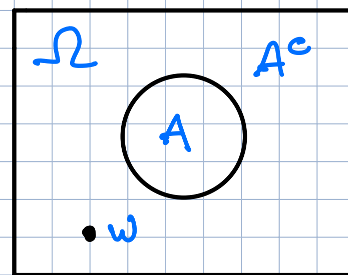
Lecture 1 – Probability Basics

A. Set Notation (a brief review)

Venn diagram



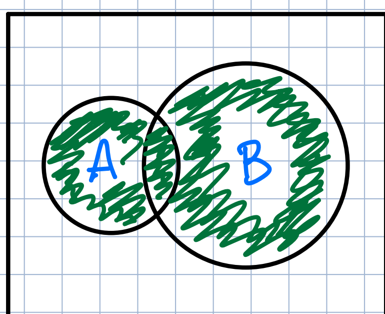
Let Ω be a set of points

A is a subset of Ω, $A \subset \Omega$

w is a point in Ω, $\omega \in \Omega$

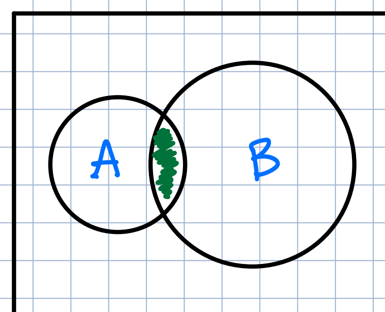
$A^c = \{\omega \in \Omega : \omega \notin A\}$, the compliment

Suppose $A \subset \Omega$ and $B \subset \Omega$.



$A \cup B$ "or"

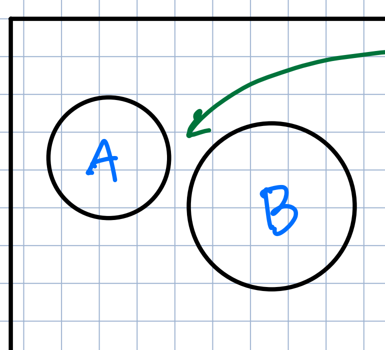
union



$A \cap B$ "and"

intersection

If $A \cap B$ is $\emptyset$, the null set, then we say that A and B are disjoint or mutually exclusive



no area of overlap

B. Probability Axioms

There are several different ways to define probability. By probability I mean a number $P(A)$ assigned to event A that corresponds to how frequently A occurs.

1. classical probability: (e.g. we know beforehand) $P(A) = \frac{|A|}{|\Omega|}$ ← "cardinality" = number

2. statistical probability: (e.g. we measure) $P(A) = \lim_{n\to\infty} \frac{n_A}{n}$ where n are repeated trials in Ω and $n_A \in A$.

3. subjective probability: "degree of belief"
   - Can apply to non-random processes (e.g. a presidential election). This is the domain of philosophers and social scientists. Can be useful, but not our focus here.

4. Axiomatic probability: mathematically rigorous "measure" originally formulated by Kolmogorov

(i.) $P(\emptyset) = 0$, "impossible event"

(ii.) $P(\Omega) = 1$, "sure event"

(iii.) $P(A) \geq 0$, non-negative probabilities

(iv.) If $A \cap B = \emptyset$, then "countable additivity"

$$P(A \cup B) = P(A) + P(B)$$ The probabilities of mutually exclusive events add.

these axioms give probabilities certain properties that must always be true. These include:

$$P(A^c) = 1 - P(A)$$

Kazmessis notation:

$$P(A) \le P(B) \quad \text{if } A \subset B$$

"joint probability"

$P(A + B)$ A or B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$P(A, B)$ A and B

Example: Deck of cards

$P(\text{ace of spades}) = 1/52$

$P(\text{ten}) = 4/52$

$$P(\text{ace of spades or ten}) = P(\text{ace of spades}) + P(\text{ten}) - \underbrace{P(\text{ace of spades} \cap \text{ten})}_{P(\emptyset) = 0}$$

$$= 1/52 + 4/52 = \frac{5}{52}$$

A final comment on the meaning of "probability" and "randomness": Sometimes the ideas of probability and randomness are misunderstood to mean that no order or laws apply. In fact, the lack of "determinism" does not imply this at all. The theories of probability, validated by many experiments, give very precise and orderly laws that predict the range of behavior of single events and the frequency of many events. Even if the event is not deterministic, the probability is.

C. Conditional Probability

Sometimes we want to know the probability of something, given that another event happened. This is called a conditional probability. It is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

"A given B"

$P(A \cap B)$ ↔ $P(A \text{ and } B)$ "joint probability"

Note that:

$$P(A|B) \neq P(B|A)$$

The probability that (for example) I am wet given that it rained is not the same as the probability that it rained, given I am wet. (I might have taken a shower!)

Also, note that $P(B) \neq 0$, or the conditional probability is undefined.

Finally, in the case that the two events are independent, the following holds:

$$P(A|B) = P(A)$$ ← it doesn't matter what B is on A.

Using the definition, independence implies:

$$P(A \cap B) = P(A)\,P(B)$$ Joint probability / "and"

Contrast this with mutual exclusivity:

$$P(A \cap B) = 0 \Rightarrow P(A \cup B) = P(A) + P(B) \quad \text{"or"}$$

It often occurs that I don't know the joint probability, $P(A \cap B)$. But a little algebra helps us:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

$$\boxed{P(A|B) = \frac{P(B|A)}{P(B)} P(A)} \quad \text{Bayes' Theorem}$$

"posterior" "informativeness" $= \frac{\text{likelihood}}{\text{marginal probability}}$ "prior"

Also, it is often the case that I might not know $P(B)$, but I know other conditional probabilities. In this case, I can use the law of total probability:

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) \quad \text{"not A"}$$

Combining with Bayes' theorem gives

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Example: Suppose I go in for a covid test. Only 1% of people have covid. The test is 95% accurate (sensitivity) if one has covid, but has a 10% false positive (specificity) rate. The test comes back positive. What is the probability that I actually have covid?

95% = sensitivity
true positive rate
$\beta = 5\%$
false negative rate
(type II error)

90% = specificity
true negative rate
$\alpha = 10\%$
false positive rate
(type I error)

The question is asking for $P(\text{covid} \mid +)$

Bayes theorem says:

$$P(\text{covid} \mid +) = \frac{P(+ \mid \text{covid})}{P(+)} P(\text{covid})$$

$P(\text{covid}) = 0.01 \qquad P(\text{not covid}) = 0.99$

$P(+ \mid \text{covid}) = 0.95 \qquad P(+ \mid \text{not covid}) = 0.1$

$$P(+) = P(+ \mid \text{covid}) P(\text{covid}) + P(+ \mid \text{not covid}) P(\text{not covid})$$

$$= 0.95 \cdot 0.01 + 0.1 \cdot 0.99 = 0.1085$$

$$P(\text{covid} \mid +) = \frac{0.95}{0.1085} \cdot \underbrace{0.01}_{\text{prior}} = 0.0875 \approx 8.8\%$$

prior was low. + test makes it 8x more likely.

What if false positive rate was 0.5%?

$P(+ \mid \text{not covid}) = 0.005 \Rightarrow P(+) = 0.01445$

$$P(\text{covid} \mid +) = \frac{0.95}{0.01445} \cdot 0.01 = 0.657 \approx 66\%$$

much better test.

D. Combinatorics

Combinatorics is the mathematical discipline concerned with counting. We can combine some of the more elementary insights in combinatorics with the classical model of probability: $P(A) = |A|/|\Omega|$ to obtain probabilities. This is often useful in games, but is also insightful for molecules.

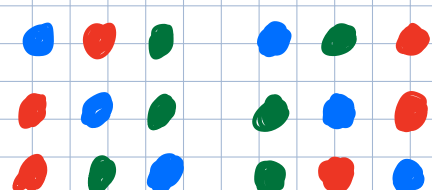
The fundamental principle of combinatorics is that if event A has cardinality n_A (number of ways to do event A) and event B has cardinality n_B, then the cardinality of event A+B is $n_A \times n_B$.

or

Case 1: arranging (ordering) objects

(1a) order n different objects = $n!$ (distinuishable)

Example:



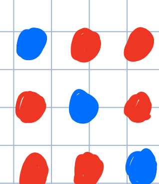
$3! = 6$

(1b) order n objects of which p are identical = $n!/p!$
(indistinguishable)

Example:

$n=p=3$

$3!/3! = 1$



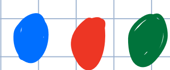
$n=3$

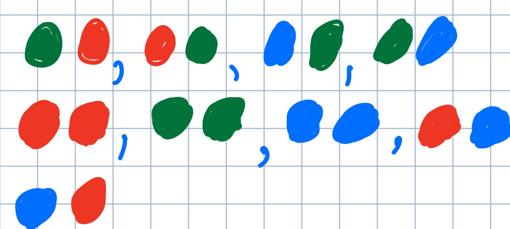
$p=2$

$\frac{3!}{2!} = 3$

General formula : $\dfrac{n!}{p!\,q!\,r!\ldots}$ for objects $p, q, r, \ldots$

(1c) order k objects from n with replacement $= n^k$

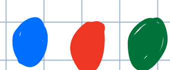
Example:  $n=3$, $k=2$

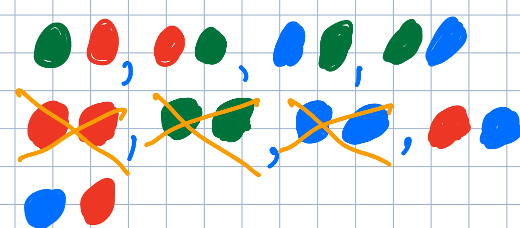


$3^2 = 9$

(1d) order k objects from n w/out replacement $= \dfrac{n!}{(n-k)!}$

"Permutations", ${}_nP_k$

Example:  $n=3$, $k=2$



$$\frac{3!}{(3-2)!} = \frac{3\cdot 2\cdot 1}{1} = 6$$

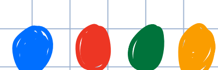
case 2: Selecting objects (no ordering)

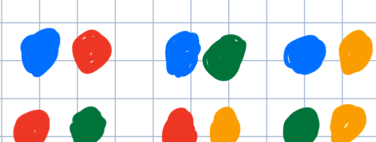
(2a) select k objects from n w/out replacement $= \dfrac{n!}{k!\,(n-k)!}$

"Combinations", ${}_nC_k$

$= \dbinom{n}{k}$

binomial coefficient

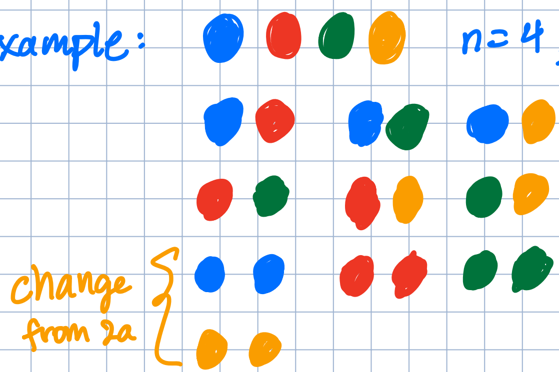
Example:  $n=4$, $k=2$



$$\frac{4!}{2!\,(4-2)!} = \frac{4\cdot 3\cdot 2\cdot 1}{2\cdot 1\cdot 2\cdot 1} = 6$$

(2b) select k objects from n with replacement $= \binom{n-1+k}{k}$

Example: $n=4$, $k=2$



$$\frac{(4-1+2)!}{2!\,(4-1)!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot3\cdot2\cdot1} = 10$$

summary: 3 different types of properties

- ordering vs. not ordering
- distinguishable vs. indistinguishable
- replacement vs. no replacement

E. Stirling's Approximation

An important question for molecules: How do we calculate $n!$ when n is a big number?

$$\lim_{n\to\infty} n! = \sqrt{2\pi n}\, n^n e^{-n}$$

more often we write this as a log:

$$\ln n! = \ln\left[\sqrt{2\pi n}\, n^n e^{-n}\right] = \frac{1}{2}\ln(2\pi n) + n\ln n - n\ln e$$

$$= \frac{1}{2}\ln 2\pi + (n+\tfrac{1}{2})\ln n - n$$

$$\approx n\ln n - n$$

(when $n\to\infty$)

$$\boxed{\ln n! \approx n\ln n - n}$$