Probability

Probability is the mathematical model to analyze experiments physical, chemical, sociological, etc. that don't have a predictable outcome, for example measuring fluctuating voltages or polling voters.

Examples: **You toss a coin; H or T? It is one or the other with probability $1/2$, provided the coin is fair.**
You throw two dice; what is the sum of the faces? It is $2, \ldots, 7, \ldots, 12$ with probability $1/36, \ldots, 1/6, \ldots, 1/36$, respectively. [How do I know?]

You wade into your corn field and pick an ear. What is its length, its circumference, the number of kernels on it?
That depends on which ear you picked.
What is the average number of kernels on an ear?
Safeway buys your crop only if 90% of your ears have more than 200 kernels on it; how big a sample of ears do you have test to make reasonably sure your crop meets this criterion?
You take a pregnancy test, and it is positive. What is the probability that you are actually pregnant? [Hint: It is zero if you are a boy.]
You ask 1052 voters whom they are going to vote for. 53% **say candidate X. What**

is the probability X will win? What does the question and its answer mean?

The preceding examples had this in common: you performed an experiment (toss coin, roll dice, pick ear, take test, ask voter) and made a decision about what to observe (H or T, sum of faces, number of kernels, test positive or negative, candidate preferred). The question was as to the probability of the possible outcomes, or a set of outcomes.

The common mathematical model for this is a pair $(S, \mathbb{P})$, where S is the State Space whose elements are called the Outcomes and whose subsets are called the Events,

and a function $\mathbb{P}$, called the Probability, on the power set of S that satisfies the axioms

$$0 \leq \mathbb{P}[A] \leq \mathbb{P}[B] \leq 1 \, , \qquad A \subseteq B \subseteq S \, ,$$

$$\mathbb{P}[\emptyset] = 0 \, , \; \mathbb{P}[S] = 1 \, ,$$

$$S \supseteq A, B \text{ disjoint } \implies \mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] \, .$$

For technical reasons we will assume that S is finite [or at most countable] and introduce the Probability Function p on S by

$$p(s) \stackrel{\text{def}}{=} \mathbb{P}[\{s\}] \qquad s \in S \, .$$

$p(s)$ is the probability of the outcome s, and clearly

$$\mathbb{P}[A] = \sum_{s \in A} p(s) \, , \qquad A \subseteq S \, .$$

Given a particular experiment (including the decision what to observe), it is most often quite clear what the model of the state space should be. In the examples above S should be $\{H, T\}$, $\{2, \ldots, 12\}$, $\{0, \ldots, 400\}$, $\{TP, TN\}$, list of candidate names. It is perhaps not so clear how to come by $\mathbb{P}$. Let us agree on what the probability $\mathbb{P}$ tells us:

for $A \subseteq S$ $\mathbb{P}[A]$ is the limiting frequency with which A occurs in ever more numerous independent repetitions of the experiment.

In the example of the ear of corn, the probability p of the event E that an ear

has 200 kernels or more,

$$\mathbb{P}[E] = \lim_{N\to\infty} \frac{\textbf{\# of ears with} \geq \textbf{200 kernels}}{\textbf{total \#}\ N\ \textbf{of ears harvested}}\ .$$

This is of course practically impossible. We should harvest only a few ears and come up with a useful estimate for $\mathbb{P}$.

A Priori Probability

In a large number of models with finite state space S any outcome is just as likely as any other, say $p = p(s) \quad \forall\, s \in S$. Then from

$$\mathbb{P}[S] = \sum_{s \in S} p(s) = \sum_{s \in S} p = |S| \times p = 1$$

we get

$$p = \frac{1}{|S|} \text{ and } \mathbb{P}[A] = \frac{|A|}{|S|} \qquad \forall A \subseteq S \;:$$

Determining $\mathbb{P}$ is reduced to counting.

Example: **Roll two distinguishable dice, die_1 and die_2, and observe both faces (not merely their sum). A suitable state space**

consists of all pairs (f_1, f_2), f_i denoting the face of die$_i$:

$$S \stackrel{\text{def}}{=} \left\{ \begin{array}{llllll} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

This state space

$$S \stackrel{\text{def}}{=} \{(f_1, f_2) : 1 \le f_i \le 6\}$$

has 36 elements, and there is *a priori* **no reason that any outcome is likelier than any other. Let us look at the event E that the sum of the faces equals 7:**

$$E = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} \ .$$

Its probability is

$$\mathbb{P}[E] = \frac{|E|}{|S|} = \frac{6}{36} = \frac{1}{6}\ .$$

Example: **What is the probability of getting a full house in 5–card poker?**

Its probability is

$$\mathbb{P}[E] = \frac{|E|}{|S|} = \frac{6}{36} = \frac{1}{6}\,.$$

Example: **What is the probability of getting a full house in 5–card poker?**
The experiment is dealing a hand. A suitable state space S has $\binom{52}{5}$ elements.

$|S| = \binom{52}{5}$**, not** $S = \binom{52}{5}$**!**

The event in question is the subset of all full houses. There are $\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}$ of them. The probability in question is

$$\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2} \Big/ \binom{52}{5}\,.$$

Lemma: **Let** $A, B \subseteq S$

$$\mathbb{P}[\overline{A}] = 1 - \mathbb{P}[A]$$
$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$$

Proof: $A \cap \overline{A} = \emptyset \implies \mathbb{P}[A] + \mathbb{P}[\overline{A}] = \mathbb{P}[S] = 1$.

$A \cup B = A \dot\cup (B \setminus A) \implies \mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B \setminus A]$.

$B = (B \setminus A) \dot\cup (A \cap B) \implies \mathbb{P}[B \setminus A] = \mathbb{P}[B] - \mathbb{P}[A \cap B]$.

[$C = A \dot\cup B$ **means** C **is the disjoint union of** A **and** B.]

Lemma: **For** $A, B, C \subseteq S$

$$\begin{aligned} \mathbb{P}[A \cup B \cup C] &= \mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] \\ &- \mathbb{P}[A \cap B] - \mathbb{P}[A \cap C] - \mathbb{P}[B \cap C] \\ &+ \mathbb{P}[A \cap B \cap C] . \end{aligned}$$

Lemma is Greek for Horn, like a horn to hang your hat on. What is a dilemma?

Example: **You toss n times a coin whose Head comes up with probability p.**
a) Describe a suitable state space.

Example: **You toss n times a coin whose Head comes up with probability p.**
a) Describe a suitable state space.

$$S \stackrel{\text{def}}{=} \{(C_1, \ldots, C_n) : C_i \in \{H, T\}\} .$$

b) What is the probability that any given n–tuple like

$$(H, H, T, T, T, H, \ldots) ,$$

appears?

Example: You toss n times a coin whose Head comes up with probability p.
a) Describe a suitable state space.

$$S \stackrel{\text{def}}{=} \{(C_1, \ldots, C_n) : C_i \in \{H, T\}\} \ .$$

b) What is the probability that any given n–tuple like

$$(H, H, T, T, T, H, \ldots) \ ,$$

appears?

$$p^{\#\ of\ H}(1-p)^{\#\ of\ T}$$

c) What is the probability that exactly k Heads appear in the n tosses?

c) What is the probability that exactly k Heads appear in the n tosses?

$$\binom{n}{k} p^k (1-p)^{1-k} \ .$$

This is the Binomial Distribution.

d) What is the probability that H appears for the first time on the k^{th} toss?

c) What is the probability that exactly k Heads appear in the n tosses?

$$\binom{n}{k} p^k (1-p)^{1-k} \ .$$

This is the Binomial Distribution.

d) What is the probability that H appears for the first time on the k^{th} toss?

$$(1-p)^{k-1} p \ .$$

This does not depend on n!
This is the Geometric Distribution.

Conditional Probability

Example: **Roll two distinguishable dice, and consider the events**

$A:$ **Faces add to 9 or more** $\quad \mathbb{P}[A] = \frac{5}{18}$

$B:$ **A double appears** $\quad \mathbb{P}[B] = \frac{1}{6}$

$A \cap B:$ **both events occur** $\quad \mathbb{P}[A \cap B] = \frac{1}{18}$

$$S \stackrel{\text{def}}{=} \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6), \end{array} \right\}$$

Now we ask: what is the probability of B **given that** A **has occurred?**

it is $$\mathbb{P}[B|A] \stackrel{\text{def}}{=} \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]} = \frac{1/18}{5/18} = \frac{1}{5},$$

"the probability of B given A."
The question amounts to reducing the state space to A and asking for the occurrence of $B \cap A$.

This example contains some red herrings; that S is finite and the probability function is constant is immaterial:

Definition: Let $A, B \subseteq S$, with $\mathbb{P}[B] \neq 0$. The Conditional Probability of A given B is defined by

$$\mathbb{P}[A|B] \stackrel{\text{def}}{=} \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \, .$$

Note this consequence:

$$\mathbb{P}[A \cap B] = \mathbb{P}[B|A] \cdot \mathbb{P}[A] \, .$$

Bayes' Theorem: Let $C_1, \ldots, C_n$ be mutually disjoint events that cover the state space S, a Paving of S, and A another event. Then

$$\mathbb{P}[C_1|A] = \frac{\mathbb{P}[A|C_1] \cdot \mathbb{P}[C_1]}{\sum_i \mathbb{P}[A|C_i] \cdot \mathbb{P}[C_i]} \, .$$

Proof:

$$\mathbb{P}[C_1|A] = \frac{\mathbb{P}[C_1 \cap A]}{\mathbb{P}[A]} = \frac{\mathbb{P}[A|C_1] \cdot \mathbb{P}[C_1]}{\mathbb{P}[\bigcup_i A \cap C_i]}$$

$$= \frac{\mathbb{P}[A|C_1] \cdot \mathbb{P}[C_1]}{\sum_i \mathbb{P}[A \cap C_i]} = \frac{\mathbb{P}[A|C_1] \cdot \mathbb{P}[C_1]}{\sum_i \mathbb{P}[A|C_i] \cdot \mathbb{P}[C_i]}$$

In the case that the paving consists of a set C and its complement $\overline{C}$, Bayes' formula takes the form

$$\mathbb{P}[C|A] = \frac{\mathbb{P}[A|C] \cdot \mathbb{P}[C]}{\mathbb{P}[A|C] \cdot \mathbb{P}[C] + \mathbb{P}[A|\overline{C}] \cdot \mathbb{P}[\overline{C}]} .$$

Example: The stage shows three closed doors, and you know that behind one of them is a car, behind the other two a goat each. You are allowed to pick one of the doors. Once you have picked one, the moderator opens one of the other two doors, one that has a goat behind it; you are then allowed to either stick to your original pick or to switch to the other unopened door. If your final door chosen has the car behind it, you win it.
Should you switch, or does it make no difference?

Example: The stage shows three closed doors, and you know that behind one of them is a car, behind the other two a goat each. You are allowed to pick one of the doors. Once you have picked one, the moderator opens one of the other two doors, one that has a goat behind it; you are then allowed to either stick to your original pick or to switch to the other unopened door. If your final door chosen has the car behind it, you win it.
Should you switch, or does it make no difference?
Let us compare the winning chances of two players, the Switcher and the Sticker.
Let C be the event that the first door picked had the car behind it, W the event

that the player wins the car.

The Sticker:

$$\begin{aligned}\mathbb{P}[W] &= \mathbb{P}[W \cap C] && + \mathbb{P}[W \cap \overline{C}] \\ &= \mathbb{P}[W|C] \times \mathbb{P}[C] && + \mathbb{P}[W|\overline{C}] \times \mathbb{P}[\overline{C}] \\ &= 1 \times 1/3 && + 0 \times 2/3 \\ &= 1/3 \ .\end{aligned}$$

The Switcher:

$$\begin{aligned}\mathbb{P}[W] &= \mathbb{P}[W \cap C] && + \mathbb{P}[W \cap \overline{C}] \\ &= \mathbb{P}[W|C] \times \mathbb{P}[C] && + \mathbb{P}[W|\overline{C}] \times \mathbb{P}[\overline{C}] \\ &= 0 \times 1/3 && + 1 \times 2/3 \\ &= 2/3 \ .\end{aligned}$$

The Switcher doubles his chance!

Example: **You go to the doctor for a routine checkup, in which she also administers a test for the condition C. The test is not 100% accurate; instead, the test shows positive (event TP) only with probability $.95$ if you actually do have the condition, and with probability $.05$ even if you don't:**

$$\mathbb{P}[TP|C] = 0.95 \text{ and } \mathbb{P}[TP|\overline{C}] = 0.05 \ .$$

In addition, the CDC estimates that 2% of the people going for a routine checkup have the condition:

$$\mathbb{P}[C] = 0.02 \ .$$

Panic: The test shows positive!
What to do?

Example: **You go to the doctor for a routine checkup, in which she also administers a test for the condition C. The test is not 100% accurate; instead, the test shows positive (event TP) only with probability $.95$ if you actually do have the condition, and with probability $.05$ even if you don't:**

$$\mathbb{P}[TP|C] = 0.95 \textbf{ and } \mathbb{P}[TP|\overline{C}] = 0.05 \ .$$

In addition, the CDC estimates that 2% of the people going for a routine checkup have the condition:

$$\mathbb{P}[C] = 0.02 \ .$$

Panic: The test shows positive!
What to do?

You compute the probability that you actually have the condition:

$$\mathbb{P}[C|TP] = \frac{\mathbb{P}[TP|C]\cdot\mathbb{P}[C]}{\mathbb{P}[TP|C]\cdot\mathbb{P}[C]+\mathbb{P}[TP|\overline{C}]\cdot\mathbb{P}[\overline{C}]}$$

$$= \frac{.95\times .02}{.95\times .02 + .05\times .98} \approx 28\% \ .$$

You feel a little relief but still want to know. You can go for a more accurate (and presumably more expensive) test, or take the same one over. The difference the second time around is that you now belong to a population having $\mathbb{P}[C] = .28$. Let $T2P$ denote the event that the Second Test is Positive and compute:

$$\mathbb{P}[C|T2P] = \frac{\mathbb{P}[T2P|C]\cdot\mathbb{P}[C]}{\mathbb{P}[T2P|C]\cdot\mathbb{P}[C]+\mathbb{P}[T2P|\overline{C}]\cdot\mathbb{P}[\overline{C}]}$$

$$= \frac{.95\times .28}{.95\times .28 + .05\times .72} \approx 88\% \ .$$

Now you are really worried. You may want to take the test a third time:

$$\mathbb{P}[C|T3P] = \frac{\mathbb{P}[T3P|C]\cdot\mathbb{P}[C]}{\mathbb{P}[T3P|C]\cdot\mathbb{P}[C]+\mathbb{P}[T3P|\overline{C}]\cdot\mathbb{P}[\overline{C}]}$$

$$= \frac{.95\times.88}{.95\times.88 + .05\times.12} \approx 99\% \ .$$

Suppose however, the second test is negative. Then

$$\mathbb{P}[C|\overline{T2P}] = \frac{\mathbb{P}[\overline{T2P}|C]\cdot\mathbb{P}[C]}{\mathbb{P}[\overline{T2P}|C]\cdot\mathbb{P}[C]+\mathbb{P}[\overline{T2P}|\overline{C}]\cdot\mathbb{P}[\overline{C}]}$$

$$= \frac{.05\times.28}{.05\times.28 + .95\times.72} \approx 2\% \ .$$

You should be just as [un–]concerned about having C as you were before the first visit.

Independence

Definition: **Two events $A, B \subseteq S$ are independent, written $A \perp\!\!\!\perp B$, if**

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B] \; .$$

If $\mathbb{P}[B] \neq 0$ we divide by it and find that A, B are independent if

$$\mathbb{P}[A] = \mathbb{P}[A|B] \quad \Big(\textbf{ or } \mathbb{P}[B] = \mathbb{P}[B|A] \Big) \; :$$

i.e., if knowledge that B has occurred does not alter the probability that A will occur and vice versa.

Example: **Roll two distinguishable dice. The events $[\text{die}_1 = 2]$ and $[\text{die}_2 = 5]$ are**

independent. So are the events [die$_1 = 2$] and [sum of faces $= 7$].

$$\left\{\begin{array}{llllll}
(1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\
(2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\
(3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\
(4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\
(5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\
(6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6)
\end{array}\right\}$$

Two disjoint events of strictly positive probability are NOT independent.

Lemma: **If** $A \perp\!\!\!\perp B$ **then**

$$A \perp\!\!\!\perp \overline{B} \text{ and } \overline{A} \perp\!\!\!\perp B \text{ and } \overline{A} \perp\!\!\!\perp \overline{B} \,.$$

Proof:

$$\begin{aligned}\mathbb{P}[A \cap \overline{B}] &= \mathbb{P}[A] - \mathbb{P}[A \cap B] = \mathbb{P}[A] - \mathbb{P}[A] \cdot \mathbb{P}[B] \\ &= \mathbb{P}[A](1 - \mathbb{P}[B]) = \mathbb{P}[A] \cdot \mathbb{P}[\overline{B}] \,.\end{aligned}$$

Use with $A \longrightarrow B$ and $A \longrightarrow \overline{A}$. QED

Definition: **The family $\{A_1, A_2, \ldots, A_n\} \subseteq \mathcal{P}[S]$ is independent if for every subfamily $\{A_{n_1}, \ldots, A_{n_k}\}$**

$$\mathbb{P}\Big[\bigcap_{i=1}^{k} A_{n_i}\Big] = \prod_{i=1}^{k} \mathbb{P}[A_{n_i}] \,.$$

This means that knowledge that any number of the A_i have occurred does not help reassessing the probability of any event A_j not in that number.

Let $E \subset S$. Are E and its complement $\overline{E}$ independent?

Let $E \subset S$**. Are** E **and its complement** $\overline{E}$ **independent?**
E **and its complement** $\overline{E}$ **are independent iff** $\mathbb{P}[E] = 0$ **or** $\mathbb{P}[\overline{E}] = 0$**.**

In our example of two dice the three events $[\textbf{die}_1 = 2]$**,** $[\textbf{die}_2 = 5]$**,** $[\textbf{sum} = 7]$ **are not independent, yet any pair among them is.**

Example, Independent Trials: Suppose you have an experiment with finite state space S and probability function

$$p(s) = \mathbb{P}[\{s\}] \,, \qquad\qquad s \in S$$

— the "primitive" experiment — and you repeat it independently N times – independently means that the outcomes of any repetition are not influenced by the outcomes of any or all earlier or later repetitions. Then a reasonable state space for the new experiment that consists of the N repetitions of the "primitive" experiment is the N–fold cartesian Product

$$S^N \stackrel{\text{def}}{=} \{(s_1, s_2, \ldots, s_N) : s_i \in S\}$$

and the probability function on it is given

by

$$p\big((s_1, s_2, \ldots, s_N)\big) \stackrel{\text{def}}{=} p(s_1) \cdot p(s_2) \cdots p(s_N) \ .$$

Example, Independent Bernoulli Trials: A Bernoulli trial is an experiment with only two outcomes H and T such that $\mathbb{P}[\{H\}] = p$ and $\mathbb{P}[\{T\}] = 1 - p$. Repeat it N times. A typical outcome is

$$\underbrace{HHT \cdots TTTH}_{N \text{ letters}} \ .$$

What is the probability of $[\# \text{ of } H = k]$?

by

$$p\big((s_1, s_2, \ldots, s_N)\big) \stackrel{\text{def}}{=} p(s_1) \cdot p(s_2) \cdots p(s_N) \ .$$

Example, Independent Bernoulli Trials: A Bernoulli trial is an experiment with only two outcomes H and T such that $\mathbb{P}[\{H\}] = p$ and $\mathbb{P}[\{T\}] = 1 - p$. Repeat it N times. A typical outcome is

$$\underbrace{HHT\cdots TTTH}_{N \text{ letters}} \ .$$

What is the probability of $[\#$ of $H = k]$? An outcome like

$$\underbrace{HHT\cdots TTTH}_{k\ H\text{s and } N-k\ T\text{s}} \ .$$

has probability $p^k(1-p)^{N-k}$. There are $\binom{N}{k}$

of them:

$$\mathbb{P}[\#\ \textbf{of}\ H = k] = \binom{N}{k} p^k (1-p)^{N-k}\ .$$

This is the Binomial Distribution.

Keep repeating the Bernoulli experiment infinitely often. What is the probability of the event E that the outcome H appears first on the $k^{\textbf{th}}$ repetition?

of them:

$$\mathbb{P}[\#\ \textbf{of}\ H = k] = \binom{N}{k} p^k (1-p)^{N-k}\ .$$

This is the Binomial Distribution.

Keep repeating the Bernoulli experiment infinitely often. What is the probability of the event E that the outcome H appears first on the $k^{\textbf{th}}$ repetition?

$$\mathbb{P}[E] = (1-p)^{k-1} p\ .$$