Note: ProbSet 1 due Thu night, 8 Feb 2018
Recall the probability space

Finite probability space

1) a set \( S \)

2) a function \( p : S \to [0, 1] \)

such that \( p(s) > 0 \) (\( \forall s \in S \))

and \( \sum_{s \in S} p(s) = 1. \)

\( S \) is the sample space, subsets of \( S \) are events, and \( p \) is the probability distribution.

Probability of 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 \( p(A \cup B) = p(A) + p(B) - p(A \cap B) \)
Birthday “paradox”

**Example:** a) What is the probability that in a group of \( n \) people, at least two have the same birthday?

(Simplifications: assume no leap years, and assume that all birthdays are equally likely.)

Again consider the complement problem, the probability that no two birthdays coincide:

Total number of possibilities with no coincidences is \( 365 \cdot 364 \cdot \ldots \cdot (366 - n) \)

(i.e., \( n \) factors each successive one with one fewer choice of day).

Total number of possibilities for \( n \) choices of birthdays is \( 365^n \),

so the probability of no coincidences is \( 365 \cdot 364 \cdot \ldots \cdot (366 - n)/365^n \).

The probability that at least two coincide is therefore \( 1 - 365 \cdot 364 \cdot \ldots \cdot (366 - n)/365^n \).

This probability is rapidly increasing as a function of \( n \) and turns out to be greater than .5 for \( n = 23 \).
Red (upper): The probability $1 - \frac{365!}{(365-n)!365^n}$ that at least two birthdays coincide within a group of $n$ people, as function of $n$.

Green (lower): The probability $1 - \left(\frac{364}{365}\right)^{n-1}$ of a birthday coinciding with yours within a group of $n$ people including you.
We can estimate the number of coincident pairs as

\[ \frac{n(n-1)}{2} \cdot \frac{1}{365} \]

since \( \frac{n(n-1)}{2} = \binom{n}{2} \) is the number of pairs, and \( 1/365 \) is the probability that any pair has a coincident birthday.

(It still works for \( n \) large, though has to be corrected for pairs contained in triples, and so on.)

(For small \( n \), this formula also gives the probability of at least one colliding pair, and starts deviating by around \( n \approx 15 \).)

Example: In a group of 141 people, we expect

\[ \frac{141 \times 140}{2} / 365 = 27 \] “pairs” of coincident birthdays.

(where a triple counts as 3 pairs, a quadruple as 6, …)
b) In a group of 23 people, what is the probability that at least one person has a birthday coincident specifically with yours?

First calculate probability that none of the 22 others (again under the above simplifications) has a birthday coincident with a given day:

\[ \left(\frac{364}{365}\right)^{22} \]

Probability that at least one coincides with that day is therefore \( 1 - \left(\frac{364}{365}\right)^{22} \approx .059 \), so a roughly 6% chance.

This probability increases more slowly as a function of the size of the group.

Note that \( 1 - \left(\frac{364}{365}\right)^n \) is well approximated by \( 1 - \exp(-n/365) \) for all \( n \).
Red (upper): The probability \( 1 - \frac{365!}{(365-n)!365^n} \) that at least two birthdays coincide within a group of \( n \) people, as function of \( n \).

Green (lower): The probability \( 1 - \left(\frac{364}{365}\right)^{n-1} \) of a birthday coinciding with yours within a group of \( n \) people including you.

\[ 1 - \frac{1}{e} \approx 0.632 \]
Red (upper): The probability \(1 - \frac{365!}{(365-n)!365^n}\) that at least two birthdays coincide within a group of \(n\) people, as function of \(n\).

Green (lower): The probability \(1 - \left(\frac{364}{365}\right)^{n-1}\) of a birthday coinciding with yours within a group of \(n\) people including you.
This class

<table>
<thead>
<tr>
<th>Date</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>01/04</td>
<td>2</td>
</tr>
<tr>
<td>01/09</td>
<td>2</td>
</tr>
<tr>
<td>02/05</td>
<td>2</td>
</tr>
<tr>
<td>02/19</td>
<td>3</td>
</tr>
<tr>
<td>03/04</td>
<td>3</td>
</tr>
<tr>
<td>04/02</td>
<td>3</td>
</tr>
<tr>
<td>04/04</td>
<td>2</td>
</tr>
<tr>
<td>04/19</td>
<td>2</td>
</tr>
<tr>
<td>05/04</td>
<td>3</td>
</tr>
<tr>
<td>05/22</td>
<td>2</td>
</tr>
<tr>
<td>06/10</td>
<td>2</td>
</tr>
<tr>
<td>07/03</td>
<td>3</td>
</tr>
<tr>
<td>07/20</td>
<td>2</td>
</tr>
<tr>
<td>08/03</td>
<td>2</td>
</tr>
<tr>
<td>08/08</td>
<td>2</td>
</tr>
<tr>
<td>08/22</td>
<td>2</td>
</tr>
<tr>
<td>08/26</td>
<td>2</td>
</tr>
<tr>
<td>09/07</td>
<td>3</td>
</tr>
<tr>
<td>09/25</td>
<td>2</td>
</tr>
<tr>
<td>10/05</td>
<td>2</td>
</tr>
<tr>
<td>10/20</td>
<td>2</td>
</tr>
<tr>
<td>10/22</td>
<td>2</td>
</tr>
<tr>
<td>10/23</td>
<td>2</td>
</tr>
<tr>
<td>10/25</td>
<td>2</td>
</tr>
<tr>
<td>11/12</td>
<td>3</td>
</tr>
<tr>
<td>12/05</td>
<td>2</td>
</tr>
</tbody>
</table>

(See PS1 #4)
3M fraudulent voters our of 130M?

Same voter registered in different states...
But how do we know they're the same?

They have same name and same birthday — "Must" be fraudulent!

First eliminate obvious errors (e.g., given default 1 Jan due to lack of info), leaves 750k.

Now for the “birthday paradox”. Example: 8,575 ballots cast under the name John Smith in 2012, 141 of those were born in 1970 so expect \( \frac{141 \times 140}{2} / 365 = 27 \) pairs with same birthdate.

Elementary statistics, not evidence of fraud.

Take into account all similarly multiply occurring names explains > 720k of remaining “fraud”

But names and birthdays are not uniformly distributed, so even more chance of correlated collisions. [e.g. Carol, Christine, Jesús more likely on 25 Dec, Josefina more likely on 19 Mar [St. Joseph's day], ...].

That plus sampling to detect recording errors eliminates all discrepancies. No evidence of voter fraud, widespread or otherwise. End of story?

Fake Math? Sharad Goel et al:
https://www.thisamericanlife.org/630/things-i-mean-to-know
http://www.slate.com/articles/news_and_politics/jurisprudence/2016/11/we_looked_at_130_million_ballots_from_the_2012_election_and_found_zero_fraud.html
https://5harad.com/papers/1p1v.pdf
Conditional Probability

Suppose we know that one event has happened and we wish to ask about another. For two events $A$ and $B$, the joint probability of $A$ and $B$ is defined as

$$p(A, B) = p(A \cap B)$$

the probability of the intersection of events $A$ and $B$ in the sample space, equivalently the probability that events $A$ and $B$ both occur.

The conditional probability of $A$ relative to $B$ is

$$p(A|B) = p(A \cap B)/p(B)$$

“the probability of $A$ given $B$”
The sample space is divided into disjoint pieces. If \( a \) is a general rule, it can be rewritten equivalently as the probability that events \( A \) and \( B \) both occur:

\[
p(A \cap B) = p(A|B) \cdot p(B)
\]

Notice that the definition of conditional probability also gives us the formula for the joint probability:

\[
p(A \cap B) = p(A|B) \cdot p(B)
\]

We can also use conditional probabilities to find the probability of an event given the occurrence of another. For example, if \( A \) and \( B \) are independent events:

\[
p(A|B) = p(A)
\]

For two events \( A \) and \( B \), the conditional probability of \( A \) given \( B \) is:

\[
p(A|B) = \frac{p(A \cap B)}{p(B)}
\]

This immediately gives:

\[
p(A|B) = \frac{p(A \cap B)}{p(B)}
\]

For three events we have:

\[
p(A|B) = \frac{p(A \cap B)}{p(B)}
\]

Suppose we flip a fair coin twice. Let \( A \) be the outcomes where the first flip is heads, \( B \) be the outcomes where the first flip is tails, and \( C \) be the outcomes where the second flip is heads.

Two events are independent if:

\[
p(A|B) = p(A)
\]

In this case, \( A \) and \( B \) are independent:

\[
p(A|B) = p(A)
\]

The conditional probability is defined as:

\[
p(A|B) = \frac{p(A \cap B)}{p(B)}
\]

“the probability of \( A \) given \( B \)”
Example: Flip a fair coin 3 times.

\[ B = \text{event that we have at least one } H \]

\[ A = \text{event of getting exactly 2 } H \text{s} \]

What is the probability of \( A \) given \( B \)?

In this case, \((A \cap B) = A\), \(p(A) = \frac{3}{8}\), \(p(B) = \frac{7}{8}\), and therefore \(p(A|B) = \frac{3}{7}\).

Equivalently \( p(A | B) = \frac{(A \cap B)}{p(B)} = \frac{3/8}{7/8} = 3/7 \)
Two events $A$ and $B$ are *independent* if $p(A \cap B) = p(A)p(B)$.

Since $p(A \cap B) = p(A|B)p(B)$:

$$A \text{ and } B \text{ are independent iff } p(A|B) = p(A).$$

If $p(A \cap B) > p(A)p(B)$ then $A$ and $B$ are said to be *positively correlated.*

(equivalently, $p(A|B) > p(A)$)

If $p(A \cap B) < p(A)p(B)$ then $A$ and $B$ are said to be *negatively correlated.*

($p(A|B) < p(A)$)

Alternate notation for joint probability: $P(A, B) = P(A \cap B)$

Note that it is symmetric: $p(A, B) = p(B, A)$. 
**Example:** flip 3 coins

Recall $B = \text{at least one } H$

$A = \text{exactly 2 } Hs$

$p(A) = 3/8$, $p(B) = 7/8$, and $p(A|B) = 3/7$

$p(A|B) \neq p(A)$, so the two events are **not** independent

(Since $p(A)p(B) = (3/8)(7/8) < 3/8 = p(A, B)$, they’re **positively** correlated.)

$B$

HHH HHT HTH HTT THH THT TTH TTT
Example: flip 3 coins (cont’d)

$C =$ at least one $H$ and at least one $T$.

$D =$ at most one $H$

$p(C) = 6/8$, $p(D) = 4/8$, and $p(C \cap D) = 3/8$.

Therefore events $C$ and $D$ are independent.

Whereas $p(B)p(D) = (7/8)(1/2) > 3/8 = p(B, D)$

so the events $B$ and $D$ are negatively correlated

(not surprising for “at least one $H$” and “at most one $H$”).
63 Information Science (41%)
43 Undeclared (28%)
4 Biological Sciences (3%)
3 Applied Economics and Mgmt (2%)
3 Economics, Information Science (2%)
3 Computer Science (2%)
3 Communication, Information Science (2%)
3 Oper Research & Engineering (2%)
2 Biology and Society (1%)
2 Interdisciplinary Study in ALS (1%)
2 Policy Analysis and Management (1%)
2 Economics (1%)
2 Development Sociology (1%)
2 Biometry & Statistics (1%)
1 Linguistics (1%)
1 Information Science Sys & Tech (1%)
1 Animal Science (1%)
1 Animal Science, Information Science (1%)
1 Information Science, Performing and Media Arts (1%)
1 Urban & Regional Studies (1%)
1 Inter Agriculture & Rural Dev (1%)
1 Astronomy, Planetary Studies, Theoretical Astrophysics, Applied Physics (1%)
1 Government (1%)
1 College Scholar, Psychology (1%)
1 Information Science, Comparative Literature (1%)
1 Chemical Engineering (1%)
1 German, Information Science (1%)
1 Computer Science, Philosophy (1%)
1 Communication (1%)
1 College Scholar (1%)
1 Psychology, Information Science (1%)
1 Fine Arts (1%)
\[
p(\text{info major} | \text{CALS}) = \frac{51}{64} \\
p(\text{info major} | \text{A&S}) = \frac{23}{52} \\
p(\text{info major} | \text{ENG}) = \frac{1}{7} \\
p(\text{info major}) = \frac{75}{155} = 0.484
\]

\[
p(\text{CALS} | \text{info major}) = ? \\
\text{(calculate directly or via Bayes’ thm)}
\]

Class data for 155 Total 31 Jan 2018

<table>
<thead>
<tr>
<th>college</th>
<th>u1</th>
<th>u2</th>
<th>u3</th>
<th>u4</th>
<th>IM</th>
<th>IM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CALS</td>
<td>3</td>
<td>34</td>
<td>20</td>
<td>7</td>
<td>51</td>
<td>13</td>
</tr>
<tr>
<td>A&amp;S</td>
<td>4</td>
<td>38</td>
<td>28</td>
<td>5</td>
<td>23</td>
<td>52</td>
</tr>
<tr>
<td>ENG</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>other</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>
**Example:** two children

\( E = 2 \) 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:** 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’re independent: \( p(E|F) = p(E) = 3/4 \)
**Example:** flip a coin 3 times

\[ \begin{align*} A &= \text{1st flip is H}, \quad B = \text{at least two H}, \quad C = \text{at least two T} \\
\text{Verify that } p(A) &= p(B) = p(C) = \frac{1}{2} \\
\text{but the probability }\frac{1}{2} \text{ events can be correlated or uncorrelated} \\
p(A, B) &= \frac{3}{8} \text{ so } A, B \text{ positively correlated} \\
&\quad \text{(makes sense, since 1st being H makes more likely there are at least two H).} \\
p(A, C') &= \frac{1}{8} \text{ so } A, C \text{ negatively correlated} \\
&\quad \text{(again makes sense, since 1st being H makes less likely there are at least two T).} \\
p(B, C') &= 0, \text{ disjoint events} \\
&\quad \text{(maximally negatively correlated, can’t have both two T and two H in three rolls)} \end{align*} \]
**Example:** 4 bit number

\[ E = \text{at least two consecutive 0's} \]

\[ F = \text{first bit is 0.} \]

\[ (E \cap F = \{0000, 0001, 0010, 0011, 0100\}) \]

\[ p(E \cap F) = \frac{5}{16}, \quad p(F) = \frac{8}{16}, \quad p(E|F) = \frac{(5/16)}{(1/2)} = \frac{5}{8} \]
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 probabilities are non-zero,

they are necessarily negatively correlated $(p(A, B) < p(A)p(B))$.
Bayes’ Rule

A simple formula follows from the above definitions and symmetry of the joint probability:

$$p(A|B)p(B) = p(A, B) = p(B, A) = p(B|A)p(A):$$

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

Called “Bayes’ theorem” or “Bayes’ rule” — connects inductive and deductive inference

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

For mutually disjoint sets $A_i$ with $\bigcup_{i=1}^{n} A_i = S$, 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 ($L$), probability of rolling a 6 is 50%:

$$p(6|L) = 1/2, \text{ and } p(i|L) = 1/10 \ (i = 1, \ldots, 5)$$

For a fair die ($\bar{L}$), the probabilities are $p(i|\bar{L}) = 1/6 \ (i = 1, \ldots, 6)$.

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

\[
p(L|6, 6, 6) = \frac{p(6, 6, 6|L)p(L)}{p(6, 6, 6)} = \frac{p(6|L)³p(L)}{p(6|L)³p(L) + p(6|\bar{L})³p(\bar{L})}
\]

\[
= \frac{(1/2)³ \cdot (1/100)}{(1/2)³ \cdot (1/100) + (1/6)³ \cdot (99/100)}
\]

\[
= \frac{1}{1 + (1/3)³ \cdot 99} = \frac{1}{1 + 11/3} = \frac{3}{14} \approx .21 ,
\]

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 expected probability before collecting the data from actual rolls, and significantly affects the inferred posterior probability.)

Example 2: Duchenne Muscular Dystrophy (DMD) can be regarded as a simple recessive sex-linked disease caused by a mutated X chromosome (\(e\)X). An \(e\)XY male expresses the disease, whereas an \(e\)XX 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 \(a\)prior\(i\) 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

\[
p(C|h.s.) = \frac{p(h.s.|C)p(C)}{p(h.s.|C)p(C) + p(h.s.|\bar{C})p(\bar{C})}
\]

\[
= \frac{1}{1 + (1/3)³ \cdot 99} = \frac{1}{1 + 11/3} = \frac{3}{14} \approx .21 ,
\]

so 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 \(\bar{C}\), 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 8

INFO 2950, 9–11 Feb 16