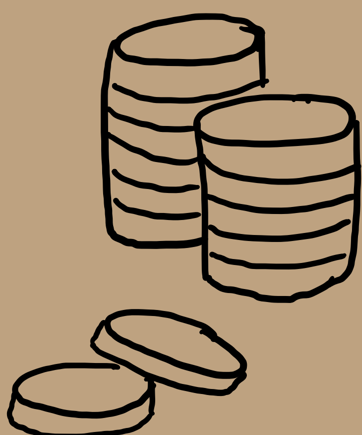
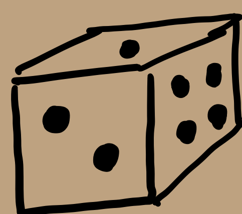
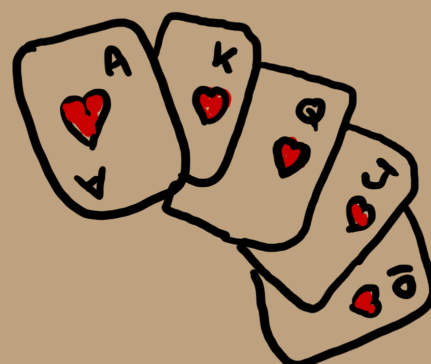
Probability







John Rachlin
CS 5002
Northeastern

(A) Introduction

1. Probability theory is the foundation for modern statistics, but interest in probability was originally motivated by gambling, and thus we'll encounter problems about coins, dice, and playing cards!

2. Historians trace probability theory to the French gambler Chevalier de Mere who would bet:
   - Roll at least one 6 in four rolls of a single die. (Winning $\approx$ 52% of the time)
   - Roll double-6 in twenty-four rolls of a pair of dice. (Losing $\approx$ 51% of the time)

3. Chevalier de Mere brought these problems to Blaise Pascal (binomial theorem, pascal's triangle) who also communicated with Pierre de Fermat, (number theory), leading to modern probability theory.

4. Today, probability and statistics are key to many areas of CS/AI: machine learning, robotics, expert systems, NLP, simulation.

(B) Definitions

1. A <u>trial</u> refers to an event whose outcome is unknown.

   - Pick a card
   - flip a coin.
   - test a drug on one patient

2. An <u>experiment</u> involves multiple trials.

   - multiple coin flips
   - clinical trials: an experiment to test a drug on many patients.

3. <u>Sample Space</u> defines all possible outcomes or alternatives

   Coin flips: $S = \{H, T\}$

   2 coin flips: $S = \{(H,H), (H,T), (T,H), (T,T)\}$

   Roll 1 die : $S = \{1, 2, 3, 4, 5, 6\}$

   Take an aspirin: $S = \{$headache goes away, headache doesn't go away$\}$

4. An <u>event</u>, E, specifies a particular outcome or set of outcomes.

   Flipped a coin and got heads: $E = \{H\}$

   Rolled a die and got odd: $E = \{1, 3, 5\}$

   Rolled two dice and got less than 6: $E = \{2, 3, 4, 5\}$ or $E = \{\text{sum} < 6\}$

   n people in room: $E = \{\text{number with same b-day} \geq 2\}$

5. The <u>probability</u> of an event, $P(E)$, is the liklihood that the event occurs in repeated trials.

Flipping a fair coin $P(E = \{H\}) = 0.5$
"50% probability"

Rolling a 6 in a fair die: $P(E = \{6\}) = \frac{1}{6}$

When we say "30% chance of rain tomorrow", we mean: If I simulate the complex dynamics of the atmosphere, with some randomness built in to reflect uncertainties, 30% of the time it rains the next day.**

6. Some basic facts:

$0 \leq P(E) \leq 1$

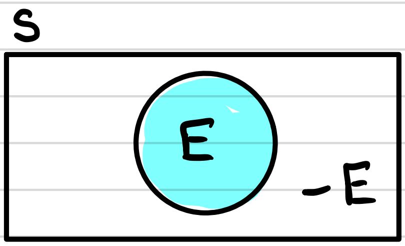
$P(S) = 1$ (all possible outcomes)

$P(E) + P(\neg E) = 1$ Event happens, or it doesnt

$P(\neg E) = 1 - P(E)$

** Due to turbulence, chaos, and sensitivity to initial conditions, local forecasts out beyond about 10-14 days are pretty useless.

Probability visualized w/ Venn diagrams



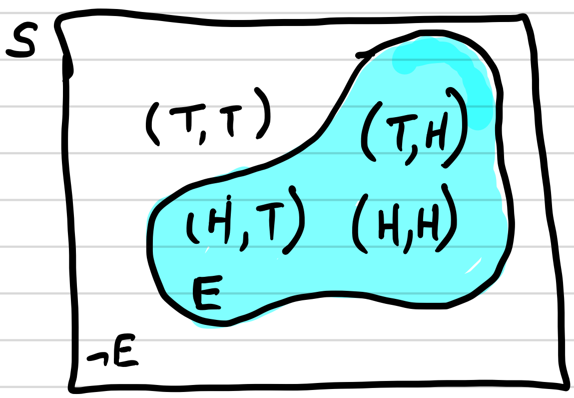
Events are a subset of outcomes drawn from the sample space.

Experiment : Flip coin twice

$S = \{(H,H), (H,T), (T,H), (T,T)\}$

$E = \{\text{At least one head}\} = \{(H,H), (H,T), (T,H)\}$

$\therefore P(E) = \frac{3}{4} = 0.75$



Note : reducing probabilities to counting problems usually makes sense only when each outcome is equally likely.

In that case,

$P(E) = |E| / |S|$

But an extreme (and silly) counter-example, if I take an aspirin for my headache and

$S = \{\text{headache cured}, \text{headache remains}\}$

$P(E = \{\text{headache cured}\}) \neq 0.50$!

Flipping 4 Coins:

$S = \{TTTT, TTTH, TTHT, TTHH, \dots, HHHH\}$

$|S| = 2^4 = 16$ possibilities.

a) $P(E = \{\text{exactly two heads}\})$

$$|E| = \binom{4}{2} = \frac{4\cdot3\cdot2\cdot1}{2\cdot1\cdot2\cdot1} = 6, \text{ so } P(E) = \frac{6}{16} = \frac{3}{8}$$

b) $P(E = \{\geq 1 \text{ head}\})$

$= 1 - P(E = \{\text{No Heads}\})$

$= 1 - P(\{TTTT\}) = 1 - 1/16 = 15/16$

c) Flipping n coins : $|S| = 2^n$

$$P(E = \{k \text{ heads}\}) = \frac{\binom{n}{k} \quad \}\ \text{\# ways to get } k \text{ heads}}{2^n \quad \}\ \text{\# of possible sequences}}$$

OPTIONAL

This should remind us of Pascal's $\triangle$:

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
\Downarrow & & \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow \\
0.03 & & 0.16 & & 0.31 & & 0.31 & & 0.16 & & 0.03
\end{array}$$

$$\begin{array}{ccccccccccc|c}
 & & & & & \binom{0}{0} & & & & & & 2^0 \\
 & & & & \binom{1}{0} & & \binom{1}{1} & & & & & 2^1 \\
 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & & 2^2 \\
 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & & 2^3 \\
 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & & 2^4 \\
\binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} & 2^5
\end{array}$$

with the last column headed Σ.

Probability of Multiple events

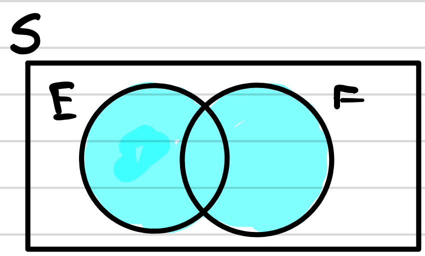
Two Events E, F.

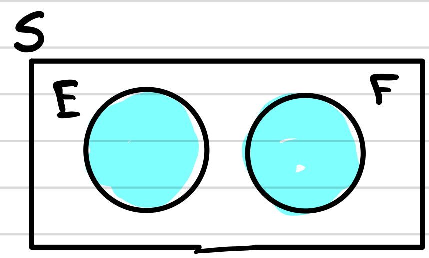
Either event can occur

$$P(E \cup F) = P(E \text{ or } F)$$
$$= P(E) + P(F) - P(E \cap F)$$
$$= P(E) + P(F) \quad \text{if } E, F \text{ are}$$

"mutually exclusive"

(sets are disjoint)





$P(E \vee F) = P(E) + P(F) - P(E \wedge F)$

$P(E \cup F) = P(E) + P(F)$

Suppose roll 1 die

$E = \{\text{Roll odd}\}$ $\quad |E| = 3$

$F = \{\text{Roll} \geq 4\}$ $\quad |F| = 3$

$$P(E \cup F) = \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6}$$

Suppose double majors aren't allowed.

$P(\{CS\}) = 0.6$

$P(\{\text{art history}\}) = 0.05$

$P(CS \vee Art) = 0.65$

Probability of Multiple events - continued

Two events E, F $P(E \cap F)$ = "Joint Probability"

All events occur. Now it depends on whether the two events are linked conditionally or independent.

<u>Independence</u>: Two events are independent if the outcome of one event has no impact on the outcome of the other.

Then, $P(E \cap F) = P(E) \cdot P(F)$

E.g. $E = \{\text{heads on flip \#1}\}$

$F = \{\text{heads on flip \#2}\}$

Then: $P(\text{two heads}) = P(E \cap F) = P(E) \cdot P(F)$

$= 0.50 \times 0.50 = 0.25$

(Coin flips are independent trials.)

P(Roll 4 - sixes on 4 rolls of a die)

$= \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{1296} \approx 0.00077$

Revisiting Chevalier de Meres Bets:

Roll 4 Dice and get at least one ⚅

$S = \{1111, 1112, 1113, \dots 6665, 6666\}$

$|S| = 6^4 = 1296$ outcomes

$E = \{\text{at least one } 6\}$

$$P(E) = 1 - P(\neg E) = 1 - P(\{\text{no } 6\})$$

$$= 1 - P(\{1111, 1112, \dots, 5554, 5555\})$$

$$= 1 - \frac{|\{\text{no } 6\}|}{|S|} = 1 - \frac{5^4}{6^4}$$

$$= 1 - \frac{625}{1296} = 0.5177$$

OR:

$$= 1 - \underbrace{\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)}_{\text{dont roll a 6 in all 4 attempts}}$$

$$= 1 - (5/6)^4 = 0.5177 \quad \text{(Winning Bet)}$$

Roll 2 Dice 24 times and get at least one: ⚅⚅

E_i = roll double six on ith roll

$$P(E) = 1 - P(\neg E) = 1 - P(\neg E_1) \cdot P(\neg E_2) \cdot P(\neg E_3) \cdots P(\neg E_{24})$$

$$P(\geq 1 \text{ ⚅⚅}) = 1 - P(0 \text{ ⚅⚅}) = 1 - \left(\frac{35}{36}\right)^{24} = 0.4914 \quad \text{(Losing Bet)}$$

Conditional Probability

We want to know the probability of one event given that another event has occurred:

$P(\text{lung cancer})$: overall liklihood of having lung cancer in the general population.

But we want to try to understand the cause or factors that contribute to Risk:

$P(\text{lung cancer} \mid \text{smoke})$ = Probability of getting lung cancer given that person also smokes.

Studies have shown:

$$P(\text{lung cancer} \mid \text{smoke}) > P(\text{lung cancer}) > P\left(\text{lung caner} \,\middle|\, \begin{matrix}\text{don't} \\ \text{smoke}\end{matrix}\right)$$

For non-independent events:

$$\boxed{P(E \cap F) = P(E) \cdot P(F \mid E)}$$

For independent events we had

$$P(E \cap F) = P(E) \cdot P(F)$$

$\therefore$ Saying that Events E and F are independent means E has no impact on F, or:

$$\boxed{P(F \mid E) = P(F)}$$

Example: Draw two spades from a 52 card deck.

E = draw spade on 1st draw
F = draw spade on 2nd draw

<u>With replacement</u>: Draws are independent

$$P(E \cap F) = P(E) \cdot P(F) = \underset{13/52}{0.25} \times \underset{13/52}{0.25} = 0.0625$$

<u>Without replacement</u>. 2nd draw affected by 1st

$$P(E \cap F) = P(E) \cdot P(F \mid E)$$

$$= \frac{13}{52} \times \frac{12}{51} \approx 0.0588$$

Again, we can think of this as a counting problem:

\# 2-card hands: $\binom{52}{2}$

\# 2-spade combinations $\binom{13}{2}$

$$P(E \wedge F) = \binom{13}{2} \Big/ \binom{52}{2} = \frac{13 \cdot 12}{52 \cdot 51} \approx 0.0588$$

Revisiting:

Independent -vs- Non Independent Events.

Independence means that the outcome of one event doesnt affect the outcome of another.

Independent	Not Independent
• Separate rolls of dice • Separate coin flips	E(roll is even), F(roll > 3) (If I know outcome is even, liklihood of outcome >3 increases $\frac{1}{2} \to \frac{2}{3}$)
• Card draw with replacement	Card draw without replacement
• Weather on Mars -vs- Weather on Venus	Weather in Boston today Weather in Boston tomorrow
• Two unrelated people's birthday	Δ stock price of two stocks (This might be hard to find).
$P(E \cap F) = P(E) \cdot P(F)$	$P(E \cap F) \neq P(E) \cdot P(F)$ $= P(E) \cdot \underbrace{P(F \mid E)}_{\text{conditional probability}}$
E and F are distinct E has no affect on F and vice versa	E informs F E influences F Possible causal connection

Are Two events Independent?

Sometimes its unclear, so we calculate:

If $P(E) \cdot P(F) = P(E \cap F)$

Then independent, otherwise not.

Roll Two Dice

E = roll doubles
F = roll at least 1 six

$P(E) = 1/6$ (6 possible doubles / 36 outcomes)

$P(F) = 1 - P(\neg F) = 1 - P(\text{no sixes}) = 1 - \left(5/6\right)^2 = \frac{36}{36} - \frac{25}{36} = \frac{11}{36}$

$P(E \cap F) = P(\text{double } 6) = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{11}{36}$ $\left(1 \neq \frac{11}{6}\right)$

$\therefore$ not independent. Knowing that we rolled doubles reduces chances of at least 1 six.

But in this example, E and F are independent:

E = roll doubles $P(E) = 1/6$
F = roll 6 on die #1 $P(F) = 1/6$

$P(E \cap F) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$ ✓ ($\therefore$ independent)

Knowing that we rolled doubles doesn't inform us as to the outcome of the 1st die.

Monopoly: Roll at least 1 double in three rolls to get out of jail for free.

What are your chances?

The rolls are <u>independent</u>.

$P(\geq 1 \text{ double in } 3 \text{ rolls})$

$= 1 - P(\text{no doubles in three rolls})$

$= 1 - P(\text{no double on one roll})^3$

$= 1 - (1 - P(\text{double on one roll}))^3$

$= 1 - (1 - 1/6)^3 = 1 - (5/6)^3 = 1 - \frac{125}{216} = \frac{91}{216} \approx 0.42$

Full House:

3 cards of 1 Rank

2 cards of a different rank

example: K♡ K◇ K♤ 5◇ 5♧

P(dealt full house) = ?

S = 5 card hands, $|S| = \binom{52}{5} = 2598960$

$|E|$ = # full houses

Pick first rank : 13 (2, 3, ..., 10, J, Q, K, A)

Pick 2nd rank : 12

Pick suits of 1st Rank: $\binom{4}{3} = 4$

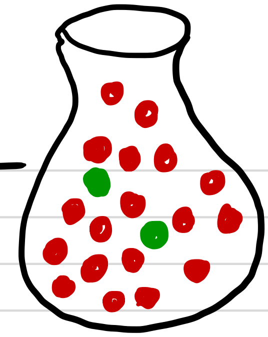
Pick suits of 2nd Rank $\binom{4}{2} = 6$

$\overline{3744} = 13 \cdot 12 \cdot 4 \cdot 6$

$\therefore P(E) = |E| / |S| \simeq 0.0014$

(0.14 %)

Colored Balls in an Urn



20 balls
18 red
2 green

With replacement: After each draw, we put the ball back (we might choose it again)

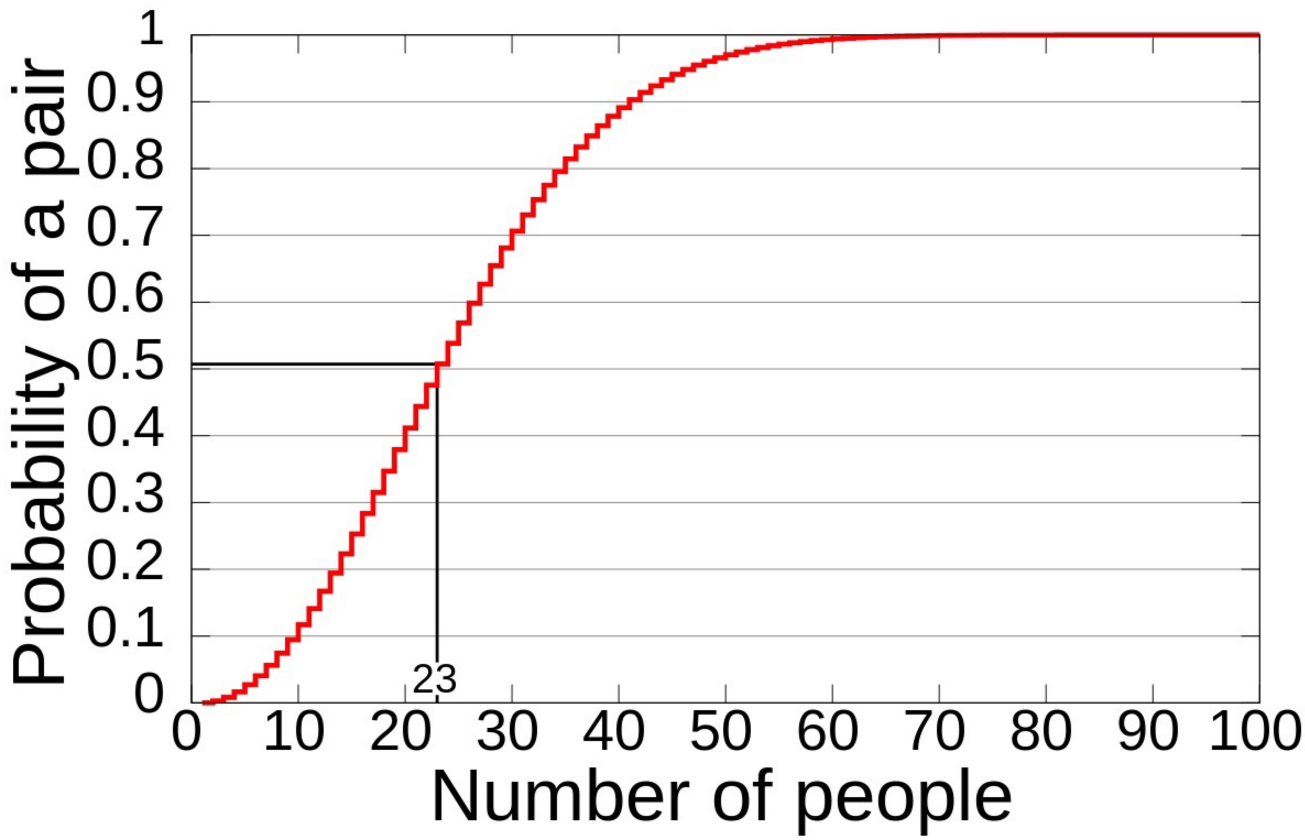
Without replacement: Once drawn, a ball is kept out of the urn, and can't be.

Pick 3 balls.

n_{Green}	With Replacement $S = 20 \times 20 \times 20$ $\lvert S\rvert = 8000$	Without Replacement $\lvert S\rvert = \binom{20}{3} = 1140$
0	$\frac{18}{20}\,\frac{18}{20}\,\frac{18}{20} = 0.729$	$\binom{18}{3}\Big/\binom{20}{3} = .716$ OR: $\frac{18}{20}\cdot\frac{17}{19}\cdot\frac{16}{18}$
1	R R G G-Location $\frac{18 \times 18 \times 2 \times 3}{8000} = 0.243$	$\binom{18}{2}\binom{2}{1}\Big/\binom{20}{3} = 0.268$
2.	R G G R-location $\frac{18 \times 2 \times 2 \times 3}{8000} = 0.027$	$\binom{18}{1}\binom{2}{2}\Big/\binom{20}{3} = \frac{18}{1140} = 0.016$
3.	$\frac{2}{20}\cdot\frac{2}{20}\cdot\frac{2}{20} = \frac{8}{8000} = 0.001$	$\emptyset$ (only two greens)

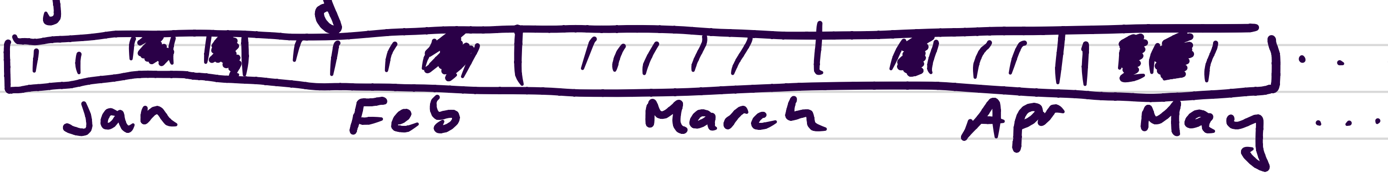
OPTIONAL

The birthday paradox: If $n > 23$ people in a room, chances are at least two people have the same birthday.



It seems counter-intuitive that it only requires 23 people. Most people are thinking: "How many people are needed to make it likly someone has the same B-Day as <u>me</u>"

Think of calendar days being filled gradually:



OPTIONAL

let P_k = probabilty of <u>no</u> common birthdays with k random people in the room.

Then:

$$\left(\frac{365}{365}\right) = 1.0$$

(only one person, so there are 365/365 choices)

$$P_2 = \left(\frac{365}{365}\right) \cdot \left(\frac{364}{365}\right)$$

2nd person has 364/365 choices to ensure no common birthday

$$P_3 = \left(\frac{365}{365}\right) \cdot \left(\frac{364}{365}\right) \cdot \left(\frac{363}{365}\right)$$

3rd Person has 363 choices

$$P_k = \left(\frac{365}{365}\right) \cdot \left(\frac{364}{365}\right) \cdot \left(\frac{363}{365}\right) \cdots \left(\frac{365-k+1}{365}\right)$$

$$= \frac{{}_{365}P_k}{365^k}$$ ← k-permutation

e.g. $P_{23} = \left(\frac{365}{365}\right)\left(\frac{364}{365}\right) \cdots \left(\frac{343}{365}\right) \approx 0.4927 < 50\%$

Probability Examples

1. Roll 3 6-sided dice.

$P(\Sigma = 18) = ?$ $\quad E = \{(6,6,6)\} = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{6^3} = \frac{1}{216}$

Note: $|E| = 1$, $|S| = 6^3$ and all outcomes are equally likly so $P(E) = |E|/|S|$

2. $P(\Sigma = 17) = ?$

$E = \{(5,6,6), (6,5,6), (6,6,5)\}$
$|E| = 3$

$\therefore P = 3/216 = 1/72$

3. $P(\text{Roll} = \{1,2,3\})$ (in any order)

$|E| = 3! = 6 \quad \therefore P = 6/216 = 1/36$

4. Dealt 2 cards in range 2..10 ?

$|S| = \binom{52}{2}$ (# ways to be dealt 2 cards)

$|E| = \binom{36}{2}$ 9 ranks (2..10) X 4 suits/rank $= 36$

$\therefore |E|/|S| = \binom{36}{2}/\binom{52}{2} \approx 0.475$

$=$

OPTIONAL

5. Deck with 7 cards: A B C 1 2 3 4

After shuffling, letters and numbers are in order but not necessarily contiguous.

Valid: A B 1 2 C 3 4 $\Rightarrow$ L L N N L N N

1 2 3 A 4 B C $\Rightarrow$ N N N L N L L

There are $7!$ orderings.

The number of valid possibilities = # of choices for where the letters (or numbers) go:

$$\binom{7}{3} = \binom{7}{4}$$

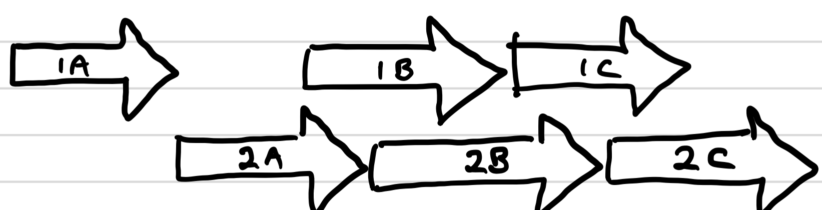
$$\therefore P = \frac{\binom{7}{3}}{7!} = \frac{\frac{7!}{3!\,4!}}{7!} = \frac{1}{3!\,4!}$$

$$\boxed{\text{Question: How would you extend this to 3 classes of cards ??}}$$

This has a real world application: A database must interleave operations of two transactions while maintaining the relative order of each transaction's operations.

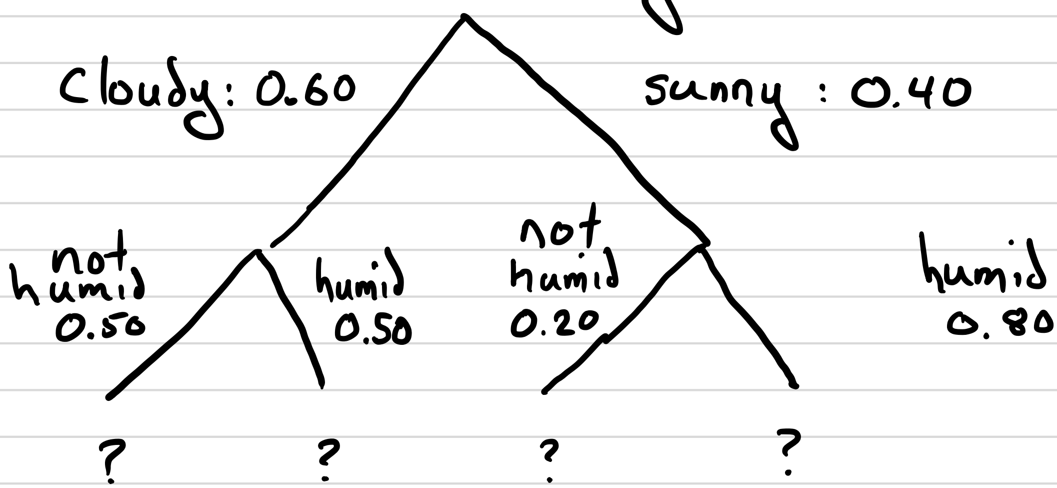
Transaction #1 :

Transaction #2 :



There are, however, additional constraints that further restrict transactional concurrency.

Conditional Probability and Weather



$$P(\text{humid}) = P(\text{cloudy } \& \text{ humid}) + P(\text{sunny } \& \text{ humid})$$

$$= P(\text{humid} \mid \text{cloudy}) \cdot P(\text{cloudy}) + P(\text{humid} \mid \text{sunny}) \cdot P(\text{sunny})$$

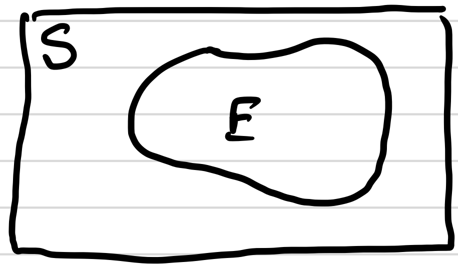
$$= (.50)(.60) + (.80)(.40)$$

$$= .30 + .32$$

$$= .62 \quad (62\% \text{ of the time})$$

Conditional Probability : Interpretation

$P(E) = |E| / |S|$



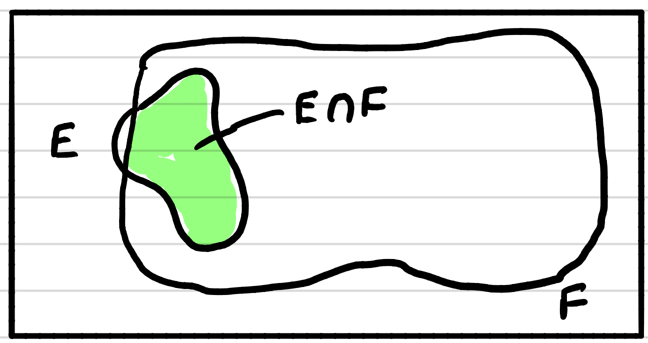
With dependent events E, F :

$P(E \cap F) = P(E) \cdot P(F|E)$

OR

$P(F|E) = \dfrac{P(E \cap F)}{P(E)}$

$= |E \cap F| / |E|$



Read: "What fraction of E is also in F ?

$P(F|E) \sim 1.0$

$P(E|F) \ll 1$

It follows that $P(E/F) = P(E \cap F) / P(F) = \dfrac{|E \cap F|}{|F|}$

And therefore:

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

So $\boxed{P(F|E) = \dfrac{P(E|F) \cdot P(F)}{P(E)}}$ } Bayes' Theorem (LATER!)