2. Ordered sampling without replacement

Example

You have 6 different dinosaur toys in a bag. You reach in, remove 4 of them one at a time, and line them up in front of you.

- Suppose you don't put them back in the bag before picking the next one (no replacement)
- Suppose that you don't look in the bag and that we can model this as uniform random sampling

How many outcomes are there? Define the sample space Ω and compute $|\Omega|$.

$\underline{6} \cdot \underline{5} \cdot \underline{4} \cdot \underline{3}$

$$\Omega = \{(x_1, x_2, x_3, x_4) : x_i \in \{1, 2, \ldots, 6\}\ \forall i = 1, \ldots, 4 \text{ and } x_i \neq x_j\ \forall i \neq j\}$$

$$|\Omega| = 6 \cdot 5 \cdot 4 \cdot 3 = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = \frac{6!}{2!}$$

2. Ordered sampling without replacement

Example

If instead you draw and line up all six dinosaur toys, how many outcomes are there? If all the dinosaurs are different species, what is the probability that triceratops is either second or third in line?

$$|\Omega| = \underline{6} \cdot \underline{5} \cdot \underline{4} \cdot \underline{3} \cdot \underline{2} \cdot \underline{1} = 6!$$

union

Let $A = \{\text{triceratops is 2nd or 3rd in line}\}$

$$P(A) = \underbrace{P(\underline{5}\ \underline{T}\ \underline{4}\ \underline{3}\ \underline{2}\ \underline{1})}_{\frac{5!}{6!}} + \underbrace{P(\underline{5}\ \underline{4}\ \underline{T}\ \underline{3}\ \underline{2}\ \underline{1})}_{\frac{5!}{6!}}$$

$$= \frac{2 \cdot 5!}{6!} = \frac{2 \cdot \cancel{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}}{6 \cdot \cancel{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}} = \frac{1}{3}$$

2. Ordered sampling without replacement

In general, these are called permutations or arrangements:

Arrangements: If you draw k out of n objects without replacement, the number of ways to order them is

$$|\Omega| = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

Permutations: In the special case that $k = n$,

$$|\Omega| = n \cdot \cdot (n-1) \cdot (n-2) \cdots 1 = n!$$

2. Ordered sampling without replacement

Example

You have a bag containing 3 red balls and 5 orange balls. You remove 4 of them, one at a time, without putting them back in the bag. What is the probability that you draw two red balls followed by two orange balls?

$\equiv A$

$A: \quad 3 \cdot 2 \cdot 5 \cdot 4$

$\Omega: \quad \underset{\text{ball 1}}{\underline{8}} \cdot \underset{2}{\underline{7}} \cdot \underset{3}{\underline{6}} \cdot \underset{4}{\underline{5}}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\cancel{3} \cdot \cancel{2} \cdot \cancel{5} \cdot \cancel{4}}{\cancel{8} \cdot 7 \cdot \underset{2}{\cancel{6}} \cdot \cancel{5}} = \frac{1}{14}$$

3. Unordered sampling without replacement

Example

A child wants to choose three of her six dinosaur toys to bring to a friend's house. How many choices does she have? Again assuming that all the dinosaur species of the toys are different, what is the probability that triceratops and ankylosaurus are among the three toys she brings if we assume each set of three toys is equally likely? $\equiv A$

1. draw 3 toys in an ordered seq. w/o replacement:

$$|\Omega| = \underline{6} \cdot \underline{5} \cdot \underline{4} = \frac{6!}{3!} \leftarrow$$

2. for each choice of 3 toys, there are $3!$ orderings

3. divide: $\dfrac{6!}{3!} \div 3! = \dfrac{6!}{3!\,3!} = \binom{6}{3}$ binomial coefficient

$\underline{T}$ $\underline{A}$ $\underline{4}$ $\quad \binom{4}{1} = \dfrac{4!}{3!\,1!} = 4$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{4}{\binom{6}{3}} = \frac{1}{5}$$

3. Unordered sampling without replacement

$k! \leftarrow 5!$

$n = 20$ $k = 5$

1 2 3 4 5

2 3 1 4 5

5 2 3 1 4

$\{1, 2, 3, 4, 5\}$

In general, these are called **combinations** or sets

The number of ways to draw a subset k out of n objects is given by the **binomial coefficient**:

$$|\Omega| = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

"n choose k"

This is the same as dividing (1) the number of arrangements of k out of n things by (2) the number of ways to order k things:

$$\frac{n!}{(n-k)!} \div k!$$

$\underbrace{\underline{n} \quad \underline{n-1} \quad \underline{n-2} \quad \cdots \quad \underline{n-k+1}}_{k}$

$$= \frac{\underbrace{n \cdot (n-1) \cdots (n-k+1)} \cdot \underbrace{(n-k) \cdot (n-k-1) \cdots 2 \cdot 1}}{(n-k)(n-k-1) \cdots 2 \cdot 1} = \boxed{\frac{n!}{(n-k)!}}$$

Practice: Identifying an appropriate sampling mechanism

Which method would you use? Then solve the problem.

1. 9 piano students are nervous about their upcoming recital, so they put their names in a bag and draw one at a time to determine who plays when. What is the probability of any particular outcome (recital program)?
2. How many different lunches can you make by choosing 1 main and 1 side from 3 different main dishes and 5 different side dishes?
3. You flip a fair coin 10 times. What is the probability of getting exactly three heads?
4. How many ways can a team of 12 basketball players choose a captain and co-captain?

1. ordered w/o replacement (permutation): $\underline{9} \cdot \underline{8} \cdot \underline{7} \cdot \underline{\ } \ \underline{\ } \ \underline{\ } \ \underline{\ } \ \underline{\ } \ \underline{\ }$

$|\Omega| = 9!$ $\quad P(\omega) = 1/9! \quad \forall\, \omega \in \Omega$

for all outcomes ω

for each outcome, $P(\text{outcome}) = \frac{1}{9!}$

2.

3.

4.

2. ordered (tuple)

$$\underset{\text{main}}{\underline{3}} \cdot \underset{\text{side}}{\underline{5}} = 15$$

$M = \{1, 2, 3\}$

$S = \{A, B, C, D, E\}$

$|\Omega| = |M| \cdot |S|$

$\binom{3}{1} \cdot \binom{5}{1} = 3 \cdot 5$

3. ordered w/ replacement

$\Omega = \{(f_1, \ldots, f_{10}) : f_i \in \{H, T\} \ \forall i = 1, \ldots, 10\}$

$A = \{(f_1, \ldots, f_{10}) : \text{exactly } 3 \text{ of } f_i \text{ are } H\}$

$|A| = \binom{10}{3}$ $\quad \underline{H}\ \underline{H}\ \underline{\ \ }\ \underline{\ \ }\ \underline{H}\ \ldots\ \underline{\ \ }$

$|\Omega| = |\{H, T\}^{10}| = 2^{10}$.

$P(A) = \binom{10}{3} / 2^{10} \approx 0.1$

4.

$\underline{\quad}_{12}\ \underline{\quad}_{11} = 12 \cdot 11$

$$\binom{12}{2} \cdot 2 = \frac{12 \cdot 11}{2} \cdot 2 = 12 \cdot 11$$

arrangement

pick 2 leaders, $\rightarrow \binom{12}{2} = \frac{12!}{10!\,2!} = \boxed{\frac{12 \cdot 11}{2}}$

$\boxed{\text{then pick which leader will be captain}} \rightarrow \cdot 2$

$$\frac{12 \cdot 11 \cdot \cancel{10} \cdots}{\cancel{10 \cdot 9} \cdots}$$

$$\boxed{\binom{12}{1}} \cdot \boxed{\binom{11}{1}} = 12 \cdot 11$$

$\boxed{12} \cdot \boxed{11}$

captain cocaptain

$$\binom{12}{2} \cdot 2 = \frac{12!}{10!\,2!} \cdot 2$$

More practice: Identifying an appropriate sampling mechanism

Both of these are technically ordered sampling w/o replacement but see below.

Which method would you use? Then solve the problem.

1. The four types of nucleotide bases in DNA are A, T, C, and G. Suppose (1) you have a solution in which all four are present in equal concentrations and the bases form a sequence of three base pairs uniformly at random. What is the probability of getting the sequence ATC?
   - (2) What if instead of molecules in solution you had a bag with scrabble tiles (2 A's, 2 T's, 2 C's, 2 G's)?

(1) Even though there are technically finitely many A's T's C's & G's in solution, there are so many that we would probably better model this problem as sampling with replacement from an infinite supply of bases. A, T, C, G are equally likely & the sequence is short: $\frac{1\cdot 1\cdot 1}{4\cdot 4\cdot 4} = \frac{1}{64}$

(2) With scrabble tiles, there are few enough of each letter that we model the problem as sampling w/o replacement. Then what is the probability of getting the sequence (A, T, C)?

* What is the probability $P(\{A, A, A\})$ in both cases here?

What about unordered sampling with replacement?

Let's revisit this fourth sampling method.

Example

Draw two numbers from the set $\{1, 2, 3\}$ with replacement, but then disregard order so that you define the outcomes to be the possible combinations of numbers you can draw. What are the possible outcomes, and what are their probabilities? Are they equally likely?

This is a good problem to think about, but I'll demonstrate another on the next page.

You flip a coin 3 times. What is the probability of getting 2 heads? 0 heads?

We can either think of this as ① ordered sampling without replacement or ② unordered sampling with replacement. The difficulty with approach ② right now is that the outcomes are not equally likely; we will see later in the class how to handle this type of problem. For now, we solved this (for 10 flips) using approach ①.

① ordered w/o replacement

Ω: all outcomes equally likely

② unordered with replacement

Ω: 1 & 2 are more likely than 0 & 3.

①	②
(H, H, H)	3 heads
(H, H, T)	2 heads
(H, T, H)	2 heads
(T, H, H)	2 heads
(H, T, T)	1 head
(T, H, T)	1 head
(T, T, H)	1 head
(T, T, T)	0 heads

Outline

Infinitely many outcomes

All our examples until now dealt with finitely many outcomes. This example

- involves infinitely many outcomes
- demonstrates we can sometimes turn a problem with unequally likely outcomes into a series of problems with equally likely ones

Example

Define a probability model (~~all three parts~~) for this problem: you flip a coin until it comes up tails. The outcome is the number of coin flips up to and including the first tails. What is the probability that you get tails on exactly the third coin toss?

$\Omega = \{\#$ flips up to & including 1st tails$\} = \{1, 2, 3, \dots\} = \mathbb{N}$

$A = \{3\}$ $\quad$ ~~$P(A) = \frac{|A|}{|\Omega|}$~~ ← finitely many outcomes

outcome $= k \iff \boxed{\underbrace{H, \dots, H}_{k-1}, T}$ $\quad$ how many sequences could you get from k tosses? 2^k

$P(k) = \frac{1}{2^k} \quad \forall k \in \mathbb{N}$

$P(A) = \frac{1}{2^3} = \frac{1}{8}$.

$$\Omega = \boxed{\mathbb{N}} \cup \boxed{\{\infty\}}$$

$$\sum_{\omega \in \Omega} P(\omega) = 1$$

$$\sum_{\omega \in \mathbb{N}} P(\omega) = \boxed{\sum_{\omega \in \mathbb{N}} \frac{1}{2^\omega}} = 1$$

$$\Rightarrow P(\infty) = 0.$$

Different kinds of infinite sets

Consider two other examples with infinite outcomes:

- What is the probability that you get a bullseye if you throw a dart uniformly at random at a target?
- If you choose a real number x uniformly at random from the interval [0,1], what is the probability that $x \in (0.25, 0.5)$?

The previous example (coin flips) is inherently different from the two examples above in an important way:

The sets involved in the coin flips example are **countable** or **discrete**

- Elements can be labeled in one-to-one correspondence with the positive integers (natural numbers) $\mathbb{N}$ (infinite) or a finite subset of them (finite)

Sets in the two examples above are **uncountable** or **continuous**

- These sets cannot be covered by or mapped to $\mathbb{N}$

Measuring uncountable sets

For an uncountable set A such as the real numbers,

$$P(x) = 0 \quad \forall\, x \in A.$$

Therefore a probability measure much be based not on points or counting the number of elements, but rather on lengths, areas, etc.

We can generalize our previous result on equally likely outcomes:

Fact

If the sample space Ω has uncountably many elements and equally likely outcomes, then for any event A we have

$$P(\omega) = 0 \textit{ for any outcome } \omega \in \Omega \textit{ and } P(A) = \frac{\nu(A)}{\nu(\Omega)} \textit{ for any event } A,$$

where $\nu(A)$ (pronounced "nu of A") is an appropriately defined **measure** *of A for the problem of interest (e.g. length or area).*

This is somewhat imprecise but suffices for this course; a measure theoretic probability course will make this more precise

A note about Axiom iii

Note that Axiom iii in our statement of the probability model is only required to hold for countable sets: If $A_1, A_2, A_3, ...$ is a sequence of **pairwise disjoint** events (i.e. mutually exclusive, $A_i \cap A_j = \emptyset$ for $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

This is called countable additivity.

An aside: σ-algebras

For uncountably infinite sets, we actually cannot use the power set as the event space

- Instead, replace this with a **$\boldsymbol{\sigma}$-algebra**, any collection of sets satisfying three properties (see textbook §1.6)

Even for countable sets (finite or infinite), it can be useful to define the event space as a σ-algebra that is not the power set

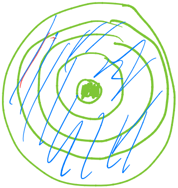
- Allows you to model information available to an observer/experimenter
- Arises in advanced probability concepts such as filtrations, martingales, optional stopping theorem
- Applications include selective inference, mathematical finance

Back to an uncountably infinite example

Example

What is the probability that you hit the bullseye if you throw a dart uniformly at random at a target?

- What additional information do you need? *This is an important part of research and consulting in which the problems are not handed to you fully defined*
- Define the sample space and the event of interest (not the whole event space)
- Answer the original question above



need areas of target & bullseye

Suppose radii are 1 (bullseye) & 3 (overall)

$$\Omega = \{(x, y) : x^2 + y^2 < 9\} \leftarrow$$

$$A = \{(x, y) : x^2 + y^2 < 1\} \leftarrow$$

$$P(A) = \frac{\text{area}(A)}{\text{area}(\Omega)} = \frac{\pi}{9\pi} = \frac{1}{9}.$$