A collection of quantitative finance interview questions and solutions

Krish Nigam

Introduction

My goal with this document is to maintain a regularly updated collection of some of my favourite interview questions from either my own interviews, or having heard those of my peers and friends. These questions are aimed predominantly at interviews in quantitative finance.

The format of this document is such that each question is in a grey box followed by the complete solution immediately below it. Therefore, when scrolling through this document, please bear this in mind if you'd prefer to independently attempt the problem, before going through the solution (*recommended*). Moreover, with regard to the solution, I try to explain, as much as possible, the full reasoning and intuition as to how one may approach the problem and reasonably 'come up' with the solution.

Please feel free to email any corrections and suggestions to me.[1]

Contents

[1] krish.nigam@princeton.edu

Problem 1

We play a game where we roll a fair dice for a number of rounds. If the dice lands on a number between 2 to 6, we can decide whether to accept this score (in which case we receive the dollar amount of the score rolled) or roll again, however if the dice lands on 1 our game stops there and we are forced to pay $\$n$ where n is the current round we are on.

For example in our first round, rolling a 1 gives payoff of $-\$1$ and 2 to 6 gives the score, and if we decide to continue to the next round after rolling 2 to 6 in round 1, rolling a 1 means our payoff here would be $-\$2$, and so on *etc.* What is the expected value of this game if we have n rounds to play, and after how many rounds does it become unprofitable to keep playing?

Solution.

We analyse the game by working backwards. Let there be n total rounds available.

Let V_k denote the expected value of the game when we are about to play round k (so round k has penalty $-k$ if we roll a 1). Our goal is to compute V_1, the value at the start. First consider the simplest possible version.

For the case $n = 1$, if we only have one round, we must accept whatever happens:

$$V_1 = \frac{1}{6}(-1) + \frac{1}{6}(2+3+4+5+6) = -\frac{1}{6} + \frac{20}{6} = \frac{19}{6} \approx 3.17.$$

For $n = 2$, on the second (final) round we have no decisions left, so V_2 is just the one–round value:

$$V_2 = \frac{19}{6}.$$

Now consider round 1. If we roll 1, we get -1.

If we roll $r \in \{2, 3, 4, 5, 6\}$, we may either: take r or continue and expect V_2. So, the expected value is:

$$V_1 = \frac{1}{6}\left(-1 + \sum_{r=2}^{6} \max(r,\, V_2)\right).$$

Since $V_2 = 19/6 \approx 3.17$, we compare each r to 3.17.

We take 2 and 3 immediately, but for $4, 5, 6$ we would rather roll again because V_2 is larger than 3 but less than 4:

$$\max(2, V_2) = V_2, \quad \max(3, V_2) = V_2, \quad \max(4, V_2) = 4, \quad \max(5, V_2) = 5, \quad \max(6, V_2) = 6.$$

Thus,

$$V_1 = \frac{1}{6}\left(-1 + 2V_2 + 4 + 5 + 6\right) = \frac{1}{6}\left(-1 + 2 \cdot \frac{19}{6} + 15\right) \approx 3.28.$$

For the case $n = 3$, we already have $V_3 = \frac{19}{6}$ from the one–round case.

To compute V_2, we use:

$$V_2 = \frac{1}{6}\left(-2 + \sum_{r=2}^{6} \max(r,\, V_3)\right),$$

and since $V_3 \approx 3.17$, the same comparisons apply as before (take $4, 5, 6$; continue on $2, 3$). This gives, $V_2 \approx \frac{1}{6}\left(-2 + 2V_3 + 4 + 5 + 6\right) \approx 3.16$. Then, $V_1 = \frac{1}{6}\left(-1 + \sum_{r=2}^{6} \max(r,\, V_2)\right)$. Since $V_2 \approx 3.16$ is slightly below 4, the rule remains the same; so, $V_1 \approx \frac{1}{6}\left(-1 + 2V_2 + 4 + 5 + 6\right) \approx 3.39$.

In the *general case*, we have the recursion form:

$$V_k = \frac{1}{6}\left(-k + \sum_{r=2}^{6} \max(r,\, V_{k+1})\right), \qquad k = 1, 2, \ldots, n,$$

with terminal condition $V_{n+1} = 0$. Computing backwards numerically, V_1 stabilises quickly as n grows. The limit value is, $V_1 \approx 3.39$.

Thus the expected value of the game with a large number of available rounds is approximately, $\mathbb{E}[\text{game value}] \approx 3.39$.

For the next part of the problem of when it becomes unprofitable to continue, as k increases, the penalty for rolling 1 grows, so V_k decreases. Numerical calculation[2] shows that V_k becomes negative around $k \approx 20$.

Therefore, after roughly the 20th round, the expected value of continuing is negative (one should never let the game continue that far in an optimal strategy).

Intuitively, one can also see it if one considered each round independently, then the payoff of round n is $\frac{20-n}{6}$ and this becomes negative at $n = 20$.

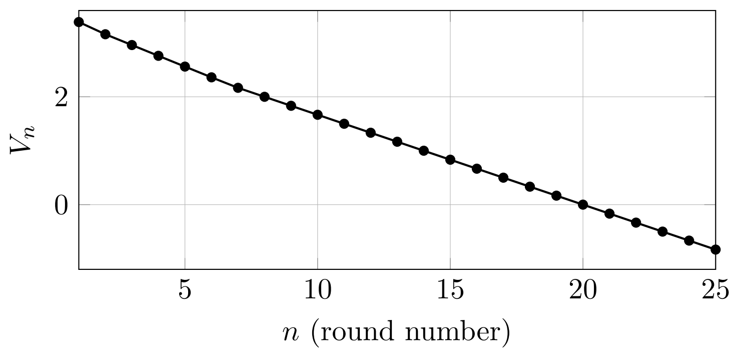

Figure 1: Expected value V_n as a function of round number n. Note it looks fully linear, but is not (although it is piecewise linear by the discrete nature of the problem).

□

[2] One can easily code this recursive solution up to view the expected payoffs for each n and can see how this changes.

Problem 2

We have a 100 sided die. We roll it and each time we have the choice of accepting the score as our payoff, or re-rolling which costs 1 each time to re-roll. Assuming optimal play (*i.e.* we want to maximise our payoff), what is our expected payoff?

(In an interview, they may ask for a rough, quick approximate answer first, so think about how you can roughly approximate this first before solving for it exactly.)

Solution.

We are allowed to roll a 100-sided die repeatedly. Each re-roll costs 1. When we roll a value r, we may either take r as our payoff and stop, or pay 1 and roll again. The goal is to choose a stopping rule that maximises our expected payoff.

Rough intuition: If we keep re-rolling forever on small values, we are effectively paying cost 1 each time in order to "search" for a large number. Obviously we should accept very large rolls (e.g. 95, 96, ..., 100) and clearly we should re-roll on very small values (e.g. 1, 2, 3). So there must be some cutoff value k such that:

$$\text{If } r \geq k, \text{ accept.} \quad \text{If } r < k, \text{ re-roll.}$$

This is a standard optimal stopping structure.

We expect k to be fairly high, because good rolls are quite valuable (near 100), and the re-roll cost is only 1. A quick mental estimate might suggest the cutoff is somewhere in the high 80s or low 90s.

Formal analysis: Let V be the expected payoff *before* rolling (i.e. the value of the game). If we roll a number r, then:

$$\text{If we accept: payoff } r. \quad \text{If we re-roll: payoff } V - 1 \text{ (pay cost 1, restart).}$$

So the optimal decision is:

$$\text{Accept } r \quad \text{if } r \geq V - 1.$$

This means the cutoff k must satisfy $k \approx V - 1$, so we take all $r \geq k$ and re-roll otherwise.

Assume the optimal cutoff is an integer k. Then:

$$V = \frac{1}{100}\left(\sum_{r=1}^{k-1}(V-1) + \sum_{r=k}^{100} r\right).$$

The first sum has $k-1$ terms, each equal to $V-1$. The second sum is the sum of integers $k + (k+1) + \cdots + 100$:

$$\sum_{r=k}^{100} r = \frac{(100+k)(100-k+1)}{2}.$$

Thus,

$$V = \frac{(k-1)(V-1) + \frac{(100+k)(101-k)}{2}}{100}.$$

Solving this for V gives:

$$V = \frac{100+k}{2} - \frac{k-1}{101-k}.$$

The cutoff must satisfy:

$$k - 1 \leq V - 1 < k.$$

Checking integer values k from 1 to 100, the unique consistent solution is:

$$k = 87, \qquad V \approx 87.36.$$

Thus, the optimal strategy is: Accept the roll if it is 87 or higher, otherwise re-roll.

The expected payoff under this optimal strategy is ≈ 87.36. $\square$

Problem 3

We have N coins arranged in a circular shape. We start with coin 1 and toss it. If it lands on heads, we remove it from the circle. If it lands on tails we keep it in the circle. We then toss coin 2 (the next coin going clockwise from coin 1) and if it is heads remove it from the circle, otherwise if tails then keep it in the circle. We then move to coin 3 and do the same and so on going round the circle of coins again, until there is only one coin left.

Which coin has the highest probability of being the last coin left, and, more generally, what is the probability that the i-th coin is the last coin remaining?

Solution.

Label the coins $1, 2, \ldots, N$ clockwise and start at coin 1. A coin is removed the first time it shows heads; otherwise (tails) it stays and we move on. We stop when one coin remains. Intuitively, we'd expect coin N in the circle to have highest probability of being last, since each coin has an i.i.d. 'chance' of being taken out before the last coin does, as each of their first tosses come before.

Consider the 2-coin problem: with $N = 2$, let P_2 be the probability coin 2 is last (so coin 1 is last with probability $1 - P_2$). Consider the next single toss at the current coin:

After tossing coin 1, with probability $\frac{1}{2}$ it is H and coin 2 is instantly last, and with probability $\frac{1}{2}$ it is T and coin 1 stays. If coin 1 is T and coin 2 is H, then coin 1 is instantly last. Otherwise, if both are T with probability $\frac{1}{4}$, we are back to the *same state* as the beginning. Thus, we can write an equation for this in P_2.

$$P_2 = \tfrac{1}{2} + \tfrac{1}{4} P_2 \implies P_2 = \frac{2}{3}, \qquad \mathbb{P}(\text{coin 1 last}) = \frac{1}{3}.$$

Now we generalize this for an arbitrary number of coins, n. Define, for integers $n \geq 1$ and $i \in \{1, \ldots, n\}$,

$$p(i, n) := \Pr(\text{coin } i \text{ is the last remaining, starting with } n \text{ coins}).$$

We can now write a *two-case* recursion directly in terms of (i, n) by conditioning on the next toss of coin 1.

Consider tossing coin 1; there are two cases (it lands heads or tails).

- If heads (probability $\frac{1}{2}$), coin 1 is removed. The game restarts on the remaining $n - 1$ coins, with the next coin to toss being the original coin 2. After relabelling (rotation), each original coin $i > 1$ becomes index $i - 1$ among $n - 1$ coins; original coin $i = 1$ is gone. Thus

$$\Pr(\text{coin } i \text{ wins} \mid \text{H at coin 1}) = \begin{cases} 0, & i = 1, \\ p(i-1,\, n-1), & i = 2, \ldots, n. \end{cases}$$

- If tails (probability $\frac{1}{2}$), coin 1 stays, and we advance to coin 2 with all n coins present. After rotation, each original coin $i > 1$ becomes index $i - 1$ among n coins, while the original coin 1 moves to the end (index n):

$$\Pr(\text{coin } i \text{ wins} \mid \text{T at coin 1}) = \begin{cases} p(n,\, n), & i = 1, \\ p(i-1,\, n), & i = 2, \ldots, n. \end{cases}$$

This generalises to the following recursion for $p(i, n)$ (for all $n \geq 2$).

$$\begin{aligned} p(1,n) &= \tfrac{1}{2}\, p(n,n), \\ p(i,n) &= \tfrac{1}{2}\, p(i-1,n-1) \;+\; \tfrac{1}{2}\, p(i-1,n), \qquad i = 2,\ldots,n, \end{aligned}$$

with base case $p(1, 1) = 1$.

Intuitively, why does this work? Each step either removes the current coin (heads) or keeps it and advances (tails). In either branch we rotate the labels so the "next to toss" becomes index 1. That rotation maps original index i to $i-1$ (or to n if $i = 1$ on a tails), which is exactly what the right-hand sides record.

The question now becomes how can this be solved. We can implement this recursion through code as a dynamic programming problem.Compute for $n = 1, 2, \ldots, N$:

$$p(1,1) = 1, \qquad \begin{cases} p(1,n) = \frac{1}{2}\, p(n,n), \\ p(i,n) = \frac{1}{2}\, p(i-1,n-1) + \frac{1}{2}\, p(i-1,n), \quad i = 2,\ldots,n. \end{cases}$$

This fills each row n left-to-right using the already-computed row $n - 1$. Total time $O(N^2)$.

For fixed n, the recursion provides exactly n linear equations (one for each i), and we have n unknowns; $p(1, n), \ldots, p(n, n)$. Moreover the system is triangular in i: each $p(i, n)$ depends only on $p(i-1, n)$ (same row, earlier index) and on values from the previous row $p(\cdot, n-1)$, which are already known if you fill the table row-by-row. Concretely:

$$\begin{aligned} p(2,n) &= \tfrac{1}{2}p(1,n-1) + \tfrac{1}{2}p(1,n), \\ p(3,n) &= \tfrac{1}{2}p(2,n-1) + \tfrac{1}{2}p(2,n), \;\; \textit{etc.} \end{aligned}$$

Iterating up to $i = n$ yields $p(n, n) = A + B\, p(1, n)$ (with known A, B from the previous row). Together with $p(1, n) = \frac{1}{2}p(n, n)$ you solve $p(1, n)$, then back-substitute to get $p(2, n), \ldots, p(n, n)$.

From the recursion, we also clearly see that $p(i, n)$ increases in i for fixed n. Hence,

$$p(1,n) \leq p(2,n) \leq \cdots \leq p(n,n),$$

so coin n (the last in the order) has the highest probability of being the final survivor, in line with our intuition.

$n = 2$	$p(1,2)$	$p(2,2)$
	$\frac{1}{3}$	$\frac{2}{3}$

$n = 3$	$p(1,3)$	$p(2,3)$	$p(3,3)$
	$\frac{5}{21}$	$\frac{2}{7}$	$\frac{10}{21}$

$n = 4$	$p(1,4)$	$p(2,4)$	$p(3,4)$	$p(4,4)$
	$\frac{19}{105}$	$\frac{22}{105}$	$\frac{26}{105}$	$\frac{38}{105}$

Figure 2: Computed $p(i, n)$ for small values of n.

$\square$

Problem 4

I think of a polynomial $p(x)$ with non-negative integer coefficients. You can provide any input number a and I will tell you the value $p(a)$. What is the minimum number of inputs to the polynomial you require in order to guess what my original polynomial was?

Solution.

Rough intuition: If we could somehow guarantee that all coefficients c_i were strictly smaller than our chosen input a, then the value $p(a) = c_0 + c_1 a + c_2 a^2 + \cdots$ would be the base-a expansion whose "digits" are exactly the coefficients c_i. In other words, evaluating p at such an a would encode the entire polynomial in one number. The only thing is that we don't know how large the coefficients are.

The trick is to first get an upper bound on the sum of coefficients, then choose a larger than that sum so that each $c_i < a$. After that, we can recover the coefficients by taking remainders (or *mods*) by $a, a^2, \ldots$.

Formal analysis: We begin by querying the polynomial at $x = 1$, *i.e.* we want the value $p(1) = \sum_{i=0}^{d} c_i = S$. This gives the total sum of all the coefficients, and observe that this forces that each $x_i \leq S$.

We now require a second input to the polynomial, and since we want to do it in the minimum number of inputs, it makes sense that our second input should somehow utilise the result of our first input. Hence, it would be reasonable to try $x = S = \sum_{i=0}^{d} c_i$ as our second input.

Computing $p(S)$ gives,

$$\begin{aligned} p(S) = p(p(1)) &= c_0 + c_1 \left(\sum_{i=0}^{d} c_i \right) + c_2 \left(\sum_{i=0}^{d} c_i \right)^2 + \cdots + c_d \left(\sum_{i=0}^{d} c_i \right)^d \\ &= c_0 + c_1 S + c_2 S^2 + \cdots + c_d S^d. \end{aligned}$$

Immediately from the above equation, we see that this 'suspiciously' looks like the base-S expansion of $p(S)$ with digits c_i. However, there are now two cases to consider, which are automatically handled and give back the coefficients of p.

The first case is if some coefficient equals S. Since all $c_i \geq 0$ and $\sum_{i=0}^{d} c_i = S$, this can only occur if exactly one coefficient is nonzero. Therefore, $p(x)$ would have the form $p(x) = Sx^m$ for some m, and we know m directly from the value of $p(S) = S^{m+1}$.

The second more general case is where all the coefficients satisfy $c_i < S$. Then, $p(S) = c_0 + c_1 S + c_2 S^2 + \cdots + c_d S^d$ is exactly the base-S expansion of $p(S)$. Thus, we can recover the coefficients by successive remainder (or mods).

$$\begin{aligned} c_0 &\equiv p(S) \pmod{S}, \\ c_1 &\equiv \frac{p(S) - c_0}{S} \pmod{S}, \\ c_2 &\equiv \frac{p(S) - c_0 - c_1 S}{S^2} \pmod{S}, \quad \text{etc.} \end{aligned}$$

Equivalently, we just write $p(S)$ in base S; its digits are $c_0, c_1, \ldots, c_d$ and beyond d all digits are zero, which also reveals the degree.

Why two inputs to the polynomial are necessary? A single value never pins down the polynomial, since many distinct polynomials take the same value at a single input (*e.g.* at $x = 1$, both $p(x) = 5$ and $p(x) = 2 + 3x$ give 5). Hence, at least two are required, and the pair of inputs $\{p(1), p(p(1))\}$ suffices as shown. $\square$

Problem 5

If we pick 3 points uniformly on the unit circle centered at $(0,0)$, this forms three arcs on the circle. What is the expected length of the arc which contains the point $(1,0)$?

Solution.

There are two main solutions of doing this problem. The first, and typical one which most people would do, is through using integration. The second, much more intuitive, solution simply requires a subtle *change of perspective.*

By rotational symmetry the answer does not depend on the specific reference point on the circle, be it $(1,0)$ or $(0,1)$ or any other point on the circle, the expectation remains the same. A naive solution to this problem would be $\frac{2\pi}{3}$, however it is not the case that the three regions/arcs have equal probability of containing the point, and in fact the largest arc has a greater probability of containing the reference point.

Solution 1 (integration for three points): the circumference of the circle is 2π. Place an angular coordinate $\theta \in [0, 2\pi)$ with $\theta = 0$ at $(1,0)$. Let $X_1, X_2, X_3 \overset{\text{i.i.d.}}{\sim} \text{Unif}(0, 2\pi)$ be the three sampled angles.

Consider the (clockwise) arc length M from $(1,0)$ to the *nearest* sampled point encountered when moving clockwise. Then

$$\Pr(M > x) = \Pr\left(X_1 > x,\ X_2 > x,\ X_3 > x\right) = \left(1 - \frac{x}{2\pi}\right)^3, \qquad 0 \le x \le 2\pi.$$

Hence, using the tail integral formula $\mathbb{E}[M] = \int_0^{2\pi} \Pr(M > x)\, dx$,

$$\mathbb{E}[M] = \int_0^{2\pi} \left(1 - \frac{x}{2\pi}\right)^3 dx = \left[-\frac{2\pi}{4}\left(1 - \frac{x}{2\pi}\right)^4\right]_0^{2\pi} = \frac{2\pi}{4} = \frac{\pi}{2}.$$

By symmetry, if M_{cw} is the clockwise gap from $(1,0)$ to the next sampled point and M_{ccw} is the counterclockwise gap to the next sampled point, then $\mathbb{E}[M_{\text{cw}}] = \mathbb{E}[M_{\text{ccw}}] = \pi/2$. The arc that *contains* the cut point $(1,0)$ is precisely the union of these two nearest gaps, so its length is $L = M_{\text{cw}} + M_{\text{ccw}}$. Therefore, by linearity of expectation,

$$\mathbb{E}[L] = \mathbb{E}[M_{\text{cw}}] + \mathbb{E}[M_{\text{ccw}}] = \frac{\pi}{2} + \frac{\pi}{2} = \pi.$$

Thus the expected length of the arc containing the reference point $(1,0)$ is $\mathbb{E}[L] = \pi$.

Intuitively, the expected clockwise distance from a fixed cut to the nearest of three uniform points is $\frac{1}{4}$ of the full circumference (here: $\pi/2$), and the counterclockwise distance has the same expectation; summing gives exactly half the circle on average.

Solution 2 (unwrap the circle to a line / order statistics, for three points): unwrap the circle at $(1,0)$ to the unit interval $[0,1)$ (arc length normalized by 2π), such that both end points of the $[0,1]$ interval line correspond to the point $(1,0)$. The three points picked become i.i.d. $U_1, U_2, U_3 \sim \text{Unif}(0,1)$. The arc on the circle that contains the cut point corresponds, on the line, to the 'wraparound gap', whose (normalized) length is

$$L_{\text{line}} \;=\; \left(1 - U_{(3)}\right) + U_{(1)} \;=\; 1 - \left(U_{(3)} - U_{(1)}\right),$$

where $U_{(1)} < U_{(2)} < U_{(3)}$ are the order statistics (*i.e.* $U_{(1)} = \min\{U_1, U_2, U_3\}$ and $U_{(3)} = \max\{U_1, U_2, U_3\}$ *etc.*).

Taking expectations,

$$\mathbb{E}[L_{\text{line}}] = 1 - \big(\mathbb{E}[U_{(3)}] - \mathbb{E}[U_{(1)}]\big).$$

For three i.i.d. uniforms on $[0, 1]$,

$$\mathbb{E}[U_{(1)}] = \frac{1}{4}, \qquad \mathbb{E}[U_{(3)}] = \frac{3}{4}.$$

Hence

$$\mathbb{E}[L_{\text{line}}] = 1 - \left(\frac{3}{4} - \frac{1}{4}\right) = \frac{1}{2}.$$

Scaling back to arc length on the circle (multiply by the circumference 2π) gives

$$\mathbb{E}[L] = 2\pi \cdot \frac{1}{2} = \pi \ .$$

Intuitively, after unwrapping, the arc containing the cut point is simply the complement of the span between the smallest and largest sampled positions. On average, for three points, the span occupies half of the interval, leaving half for the containing arc; rescaling gives π.

Problem 6

I have two stocks A and B and a call option. How does the price of a call option vary with the correlation of stocks A and B if the underlying of the option is:

a) the average price of the two stocks?
b) the maximum return of the two stocks (in each constant time interval)?

Solution.

We compare how the value of a call option varies with the correlation between two assets A and B, in two cases: (a) the underlying is the average price $S_{\text{avg}} = \frac{A+B}{2}$; (b) the underlying is the maximum *return* of the two stocks over each (fixed) time step, i.e. $S_{\max} = \max(R_A, R_B)$ where R_A, R_B are the period returns.

In both parts the central idea is that call values increase with the volatility or dispersion of the underlying payoff (under standard models (*e.g.* Black-Scholes) the vega of a call is positive; more generally, under a martingale measure, the price of a convex payoff increases with mean-preserving spreads of the underlying).

a) For the option on the average price, we write $\sigma_A^2 = \mathrm{Var}(A)$, $\sigma_B^2 = \mathrm{Var}(B)$, and $\rho = \mathrm{Corr}(A, B)$. Then

$$\mathrm{Var}\left(\frac{A+B}{2}\right) = \frac{1}{4}\Big(\mathrm{Var}(A) + \mathrm{Var}(B) + 2\,\mathrm{Cov}(A,B)\Big) = \frac{1}{4}\Big(\sigma_A^2 + \sigma_B^2 + 2\rho\,\sigma_A\sigma_B\Big).$$

Hence, holding the marginal distributions (*i.e.* the volatilities) fixed and varying only ρ, the variance of the average is *increasing* in ρ.

In the symmetric case $\sigma_A = \sigma_B = \sigma$, this reduces to $\mathrm{Var}\big(\frac{A+B}{2}\big) = \frac{\sigma^2}{2}(1+\rho)$, so that $\rho = 1 \implies \mathrm{Var} = \sigma^2$; $\rho = 0 \implies \mathrm{Var} = \frac{\sigma^2}{2}$; $\rho = -1 \implies \mathrm{Var} = 0$. Thus, when $\rho = 1$, the two stocks move together with no diversification, and so the average is as volatile as either stock. When $\rho = 0$, the average is 'stabilized' by diversification. When $\rho = -1$, the fluctuations cancel and the average can even be (locally) constant.

Since a call on $\frac{A+B}{2}$ is convex in the underlying, higher volatility of $\frac{A+B}{2}$ (with its mean held fixed under the pricing measure) increases its value. Therefore, the call price on the average roughly *increases* with correlation.

b) For the option on the maximum return, we now view one 'rebalancing' period and let R_A, R_B be the simple returns over that period. The underlying is $S_{\max} = \max(R_A, R_B)$, and the call payoff is a convex, increasing function of $S_{\max}$. Fix the marginal laws of R_A and R_B and vary only their dependence via ρ.

Low correlation means the two returns move more independently; this increases the chance that at least one of them experiences a large positive realization in the period. Consequently $\max(R_A, R_B)$ is stochastically larger (its upper tail is heavier), which raises both its mean and its dispersion.

As correlation increases toward 1, the two returns tend to move together and the maximum behaves more like either return alone, reducing the benefit from 'picking the better of the two', and instead one return usually dominates. Thus, as ρ increases, the distribution of $\max(R_A, R_B)$ becomes less spread out to the upside. Hence, since the call payoff is convex in $S_{\max}$, its price *decreases* with correlation.

Problem 7

There are 5 cards (numbered $1-5$), and you have to create a deck of n cards using any combination of these five (you may repeat instances of each number $1-5$). I can take a look at your deck, and then you will shuffle the deck and I will draw a card at random. I will make a guess at the card I picked, and if the number I guessed matches the number I picked, you have to pay me the number on the card I picked. What deck should you make, assuming optimal play (*i.e.* you want to minimize the potential amount of money you can lose)?

Solution.

Problem 8

A stock is priced at 100, and has 50% chance of going to 160 and 50% chance of going to 80. What is the price of a call option on this stock?

Solution.

Problem 9

(Geometric interpretation of linear regression:) If I regress Y on X_1 using OLS linear regression, I obtain a positive β-coefficient. If I now introduce a new factor X_2, will regressing Y on X_1 and X_2 necessarily maintain the β-coefficient of X_1 to be positive? Explain why.

Solution.

Problem 10

(Tricky martingales:) We have a martingale, which goes up $+1$ 50% of the time, and down -1 50% of the time. What is the expected time for the martingale, starting at 0, to hit a threshold $\pm k$ (for $k \in \mathbb{Z}$)?

What about the expected time if we now have a dynamic threshold instead; the upper bound is $k - \min\{X_1, \ldots, X_t\}$ and the lower bound is $k + \max\{X_1, \ldots, X_t\}$ (where $X_1, \ldots, X_t$ are realizations of the martingale up to time t, and at each time step t, this upper and lower bound changes accordingly, based on the values of the process already observed)?

Solution.