We can generalize the concept of measurement by introducing a unique number as a *label* to characterize each state ω in the set Ω. This would provide a label function $L(\omega)$ as a random variable. Our condition of uniqueness makes $L(\omega)$ isomorphic, so that the label can be used interchangeably with the full state,

$$\omega \longleftrightarrow L(\omega).$$

While the measurements $X(\omega)$ we consider will generally not produce a unique number for each ω, we will design them precisely to remove irrelevant information that doesn't interest us. Ignoring that irrelevant information leaves us free to interchange the set of outcomes A for the set of states Ω. (Some textbooks may never distinguish between A vs Ω in the first place, though this can be a source of confusion.)

Only a couple of definitions remain. The next is to define an **event** to be any subset of the set of all outcomes A. For example, events resulting from rolling a die could include (i) rolling a 6, (ii) rolling anything but a 6, (iii) rolling any even number, and many more. Collecting all events of interest defines the **set of events** (or **event space**) $\mathcal{F}$.

We are now prepared for the final foundational definition in this section, the **probability** P of an event in the set $\mathcal{F}$. Mathematically, P is a *measure function*,

$$P : \mathcal{F} \to [0, 1],$$

which must satisfy the following two requirements:

1. The probability of a countable union of mutually exclusive events must equal the sum of the probabilities of each of these events.
2. The probability of the outcome space ($\mathcal{F} = A$) must equal 1 (even if A is uncountable). This simply means that the experiment $\mathcal{E}$ must produce an outcome. If no outcome were produced, it would not make sense to say that the experiment had occurred.

Combining the outcome space, event space and probability measure gives us a *probability space* $(A, \mathcal{F}, P)$.

For example, consider an experiment that can only produce N possible states, so that

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}.$$

If two states are identical, $\omega_i = \omega_j$, they must produce the same measurement outcomes $X(\omega_i) = X(\omega_j)$, which implies the contrapositive

$$X(\omega_i) \neq X(\omega_j) \qquad \Longrightarrow \qquad \omega_i \neq \omega_j.$$

On the other hand, as described above, it is possible to have $X(\omega_i) = X(\omega_j)$ even when $\omega_i \neq \omega_j$. This means that the size n of the outcome space A may be smaller than the size of Ω, $n \leq N$. We can write

$$A = \{X_1, X_2, \cdots, X_n\},$$

where each X_α is distinct and its index does *not* necessarily correspond to that on ω_i. We can take the individual X_α themselves to be the events we're interested in, choosing the event space

$$\mathcal{F} = \{X_1, X_2, \cdots, X_n\} = A. \tag{1}$$

These events are all mutually exclusive by construction, so if we assign them probabilities

$$P(X_\alpha) \equiv p_\alpha \qquad \text{for } \alpha = 1, \cdots, n,$$

then the above requirements on probabilities demand that for any $\alpha \neq \beta$ we have

$$P(X_\alpha \text{ or } X_\beta) = p_\alpha + p_\beta$$

$$P(A) = P(X_1 \text{ or } X_2 \text{ or } \cdots \text{ or } X_n) = \sum_{\alpha=1}^{n} p_\alpha = 1.$$

Similarly choosing an event space $\mathcal{F} = A$ for the six-sided die considered in an earlier gap, what are the probabilities p_1 through p_6 that result from assuming the die is fair?

"Fair" $= p_\alpha = p = 1/6$

$$\sum_{\alpha=1}^{6} p_\alpha = 6p = 1$$

Again taking $\mathcal{F} = A$ for the case of tossing a coin four times, what are the probabilities p_α that result from assuming the coin is fair? If we instead consider the event space

$$\mathcal{F} = \{\text{equal number of } H \text{ and } T,\ \text{different numbers of } H \text{ and } T\},$$

what are the probabilities p_{equal} and p_{diff} for the two events in this $\mathcal{F}$?

$p_\alpha = \frac{1}{16}$

HHTT TTHH
HTHT THTH
HTTH THHT

$p_{\text{equal}} = \frac{6}{16} = \frac{3}{8}$

$p_{\text{diff}} = 1 - p_{\text{equal}} = \frac{5}{8}$

The standard European roulette wheel shown to the left (source) has 37 pockets labelled "0" through "36". 18 of these pockets are coloured red, 18 are coloured black and 1 (pocket "0") is coloured green. Note that measuring the label automatically provides the colour.

What is the outcome space A for a spin of the roulette wheel? With $\mathcal{F} = A$, what are the probabilities p_α for a fair wheel? With

$$\mathcal{F} = \{\text{ball in a red pocket, ball in a black pocket, ball in the green pocket}\},$$

what are the corresponding probabilities p_{red}, p_{black} and p_{green}?

$A = \{0, 1, 2, \cdots 36\}$

$\sum_{\alpha=0}^{36} p_\alpha = 1 \rightarrow p_\alpha = 1/37$

$p_{\text{red}} = p_1 + p_3 + \cdots + p_{34} = \frac{18}{37}$

$p_{\text{black}} = \frac{18}{37} = p_2 + p_4 + \cdots + p_{35}$

$p_{\text{green}} = \frac{1}{37}$

The process of assigning probabilities to events is called *modelling*. In the gaps above we saw above that *symmetries* are a powerful way to constrain probabilities. The symmetry between the six sides of a fair die, the two sides of a fair coin, and the 37 pockets of a fair roulette wheel each sufficed to completely fix the corresponding probabilities p_α.

Modelling can also be guided by empirical data obtained by repeating an experiment many times. For example, if we don't know whether a set of dice are fair, we will be able to infer their probabilities p_α (with a certain confidence level) by rolling them enough times. The need to repeat the experiment many times comes from the law of large numbers, to which we now turn.

1.2 Law of large numbers

Let's return to the setup leading to Eq. 1 above, with

$$\mathcal{F} = A = \{X_1, X_2, \cdots, X_n\}$$

for finite n, and probabilities $p_\alpha = P(X_\alpha)$ that obey

$$p_\alpha \in [0, 1] \qquad \sum_{\alpha=1}^{n} p_\alpha = 1.$$

We can generalize this notation by writing instead

$$\sum_{X \in A} P(X) = 1,$$

which provides simple expressions for the **mean** μ and **variance** σ^2 of the probability space,

$$\mu = \langle X \rangle = \sum_{X \in A} X\, P(X) \tag{2}$$

$$\sigma^2 = \left\langle (X-\mu)^2 \right\rangle = \sum_{X \in A} (X-\mu)^2\, P(X). \tag{3}$$

The angle bracket notation indicates the **expected** (or **expectation**) **value** with general definition

$$\langle f(X) \rangle = \sum_{X \in A} f(X)\, P(X), \tag{4}$$

which is a linear operation,

$$\langle c \cdot f(X) + g(X) \rangle = c \langle f(X) \rangle + \langle g(X) \rangle .$$

The square root of the variance, $\sqrt{\sigma^2} = \sigma$, is the **standard deviation**. What is σ expressed in terms of $\langle X^2 \rangle$ and $\langle X \rangle^2$?

$\mu = \langle X \rangle$

$$\sigma^2 = \langle (X-\mu)^2 \rangle = \langle X^2 - 2X\mu + \mu^2 \rangle = \langle X^2 \rangle - 2\mu\langle X \rangle + \mu^2$$
$$= \langle X^2 \rangle - 2\langle X \rangle^2 + \langle X \rangle^2$$
$$= \langle X^2 \rangle - \langle X \rangle^2$$
$$\sigma = \sqrt{\langle X^2 \rangle - \langle X \rangle^2}$$

We now define a new experiment that consists of *repeating* the original experiment R times, with each repetition independent of all the others. Using the same measurement as before for each repetition, we obtain a new outcome space that we can call B. For $R = 4$, what are some representative outcomes in the set B? What is the total size of B?

$A = \{X_1, X_2, \ldots X_n\}$ $\quad R = 4$

$B \supset \{ X_1X_1X_1X_1,$
$X_4X_1X_{42}X_1,$
$X_nX_1X_1X_1 \ldots \}$

$\#A = n$

$\#B = n \cdot n \cdot n \cdot n \cdots n$
$= n^R$

Each outcome in B contains R different $X^{(r)} \in A$, one for each repetition $r = 1, \cdots, R$, and each with mean $\langle X^{(r)} \rangle = \mu$ and variance $\langle (X^{(r)} - \mu)^2 \rangle = \sigma^2$. Considering the case $R = 4$ for simplicity, any element of B can be written as $X_i^{(1)} X_j^{(2)} X_k^{(3)} X_l^{(4)} \in B$ with corresponding probability

$$P_B\left(X_i^{(1)} X_j^{(2)} X_k^{(3)} X_l^{(4)}\right) = P_A\left(X_i^{(1)}\right) P_A\left(X_j^{(2)}\right) P_A\left(X_k^{(3)}\right) P_A\left(X_l^{(4)}\right),$$

using subscripts to distinguish between the single-experiment (A) and repeated-experiment (B) probability spaces.

Averaging over all R repetitions defines the *arithmetic mean*

$$\overline{X}_R = \frac{1}{R} \sum_{r=1}^{R} X^{(r)}. \tag{5}$$

Unlike the true mean μ, the arithmetic mean $\overline{X}_R$ is a random variable—a number that may be different for each element of B. That said, $\overline{X}_R$ and μ are certainly related, and so long as the standard deviation exists—that is, so long as σ^2 is finite—this relation can be proved rigorously in the limit $R \to \infty$.[2]

1 Feb.

Here we will not be fully rigorous, and take it as given that

$$\left\langle \left(X^{(i)} - \mu\right)\left(X^{(j)} - \mu\right) \right\rangle = \sigma^2 \delta_{ij} = \begin{cases} \sigma^2 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases},$$

where the *Kronecker delta* $\delta_{ij} = 1$ for $i = j$ and vanishes for $i \neq j$. This is a consequence of the assumed independence of the different repetitions. Using this result and the relation $\left(\sum_i a_i\right)\left(\sum_j b_j\right) = \sum_{i,j} (a_i b_j)$, express the following quantity in terms of σ and R:

$$\left\langle \left(\frac{1}{R} \sum_{r=1}^{R} X^{(r)} - \mu \right)^2 \right\rangle =$$

You should find that your result vanishes in the limit $R \to \infty$, so long as σ^2 is finite. Since the square makes this expectation value a sum of non-negative terms, it can vanish only if every one of those terms is individually zero.

[2] In the computer project we will numerically investigate a situation where σ^2 diverges.