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 $\overset{(n-k)}{4}$ 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 $n! = 5! = 120$ 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

binomial coefficient

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

\# of T-S babies

valid for all $n \geq 1$ and $y = 0, 1, \ldots, n$, $0 \leq p \leq 1$.

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

Case study: the birthday problem (84) (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 → ~~[illegible]~~ 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, (birthdates)

each of which has n possible values,

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 $P(A) = 1 - P(\text{not } A) = 1 - \frac{365!}{272! \cdot 365^{93}}$

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

$\uparrow$ (123) $\quad \uparrow$ 242

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

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: 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, $\sum_{j=1}^{k} n_j = n$;

[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 **bi**nomial coefficients $\binom{n}{y}$ when $k=2$:

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

Definition: A **multi**nomial coefficient is of the form

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

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

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 US. ("the population" *);

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(1^{st} person favors candidate i_1, (&) and 2^{nd} 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 (1), ..., (4)

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

$$P(X_1 = x_1, \ldots, X_n = x_n) \overset{(1)\ldots(4)}{=} \frac{n!}{x_1! \ldots, x_k!}\, p_1^{x_1} \cdots p_k^{x_k}.$$

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?