Basic ideas of probability

What *is* probability???

To say that the *probability* of something happening is the *chance* or it happening, or the *likelihood*, or any other synonym, does not address the issue.

Can we *measure* it? This would be a problem of applied statistics, and is a profound and confusing issue in itself. We will mainly ignore it.

Can we make *inferences* about probability?

Yes, and without knowing what it *truly* is and without worry about subtleties of *measuring* it.

(Note: the conversion between *chances-of* and *probability* is that chances-of is a percentage, while probability is a number between 0 and 1. So 34% chance is probability of 0.34)

Example: fairness. A **fair coin** is a coin with `heads` and *tails* **equally likely**. That is,

$$P(\texttt{heads}) = P(\texttt{tails})$$

It is merely a *normalization* that the sum of the probabilities of all the possible outcomes is 1, so

$$P(\texttt{heads}) + P(\texttt{tails}) = 1$$

That is, we have a system of the form

$$\begin{cases} x & = y \\ x + y & = 1 \end{cases}$$

which we solve (without knowing what probability is)

$$P(\texttt{heads}) = P(\texttt{tails}\ ) = \frac{1}{2}$$

We have completed a numerical computation without being able to answer any philosophical or other deeper questions.

Example: urns.

Suppose there are 4 red balls and 8 green balls in an *urn*, otherwise indistinguishable. (An **urn** is a large deep container, possibly for use as a large flower pot outdoors.)

As with the coin, we infer that the probabilities of drawing the $12 = 4 + 8$ balls are all the same, and add up to 1, so are all $1/12$.

It is a small leap to infer that, since drawing one of the balls precludes drawing any other, that the probability of drawing a *red* ball is

$$P(\texttt{red}) = P(\texttt{red}_1)+P(\texttt{red}_2)+P(\texttt{red}_3)+P(\texttt{red}_4)$$

$$= \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = 4 \cdot \frac{1}{12}$$

and that of drawing a `green` ball is, similarly,

$$P(\texttt{green}) = 8 \cdot \frac{1}{12}$$

Note: The fact that we can correctly analyze the situation by distinguishing $\texttt{red}_1$ from $\texttt{red}_2$, etc., appears to be correct in macroscopic phenomena.

It is false for subatomic particles called **bosons**.

It is correct for **fermions**.

Example: independence. The apparent fact that the outcome of *one* flip of a coin has no effect on the outcome of *another* flip of a coin is the **independence** of the two events.

(This has nothing to do with the *fairness* (or not) of the coin.)

That is, neither the coin nor the universe remember prior flips, and do not try to compensate or make up for too many heads in the past, etc.

The apparent fact that the outcome of *one* drawing (with replacement) of a ball from an urn has no effect on the outcome of *another* draw of a ball from that urn is the **independence** of the two events.

In a different world this could have been otherwise.

(One way to formally model a sequence of events in which the outcome of the next event *can* be affected by the previous one is (roughly) a **Markov process**.)

NOT the definition of probability

It *turns out* **not** wise to define the probability of an outcome of an event as

$$P(\texttt{outcome})$$

$$= \lim_{\textit{trials} \to \infty} \frac{\textit{no. times}\ \texttt{outcome}\ \textit{occurs}}{\textit{total no. trials}}$$

(Yet this statement *is true*, and is a *theorem*, the Law of Large Numbers.)

Cannot do infinitely many tests.

Do not know how rapidly the result of a finite number of tests approaches the limit.

Do not know that the limit exists in any sense.

Might evaluate the limit on different days and get different answers?

Different people might get different approximations?

Formalizing ideas of probability

A **probability space** or **event space** is a set Ω together with a **probability measure** P on it. This means that to each subset $A \subset \Omega$ we associate the **probability**

$$P(A) = \text{ probability of } A$$

with $0 \leq P(A) \leq 1$. The probability of the whole space is normalized to be $P(\Omega) = 1$, and $P(\phi) = 0$.

A subset $A \subset \Omega$ is called an **event**.

For an element $\omega \in \Omega$ we may call ω an **atomic event**, and write

$$P(\omega) = P(\{\omega\})$$

For a *compound event* $A = \{\omega_1, \ldots, \omega_n\} \subset \Omega$

$$P(A) = P(\omega_1) + \ldots P(\omega_n)$$

For two *disjoint* subsets A and B of Ω, say that A and B are **disjoint events**. For disjoint events A and B we take an *axiom*

$$P(A \cup B) = P(A) + P(B)$$

Two events A, B are **independent** if

$$P(A \cap B) = P(A) \cdot P(B)$$

Union of events is **'or'**, and **intersection** of events is **'and'**:

$$P(A \text{ or } B) = P(A \cup B)$$

$$P(A \text{ and } B) = P(A \cap B)$$

We do not try to say *what* probability is, nor how to *measure* it.

Re-interpretation of real-life questions into this formalism is a significant issue.

Example: The probability space for flipping a fair coin is

$$\Omega = \{\text{heads}, \text{tails}\}$$

with

$$P(\text{heads}) = \frac{1}{2} \qquad P(\text{tails}) = \frac{1}{2}$$

Little is accomplished by the formalization in this example.

Example: The probability space for drawing (with replacement) a ball from an urn containing 3 red balls and 4 green balls (otherwise indistinguishable) is

$$\Omega = \{r_1, r_2, r_3, g_1, g_2, g_3, g_4\}$$

where the r_is are the red balls and the g_is are the green ones. The probability measure $P()$ is

$$P(\text{any single ball}) = \frac{1}{3+4} = \frac{1}{7}$$

The probability of drawing *some* red ball is

$$P(\{r_1, r_2, r_3\}) = P(r_1) + P(r_2) + P(r_3)$$

$$= \frac{1}{7} + \frac{1}{7} + \frac{1}{7}$$

since the (atomic) events r_1, r_2, r_3 are *disjoint*.

Example: The probability space for flipping a fair coin 3 times is

$$\begin{aligned} \Omega &= \{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH}, \\ &\quad \text{HTT}, \text{THT}, \text{TTH}, \text{TTT}\} \end{aligned}$$

The event

$$A = \text{ get an H on the first flip}$$

is

$$A = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}, \}$$

The event

$$B = \text{ get an H on the second flip}$$

is

$$B = \{\text{HHH}, \text{HHT}, \text{THH}, \text{THT}\}$$

The assumed *independence* of the different flips says things like

$$P(\text{H on first } \textit{and} \text{ second flip})$$

$$= P(\text{H on first flip}) \cdot P(\text{H on second flip})$$

$$= \frac{1}{2} \cdot \frac{1}{2}$$

and

$$P(\text{HHH}) = P(\text{H on first, second, third})$$

$$= P(\text{H on first}) \cdot P(\text{H on first}) \cdot P(\text{H on first})$$

$$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$

The fairness and independence together imply that

$$P(\text{any 3-flip pattern of H's and T's}) = \frac{1}{2^3}$$

Example: What is the probability of at exactly 2 heads in 3 flips of a fair coin? (Use the previous set-up.)

$$P(\text{exactly two H's in 3 flips})$$
$$= P(\{\text{HHT}, \text{HTH}, \text{THH}\})$$
$$= P(\text{HHT}) + P(\text{HTH}) + P(\text{THH}) = \frac{3}{8}$$

by disjointness.

But this explicit *listing* approach **scales badly**.

For example, the event space for the question of the probability of getting exactly 17 heads in 30 flips of a fair coin has

$$2^{30} \sim 1{,}000{,}000{,}000$$

elements.

Example: What is the probability of getting exactly 17 heads in 30 flips of a fair coin?

The independence and fairness together imply that the probability of each *single* pattern of 30 H's and T's has probability $1/2^{30}$.

The different patterns of 17 heads from among an ordered list of 30 are counted as the number of choices of 17 locations from among 30. There are 30 choices for the location of the 'first' H, $30-1$ for the second, etc. up to $30-(17-1)$ for the 17^{th}. And then divide by 17! since the order of the selections does not matter, giving

$$\text{number of patterns with 17 H's} = \binom{30}{17}$$

Since each has probability $1/2^{30}$ and they are disjoint,

$$P(\text{17 heads in 30}) = \binom{30}{17} \cdot \frac{1}{2^{30}}$$

Example: What is the probability of getting *at least* 4 heads in 6 flips of a fair coin?

The new idea here is to break the compound event into convenient smaller disjoint ones

$$P(\text{at least 4 in 6 flips})$$

$$= P(\text{exactly } 4) + P(\text{exactly } 5) + P(\text{exactly } 6)$$

Then, as in the previous example, this is

$$\binom{6}{4} \cdot \frac{1}{2^6} + \binom{6}{5} \cdot \frac{1}{2^6} + \binom{6}{6} \cdot \frac{1}{2^6}$$

$$= \frac{15 + 6 + 1}{64}$$

Example: There are 3 blue balls and 2 red balls in an urn. What is the probability of drawing at exactly 4 blue balls out of 7 draws (with replacement)?

As usual, we assume that the different draws are *independent*. The probability of drawing a blue ball in a single draw is $3/5$, and the probability of drawing a red ball in a single draw is $2/5$, since the total number of balls is $5 = 3 + 2$ and we assume that they have the same probability of being drawn.

The independence means that the probability of any pattern of colors is the product of the individual probabilities. For example,

$$P(\mathrm{RRB}) = P(\mathrm{R}) \cdot P(\mathrm{R}) \cdot P(\mathrm{B})$$

$$P(\mathrm{RRBR}) = P(\mathrm{R}) \cdot P(\mathrm{R}) \cdot P(\mathrm{B}) \cdot P(\mathrm{R})$$

Thus, for any pattern with 4 B's and 3 R's,

$$P(\text{BBBRRRR}) = P(\text{BBRBRRR})$$

$$= P(\text{RRBBBRR}) = \ldots = P(B)^4 \cdot P(R)^3$$

The *number* of ways to draw exactly 4 blue balls in 7 draws is equal to the number of ways of choosing 4 things from 7, $\binom{7}{4}$.

Together, the probability of drawing exactly 4 blue balls in 7 draws from an urn with 3 blue and 2 red balls is

$$\binom{7}{4} \cdot P(B)^4 \cdot P(R)^3 = \binom{7}{4} \cdot \left(\frac{3}{5}\right)^4 \cdot \left(\frac{2}{5}\right)^3$$

Example: There are 3 blue balls and 2 red balls in an urn. What is the probability of drawing at least 7 blue balls out of 9 draws (with replacement)?

We start where we left off in the previous problem. To draw *at least* 7 blue balls means to draw *exactly* either $7, 7+1, 7+2$, which are *disjoint* events, so the probability of their union is the sum of their probabilities

$$P(\text{at least 7 in 9})$$
$$= P(\text{exactly 7}) + P(\text{exactly 8}) + P(\text{exactly 9})$$

The *number* of ways to draw exactly ℓ blue balls in 9 draws is equal to the number of ways of choosing ℓ things from 9, the binomial coefficient $\binom{9}{\ell}$.

As in the previous problem, the probability of drawing exactly ℓ blue balls in 9 draws is

$$\binom{9}{\ell}(\frac{3}{5})^{\ell}(\frac{2}{5})^{9-\ell}$$

Adding up these probabilities of *disjoint* events, the desired total probability is

$$P(\text{at least 7 blue in 9})$$

$$= P(\text{exactly } 7) + P(\text{exactly } 8) + P(\text{exactly } 9)$$

$$= \binom{9}{7}\left(\frac{3}{5}\right)^{7}\left(\frac{2}{5}\right)^{9-7} + \binom{9}{8}\left(\frac{3}{5}\right)^{8}\left(\frac{2}{5}\right)^{9-8}$$

$$+\binom{9}{9}\left(\frac{3}{5}\right)^{9}\left(\frac{2}{5}\right)^{9-9}$$

Random variables

A **random variable** X is really just a real-valued function on a probability space Ω (which, recall, is basically a set with a probability measure on its subsets).

For a real number x, the **probability that** X **takes value** x is denoted $P(X = x)$, and by definition is

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

Example: For $\Omega = \{H, T\}$ the sample space for flipping a fair coin, define a random variable X for $\omega \in \Omega$ by

$$X(\omega) = \text{ no. heads when } \omega \text{ occurs}$$

Yes,

$$X(\text{H}) = 1 \qquad X(\text{T}) = 0$$

For Ω the set of outcomes ω of 4 flips of a fair coin we could similarly define

$$X(\omega) = \text{no. H's occuring in } \omega$$

In this example, the notation means

$$\begin{aligned}
P(X = 0) &= P(\text{zero Hs in 4 flips}) \\
P(X = 1) &= P(\text{exactly one H in 4 flips}) \\
P(X = 2) &= P(\text{exactly two Hs in 4 flips}) \\
P(X = 3) &= P(\text{exactly three H in 4 flips}) \\
P(X = 4) &= P(\text{exactly four H in 4 flips})
\end{aligned}$$

For any other real value x, $P(X = x) = 0$, since we can't get any other number of Hs in 4 flips.

Expected values

The **expected value** $E(X)$ or EX of a random variable X on a probability space Ω is a kind of *weighted average* of the values of X, with the weights being the probabilities of the different inputs/outputs. The precise definition is

$$\text{expected value of } X = E(X)$$

$$= \sum_{\omega \in \Omega} P(\omega) \cdot X(\omega)$$

We can *group* the inputs according to the output value produced, so this is also equal to

$$E(X) = \sum_{\text{values } x \text{ of } X} P(X = x) \cdot x$$

where (again) the notation $P(X = x)$ means the probability that X takes value x:

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

About notation

Yes, the notation and terminology for random variables is different from, and in conflict with, the notation used for functions and their values in calculus and differential equations.

First, and most importatly, yes, random *variables* are actually *functions*.

Yes, the random variable's name is often X, unlike the f or g in calculus.

Yes, usually the *input* to a function is called x, not the *output*, as in $X(\omega) = x$.

Examples of expected values

With X being the random variable counting `H`s in a single flip of a fair coin,

$$E(X) = \sum_{\text{values } x \text{ of } X} P(X = x) \cdot x$$

$$= P(X = 0) \cdot 0 + P(X = 1) \cdot 1$$

$$= \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}$$

Note that we will never actually get 1/2 head in a flip of a fair coin.

But, as with many averages, the average or weighted average of integer values may be a non-integer.

That's ok.

With X being the random variable counting `H`s in 3 flips of a fair coin,

$$E(X) = \sum_{\text{values } x \text{ of } X} P(X = x) \cdot x$$

$$= P(X = 0) \cdot 0 + P(X = 1) \cdot 1$$

$$+ P(X = 2) \cdot 2 + P(X = 3) \cdot 3$$

$$= \binom{3}{0} 2^{-3} \cdot 0 + \binom{3}{1} 2^{-3} \cdot 1$$

$$+ \binom{3}{2} 2^{-3} \cdot 2 + \binom{3}{3} 2^{-3} \cdot 3$$

$$= \frac{0 + 3 \cdot 1 + 3 \cdot 2 + 1 \cdot 3}{8} = \frac{3}{2}$$

This may be an intuitively appealing answer, if we imagine that we get an *average* of $1/2$ head per flip in 3 flips.

But notice that the *definition* hands us an expression whose value is not obviously the answer what we expect, though it turns out to be so.
