odds of winning. This is the subject of the movie “21”. Now, don’t go and watch the movie because the movie does not say much about the strategy.

- (c) What is the expected number of correct guesses under this intelligent strategy? It is hard (but possible) to come up with an analytical solution. However, it is easy to do a simulation. Write a program to play this game a 1000 times and note down the number of correct guesses each time. Based on this simulation calculate the estimate number of correct guesses.

Solution: First paste your code here and then write down the expected number of correct guesses. I am not looking for the analytical solution.

19. Suppose every morning the front page of Chennai Times contains a photo of exactly one of the n celebrities of Kollywood. You are a movie buff and collect these photos. Of course, on some days the paper may publish the photograph of a celebrity which is already in your collection. Suppose you have already obtained photos of $k - 1$ celebrities. Let X_k be the random variable indicating that number of days you have to wait before you obtain the next new picture (after obtaining the first $k - 1$ pictures).
- (a) Show that X_k has a geometric distribution with $p = (n - k + 1)/n$

Solution:

- (b) Simulate this experiment with 50 celebrities. Carry out a large number of simulations and estimate the expected number of days required to get the photos of the first 25 celebrities and the next 25 celebrities. Paste your code and estimates of the two expected values below.

Solution:

20. You want to test a large population of N people for COVID19. The probability that a person may be infected is p and it is the same for every person in the population. Instead of independently testing each person you decide to do pool testing wherein you collect the blood samples of k people and test them together (N is divisible by k). If the test is negative then you conclude that all are negative and no further tests are required for these k people. However, if the test is positive then you do k more tests (one for each person).

- (a) What is the probability that the test for a given pool of k people will be positive?

Solution:

- (b) What is the expected number of tests required under this strategy to conclusively test the entire population?

Solution:

- (c) When would such a pooling strategy be beneficial?

Solution:

21. A family decides to have children until they have a girl or until there are 3 children, whichever happens first. Let X be the random variable indicating the number of girls in the family and let Y be the random variable indicating the number of boys in the family. Assume that the probability of having a girl child is the same as that of having a boy child.

- (a) Find $E[X]$ and $Var[X]$.

Solution:

- (b) Find $E[Y]$ and $Var[Y]$.

Solution:

22. Suppose n people bring their umbrellas to a meeting. While returning back each one randomly picks up an umbrella and walks out.

Let X_i be the random variable indicating whether the i -th person walked out with his own umbrella (i.e., the umbrella that he walks out with is the same as the umbrella that he walked in with).

- (a) Find $E[X_i^2]$

Solution:

- (b) Find $E[X_i X_j]$ (for $i \neq j$)

Solution:

Let S be the random variable indicating the number of people who walk out with their own umbrella.

- (c) Find $E[S]$

Solution:

- (d) $\text{Var}[S]$

Solution:

23. The covariance of two random variables X and Y is defined as $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$. Show that if X and Y are independent then $\text{Cov}(X, Y) = 0$.

Solution:

24. Consider a die which is loaded such that the probability of a face is proportional to the number on the face. Let X be the random variable indicating the outcome of the die.

- (a) Find $E[X]$ and $\text{Var}(X)$

Solution:

- (b) Write code to simulate such a die. Run 1000 simulations and note down the number of times each face shows. Calculate the empirical expectation and see if it matches the theoretical expectation you have computed above.

Solution: