MATH 323: PROBABILITY
SOME COMBINATORIAL CALCULATIONS

THE HYPERGEOMETRIC FORMULA

There are two alternate forms for the probability that results from a 'hypergeometric' sampling experiment; if a finite population of size N comprises R Type I objects and $N - R$ Type II objects, and a sample of size $n \leq N$ is obtained from the population, then the probability that the sample contains r Type I objects and $n - r$ Type II objects is either expressed

$$\frac{\binom{n}{r}\binom{N-n}{R-r}}{\binom{N}{R}} \qquad \text{or} \qquad \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

These two calculations originate from different methods of viewing the sampling mechanism:

- **Method A:** Consider choosing items in the population to label Type I and Type II, broken down by whether they are IN the sample or OUT of the sample. The denominator is the total number of ways of choosing R from N items to label Type I. The numerator is the number of ways of choosing r from n items IN the sample to be Type I items, multiplied by the number of ways of choosing the $R - r$ from $N - n$ items OUT of the sample to be Type I items.
- **Method B:** Consider items to be IN the sample, broken down by Type. The denominator is the total number of ways of choosing n items from N to be IN the sample. The numerator is the number of ways of choosing r Type I items from R to be IN the sample, multiplied by the number of ways of choosing the $n - r$ Type II items from $N - R$ to be IN the sample.

THE BIRTHDAY PROBLEM

The *Birthday Problem* can be stated as follows: in a class of N pupils, what is the probability that no two pupils share the same birthday ? Assuming that birthdays are uniformly spread across the 365 days of the year, then the total number of possible ways of selecting N birthdays is, using the multiplication principle,

$$n_S = 365 \times 365 \times 365 \times \cdots \times 365 = 365^N.$$

For those birthdays to all be different, we must sample birthdays from the set of days without replacement N times, and the number of ways of doing that is

$$n_A = 365 \times 364 \times 363 \times \cdots \times (365 - N + 1) = P_N^{365}$$

$$P(A) = \frac{n_A}{n_S} = \frac{365 \times 364 \times \ldots \times (365 - N + 1)}{365 \times 365 \times \ldots \times 365} = \frac{365!}{(365 - N)!}\frac{1}{365^N} = \frac{P_N^{365}}{365^N}$$

We can equivalently write this as

$$P(A) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{365 - N + 1}{365} = 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{N-1}{365}\right)$$

This probability decreases remarkably rapidly:

N	5	10	15	20	25	30	35	40	450	50
Prob.	0.9729	0.8831	0.7471	0.5886	0.4313	0.2937	0.1856	0.1088	0.059	0.0296

Thus, before N reaches 25, the probability that all the birthdays are distinct drops below 0.5 (see Figure 1).

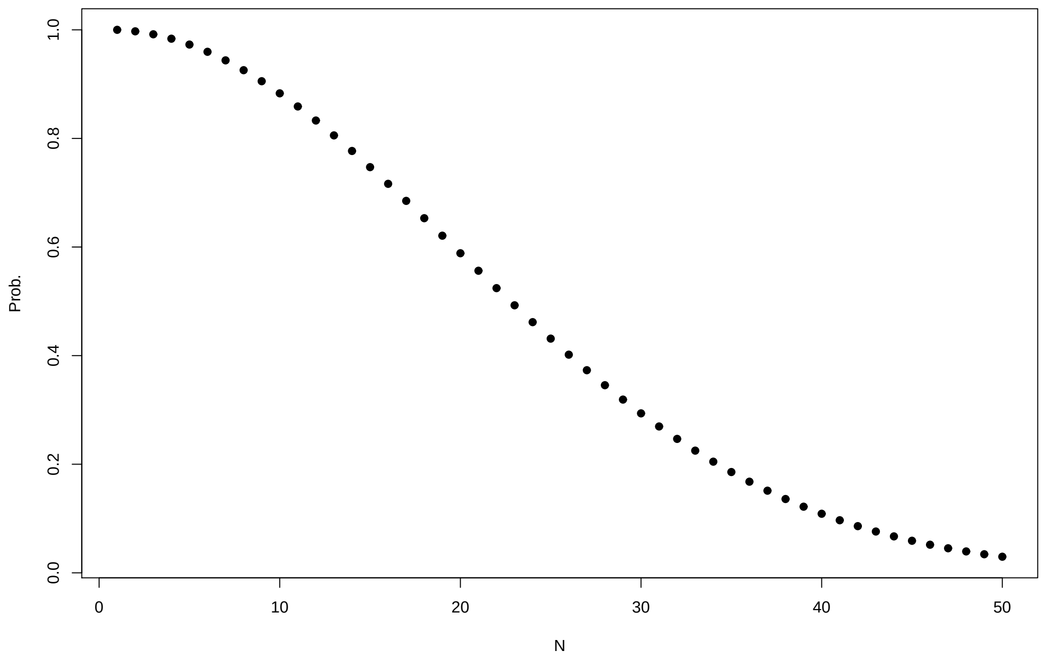

Figure 1: The probability that all N birthdays selected are distinct for $N = 1, \ldots, 50$

Occupancy Problems

Consider the combinatorial problem of allocating r items (objects, balls) to n boxes (cells): for example, can we enumerate *the number of ways of allocating 6 items to 4 boxes*.

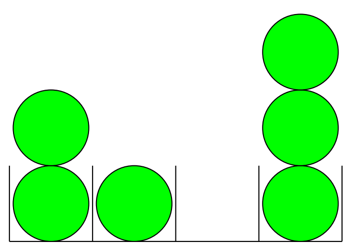

Figure 2: Allocating 6 balls to 4 boxes

To count the possible number of allocations, we consider the cases of *DISTINGUISHABLE* and *INDISTINGUISHABLE* items separately.

- **DISTINGUISHABLE ITEMS**

	ITEM	1	2	3	4	5	6
ALLOCATION 1	BOX LABEL SEQUENCE	2	4	4	1	4	1
ALLOCATION 2	BOX LABEL SEQUENCE	1	2	4	1	4	4

  and, essentially, we have selected r box labels from n **with replacement**, where the box labels are **ordered**. Hence the number of possible allocations is n^r, by a previous result using the multiplication theorem.

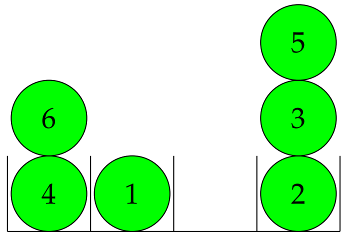

Figure 3: Allocating 6 distinguishable balls to 4 boxes: Allocation 1

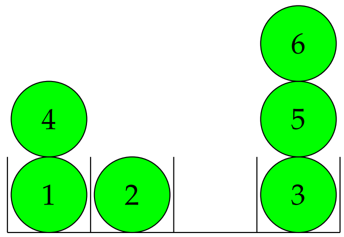

Figure 4: Allocating 6 distinguishable balls to 4 boxes: Allocation 2

If we require

$$r_1 \text{ items in BOX 1}, r_2 \text{ items in BOX 2}, \ldots, r_n \text{ items in BOX } n$$

then we must **partition** the box label sequence to contain r_1 1s, r_2 2s,...,r_n ns. Hence the number of possible allocations is given by the partition formula

$$\frac{r!}{r_1!r_2!...r_n!} \qquad \text{where} \qquad \sum_{i=1}^{n} r_i = r$$

- **INDISTINGUISHABLE ITEMS**

  If the items are indistinguishable, that is, here completely identical, and we wish to consider **distinct** allocation patterns, we must consider *unordered* arrangements; the two allocations in Figure 3 and Figure 4 are regarded as identical, and identical to the allocation in Figure 2, as the items are not labelled. For example, consider forming the allocation pattern by dropping the items into the boxes in sequence:

ITEM	1	2	3	4	5	6
SEQUENCE (1)	2	1	2	4	1	4
SEQUENCE (2)	3	2	4	1	3	3
SEQUENCE (3)	2	4	4	1	1	2

Then, we have that the patterns obtained by sequences (1) and (3) are both of the form

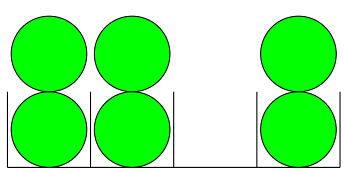

Figure 5: Allocation patterns for sequences (1) and (3)

which is distinct from the pattern for (1) and (3).

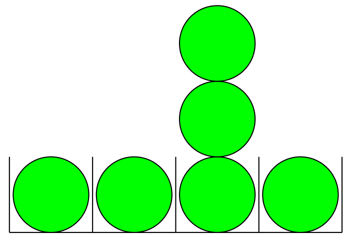

Figure 6: Allocation pattern for sequence (2)

To enumerate the number of possible allocation patterns, we utilize a binary sequence representation. We code an allocation pattern by reading from left to right, and writing a 1 for a box edge, and 0 for an item, so that the pattern in Figure 5 is coded

$$1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 1$$

and the pattern in Figure 6 is coded

$$1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1$$

The number of possible allocation patterns is equal to the number of binary sequences that correspond to them, and these sequences are composed as follows; they contain $n+1$ 1s (for the box edges) and r 0s (for the items), but also they begin with a 1, and end with a 1. The number of sequences like this is therefore equal to the number of ways of arranging a sequence of $(n+1)+r-2 = n+r-1$ binary digits containing precisely $n-1$ 1s and r 0s. This number is

$$\binom{n+r-1}{n-1} = \binom{n+r-1}{r}$$

from combination/binomial coefficient definition, and this is therefore the total number of distinct allocation patterns.

EXAMPLE 1 If r identical dice are rolled, with $n=6$ possible scores for each die, the total number of distinct score patterns is

$$\binom{n+r-1}{r} = \binom{6+r-1}{r} = \binom{5+r}{r}$$

for example, with $r=4$, we could have

DICE NUMBER				SCORE PATTERN					
1	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**5**	**6**
6	1	6	2	1	1	0	0	0	2
3	2	3	4	0	1	2	1	0	0
6	2	1	6	1	1	0	0	0	2

and the number of distinct patterns is $\binom{9}{4} = 126$.

EXAMPLE 2 Allocate n items to n boxes. Evaluate the probability of event A that no box is empty.

SOLUTION: Probability is

$$\mathrm{P}(A) = \frac{n_A}{n_S} = \frac{n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1}{n \times n \times n \times \ldots \times n \times n \times n} = \frac{n!}{n^n}$$

as we allocate by sampling n boxes **with** and **without** replacement for denominator and numerator respectively (each of which are a product of n terms).

EXAMPLE 3 Allocate r items to n boxes. Evaluate the probability of event A that no box is contains more than one item.

SOLUTION: Probability is

$$\mathrm{P}(A) = \frac{n_A}{n_S} = \frac{n \times (n-1) \times (n-2) \times ... \times (n-r+1)}{n \times n \times n \times ... \times n} = \frac{n!/(n-r)!}{n^r} = \frac{P_r^n}{n^r}$$

as we allocate by sampling r boxes **with** and **without** replacement for denominator and numerator respectively (each of which are a product of r terms).

NOTE: we can re-write P(A) using a conditional probability/chain rule argument corresponding to a sequence of selection probabilities:

$$\mathrm{P}(A) = 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times ... \left(1 - \frac{r-1}{n}\right)$$

where each term is the conditional probability of choosing a currently empty box, given the allocations at that instant.

Note: The Birthday Problem can be viewed as an Occupancy Problem: in a group of r people, what is the probability that no two people have the same birthday ? Assuming that all of the $n = 365$ days in the year are equally likely to be a birthday, then we identify in EXAMPLE 2 the "boxes" as days, and "items" as people, and evaluate the probability as

$$\frac{P_r^n}{n^r} = \frac{P_r^{365}}{365^r}.$$

EXAMPLE 4 Allocate r items to n boxes. Evaluate the probability of event A that box 1 contains precisely k items.

SOLUTION: For $0 \le k \le r$, probability is

$$\mathrm{P}(A) = \frac{n_A}{n_S} = \frac{\binom{r}{k}(n-1)^{r-k}}{n^r} = \binom{r}{k}\left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{r-k}$$

For the numerator, we first select the k **items** from r (**without** replacement) to place in box 1, and then select $(r-k)$ **boxes** from the remaining $(n-1)$ (**with** replacement) to house the remaining $(r-k)$ items. For the denominator, we merely select r boxes from n with replacement.

POKER HANDS

To compute the configuration of hands in poker, where five cards are dealt from a standard deck of 52 cards. In the probability, the denominator is always

$$n_S = \binom{52}{5}.$$

For the numerator, using generic notation let x, y be the "scoring" cards and a, b, c be the remaining ones. In calculating the probabilities, list the scoring denominations in descending order, then for each in turn multiply the number of remaining ways of choosing the denomination by the number of ways of choosing the suits for that denomination. For example, for Full House ($xxxyy$), you get

$$\binom{13}{1}\binom{4}{3}\times\binom{12}{1}\binom{4}{2}$$

In the case of ties in the list of scoring denominations (i.e. two x and two y in the hand), remember that you have to choose two **different** denominations so that for Two Pairs ($xxyya$) the number of ways of choosing the scoring cards is

$$\binom{13}{2}\binom{4}{2}\binom{4}{2}$$

Hand	Configuration	n_A	Prob.
ONE PAIR	$xxabc$	$\binom{13}{1}\binom{4}{2}\times\binom{12}{3}\binom{4}{1}\binom{4}{1}\binom{4}{1}$	0.42
TWO PAIRS	$xxyya$	$\binom{13}{2}\binom{4}{2}\binom{4}{2}\times\binom{11}{1}\binom{4}{1}$	0.048
THREE OF A KIND	$xxxab$	$\binom{13}{1}\binom{4}{3}\times\binom{12}{2}\binom{4}{1}\binom{4}{1}$	0.021
STRAIGHT	Run of five cards	$\binom{10}{1}\binom{4}{1}^5-\binom{10}{1}\binom{4}{1}$	0.0039
FLUSH	Five cards in same suit	$\binom{13}{5}\binom{4}{1}-\binom{10}{1}\binom{4}{1}$	0.0020
FULL HOUSE	$xxxyy$	$\binom{13}{1}\binom{4}{3}\times\binom{12}{1}\binom{4}{2}$	0.0014
FOUR OF A KIND	$xxxxa$	$\binom{13}{1}\binom{4}{4}\times\binom{12}{1}\binom{4}{1}$	0.00024
STRAIGHT FLUSH	Run of five cards in same suit	$10\binom{4}{1}-\binom{4}{1}$	0.000014
ROYAL FLUSH	AKQJ10 in any suit	$1\binom{4}{1}$	0.0000015

Table 1: Probabilities for different poker hands. For STRAIGHT and FLUSH, we must remember to subtract the STRAIGHT FLUSHes (which are counted separately); similarly, for STRAIGHT FLUSH, must remember to subtract ROYAL FLUSHes.