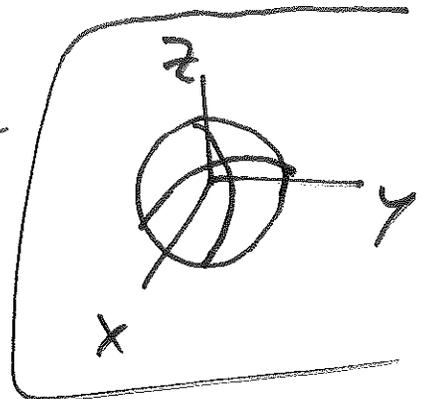


# Theorem (Banach-Tarski paradox (1924)) | ①

Given a sphere (solid ball) in 3-dimensional space of radius 1, you can break up the sphere into a finite number of non-overlapping subsets, <sup>("pieces")</sup> ~~and~~ move the pieces around, by rotating them and shifting them in the  $x, y$  or  $z$  directions, and reassemble



them into 2 identical copies of the original ball (!).

Why this matters to us

Later in this course we will want to work on problems where the sample space  $S$  is the positive integers  $\mathbb{N}$  <sup>(countable)</sup>

or the real numbers  $\mathbb{R}$  (uncountable). <sup>(12)</sup>

Because of weird results like the Banach-

Tarski paradox, Kolmogorov found that

when  $S$  is infinite, the set  <sup>$2^S$</sup>  of all subsets of  $S$  is "too big" and "too strange"

to permit the assignment of probabilities to all the sets in  $2^S$  in a logically-

internally-consistent way. when  $S$

is infinite, Kolmogorov was forced to restrict attention to a smaller collection

of subsets of  $S$  <sup>than  $2^S$</sup>  in which nothing weird

can happen. (see p. 7 of DS). The sets

in this smaller collection  <sup>$\subset$</sup>  have to

satisfy 3 simple rules to avoid the 13  
weirdness.

Rule 1:  $\mathcal{E}$  includes the entire sample space.

Rule 2: If an event  $A$  is in  $\mathcal{E}$  then so is its complement  $A^c$ .

Rule 3 requires a Definition Given any

two sets  $A$  and  $B$ , the union of  $A$  and  $B$  (written  $A \cup B$  or  $B \cup A$ ) is the set formed by throwing all the elements of  $A$  and all the elements of  $B$  <sup>together</sup> into one (potentially bigger) set (and discarding any and all duplicates).

This idea can be extended to more (14)  
than 2 sets: if  $A_1, A_2, \dots, A_n$  are  
events, we can talk about

$$(A_1 \cup A_2 \cup \dots \cup A_n) \stackrel{\Delta}{=} \bigcup_{i=1}^n A_i; \text{ and}$$

$\stackrel{\Delta}{=}$  is defined  
to be

if  $A_1, A_2, \dots$  is a countable collection  
of events we can even talk about

$$(A_1 \cup A_2 \cup \dots) \stackrel{\Delta}{=} \bigcup_{i=1}^{\infty} A_i$$

Rule 3:

If  $A_1, A_2, \dots$  are all in  $\mathcal{C}$  then

so is  $\bigcup_{i=1}^{\infty} A_i$ .

Example

whenever

$|\mathcal{S}| < \infty$  we can take  $\mathcal{C} = 2^{\mathcal{S}}$  with  
no weirdness arising; in other

words, if the sample space  $S$  is finite, <sup>(13)</sup>  
we can meaningfully assign probabilities  
to all of the subsets of  $S$ .

Some  
more  
basic facts  
about sets

① For any event  $A$ ,

$$(A^c)^c = A.$$

②  $\phi^c = S$

and  $S^c = \phi$ .

For any events  $A, B$ :

③  $A \cup B = B \cup A$ ,

$$A \cup A = A, \quad A \cup A^c = S, \quad A \cup \phi = A,$$

$$A \cup S = S, \quad \text{and if } A \subset B \text{ then } A \cup B = B.$$

④ For any events  $A, B, C$ ,

$$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C)$$

(This is called associativity of the  
 $\cup$  operation)

Definition

with  $A$  and  $B$  any (16)

two sets, the intersection  $A \cap B$

is the set containing all, and only, those elements belonging both to  $A$

and to  $B$ .

If  $A$  is  
an event  
(set: a  
subset of  $S$ ),

(sets) set operation	(true/false propositions) logical operation
$A^c$	not $A$
$A \cup B$	$A \text{ or } B$
$A \cap B$	$A \text{ and } B$

we can

equivalently talk either about

the set  $A$  or the true/false

proposition that one of the elements

in  $A$  (15) the outcome of the experiment  $E$ .

Example (T-S discourse)  $A = \{NNNNN\}$  (17)

as a set is equivalent to the true/false proposition (exactly 0 T-S beliefs) ~~being~~ <sup>being</sup> true.

Even more basic facts about sets

(5) It's meaningful to talk about the intersection of more than 2 sets: with

$A_1, \dots, A_n$  the set  $A_1 \cap \dots \cap A_n = \bigcap_{i=1}^n A_i$

is meaningful, and with  $A_1, A_2, \dots$

so is  $\bigcap_{i=1}^{\infty} A_i$ .

(6)  $A, B, C$  any events:

$$A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C)$$

(associativity of the  $\cap$  operation)

Definition Two sets  $A, B$  are disjoint  $\equiv$  mutually exclusive if

$A \cap B = \emptyset$  (if they have no outcomes

in common).  $n$  sets  $A_1, \dots, A_n$  are disjoint if all <sup>distinct</sup> pairs are

disjoint:  $A_i \cap A_j = \emptyset$  for  $i \neq j$ .

logic equivalent | propositions  $A, B$

mutually exclusive  $\leftrightarrow$  they cannot both be true simultaneously

(Exactly 1 TS baby), (Exactly 2 TS babies) are mutually exclusive. Example (TS disease)

Still more  
basic facts  
about sets

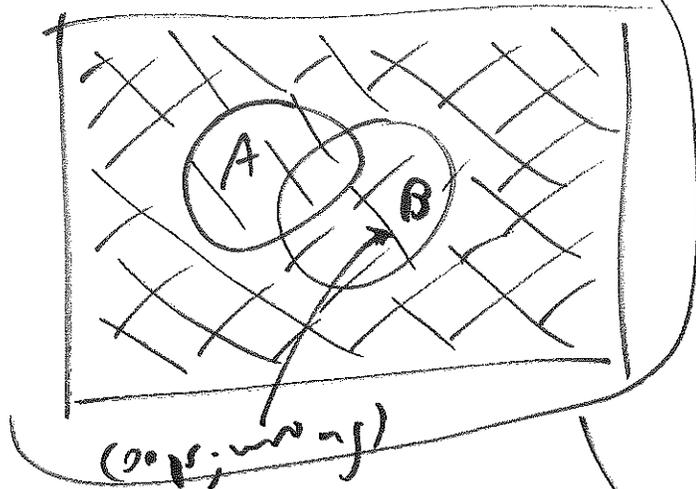
⑦ (attributed to Augustus <sup>①⑨</sup>  
de Morgan (1806 - 1871), a  
British logician):

De Morgan's  
Laws

$A, B$  any two sets:

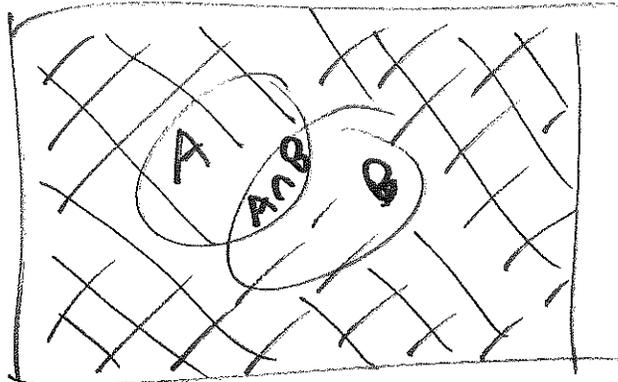
(a)  $(A \cup B)^c = A^c \cap B^c$

//  $A^c$  //  $B^c$



and (b)  $(A \cap B)^c = A^c \cup B^c$

//  $A^c$  //  $B^c$



logical v/s to parent:

(a) if  $(A \cup B)^c$  is true, then  $(A \cup B)$  is  
false, which can only occur if  $A$  and  
 $B$  are both false, making  $A^c \cap B^c$   
true.

or

and

(b) if  $(A \cap B)^c$  is true, then  $A \cap B$  is 20  
 false, which will occur if either one  
 (or both) of  $A, B$  are false, making

$$A^c \cup B^c \text{ true.}$$

⑧  $A, B, C$  any sets:

(a)

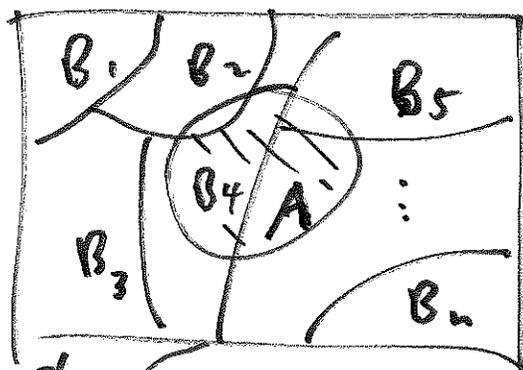
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

and (b)  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ .

(this is called the distributive property of  $\cap$  and  $\cup$ )

⑨ (important property for probability)

Definition: If you can find events



$B_1, \dots, B_n$  such that

(a) the  $B_i$  are mutually exclusive, and (b) the

$B_i$  are exhaustive, in the sense that

$\bigcup_{i=1}^n B_i = S'$ , then  $(B_1, \dots, B_n)$  forms a partition of  $S'$ .

The idea of a partition is that  $\textcircled{2}$   
every outcome in  $\mathcal{S}$  lives inside one,  
and only one, of the partition sets.

---

If you look at the Venn diagrams on  
p.  $\textcircled{20}$ , you'll see that (for any event  $A$ )

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n);$$

in other words,  $A = \bigcup_{i=1}^n (A \cap B_i)$ :

the partition chops  $A$  up into  $n$   
mutually exclusive pieces (some of  
which may be empty) whose union is  $A$ .

---

we're now ready to look at	Kolmogorov's
Kolmogorov wants to define	probability
$P_K(A)$ - what Axioms should be we!	Axioms

It was clear to Kolmogorov that  $\mathbb{P}_k(A)$  needs to be a function from  $\mathcal{C}$  (the collection of non-weird subsets of the sample space  $S$ ) to the real number line  $\mathbb{R}$ ; but what else should we assume about  $\mathbb{P}_k$ ? (22)

Axiom 1:

For all events  $A \in \mathcal{C}$ ,  $\mathbb{P}_k(A) \geq 0$   
(motivated by relative frequency)

Axiom 2:  $\mathbb{P}_k(S) = 1$  (again

motivated by relative frequency)

Axiom 3: For every countable collection of disjoint events  $A_1, A_2, \dots \in \mathcal{C}$ ,

$$P_k \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P_k(A_i) \quad (*) \quad (23)$$

turns out to be absolutely necessary but is hard to motivate: it's a small piece of genius on Kolmogorov's part that he assumed this not just for a finite number of disjoint events — and

if  $A_1, \dots, A_n$  are disjoint then

$$P_k \left( \bigcup_{i=1}^n A_i \right) = \sum_{i=1}^n P_k(A_i) \text{ follows from } (*)$$

— but also for a countable collection.

Consequences that follow

from Kolmogorov's

Axioms

(From now on I'll drop the subscript  $k$ .)

$$① \quad P(\emptyset) = 0$$

$\mathbb{D}r: P_r$

$P$

②  $P(A^c) = 1 - P(A)$  | ③ If  $A \subset B$  ②④  
then  $P(A) \leq P(B)$ .

④ For all events  $A$ ,  
 $0 \leq P(A) \leq 1$  ← (the easy rule)

⑤ For all events  $A, B$ , general addition rule for  $\square$   
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

⑥ (attributed to the Italian mathematician, Carlo Bonferroni (1892 - 1960)): For any events  $A_1, A_2, \dots, A_n$ ,

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) \quad \text{and}$$

$$P\left(\bigcap_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n P(A_i^c)$$

useful in statistics

# Tay-Sachs disease in more detail

NNNNN	0
TNNNN	1
NTNNN	
NNTNN	
NNNTN	
NNNNT	
TTNNN	2
TNTNN	
TNNTN	
TNNNT	
NTTNN	
NTNTN	
NTNNT	
NNTTN	
NNTNT	
NNNTT	
⋮	⋮
TTTTT	5

$\# \text{ of T-S babies} = Y$  Let's

see if we can work out  
 $P(Y=1), P(Y=2), \dots,$   
 $P(Y=5)$ ; we already  
 worked out

$$P(Y=0) = P(\text{exactly 0 T-S babies})$$

$$= P(\begin{matrix} \text{1st} \\ \text{baby} \\ \text{not} \\ \text{T-S} \end{matrix} \& \begin{matrix} \text{2nd} \\ \text{baby} \\ \text{not} \\ \text{T-S} \end{matrix} \& \dots \& \begin{matrix} \text{5th} \\ \text{baby} \\ \text{not} \\ \text{T-S} \end{matrix})$$

independence

$$= P(\begin{matrix} \text{1st} \\ \text{baby} \\ \text{not} \\ \text{T-S} \end{matrix}) \cdot P(\begin{matrix} \text{2nd} \\ \text{not} \\ \text{T-S} \end{matrix}) \cdot \dots \cdot P(\begin{matrix} \text{5th} \\ \text{not} \\ \text{T-S} \end{matrix})$$

identical distribution

$$= [1 - P(\begin{matrix} \text{1st} \\ \text{baby} \\ \text{T-S} \end{matrix})] \cdot \dots \cdot [1 - P(\begin{matrix} \text{5th} \\ \text{baby} \\ \text{T-S} \end{matrix})]$$

$$= 1 \cdot P^0 (1-p)^5 = (1-p)^5 = 24\%$$

5 with  $p = \frac{1}{4}$

A similar line of reasoning gives (26)

$$P(\bar{Y}=5) = P(TTTTT) = p^5 = \frac{1}{1 + p^5(1-p)^0}$$

what about  $P(\bar{Y}=1)$ ? The table

on the previous page lists all of the

outcomes with 1 T-5 baby: they

all have 1 T and 4 Ns, so each one

has probability  $p(1-p)^4$ , and there

are 5 of them, so  $P(\bar{Y}=1) = 5p^1(1-p)^4$ .

By similar reasoning  $P(\bar{Y}=2) = 10p^2(1-p)^3$

The outcomes with  $(\bar{Y}=3)$  are minor

images of those with  $(\bar{Y}=2)$ :

TTNNN

NNTTT

So there must also be 10 elements of  $S$  with  $(\Sigma=3)$  and  $P(\Sigma=3) = 10 p^3 (1-p)^2$

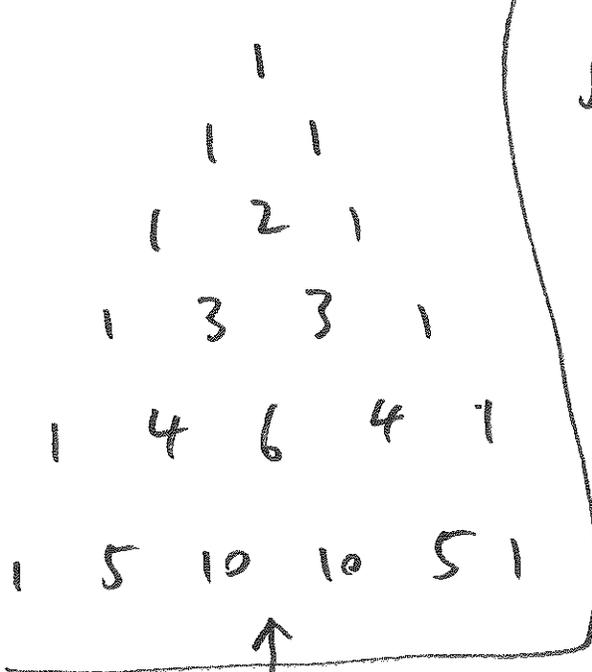
And finally,  $(\Sigma=4)$  is a mirror image of  $(\Sigma=1)$  so  $P(\Sigma=4) = 5 p^4 (1-p)^1$

# of T-S beliefs $y$	$P(\Sigma=y)$	with $p = \frac{1}{4}$
0	$1 p^0 (1-p)^5$	0.2373
1	$5 p^1 (1-p)^4$	0.3955
2	$10 p^2 (1-p)^3$	0.2637
3	$10 p^3 (1-p)^2$	0.0879
4	$5 p^4 (1-p)^1$	0.0146
5	$1 p^5 (1-p)^0$	0.0010
	1	1.0000

Soon we'll call  $\Sigma$  a random variable (symbolizing the data generating process) and use  $y$  to stand for a possible value of  $\Sigma$ .

So it looks like

$$P(Y=y) = \boxed{?} p^y (1-p)^{5-y}$$



we could even be a bit more symbolic and note

that  $n=5$  is the number of times the basic dichotomy (T vs. N) occurs in this case study, so  $P(Y=y) = \boxed{?} p^y (1-p)^{n-y}$

What about  $\boxed{?}$

You can see that the

multiplicands  $\boxed{?}$  come from Pascal's Triangle, but can we write down a formula for them?

Permutations & Combinations

You have an ordinary deck of  $n=52$  playing cards.

How many possible poker hands of  $k=5$  cards can you draw at random without replacement from the deck?

---

It's like filling in 5 slots:  $\underline{\quad} \_ \_ \_ \_ \_$

the first slot can be filled in  $n=52$  ways, and the second in  $(n-1)=51$  ways, ..., the  $5^{\text{th}}$  slot in  $(n-k+1)=48$  ways; so the total # of ways you can do this is  $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48$

$$= n(n-1) \cdots (n-k+1) = 311,875,200$$

ways. This is called the number of permutations of 52 things taken 5 at a time.

Definition

The number of permutations of  $n$  distinct things taken  $k$  at a time

is written  $P_{-n,k} = n(n-1)\dots(n-k+1)$ .

How many possible orderings of a 52-card deck are there? Now there are 52

slots, e.g.,  $\frac{J}{52} \frac{3}{51} \dots \frac{A}{1}$ , so the total

number must be  $52 \cdot 51 \cdot \dots \cdot 1 =$  Def.

$n(n-1)\dots 1 \stackrel{\Delta}{=} n!$  read  $n$  factorial

$$= 806581751709438785716606368564037669752895054408832778240000000000000000 = 8.1 \cdot 10^{67}$$

with this notation you can see that ③

$$P_{n,k} = \frac{n(n-1)(n-k+1)(n-k)!}{(n-k)!} = \frac{n!}{(n-k)!}$$

Convention	$0! \triangleq 1$	Combinations
------------	-------------------	--------------

In the T-S case study we want to fill  $n=5$  slots, each either a T or an N.

Consider the special case in which the family ends up with exactly  $\binom{k}{1}$  T's total, i.e.,  $\binom{k}{1}$  T and  $\binom{n-k}{4}$  N's. Let's initially

imagine that all 5 of these T and N

symbols are different (like different

playing cards), by denoting them  $\left\{ \begin{matrix} T_1 \\ N_1, N_2, N_3, N_4 \end{matrix} \right\}$ .

There would then be  $n! = 5! = 120$  ~~30~~  
ways to arrange them in order left to

right, e.g.  $\frac{N_3}{-} T_1 \frac{N_4}{-} \frac{N_1}{-} \frac{N_2}{-}$ . Now take

the subscripts away: there are  $4!$  ways

to rearrange the  $N$ s among themselves

and  $1! = 1$  way to "rearrange" the  $T$ s

among themselves, so  $5!$  is way too

big and needs to be divided by  $4! \cdot 1!$ :

$$\frac{5!}{1! \cdot 4!} = \frac{n!}{k! \cdot (n-k)!} = \frac{5 \cdot 4!}{4!} = 5 \text{ (the right answer)}$$

Definition Given a set with  $n$  <sup>distinct</sup> elements,  
each distinct subset of size

$k$  is called a combination of elements,

and there are  $C_{n,k} = \frac{n!}{k! \cdot (n-k)!}$  ways to do this

Notation Everybody in the world other than De Groot & Schervish uses a different notation:  $\frac{n!}{k!(n-k)!} = \binom{n}{k}$ , read out loud as "n choose k"

Back to T-S So what we have shown

is  $P(I=y) = \binom{n}{y} p^y (1-p)^{n-y}$

# of T-S beliefs valid for all  $n \geq 1$  and  $y=0, 1, \dots, n$   $0 \leq p \leq 1$ .

Later we'll refer to this as the binomial distribution.