CH. 1, THE BEAR IN THE MOONLIGHT: ANSWERS

1. **Which approach to probability—classical, empirical, or subjective—rings truest to your intuitions about what probability is?**

   This is up to you, obviously! I find "subjective" is the least common answer.

   If you think of probability as a limited game, played on chalkboards and in classrooms, with little relevance to daily life (except perhaps casinos)… then you're probably picturing classical probability.

   If you think of probability as a science of real-world predictions, and the examples that come to mind are survival rates, actuarial tables, and telephone polls… then you're probably picturing empirical probability.

2. **Most probability teachers (myself included, honestly) prefer to sweep "subjective probability" under the rug. Why do you think that is?**

   I can't speak for everyone, but I can give you my reasons. First: There aren't many good questions to ask—and certainly no straightforward questions—so it's tough to assess whether students understand the idea. Second: It doesn't "feel" like math; it feels like wishy-washy philosophy. Third: Far from actually being wishy-washy, subjective probability opens up some tricky philosophical questions about the nature of cognition and subjective experience, questions that, as a math teacher, I feel ill-equipped to handle.

3. **Why, then, would the teacher in our fable want to introduce the student to probability via the subjective approach?**

   Here's another open-ended question with no right answer. (You see why teachers like to steer clear of subjective probability!) This is my take.

   All of probability—classical, empirical, and subjective—can be interpreted as an expression of uncertainty. In every probability problem, there are things we know, and things we don't know, and the goal is often to quantify "how likely" the things we don't know are.

The only thing that changes, from one approach to the next, is how *we know what we know. Is it from real-world experience? That's empirical. Is it from built-in assumptions? That's classical. Or is it from someplace else? That's subjective.*

In this sense, all *probability is subjective, which makes it a good starting point.*

4. **After the student hears the teacher say "40%," should she adjust her own guess to something higher than 20%—like, say, 25% or 30%?**

   If I were her, I probably would. Knowing the teacher's guess gives her new information about the situation, so her probability estimate should probably change. I wouldn't go all the way up to 40%, but I might say 25%.

5. **So the student says "20%," then the teacher says "40%," and then let's suppose the student says, "I change my answer to 25%." Should the *teacher* now change *her* answer? How long could this process go on? How long *should* it go on?**

   Now we get into some tricky territory. As I said, when the student hears the teacher's estimate, she gains new information, so she updates her probability.

   But then, we the teacher hears the student's update, doesn't she *gain new information? For example, if the student only changes slightly (from, say, 20% to 21%), then the teacher knows that the student had a high level of confidence in her original answer, and isn't putting much stock in the teacher's answer. Learning about that confidence is new information! It might inspire the teacher to put more stock in the student's answer, too, and perhaps adjust her own answer down to, say, 35%.*

   This could go on and on, but it probably shouldn't. After a few iterations of this, it's hard to imagine that the teacher and student would be gleaning much helpful data from hearing each other's estimates.

6. **Suppose the student says the probability it's a bear is 20%, and the probability it's a child is 90%. What's the problem with this? How does this illustrate a danger of subjective probability?**

   The problem is that 20% + 90% = 110%, and probabilities should sum to 100%. The student's answers are not internally consistent.

This is perhaps the biggest danger of subjective probability. When we allow our minds to roam freely, making probabilistic predictions here and there, they don't tend to follow consistent mathematical rules. You can wind up mired in a self-contradictory swamp, which does no one any good.

CH. 2, THE BLINDFOLD AND THE CHESTNUTS: ANSWERS

1. **Was the student wrong to reach for the third bowl? What would you have done?**

   Personally? I'd have gone for the third bowl, for similar reasons to the student's.

   But if I had a specific reason to believe the teacher might trick me, I might go for the first bowl, since it's counterintuitive (and perhaps less likely to be booby-trapped, the way the third bowl was in the story).

   Or, if the teacher knew me well, she might anticipate my going for the counterintuitive first bowl, and booby-trap that one instead! In that case, I might go for the second bowl, as a risk-minimizing approach. Unless she anticipated that *decision, in which case… you get the idea.*

2. **More generally: What *should* a sensible person do in the student's place, and what does this tell us about how we should behave under uncertainty?**

   There's no right answer. We don't have enough information to know which bowl is best. But still, we have to pick. Gaps in our knowledge can't stop us from acting, nor should the need for action make us forget the gaps in our knowledge.

3. **Give a way of dividing the 120 good chestnuts so that…**
   a. **The third bowl would be the best choice**

      Many possibilities. Here's one: put all *the good nuts in the third bowl.*

   b. **The second bowl would be the best choice**

      A simple solution: put all *the good nuts in the second bowl.*

c. **All three bowls would be equally good choices**

We need to divide up the good nuts in the same ratio as the wasabi "nuts."

That's a 6:3:1 ratio, so we first divide the chestnuts into ten "parts." Since there are 120 in total, each part will have 12 nuts. Then we put 6 parts in the first bowl, 3 in the second, and 1 in the third.

The result: 72 nuts in bowl one, 36 nuts in bowl two, and 12 in bowl three.

4. **Hidden assumptions are always dangerous in math. But why are they *especially* dangerous in solving probability problems?**

In short: because probability problems tend to take that dreaded format known as the "word problem."

Many probability problems involve some physical system (a deck of cards, a bag of marbles, a room of people…) and some random process (drawing a card, picking a marble, selecting people for a committee…). Such systems and processes are complex! Are the cards replaced after being selected? Can you catch a glimpse in the bag? What exactly is the composition of our committee?

In much of math, we deal with abstractions. The equation "$y = x^2$" is not physical; it exists only in our imagination, and has only the properties we imagine for it. But marbles and committees are real-world objects. Their features vary—some committees have leaders, for example, while others do not. It's important to lock down these features at the start, or you may solve a different problem than the one you're going for.

Ch. 3, The Riddle of the Odorless Incense: Answers

1. **Suppose the student picks a bag, then picks a random stick from inside and smells it alone. It's *Forest*. What's the probability that both sticks in the bag are *Forest*?**

The stick in her hand is Forest. What about the stick in the bag? There's a 50% chance that it's also Forest, and thus, a 50% chance that the bag is double-Forest.

2. How is the scenario in #1 different from the scenario in the story? Does this matter?

In the story, we smell both sticks together, as a pair. In question #1, we smell one stick alone (and learn nothing about the other stick). This yields different information, and since probability is all about information, we get different answers for the two scenarios. In the story, there's a 1-in-3 probability they're both Forest, whereas in question #1, it's 1-in-2.

3. *A case study in assumptions*: The student picks up a bag of two incense sticks. "Does this have any *Forest* in it?" she asks the vendor. He says yes. What is the probability that both sticks are *Forest*?

Trick question! Sorry to do that to you. I hate trick questions, too, though I never let a little hypocrisy stop me in life.

The problem is that we don't know how the vendor obtained his information. Did he sniff the bag as a whole (like in the story), or pull out just one stick and sniff it (like in question #1)? The different scenarios yield different answers, and we have no way of knowing which one to analyze.

It goes to show: Hidden assumptions matter!

4. The vendor is now making sure that there are precisely two *Forest* sticks in every bag. But the bags now vary in size. Some have two sticks, but some have more. He just fills them, one stick at a time, until he reaches the second *Forest* stick. Then he moves on to the next bag. So by definition, the last stick he puts into each bag must be *Forest*. What's the probability that the second-to-last stick in the bag is also *Forest*?

At some point, as he fills the bag, he puts in the first stick of Forest. What happens next is crucial. There's a 50% chance he puts in another Forest right away. In that case, the second-to-last stick will be Forest.

There's also a 50% chance that, after the first Forest, he puts in a Tea Garden. After that, he'll keep going until he gets a Forest. In this case, the second-to-last stick will have to be Tea Garden.

So the answer is 50%.

5. **Suppose you're picking sticks of incense, one after the other, until you obtain one of the following sequences: two straight *Forest*, or a *Tea Garden* followed by a *Forest*. When you hit one of those sequences, you stop picking. By definition, the last stick you pick must be *Forest*. But what's the probability that the second-to-last stick is also *Forest*?**

   Let's look at the four (equally likely) possibilities for the first two sticks.

   If we begin F-F: We're already done. The second-to-last stick is Forest.

   If we begin F-TG: As soon as we get another Forest, we'll be done. So there's no way to get two Forests in a row. The second-to-last stick will be Tea Garden.

   If we begin TG-F: We're already done. The second-to-last stick is Tea Garden.

   If we begin TG-TG: As soon as we get our first Forest, we'll be done. The second-to-last stick will be Tea Garden.

   Looking over those possibilities, we see that the only way for the second-to-last stick to be Forest is if both of the first two sticks are Forest. This happens with probability ¼.

CH. 4, THE SWINDLER'S COIN: ANSWERS

1. **Suppose you're watching someone flip coins, and they keep getting heads. How many heads in a row would it take for you to believe that they're cheating somehow?**

   The simple, unsatisfying answer: It really depends.

   For me, the answer relies heavily on the flipper's behavior. Did they promise to show me "something cool"? Are they grinning mischievously as they flip? If so, it's probably a trick. Or, are they acting just as surprised as I am? Were they planning to stop after five flips, but kept going because they wanted to see how long it would last? If so, they're more likely to be honest.

   Still, even if they appear innocent, I'd be confident they're cheating somehow if we got to 12 heads in a row (a 1-in-4000 event).

2. **Suppose you're hearing somebody report the results of their coin flips (but you can't see for yourself). How many heads in a row would it take for you to believe they're lying?**

   Again, the circumstances—who the person is, how they're behaving, their stated reason for flipping, etc.—have a strong pull on my answer. But this time, I'd start getting suspicious earlier, around 6 flips, which (with a fair coin) is a 1-in-64 event.

3. **Suppose you're flipping coins by yourself. How many heads in a row would it take for you to believe that someone is controlling the flips somehow?**

   If I'm the one flipping, it seems pretty unlikely that someone would be controlling the flips somehow. After 8 or 9 heads, I'd start checking the coin for oddities, but if I found nothing, I'd conclude that I'd just been lucky.

   After, say, 15 heads in a row—a 1-in-30,000 event—I'd probably believe that something fishy was going on.

4. **Did you give the same answer to #1, #2, and #3? If not, what key factor(s) determine your willingness to believe in the coin's fairness?**

   To me, there isn't one key factor. Probability, remember, is just an expression of uncertainty. A good probabilist uses all the available information, which may take many forms—a glimpse of the coin as it flips, the tone of voice in the person flipping it, the expressions on the faces of other bystanders… In the real world, we're constantly updating our subjective probabilities on the basis of new information.

5. **What is Bayes' Law, and what does it have to do with this whole discussion?**

   Bayes' law says the following: P(A given B) = P(A and B)/P(B).

   It's about conditional probability (discussed further in Ch. 6). The idea here is to consider the following two events: (A) The coin-flipper is lying; and (B) The coin-flipper gets 30 heads in a row. We want to find the probability that A occurs, given that B has occurred.

   Bayes' law gives us a rigid framework for doing so. In fact, the student is applying a mental version of Bayes' law in the story.

Ch. 5, The Wise Monkey: Answers

1. **Let's say 1 in every 10,000 monkeys really *is* wise.**
   a. **If you give a ten-question quiz to 10,000 monkeys, how many would you expect to get 100% on the quiz? (Include wise ones *and* lucky ones.)**

      Of the 10,000 monkeys, 1 should be wise, and therefore get a perfect score. As for the others, about 1 in 1000 should get all ten questions by luck (just like our lucky monkey in the story). That yields 10 more perfect scores, for a total of 11.

   b. **Suppose we've got a monkey that aced the 10-question quiz. What's the probably that it's a wise one?**

      Of roughly 11 perfect scorers, only 1 is truly wise. So the probability is roughly 1 in 11. (To be precise, it's slightly larger—about 9.3%.)

   c. **How many yes-or-no questions would a monkey need to get right before you'd be confident that it's a wise monkey, and not just a lucky imposter?**

      Our answer to (b) showed that 10 questions clearly aren't enough—even if you get them all right, there's a less than 10% chance you're a wise monkey!

      What if we up it to 14 questions? Now, there's only a 1-in-16,000 chance of acing the test through luck. That's rarer than the wise monkeys themselves. So if a monkey passes the test, it's probably wise (with probability about 62%).

      In general, the longer you make the quiz, the less likely that a monkey will pass it by dumb luck, and so the more confident you can be that a perfect-scoring monkey is a wise one. For example, a monkey who aces an 18-question quiz is wise with probability 96%. And if a monkey aces a 25-question quiz, it's wise with probability 99.97%.

2. **What if, instead of asking yes-or-no questions, the student asked the monkeys multiple-choice questions? (Assume the monkeys can make enough noises to answer these.) Out of 1000 (not-wise) monkeys, how many would get 100% on…**
   a. **A five-question quiz with three answer choices per question (A, B, and C)?**

Starting with 1000 monkeys, how many monkeys will still have a perfect score after each question?

After 1: About 333.
After 2: About 111.
After 3: About 37.
After 4: About 12.
After 5: About 4.

b. A five-question quiz with four answer choices per question (A, B, C, and D)?

Starting with 1000 monkeys, how many monkeys will still have a perfect score after each question?

After 1: About 250.
After 2: About 63.
After 3: About 16.
After 4: About 4.
After 5: Probably just 1.

c. A five-question quiz with five answer choices per question (A, B, C, D, and E)?

Starting with 1000 monkeys, how many monkeys will still have a perfect score after each question?

After 1: About 200.
After 2: About 40.
After 3: About 8.
After 4: Just 1 or 2.
After 5: Probably none!

d. Would a multiple-choice quiz have made it easier for the student to locate a wise monkey? Why or why not?

Definitely! It's much harder to "guess" the right answer when there are 4 or 5 options than when there are only 2. A multiple-choice test would have exposed the lucky monkeys as imposters more quickly.

3. **Here's a classic scam: Send out a letter to 32,000 people, promising to name a stock each week, and to state whether it will rise or drop in value. Give free tips for the first five weeks, and then charge $1000 for the sixth week's tips. If you play your cards right, you could make up to $1,000,000 that sixth week (minus the cost of postage). How does this scam work?**

   Before you read on: Promise you won't try this at home!

   Here's the idea. The first week, I pick a stock—say, Apple. To half of my recipients (16,000 people), I tell them its price will rise. To the other half, I tell them the price will fall. By necessity, 16,000 people will receive an accurate prediction.

   The next week, I follow up with the 16,000 who received an accurate prediction. To half of them (8,000 people), I predict that another stock (say, Boeing) will rise. To the other half, I predict Boeing will fall. Inevitably, 8,000 people will receive an accurate prediction. From their perspective, I'm now 2 for 2.

   We repeat this. A week later, 4,000 people have seen me go 3 for 3. After another week, 2,000 people have seen me go 4 for 4. And at the end of the fifth week, 1,000 people have seen me go a remarkable 5 for 5 on my predictions.

   That's when I start asking for money. If all of those people pay me $1000 each for my next tip (a big "if"!) then I'll pull a cool $1,000,000. An even savvier scam would be to start charging a smaller amount (say, $50) after the first week, and to increase your prices the "correct" predictions mount.

4. **Why do humans often assign so much meaning to coincidences? Is this a bad impulse, a neutral one, or somehow beneficial?**

   You can be the philosopher here.

   Lots of explanations employ the logic of evolutionary psychology, a fascinating (and, in some cases, dubiously scientific) effort to account for human thought and behavior in terms of evolutionary pressures. The story might go something like this: In humans' ancestral environment, there was a high cost to missing a pattern—like the signs of a predator or a rival tribe, where missing the clues could spell death. But there was little cost to seeing patterns where none really exist. So our brains adapted to read meaning *into our surroundings and to seek out patterns, erring on the side of finding too many rather than too few.*

Whatever the ultimate cause, it's clear that humans prefer stories and explanations to imagining a world of unexplained and inexplicable coincidences.

CH. 6, THE MOUNTAIN WHERE RAIN NEVER FALLS: ANSWERS

1. **Some of our idioms for "That'll never happen!" can be reinterpreted as statements about conditional probability. Take your favorite low-probability event, and call it X. Then explain what each phrase has to do with conditional probability.**
    a. **"If X happens, then pigs fly."**

    In other words: the probability of X is so low that the probability of pigs flying is relatively large by comparison.

    Or: Given that X has occurred, the probability that pigs fly becomes very high.

    b. **"X happening? That'll be the day hell freezes over."**

    Similarly, "The probability of X is so low that, given its occurrence, the probability of hell freezing over might become high."

    Obviously, you can understand these phrases without resorting to the language of conditional probability, but where's the fun in that?

2. **We say that A and B are *dependent* if knowing that one has occurred changes the probability of the other occurring. For example, when you flip one coin, getting heads and getting tails are *dependent*. Let's see why.**
    a. **What's the probability that a coin flip comes up heads?**

    ½. Hopefully we're cool on that fact by now.

    b. **What's the probability that a coin flip comes up heads, *given that it's tails*? (Don't overthink it.)**

Well, given that it's tails, we know it's not heads, so the probability is 0. (I told you not to overthink it.)

c. **Why, then are "getting heads" and "getting tails" on a coin flip dependent?**

Knowing that we've gotten tails changes the probability of getting heads. Specifically, the probability of heads drops from ½ to 0.

3. **Meanwhile, A and B are *independent* if knowing that one has occurred *doesn't* change the probability of the other. For example, suppose we roll two dice, and we're wondering two things: (A) Was the first die a 3? and (B) Was the sum of the dice 7? These might seem like dependent events, but they're not. Let's see why.**

a. **What's the probability that the sum of two dice is 7?**

1/6. Of the 36 ways for two dice to come up (taking the order of the dice into account, which you should), 6 ways give a sum of seven (1-6, 2-5, 3-4, 4-3, 5-2, and 6-1).

b. **What's the probability that the sum of two dice is 7, *given the first die is 3*?**

If the first die is 3, then the sum will be 7 if and only if the second die is 4. That happens with probability 1/6.

c. **Why, then, are "first die is a 3" and "sum of dice is 7" independent events?**

Knowing the first die came up 3 doesn't change the probability that the sum of the dice is 7. We've gotten new information, but it doesn't actually affect our probabilistic outlook.

4. **Now, for each pair of events A and B, say whether they're dependent or independent:**

a. **I pull one card from a deck. A = "the card is a diamond." B = "the card is red."**

Dependent. Knowing the card is red lifts the probability that it's a diamond from ¼ to ½.

Or, looking at it the other way: Knowing the card is a diamond lifts the probability that it's red from ½ to 1.

b. **I pull one card from a deck. A = "the card is a diamond." B = "the card is black."**

Dependent. The probability it's black is ½. The probability it's black given that it's a diamond is 0.

c. **I pull two cards from a deck. A = "both cards are the same suit." B = "the first card is a diamond."**

Independent. Whatever suit the first card is, the probability that the second card is the same will be roughly 1 in 4 (technically, it's 12 in 51, since we've taken one card out of the deck already).

d. **I pull two cards from a deck. A = "both cards are the same suit." B = "both cards are the same number."**

Dependent. If they're the same suit—say, spades—then they CAN'T be the same number (because, for example, there's only one 7 of spades).

So knowing that they're the same suit drops the probability of being the same number from 1 in 17 down to 0.

5. **Some people define "independent events" as "unrelated events." It's not a terrible shortcut, but it's a little sloppy. Let's see why.**
   a. **Give an example of independent events that aren't quite "unrelated."**

   Take question #4c above. "Two cards being the same suit" and "the first card being a diamond" seem like related events—both have to do with the suits of the cards we're picking. But they're independent.

   b. **Give an example of "unrelated" events that might not be independent. (*Note:* This obviously hinges a little on your definition of "unrelated.")**

   "I'm sleeping" and "Arnold Schwarzenegger is sleeping" are dependent events. We both live in California, so if he's sleeping, it's probably night time, and I'm more likely to be sleeping, too. Knowing that one of us is asleep increases the probability that the other also is.

Whether you'd call these events "unrelated" depends on your definition of the word. But I can assure you that Arnold and I make no special effort to coordinate our sleep schedules.

CH. 7, PATTERNS IN THE STONEWORK: ANSWERS

1. **Why do you think it's so hard for humans to generate random numbers or data? Why does everything we do—even when we're *trying* to be random—come out patterned?**

   My answer goes something like this: Our minds are great connection-making machines—so great, in fact, that it's hard to turn off that connection-making process.

2. **Conversely: Why is it so hard for humans to accept "It's just random" as an explanation for patterns? Why do we find such non-explanations so unsatisfying?**

   Humans love stories and explanations. Our discomfort with ignorance is the engine driving science, exploration, and some of religion. As Kurt Vonnegut so beautifully put it:

   Tiger got to hunt, bird got to fly;
   Man got to sit and wonder, "Why, why, why?"

   Tiger got to sleep, bird got to land;
   Man got to tell himself he understand.

   As to why *this is the case… well, it beats me. Let me know if you've got a good answer.*

3. **Let's take the basketball example. (Choose a different sport if you prefer—it works out the same.)**
   a. **Do you believe there's such a thing as "in the zone," or "on a cold streak"? In other words, when playing a sport, are there times when you're more than just lucky or unlucky, but actually playing *better* or *worse* than average?**

      You probably said yes. Heck, I say yes, and I'm the one about to offer evidence to the contrary. The question is: Do we miss shots because we're playing poorly? Or do we feel like we're playing poorly because we miss shots?

b. **Let's suppose you answered "yes" to part (a). You think there *are* times when an athlete plays better or worse. (It seems plausible.) If that's the case, then when would you expect a basketball player to do better—right after he's missed his last two shots, or right after he's made his last two shots?**

   If there's such a thing as meaningful hot and cold streaks, then I should have better luck right after making two shots (when I'm hot) then right after missing two shots (when I'm cold).

c. **This gives us a testable prediction. If there's such a thing as "in the zone," then a basketball player should have a better chance of scoring right after he's *made* two shots than after he's *missed* two shots. Take a guess: Do you think the data confirms this prediction?**

   They've run this study. The data don't confirm the prediction. Look at a player's third shot right after two makes, and his third shot right after two misses, and you'll see no meaningful difference.

4. **An interesting case study: In the United States, the highest rates of kidney cancer (per capita in a town) occur in small, rural towns.**

   a. **Come up with some plausible explanations for this fact.**

      Maybe their medical care is worse, so the signs aren't detected early. Maybe there are pollutants in fertilizer or pesticides that contaminate their local water supply. We can imagine plenty of stories!

   b. **Here's another fact: In the United States, the *lowest* rates of kidney *also* occur in small, rural towns. Try explaining *that*.**

      Hmm… maybe it's only SOME towns that use carcinogenic pesticides, and those ones have high cancer rates? But the rest do better because they're free of the carcinogens of urban life?

      Clearly, this new fact challenges the plausibility of our stories.

   c. **Finally, here's a hint: This is nothing more than randomness. What's going on?**

      They key word here isn't "rural." It's "small."

Let’s suppose the rate of kidney cancer is 1 in 10,000 people. Now, imagine we’ve got a big city (1 million people) and a bunch of tiny towns (only 10 people each).

In the city, we expect roughly 100 kidney cancer cases in the population. Maybe a little higher, or a little lower, but the final result should adhere pretty closely to the 1-in-10,000 baseline rate.

Now look at those little towns. Most will have no cancer cases—their rate is a very healthy 0%. But a few will have a single case of kidney cancer, giving them a terrifying kidney cancer rate of 10%.

There’s nothing special about the towns themselves. It’s just that small samples yield extreme outcomes, whereas in large samples, the randomness of fortune tends to balance out.

The same observation has been made about small schools. Many of the highest-achieving schools in America are small—as are many of the lowest-achieving. Whatever the merits of small schools, it’s clear that part of the explanation for their anomalous successes and struggles is nothing more than the probabilistic consequences of their smallness itself.