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 <u>all</u> 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 <u>finite</u>, it turns out that nothing nasty can happen.

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

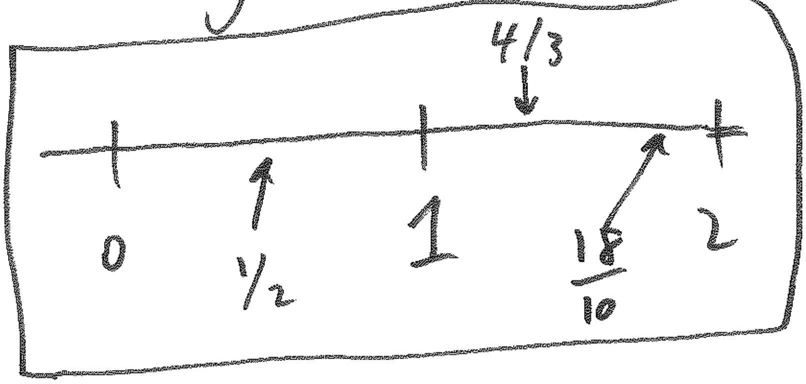
Definition

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

Definition If the elements of an infinite set A can be placed in 1-to-1 correspondence with the positive integers $\mathbb{N} = \{1, 2, 3, \ldots\}$, 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 <u>countable</u>. He also showed something even more surprising: the number of distinct values on the real number line is an <u>order of infinity greater</u> than the number of integers or rationals.

$\boxed{\text{Definition}}$ An infinite set that is not countable is called <u>uncountable</u>.

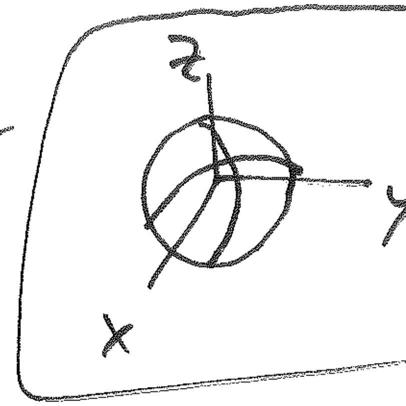
$\boxed{\text{Example}}$ $\mathbb{N} = \{1, 2, 3, \ldots\}$ is countable, but $\mathbb{R} = \{\text{all real numbers}\}$ is uncountable.

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 $\mathcal{S}$ is infinite, the set $2^{\mathcal{S}}$ of all subsets of $\mathcal{S}$ is "too big" and "too strange" to permit the assignment of probabilities to all the sets in $2^{\mathcal{S}}$ in a logically-internally-consistent way. When $\mathcal{S}$ is infinite, Kolmogorov was forced to restrict attention to a smaller collection of subsets of $\mathcal{S}$ than $2^{\mathcal{S}}$ 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.

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

<u>Rule 2:</u> If an event A is in $\mathbb{C}$ then so is its complement A^c.

<u>Rule 3 requires a</u> $\boxed{\text{Definition}}$ Given any two sets A and B, the <u>union</u> 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 (14)

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

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

$A \cup S = S$, 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 <u>associativity</u> of the $\cup$ operation)

Definition with A and B any (16)

two sets, the <u>intersection</u> $A \cap B$ 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) $A = \{NNNNN\}$ (17)

as a set is equivalent to the true/false proposition (exactly 0 T-S lolling) ~~~~ 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

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

in common). n sets $A_1, \ldots, A_n$

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)

$\begin{pmatrix} \text{Exactly} \\ \text{1 T-S} \\ \text{baby} \end{pmatrix}$, $\begin{pmatrix} \text{Exactly} \\ \text{2 T-S} \\ \text{babies} \end{pmatrix}$ are mutually exclusive.

Still more basic facts about sets

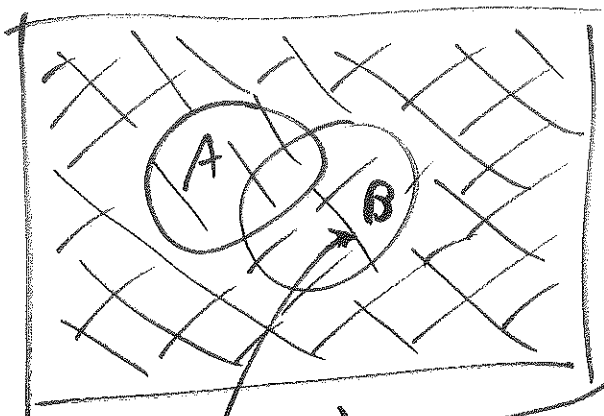
(7) (attributed to Augustus de Morgan (1806 - 1871), a British logician):

De Morgan's laws

A, B any two sets:

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

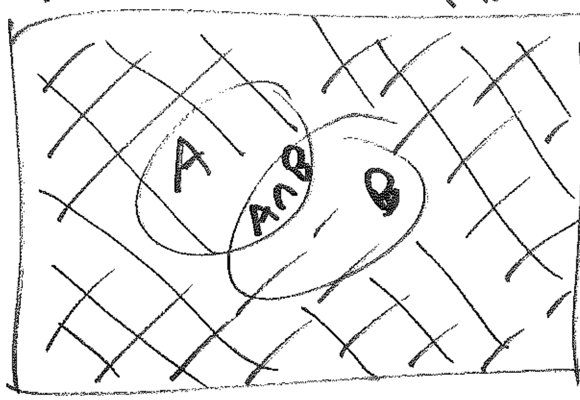
// A^c \\\\ B^c



(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:

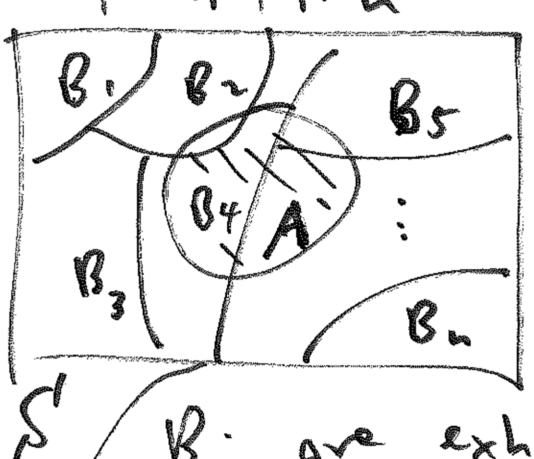
(a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

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



$B_1, \ldots, B_n$ 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 $(B_1, \ldots, B_n)$ 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)

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

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 be use?

It was clear to Kolmogorov that $P_K(A)$ 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 he assume about P_K?

$\boxed{\text{Axiom 1:}}$ For all events $A \in \mathcal{C}$, $P_K(A) \geq 0$ (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 $A_1, A_2, \ldots \in \mathcal{C}$,

$$P_K\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P_K(A_i) \quad (*)$$

disjoint

(this 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 $A_1, \ldots, A_n$ are disjoint then

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

— but also for a countable collection.

Consequences that follow from Kolmogorov's Axioms

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

① $P(\phi) = 0$.

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 $A_1, A_2, \ldots, A_n$,

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

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

useful in statistics

Tay-Sachs disease in more detail

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

$\boxed{\#\text{ of T-S babies} = Y}$ Let's see if we can work out

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

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

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

independence

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

identical distribution

$$= \left[1 - P\begin{pmatrix}\text{1st baby} \\ \text{T-S}\end{pmatrix}\right] \cdot \dots \cdot \left[1 - P\begin{pmatrix}\text{5th baby} \\ \text{T-S}\end{pmatrix}\right] = (1-p)^5 = 1\, p^0 (1-p)^5 \doteq 24\% \text{ with } p = \frac{1}{4}$$

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-S 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

$\boxed{\text{upper case}}$

Soon we'll call Y a random variable (symbolizing the data generating process) and $\boxed{\text{lower case}}$ use y to stand for a possible value of Y.

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

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 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?

(8 of diamonds)

It's like filling in 5 slots: $\underset{\blacklozenge}{8}$ _ _ _ _

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

capital

How many possible orderings of a 52-card deck are there? Now there are 52 slots, e.g.,

$$\begin{array}{cccc} J \text{ of } \blacklozenge & 3 \text{ of } \heartsuit & \cdots & A \text{ of } \clubsuit \\ \hline 52 \cdot & 51 & \cdots\cdots & 1 \end{array}$$

, so the total number must be $52 \cdot 51 \cdot \cdots \cdot 1 =$ $\boxed{\text{Def.}}$

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

$= 80658175170943878571660636856403766975289505440883277824000000000000 \doteq 8.1 \cdot 10^{67}$

with their notation you can see that 31

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

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 $\begin{Bmatrix} T_1 \\ N_1\ N_2\ N_3\ N_4 \end{Bmatrix}$.

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 $4! \cdot 1!$:

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

$\boxed{\text{Definition}}$ 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 is

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

binomial coefficient

# of T-s believers

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.

Case study: the birthday problem

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 7 = 93 (132) as of 29 Jul 2016 (2 Aug 2017), and

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 k (no longer random) has $(n-k+1) = 273$ 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 $P(A) = 1 - P(\text{not } A) = 1 - \frac{365!}{272! \cdot 365^{93}}$

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

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.

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

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

$$\doteq 0.9999974.$$

$\boxed{\Gamma(x)}$

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: T-S baby; N: 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, $\sum_{j=1}^{k} n_j = n$;