**Probability for Data Science Interviews - Summary**

**Introduction to Probability**

Probability is essential in data science, helping quantify uncertainty and make informed decisions. Interviews often test candidates on probability concepts, focusing on topics like **conditional probability, distributions, counting techniques, and Markov chains**, which have practical applications in finance, healthcare, and A/B testing.

**Key Probability Concepts for Interviews**

**1. Conditional Probability**

Conditional probability measures the likelihood of an event occurring given that another event has already happened. This concept is widely applied in fraud detection, medical testing, and recommendation systems.

**Example:**

* Given that a person tests positive for a rare disease, what is the probability they actually have it?

**Key Takeaways:**

* Helps refine predictions when new information is introduced.
* Commonly tested through questions related to **Bayes' Theorem**.
* Essential for understanding dependent and independent events.

**Interview Tip:** Expect questions requiring updates to probabilities based on newly observed data.

**2. Law of Total Probability**

This principle helps break down complex problems by considering all possible outcomes and their associated probabilities.

**Example:**

* Estimating customer churn probability based on different user segments.

**Key Takeaways:**

* Useful for decomposing a problem into smaller, manageable parts.
* Frequently applied in business scenarios like forecasting and risk analysis.

**Interview Tip:** Be prepared to segment data and analyze probability contributions from different sources.

**3. Counting Principles (Combinatorics)**

Counting principles help calculate the number of ways events can occur, critical in scenarios like team formation, permutations, and combinations.

**Example:**

* How many ways can a team of 3 be selected from a pool of 10 candidates?

**Key Takeaways:**

* Permutations: When order matters.
* Combinations: When order does not matter.

**Interview Tip:** Clarify whether the order of selection is relevant before solving.

**4. Random Variables**

Random variables quantify uncertainty and can be either:

* **Discrete:** Countable values (e.g., number of website visitors).
* **Continuous:** Infinite possible values (e.g., customer browsing time).

**Example:**

* Tracking the number of daily purchases on an e-commerce platform.

**Key Takeaways:**

* Understand when to use discrete vs. continuous models.
* Important for statistical modeling and machine learning applications.

**Interview Tip:** Be ready to discuss real-world applications and select the appropriate type of random variable.

**5. Probability Distributions**

Distributions describe how values of a random variable are spread. Important types include:

* **Binomial Distribution:** For binary outcomes (e.g., coin flips).
* **Poisson Distribution:** For rare event occurrences (e.g., system failures).
* **Normal Distribution:** For naturally occurring data (e.g., test scores).

**Example:**

* Predicting the number of website visits per hour using Poisson distribution.

**Key Takeaways:**

* Distributions help model real-world phenomena accurately.
* Recognizing the correct distribution is crucial for problem-solving.

**Interview Tip:** Be ready to identify the correct distribution for different data types.

**6. Markov Chains**

Markov Chains model processes where the next state depends only on the current state, commonly used in user behavior modeling and predictive analytics.

**Example:**

* Modeling user engagement transitions from 'new' to 'active' to 'churned'.

**Key Takeaways:**

* Useful in predicting future states based on current data.
* Involves recurrent and transient states.

**Interview Tip:** Expect questions on state transition probabilities and their applications in business forecasting.

**Common Interview Questions**

1. **Conditional Probability:**
   * "Given 70% of customers are likely to buy and 30% of those click an ad, what's the probability of a purchase after clicking?"
2. **Counting Problems:**
   * "How many ways can a committee of 4 be chosen from a group of 12?"
3. **Random Variables:**
   * "Explain the difference between discrete and continuous random variables with examples."
4. **Distributions:**
   * "What distribution would you use for modeling customer complaints arriving at a call center?"
5. **Markov Chains:**
   * "How would you use a Markov Chain to model customer retention patterns?"

**Best Practices for Interview Success**

* **Think Conceptually:** Focus on understanding real-world applications rather than memorizing formulas.
* **Clarify Assumptions:** Verify if events are independent or dependent.
* **Decompose Problems:** Break complex scenarios into smaller parts.
* **Practice Practical Scenarios:** Apply concepts to business cases to enhance understanding.

**Key Takeaways**

* Probability concepts are essential for data science interviews, focusing on conditional probability, distributions, and counting principles.
* Mastery of these concepts helps in solving complex business and technical problems.
* Focus on understanding practical applications rather than theoretical derivations.

**Probability Interview Questions and Answers**

**5.1 Google: Two teams play a series of games (best of 7) in which each team has a 50% chance of winning any given round. What is the probability that the series goes to 7 games?**

**How to Start:**

*"This problem can be solved using the concept of binomial distribution, as we need exactly 3 wins for one team before the final game."*

**Solution:**

1. **Required Condition:**
   * The series must reach 3-3 before the final 7th game.
   * The first 6 games must result in 3 wins and 3 losses for each team.
2. **Binomial Distribution Application:**
   * The probability of a 3-3 tie after 6 games follows a binomial distribution: P(X=3)=(63)(0.5)3(0.5)3P(X = 3) = \binom{6}{3} (0.5)^3 (0.5)^3P(X=3)=(36​)(0.5)3(0.5)3
   * Evaluating the binomial coefficient: (63)=20\binom{6}{3} = 20(36​)=20
   * The probability calculation: 20×(0.5)6=20×164=2064=0.312520 \times (0.5)^6 = 20 \times \frac{1}{64} = \frac{20}{64} = 0.312520×(0.5)6=20×641​=6420​=0.3125

**Conclusion:**

*"The probability that the series goes to 7 games is 31.25%."*

**5.2 JP Morgan: Say you roll a die three times. What is the probability of getting two sixes in a row?**

**How to Start:**

*"We analyze the possible positions of two consecutive sixes among the three dice rolls."*

**Solution:**

1. **Possible Cases:**
   * The sixes can appear in positions (1,2), (2,3).
   * The third position should not contain a six.
2. **Probability Calculation:**
   * Probability of rolling a six: 16\frac{1}{6}61​.
   * Probability of not rolling a six: 56\frac{5}{6}65​.
   * The total probability is: 2×(16×16×56)=2×5216=10216≈0.0462 \times \left(\frac{1}{6} \times \frac{1}{6} \times \frac{5}{6}\right) = 2 \times \frac{5}{216} = \frac{10}{216} \approx 0.0462×(61​×61​×65​)=2×2165​=21610​≈0.046

**Conclusion:**

*"The probability of getting two consecutive sixes in three rolls is approximately 4.6%."*

**5.3 Uber: You roll three dice, one after another. What is the probability that you obtain three numbers in a strictly increasing order?**

**How to Start:**

*"The key observation here is that every ordering of the three dice outcomes is equally likely."*

**Solution:**

1. **Total Outcomes:**
   * There are 63=2166^3 = 21663=216 total outcomes.
2. **Counting Favorable Outcomes:**
   * Any set of three distinct numbers can appear in strictly increasing order in only **one** way out of the 3!=63! = 63!=6 possible arrangements.
   * The number of ways to choose 3 unique values from 6 is (63)=20\binom{6}{3} = 20(36​)=20.
3. **Probability Calculation:**

20216=554≈0.0926\frac{20}{216} = \frac{5}{54} \approx 0.092621620​=545​≈0.0926

**Conclusion:**

*"The probability of rolling three numbers in strictly increasing order is approximately 9.26%."*

**5.4 Zenefits: Assume you have a deck of 100 cards with values ranging from 1 to 100, and that you draw two cards at random without replacement. What is the probability that the number of one card is precisely double that of the other?**

**How to Start:**

*"We need to count all favorable cases where one card is exactly twice the value of the other."*

**Solution:**

1. **Counting Favorable Pairs:**
   * Possible pairs include (1,2), (2,4), (3,6), ..., (50,100).
   * There are 50 such favorable pairs.
2. **Total Ways to Draw Two Cards:**

(1002)=100×992=4950\binom{100}{2} = \frac{100 \times 99}{2} = 4950(2100​)=2100×99​=4950

1. **Probability Calculation:**

504950≈0.0101\frac{50}{4950} \approx 0.0101495050​≈0.0101

**Conclusion:**

*"The probability is approximately 1.01%."*

**5.5 JP Morgan: Imagine you are in a 3D space. From (0,0,0) to (3,3,3), how many paths are there if you can move only up, right, and forward?**

**How to Start:**

*"This problem can be solved using combinatorics, treating the movements as arranging sequences of steps."*

**Solution:**

1. **Number of Steps Required:**
   * To reach (3,3,3), we must take 3 steps in each direction (U, R, F), totaling 9 steps.
2. **Counting Unique Paths:**
   * The number of unique paths is the number of ways to arrange 3 U’s, 3 R’s, and 3 F’s: 9!3!3!3!=9×8×73×2×1×66=168\frac{9!}{3!3!3!} = \frac{9 \times 8 \times 7}{3 \times 2 \times 1} \times \frac{6}{6} = 1683!3!3!9!​=3×2×19×8×7​×66​=168

**Conclusion:**

*"The total number of paths is 168."*

**5.6 Amazon: If someone tests positive for a disease, what are the odds they actually have it given the test accuracy and disease rate?**

**How to Start:**

*"This is a classic Bayes' theorem problem where we must update prior probabilities based on new evidence."*

**Solution:**

1. **Given Information:**
   * Disease prevalence: P(D)=11000P(D) = \frac{1}{1000}P(D)=10001​.
   * Test sensitivity: P(+∣D)=0.98P(+ | D) = 0.98P(+∣D)=0.98.
   * False positive rate: P(+∣¬D)=0.05P(+ | \neg D) = 0.05P(+∣¬D)=0.05.
2. **Apply Bayes' Theorem:**

P(D∣+)=P(+∣D)P(D)P(+∣D)P(D)+P(+∣¬D)P(¬D)P(D | +) = \frac{P(+ | D) P(D)}{P(+ | D) P(D) + P(+ | \neg D) P(\neg D)}P(D∣+)=P(+∣D)P(D)+P(+∣¬D)P(¬D)P(+∣D)P(D)​ =0.98×11000(0.98×11000)+(0.05×9991000)= \frac{0.98 \times \frac{1}{1000}}{(0.98 \times \frac{1}{1000}) + (0.05 \times \frac{999}{1000})}=(0.98×10001​)+(0.05×1000999​)0.98×10001​​ =0.000980.00098+0.04995≈0.019= \frac{0.00098}{0.00098 + 0.04995} \approx 0.019=0.00098+0.049950.00098​≈0.019

**Conclusion:**

*"If someone tests positive, there is only a 1.9% chance they actually have the disease."*

**5.7 Facebook: Assume two coins, one fair (having one side heads and one side tails) and one unfair (having both sides tails). You pick one at random, flip it five times, and observe that it comes up tails all five times. What is the probability that you are flipping the unfair coin?**

**How to Start:**

*"We can solve this problem using Bayes' Theorem by updating our prior belief based on observed evidence."*

**Solution:**

1. **Define Events:**
   * Let FFF represent the fair coin and UUU the unfair coin.
   * Prior probabilities: P(F)=P(U)=12P(F) = P(U) = \frac{1}{2}P(F)=P(U)=21​.
   * Probability of observing 5 tails with the fair coin: P(T∣F)=(12)5=132P(T|F) = \left(\frac{1}{2}\right)^5 = \frac{1}{32}P(T∣F)=(21​)5=321​
   * Probability of observing 5 tails with the unfair coin: P(T∣U)=1P(T|U) = 1P(T∣U)=1
2. **Applying Bayes' Theorem:**

P(U∣T)=P(T∣U)P(U)P(T∣U)P(U)+P(T∣F)P(F)P(U | T) = \frac{P(T | U) P(U)}{P(T | U) P(U) + P(T | F) P(F)}P(U∣T)=P(T∣U)P(U)+P(T∣F)P(F)P(T∣U)P(U)​ =1×121×12+132×12= \frac{1 \times \frac{1}{2}}{1 \times \frac{1}{2} + \frac{1}{32} \times \frac{1}{2}}=1×21​+321​×21​1×21​​ =1212+164= \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{64}}=21​+641​21​​ =3233= \frac{32}{33}=3332​

**Conclusion:**

*"The probability that the unfair coin was chosen given the observed data is 3233≈97%\frac{32}{33} \approx 97\%3332​≈97%."*

**5.8 Goldman Sachs: Players A and B are playing a game where they take turns flipping a biased coin with probability ppp of landing on heads. Player A starts the game, and the players pass the coin back and forth until one person flips heads and wins. What is the probability that A wins?**

**How to Start:**

*"This is a classic example of a geometric series in a turn-based probability game. Let's analyze it step by step."*

**Solution:**

1. **Winning Conditions for A:**
   * A wins if they get heads on their first flip (ppp).
   * If A fails (1−p1-p1−p), B gets a turn, and if B also fails, the game resets to A again.
2. **Recursive Formula:**
   * Let PAP\_APA​ be the probability of A winning. Then: PA=p+(1−p)(1−p)PAP\_A = p + (1-p)(1-p) P\_APA​=p+(1−p)(1−p)PA​
   * Solving for PAP\_APA​: PA=p1−(1−p)2P\_A = \frac{p}{1 - (1-p)^2}PA​=1−(1−p)2p​
3. **Simplified Result:**

PA=pp+(1−p)2P\_A = \frac{p}{p + (1-p)^2}PA​=p+(1−p)2p​

**Conclusion:**

*"Player A's winning probability depends on the bias ppp and follows the formula pp+(1−p)2\frac{p}{p + (1-p)^2}p+(1−p)2p​."*

**5.9 Microsoft: Three friends in Seattle each told you it is rainy, and each person has a 1/3 probability of lying. What is the probability that Seattle is really rainy, assuming that the likelihood of rain on any given day is 0.25?**

**How to Start:**

*"We can solve this problem using Bayes' theorem by updating our belief based on the reliability of the information received."*

**Solution:**

1. **Define Events:**
   * Let RRR be the event that it rains, and LLL be the event of lying.
   * Prior probability of rain: P(R)=0.25P(R) = 0.25P(R)=0.25, P(¬R)=0.75P(\neg R) = 0.75P(¬R)=0.75.
   * Probability of truth: P(Truth)=23P(\text{Truth}) = \frac{2}{3}P(Truth)=32​, probability of lying: P(L)=13P(L) = \frac{1}{3}P(L)=31​.
2. **Likelihood Calculations:**
   * If it is raining: P(3 say rain∣R)=(23)3=827P(\text{3 say rain} | R) = \left(\frac{2}{3}\right)^3 = \frac{8}{27}P(3 say rain∣R)=(32​)3=278​.
   * If it is not raining: P(3 say rain∣¬R)=(13)3=127P(\text{3 say rain} | \neg R) = \left(\frac{1}{3}\right)^3 = \frac{1}{27}P(3 say rain∣¬R)=(31​)3=271​.
3. **Applying Bayes' Theorem:**

P(R∣3 say rain)=P(3 say rain∣R)P(R)P(3 say rain∣R)P(R)+P(3 say rain∣¬R)P(¬R)P(R | \text{3 say rain}) = \frac{P(\text{3 say rain} | R) P(R)}{P(\text{3 say rain} | R) P(R) + P(\text{3 say rain} | \neg R) P(\neg R)}P(R∣3 say rain)=P(3 say rain∣R)P(R)+P(3 say rain∣¬R)P(¬R)P(3 say rain∣R)P(R)​ =827×0.25(827×0.25)+(127×0.75)= \frac{\frac{8}{27} \times 0.25}{\left(\frac{8}{27} \times 0.25\right) + \left(\frac{1}{27} \times 0.75\right)}=(278​×0.25)+(271​×0.75)278​×0.25​ =23= \frac{2}{3}=32​

**Conclusion:**

*"Given the three friends' statements, the probability that it is actually raining is 23\frac{2}{3}32​ or 66.67%."*

**5.10 Bloomberg: You draw a circle and choose two chords at random. What is the probability that those chords will intersect?**

**How to Start:**

*"This problem requires an understanding of geometric probability and random chord placement."*

**Solution:**

1. **Random Chord Selection:**
   * A random chord can be defined by picking two random points on the circumference.
2. **Intersection Condition:**
   * Two chords intersect if and only if the four points defining them are arranged in alternating order around the circumference.
3. **Calculation:**
   * Consider the four chosen points; they can be arranged in 4!=244! = 244!=24 possible ways.
   * Among these arrangements, exactly 8 correspond to an intersecting configuration.
4. **Probability Calculation:**

P(intersection)=824=13P(\text{intersection}) = \frac{8}{24} = \frac{1}{3}P(intersection)=248​=31​

**Conclusion:**

*"The probability that two randomly chosen chords will intersect inside the circle is 13\frac{1}{3}31​ or approximately 33.33%."*

**5.11 Morgan Stanley: You and your friend are playing a game where you toss a coin until the sequence HH or TH shows up. What is the probability of you winning?**

**How to Start:**

*"We will solve this problem using conditional probabilities and recursion based on the stopping conditions."*

**Solution:**

1. **Possible sequences:**
   * If the first toss is H, the next toss determines the outcome.
     + HH (you win), probability = 12\frac{1}{2}21​
     + HT (game continues)
   * If the first toss is T, the next toss determines the outcome.
     + TH (friend wins), probability = 12\frac{1}{2}21​
     + TT (game continues)
2. **Recursive approach:**

P(win)=12P(win)+12(0)P(\text{win}) = \frac{1}{2} P(\text{win}) + \frac{1}{2}(0)P(win)=21​P(win)+21​(0)

Solving:

P(win)=13P(\text{win}) = \frac{1}{3}P(win)=31​

**Conclusion:**

*"The probability that you win the game is 13\frac{1}{3}31​ or approximately 33.33%."*

**5.12 JP Morgan: Rolling a 6-sided die up to two times. Expected payout analysis.**

**How to Start:**

*"This is an optimal stopping problem where we decide whether to stop or roll again based on maximizing expected value."*

**Solution:**

1. **Expected value from one roll:**

E[X]=1+2+3+4+5+66=3.5E[X] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5E[X]=61+2+3+4+5+6​=3.5

1. **Decision Strategy:**
   * If first roll ≥4\geq 4≥4, it's better to stop.
   * If first roll <4< 4<4, expected value of rolling again is 3.5.
2. **Expected Value Calculation:**

E[optimal strategy]=16(4+5+6)+16(3.5+3.5+3.5)=4.25E[\text{optimal strategy}] = \frac{1}{6}(4 + 5 + 6) + \frac{1}{6}(3.5 + 3.5 + 3.5) = 4.25E[optimal strategy]=61​(4+5+6)+61​(3.5+3.5+3.5)=4.25

**Conclusion:**

*"You should be willing to pay up to $4.25 to play this game."*

**5.13 Facebook: Probability that a rater is a diligent rater given their decisions.**

**How to Start:**

*"We can use Bayes' theorem to calculate the probability of being diligent based on observed labeling decisions."*

**Solution:**

1. **Given Probabilities:**
   * Diligent rater: 90%, marks correctly 80%.
   * Non-diligent rater: 10%, always marks incorrectly.
2. **Applying Bayes' Theorem:**

P(diligent∣correct label)=P(correct∣diligent)P(diligent)P(correct)P(\text{diligent} | \text{correct label}) = \frac{P(\text{correct} | \text{diligent}) P(\text{diligent})}{P(\text{correct})}P(diligent∣correct label)=P(correct)P(correct∣diligent)P(diligent)​ =0.8×0.9(0.8×0.9)+(0×0.1)=0.720.72+0=1= \frac{0.8 \times 0.9}{(0.8 \times 0.9) + (0 \times 0.1)} = \frac{0.72}{0.72 + 0} = 1=(0.8×0.9)+(0×0.1)0.8×0.9​=0.72+00.72​=1

**Conclusion:**

*"The probability that the rater is diligent given their correct label is 1 (or 100%)."*

**5.14 D.E. Shaw: Given one child is a boy, what is the probability the second child is also a boy?**

**How to Start:**

*"This problem involves conditional probability considering all possible cases."*

**Solution:**

1. **Sample Space:**
   * BB, BG, GB, GG. (Each equally likely, 14\frac{1}{4}41​).
   * Given one child is a boy, eliminate GG.
2. **Favorable Cases:**
   * BB, BG, GB (boy in the pair).
   * Probability: P(both boys∣one boy)=13P(\text{both boys} | \text{one boy}) = \frac{1}{3}P(both boys∣one boy)=31​

**Conclusion:**

*"The probability that the second child is also a boy is 13\frac{1}{3}31​."*

**5.15 JP Morgan: Probability that the last drawer contains a letter after checking 7 empty ones.**

**How to Start:**

*"This is a conditional probability problem where we're updating probabilities based on observed information."*

**Solution:**

1. **Initial Probability:**
   * Letter in any drawer =12= \frac{1}{2}=21​.
2. **Conditional Calculation:**
   * Probability of no letter in 7 drawers: P(7 empty)=(12)7P(\text{7 empty}) = \left(\frac{1}{2}\right)^7P(7 empty)=(21​)7
   * Probability of letter in the 8th drawer: P(8th has letter∣7empty)=11+27=1129P(\text{8th has letter} | 7 empty) = \frac{1}{1 + 2^7} = \frac{1}{129}P(8th has letter∣7empty)=1+271​=1291​

**Conclusion:**

*"The probability the letter is in the last drawer is 1129\frac{1}{129}1291​."*

**5.16 Optiver: Tennis match at deuce, probability of winning.**

**How to Start:**

*"We will solve this problem using recursive probability models."*

**Solution:**

1. **Recursive Equation:**

PA=0.6PA+0.4(1−PA)P\_A = 0.6 P\_A + 0.4 (1 - P\_A)PA​=0.6PA​+0.4(1−PA​)

1. **Solving:**

PA=0.60.6+0.4=0.6P\_A = \frac{0.6}{0.6 + 0.4} = 0.6PA​=0.6+0.40.6​=0.6

**Conclusion:**

*"The probability that the first player wins is 60%."*

**5.17 Facebook: Probability two cards have different colors and numbers.**

**How to Start:**

*"This problem involves combinatorics and probability of selecting cards with specific attributes."*

**Solution:**

1. **Total possible selections:**

(502)=1225\binom{50}{2} = 1225(250​)=1225

1. **Favorable outcomes:**
   * Number of ways to choose different colors and numbers.
2. **Probability Calculation:**

20001225≈0.816\frac{2000}{1225} \approx 0.81612252000​≈0.816

**Conclusion:**

*"The probability is approximately 81.6%."*

**5.18 SIG: Probability that sum of ten dice is divisible by 6.**

**How to Start:**

*"We solve this using modular arithmetic and symmetry principles of uniform distributions."*

**Solution:**

1. **Divisibility Rules:**
   * Each die has values {1,2,3,4,5,6}\{1,2,3,4,5,6\}{1,2,3,4,5,6}.
   * Probability of divisibility by 6: 16\frac{1}{6}61​.

**Conclusion:**

*"The probability is 16\frac{1}{6}61​ or 16.67%."*

**medium-level** probability interview questions.

**5.19 Morgan Stanley: A and B play a game where they roll a die until the first person rolls a six. How much is A willing to pay to play first if the reward is $100?**

**How to Start:**

*"This problem can be solved using expected value analysis and geometric probability distribution."*

**Solution:**

1. **Winning Probabilities:**
   * A wins if they roll a 6 first: P(A wins)=16+(56)2P(A wins)P(\text{A wins}) = \frac{1}{6} + \left(\frac{5}{6}\right)^2 P(\text{A wins})P(A wins)=61​+(65​)2P(A wins)
   * Solving the equation: P(A wins)=16+2536P(A wins)P(\text{A wins}) = \frac{1}{6} + \frac{25}{36} P(\text{A wins})P(A wins)=61​+3625​P(A wins)
   * Simplifying: P(A wins)=16×3611=611P(\text{A wins}) = \frac{1}{6} \times \frac{36}{11} = \frac{6}{11}P(A wins)=61​×1136​=116​
2. **Expected Value Calculation:**
   * If A wins with probability 611\frac{6}{11}116​, the expected payout is: E[value]=611×100=54.55E[\text{value}] = \frac{6}{11} \times 100 = 54.55E[value]=116​×100=54.55

**Conclusion:**

*"A should be willing to pay up to $54.55 to play first."*

**5.20 Airbnb: How can you generate fair odds using an unfair coin with unknown bias?**

**How to Start:**

*"We can generate fair outcomes using the von Neumann trick, which leverages symmetry."*

**Solution:**

1. **Strategy to Create Fair Odds:**
   * Flip the coin twice:
     + If HT appears → assign it as "heads" (fair outcome).
     + If TH appears → assign it as "tails" (fair outcome).
     + If HH or TT appears, discard and flip again.
2. **Probability Justification:**
   * Since the events HT and TH have equal probability regardless of bias ppp, they create fair 50-50 odds.

**Conclusion:**

*"Using the von Neumann method, fair odds can be generated from an unfair coin."*

**5.21 SIG: You randomly pick a small cube from a larger 3x3x3 cube where 6 outside faces were painted green. What is the probability the bottom face is white?**

**How to Start:**

*"We analyze the cube structure to calculate the probability of selecting an internal (unpainted) cube."*

**Solution:**

1. **Total Small Cubes:**
   * A 3×3×33 \times 3 \times 33×3×3 cube contains 27 small cubes.
2. **Painted Cubes Calculation:**
   * Only the external layer of the cube is painted. The interior 1x1x1 cube remains unpainted.
   * The number of unpainted cubes in the bottom face is 1 (center).
3. **Probability Calculation:**
   * Selecting the bottom face white cube: P(bottom white)=127P(\text{bottom white}) = \frac{1}{27}P(bottom white)=271​

**Conclusion:**

*"The probability of picking a white cube on the bottom face is 127\frac{1}{27}271​."*

**5.22 Goldman Sachs: Suppose you take a stick of length 1 and break it uniformly at two random points. What is the probability the three pieces can form a triangle?**

**How to Start:**

*"This is a geometric probability problem where we analyze the triangle inequality condition."*

**Solution:**

1. **Triangle Inequality Condition:**
   * To form a triangle, the sum of the two smaller pieces must be greater than the largest piece.
2. **Uniform Breaks:**
   * If the stick is broken at points XXX and YYY (where X<YX < YX<Y), the three segments are X,Y−X,1−YX, Y-X, 1-YX,Y−X,1−Y.
3. **Valid Condition:**
   * The condition for a triangle: X+(Y−X)>(1−Y),X+(1−Y)>(Y−X),(Y−X)+(1−Y)>XX + (Y - X) > (1 - Y), \quad X + (1 - Y) > (Y - X), \quad (Y - X) + (1 - Y) > XX+(Y−X)>(1−Y),X+(1−Y)>(Y−X),(Y−X)+(1−Y)>X
   * Solving this using Monte Carlo simulation or known results yields: 14\frac{1}{4}41​

**Conclusion:**

*"The probability that the three pieces can form a triangle is 14\frac{1}{4}41​ or 25%."*

**5.23 Lyft: What is the probability that in a random sequence of H’s and T’s, HHT shows up before HTT?**

**How to Start:**

*"This problem can be modeled using Markov chains or probabilistic recursion techniques."*

**Solution:**

1. **State Transitions:**
   * If the sequence starts with H, the sequence could evolve to HHT or HTT.
   * Use state analysis to count possible paths.
2. **Probabilistic Approach:**
   * Defining winning states: P(HHT first)=23P(\text{HHT first}) = \frac{2}{3}P(HHT first)=32​
   * This is derived using symmetric probability approaches based on the structure of sequences.

**Conclusion:**

*"The probability that HHT appears before HTT is 23\frac{2}{3}32​ or approximately 66.67%."*

**5.24 Uber: A fair coin is tossed twice. Is it more likely that at least one heads appeared or that the second toss was heads?**

**How to Start:**

*"We analyze both events separately using the sample space of coin tosses and compare their probabilities."*

**Solution:**

1. **Sample Space:**
   * The possible outcomes of two tosses: {HH,HT,TH,TT}\{HH, HT, TH, TT\}{HH,HT,TH,TT}.
   * Total possible outcomes = 4.
2. **Case 1: At least one heads (H):**
   * Favorable outcomes: HH,HT,THHH, HT, THHH,HT,TH (3 out of 4).
   * Probability = 34\frac{3}{4}43​.
3. **Case 2: Second toss is heads:**
   * Favorable outcomes: HT,HHHT, HHHT,HH (2 out of 4).
   * Probability = 24=12\frac{2}{4} = \frac{1}{2}42​=21​.

**Conclusion:**

*"It is more likely that at least one heads appeared (34\frac{3}{4}43​) than the second toss being heads (12\frac{1}{2}21​). If the coin were biased, the probabilities would change accordingly."*

**5.25 Facebook: Three ants start at the corners of an equilateral triangle and move randomly along the edges. What is the probability that none of the ants meet?**

**How to Start:**

*"We solve this problem by considering the symmetry and possible movement patterns of the ants."*

**Solution:**

1. **Possible Movement Directions:**
   * Each ant can move clockwise (CW) or counterclockwise (CCW), each with a probability of 12\frac{1}{2}21​.
2. **Total Outcomes:**
   * The ants can move in 23=82^3 = 823=8 possible ways.
3. **Favorable Cases (No Meeting):**
   * All ants move CW or all ants move CCW, which results in no collision.
   * Favorable outcomes = 2.
4. **Probability Calculation:**

P(no meeting)=28=14P(\text{no meeting}) = \frac{2}{8} = \frac{1}{4}P(no meeting)=82​=41​

**Conclusion:**

*"The probability that none of the ants meet is 14\frac{1}{4}41​ or 25%. The problem generalizes for k ants on an equilateral polygon."*

**5.26 Robinhood: A biased coin (probability ppp of heads) is tossed nnn times. Derive a recurrence relation for total number of heads.**

**How to Start:**

*"Let’s define the recurrence relation based on the outcomes of each toss."*

**Solution:**

1. **Define Random Variable:**
   * Let H(n)H(n)H(n) be the total number of heads after nnn tosses.
   * H(n)H(n)H(n) depends on whether the last toss is a head or tail.
2. **Recurrence Relation:**

H(n)=H(n−1)+XnH(n) = H(n-1) + X\_nH(n)=H(n−1)+Xn​

* + Where XnX\_nXn​ is 1 with probability ppp and 0 with probability 1−p1-p1−p.

1. **Expectation Formula:**
   * Taking expectations on both sides: E[H(n)]=E[H(n−1)]+pE[H(n)] = E[H(n-1)] + pE[H(n)]=E[H(n−1)]+p
   * Expanding recursively: E[H(n)]=npE[H(n)] = npE[H(n)]=np

**Conclusion:**

*"The recurrence relation for the expected number of heads is H(n)=H(n−1)+XnH(n) = H(n-1) + X\_nH(n)=H(n−1)+Xn​, leading to E[H(n)]=npE[H(n)] = npE[H(n)]=np."*

**5.27 Citadel: Alice and Bob play a game. Alice wins a round with probability ppp. What is the probability that Bob wins the overall series if Alice needs to win two more rounds than Bob?**

**How to Start:**

*"This problem can be analyzed using recursive probability based on geometric distribution."*

**Solution:**

1. **Define Probability Terms:**
   * Alice needs to win 2 more rounds than Bob.
2. **Recursive Formula:**

P(Bob wins)=(1−p)P(Bob wins)+p(1−p)P(Bob wins)+p2P(\text{Bob wins}) = (1-p)P(\text{Bob wins}) + p(1-p)P(\text{Bob wins}) + p^2P(Bob wins)=(1−p)P(Bob wins)+p(1−p)P(Bob wins)+p2

1. **Solving for Probabilities:**
   * Using geometric sum properties and combinatorics, the probability is: (1−p)21−2p(1−p)\frac{(1-p)^2}{1 - 2p(1-p)}1−2p(1−p)(1−p)2​

**Conclusion:**

*"The probability that Bob wins the overall series is (1−p)21−2p(1−p)\frac{(1-p)^2}{1 - 2p(1-p)}1−2p(1−p)(1−p)2​."*

**5.28 Google: You have three draws from a uniform distribution U(0,2)U(0,2)U(0,2). What is the probability that the median is greater than 1.5?**

**How to Start:**

*"We analyze the order statistics to find the median probability."*

**Solution:**

1. **Sample Space and Order:**
   * Sort the three values: X1≤X2≤X3X\_1 \leq X\_2 \leq X\_3X1​≤X2​≤X3​.
   * The median is X2X\_2X2​.
2. **Probability Calculation:**
   * The probability that the median exceeds 1.5 is the event P(X2>1.5)P(X\_2 > 1.5)P(X2​>1.5).
   * The median follows the Beta distribution B(2,2)B(2,2)B(2,2) in the range [0,2].
3. **Final Computation:**

P(X2>1.5)=∫1.526x(2−x)dx=78P(X\_2 > 1.5) = \int\_{1.5}^{2} 6x(2-x) dx = \frac{7}{8}P(X2​>1.5)=∫1.52​6x(2−x)dx=87​

**Conclusion:**

*"The probability that the median is greater than 1.5 is 78\frac{7}{8}87​ or 87.5%."*

**Hard-Level Probability Interview Questions and Answers**

**5.29 D.E. Shaw: Say you have 150 friends, and 3 of them have phone numbers with a permutation of the digits 0, 1, 4, and 9. Is this just a chance occurrence? Why or why not?**

**How to Start:**

*"We will analyze whether this observation is statistically significant using probability principles."*

**Solution:**

1. **Possible Permutations:**
   * The number of ways to arrange four digits (0,1,4,9): 4!=244! = 244!=24
   * Assuming 10-digit phone numbers, the probability of a random number containing a specific permutation is low.
2. **Expected Occurrences:**
   * With 150 friends, assuming uniformly distributed numbers, we can calculate the probability based on the total possible phone numbers (101010^{10}1010).
3. **Statistical Insight:**
   * Using the Poisson approximation for rare events: λ=150×P(having permutation)\lambda = 150 \times P(\text{having permutation})λ=150×P(having permutation)
   * If the observed count significantly exceeds expected occurrences, it may not be due to chance.

**Conclusion:**

*"While it could be a coincidence, further statistical tests (e.g., binomial or Poisson tests) could determine significance."*

**5.30 Spotify: A fair die is rolled nnn times. What is the probability that the largest number rolled is kkk for each k∈{1,2,3,4,5,6}k \in \{1,2,3,4,5,6\}k∈{1,2,3,4,5,6}?**

**How to Start:**

*"This problem involves order statistics and calculating the cumulative probability distribution."*

**Solution:**

1. **Cumulative Distribution Approach:**
   * The probability that all rolls are ≤k\leq k≤k is: (k6)n\left(\frac{k}{6}\right)^n(6k​)n
   * The probability that the largest number is exactly kkk is obtained by subtracting lower probabilities: P(max⁡=k)=(k6)n−(k−16)nP(\max = k) = \left(\frac{k}{6}\right)^n - \left(\frac{k-1}{6}\right)^nP(max=k)=(6k​)n−(6k−1​)n

**Conclusion:**

*"The probability that the maximum roll is kkk is (k6)n−(k−16)n\left(\frac{k}{6}\right)^n - \left(\frac{k-1}{6}\right)^n(6k​)n−(6k−1​)n."*

**5.31 Goldman Sachs: Amoeba splits into 0, 1, 2, or 3 amoebas with equal probability every minute. What is the probability of eventual extinction?**

**How to Start:**

*"This is a branching process problem where we analyze extinction probability recursively."*

**Solution:**

1. **Define Recurrence:**
   * Let ppp be the extinction probability. We have: p=14+14p+14p2+14p3p = \frac{1}{4} + \frac{1}{4} p + \frac{1}{4} p^2 + \frac{1}{4} p^3p=41​+41​p+41​p2+41​p3
   * Solving the equation numerically, p≈0.796p \approx 0.796p≈0.796.

**Conclusion:**

*"The probability of eventual extinction is approximately 0.796 (79.6%)."*

**5.32 Lyft: A fair coin is tossed nnn times. Given that there were kkk heads, what is the probability that the first toss was heads?**

**How to Start:**

*"This problem can be solved using conditional probability and symmetry."*

**Solution:**

1. **Symmetry Argument:**
   * Since each outcome of kkk heads is equally likely across nnn tosses, the probability is evenly distributed.
2. **Combinatorial Approach:**
   * Total ways to choose kkk heads from nnn tosses: (nk)\binom{n}{k}(kn​)
   * Ways in which the first toss is a head: (n−1k−1)\binom{n-1}{k-1}(k−1n−1​)
3. **Final Probability:**

P(first toss is head∣k heads)=(n−1k−1)(nk)=knP(\text{first toss is head} | k \text{ heads}) = \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}P(first toss is head∣k heads)=(kn​)(k−1n−1​)​=nk​

**Conclusion:**

*"The probability that the first toss was heads given kkk total heads is kn\frac{k}{n}nk​."*

**5.33 Quora: You have NNN i.i.d. draws of numbers following a normal distribution with mean μ\muμ and variance σ2\sigma^2σ2. What is the probability that kkk of these draws are larger than some value YYY?**

**How to Start:**

*"This can be modeled using the binomial distribution after standardizing the normal distribution."*

**Solution:**

1. **Standardization:**
   * Convert normal values to standard normal: Z=Y−μσZ = \frac{Y - \mu}{\sigma}Z=σY−μ​
   * Let ppp represent the probability that a single draw exceeds YYY: p=P(X>Y)=1−Φ(Y−μσ)p = P(X > Y) = 1 - \Phi\left(\frac{Y - \mu}{\sigma}\right)p=P(X>Y)=1−Φ(σY−μ​)
2. **Binomial Probability:**
   * The number of draws greater than YYY follows a binomial distribution: P(K=k)=(Nk)pk(1−p)N−kP(K = k) = \binom{N}{k} p^k (1-p)^{N-k}P(K=k)=(kN​)pk(1−p)N−k

**Conclusion:**

*"The probability that exactly kkk draws are greater than YYY is (Nk)pk(1−p)N−k\binom{N}{k} p^k (1-p)^{N-k}(kN​)pk(1−p)N−k."*

**5.34 Akuna Capital: Three random points are chosen on a unit circle. What is the probability that they form an acute-angled triangle?**

**How to Start:**

*"This problem is approached by considering geometric properties of triangles within a circle."*

**Solution:**

1. **Geometric Considerations:**
   * A triangle is acute if the largest angle is <90∘< 90^\circ<90∘.
   * If any of the points lie on opposite semicircles, the triangle is not acute.
2. **Monte Carlo Approximation:**
   * Simulations suggest the probability is approximately 14\frac{1}{4}41​.

**Conclusion:**

*"The probability that the three points form an acute triangle is 14\frac{1}{4}41​."*

**5.35 Citadel: You have rrr red balls and www white balls in a bag. What is the probability that a randomly selected subset contains only red balls?**

**How to Start:**

*"We calculate the probability by considering combinations of red and white balls."*

**Solution:**

1. **Total Subsets:**
   * The total number of ways to choose any subset: 2r+w2^{r+w}2r+w
2. **Favorable Outcomes (Only Red):**
   * The number of ways to choose only from red balls: 2r2^r2r
3. **Final Probability Calculation:**

P(all red)=2r2r+w=2−wP(\text{all red}) = \frac{2^r}{2^{r+w}} = 2^{-w}P(all red)=2r+w2r​=2−w

**Conclusion:**

*"The probability that a randomly chosen subset contains only red balls is 2−w2^{-w}2−w."*