Definition An *experiment* $\mathcal{E}$ is a data-generating process in which all possible outcomes can be listed before $\mathcal{E}$ is performed.

Definition An *event* E is a set of possible outcomes of an experiment $\mathcal{E}$.

Example Tay-Sachs (T-S) disease

$\mathcal{E}$ = (the process by which the husband & wife end up with 5 children, each a T-S baby or not) the E of interest is $E = \{$at least 1 T-S baby$\}$

Definition The sample space $S^{(\Omega)}$ is the set of all possible outcomes (elements) of an experiment $\mathcal{E}$.

Example: T-S

Let T = (T-S baby) and N = (not T-S baby)

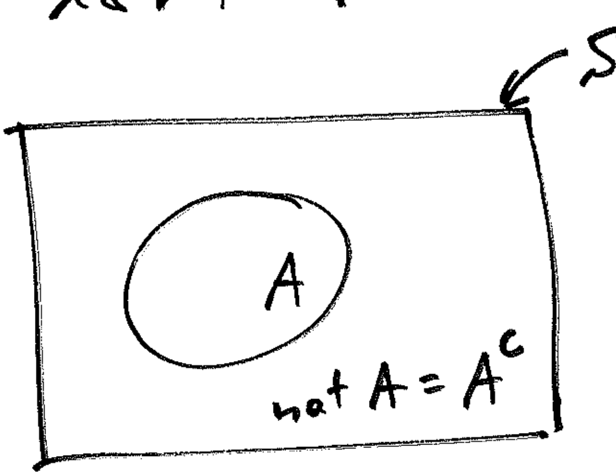
NNNNN	Here $S = \{NNNNN, \ldots, TTTTT\}$
TNNNN	
NTNNN	Since there are 2 possibilities for each baby (T, N) and 5 babies, the number of elements in S is $2^5 = 32$.
NNTNN	
NNNTN	
NNNNT	
TTNNN	
TNTNN	S is an example of a product space:
⋮	
TTTTT	$\underbrace{\{T,N\} \times \{T,N\} \times \cdots \times \{T,N\}}_{5} = \{T,N\}^5$.

E

Here $E = \{TNNNN, \ldots, TTTTT\}$.

Notation Let's use s to stand for the individual outcomes (elements) of S.

The theory of probability we'll look at in this class was developed by Kolmogorov (1933) in an attempt to rigorize the hypothetical process of repeatedly throwing a dart at a Venn diagram (rectangle)



The rules of this dart-throwing were simple: ① the dart must land somewhere inside (or on the boundary of) the rectangle S, which symbolically stands for the sample space, ④

and ② all the points where the dart might land in S are "equally likely" (as yet, an undefined primitive concept).

Definition The complement A^c of a set A contained in S is the set that contains all elements of S not in A

(You can see from the Venn diagram on p. ③ that the dart has to fall either in A or in A^c, which we could also call $\boxed{\text{not } A}$.) **Notation**

$s \in S$ means that $\left\{ \begin{array}{c} \text{outcome} \\ \text{element} \end{array} \right\}$ s belongs to S

$\in$ is an element of

Definition | A set A is contained in another set B (A is a subset of B) (written $A \subset B$) if every element of A is also in B; can also say that B contains A ($B \supset A$).

Evidently, if A and B are events, $A \subset B \leftrightarrow$ (iff) (if and only if) if A occurs then so does B

Consequences (theorem) | If A, B, C are events then (a) $A \subset B$ and $B \subset A \leftrightarrow A = B$ and (b) $A \subset B$ and $B \subset C \rightarrow A \subset C$.

Definition | The cardinality of a set A (written $|A|$) is the number of distinct elements in A.

Example (Tay-Sachs) $|S| = 32$ (see p. 2)

Definition The set of all subsets of a given set S is called the <u>power set</u> of S, denoted by 2^S; this notation was chosen because, if $|S| = n$, then $|2^S| = 2^n$ (in other words, if S has n distinct elements then there are 2^n distinct subsets of S.

Definition It's convenient to have a symbol for the set that has no elements in it: $\emptyset$, the <u>empty set</u>.

Example If $S = \{a, b, c\}$ then

$|S| = 3$ and the power set has $2^3 = 8$ sets in it.

$\emptyset$ 1
$\{a\}$
$\{b\}$ 3
$\{c\}$
$\{a, b\}$
$\{a, c\}$ 3
$\{b, c\}$
$\{a, b, c\} = S$ 1

Given any set (sample space) S, Kolmogorov (1933) wanted to be able to define probabilities in a logically-internally-consistent manner (in other words, free from contradictions or paradoxes) to all of the sets in 2^S.

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 1 & & 1 & & \\ & 1 & & 2 & & 1 & \\ 1 & & 3 & & 3 & & 1 \end{array}$$

Pascal's triangle

If $|S|$ is finite, it turns out that nothing nasty can happen.

But if $|S|$ is infinite, nasty things can unfortunately happen.

Definition

A set with an infinite number of distinct elements is called an infinite set.

(4 Apr 19)

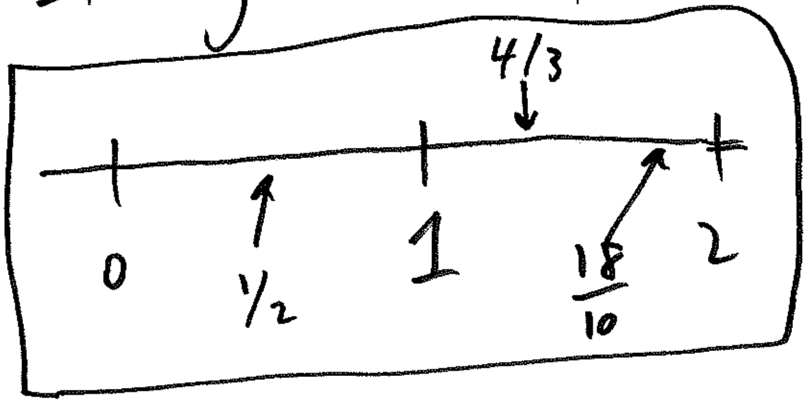
Definition

If the elements of an infinite set A can be placed in 1-to-1 correspondence with the positive integers $\mathbb{N} = \{1, 2, 3, \dots\}$, A is said to be countably infinite.

Example

The rational numbers are those real numbers that can be expressed as ratios of integers (ex. $\frac{1}{2}$, $\frac{14}{13}$, $-\frac{89}{2121}$, ...)

It might seem that there are a lot more



rational numbers than integers, but Cantor (1878) showed that the rational numbers are *countable*. He also showed something even more surprising: the number of distinct values on the real number line is an *order of infinity greater* than the number of integers or rationals.

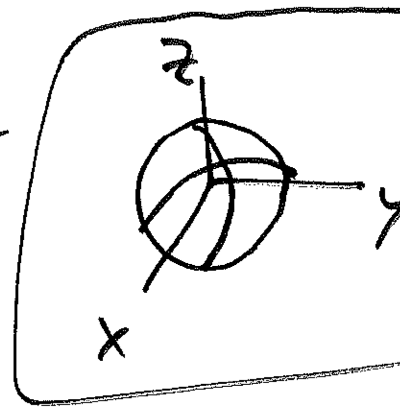
Definition An infinite set that is not countable is called *uncountable*.

Example $\mathbb{N} = \{1, 2, 3, \ldots\}$ is countable, but $\mathbb{R} = \{\text{all real numbers}\}$ is uncountable.

The mathematical foundation Kolmogorov chose for his development of probability theory is a part of mathematics called <u>measure theory</u>: an attempt to make rigorous the informal concepts of <u>length</u>, <u>area</u> and <u>volume</u> introduced by ancient Greek mathematicians including Euclid (about 2,300 years ago) and Pythagoras (about 2,500 years ago). However, people discovered in the early 1900s that infinity is a weird thing when you try to make an idea like volume of a sphere in 3-dimensional space rigorous.

Theorem (Banach–Tarski paradox (1924))

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 ("pieces"), 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 $\mathbb{N}$ (countable) or the real numbers $\mathbb{R}$ (uncountable). Because of weird results like the Banach-Tarski paradox, Kolmogorov found that when S is infinite, the set 2^S 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 <u>smaller</u> collection of subsets of S than 2^S in which nothing weird can happen. (see p. 7 of DS). The sets in this smaller collection $\mathcal{C}$ have to satisfy 3 simple rules to avoid the weirdness.

Rule 1: $\mathbb{C}$ includes the entire sample space S.

Rule 2: If an event A is in $\mathbb{C}$ 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 together into one (potentially bigger) set (and discarding any and all duplicates).

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

$\triangleq$ is defined to be

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

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

$$(A_1 \cup A_2 \cup \ldots) \triangleq \bigcup_{i=1}^{\infty} A_i$$

Rule 3:

If $A_1, A_2, \ldots$ are all in $\mathcal{C}$ then so is $\bigcup_{i=1}^{\infty} A_i$.

Example

whenever $|S| < \infty$ we can take $\mathcal{C} = 2^S$ with no weirdness arising; in other words, if the sample space S is finite,

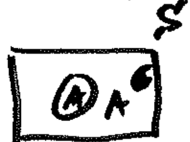
we can meaningfully assign probabilities

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) $A \cup B = B \cup A$,

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

$A \cup S = S$, and if $A \subset B$ 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)

(16)

Definition | with A and B any two sets, the intersection $A \cap B$ is the set containing all, and only, those elements belonging both to A and to B.

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

If A is an event (a set: a subset of S), we can equivalently talk either about the set A or the true/false proposition that one of the elements in A is the outcome of the experiment $\mathcal{E}$.

Example (T-s disclosure) $A = \{NNNNN\}$

as a set is equivalent to the true/false proposition (exactly 0 T-s lobbying) being true.

Even more basic facts about sets

(5) It's meaningful to talk about the intersection of more than 2 sets: with $A_1, \ldots, A_n$ the set $A_1 \cap \ldots \cap A_n = \bigcap_{i=1}^{n} A_i$ is meaningful, and with $A_1, A_2, \ldots$ 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 <u>disjoint</u> $\triangleq$ <u>mutually exclusive</u> if (no overlap)

$A \cap B = \phi$ (if they have no outcomes in common). n sets $A_1, \dots, A_n$ are disjoint if all distinct pairs are disjoint: $A_i \cap A_j = \phi$ for $i \neq j$.

logic equivalent propositions A, B mutually exclusive $\leftrightarrow$ they cannot both be true simultaneously

Example (T-S disease)

(Exactly 1 TS baby), (Exactly 2 T-S babies) are mutually exclusive.

Still more basic facts about sets

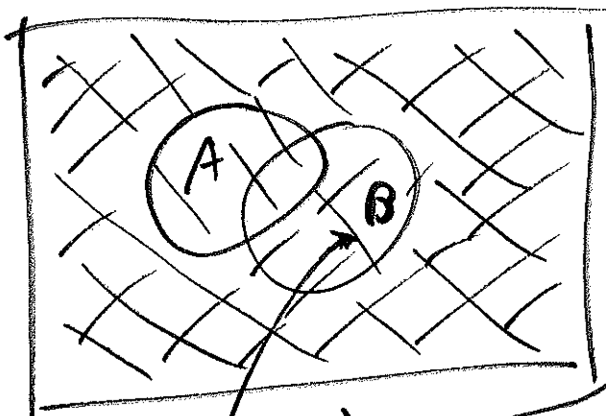
(7) (attributed to Augustus de Morgan (1806 - 1871), a British logician):

De Morgan's laws

A, B any two sets:

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

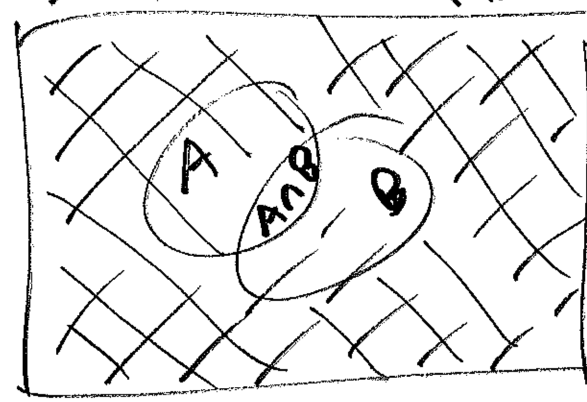
// A^c \\\\ B^c



(oops; wrong)

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

// A^c \\\\ B^c



logical restatement:

$\boxed{\text{or}}$

(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.

$\boxed{\text{and}}$

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

(8) A, B, C any sets:

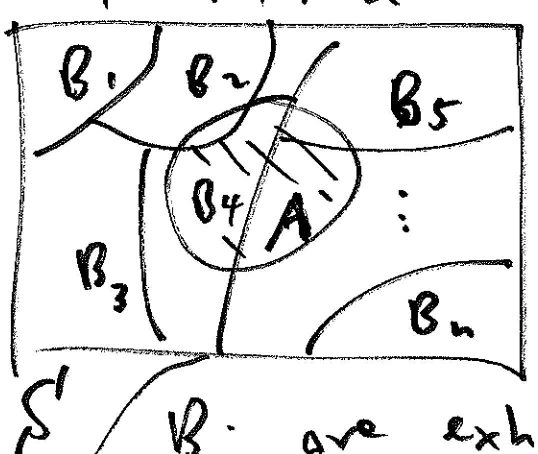
(a)
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

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 every outcome in S lives inside one, and only one, of the partition sets.

If you look at the Venn diagram on p. 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 probability Axioms

Kolmogorov wants to define $P_K(A)$ – what Axioms should he use?

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

$\boxed{\text{Axiom 1:}}$

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

$\boxed{\text{Axiom 2:}}$ $P_K(S) = 1$ (again motivated by relative frequency)

$\boxed{\text{Axiom 3:}}$ For every countable collection of disjoint events $A_1, A_2, \ldots \in \mathbb{C}$,

$$P_K\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P_K(A_i) \;\circledast$$

disjoint

(this countable additivity 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

$P_K\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P_K(A_i)$ follows from $\circledast$

— but also for a countable collection.

(9 Apr 19)

Consequences that follow from Kolmogorov's Axioms

(From now on I'll drop the subscript K.) (Kolmogorov)

① $P(\emptyset) = 0$.

DS: Pr

p

(2) $P(A^c) = 1 - P(A)$.

(3) If $A \subset B$ then $P(A) \le P(B)$.

(4) For all events A,

$0 \le P(A) \le 1$ (the easy rule)

(5) For all events A, B, (general addition rule for $\boxed{\text{or}}$)

$$P(\underset{\substack{\uparrow \\ \text{or}}}{A \cup B}) = P(A) + P(B) - P(\underset{\substack{\uparrow \\ \text{and}}}{A \cap B}).$$

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

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

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

useful in statistics

Tay-Sachs disease in more detail

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

$\boxed{\text{\# of T-S babies} = Y}$ Let's see if we can work out $P(Y=1)$, $P(Y=2)$, …, $P(Y=5)$; we already worked out

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

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

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

$$\overset{\text{identical distribution}}{=} \left[1 - P\left(\begin{matrix}\text{1st}\\\text{baby}\\\text{T-S}\end{matrix}\right)\right] \cdot \dots \left[1 - P\left(\begin{matrix}\text{5th}\\\text{baby}\\\text{T-S}\end{matrix}\right)\right] = (1-p)^5 = 1\, p^0 (1-p)^5 \doteq 24\%$$

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

A similar line of reasoning gives

$P(Y=5) = P(TTTTT) = p^5 = 1\,p^5(1-p)^0$

$\boxed{\text{What about } P(Y=1)?}$ 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 $P(Y=1) = 5\,p^1(1-p)^4$.

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

The outcomes with $(Y=3)$ are mirror images of those with $(Y=2)$: $\left\{\begin{matrix} TTNNN \\ NNTTT \end{matrix}\right\}$

so there must also be 10 elements of S with $(Y = 3)$ and $p(Y = 3) = 10\, p^3 (1-p)^2$.

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

# of T-S babies y	$p(Y = y)$	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

upper case

Soon we'll call Y a <u>random variable</u> (symbolizing the data generating process) and

lower case

use y to stand for a possible value of Y.

$$
\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1
\end{array}
$$

So it looks like

n = # children

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

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

Permutations & Combinations

$\boxed{\text{Ex.}}$ 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: $\underset{\diamond}{\overset{8}{}}$ _ _ _ _ (8 of diamonds)

The first slot can be filled in $n=52$ ways, and the second in $(n-1)=51$ ways, ..., the 5^{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 <u>permutations</u> of 52 things taken 5 at a time.

Definition The number of permutations of n distinct things taken k at a time is written $\underline{P}_{n,k} = n(n-1)\cdots(n-k+1)$.

capital

How many possible orderings of a 52-card deck are there? Now there are 52 slots, e.g., $\underset{52}{\underline{\text{J of }\blacklozenge}}\ \cdot\ \underset{51}{\underline{\text{3 of }\heartsuit}}\ \cdots\cdots\ \underset{1}{\underline{\text{A of }\clubsuit}}$, so the total number must be $52 \cdot 51 \cdot \cdots \cdot 1 =$

$$\boxed{\text{Def.}\quad n(n-1)\cdots 1 \overset{\Delta}{=} n! \quad \text{read } n \text{ factorial}}$$

$= 80658175170943878571660636856403766975289$
$505440883277824000000000000 \doteq 8.1 \cdot 10^{67}$

free
wolfram alpha

8
(maple) (11 Apr 19)

with this notation you can see that

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

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 $k=1$ T-S baby, i.e., $\overset{k}{1}$" T and $\overset{(n-k)}{4}$" Ns. 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 $n! = 5! = 120$ ways to arrange them in order left to right, e.g. $\underline{N_3}\ \underline{T_1}\ \underline{N_4}\ \underline{N_1}\ \underline{N_2}$. Now take the subscripts away: there are $4!$ ways to rearrange the Ns among themselves and $1! = 1$ ways to "rearrange" the Ts among themselves, so $5!$ is way too big and needs to be divided by $4! \cdot 1!$:

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

Definition: Given a set with n distinct elements, each distinct subset of size k is called a <u>combination</u> of elements, and there are $C_{n,k} = \frac{n!}{k!\,(n-k)!}$ ways to do this

Notation Everybody in the world other than DeGroot & 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

binomial coefficient

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

# of T-s babies

valid for all $n \ge 1$ and $0 \le p \le 1$.

$y = 0, 1, \ldots, n$

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

Case study: the birthday problem (34) (extra notes)

(A)

P(at least 2 people registered for AMS 131 this term have the same birthday) = ?

(day & month of the year not counting birth year)

Simplifying assumptions:

(1) birth rate constant from 1 Jan to 31 Dec; (2) Feb 29 → randomize to another day

Let k = # people registered for AMS 131 = 93 as of 29 Jul 2016, and (132) (141) (242) (2 Aug 2017) (29 Jul 18) (16 Apr 19)

let $n = 365$ = # possible birthdays. Building

the sample space S is like filling in k slots, each of which has n possible values, (birthdates)

so S contains n^k equally likely outcomes.

Turns out to be hard to count the number of those outcomes that make A true, so let's try to work out $P(\text{not } A)$:

If nobody has the same birthday, then a randomly chosen person 1 has $n = 365$ possibilities, a randomly chosen person 2 (distinct from person 1) has $(n-1) = 364$ possibilities, ..., and finally the last person $\overset{(242)}{k}$ (no longer random) has $(n-k+1)$ $= \overset{(124)}{273}$ possibilities, so all together (not A) has $n(n-1)\cdots(n-k+1) = P_{n,k} = \frac{n!}{(n-k)!}$ equally likely outcomes favorable to it

and $P(A) = 1 - P(\text{not } A) = 1 - \frac{365!}{\underset{\uparrow\,(123)}{272!} \cdot 365^{\overset{93}{\uparrow 242}}}$

$= 1 - \frac{n!}{(n-k)! \cdot n^k} \doteq ?$

this number is hard to compute with an ordinary pocket calculator; for example, $365! \doteq 2.5 \cdot 10^{778}$; so we need to be a bit clever. 36

$\boxed{\text{Three methods:}}$

① Don't evaluate numerator & denominator separately & then divide; both are ginormous. Instead, cancel them against each other:

$$1 - \frac{365!}{272! \, 365^{93}} = 1 - \frac{(365)(364) \cdots (273)}{(365)(365) \cdots (365)}$$

$\doteq 0.999997$

② Stirling's approximation:

$$\log n! \doteq \frac{1}{2} \log 2\pi + \left(n + \frac{1}{2}\right) \log n - n$$

(attributed to James Stirling (1692–1770), Scottish, but first stated by Abraham de Moivre (1667–1754)) French/English so $P(A) = 1 - \exp\left\{ \log\left[\frac{n!}{(n-k)!\,n^k} \right] \right\}$

stirling + simplification

for any $x > 0$, $x = \exp[\log(x)]$

$$= 1 - \exp\left\{ \left(n - k + \tfrac{1}{2}\right)\left[\log(n) - \log(n-k)\right] - k \right\}$$

$$\doteq 0.9999974.$$

$\Gamma(x)$

(3) The Gamma function is a generalization of $n!$, n integer, to all positive real numbers: $n! = \Gamma(n+1)$.

Many mathematical packages (R, matlab, ...) have a log-gamma function built-in.

$$P(A) = 1 - \exp\left[\log n! - \log(n-k)! - k \log n\right]$$

$$= 0.9999974$$

$$= 1 - \exp\left[\log \Gamma(n+1) - \log \Gamma(n-k+1) - k \log n\right].$$

You can play around with $P(A)$ as a function of k for fixed $n = 365$ & find that $P(A) > 0.5$ for $k \geq 23$, which many people find surprisingly low.

(2 Aug 17)

Generalizing the binomial coefficients

(p.33) what if there are more than 2 possible outcomes $\binom{n}{y}$

in a generalization of the Tay-Sachs case study (T, N)? T-S baby, not T-S baby

we want n distinct elements to be divided into k different groups $(k \geq 2)$ so that n_j elements fall into group j, $\sum_{j=1}^{k} n_j = n$;

Q In how many different ways can this be done? Follow the argument in DS pp. 42-43, which generalizes the line of reasoning leading to the <u>bi</u>nomial coefficients $\binom{n}{y}$ when $k = 2$:

$$\frac{n!}{y!\,(n-y)!}$$

Definition: A <u>multi</u>nomial coefficient is of the form

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1!\, n_2! \cdots n_k!}$$

$n \geq 1$
$k \geq 2$
$1 \leq n_j \leq k$
$\sum_{j=1}^{k} n_j = n$

This answers the how many different ways question Q above.

Example: (2016 presidential election) (see DS pp. 333-334) (with replacement)

Imagine randomly sampling n eligible prospective voters from all such people in the U.S. "the population" (*) ~~[illegible]~~ (~~[illegible]~~);

possible outcome (k=5)
Clinton (Democrat)
Trump (Republican)
Johnson (Libertarian)
Stein (Green)
Undecided / no comment / other candidate

let X_i = # people in sample who say they will vote for candidate i, $i = 1, \dots, k = 5$ as in this table.

Suppose (unknown to us) that the proportion of voters who favor candidate i in the population (*) above is p_i, where $0 < p_i < 1$ and $\sum_{i=1}^{k} p_i = 1$.

Because the people are chosen with independent identically distributed (IID) sampling (i.e., at random with replacement), each person's outcome will be independent of all the other outcomes. Thus

P(1st person favors candidate i_1, (&) and 2nd person favors i_2, (&) ..., (&) person n favors i_n) =

$p_{i_1} p_{i_2} \cdots p_{i_n}$. Therefore P(the sample, listed in a prespecified order, has x_1 people favoring candidate 1, x_2 people choosing candidate 2, ..., x_k people favoring candidate k) = $p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$, with $0 \le x_i \le n$ and $\sum_{i=1}^{k} x_i = n$. Thus

$P($exactly x_1 people favor Clinton, ..., x_k people "favor" Undecided$) = \boxed{?}\, p_1^{x_1} \cdots p_k^{x_k}$,

where $\boxed{?}$ is the total # of different ways the order of the n people in the sample can be listed. But this $\boxed{?}$ is precisely

$$\binom{n}{x_1, \ldots, x_k} = \frac{n!}{x_1!\, x_2! \cdots x_k!},$$

the multinomial coefficient defined on p. 39 above. Thus

$$P(X_1 = x_1, \ldots, X_n = x_n) \overset{(*)\cdots(*)}{=} \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}.$$

Later in the course we'll refer to this as the multinomial (probability) distribution

(16 Apr 19)

How to work with $\boxed{OR}$ when you have more than 2 events (union) $\updownarrow$ $\cup$

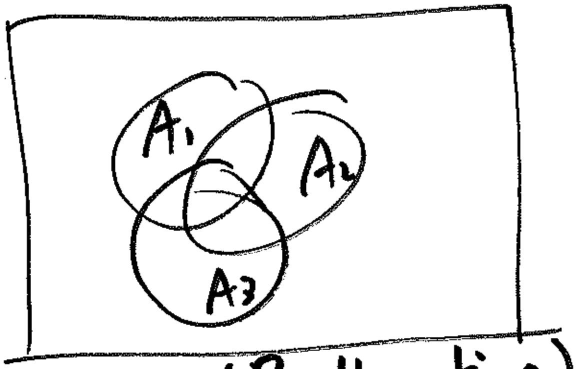
We already worked out that

$$P(A_1 \text{ or } A_2) = P(A_1 \cup A_2)$$

$$= P(A_1) + P(A_2) - P(A_1 \overset{\text{and}}{\cap} A_2).$$

We also know from Kolmogorov's 3rd Axiom that if events $A_1, \dots, A_n$ are disjoint then $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$.

How do these 2 things generalize?



(Ballantine)

By (tedious) enumeration you can show that with 3 events,

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$$

$$- \left[P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3)\right]$$

$$+ P(A_1 \cap A_2 \cap A_3).$$

You probably can now (or guess) see how this generalizes: for any events $A_1, \ldots, A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j)$$

$$+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n+1} P(A_1 \cap \ldots \cap A_n).$$

Example Get 2 decks of ordinary playing cards; order deck 1 from (1 to 52) using any sequence you like, e.g.

1 = 2♣

⋮

13 = A♣

14 = 2♦

⋮

26 = A♦

27 = 2♥

⋮

39 = A♥

40 = 2♠

⋮

52 = A♠

Shuffle deck 2 until (practically speaking) all 52! orderings are equally likely.

Now turn the first card of each deck over; do they match? Continue through all 52 cards; P(at least one match) = ?

let $n = 52$

Let A_i = (a match occurs on card i), we want $P\left(\bigcup_{i=1}^{n} A_i \right)$, which can be computed with the complicated formula on the previous page.

Follow the logic detailed on DS pp. 49-50 to obtain

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n+1}\frac{1}{n!}$$

wolfram alpha

limit (sum (-1)^(i+1)/i!, i = 1 to n) as n -> infinity

calculus result: $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{(-1)^{i+1}}{i!} = 1 - \frac{1}{e} \doteq 0.63$

This sum approaches its limit quickly; already with $n = 7$ you have the first 4 significant figures: 0.6321

DS ch.2

Conditional probability

Note that Kolmogorov's probability axioms defined the function $P_K(A)$, where A is a set in the collection $\mathbb{C}$ of subsets of the sample space S in which nothing weird can occur; in other words, $P_K(A)$ is a function of a single argument A.

To include the extremely useful idea of conditional probability in his setup, Kolmogorov has to define it using P_K. $\boxed{\text{Definition}}$ Given any two events A, B in $\mathbb{C}$, the conditional probability of A given B is

$$P(A \mid B) = \begin{cases} \dfrac{P(A \cap B)}{P(B)} & \text{if } P(B) > 0 \\ \text{undefined} & \text{if } P(B) = 0 \end{cases}.$$