Lecture 8: Probability

The idea of probability is well-known. The flipping of a balanced coin can produce one of two *outcomes*: T (tail) and H (head) and the symmetry between the two outcomes means, of course, that the probability of each is $\frac{1}{2}$. Briefly,

$$P(T) = P(H) = \frac{1}{2}.$$

We cannot predict apriory (i.e., beforehand) the result of a specific experiment, but for a long series of coin tossing we can formulate a non-trivial "deterministic" law. Let $t(n)$ denote the number of times T results in n flips of a coin, and $v_T(n) = \frac{t(n)}{n}$ be the corresponding frequency (which is random and different for different sequences of flips!), then we can expect, that

$$\lim_{n\to\infty} v_T(n) = P(T) = \frac{1}{2}.$$

This is an "experimental fact", rather than a theoretical statement. Similar experiments in biology (the sex of a human child) were repeated throughout history many billions of times and gave the same result:

$$P(boy) = P(girl) = \frac{1}{2}.$$

The main object of study in probability theory is the "random experiment", which preserves all main features of coin tossing.

Random experiments.

We will discuss three important ideas associated with probability.

1. *Randomness.* The result (for given and fixed external conditions) cannot be predicted in advance. The set of all possible outcomes must be known. The "*sample space*" S of an experiment is the set of all possible "elementary" (irreducible) results of the experiment. Suppose for simplicity that S contains a finite number N of outcomes, which can be enumerated as $\{s_1, s_2, ..., s_N\} = S$. Every experiment produces one (and only one) outcome $s_i, i = 1, 2, ...N$.

   Example 1. A classic experiment is dice tossing (6 elementary outcomes, the possible numbers on the upper face). The event that the number on the upper face is even is not elementary because it contains 3 elementary outcomes: 2,4, and 6.

2. *Reproducibility.* The experiment can be repeated under fixed conditions an infinite number of times, at least in principle. Games (coins, dice, cards, roulette) are classic examples. The question about probability of the existence of life on Mars or somewhere else in the Universe cannot be the subject of probability theory or statistics, because such an "experiment" is not reproducible.

3. *Symmetry.* All outcomes $S_i, i = 1, 2, ..., N$ must be "equally likely", meaning that in a very long series of experiments

   $$P = \lim_{n\to\infty} v_i(n) = \lim \frac{t_i(n)}{n} = \frac{1}{N}.$$

   In other words, each outcome occurs just as often as the other outcomes.

   Here, just as in the earlier case of coin flipping, t_i represents the number of times outcome S_i occurs during the first n trials of the experiment and $v_i(n) = \frac{t_i(n)}{n}$ is the corresponding frequency. Usually, the equal likelihood of the outcomes is a consequence of some sort of symmetry of the random experiment.

   The set $S = \{S_1, S_2, ...S_N\}$ (the sample space with corresponding equal probabilities $P_i = \frac{1}{N},\ i = 1, 2, ...N$) is known as the classical or combinatorial probabilistic space.

   The table

   $$S = \left\{ \begin{array}{cccc} S_1 & S_2 & \dots & S_N \\ \frac{1}{N} & \frac{1}{N} & \dots & \frac{1}{N} \end{array} \right\}$$

   gives a complete description of the statistical experiment. *Example 2.* Dice tossing:

   $$S_{P_i} \quad = \left\{ \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array} \right\}$$

   (because of symmetry!)

An *event* is a collection of outcomes; that is, a subset of the sample space. Every event A can be described by a *finite subset* of S (the same notation A).

Logical language: statements $A, B, \ldots$. Geometrical language: subsets $A, B, \ldots$ By the definition

$$P(A) = \frac{|A|}{|S|} = \frac{|A|}{N} = \sum_{S_i \in A} \frac{1}{N}.$$

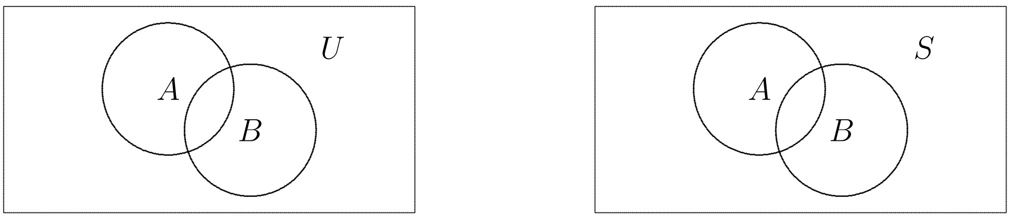
In other words, to find the probability of an event, count the number of outcomes belonging to the event and add up their individual probabilities. In the equally likely sample spaces we are discussing here, the probability is just 1 divided by the number N of elements of the sample space. The notation $S_i \in A$ means that S_i belongs to the subset A or, in logical terms, it means that the outcome S_i implies statement (event) A. This is the classical definition of probability.

Example 3. Experiment: selection of one card from the full deck of 54 playing cards. Such a deck consists of two jokers together with 13 denominations $\{A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\}$ in each of four suits $\clubsuit, \diamondsuit, \heartsuit, \spadesuit$. Again we assume that each outcome is equally likely:

$$P_i = \frac{1}{54},\ i = 1, 2, ..., 54.$$

To find the probability of the event that the card selected is a heart, write

$$P(\heartsuit) = \sum_{S_i \in \heartsuit} \frac{1}{54} = \frac{13}{54}.$$



Algebra of events.

Using the given system of events one can construct new events using logical (set theory) operations. Two equivalent languages are admissible, the second one is already known.

Logic	**Set theory**
A the statement	A is a subset of S.
Certain event, T.	Full sample space, S.
Impossible event, F.	empty set, $\emptyset$.
Not A, $\overline{A}$.	complementary set $\overline{A}$, containing all points of S outside A.
A but not B, $A \setminus B$.	$A \setminus B$, all points of A outside of B.
Or A or B, $A \cup B$.	Union of A and B, notation $A \cup B$.
Both A and B, $A \cap B$.	Intersection $A \cap B$, or A$\cdot B$, points common to both A and B
B implies A, $B \to A$.	$A \supseteq B$, inclusion.
A and B are not both true	$A \cdot B = \emptyset$.

The notion of probability is very similar to the notions of mass, weight, area, etc. Sometimes in applications, instead of using the term “probability of event A” you will see the phrase “statistical weight of A”. The following properties of probabilities are provable, and sometimes the first three are taken as axioms.

1. $P(S) = 1,\ P(\emptyset) = 0$

2. $P(\overline{A}) = 1 - P(A)$, or $P(A) + P(\overline{A}) = 1$ (every elementary event is either in $\overline{A}$, or in A, but not in both).

3. If $AB = \emptyset$ (non intersecting, mutually exclusive events), then $P(A \cup B) = P(A) + P(B)$, additive property similar to additivity of area, volume, etc.

3a. Moreover, if $A_1, A_2, ..., A_n$ satisfy

$$\begin{aligned} A_i A_j &= \emptyset \quad \text{for all } i, j = 1, 2, ..., n,\ i \neq j \text{ then} \\ P(\cup_{i=1}^n A_i) &= \sum_{i=1}^n P(a_i). \end{aligned}$$

The hypothesis $A_i A_j = \emptyset \quad i, j = 1, 2, ..., n$ is called *pairwise disjointness.*

4. If $A \subseteq B$ then $P(A) \leq P(B)$.

5. $P(A \cup B) = P(A) + P(B) - P(AB)$ (for intersecting A,B). The formula is simple: if we take the sum of probabilities of all the elementary events from A and from B, then we'll get all the S_i from $A \cup B$, but the S_i belonging to the intersection get double counted.

6.

$$P(A \setminus B) = P(A) - P(AB).$$

(See the Venn diagram for $A \setminus B$).

The proofs of all these formulas are straightforward:

$$P(\overline{A}) = \frac{|\overline{A}|}{|S|}\ ,\ \text{but } |\overline{A}| = |S| - |A|$$

$$P\left(\overline{A}\right) = \frac{|S| - |A|}{|S|} = 1 - \frac{|A|}{|S|} = 1 - P\left(A\right)$$

or

$$P\left(A \cup B\right) = \frac{|A \cup B|}{|S|} = \frac{|A| + |B| - |AB|}{|S|} \text{ (we used the addition principle)}$$

$$= \frac{|A|}{|S|} + \frac{|B|}{|S|} - \frac{|AB|}{|S|} = P(A) + P(B) - P(AB).$$

Using formula 5, i.e. the addition principle for probabilities, and standard facts of set theory, one can prove more general formulas. For instance,

$$\begin{aligned} P(A \cup B \cup C) &= P((A \cup B) \cup C) = P(A \cup B) + P(C) - P((A \cup B)C) \\ &= P(A) + P(B) - P(AB) + P(C) - P(AC \cup BC) \\ &= P(A) + P(B) + P(C) - P(AB) - [P(AC) + P(BC) - P(ABC)] \\ &= P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \end{aligned}$$

using the addition principle, distributivity and associativity of the set operations, and the formula $AC \cdot BC = ABC$.

Example 4. So-called DeMere paradox. The experiment consists of tossing two identical dice. There are two possibilities for the sample space S, one consisting of 36 ordered pairs and the other with 21 unordered pairs.

$$\begin{array}{cc} \text{I (ordered pairs)} & \text{II (unordered pairs)} \\ (1,1)(1,2)(1,3)(1,4)(1,5)(1,6) & (1,1) \\ (2,1)(2,2)(2,3)(2,4)(2,5)(2,6) & (2,1)(2,2) \\ (3,1)(3,2)(3,3)(3,4)(3,5)(3,6) & (3,1)(3,2)(3,3) \\ \ldots\ldots & \ldots\ldots \\ (6,1)(6,2)(6,3)(6,4)(6,5)(6,6) & (6,1)\ (6,2)\ldots(6,6) \\ N = 36 & N = 1 + 2 + \cdots + 6 = 21 \end{array}.$$

Now suppose A is the set of doublets, $A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$. Clearly $N(A) = 6$. In the first case $P_I(A) = \frac{6}{36} = \frac{1}{6}$. In the second case $P_{II}(A) = \frac{6}{21} = \frac{2}{7} > \frac{1}{6}$.

There are no theoretical reasons to prefer one of these possibilities over the other. But experimental data shows that the *ordered pairs* model is correct!

This general idea works in many cases: even for coins, dice and some other objects that look identical. We can always consider them as distinguishable, and usually that is the best perspective.

Example 5. (Birthdays, due to Euler.) What is the probability, that in a class with 25 students, two of them have a common birthday? The answer is surprising: 0.56 (i.e. 56%).

Solution. Let's consider the following experiment. Randomly choose a class among all classes of such size. We can interview students (say, in alphabetical order) to get information about their birthdays (*bd*). The corresponding sequence:

$(bd_1,\ bd_2, \ldots.., \ bd_{25})$ is an elementary outcome of the experiment. According to the product rule the total number of outcomes is very large:

$$|S| = N = \underbrace{365 \cdot 365 \cdots 365}_{25 \text{ times}} = 365^{25}.$$

For simplicity, we suppose that every year has 365 days. Let A_{25} denote the event that two students in the class have a common birthday. Then

$$P(A_{25}) = 1 - P(\overline{A}_{25})$$

and $\overline{A}_{25}$ is the event that all 25 students have different birthdays. It is easy to see that

$$|\overline{A}_{25}| = N(\overline{A}) = \underbrace{365 \cdot 364 ... 341}_{25 \text{ factors}} = P_{25}^{365} \approx 4.92 \times 10^{63}.$$

We count this set by noting that the first student has 365 possible birthdays, the second has 364 opportunities, because his birthday is distinct from the first one, etc.

As a result

$$P(\overline{A_{25}}) = \frac{365 \cdot 364 \cdots 341}{365 \cdot 365 \cdots 365} = 1 \cdot \left(1 - \frac{1}{365}\right) \cdots \left(1 - \frac{24}{365}\right)$$

Direct calculation gives

$$\begin{aligned} P(\overline{A}_{25}) &\approx .44, \text{ i.e.} \\ P(A_{25}) = 1 - P(\overline{A}_{25}) &\approx 1 - .44 = .56 \quad (56\%)\,. \end{aligned}$$

Now suppose the class has r students.

$$\begin{aligned} P\left(\overline{A}_r\right) &= 1\left(1 - \frac{1}{365}\right) \cdots \left(1 - \frac{r-1}{365}\right) \\ P\left(A_r\right) &= 1 - P\left(\overline{A}_r\right). \end{aligned}$$

One can use instead of direct calculations the following approximation. Recall that $e^{\ln x} = x$ and $\ln ab = \ln a + \ln b$. Since the derivative of $\ln x$ at $x = 1$ is 1, it follows that

$$\ln(1 - \xi) = -\xi, \text{ for small } \xi.$$

As a result,

$$P\left(\overline{A}_r\right) = 1\left(1 - \frac{1}{365}\right) \cdots \left(1 - \frac{r-1}{365}\right)$$

$$
\begin{aligned}
&= e^{\ln(1)+\ln\left(1-\frac{1}{365}\right)+\cdots+\ln\left(1-\frac{r-1}{365}\right)} \\
&\approx e^{0-\frac{1}{365}-\frac{2}{365}\cdots\frac{(r-1)}{365}} \\
&= e^{-\frac{1+2+\cdots+(r-1)}{365}} \\
&= e^{-\frac{r(r-1)}{2\cdot 365}} \text{ (using the formula } 1+2+\cdots+n=\frac{n(n+1)}{2}\text{).}
\end{aligned}
$$

If $r = 23$ then $P(A_{23}) = P\left(\overline{A}_{23}\right) \approx 0.50$. If $r = 30$ then $P(\overline{A}_{30}) \approx 0.30$, $P\left(A_{30}\right) \approx 0.70$. If $r = 40$ then $P(\overline{A}_{40}) \approx 0.10$, $P\left(A_{40}\right) \approx 0.90$.