175

L5 Compositions 2.8.1 HMM

Q. What is the number of ways to write 5 as a sum of three positive numbers, where order matters?

A The addends available to us are $\{1, 2, 3, 4\}$.

Brute-force enumeration

$1+1+3$	$2+1+2$	$3+1+1$
$1+2+2$	$2+2+1$	
$1+3+1$		

Each of those is said to be a "composition" of the number 5.

To develop a formula for the solution, you might be tempted to count # possibilities for first addend and multiply by # possibilities for second addend, etc.

The problem is that # 2-nd place addends depends on the identity of the 1st-place addend

Instead of grouping by identity of first addend, let's group by identity of all addends:

$\{1, 1, 3\}$ $\quad$ $\{1, 2, 2\}$ $\quad$ "partitions" of 5.

$1+1+3$	$1+2+2$
$1+3+1$	$2+1+2$
$3+1+1$.	$2+2+1$

Yet we still don't see a convenient way to count compositions.

The trick is to break everything into 1's:

e.g. $1+1+3 = \langle 1, 1, 1+1+1 \rangle$

$1+2+2 = \langle 1, 1+1, 1+1 \rangle$

Now if we squint at the right-hand sides, we realize that all compositions of 5 into 3 parts are obtained by filling 4 slots w/ 2 commas (putting "+"s in the remainder):

$1+1+3 = 1 \underline{,} 1 \underline{,} 1 \underline{+} 1 \underline{+} 1$

$1+2+2 = 1 \underline{,} 1 \underline{+} 1 \underline{,} 1 \underline{+} 1$

Thus # compositions is the # ways of choosing 2 slots (to place the commas in) from a total of 4 slots.

You'll notice that this is essentially the "stars and bars" approach where the stars are "+"s and the bars are ","s.

Since the # slots is 1 less than 5 and # "bars" is 1 less than # terms in composition, we have the general result:

Number of compositions of n into k parts is

$$\binom{n-1}{k-1}.$$

What about the number of compositions of n regardless of the # parts? Use the stars and bars approach! Every composition takes the form

$$\overbrace{1 \_ 1 \_ \cdots \_ 1 \_ 1}^{n}$$

where there are $n-1$ slots each of which is filled by either a "+" or a comma.

More concretely:

- each assignment of "+"'s and commas determines a unique composition
- each composition determines a unique assignment of "+"'s and commas.

Since there are $n-1$ independent binary choices to be made, there are 2^{n-1} compositions.

Thus it must be the case that

$$\sum_{k=1}^{n} \binom{n-1}{k-1} = 2^{n-1}.$$

which is true since

$$LHS = \sum_{m=0}^{n-1} \binom{n-1}{m} = 2^{n-1}.$$

because:

$$\sum_{m=0}^{\ell} \binom{\ell}{m} = 2^{\ell} \qquad \text{(cf L3)}$$

What if we now permit the integers in a composition to be 0? How many compositions into k parts are there of n?

EX $n=7,\ k=5$:

$$7 = 4+0+1+2+0.$$

Again breaking everything into 1^s, one encoding is:

$$\begin{array}{ccccc} **** & | & | * & | ** & | \\ 4 & 0 & 1 & 2 & 0 \end{array}$$

Again, there are $4+0+1+2+0 = 7$ stars and $5-1$ bars, making a total of $7+5-1$ symbols:

$$\underline{*}\ \underline{*}\ \underline{*}\ \underline{*}\ \underline{|}\ \underline{|}\ \underline{*}\ \underline{|}\ \underline{*}\ \underline{*}\ \underline{|}$$

Thus all we need to do is to count the # distinct ways of placing $5-1$ bars into $7+5-1$ slots, which is

$$\binom{7+5-1}{5-1}$$

In general, there are

$$\binom{n+k-1}{k-1}$$

"weak" compositions (those in which 0 is allowed) into k parts of an integer n.

§2.5 LVP

Q. Probability that a word of length k has no repeated letter? Assume letters come from an alphabet of size n. Assume all words are equally likely.

A. How many words are there in total?

$$\_\ \_\ \_\ \cdots\ \_\ \_$$

#choices: $\underbrace{n\ n\ n\ \ \ \ n\ n}_{k \text{ times}}$.

$\Rightarrow$ #words $= n^k$

How many of these have no repeated letter?

$$\_\ \_\ \_\ \cdots\ \_$$

eg $k=3$ ABC, BAC, $\vdots$

#choices: $n \quad n-1 \quad n-2 \quad\quad n-(k-1)$

$\Rightarrow$ #words w/ no repeats

$$= \begin{cases} n(n-1)\cdots(n-(k-1)) & k \leq n \\ 0 & k > n \end{cases}$$

$$= \begin{cases} n!/(n-k)! & k \leq n \\ 0 & k > n \end{cases}$$

Since all words are equally likely, the prob. that a randomly chosen word has no repeated letter is:

$$\frac{\#(\text{words w/ no repeats})}{\#\text{words}}$$

$$= \begin{cases} \dfrac{n!}{(n-k)!} \Big/ n^k & k \le n \\ 0 & k > n \end{cases}$$

Q. What is the smallest word length k s.t. words are more likely to contain a repeat than not? Assume $n = 26$.

A.

word length, k :	1	2	3	4	5	6	7
P(no repeat) :	1	$\frac{25}{26}$	.89	.79	.66	.53	.41

↑ P(repeat) > P(no repeat)

k :	...	n	$n+1$	$n+2$	...
P(no repeat) :	...	$\frac{n!}{n^n}$	0	0	...

Birthday Paradox

Q How many people need to be in room before there is more than a 50% chance of a shared birthday?

A To establish a connection with the previous counting problem, imagine asking for birthdays, one by one:

$\underline{\text{Jan 17}}$ $\underline{\text{Feb 9}}$ ... $\underline{\text{June 4}}$
1st person 2nd person 46th person

This is a word of length k (we don't specify the value of k yet) chosen from an alphabet of size $n = 365$.

Thus, from the previous problem,

P(k people don't share a birthday)

$$= \frac{365!}{(365-k)!} \Big/ 365^k$$

For $k = 23$, this is $\approx 49\%$.

This problem is a "paradox" because the # of people needed is surprisingly small compared to the # required to guarantee a shared birthday — 365 (by Pigeonhole principle).

Now, how do things change if birthdays are not equally likely (eg birthdates of professional hockey players). Turns out that this just increases the chance of birthday collisions. To see this, we need a new (probabilistic) approach:

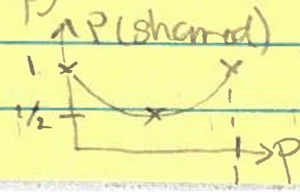
Q Two people. Two possible birthdays $\{A, B\}$. What is the prob. they share a birthday?

A. Suppose $P(A)$ = prob of being born on A $= p$

Then $P(B) = 1-p$.

birthdays:	AA	AB	BA	BB
prob:	p^2	$p(1-p)$	$(1-p)p$	$(1-p)^2$

$P(\text{shared b'day}) = p^2 + (1-p)^2$



Q. For what p is P(shared b'day) smallest?

A. $$\frac{d}{dp} P(\text{shared}) = 0.$$

$$\Rightarrow \quad 2p + 2(1-p)(-1) = 0.$$

$$\Rightarrow \quad p = 1/2.$$

Note:
$$\frac{d}{dp} P(\text{shared})$$
$$= 2[p-(1-p)]$$
$$= 2[2p-1]$$
$$\begin{cases} <0 & p<1/2 \\ 0 & p=1/2 \\ >0 & p>1/2 \end{cases}$$

Thus, when birthdays are not equally likely (as we initially assumed), then prob. of collisions only becomes larger. A consequence is that it is a really good bet that two people share the same birthday if there are only two possibilities for those birthdays! Please see §25 of LVP to see how a professor used this fact to make money over the course of his career!

Summary of L2-L5:

Four kinds of counting:

	sequences (order matters)	sets (order doesn't matter)
repetition allowed	n^k	$\binom{k+n-1}{n-1} = \binom{n-1+k}{k}$
no repetition	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$