Mathematical Preliminaries

Introduction to Probability

Preliminaries

Randomness lies at the heart of finance and whether terms uncertainty or risk are used, they refer to the random nature of the financial markets. Probability theory provides the necessary structure to model the uncertainty that is central to finance. We begin by defining some basic mathematical tools.

The set Ω of all possible outcomes of some given experiment is called the *sample space*. A particular outcome $\omega \in \Omega$ is called a *sample point*.

An *event* Ψ is a set of outcomes, i.e. $\Psi \subset \Omega$.

To a set of basic outcomes ω_i we assign real numbers called probabilities, written $\mathbb{P}(\omega_i) = p_i$. Then for any event E,

$$\mathbb{P}(E) = \sum_{\omega_i \in E} p_i$$

Example 1

Experiment: A dice is rolled and the number appearing on top is observed. The sample space consists of the 6 possible numbers:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

If the number 4 appears then $\omega = 4$ is a sample point, clearly $4 \in \Omega$.

Let Ψ_1, Ψ_2, Ψ_3 = events that an even, odd, prime number occurs respectively.

So

$$\Psi_1 = \{2,\ 4,\ 6\}, \ \Psi_2 = \{1,\ 3,\ 5\}, \ \ \Psi_3 = \{2,\ 3,\ 5\}$$

$\Psi_1 \cup \Psi_3 = \{\ 2,\ 3,\ 4,\ 5,\ 6\}$ – event that an even <u>or</u> prime number occurs.

$\Psi_2 \cap \Psi_3 = \{3,\ 5\}$ – event that odd <u>and</u> prime number occurs.

$\Psi_3^c = \{1,\ 4,\ 6\}$ – event that prime number does not occur (complement of event).

Example 2

Toss a coin twice and observe the sequence of heads (H) and tails (T) that appears. Sample space

$$\Omega = \{\text{HH, TT, HT, TH}\}$$

Let Ψ_1 be event that at least one head appears, and Ψ_2 be event that both tosses are the same:

$$\Psi_1 = \{\text{HH, HT, TH}\}, \ \Psi_2 = \{\text{HH, TT}\}$$

$\Psi_1 \cap \Psi_2 = \{\text{HH}\}$

Events are subsets of Ω, but not all subsets of Ω are events.

The basic properties of probabilities are

1. $0 \le p_i \le 1$
2. $\mathbb{P}(\Omega) = \sum_i p_i = 1$ (the sum of the probabilities is always 1).

Random Variables

Outcomes of experiments are not always numbers, e.g. 2 heads appearing; picking an ace from a deck of cards. We need some way of assigning real numbers to each random event. Random variables assign numbers to events.

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

$$X : \omega \in \Omega \to \mathbb{R},$$

i.e. it associates a number $X(\omega)$ 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 $\mathbb{P}(\mathrm{H}) = \mathbb{P}(\mathrm{T}) = \frac{1}{2}$. So now we can assign the following outcomes

$$\begin{aligned} \mathbb{P}(1) &= \frac{1}{2} \\ \mathbb{P}(-1) &= \frac{1}{2} \end{aligned}$$

Mathematically, if our random variable is X, then

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

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

The probability that the RV takes on each possible value is called the *probability distribution.*

If X is a RV then

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

is the probability that a occurs (or X maps onto a).

$\mathbb{P}(a \le X \le b) =$ probability that X lies in the interval $[a, b] =$

$$\mathbb{P}(\{\omega \in \Omega : a \le X(\omega) \le b\})$$

$$X : \underset{\text{Domain}}{\Omega} \longrightarrow \underset{\text{Range (finite)}}{\mathbb{R}}$$

$X(\Omega) = \{x_1, ..., x_n\} = \{x_i\}_{1 \le i \le n}$

$\mathbb{P}[x_i] = \mathbb{P}[X = x_i] = f(x_i) \ \forall\, i.$

So the earlier coin tossing example gives

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

$$f(x_i) \quad \text{is the probability distribution of } X.$$

This is called a *discrete probability distribution.*

x_i	x_1	x_2	....	x_n
$f(x_i)$	$f(x_1)$	$f(x_2)$	....	$f(x_n)$

There are two properties of the distribution $f\ (x_i)$

(i) $f(x_i) \geq 0 \ \ \forall \ i \in [1, \ n]$

(ii) $\sum_{i=1}^{n} f(x_i) = 1,$ i.e. sum of all probabilities is one.

Mean/Expectation

The *mean* μ measures the centre (average) of the distribution

$$\begin{aligned} \mu &= \mathbb{E}[X] = \sum_{i=1}^{n} x_i \ f\ (x_i) \\ &= x_1 f(x_1) + x_2 f(x_2) + \ldots.. + x_n f(x_n) \end{aligned}$$

which is equal to the weighted average of all possible values of X together with associated probabilities.

This is also called the *first moment.*

Example:

x_i	2	3	8
$f(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$$\begin{aligned} \mu &= \mathbb{E}[X] = \sum_{i=1}^{3} x_i f(x_i) = 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{2}\right) + 8\left(\frac{1}{4}\right) \\ &= 4 \end{aligned}$$

Variance/Standard Deviation

This measures the spread (dispersion) of X about the mean.

Variance $\mathbb{V}[X] =$

$$\mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}\left[X^2\right] - \mu^2 = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2 = \sigma^2$$

$\mathbb{E}\left[(X-\mu)^2\right]$ is also called the *second moment about the mean.*

From the previous example we have $\mu = 4,$ therefore

$$\begin{aligned} \mathbb{V}[X] &= \left(2^2\left(\frac{1}{4}\right) + 3^2\left(\frac{1}{2}\right) + 8^2\left(\frac{1}{4}\right)\right) - 16 \\ &= 5.5 = \sigma^2 \rightarrow \sigma = 2.34 \end{aligned}$$

Rules for Manipulating Expectations

Suppose $X, \ Y$ are random variables and $\alpha, \ \beta, \ \lambda \in \mathbb{R}$ are constant scalar quantities. Then

- $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, (linearity)
- $\mathbb{V}[\alpha X + \beta] = \alpha^2 \mathbb{V}[X]$
- $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$,
- $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$

The last two are provided $X,\ Y$ are independent.

Continuous Random Variables

As the number of discrete events becomes very large, individual probabilities $f(x_i) \to 0$. Now look at the continuous case.

Instead of $f(x_i)$ we now have $p(x)$ which is a continuous distribution called as *probability density function, PDF.*

$$P(a \le X \le b) = \int_a^b p(x)\,dx$$

The *cumulative distribution function* $F(x)$ of a RV X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} p(x)\,dx$$

$F(x)$ is related to the PDF by

$$p(x) = \frac{dF}{dx}$$

(fundamental theorem of calculus) provided $F(x)$ is differentiable. However unlike $F(x)$, $p(x)$ may have singularities (and may be unbounded).

Special Expectations:

Given any PDF $p(x)$ of X.

Mean $\mu = \mathbb{E}[X] = \int_{\mathbb{R}} xp(x)\,dx.$

Variance $\sigma^2 = \mathbb{V}[X] = \mathbb{E}\left[(X-\mu)^2\right] = \int_{\mathbb{R}} x^2 p(x)\,dx - \mu^2$

(2^{nd} moment about the mean).

The n^{th} moment about zero is defined as

$$\begin{aligned} \mu_n &= \mathbb{E}[X^n] \\ &= \int_{\mathbb{R}} x^n p(x)\,dx. \end{aligned}$$

In general, for any function h

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)\,p(x)\,dx.$$

where X is a RV following the distribution given by $p(x)$.

Moments about the mean are given by

$$\mathbb{E}[(X-\mu)^n]\,;\ n = 2,\ 3,\ \ldots$$

The special case $n = 2$ gives the variance σ^2.

Skewness and Kurtosis

Having looked at the variance as being the second moment about the mean, we now discuss two further moments centred about μ, that provide further important information about the probability distribution.

Skewness is a measure of the asymmetry of a distribution (i.e. lack of symmetry) about its mean. A distribution that is identical to the left and right about a centre point is symmetric.

The third central moment, i.e. third moment about the mean scaled with σ^3. This scaling allows us to compare with other distributions.

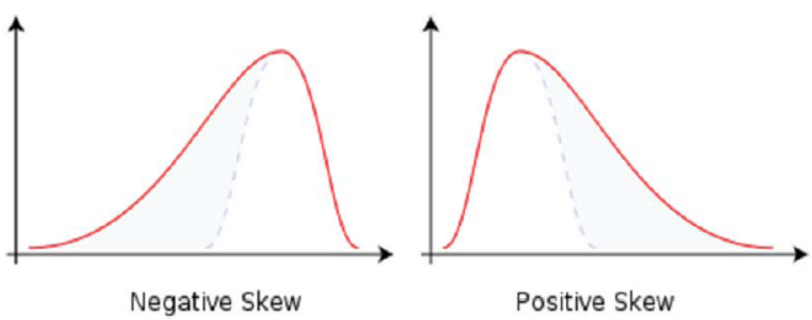
$$\frac{\mathbb{E}\left[(X-\mu)^3\right]}{\sigma^3}$$

is called the *skew* and is a measure of the skewness (a non-symmetric distribution is called *skewed*).

Any distribution which is symmetric about the mean has a skew of zero.

Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right.

By skewed left, we mean that the left tail is long relative to the right tail. Similarly, skewed right means that the right tail is long relative to the left tail.



The fourth centred moment scaled by the square of the variance, called the *kurtosis* is defined

$$\frac{\mathbb{E}\left[(X-\mu)^4\right]}{\sigma^4}.$$

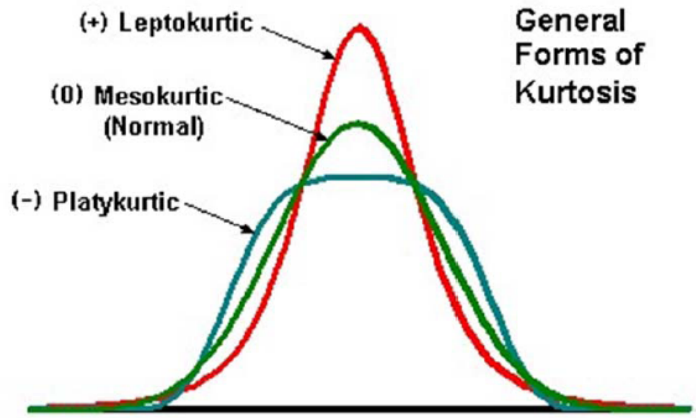
This is a measure of how much of the distribution is out in the tails at large negative and positive values of X.

The 4^{th} central moment is called Kurtosis and is defined as

$$\text{Kurtosis} = \frac{\mathbb{E}\left[(X-\mu)^4\right]}{\sigma^4}$$

normal random variable has Kurtosis of 3 irrespective of its mean and standard deviation. Often when comparing a distribution to the normal distribution, the measure of excess Kurtosis is used, i.e. Kurtosis of distribution -3.

If a random variable has Kurtosis greater than 3 is called Leptokurtic, if is has Kurtosis less than 3 it is called platykurtic Leptokurtic is associated with PDF's that are simultaneously peaked and have fat tails.



Normal Distribution

The *normal* (or *Gaussian*) distribution $N(\mu, \sigma^2)$ with mean and standard deviation μ and σ^2 in turn is defined in terms of its density function

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

For the special case $\mu = 0$ and $\sigma = 1$ it is called the *standard normal* distribution $N(0,\ 1)$.

This is also verified by making the substitution

$$\phi = \frac{x-\mu}{\sigma}$$

in $p(x)$ which gives

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

and clearly has zero mean and unit variance:

$$\mathbb{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\mathbb{E}[X-\mu] = 0,$$

$$\mathbb{V}\left[\frac{X-\mu}{\sigma}\right] = \mathbb{V}\left[\frac{X}{\sigma} - \frac{\mu}{\sigma}\right]$$

Now $\mathbb{V}[\alpha X + \beta] = \alpha^2 \mathbb{V}[X]$ (standard result), hence

$$\frac{1}{\sigma^2}\mathbb{V}[X] = \frac{1}{\sigma^2}.\sigma^2 = 1$$

Its cumulative distribution function is

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\phi^2} d\phi = P(-\infty \leq X \leq x).$$

The skewness of $N(0,\ 1)$ is zero and its kurtosis is 3.

Correlation

The covariance is useful in studying the statistical dependence between two random variables. If $X,\ Y$ are RV's, then their covariance is defined as:

$$\begin{aligned}\mathrm{Cov}\left(X,\ Y\right) &= \mathbb{E}\left[\left(X-\underbrace{\mathbb{E}\left(X\right)}_{=\mu_x}\right)\left(Y-\underbrace{\mathbb{E}\left(Y\right)}_{=\mu_y}\right)\right] \\ &= \mathbb{E}\left[XY\right]-\mu_x\mu_y\end{aligned}$$

which we denote as σ_{XY}. **Note:**

$$\mathrm{Cov}\left(X,\ X\right)=\mathbb{E}\left[\left(X-\mu_x\right)^2\right]=\sigma^2.$$

$X,\ Y$ are *correlated* if

$$\mathbb{E}\left[\left(X-\mu_x\right)\left(Y-\mu_y\right)\right]\neq 0.$$

We can then define an important dimensionless quantity (used in finance) called the *correlation coefficient* and denoted as $\rho_{XY}\left(X,\ Y\right)$ where

$$\rho_{XY}=\frac{\mathrm{Cov}\left(X,\ Y\right)}{\sigma_x\sigma_y}.$$

The correlation can be thought of as a normalised covariance, as $\left|\rho_{XY}\right|\leq 1$, for which the following conditions are properties:

i. $\rho\left(X,\ Y\right)=\rho\left(Y,\ X\right)$

ii. $\rho\left(X,\ \pm X\right)=\pm 1$

iii. $-1\leq\rho\leq 1$

$\rho_{XY}=-1\Rightarrow$ perfect negative correlation

$\rho_{XY}=1\Rightarrow$perfect correlation

$\rho_{XY}=0\Rightarrow X,\ Y$ uncorrelated

Why is the correlation coefficient bounded by ± 1? Justification of this requires a result called the *Cauchy-Schwartz inequality*. This is a theorem which most students encounter for the first time in linear algebra (although we have not discussed this). Let's start off with the version for random variables (RVs) X and Y, then the Cauchy-Schwartz inequality is

$$\left[\mathbb{E}\left[XY\right]\right]^2\leq\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right].$$

We know that the covariance of X,Y is

$$\sigma_{XY}=\mathbb{E}\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]$$

If we put

$$\begin{aligned}\mathbb{V}\left[X\right] &= \sigma_X^2=\mathbb{E}\left[\left(X-\mu_X\right)^2\right] \\ \mathbb{V}\left[Y\right] &= \sigma_Y^2=\mathbb{E}\left[\left(Y-\mu_Y\right)^2\right].\end{aligned}$$

From Cauchy-Schwartz we have

$$\left(\mathbb{E}\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]\right)^2 \leq \mathbb{E}\left[\left(X-\mu_X\right)^2\right] \mathbb{E}\left[\left(Y-\mu_Y\right)^2\right]$$

or we can write

$$\sigma_{XY}^2 \leq \sigma_X^2 \sigma_Y^2$$

Divide through by $\sigma_X^2 \sigma_Y^2$

$$\frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} \leq 1$$

and we know that the left hand side above is ρ_{XY}^2, hence

$$\rho_{XY}^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} \leq 1$$

and since ρ_{XY} is a real number, this implies $|\rho_{XY}| \leq 1$ which is the same as

$$-1 \leq \rho_{XY} \leq +1.$$

Central Limit Theorem

This concept is fundamental to the whole subject of finance.

Let X_i be any independent identically distributed (i.i.d) random variable with mean μ and variance σ^2, i.e. $X \sim D\left(\mu,\ \sigma^2\right)$, where D is some distribution. If we put

$$S_n = \sum_{i=1}^{n} X_i$$

Then $\dfrac{\left(S_n - n\mu\right)}{\sigma\sqrt{n}}$ has a distribution that approaches the standard normal distribution as $n \to \infty$.

The distribution of the sum of a large number of independent identically distributed variables will be approximately normal, regardless of the underlying distribution. That is the beauty of this result.

Conditions:

The Normal distribution is the limiting behaviour if you add many random numbers from any basic-building block distribution provided the following is satisfied:

1. Mean of distribution must be finite and constant

2. Standard deviation of distribution must be finite and constant

This is a measure of how much of the distribution is out in the tails at large negative and positive values of X.

Moment Generating Function

The *moment generating function* of X, denoted $M_X(\theta)$ is given by

$$M_X(\theta) = \mathbb{E}\left[e^{\theta X}\right] = \int_{\mathbb{R}} e^{\theta x} p(x)\, dx$$

provided the expectation exists. We can expand as a power series to obtain

$$M_X(\theta) = \sum_{n=0}^{\infty} \frac{\theta^n \mathbb{E}(X^n)}{n!}$$

so the n^{th} moment is the coefficient of $\theta^n/n!$, or the n^{th} derivative evaluated at zero.

How do we arrive at this result?

We use the Taylor series expansion for the exponential function: $\int_{\mathbb{R}} e^{\theta x} p(x)\, dx =$

$$\int_{\mathbb{R}} \left(1 + \theta x + \frac{(\theta x)^2}{2!} + \frac{(\theta x)^3}{3!} + ...\right) p(x)\, dx$$

$$\begin{aligned}
&= \underbrace{\int_{\mathbb{R}} p(x)\, dx}_{1} + \theta \underbrace{\int_{\mathbb{R}} x p(x)\, dx}_{\mathbb{E}(X)} + \tfrac{\theta^2}{2!} \underbrace{\int_{\mathbb{R}} x^2 p(x)\, dx}_{\mathbb{E}(X^2)} + \\
&\quad \tfrac{\theta^3}{3!} \underbrace{\int_{\mathbb{R}} x^3 p(x)\, dx}_{\mathbb{E}(X^3)} + ... \\
&= 1 + \theta \mathbb{E}(X) + \tfrac{\theta^2}{2!} \mathbb{E}\left(X^2\right) + \tfrac{\theta^3}{3!} \mathbb{E}\left(X^3\right) + ... \\
&= \sum_{n=0}^{\infty} \frac{\theta^n \mathbb{E}(X^n)}{n!}.
\end{aligned}$$

Calculating Moments

The k^{th} moment m_k of the random variable X can now be obtained by differentiating, i.e.

$$\begin{aligned} m_k &= M_X^{(k)}(\theta)\,; \quad k = 0, 1, 2, ... \\ M_X^{(k)}(\theta) &= \left.\frac{d^k}{d\theta^k} M_X(\theta)\right|_{\theta=0} \end{aligned}$$

So what is this result saying? Consider $M_X(\theta) = \sum_{n=0}^{\infty} \frac{\theta^n \mathbb{E}(X^n)}{n!}$

$$M_X(\theta) = 1 + \theta \mathbb{E}[X] + \frac{\theta^2}{2!}\mathbb{E}\left[X^2\right] + \frac{\theta^3}{3!}\mathbb{E}\left[X^3\right] + ... + \frac{\theta^n}{n!}\mathbb{E}[X^n]$$

As an example suppose we wish to obtain the second moment; differentiate twice with respect to θ

$$\frac{d}{d\theta} M_X(\theta) = \mathbb{E}[X] + \theta \mathbb{E}\left[X^2\right] + \frac{\theta^2}{2}\mathbb{E}\left[X^3\right] + ... + \frac{\theta^{n-1}}{(n-1)!}\mathbb{E}[X^n]$$

and for the second time

$$\frac{d^2}{d\theta^2} M_X(\theta) = \mathbb{E}\left[X^2\right] + \theta \mathbb{E}\left[X^3\right] + ... + \frac{\theta^{n-2}}{(n-2)!}\mathbb{E}[X^n].$$

Setting $\theta = 0$, gives

$$\frac{d^2}{d\theta^2} M_X(0) = \mathbb{E}\left[X^2\right]$$

which captures the second moment $\mathbb{E}[X^2]$. Remember we will already have an expression for $M_X(\theta)$.

A useful result in finance is the MGF for the normal distribution. If $X \sim N(\mu, \sigma^2)$, then we can construct a standard normal $\phi \sim N(0, 1)$ by setting $\phi = \dfrac{X - \mu}{\sigma} \implies X = \mu + \sigma\phi$.

The MGF is

$$\begin{aligned} M_X(\theta) &= \mathbb{E}\left[e^{\theta x}\right] = \mathbb{E}\left[e^{\theta(\mu + \phi\sigma)}\right] \\ &= e^{\theta\mu}\mathbb{E}\left[e^{\theta\sigma\phi}\right] \end{aligned}$$

So the MGF of X is therefore equal to the MGF of ϕ but with θ replaced by $\theta\sigma$.This is much nicer than trying to calculate the MGF of $X \sim N(\mu, \sigma^2)$.

$$\begin{aligned} \mathbb{E}\left[e^{\theta\phi}\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\theta x} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\theta x - x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(x^2 - 2\theta x + \theta^2 - \theta^2\right)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2 + \frac{1}{2}\theta^2} dx \\ &= e^{\frac{1}{2}\theta^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} dx \end{aligned}$$

Now do a change of variable - put $u = x - \theta$

$$\begin{aligned} \mathbb{E}\left[e^{\theta\phi}\right] &= e^{\frac{1}{2}\theta^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du \\ &= e^{\frac{1}{2}\theta^2} \end{aligned}$$

Thus

$$\begin{aligned} M_X(\theta) &= e^{\theta\mu}\mathbb{E}\left[e^{\theta\sigma\phi}\right] \\ &= e^{\theta\mu+\frac{1}{2}\theta^2\sigma^2} \end{aligned}$$

To get the simpler formula for a standard normal distribution put $\mu = 0,\ \sigma = 1$ to get $M_X(\theta) = e^{\frac{1}{2}\theta^2}$.

We can now obtain the first four moments for a standard normal

$$\begin{aligned} m_1 &= \left.\frac{d}{d\theta}e^{\frac{1}{2}\theta^2}\right|_{\theta=0} \\ &= \left.\theta e^{\frac{1}{2}\theta^2}\right|_{\theta=0} = 0 \end{aligned}$$

$$\begin{aligned} m_2 &= \left.\frac{d^2}{d\theta^2}e^{\frac{1}{2}\theta^2}\right|_{\theta=0} \\ &= \left.\left(\theta^2+1\right)e^{\frac{1}{2}\theta^2}\right|_{\theta=0} = 1 \end{aligned}$$

$$\begin{aligned} m_3 &= \left.\frac{d^3}{d\theta^3}e^{\frac{1}{2}\theta^2}\right|_{\theta=0} \\ &= \left.\left(\theta^3+3\theta\right)e^{\frac{1}{2}\theta^2}\right|_{\theta=0} = 0 \end{aligned}$$

$$\begin{aligned} m_4 &= \left.\frac{d^4}{d\theta^4}e^{\frac{1}{2}\theta^2}\right|_{\theta=0} \\ &= \left.\left(\theta^4+6\theta^2+3\right)e^{\frac{1}{2}\theta^2}\right|_{\theta=0} = 3 \end{aligned}$$

The latter two are particularly useful in calculating the skew and kurtosis.

If X and Y are independent random variables then

$$\begin{aligned} M_{X+Y}(\theta) &= \mathbb{E}\left[e^{\theta(x+y)}\right] \\ &= \mathbb{E}\left[e^{\theta x}e^{\theta y}\right] = \mathbb{E}\left[e^{\theta x}\right]\mathbb{E}\left[e^{\theta y}\right] \\ &= M_X(\theta)\,M_Y(\theta)\,. \end{aligned}$$

Calculus Refresher

Taylor for two Variables

Assuming that a function $f(x,t)$ is differentiable enough, near $x = x_0$, $t = t_0$,

$$\begin{aligned} f(x,t) &= f(x_0,t_0) + (x - x_0) f_x(x_0,t_0) + \\ &\quad (t - t_0) f_t(x_0,t_0) \\ &\quad + \frac{1}{2} \begin{bmatrix} (x-x_0)^2 f_{xx}(x_0,t_0) \\ +2(x-x_0)(t-t_0) f_{xt}(x_0,t_0) \\ +(t-t_0)^2 f_{tt}(x_0,t_0) \end{bmatrix} + \ .... \end{aligned}$$

That is,

$$\begin{aligned} f\ (x,t) &= \text{constant} + \ \text{linear} + \text{quadratic} \\ &\quad +.... \end{aligned}$$

The error in truncating this series after the second order terms tends to zero faster than the included terms. This result is particularly important for Itô's lemma in Stochastic Calculus.

Suppose a function $f = f(x,y)$ and both x, y change by a small amount, so $x \longrightarrow x + \delta x$ and $y \longrightarrow y + \delta y$, then we can examine the change in f using a two dimensional form of Taylor

$$\begin{aligned} f(x+\delta x, y+\delta y) &= f(x,y) + f_x \delta x + f_y \delta y + \\ &\quad \frac{1}{2} f_{xx} \delta x^2 + \frac{1}{2} f_{yy} \delta y^2 + \\ &\quad f_{xy} \delta x \delta y + O\left(\delta x^2, \delta y^2\right). \end{aligned}$$

By taking $f(x,y)$ to the lhs, writing

$$df = f(x+\delta x, y+\delta y) - f(x,y)$$

and considering only linear terms, i.e.

$$df = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y$$

we obtain a formula for the *differential* or *total change* in f.

Integration

There are two ways to show the following important result

$$\int_{\mathbb{R}} e^{-x^2} = \sqrt{\pi}.$$

The first can be thought of as the 'poor man's' derivation.

The **CDF** for the Normal Distribution is

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2} ds$$

If $x \longrightarrow \infty$ then we know (by the fact that the area under a PDF has to sum to unity) that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-s^2/2} ds = 1.$$

Make the substitution $x = s/\sqrt{2}$ to give $dx = ds/\sqrt{2}$, hence the integral becomes

$$\sqrt{2} \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{2\pi}$$

and hence we obtain

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

From this we also note that $\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$ because e^{-x^2} is an even function.

The second requires double integration. Put $I = \int_{\mathbb{R}} e^{-x^2} dx$ so that

$$\begin{aligned} I^2 &= \int_{\mathbb{R}} e^{-x^2} dx \int_{\mathbb{R}} e^{-y^2} dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\left(x^2+y^2\right)} dx dy \end{aligned}$$

The region of integration is a square centered at the origin of infinite dimension

$$\begin{aligned} x &\in (-\infty, \infty) \\ y &\in (-\infty, \infty) \end{aligned}$$

i.e. the complete 2D plane. Introduce plane polars

$$\left.\begin{aligned} x = r\cos\theta \\ y = r\sin\theta \end{aligned}\right\} dxdy \rightarrow rdrd\theta$$

The region of integration is now a circle centred at the origin of infinite radius

$$\begin{aligned} 0 &\leq r < \infty \\ 0 &\leq \theta \leq 2\pi \end{aligned}$$

so the problem becomes

$$\begin{aligned} I^2 &= \int_0^{2\pi} \int_0^\infty e^{-r^2} r dr d\theta \\ &= \frac{1}{2} \int_0^{2\pi} d\theta = \pi \end{aligned}$$

Hence

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

Review of Differential Equations

Cauchy Euler Equation

An equation of the form

$$Ly = ax^2\frac{d^2y}{dx^2} + \beta x\frac{dy}{dx} + cy = g(x)$$

is called a Cauchy-Euler equation.

To solve the homogeneous part, we look for a solution of the form

$$y = x^\lambda$$

So $y' = \lambda x^{\lambda-1} \rightarrow y'' = \lambda(\lambda-1)x^{\lambda-2}$, which upon substitution yields the quadratic, A.E.

$$a\lambda^2 + b\lambda + c = 0,$$

where $b = (\beta - a)$ which can be solved in the usual way - there are 3 cases to consider, depending upon the nature of $b^2 - 4ac$.

Case 1: $b^2 - 4ac > 0 \rightarrow \lambda_1,\ \lambda_2 \in \mathbb{R}$ - 2 real distinct roots

$$\text{GS } y = Ax^{\lambda_1} + Bx^{\lambda_2}$$

Case 2: $b^2 - 4ac = 0 \rightarrow \lambda = \lambda_1 = \lambda_2 \in \mathbb{R}$ - 1 real (double fold) root

$$\text{GS } y = x^\lambda(A + B\ln x)$$

Case 3: $b^2 - 4ac < 0 \rightarrow \lambda = \alpha \pm i\beta \in \mathbb{C}$ - pair of complex conjugate roots

$$\text{GS } y = x^\alpha(A\cos(\beta\ln x) + B\sin(\beta\ln x))$$

Example

Consider the following **<u>Euler</u>** type problem

$$\frac{1}{2}\sigma^2S^2\frac{d^2V}{dS^2} + rS\frac{dV}{dS} - rV = 0,$$
$$V(0) = 0, \quad V(S^*) = S^* - E$$

where the constants $E, S^*, \sigma, r > 0$. We are given that the roots of A.E $m_\pm$ are real with $m_- < 0 < m_+$.

Look for a solution of the form General Solution is

$$V(S) = AS^{m_+} + BS^{m_-}.$$

$$V(0) = 0 \Longrightarrow B = 0 \text{ else we have division by zero}$$
$$V(S) = AS^{m_+}$$

To find A use the second condition $V(S^*) = S^* - E$

$$V(S^*) = A(S^*)^{m_+} = S^* - E \longrightarrow A = \frac{S^* - E}{(S^*)^{m_+}}$$

hence

$$V(S) = \frac{S^* - E}{(S^*)^{m_+}}(S)^{m_+} = (S^* - E)\left(\frac{S}{S^*}\right)^{m_+}.$$

Similarity Methods

$f(x,y)$ is **homogeneous of degree** $t \geq 0$ if $f(\lambda x, \lambda y) = \lambda^t f(x,y)$.

1. $f(x,y) = \sqrt{(x^2+y^2)}$

$$f(\lambda x, \lambda y) = \sqrt{\left[(\lambda x)^2 + (\lambda y)^2\right]} = \lambda\sqrt{[(x^2+y^2)]} = \lambda f(x,y)$$

   $g(x,y) = \frac{x+y}{x-y}$ then

$$g(\lambda x, \lambda y) = \tfrac{\lambda x + \lambda y}{\lambda x - \lambda y} = \lambda^0\left(\tfrac{x+y}{x-y}\right) = \lambda^0 g(x,y)$$

2. $h(x,y) = x^2 + y^3$

$$h(\lambda x, \lambda y) = (\lambda x)^2 + (\lambda y)^3 = \lambda^2 x^2 + \lambda^3 y^3 \neq \lambda^t\left(x^2 + y^3\right)$$

   for any t. So h is not homogeneous.

Consider the function

$$F(x,y) = \tfrac{x^2}{x^2+y^2}$$

If for any $\lambda > 0$ we write

$$x' = \lambda x, \;\; y' = \lambda y$$

then

$$\frac{dy'}{dx'} = \frac{dy}{dx}, \;\; \tfrac{x^2}{x^2+y^2} = \tfrac{x'^2}{x'^2+y'^2}.$$

We see that the equation is *invariant* under the change of variables. It also makes sense to look for a solution which is also invariant under the transformation. One choice is to write

$$v = \frac{y}{x} = \frac{y'}{x'}$$

so write

$$y = vx.$$

Definition The differential equation $\dfrac{dy}{dx} = f(x,y)$ is said to be *homogeneous* when $f(x,y)$ is homogeneous of degree t for some t.

Method of Solution

Put $y = vx$ where v is some (as yet) unknown function. Hence we have

$$\begin{aligned}\frac{dy}{dx} &= \frac{d}{dx}(vx) = x\frac{dv}{dx} + v\frac{dx}{dx} \\ &= v'x + v\end{aligned}$$

Hence

$$f(x,y) = f(x,vx)$$

Now f is homogeneous of degree t - so

$$f(x,vx) = x^t f(1,v)$$

The differential equation now becomes

$$v'x + v = x^t f(1,v)$$

which is not always solvable - the method may not work. But when $t = 0$ (homogeneous of degree zero) then $x^t = 1$. Hence

$$v'x + v = f(1,v)$$

or

$$x\frac{dv}{dx} = f(1,v) - v$$

which is separable, i.e.

$$\int \frac{dv}{f(1,v) - v} = \int \frac{dx}{x} + c$$

and the method is guaranteed to work.

Example

$$\frac{dy}{dx} = \frac{y-x}{y+x}$$

First we check:

$$\frac{\lambda y - \lambda x}{\lambda y + \lambda x} = \lambda^0 \left(\frac{y-x}{y+x}\right)$$

which is homogeneous of degree zero. So put $y = vx$

$$v'x + v = f(x,yx) = \frac{vx - x}{vx + x} = \frac{v-1}{v+1} = f(1,v)$$

therefore

$$\begin{aligned} v'x &= \frac{v-1}{v+1} - v \\ &= \frac{-\left(1+v^2\right)}{v+1} \end{aligned}$$

and the D.E is now separable

$$\begin{aligned} \int \frac{v+1}{v^2+1} dv &= -\int \frac{1}{x} dx \\ \int \frac{v}{v^2+1} dv + \int \frac{1}{v^2+1} dv &= -\int \frac{1}{x} dx \\ \frac{1}{2}\ln\left(1+v^2\right) + \arctan v &= -\ln x + c \\ \frac{1}{2}\ln x^2\left(1+v^2\right) + \arctan v &= c \end{aligned}$$

Now we turn to the original problem, so put $v = \dfrac{y}{x}$

$$\frac{1}{2}\ln x^2\left(1+\frac{y^2}{x^2}\right) + \arctan\left(\frac{y}{x}\right) = c$$

which simplifies to

$$\frac{1}{2}\ln\left(x^2+y^2\right) + \arctan\left(\frac{y}{x}\right) = c.$$

The Error Function

We begin by solving the following initial value problem (IVP)

$$\frac{dy}{dx} - 2xy = 2, \;\; y(0) = 1.$$

which is clearly a linear equation. The integrating factor is $R(x) = \exp(-x^2)$ which multiplying through gives

$$\begin{aligned} e^{-x^2}\left(\frac{dy}{dx} - 2xy\right) &= 2e^{-x^2} \\ \frac{d}{dx}\left(e^{-x^2}y\right) &= 2e^{-x^2} \\ \int_0^x d\left(e^{-t^2}y\right) &= 2\int_0^x e^{-t^2}dt \end{aligned}$$

Concentrate on the lhs and noting the IC $y(0) = 1$

$$e^{-t^2}y\Big|_0^x = e^{-x^2}y(x) - y(0) = e^{-x^2}y(x) - 1$$

hence

$$\begin{aligned} e^{-x^2}y(x) - 1 &= 2\int_0^x e^{-t^2}dt \\ y(x) &= e^{x^2}\left(1 + 2\int_0^x e^{-t^2}dt\right) \end{aligned}$$

We cannot simplify the integral on the rhs any further if we wish this to remain as a closed form solution.

However we note the following non-elementary integrals

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-s^2}ds,$$

$$\operatorname{erf}c(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-s^2}ds.$$

This is the *error function* and *complimentary error function*, in turn.

The solution to the IVP can now be written

$$y(x) = e^{x^2}\left(1 + \sqrt{\pi}\operatorname{erf}(x)\right)$$

So, for example

$$\begin{aligned} \int_{x_0}^{x_1} e^{-x^2}dx &= \int_0^{x_1} e^{-x^2}dx - \int_0^{x_0} e^{-x^2}dx \\ &= \frac{\sqrt{\pi}}{2}\left(\operatorname{erf}(x_1) - \operatorname{erf}(x_0)\right). \end{aligned}$$

Working: We are using $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-s^2}ds$ which rearranges to give

$$\int_0^x e^{-s^2}ds = \frac{\sqrt{\pi}}{2}\operatorname{erf}(x)$$

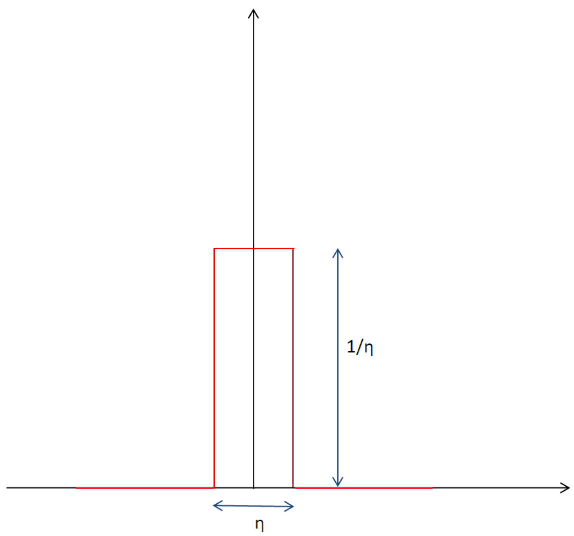
then

$$
\begin{aligned}
\int_{x_0}^{x_1} &\equiv \int_{x_0}^{0} + \int_{0}^{x_1} = -\int_{0}^{x_0} + \int_{0}^{x_1} \\
&= \int_{0}^{x_1} e^{-x^2} dx - \int_{0}^{x_0} e^{-x^2} dx \\
&= \frac{\sqrt{\pi}}{2} \left(\operatorname{erf}(x_1) - \operatorname{erf}(x_0) \right)
\end{aligned}
$$

The Dirac delta function

The *delta* function denoted $\delta(x)$; is a very useful 'object' in applied maths and more recently in quant finance. It is the mathematical representation of a point source e.g. force, payment. Although labelled a function, it is more of a distribution or *generalised function.* Consider the following definition for a piecewise function

$$f_\eta(x) = \begin{cases} \frac{1}{\eta}, & x \in \left(-\frac{\eta}{2}, \frac{\eta}{2}\right) \\ 0, & \text{otherwise} \end{cases}$$



Now put the delta function equal to the above for the following limiting value

$$\delta_\eta(x) = \lim_{\eta \longrightarrow 0} f_\eta(x)$$

What is happening here? As η decreases we note the 'hat' narrows whilst becoming taller eventually becoming a spike. Due to the definition, the area under the curve (i.e. rectangle) is fixed at 1, i.e. $\eta \times \frac{1}{\eta}$; which is independent of the value of η. So mathematically we can write in integral terms

$$\begin{aligned} \int_{-\infty}^{\infty} f_\eta(x)\,dx &= \int_{-\infty}^{-\frac{\eta}{2}} f_\eta(x)\,dx + \int_{-\frac{\eta}{2}}^{\frac{\eta}{2}} f_\eta(x)\,dx + \int_{\frac{\eta}{2}}^{\infty} f_\eta(x)\,dx \\ &= \eta \times \frac{1}{\eta} = 1 \text{ for all } \eta. \end{aligned}$$

Looking at what happens in the limit $\eta \longrightarrow 0$, the spike like (singular) behaviour at the origin gives the following definition

$$\delta_\eta(x) = \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases}$$

with the property

$$\int_{-\infty}^{\infty} \delta_\eta(x)\,dx = 1.$$

There are many ways to define $\delta(x)$. Consider the Gaussian/Normal distribution with pdf

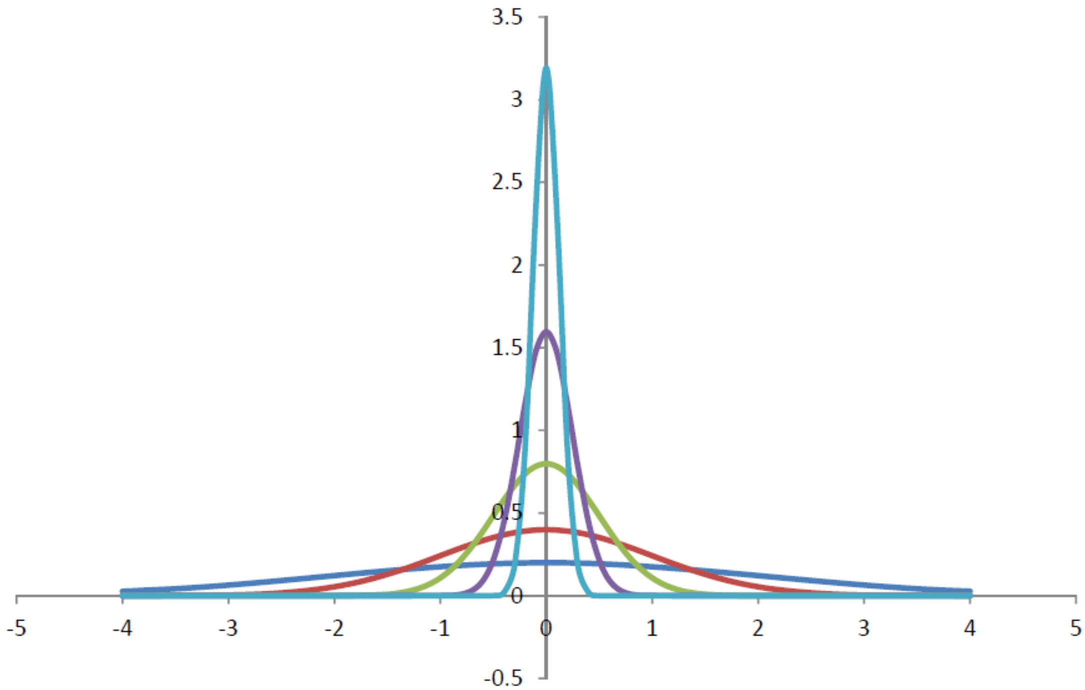
$$G_\eta(x) = \frac{1}{\eta\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\eta^2}\right).$$

The function takes its highest value at $x = 0$; as $|x| \to \infty$ there is exponential decay away from the origin. If we stay at the origin, then as η decreases, $G_\eta(x)$ exhibits the earlier spike (as it shoots up to infinity), so

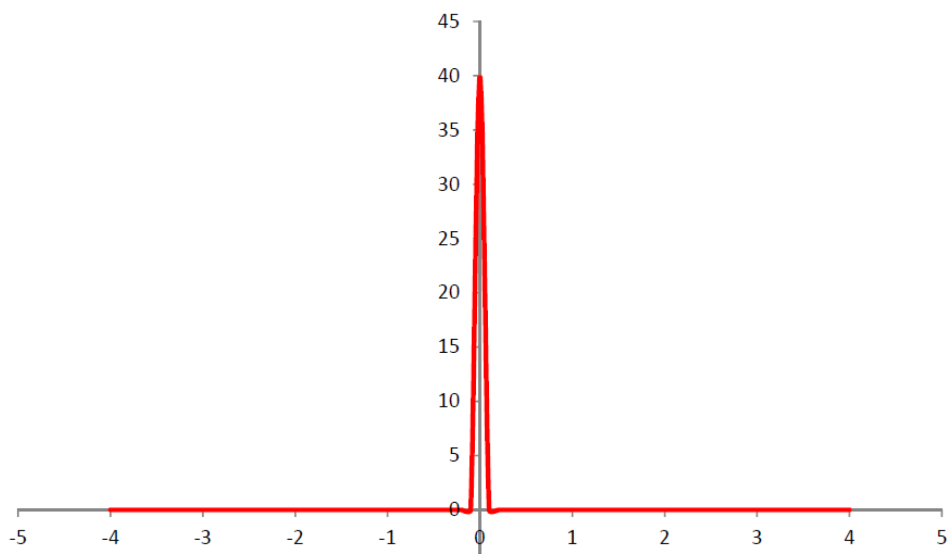
$$\lim_{\eta \longrightarrow 0} G_\eta(x) = \delta(x).$$

The normalising constant $\frac{1}{\eta\sqrt{2\pi}}$ ensures that the area under the curve will always be unity.

The graph below shows $G_\eta(x)$ for values $\eta = 2.0$ (royal blue), 1.0 (red), 0.5 (green), 0.25 (purple), 0.125 (turquoise); the Gaussian curve becomes slimmer and more peaked as η decreases.



$G_\eta(x)$ is plotted for $\eta = 0.01$



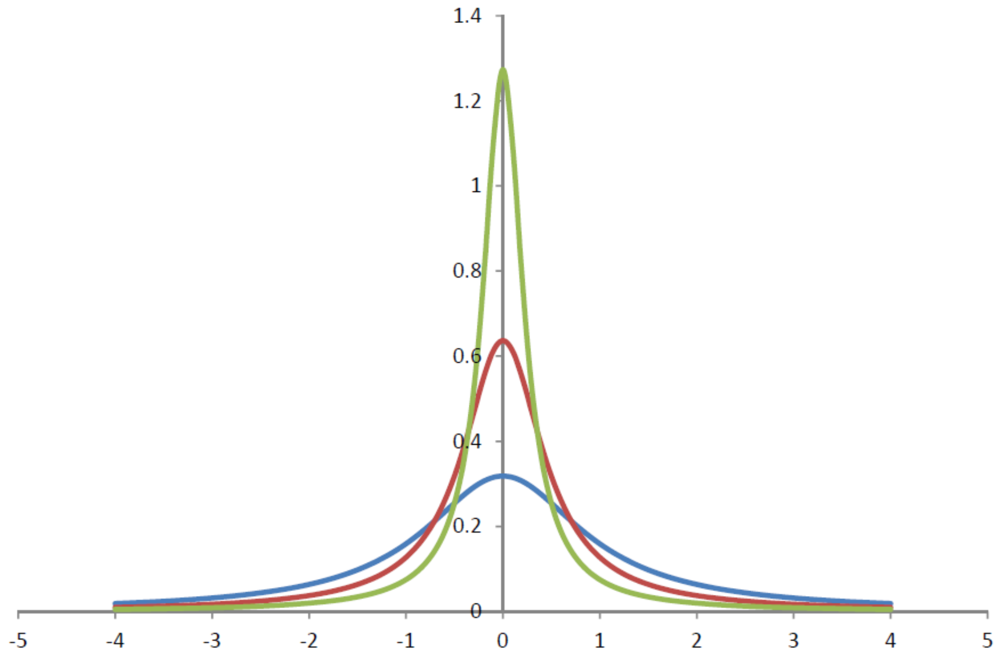
Now generalise this definition by centring the function $f_\eta(x)$ at any point x'. So

$$\delta(x-x') = \lim_{\eta\longrightarrow 0} f_\eta(x-x')$$

$$\int_{-\infty}^{\infty} \delta(x-x')\,dx = 1.$$

The figure will be as before, except that now centered at x' and not at the origin as before. So we see two definitions of $\delta(x)$. Another is the Cauchy distribution

$$L_\epsilon(x) = \frac{1}{\pi}\frac{\epsilon}{x^2+\epsilon^2}$$



So here

$$\delta(x) = \lim_{\epsilon\longrightarrow 0}\frac{1}{\pi}\frac{\epsilon}{x^2+\epsilon^2}$$

Now suppose we have a smooth function $g(x)$ and consider the following integral problem

$$\int_{-\infty}^{\infty} g(x)\,\delta(x-x_0)\,dx = g(x_0)$$

This sifting property of the delta function is a very important one.

Heaviside Function

The **Heaviside function**, denoted by $H()$, is a discontinuous function whose value is zero for negative parameters and one for positive arguments

$$\mathcal{H}(x) = \begin{cases} 1 & x>0 \\ 0 & x<0 \end{cases}$$

Some definitions have

$$\mathcal{H}(x) = \begin{cases} 1 & x>0 \\ \frac{1}{2} & x=0 \\ 0 & x<0 \end{cases}$$

and

$$\mathcal{H}(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

It is an example of the general class of step functions.

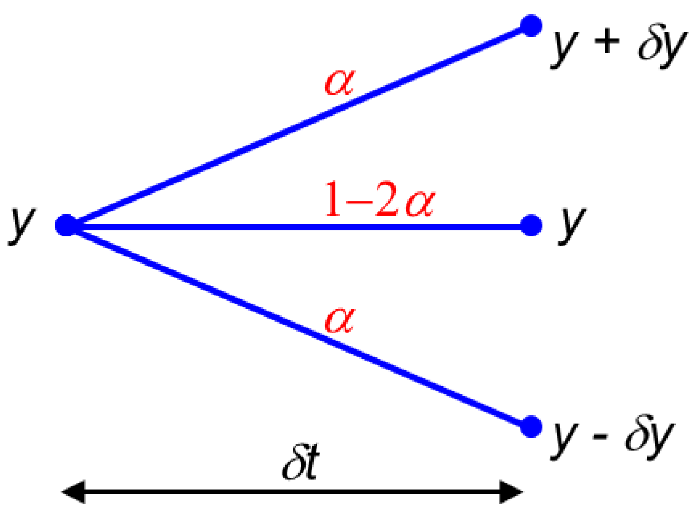
Probability Distributions

At the heart of modern finance theory lies the uncertain movement of financial quantities. For modelling purposes we are concerned with the evolution of random events through time.

A *diffusion process* is one that is continuous in space, while a *random walk* is a process that is discrete. The random path followed by the process is called a realization. Hence when referring to the path traced out by a financial variable will be termed as an asset price realization.

The mathematics can be achieved by the concept of a transition density function and is the connection between probability theory and differential equations.

Trinomial Random Walk



A trinomial random walk models the dynamics of a random variable, with value y at time t. α is a probability. δy is the size of the move in y.

The Transition Probability Density Function

The transition pdf is denoted by

$$p\left(y,t;y',t'\right)$$

We can gain information such as the centre of the distribution, where the random variable might be in the long run, etc. by studying its probabilistic properties. So the density of particles diffusing from (y,t) to (y',t').

Think of (y,t) as current (or backward) variables and (y',t') as future ones.

The more basic assistance it gives is with

$$\mathbb{P}\left(a<y'<b \text{ at } t' \middle| \, y \text{ at } t\right)=\int_a^b p\left(y,t;y',t'\right)dy'$$

i.e. the probability that the random variable y' lies in the interval a and b,at a future time t', given it started out at time t with value y.

$p(y,t;y',t')$ satisfies two equations:

Forward equation involving derivatives with respect to the future state (y',t'). Here (y,t) is a starting point and is 'fixed'.

Backward equation involving derivatives with respect to the current state (y,t). Here (y',t') is a future point and is 'fixed'. The backward equation tells us the probability that we were at (y,t) given that we are now at (y',t'), which is fixed.

The mathematics: Start out at a point (y,t). We want to answer the question, what is the probability density function of the position y' of the diffusion at a later time t'?

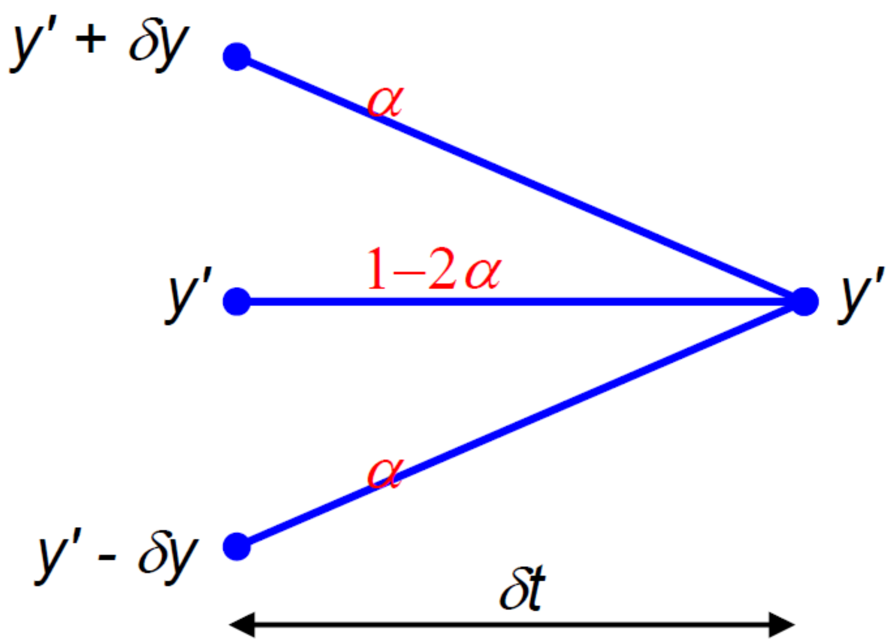
This is known as the **transition density function** written $p(y,t;y',t')$ and represents the density of particles diffusing from (y,t) to (y',t'). How can we find p?

Forward Equation

Starting with a trinomial random walk which is discrete we can obtain a continuous time process to obtain a partial differential equation for the transition probability density function (i.e. a time dependent PDF).

So the random variable can either rise or fall with equal probability $\alpha < \frac{1}{2}$ and remain at the same location with probability $1-2\alpha$.

Suppose we are at (y',t'), how did we get there?



At the previous step time step we must have been at one of $(y'+\delta y, t'-\delta t)$ or $(y'-\delta y, t'-\delta t)$ or $(y', t'-\delta t)$.

So

$$\begin{aligned} p(y,t,y',t') &= \alpha p(y,t,y'+\delta y,t'-\delta t) + (1-2\alpha)p(y,t,y',t'-\delta t) \\ &\quad + \alpha p(y,t,y'-\delta y,t'-\delta t) \end{aligned}$$

Taylor series expansion gives (omit the dependence on (y,t) in your working as they will not change)

$$
\begin{aligned}
p\left(y'+\delta y, t'-\delta t\right) &= p\left(y', t'\right)-\frac{\partial p}{\partial t'}\delta t+\frac{\partial p}{\partial y'}\delta y+\tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}\delta y^2+\ldots \\
p\left(y', t'-\delta t\right) &= p\left(y', t'\right)-\frac{\partial p}{\partial t'}\delta t+\ldots \\
p\left(y'-\delta y, t'-\delta t\right) &= p\left(y', t'\right)-\frac{\partial p}{\partial t'}\delta t-\frac{\partial p}{\partial y'}\delta y+\tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}\delta y^2+\ldots
\end{aligned}
$$

Substituting into the above

$$
\begin{aligned}
p\left(y', t'\right) &= \alpha\left(p\left(y', t'\right)-\frac{\partial p}{\partial t'}\delta t+\frac{\partial p}{\partial y'}\delta y+\tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}\delta y^2\right)+ \\
&\quad (1-2\alpha)\left(p\left(y', t'\right)-\frac{\partial p}{\partial t'}\delta t+\right) \\
&\quad +\alpha\left(p\left(y', t'\right)-\frac{\partial p}{\partial t'}\delta t-\frac{\partial p}{\partial y'}\delta y+\tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}\delta y^2\right)
\end{aligned}
$$

$$
0=-\frac{\partial p}{\partial t'}\delta t+\alpha\frac{\partial^2 p}{\partial y'^2}\delta y^2
$$

$$
\frac{\partial p}{\partial t'}=\alpha\frac{\delta y^2}{\delta t}\frac{\partial^2 p}{\partial y'^2}
$$

Now take limits. This only makes sense if $\alpha\frac{\delta y^2}{\delta t}$ is $O(1)$, i.e. $\delta y^2 \sim O(\delta t)$ and letting δy, $\delta t \longrightarrow 0$ gives the equation

$$
\frac{\partial p}{\partial t'}=c^2\frac{\partial^2 p}{\partial y'^2},
$$

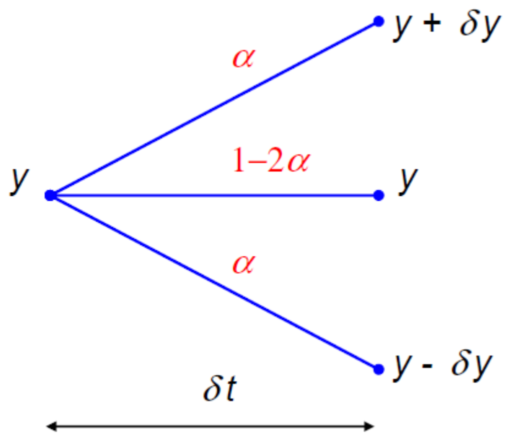
where $c^2=\alpha\frac{\delta y^2}{\delta t}$. This is called the **forward Kolmogorov equation.** Also called Fokker Planck equation.

It shows how the probability density of future states evolves, starting from (y,t).

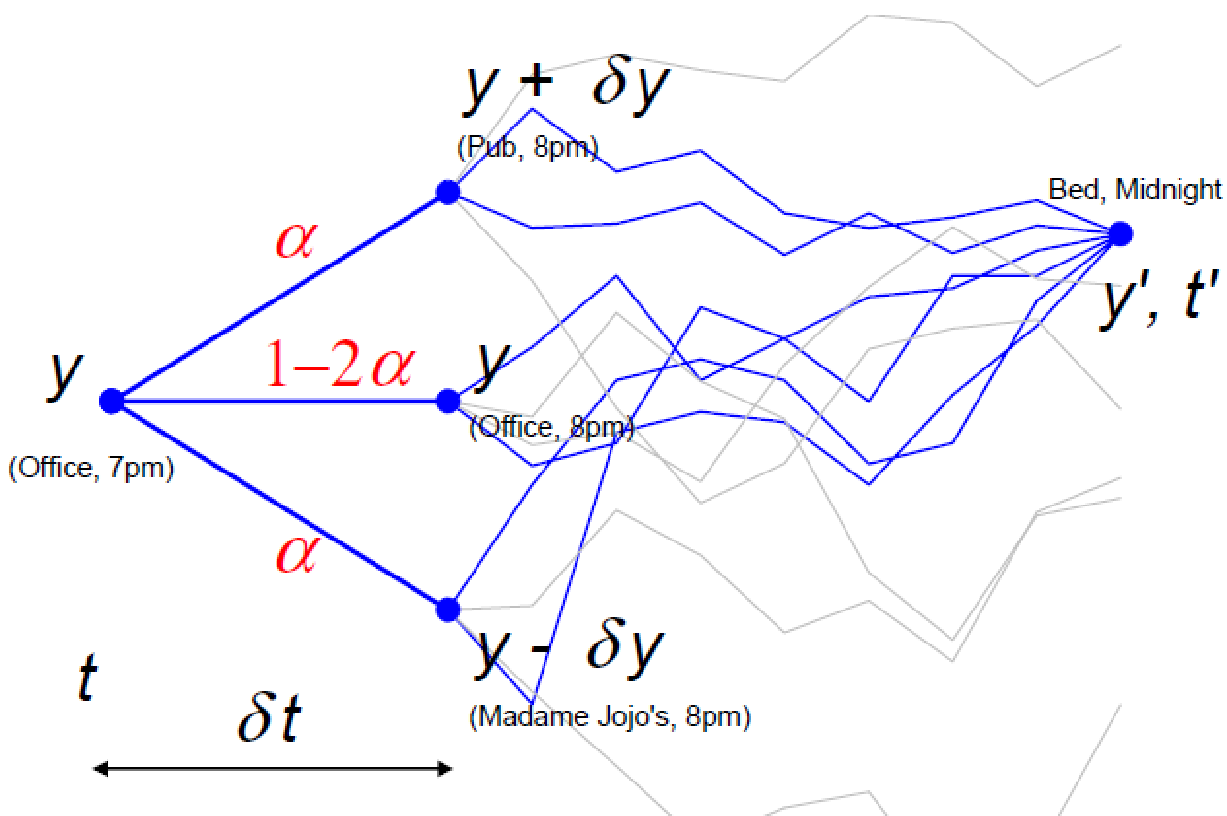
The Backward Equation

The backward equation is particularly important in the context of finance, but also a source of much confusion. Illustrate with the 'real life' example that Wilmott uses.

Wilmott uses a *Trinomial* Random Walk



So 3 possible states at the next time step. Here $\alpha < 1/2$.



- At 7pm you are at the office - this is the point (y,t)
- At 8pm you will be at one of three places:

  § The Pub - the point $(y+\delta y, t+\delta t)$;
  § Still at the office - the point $(y, t+\delta t)$;
  § Madame Jojo's - the point $(y-\delta y, t+\delta t)$

We are interested in the probability of being tucked up in bed at midnight (y', t'), given that we were at the office at 7pm.

Looking at the earlier figure, we can only get to bed at midnight via either

- the pub
- the office
- Madame Jojo's

at 8pm.

What happens after 8pm doesn't matter - we don't care, you may not even remember! We are only concerned with being in bed at midnight.

The earlier figure shows many different paths, only the ones ending up in 'our' bed are of interest to us.

In words: The probability of going from the office at 7pm to bed at midnight is

- the probability of going to the pub from the office and then to bed at midnight plus
- the probability of staying in the office and then going to bed at midnight plus
- the probability of going to Madame Jojo's from the office and then to bed at midnight

The above can be expressed mathematically as

$$\begin{aligned} p(y,t;y',t') &= \alpha p(y+\delta y, t+\delta t; y', t') + (1-2\alpha)\, p(y, t+\delta t; y', t') + \\ & \quad \alpha p(y-\delta y, t+\delta t; y', t'). \end{aligned}$$

Performing a Taylor expansion gives dropping y', t'

$$\begin{aligned} p(y,t) &= \alpha\left(p + \frac{\partial p}{\partial t}\delta t + \frac{\partial p}{\partial y}\delta y + \frac{1}{2}\frac{\partial^2 p}{\partial y^2}\delta y^2 + ..\right) \\ & \quad + (1-2\alpha)\left(p - \frac{\partial p}{\partial t}\delta t + ..\right) \\ & \quad \alpha\left(p + \frac{\partial p}{\partial t}\delta t - \frac{\partial p}{\partial y}\delta y + \frac{1}{2}\frac{\partial^2 p}{\partial y^2}\delta y^2 + ..\right). \end{aligned}$$

Most of the terms cancel and leave

$$0 = \delta t \frac{\partial p}{\partial t} + \alpha \delta y^2 \frac{\partial^2 p}{\partial y^2} + ...$$

which becomes

$$0 = \frac{\partial p}{\partial t} + \alpha \frac{\delta y^2}{\delta t}\frac{\partial^2 p}{\partial y^2} + ...$$

and letting $\alpha \frac{\delta y^2}{\delta t} = c^2$ where c is non-zero and finite as $\delta t, \delta y \longrightarrow 0$, we have

$$\frac{\partial p}{\partial t} + c^2 \frac{\partial^2 p}{\partial y^2} = 0$$

Solving the Forward Equation

The equation is

$$\frac{\partial p}{\partial t'} = c^2 \frac{\partial^2 p}{\partial y'^2}$$

for the unknown function $p = p(y', t')$. The idea is to obtain a solution in terms of Gaussian curves. Let's drop the primed notation.

We assume a solution of the following form exists:

$$p(y,t) = t^a f\left(\frac{y}{t^b}\right)$$

where a, b are constants to be determined. So put

$$\xi = \frac{y}{t^b} = yt^{-b},$$

which is a dimensionless variable. We have the following derivatives

$$\frac{\partial \xi}{\partial y} = t^{-b}; \quad \frac{\partial \xi}{\partial t} = -byt^{-b-1}$$

we can now say

$$p(y,t) = t^a f(\xi)$$

therefore

$$\frac{\partial p}{\partial y} = \frac{\partial p}{\partial \xi}\frac{\partial \xi}{\partial y} = t^a f'(\xi).t^{-b} = t^{a-b} f'(\xi)$$

$$\begin{aligned}\frac{\partial^2 p}{\partial y^2} &= \frac{\partial}{\partial y}\left(\frac{\partial p}{\partial y}\right) = \frac{\partial}{\partial y}\left(t^{a-b} f'(\xi)\right) \\ &= \frac{\partial \xi}{\partial y}\frac{\partial}{\partial \xi}\left(t^{a-b} f'(\xi)\right) \\ &= t^{a-b}\frac{1}{t^b}\frac{\partial}{\partial \xi} f'(\xi) = t^{a-2b} f''(\xi)\end{aligned}$$

$$\frac{\partial p}{\partial t} = t^a \frac{\partial}{\partial t} f(\xi) + at^{a-1} f(\xi)$$

we can use the chain rule to write

$$\frac{\partial}{\partial t} f(\xi) = \frac{\partial f}{\partial \xi}.\frac{\partial \xi}{\partial t} = -byt^{-b-1} f'(\xi)$$

so we have

$$\frac{\partial p}{\partial t} = at^{a-1} f(\xi) - byt^{a-b-1} f'(\xi)$$

and then substituting these expressions in to the pde gives

$$at^{a-1}f(\xi) - byt^{a-b-1}f'(\xi) = c^2 t^{a-2b} f''.$$

We know from ξ that

$$y = t^b \xi$$

hence the equation above becomes

$$at^{a-1}f(\xi) - b\xi t^{a-1}f'(\xi) = c^2 t^{a-2b} f''.$$

For the similarity solution to exist we require the equation to be independent of t, i.e. $a-1 = a-2b \Longrightarrow b = 1/2$, therefore

$$af - \frac{1}{2}\xi f' = c^2 f''$$

thus we have so far

$$p = t^a f\left(\tfrac{y}{\sqrt{t}}\right)$$

which gives us a whole family of solutions dependent upon the choice of a.

We know that p represents a pdf, hence

$$\int_{\mathbb{R}} p(y,t)\, dy = 1 = \int_{\mathbb{R}} t^a f\left(\tfrac{y}{\sqrt{t}}\right) dy$$

change of variables $u = y/\sqrt{t} \longrightarrow du = dy/\sqrt{t}$ so the integral becomes

$$t^{a+1/2}\int_{-\infty}^{\infty} f(u)\, du = 1$$

which we need to normalize independent of time t. This is only possible if $a = -1/2$.

So the D.E becomes

$$-\frac{1}{2}(f + \xi f') = c^2 f''.$$

We have an exact derivative on the lhs, i.e. $\dfrac{d}{d\xi}(\xi f) = f + \xi f'$, hence

$$-\frac{1}{2}\frac{d}{d\xi}(\xi f) = c^2 f''$$

and we can integrate once to get

$$-\frac{1}{2}(\xi f) = c^2 f' + K.$$

We obtain K from the following information about a probability density, as $\xi \to \infty$

$$\begin{array}{rcl} f(\xi) & \to & 0 \\ f'(\xi) & \to & 0 \end{array}$$

hence $K = 0$ in order to get the correct solution, i.e.

$$-\frac{1}{2}(\xi f) = c^2 f'$$

which can be solved as a simple first order variable separable equation:

$$f(\xi) = A \exp\left(-\tfrac{1}{4c^2}\xi^2\right).$$

A is a normalizing constant, so write

$$A \int_{\mathbb{R}} \exp\left(-\tfrac{1}{4c^2}\xi^2\right) d\xi = 1.$$

Now substitute $x = \xi/2c$, so $2c dx = d\xi$

$$2c \underbrace{A \int_{\mathbb{R}} \exp\left(-x^2\right) dx}_{=\sqrt{\pi}} = 1,$$

which gives $A = 1/2c\sqrt{\pi}$. Returning to

$$p\left(y,t\right) = t^{-1/2} f\left(\xi\right)$$

becomes

$$p\left(y',t'\right) = \frac{1}{2c\sqrt{\pi t'}} \exp\left(-\frac{{y'}^2}{4t'c^2}\right).$$

This is a pdf for a variable y that is normally distributed with mean zero and standard deviation $c\sqrt{2t}$, which we ascertained by the following comparison:

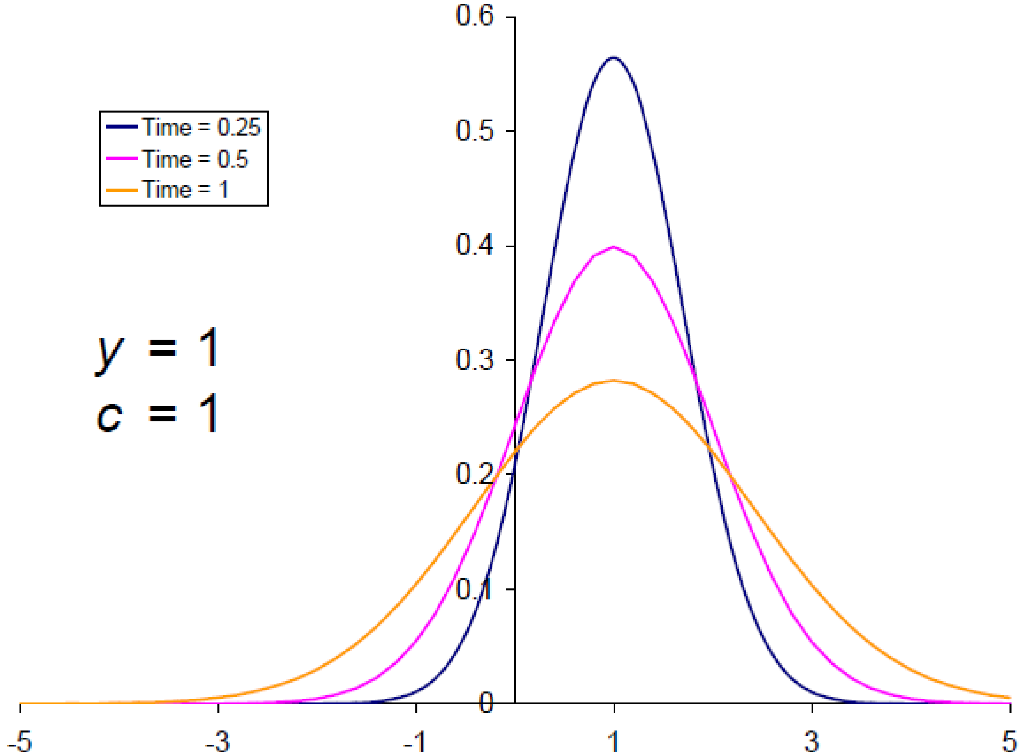
$$-\frac{1}{2}\frac{{y'}^2}{2t'c^2} : -\frac{1}{2}\frac{\left(x-\mu\right)^2}{\sigma^2}$$

i.e. $\mu \equiv 0$ and $\sigma^2 \equiv 2t'c^2$.

This solution is also called the **Source Solution** or **Fundamental Solution**.

If the random variable y' has value y at time t then we can generalize to

$$p\left(y,t;y',t'\right) = \frac{1}{2c\sqrt{\pi\left(t'-t\right)}} \exp\left(-\frac{\left(y'-y\right)^2}{4c^2\left(t'-t\right)}\right)$$



At $t' = t$ this is now a Dirac delta function $\delta(y' - y)$. This particle is known to start from (y, t) and diffuses out to (y', t') with mean y and variance $(t' - t)$

Recall this behaviour of decay away from one point y, unbounded growth at that point and constant area means that $p(y, t; y', t')$ has turned in to a **Dirac delta function** $\delta(y' - y)$ as $t' \to t$.

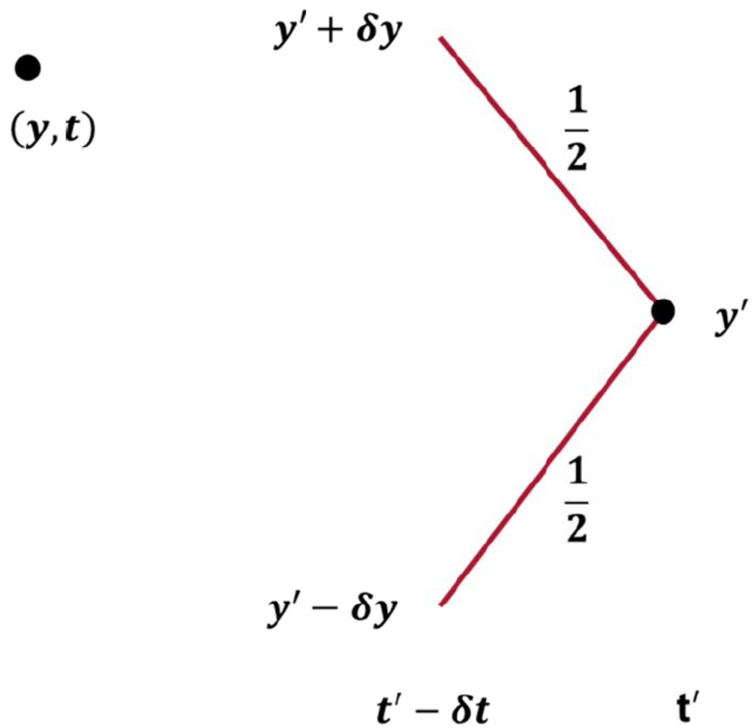
Using a Binomial random walk

The earlier results can also be obtained using a symmetric random walk. Consider the following (two step) binomial random walk. So the random variable can either rise or fall with equal probability.

y is the random variable and δt is a time step. δy is the size of the move in y.

$\mathbb{P}[\delta y] = \mathbb{P}[-\delta y] = 1/2.$

Suppose we are at (y', t'), how did we get there? At the previous step time step we must have been at one of $(y' + \delta y, t' - \delta t)$ or $(y' - \delta y, t' - \delta t)$.



So

$$p(y', t') = \tfrac{1}{2} p(y' + \delta y, t' - \delta t) + \tfrac{1}{2} p(y' - \delta y, t' - \delta t)$$

Taylor series expansion gives

$$\begin{aligned} p(y' + \delta y, t' - \delta t) &= p(y', t') - \frac{\partial p}{\partial t'}\delta t + \frac{\partial p}{\partial y'}\delta y + \tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}\delta y^2 + ... \\ p(y' - \delta y, t' - \delta t) &= p(y', t') - \frac{\partial p}{\partial t'}\delta t - \frac{\partial p}{\partial y'}\delta y + \tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}\delta y^2 + ... \end{aligned}$$

Substituting into the above

$$\begin{aligned} p(y', t') &= \tfrac{1}{2}\left(p(y', t') - \frac{\partial p}{\partial t'}\delta t + \frac{\partial p}{\partial y'}\delta y + \tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}\delta y^2\right) \\ &\quad + \tfrac{1}{2}\left(p(y', t') - \frac{\partial p}{\partial t'}\delta t - \frac{\partial p}{\partial y'}\delta y + \tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}\delta y^2\right) \end{aligned}$$

$$0 = -\frac{\partial p}{\partial t'}\delta t + \tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}\delta y^2$$

$$\frac{\partial p}{\partial t'} = \tfrac{1}{2}\frac{\delta y^2}{\delta t}\frac{\partial^2 p}{\partial y'^2}$$

Now take limits. This only makes sense if $\frac{\delta y^2}{\delta t}$ is $O(1)$, i.e. $\delta y^2 \sim O(\delta t)$ and letting $\delta y,\ \delta t \longrightarrow 0$ gives the equation

$$\frac{\partial p}{\partial t'} = \tfrac{1}{2}\frac{\partial^2 p}{\partial y'^2}$$

This is called the **forward Kolmogorov equation.** Also called Fokker Planck equation.

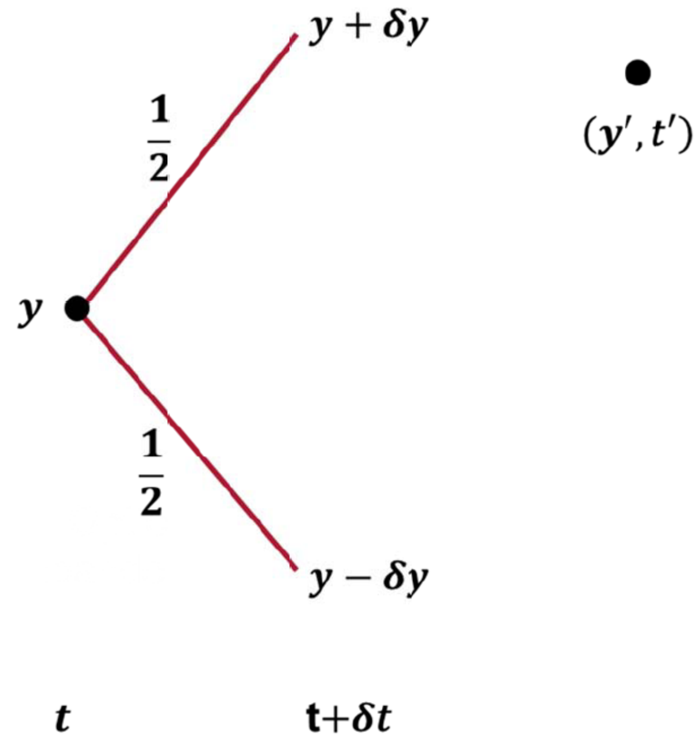
It shows how the probability density of future states evolves, starting from (y, t).

A particular solution of this is

$$p(y, t; y', t') = \frac{1}{\sqrt{2\pi(t'-t)}} \exp\left(-\frac{(y'-y)^2}{2(t'-t)}\right)$$

At $t' = t$ this is equal to $\delta(y' - y)$. The particle is known to start from (y, t) and its density is normal with mean y and variance $t' - t$.

The **backward equation** tells us the probability that we are at (y,t) given that we are at (y',t') in the future. So (y',t') are now fixed and (y,t) are variables. So the probability of being at (y,t) given we are at y' at t' is linked to the probabilities of being at $(y+\delta y, t+\delta t)$ and $(y-\delta y, t+\delta t)$.



$$p\left(y,t;y',t'\right)=\tfrac{1}{2}p\left(y+\delta y,t+\delta t;y',t'\right)+\tfrac{1}{2}p\left(y-\delta y,t+\delta t;y',t'\right)$$

Since (y',t') do not change, drop these for the time being and use a TSE on the right hand side

$p\left(y,t\right)=$

$$\begin{array}{l}\tfrac{1}{2}\left(p\left(y,t\right)+\dfrac{\partial p}{\partial t}\delta t+\dfrac{\partial p}{\partial y}\delta y+\tfrac{1}{2}\dfrac{\partial^2 p}{\partial y^2}\delta y^2+...\right)+\\ \tfrac{1}{2}\left(p\left(y,t\right)+\dfrac{\partial p}{\partial t}\delta t-\dfrac{\partial p}{\partial y}\delta y+\tfrac{1}{2}\dfrac{\partial^2 p}{\partial y^2}\delta y^2+...\right)\end{array}$$

which simplifies to

$$0=\frac{\partial p}{\partial t}+\tfrac{1}{2}\frac{\delta y^2}{\delta t}\frac{\partial^2 p}{\partial y^2}.$$

Putting $\frac{\delta y^2}{\delta t}=O\left(1\right)$ and taking limit gives the **backward equation**

$$-\frac{\partial p}{\partial t}=\tfrac{1}{2}c^2\frac{\partial^2 p}{\partial y^2}.$$

or commonly written as $\dfrac{\partial p}{\partial t}+\dfrac{1}{2}\dfrac{\partial^2 p}{\partial y^2}=0$

Further Solutions of the heat equation

We know the one dimensional heat/diffusion equation

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}$$

can be solved by seeking a solution of the form $u(x,t)=t^{\alpha}\phi\left(\frac{x}{t^{\beta}}\right)$. The corresponding solution derived using the similarity reduction technique is the *fundamental solution*

$$u(x,t)=\frac{1}{2\sqrt{\pi t}}\exp\left(-\frac{x^2}{4t}\right).$$

Some books refer to this as a *source solution.*

Let's consider the following integral

$$\lim_{t\longrightarrow 0}\int_{-\infty}^{\infty}u(y,t)\,f(y)\,dy$$

which can be simplified by the substitution

$$s=\frac{y}{2\sqrt{t}}\Longrightarrow 2\sqrt{t}ds=dy$$

to give

$$\lim_{t\longrightarrow 0}\frac{1}{2\sqrt{\pi t}}\int_{-\infty}^{\infty}\exp\left(-s^2\right)f\left(2\sqrt{t}s\right)2\sqrt{t}ds.$$

In the limiting process we get

$$\begin{aligned} f(0)\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp\left(-s^2\right)ds &= f(0)\frac{1}{\sqrt{\pi}}\sqrt{\pi} \\ &= f(0). \end{aligned}$$

Hence

$$\lim_{t\longrightarrow 0}\int_{-\infty}^{\infty}u(y,t)\,f(y)\,dy=f(0).$$

A slight extension of the above shows that

$$\lim_{t\longrightarrow 0}\int_{-\infty}^{\infty}u(x-y,t)\,f(y)\,dy=f(x),$$

where

$$u(x-y,t)=\frac{1}{2\sqrt{\pi t}}\exp\left(-\frac{(x-y)^2}{4t}\right).$$

Let's derive the result above. As earlier we begin by writing $s=\dfrac{x-y}{2\sqrt{t}}\Longrightarrow y=x-2\sqrt{t}s$ and hence $dy=-2\sqrt{t}ds$. Under this transformation the limits are

$$\begin{aligned} y &= \infty\longrightarrow s=-\infty \\ y &= -\infty\longrightarrow s=\infty \end{aligned}$$

$$\tfrac{1}{2\sqrt{\pi t}}\int_{\infty}^{-\infty}\exp\left(-s^2\right)f\left(x-2\sqrt{t}s\right)\left(-2\sqrt{t}ds\right)ds$$

$$\begin{aligned}
&\lim_{t\longrightarrow 0}\tfrac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp\left(-s^2\right)f\left(x-2\sqrt{t}s\right)ds\\
&= f\left(x\right)\tfrac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp\left(-s^2\right)ds\\
&= f\left(x\right)\tfrac{1}{\sqrt{\pi}}\sqrt{\pi}
\end{aligned}$$

and

$$\lim_{t\longrightarrow 0}\int_{-\infty}^{\infty}u\left(x-y,t\right)f\left(y\right)dy=f\left(x\right).$$

Since the heat equation is a constant coefficient PDE, if $u\left(x,t\right)$ satisfies it, then $u\left(x-y,t\right)$ is also a solution for any y.

Recall what it means for an equation to be linear:

Since the heat equation is linear,

1. if $u\left(x-y,t\right)$ is a solution, so is a multiple $f\left(y\right)u\left(x-y,t\right)$
2. we can add up solutions. Since $f\left(y\right)u\left(x-y,t\right)$ is a solution for any y, so too is the integral

$$\int_{-\infty}^{\infty}u\left(x-y,t\right)f\left(y\right)dy.$$

Recall, adding can be done in terms of an integral. So we we can summarize by specifying the following initial value problem

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \frac{\partial^2 u}{\partial x^2}\\
u\left(x,0\right) &= f\left(x\right)
\end{aligned}$$

which has a solution

$$u\left(x,t\right)=\tfrac{1}{2\sqrt{\pi t}}\int_{\infty}^{-\infty}\exp\left(-\frac{\left(x-y\right)^2}{4t}\right)f\left(y\right)dy.$$

This satisfies the initial condition at $t=0$ because we have shown that at that point the value of this integral is $f\left(x\right)$. Putting $t<0$ gives a non-existent solution, i.e. the integrand will blow up.

Example 1 *Consider the IVP*

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \frac{\partial^2 u}{\partial x^2}\\
u\left(x,0\right) &= \begin{cases}0 & \text{if } x>0\\ 1 & \text{if } x<0\end{cases}
\end{aligned}$$

We can write down the solution as

$$\begin{aligned}
u\left(x,t\right) &= \tfrac{1}{2\sqrt{\pi t}}\int_{\infty}^{-\infty}\exp\left(-\frac{\left(x-y\right)^2}{4t}\right)\underbrace{u\left(y,0\right)}_{=f(y)}dy\\
&= \tfrac{1}{2\sqrt{\pi t}}\int_{-\infty}^{0}\exp\left(-\frac{\left(x-y\right)^2}{4t}\right).1dy
\end{aligned}$$

put

$$s = \frac{y - x}{\sqrt{2t}}$$

$\int_{-\infty}^{0}$ *becomes* $\int_{-\infty}^{\frac{-x}{\sqrt{2t}}}$

$$\begin{aligned} & \tfrac{1}{2\sqrt{\pi t}} \int_{-\infty}^{\frac{-x}{2\sqrt{t}}} \exp\left(-s^2/2\right) \sqrt{2t} ds \\ = & \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-x}{\sqrt{2t}}} \exp\left(-s^2/2\right) ds \\ = & N\left(\tfrac{-x}{\sqrt{2t}}\right) \end{aligned}$$

So we have expressed the solution in terms of the CDF.

This can also be solved by using the substitution

$$\widehat{s} = \frac{-(y - x)}{2\sqrt{t}} \longrightarrow -dy = 2\sqrt{t} d\widehat{s}$$

$\int_{-\infty}^{0}$ becomes $\int_{\infty}^{\frac{x}{2\sqrt{t}}}$

$$\begin{aligned} & -\tfrac{1}{2\sqrt{\pi t}} \int_{\infty}^{\frac{x}{2\sqrt{t}}} \exp\left(-\widehat{s}^2\right) 2\sqrt{t} d\widehat{s} \\ = & \tfrac{1}{2} \cdot \tfrac{2}{\sqrt{\pi}} \int_{\frac{x}{2\sqrt{t}}}^{\infty} \exp\left(-\widehat{s}^2\right) d\widehat{s} \\ = & \tfrac{1}{2} \operatorname{erf} \mathrm{c}\left(\tfrac{x}{2\sqrt{t}}\right) \end{aligned}$$

so now we have a solution in terms of the complimentary error function.