**Key Statistical Concepts for Data Science Interviews**

**1. Overview of Core Statistical Concepts**

**Random Variables and Their Properties**

A random variable represents a numerical outcome of a random process. In statistics, it is essential to understand key properties such as:

* **Expectation (Mean):**  
  The long-run average value of the random variable over many trials. It represents the expected outcome.
* **Variance and Standard Deviation:**  
  Variance measures how far the values spread out from the mean, while the standard deviation quantifies dispersion and is more interpretable in real-world applications.

**Real-world Example:**  
In business, random variables can represent the daily sales of a product, where the mean provides an estimate of average daily revenue.

**Law of Large Numbers (LLN)**

The LLN states that as the number of trials increases, the sample mean will converge to the population mean. This is critical in experimental analysis where repeated measurements ensure accurate results.

**Example:**  
If you keep flipping a fair coin, the proportion of heads will get closer to 50% as the number of flips increases.

**Central Limit Theorem (CLT)**

The CLT states that the distribution of the sample mean approximates a normal distribution, regardless of the population’s original distribution, given a large enough sample size.

**Why It Matters:**  
It allows us to apply statistical methods to many real-world problems, such as A/B testing, where sample means are analyzed.

**Example:**  
In web analytics, the average session duration across thousands of users tends to follow a normal distribution, even if individual session times are skewed.

**2. Hypothesis Testing and Statistical Inference**

**Hypothesis Testing Process:**

Hypothesis testing helps determine whether there is enough evidence to support a claim about a population. The steps include:

1. Formulating a **null hypothesis (H₀)** (e.g., no effect) and an **alternative hypothesis (H₁)** (e.g., an effect exists).
2. Choosing an appropriate statistical test.
3. Calculating the test statistic and p-value.
4. Comparing the p-value to the significance level (e.g., 0.05) to make a decision.

**Common Statistical Tests:**

* **Z-Test:** Used when the sample size is large or the population variance is known.
* **T-Test:** Suitable for small sample sizes where variance is unknown.
* **Chi-Square Test:** Evaluates relationships between categorical variables (e.g., customer preferences by region).

**Example:**  
A company may use a T-test to compare the average sales of two stores to determine if there is a significant difference.

**3. Confidence Intervals and P-values**

**Confidence Intervals:**

A confidence interval provides a range within which the true population parameter is expected to fall with a given confidence level (e.g., 95%).

**Example:**  
If the average revenue per user is $50 with a 95% confidence interval of [$45, $55], we can be confident that the true average lies within this range.

**P-values:**

The p-value indicates the likelihood of obtaining the observed results under the assumption that the null hypothesis is true. A lower p-value suggests stronger evidence against the null hypothesis.

**Key Point:**  
A p-value below 0.05 is commonly considered statistically significant, meaning the results are unlikely due to chance.

**4. Practical Applications in Data Science Interviews**

**A/B Testing:**

A practical use of hypothesis testing is A/B testing, which is used to compare two product variations to determine which performs better based on user engagement metrics.

**Example:**  
A company might test two different website layouts and compare conversion rates to decide which design is more effective.

**5. Estimation Techniques: MLE vs. MAP**

**Maximum Likelihood Estimation (MLE):**

MLE estimates parameters by maximizing the likelihood of observing the given data. It assumes no prior knowledge about the parameters.

**Example:**  
Predicting customer churn based on past data by finding the probability that maximizes observed behaviors.

**Maximum A Posteriori (MAP):**

MAP incorporates prior knowledge into the estimation process, making it useful when domain expertise or historical data is available.

**Example:**  
If we already have data from past marketing campaigns, MAP can help refine predictions using prior insights.

**6. Interview Tips and Preparation Strategy**

**Common Interview Questions:**

1. Explain the Central Limit Theorem and its significance in data analysis.
2. When should you use a T-test instead of a Z-test?
3. How do you interpret a confidence interval in business decision-making?
4. What is the difference between Type I and Type II errors?
5. How would you approach an A/B testing problem?

**Best Practices for Interviews:**

* **Understand Assumptions:** Always verify the assumptions of statistical tests before applying them.
* **Use Intuitive Explanations:** Frame responses with real-world analogies.
* **Communicate Clearly:** Avoid unnecessary jargon and focus on clear, structured answers.

**7. Key Takeaways for Interview Success**

1. **Master Core Concepts:** Focus on understanding statistical fundamentals rather than memorizing formulas.
2. **Think Practically:** Be prepared to apply statistical concepts to solve business problems.
3. **Explain Confidently:** Practice explaining complex concepts in simple terms.
4. **Avoid Common Pitfalls:** Ensure you distinguish between statistical significance and practical significance.

**40 Real Statistics Interview Questions**

**6.1. Uber: Explain the Central Limit Theorem. Why is it useful?**  
**Answer:**  
The Central Limit Theorem (CLT) states that, regardless of the population's distribution, the distribution of the sample mean will approximate a normal distribution if the sample size is sufficiently large.

**Why it is useful:**

* It allows us to apply inferential statistics even when the population distribution is unknown.
* It helps in hypothesis testing and constructing confidence intervals.
* It is widely used in A/B testing, quality control, and other business analytics applications.

**Example:**  
If we measure customer wait times across multiple stores, their distribution may be skewed. However, with a large enough sample, the average wait times will follow a normal distribution, making predictions and analyses easier.

**6.2. Facebook: How would you explain a confidence interval to a non-technical audience?**  
**Answer:**  
A confidence interval is a range of values that likely contains the true value of a population parameter (such as an average). For example, if we estimate the average delivery time of food is 30 minutes with a 95% confidence interval of [28, 32], it means we're 95% confident that the actual average falls within this range.

**Key points to convey:**

* It accounts for uncertainty in our estimates.
* A wider interval means less precision; a narrower one indicates higher precision.
* The confidence level (e.g., 95%) means that if we repeat the study multiple times, 95% of the intervals will contain the true value.

**6.3. Twitter: What are some common pitfalls encountered in A/B testing?**  
**Answer:**  
Some common pitfalls include:

* **Insufficient sample size:** Drawing conclusions too early without enough data leads to unreliable results.
* **Multiple comparisons problem:** Running too many tests increases the chance of false positives.
* **Lack of randomization:** Non-randomized groups can introduce bias.
* **Ignoring seasonality:** Comparing data across different time periods can produce misleading results.
* **Misinterpreting statistical significance:** Statistical significance doesn’t always mean practical significance.

**6.4. Lyft: Explain both covariance and correlation formulaically, and compare and contrast them.**  
**Answer:**

* **Covariance formula:**

Cov(X,Y)=∑(Xi−Xˉ)(Yi−Yˉ)nCov(X, Y) = \frac{\sum (X\_i - \bar{X})(Y\_i - \bar{Y})}{n}Cov(X,Y)=n∑(Xi​−Xˉ)(Yi​−Yˉ)​

It measures how two variables change together. A positive value means they increase together, while a negative value means they move in opposite directions.

* **Correlation formula:**

Cor(X,Y)=Cov(X,Y)σXσYCor(X, Y) = \frac{Cov(X, Y)}{\sigma\_X \sigma\_Y}Cor(X,Y)=σX​σY​Cov(X,Y)​

It standardizes covariance by dividing it by the product of the standard deviations of both variables, resulting in a value between -1 and 1.

**Comparison:**

* **Covariance** is scale-dependent, while **correlation** is scale-free and easier to interpret.
* Correlation provides direction and strength, while covariance only indicates the direction of the relationship.

**6.5. Facebook: Say you flip a coin 10 times and observe only one heads. What would be your null hypothesis and p-value for testing whether the coin is fair or not?**  
**Answer:**

* **Null Hypothesis (H₀):** The coin is fair, meaning the probability of heads is 0.5.
* **Alternative Hypothesis (H₁):** The coin is biased, meaning the probability of heads is not 0.5.

To calculate the p-value, we use the binomial distribution:

P(X≤1)=P(X=0)+P(X=1)P(X \leq 1) = P(X = 0) + P(X = 1)P(X≤1)=P(X=0)+P(X=1)

Using a binomial calculator, if n=10,p=0.5n = 10, p = 0.5n=10,p=0.5, the probability of getting one or fewer heads is calculated. If this probability (p-value) is less than 0.05, we reject the null hypothesis, suggesting the coin may not be fair.

**6.6. Uber: Describe hypothesis testing and p-values in layman’s terms.**  
**Answer:**  
Hypothesis testing is a way to check if a claim about data is true. Imagine a restaurant claiming their delivery time is 30 minutes. We collect sample data and check if it’s significantly different from the claim.

**Steps:**

1. Assume the claim (null hypothesis) is true.
2. Collect data and compute a test statistic.
3. Find the p-value, which tells us how likely the observed data is if the claim is true.
4. If the p-value is very low (e.g., below 0.05), we reject the claim. Otherwise, we accept it for now.

**6.7. Groupon: Describe what Type I and Type II errors are, and the trade-offs between them.**  
**Answer:**

* **Type I Error (False Positive):** Rejecting a true null hypothesis (e.g., detecting a website improvement effect that doesn’t exist).
* **Type II Error (False Negative):** Failing to reject a false null hypothesis (e.g., missing a real effect of an improvement).

**Trade-offs:**  
Reducing the probability of a Type I error (using a lower significance level) increases the chance of a Type II error. Businesses must balance these depending on the situation—critical applications (like medicine) minimize Type I errors, while marketing might focus on reducing Type II errors.

**6.8. Microsoft: Explain the statistical background behind power.**  
**Answer:**  
Statistical power is the probability of correctly rejecting the null hypothesis when it is false. It measures a test’s sensitivity to detect real effects. Power depends on:

* **Sample size:** Larger samples increase power.
* **Effect size:** Bigger differences are easier to detect.
* **Significance level:** Lower alpha (e.g., 0.01) reduces power.
* **Variability in data:** Less variation leads to higher power.

A high-power test ensures fewer Type II errors and reliable conclusions in experiments.

**6.9. Facebook: What is a Z-test and when would you use it versus a T-test?**  
**Answer:**

* **Z-test:** Used when the population variance is known and the sample size is large (typically n>30n > 30n>30). It assumes normal distribution and is commonly applied in quality control and A/B testing.
* **T-test:** Used when the sample size is small and the population variance is unknown. It accounts for more uncertainty and is widely used in academic research and clinical trials.

**Example:**  
If analyzing the average height of students from a large dataset with known variance, a Z-test is appropriate. If we only have a small sample without known variance, a T-test is better.

###

**6.10 Amazon: Say you are testing hundreds of hypotheses, each with a t-test. What considerations would you take into account when doing this?**

**How to Start:**  
*"When testing multiple hypotheses, it is crucial to address the challenges associated with multiple comparisons to ensure the validity of our findings. Let me break down key considerations we should keep in mind."*

**Key Points to Cover:**

1. **Multiple Testing Problem:**
   * Each hypothesis test has a small probability (e.g., 5%) of producing a false positive.
   * With hundreds of tests, the cumulative chance of false positives increases.
   * **Solution:** Apply corrections such as the Bonferroni correction to control the family-wise error rate.
2. **False Discovery Rate (FDR):**
   * Instead of controlling Type I error across all tests, we control the proportion of false positives.
   * Methods such as the Benjamini-Hochberg procedure are commonly used.
3. **Power and Sample Size Considerations:**
   * Ensure each test has sufficient power to detect real effects without inflating Type II errors.
   * Balance between the significance level and the power of tests.
4. **Correlation Among Tests:**
   * If the hypotheses are correlated, adjusting them independently may lead to incorrect inferences.

**Conclude Strongly:**  
*"To summarize, multiple hypothesis testing requires thoughtful adjustments to error rates, adequate sample sizes, and an understanding of dependencies between tests to ensure reliable and interpretable results."*

**6.11 Google: How would you derive a confidence interval for the probability of flipping heads in a series of coin tosses?**

**How to Start:**  
*"To estimate the probability of flipping heads, we can use confidence intervals based on the binomial distribution. Let me explain the step-by-step process for deriving it."*

**Step-by-Step Approach:**

1. **Define the Problem:**
   * Let ppp be the unknown probability of getting heads, and nnn be the number of coin flips.
   * The sample proportion p^=xn\hat{p} = \frac{x}{n}p^​=nx​, where xxx is the number of heads.
2. **Normal Approximation (For Large Samples):**
   * Using the Central Limit Theorem (CLT), the sample proportion follows an approximate normal distribution.
   * The formula for the confidence interval is: p^±Zα/2p^(1−p^)n\hat{p} \pm Z\_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}p^​±Zα/2​np^​(1−p^​)​​
   * Zα/2Z\_{\alpha/2}Zα/2​ is the critical value for the chosen confidence level (e.g., 1.96 for 95%).
3. **Exact Confidence Interval (For Small Samples):**
   * For smaller sample sizes, the Clopper-Pearson method provides an exact confidence interval using the binomial distribution.

**Conclude Strongly:**  
*"The choice of method depends on the sample size. For large samples, the normal approximation works well, while for small samples, exact methods are preferred to ensure accuracy."*

**6.13 Citadel: What is the expected number of rolls needed to see all 6 sides of a fair die?**

**How to Start:**  
*"This problem is a variation of the well-known Coupon Collector’s Problem, where we collect unique outcomes from independent trials. Let's analyze how we can solve it mathematically."*

**Explanation:**

1. **Understanding the Problem:**
   * Each time we roll the die, we have a 1/6 probability of getting a specific number.
   * Once we’ve seen some sides, the probability of seeing a new number decreases.
2. **Mathematical Expectation:**
   * The expected number of rolls is calculated as: 6(1+12+13+⋯+16)≈14.76 \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{6}\right) \approx 14.76(1+21​+31​+⋯+61​)≈14.7
   * The terms represent the increasing difficulty of finding new numbers as more are collected.
3. **Practical Implications:**
   * This concept is useful in simulations, A/B testing experiments, and data exploration tasks.

**Conclude Strongly:**  
*"Thus, on average, we need approximately 14.7 rolls to observe all six sides at least once, making it a useful principle in probability-based applications."*

**6.15 D.E. Shaw: A coin was flipped 1000 times, and 550 times it showed heads. Do you think the coin is biased?**

**How to Start:**  
*"To determine whether the coin is biased, we need to conduct a hypothesis test to compare the observed proportion with the expected proportion under fairness assumptions."*

**Step-by-Step Approach:**

1. **Define Hypotheses:**
   * Null Hypothesis H0:p=0.5H\_0: p = 0.5H0​:p=0.5 (the coin is fair).
   * Alternative Hypothesis H1:p≠0.5H\_1: p \neq 0.5H1​:p=0.5 (the coin is biased).
2. **Use of Statistical Test:**
   * Compute the test statistic using: Z=p^−0.50.5(1−0.5)1000Z = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{1000}}}Z=10000.5(1−0.5)​​p^​−0.5​
   * Plugging in values: Z=0.55−0.50.5×0.51000=3.16Z = \frac{0.55 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = 3.16Z=10000.5×0.5​​0.55−0.5​=3.16
   * Compare with critical value (e.g., 1.96 for 95% confidence level).
3. **Interpretation:**
   * Since the Z-score is greater than 1.96, we reject the null hypothesis and conclude the coin is likely biased.

**Conclude Strongly:**  
*"Given our analysis, the coin appears biased with statistical significance, but further investigation into potential external factors should be considered."*

**6.18 Google: What is the expected value of the minimum of two independent uniform random variables on [0,1]?**

**How to Start:**  
*"This is a problem involving order statistics. When taking the minimum of two uniform random variables, we can derive its expected value using probability theory."*

**Solution:**

1. **Definition:**
   * If XXX and YYY are independent and uniformly distributed on [0,1], their minimum follows the cumulative distribution: P(min⁡(X,Y)≤t)=1−P(X>t)P(Y>t)=1−(1−t)2P(\min(X, Y) \leq t) = 1 - P(X > t)P(Y > t) = 1 - (1-t)^2P(min(X,Y)≤t)=1−P(X>t)P(Y>t)=1−(1−t)2
   * Differentiating to get the density function: fmin⁡(t)=2(1−t)f\_{\min}(t) = 2(1-t)fmin​(t)=2(1−t)
2. **Expected Value Calculation:**
   * Compute the expectation: E[min⁡(X,Y)]=∫01t⋅2(1−t) dt=13E[\min(X, Y)] = \int\_0^1 t \cdot 2(1-t) \, dt = \frac{1}{3}E[min(X,Y)]=∫01​t⋅2(1−t)dt=31​

**Conclude Strongly:**  
*"Thus, the expected value of the minimum of two independent uniform random variables is 1/3, which has applications in scheduling and reliability analysis."*

**6.19 Morgan Stanley: Say you have an unfair coin which lands on heads 60% of the time. How many coin flips are needed to detect that the coin is unfair?**

**How to Start:**

*"To determine the number of flips required to detect the bias of an unfair coin, we can apply hypothesis testing techniques and power analysis."*

**Solution:**

1. **Hypothesis Formulation:**
   * Null hypothesis (H0H\_0H0​): The coin is fair (p=0.5p = 0.5p=0.5).
   * Alternative hypothesis (H1H\_1H1​): The coin is biased (p=0.6p = 0.6p=0.6).
2. **Z-Test for Proportions:**
   * The test statistic is: Z=p^−p0p0(1−p0)nZ = \frac{\hat{p} - p\_0}{\sqrt{\frac{p\_0(1-p\_0)}{n}}}Z=np0​(1−p0​)​​p^​−p0​​
   * Setting significance level α\alphaα (e.g., 0.05) and power (e.g., 0.8).
3. **Sample Size Calculation:**
   * Using power analysis formula: n=(Zα+Zβ)2⋅p(1−p)δ2n = \frac{(Z\_{\alpha} + Z\_{\beta})^2 \cdot p(1-p)}{\delta^2}n=δ2(Zα​+Zβ​)2⋅p(1−p)​
   * Where δ\deltaδ is the effect size ∣0.6−0.5∣|0.6 - 0.5|∣0.6−0.5∣.

**Conclusion:**

*"With a significance level of 0.05 and power of 0.8, approximately 100–200 flips might be required to detect bias reliably."*

**6.20 Uber: Say you have nnn numbers from 1 to nnn, and you uniformly sample from this distribution with replacement kkk times. What is the expected number of distinct values you would draw?**

**How to Start:**

*"This problem is related to the coupon collector's problem, which determines the expected number of unique elements collected in repeated sampling."*

**Solution:**

1. **Expected Distinct Values Formula:**
   * The probability of drawing a new unique number in each trial is: E[distinct values]=n(1−(1−1n)k)E[\text{distinct values}] = n \left( 1 - \left(1 - \frac{1}{n}\right)^k \right)E[distinct values]=n(1−(1−n1​)k)
2. **Approximation for Large nnn:**
   * Using the harmonic series approximation: E[distinct values]≈n(1−e−k/n)E[\text{distinct values}] \approx n \left(1 - e^{-k/n}\right)E[distinct values]≈n(1−e−k/n)
3. **Special Case Insights:**
   * If k=nk = nk=n, the expected distinct values ≈n(1−e−1)≈0.63n\approx n (1 - e^{-1}) \approx 0.63n≈n(1−e−1)≈0.63n.

**Conclusion:**

*"On average, if kkk equals nnn, about 63% of the numbers will be unique in the sample."*

**6.21 Goldman Sachs: There are 100 noodles in a bowl. At each step, you randomly select two noodle ends and tie them together. What is the expectation on the number of loops formed?**

**How to Start:**

*"This problem is related to random pairing in combinatorics. Let’s analyze the expected number of loops created."*

**Solution:**

1. **Understanding the Process:**
   * Initially, there are 100 noodles (200 ends). Each pairing reduces the count by 1.
2. **Expectation Calculation:**
   * After each step, the number of loops follows a recurrence relation: E[loops]=1+n−1nE[remaining loops]E[\text{loops}] = 1 + \frac{n-1}{n}E[\text{remaining loops}]E[loops]=1+nn−1​E[remaining loops]
3. **Final Result:**
   * The expected number of loops is: ∑i=11001i≈4.6\sum\_{i=1}^{100} \frac{1}{i} \approx 4.6i=1∑100​i1​≈4.6

**Conclusion:**

*"On average, about 4.6 loops will be formed when all noodles are tied randomly."*

**6.22 Morgan Stanley: What is the expected value of the max of two independent uniform random variables?**

**How to Start:**

*"The maximum of independent uniform random variables follows an order statistics distribution. Let's derive the expected value."*

**Solution:**

1. **Distribution of the Maximum:**
   * Given X,Y∼Uniform(0,1)X, Y \sim Uniform(0,1)X,Y∼Uniform(0,1), the CDF of the maximum is: P(max⁡(X,Y)≤t)=P(X≤t)P(Y≤t)=t2P(\max(X, Y) \leq t) = P(X \leq t) P(Y \leq t) = t^2P(max(X,Y)≤t)=P(X≤t)P(Y≤t)=t2
2. **Expected Value Calculation:**
   * Differentiate to get the PDF: fmax⁡(t)=2tf\_{\max}(t) = 2tfmax​(t)=2t
   * Compute expectation: E[max⁡(X,Y)]=∫012t2dt=23E[\max(X, Y)] = \int\_0^1 2t^2 dt = \frac{2}{3}E[max(X,Y)]=∫01​2t2dt=32​

**Conclusion:**

*"The expected value of the maximum of two uniform [0,1] variables is 23\frac{2}{3}32​, a useful concept in optimization and simulations."*

**6.23 Lyft: Derive the mean and variance of the uniform distribution U(a,b)U(a, b)U(a,b).**

**How to Start:**

*"For a uniform distribution over an interval [a, b], all values have equal probability. Let's derive the mean and variance."*

**Solution:**

1. **Mean Calculation:**

E[X]=a+b2E[X] = \frac{a + b}{2}E[X]=2a+b​

1. **Variance Calculation:**

Var(X)=(b−a)212Var(X) = \frac{(b-a)^2}{12}Var(X)=12(b−a)2​

**Conclusion:**

*"Uniform distributions are simple yet powerful models for modeling equal-likelihood scenarios."*

**6.24 Citadel: How many cards would you expect to draw from a standard deck before seeing the first ace?**

**How to Start:**

*"This problem follows the geometric distribution, as we're waiting for the first occurrence of an event (drawing an ace)."*

**Solution:**

1. **Geometric Distribution Formula:**
   * Probability of success (drawing an ace): p=452p = \frac{4}{52}p=524​.
   * Expected value of the first success: E[X]=1p=524=13E[X] = \frac{1}{p} = \frac{52}{4} = 13E[X]=p1​=452​=13

**Conclusion:**

*"On average, it takes 13 draws to encounter the first ace in a shuffled deck."*

**6.25 Spotify: Say you draw nnn samples from a uniform distribution U(a,b)U(a, b)U(a,b). What are the MLE estimates of aaa and bbb?**

**How to Start:**

*"Maximum Likelihood Estimation (MLE) helps estimate parameters of distributions from observed data. Let's derive estimates for a uniform distribution."*

**Solution:**

1. **Likelihood Function:**
   * Given the PDF of a uniform distribution: f(x;a,b)=1b−a,a≤x≤bf(x; a, b) = \frac{1}{b-a}, \quad a \leq x \leq bf(x;a,b)=b−a1​,a≤x≤b
   * The likelihood function for samples X1,X2,...,XnX\_1, X\_2, ..., X\_nX1​,X2​,...,Xn​: L(a,b)=(1b−a)nL(a, b) = \left(\frac{1}{b-a}\right)^nL(a,b)=(b−a1​)n
2. **MLE Estimates:**
   * The MLE for aaa is: a^=min⁡(Xi)\hat{a} = \min(X\_i)a^=min(Xi​)
   * The MLE for bbb is: b^=max⁡(Xi)\hat{b} = \max(X\_i)b^=max(Xi​)

**Conclusion:**

*"The MLE estimates for a uniform distribution correspond to the minimum and maximum observed values in the sample."*

**6.26 Google: Assume you are drawing from an infinite set of i.i.d. random variables uniformly distributed from (0,1). You keep drawing as long as the sequence you are getting is monotonically increasing. What is the expected length of the sequence you draw?**

**How to Start:**

*"This problem is related to the concept of records in sequences of uniform random variables. Let's analyze how long the sequence is expected to continue before it stops increasing."*

**Solution:**

1. **Understanding the Problem:**
   * Since the sequence is increasing, each new element must be larger than the previous one.
   * Each new number has a probability 1k+1\frac{1}{k+1}k+11​ of being greater than the previous maximum after kkk draws.
2. **Expectation Calculation:**
   * The length of the increasing sequence follows the **harmonic series:** E[length]=∑k=1∞1kE[\text{length}] = \sum\_{k=1}^{\infty} \frac{1}{k}E[length]=k=1∑∞​k1​
   * The sum of the harmonic series asymptotically approaches: E[n]≈e≈2.718E[n] \approx e \approx 2.718E[n]≈e≈2.718

**Conclusion:**

*"On average, the expected length of the longest increasing subsequence is eee, which is approximately 2.718."*

**6.27 Facebook: There are two games involving dice that you can play. In the first game, you roll one die and receive a dollar amount equal to the product of the rolls. In the second game, you roll one die and get the dollar amount equivalent to the square of that value. Which game has the higher expected value and why?**

**How to Start:**

*"We can determine which game has the higher expected value by calculating the expected payout for both scenarios."*

**Solution:**

**Game 1: Expected Value of Product of Two Rolls**

1. The expected value of rolling a fair die (1 to 6) is: 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
2. The expected value of the product of two independent rolls: E[X⋅Y]=E[X]⋅E[Y]=3.5×3.5=12.25E[X \cdot Y] = E[X] \cdot E[Y] = 3.5 \times 3.5 = 12.25E[X⋅Y]=E[X]⋅E[Y]=3.5×3.5=12.25

**Game 2: Expected Value of Squaring a Single Roll**

1. The expected value of the squared value of a die roll: E[X2]=12+22+32+42+52+626E[X^2] = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6}E[X2]=612+22+32+42+52+62​ =1+4+9+16+25+366=916≈15.17= \frac{1 + 4 + 9 + 16 + 25 + 36}{6} = \frac{91}{6} \approx 15.17=61+4+9+16+25+36​=691​≈15.17

**Comparison:**

* Expected value of Game 1 = 12.25
* Expected value of Game 2 = 15.17

**Game 2 has a higher expected value.**

**Conclusion:**

*"The second game (square of the roll) has a higher expected value because squaring amplifies larger numbers more than multiplication does in the first game."*

**6.28 Google: What does it mean for an estimator to be unbiased? What about consistent? Provide examples.**

**How to Start:**

*"Estimators are functions of sample data used to infer population parameters. Unbiasedness and consistency are two crucial properties for evaluating estimators. Let me explain."*

**Key Points:**

1. **Unbiased Estimator:**
   * An estimator is **unbiased** if its expected value equals the true population parameter. E[θ^]=θE[\hat{\theta}] = \thetaE[θ^]=θ
   * Example: Sample mean is an unbiased estimator of the population mean.
2. **Consistent Estimator:**
   * An estimator is **consistent** if it converges to the true parameter value as the sample size increases. θ^n→θas n→∞\hat{\theta}\_n \to \theta \quad \text{as } n \to \inftyθ^n​→θas n→∞
   * Example: Maximum Likelihood Estimation (MLE) of variance.

**Examples:**

* **Unbiased but not consistent:** Sample variance with a small sample size.
* **Biased but consistent:** MLE of variance (without n−1n-1n−1 correction).

**Conclusion:**

*"Unbiasedness ensures accuracy for finite samples, while consistency ensures long-term reliability in large samples."*

**6.29 Netflix: What are MLE and MAP? What is the difference between the two?**

**How to Start:**

*"Both Maximum Likelihood Estimation (MLE) and Maximum A Posteriori Estimation (MAP) are parameter estimation methods. Let's discuss their differences."*

**Explanation:**

1. **MLE (Maximum Likelihood Estimation):**
   * Finds the parameter that maximizes the likelihood function: θ^MLE=arg⁡max⁡θP(X∣θ)\hat{\theta}\_{MLE} = \arg \max\_{\theta} P(X | \theta)θ^MLE​=argθmax​P(X∣θ)
   * Relies solely on observed data.
2. **MAP (Maximum A Posteriori Estimation):**
   * Incorporates prior knowledge through Bayes’ theorem: θ^MAP=arg⁡max⁡θP(θ∣X)=arg⁡max⁡θP(X∣θ)P(θ)\hat{\theta}\_{MAP} = \arg \max\_{\theta} P(\theta | X) = \arg \max\_{\theta} P(X | \theta) P(\theta)θ^MAP​=argθmax​P(θ∣X)=argθmax​P(X∣θ)P(θ)
   * Uses a prior distribution P(θ)P(\theta)P(θ).

**Conclusion:**

*"MLE is data-driven, while MAP incorporates prior knowledge, making it useful when prior information is available."*

**6.30 Uber: How would you generate values from a standard normal distribution using a Bernoulli generator?**

**How to Start:**

*"We can use transformation techniques such as the Box-Muller method to generate standard normal values from a Bernoulli generator. Here's how."*

**Solution:**

1. **Box-Muller Transform:**
   * Generate two independent uniform random variables U1,U2U\_1, U\_2U1​,U2​.
   * Apply the transformation: Z1=−2ln⁡U1cos⁡(2πU2)Z\_1 = \sqrt{-2 \ln U\_1} \cos(2\pi U\_2)Z1​=−2lnU1​​cos(2πU2​) Z2=−2ln⁡U1sin⁡(2πU2)Z\_2 = \sqrt{-2 \ln U\_1} \sin(2\pi U\_2)Z2​=−2lnU1​​sin(2πU2​)
   * The values Z1Z\_1Z1​ and Z2Z\_2Z2​ will be standard normal.

**Conclusion:**

*"The Box-Muller transform is an efficient way to generate normal values using uniform random variables."*

**6.31 Facebook: Derive the expectation for a geometric random variable.**

**How to Start:**

*"A geometric random variable models the number of trials until the first success in Bernoulli trials. Let's derive its expectation."*

**Solution:**

1. **Probability Mass Function (PMF):**

P(X=k)=(1−p)k−1p,k=1,2,3,...P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, ...P(X=k)=(1−p)k−1p,k=1,2,3,...

1. **Expectation Calculation:**

E(X)=∑k=1∞k(1−p)k−1pE(X) = \sum\_{k=1}^{\infty} k (1-p)^{k-1} pE(X)=k=1∑∞​k(1−p)k−1p

* + Using the sum of an infinite series, the expectation is: E(X)=1pE(X) = \frac{1}{p}E(X)=p1​

**Conclusion:**

*"The expected value of a geometric random variable is 1/p1/p1/p, useful in reliability analysis and waiting time modeling."*

**6.34 Tesla: What is the best estimator for an exponential distribution parameter?**

**How to Start:**

*"The exponential distribution models time between events. The best estimator for the rate parameter λ\lambdaλ can be derived using the Maximum Likelihood Estimation (MLE)."*

**Solution:**

1. **Likelihood Function:** L(λ)=∏i=1Nλe−λxiL(\lambda) = \prod\_{i=1}^{N} \lambda e^{-\lambda x\_i}L(λ)=i=1∏N​λe−λxi​
2. **Log-Likelihood:** ln⁡L(λ)=Nln⁡λ−λ∑xi\ln L(\lambda) = N \ln \lambda - \lambda \sum x\_ilnL(λ)=Nlnλ−λ∑xi​
3. **MLE Solution:**
   * Differentiating and solving: λ^=N∑xi\hat{\lambda} = \frac{N}{\sum x\_i}λ^=∑xi​N​

**Conclusion:**

*"The best estimator for λ\lambdaλ is the reciprocal of the sample mean, making it efficient and unbiased."*

**6.36 Google: How to calculate the blended mean and standard deviation of two datasets?**

**How to Start:**

*"When combining datasets, the overall mean and variance must be computed considering their sample sizes."*

**Solution:**

1. **Blended Mean:**

μ=n1μ1+n2μ2n1+n2\mu = \frac{n\_1 \mu\_1 + n\_2 \mu\_2}{n\_1 + n\_2}μ=n1​+n2​n1​μ1​+n2​μ2​​

1. **Blended Variance:**

σ2=n1(σ12+μ12)+n2(σ22+μ22)n1+n2−μ2\sigma^2 = \frac{n\_1(\sigma\_1^2 + \mu\_1^2) + n\_2(\sigma\_2^2 + \mu\_2^2)}{n\_1 + n\_2} - \mu^2σ2=n1​+n2​n1​(σ12​+μ12​)+n2​(σ22​+μ22​)​−μ2

**Conclusion:**

*"This method allows effective aggregation of statistical measures across datasets."*

**6.39 Lyft: How do you uniformly sample points from a circle with radius RRR?**

**How to Start:**

*"To uniformly sample points within a circle, it's essential to carefully sample both radius and angle. Here’s how we achieve uniformity."*

**Solution:**

1. **Generate random angle:**
   * Sample θ∼Uniform(0,2π)\theta \sim Uniform(0, 2\pi)θ∼Uniform(0,2π)
2. **Generate random radius:**
   * Use the square root transformation to ensure uniformity: r=RU,U∼Uniform(0,1)r = R \sqrt{U}, \quad U \sim Uniform(0,1)r=RU​,U∼Uniform(0,1)
3. **Convert to Cartesian coordinates:**

x=rcos⁡θ,y=rsin⁡θx = r \cos \theta, \quad y = r \sin \thetax=rcosθ,y=rsinθ

**Conclusion:**

*"This technique ensures uniform sampling in circular regions, applicable in simulations and spatial statistics."*

**Final Interview Tips:**

1. **Start with a concise definition to show clarity.**
2. **Explain with formulas or step-by-step reasoning.**
3. **Provide a real-world application to demonstrate practical relevance.**
4. **Conclude with a confident summary.**