Lecture 9 Counting - basic principle, permutations, combinations & partitions

BIO210 Biostatistics

Xi Chen

Spring, 2022

School of Life Sciences
Southern University of Science and Technology

南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

Counting

Aims

- **Recap from the math class**
- **Be familiar with the notations**

Principles of counting

- Basic principles of counting
- permutations
- k-permutations
- combinations
- partitions

Discrete Uniform Law

- **Let all outcomes be equally likely**
- **Then**:

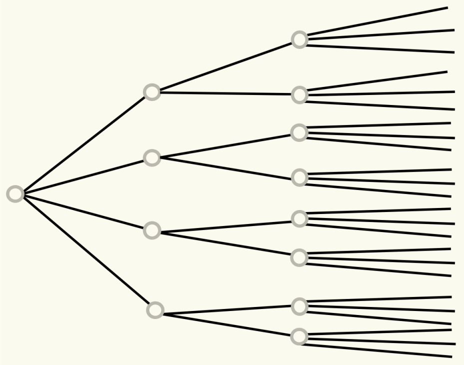
$$P(A) \ = \frac{\text{number of outcomes in } A}{\text{total number of outcomes}} = \frac{|A|}{|\Omega|}$$

$$|A|, |\Omega| : \text{cardinality of the set}$$

- All you need to do is: **counting** !

Basic counting principle

- **Basic scenario**: r stages, n_i choices at stage i



Number of choices is:

$$\prod_{i=1}^{r} n_i = n_1 \cdot n_2 \cdot n_3 \cdots n_r$$

1. **Permutations**: Number of ways of ordering n elements is: $n!$
2. Number of subsets of n elements: 2^n

Example: DNA hexamer

Letter	Base
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base

- DNA synthesis: how many different sequences can a **random hexamer** represent?
- How many of them satisfy the requirement that two adjacent bases cannot be the same?

The Birthday Problem

- What is chance that in a group of 25 randomly selected people two or more will be found to share the same birthday?

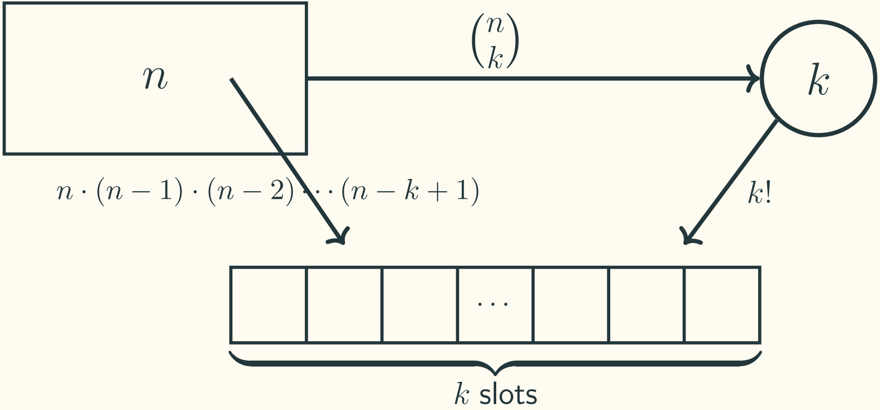
- $|\Omega| = ?$
- $A =$ { two or more will be found to share the same birthday }. What is $|A|$?

$$\underbrace{\boxed{}\boxed{}\boxed{}\boxed{\;\cdots\quad\cdots\quad\cdots\;}\boxed{}\boxed{}\boxed{}}_{\text{25 people}}$$

- $A^C =$ { all of them have distinct birthday }. What is $|A^C|$?

Combinations

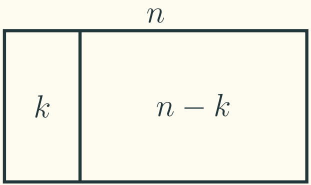
- Number of k-element subsets of a given n-element set - how many ways are there of choosing k-elements from an n-element set without replacement.



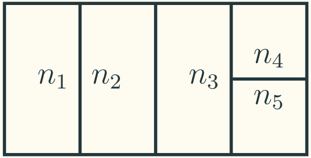
Coin flip example

- Event $A = \{$ 3 out of 10 independent flips were Hs $\}$
- Given that A has occurred, what is the conditional probability that the first two flips were Hs?

Partitions



$$n = n_1 + n_2 + n_3 + n_4 + n_5$$



$$|\Omega| = \frac{n!}{n_1!n_2!n_3!n_4!n_5!}$$

Example (ABO blood groups): In a room with 100 people, we know that there are 20 people with blood type A, 10 with B, 20 with AB and 50 with O, but we don't have the information of the blood type of each individual person.

1. How many total possible observations are there?
2. If we only know Adam, Bob, Charlie and Dave have different four different blood types, how many total possible observations are there for all people in the room?

Partitions

Let $n = n_1 + n_2 + n_3 + \cdots + n_k$, then the **multinomial coefficient** $\binom{n}{n_1, n_2, n_3, \ldots, n_k}$ can be interpreted as:

1. The number of ways of putting n **interchangeable/exchangeable** objects into k different boxes, so that box i has n_i objects in it for $1 \leqslant i \leqslant k$.
2. The number of unique permutations of a word with n letters and k distinct letters, such that the $i\text{-}th$ letter occurs n_i times.
3. The follows:

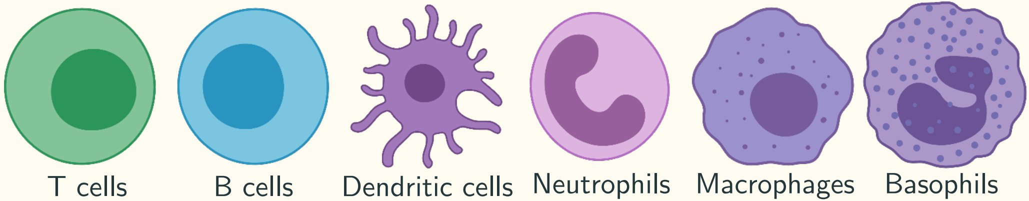
$$\binom{n}{n_1} \cdot \binom{n - n_1}{n_2} \cdot \binom{n - n_1 - n_2}{n_3} \cdots \binom{n_k}{n_k} = \frac{n!}{n1!n2!n3! \cdots n_{k-1}!n_k!}$$

The Multinomial Theorem

For a positive integer k and non-negative integer n :

$$(x_1 + x_2 + x_3 + \cdots + x_k)^n = \sum_{n_1+n_2+\cdots+n_k=n} \binom{n}{n_1, n_2, ..., n_k} \prod_{i=1}^{k} x_i^{n_i}$$

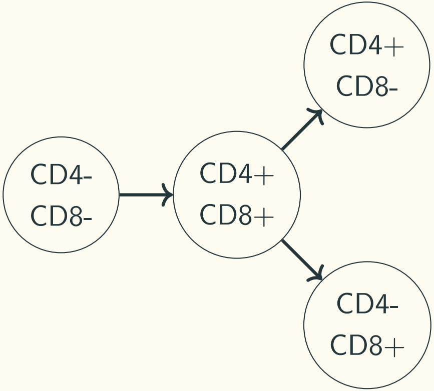
Multinomial Analysis of CD4 and CD8 Expressions In T Cells



T cells: T cells are a type of immune cells, and they are one of the important white blood cells of the immune system. T cells play a central role in the adaptive immune response. T cells can be easily distinguished from other lymphocytes by the presence of a T-cell receptor (TCR) on their cell surface. Within T cells, there are also many subtypes, based on different proteins on their cell surface, $e.g.$ CD4 and CD8.

Multinomial Analysis of CD4 and CD8 Expressions In T Cells

T cell development in Thymus



Proportions of different T cell subtypes in the thymus

CD4	CD8	Proportions
$-$	$-$	0.065
$+$	$+$	0.75
$+$	$-$	0.095
$-$	$+$	0.09

Multinomial Analysis of CD4 and CD8 Expressions In T Cells

Sample 1 ($n = 100$ cells)

CD4	CD8	# of cells
$-$	$-$	7
$+$	$+$	72
$+$	$-$	10
$-$	$+$	11

Sample 2 ($n = 100$ cells)

CD4	CD8	# of cells
$-$	$-$	5
$+$	$+$	5
$+$	$-$	45
$-$	$+$	45

Question: which sample is more likely to be taken from the thymus?

$$P(\text{Sample 1 comes from thymus}) = \binom{100}{7, 72, 10, 11} \cdot 0.065^{7} \cdot 0.75^{72} \cdot 0.095^{10} \cdot 0.09^{11}$$

$$P(\text{Sample 2 comes from thymus}) = \binom{100}{5, 5, 45, 45} \cdot 0.065^{5} \cdot 0.75^{5} \cdot 0.095^{45} \cdot 0.09^{45}$$