

## III. Randomness over time: Stochastic Process

### 1. Reminders

- First recall basic randomness without time involved: a *random variable* (a quantity which can take various values depending on some random outcome)
- Expectation (or mean):  
Weighted average of the possible values, weighted according to their probabilities.

Discrete case:

$$EX := \sum_i x_i f(x_i)$$

Continuous case:

$$EX := \int x f(x) dx ,$$

where  $f(x)$  is the density of the random variable  $X$ .

- Expectation is linear:

$$E(aX + bY) = aEX + bEY.$$

- Variance: how variable is  $X$ ?

$$\text{var}(X) := E[(X - EX)^2]$$

- By linearity of the expectation, we have:

$$\begin{aligned} \text{var}(X) &= E[X^2 - 2(EX)X + (EX)^2] \\ &= E[X^2] - 2EX \cdot EX + (EX)^2 \\ &= E[X^2] - [EX]^2 \end{aligned}$$

- Note that the variance measure how widely scattered the values of  $X$  are about the mean.

- Also:  $\text{var}(X) = 0$  iff  $(X - EX)^2 = 0$   
i.e.  $X = EX$ .

A random variable has variance 0 iff it is constant (hence non-random!).

- Also use the square root of the variance, the *standard deviation* (SD), which has the same dimensions as the mean.
- Standard notation:

$\mu$  for  $EX$ ,  $\sigma^2$  for  $var(X)$ ,  $\sigma$  for the SD

- Under affine transformation:

$$E[aX+b] = aEX+b$$

$$var(aX+b) = a^2 var(X)$$

$$SD(aX+b) = |a|.SD(X)$$

- The distribution function of  $X$  tells us how the values of  $X$  are distributed:

$$F(x) := P(X \leq x)$$

Discrete case:

$$F(x) = \sum_{\{i: x_i \leq x\}} f(x_i)$$

Continuous case:

$$F(x) = \int_{-\infty}^x f(u) \, du$$

- The Normal distribution:

Most important density in Statistics is the *standard normal density*

$$\phi(x) := \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\},$$

with distribution:

$$\Phi(x) = \int_{-\infty}^x \phi(u) \, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{1}{2}u^2\right\} \, du$$

We write  $N(0,1)$  to denote the normal distribution with mean 0 and variance 1

Note:

If  $X \sim N(0, 1)$

and  $\mu, \sigma > 0$  are constant,

then:  $\mu + \sigma X \sim N(\mu, \sigma^2)$  - the normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

Density is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \cdot \frac{(x - \mu)^2}{\sigma^2}\right\}$$

The normal distribution is the signature in Nature of a phenomenon which arises from the cumulative effect of a large number of individually small cause.

*e.g.*

- 1) Heights of male adults in the UK vary for genetic reasons (large number of genes each have a small effect).
- 2) Price movement of heavily traded stock under normal market conditions: look continuous from a distance but move by lots of small jumps in detail (caused by the very act of trading).

## **Two fundamental principles:**

1. Law of Large Numbers: if we average lots of independent values of  $X$ , the results tends to the mean.

$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$  converges to the expected value

$\bar{X}_n \rightarrow \mu$  for  $n \rightarrow \infty$ , where  $X_1, X_2, \dots$  is an infinite sequence of i.i.d integrable random variables with  $E[X_1] = E[X_2] = \dots = \mu$

2. Central Limit Theorem: errors are normally distributed about the mean.

For large enough  $n$ , the distribution of  $\bar{X}_n$  is close to the normal distribution with mean  $\mu$  and variance  $\sigma^2/n$ .

- The Poisson distribution (recall EPM944!)

The most important discrete distribution for us – discrete distribution of non-negative integers, defined for  $\lambda > 0$  by

$$P(X = k) = e^{-\lambda} \lambda^k / k! \quad (k = 0, 1, 2, \dots)$$

**Question 1:** This is a probability distribution – prove it.

**Question 2:** Show that

$$EX = \lambda, \quad \text{var}(X) = \lambda$$

- Higher dimensions

Let's consider the following random vector:  $X = (X_1, \dots, X_n)$

with mean vector:  $\mu = (\mu_1, \dots, \mu_n)$

Covariance matrix  $\Sigma = (\sigma_{ij})$ , where:

$$\sigma_{ij} = \text{cov}(X_i, X_j) := E[(X_i - EX_i)(X_j - EX_j)]$$

**Question:** What are the diagonal elements of the covariance matrix?

- Joint and marginal distribution function:

$$F(x) := P(X \leq x) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

This is called the *joint distribution function* of  $X = (X_1, \dots, X_n)$ , it tells us how  $X_1, \dots, X_n$  are distributed together.

On the other hand, each  $X_i$  has a distribution function:

$$F_i(x_i)$$

called a *marginal distribution* and these tell us how  $X_1, \dots, X_n$  are distributed separately.

Aside on Copula: The joint distribution determines the marginals (**Question:** How?).

In the other direction we need more information – contained in the *Copula function*. It couples the marginals together to give the joint distribution, via *Sklar's Theorem*:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

$X_1, \dots, X_n$  are independent if and only if:

$$F(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n)$$

Copulas are used in credit derivative pricing, where correlation between default times is often required (pricing of CDO tranches).

- Conditional distribution

Recall that if  $P(B) > 0$ :

$$P(A|B) := P(A \cap B) / P(B)$$

is the conditional probability of A given B.

- If  $(X, Y)$  are discrete with joint distribution

$$f(x_i, y_j) = P(X = x_i, Y = y_j)$$

Then, since:

$$\begin{aligned} f(x_i) &= P(X = x_i) = \sum_j P(X = x_i, Y = y_j) \\ &= \sum_j f(x_i, y_j) \end{aligned}$$

$$\begin{aligned} P(Y = y_j | X = x_i) &= P(X = x_i, Y = y_j) / P(X = x_i) \\ &= f(x_i, y_j) / \sum_j f(x_i, y_j) \end{aligned}$$

- Similarly, if  $(X, Y)$  are continuous, the conditional density of  $Y$  given  $X$  is:

$$f(y|x) = f(x, y) / \int f(x, u) du$$

Note: for random vector, one can similarly form the conditional distribution of some components given the values of others (see EPM944 lecture notes, p.10)

## 2. Stochastic Process

- Suppose: we are trying to model the evolution of some random phenomenon over time.

At each time  $t$  we have a random variable  $X_t$  .

So, for any finite set of  $n$  time points  $(t_1, \dots, t_n)$ , we have a random  $n$ -vector:

$$(X(t_1), \dots, X(t_n)) ,$$

which has a distribution (the collection of all of these give the *finite-dimensional distributions* of the random phenomenon we are trying to model).

- Filtrations

As time evolves, we learn more!

Assume that information is never forgotten.

Write:

$$\mathcal{F}_t$$

for the information available at time  $t$ , that is the values:

$$\{X_s : 0 \leq s \leq t\}$$

As  $t$  increases, the family  $\{\mathcal{F}_t\}$  is called a ***filtration***.

**Note:** in this course we confine ourselves to situations where all relevant information is publicly available

- Stochastic process

Formal definition involves a *filtered probability space* (a filtration + a probability space) – the relevant mathematics is **Measure Theory** (see references p1 of Lecture Notes).

So: a *stochastic process*

$$X = \{X_t : t \in I\}$$

is a family of random variables, defined on some common probability space, indexed by an index-set  $I$  (always in this course  $I$  will represent time).

### 3. Brownian Motion (also called *Wiener process*)

A *standard Brownian motion*  $B = (B_t)$  is a stochastic process which satisfies the following:

- I. each  $B_t$  is normally distributed with mean 0 and variance  $t$ :

$$B_t \sim N(0, t)$$

- II.  $B$  has independent increments: for  $s < t$ , the increment

$$B_t - B_s \text{ is independent of all } B_u, 0 \leq u \leq s.$$

- III. Increments  $B_t - B_s$  are normally distributed:  $B_t - B_s \sim N(0, t - s), t > s$

- IV.  $B$  has *continuous paths*: the random function (path)  $t \mapsto B_t$  is continuous.

Brownian motion are used in mathematical finance as a *driving noise* process.

It is the dynamic counterpart of the occurrence of the normal distribution in the *Central Limit Theorem*.

The future evolution of an individual stock price is uncertain, because of this we model it as random, *i.e.* as a stochastic process (what we see is the net outcome of a large number of causes, each reflecting some aspect of uncertainty).

## 4. Martingale

In Regression in Statistics:

$$\text{data} = \text{signal} + \text{noise}$$

Here, similarly we decompose the price movement into:

$$\text{Price movement} = \text{systematic component} + \text{random noise}$$

We turn now to “how to model pure random noise?”

Call a stochastic process  $M = (M_t)$  a *martingale* with respect to the a filtration  $\{\mathcal{F}_t\}$  if

- I.  $E|M_t| < \infty$  for all  $t$
- II. for  $s < t$ ,  $E[M_t|\mathcal{F}_s] = M_s$

This says: “the best prediction of  $M$  at time  $t$  given the information available at some earlier time  $s$  is the value of  $M$  at time  $s$ ”.

Martingales are constant on average – they model fair games.

Martingales can be studied in either discrete time ( $t = 0, 1, 2, \dots, n, \dots$ ) or continuous times ( $t \geq 0$ ).

In either way there may be some time-horizon,  $T$  say, beyond which we need not look (*think of the expiry time of an option!!*).

Among martingales, we will concentrate on “*regular*” or “*uniformly integrable*” ones (nice convergence properties...). They are of the form:

$$M_t = E[X | \mathcal{F}_t]$$

for some random variable  $M$  with  $E|X| < \infty$ .

Note that  $M_t = E[X|\mathcal{F}_t]$  is a martingale as:

$$\begin{aligned} E[M_t|\mathcal{F}_s] &= E[E[X|\mathcal{F}_t]|\mathcal{F}_s] \\ &= E[X|\mathcal{F}_s] \\ &= M_s \end{aligned}$$

The first equality is an “*iterated conditional expectation*” (the inner one with respect to the information known at time  $t$ , the outer one w.r.t. the lesser amount of information at time  $s < t$ ). We use the tower property – the iterated conditional expectation is just the single conditional expectation w.r.t. the information at time  $s$ .

If  $t \mapsto M_t$  is continuous, we say that  $M$  is a continuous martingale (one example of a continuous martingale is... a Brownian Motion!!).

## 5. Quadratic Variation

Recall: for  $B$  a Brownian motion, we have:

$$EB_t = 0$$

$$\text{var}(B_t) = E[B_t^2] = t$$

**Question:** Why?

And one can show that:

$$\text{var}[B_t^2] = 2t^2$$

Note that for  $t$  small,  $2t^2$  is negligible w.r.t.  $t$ , *i.e.* the variance of  $B_t^2$  is negligible w.r.t  $t$ .

Recall that the variance measures how random the random variable is (here our RV is  $B_t^2$ ). This suggests that for  $t$  very small,  $B_t^2$  is nearly constant, equal to  $t$ .

Let's take a time  $t$  and increase it a little to  $t+dt$ , and write:

$$dB_t := B_{t+dt} - B_t$$

Using the argument above we get:

$$(dB_t)^2 = dt \quad (*)$$

at least to first order.

**Question:** Do you see why?

This can be made more precise in terms of *quadratic variation*.

Given an interval  $[0, t]$ , divide it by a sequence of partition points  $t_{n,i}$ :

$$0 = t_{n,0} < t_{n,1} < \dots < t_{n,k(n)} = t$$



The sum:

$$\sum_i |B(t_{n,i+1}) - B(t_{n,i})|^2$$

is called the quadratic variation of  $B$  over the  $n$ th partition  $\{t_{n,i}\}$ .

It turns out that if:

$$\max_i (t_{n,i+1} - t_{n,i}) \rightarrow 0 \quad (n \rightarrow \infty)$$

then:

$$\sum_i |B(t_{n,i+1}) - B(t_{n,i})|^2 \rightarrow t \quad a.s. \quad (**)$$

The quadratic variation of Brownian motion over  $[0,t]$  is  $t$ .

**Question:** use these statements for  $t+dt$  and  $t$  and show that we get (\*).

Recall that the Brownian path is continuous, *i.e.*:

$$\max_i |B(t_{n,i+1}) - B(t_{n,i})| \rightarrow 0 \quad (n \rightarrow \infty)$$

Since the sum of the squares tends to  $t$  - see (\*\*) – this suggests that:

$$\sum_i |B(t_{n,i+1}) - B(t_{n,i})|^2 \rightarrow t \quad a.s.$$

This is true!

To summarize:

Brownian motion has finite quadratic variation over  $[0, t]$  but infinite variation.

## 6. Itô integrals

Write:  $I_{(a,b]}(x) := 1 \quad (x \in (a, b]), \quad 0 \quad \text{otherwise}$

$I_{(a,b]}(x)$  is called the *indicator function* of the interval  $(a,b]$ .

Itô defined the following integral (*Itô integral*):

$$\boxed{\int I_{(a,b]} dB := B(b) - B(a)} \quad (***)$$

By a simple linearity argument we have:

$$\int \sum_i c_i I_{(a_i, b_i]} dB := \sum_i c_i [B(b_i) - B(a_i)]$$

Such functions - which are expressed as a linear combination of indicator functions - are called *simple functions*.

It turns out that continuous functions may suitably be approximated by simple functions.

Hence, the definition of the *Itô integral* above - see (\*\*\*) - may be extended to continuous functions.

In ordinary Calculus we usually neglect differential of the second or higher order.

By (\*) above -  $(dB_t)^2 = dt$  - we cannot neglect second order Brownian differentials!

What you need to remember:

Let  $Y$  be FV and  $B$  a Brownian motion, then:

$$(dB)^2 = dt, \quad (dY)^2 = 0, \quad dB.dY = 0$$

We will see later how we use this!

**Question:** if  $B_i$  and  $B_j$  are independent Brownian motions what can you say about

$$dB_i dB_j$$

## 7. Stochastic Differential Equations (SDEs)

Long pre-Itô, Bachelier's model for a stock price  $S = (S_t)$  can be summarised as:

$$dS = \mu dt + \sigma dB$$

This is a simple SDE, which we can solve on sight as:

$$S_t = S_0 + \boxed{\mu}t + \boxed{\sigma}\boxed{B_t}$$

drift

diffusion coefficient

Brownian motion

### Interpretation:

$\mu$  : gives the mean growth rate,

$\sigma$  : measures how random the stock price is.

However, this model is unrealistic. Why?

The stock price can go **negative**!! The stock price of a company is positive for as long as the company is solvent. If the company goes bankrupt, stock price hits 0 and stays there.

The stock is part ownership of a *limited liability company*: the meaning of limited liability is that the worst that can happen to an investor is that he only loses his investment if the company goes bankrupt.

Now consider the other SDE:

$$dS/S = \mu dt + \boxed{\sigma} dB \quad \text{(GBM)}$$

  
*volatility*

Interpretation:

$dS = S(t+dt) - S(t)$ : is the change in stock price over a short time-interval  $[t, t+dt]$ ,

$dS/S$ : is the change relative to the starting price – *i.e.* the return over  $[t, t+dt]$ .

**(GBM)** says that this return is the sum of two components:

- 1) a *systemic part*  $\mu dt$  with rate  $\mu$  , and
- 2) a *random part*  $\sigma dB$  , where  $\sigma$  measures how random the stock price is.

## 8. Itô's Lemma

Suppose we have a stochastic process  $f$ , which is a function of a Brownian motion  $B$  and time  $t$ , *i.e.*:  $f = f(t, B)$ .

As in ordinary calculus of several variable we can take the differential:

$$df = f_1 dt + f_2 dB + \frac{1}{2}[f_{11}(dt)^2 + 2f_{12}dt dB + f_{22}(dB)^2] + \dots \quad (1)$$

where:

- $f_1, f_2$  denote the first partial derivative of  $f$  w.r.t.  $t$  and  $B$ .
- $f_{11}$  the second derivative w.r.t.  $t$ ,  $f_{12}$  the second mixed partial derivative, etc...

Now,  $t$  is FV, so:

$$(dt)^2 = 0 \qquad dt dB = 0$$

And  $B$  has finite quadratic variation, recall:

$$(dB)^2 = dt$$

All higher differential are 0. Hence we can rewrite equation (1) above as:

$$df = (f_1 + \frac{1}{2}f_{22})dt + f_2dB$$

This is a simple of *Itô's Lemma*. It can be extended to functions of more than 2 variables, along the same lines.

- Integration by part

Consider the special case:  $f(x, y) = xy$ .

Then:  $f_1 = y$ ,  $f_2 = x$ ,  $f_{11} = 0$ ,  $f_{22} = 0$ ,  $f_{12} = 1$ .

Itô's Lemma above gives:

$$d(XY) = XdY + YdX + dX.dY.$$

This is the product rule for Itô's Lemma.

**Question:** How is this different to the “usual” product rule?

## 9. Geometric Brownian motion (GBM)

Recall the SDE:

$$dS/S = \mu dt + \sigma dB \quad (\text{GBM})$$

We can now solve it and the solution is:

$$S_t = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma B_t\right\} \quad (2)$$

We now use Itô's Lemma to check that this is the solution to our SDE. The R.H.S. of (2) is  $f(t, B)$  with:

$$f(t, x) := S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma x\right\}$$

So:  $f_1 = (\mu - \frac{1}{2}\sigma^2)f, \quad f_2 = \sigma f, \quad f_{22} = \sigma^2 f$

Hence, using Itô's Lemma, we have:

$$\begin{aligned} df &= (f_1 + \frac{1}{2}f_{22})dt + f_2dB \\ &= (\mu - \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2)f dt + \sigma f dB \\ &= f(\mu dt + \sigma dB) \end{aligned}$$

And so  $f$  satisfies the SDE (GBM) with the initial condition:  $f(0, B) = S_0$

This process is called a Geometric Brownian motion (GBM). It is the model of stock prices used in the *Black Sholes theory* (see Chapter 4).

## 10. Change of Probability measure (Girsanov's theorem)

Randomness is governed by probability: the function that takes an event  $A$  (a set whose probability is defined) to its probability  $P(A)$  is called a *probability measure*.

Sometimes we need to compare two probability measures,  $P$  and  $Q$  say. We call  $P$  and  $Q$  equivalent if the same events are possible and impossible for both:

$$P(A) = 0 \text{ iff } Q(A) = 0$$

$$P(A) > 0 \text{ iff } Q(A) > 0$$

In Ch. 4 we will need to *change the measure*; only a change to equivalent measure is useful. Change to a suitable equivalent measure is the key technique to mathematical finance.

**Example:** insurance company (more weights to large losses).

## 11. Random walks, trees

Main reason for using Brownian motion: treat time as continuous and we can use some powerful mathematical tools.

We also need to treat time as discrete (*i.e.* looking at a sequence of opening or closing prices, high frequency data sampled every 10 seconds say).

We need a discrete analogue of the Brownian motion: this is a *random walk*.

Take a sequence of independent tosses of a coin, bet +\$1 on heads and keep track of the total gain (or loss!!) to date at time  $n$ . Call this:

$$S_n := \sum_1^n X_i$$

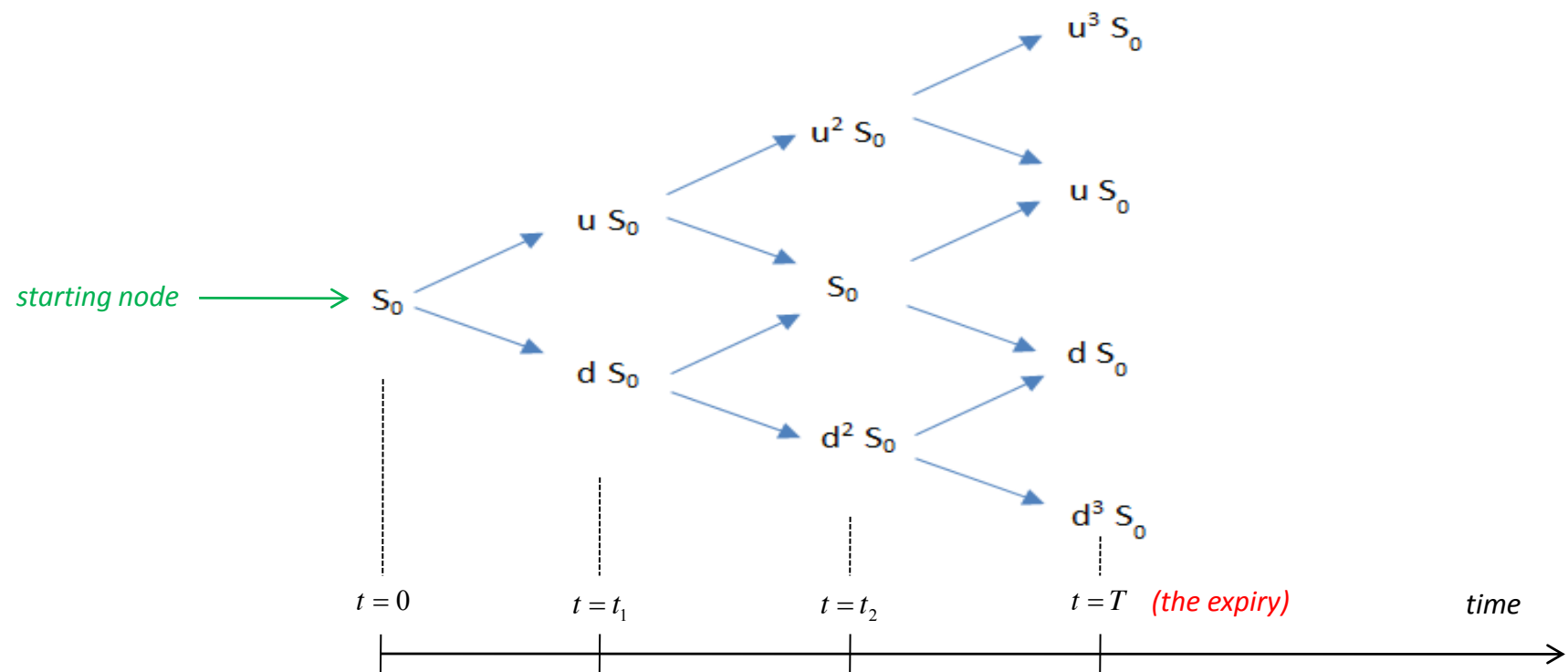
This process  $S = (S_n)$  is a *random walk*, the discrete analogue of Brownian motion.

Note: this is written additively but may need to work multiplicatively instead.

Simplest case: stock price which in each time interval may go up or down. Represent the price evolution by a *recombining tree*.

## Example of a recombining tree:

Note that this is a special example where  $u.d = 1$  (it doesn't have to!).



## How do we use such trees:

When dealing with option with expiry  $T$ :

- i. fill in the value of the stock at each terminal node (to do this you may need to compute the value of the stock at all of the previous nodes),
- ii. fill in the value of the option at each terminal node,
- iii. recur backward in time (taking into consideration the appropriate discounting), eventually filling in the value at the starting node - this gives the value of the option when it is initially bought or sold.

This technique is called *backward induction* or *dynamic programming*.

## 12. Poisson process

Remember the Poisson distribution?

Recall: BM is the dynamic counterpart of the Normal distribution.

The Poisson process is the dynamic counterpart of the Poisson distribution. We can build this as follows.

Take a sequence of independent and identically distributed (i.i.d.) random variables  $X_i$ , with distribution the exponential distribution with parameter  $\lambda > 0$ , hence density is:

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0$$

Think of the  $X_i$  as the lifetimes of light bulbs, replaced on failure.

Write  $N_t$  for the number of failures by time  $t$ . Then:

$$N = (N_t)$$

is called the *Poisson point process* with rate/intensity  $\lambda > 0$ , write  $Ppp(\lambda)$ .

It has independent increments (from the lack-of-memory property of the exponential distribution - remember?).

**It counts the number of jumps to date.**

Since at time  $t$ , there is jump 0 (*i.e.* no jump) or 1 (*i.e.* 1 jump), we may write:

$$(dN_t)^2 = dN_t$$

this is the counterpart of  $(dB)^2 = dt$  for Brownian motion!!

Hence one can extend Itô calculus to the Poisson process, and more generally to the *compound Poisson process*, where the jumps at the points of a Poisson process are independent and random with distribution  $F$  say.

Used in insurance a lot to model the total claims to date.

Brownian motion and compound Poisson processes are examples of ***Lévy processes***.

Lévy processes can have infinitely many jumps in finite time – useful for modelling “*jitter*” in stock price movement.