1 The probability measure

One **goal of probability** is to make sense of uncertainty (**randomness**). How to use it for prediction?

Example. If one flips a fair coin then the events H (heads) and T (tails) are equally likely. Flipping it many times, about half of the results will be H. Unfair coin: long term proportion $p \neq 1/2$.

Randomness is uncertainty that is statistically predictable, i.e. "the proportion of times an event occurs" will converge to a specific value, namely to the likelihood the event will occur at any given time.

1.1 Sample spaces and σ-algebras

Definition
(i) A **random experiment** is a well-defined, repeatable investigation in which exactly one of a set of possible outcomes is the experimental result, but just which outcome that is, is a matter of randomness.
(ii) Call the set of possible outcomes **sample space** $\mathcal{S}$.
(iii) An **event** (E) is any subset of $\mathcal{S}$, including $\mathcal{S}$ and $\emptyset$.

Example (two dice)

11	**12**	13	14	15	16	**E** = "sum ≤ 3"
21	22	23	24	25	26	$\mathcal{S} = \{(i,j) : 1 \leq i,j \leq 6\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
61	62	63	64	65	66	

Example (coin flips)
Exp. 1: one coin flip, $\mathcal{S} = \{0,1\}$ with $0 = T, 1 = H$.
Exp. 2: n coin flips, $\mathcal{S} = \{s = (x_1, \ldots, x_n) : x_i \in \{0,1\}\}$ has 2^n outcomes. Some possible events:

$A =$ "the first two flips are H" $= \{(x_1, \ldots, x_n) : x_1 = x_2 = 1\}$,
$B =$ "the total number of H is 12" $= \{(x_1, \ldots, x_n) : \sum_{i=1}^n x_i = 12\}$.

Exp. 3: Flip the coin until H occurs, $\mathcal{S}$ is countably infinite, $\mathcal{S} = \{(1, x_2, x_3, x_4 \ldots),$ $(0, 1, x_3, x_4, \ldots),$ $(0, 0, 1, x_4, \ldots)\}$ or, briefly, $\mathcal{S} = \{1, 2, 3, \ldots\}$.
Exp. 4: An infinite sequence of flips, $\mathcal{S} = \{s = (x_1, x_2, \ldots) : x_i \in \{0,1\}\}$. An important event is $C =$"the limiting proportion of H's is 1/2" $= \{(x_1, x_2, \ldots) \in \mathcal{S} : \lim_{n\to\infty} \sum_{i=1}^n x_i/n = 1/2\}$.

Events / notation: Write $A \cup B, \bigcup_{i=1}^n A_i$ (union), $A \cap B$ (or short AB), $\bigcap_{i=1}^n A_i$ (intersection) and A^c or $\bar{A}$ (complement). Say "A occurs" if the actual outcome s is in A ($s \in A$).

Useful events: if the outcome is in $\bigcup_n A_n, \bigcap_n A_n, (\bigcup_n A_n)^c$ or in $(\bigcap_n A_n)^c$ then we say that A_n occurs for some n, for all n, for no n, or that some A_n does not occur.

Definition. A collection of events $\mathcal{A}$ is a σ**-algebra** if
(i) $\mathcal{A}$ contains $\mathcal{S}$, $\mathcal{S} \in \mathcal{A}$,
(ii) $\mathcal{A}$ is closed under complementation, $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
(iii) $\mathcal{A}$ is closed under countable unions, $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{n=1}^\infty A_n \in \mathcal{A}$.

Note. $\mathcal{S} \in \mathcal{A}$ and $\mathcal{S}^c = \emptyset \in \mathcal{A}$ by (i), (ii) for any $\mathcal{A}$. The simplest σ-algebra is $\{\emptyset, \mathcal{S}\}$.

Example (coin flips)
Exp. 1: flip once: $\mathcal{S} = \{0,1\}, \mathcal{A} = \{\emptyset, \mathcal{S}, \{0\}, \{1\}\}$ all possible subsets.
Exp. 2: flip n times and count only the number of heads; $\mathcal{A}$ contains $\emptyset$, $\{x_1 + \ldots + x_n = k\}, k = 0, 1, \ldots, n$, all unions.

Example. Consider the lifetime X of some component. Then $\mathcal{S} = [0, \infty)$ and $\mathcal{A}$ includes intervals $[a, b] =$ "$a \leq X \leq b$". The smallest σ-algebra is called Borel σ-algebra (also if $\mathcal{S} = (-\infty, \infty)$).

1.2 Axioms and properties

Definition. A probability space is $(\mathcal{S}, \mathcal{A}, P)$ where P is a probability measure on $\mathcal{A}$ satisfying Kolmogorov's axioms:
(i) $P(A) \geq 0$,
(ii) $P(\mathcal{S}) = 1$,
(iii) if $A_1, A_2, \ldots$ are disjoint, i.e. $A_m \cap A_n = \emptyset$ for $m \neq n$, then $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ (countably additive).

Theorem. P is finitely additive:

$$A_1, \ldots, A_n \text{ disjoint } \Rightarrow P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

Proof. $P(\bigcup_{i=1}^n A_i) = P(\cup_{i=1}^n A_i \cup \emptyset \cup \emptyset \cup \ldots) = P(\bigcup_{i=1}^\infty A_i)$ with $A_i = \emptyset$ for $i \geq n+1$. By Def. (iii) this is $\sum_{i=1}^n P(A_i) + \sum_{i=n+1}^\infty P(\emptyset) = \sum_{i=1}^n P(A_i)$ where we used that $P(\emptyset) = 0$ which follows from

$$1 = P(\mathcal{S}) = P(\mathcal{S} \cup \bigcup_{n=2}^{\infty} \emptyset) = \underbrace{P(\mathcal{S})}_{=1} + \sum_{n=2}^{\infty} \underbrace{P(\emptyset)}_{\geq 0}$$

using axioms (i), (ii) and (iii). □

Corollary (consequences of finite additivity).
(i) $P(\emptyset) = 0$,
(ii) $P(A^c) = 1 - P(A)$,
(iii) $P(A) \leq 1$,
(iv) $P(AB) + P(AB^c) = P(A)$
(v) $P(A) + P(B) = P(AB) + P(A \cup B)$,
(vi) $A \subset B \Rightarrow P(A) \leq P(B)$.

Proof of (v) (rest: exercise):
The left-hand side is $P(A)+P(B) = [P(AB)+P(AB^c)]+[P(AB)+P(A^cB] = (*)$;
now use $A \cup B = AB + AB^c + A^cB$ to obtain for the right-hand side
$P(AB) + P(A \cup B) = 2P(AB) + P(AB^c) + P(BA^c) = (*)$. □

Theorem
(i) $P(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$ (**Boole**),
(ii) If $P(A_i) \geq 1 - a_i$ then $P(\cap_{i=1}^n A_i) \geq 1 - \sum_{i=1}^n a_i$ (**Bonferroni**).

Proof (i): Use Cor. (v) to obtain $P(\cup_{i=1}^n A_i) = P(A_1) + P(\cup_{i=2}^n A_i) - P(A_1 \cap \bigcup_{i=2}^n A_i) \leq P(A_1) + P(\cup_{i=2}^n A_i) = P(A_1) + P(A_2) + P(\cup_{i=3}^n A_i) - P(A_2 \cap \bigcup_{i=3}^n A_i) \leq P(A_1) + P(A_2) + P(\cup_{i=3}^n A_i) \leq \ldots \leq \sum_{i=1}^n P(A_i)$.

(ii): Use *De Morgan*, the above corollary and *Boole* to obtain $P(\cap_{i=1}^n A_i) = P\{(\cup_{i=1}^n A_i^c)^c\} = 1 - P(\cup_{i=1}^n A_i^c) \geq 1 - \sum_{i=1}^n P(A_i^c) = 1 + \sum_{i=1}^n \{P(A_i) - 1\} \geq 1 + \sum_{i=1}^n (1 - a_i - 1)$ by assumption. □

Example. Suppose three statistical methods each have 95% confidence. What is the chance that all three are accurate (events A_1, A_2, A_3 occur)?
Answer: Use $P(A_i) \geq 95\% = 0.95 = 1 - a_i$ with $a_i = 0.05$ and Bonferroni to obtain $P(A_1 \cap A_2 \cap A_3) \geq 1 - 3 \cdot 0.05 = 0.85 = 85\%$.

Example. Flip a fair coin five times. Then $\mathcal{S}$ has 32 equally likely outcomes. To compute the chance of the event "heads at least once" it is easier to consider the complement: $P(x_1 + \ldots + x_5 \geq 1) = 1 - P(x_1 + \ldots + x_5 = 0) = 1 - P(\{0,0,0,0,0\}) = 1 - 1/32 = 31/32$.

Theorem. Suppose $\mathcal{S}$ is finite or countable and $A \subset \mathcal{S}$. Then

$$P(A) = \sum_{n: s_n \in A} P(\{s_n\}).$$

Proof. By countable additivity since $A = \cup_{n: s_n \in A} \{s_n\}$ is a union of disjoint events. □

What if $\mathcal{S}$ is uncountable, for **example** if X is a lifetime? Then $P(10 < X \leq 20)$ can be discussed but we cannot add up probabilities (∞ possibilities). Suppose (model!) $P(X > t) = e^{-t/10}$ for all $t > 0$, then $P(10 < X \leq 20) = P(X > 10) - P(X > 20) = e^{-1} - e^{-2}$. [Note that we can decompose $\mathcal{S}$ into countably many disjoint intervals, e.g. $\mathcal{S} = [0,1) \cup [1,2) \cup \ldots$]

Theorem. Let $B_1, B_2, \ldots$ be a partition of $\mathcal{S}$, i.e. $B_1, B_2, \ldots$ are disjoint and their union is $\mathcal{S}$. Then

$$P(A) = \sum_n P(AB_n) \quad \text{for any event } A.$$

Proof. $A = \cup_n AB_n$ with $AB_1, AB_2, \ldots$ disjoint, so it follows by countable additivity. □

Example. Roll a dice until a 6 appears and then stop. What is $P(A)$ for $A =$ "1 never appears"?
Let B_n be "the number of rolls is n". The B_n's partition $\mathcal{S}$ and $AB_n =$ "$n-1$ rolls of $2, 3, 4$ or 5 followed by a 6". Then, using the geometric series,

$$\begin{aligned} P(AB_n) &= \frac{4 \cdot 4 \cdot \cdots \cdot 4 \cdot 1}{6^n} = \frac{1}{6} \cdot \frac{4^{n-1}}{6^{n-1}} = \frac{1}{6}(\frac{2}{3})^{n-1} \\ P(A) &= \sum_{n=1}^\infty P(AB_n) = \frac{1}{6} \sum_{n=0}^\infty (\frac{2}{3})^n = \frac{1}{6} \cdot \frac{1}{1 - \frac{2}{3}} = \frac{1}{2}. \end{aligned}$$

Theorem (continuity).
(i) Let $A_1 \subset A_2 \subset \dots$ (increasing series of events) and $A = \cup_n A_n$. Then $P(A) = \lim_{n\to\infty} P(A_n)$.
(ii) Suppose $A_1 \supset A_2 \supset \dots$ (decreasing series of events) and $A = \cap_n A_n$. Then $P(A) = \lim_{n\to\infty} P(A_n)$.

Proof of (i). Let $B_1 = A_1$, $B_n = A_n \cap (\bigcup_{i=1}^{n-1} A_i)^c$ for $n > 1$. The B_n's are disjoint and $A_n = \cup_{i=1}^n B_i$. Thus $\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} P(\cup_{i=1}^n B_i) = \lim_{n\to\infty} \sum_{i=1}^n P(B_i) = \sum_{i=1}^\infty P(B_i)$. Use $s \in \cup_n B_n \iff s \in \cup_n A_n = A$ and additivity of P to obtain $\lim_{n\to\infty} P(A_n) = \sum_{i=1}^\infty P(B_i) = P(\cup_{i=1}^\infty B_i) = P(A)$. $\square$

Example (cont.). What is P("6 never appears")?
Let $A_n =$ "no 6 in the first n rolls". Then $A_1 \supset A_2 \supset \dots$, $A = \cap_n A_n =$ "no 6's ever" and $P(A) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} (5/6)^n = 0$.

1.3 Counting rules

Theorem.
(i) *Product rule*: if selection i ($i = 1, \dots, k$) has n_i possibilities irrespective of the other selections then the total number of possibilities is $\prod_{i=1}^k n_i$.
(ii) *Ordering rule*: there are $n!$ ways to order n items.
(iii) *Permutations rule*: the number of ordered (order identified) selections of n items from N is $\frac{N!}{(N-n)!}$.
(iv) The number of subsets (unordered selections) of n items from N is $\binom{N}{n} = \frac{N!}{(N-n)!n!}$.

Example (quality control). Consider a production lot of N components (M/N defective); take a random sample of size $n << N$ to inspect. Observe X defectives in the sample; estimate $p = M/N$ by $\hat{p} = X/n$. What can one predict about $\hat{p}$ (or X), which value is most likely?
Solution: consider sampling without replacement (SWOR), then $\mathcal{S}$ consists of $\binom{N}{n}$ equally likely outcomes (ignoring order). As there are $\binom{M}{x}$ possibilities to choose x defectives from M and $\binom{N-M}{n-x}$ possibilities to choose $n-x$ non-defectives, $P(X = x) = \binom{M}{x}\binom{N-M}{n-x}/\binom{N}{n}$ where $0 \le x \le n$, $n - N + M \le x \le M$. The most likely $\hat{p}$ is

$$\hat{p} = \frac{(M+1)(n+1)}{N+2} \cdot \frac{1}{n} = \frac{n+1}{n} \cdot \frac{M+1}{N+2} \approx \frac{M}{N} = p$$

which can be seen as follows:

$$\begin{aligned} 1 > \frac{P(X = x-1)}{P(X = x)} &= \frac{\binom{M}{x-1}\binom{N-M}{n-x+1}}{\binom{M}{x}\binom{N-M}{n-x}} = \frac{x}{M-x+1} \cdot \frac{N-M-n+x}{n-x+1} \\ &= 1 + \frac{(N+2)x - (M+1)(n+1)}{(M-x+1)(n-x+1)} \end{aligned}$$

$$\iff \frac{(N+2)x - (M+1)(n+1)}{(M-x+1)(n-x+1)} < 0 \iff x < \frac{(M+1)(n+1)}{N+2};$$

now choose the largest integer x below this bound.