Elementary Probability

Games of chance have a long history...

Cubical Die from Tepe Gawra. Photo of Object 31-52-309 of the Penn Museum. Online: https://www.penn.museum/collections/object/332432. Described in Brown, W. N. “Indian Games of Pachisi, Chaupar, and Chausar”. Expedition: The Magazine of the University of Pennsylvania Museum of Archaeology and Anthropology. Philadelphia. The University Museum. 1964. Vol. 6. no. 3. Pages 32-35

… but probability in mathematics does not. Why?

- ***Platonism***: real-life objects are imperfect representations of ideal "Platonic Forms". A six-sided die is a representation of an ideal cube.
- But randomness appears tied to real-life processes - no platonic form.
- Randomness was not considered amenable to mathematics.

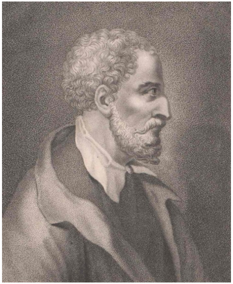
Detail of the "School of Athens" by Rafael. Wall Fresco in the Vatican, Stanza della Segnatura. 1509. File:Raffael 058.jpg. (2019, September 1). Wikimedia Commons, the free media repository.

… but probability in mathematics does not. Why?

- ***Divination***: randomness was often used to predict the future.
- Predicting randomness = interfering with the will of the gods.

Runestones. Online: https://www.needpix.com/photo/download/1235148/divination-background-krupnyj-plan-the-consignment-a-few-runes-stones-scrying-stones-bone.

Girolamo Cardano (1501-1576)



Girolamo Cardano (1501-1576). Etching by Carl Mayer. Dated 1813/1863 Online: https://picryl.com/media/cardano-girolamo-1f9a09.

- Invented cardan shaft
- Published solutions to cubic and quartic equations
- First systematic use of negative numbers in Europe; acknowledged imaginary numbers
- Heavy gambler, known to be short of money
- Published first systematic treatment of probability

1.1. Cardano's Principle. Let A be a random outcome of an experiment that may proceed in various ways. Assume each of these ways is equally likely. Then the probability $P[A]$ of the outcome A is

$$P[A] = \frac{\text{number of ways leading to outcome } A}{\text{number of ways the experiment can proceed.}}$$

Two Die Rolls

It is clear that the probability is a real number between 0 and 1.

1.2. Example. Two six-sided dice are rolled. Both are fair dice and have equal probability of returning any given number.

What is the probability that the sum of the results is 11 or 12?

There are six possible results for the first die and 6 possible results for the second die, so there are a total of $6 \cdot 6 = 36$ ***possible outcomes***.

The outcomes that give a result of 11 or 12, writing the outcomes as (first die, second die), are:

- outcome $(6, 6)$ gives the sum 12;
- outcome $(5, 6)$ gives the sum 11;
- outcome $(6, 5)$ gives the sum 11;
- all other outcomes will give a sum of 10 or less.

Two Die Rolls

The probability that the sum of the results is at least 11 is

$$\frac{\text{number of outcomes leading to a sum of 11 or 12}}{\text{number of possible outcomes}} = \frac{3}{36} = \frac{1}{12}.$$

When applying Cardano's principle, it is crucial that all outcomes are equally likely!

Tossing a Coin 10 Times

A fair coin is tossed ten times and the result of “heads” (h) or “tails” (t) is recorded each time. Which of the following sequences of results is the most likely:

(a) $h, \ h, \ t, \ h, \ t, \ t, \ t, \ h, \ h, \ t;$

(b) $h, \ h, \ h, \ h, \ h, \ t, \ t, \ t, \ t, \ t;$

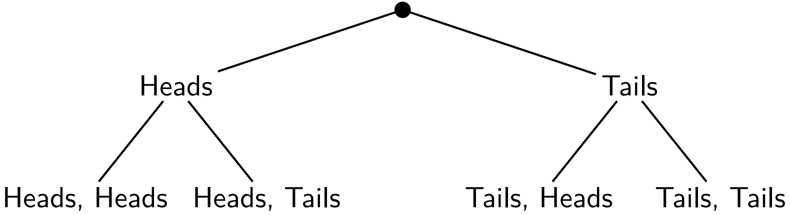
(c) $h, \ h, \ h, \ h, \ h, \ h, \ h, \ h, \ h, \ h;$

(d) None of them (they are all equally likely).

Two Coin Tosses

1.3. Example. A fair coin is tossed twice and the result of "heads" (h) or "tails" (t) is recorded each time. What is the probability of obtaining at least one head?

We use a tree diagram to visualize the possible outcomes:



We see that

$$P[\text{at least 1 Head}] = \frac{3}{4}.$$

D'Alembert's Error

Portrait de Jean Le Rond d'Alembert. de la Tour, Maurice Quentin. 1753. Painting. Musee de Louvre. Paris. File:Alembert.jpg. (2020, January 29). Wikimedia Commons, the free media repository.

D'Alembert asserted in 1754 that it is erroneous to consider the four cases (h, h), (h, t), (t, h), (t, t) since the experiment can be stopped immediately if heads comes up on the first toss.

Therefore, he claimed, there are only three outcomes:

- heads;
- tails, then heads;
- tails, then tails;

so the probability of obtaining at least one head should be $2/3$, not $3/4$.

Of course, the error in his thinking is that not all of the three outcomes that he cites are equally likely.

Literature: http://www.cs.xu.edu/math/Sources/Dalembert/croix_ou_pile.pdf

Bose's inspiration

If the experiment were modified so that two coins were tossed at the same time, the result remains the same.

But if these two coins were ***indistinguishable***, so that the results (h, t) and (t, h) could not be told apart, then d'Alembert's reasoning would be correct.

In the early 1920's, the Indian physicist Satyendra Nath Bose was working on the energy distribution of elementary particles such as photons. Contemporary theory could not explain the experimental data.

Satyendra Nath Bose (1894-1974) in Paris. 1925. Photography. Siliconeer, August 2000, Vol. 1,7 File:SatyenBose1925.jpg. (2020, January 27). Wikimedia Commons, the free media repository.

In a calculation during a lecture, Bose made a mistake similar to the one described here. He discovered that, based on the mistake, the calculations turned out to correctly describe the data. From this he deduced that photons (and related particles) are indistinguishable – there is in principle no physical way to tell two photons apart.

Basic Principles of Counting

Suppose a set A of n objects is given.

- There are $\frac{n!}{(n-k)!}$ different ways of choosing an ordered tuple of k objects from A.

  Such a choice is called a ***permutation of k objects*** from A.

- There are $\frac{n!}{k!(n-k)!}$ different ways of choosing an unordered set of k objects from A.

  Such a choice is called a ***combination of k objects*** from A.

- There are $\frac{n!}{n_1!n_2!\ldots n_k!}$ ways of partitioning A into k disjoint subsets $A_1, \ldots, A_k$ whose union is A, where each a_i has n_i elements.

  This is called a ***permutation of k indistinguishable objects*** from A.

Binomial Coefficients

We define binomial coefficients by

$$\binom{\alpha}{0} := 1, \qquad \text{for } \alpha \in \mathbb{R} \tag{1.1}$$

and, for $n \in \mathbb{N} \setminus \{0\}$ and $\alpha \in \mathbb{R}$,

$$\binom{\alpha}{n} := \frac{\alpha \cdot (\alpha - 1) \cdot (\alpha - 2) \cdots (\alpha - n + 1)}{n!}. \tag{1.2}$$

If $\alpha \in \mathbb{N}$, this may be expressed as the perhaps more familiar

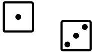
$$\binom{\alpha}{n} = \frac{\alpha!}{(\alpha - n)!n!}.$$

The definition (1.2) also implies that

$$\binom{m}{n} = 0 \qquad \text{whenever } n > m \text{ and } m, n \in \mathbb{N}. \tag{1.3}$$

Sample Spaces and Sample Points

We want to translate physical outcomes into mathematical objects, for example:



$$\longmapsto \qquad (1, 3)$$

or

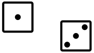
$$\longmapsto \qquad 173\,\text{cm}$$

The mathematical objects are called ***sample points***. They can be numbers, pairs of numbers or any sort of abstract object.

We need to define a ***sample space***, often denoted S, large enough to accommodate all sample points.

Events

The sample space can be larger than seems necessary:



$$\longmapsto \qquad (1,3) \in \mathbb{N}^2 = \mathbb{N} \times \mathbb{N}$$

physical outcome ***sample point*** ***sample space***

An outcome in the sense of Cardano's principle is then interpreted as a subset A of a sample space S and called an ***event***.

Two events A_1, A_2 are called ***mutually exclusive*** if $A_1 \cap A_2 = \emptyset$.

Events

1.4. Example. A six-sided die is rolled four times. The sample space can be taken to be $S = \mathbb{N}^4$ and a sample point is a 4-tuple, for example $(1, 2, 5, 2) \in \mathbb{N}^4$. This sample point would correspond to first rolling a 1, then a 2, next a 5, followed by a 2.

Many tuples, such as $(7, 20, 2, 3) \in S$ do not correspond to any physical outcome of the experiment.

An event might be "rolling at least two fours" in which case this would be a subset $A \subset S$ such that each 4-tuple in A has at least two entries equal to 4. For example, $(1, 3, 4, 4) \in A$ but $(1, 1, 3, 4) \notin A$.

We can then apply counting principles to subsets of sample spaces in order to find the probabilities of events by Cardano's principle.

Probabilities of Events

1.5. Example. We roll a four-sided die 10 times. What is the probability of obtaining 5 ones, 3 twos, 1 three and 1 four?

There are $4^{10} = 1048576$ possibilities for the 10-tuple of results of the die rolls, corresponding to that many sample points in $S = \mathbb{N}^{10}$ that correspond to physical results. The event A consists of all ordered 10-tuples containing 5 ones, 3 twos, 1 three and 1 four. There are

$$\frac{10!}{5!3!1!1!} = 5040$$

possible ways of obtaining 5 ones, 3 twos, 1 three and 1 four, so there are that many elements in A. The probability is

$$\frac{5040}{1048576} \approx 0.00481 \approx 0.5\%.$$

A Dartboard

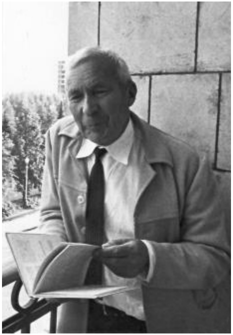
Consider the dartboard shown below:

The red and white rings have equal width. Suppose a dart is launched and hits a point on the board entirely at random. Hitting a red ring is the event A_1 and hitting a white ring is the event A_2.

Is $P[A_1] = P[A_2]$? Why or why not? Does the radial symmetry play a role?

An Axiomatic Approach

Clearly, for more complicated situations of randomness that go beyond simple counting, a more formal model of probability is needed. In 1933, the Russian mathematician Kolmogorv introduced an axiomatic approach.



Andrej N. Kolmogorov (1903-1987). File:Andrej Nikolajewitsch Kolmogorov.jpg. (2018, December 28). Wikimedia Commons, the free media repository.

Given a sample space S, we first need to determine the set of permissible events.

If S has a finite number of elements, then we can simply allow any subset of S to be an event. However, if S is very large (for example, if $S = \mathbb{R}$) then a more careful approach is needed.

Not every subset of S may be an allowable event. However, we need to choose “allowable” subsets in a consistent way.

A σ-Field of Subsets

Suppose that a non-empty set S is given. A ***σ-field*** $\mathscr{F}$ on S is a family of subsets of S such that

(i) $\emptyset \in \mathscr{F}$;

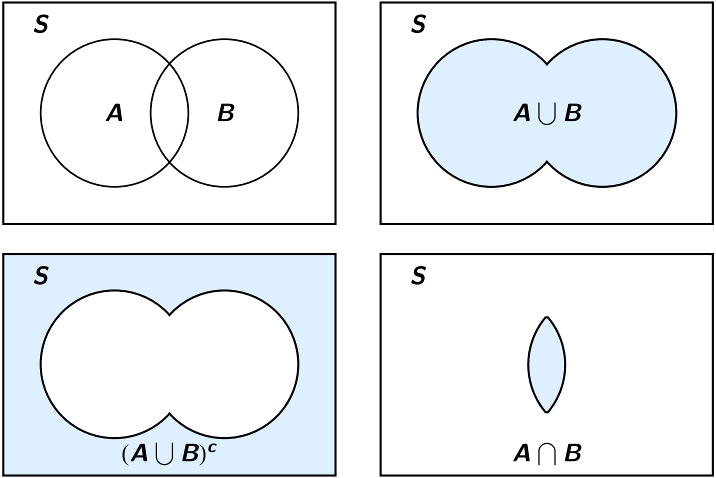
(ii) if $A \in \mathscr{F}$, then $S \setminus A \in \mathscr{F}$;

(iii) if $A_1, A_2, A_3, \ldots \in \mathscr{F}$ is a finite or countable sequence of subsets, then the union $\bigcup_k A_k \in \mathscr{F}$.

In probability, we consider families of events that are σ-fields. This is clearly reasonable, since the above properties guarantee that if a subset is an event, then so is the complement and if two subsets are events, then their union must also be an event. Furthermore, the entire set S is an event.

A σ-Field of Subsets

This is illustrated by the diagrams below:



Examples of σ-Fields of Events

- If S is finite, one can take $\mathscr{F} = \mathscr{P}(S)$ (the power set of S) without problems. This is also the case if S is countable.
- For any set S, the smallest possible σ-field is $\mathscr{F} = \{\emptyset, S\}$.
- One of the most important σ-fields in practice is the set $\mathscr{B}(I)$, the set of ***Borel sets*** on an interval $I \subset \mathbb{R}$. This is the smallest σ-family containing all subintervals of I. (We do not give an explicit definition here.)

Now that we have a sample space and a set of permissible events, we need a probability function that assigns in principle to every event the probability of that event occurring.

Probability Measures and Spaces

Let S be a sample space and $\mathscr{F}$ a σ-field on S. Then a function

$$P\colon \mathscr{F} \to [0,1], \qquad\qquad A \mapsto P[A],$$

is called a ***probability measure*** (or ***probability function*** or just ***probability***) on S if

(i) $P[S] = 1$,

(ii) For any set of events $\{A_k\} \subset \mathscr{F}$ such that $A_j \cap A_k = \emptyset$ for $j \neq k$,

$$P\Big[\bigcup_k A_k\Big] = \sum_k P[A_k].$$

The triple $(S, \mathscr{F}, P)$ is called a ***probability space***.

Rolling a Die Twice

1.6. Example. Suppose we roll a six-sided die twice. The we can take the sample space to have 36 elements as follows

$$\begin{aligned} S &= \{(j, k) \colon j, k = 1, \ldots, 6\} \\ &= \big\{(1,1),\ (1,2),\ \ldots,\ (6,5),\ (6,6)\big\}. \end{aligned}$$

We take the σ-field to be the power set $\mathscr{P}(S)$. Following Cardano's approach, we then assign the probability

$$P[\{(i,j)\}] = \frac{1}{36}, \qquad\qquad \text{for } i,j = 1,2,3,4,5,6$$

for each individual sample point. This allows us to define probabilities for an arbitrary event in $\mathscr{P}(S)$.

Rolling a Die Twice

Let A_1 be the event that corresponds to the outcome "the sum of the two die rolls is at most 3" and A_2 correspond to the outcome "the two die rolls give the same number". Then

$$A_1 = \{(1,1),(1,2),(2,1)\}, \quad A_2 = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}.$$

The probability of these events is calculated as

$$P[A_1] = P[\{(1,1)\}] + P[\{(1,2)\}] + P[\{(2,1)\}] = \frac{3}{36},$$
$$P[A_2] = \frac{6}{36} = \frac{1}{6}.$$

The event "the sum of two die rolls is at most three and both rolls are the same" is given by the set containing those sample points both in A_1 and in A_2. We calculate its probability to be

$$P[A_1 \cap A_2] = P[\{1,1\}] = \frac{1}{36}.$$

Almost Sure Occurrence

An event $A \in \mathscr{F}$ is said to occur ***almost surely*** if $P[A] = 1$.

1.7. Example. Suppose we toss a fair coin repeatedly. If it turns heads up, we stop, otherwise we continue to toss. The sample space may be taken to record the tosses as strings of "t" (for tails) and "h" (for heads), i.e.,

$$S = \big\{h,\ th,\ tth,\ ttth, \dots\big\} \cup \big\{t^\infty\big\}.$$

where "t^∞" stands for an infinite sequence of tosses yielding tails. This a countable set and we can simply take $\mathscr{F} = \mathscr{P}(S)$.

In order to define a probability function, we set

$$P[\{\underbrace{t \cdots t}_{n \text{ times}} h\}] = \frac{1}{2^{n+1}}$$

and $P[\{t^\infty\}] = 0$.

Almost Sure Occurrence

Then the event "Eventually the coin turns up heads." is given by taking the union of all sample points that include h. We calculate

$$\begin{aligned} P[A] &= P[\{h\}] + P[\{th\}] + P[\{tth\}] + \cdots \\ &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots \\ &= 1. \end{aligned}$$

We say that "Almost surely, the coin will turn up heads eventually." However, it is not in principle inconceivable that we toss the coin forever and never see heads turn up. However, the probability of this happening is zero.

Basic Properties of Probabilities of Events

We end by listing some general properties that follow immediately from the definition of a probability space $(S, \mathscr{F}, P)$:

$$
\begin{aligned}
P[S] &= 1, \\
P[\emptyset] &= 0, \\
P[S \setminus A] &= 1 - P[A], \\
P[A_1 \cup A_2] &= P[A_1] + P[A_2] - P[A_1 \cap A_2],
\end{aligned}
$$

where $A, A_1, A_2 \in S$ are any events.