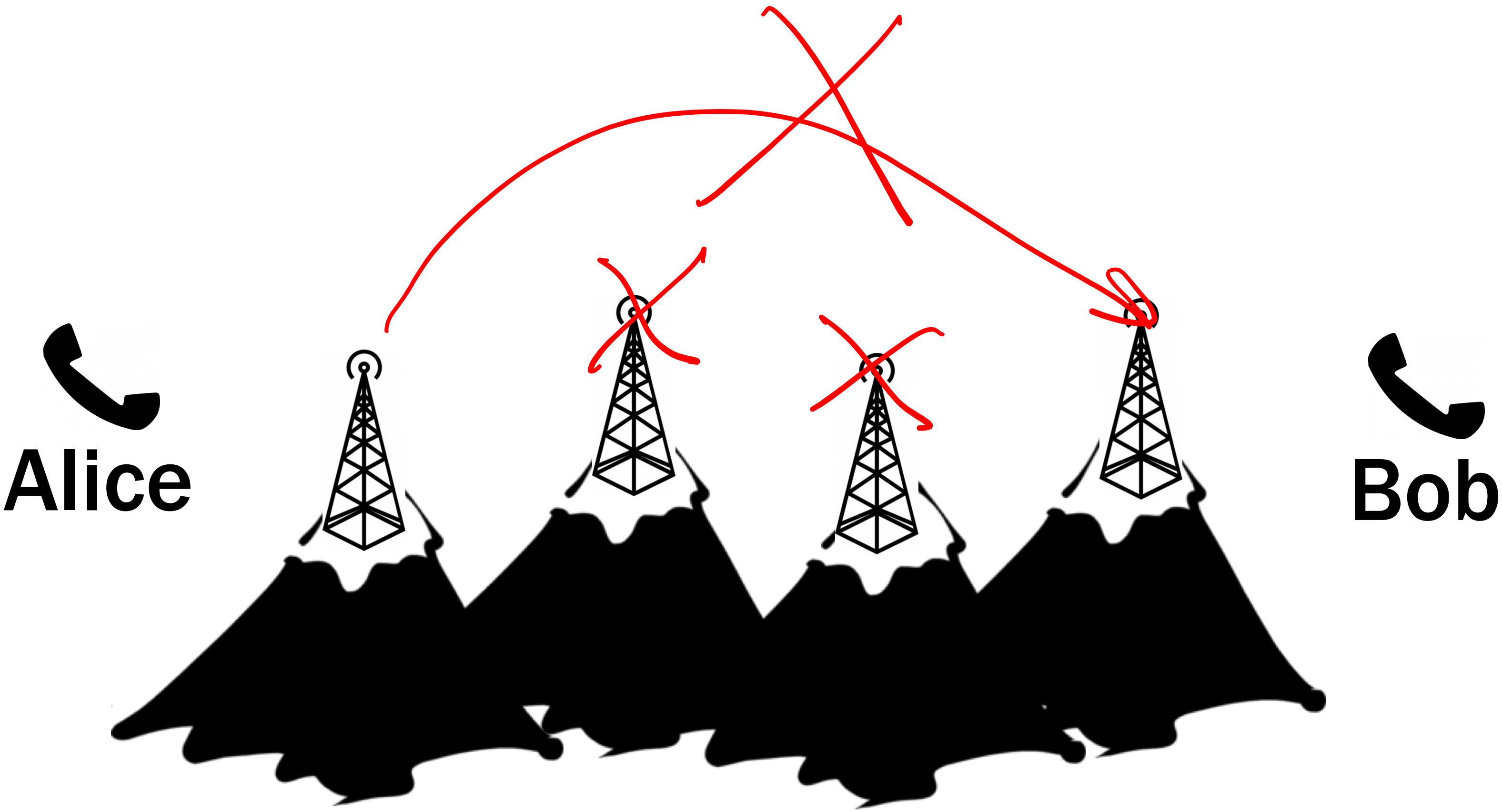
ENGG 2430 / ESTR 2004: Probability and Statistics

Spring 2019

1. Probabilistic Models

Andrej Bogdanov


Alice
Bob

Can Alice and Bob make a connection?

In uncertain situations we want a number saying how likely something is

probability

The cheat sheet

1. Specify all possible **outcomes**

2. Identify **event(s)** of interest

3. Assign **probabilities**

4. Shut up and **calculate**!

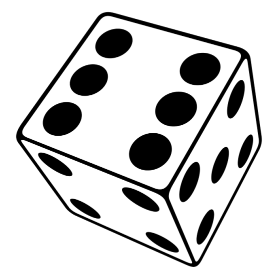
$\frac{8}{16} = \frac{1}{2}$


D
A more likely than B
A
C
B

Sample spaces

The sample space is the set of all possible outcomes.

Examples



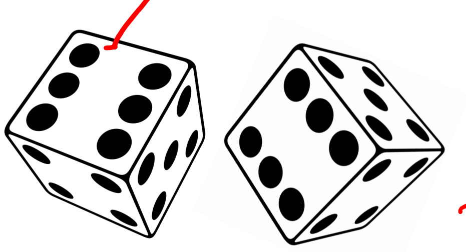
$\Omega = \{H, T\}$

$\Omega = \{1, 2, 3, 4, 5, 6\}$

Sample spaces

three coin tosses

$$\Omega = \{ \text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}, \text{THH}, \text{THT}, \text{TTH}, \text{TTT} \}$$



a pair of dice

$$\Omega = \{ \begin{array}{l} 11, 12, 13, 14, 15, 16, \\ 21, 22, 23, 24, 25, 26, \\ 31, 32, 33, 34, 35, 36, \\ 41, 42, 43, 44, 45, 46, \\ 51, 52, 53, 54, 55, 56, \\ 61, 62, 63, 64, 65, 66 \end{array} \}$$

Events

An event is a subset of the sample space.

$$\Omega = \{\ \mathtt{HHH}, \mathtt{HHT}, \mathtt{HTH}, \cancel{\mathtt{HTT}}, \\ \mathtt{THH}, \mathtt{THT}, \cancel{\mathtt{TTH}}, \cancel{\mathtt{TTT}}\ \}$$

Exactly two heads: $A = \{HHT, HTH, THH\}$

No consecutive tails: $B = \{HHH, HHT, HTH, THH, THT\}$

Discrete probability

A probability model is an assignment of probabilities to elements of the sample space.

Probabilities are nonnegative and add up to one.

Example: three fair coins

$$\Omega = \{ \text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}, \text{THH}, \text{THT}, \text{TTH}, \text{TTT} \}$$

1/8 1/8 . . . . 1/8

EQUALLY LIKELY OUTCOMES

Calculating probabilities

Exactly two heads:

$\mathcal{A} = \{$ HHT, HTH, THH $\}$

1/8 1/8 1/8

$\mathbf{P}(\mathcal{A}) =$ 3/8

No consecutive tails:

$B = \{$ HHT, HTH, THH, THH, THT $\}$

1/8 ... 1/8

$\mathbf{P}(B) =$ 5/8

Uniform probability law

If all outcomes are equally likely, then...

$$\mathbf{P}(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } \Omega} = \frac{|A|}{|\Omega|}$$

...and probability amounts to counting.

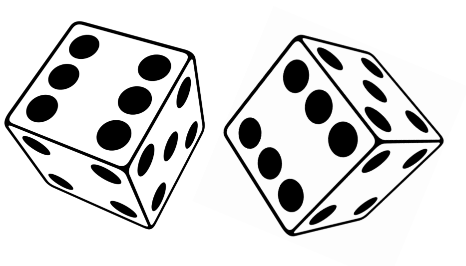
Product rule for counting

Experiment 1 has n possible outcomes.

Experiment 2 has m possible outcomes.

Together there are nm possible outcomes.

Examples



$6 \times 6 = 36$

$2 \times 2 \times 2 = 8$

$6 \times 2 = 12$

12

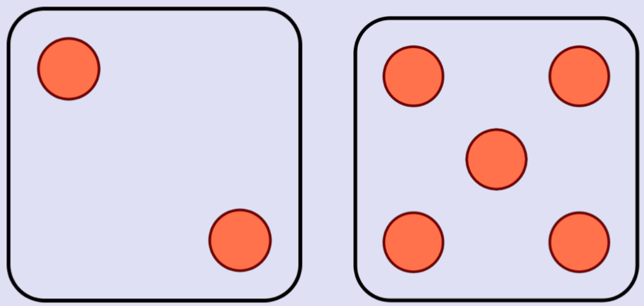
Generalized product rule

Experiment 1 has n possible outcomes.

For each such outcome,
experiment 2 has m possible outcomes.

Together there are nm possible outcomes.

You toss two dice. How many ways are there for the two dice to come out different?



A

15 ways

B

25 ways

C

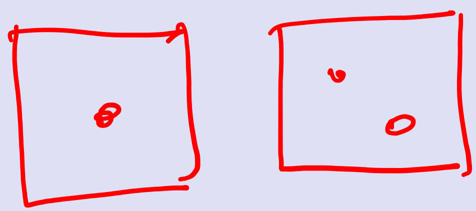
30 ways

Solution 1:

11, 12, 13, 14, 15, 16,
21, 22, 23, 24, 25, 26,
31, 32, 33, 34, 35, 36,
41, 42, 43, 44, 45, 46,
51, 52, 53, 54, 55, 56,
61, 62, 63, 64, 65, 66

$36-6$
$=30$

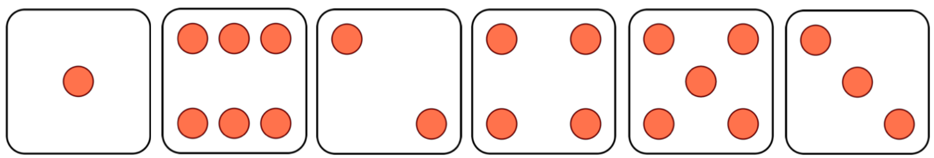
Solution 2:



$6 \times 5 = 30$

Permutations

You toss **six dice**. How many ways are there for **all** six to come out **different**?



$$6 \times 5 \times 4 \times 3 \times 2 \times 1$$

The number of **permutations** of n different objects is

$$n \times (n-1) \times \dots \times 1 = n!$$

Equally likely outcomes

For two dice, the chance both come out different is

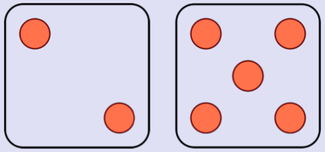
$$\Pr(A) = \frac{|A|}{|\Omega|} = \frac{30}{36} = \frac{5}{6} \approx 83.5\%$$

For six dice, the chance they all come out different is

$$\Pr(B) = \frac{|B|}{|\Omega|} = \frac{6!}{6^6} \approx 1.5\%$$

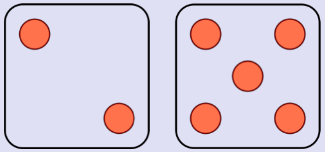
Toss two fair dice. What are the chances that…

(a) The second one is bigger?



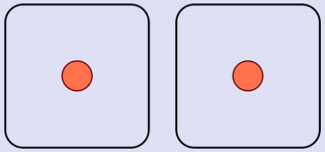
$$\frac{15}{36}$$

(b) The sum is equal to 7?



$$\frac{6}{36} = \frac{1}{6}$$

(c) The sum is even?



$$\frac{6 \times 3}{36} = \frac{18}{36} = \frac{1}{2}.$$

11, 12, 13, 14, 15, 16,
21, 22, 23, 24, 25, 26,
31, 32, 33, 34, 35, 36,
41, 42, 43, 44, 45, 46,
51, 52, 53, 54, 55, 56,
61, 62, 63, 64, 65, 66

There are 3 brothers. What is the probability that their birthdays are

(a) All on the same day of the week?

M T W T F S S

$$\Omega = \{M, T, W, R, F, S, U\}^3 \qquad |\Omega| = 7^3$$

$$= \{(b,c,d) : \ldots\}$$

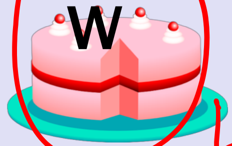
$$E = \{(b,c,d) : b=c=d\}$$

$$= \{MMM, TTT, \ldots, UUU\}$$

$$\boxed{P(E) = \frac{7}{7^3} = \frac{1}{7^2}}$$

(b) All on different days of the week?

M T W T F S S



b

c

$$F = \{(b,c,d) : b,c,d \text{ ALL DIFFERENT}\}$$

$$P(F) = \frac{|F|}{|\Omega|} = \frac{7 \times 6 \times 5}{7^3} = \frac{30}{49}$$

a **classical,** b **jazz, and** c **pop CDs are arranged at random. What is the probability that all CDs of the same type are contiguous?**

$$\underbrace{b\text{ CDs}}_{J}\ \underbrace{a\text{ CDs}}_{C}\ \underbrace{c\text{ CDs}}_{P}$$

Ω = PERMUTATIONS OF ALL CDs

$|\Omega| = (a+b+c)!$

E = CONTIGUOUS

TYPES

$$\boxed{\Pr(E) = \frac{|E|}{|\Omega|} = \frac{3!\cdot a!\cdot b!\cdot c!}{(a+b+c)!}}$$

Partitions

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

$\{\underbrace{1,2}, \underbrace{3,4,5}\}$

$\{1,2\}$ $\{3,4,5\}$

is the number of size-k subsets of a size-n set

In how many ways can you partition a size-n set into three subsets of sizes n_1, n_2, n_3?

$\{1,2,3,4,5\}$

A B C

SIZE 1 SIZE 2 SIZE 2

$\{3\}$ $\{5\}$ $\{1,2,4\}$

$$\binom{n}{n_1}\binom{n-n_1}{n_2} = \frac{n!}{n_1!\,\cancel{(n-n_1)!}} \cdot \frac{\cancel{(n-n_1)!}}{n_2!\,(n-n_1-n_2)!}$$

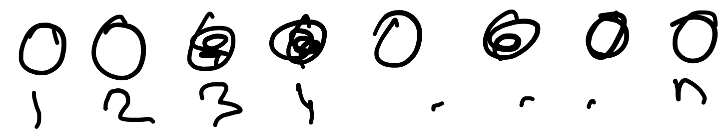
$$= \frac{n!}{n_1!\cdot n_2!\cdot n_3!}$$

Partitions and arrangements

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

size-k subsets of a size-n set

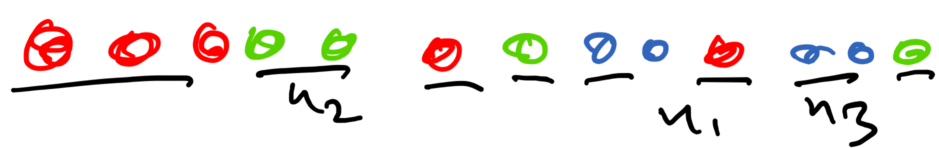
arrangements of k white and n - k black balls



$$\binom{n}{n_1, \ldots, n_t} = \frac{n!}{n_1! \cdot n_2! \cdots n_t!}$$

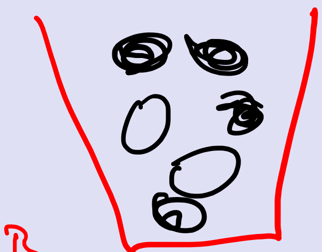
partitions of a size-n set into t subsets of sizes $n_1, \ldots, n_t$

arrangements of n_1 red, n_2 blue, …, n_t green balls



An urn has 10 white balls and 20 black balls. You draw two at random. What is the probability that their colors are different?

Ω = ARRANGEMENTS OF 10 W & 20 B BALLS



● ● ○ ○ ● ● ● ○ $\cdots$ ○ $\in \Omega$

1 2 3 $\ldots$ 30

$$E = \{ ○ ● \underbrace{\cdots}_{9W\ 19B},\ ● ○ \underbrace{\cdots}_{9W\ 19B} \}$$

EQUALLY LIKELY OUTCOMES

$$P(E) = \frac{|E|}{|\Omega|} = \frac{\binom{28}{9} + \binom{28}{9}}{\binom{30}{10}} = \frac{2 \times \binom{28}{9}}{\binom{30}{10}} = \frac{2 \times \frac{28!}{9! \cdot 19!}}{\frac{30!}{10! \cdot 20!}}$$

$$= \frac{2 \cdot 10 \cdot 20}{29 \cdot 30} = \frac{20}{29} \cdot \frac{20}{30}$$

12 HK and 4 mainland students are randomly split into four groups of 4. What is the probability that each group has a mainlander?

$S = \{M_1, \ldots, M_4, H_1, \ldots, H_{12}\}$

Ω = ALL PARTITIONS OF S INTO 4 SETS OF 4.

$E = \{S_1, S_2, S_3, S_4$; EACH S_i CONTAINS SOME $M_?\}$

$\underbrace{\{M_3, H_1, H_7, H_{11}\}}_{S_1} \underbrace{\{M_1, H_2, H_3, H_4\}}_{S_2} \{\ldots\} \{\ldots\}$

EQUALLY LIKELY OUTCOMES

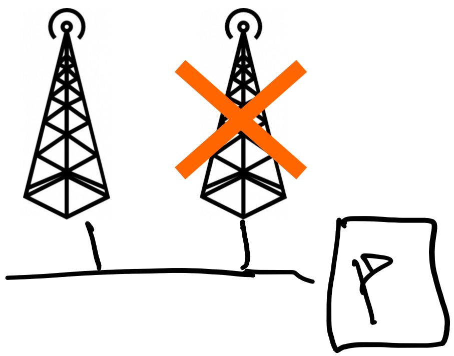
$$\Pr(E) = \frac{|E|}{|\Omega|} = \frac{4! \cdot \frac{12!}{3! \cdot 3! \cdot 3! \cdot 3!}}{\frac{16!}{4! \cdot 4! \cdot 4! \cdot 4!}}$$

MAINLAND

— PARTITIONS OF HK STUDENTS.

— PARTITIONS OF ALL

How to come up with a model?



a pair of antennas

each can be working or defective

$\Omega = \{ \mathtt{WW}, \mathtt{WD}, \mathtt{DW}, \mathtt{DD} \}$

Model 1: Each antenna defective 10% of the time

Defects are "independent"

WW WD DW DD

$.81 + .09 + .09 + .01 = 1$

Model 2: Dependent defects

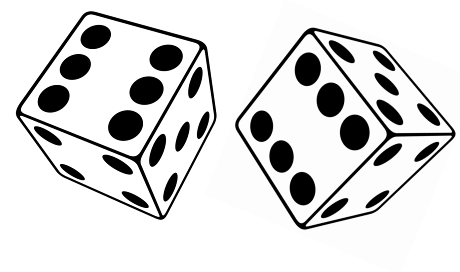
e.g. both antennas use same power supply

WW WD DW DD

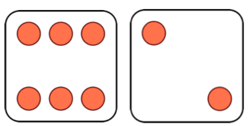
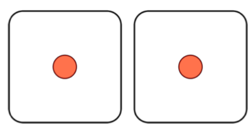
$.9 + 0 + 0 + .1 = 1$

How to come up with a model?

Option 1: Use common sense



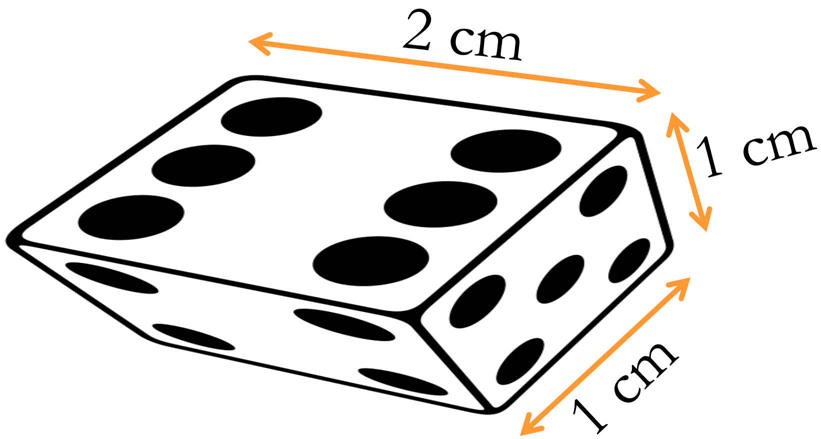
If there is no reason to favor one outcome over another, assign same probability to both

E.g.  and  should get same probability

1/36 1/36

So every outcome must be given probability $1/36$

The unfair die



$$\Omega = \{ 1, 2, 3, 4, 5, 6 \}$$

Common sense model: Probability $\propto$ surface area

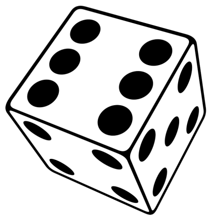
outcome	1	2	3	4	5	6
surface area (in cm^2)	2	1	2	2	1	2
probability	.2	.1	.2	.2	.1	.2

How to come up with a model?

Option 2: Frequency of occurrence

The probability of an outcome should equal the fraction of times that it occurs when the experiment is performed many times under the same conditions.

Frequency of occurrence



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

toss 50 times

44446163164351534251412664636216266362223241324453

outcome	1	2	3	4	5	6
occurrences	7	9	8	11	8	11
probability	.14	.18	.16	.22	.16	.22

Frequency of occurrence

The more times we repeat the experiment, the more accurate our model will be

toss 500 times

```
13565325111323652264346346233456634535436335145464236235511614456134412624621345412556566164361454 65
55645444326665111154232261536556643356223165166252534243112631124661334431221134562442223241526256 54
24351425655126532455545544352441532345351122324516565555514314353422253114533666524166215556636451 55
14665654234511546115561566231521422243262656542635222341452143134531552215615231352622556331446134 11
11151461136561562642553263315632116223556635451161446552161226565153622634563552321155655335212455 36
```

outcome	1	2	3	4	5	6
occurrences	81	79	73	72	110	85
probability	.162	.158	.147	.144	.220	.170

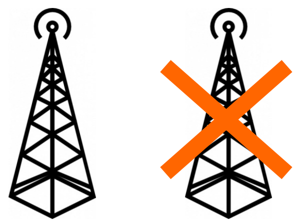
Frequency of occurrence

The more times we repeat the experiment, the more accurate our model will be

toss 5000 times

outcome	1	2	3	4	5	6
occurrences	797	892	826	817	821	847
probability	.159	.178	.165	.163	.164	.169

Frequency of occurrence



$S = \{ \texttt{WW}, \texttt{WD}, \texttt{DW}, \texttt{DD} \}$

	M	T	W	T	F	S	S	M	T	W	T	F	S	S
WW	x	x	x	x		x		x			x	x		
WD														
DW														x
DD					x		x		x	x			x	

outcome	WW	WD	DW	DD
occurrences	8	0	1	5
probability	8/14	0	1/14	5/14

Frequency of occurrence

Give a probability model for the gender of Hong Kong young children.

sample space = { boy, girl }

Model 1: common sense 1/2 1/2

Model 2: .51966 .48034

1.2 按年齡組別及性別劃分的年中人口
Mid-year population by age group and sex

人數
Number of persons

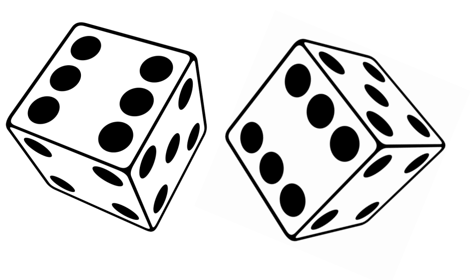
年齡組別（歲） Age group (years)	性別 Sex		2001	2006	2007	2008	2009	2010	2011
0 - 4	男性	M	142 000	110 400	111 300	114 000	117 700	124 200	129 500
	女性	F	130 800	102 600	103 200	105 200	108 300	113 800	119 700

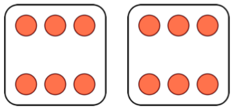
from *Hong Kong annual digest of statistics, 2012*

How to come up with a model

Option 3: Ask the market

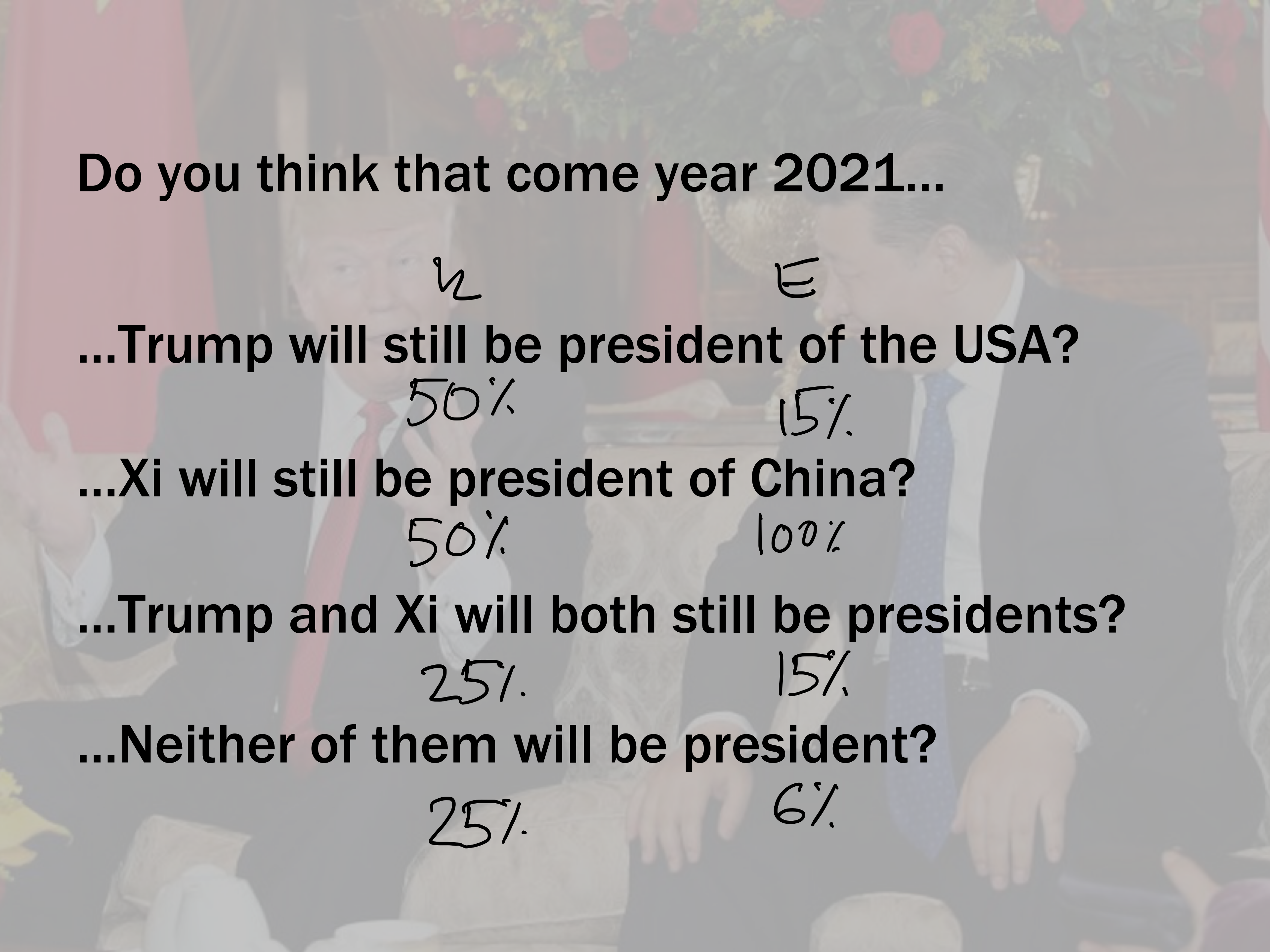
The probability of an outcome should be proportional to the amount of money you are willing to bet on it.



Will you bet on 

... if the casino's odds are 35:1?

... how about 37:1? 36:1


Do you think that come year 2021…
½
E
…Trump will still be president of the USA?
50%
15%
…Xi will still be president of China?
50%
100%
…Trump and Xi will both still be presidents?
25%
15%
…Neither of them will be president?
25%
6%

Events

An event is a subset of the sample space.

Examples

$\Omega = \{\ \mathrm{HH},\ \mathrm{HT},\ \mathrm{TH},\ \mathrm{TT}\ \}$

both coins come out heads $E_1 = \{HH\}$

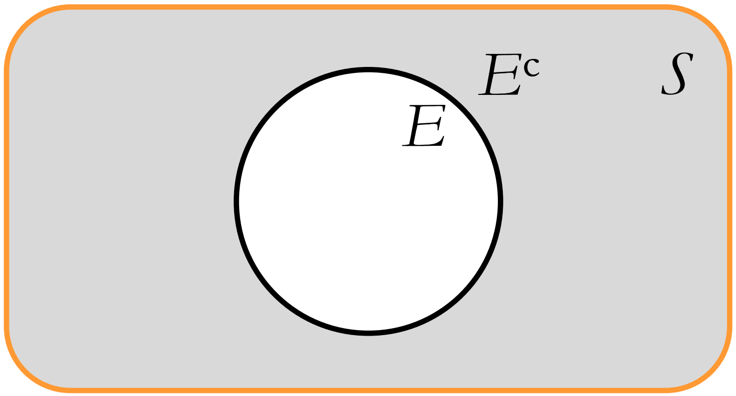
first coin comes out heads $E_2 = \{HH, HT\}$

both coins come out same $E_3 = \{HH, TT\}$

Events

The complement of an event is the opposite event.

both coins come out heads	$E_1 = \{\ \mathtt{HH}\ \}$
both coins do not come out heads	$E_1{}^c = \{HT, TH, TT\}$



Events

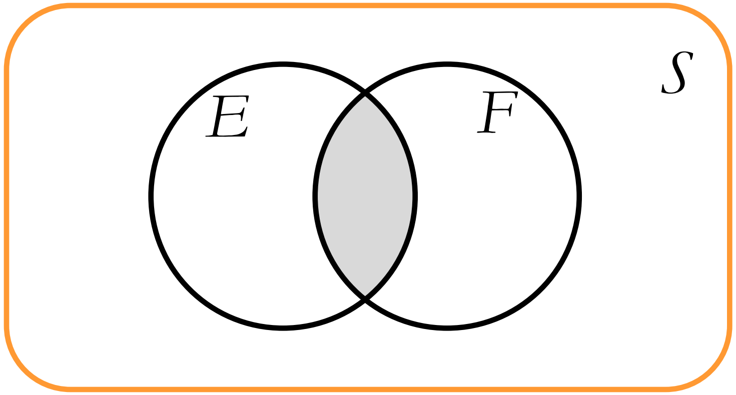
The intersection of events happens when all events happen.

(a) first coin comes out heads $E_2 = \{\ \mathtt{HH}, \mathtt{HT}\ \}$

(b) both coins come out same $E_3 = \{\ \mathtt{HH}, \mathtt{TT}\ \}$

---

both (a) and (b) happen $E_2 \cap E_3 = \{HH\}$



Events

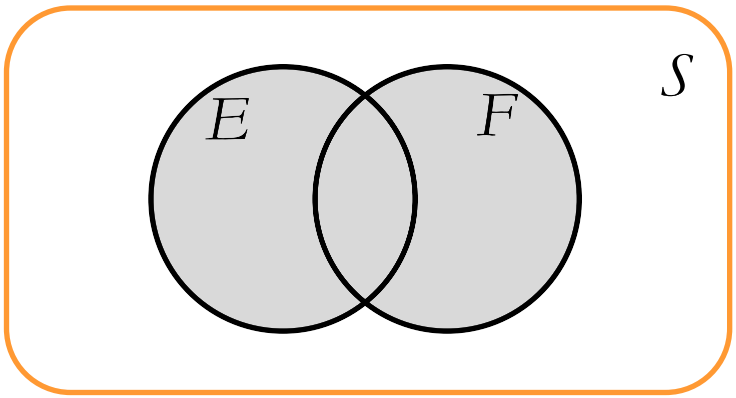
The union of events happens when at least one of the events happens.

(a) first coin comes out heads $E_2 = \{\ \mathtt{HH}, \mathtt{HT}\ \}$

(b) both coins come out same $E_3 = \{\ \mathtt{HH}, \mathtt{TT}\ \}$

---

at least one happens $E_2 \cup E_3 = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TT}\}$



Probability for finite spaces

The probability of an event is the sum of the probabilities of its elements

Example

$$\Omega = \{ \underset{1/4}{\mathtt{HH}}, \underset{1/4}{\mathtt{HT}}, \underset{1/4}{\mathtt{TH}}, \underset{1/4}{\mathtt{TT}} \}$$

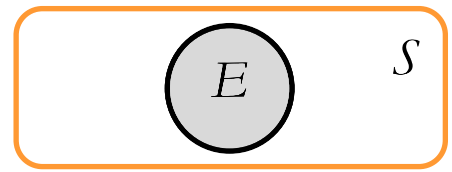
both coins come out heads	$E_1 = \{ \mathtt{HH} \}$	$P(E_1) = 1/4$
first coin comes out heads	$E_2 = \{ \mathtt{HH}, \mathtt{HT} \}$	$P(E_2) = 1/2$
both coins come out same	$E_3 = \{ \mathtt{HH}, \mathtt{TT} \}$	$P(E_3) = 1/2$

Axioms of probability

A sample space Ω.

For every event E, a **probability** $\mathbf{P}(E)$ such that

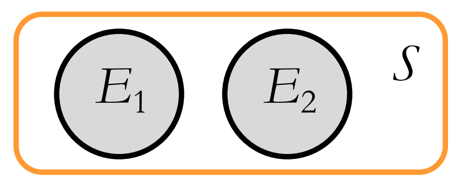
1. for every E: $0 \leq \mathbf{P}(E) \leq 1$



2. $\mathbf{P}(\Omega) = 1$

3. If $E_1, E_2, \ldots$ disjoint then

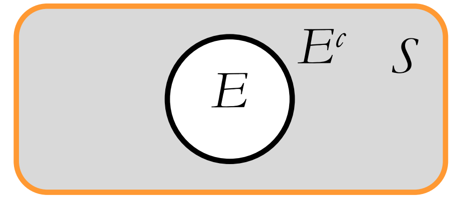
$$\mathbf{P}(E_1 \cup E_2 \cup \ldots) = \mathbf{P}(E_1) + \mathbf{P}(E_2) + \ldots$$



Rules for calculating probability

Complement rule:

$$\mathbf{P}(E^c) = 1 - \mathbf{P}(E)$$

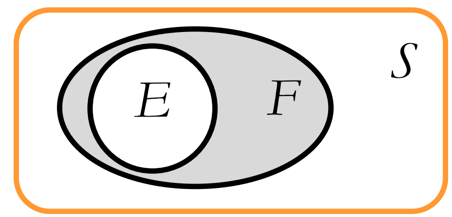


$P(E)+P(E^c)=1$

Difference rule: If $E \subseteq F$

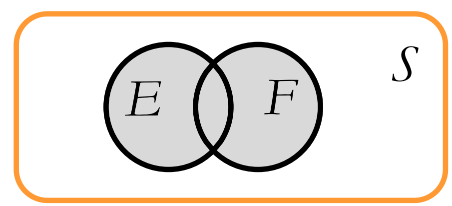
$$\mathbf{P}(F \cap E^c) = \mathbf{P}(F) - \mathbf{P}(E)$$

in particular, $\mathbf{P}(E) \leq \mathbf{P}(F)$



Inclusion-exclusion:

$$\mathbf{P}(E \cup F) = \mathbf{P}(E) + \mathbf{P}(F) - \mathbf{P}(E \cap F)$$



You can prove them using the axioms.

In some town 10% of the people are rich, 5% are famous, and 3% are rich and famous. For a random resident of the town what are the chances that:

(a) The person is not rich?

$P(R^c) = 1 - P(R) = 90\%$

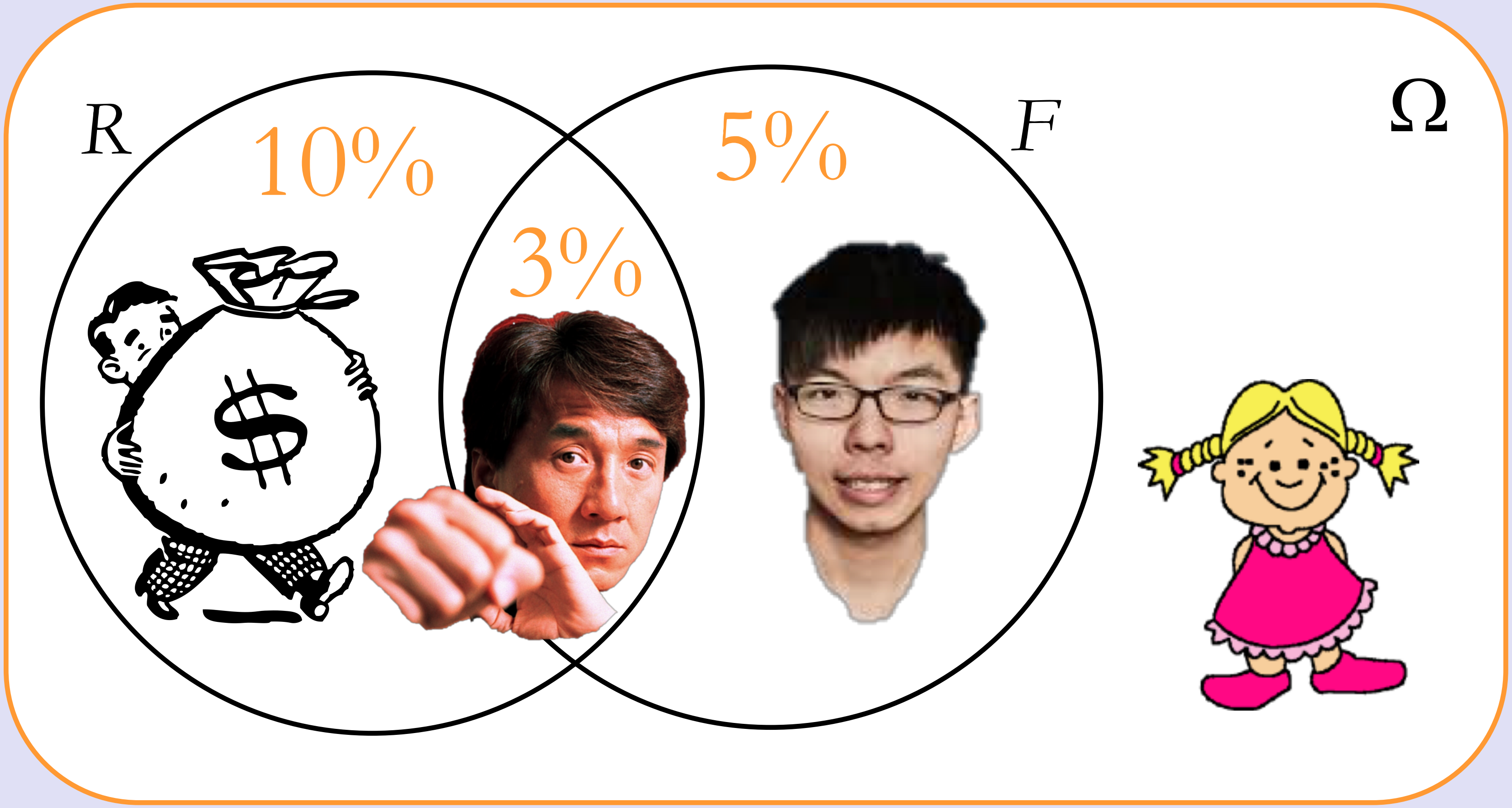
(b) The person is rich but not famous?

$P(R \cup F^c) = P(R) - P(R \cap F) = 10\% - 3\% = 7\%$

(c) The person is neither rich nor famous?

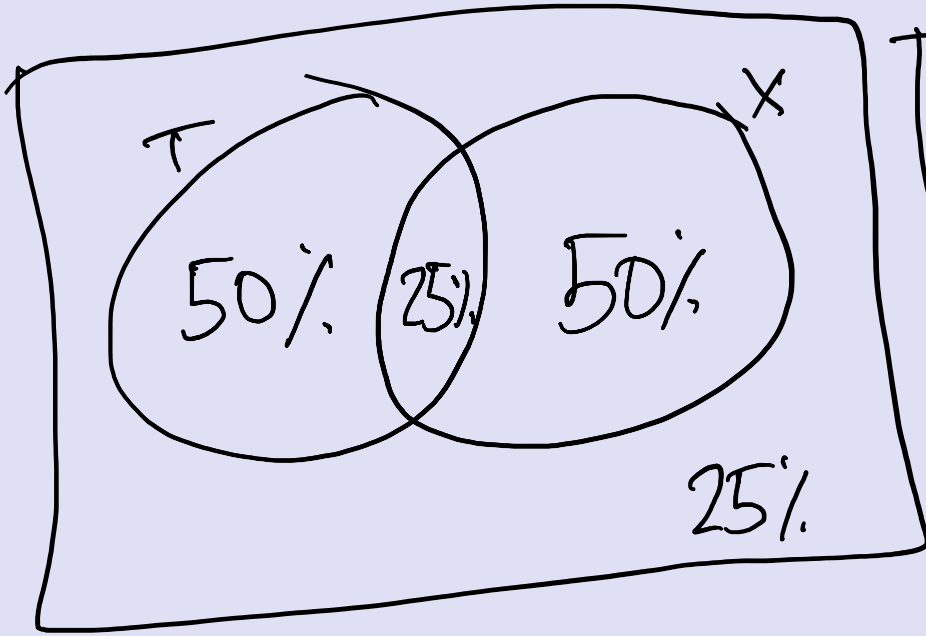
$P(R \cup F) = P(R) + P(F) - P(R \cap F) = 12\%$

$P((R \cup F)^c) = 88\%$



R
10%
3%
5%
F
Ω

KITHANIA

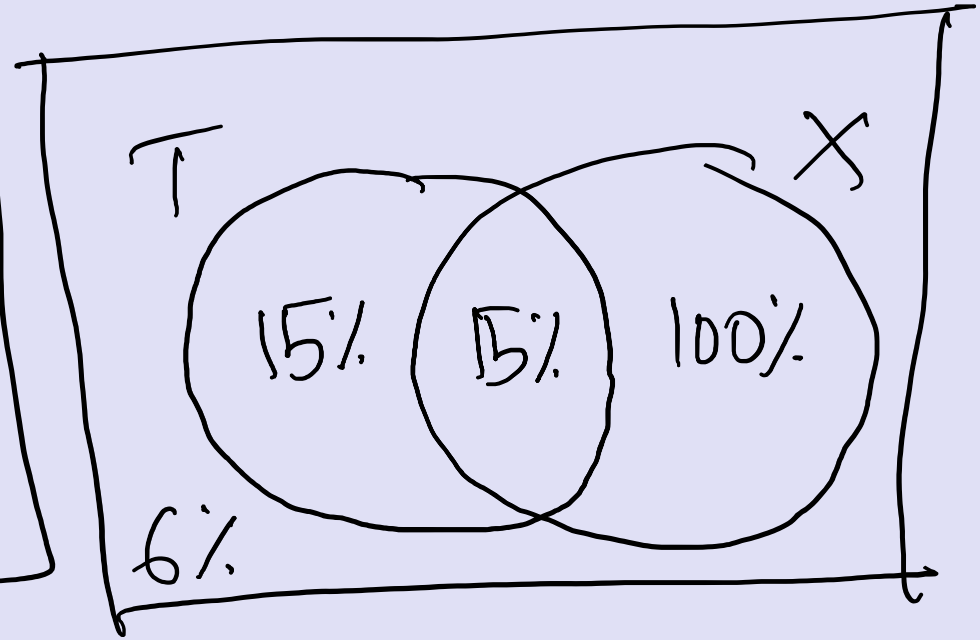


$P(T \cup X) = 50\% + 50\% - 25\% = 75\%$

$P((T \cup X)^c) = 1 - P(T \cup X) = 25\%$

CONSISTENT

ERIC



$P(T \cup X) = 100\% + 15\% - 15\% = 100\%$

$P((T \cup X)^c) = 1 - P(T \cup X) = 0\%$

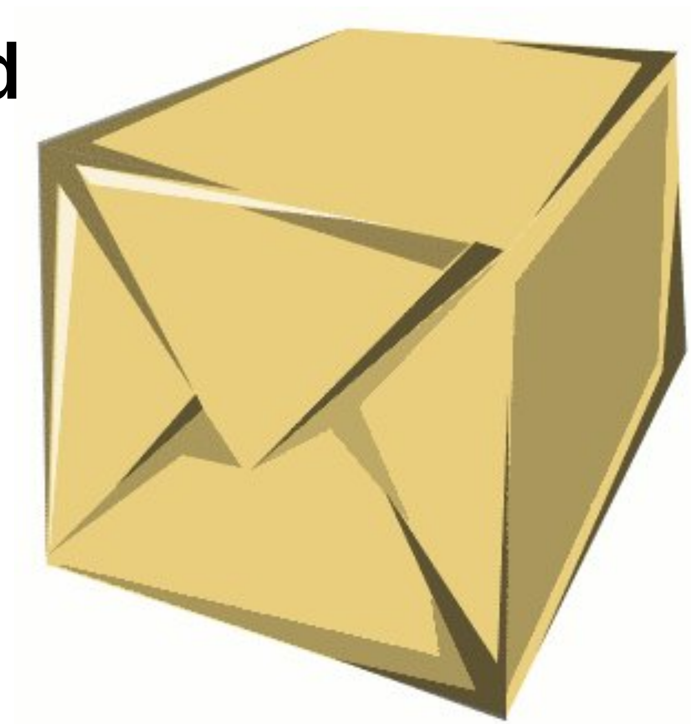
INCONSISTENT

Delivery time

A package is to be delivered between noon and 1pm.

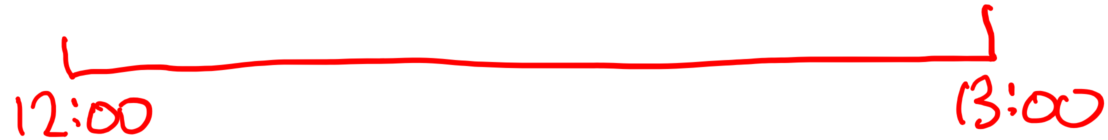
You go out between 12:30 and 12:45.

What is the probability you missed the delivery?



Delivery time

1. Sample space: $\Omega = [12{:}00, 13{:}00]$



2. Event of interest: $E = [12{:}30, 12{:}45]$

3. Probabilities?

$$P([a,b]) = \frac{b-a}{60}$$

$$P(E) = \frac{45-30}{60} = \frac{1}{4}$$

Uncountable sample spaces

In Lecture 2 we said:

> "*The probability of an event is the sum of the probabilities of its elements*"

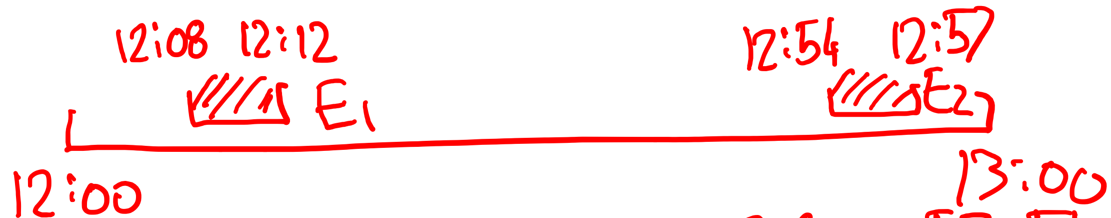
...but all elements have probability zero!

To specify and calculate probabilites, we have to work with the axioms of probability

Delivery time

From 12:08 - 12:12 and 12:54 - 12:57 the doorbell wasn't working.

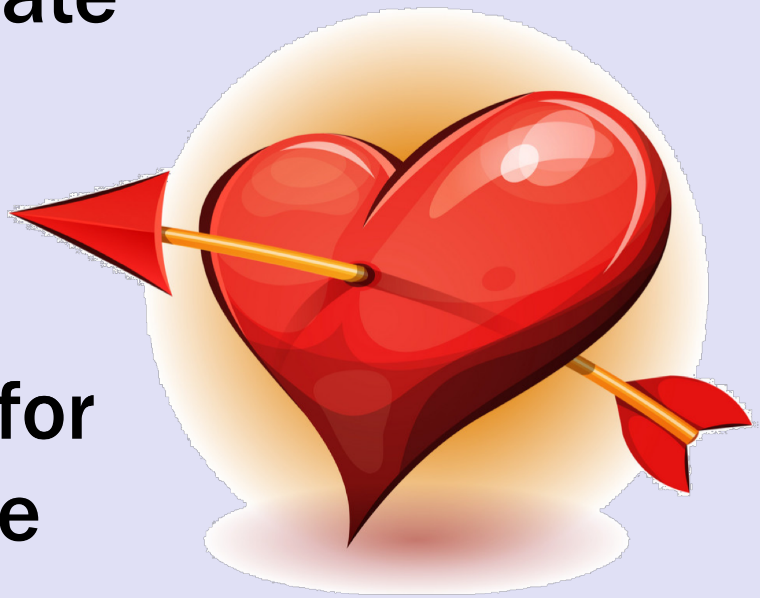
Event of interest: $E =$



Probability: $P(E) = P(E_1) + P(E_2) = \frac{12-8}{60} + \frac{57-54}{60}$

$$= \frac{7}{60}.$$

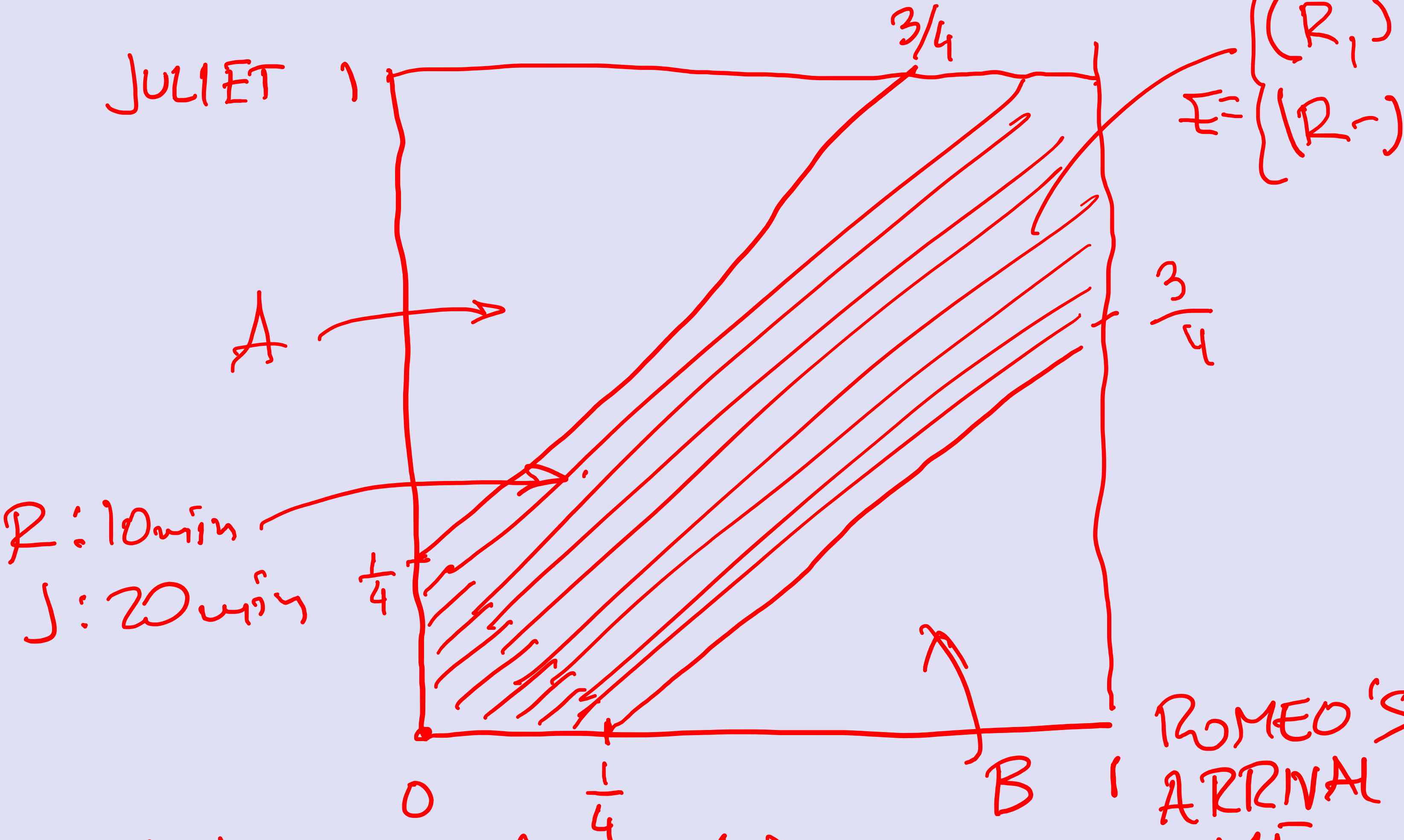
Romeo and Juliet have a date between 9 and 10.



The first to arrive will wait for 15 minutes and leave if the other isn't there.

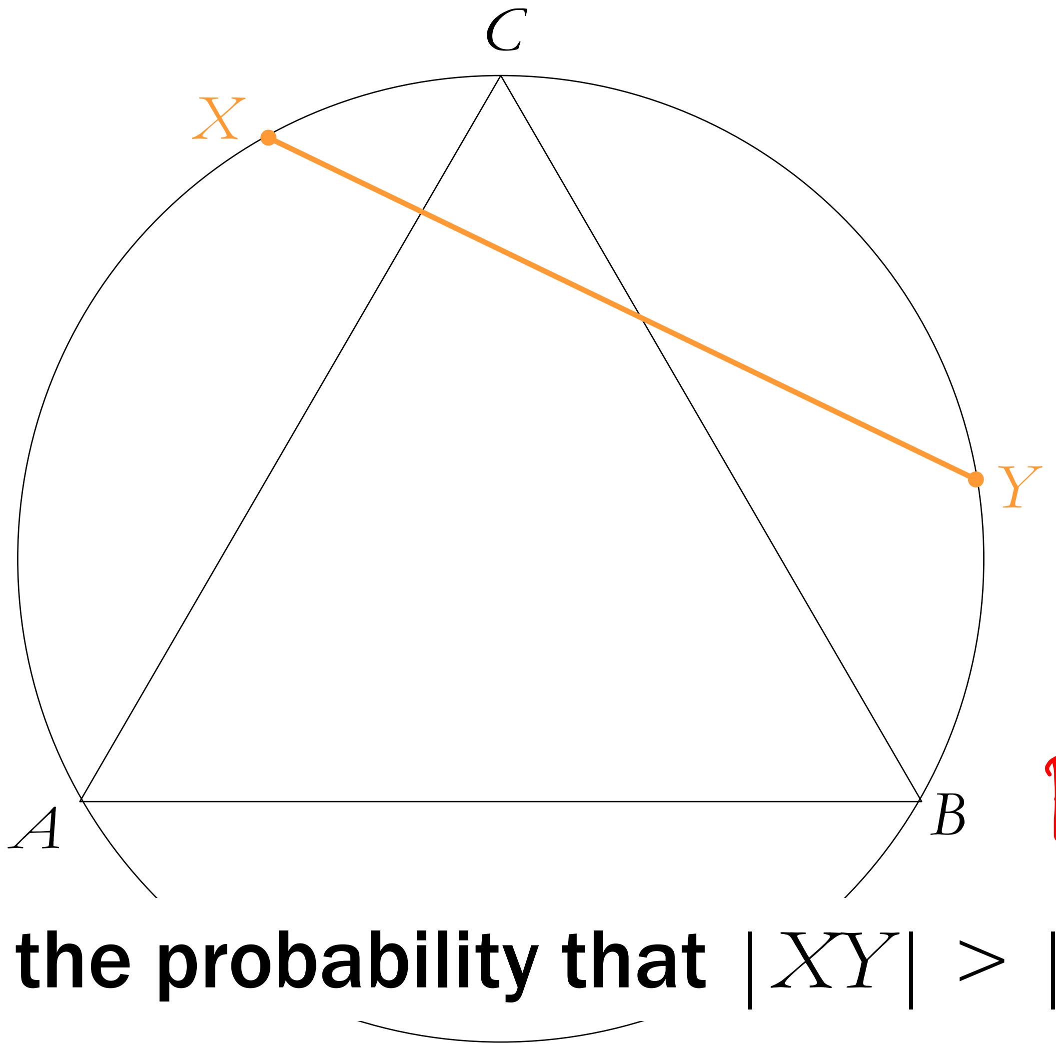
What is the probability they meet?

$$E = \left\{ (R, J) : |R - J| \le \tfrac{1}{4} \right\}$$



$$P(E) = 1 - P(A) - P(B) = 1 - \left(\tfrac{3}{4}\right)^2 = \boxed{7/16}$$

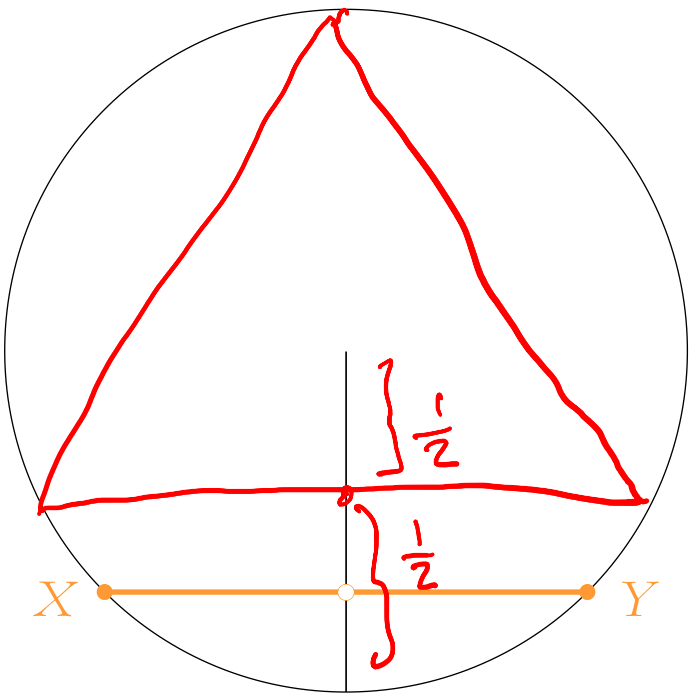
Bertrand's paradox



$P(|XY| > |AB|)$

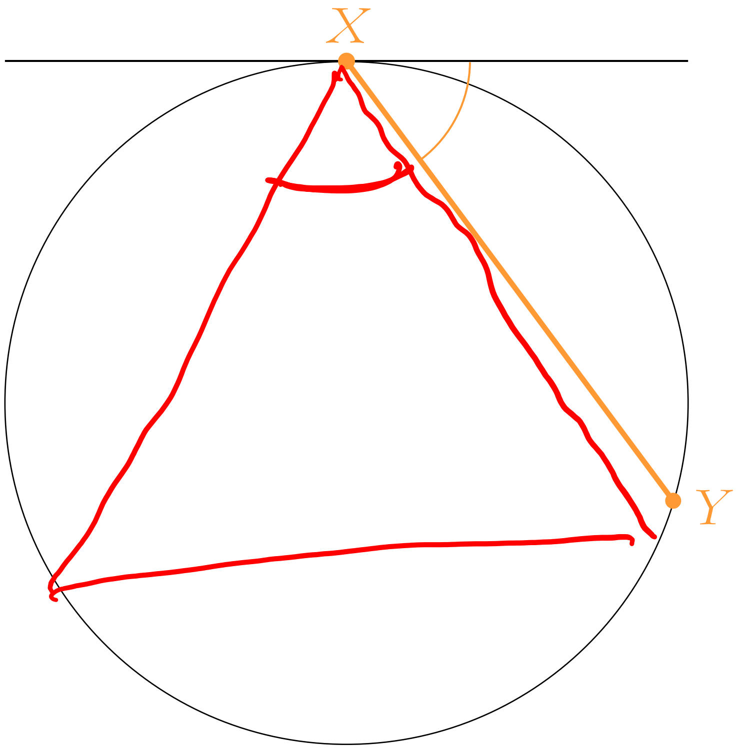
What is the probability that $|XY| > |AB|$?

Model 1



$P(|XY| > |AB|) = \frac{1}{2}$

Model 2



$P(|XY| > |AB|) = \frac{1}{3}$