Brain Teasers for interview preparation

2024-07-04

Many problems are taken from A Practical Guide to Quantitative Finance Interviews by Xinfeng Zhou.

Contents

1   Useful facts learnt along the way

- The Cayley Hamilton theorem states that a square matrix A satisfies its own characteristic equation. Useful for finding the inverse of a matrix.

2   Brain Teasers

2.1   Problem simplifications

I had already seen similar problems, the solutions are not surprising nor complicated.

Q: Five pirates looted a chest full of 100 gold coins. Being a bunch of democratic pirates, they agree on the following method to divide the loot. The most senior pirate will propose a distribution of the coins. All pirates, *including the most senior pirate*, will then vote. If at least 50% (inclusive?) of the pirates accept the proposal, the gold is divided as proposed. If not, the most senior pirate will be fed to shark and then process starts over with the next most senior pirate... The process is repeated until a plan is approved. All pirates are perfectly rational: above all, they want to stay alive, and then get as much gold as possible. Being blood thirsty pirates, they want to have fewer pirates on the boat if given a choice between otherwise equal outcomes. How will the gold coins be divided in the end?
A: It is probably better to start with simpler scenarios. I will assume that the "at least" is not inclusive.

1 pirate The *senior pirate* will get all the 100 gold coins.

2 pirates The 2nd senior pirate will have no interest in any offer that doesn't give him 100 gold coins: he will always vote against the proposal (even if the senior pirate proposes to get give him all the gold).

3 pirates The 2nd senior pirate has no interest in becoming the next senior pirate and will always vote in favour of what's proposed. The senior pirate will therefore propose to keep all the gold for himself, and win.

By induction, if the number of pirates is odd, the senior pirate will get all the gold, and if the number of pirates is even, the second senior pirate will get it.
The solution actually assumed an inclusive vote, where 50% votes is enough to win. This makes the game more interesting, and makes it so that

Q: 100 tigers and 1 sheep are put in a magic island that only has grass. Tigers can eat grass, but prefer to eat sheep. Assume that: only one tiger can eat the sheep, and that a tiger is transformed into a sheep after eating it. All tigers are perfectly rational and want to survive. Will the sheep be eaten?

A: if there were only one tiger, the sheep would be eaten. If there were two tigers, no tiger would eat the sheep, as the other tiger would then eat them afterwards. By induction, we see that the sheep will be eaten if there are an odd number of tigers, and will survive if there are an even number of tigers (e. g. 100).

2.2 Logic Reasoning

Q: Four people, A, B, C and D need to get across a river. The only way to cross the river is by an old bridge, which holds at most 2 people at a time. Being dark, they can't cross the bridge without a torch, of which they only have one. So each pair can only walk at the speed of the slower person. They need to get all of them across to the other side as quickly as possible. A is the slowest and takes 10 minutes to cross; B takes 5 minutes; C takes 2 minutes; and D takes 1 minute. What is the minimum time to get all of them across the other side?
A: Idea: we should always send the fastest person back. A and B should go together, but not first. So we'd have C, D go together, D goes back and brings the torch to A, B, which go together, sending C back, and then having C, D go together. It wouldn't make a difference to bring C back the first time. This takes $2 + 1 + 10 + 2 + 2 = 17$ minutes.

Q: Long description
A: Not hard, nor interesting. Just go step by step.

Q: A casino offers a card game using a normal deck of 52 cards. The rule is that you turn over two cards each time. For each pair, if both are black, the go to the dealer's pile; if both are red, they go to your pile; if one black and one red, they are discarded. The process is repeated until you two go through all 52 cards. If you have more cards in your pile, you win \$100; otherwise (including ties) you get nothing. The casino allows you to negotiate the price you want to pay for the game. How much would you be willing to play this game?
A: Intuitively, the expected value of the game is 0. For any game, a number $n \leq 26$ of pairs will be discarded. There will then remain $26 - n$ pairs that are of the same colour. Half of these will go to the dealer, and half to the player, resulting in a draw.
Not complicated.

Q: You have two ropes, each of which takes 1 hour to burn. But either rope has different densities at different points, so there's no

guarantee of consistency in the time it takes different sections within the rope to burn. How do you use these two ropes to measure 45 minutes?
A: Consider all the options you have. You can burn both ends of the ropes, hence halving the burning time to 30 minutes. We are still not there, and we do not want to burn all our ropes by minute 30. Instead, burn one rope from both ends, and burn the other from only one ends. After 30 minutes, the first rope will have burnt completely, while the second one will be halfway. Burn the other end of the second rope, to half the remaining time to 15 minutes. In total, this takes 45 minutes.

Q: You have 12 identical balls. One of the balls is heavier OR lighter than the rest. Using just a balance that can only show you which side of the tray is heavier, how can you determine which ball is the defective one with 3 measurements?
A: I think it is better to answer this question with a nested list. Split the balls into three groups of 4. Compare the weight of the first two groups.

- If they are identical, the weird ball is in the third group, and we have 2 measurements to find it. We further split the third group into two subgroups of 2 and we use one measurement to compare the weight of the two balls in the first subgroup.

    - If they are the same, then the weird ball is in the second subgroup. We compare one ball in the second subgroup with a normal ball.

        * If the weight differs, the ball from the subgroup is the weird one.
        * If not, the other ball in the subgroup is the weird one.

    - If they are not the same, then the weird ball is in the first subgroup Compare one of these balls with a normal ball and proceed as before.

- If they are not the same, then the weird ball is in one of the first two groups. ... *need to think about this more clearly. Read the solution and I will sleep on it.*

Q: How many trailing zeroes are there in $100!$?
A: Using the prime factorisation of $x = 100!$, we can write:

$$x = \prod_{p \notin \{2,5\}} p_p^n \cdot 2^{n_2} \cdot 5^{n_5}.$$

The number of trailing zeroes is then $\min(n_2, n_5)$. As multiple of twos are more common than multiple of fives, we simply have to count how many 5 factors there are among the first 100 integers. These are $20 + 4 = 24$, so there are 24 trailing zeroes in $100!$.

Q: There are 25 horses, each of which runs at a constant speed that is different from the other horses'. Since the track only has 5 lanes, each race can have at most 5 horses. If you need to find the 3 fastest horses, what is the minimum number of races needed to identify them?
A: I am not sure whether we can measure times. In this scenario, I think we'd need 5 races (one race for each horse). Therefore, assume we do not measure them. I can prove a solution with 7 races exists. I would only be left to prove that a solution with 6 rates exists. To prove 7 races exist: 5 races to make 5 groups of 5 race. Then 1 race for the winners (this excludes the groups where the 4th and 5th horses previously won, as well as the 2nd and 3rd of the $\ldots$). The first position is known, and 5 horses are left with unclear positions just after: let them race.

Q: If $\hat{x}\hat{x}\hat{x}\cdots = 2$, what is x?
A: The idea with these kind of questions is to exploit the infiniteness of these series. We have $x^2 = 2$ which implies $x = \sqrt{2}$, as we exclude the negative solution.

2.3   Thinking out of the box

Q: Can you pack 53 bricks of dimensions 1 $1 \times 1 \times 4$ into a $6 \times 6 \times 6$ cube?
A: Intuitively, I think that $1 + 1 + 4 = 6$, which may hint some strategy. In terms of maths, let us make basic checks. The volume of the cube is $6^3 = 216$, while the volume of the bricks is $4 \cdot 53 = 212$. We have 4 units of volume that we can leave empty.
The solution tells us to think of the 27 $2 \times 2 \times 2$ sub-cubes that make up the bigger cube. We can color these small cubes in 2 colours, such that 2 adjacent small cubes have different colours. Without loss of generality, 14 will be white and 13 will be black. Now, each brick in the configuration will be placed in a way that occupies the same volume of white and black cubes. This means that we can at most cover 13 black cubes and white cubes, leaving a volume equivalent to a white cube empty (that is 8 units). We would still need to back an extra brick in the cube.
Once understood, it's easy. However it is hard to think about this from scratch. The solution is very elegant, creating an "imbalance" that can't be broken.

Q: You just had two dice custom-made. Instead of the numbers 1 to 6, you place single-digit numbers on the faces of each dice so that every morning you can arrange the dice in a way as to make the two front faces show the current day of the month. You must use both dice (1 must be shown as 01), but you can switch the order of the

dice if you want. What numbers do you have to put on the six faces of each of the two dice to achieve that?
A: How many combinations of values do we need? We need to count until 31, but to avoid double counting, we only consider numbers with digits reshuffled in increasing order. There are 9 combinations starting with 0 (from 01 to 09); 9 combinations starting with 1 (from 11 to 19); 8 combinations starting with 2 (from 22 to 29); 0 combinations starting with 3. This makes up a total of 26 combinations. Both dices must include the numbers 1 and 2, else we cannot represent the numbers 11 and 12, and both must include 0, else we can only represent 6 numbers leading with 0. We are left with 3 spots per dice. This is not enough: we have 7 spots for 6 numbers. The stupid solution is just to use the 6 as a 9, stupid but smart.

Q: You are facing two doors. One leads to your job offer and the other leads to exit. In front of either door is a guard. One guard always tells lies and the other always tells the truth. You can only ask one guard one yes/no question. Assuming you do want to get the job offer, what question will you ask.
A: We need to find a question such that both guards would give the same answer if the job offer is behind either door. We can maybe have a double condition: such as "is this the door, and are you honest?"

2.4 Application of Symmetry

Q: Suppose that you are blind folded in a room and are told that there are 1000 coins on the floor. 980 of the coins have tails up and the other 20 coins have heads up. Can you separate the coins into two piles so to guarantee both piles have equal number of heads? Assume that you cannot tell a coin's side by touching it, but you are allowed to turn over any number of coins.
A: Take 20 coins, and flip them. (this works in general, write proof).

Q: You are given 3 bags of fruits. One has apples in it; one has oranges in it; and one has a mix of apples and oranges in it. Each bag has a label on it (apple, orange or mix). Unfortunately, your manager tells you that ALL bags are mislabeled. Develop a strategy to identify the bags by taking out minimum number of fruits? You can take any number of fruits from any bag.
A: Pick a single fruit from the mixed bag. Proceed by exclusion.

Q: A sultan has captured 50 wise men. He has a glass currently standing bottom down. Every minute he calls one of the wise men

who can choose either to turn it over (set it upside down or bottom down) or to do nothing. The wise men will be called randomly, possibly for an infinity number of times. When someone called to the sultan correctly states that all wise men have already been called to the sultan at least once, everyone goes free. But if his statement is wrong, the sultan puts everyone to death. The wise men are allowed to communicate only once before they get imprisoned into separate rooms. Design a strategy that lets the wise men go free.

A: 49 YOFO (you only flip once) and one counter which counts how many times the cup has been flipped. *Very clean use of simmetry in the strategy.*

2.5 Series summations

Q: A clock (numbered 1 to 12) fell off the wall and broke into three pieces. You find that the sums of the numbers on each piece are equal. What are the numbers on each piece?

A: Each piece must have a sum of $\frac{12 \cdot 13}{2 \cdot 3} = 26$. As $\frac{26}{3} = 12$, it is impossible to achieve this sum with three numbers on the clock, so we must have four numbers on each piece. Now we need a bit of inspection: $11 + 12 + 1 + 2 = 26$ and $5 + 6 + 7 + 8 = 26$. We are left with $3 + 4 + 9 + 10$. (Nice to think of it as having pairs of numbers that sum to 13).

Q: Suppose we have 98 distinct integers from 1 to 100. What is a good way to find out the two missing numbers

A: I am confused about what "good" means: time complexity? The solutions compute $x + y$ and $(x + y)^2$ as a function of the sum of the 98 numbers and the sum of the squares.

Q: There are 10 bags with 100 identical coins in each bag. In all bags but one, each coin weighs 10 grams. However, all the coins in the counterfeit bag weigh either 9 or 11 grams. Can you find the counterfeit bag in only one weighting, using a digital scale that tells the exact weight? (allowed to take any number of coins from any bag)

A: The weighting strategy needs to discriminate the original bag from which the counterfeit bag is. This means that we have to take a different amount of coins from each bag. The pigeonhole principle tells us the only option is to take n coins from the nth bag. Denote the index of the counterfeit bag as i, and let w be equal to 1 if the coins in the bag weigh 11 grams, and -1 otherwise. The sum of the weights of this coins will be

$$\sum_{n=1}^{10} 10n + iw = \frac{10 \cdot 10 \cdot 11}{2} - iw = 550 - iw.$$

Is is easy to see that subtracting 500 from the scale reading will reveal both the "sign" of the difference in counterfeit weight, and the bag.

Q: You are holding two balls in a 100-story building. If a ball is thrown out of the window, it will not break if the floor number is less than X, and it will always break if the floor number is equal to or greater than X. You would like to determine X. What is the strategy that will minimize the number of drops for the worst case scenario?
A: For any strategy, we need to compute the worst case scenario. Mathematically, we interested in finding a number in the range $N = \{1, 2, \dots, 100\}$ asking questions such as: is $X \geq n$ for any $n \in N$. It should be possible to get this in: $\lceil \log_2 100 \rceil = 7$ iterations, by always throwing the ball from the middle floor of the remaining range.
This is not the solution but honestly I don't understand what they are doing. Feels like we are answering different questions?

3   Calculus and Linear Algebra

- logdiff trick can be useful when computing derivatives of functions of the form $f(x)^{g(x)}$.

- sec and tan work together.

3.1   Limits and derivatives

Q: What is the derivative of $y = (\log x)^{\log x}$

A: Taking the natural logarithm of both sides, we have:

$$\log y = \log x \cdot \log(\log x).$$

Where we recall that $\frac{d}{dx}\log x = \frac{1}{x}$. Thus, taking derivatives on both sides with respect to x and using the product rule:

$$\frac{1}{y}\frac{dy}{dx} = \frac{d}{dx}(\log x) \cdot \log(\log x) + \log x \cdot \frac{d}{dx}(\log(\log x))$$

$$= \frac{1}{x}\log(\log x) + \log x \cdot \frac{1}{x}\frac{1}{\log x}.$$

So:

$$\frac{dy}{dx} = \frac{(\log x)^{\log x}}{x}\left(\log(\log x) + 1\right)$$

This requires some care, I did get some mistakes

Q: Without calculating the numerical results, can you tell me which number is larger: e^{π} **or** π^{e}**?**

A: We are interested in determining whether $e^{\pi} > \pi^{e}$, which holds if and only if $\pi \log e > e \log \pi$ or $\frac{\log e}{e} > \frac{\log \pi}{\pi}$. As such, consider the function $f(x) = \frac{\log x}{x}$, which we want to maximize. The function tends to 0 at $x = 0$ and at $x = \infty$, it is continuous and differentiable, so it must have at least a local maximum. We can find this:

$$0 = \frac{d}{dx}f(x) = \frac{\frac{1}{x}\cdot x - 1 \cdot \log x}{x^2}.$$

For $x \neq 0$, this holds for $1 = lnx$, which implies $x = e$ is the only local (and hence global) maximum. This implies $e^{p}i > \pi^{e}$.

Again, I got it wrong, but reading the first part of the solution helps me remember how to derive the relevant function to maximize

3.2   Integration

Q: What is the integral of $\log x$?
A: We proceed with integration by parts:

$$\int 1 \cdot \log x \, \mathrm{d}x = x \log x - \int x \frac{1}{x} \mathrm{d}x + c = x(\log x - 1) + c.$$

Q: What is the integral of $\sec(x)$ from $x = 0$ to $x = \frac{\pi}{6}$?
A: It is hard to remember the solution from scratch, so we remember to derive relationships of sec and tan. We have:

$$\frac{d \sec(x)}{dx} = \frac{d(\cos x)^{-1}}{dx} = -(-1)\sin x(\cos x)^{-2} = \sec x \tan x$$

$$\frac{d \tan(x)}{dx} = \frac{d \frac{\sin x}{\cos x}}{dx} = \frac{\cos^2 x - (-\sin^2 x)}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x.$$

The relationship appears when we consider the sum $g(x) := \sec(x) + \tan(x)$.

$$\frac{dg(x)}{dx} = \sec x(\tan x + \sec x) = \sec x g(x).$$

What does this mean exactly? Can we step away for one second? We know that:

$$\sec(\pi/6) - \sec(0) = \int_0^{\frac{\pi}{6}} \frac{\mathrm{d} \sec x}{\mathrm{d}x} \mathrm{d}x$$

Here I haven't got an expression in $g(x)$ even though I could add and subtract $\frac{\mathrm{d} \tan x}{\mathrm{d}x}$. *So maybe let us go and find $g(x)$.* We know:

$$0 = \frac{\mathrm{d}g(x)}{\mathrm{d}x} - \sec x g(x)$$
$$= \exp\left(-\int_0^x \sec t \mathrm{d}t\right)\left(\frac{\mathrm{d}g(x)}{\mathrm{d}x} - \sec x g(x)\right)$$
$$= \frac{\mathrm{d}\exp\left(-\int_0^x \sec t \mathrm{d}t\right) g(x)}{\mathrm{d}x}$$

Integrating both sides yields:

$$c = \exp\left(-\int_0^x \sec t \mathrm{d}t\right) g(x).$$

From which:

$$\int_0^x \sec t \mathrm{d}t = \log g(x) - \log c.$$

As $\cos(\pi/6) = \frac{\sqrt{3}}{2}$, we have $\sec(\pi/6) = \frac{2\sqrt{3}}{3}$ and $\tan(\pi/6) = \frac{\sqrt{3}}{3}$. Thus $g(\pi/6) = \sqrt{3}$. Also, $c = g(0) = 1$. In summary:

$$\int_0^{\frac{\pi}{6}} \sec t \mathrm{d}t = \log g(\pi/6) - 0 = \log\sqrt{3}.$$

Ok, this is good, but Hugo suggested a faster way to solve this through separation:

$$I := \int_0^{\pi/6} \sec x \mathrm{d}x = \int_0^{\pi/6} \frac{\mathrm{d}g(x)}{\mathrm{d}x}/g(x)\mathrm{d}x.$$

Substituting $u = g(x)$, we have:

$$I = \int_{g(0)}^{g\left(\frac{\pi}{6}\right)} \frac{1}{u}\mathrm{d}u = \ln g(\frac{\pi}{6}) - \ln g(0) = \ln\sqrt{3}.$$

Q: Suppose that two cylinders each with radius 1 intersect at right angles and their centres also intersect. What is the volume of the intersection?

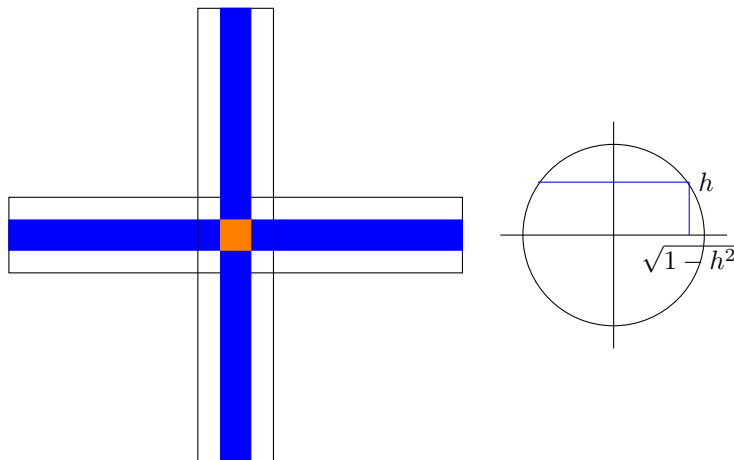**A:** The best thing would be to visualise this. Cross sectionally, from above:

Figure 1: In blue: the region of each cylinder which is above h. In yellow, the region intersecting both cylinders which is above h.

Without loss of generality, we can focus on the upper-part of the intersection (by symmetry, the top half and the bottom half will have same volume). We need to understand the shape of the cross-section at height $h < 1$ which is in the intersection of the two cylinders. Apparently, this is a square with side $2\sqrt{1 - h^2}$, and so:

$$\frac{V}{2} = \int_0^1 4(1 - h^2)\mathrm{d}h = 4\left(1 - \frac{1}{3}\right) = \frac{8}{3} \implies V = \frac{16}{3}.$$

Q: The snow began to fall some time before noon at a constant rate. The city of Cambridge sent out a snow plow at noon to clear Massachsusetts Avenue from MIT to Harvard. The plow removed snow at a constant volume per minute. At 1 pm, it had moved 2 miles and at 2 pm, 3 miles. When did the snow begin to fall?

A: Let us start from scratch. Denote by $T < 0$ the time at which it started to snow, and let 0 denote the time at which the plow was started. For $t > 0 > T$, the speed of the plow at time t is inversely proportional to the amount of snow in front of it: $\frac{c_1}{(t-T)}$. The distance travelled up to time t will then be:

$$D(t) := \int_0^t \frac{c_1}{u - T} du = c_1(\log(t - T) - \log(-T))$$

We know $D(1) = 2$ and $D(2) = 3$, so we have 2 equations for 2 unknowns:

$$\begin{cases} \frac{2}{c_1} = \log\left(\frac{1-T}{-T}\right) \\ \frac{3}{c_1} = \log\left(\frac{2-T}{-T}\right). \end{cases}$$

from which:

$$\left(\frac{-T}{1-T}\right)^{\frac{1}{2}} = \left(\frac{-T}{2-T}\right)^{\frac{1}{3}} \implies -T(2-T)^2 = (1-T)^3 \implies -4T+4T^2-T^3 = 1-3T+3T^2-T^3$$

Which simplifies to $T^2 - T - 1 = 0$ with only negative solution:

$$T_1 = \frac{1 - \sqrt{1+4}}{2} = \frac{1 - \sqrt{5}}{2}.$$

Q: If X is a standard normal random variable $X \sim \textbf{Normal}(0, 1)$, what is $\mathbf{E}[X|X > 0]$?

A: By definition of expectation:

$$I := \mathbf{E}[X|X > 0] = \int_0^\infty x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx.$$

We substitute $u = x^2/2$ so that $du = x dx$:

$$I = \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp(-u) du = \frac{1}{\sqrt{2\pi}}.$$

where we can recognize the integral of the exponential pdf. In general, for Gaussian random variables X_σ with mean 0 and standard deviation σ, we have

$$\mathbf{E}[X_\sigma|X_\sigma > 0] = \sigma E[X_1|X_1 > 0] = \frac{\sigma}{\sqrt{2\pi}}.$$

3.3 Partial derivatives and multiple integrals

Q: Calculate $I = \int_0^\infty \exp(-x^2/2)\mathrm{d}x$.
A: We can either relate this to the pdf of a standard normal distribution, or alternatively we can use polar coordinates. We have:

$$(2I)^2 = \int_{-\infty}^\infty \exp(-x^2/2)\,\mathrm{d}x \int_{-\infty}^\infty \exp(-y^2/2)\,\mathrm{d}y = \int_0^\infty \int_0^\infty \exp(-(x^2+y^2)/2)\,\mathrm{d}x\mathrm{d}y$$

We use the transformation:

$$\begin{cases} x & = r\cos\theta \\ y & = r\sin\theta \end{cases},$$

with $\mathrm{d}x\mathrm{d}y = r\mathrm{d}r\mathrm{d}\theta$. As such:

$$(2I)^2 = \int_0^\infty \int_0^{2\pi} \exp\left(-r^2\right) r\,\mathrm{d}\theta\,\mathrm{d}r = 2\pi \implies I = \sqrt{\frac{\pi}{2}}.$$

3.4 Important Calculus Methods

These include Taylor Series, Newton's method ($x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$), and the method of Lagrange multipliers. We also recall the bisection method and the secant method. In the bisection method we start from an interval whose boundaries are an overestimate and an underestimate of the root. We compute the sign of the function at the midpoint and restart, proceeding by induction. The secant method is similar to Newton's method, but we approximate the derivative by the slope of the secant line between the last two points.

Q: What is i^i?
A: The confusion is that even though we are used to work with imaginary numbers as bases, we can struggle with imaginary exponents. However, it is always possible to move the imaginary part to the exponent thanks to Euler's formula. As $i = \exp\left(i\frac{\pi}{2}\right)$, taking exponents yields:

$$i^i = \exp(-\pi/2).$$

Q: Prove that $(1+x)^n \geq 1 + nx$ for all $x > -1$ and for all integers $n \geq 2$.

A: The simplest way to prove this is through the binomial theorem:

$$(1+x)^n = \sum_{j=0}^{n} \binom{n}{j} x^j = 1 + nx + x^2 \sum_{j=0}^{n-2} \binom{n}{j+2} x^j.$$

We'd need to show that the last sum is non-negative. Not sure exactly how, but induction could work. (obvious for n =2 ...)

Another solution involves Taylor's expansions with intermediate value theorem. We consider the function $f(x) = (1+x)^n$, with first two derivatives equal to: $f'(x) = n(1+x)^{n-1}$, $f''(x) = n(n-1)(1+x)^{n-2}$. For every $x > -1$, there exists $\tilde{x}$ with $|\tilde{x}| \leq x$ such that:

$$f(x) = f(0) + xf'(0) + x^2 \frac{f''(\tilde{x})}{2!} = 1 + nx + x^2 n(n-1)(1+\tilde{x}).$$

All the factors of the last term are non-negative as $n \geq 2$ and $|\tilde{x}| \leq 1$, which implies result.

Q: Solve $x^2 = 37$ to the third digit.
A: Here, we use Newton's method for the function $f(x) = x^2 - 37$, with $f'(x) = 2x$ We know the solution must be between 6 and 7, and we start with our initial guess $x_0 = 6$. Netwon's method proceeds by iterating: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ (analogously as Taylors).

$$x_1 = 6 - \frac{-1}{12} = \frac{71}{12}.$$

The error of this solution is low.

$$\left(6 + \frac{1}{12}\right)^2 - 37 = 36 + 1 + \frac{1}{144} - 37 = \frac{1}{144} < \frac{1}{100}.$$

3.5   Lagrange multipliers

Let $f \colon R^n \to$ with gradient $\nabla f(x) = \left\langle \frac{\mathrm{d}f}{\mathrm{d}x_1}, \frac{\partial f}{\partial x_2}, \ldots \frac{\partial f}{\partial x_n} \right\rangle$. The necessary condition for maximizing or minimizing $f(x)$ subject to a set of k constraints:

$$\begin{cases} g_1(x_1, x_2, \ldots, x_n) = 0, \\ g_2(x_1, x_2, \ldots, x_n) = 0, \\ \ldots \\ g_k(x_1, x_2, \ldots, x_n) = 0, \end{cases}$$

is that $\nabla f(x) + \lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x) + \cdots + \lambda_k \nabla g_k(x) = 0$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are called Lagrange multipliers.

Q: What is the distance from the origin to the plane $2x + 3y + 4z = 12$?
A: The distance from the origin of a point (x, y, z) is $f(x) = \sqrt{x^2 + y^2 + z^2}$.
There is one constraint $g(x, y, z) = 0$ where $g(x, y, z) = 2x + 3y + 4z - 12$. At
the minimizer, we must have $\nabla f = \lambda \nabla g$, implying:

$$
\begin{cases}
\frac{2x^*}{f(x^*, y^*, z^*)} = 2\lambda, \\
\frac{2y^*}{f(x^*, y^*, z^*)} = 3\lambda, \\
\frac{2z^*}{f(x^*, y^*, z^*)} = 4\lambda, \\
2x^* + 3y^* + 4z^* = 12.
\end{cases}
\implies
\begin{cases}
3x^* = 2y^* \\
2x^* = z^* \\
2x^* + \frac{9}{2}x^* + 8x^* = 12.
\end{cases}
\implies
\begin{cases}
x^* = \frac{24}{29} \\
y^* = \frac{36}{29} \\
z^* = \frac{48}{29}
\end{cases},
$$

where the first implication comes from finding expressions of each variable equal
to $\lambda f(x^*, y^*, z^*)$.

3.6   Ordinary Differential Equations

Four ODE pattern commonly seen in interviews:

Separable $\frac{dy}{dx} = f(x)g(y)$ which has solution $\int \frac{dy}{h(y)}\, dy = \int g(x)\, dx$.

First-order $\frac{dy}{dx} + P(x)y = Q(x)$, which are solved through the integrating factor:
$I(x) =$

Q: Solve the ODE: $y' + 6xy = 0$.
A: This equation is separable as: $\frac{dy}{dx}\frac{1}{y} = -6x$. Thus, $\log y = -3x^2 + c$ and
$y \propto \exp\left(-3x^2\right)$.

Q: Solve the ODE: $y' = \frac{x-y}{x+y}$.
A: Here the separation is not trivial, because algebra doesn't allow us to find
the required form. Instead, we can use change of variables to help us. Consider
$z = y + x$ so that $z' = y' + 1$. We can rewrite the ODE as:

$$
z' - 1 = \frac{2x - z}{z} = \frac{2x}{z} - 1. \implies z'z = 2x \implies \int z\, dz = \int 2x\, dx + c
$$

$$
2x^2 + c = z^2 = (x + y)^2 = x^2 + 2xy + y^2 \implies x^2 - 2xy - y^2 = c
$$

3.7   Linear Algebra

To remember:

- *QR decomposition.* For each non-singular $n \times n$ matrix A, there is a unique pair of orthogonal matrix Q and upper-triangular matrix R with positive diagonal elements such that $A = QR$ [1].

Q: There are 3 random variables X, Y, Z. The correlation between X and Y is 0.8, and the correlation between X and Z is 0.8. What is maximum and minimum correlation between Y and Z?
A: The variance covariance matrix is given by:

$$\Sigma = \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & \rho \\ 0.8 & \rho & 1 \end{pmatrix}$$

It is obvious that the determinant of the first two leading principals are positive, and so we simply need to find the minimum ρ such that the determinant of the matrix is non-negative. The determinant is a quadratic function of ρ, and we know $\rho = 1$ is a root. The minimum ρ satisfying the condition is hence the other root. We solve:

$$0 = \det(\Sigma) = (1 - \rho^2) - 0.8^2(1 - \rho) + 0.8^2(\rho - 1)$$
$$= (1 - \rho)(1 + \rho - 2 \cdot 0.8^2)$$

which gives $\rho = 1.28 - 1 = 0.28$ as the smaller root.

Q: If the programming language you are using doesn't have a function of the linear least squares regression, how would you design an algorithm to compute it?
A: Given a design matrix X, I would compute:

$$(X^T X)\beta = X^T y \implies \beta = (X^T X)^{-1} Xy.$$

Probably, they want me to use the QR decomposition, so let us write $X^T X = QR$, which means:

$$\beta = R^{-1} Q^T X^T y,$$

where R^{-1} is easy to compute as R is upper triangular.
Just treat the equation as

$$Ax = y \implies QRx = y \implies Rx = Q^T y$$

, where the system is easy to solve from "the bottom up" due to upper-triangularity of R.

---

[1] Q corresponds to an orthonormal basis constructed from the original A basis using the Gram-Schmidt process. You can see R as a change of basis, from Q to A. Its triangular shape is explained by the fact that vectors from A are incorporated one at a time.

3.8   Determinant, eigenvalue and eigenvector

Q: What are the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$**?**

A: *By inspection (1,1) is an eigenvector, but let's keep going.*

To find eigenvalues, we solve the characteristic equation:

$$0 = \det(A) = (2 - \lambda)^2 - 1^2 = (2 - \lambda + 1)(2 - \lambda - 1) = 0.$$

so that the eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = 1$. For the eigenvectors, we solve $(A - \lambda_i I)v_i = 0$. It is not hard to see that $\lambda_1 = 3$ yields $v_2 = (1, 1)$ and $\lambda_2 = 1$ yields $v_2 = (1, -1)$.

For 2×2 matrices, we can also solve $\lambda_1 \lambda_2 = det(A)$ and $\lambda_1 + \lambda_2 = tr(A)$.

4   Basic probability

Q: An ant is placed in a corner of a cube and cannot move. A spider starts from the opposite corner, and can move along the cube's edges in any direction (x, y, z) with equal probability $(\frac{1}{3})$. On average, how many steps will the spider need to get to the ant?

A: First thoughts would be to set up a difference equation, i.e. a function mapping each vertex to the expected number of steps to the reach the ant. There exists some symmetry in the cube vertices, meaning that two vertices which are at the same distance from the ant will have the same expected number of steps. (this means the difference equation is set-up at a "higher level" than the individual vertices, a group level). There are 4 types of vertices: one at distance 0, 3 at distance 1, 3 at distance 2 and 1 at distance 3. We have:

$$S(0) = 0$$
$$3 \cdot S(1) = 3 + 1 \cdot 0 + 2 \cdot S(2)$$
$$3 \cdot S(2) = 3 + 1 \cdot S(3) + 2 \cdot S(1)$$
$$S(3) = 1 + S(2).$$

Substituting the last equation into the previous we obtain $2S(2) = 4 + 2S(1)$, which implies $S(1) = 7$, $S(2) = 9$ and $S(3) = 10$.
My solution is way cleaner than what is on Stack Overflow!

Q: Consider you have two random variables, X and Y, which are joint normal with zero means, unit variances and correlation of rho. Describe/derive distribution of X given $Y = y$.

A: My first approach uses the definition of conditional probability. Using the definition of the multivariate normal in the equation $p(x|y) \propto p(x, y)$ yields:

$$p(x|y) \propto \exp\left(-\frac{1}{2}(x, y) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix}\right).$$

To invert the matrix, we recall the formula for the inverse of a 2x2 matrix:

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

Computing the quadratic form of the matrix gives $x^2 + y^2 - 2\rho xy$. This yields:

$$p(x|y) \propto \exp\left(-\frac{(x - \rho y)^2}{2(1 - \rho^2)}\right)$$

which is proportional to a normal distribution with mean ρy and variance $1 - \rho^2$. In conclusion:

$$X|Y = y \sim \text{Normal}(\rho y, 1 - \rho^2)$$

I believe this answer is correct, see link. I am not sure there is anything more elegant? In terms of generalisation, we can reduce the general case to this one through linear transformations(?).

References