5. independence
Two events $E$ and $F$ are *independent* if
\[ P(EF) = P(E) \cdot P(F) \]
equivalently: \[ P(E|F) = P(E) \]
otherwise, they are called *dependent*
Roll two dice, yielding values $D_1$ and $D_2$

$E = \{ D_1 = 1 \}$

$F = \{ D_2 = 1 \}$

$P(E) = 1/6, \ P(F) = 1/6, \ P(EF) = 1/36$

$P(EF) = P(E) \cdot P(F) \Rightarrow E$ and $F$ independent

$G = \{D_1 + D_2 = 5\} = \{(1,4),(2,3),(3,2),(4,1)\}$

$P(E) = 1/6, \ P(G) = 4/36 = 1/9, \ P(EG) = 1/36$

not independent!

$E, G$ dependent events
Two events $E$ and $F$ are *independent* if
\[ P(EF) = P(E) P(F) \]
equivalently:  \[ P(E|F) = P(E) \]
otherwise, they are called *dependent*

Three events $E$, $F$, $G$ are independent if
\[ P(EF) = P(E)P(F), \quad P(EG) = P(E)P(G), \quad P(FG) = P(F)P(G) \]
and \[ P(EFG) = P(E) P(F) P(G) \]

*Example*: Let $X, Y$ be each $\{-1,1\}$ with equal prob
$E = \{X = 1\}$, $F = \{Y = 1\}$, $G = \{XY = 1\}$
\[ P(EF) = P(E)P(F), \quad P(EG) = P(E)P(G), \quad P(FG) = P(F)P(G) \]
but $P(EFG) = 1/4$ !!!  (because $P(G|EF) = 1$)
In general, events $E_1, E_2, \ldots, E_n$ are independent if for every subset $S$ of \{1,2,\ldots, n\}, we have

$$P \left( \bigcap_{i \in S} E_i \right) = \prod_{i \in S} P(E_i)$$

(Sometimes this property holds only for small subsets $S$. E.g., $E,F,G$ on the previous slide are pairwise independent, but not fully independent.)
Theorem: E, F independent $\Rightarrow$ E, $F^c$ independent

Proof: $P(\overline{EF}) = P(E) - P(EF)$

$= P(E) - P(E)P(F)$

$= P(E)(1 - P(F))$

$= P(E)P(\overline{F})$

Theorem:
E, F independent $\iff P(E|F) = P(E)$ $\iff P(F|E) = P(F)$

Proof: Note $P(EF) = P(E|F)P(F)$, regardless of in/dep.
Assume independent. Then

$P(E)P(F) = P(EF) = P(E|F)P(F) \Rightarrow P(E|F) = P(E)$ ($\div$ by $P(F)$)

Conversely, $P(E|F) = P(E) \Rightarrow P(E)P(F) = P(EF)$ ($\times$ by $P(F)$)
Biased coin comes up heads with probability $p$.

\[ P(\text{heads on } n \text{ flips}) = p^n \]
\[ P(\text{tails on } n \text{ flips}) = (1-p)^n \]
\[ P(\text{exactly } k \text{ heads in } n \text{ flips}) = \binom{n}{k} p^k (1-p)^{n-k} \]

Aside: note that the probability of some number of heads $= \sum_k \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1$ as it should, by the binomial theorem.
m strings hashed (uniformly) into a table with n buckets
  Each string hashed is an independent trial
  E = at least one string hashed to first bucket
What is P(E) ?
Solution:
  \( F_i = \) string \( i \) not hashed into first bucket (\( i=1,2,\ldots,m \))
  \( P(F_i) = 1 - 1/n = (n-1)/n \) for all \( i=1,2,\ldots,m \)
Event \( (F_1 \cap F_2 \cdots \cap F_m) = \) no strings hashed to first bucket
  \[ P(E) = 1 - P(F_1 \cap F_2 \cdots \cap F_m) \]
  \[ = 1 - P(F_1) \cdot P(F_2) \cdot \cdots \cdot P(F_m) \]
  \[ = 1 - ((n-1)/n)^m \]
m strings hashed (non-uniformly) to table w/ n buckets
Each string hashed is an independent trial, with probability p_i of getting hashed to bucket i
E = At least 1 of buckets 1 to k gets ≥ 1 string
What is P(E) ?
Solution:
F_i = at least one string hashed into i-th bucket
P(E) = P(F_1 \cup \cdots \cup F_k) = 1 - P((F_1 \cup \cdots \cup F_k)^c)
= 1 - P(F_1^c F_2^c \cdots F_k^c)
= 1 - P(no strings hashed to buckets 1 to k)
= 1 - (1 - p_1 - p_2 - \cdots - p_k)^m
Consider the following parallel network

\[ P(\text{there is functional path}) = 1 - P(\text{all routers fail}) \]

\[ = 1 - p_1 p_2 \cdots p_n \]

\( n \) routers, \( i^{th} \) has probability \( p_i \) of failing, independently.
Contrast: a series network

\[ P(\text{no routers fail}) = (1 - p_1)(1 - p_2) \cdots (1 - p_n) \]
Recall: Two events $E$ and $F$ are independent if
$$P(EF) = P(E) P(F)$$

If $E$ & $F$ are independent, does that tell us anything about
$$P(EF|G), P(E|G), P(F|G),$$
when $G$ is an arbitrary event? In particular, is
$$P(EF|G) = P(E|G) P(F|G) ~?$$

In general, no.
Roll two 6-sided dice, yielding values $D_1$ and $D_2$

- $E = \{ D_1 = 1 \}$
- $F = \{ D_2 = 6 \}$
- $G = \{ D_1 + D_2 = 7 \}$

$E$ and $F$ are independent

- $P(E|G) = 1/6$
- $P(F|G) = 1/6$, but
- $P(EF|G) = 1/6$, not $1/36$

so $E|G$ and $F|G$ are not independent!
Two events $E$ and $F$ are called *conditionally independent given* $G$, if

$$P(EF|G) = P(E|G) \cdot P(F|G)$$

Or, equivalently,

$$P(E|FG) = P(E|G)$$
Say you are in a dorm with 100 students

10 are CS majors: \( P(CS) = 0.1 \)

30 get straight A’s: \( P(A) = 0.3 \)

3 are CS majors who get straight A’s

\( P(CS,A) = 0.03 \)

\( P(CS,A) = P(CS) \times P(A) \), so CS and A independent

At faculty night, only CS majors and A students show up

So 37 students arrive

Of 37 students, 10 are CS \( \Rightarrow \)

\( P(CS \mid CS \text{ or } A) = 10/37 = 0.27 < .3 = P(A) \)

Seems CS major lowers your chance of straight A’s 😞

Weren’t they supposed to be independent?

In fact, CS and A are conditionally dependent at fac night

do CSE majors get fewer A’s?
Say you have a lawn
    It gets watered by rain or sprinklers
These two events are independent
You come outside and the grass is wet.
    You know that the sprinklers were on
Does that lower the probability that it rained?

This is a phenomenon is called “explaining away” –
    One cause of an observation makes another
cause less likely

Only CS majors and A students come to faculty night
    Knowing you came because you’re a CS major makes it
less likely you came because you get straight A’s
Randomly choose a day of the week
A = \{ \text{It is not a Monday} \}
B = \{ \text{It is a Saturday} \}
C = \{ \text{It is the weekend} \}
A and B are dependent events
P(A) = \frac{6}{7}, \quad P(B) = \frac{1}{7}, \quad P(AB) = \frac{1}{7}.
Now condition both A and B on C:
P(A|C) = 1, \quad P(B|C) = \frac{1}{2}, \quad P(AB|C) = \frac{1}{2}
P(AB|C) = P(A|C) P(B|C) \Rightarrow A|C \text{ and } B|C \text{ independent}

Dependent events can become independent by conditioning on additional information!
2 Gamblers: Alice & Bob.
A has i dollars; B has (N-i)
Flip a coin. Heads – A wins $1; Tails – B wins $1
Repeat until A or B has all N dollars
What is \( P(A \text{ wins}) \)?
Let \( E_i \) = event that A wins starting with \( i \) dollars
Approach: Condition on outcome of 1st flip; \( H = \text{heads} \)
\[
P_i = \frac{1}{2} (p_{i+1} + p_{i-1})
\]
\[
2p_i = p_{i+1} + p_{i-1}
\]
\[
p_{i+1} - p_i = p_i - p_{i-1}
\]
\[
p_{i+1} - p_i = p_i - p_{i-1}
\]
\[
p_2 - p_1 = 2p_1
\]
\[
2p_1 = 2p_1
\]
\[
p_i = i p_1
\]

aka “Drunkard’s Walk”

nice example of the utility of conditioning: future decomposed into two crisp cases instead of being a blurred superposition thereof
Theorem

Let $S$ be any sample space and $F$ be any event in $S$ with $P(F) 
eq 0$. Then “$P(-1|F)$”, conditioned probabilities given $F$, satisfy the axioms of probability:

(a) $0 \leq P(E|F) \leq 1$
(b) $P(S|F) = 1$
(c) If $E_i$ are mutually exclusive, then

$$P(\bigcup E_i|F) = \sum_i P(E_i|F)$$

Proof: See book (some algebra + some set theory) but the idea is very simple: Every event of interest is “$\cap F$”, so just as if $S$ shrinks to $F$. 

$P(\bullet|F)$ is a probability

Ross 3.5
Child is born with (A,a) gene pair (event $B_{A,a}$) 

Mother has (A,A) gene pair 

Two possible fathers: $M_1 = (a,a)$, $M_2 = (a,A)$ 

$P(M_1) = p$, $P(M_2) = 1-p$ 

What is $P(M_1 \mid B_{A,a})$? 

Solution: 

$$P(M_1 \mid B_{Aa}) = \frac{P(B_{Aa} \mid M_1)P(M_1)}{P(B_{Aa} \mid M_1)P(M_1) + P(B_{Aa} \mid M_2)P(M_2)}$$

$$= \frac{1 \cdot p}{1 \cdot p + 0.5(1 - p)} = \frac{2p}{1 + p} > \frac{2p}{1 + 1} = p$$
Events E & F are *independent* if

\[ P(EF) = P(E) \cdot P(F), \text{ or, equivalently } P(E|F) = P(E) \]

More than 2 events are indp if, for *all subsets*, joint probability = product of separate event probabilities

Independence can greatly simplify calculations

For fixed G, conditioning on G gives a probability measure, P(E|G)

But “conditioning” and “independence” are orthogonal:

- Events E & F that are (unconditionally) independent may become dependent when conditioned on G
- Events that are (unconditionally) dependent may become independent when conditioned on G