Week 2

Recall:

Probability Functions

Def: Given a sample space S,
a sigma algebra $\mathcal{B}$ associated w/ S,

a probability function (distribution, measure)
is a function P on $\mathcal{B}$ that satisfies

1. $P(A) \geq 0 \quad \forall A \in \mathcal{B}$
2. $P(S) = 1$
3. If $A_1, A_2, \ldots, \in \mathcal{B}$ are pairwise disjoint
   then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Remark:
- A and B are disjoint i.f.f. $A \cap B = \emptyset$.
- the three properties termed Kolmogorov's Axioms of Prob.
- (3) is the axiom of countable additivity

recall: Finite additivity := $A_1, \ldots, A_n$ pairwise disjoint.
then $P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) + \cdots + P(A_n)$

Ex: Tossing a fair coin: $S = \{H, T\}$.
$P(\{H\}) := P(\{T\}) := \frac{1}{2}$

Tossing an unfair coin: $\begin{cases} P(\{H\}) = q \\ P(\{T\}) = 1-q \end{cases}$ for some $q \in [0,1]$

Tossing two coins in a row:
$S = \{HH, HT, TH, TT\}$

Take $P(HH) = q_1, \ldots, P(TT) = q_4$
where $\begin{cases} q_1 + \cdots + q_4 = 1 \\ q_1, \ldots, q_4 \geq 0 \end{cases}$

Recall: a set can be finite or infinite
→ an infinite set is either countable or uncountable
→ a countable set is either finite or countably infinite

Probabilities on countable sets

- relatively easy to define

Thm. Let $S = \{s_1, \ldots, s_n\}$ (finite),
$\mathcal{B}$ any sigma-algebra of subsets of S.
Let $p_1, \ldots, p_n \in [0,1]$ s.t. $\sum_{i=1}^{n} p_i = 1$.
Define, for any $A \in \mathcal{B}$ $\boxed{P(A) := \sum_{i: s_i \in A} p_i}$
then, P is a valid prob function on $\mathcal{B}$

Proof. check finite additivity, i.e. $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$.

$$P(A \cup B) = \sum_{i: s_i \in A \cup B} \underset{A \cap B = \emptyset}{=} \sum_{s_i \in A} p_i + \sum_{s_i \in B} p_i$$

$$= P(A) + P(B). \blacksquare$$

Thm. Let $S = \{s_1, s_2, \ldots, \}$ (countably infinite)
$\mathcal{B}$ any sigma-algebra of subsets of S.
Let $p_1, p_2, \ldots \in [0,1]$ s.t. $\sum_{i=1}^{\infty} p_i = 1$.
Define, for any $A \in \mathcal{B}$ $\boxed{P(A) := \sum_{i: s_i \in A} p_i}$
then, P is a valid prob function on $\mathcal{B}$

Proof. check infinite additivity, i.e. $P(A \cup B \cup \cdots) = P(A) + P(B) + \cdots$ if $A \cap B \cap \cdots = \emptyset$.

$$P(A \cup B \cup \cdots) = \sum_{i: s_i \in A \cup B \cdots} \underset{\text{pairwise disjoint}}{=} \sum_{s_i \in A} p_i + \sum_{s_i \in B} p_i + \cdots$$

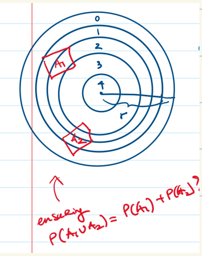
$$= P(A) + P(B) + \cdots. \blacksquare$$

Assigning Probabilities to uncountable sets

- often/can be a non-trivial pursuit

Ex: Dart throwing experiment

Let $S = \{$dart positions on the circle based on radius $r\}$
$\in \mathbb{R}^2$
- S is uncountable
- valid P must satisfy countable additivity, among other axioms
- Perhaps we can try to to define an "unforseen" prob. dist.



we may define $P(\{i\}) := \dfrac{\text{area on ring } i}{\text{area of dart board}}$

$$= \frac{\pi\left(1-\frac{i}{5}\right)^2 r^2 - \pi\left(1-\frac{i+1}{5}\right)^2 r^2}{\pi r^2}$$

$$= \left(1-\frac{i}{5}\right)^2 - \left(1-\frac{i+1}{5}\right)^2$$

General Approach

+ Placing probabilities on (countably) many 'simple' subsets
+ using axioms of Prob. (AOP) to derive probabilities of all remaining subsets of the sigma algebra $\mathcal{B}$.

Tools

A prob. function P on $\mathcal{B}$ must satisfy the following facts:

Thm. $\forall\, A \in \mathcal{B}$,

$0 \le P(A) \le 1$,

$\boxed{P(A^c) = 1 - P(A)}$

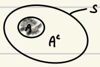
Proof. $A \cap A^c = \emptyset$ by def.

$A \cup A^c = S$

By AOP(3), $P(S) = P(A) + P(A^c)$

AOP(2) states $P(S) = 1$

$\Rightarrow P(A^c) = 1 - P(A)$. ∎

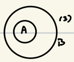


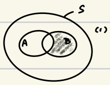
Thm. $\forall\, A, B \in \mathcal{B}$,

1. $P(B \cap A^c) = P(B) - P(A \cap B)$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. If $A \subset B$, then $P(A) \le P(B)$

Proof (3). $P(B) \overset{\text{AOP3}}{=} P(A) + \underbrace{P(B \cap A^c)}_{\ge 0 \text{ by AOP1}}$

$P(B) \ge P(A)$. ∎





Remark: Proofs follow from AOP

Corollary (Bonferroni's Inequality):

$P(A \cap B) \ge P(A) + P(B) - 1$

Proof. By (3), $P(A \cap B)$

$= P(A) + P(B) - \underbrace{P(A \cup B)}_{\le 1}$

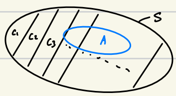
$P(A \cap B) \ge P(A) + P(B) - 1$

More Powerful Tool: countable → uncountable

i.e. using probabilities on a countable collection of subsets to determine prob. on an uncountable collection

Thm.

1. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $(C_1, C_2, \dots)$ of S
2. $P\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} P(A_i)$ $\forall\, A_1, A_2, \dots \in \mathcal{B}$

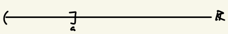


Remarks:

- A partition $(C_1, C_2, \dots)$ of S means $C_1, C_2, \dots$ are pairwise disjoint, and $\bigcup_{i=1}^{\infty} C_i = S$

Q: what is an ex. of assigning prob.'s to the Borel sigma-algebra of subsets in $\mathbb{R}$?

A: Suppose $\{(-\infty, a] \mid a = \frac{x}{y},\ x, y \in \mathbb{Z}\}$



1.2: Counting

- To place prob.'s on countably many possibilities (events), we must know how to count all such numbers

Thm. (Fundamental Thm of Counting)

If a job consists of k separate tasks, the i^{th} of which can be done by n_i ways, for $i = 1, \dots, k$.

Then, the entire job can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways

Ex (Lottery): From $\{1, \dots, 44\}$, pick six different numbers for a ticket

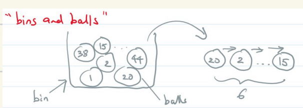
Q: How many possible tickets?

e.g. $\{10, 2, 35, 7, 49, 13\}$

Sampling terminology: ordered, w/o replacement

A: $44 \times 43 \times 42 \times 41 \times 40 \times 39$

$= \frac{44 \times \cdots \times 2 \times 1}{38 \times \cdots \times 2 \times 1} = \frac{44!}{38!}$



Here, suppose $n = 44$, $r = 6$

$\Rightarrow A = \frac{n!}{(n-r)!}$

Ex (ordered w/ replacement): $44 \times 44 \times \cdots \times 44 = 44^6 = n^r$

Ex (unordered w/o replacement): $\frac{44 \times 43 \times \cdots \times 39}{6 \times 5 \times \cdots \times 1} = \frac{44!}{38!\,6!} = \binom{44}{6} = \binom{n}{r} := \frac{n(n-1)\cdots(n-r+1)}{r!\,(n-r)!} = \frac{n!}{r!\,(n-r)!}$

"n choose r"

Ex (unordered w/ replacement):

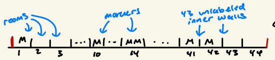
the six numbers need not be distinct, order irrelevant

non-trivial, and answer $\ne \frac{44^6}{6!}$

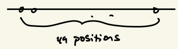
Reasoning

+ Lay out the 44 numbers in a fixed order and create 44 "rooms", one for each number
+ Rooms are separated by walls (45 in total)

+Place the 6 markers "M" in rooms where the numbers have been chosen



$\Leftrightarrow$ Given $\begin{cases} 6 \text{ marks} \\ 43 \text{ inner walls} \end{cases}$



Above: $\{1, 10, 14, 14, 41, 42\}$ were chosen

★ **key obs**: # tickets $\equiv$ # unique ways to place 6 markers to 44 bins

= # ways to arrange 6 markers w/ 43 inner walls

$= \binom{6+43}{6}$

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

Possible methods of counting (Summary)

Number of possible arrangements of size r from n objects

	w/o replacement	w/ replacement
ordered	$P^n_r = \frac{n!}{(n-r)!}$	n^r
unordered	$C^n_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$	$\binom{n+r-1}{r}$

Remark: $\binom{n}{r}$ are referred to as binomial coefficients

---

Counting (cont.)

· If we can count the number of outcomes of an experiment, say N, and if such outcomes are equally probable, then the prob. assigned to each outcome can be set to $1/N$, i.e.

> **Prop.** If each outcome (a) is equally probable
> and $|S| = N$,
> then let $P(\{a\}) = \frac{1}{N}$ $\forall a \in S$

Ex: How many ways to choose a 5-card hand from a deck of 52 cards?

A: $\binom{52}{5}$ since unordered w/o replacement

Assuming the cards are well-shuffled and randomly dealt

$P(\text{a hand}) = \frac{1}{\binom{52}{5}} = 1/2{,}598{,}960.$

Ex: What is the prob. of obtaining 4 aces from the 5-cards?

$P(4 \text{ aces}) = \frac{48}{\binom{52}{5}}$ → ways to select the fifth card

$\approx 1/50{,}000$ A A A A *

Ex: What is the prob. of "four of a kind"?

$P(4 \text{ of a kind}) = 13 \times 48 \times \frac{1}{\binom{52}{5}}$ 8D 8H 8S 8C *

13 (number) possibilities

ways to select fifth card

Ex: Prob. of having exactly one pair

such hands is $m = 13\binom{4}{2}\binom{12}{3}4^3$

e.g. $(3D, 3♡, \underbrace{2, J, 10}_{13-1=12})$ # possibilities   suit for final 3 cards

$\Rightarrow P(\text{having exactly one pair in 5-card}) = \frac{m}{\binom{52}{5}}$

Ex (sampling w/ replacement)

"uniform sampling" $r=2$ items from $n=3$ items w/ replacement

ordered	(1,1)	(2,2)	(3,3)	(1,2)(2,1)	(1,3)(3,1)	(2,3)(3,2)
unordered	$\{1,1\}$	$\{2,2\}$	$\{3,3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
Prob.	1/9	1/9	1/9	2/9	2/9	2/9

Q: Is there a sampling mechanism according to which the unordered outcomes are given equal probabilities (1/6)?

A: Rolling a fair dice

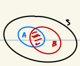
1.3: Conditional Probabilities

> **Def**: If A, B are events in sample space S
> and $P(B) > 0$
> then the conditional prob of A given B
> is $\boxed{P(A|B) := \frac{P(A \cap B)}{P(B)}}$

Remarks: · Equivalently, $P(A \text{ true} \mid B \text{ true}) = \frac{P(\text{both } A \text{ and } B \text{ true})}{P(B \text{ true})}$ (*)

where (*) is 'renormalizing' under (new) info B

· Equivalently $\underbrace{P(A \cap B)}_{\text{joint prob}} = \underbrace{P(B)}_{\text{marginal}} \cdot \underbrace{P(A|B)}_{\text{conditional}}$



> **Prop.** If $P(A) > 0$, then $P(B \cap A) = P(A)P(B|A)$
> $= P(B)P(A|B)$

$\Rightarrow$ Hence, $\boxed{\text{Bayes' Formula}: = P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$

Prop. Generally, $P(A \cap B \cap C) = P(A) \times P(B \cap C \mid A)$
$= P(A) \times P(B \mid A) \times P(C \mid B, A)$

$\rightarrow$ Hence, $\boxed{\underline{\text{Chain Formula}} := P(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} P(A_i \mid A_1 \cap \cdots \cap A_{i-1})}$

Ex: Suppose 4 cards are dealt from the top of a well-shuffled deck.

Then $P(\{4 \text{ aces}\}) = \dfrac{\binom{4}{4}}{\binom{52}{4}} = \dfrac{1}{\binom{52}{4}} = \left(\dfrac{52 \times 51 \times 50 \times 49}{4 \times 3 \times 2 \times 1}\right)^{-1} = \dfrac{4 \times 3 \times 2 \times 1}{52 \times 51 \times 50 \times 49}$

$\Leftrightarrow P(\{4 \text{ aces}\}) = P(\text{first card ace}) \cdot P(2^{nd} \text{ card is ace} \mid \text{first ace})$
$\cdot P(3^{rd} \text{ card ace} \mid \text{first two aces})$
$\cdot P(4^{th} \text{ card ace} \mid \text{first three aces})$
$= \dfrac{4}{52} \times \dfrac{3}{51} \times \dfrac{2}{50} \times \dfrac{1}{49}$

· Conditional prob.'s can 'surprise' us
(due to how they are defined and how we use them)

e.g. Three prisoners A, B, and C
one of them chosen (randomly) to be pardoned.

A asks the warden, who is supposed to keep secret:
· which among B and C will be executed?
to which the warden responds
· B is to be executed

Warden's thinking: The prob. that either A or B or C gets pardoned is 1/3.
Between B and C, at least one is executed.
So, "I" gave A no new information on his life.

A's thinking: Since either C or I gets the pardon, my chance of being alive has gone up to 1/2

Let A, B, C be pardon events, i.e.
$P(A) = P(B) = P(C) = 1/3$.

If Prisoner is Pardoned	Then, warden tells A	Prob
A	$\rightarrow$ B dies	1/2
	$\rightarrow$ C dies	1/2
B	$\rightarrow$ C dies	1
C	$\rightarrow$ B dies	1

Let W: event that warden tells A that B dies
· The warden's reasoning is as follows:
$$P(A \mid W) = \frac{P(A \cap W)}{P(W)} = \frac{P(A)P(W \mid A)}{P(W)} = \frac{(1/3)(1/2)}{(\frac{1}{3})(\frac{1}{2}) + (\frac{1}{3})(1)} = \frac{1/6}{1/2} = \frac{1}{3}. \blacksquare$$
· A's reasoning comes from
$$P(A \mid B^c) = \frac{P(A \cap B^c)}{P(B^c)} \overset{A \subset B^c}{=} \frac{P(A)}{1 - P(B)} = \frac{1/3}{2/3} = \frac{1}{2}. \blacksquare$$

Remarks:

* conditional probabilities provide the means to quantify how we may have different assessments of the same event
  (due to different info we use or have access to)
  * what event/info we choose to condition on can affect the answer

* conditional prob.'s provide the mathematical machinery of Bayesian statistics, which bases the inference on the posterior distribution, i.e. cond. prob. of quantities of interest conditionally given observed data

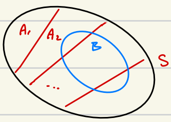
* cond. Prob.'s are obtained usually via the Bayes' Formula

Recall: Bayes Formula: $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)}$ for $P(B) > 0$, $A, B \in \mathcal{B}$.

Prop. Let $A_1, A_2, \ldots$ be a partition of S, i.e. $S = \bigcup_{i=1}^{\infty} A_i$

Then $P(B) = \sum_i P(B \cap A_i) = \sum_i P(B \mid A_i) P(A_i)$

$\Rightarrow \boxed{P(A_i \mid B) = \dfrac{P(B \mid A_i)P(A_i)}{(*) \sum_i P(B \mid A_i) P(A_i)}}$



Corollary: $\forall i \quad \underbrace{P(A_i \mid B)}_{\text{Posterior dist.}} \propto \underbrace{P(B \mid A_i)}_{\text{likelihood}} \underbrace{P(A_i)}_{\text{Prior dist.}}$

$\propto$: proportional to (up to multiplying constant)

(Bayesian Stats): observed data

Remark: $(*)$ is the law of total prob.

Independence

We want to be able to convey that an event B has no effect on A by insisting that

$$\underline{P(A|B) = P(A)} \quad \text{(provided } P(B)>0\text{)}$$

$$\Rightarrow P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{\cancel{P(A)}P(B)}{\cancel{P(A)}} = P(B), \text{ provided } P(A)>0.$$

Moreover,

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

Def: two events are statistically independent if $\boxed{P(A\cap B) = P(A)P(B)}$

Remarks:
- does not require $P(A)$ or $P(B) > 0$
- A & B independent does not mean $A\cap B = \emptyset$ (disjoint def.)