Quant, FM, and Data Science Interview Compilation

Aaron Cao

Contents

Introduction for LSU Students

There is a growing popularity of financial mathematics among the LSU undergraduate population. Many of these students are Mathematics majors concentrating in Actuarial Sciences or Statistics, but there are also quite a number in the Finance and Computer Science departments. The topics and problems compiled in this handout will invariably be encountered by students in all of these fields, and it is important that enthusiasts of the quant/FM/data science career path have a firm grasp of the technical interview

Throughout my undergraduate years, I was able to compile a good number of popular and frequently-asked interview problems from both my own interview experiences and from friends in the field. The problems listed here primarily come from top-tier hedge funds, trading firms, software companies, and even graduate schools. These are attractive career choices for many of the aforementioned students, and along all of these career paths one will inevitably face a technical interview. I hope that the content provided in this handout will help many generations of LSU students to prepare for such a task.

Though the problems in this handout are by no means exhaustive, they do provide a good road map for the general knowledge that is expected from professionals in the industry. I provide link solutions and occasionally written solutions to every problem, so some research on the readers' part to fill in knowledge gaps will be necessary. For baseline proficiency I would recommend knowing the undergraduate curriculum from at least the first two introductory statistics courses, the data structures course, the mathematical statistics course, and the stochastics course.

On a closing note, I encourage any LSU reader to participate in the Problem Solving Seminar and Putnam Competition during their undergraduate years. Beyond the specialized mathematics and coding skills, the most important ability any person in this industry should have is the ability to solve problems, and honing one's problem-solving ability pays large dividends throughout life.

Good Resources

- Art of Problem Solving (https://artofproblemsolving.com/)
- Brilliant (https://brilliant.org/)
- 3Blue1Brown (http://www.3blue1brown.com/)
- Mark S. Joshi (https://en.wikipedia.org/wiki/Mark_S._Joshi)
- Xinfeng Zhou (https://www.biblio.com/xinfeng-zhou/author/1230731)
- Competitive Programming (https://cpbook.net/)
- Cracking the Coding Interview (http://www.crackingthecodinginterview.com/)
- DarthPrince (http://codeforces.com/blog/entry/15729)
- Inishan (https://codeforces.com/blog/entry/23054)
- Matei Zaharia (https://cs.stanford.edu/~matei/programming_contests/)

For questions and concerns feel free to contact me at aarontcao@gmail.com.

Interview Problems

1. Discuss algorithms for parallel matrix multiplication.
   - https://cse.buffalo.edu/faculty/miller/Courses/CSE633/Ortega-Fall-2012-CSE633.pdf
   - https://www.tutorialspoint.com/parallel_algorithm/matrix_multiplication.htm
   - https://www3.nd.edu/ zxu2/acms60212-40212/Lec-07-3.pdf
2. Design a risk and asset pricing model for tech startup equity.
   - https://www.startups.co/articles/startup-equity-101
   - https://www.investopedia.com/terms/c/capm.asp
   - https://www.investopedia.com/articles/personal-finance/050515/how-calculate-beta-private-company.asp
   - https://www.financierworldwide.com/are-we-pricing-private-equity-risk-properly/s
3. Discuss ordinary least squares (OLS), maximum likelihood (MLE), and maximum a posteriori (MAP) estimation.
   - https://en.wikipedia.org/wiki/Ordinary_least_squares
   - https://en.wikipedia.org/wiki/Proofs_involving_ordinary_least_squares
   - https://en.wikipedia.org/wiki/Maximum_likelihood_estimation
   - https://en.wikipedia.org/wiki/Maximum_a_posteriori_estimation
   - https://stats.stackexchange.com/questions/80424/sum-of-squared-difference-and-gaussian-noise-model/80564#80564
4. A horizontal stick is one metre long. Fifty ants are placed in random positions on the stick, pointing in random directions. The ants crawl head first along the stick, moving at one metre per minute. If an ant reaches the end of the stick, it falls off. If two ants meet, they both change direction. How long do you have to wait to be sure that all the ants have fallen off the stick?
   - https://math.stackexchange.com/questions/1036902/interesting-question-on-ants
   - https://math.stackexchange.com/questions/1418351/random-ants-probability-question
5. Design a data structure for storing an integer-weighted probability distribution. Discuss runtime tradeoffs for different designs.
   - https://www.geeksforgeeks.org/random-number-generator-in-arbitrary-probability-distribution-fashion/
   - https://softwareengineering.stackexchange.com/questions/150616/get-weighted-random-item
   - https://cs.stackexchange.com/questions/59690/is-this-probability-distribution-data-structure-already-discovered

6. Given a signal, which is regularly sampled over time and is "noisy", how can the noise be reduced while minimizing the changes to the original signal? The standard method is with a Fourier transform. What is the intuition for why it works? Can one optimize hyperparameters for terms of the transform that sit inside the sum?

   - https://exnumerus.blogspot.com/2011/12/how-to-remove-noise-from-signal-using.html

   Solution. The simple answer is no. Sine and cosine are periodic, so one can not use convex optimization techniques.

7. There are N lions and 1 sheep in a field. All the lions really want to eat the sheep, but the problem is that if a lion eats a sheep, it becomes a sheep. A lion would rather stay a lion than be eaten by another lion. There is no other way for a lion to die than to become a sheep and then be eaten. When is it safe for any lion to eat?

   - https://math.stackexchange.com/questions/937410/understanding-the-solution-of-a-riddle-about-lions-and-sheep

8. What are some pros and cons of using daily returns data versus monthly returns data?

   - https://stats.stackexchange.com/questions/124404/why-monthly-stock-returns-instead-of-daily-returns-in-multiple-regressions
   - https://www.fields.utoronto.ca/programs/scientific/09-10/bachelier/ talks/Sat/Varley/bfs80groth.pdf
   - https://www.quora.com/Should-I-use-daily-monthly-or-yearly-returns-in-portfolio-variance-calculations-when-calculating-relevant-means-variances-exces-returns-covar

9. What is the geometric interpretation of regression?

   - https://www.datasciencecentral.com/profiles/blogs/linear-regression-geometry
   - https://www.youtube.com/watch?v=oWuhZuLOEFY
   - https://www.youtube.com/watch?v=PbyP3goun2Y
   - https://www.youtube.com/watch?v=444ZkgiHI3Q
   - https://en.wikipedia.org/wiki/Regression_analysis

10. Calculate the correlation of two vectors and write the code in Python.

    - https://en.wikipedia.org/wiki/Pearson_correlation_coefficient
    - https://en.wikipedia.org/wiki/Correlation_and_dependence
    - https://stackoverflow.com/questions/19428029/how-to-get-correlation-of-two-vectors-in-python

11. Design a research that uses dividend cut as a variable of interest.

    - https://www.investopedia.com/ask/answers/06/dividendpaymentcut.asp
    - https://www.investopedia.com/trading/dividends-interest-rates-effect-stock-options/

12. A population dies out with $p = 0.2$. It remains stable with $p = 0.5$. It doubles with $p = 0.3$. What is the expected long term behavior of the population?

- https://en.wikipedia.org/wiki/Markov_chain
- https://en.wikipedia.org/wiki/Stochastic_matrix
- https:// drive.google.com/file/d/1BydxM4mZNc4rUHF5VAZQDshY8xbfWzmL/view?usp=sharing

Solution. This is a classic Markov chain problem. A transition probability matrix solves this easily.

13. Given a matrix of correlations, write an algorithm to cluster the stocks with correlation equal to a certain value.

    - https://www.cs.princeton.edu/sites/default/files/uploads/karina_marvin.pdf
    - https://arxiv.org/pdf/1511.07945.pdf
    - http://www.diva-portal.org/smash/get/diva2:196577/FULLTEXT01.pdf
    - https://en.wikipedia.org/wiki/Correlation_clustering
    - https://quant.stackexchange.com/questions/2263/how-to-cluster-stocks-and-construct-an-affinity-matrix
    - https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

    Solution. A basic idea is to use the k-nearest-neighbors algorithm.

14. Discuss correlation matrices with panel data.

    - https://stackoverflow.com/questions/47886162/correlation-matrix-for-panel-data
    - https://www.researchgate.net/post/correlation_matrix_for_variables_in_panel_data
    - https://www.statalist.org/forums/forum/general-stata-discussion/general/1432235-correlation-matrix-in-panel-data-model

15. Discuss problems with multiple inheritance in Java.

    - http://www.lambdafaq.org/what-about-the-diamond-problem/
    - https://www.geeksforgeeks.org/java-and-multiple-inheritance/
    - https://www.journaldev.com/1775/multiple-inheritance-in-java
    - https://javapapers.com/core-java/why-multiple-inheritance-is-not-supported-in-java/

16. What are the differences between Python lists and Java arrays?

    - https://www.pythoncentral.io/the-difference-between-a-list-and-an-array/
    - https://stackoverflow.com/questions/27769511/python-list-vs-java-array-efficiency
    - https://stackoverflow.com/questions/33978318/arraylist-in-java-vs-list-in-python
    - https://www.quora.com/How-do-I-think-about-an-array-in-Java-or-a-list-in-Python

17. If someone comes up to you with a new factor, how would you consider incorporating it into an existing factor model?

    - https://www.investopedia.com/terms/m/multifactor-model.asp

- https://en.wikipedia.org/wiki/Multiple_factor_models
- https://ocw.mit.edu/courses/mathematics/18-s096-topics-in-mathematics-with-applications-in-finance-fall-2013/lecture-notes/MIT18_S096F13_lecnote15.pdf
- https://faculty.washington.edu/ezivot/research/factormodellecture_handout.pdf
- https://web.stanford.edu/ wfsharpe/mia/fac/mia_fac3.htm

18. Design an algorithm to play Connect 4.
    - https://connect4.gamesolver.org/
    - http://blog.gamesolver.org/solving-connect-four/01-introduction/
    - http://web.mit.edu/sp.268/www/2010/connectFourSlides.pdf
    - https://roadtolarissa.com/connect-4-ai-how-it-works/

19. Write a recursive function to compute the number of partitions of a natural number.
    - https://stackoverflow.com/questions/14053885/integer-partition-algorithm-and-recursion
    - This is also a classic generating function problem.
    - https://www.overleaf.com/read/sbtmxdddtnzy

20. Given some regression filters, talk about their upsides and downside versus principal component analysis and other dimensionality reduction techniques.
    - https://en.wikipedia.org/wiki/Principal_component_analysis
    - https://en.wikipedia.org/wiki/Dimensionality_reduction
    - https://dsp.stackexchange.com/questions/19962/linear-regression-filter-properties
    - http://faculty.chicagobooth.edu/bryan.kelly/research/pdf/forecasting_theory.pdf

21. Design a neural network, hidden Markov model, or state machine to solve the knight's tour problem.
    - https://dmitrybrant.com/knights-tour
    - https://math.stackexchange.com/questions/87991/knights-tour-as-a-neural-network
    - http://www.jamesphoughton.com/2013/09/14/knights-hidden-path-0-hidden-markov.html
    - http://stanford.edu/ cpiech/cs221/handouts/practiceMidterms.html
    - https://community.computingatschool.org.uk/files/6118/original.pdf
    - https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=8383&context=etd_theses
    - https://en.wikipedia.org/wiki/Finite-state_machine

22. Given a dictionary of words and a string, generate all valid anagrams.
    - https://stackoverflow.com/questions/20680145/best-algorithm-to-find-anagram-of-word-from-dictonary

- https://stackoverflow.com/questions/25298200/given-a-dictionary-and-a-list-of-letters-find-all-valid-words-that-can-be-built

23. You have 12 balls that appear identical. However, one is a different weight from the others (could be either lighter or heavier). You also have a balance scale. With only three weighs on the scale, devise a method to find the odd ball and determine if it is heavier or lighter.

    - https://www.mathsisfun.com/pool_balls_solution.html

24. You are given two unfair coins. You flip both of them and ones comes up heads $\frac{2}{3}$ of the time while the other comes up heads $\frac{1}{3}$ of the time. Given you had a uniform prior on the bias before flipping, what is the probability that the first coin is more biased than the second coin?

    - https://math.stackexchange.com/questions/1114093/why-would-a-uniform-prior-distribution-give-a-different-result-than-a-purely-fre
    - https://www.probabilisticworld.com/calculating-coin-bias-bayes-theorem/
    - https://stats.stackexchange.com/questions/291955/bayesian-biased-prior-formula
    - https://math.stackexchange.com/questions/1689448/statistical-testing-of-a-biased-coin

25. What is the expected number of draws from a standard deck until you see an ace?

    - https://math.stackexchange.com/questions/1138853/expected-number-of-cards-you-should-turn-before-finding-an-ace

26. You are given two eggs, and access to a 100-story building. Both eggs are identical. The aim is to find out the highest floor from which an egg will not break when dropped out of a window from that floor. If an egg is dropped and does not break, it is undamaged and can be dropped again. However, once an egg is broken, thats it for that egg. Generalize for any number of eggs and floors and code the problem with dynamic programming.

    - https://www.geeksforgeeks.org/egg-dropping-puzzle-dp-11/
    - http://datagenetics.com/blog/july22012/index.html

27. Aaron samples from the Uniform(0,1) distribution. Then Brooke repeatedly samples from the same distribution until she obtains a number higher than Aaron's. How many samples is she expected to make?

    Solution. Call Aaron's number a (with associated random variable A) and let B be a random variable that represents Brooke's sample at any given point. We know that $P(B < a) = a$, $P(B > a) = 1 - a$, and of course $P(B = a) = 0$, as it is a continuous distribution. We now find the conditional expectation from the geometric distribution. Since the probability of getting a number higher than Aaron's is $1 - a$, we expect $\frac{1}{1-a}$ samples. Note that this is the conditional expectation of the number of samples given the event $A = a$, not the final answer. We now find the unconditional expectation of the number of samples. To do this, let N be a random variable that represents the unconditional number of samples before exceeding A.

We have calculated $\mathrm{E}[N|A=a] = \frac{1}{1-a}$. To find $\mathrm{E}[N]$ explicitly, we must invoke the **Law of Iterated Expectation**. This can be done by noting $\mathrm{E}[N|A] = \frac{1}{1-A}$. Thus,

$$\mathrm{E}[N] = \mathrm{E}[\mathrm{E}[N|A]] = \mathrm{E}\left[\frac{1}{1-A}\right].$$

Since the function $N = \frac{1}{1-A} = g(A)$ is monotonically increasing on $(0,1)$, it has an inverse on the interval. We can use the inverse function method to find the distribution of $N = g(A)$.

$$f_N(n) = f_A(g^{-1}(n))\frac{d}{dn}g^{-1}(n) = 1 \cdot \frac{d}{dn}\left(\frac{n-1}{n}\right) = \frac{1}{n^2} \text{ for } 0 < n < \infty.$$

Thus, we can conclude

$$\mathrm{E}[N] = \sum_{n=1}^{\infty} \frac{1}{n^2} = \boxed{\frac{\pi^2}{6}}.$$

The last line arises from the classic **Basel problem** result.

28. You are given n unit vectors in n-dimensional space. Find a vector that forms the same angle to all of them.

    Solution. We first present a less efficient solution. Consider the $(n-1)$-sphere that intersects the end of all of the vectors. Find the center of this sphere and then solve n equations in n variables.

    Now for a more clever solution. Call our desired vector w and the vectors v_i for $i = 1, \ldots, n$. Note that $w \cdot (v_i - v_j) = 0$ for all $i \neq j$. This is because w makes the same angle to all v_i's. Thus, w is orthogonal to $\mathrm{span}(v_i - v_1)$ for all $2 \leq i \leq n$. Compute this subspace you are done, and it can be done in $O(n)$ time with the Gram-Schmidt process to find a fully orthogonal vector.

29. You have a strategy with supposed Sharpe ratio 8. After n days it has lost money. What does n have to be before you reject the hypothesis that the Sharpe is 8?

    - https://www.investopedia.com/terms/s/sharperatio.asp
    - https://stats.stackexchange.com/questions/155223/testing-sharpe-ratio-significance
    - http://www.econ.uzh.ch/static/wp_iew/iewwp320.pdf

30. What are some methods to prevent overfitting?

    Solution. Add more data, use data augmentation, use models that generalize well, add regularization, reduce model complexity, cross-validation, early stopping, ensemble methods such as bootsrapping or bagging (not going to add links for all of these but one should be familiar with all of them).

31. Given two sorted arrays of integers of lengths m and n where $m << n$, find their intersection in $O\left(n \log\left(\frac{m}{n}\right)\right)$ time.

    - https://articles.leetcode.com/here-is-phone-screening-question-from/
    - https://siderite.blogspot.com/2016/08/finding-intersection-of-two-large.html
    - https://www.geeksforgeeks.org/union-and-intersection-of-two-sorted-arrays-2/

32. Find the expected number of cycles of length greater than $\frac{n}{2}$ in a random permutation of $\{1, \ldots, n\}$.

    - https://math.stackexchange.com/questions/73550/the-limit-of-truncated-sums-of-harmonic-series-lim-limits-k-to-infty-sum-n

33. Devise a way to uniformly sample from a disk. Then do it without square rooting.

    - http://mathworld.wolfram.com/DiskPointPicking.html
    - https://667-per-cm.net/2016/09/23/uniform-sampling-of-a-disk-and-implications-for-sampling-the-internet/
    - https://math.stackexchange.com/questions/927347/uniform-distribution-over-disk

34. Find the eigenvalues of an $n \times n$ matrix with $n's$ on the diagonal and 1's everywhere else

    - https://math.stackexchange.com/questions/175228/suppose-a-is-an-n-by-n-matrix-with-its-diagonal-entries-are-n-and-other-entries

35. Given i.i.d. random variables $X, Y \sim N(0, 1)$, find the conditional distribution of X given that $X + Y > 0$. Prove this is a valid probability distribution.

    Solution. The sum of i.i.d. normal variables is also normal. One can prove this with moment generating functions. The distribution of $X + Y$ is $N(0 + 0, 1 + 1) = N(0, 2)$. Use this distribution to find the conditional distribution.

36. Implement a stack with two queues. Then do it with one queue (the dumb way).

    Solution. Move everything from one queue to the next except for the last element, then return the last element. Continue doing this by alternating the queues. This can be done with two loops. The dumb way with one queue is to try `queue.push(queue.pop())` $n - 1$ times and then do `queue.pop()`.

37. Efficiently find (and then program) a way to find a number that uses each of the digits $1, \ldots, 9$ exactly once and such that the number determined by the first k digits is divisible by k for all $k \in \{1, \ldots, 9\}$.

    - http://mathforum.org/library/drmath/view/56742.html

38. Create a classifier for data using a support-vector machine.

    - https://en.wikipedia.org/wiki/Support-vector_machine
    - https://blog.statsbot.co/support-vector-machines-tutorial-c1618e635e93
    - http://web.mit.edu/zoya/www/SVM.pdf
    - https://www.svm-tutorial.com/

39. Your friend claims he can tell the five colors of skittles apart by taste alone. The probability of a skittle being any particular color is $\frac{1}{5}$. You give your friend 3 skittles and he gets 2 correct. Should you believe him? What if you give him 100 and he gets 40 correct?

    - https://www.channelfireball.com/articles/magic-math-how-many-games-do-you-need-for-statistical-significance-in-playtesting/

Solution. This is a classic hypothesis testing problem. Test the null of $p = 0.2$ versus $p > 0.2$. Then use the normal approximation from the binomial distribution or a binomial distribution calculator.

40. Fast mental mathematics questions:

    - What is $0.5382 - 0.332$?
    - Calculate $\frac{1}{2} + \frac{1}{4} + \frac{1}{16}$.
    - Calculate $23 \cdot 21$.
    - https://www.intmath.com/blog/letters/intmath-newsletter-wus-squaring-trick-patrickjmt-google-calculus-8858
    - https://www.quora.com/Whats-a-math-trick-that-is-not-very-well-known

41. Discuss how you would model the acquaintance graph of the United States. Use this model to guess the average degree of a vertex over this graph.

    - https://www.princeton.edu/ mjs3/mccormick_salganik_zheng10.pdf
    - http://www.stat.columbia.edu/ gelman/research/published/DiPreteetal.pdf

42. Given two data sets X and Y, we run two linear regressions to obtain $y \sim ax + b$ and $x \sim cy + d$. What are the bounds on ac?

    - https://en.wikipedia.org/wiki/Simple_linear_regression

    Solution. For simplicity one can assume that both data sets have mean 0, as they can always be scaled to mean 0. Then we simply use the covariance definition of slope and finish with the Cauchy-Schwarz inequality on the sums. The bounds will be between 0 and 1.

43. Give an example of two variables that are uncorrelated but dependent.

    - https://stats.stackexchange.com/questions/85363/simple-examples-of-uncorrelated-but-not-independent-x-and-y

44. Choose $n - 1$ points randomly on a line segment and break the segment at those points. What is the probability that the resulting n segments form an n-gon?

    - https://math.stackexchange.com/questions/2848881/the-probability-of-those-n-broken-parts-of-sticks-to-form-a-closed-polygon

45. Estimation questions galore:

    - Weight of an average adult giraffe,
    - Number of US businesses that went bankrupt in 2014,
    - Largest change of temperature over one day in the US,
    - Weight of your phone if it was converted to solid gold,
    - Average global life expectancy of a woman,
    - 0.99^{100},
    - Smallest n such that $n!$ has 100 digits,

- $3^{3.8}$,
- $x^x = 1000$ estimate x,
- $\ln(314)$.

Solution. The majority of these are not intended for one to obtain an exact answer, but for discussion and mental evaluation. The numerical ones are a bit more interesting.

For exponents, it is generally a good idea to use logarithms and then observe behaviors at small or large values. For instance, when estimating $x = 0.99^{100}$, one can perform $\ln(x) = 100\ln(0.99) = 100\ln(1 - 0.01) \approx 100(-0.01)$. This is because for small x, $\ln(1 + x)$ is very close to x. Thus, $\ln(x) \approx -1$ and $x \approx e^{-1}$. This is very close to the actual answer.

For the $n!$ problem, the number of base-10 digits of an integer m is $\lfloor \log_{10}(m) \rfloor + 1$. We are looking for the smallest n such that $\lfloor \log_{10}(n!) \rfloor \geq 99$. We can then change our base from 10 to e and use Stirling's approximation to observe the growth of $\ln(n!)$.

For $3^{3.8}$, we will approach the problem two ways. Don't fall into the trap of thinking it is particularly close to 3^4, as exponentiation increases the number very quickly.

1) We first note that $3^{3.8} = \frac{3^4}{3^{0.2}} = \frac{3^4}{\sqrt[5]{3}} = \frac{81}{\sqrt[5]{3}}$. We can estimate fifth roots with a derivative trick. Let $y = \sqrt[5]{3}$. Let x be the nearest fifth power to 3, so $x = 1$ and we write $y = (1+2)^{\frac{1}{5}}$. Here we can consider $\Delta x = 2$ and $y = x^{\frac{1}{5}} = 1$. We are essentially taking the derivative of $y = x^{\frac{1}{5}}$ which is $y' = 5x^{-\frac{4}{5}}$ so we can note that $\Delta y = \frac{\Delta x}{5x^{\frac{4}{5}}} = \frac{2}{5} = 0.4$. Thus, $y + \Delta y = 1 + 0.4 = 1.4$ so $\sqrt[5]{3} \approx 1.4$ and $3^{3.8} \approx \frac{81}{1.4} \approx 57.857$. However, this is not a good estimate, as our Δx is twice as large x itself.

2) Observe the derivative of 3^x directly, or $3^x \ln(3)$. Our estimate will be $3^4 - 3^4 \ln(3)$. To estimate $\ln(3)$, we can use a numerical method (https://math.stackexchange.com/questions/1179348/estimate-ln3-using-taylor-expansion-up-to-3rd-order) and we are done. This gives a much better approximation.

We use a similar method for x^x. Note that $4^4 = 256$ and $5^5 = 3125$ so $4 < x < 5$. Also note the derivative is $x^x(\ln(x) + 1)$. We essentially wish to solve for Δx in $4^4 + 4^4(\ln(4) + 1)\Delta x = 1000$. We can estimate $\ln(4)$ with the same method as the previous problem and we are done.

For $\ln(314)$ we can again use a numerical method to approximate, but for a large number it can be tedious. Instead, note that ln does not grow particularly fast, and that $7^3 = 343$ is close to 314. Thus, $\ln(314) \approx \ln(7^3) = 3\ln(7)$. We can then approximate $\ln(7)$ much easier.

46. You have \$100 and are betting on a fair coin flip. You can bet any percentage of the \$100. If you win, you gain 1.2 times your bet (and your bet back), but if you lose, you lose your bet. What is the optimal bet size to maximize long-run expected earnings?

    - https://en.wikipedia.org/wiki/Kelly_criterion

    Solution. We approach this problem analytically. Let $f(x, y)$ represent the maximum expected money the player can get starting with x red cards and y blue cards. Observing simpler cases, we can see that $f(n, 0) = 2^n$, as there are no blue cards and the player simply doubles their money n times. Our goal is to find a recurrence for $f(n, m)$ with this base case.

    For the first card, the player bets a proportion p of their money on red. p can be negative and this signifies a bet for blue. Then there are two cases: the flipped card is either red or blue.

Now let's observe $f(n,1)$. The profit if red is thus $(1+p)f(n-1,1)$, as the player wins and has $1+p$ of what the player had before, then the process repeats sans one red card. The profit if blue is $(1-p)f(n,0)$ since the player gains nothing and the process repeats sans one blue card. We have $f(n,1) = (1+p)f(n-1,1) + (1-p)f(n,0)$ and in general $f(n,m) = (1+p)f(n-1,m) + (1-p)f(n,m-1)$.

Plugging in $f(n,0) = 2^n$, we get that $f(n,1) = \frac{2^{n+1}}{n+1}$. Note that the proportion p does not even matter! In fact, the general form is

$$f(n,m) = \frac{2^{n+m}}{\binom{n+m}{n}}.$$

Before we prove this, note how interesting it is that p does not matter. The only thing that matters is that when there is only one color remaining the player bets all their money. In fact, the other bets the player makes do not matter either.

Observe $\binom{n+m}{n}$. This is the number of possible orderings of the cards. The term $\frac{2^{n+m}}{\binom{n+m}{n}}$ represents picking a random ordering and betting all of the player's money on it at every stage. If the player wins, they obtain \$$2^{n+m}$ and the probability that they win is $\frac{1}{\binom{n+m}{n}}$. Obviously this strategy has high variance, but it can be modified.

The idea arises from observing the original recurrence $f(n,m) = (1+p)f(n-1,m) + (1-p)f(n,m-1)$. Note that setting $p = \frac{n-m}{m+n}$ gives our desired result of 0 variance, and we can show that our expected profit stays the same. But this method is not particularly motivated.

Instead, let us change our betting method. Previously we picked a random ordering of cards and bet on it, but we can also bet on all the ordering of cards equally and simultaneously. With n red and m blue, $\frac{n}{n+m}$ of the orderings start with red and $\frac{m}{n+m}$ start with blue. Betting the difference of $\frac{n-m}{n+m}$ on red every time clearly has 0 variance, as exactly one ordering will be correct every time. This correct bet will make \$$2^{n+m}$ and $\frac{1}{\binom{n+m}{n}}$ of your money will be bet on it. This is exactly the result obtained from the Kelly criterion.

Miscellaneous Statistics Problems

1. X and Y are i.i.d. $N(0,1)$ random variables. You are given that $X > 0$ and $Y > 0$. What is the probability that $Y > X$?

   Solution. The conditions make this very easy, as by symmetry the answer is simply $\boxed{\frac{1}{2}}$.

2. X and Y are i.i.d. $N(0,1)$ random variables. What's the probability that $Y > 3X$?

   Solution. Rearrange and see that we want $P(Y - 3X > 0)$. Note that the linear combination of i.i.d. normal variables is normal, so $Y - 3X \sim N(0,10)$. Thus, the probability is $\frac{1}{2}$ by symmetry.

3. X and Y are i.i.d. $N(0,1)$ random variables. You are given that $Y > 0$. What is the probability that $Y > 3X$?

   Solution. The key is that $N(0,1)^2$ is cyclically symmetric. When plotting the distributions, the pdf will be cyclically symmetric about the origin. Then one can perform a geometric probability calculation to obtain an answer in terms of arctan.

4. How can one use the normal distribution to sample points uniformly from a disk? How can one use a uniform disk to sample points from a normal distribution?

   - https://en.wikipedia.org/wiki/Box%E2%80%93Muller_transform

   Solution. For the first part, sample x and y from $N(0,1)$. The ordered pair (x, y) can then be normalized. The points are cyclically symmetric, as probability of a point is proportional to $e^{-\frac{x^2+y^2}{2}}$ so for fixed $x^2 + y^2$ all points should have the same probability. For the second part, use the Box-Muller transformation.

5. I flip 10,000 identical coins and 5200 come up heads. Are my coins fair?

   Solution. Another classic hypothesis testing question. Test the null $p = 0.5$ against the alternative $p > 0.5$. Use a binomial distribution calculator or the normal approximation to finish. The coins have very low probability of being fair.

6. Derive the analytical solution for linear regression $\beta = (X^T X)^{-1} X^T y$.

   - https://towardsdatascience.com/analytical-solution-of-linear-regression-a0e870b038d5

   Solution. The goal is to minimize the cost function $J(\beta) = (y - X\beta)^T (y - X\beta)$. Expand and differentiate with respect to β.

7. I take n samples from a distribution. Why is the canonical best estimator for the mean the sample mean? Discuss estimators for the variance and standard deviation. What are the estimators that minimize bias? Are they different from the ones that minimize MSE?

   - https://en.wikipedia.org/wiki/Mean_squared_error
   - https://en.m.wikipedia.org/wiki/Bessel%27s_correction

   Solution. The expected value of the sample mean is the mean. Furthermore, the MSE of the sample mean is always equal to the variance of the sample mean. Estimators that minimize bias are often different from the ones that minimize MSE. For particular estimators such as the one given by Bessel's correction, one may need to invoke the Central Limit Theorem.

8. I have three random variables X, Y, and Z with pairwise correlations all equal to r. Whar are the bounds on r? What are the bounds on $\text{corr}(X, Z)$ if $\text{corr}(X, Y) = a$ and $\text{corr}(Y, Z) = b$?

   - https://math.stackexchange.com/questions/284877/correlation-between-three-variables-question

   Solution. The Cauchy-Schwarz inequality can give us the answer with the classic "correlations are cosines" idea. However, this particular problem can also be solved with the correlation matrix, as correlation matrices are positive-semidefinite. Our particular 3×3 correlation matrix has 1's along the diagonal and r's everywhere else. The eigenvalues are $-\frac{1}{2}$ and 1 which correspond to the minimum and maximum, respectively.

9. Suppose one has two covariance matrices A and B. Is AB also a covariance matrix? What if $AB = BA$?

   - https://math.stackexchange.com/questions/982797/prove-that-the-product-of-two-positive-semidefinite-and-symmetric-matrices-has-n

Solution. The essential argument is that AB must be symmetric for it to be a covariance matrix. Thus the answer to the first part is no. $AB = BA$ satisfies symmetry, so the problem boils down to whether it is positive-semidefinite. One can establish that AB is similar to a positive semidefinite matrix and therefore must be positive semidefinite

10. What happens to the coefficient of determination (R^2) when more independent variables are added to a regression model?

    Solution. Having more covariates will in general give a better fit, however, this does not necessarily mean a better model (in terms of generalization). Thus, model comparison should be carried out at the end in terms of how the model explains the data (e.g., likelihood, R^2, etc), and how simple the model is (i.e., Occam's razor). Overfitting can also be potentially induced, so one can look at the BIC model selection criterion where having too many variables is penalized.

Additional Problems on Glassdoor

https://www.glassdoor.com/Interview/quant-interview-questions-SRCH_KO0,5.htm