Probability

CHAPTER 5

One of the most crucial skills a data scientist needs to have is the ability to think probabilistically. Although probability is a broad field and ranges from theoretical concepts such as measure theory to more practical applications involving various probability distributions, a strong foundation in the core concepts of probability is essential.

In interviews, probability's foundational concepts are heavily tested, particularly conditional probability and basic applications involving PDFs of various probability distributions. In the finance industry, interview questions on probability, including expected values and betting decisions, are especially common. More in-depth problems that build off of these foundational probability topics are common in statistics interview problems, which we cover in the next chapter. For now, we'll start with the basics of probability.

Basics

Conditional Probability

We are often interested in knowing the probability of an event A given that an event B has occurred. For example, what is the probability of a patient having a particular disease, given that the patient tested positive for the disease? This is known as the conditional probability of A given B and is often found in the following form based on **Bayes' rule:**

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Under Bayes' rule, $P(A)$ is known as the prior, $P(B|A)$ as the likelihood, and $P(A|B)$ as the posterior.

If this conditional probability is presented simply as $P(A)$—that is, if $P(A|B) = P(A)$—then A and B are independent, since knowing about B tells us nothing about the probability of A having also occurred. Similarly, it is possible for A and B to be conditionally independent given the occurrence of another event C: $P(A \cap B|C) = P(A|C)P(B|C)$.

The statement above says that, given that C has occurred, knowing that B has also occurred tells us nothing about the probability of A having occurred.

If other information is available and you are asked to calculate a probability, you should always consider using Bayes' rule. It is an incredibly common interview topic, so understanding its underlying concepts and real-life applications involving it will be extremely helpful. For example, in medical testing for rare diseases, Bayes' rule is especially important, since it is may be misleading to simply diagnose someone as having a disease—even if the test for the disease is considered "very accurate"—without knowing the test's base rate for accuracy.

Bayes' rule also plays a crucial part in machine learning, where, frequently, the goal is to identify the best conditional distribution for a variable given the data that is available. In an interview, hints will often be given that you need to consider Bayes' rule. One such strong hint is an interviewer's wording in directions to find the probability of some event having occurred "given that" another event has already occurred.

Law of Total Probability

Assume we have several disjoint events within B having occurred; we can then break down the probability of an event A having also occurred thanks to the law of total probability, which is stated as follows: $P(A) = P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n)$.

The equation above provides a handy way to think about partitioning events. If we want to model the probability of an event A happening, it can be decomposed into the weighted sum of conditional probabilities based on each possible scenario having occurred. When asked to assess a probability involving a "tree of outcomes" upon which the probability depends, be sure to remember this concept. One common example is the probability that a customer makes a purchase, conditional on which customer segment that customer falls within.

Counting

The concept of counting typically shows up in one form or another in most interviews. Some questions may directly ask about counting (e.g., "How many ways can five people sit around a lunch table?"), while others may ask a similar question, but as a probability (e.g., "What is the likelihood that I draw four cards of the same suit?").

Two forms of counting elements are generally relevant. If the order of selection of the n items being counted k at a time matters, then the method for counting possible permutations is employed:

$$n * (n-1) * \ldots * (n-k+1) = \frac{n!}{(n-k)!}$$

In contrast, if order of selection does not matter, then the technique to count possible number of combinations is relevant:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Knowing these concepts is necessary in order to assess various probabilities that involve counting procedures. Therefore, remember to determine when selection does versus does not matter.

For some real-life applications of both, consider making up passwords (where order of characters matters) versus choosing restaurants nearby on a map (where order does not matter, only the options). Lastly, both permutations and combinations are frequently encountered in combinatorial and graph theory-related questions.

Random Variables

Random variables are a core topic within probability, and interviewers generally verify that you understand the principles underlying them and have a basic ability to manipulate them. While it is not necessary to memorize all mechanics associated with them or specific use cases, knowing the concepts and their applications is highly recommended.

A random variable is a quantity with an associated probability distribution. It can be either discrete (i.e., have a countable range) or continuous (have an uncountable range). The probability distribution associated with a discrete random variable is a probability mass function (PMF), and that associated with a continuous random variable is a probability density function (PDF). Both can be represented by the following function of x: $f_X(x)$

In the discrete case, X can take on particular values with a particular probability, whereas, in the continuous case, the probability of a particular value of x is not measurable; instead, a "probability mass" per unit per length around x can be measured (imagine the small interval of x and $x + \delta$).

Probabilities of both discrete and continuous random variables must be non-negative and must sum (in the discrete case) or integrate (in the continuous case) to 1:

$$\text{Discrete: } \sum_{x \in X} f_X(x) = 1, \text{ Continuous: } \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

The cumulative distribution function (CDF) is often used in practice rather than a variable's PMF or PDF and is defined as follows in both cases: $F_X(x) = p(X \le x)$

For a discrete random variable, the CDF is given by a sum: $F_X(x) = \sum_{k \le x} p(k)$; whereas, for a continuous random variable, the CDF is given by an integral:

$$F_X(x) = \int_{-\infty}^{x} p(y)\,dy$$

Thus, the CDF, which is non-negative and monotonically increasing, can be obtained by taking the sums of PMFs for discrete random variables, and the integral of PDFs for continuous random variables.

Knowing the basics of PDFs and CDFs is very useful for deriving properties of random variables, so understanding them is important. Whenever asked about evaluating a random variable, it is essential to identify both the appropriate PDF and CDF at hand.

Joint, Marginal, and Conditional Probability Distributions

Random variables are often analyzed with respect to other random variables, giving rise to *joint PMFs* for discrete random variables and *joint PDFs* for continuous random variables. In the continuous case, for the random variables X and Y varying over a two-dimensional space, the integration of the joint PDF yields the following:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

This is useful, since it allows for the calculation of probabilities of events involving X and Y.

From a joint PDF, a marginal PDF can be derived. Here, we derive the marginal PDF for X by integrating out the Y term:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

Similarly, we can find a joint CDF where $F_{X,Y}(x, y) = P(X \le x, Y \le y)$ is equivalent to the following:

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du$$

It is also possible to condition PDFs and CDFs on other variables. For example, for random variables X and Y, which are assumed to be jointly distributed, we have the following conditional probability:

$$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) f_{X|Y}(x \mid y)\,dy$$

where X is conditioned on Y. This is an extension of Bayes' rule and works in both the discrete and continuous case, although in the former, summation replaces integration.

Generally, these topics are asked only in very technical rounds, although a basic understanding helps with respect to general derivations of properties. When asked about more than one random variable, make it a point to think in terms of joint distributions.

Probability Distributions

There are many probability distributions, and interviewers generally do not test whether you have memorized specific properties on each (although it is helpful to know the basics), but, rather, to see if you can properly apply them to specific situations. For example, a basic use case would be to assess the probability that a certain event occurs when using a particular distribution, in which case you would directly utilize the distribution's PDF. Below are some overviews of the distributions most commonly included in interviews.

Discrete Probability Distributions

The *binomial distribution* gives the probability of k number of successes in n independent trials, where each trial has probability p of success. Its PMF is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

and its mean and variance are: $\mu = np$, $\sigma^2 = np(1 - p)$.

The most common applications for a binomial distribution are coin flips (the number of heads in n flips), user signups, and any situation involving counting some number of successful events where the outcome of each event is binary.

The *Poisson distribution* gives the probability of the number of events occurring within a particular fixed interval where the known, constant rate of each event's occurrence is λ. The Poisson distribution's PMF is

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

and its mean and variance are: $\mu = \lambda$, $\sigma^2 = \lambda$.

The most common applications for a Poisson distribution are in assessing counts over a continuous interval, such as the number of visits to a website in a certain period of time or the number of defects in a square foot of fabric. Thus, instead of coin flips with probability p of a head as a use case of the binomial distribution, applications on the Poisson will involve a process X occurring at a rate λ.

Continuous Probability Distributions

The *uniform distribution* assumes a constant probability of an X falling between values on the interval a to b. Its PDF is

$$f(x) = \frac{1}{b-a}$$

and its mean and variance are:

$$\mu = \frac{a+b}{2}, \ \sigma^2 = \frac{(b-a)^2}{12}$$

The most common applications for a uniform distribution are in sampling (random number generation, for example) and hypothesis testing cases.

The *exponential distribution* gives the probability of the interval length between events of a Poisson process having a set rate parameter of λ. Its PDF is $f(x) = \lambda e^{-\lambda x}$ and its mean and variance are:

$$\mu = \frac{1}{\lambda}, \ \sigma^2 = \frac{1}{\lambda^2}$$

The most common applications for an exponential distribution are in wait times, such as the time until a customer makes a purchase or the time until a default in credit occurs. One of the distribution's most useful properties, and one that makes for natural questions, is the property of memorylessness the distribution.

The *normal distribution* distributes probability according to the well-known bell curve over a range of X's. Given a particular mean and variance, its PDF is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \left(\frac{(x-\mu)^2}{2\sigma^2} \right)$$

and its mean and variance are given by: $\mu = \mu$, $\sigma^2 = \sigma^2$

Many applications involve the normal distribution, largely due to (a) its natural fit to many real-life occurrences, and (b) the Central Limit Theorem (CLT). Therefore, it is very important to remember the normal distribution's PDF.

Markov Chains

A Markov chain is a process in which there is a finite set of states, and the probability of being in a particular state is only dependent on the previous state. Stated another way, the Markov property is such that, given the current state, the past and future states it will occupy are conditionally independent.

The probability of transitioning from state i to state j at any given time is given by a transition matrix, denoted by P:

$$\begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \cdots & & \cdots \\ p_{m1} & \cdots & p_{mn} \end{pmatrix}$$

Various characterizations are used to describe states. A *recurrent state* is one whereby, if entering that state, one will always transition back into that state eventually. In contrast, a transient state is one in which, if entered, there is a positive probability that upon leaving, one will never enter that state again.

A stationary distribution for a Markov chain satisfies the following characteristic: $\pi = \pi P$, where P is a transition matrix, and remains fixed following any transitions using P. Thus, P contains the long-run proportions of the time that a process will spend in any particular state over time.

Usual questions asked on this topic involve setting up various problems as Markov chains and answering basic properties concerning Markov chain behavior. For example, you might be asked to model the states of users (new, active, or churned) for a product using a transition matrix and then be asked questions about the chain's long-term behavior. It is generally a good idea to think of Markov chains when multiple states are to be modeled (with transitions between them) or when questioned concerning the long-term behavior of some system.

Probability Interview Questions

Easy

5.1. Google: Two teams play a series of games (best of 7 — whoever wins 4 games first) in which each team has a 50% chance of winning any given round (no draws allowed). What is the probability that the series goes to 7 games?

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

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

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?

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?

5.6. Amazon: One in a thousand people have a particular disease, and the test for the disease is 98% correct in testing for the disease. On the other hand, the test has a 1% error rate if the person being tested does not have the disease. If someone tests positive, what are the odds they have the disease?

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

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

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 rainy, assuming that the likelihood of rain on any given day is 0.25?

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

5.11. Morgan Stanley: You and your friend are playing a game. The two of you will continue to toss a coin until the sequence HH or TH shows up. If HH shows up first, you win. If TH shows up first, your friend wins. What is the probability of you winning?

5.12. JP Morgan: Say you are playing a game where you roll a 6-sided die up to two times and can choose to stop following the first roll if you wish. You will receive a dollar amount equal to the final amount rolled. How much are you willing to pay to play this game?

5.13. Facebook: Facebook has a content team that labels pieces of content on the platform as either spam or not spam. 90% of them are diligent raters and will mark 20% of the content as spam and 80% as non-spam. The remaining 10% are not diligent raters and will mark 0% of the content as spam and 100% as non-spam. Assume the pieces of content are labeled independently of one another, for every rater. Given that a rater has labeled four pieces of content as good, what is the probability that this rater is a diligent rater?

5.14. D.E. Shaw: A couple has two children. You discover that one of their children is a boy. What is the probability that the second child is also a boy?

5.15. JP Morgan: A desk has eight drawers. There is a probability of 1/2 that someone placed a letter in one of the desk's eight drawers and a probability of 1/2 that this person did not place a letter in any of the desk's eight drawers. You open the first 7 drawers and find that they are all empty. What is the probability that the 8th drawer has a letter in it?

5.16. Optiver: Two players are playing in a tennis match, and are at deuce (that is, they will play back and forth until one person has scored two more points than the other). The first player has a 60% chance of winning every point, and the second player has a 40% chance of winning every point. What is the probability that the first player wins the match?

5.17. Facebook: Say you have a deck of 50 cards made up of cards in 5 different colors, with 10 cards of each color, numbered 1 through 10. What is the probability that two cards you pick at random do not have the same color and are also not the same number?

5.18. SIG: Suppose you have ten fair dice. If you randomly throw these dice simultaneously, what is the probability that the sum of all the top faces is divisible by 6?

Medium

5.19. Morgan Stanley: A and B play the following game: a number k from 1-6 is chosen, and A and B will toss a die until the first person throws a die showing side k, after which that person is awarded $100 and the game is over. How much is A willing to pay to play first in this game?

5.20. Airbnb: You are given an unfair coin having an unknown bias towards heads or tails. How can you generate fair odds using this coin?

5.21. SIG: Suppose you are given a white cube that is broken into $3 \times 3 \times 3 = 27$ pieces. However, before the cube was broken, all 6 of its faces were painted green. You randomly pick a small cube and see that 5 faces are white. What is the probability that the bottom face is also white?

5.22. Goldman Sachs: Assume you take a stick of length 1 and you break it uniformly at random into three parts. What is the probability that the three pieces can be used to form a triangle?

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

5.24. Uber: A fair coin is tossed twice, and you are asked to decide whether it is more likely that two heads showed up given that either (a) at least one toss was heads, or (b) the second toss was a head. Does your answer change if you are told that the coin is unfair?

5.25. Facebook: Three ants are sitting at the corners of an equilateral triangle. Each ant randomly picks a direction and begins moving along an edge of the triangle. What is the probability that none of the ants meet? What would your answer be if there are, instead, k ants sitting on all k corners of an equilateral polygon?

5.26. Robinhood: A biased coin, with probability p of landing on heads, is tossed n times. Write a recurrence relation for the probability that the total number of heads after n tosses is even.

5.27. Citadel: Alice and Bob are playing a game together. They play a series of rounds until one of them wins two more rounds than the other. Alice wins a round with probability p. What is the probability that Bob wins the overall series?

5.28. Google: Say you have three draws of a uniformly distributed random variable between (0, 2). What is the probability that the median of the three is greater than 1.5?

Hard

5.29. D.E. Shaw: Say you have 150 friends, and 3 of them have phone numbers that have the last four digits with some permutation of the digits 0, 1, 4, and 9. What's the probability of this occuring?

5.30. Spotify: A fair die is rolled n times. What is the probability that the largest number rolled is r, for each r in 1,...,6?

5.31. Goldman Sachs: Say you have a jar initially containing a single amoeba in it. Once every minute, the amoeba has a 1 in 4 chance of doing one of four things: (1) dying out, (2) doing nothing, (3) splitting into two amoebas, or (4) splitting into three amoebas. What is the probability that the jar will eventually contain no living amoeba?

5.32. Lyft: A fair coin is tossed n times. Given that there were k heads in the n tosses, what is the probability that the first toss was heads?

5.33. Quora: You have N i.i.d. draws of numbers following a normal distribution with parameters μ and σ. What is the probability that k of those draws are larger than some value Y?

5.34. Akuna Capital: You pick three random points on a unit circle and form a triangle from them. What is the probability that the triangle includes the center of the unit circle?

5.35. Citadel: You have r red balls and w white balls in a bag. You continue to draw balls from the bag until the bag only contains balls of one color. What is the probability that you run out of white balls first?

Probability Interview Solutions

Solution #5.1

For the series to go to 7 games, each team must have won exactly three times for the first 6 games, an occurrence having probability

$$\frac{\binom{6}{3}}{2^6} = \frac{20}{64} = \frac{5}{16}$$