Probability - Part 1

Harrison B. Prosper

Department of Physics, Florida State University

September 18, 2017

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 in a row from a standard deck of cards is 1 in 270,725.
- In 2016, the a priori chance of dying in a car crash in the United States was 1 in 8000.
- 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 probability is the foundational concept in statistics, we should spend a bit of time learning about it.

So let's get started!

On specialists:

"Most of them do not instruct at all since they rely upon obscure, sometimes even false, principles, and instead of seeking the truth to clarify it they aim to embarrass each other, also by terms which they do not even themselves understand, and by chimerical distinctions."

Antoine Gombaud (le chevalier de Méré, 1607-1684)[1]

In this course, we hope to avoid being that kind of specialist!

[1] *Pascal and the Invention of Probability Theory*, Oystein Ore, The American Mathematical Monthly, Vol. 67, No. 5 (May, 1960), pp. 409-419

Example (1.1 de Méré's Problem)

Which of the following 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?

Antoine Gombaud brought this problem to the attention of Blaise Pascal who, in 1654, wrote to Pierre de Fermat (1601-1665). The Pascal-Fermat correspondence along with previous work by Gerolamo Cardano (1501 - 1576) constitute the foundation of the mathematical theory of probability.

Blaise Pascal (1623 - 1662)

Let's review the solution to this problem.

Example (1.1 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., the only possible outcomes.
2. There are 6 possible elementary outcomes[2] of which only 1 is a six.
3. The 6 elementary outcomes are equally probable.
4. The elementary outcomes remain equally probable for every throw and every throw is independent of the others.

Assumption 1 $\implies$ $p + q = 1$, or $p = 1 - q$.
Assumption 2 $\implies$ there are 5 ways not to get a six.
Assumption 3 $\implies$ the probability of each of the elementary outcomes is $1/6$, therefore, given assumption 2 the probability not to get a six is $5/6$.
Assumption 4 $\implies$ the probability not to get a six in 4 throws of the die is $q = (5/6) \times (5/6) \times (5/6) \times (5/6)$. Therefore, $p = 1 - (5/6)^4$.

[2] 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;
- count the number of favorable outcomes and
- take the ratio of the two counts.

The outcome of the experiment can be represented by the 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 $6 \times 6 \times 6 \times 6$. The total number of outcomes without a six is $5 \times 5 \times 5 \times 5$, 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 the 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 will be $k!$ permutations that will consist of rearrangements of the items in the k slots. If the order of the items is irrelevant (perhaps because the items are indistinguishable) then the number of *distinct* arrangements is smaller by $k!$, that is, the number of arrangements is now

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

This is called the number of combinations.

Example (1.2 The Birthday Problem)

A crowd of people is randomly assembled. How large must the crowd be so that the chance of finding at least two people with the same birthday is $\geq 50\%$?

Assumptions

1. There are 365 possible birthdays (ignoring leap years).
2. Every birthday is equally probable.

Solution

Consider a crowd of size n. The possible outcomes of the experiment is an n-tuple each element of which is one of 365 possible birthdays.

Let M be the number of possible n-tuples and N the number of n-tuples with at least two identical entries. In such problems, as in the previous example, it is typically easier to count the number of n-tuples K with no duplicates, and then compute the desired probability using $p = N/M = (M - K)/M$ assuming each n-tuple to be equally likely.

Solution contd.

M What is the cardinality (size) of the set Ω of n-tuples?
Each slot in an n-tuple can be filled in 365 ways. Therefore, there are $M = 365^n$ n-tuples in Ω.[3]

K What is the size 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, which can be filled in $365 - (n - 1)$ ways. The size of set A is therefore $K = 365 \times 364 \times \cdots \times (365 - n + 1) = 365!/(365 - n)!$. Consequently, since the experimental outcomes are considered equally likely, the probability of at least one duplicate birthday in a crowd of size n is

$$\begin{aligned} p &= [365^n - 365!/(365 - n)!]/365^n \\ &= 1 - \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{(365 - n + 1)}{365} \geq 0.50. \end{aligned}$$

[3] This is akin to sampling with replacement from a box with 365 distinguishable items.

Example (1.3 Of Photons and Boxes)

In 1900, in an inspired guess, Max Planck arrived at the correct formula for the black body radiation spectrum. But, since it would have been a tad embarrassing for Herr Professor Dr. Planck to admit to the German Physical Society, "I guessed it", he desperately wanted to derive the formula theoretically.

Later that year, in "an act of desperation", he invented the quantum and calculated the average energy per oscillator for a model he also invented: m indistinguishable quanta with energy $\epsilon = h\nu$ are in equilibrium with n indistinguishable oscillators in a cavity at temperature T. The number of quanta per oscillator can vary from zero up to m.

The problem reduced to counting: how many ways Ω can m indistinguishable items be distributed among the n indistinguishable boxes? Given this number, the entropy S of the system can be calculated using Ludwig Boltzmann's formula $S = k \log \Omega$, which provides a fundamental link between statistical mechanics and thermodynamics.

Example (1.3 Of Photons and Boxes)

Solution

Imagine arranging the m items in a row with $n-1$ partitions between them. In the figure below, we have $m=10$ and $n-1=4$.

The partitions divide the items into n boxes and yield a total of $m+n-1$ items and partitions that can be arranged in $(m+n-1)!$ ways. But, we need to count only the *distinguishable* permutations. Since the items and partitions are indistinguishable, we must divide by $m! \times (n-1)!$. Therefore, we conclude that

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

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 (Crazy Boris)

"as crazy as" is an equivalence relation $\sim$[a]

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

[a] With apologies to all persons named Boris, Aardvark, or Zorg!

Axioms of Boolean Algebra (Huntington)

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

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

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

Here are some useful lemmas and theorems:

$$
\begin{aligned}
A + A &= A & \bar{0} &= 1\\
A + 1 &= 1 & \bar{1} &= 0\\
A0 &= 0 & A + AB &= A\\
AA &= A & A(A+B) &= A\\
(A+B)+C &= A+(B+C) & (AB)C &= A(BC)
\end{aligned}
$$

De Morgan's Laws

$$
\begin{aligned}
\overline{A+B} &= \bar{A}\,\bar{B}\\
\overline{AB} &= \bar{A} + \bar{B}
\end{aligned}
$$

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.

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

□

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 basic theorems that can be derived 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,

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

$\square$

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

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}. \tag{13}$$

Example (1.3 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 its power set S[a] 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 $\emptyset$ and Ω with cardinality (size) 2^n.

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

$$
\begin{array}{ccc}
\{\} & \{A, B, C\} & \\
\{A\} & \{B\} & \{C\} \\
\{A, B\} & \{A, C\} & \{B, C\}
\end{array}
$$

Given events

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

$\in S$, 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 $P(A|B) = P(AB)/P(B)$?

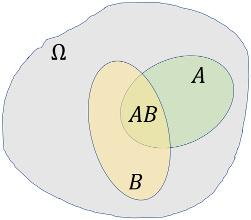
It tells us to restrict the set of elementary events to those in event B. In effect, B plays the rôle of Ω, but with fewer possibilities. Then, determine what fraction of the events in B are also in A, that is, in the event AB.

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 (1.4 Are You Doomed?)

According to WHO[a] there is a test that correctly identifies 92% of people with Ebola and correctly identifies 85% who are Ebola free. During a visit to your doctor in the US you suddenly become paranoid and insist that you be tested for Ebola. The test result is positive $(+)$. Are you doomed?

Well, it depends...

During the 2014 - 2016 Ebola outbreak in Africa, 4 cases of Ebola infection were reported in the United States. Consider the following mutually exclusive events and associated probabilities:

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

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

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

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

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

Example (1.4 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

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

No, you are not doomed! You are merely paranoid!

If you were in the US in 2016, you had more than a 1000 times higher *a priori* chance of meeting your maker via a car crash!

[a] $\sum_i B_i = 1$

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[4] 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 (in some suitably defined units) 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.

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

On Wednesday, we'll discuss probability distributions.

Have fun with the Assignment 1!

The due date and time is next Monday at 1pm.

Solutions will be posted about an hour after the deadline.