Question 1 – Unfair Coin

This problem was asked by Facebook.

There is a fair coin (one side heads, one side tails) and an unfair coin (both sides tails). You pick one at random, flip it 5 times, and observe that it comes up as tails all five times. What is the chance that you are flipping the unfair coin?

Solution:

This problem can be solved using Bayes Rule. We are asked to calculate the following quantity: $P(\text{Unfair} \mid \text{TTTTT})$.

We have that

$$\begin{aligned}
P(\text{Unfair} \mid \text{TTTTT}) &= \frac{P(\text{TTTTT} \mid \text{Unfair}) \cdot P(\text{Unfair})}{P(\text{TTTTT})} \\
&= \frac{P(\text{TTTTT} \mid \text{Unfair}) \cdot P(\text{Unfair})}{P(\text{TTTTT} \mid \text{Unfair}) \cdot P(\text{Unfair}) + P(\text{TTTTT} \mid \text{Fair}) \cdot P(\text{Fair})} \\
&= \frac{1^5 \cdot 0.5}{1^5 \cdot 0.5 + 0.5^5 \cdot 0.5} \\
&= \frac{0.5}{0.5 + 0.015625} \\
&= \boxed{0.9697}
\end{aligned}$$

Question 2 – Flips until two heads

This problem was asked by Lyft.

What is the expected number of coin flips needed to get two consecutive heads?

Solution:

This is a slighly more complicated version of the classic problem – expected number of coin flips needed to get heads (which is simply the expected value of the geometric random variable with $p = 0.5$)

We can represent this sytem as a Markov chain as follows:

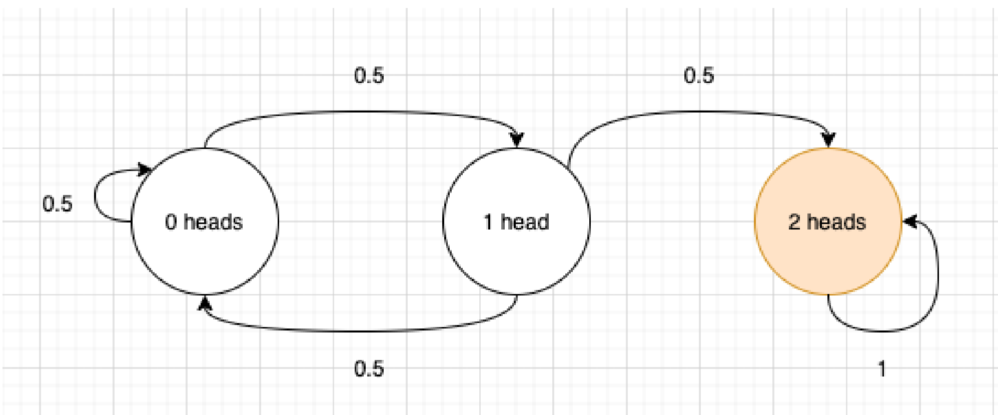


Figure 1: We start with 0 heads. With 0.5 probability, we see 1 head. After we have seen 1 head, if we see another head (which happens with $p = 0.5$), we are in the absorbing state 2. If we see a tails ($p = 0.5$), we go back to state 0.

If we are in state 2, the expected time to see two heads is $E[2] = 0$ as we have already seen two heads.

If we are in state 1, the expected time to see two heads is given by (using the law of total expectation):

$$
\begin{aligned}
E[1] &= 1 + \frac{1}{2} \cdot E[0] + \frac{1}{2} \cdot E[2] && (1) \\
&= 1 + \frac{1}{2} \cdot E[0] && (2)
\end{aligned}
$$

If we are in state 0, the expected time to see two heads is given by:

$$E[0] = 1 + \frac{1}{2} \cdot E[0] + \frac{1}{2} \cdot E[1] \quad (3)$$

Substituting (2) in (3), we have:

$$E[0] = 1 + \frac{1}{2} \cdot E[0] + \frac{1}{2} \cdot \left(1 + \frac{1}{2} \cdot E[0]\right) \tag{4}$$

$$E[0] = \boxed{6} \tag{5}$$

Thus, if we are in state 0, that is when we start the experiment, the expected number of flips to see 2 heads is 6.

Question 3 – Drawing normally

This problem was asked by Quora.

You are drawing from a normally distributed random variable $X \sim \mathcal{N}(0, 1)$ once a day. What is the approximate expected number of days until you get a value of more than 2?

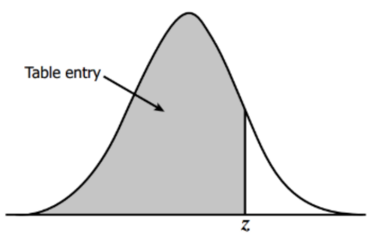
Solution

We can look at this problem as follows. Each day we carry out an experiment in which we draw from a standard unit normal. If the value sampled is greater than 2, then the experiment is successful. We want to know the average number of days in which we can expect a success.

The second part of the problem – average number of days in which we can expect a success sounds like a geometric random variable with parameter $Y \sim \text{Geom}(p)$ All we need to do is to find the value of parameter p and then the mean of the geometric random variable is $\frac{1}{p}$ which will give us the average number of days until we see the first "success" as we have defined it.

Now, to find p, consider the experiment that we perform every day. For it to be a "success", i.e., have a value more extreme than 2, it needs to be 2 standard deviations above the mean. In other words, the $z-$score is 2.0. We can look up the value for a $z-$score of 2.0, (from a $z-$table such as 2) which will give us the probability of getting a value less than 2.0. We can then subtract this from 1 to give us the probability of getting a value more extreme than 2.0.

The value here is 0.9772. Thus the probability we want is $1 - 0.9772 \approx 0.0228$. Thus, we have that $Y \sim \text{Geom}(0.0228)$. Thus, the average number of days that we have to wait are $\frac{1}{0.0228} \approx \boxed{43.86.}$



Find values on the right of the mean in this z-table. Table entries for z represent the area under the bell curve to the left of z. Positive scores in the Z-table correspond to the values which are greater than the mean.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Figure 2: We need to find the value for $z-$score equal to 2.0. Figure is from here.

Question 4 – Is this coin biased?

This problem was asked by Google.

A coin was flipped 1000 times, and 550 times it showed up heads. Do you think the coin is biased? Why or why not?

Solution

We could approach this the frequentist way, since the question demands a specific answer, that is, is the coin biased or not. We can formulate the following hypotheses:

$$H_0 : \text{Coin is not biased } (p = 0.5)$$
$$H_1 : \text{Coin is biased } (p \neq 0.5)$$

Let the random variable X denote the number of heads obtained where $X \sim \text{Binom}(p)$. The PMF looks like so:

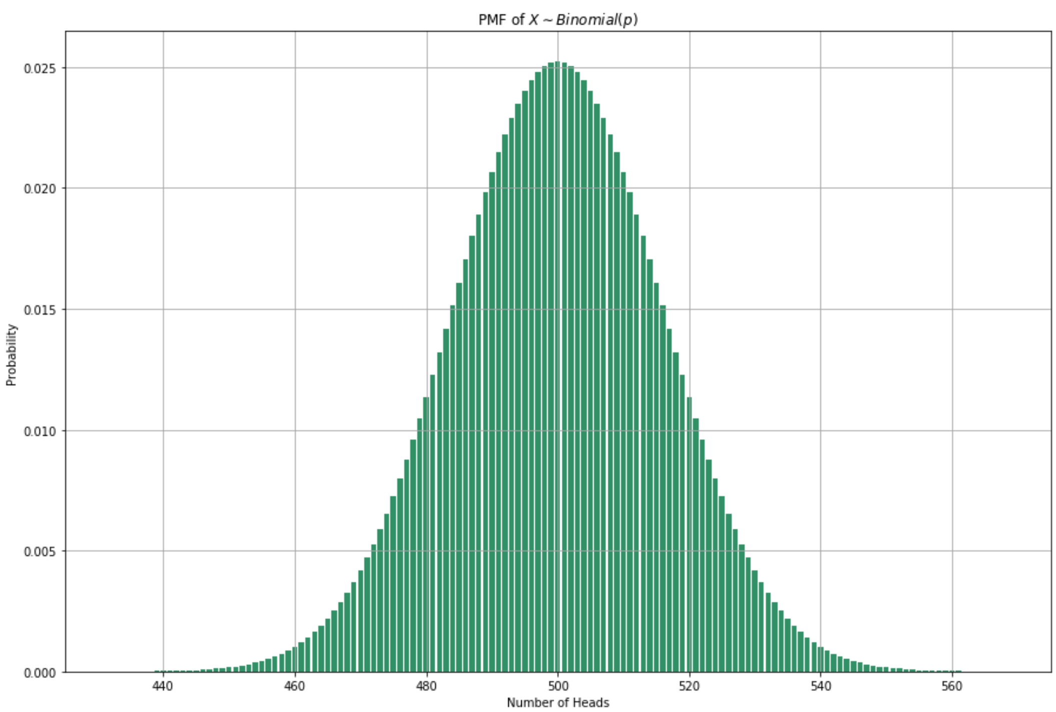


Figure 3: PMF of $X \sim \text{Binom}(0.5)$

We can formulate this as a one-sided test, that is, we want to find the value of $p(X \geq 550 \mid p = 0.5)$ This is the shaded area (in red) in 4.

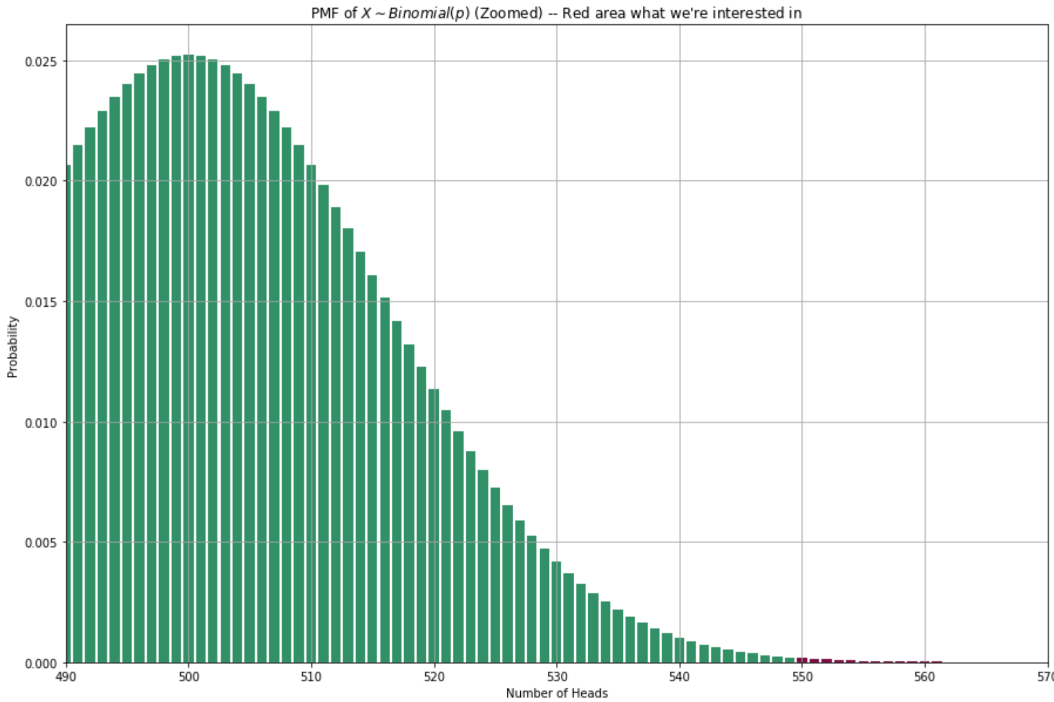


Figure 4: PMF of $X \sim \text{Binom}(0.5)$. We want to find the value of the red shaded area.

Now,

$$\begin{aligned} p(X \geq 550 \mid p = 0.5) &= \sum_{i=550}^{1000} \binom{1000}{i} \cdot 0.5^{1000} \\ &\approx 0.00086 \end{aligned}$$

For a reasonable threshold of $\alpha = 0.05$, since $0.00086 < \alpha = 0.05$, we can reject the null hypothesis and **conclude that the coin is biased.** Note that the value of 0.00086 was obtained computationally and we could also obtain this value analytically using the normal approximation to the binomial since our sample size is large.

To find the value using the normal approximation, we find the $z-$score of X. The standard deviation of the binomial distribution is $\sqrt{n \cdot p \cdot (1-p)} = \sqrt{1000 \cdot 0.5 \cdot 0.5} \approx 15.8113$ and thus the $z-$score is $\dfrac{550 - 500}{15.8113} \approx 3.162$. The probability is then $1 - 0.9992 \approx 0.0008$ (the value 0.9992 is obtained by looking up the value of 3.16 in a $z-$table like the one in 2) which is very close to the computational value that we obtained.

Question 5 – Rolls to see all sides

What is the expected number of rolls needed to see all 6 sides of a fair die?

Solution

During the first roll, we are guaranteed to see an unseen side. For the second roll, there is a probability of $\frac{5}{6}$ to see an unseen side. The expected number of rolls to see an unseen side is a geometric random variable $X \sim \text{Geom}\left(\frac{5}{6}\right)$ with a mean value of $\frac{6}{5}$. Similary, for the third unseen side, there is a probability of $\frac{4}{6}$ to see an unseen side. The expected number of rolls to see an unseen side is again a geometric random variable $X \sim \text{Geom}\left(\frac{4}{6}\right)$ with a mean value of $\frac{6}{4}$. This goes on for all the unseen sides until we've seen all of them which gives us:

$$\begin{aligned} E[Y] &= 1 + \frac{6}{5} + \frac{6}{4} + \cdots + \frac{6}{1} \\ &= \boxed{14.7} \end{aligned}$$

Thus, on average we will have to roll the die 14.7 times to see all the sides.

Question 6 – Picking between two dice games

This problem was asked by Facebook.

There are two games involving dice that you can play. In the first game, you roll two die at once and get the dollar amount equivalent to the product of the rolls. In the second game, you roll one die and get the dollar amount equivalent to the square of that value. Which has the higher expected value and why?

Solution

Consider the first game. Let the outcomes of the two rolls be represented by the random variables X and Y. The quantity to be computed here is $E[XY]$. Since the two die rolls are independent the expected value of the product of random variables is equal to the product of their expectation. Individually, X and Y are discrete random variables distributed as $X \sim \text{Uniform}(1, 6)$ and $Y \sim \text{Uniform}(1, 6)$.

$$\begin{aligned} E[XY] &= E[X] \cdot E[Y] \\ &= 3.5 \cdot 3.5 \\ &= 12.25 \end{aligned}$$

For the second game, let the outcome of the first die roll be denoted by the random variable Z which is also a discrete uniform random variable distributed as $Z \sim \text{Uniform}(1, 6)$. For a discrete uniform random variable with parameters a and b, the variance of Z is given by:

$$\begin{aligned} \text{Var}(Z) &= \frac{(b - a + 1)^2 - 1}{12} \\ &= \frac{(6 - 1 + 1)^2 - 1}{12} \\ &= \frac{35}{12} \\ &\approx 2.92 \end{aligned}$$

The quantity of interest here is $E[Z^2]$. We have the following relation:

$$\begin{aligned} \text{Var}(Z) &= E[Z^2] - (E[Z])^2 \\ E[Z^2] &= \text{Var}(Z) + (E[Z])^2 \\ &= 2.92 + 12.25 = 15.17 \end{aligned}$$

Thus, the expected value of the second game (15.17) is higher than that of the first game (9). [1]

[1] Thank you Kevin Yang for correcting my calculation of $E[X]$ and $E[Y]$

Question 7 – Fair odds from unfair coin

This problem was asked by Airbnb.

Say you are given an unfair coin, with an unknown bias towards heads or tails. How can you generate fair odds using this coin?

Solution

Let the coin be biased with a probability of heads equal to p. Thus, the probability of tails is $(1-p)$. Instead of flipping it once, consider flipping it twice. There are four possible outcomes:

Sequence	**Probability**
HH	p^2
HT	$p \cdot (1-p)$
TH	$(1-p) \cdot p$
TT	$(1-p)^2$

Table 1: Four outcomes with their corresponding probability

We use a form of **rejection sampling.** The probability of HT and TH is the same, equal to $p \cdot (1-p)$. Thus, we change our experiment to flip the coin twice instead of once. If the outcome is HH or TT, we repeat the experiment. Otherwise, we can arbitrarily denote HT to denote heads and TH to denote tails before the experiment and record the outcome.

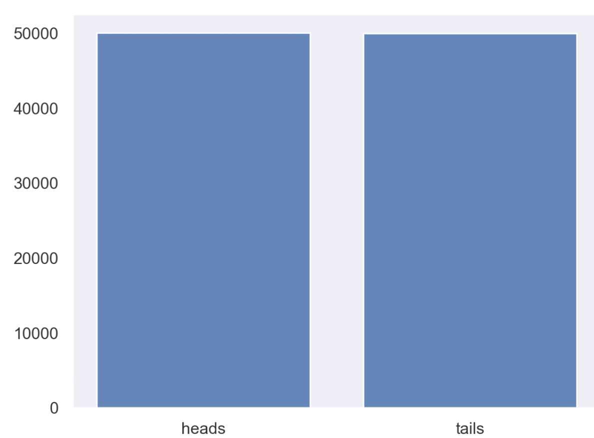


Figure 5: We ran a simulation of this technique with $p = 0.7$ and found 50036 heads and 49964 tails in 100000 experiments. These appear to be roughly equal.

Question 8 – Ant Collision

This problem was asked by Facebook.

Three ants are sitting at the corners of an equilateral triangle. Each ant randomly picks a direction and starts moving along the edge of the triangle. What is the probability that none of the ants collide? Now, what if it is k ants on all k corners of an equilateral polygon?

Solution

The ants will not collide if they all go in the same direction. They can either all go left or all go right. As each ant independently randomly picks a direction, the probability of this happening is $0.5^3 + 0.5^3 = 0.25$.

For k ants on k corners, the probability is $0.5^k + 0.5^k$.

Question 9 – Classification metrics

This problem was asked by Uber.

Say you need to produce a binary classifier for fraud detection. What metrics would you look at, how is each defined, and what is the interpretation of each one?

Solution

Consider the following table for classification:

		True	
		Fradulent	Authentic
Predicted	Fradulent	True Positive (TP)	False Positive (FP)
	Authentic	False Negative (FN)	True Negative (TN)

Table 2: Binary Confusion Matrix

True positives and true negatives are when we correctly identify frauds and geniune transactions respectively. From a business perspective, we would want to minimize False Negatives, or when we incorrectly classify a fradulent transaction as authentic. These could cost the company a lot of money. False positives typically would be less expensive (but still important, for example for the reputation of the company and to maintain a positive customer relationship), as these are instances when we incorrectly classify a geniune transaction as a fraud. In a realistic scenario, most of the transactions would be authentic and only a small number would be fradulent. As a consequence, we would expect a smaller number of False Positives, and we can have additional measures of safety like asking the user for additional documentation or adding a human in the loop to verify the transaction. Considering the above, the following metrics could be useful:

- **Accuracy:** The overall accuracy of the model, that is how often is the model making correct predictions (identifying both fradulent and authentic transactions). This is defined as $\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$.
- **False Negative Rate:** This is defined as $\frac{\text{FN}}{\text{TP} + \text{FN}}$, that is, what proportion of fradulent transactions were (incorrectly) classified as authentic.
- **Recall (True Positive Rate):** This is defined as $\frac{\text{TP}}{\text{TP} + \text{FN}}$, that is, what proportion of fradulent transactions were correctly classified as fradulent.

- **Precision:** This is defined as $\dfrac{\text{TP}}{\text{TP} + \text{FP}}$, that is, what proportion of transactions that were classified as fradulent are actually fradulent.

Question 10 - Flipping game

This problem was asked by Facebook.

You and your friend are playing a game. The two of you will continue to toss a coin until the sequence HH or TH shows up. If HH shows up first, you win. If TH shows up first, your friend wins. What is the probability of you winning?

Solution

If the first flip is T, there is no chance for me to win because if the next flip is H, then the friend wins, and if the next flip is T, we are still in the same position.

Therefore, the only way for me to win is to get two consecutive heads, which happens with a $0.5 \cdot 0.5 = 0.25$ probability. [2]

[2] Thank you Kevin Yang for correcting my solution

Question 11 - First to roll side k

This problem was asked by Lyft.

A and B are playing the following game: a number k from 1-6 is chosen, and A and B will toss a die until the first person sees the side k, and that person gets \$100. How much is A willing to pay to play first in this game?

Solution

Assuming player A plays first, the probability that they will see the side k on their first move is $\frac{1}{6}$. Thereafter, it is B's turn. The probability that A sees the side k on their second attempt (if B fails during their turn) is $\underbrace{\frac{5}{6}}_{\text{1st attempt missed by A}} \cdot \underbrace{\frac{5}{6}}_{\text{1st attempt missed by B}} \cdot \underbrace{\frac{1}{6}}_{\text{A succeeds at 2nd attempt}}$

Thus, in the general case, A alternates flipping the coin and the probability that they win during the i^{th} attempt is $\frac{5}{6}^{2i} \cdot \frac{1}{6}$ where i is zero indexed. Thus, the total probability of A winning if they start first is given by

$$P(\text{A starts first and wins}) = \sum_{i=0}^{\infty} \left(\frac{5}{6}^{2i} \right) \cdot \frac{1}{6}$$

Thus, we need to find the quantity $\sum_{i=0}^{\infty} \left(\frac{5}{6}^{2i} \right) \cdot \frac{1}{6}$. This is a geometric series with the first term $a = \frac{1}{6}$ and the common ratio $r = \frac{5}{6}^{2}$ as each subsequent value in the series increases by a factor of $\frac{5}{6}^{2}$. Since $|r| < 1$, the series is convergent and the sum is equal to $\frac{a}{1-r} = 0.545454$. Let X be the random variable that denotes if A wins. Thus, $P(\text{A starts first and wins}) = P(X) = 0.545454$. The expected amount of money that A earns if they begin first $E[X]$ is given by

$$\begin{aligned} E[X] &= 0.5454 \cdot 100 + (1 - 0.545454) \cdot 0 \\ &= 54.5454 \end{aligned}$$

Thus, A could be willing to bet $\approx$ \$54.54 or any amount lower than that.

Question 12 - One extra coin toss

This problem was asked by Robinhood.

A and B are playing a game where A has $n+1$ coins, B has n coins, and they each flip all of their coins. What is the probability that A will have more heads than B?

Solution

Up till the point that both A and B have flipped n coins, there is an equal probability of either one of them having more heads. Thus we can simply ignore this history. Now since the process has the property of *memorylessness*, at the $(n+1)^{\text{th}}$ flip, assuming that it is a fair coin, the probability that A gets heads is $\frac{1}{2}$. This is the only scenario in which A will have more heads than B, and thus the probability that A has more heads than B is $\frac{1}{2} = 0.5$.

Question 13 - Labeling content

This problem was asked by Facebook.

Facebook has a content team that labels pieces of content on the platform as spam or not spam. 90% of them are diligent raters and will label 20% of the content as spam and 80% as non-spam. The remaining 10% are non-diligent raters and will label 0% of the content as spam and 100% as non-spam. Assume the pieces of content are labeled independently from one another, for every rater. Given that a rater has labeled 4 pieces of content as good, what is the probability that they are a diligent rater?

Solution

Let S denote if the article was spam and $\neg S$ if the article is not spam. Let D denote a diligent rater and $\neg D$ a non diligent one. Therefore, we are asked to find $P(D \mid \neg S \neg S \neg S \neg S)$. Using Bayes rule we have

$$\begin{aligned} P(D \mid \neg S \neg S \neg S \neg S) &= \frac{P(\neg S \neg S \neg S \neg S \mid D) \cdot P(D)}{P(\neg S \neg S \neg S \neg S \mid D) \cdot P(D) + P(\neg S \neg S \neg S \neg S \mid \neg D) \cdot P(\neg D)} \\ &= \frac{0.8^4 \cdot 0.9}{0.8^4 \cdot 0.9 + 1^4 \cdot 0.1} \\ &\approx 0.787 \end{aligned}$$

Thus, given that a rater has labeled 4 pieces of content as not spam, there is a 78.7% probability that they are a diligent rater.

Question 14 - Coin flips needed to detect bias

This problem was asked by Lyft.

Say you have an unfair coin which will land on heads 60% of the time. How many coin flips are needed to detect that the coin is unfair?

Solution

Let us formalize the problem. Let us denote each coin flip i as a Bernoulli random variable $X_i \sim \text{Bernoulli}(p)$. In this case, we know that $X_i \sim \text{Bernoulli}(0.6)$. We cannot be a 100% certain that the coin in unfair for a finite number of coin flips, but rather express this in a probabilistic sense.

For n identically distributed independent random variables X_i, each having a mean μ and variance σ^2, we have the following relation (called Chebyshev's inequality)

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

In our case, the value of $\mu = 0.6$ and $\sigma^2 = p \cdot (1-p) = 0.6 \cdot 0.4 = 0.24$. For us to be reasonably confident that the sample mean (fraction of coin flips that turned up heads in our experiment) is in the interval $[0.5, 0.7]$ (which implies that it is not fair), we can choose $\epsilon = 0.1$. Thus, we have that

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| \geq 0.1\right) \leq \frac{0.24}{n \cdot 0.1^2}$$

Now, we want the right hand side to be ≤ 0.05, (this number is an arbitrary choice, and expresses our confidence about our estimates. In this case, we want to be 95% confident, giving us $1 - 0.95 = 0.05$) thus we have

$$\begin{aligned} \frac{0.24}{n \cdot 0.1^2} &\leq 0.05 \\ n &\geq 480 \end{aligned}$$

Thus, we need to flip the coin atleast 480 times to be 95% sure that the probability of heads of the coin is in the range $[0.5, 0.7]$ (which implies that it is unfair). We can have tighter bounds using other inequalities like Chernoff's inequality. And if we want to make the range shorter, for instance $[0.55, 0.65]$, we could use a different value of ϵ, which will in turn increase the number of flips n required to have the same level of confidence which is currently at 95%.

Question 15 - Max Dice Roll

This question was asked by Spotify.

A fair die is rolled n times. What is the probability that the largest number rolled is r, for each r in 1..6?

Solution

Let us consider the simple case, when $r = 1$. The probability that the largest number is $r = 1$ when the **fair** die is **independently** rolled n times is if we roll 1 for all the n runs, i.e., $\left(\frac{1}{6}\right)^n$.

If $r = 2$, we would have to roll either 1 or 2 for the n runs, i.e., $\left(\frac{2}{6}\right)^n$

In a general sense, if X_i denotes the random variable in the i^{th} roll of a total of n rolls, $P\Big(\max\{X_1, \cdots, X_n\} \leq r\Big) = \left(\frac{r}{6}\right)^n$ for each r in [1..6].

Question 16 - Customer Churn MLE

This question was asked by Airbnb.

Say you model the lifetime for a set of customers using an exponential distribution with parameter λ, and you have the lifetime history (in months) of n customers. What is the MLE for λ?

Solution

We are given $\{x_1, x_2, ..., x_n\}$ which denote the lifetime history (in months) of n customers. This is modelled using an exponential distribution parameterized by λ. Thus, we have

$$\begin{aligned} p(x_i \mid \lambda) &\sim \text{Exponential}(\lambda) \qquad i = \{1, ..., n\} \\ &\sim \lambda e^{-\lambda x_i} \end{aligned}$$

Our task is to infer λ. Assuming that the n observations are independent and identically distributed, we the combined likelihood is given as

$$p(x_1, x_2, ..., x_n \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

For MLE, we find find the value of the parameter λ that maximizes the likelikhood function. Additionally, we can maximize the log likelihood since log is a monotonic function and the products can be converted to sums, which are easier to work with.

$$\begin{aligned} p(x_1, x_2, ..., x_n \mid \lambda) &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \\ &= \lambda^n \prod_{i=1}^{n} e^{-\lambda x_i} \\ &= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \\ &= n \log \lambda - \lambda \cdot \sum_{i=1}^{n} x_i \qquad \text{taking the log} \\ 0 &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \qquad \text{taking derivative w.r.t } \lambda \text{ \& setting to } 0 \\ \lambda_{\text{MLE}} &= \frac{n}{\sum_{i=1}^{n} x_i} \end{aligned}$$

Question 17 - Server Wait Time

This problem was asked by Dropbox.

Dropbox has just started and there are two servers that service users: a faster server and a slower server. When a user is on the website, they are routed to either server randomly, and the wait time is exponentially distributed with two different parameters. What is the probability density of a random user's waiting time?

Solution

Let the wait time for the fast server be distributed according to $x \sim \text{Exponential}(\lambda_f)$ and the slow server $x \sim \text{Exponential}(\lambda_s)$. When the user is on the website, they are randomly routed to either of these servers. This means there is a 0.5 probability of reaching the fast server and 0.5 probability of the slow server. Thus, the overall distribution is a convex combination of the two.

$$
\begin{aligned}
x &\sim \frac{1}{2} \cdot \text{Exponential}(\lambda_f) + \frac{1}{2} \cdot \text{Exponential}(\lambda_s) \\
&\sim \frac{1}{2} \cdot \left(\lambda_f \cdot e^{-\lambda_f x}\right) + \frac{1}{2} \cdot \left(\lambda_s \cdot e^{-\lambda_s x}\right) \\
&\sim \frac{1}{2}\left(\lambda_f \cdot e^{-\lambda_f x} + \lambda_s \cdot e^{-\lambda_s x}\right)
\end{aligned}
$$

Question 18 - First Toss

This problem was asked by Lyft.

A fair coin is tossed n times. Given that there were k heads in the n tosses, what is the probability that the first toss was heads?

Solution

We are asked to find $P(\text{first toss} = \text{heads} \mid k \text{ heads in } n \text{ tosses})$. If the first toss was heads (since this is a fair coin, this happens with a probability of 0.5), then for us to get a total of k heads in n tosses, we would need to get $(k-1)$ heads in the remaining $(n-1)$ tosses. If the first toss was not heads ($p = 0.5$), we would need to get k heads in $(n-1)$ tosses. Thus, using the binomial distribution and bayes theorem we have

$$
\begin{aligned}
&= \frac{\binom{n-1}{k-1} \cdot 0.5^{(n-1)} \cdot 0.5}{\binom{n-1}{k-1} \cdot 0.5^{(n-1)} \cdot 0.5 + \binom{n-1}{k} \cdot 0.5^{(n-1)} \cdot 0.5} \\
&= \frac{0.5^n}{0.5^n} \cdot \frac{\binom{n-1}{k-1}}{\binom{n-1}{k-1} + \binom{n-1}{k}} \\
&= \frac{\binom{n-1}{k-1}}{\binom{n}{k}} \\
&= \frac{k}{n}
\end{aligned}
$$

Question 19 - Coin Recursion

This problem was asked by Robinhood.

A biased coin, with probability p of landing on heads, is tossed n times. Write a recurrence relation for the probability that the total number of heads after n tosses is even.

Solution

Let us define $f(n, 1)$ to be the probability that we observe an even number of heads in n tosses. $f(n, 0)$ therefore is the probability that we observe an odd number of heads in n tosses. Thus, the recurrence relation is as follows:

$$
\begin{aligned}
&f(n,1) = \Big(p \cdot f(n-1,0)\Big) + \Big((1-p) \cdot f(n-1,1)\Big) \\
&f(0,1) = 1 \\
&f(0,0) = 0
\end{aligned}
$$

Question 20 - Random Testing

This problem was asked by Lyft.

Say that you are pushing a new feature X out. You have 1000 users and each user is either a fan or not a fan of X, at random. There are 50 users of 1000 that do not like X. You will decide whether to ship the feature or not based on sampling 5 users independently and if they all like the feature, you will ship it. What is the probability that you will ship the feature?

Solution

For the first person we pick, the probability that they like the feature is $\frac{950}{1000}$. For the second person, it is $\frac{949}{999}$ and so on until the fifth person. Thus the probability is

$$
\begin{aligned}
&= \frac{950}{1000} \cdot \frac{949}{999} \cdot \frac{948}{998} \cdot \frac{947}{997} \cdot \frac{946}{996} \\
&= 0.774
\end{aligned}
$$

Question 21 - Fan Groups

This problem was asked by Snapchat.

You are testing a new feature with various sample groups of three people. Assume that each person is equally likely to be a fan or not a fan of the feature. What is the probability that a randomly chosen group has exactly one fan, given that there is a fan among the three?

Solution

This problem is easiest solved by counting.

Let the value 1 indicate that a person is fan and 0 that they are not a fan. In a group of 3 people, there are $2^3 = 8$ total possibilities. Given that there is atleast one fan, the total possible groups are 7. The total number of groups where there are exactly one fan are 3. Therefore, the answer is

$$
\begin{aligned}
&= \frac{3}{7} \\
&\approx 0.433
\end{aligned}
$$

Question 22 - Hit Show

This problem was asked by Netflix.

Before a show is released, it is shown to several in-house raters. You assume there are two types of shows: hits, which have an 80% chance of being liked by any viewer, and misses, which have a 20% chance of being liked by any viewer. There is currently a new show which you believe has a prior distribution of 60% being a hit, and 40% being a miss. Given that 8 raters rated the show and 6 of the 8 liked the show, what is the new posterior distribution of being a hit or miss?

Solution

We are asked to find the posterior probability that a new show is a hit given that 6 out of 8 people liked this show. This can be done using the bayes rule.

$$p(\text{hit} \mid 6/8 \text{ liked}) = \frac{p(6/8 \text{ liked } \mid \text{hit}) \cdot p(\text{hit})}{p(6/8 \text{ liked} \mid \text{hit}) \cdot p(\text{hit}) + p(6/8 \text{ liked } \mid \text{miss}) \cdot p(\text{miss})}$$

The $p(6/8 \text{ liked} \mid \text{hit})$ term is distributed binomially with parameters that change depending on if the show is a hit or a miss. If it is a hit, it is $\text{Bin}(n = 8, p = 0.8)$ and if it is a miss, $\text{Bin}(n = 8, p = 0.2)$. Thus, we have

$$\begin{aligned} p(\text{hit} \mid 6/8 \text{ liked}) &= \frac{\binom{8}{6} \cdot 0.8^6 \cdot 0.2^2 \cdot 0.6}{\binom{8}{6} \cdot 0.8^6 \cdot 0.2^2 \cdot 0.6 + \binom{8}{6} \cdot 0.2^6 \cdot 0.8^2 \cdot 0.4} \\ &\approx \boxed{0.998} \end{aligned}$$

The new posterior distribution of the show, given that 6 out of 8 people liked it is that it is a hit with 0.998 probability and miss with $1 - 0.998 = 0.002$ probability.

Question 23 - Waiting Time

You are modeling the wait time a customer has for a support call as exponentially distributed with a mean of 10 minutes. Suppose a customer calls in and is told that all lines are currently busy, and the most recent last spot was occupied 5 minutes ago. What is the probability that the current customer will need to wait no more than another 5 minutes?

Solution

An exponential distribution has the property of **memorylessness**, which means that the waiting time of the event does not depend on how much previous time has passed. We are given the mean wait time as 10 minutes. In other words the parameter λ of the exponential distribution equals $\frac{1}{10}$
The CDF is given by

$$\begin{aligned} P(X \leq x) &= 1 - e^{-\lambda \cdot x} \\ P(X \leq 5) &= 1 - e^{-\frac{1}{10} \cdot 5} \\ &\approx 0.39346 \end{aligned}$$

Thus, the probability that the wait time is no more than 5 minutes given that the last known spot was 5 minutes ago, with mean wait time 10 minutes is 0.393.

Question 24 - Favorite Show

This problem was asked by Disney.

Alice and Bob are choosing their top 3 shows from a list of 50 shows. Assume that they choose independently of one another. Being relatively new to Hulu, assume also that they choose randomly within the 50 shows. What is the expected number of shows they have in common, and what is the probability that they do not have any shows in common?

Solution

Let $\{A_1, A_2, A_3\}$ be Alice's choice and $\{B_1, B_2, B_3\}$ be Bob's choice. The expected number of shows they have in common can be calculated using **linearity of expectation**.

Let $I_i = 1$ if $A_i = B_1$ or $A_i = B_2$ or $A_i = B_3$ else 0 for $i = 1, 2, 3$. $I_i \sim \text{Bern}(p)$

The total expectation is then given by $E[I_1] + E[I_2] + E[I_3]$. Each of these terms are identical.

$$\begin{aligned} p(I_i = 1) &= \frac{1}{50} + \frac{\cancel{49}}{50} \cdot \frac{1}{\cancel{49}} + \frac{\cancel{49}}{50} \cdot \frac{\cancel{48}}{\cancel{49}} \cdot \frac{1}{\cancel{48}} \\ &= \frac{3}{50} \\ E[I_i] &= \frac{3}{50} \end{aligned}$$

The total expectation is $\frac{3}{50} + \frac{3}{50} + \frac{3}{50} = \boxed{0.18}$

Secondly, the probability that they have no show in common is $\frac{47}{50} \cdot \frac{46}{49} \cdot \frac{45}{48} \approx \boxed{0.83}$

Question 25 - Bernoulli Samples

Consider a Bernoulli random variable with parameter p. Say you observe the following samples: $[1, 0, 1, 1, 1]$. What is the log likelihood function for p and what is the MLE of p?

Solution

Let us define the dataset as $\{x_1, x_2, \cdots x_n\}$. Assuming that each of these have a bernoulli distribution and are independent, we have $x_i \sim \text{Bern}(p)$ for $i = \{1 \cdots n\}$ The likelihood of the dataset of n points is given by:

$$\begin{aligned}
\mathcal{L}(p) &\propto p(x_1, x_2, \cdots x_n \mid p) \\
&\propto \prod_{i=1}^{n} p^{x_i} \cdot (1-p)^{(1-x_i)} \\
&\propto \sum_{i=1}^{n} x_i \log p + (1 - x_i) \log(1-p)
\end{aligned}$$

The MLE estimate of p is is the value of p that maximizes the above function. To find the maximum, we use calculus. Taking the derivative with respect to p and setting it to 0, we have:

$$\begin{aligned}
\sum_{i=1}^{n} \left(\frac{x_i}{p} - \frac{(1-x_i)}{(1-p)} \right) = 0 \\
\frac{1}{p} \cdot \sum_{i=1}^{n} x_i &= \frac{1}{(1-p)} \left(n - \sum_{i=1}^{n} x_i \right) \\
\frac{\sum_{i=1}^{n} x_i}{p} &= \frac{1}{(1-p)} \cdot \left(n - \sum_{i=1}^{n} x_i \right) \\
\hat{p}_{\text{MLE}} &= \boxed{\frac{\sum_{i=1}^{n} x_i}{n}}
\end{aligned}$$

Therefore, the MLE of the above dataset is $\hat{p}_{\text{MLE}} = \frac{4}{5} = \boxed{0.8}$