Probability Tutorial

In this Tutorial...

- Probability Space

- Conditional Expectations

We look at the triple $(\Omega, \mathcal{F}, \mathbb{P})$, called a *probability space.*

It forms the foundation of the *probabilistic universe*. This probability space comprises of

1. the sample space Ω

2. the filtration $\mathcal{F}$

3. the probability measure $\mathbb{P}$

Example:

The daily closing price of a risky asset, e.g. share price on the FTSE100. Over the course of a year (252 business days)

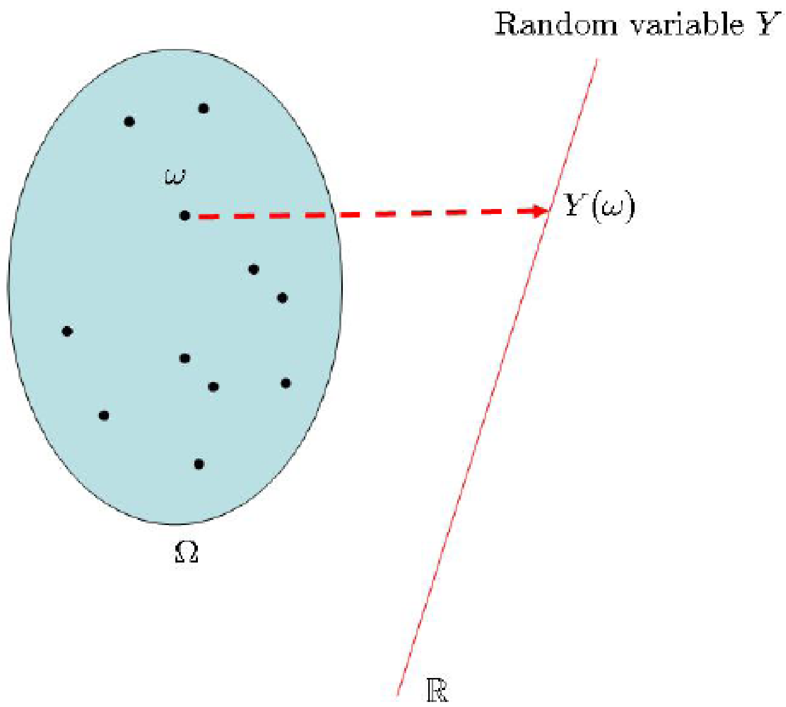
$$\Omega = \{S_1, S_2, S_3, \ldots, S_{252}\}$$

We could define an event e.g. $\Psi = \{S_i : S_i \geq 110\}$

A *random variable* (RV) Y is a function which maps from the sample space Ω to the set of real numbers

$$Y : \omega \in \Omega \longrightarrow \mathbb{R},$$

i.e. it associates a number $Y(\omega)$ with each outcome ω.



Consider the example of tossing a coin and suppose we are paid £1 for each head and we lose £1 each time a tail appears. We know that $\mathbb{P}(\mathsf{H}) = \mathbb{P}(\mathsf{T}) = \frac{1}{2}$. So now we can assign the following outcomes

$$\mathbb{P}(1) = \frac{1}{2};\ \mathbb{P}(-1) = \frac{1}{2}$$

Mathematically, if our random variable is X, then

$$Y = \begin{cases} +1 & \text{if H} \\ -1 & \text{if T} \end{cases}$$

or using the notation above $Y : \omega \in \{\mathsf{H,T}\} \to \{-1, 1\}$.

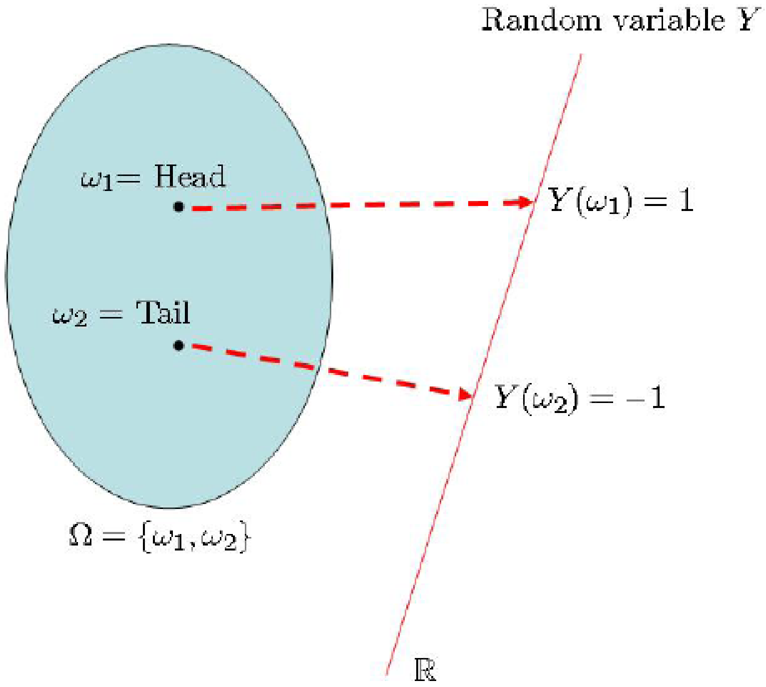
For the coin tossing game we see the sample space Ω has two events: $\omega_1 =$ Head; $\omega_2 =$ Tail. So now

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

And the P&L from this game is a RV Y defined by

$$\begin{aligned} Y(\omega_1) &= +1 \\ Y(\omega_2) &= -1 \end{aligned}$$

The earlier diagram becomes



While recording information obtained from a coin tossing game is manageable, imagine storing 100 years of S&P500 data. How can we keep track of an ever expanding sample space (in a simple yet elegant manner)?

We do this by introducing the *filtration.*

The filtration is the mathematical object that keeps track of how (the increasing flow of) information evolves; the fact that there is a binomial tree, from time 1 to time 2 there are 2 possibilities from time 2 to time 3 there are 8 possibilities in total, etc. After just 10 periods there are $2^{10} = 1024$ outcomes!

The filtration, $\mathcal{F}$, is an indication of how an increasing family of events builds up over time as more results become available, it is much more than just a family of events. The filtration, $\mathcal{F}$ is a set formed of all possible combinations of events $A \subset \Omega$, their unions and complements.

Adapted (Measurable) Process

A stochastic process S_t is said to be adapted to the filtration $\mathcal{F}_t$ (or measurable with respect to $\mathcal{F}_t$, or $\mathcal{F}_t$ -adapted) if the value of S_t at time t is known given the information set $\mathcal{F}_t$.

Probability Measures

Recap: Recall the probability density function $p(x)$ with the property that

$$\mathsf{Pr}(x < X < x + dx) = p(x)\, dx$$

for any infinitessimal interval of length dx (think of this as a limiting process starting with a small interval whose length tends to zero). It is also called a density because it is the probability of finding X on an interval of length dx divided by the length of the interval. Recall that the following are analogous

$$\begin{aligned} \int_{-\infty}^{\infty} p(x)\, dx &= 1 \\ \sum_i p_i &= 1. \end{aligned}$$

The (cumulative) distribution function of a random variable is defined by

$$P(x) = \mathsf{Pr}(X < x).$$

It is an increasing function of x with $P\left(-\infty\right) = 0$ and $P\left(\infty\right) = 1$; note that $0 \leq P\left(x\right) \leq 1$. It is related to the density function by

$$p\left(x\right) = \frac{dP\left(x\right)}{dx}$$

provided that $P\left(x\right)$ is differentiable. Unlike $P\left(x\right)$, $p\left(x\right)$ may be unbounded or have singularities such as delta functions.

$\mathbb{P}$ is the probability measure, a special type of "function", called a measure, assigning probabilities to subsets (i.e. the outcomes); the mathematics emanates from Measure Theory. Probability measures are similar to cumulative density functions (CDF); the chief difference is that where PDFs are defined on intervals (e.g. $\mathbb{R}$), probability measures are defined on general sets. We are now concerned with mapping subsets on to $\left[0, 1\right]$.

There is a very powerful relation between expectations and probabilities. In our expectation formula, choose $f(X)$ to be the *indicator function* $\mathbf{1}_{x \in A}$ for a subset $A \subset \Omega$ defined

$$\mathbf{1}_{x \in A} = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

i.e. when we are in A, the indicator function returns 1.

If two measures $\mathbb{P}$ and $\mathbb{Q}$ share the sample space Ω with $A \subset \Omega$ and if $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$, we say that Q is *absolutely continuous* with respect to $\mathbb{P}$ and denote this by $\mathbb{Q} << \mathbb{P}$.

The key point is that all impossible events under $\mathbb{P}$ remain impossible under $\mathbb{Q}$. The probability mass of the possible events will be distributed differently under $\mathbb{P}$ and $\mathbb{Q}$. Put informally, we can change the probabilities as long as we do not change the (im)possibilities.

If $\mathbb{Q} << \mathbb{P}$ and $\mathbb{P} << \mathbb{Q}$ then the two measures are said to be *equivalent*, denoted $\mathbb{P} \sim \mathbb{Q}$. This extremely important result is formalised in the Radon-Nikodym Theorem: If the measures $\mathbb{P}$ and $\mathbb{Q}$ share the same null sets, then, there exists a random variable Λ such that for all subsets $A \subset \Omega$

$$\mathbb{Q}(A) = \int_A \Lambda d\mathbb{P}$$

where

$$\Lambda = \frac{d\mathbb{Q}}{d\mathbb{P}}$$

is called the Radon-Nikodym derivative.

Conditional Expectations

What makes a conditional expectation different (from an unconditional one) is information (just as in the case of conditional probability). In our probability space, $(\Omega, \mathcal{F}, \mathbb{P})$ information is represented by the filtration $\mathcal{F}$; hence a conditional expectation with respect to the (usual information) filtration seems a natural choice.

$$Y = \mathbb{E}\left[X \mid \mathcal{F}\right]$$

is the expected value of the random variable conditional upon the filtration set $\mathcal{F}$. In general

- In general Y will be a random variable

- Y will be adapted to the filtration $\mathcal{F}$.

Conditional expectations have the following useful properties: If X, Y are integrable random variables and a, b are constants then

1. **Linearity**:

$$\mathbb{E}\left[aX + bY \middle| \mathcal{F}\right] = a\mathbb{E}\left[X \middle| \mathcal{F}\right] + b\mathbb{E}\left[Y \middle| \mathcal{F}\right]$$

2. **Tower Property (i.e. Iterated Expectations)**: if $\mathcal{F} \subset \mathcal{G}$

$$\mathbb{E}\left[\mathbb{E}\left[X \middle| \mathcal{G}\right] \middle| \mathcal{F}\right] = \mathbb{E}\left[X \middle| \mathcal{F}\right]$$

   This property states that if taking iterated expectations with respect to several levels of information, we may as well take a single expectation subject to the smallest set of available information.

3. As a special case of the Tower property, we have

$$\mathbb{E}\left[\mathbb{E}\left[X \middle| \mathcal{F}\right]\right] = \mathbb{E}\left[X\right]$$

   since "no filtration" is always a smaller information set than any filtration.

4. "Taking out what is known": If X is $\mathcal{F}-$measurable, then the value of X is known once we know $\mathcal{F}$. Therefore,

$$\mathbb{E}\left[X|\mathcal{F}\right] = X$$

5. **Independence:** If X is independent from $\mathcal{F}$, then knowing $\mathcal{F}$ does not assist in predicting the value of X. Hence

$$\mathbb{E}\left[X|\mathcal{F}\right] = \mathbb{E}\left[X\right]$$

6. **Positivity:** If $X \geq 0$ then $\mathbb{E}\left[X|\mathcal{F}\right] \geq 0$.

7. **Jensen's Inequality:** Let f be a convex function, then

$$f\left(\mathbb{E}\left[X|\mathcal{F}\right]\right) \leq \mathbb{E}\left[f\left(X\right)|\mathcal{F}\right]$$