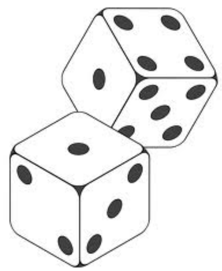
DIC15 / Review of basic probability theory
SXD 217, 219

Throwing a coin, or a die, or a pair of dice, or spinning a roulette, or betting about the value of the € against the $ tomorrow, and so on, are *random experiments.*


In each case we may obtain any of a number of outcomes, but we cannot predict with certainty which of them will occur. Probability is the science that allows us to quantify and reason about these uncertainties.

Problem. Which of the two results would you prefer to bet on?

- Getting at least one 6 in four rolls of a die?
- Getting at least a double 6 in 24 rolls of two dice?

Problem. We have three coins. Two are fair and one is biased so that heads are twice as likely as tails. We pick a coin at random and flip it three times. Assuming that we get three heads, what is the likelihood that we have chosen the biased coin?

Notation

$U = \{a_1, a_2, \dots, a_n\}$, the set of possible outcomes of a random (or stochastic) *experiment* A.

Frequencies and probabilities

If we repeat A a very great number of times N and let f_j denote the number of times that we get the outcome a_j (we say f_j is the *frequency* of a_j), then $f_j/N \to p_j$, where $p_1, \dots, p_n$ are *positive real numbers* such that

$$p_1 + \cdots + p_n = 1 \text{ (note that } f_1 + \cdots + f_n = N\text{).}$$

We say that p_j is the *probability* of the outcome a_j and write $P(a_j) = p_j$.

If $X \subseteq U$ (these subsets are called *events*), we set

$$p = P(X) = \sum_{a_j \in X} p_j$$

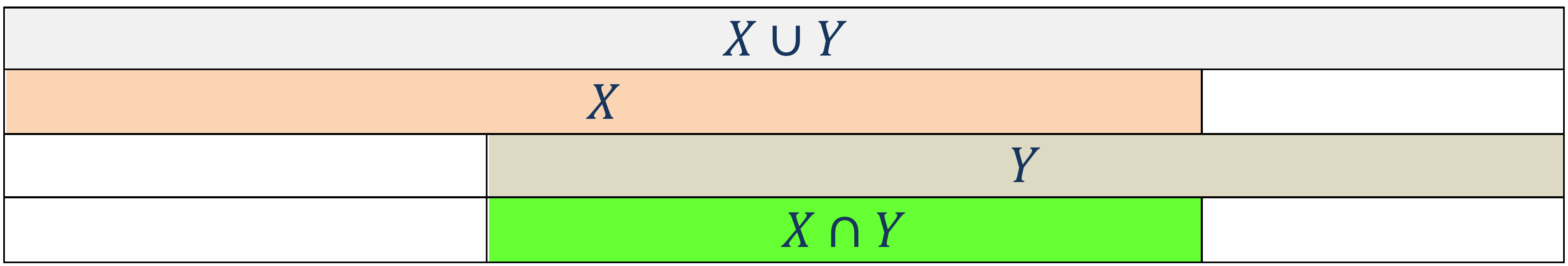
and say that p is the *probability* of X.

Main properties

- $p > 0$ unless $X = \emptyset$ (the *impossible event*).
- As $P(U) = p_1 + p_2 + \cdots + p_n = 1$, U is the *sure event*.
- If $X, Y \subseteq U$ are events, then

  $$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$

Indeed, in the sum $P(X) + P(Y)$ the outcomes in $X \cap Y$ are summed twice, and the formula follows from the definitions.



In particular we have $P(X \cup Y) = P(X) + P(Y)$ if and only if $X \cap Y = \emptyset$ (in this case we say that X and Y are *incompatible*).

- The *complement* of an event X is the event $\bar{X} = U - X$ (its outcomes are precisely the outcomes not in X). Thus $X \cap \bar{X} = \emptyset$ and $X \cup \bar{X} = U$. It follows that

$$P(\bar{X}) = 1 - P(X).$$

Important special case

If $p_1 = p_2 = \cdots = p_n$ then

$$p_k = \frac{1}{n} \text{ for } k = 1, \ldots, n$$

and

$$p(X) = \frac{m}{n}$$

where $m = \#X$ (the cardinal of X, or the *number of favorable cases*) and n is the *number of possible cases*).

Conditional probability

Let $X, Y \subseteq U$ be events and run the experiment a very large number of times N. Let $f_{xy} = f_{X\cap Y}$ be the frequency of the event $X \cap Y$. Let f_x and f_y be the frequencies of X and Y, respectively.

The quotient f_{xy}/f_x denotes the proportion of times that Y is observed when X is also observed. So $f_{xy}/f_x \to P(Y|X)$, where $P(Y|X)$ denotes the *probability of observing Y under the assumption that X has been observed* (*conditional probability*). Since

$$f_x/N \cdot f_{xy}/f_x = f_{xy}/N,$$

we finally arrive at the fundamental formula of conditional probabilities:

$$P(X \cap Y) = P(X) \cdot P(Y|X) \quad \text{or} \quad P(Y|X) = P(X \cap Y)/P(X).$$

We also can write

$$P(X \cap Y) = P(Y) \cdot P(X|Y) \quad \text{or} \quad P(X|Y) = P(X \cap Y)/P(Y).$$

Independent events

When $f_{xy}/f_x = f_y/N$, we say that X and Y are *independent*, and so this condition is equivalent to say that

$$P(X \cap Y) = P(X) \cdot P(Y), \text{ or } P(Y) = P(X \cap Y)/P(X).$$

Notation. Often we write $P(X, Y)$ instead of $P(X \cap Y)$.

Example. Suppose we have a bag with 12 marbles, 7 white and 5 red. We extract one marble and then another. Let X be the event that the first marble is red and Y the event that the second marble is white. If the first marble is returned to the bag after the observation, then X and Y are independent and $P(X \cap Y) = P(X) \cdot P(Y) = \frac{5}{12} \cdot \frac{7}{12} = \frac{35}{144}$. But if the first marble is not returned to the bag, then X and Y are not independent:

$$P(X \cap Y) = P(X)P(Y|X) = \frac{5}{12} \cdot \frac{7}{11} = \frac{35}{132}.$$

Interaction factor

Let $X, Y \subseteq U$ be two events. Then

$$P(X \cap Y) = \begin{cases} P(X) \cdot P(Y|X) \\ P(Y) \cdot P(X|Y) \end{cases}$$

This is equivalent to $\frac{P(X|Y)}{P(X)} = \frac{P(Y|X)}{P(Y)} = \frac{P(X \cap Y)}{P(X) \cdot P(Y)}$ (symmetric in X and Y).

We denote this value by $I(X, Y)$ and say that it expresses the *interaction factor* (or *reciprocal influence factor*) of X and Y.

Remark. $0 \leq P(Y|X) = \frac{P(X \cap Y)}{P(X)} \leq 1$. Its value is 0 iff $X \cap Y = \emptyset$ (when this occurs, we say that X and Y are *incompatible*). Similarly, its value is 1 iff $X = X \cap Y$, which is equivalent to $X \subseteq Y$. As a result we have

$$0 \leq I(X, Y) \leq \min\left(\frac{1}{P(Y)}, \frac{1}{P(X)}\right),$$

with equality on the right iff $X \subseteq Y$ or $Y \subseteq X$.

- $I(X, Y) = 1$ iff X and Y are independent (*neutral degree of influence*).

Bayes' rule

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} = \frac{P(Y) \cdot P(X|Y)}{P(X)} = P(Y) \cdot I(X, Y)$$

Interpretation. The probability $P(Y)$ represents the likelihood of the event Y before observing X (*prior probability*). After observing X, the updated likelihood (*posterior probability*) is measured by $P(Y|X)$. By Bayes' rule, this is the product of the *prior* $P(Y)$ times $I(X, Y)$, which therefore acts as the degree of *support* that X provides for (believing) Y.

Of course, we also have

$$P(X|Y) = P(X) \cdot I(X, Y).$$

Bayes' rule in practice. If Y_j are mutually exclusive events such that $\sum_j P(Y_j) = 1$, then

$$\sum_j P(X \cap Y_j) = P\left(\cup_j (X \cap Y_j)\right) = P\left(X \cap (\cup_j Y_j)\right) = P(X)$$

and from Bayes' rule

$$P(Y_j|X) = \frac{P(X \cap Y_j)}{P(X)} = \frac{P(Y_j) \cdot P(X|Y_j)}{P(X)}.$$

we see that

$$P(Y_j|X) \sim P(Y_j) \cdot P(X|Y_j).$$

So if we set $p_j = P(Y_j)$, $L_j = P(X|Y_j)$, $q_j = p_j L_j$ and $q = \sum_j q_j = P(X)$, then

$$P(Y_j|X) = q_j / q.$$

Example. We have three coins. Two are fair and one is biased so that heads are twice as likely than tails. We pick a coin at random and flip it three times. Assuming that we get three heads, what is the likelihood that we have chosen the biased coin?

Solution. We want to calculate $P(B|HHH)$. By Bayes' formula,

$$P(B|HHH) = \frac{P(B)\cdot P(HHH|B)}{P(HHH)}.$$

The value of the numerator is $\frac{1}{3}\left(\frac{2}{3}\right)^3 = \frac{8}{81}$. The probability in the denominator is equal to

$$\begin{aligned} P(HHH,B) + P(HHH,\bar{B}) &= P(B)\cdot P(HHH|B) + P(\bar{B})\cdot P(HHH|\bar{B}) \\ &= 8/81 + (2/3)(1/8) = 59/324. \end{aligned}$$

Therefore $P(B|HHH) = 32/59 \simeq 0.5424$, while $P(B) \simeq 0.3333$.

Note that $I(HHH,B) = \left(\frac{2}{3}\right)^3 : \frac{59}{324} = \frac{96}{59} = 1.6271$, so the prior 0.33333 and the observation HHH leads to the posterior $1.6271 \times 0.33333 = 0.5424$.

Mean with respect to a distribution (weighted average)

Suppose that for each outcome a_j of A we obtain an associated quantity E_j (a prize, an amount of energy, a bet, …). At the end of N runs of A, the total amount of this quantity will be

$$f_1 E_1 + f_2 E_2 + \cdots + f_n E_n.$$

Thus the *average amount per trial*, which we denote $\langle E \rangle$, is given by

$$\langle E \rangle = (f_1 E_1 + f_2 E_2 + \cdots + f_n E_n)/N$$

$$= \frac{f_1}{N} E_1 + \frac{f_2}{N} E_2 + \cdots + \frac{f_n}{N} E_n$$

or, taking the limit,

$$\langle E \rangle = p_1 E_1 + \cdots + p_n E_n.$$

Note that if $p_1 = \cdots = p_n$ (so $p_k = 1/n$ for all k), then

$$\langle E \rangle = (E_1 + \cdots + E_n)/n \quad (\textit{standard average}).$$