Probability And Statistics for Interview Workshop

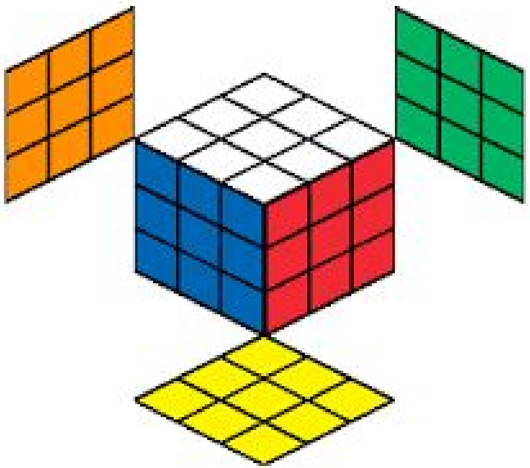
Columbia Data Science Society

Questions:

1. Say we have one 3x3x3 cube made out of 27 individual 1x1x1 cubes. Suppose I were to paint the outside of the large cube entirely red and then break it apart into the little cubes. Say we then roll one of the 27 cubes and all the sides we can see are blank. What is the probability that the remaining side is also blank?
2. In a variant of Deal or No Deal, 4 boxes each contains a single US bill. One box contains a $100, two boxes contain a $20, and the remaining box contains a $5. Say I select two bills at random from the 4 boxes, look at both of them, then put a $20 bill in my pocket. I offer you to either pick the bill in my other hand, or to choose at random a bill from the two remaining boxes. What is your strategy? What if I were to put a $5 bill in my pocket instead?
3. Say I'm auctioning a chest of gold coins. The chest contains a random integer of coins between 0 and 100, inclusive. Each coin is worth $1.50. If your bid is greater than or equal to the number of coins in the chest, you buy the chest for your bid amount of dollars. If your bid is lower than the number of the coins, nothing happens. What is your strategy?
4. Say I have 100 dice and then I roll all of them: what is the probability that I get more than 400 dots face up?

Question 1

Say we have one 3x3x3 cube made out of 27 individual 1x1x1 cubes. Suppose I were to paint the outside of the large cube entirely red and then break it apart into the little cubes. Say we then roll one of the 27 cubes and all the sides we can see are blank. What is the probability that the remaining side is also blank?

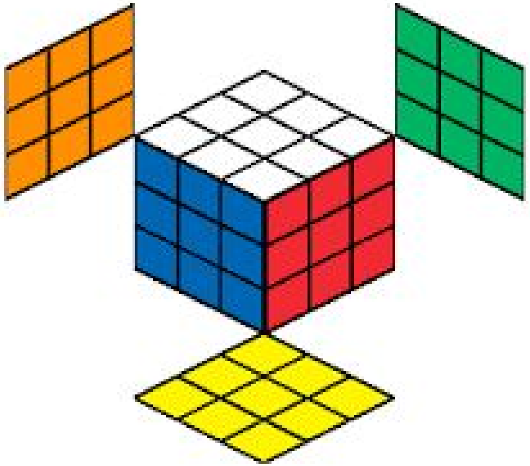


Question 1 (solution)

The answer is **not** 1/7! Why not?

The key to this problem is Bayes Theorem. If you don't remember what it is:

P(A|B) = P(B|A)*P(A)/P(B)



Question 1 (solution cont.)

In this case, we're considering the probability that the bottom face is blank, given that the five visible sides are all blank as well. **Note, we can also think of this as the probability all the sides of the cube are blank given that the five sides we see are blank.**

We can start out simple: $P(B|A) = 1$ (probability five sides shown are blank given all sides are blank)

We know that $P(A) = 1/27$, because there's only one cube (the center-most one) that is entirely unpainted.

Finally, the probability that a randomly chosen cube shows five blank sides is $P(B) = (1/6)(6/27) + 1/27 = 2/27$. Each of the 6 cubes with one painted face has a 1/6 chance of having its painted side face down.

Finally, plugging everything in, we get that the desired probability is **1/2!**

Principles: Visualize the problem, Bayes' Rule, clarify the question + ask your interviewer to repeat it

Question 2

In a variant of Deal or No Deal, 4 boxes each contains a single US bill. One box contains a $100, two boxes contain a $20, and the remaining box contains a $5. Say I select two bills at random from the 4 boxes, look at both of them, then put a $20 bill in my pocket. I offer you to either pick the bill in my other hand, or to choose at random a bill from the two remaining boxes. What is your strategy? What if I were to put a $5 bill in my pocket instead?

$100	$20	$20	$5

Question 2 (solution)

The key to this question is to look at the problem conditionally:

$100	$20	$20	$5

Sometimes, the easiest way is to list out all the possibilities! How many combinations of (L,R) - left hand bill, right hand bill - are there that contain at least one $20?

1. ($100, $20) - leaving $20 and $5 remaining
2. ($20, $100) - leaving $20 and $5 remaining
3. ($20, $5) - leaving $100 and $20 remaining
4. ($5, $20) - leaving $100 and $20 remaining
5. ($20, $20) - leaving $100 and $5 remaining.

Question 2 (solution cont.)

\$100	\$20	\$20	\$5

Now we can calculate the EV of keeping the bill versus redrawing bill for each case!

1. (\$100, \$20) - EV of keeping is \$100, redrawing is \$(20+5)/2 = \$12.5
2. (\$20, \$100) - EV of keeping is \$100, redrawing is \$(20+5)/2 = \$12.5
3. (\$20, \$5) - EV of keeping is \$5, redrawing is \$(100+20)/2 = \$60
4. (\$5, \$20) - EV of keeping is \$5, redrawing is \$(100+20)/2 = \$60
5. (\$20, \$20) - EV of keeping is \$20, redrawing is \$(100+5)/2 = \$52.5

Keeping EV = \$100+100+5+5+20 = **\$230** Redrawing EV = \$12.5+12.5+60+60+52.5 = **\$197.5**

Thus, our strategy is to keep the bill in the other hand!

Principles: Break down cases, Expected Value

Question 2 (follow up questions)

$100	$20	$20	$5

1. What if I put a $5 bill in my pocket instead?
2. What if it were a $50 bill instead of a $100 bill?
3. For what “bill” would it not matter what decision we choose?

Principles: Break down cases, Expected Value

Question 3

Say I'm auctioning a chest of gold coins. The chest contains a random integer of coins between 0 and 100, inclusive. Each coin is worth $1.50, and you want to place a bid to buy the chest of gold coins. The rules are as follows: if your bid is greater than or equal to the number of coins in the chest, you buy the chest for your bid amount of dollars; if your bid is lower than the number of the coins, nothing happens. What is your optimal strategy?

Question 3 (solution)

This is another question where we will have to figure out expected values.

The best thing to do is always to try and convert what we know into equations:

Suppose that X is the number of coins in the chest, and we know that X is uniformly distributed between 0 and 100. Now suppose that B is the number that we bid. We want to choose B so that we maximize our return.

What should we do first?

Question 3 (solution cont.)

We need to write a function for expected return in terms of X and B! The key is the break it down into the cases where we win the auction vs. we lose the auction.

$E(x) = 0$	if $B < X$	(we lose the auction so nothing happens)
$E(x) = 1.5X - B$	if $B \geq X$	(we win the auction, so we pay out bid amount for the number of coins)

Now, the key question here is: how do we maximize E(x)? It's clear we only need to consider the case where $B \geq X$.

Let's think more about expected value now. We know that X is uniformly distributed. If we know that we win the auction, **what's the expected value of X in terms of B**?

Question 3 (solution cont.)

The expected value of X if we win the auction if B/2!

This means our expected return function is now:

$E(x) = 0$ if $B < X$
$E(x) = 1.5(B/2) - B = -B/4$ if $B \geq X$

It turns out that our expected return when we win the auction is actually negative! Does this make sense intuitively?

Thus, our strategy should always be to bid 0, as that will "maximize" our return.

Principles: More Expected Value!, write out your thoughts + steps

Question 4

Say I have 100 dice and then I roll all of them: what is the probability that I get more than 400 dots face up?

Question 4 (solution)

- Number of dots face up = sum over all die of each dice's value
- Number of dots face up distributed as sum of 100 DiscreteUniform(6) variables
    - DiscreteUniform(n) means each value between 1 and n (inclusive) has probability 1 / n
- 100 ≅∞, so let's use the Central Limit Theorem
    - Central Limit Theorem says mean of **n** i.i.d. random variables is distributed normally with **mean equal to original random variable** and **variance equal to variance of original random variable divided by n**
        - i.i.d. = independent and identically distributed
- Sum is 400 ↔ Mean is 4
- Distribution of Mean of All Rolls
    - Mean of Original = Mean of One Roll: 1 * ⅙ + 2 * ⅙ + 3 * ⅙ + 4 * ⅙ + 5 * ⅙ + 6 * ⅙ = (7 + 7 + 7) * ⅙ = 21 / 6 = 7 / 2
    - Variance of Original = Variance of One Roll:
        - (1 - 7 / 2)^2 * ⅙ + (2 - 7 / 2)^2 * ⅙ + (3 - 7 / 2)^2 * ⅙ + (4 - 7 / 2)^2 * ⅙ + (5 - 7 / 2)^2 * ⅙ + (6 - 7 / 2)^2 * ⅙
        - (-5 / 2)^2 * ⅙ + (-3 / 2)^2 * ⅙ + (-1 / 2)^2 * ⅙ + (1 / 2)^2 * ⅙ + (3 / 2)^2 * ⅙ + (5 / 2)^2 * ⅙
        - (2 * 25 / 4 + 2 * 9 / 4 + 2 * 1 / 4) * ⅙ = 70 / 4 * 1 / 6 = 70 / 24 = 35 / 12 ≅ 3
    - Mean of Mean of All Rolls: 3.5
    - Variance of Mean of All Rolls: 3 / 100
        - Standard Deviation of All Rolls: sqrt(3) / 10 ≅ 0.17

Question 4 (solution cont.)

- Mean of All Dots distributed as Normal(3.5, 0.17)
- Mean > 4 ↔ Normal Variable > $\mu + 3\sigma$ ↔ 0.5%

Principles: Central Limit Theorem, don’t be afraid to do some calculations, normal bounds (68-95-99 rule)