91.673 Advanced Database Systems
Homework1

Chang Liu

chang_liu@student.uml.edu

September 12, 2015

Problem1. We flip a fair coin ten times. Find the probability of the following events.

(a) The number of heads and the number of tails are equal.

(b) There are more heads than tails.

(c) The ith flip and the $(11 - i)$th flip are the same for $i = 1, ..., 5$.

Solution:

(a) The number of heads equals to the number of tails when both of them are **5**. In that situation the head is 5(and the tail is also 5 in this case), so the probability is the selection from ten times that five is head and others are 5, since each of them are fair so the probability for one event is 1/2, so the overall probability is:

$$C_{10}^5 * (1/2)^5 * (1/2)^{(10-5)} \approx 0.246$$

(b) First list all the situation that heads is more than tails, it is:

1) 6 heads and 4 tails,

2) 7 heads and 3 tails,

3) 8 heads and 2 tails,

4) 9 heads and 1 tail,

5) 10 heads and 0 tail.

In all these situations, calculate their probability and add them all, since they're all the possible candidates. So the result should be:

$$(C_{10}^6 + C_{10}^7 + C_{10}^8 + C_{10}^9 + C_{10}^1 0) * (1/2)^{10} \approx 0.377$$

(c) All the situations are as follows.

1) for $i = 1$, it means the 1st and the 10th flip are the same, then both of them are heads or tails, the probability is 0.5

2) for other $i = 2, 3, 4, 5$, it is quite similar.

In total, the probability is that all these five requirements should be meet, so it is $(0.5)^5 = \frac{1}{2^5} = \frac{1}{32}$

Problem2. Consider the following game, played with three standard six-sided dice. If the player ends with all three dice showing the same number, she wins. The player starts by rolling all three dice. After this first roll, the player can select any one, two, or all of the three dice and re-roll them. After this second roll, the player can again select any of the three dice and re-roll them one final time. For the following questions, assume that the player used the following optimal strategy: if all three dice match, the player stops and winds; if two dice match, the player re-rolls the die that does not match; and if no dice match, the player re-rolls them all.

(a) Find the probability that all three dice show the same number on the first roll.

(b) Find the probability that exactly two of the three dice show the same number after the first roll.

(c) Find the probability that the player wins, conditioned on exactly two of the three dice showing the same number after the first roll.

Solution:

a) There're six values that can be show in one dice, and if all three dice are the same, then it should be in this situation, and the probability is the product of C_6^1 and their probability of happening in this same situation. So the result is:

$$C_6^1 * (1/6)^3 = 1/6^2 = \frac{1}{36}$$

b) There're also six values that the two dice are the same, so it is $C_6^1 = 6$ options. And in each option, the other dice should be a different value, which value is $5/6$, and the two dice should be the same value, it's probability is that $(1/6)^2 = 1/36$. And there are C_3^2 situations there. In total, the probability is:

$$C_3^2 * C_6^1 * (1/6)^2 * (1 - 1/6) = \frac{5}{12}$$

c) There are two situations that: 1) the second roll wins. Then the probability is just $\frac{1}{6}$, since the number must be the same as the first roll. 2) the third roll wins. Then the second roll fails and the probability is that $\frac{5}{6} * \frac{1}{6} = \frac{5}{36}$ So the result is:

$$\frac{1}{6} + \frac{5}{36} = \frac{11}{36}$$

Problem3. We have a function $F : \{0, ..., n - 1\} \to \{0, ..., m - 1\}$. We know that, for $0 \leq x, y \leq n - 1, F((x + y) \bmod n) = (F(x) + F(y)) \bmod m$. The only way we have for evaluating F is to use a lookup table that stores the values of F. Unfortunately, an Evil Adversary has changed the value of $1/5$ of the table entries when we were not looking.

Describe a simple randomized algorithm that, given an input z, outputs a value that equals $F(z)$ with probability at least $1/2$. Your algorithm should work for every value of z, regardless of what values the Adversary changed. Your algorithm should use as few lookups and as little computation as possible.

Suppose I allow you to repeat your initial algorithm three times. What should you do in this case, and what is the probability that your enhanced algorithm returns the correct answer?

Solution:

1) Suppose that we randomly selection one value from $\{1, ...n-1\}$ that are less then z, set it as the x, and then the other value y, we set it as $z - x$, then according to the equation's property, we can conclude that:

$$F(x + z - x \bmod n) = [F(x) + F(z - x)] \bmod m$$

Further, since z is any input from $\{1, ...n-1\}$, then it is less than n, so the above equation should be:

$$F(z) = [F(x) + F(z - x)] \bmod m$$

As $z \bmod n$ equals z, and from the condition that some value is changed, if $F(x)$ is not the real value, then the probability that it's changed is $P(F(x) \ changed) = 1/5$, and similarly, $P(F(z - x) \ changed) = 1/5$.

But there is some correlation between $F(x)$ and $F(z - x)$, they're bounded to $F(z)$, according to the first equation to calculate $F(z)$, at least one of it is changed then $F(z)$ should be error,

$$P(F(z) \ changed) = P(F(x) \ or \ F(z - x) \ changed)$$
$$= P(F(x) \ changed) + P(F(z - x) \ changed) - P(both \ changed) \qquad (1)$$
$$\leq 1/5 + 1/5 = 2/5 = 0.4$$

So the error rate for $F(z)$ is less than 0.4, and the correct rate is larger than $0.6 > 1/2$, which meets the requirement for the algorithm.

2) Do that experiment repeatably to reduce the error rate, choose the majority of the result of our judgment, for example, for the 3-times experiment, only by more than 2 times it says the value is wrong, then we can say that it's wrong.

In this way, we can calculate the probability that it's actually wrong,

a) 2/3 has predicted it's wrong, 1/3 has predict it right, then probability is

$$C_3^2 * 0.4^2 * 0.6 = 0.288$$

b) 3/3 has predicted it's wrong, then probability is

$$C_3^3 * 0.4^3 = 0.064$$

3

So the total error rate is $0.288 + 0.064 = 0.352$, the probability of estimating it right is ≥ 0.648.

Problem4. . Suppose we roll a fair k-sided die with the numbers 1 through k on the die's faces. If X is the number that appears, what is $E[X]$?

Solution:
The expectation is the sum of each value multiply its probability, so the equation should be as follows

$$1 * (1/k) + 2 * (1/k) + \ldots + k * (1/k) = \frac{1}{k} * \sum_{i=1}^{k} (i) = \frac{k+1}{2}$$

So the result expectation $E[X] = \frac{k+1}{2}$.

Problem5. A monkey types on a 26-letter keyboard that has lowercase letters only. Each letter is chosen independently and uniformly at random from the alphabet. If the monkey types 1,000,000 letters, what is the expected number of times that sequence 'proof' appears?

Solution: Here we can assume an Event X_i as seeing the word 'proof' at the index i, i should be from 0 to $999,995$, because the last chance should take all the $1,000,000$ letters, not the last one word since it has the length of 5. and the probability of 'proof' is $(1/26)^5$, so we can say that:
1) if we saw the word 'proof', then $X_i = 1$
2) otherwise, the value $X_i = 0$.
Then in this way, the expectation $E(X_i)$ is the expected number of times for the sequence 'proof' occurs.
And from this point, we can get the result:

$$E(X_i) = 999996 * (1/26)^5 \approx 0.084$$