**Conditional probability**
**18.600 Problem Set 3, due March 12**

Welcome to your third problem set! Conditional probability is defined by $P(A|B) = P(AB)/P(B)$ which implies

$$P(B)P(A|B) = P(AB) = P(A)P(B|A),$$

and dividing both sides by $P(B)$ gives Bayes' rule:

$$P(A|B) = P(A)\frac{P(B|A)}{P(B)},$$

which we may view as either a boring tautology or (after spending a few hours online reading about Bayesian epistemology, Bayesian statistics, etc.) the universal recipe for revising a worldview in response to new information. Bayes' rule relates $P(A)$ (our Bayesian *prior*) to $P(A|B)$ (our Bayesian *posterior* for $A$, once $B$ is given). If we embrace the idea that our brains have subjective probabilities for *everything* (existence of aliens, next year's interest rates, Sunday's football scores) we can imagine that our minds continually use Bayes' rule to update these numbers. Or least that they would if we were clever enough to process all the data coming our way.

By way of illustration, here's a fanciful example. Imagine that in a certain world, a *normal* person says $10^5$ things per year, each of which has a $10^{-5}$ chance (independently of all others) of being truly horrible. A *truly horrible* person says $10^5$ things, each of which has a $10^{-2}$ chance (independently of all others) of being truly horrible. Ten percent of the people in this world are truly horrible. Suppose we meet someone on the bus and the first thing that person says is truly horrible. Using Bayes' rule, we conclude that this is probably a truly horrible person.

Then we turn on cable news and see an unfamiliar politician saying something truly horrible. Now we're less confident. We don't know how the quote was selected. Perhaps the politician has made $10^5$ recorded statements and we are seeing the only truly horrible one. So we make the quote selection mechanism part of our sample space and do a more complex calculation.

The problem of selectively released information appears in many contexts. For example, lawyers select evidence to influence how judges and jurors calculate conditional probability *given* that evidence. If I'm trying to convince you that a number you don't know (but which I know to be 49) is prime, I could give you some selective information about the number without telling you exactly what it is (it's a positive integer, not a multiple of 2 or 3 or 5, less than 50) and if you don't consider my motives, you'll say "It's probably prime."

Note also that legal systems around the world designate various "burdens of proof" including *probable cause*, *reasonable suspicion*, *reasonable doubt*, *beyond a shadow of a doubt*, *clear and convincing evidence*, *some credible evidence*, and *reasonable to believe*. Usually, these terms lack clear meaning as numerical probabilities (does "beyond reasonable doubt" mean with probability at least .95, or at least .99, or something else?) but there is an exception: *preponderance of evidence* generally indicates that a probability is greater than fifty percent, so that something can be said to be "more likely than not." An interesting question (which I am not qualified to answer) is whether numerical probabilities should be assigned to the other terms as well.

A. FROM TEXTBOOK CHAPTER THREE:

1. Problem 43: There are 3 coins in a box. One is a two-headed coin, another is a fair coin, and the third is a biased coin that comes up heads 75 percent of the time. When one of the 3 coins is selected at random and flipped, it shows heads. What is the probability that it was the two-headed coin?

B. **The numbers in this problem are fictional. Google *asymptomatic covid shedding* or *covid test accuracy* for data on this complex topic.** Alice has no covid symptoms but may have the disease. Write $E_i$ for the event that Alice first acquired covid $i$ days ago. Let $C$ be the event that Alice either does not have covid or first acquired it more than 10 days ago. Based on recent activity, Alice assumes *a priori* that $P(E_1) = P(E_2) = \ldots = P(E_{10}) = .001$ and $P(C) = .99$. She then takes a mandated covid test. The likelihood of the test coming back positive depends on how much virus she is "shedding." We will assume that this particular test has a negligibly low false positive rate; it only detects live virus and not old infections. If $T$ is the event that the test is positive then $P(T|C) = 0$ and

$$\Big(P(T|E_1), P(T|E_2), \ldots, P(T|E_{10})\Big) = \Big(0,\ .2,\ .4,\ .6,\ .8,\ .8,\ .6,\ .4,\ .2,\ 0\Big).$$

Just before the covid test, Alice had a long lunch with her grandmother (who had no recent contact with anyone else). The event $G$ that her grandmother is infected satisfies $P(G|C) = 0$ and

$$\Big(P(G|E_1), P(G|E_2), \ldots, P(G|E_{10})\Big) = \Big(0,\ .2,\ .4,\ .6,\ .8,\ .8,\ .6,\ .4,\ .2,\ 0\Big).$$

Assume $G$ and $T$ are independent *given* $E_i$, i.e., $P(GT|E_i) = P(G|E_i)P(T|E_i)$ for $i \in \{1, 2, \ldots, 10\}$.

1. Compute the vector $\Big(P(E_1|T), P(E_2|T), \ldots, P(E_{10}|T)\Big)$.

2. Compute $P(G|T)$ to see how much a positive test makes Alice her worry about her grandmother.

3. Suppose that three days after Alice gets her positive test, her grandmother takes a covid test. Let $U$ be the event that this test comes back positive, and assume that the test accuracy is the same as for Alice, so that $P(U|TG) = .4$ and $P(U|TG^c) = 0$. Compute $P(G|TU^c)$. You'll find that $P(G|TU^c)$ is lower than $P(G|T)$ but still high, suggesting that when Alice's grandmother tests negative, the result is only a little reassuring. Write a sentence about how surprising this seems.

**Remark:** In reality, $P(G|E_i)$ and $P(T|E_i)$ need not be equal, as they were in the story above: "shedding enough to infect grandmother" and "shedding enough to test positive" are related but not equivalent. If someone with late-stage covid is only shedding dead virus, this may show up positive on some kinds of tests, but it would not trigger a new infection. There are various kinds of tests with different accuracy rates. And virus shedding sometimes lasts for more than 10 days. If Alice had symptoms, the nature of these would affect our probability estimates. Effective *contact tracing* (the use of probability plus detective work to warn potentially exposed individuals) is multifaceted and challenging.

C. Suppose that a fair coin is tossed infinitely many times, independently. Let $X_i$ denote the outcome of the $i$th coin toss (an element of $\{H, T\}$). Compute:

1. the conditional probability that the first toss is heads *given* that exactly 5 of the first 10 tosses are heads.

2. the conditional probability that the first two tosses are both heads *given* that exactly 5 of the first 10 tosses are heads. Is this number greater or smaller than 1/4?

3. the conditional probability that all of the first 5 tosses are heads *given* that exactly 5 of the first 10 tosses are heads. Is this number greater or smaller than 1/32?

4. the probability that the pattern HHTTH appears at least once in the sequence $X_1, X_2, X_3, \ldots$

5. the probability that TTTT appears in the sequence $X_1, X_2, X_3, \ldots$ before HH appears.

6. the probability that *every* finite-length pattern appears *infinitely many times* in the sequence $X_1, X_2, X_3, \ldots$

D. On Interrogation Planet, there are 730 suspects, and it is known that exactly one of them is guilty of a crime. It is also known that any time you ask a guilty person a question, that person will give a "suspicious-sounding" answer with probability .9 and a "normal-sounding" answer with probability .1. Similarly, any time you ask an innocent person a question, that person will give a suspicious-sounding answer with probability .1 and a normal-sounding answer with probability .9. (And these probabilities apply *regardless* of how the suspect has answered questions in the past; in other words, once a person's guilt or innocence is fixed, that person's answers are *independent* from one question to the next.)

Interrogators pick a suspect at random (all 730 people being equally likely) and ask that person nine questions. The first three answers sound normal but the next six answers all sound suspicious. The interrogators say "Wow, six suspicious answers in a row. Only a one in a million chance we'd see that from an innocent person. This person is obviously guilty." But you want to do some more thinking. Given the answers thus far, compute the conditional probability that the suspect is guilty. Give an exact numerical answer.

E. Suppose that the quantities $P[A|X_1], P[A|X_2], \ldots, P[A|X_k]$ are all equal. Check that $P[X_i|A]$ is proportional to $P[X_i]$. In other words, check that the ratio $P[X_i|A]/P[X_i]$ does not depend on $i$. (This requires no assumptions about whether the $X_i$ are mutually exclusive.)

**Remark:** This can be viewed as a mathematical version of Occam's razor. We view $A$ as an "observed" event and each $X_i$ as an event that might "explain" $A$. What we showed is that if each $X_i$ "explains" $A$ equally well (i.e., $P(A|X_i)$ doesn't depend on $i$) then the conditional probability of $X_i$ *given* $A$ is proportional to how likely $X_i$ was a *a priori*. For example, suppose $A$ is the event that there are certain noises in my attice, $X_1$ is the event that there are squirrels there, and $X_2$ is the event that there are noisy ghosts. I might say that $P(X_1|A) >> P(X_2|A)$ because $P(X_1) >> P(X_2)$. Note that after looking up online definitions of "Occam's razor" you might conclude that it refers to the above tautology *plus* the common sense rule of thumb that $P(X_1) > P(X_2)$ when $X_1$ is "simpler" than $X_2$ or "requires fewer assumptions."

F. On Cautious Science Planet, science is done as follows. First, a team of wise and well informed experts concocts a hypothesis. Experience suggests the hypotheses produced this way are correct ninety percent of the time, so we write $P(H) = .9$ where $H$ is the event that the hypothesis is true. Before releasing these hypotheses to the public, scientists do an additional experimental test (such as a clinical trial or a lab study). They decide in advance what constitutes a "positive" outcome to the experiment. Let $T$ be the event that the positive outcome occurs. The test is constructed so that $P(T|H) = .95$ but $P(T|H^c) = .05$. The result is only announced to the public if the test is positive. (Sometimes the test involves checking whether an empirically observed quantity is "statistically significant." The quantity $P(T|H)$ is sometimes called the *power* of the test.)

(a) Compute $P(H|T)$. This tells us what fraction of published findings we expect to be correct.

(b) On Cautious Science Planet, results have to be replicated before they are used in practice. If the first test is positive, a second test is done. Write $\tilde{T}$ for the event that the second test is positive, and assume the second test is like the first test, so that $P(\tilde{T}|HT) = .95$ but $P(\tilde{T}|H^cT) = .05$. Compute the reproducibility rate $P(\tilde{T}|T)$.

(c) Compute $P(H|T\tilde{T})$. This tells us how reliable the replicated results are. (Pretty reliable, it turns out—your answer should be close to 1.)

On Speculative Science Planet, science is done as follows. First creative experts think of a hypothesis that would be rather surprising and interesting if true. These hypotheses are correct only five percent of the time, so we write $P(H) = .05$. Then they conduct a test. This time $P(T|H) = .8$ (lower power) but again $P(T|H^c) = .05$. Using these new parameters:

(d) Compute $P(H|T)$.

(e) Compute the reproducibility rate $P(\tilde{T}|T)$. Assume the second test is like the first test, so that $P(\tilde{T}|HT) = .8$ but $P(\tilde{T}|H^cT) = .05$.

**Remark:** If you google Nosek reproducibility you can learn about one attempt to systematically reproduce 100 psychology studies, which succeeded a bit less than 40 perent of the time. Note that $P(\tilde{T}|T) \approx .4$ is (for better or worse) closer to Speculative Science Planet than Cautious Science Planet. The possibility that $P(H|T) < 1/2$ for real world science was famously discussed in a paper called *Why Most Published Research Findings Are False* by Ioannidis in 2005. A more recent mass replication attempt (involving just *Science* and *Nature*) allowed scientists to bet on whether a study would be replicated and found that to some extent scientists were good at predicting such things. See `https://www.nature.com/articles/d41586-018-06075-z`.

**Questions for thought:** What are the pros and cons of the two planets? Is it necessarily bad for $P(\tilde{T}|T)$ and $P(H|T)$ to be low in some contexts (assuming that people know this and don't put too much trust in single studies)? Do we need to do larger and more careful studies? What improvements can be made in fields like medicine, where controlled clinical data is sparse and expensive but life and death decisions have to be made nonetheless? And I do mean expensive. The cost of recruiting and pre-screening a *single* Alzheimer's patient for trial is $100,000, per this article
`https://www.nytimes.com/2018/07/23/health/alzheimers-treatments-trials.html`
These questions go well beyond the scope of this course, but we will say a bit more about the tradeoffs involved when we study the central limit theorem.

G. **Doomsday:** Many people think it is likely that intelligent alien civilizations exist *somewhere* (though perhaps so far separated from us in space in time that we will never encounter them). When a species becomes roughly as advanced and intelligent as our own, how long does it typically survive before extinction? A few thousand years? A few millions years? A few billion years? Closely related question: how many members of such a species typically get to exist before it goes extinct?

Let's consider a related problem. Suppose that one factory has produced 10 million baseball cards in 100,000 batches of 100. Each batch is numbered from 1 to 100. Another factory has produced 10 million baseball card in 10,000 batches of 1,000, each batch numbered from 1 to 1,000. A third factory produced a 10 million baseball card in 1000 batches of 10,000, with each batch numbered from one to 10,000. You chance upon a baseball card from one of these three factories, and *a priori* you think it is equally likely to come from each of the three factories. Then you notice that the number on it is 87.

(a) Given the number you have seen, what is the conditional probability that the card comes from the first factory? The second? The third?

Now consider the following as a variant of the card problem. Suppose that one universe contains $10^{50}$ intelligent beings, grouped into civilizations of size $10^{12}$ each. Another universe contains $10^{50}$ intelligent beings, grouped into civiliations of size $10^{15}$ each. A final universe contains $10^{50}$ intelligent beings, grouped into civilizations of size $10^{18}$ each. You pick a random one of these $3 \times 10^{50}$ beings and learn that before this being was born, exactly $141,452,234,521$ other beings were born in its civilization.

(b) What is the conditional probability that the being comes from the first universe?

**Remark:** The *doomsday argument* (google it) is that it is relatively likely that human civilization will disappear within thousands of years — as opposed to lasting millions of years — for the following reason: *if* advanced civilizations typically lasted for millions of years (with perhaps 10 billion beings born per century), then it would seem *coincidental* for us to find ourselves among the first few thousand. People disagree on what to make of this argument (what the Bayesian prior on civilization length should be, what to do with all the other information we have about our world, what measure to put on the set of alternative universes, etc.) Maybe the argument at least makes people think about the *possibility* of near-term human extinction, and whether preparing for apocalyptic scenarios (giant asteroids, incurable plagues, nuclear war, climate disaster, supervolcanos, resource depletion, the next ice age, etc.) might improve our chance of surviving a few thousand (or million or billion) more years.