Basic 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.

[0.1] **Example:** The probability space for flipping a fair coin is

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

with

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

Little is accomplished by the formalization in this example.

[0.2] **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*.

[0.3] **Example:** The probability space for flipping a fair coin 3 times is

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

The event

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

is

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

The event

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

is

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

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

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

[0.4] **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.

[0.5] **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,

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

[0.6] **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 are 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

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

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

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

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

Conditional probability

The **conditional probability** that an event A will occur **given** that an event B occurs is defined to be

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

[0.7] **Example:** The conditional probability that at a fair coin comes up heads in at least 3 of 6 flips, given that the first two flips are tails, is

$$P(\text{at least 3 H's in 6}|\text{first two T's})$$

$$= \frac{P(\text{first two T's } \textit{and} \text{ at least 3 H's})}{P(\text{first two T's})}$$

Birthday paradox

It may seem strange that *in a set of at least 23 people the probability is* $\geq 1/2$ *that two have the same birthday.*

Not 365/2, but more like $\sqrt{365}$.

For n things chosen at random with equal probabilities (and independently) from N things (with replacement), for

$$n > \sqrt{2 \ln 2} \cdot \sqrt{N} \sim \frac{17}{10} \cdot \sqrt{N}$$

the probability that two things are the same is $> \frac{1}{2}$.

This possibly counter-intuitive fact is the basis for a type of *attack* on ciphers, authentications, and signature schemes. These attacks are called **birthday attacks**.

[0.8] **Example:** Suppose that the **authenticity** of an electronic document is to be proven by computing a certain *hash function* of it.

A **hash function** or is a function that accepts arbitrarily large inputs and computes a fixed-size output in a manner that is essentially *irreversible*. The intent is that, given an output value of a hash function, it is virtually impossible to *contrive* an input to yield that value.

That is, it is intended that if the output of a hash function is t bits, allowing 2^t different output values, then to create an input with a pre-specified output would take on the average $2^t/2$ guesses.

But in some circumstances there is a *birthday attack*.

Suppose that your adversay, with whom you will sign a contract, prepares a *genuine* contract and a *bad* one (in which you give the adversary everything).

A *birthday attack* consists of the adversary making (automated) changes of spacing, punctuation, and other inessentials in *both* the genuine and the fake, until *one* of the genuine ones matches *one* of the fakes.

For a t-bit hash function, instead of taking $\sim 2^t$ documents, because of the birthday paradox principle, it will take more like $\sim \sqrt{2^t}$ real documents and $\sim \sqrt{2^t}$ fakes.

For example, with a 32-bit hash function, instead of a scam requiring

$$\sim 2^{32} > 4,000,000,000$$

alternative documents to be prepared, the attacker needs only roughly

$$\sim 2 \cdot 2^{16} \sim 130,000$$

Computation for birthday paradox

We compute the probability that *no two* outcomes are the same, and subtract this result from 1 to obtain the desired result.

After two trials, there is $1/N$ chance that the second outcome was equal to the first one, so the probability is $1 - \frac{1}{N}$ that the outcomes of two trials will be different.

After 3 trials, *given* that the first two outcomes are different, the conditional probability is $2/N$ that the third trial would give an outcome equal to *one* of the first two. Thus, given that the first two outcomes are different, the conditional probability that the third will differ from both is $1 - \frac{2}{N}$. Since the probability that the first two were different was $1 - \frac{1}{N}$, the formula above gives

$$P(\text{first 3 different}) = (1 - \frac{1}{N})\,(1 - \frac{2}{N})$$

After 4 trials, *given* that the first two outcomes are different, the *conditional probability* is $3/N$ that the third trial would give an outcome equal

to *one* of the first two. Thus, given that the first two outcomes are different, the conditional probability that the third will differ from all of the first 3 is $1-\frac{3}{N}$. Using the previous step, and the formula above,

$$P(\text{first 4 different}) = (1-\frac{1}{N})\,(1-\frac{2}{N})\,(1-\frac{3}{N})$$

Continuing, we get

$$P(n \text{ trials all different})$$

$$= (1-\frac{1}{N})\,(1-\frac{2}{N})\,(1-\frac{3}{N})\ldots(1-\frac{n-1}{N})$$

The logarithm of the probability that they're all different is

$$\ln(1-\frac{1}{N})+\ln(1-\frac{2}{N})+\cdots+\ln(1-\frac{n-1}{N})$$

The first-order Taylor expansion for $\ln(1-x)$ for $|x|<1$

$$\ln(1-x)=-(x+\frac{x^2}{2}+\frac{x^3}{3}+\frac{x^4}{4}+\cdots)$$

In particular for $0<x<1$

$$\ln(1-x)\leq -x$$

so

$$\ln(1-\frac{1}{N})+\ln(1-\frac{2}{N})+\cdots+\ln(1-\frac{n-1}{N})$$

$$\leq -(\frac{1}{N}+\frac{2}{N}+\cdots+\frac{n-1}{N})$$

Recall

$$1 + 2 + 3 + 4 + \cdots + (k-1) + k = \frac{1}{2}k(k+1)$$

Then

$$\ln\left(P(n \text{ trials all different})\right) \leq \frac{-\frac{1}{2}(n-1)n}{N}$$

As n gets larger and larger, the expression $(n-1)n$ is for practical purposes n^2. Thus, we have an *approximate* formula

$$\ln\left(P(n \text{ trials all different})\right) \leq -\frac{n^2}{2N}$$

or

$$P(n \text{ trials all different}) \leq e^{-n^2/2N}$$

$$P(2 \text{ of } n \text{ trials the same}) \geq 1 - e^{-n^2/2N}$$

The probability that some two will be the same is therefore bigger than or equal 1/2 when the probability that no two are the same is *less* than 1/2. Thus, for given N we *solve* to find the smallest n so that

$$-\frac{n^2}{2N} < \ln \frac{1}{2}$$

which gives the formula

$$n \geq \sqrt{2 \cdot \ln 2} \cdot \sqrt{N} \sim \frac{1.17}{10} \cdot \sqrt{N}$$

for the size of n to assure that the probability is bigger than 1/2 that two choices are the same.