First a few paragraphs of review from previous lectures: A *finite probability space* is a set S and a function $p : S \to [0, 1]$ such that $p(s) > 0$ $(\forall s \in S)$ and $\sum_{s\in S} p(s) = 1$. We refer to S as the *sample space*, subsets of S as *events*, and p as the *probability distribution*. The probability of an event $A \subseteq S$ is $p(A) = \sum_{a\in A} p(a)$. (And $p(\emptyset) = 0$.)

Two events are *disjoint* if their intersection is empty. In general we have $p(A \cup B) + p(A \cap B) = p(A) + p(B)$,[‡] and thus for disjoint events $p(A \cup B) = p(A) + p(B)$. (The first statement follows from the principle of *inclusion − exclusion*: $|A\cup B| = |A|+|B|-|A\cap B|$.)

The probability of the intersection of two events is also known as the *joint probability*: $p(A, B) \equiv p(A \cap B)$. Note that it is symmetric: $p(A, B) = p(B, A)$. Suppose we know that one event has happened and wish to ask about another. For two events A and B, the *conditional probability* of A given B is $p(A|B) = p(A, B)/p(B)$.

Example 1a: Suppose we flip a fair coin 3 times. Let B be the event that we have at least one H and A be the event of getting exactly 2 Hs. What is the probability of A given B? In this case, $(A \cap B) = A$, $p(A) = 3/8$, $p(B) = 7/8$, and therefore $p(A|B) = 3/7$.

Example 2: 4 bit number. E = at least two consecutive 0's. F = first bit is 0. ($E \cap F = \{0000\ 0001\ 0010\ 0011\ 0100\}$). $p(E \cap F) = 5/16$, $p(F) = 8/16$,
$p(E|F) = (5/16)/(1/2) = 5/8$.

Note that the definition of conditional probability also gives the formula: $p(A, B) = p(A|B)p(B)$. (For three events, we have $p(A \cap B \cap C) = p(A|B \cap C)p(B|C)p(C)$, with the obvious generalization to n events:
$$p(A_1 \cap A_2 \ldots \cap A_n) = p(A_1|A_2 \cap \ldots \cap A_n)p(A_2|A_3 \cap \ldots \cap A_n) \cdots p(A_{n-1}|A_n)p(A_n)\ .)$$

We can also use conditional probabilities to find the probability of an event by breaking the sample space into disjoint pieces. If $S = S_1 \cup S_2 \ldots \cup S_n$ and all pairs S_i, S_j are disjoint, then for any event A, $p(A) = \sum_i p(A|S_i)p(S_i)$.

Example 3: Suppose we flip a fair coin twice. Let S_1 be the outcomes where the first flip is H and S_2 be the outcomes where the first flip is T. What is the probability of A = getting 2 Hs? $p(A) = p(A|S_1)p(S_1) + p(A|S_2)p(S_2) = (1/2)(1/2) + (0)(1/2) = 1/4$.

Two events A and B are *independent* if $p(A, B) = p(A)p(B)$.
This immediately gives: A and B are independent iff $p(A|B) = p(A)$.
In addition, if $p(A, B) > p(A)p(B)$ then A and B are said to be *positively correlated*, and if $p(A, B) < p(A)p(B)$ then A and B are said to be *negatively correlated*.

[‡] "What is the chance of rolling a die one time and getting a 6? 1/6
Now, what is the chance of rolling a die twice and getting at least one 6? **THINK:** 1/6 + 1/6 = 2/6 = 1/3"
(From my nephew's fourth grade math text … except the two events have non-zero intersection.) Instead use any of: a) of the 36 possibilities, enumerate the 11 with at least one six = 11/36, b) $p(A \cup B) = p(A) + p(B) - p(A \cap B) = 1/6 + 1/6 - 1/36 = 11/36$, or c) probability of no sixes is $(5/6)(5/6)$, so at least one six is $1 - 25/36 = 11/36$.

Example 1b: In the example 1a of flipping 3 coins above, $p(A|B) \neq p(A)$ and therefore these two events are not independent. Let C be the event that we get at least one H and at least one T. Let D be the event that we get at most one H. We see that $p(C) = 6/8$, $p(D) = 4/8$, and $C \cap D$=1H so that $p(C, D) = 3/8$, and independence of events C, D follows from $p(C)p(D) = (6/8)(1/2) = 3/8 = p(C, D)$.]

Example 4: E= 2 boys, F= at least one boy. $p(E|F) = 1/3$ (E = BB, F= BB BG GB). Are the events independent? $p(E) = 1/4$, $p(F) = 3/4$, $p(E, F) = 1/4 \neq 3/16$, so they are positively correlated.

Example 5: now 3 children, E= at least one of each sex, F = at most one boy.
$p(E) = 6/8$, $p(F) = 4/8$, $p(E, F) = 3/8$, so they are independent: $p(E|F) = p(E) = 3/4$.

Example 6: two flips of a fair coin: A = two heads, B = first flip is heads, B'= at least one head, B''=second flip is heads, B'''=first flip or second flip is heads.

$p(A|B) = 1/2$, i.e., one of these (HH,HT) satisfying B,
or equivalently $p(A, B)/p(B) = (1/4)\big/(1/2) = 1/2$.

$p(A|B') = 1/3$, i.e., one of these three: (TH,HT,HH) satisfying B',
or equivalently $p(A, B)/p(B) = (1/4)\big/(3/4) = 1/3$.

(In each case we can calculate directly in the reduced space of event B, or we calculate $p(A, B)$ in the full space and divide by $p(B)$.)

Finally $p(A|B'') = p(A|B)$ by symmetry between flips, and $p(A|B''') = p(A|B')$ because $B''' = B'$.

Note: A minor variant (heads=girl, tails=boy) makes this equivalent to an example from the book *Innumeracy*, J.A.Paulos, p.86.*: Every family in the town has exactly two children, the probability that any given child is a girl (or boy) is the usual 50%, and a daughter, if there is one, always answers the door. You ring the doorbell of a home, and a girl comes to the door.

a) What is the probability that the family has a boy? $p(\overline{A}|B')$

b) What is the probability that the girl has a brother? $p(\overline{A}|B')$

c) You find a random girl walking around downtown. What is the probability she has a brother? $p(\overline{A}|B)$ or $p(\overline{A}|B'')$

---

* "Consider a randomly selected family of four that is known to have at least one daughter. One possible way you may come to learn this: You're in a town where every family includes a mother, father, and two children, and picking a house at random you are greeted by a girl. **You're told that in this town a daughter, if there is one, always answers the door.** In any case, given that a family has at least one daughter, what is the conditional probability that it also has a son? The perhaps surprising answer is 2/3, since there are three equally likely possibilities — older boy, younger girl; older girl, younger boy; older girl, younger girl — and in two of them the family has a son. The fourth possibility — older boy, younger boy — is ruled out by the fact that a girl answered the door. By contrast if you were simply to run into a girl on the street, the probability that her sibling is a boy would be 1/2."

Example 7: flip a coin 3 times. A= 1st flip is H, B = at least two H, C= at least two T. Then you can verify that $p(A) = p(B) = p(C) = 1/2$, but the probability 1/2 events can be correlated or uncorrelated. $p(A,B) = 3/8$ so A,B positively correlated (makes sense, since the 1st being H makes it more likely that there are at least two H). $p(A,C) = 1/8$ so A,C negatively correlated (again makes sense, since the 1st being H makes it less likely that there are at least two T). $p(B,C) = 0$, disjoint events (maximally negatively correlated, can't have both two T and two H in three rolls)

[Note that the notions of "disjoint" and "independent" events are very different. Two events A,B are disjoint if their intersection is empty, whereas they are independent if $p(A,B) = p(A)p(B)$. Two events that are disjoint necessarily have $p(A,B) = p(A \cap B) = 0$, so if their independent probabilites are non-zero they are necessarily negatively correlated ($p(A,B) < p(A)p(B)$). For example, if we flip 2 coins, and event A = exactly 1 H, and event B = exactly 2 H, these are disjoint but not independent events: they're negatively correlated since $p(A,B) = 0$ is less than $p(A)p(B) = (1/2)(1/4)$. Non-disjoint events can be positively or negatively correlated, or they can be independent. If we take event C = exactly 1T, then A and C are not disjoint (they're equal): and they're positively correlated since $p(A,C) = 1/2$ is greater than $p(A)p(C) = 1/4$. In the three coin flip of Example 1b, we saw an example of independent events C,D with $p(C)p(D) = (6/8)(1/2) = 3/8 = p(C,D)$.]

Example 8: Alice and Bob in the library. A = Alice is in the library between 6 and 10 pm. B = Bob is in the library between 6 and 10 pm. Imagine that data is collected by tracking their comings and goings by tracking the bluetooth ids on their smartphones (which they've inadvertently left in public mode). The joint probabilities are generally constrained to satisfy $p(A,B) + p(\overline{A},B) + p(A,\overline{B}) + p(\overline{A},\overline{B}) = 1$ (for the four possibilities in this sample space, both in the library, one or the other not there, or neither there). We also have $p(A) = p(A,B) + p(A,\overline{B})$ and $p(B) = p(A,B) + p(\overline{A},B)$.

Suppose they each average about two hours per night, hence $p(A) = p(B) = 1/2$. If those events are independent, then we would expect that the joint probability $p(A,B) = p(A)p(B) = 1/4$. From the data, it is also possible to determine the percentage of time they're both present. In the extreme case, the two events might be disjoint and $p(A,B) = 0$ — for some reason they never coincide. In the opposite extreme, we might have $p(A,B) = 1/2$, so they're maximally positively correlated and always coincide. For $0 \leq p(A,B) < 1/4$, they're negatively correlated, and for $1/4 < p(A,B) \leq 1/2$, they're positively correlated. If either of those two cases emerged, it might be tempting to assume they (i.e., the two events A and B) are in some causal relationship, and that one or both people are trying either to avoid or to coincide with the other. But in general when inferring structure in data, it's important to remember that "correlation $\not\Longrightarrow$ causation" necessarily. In this example, there could be some third party or effect exerting an influence on the two of them independently, resulting in the correlation, as in the possibilities suggested in class.

Bayes Theorem†

A simple formula follows from the above definitions and symmetry of the joint probability: $p(A|B)p(B) = p(A,B) = p(B,A) = p(B|A)p(A)$. The resulting relation

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \tag{Bayes}$$

is frequently called "Bayes' theorem" or "Bayes' rule", and makes the connection between inductive and deductive inference. In the case of sets A_i that are mutually disjoint, and with $\bigcup_{i=1}^{n} A_i = S$, then Bayes' rule takes the form

$$p(A_i|B) = \frac{p(B|A_i)p(A_i)}{p(B|A_1)p(A_1) + \ldots + p(B|A_n)p(A_n)} \ .$$

Example 1: Consider a casino with loaded and unloaded dice. For a loaded die, the probability of rolling a 6 is 50%: $p(6|L) = 1/2$, and $p(i|L) = 1/10$ $(i = 1, \ldots, 5)$. For a fair die the probabilities are $p(i|\overline{L}) = 1/6$ $(i = 1, \ldots, 6)$. Suppose there's a 1% probability of choosing a loaded die, $p(L) = 1/100$. If we select a die at random and roll three consecutive 6's with it, what is the posterior probability, $P(L|6,6,6)$, that it was loaded?

The probability of the die being loaded, given 3 consecutive 6's, is

$$\begin{aligned} p(L|6,6,6) &= \frac{p(6,6,6|L)p(L)}{p(6,6,6)} = \frac{p(6|L)^3 p(L)}{p(6|L)^3 p(L) + p(6|\overline{L})^3 p(\overline{L})} \\ &= \frac{(1/2)^3 \cdot (1/100)}{(1/2)^3 \cdot (1/100) + (1/6)^3 \cdot (99/100)} = \frac{3}{14} \approx .21 \ , \end{aligned}$$

so only a roughly 21% chance that it was loaded. (Note that the Bayesian "prior" in the above is $p(L) = 1/100$, giving the probability assigned prior to collecting the data from actual rolls, and note that the prior significantly affects the resulting probability inference.)

Example 2: Duchenne Muscular Dystrophy (DMD) can be regarded as a simple recessive sex-linked disease caused by a mutated X chromosome ($\widetilde{\mathrm{X}}$). An $\widetilde{\mathrm{X}}$Y male expresses the disease, whereas an $\widetilde{\mathrm{X}}$X female is a carrier but does not express the disease. Suppose neither of a woman's parents expresses the disease, but her brother does. Then the woman's mother must be a carrier, and the woman herself therefore has an *a priori* 50/50 chance of being a carrier, $p(C) = 1/2$. Suppose she gives birth to a healthy son (h.s.). What now is her probability of being a carrier?

Her probability of being a carrier, given a healthy son, is

$$p(C|\text{h.s.}) = \frac{p(\text{h.s.}|C)p(C)}{p(\text{h.s.})} = \frac{p(\text{h.s.}|C)p(C)}{p(\text{h.s.}|C)p(C) + p(\text{h.s.}|\overline{C})p(\overline{C})} = \frac{(1/2)\cdot(1/2)}{(1/2)\cdot(1/2) + 1\cdot(1/2)} = \frac{1}{3}$$

† Rev. Thomas Bayes (1763), Pierre-Simon Laplace (1812), Sir Harold Jeffreys (1939)

(where $\overline{C}$ means "not carrier"). Intuitively what is happening is that if she's not a carrier, then there are two ways she could have a healthy son, i.e., from either of her good X's, whereas if she's a carrier there's only one way. So the probability that she's a carrier is 1/3, given the knowledge that she's had exactly one healthy son.

(The other point about this example is that the woman has a hidden state, C or $\overline{C}$, determined once and for all, and she isn't making an independent coin flip each time she has a child as to whether or not she's a carrier. Prior to generating data about her son or sons, she has a "Bayesian prior" of 1/2 to be a carrier. Subsequent data permits a principled reassessment of that probability, continuously decreasing for each successive healthy son, or jumping to 1 if she has a single diseased son).

Example 3: Suppose there's a rare genetic disease that affects 1 out of a million people, $p(D) = 10^{-6}$. Suppose a screening test for this disease is 100% sensitive (i.e., is always correct if one has the disease), and 99.99% specific (i.e., has a .01% false positive rate). Is it worthwhile to be screened for this disease?

The above sensitivity and specificity imply that $p(+|D) = 1$ and $p(+|\overline{D}) = 10^{-4}$, so the probability of having the disease, given a positive test (+), is

$$p(D|+) = \frac{p(+|D)p(D)}{p(+)} = \frac{p(+|D)p(D)}{p(+|D)p(D) + p(+|\overline{D})p(\overline{D})} = \frac{1 \cdot 10^{-6}}{1 \cdot 10^{-6} + 10^{-4}(1 - 10^{-6})} \approx 10^{-2}$$

and there's little point to being screened (only once).

(We can also look at this as follows: if one million people were screened, we would expect roughly one to have the disease, but the test will give roughly 100 false positives. So a positive result would mean only roughly a 1 out of 100 chance for one of those positives to have the disease. In this case the result is biased by the small [one in a million] Bayesian prior $p(D)$.)

Example 4: Some comments on doomsday scenarios (to be added)