INFN School of Statistics 2019
Paestum, Italy

Probability Theory - 1

Harrison B. Prosper

Department of Physics, Florida State University

June 3, 2019

Outline

1. Introduction
   - Learning To Count
2. Axioms
   - Boolean Algebra
   - Kolmogorov Axioms
3. Conditional Probability
   - Bayes Theorem
4. What is Probability?

Outline

Consider the statements:

- The odds of drawing four aces one after another from a standard deck of cards is 1 in 270,725.
- In 2017, the chance of dying in a car crash in Italy was about 1 in 18,000.
- It is highly likely that William Shakespeare is the author of the inspired insult:

  "You blocks, you stones, you worse than senseless things!"

Each statement is about probability. Since the latter is the foundational concept in statistics, it makes sense to start this School with a brief survey of probability theory.

Example (1.1 Chevalier de Méré's Problem)

Which of the following outcomes is the more probable, getting *at least*

1. one 6 in 4 throws of a single 6-sided die, or
2. a double 6 in 24 throws of two 6-sided dice?

In 1654, Antoine Gombaud (aka de Méré) brought this problem to the attention of Blaise Pascal who started a correspondence with Pierre de Fermat (1601-1665). The Pascal-Fermat letters together with work a century earlier by the Italian mathematician Gerolamo Cardano (1501 - 1576) constitute the first foundation of a mathematical theory of probability.

Blaise Pascal
(1623 - 1662)

Let's review the solution to this problem.

Example (1.1 Chevalier de Méré's Problem)

What is the probability to get *at least*

1. one 6 in 4 throws of a single 6-sided die?

Let p be the probability to obtain at least 1 six in 4 throws of a die and q the probability to get 0 sixes. We shall refer to the 4 throws of the die as an **experiment**.

Now to the solution.

> *Alas, there is none! Unless ...*
>
> *...one is prepared to make a sufficient number of assumptions to render the problem well-posed. Moreover, because different people may make different assumptions, there may be different answers to the same problem.*

Solution contd.

Assumptions

1. The two experimental outcomes, either 0 sixes or ≥ 1 sixes, are exhaustive — i.e., they are the only possible outcomes.
2. The 6 elementary outcomes[1] are equally probable.
3. The elementary outcomes are equally probable for every throw and throws are independent of each other.

Assumption 1 $\implies$ $p + q = 1$.
Assumption 2 $\implies$ since there are 5 ways *not* to get a six, the probability *not* to get a six is $5/6$.
Assumption 3 $\implies$ the probability not to get a six in 4 throws of the die is $q = (5/6)^4$. Therefore, $p = 1 - (5/6)^4$. Similar reasoning for the throws of two dice yields $p = 1 - (35/36)^{24}$, which is smaller than the first probability by about 5%.

[1] From which the experimental outcomes are constructed.

Another way to approach de Méré's problem is by counting outcomes:

- break the problem down into outcomes that are considered equally likely;
- count the total number of possible outcomes T;
- count the number of favorable outcomes S and
- take the probability to be $p = S/T$.

The outcome of the experiment can be represented by a 4-tuple (z_1, z_2, z_3, z_4), where $z_i \in \{1, 2, 3, 4, 5, 6\}$. The total number of 4-tuples, that is, experimental outcomes, is 6^4. The total number of outcomes without a six is 5^4. Therefore, the number of outcomes with a six is $6^4 - 5^4$. Consequently, assuming that each experimental outcome is equally probable, the probability of a favorable outcome is $p = (6^4 - 5^4)/6^4 = 1 - (5/6)^4$.

More generally, we have to consider permutations and combinations...

Permutations

How many ways can n items be arranged in a row with k slots? The first slot can be filled in n ways, the 2nd in $(n-1)$ ways, the third in $(n-2)$ ways and so on until the last slot is reached, which can be filled in $n-k+1$ ways. This yields $n!/(n-k)!$ arrangements. When $k=n$ we get $n!$ permutations.

Combinations

For each set of k items, there are $k!$ permutations that consist of rearrangements of the items in the k slots. If the order of the items is irrelevant then the number of *distinct* arrangements is smaller by $k!$, that is, the number of combinations is

$$\frac{n!}{(n-k)!\,k!} \equiv \binom{n}{k}.$$

Example (1.2 The Birthday Problem)

A crowd of people is randomly assembled. How large must the crowd be so that the probability $p \geq 0.5$ of finding at least two people with the same birthday?

Assumptions

1. There are 365 possible birthdays (ignoring leap years).
2. Every birthday is equally probable and this remains unchanged as the crowd is assembled.

Solution Consider crowds of size n. The outcome of an experiment can be modeled as an n-tuple, with each slot in the n-tuple containing an integer from 1 to 365.

Let M be the size (cardinality) of the set of n-tuples and N the size of the subset with n-tuples having at least two identical entries.

Solution contd. As in the previous example, it is easier to count the number of n-tuples K with no duplicates, compute the probability $q = K/M$ of an outcome with no duplicates, then compute the desired probability from $p = 1 - q$.

What is the size M of the set Ω? Each slot in an n-tuple can be filled in 365 ways. Therefore, there are $M = 365^n$ n-tuples in Ω.

What is the size K of the set A of n-tuples with *no* duplicate birthdays? The 1st slot in these n-tuples can be filled in 365 ways. Since duplicates are not allowed, the 2nd slot can be filled in 364 ways, the 3rd in 363 ways, and so on until we reach the nth slot. Therefore, the size of set A is $K = \prod_{j=0}^{n-1}(365 - j)$.

Solution contd.

Consequently, the probability of *no* duplicate birthdays in a crowd of size n is

$$q = \left[\prod_{j=0}^{n-1}(365 - j)\right] / 365^n = \prod_{j=0}^{n-1}(1 - j/365),$$

while it is $p = 1 - q$ for a crowd in which at least two people have the same birthday.

Remarkably, in a crowd of 23 persons, there is about a 50:50 chance that at least two people share a birthday!

Example (1.3 An Act of Desperation)

On 19th October1900, Max Planck presented his correct guess for the formula for the black body radiation spectrum to the Berlin Academy. He then spent the next two months trying to derive the formula theoretically.

In *an act of desperation*, Planck resorted to Boltzmann's formula for the entropy $S = k \log \Omega$ and combined it with the thermodynamic relation $1/T = dS/dE$ in order to arrive at an expression for the radiation spectrum. But, in order to render the counting of microstates meaningful, the energies had to be discretized into small energy elements ϵ. In Boltzmann's work this was a purely mathematical device. But, Planck discovered that he could not take the limit $\epsilon \to 0$ if he wanted to recover his guessed formula from his derived one! Moreover, when he compared the two formulae, he found that he had to set $\epsilon = h\nu$.

Example (1.3 An Act of Desperation)

Planck computed the entropy for the following model: m indistinguishable quanta in thermal equilibrium with n indistinguishable oscillators in a system at temperature T. Planck could then compute the number of ways Ω that m indistinguishable quanta can be distributed among n indistinguishable oscillators.

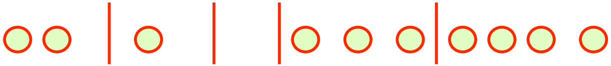
Given Ω, Planck could then use

$$\frac{1}{kT} = \frac{d \log \Omega}{dE},$$

to compute the average energy per oscillator E as a function of absolute temperature T.

Example (1.3 An Act of Desperation)

Planck's Model Imagine arranging the m quanta in a row with $n-1$ partitions between them representing the boundaries between the oscillators. The figure below depicts $m = 10$ quanta and $n - 1 = 4$ partitions.



There are $m + n - 1$ quanta plus partitions that can be arranged in $(m + n - 1)!$ ways. But, since the quanta and the partitions are *distinguishable*, we must divide by $m! \times (n-1)!$, which yields

$$\Omega = \frac{(m+n-1)!}{m!\,(n-1)!}$$ ... and the rest, as they say, is history!

Outline

In 1933, Andrey Kolmogorov published a highly influential book entitled *Foundations of the Theory of Probability* in which he developed the theory of probability starting from

1. the axioms of Boolean algebra
2. and axioms he introduced.

We first consider the axioms of Boolean algebra, then those of Kolmogorov.

A Boolean algebra is the 4-tuple $(\mathbb{B}, +, \bullet, \bar{\ })$ comprising a collection of sets $\mathbb{B}$, including the special sets 0 and 1, equipped with an equivalence relation $(=)$ and the operations OR $(+)$, AND $(\bullet)$, and NOT $(\bar{\ })$.

An equivalence relation $\sim$ obeys the rules:

1. $a \sim a$
2. $a \sim b$ and $b \sim a$
3. $a \sim b$, $b \sim c$, and $a \sim c$.

Example

"as crazy as" is an equivalence relation.

1. Boris $\sim$ Boris
2. Boris $\sim$ Aardvark and Aardvark $\sim$ Boris
3. Boris $\sim$ Aardvark, Aardvark $\sim$ Zorg, and Boris $\sim$ Zorg.

Axioms of Boolean Algebra (Huntington)

For all ($\forall$) $A,\ B,\ C \in \mathbb{B}$:

$$
\begin{aligned}
A + B &= B + A && (1) & AB &= BA && (5)\\
A + (BC) &= (A+B)(A+C) && (2) & A(B+C) &= AB + AC && (6)\\
A + 0 &= A && (3) & A1 &= A && (7)\\
A + \bar{A} &= 1 && (4) & A\bar{A} &= 0 && (8)
\end{aligned}
$$

We also assume we can replace (A) with A and vice versa.

Here are some useful lemmas and theorems:

$$A + A = A$$
$$A + 1 = 1$$
$$A0 = 0$$
$$AA = A$$
$$(A+B)+C = A+(B+C)$$

$$\bar{0} = 1$$
$$\bar{1} = 0$$
$$A + AB = A$$
$$A(A+B) = A$$
$$(AB)C = A(BC)$$

De Morgan's Laws

$$\overline{A+B} = \bar{A}\,\bar{B}$$
$$\overline{AB} = \bar{A} + \bar{B}$$

Lemma (1)

$A + A = A$

Proof.

$$
\begin{aligned}
A + (BC) &= (A + B)(A + C) && \text{(axiom 2)} \\
A + (CB) &= (A + B)(A + C) && \text{(axiom 5)} \\
A + (A\bar{A}) &= (A + \bar{A})(A + A) && C = A, B = \bar{A} \\
A + 0 &= 1(A + A) && \text{(axioms 8, 4)}, (0) \to 0, (1) \to 1 \\
A &= 1(A + A) && \text{(axiom 3)} \\
A &= A + A && \text{(axioms 6, 5, 7)}
\end{aligned}
$$

□

Lemma (2)

$A + 1 = 1$

Proof.

$$\begin{aligned}
A + (BC) &= (A + B)(A + C) && \text{(axiom 2)} \\
A + (B1) &= (A + B)(A + 1) && \text{let } C = 1 \\
A + B &= (A + B)(A + 1) && \text{(axiom 7)}, (B) \to B \\
A + \bar{A} &= (A + \bar{A})(A + 1) && B = \bar{A} \\
1 &= 1(A + 1) && \text{(axiom 4)}, (1) \to 1 \\
1 &= A + 1 && \text{(axioms 6, 5, 7)}
\end{aligned}$$

□

Exercise 1

Prove:

1. $A0 = 0$
2. $A + AB = A$
3. $\mathrm{A}(\mathrm{A} + \mathrm{B}) = \mathrm{A}$

For every step in your proofs specify which axioms you are using.

Kolmogorov Axioms

Let Ω be a set of elementary events E, S a collection of subsets of Ω called events including the empty event $\emptyset$ and the event Ω. Probability P is a real number assigned to all events $A, B \in S$ such that

$$P(A) \geq 0 \tag{9}$$
$$P(\Omega) = 1 \tag{10}$$
$$P(A+B) = P(A) + P(B) \quad \forall AB = \emptyset. \tag{11}$$

If $AB = \emptyset$, A and B are said to be mutually exclusive.

Here are a few theorems that follow from the two sets of axioms:

$$P(\emptyset) = 0$$
$$P(A+B) = P(A) + P(B) - P(AB)$$
$$P(A_1 + \cdots + A_n) = P(A_1) + \cdots P(A_n) \quad \forall A_i B_j = \emptyset, i \neq j.$$

Lemma (3)

$P(\emptyset) = 0$

Proof.

The lemma $A0 = 0$ implies $\Omega\emptyset = \emptyset$, the blindingly obvious conclusion that events Ω and $\emptyset$ are mutually exclusive! Therefore,

$$
\begin{aligned}
P(\Omega + \emptyset) &= P(\Omega) + P(\emptyset) && \text{(axiom 11)} \\
P(\Omega) &= P(\Omega) + P(\emptyset) && \text{(axiom 3)} \\
\therefore P(\emptyset) &= 0 && \text{(since } P \text{ is a finite real number)}
\end{aligned}
$$

□

Outline

Conditional Probability

Consider events A and B. The conditional probability of A given B, written as $P(A|B)$ and assuming $P(B) > 0$, is defined by

$$P(A|B) = \frac{P(AB)}{P(B)}. \tag{12}$$

This definition implies

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

provided that $P(A) > 0$. Since $BA = AB$, $P(BA) = P(AB)$. Therefore, we arrive at

Bayes' Theorem

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

Example (1.4 Two Dice)

The outcomes of an experiment, which consists of rolling two 6-sided dice, can be modeled as 2-tuples (z_1, z_2), $z_i \in \{1, 2, 3, 4, 5, 6\}$ that form the set Ω of size $n = 36$. The only probability associated with Ω is $P(\Omega) = 1$. However, if we assume each outcome of the two dice to be equally likely, then to each event E in the power set S[a] of Ω we can assign the probability $P(E) = |E|/36$, where $|E|$ denotes the cardinality of a set within the power set.

[a] The set of all subsets including Ø and Ω with cardinality (size) 2^n.

Example (The power set of $\{A, B, C\}$ contains)

$$\{\} \quad \{A, B, C\}$$
$$\{A\} \quad \{B\} \quad \{C\}$$
$$\{A, B\} \quad \{A, C\} \quad \{B, C\}$$

Example (1.4 Two Dice contd.)

Consider the sets

$$A = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\} \text{ and}$$
$$B = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$$

$\in S$, what is the conditional probability, $P(A|B)$, of A given B?

- A is the event in which each die yields an even number.
- B is the event in which the two numbers sum to 8.

What is the operational meaning of $P(A|B) = P(AB)/P(B)$?

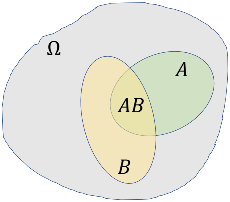
It states: restrict the set of elementary events to those in event B. Then, determine what fraction of the events in B also happen to be in A, that is, in the set AB.

The probabilities of the events A, B, and AB, given our probability model, are:

$$P(A) = 9/36$$
$$P(B) = 5/36$$
$$P(AB) = 3/36$$

Therefore,

$$\begin{aligned} P(A|B) &= P(AB)/P(B), \\ &= (3/36)/(5/36), \\ &= 3/5. \end{aligned}$$



$$\begin{aligned} A &= \{(2,2),(2,4),\mathbf{(2,6)}, \\ &\quad (4,2),\mathbf{(4,4)},(4,6), \\ &\quad \mathbf{(6,2)},(6,4),(6,6)\} \\ B &= \{\mathbf{(2,6)},(3,5),\mathbf{(4,4)},(5,3),\mathbf{(6,2)}\} \\ AB &= \{(2,6),(4,4),(6,2)\} \end{aligned}$$

Example (1.5 Are You Doomed?)

According to the WHO[a] a test is available that identifies 92% of people with Ebola and identifies 85% who are Ebola free. During an Ebola outbreak, you visit your doctor and nsist that you be tested for Ebola! The test result is positive (+). Are you doomed? During the 2014 - 2016 Ebola outbreak, 4 cases of infection were reported in the U.S. Here is a summary of what we knew then:

$$\text{event } D = \text{You are Diseased} ☹$$

$$\text{event } H = \text{You are Healthy} ☺$$

$$P(+|D) = 0.92 \qquad P(D) = 4/320,000,000$$

$$P(+|H) = 0.15 \qquad P(H) = 1 - P(D)$$

[a] http://www.who.int/medicines/ebola-treatment/1st_antigen_RT_Ebola/en/

Example (1.5 Are You Doomed?)

Bayes theorem can be generalized to

$$P(B_i|A) = \frac{P(A|B_i)\,P(B_i)}{\sum_i P(A|B_i)\,P(B_i)},$$

for mutually exclusive and exhaustive events B_i[a]. For our problem, we can write

$$P(D|+) = \frac{P(+|D)\,P(D)}{P(+|D)\,P(D) + P(+|H)\,P(H)} \approx \frac{P(+|D)}{P(+|H)}\,P(D)$$
$$\approx 1/13,000,000.$$

So NO, you are not doomed! You are merely paranoid!

[a] $\sum_i B_i = 1$

Exercise 2 – The Monte Hall Problem

A contestant on a TV show has a chance to win a car that is hidden behind one of three doors. The contestant is asked to pick a door and she picks door 1. The host, who knows which door hides the car, chooses (at random if the car is behind door 1) and opens one of the two doors without the car behind it. The host opens door 2, of course revealing no car, and offers the contestant the chance to reconsider her choice. Should she choose door 3 or stay with her first choice, door 1? The answer is she should choose door 3! Use Bayes' theorem to explain, as carefully as you can, the reasoning that leads her to this conclusion.

Hint: Denote by $H = 1, 2, \text{ or } 3$ the hypotheses that the car is behind either door 1, 2, or 3 and let $D2$ be the datum that door 2 has been opened revealing no car. The probabilities you are after are $P(H|D2)$.

Outline

- So far, we have used probability without saying how it is connected to the real world. We have sketched its abstract definition as a measure with certain properties defined on suitable collections of sets. But this is not helpful without an *interpretation*; the Stanford Encyclopedia of Philosophy[2] lists six!
- We briefly consider three of them:
  1. classical If n things can happen out of m possible things, assumed equally likely, then $P(A) = n/m$.
  2. relative frequency If n things can happen out of m possible things, then $z = n/m$ is the relative frequency with which the n things have occurred. The probability $P(A)$ is the limit as $n, m \to \infty$ (suitably defined) of the relative frequency.
  3. degree of belief If a rational agent stands to gain an amount U should she bet that A will happen or turn out to be true, then $P(A)U$ is the price she is willing to pay to place the bet.

[2] https://plato.stanford.edu/entries/probability-interpret/

In Part 2, we'll discuss a few important probability distributions.