Lecture 2: Stochastic differential equations

(a) Review of probability theory

Discrete probability distributions

Basic setup of a random experiment:

1. Sample space: Ω, consists of all elementary outcomes.
2. Set of all events $\mathcal{G}$, which consists of all subsets of Ω.
3. Probability measure: $P : \mathcal{G} \to [0, 1]$ which assigns a probability to each event.

Axioms of Probability:

1. $P(A) \geq 0$ for all $A \in \mathcal{G}$, $\quad P(\Omega) = 1$.
2. If A and B are disjoint events, then $P(A \cup B) = P(A) + P(B)$.

Important properties:

$$P(\bar{A}) = P(\Omega - A) = 1 - P(A), \tag{1}$$
$$P(\emptyset) = 0, \tag{2}$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,. \tag{3}$$

Examples:

1. Roll a dice once. Here, $\Omega = \{1, 2, 3, 4, 5, 6\}$. Each elementary outcome is equally likely, hence $P(\{1\}) = 1/6$, $P(\{2\}) = 1/6$, etc. We can use this to compute the probabilities of events that are not elementary:

$$P(\{1, 3, 5\}) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{1}{2}\,.$$

2. Roll two dice once. We can represent the sample space in the following form:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Clearly, we have $6^2 = 36$ possible outcomes which are equally likely. Hence the probability of each elementary outcome is $1/36$.

Random variables: A random variable X is a real-valued function on the sample space Ω. If the range of X is finite or countable, X is called discrete, otherwise X is called continuous.
For a discrete random variable X with the range $\{x_1, x_2, ..., x_n\}$, the expectation $\mathbb{E}(X)$ of X (often denoted as μ_X), the variance $\mathrm{Var}(X)$, and the standard deviation σ_X are defined as follows:

$$\mu_X \equiv \mathbb{E}(X) = \sum_{k=1}^{n} x_k\, p_k\,, \tag{4}$$

$$\mathrm{Var}(X) = \mathbb{E}\left((X-\mu_X)^2\right) = \sum_{k=1}^{n} (x_k-\mu_X)^2\, p_k\,, \tag{5}$$

$$\sigma_X = \sqrt{\mathrm{Var}(X)}\,. \tag{6}$$

Here, $p_k = P(X = x_k)$ is the probability that the random variable takes the value x_k. In order to compute this probability, we need to find the set $A \in \mathcal{G}$ that is mapped by X onto x_k, or

$$p_k = P(X = x_k) = P(A), \qquad A = X^{-1}(x_k)\,. \tag{7}$$

Note, that as $P(\Omega) = 1$, we have obviously

$$\sum_{k=1}^{n} p_k = 1\,. \tag{8}$$

From the definition of the expectation and the variance, we find directly that, for constants a and b we have

$$\mathbb{E}(aX+b) = a\mathbb{E}(X)+b, \qquad \mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)\,. \tag{9}$$

The proof is left as an exercise. Note that the variance is a quadratic quantity, therefore its scaling factor a^2 and not a. Moreover, the variance is insensitive to shifts, hence b does not enter on the right-hand side of the second equation.
For the variance, we have the following alternative formula:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mu^2. \tag{10}$$

Proof: We can see this by direct calculation. Set $\mu = \mathbb{E}(X)$. Then

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right)\\ &= \mathbb{E}\left(X^2\right) - 2\mu\mathbb{E}(X) + \mathbb{E}(\mu^2)\\ &= \mathbb{E}\left(X^2\right) - 2\mu^2 + \mu^2\\ &= \mathbb{E}(X^2) - \mu^2.\end{aligned}$$

Examples:

1. Consider again the random experiment of rolling a die once. If we define the random variable X as the number of points of the die, we have $x_k = k$ for $k = 1, 2, ..., 6$. Each $p_k = 1/6$ and we find the $\mu_X = 7/2$. Calculation:

$$\mu_X = \sum_{k=1}^{n} x_k p_k = \frac{1}{6}(1 + 2 + ... + 6) = \frac{7}{2}$$

2. Consider now a second random experiment of rolling two dice once. If we define the random variable X as the sum of the points, the range of X consists of the number from 2 to 12. Now it is (slightly) harder to find the corresponding p_k, but if we rewrite the matrix of elementary outcomes again (look at the ascending diagonals)

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

we find easily

$$\begin{aligned} P(X = 2) &= P((1, 1)) = \frac{1}{36}\,, \\ P(X = 3) &= P((2, 1)) + P((1, 2)) = \frac{2}{36}\,. \end{aligned}$$

Probability distributions: From the above examples, it is clear that the function $x_k \to p_k$ captures the essence of the random experiment. In the first case, we have $p_k = 1/6$ for all x_k, hence the distribution is uniform, its graph is flat (see Figure 1).

In the second case, p_k increases with k until it reaches its maximum at $x_k = 7$ and then it decreases. Here, the graph is a triangle, see Figure 2.

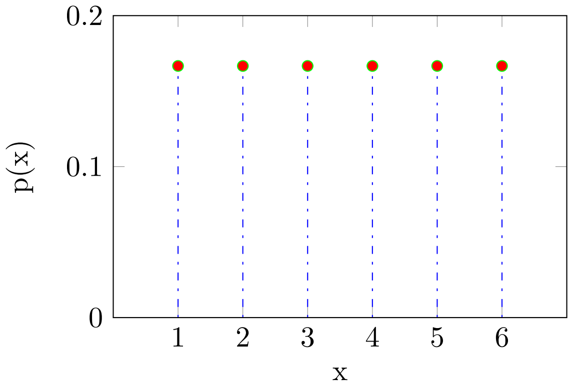


Figure 1: Probability distribution for one die.

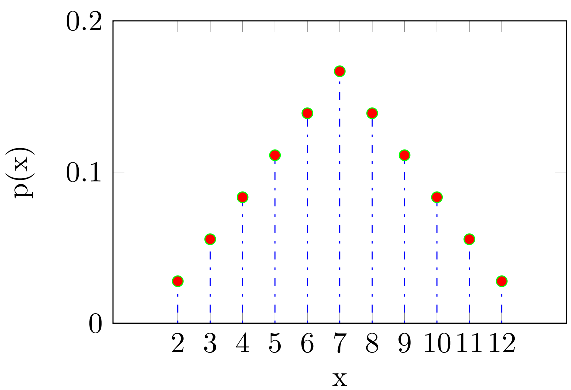


Figure 2: Probability distribution for two dice.

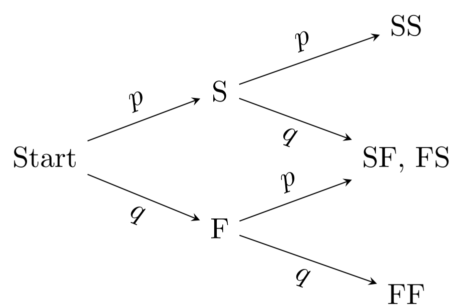
Binomial distribution: Consider a random experiment that consists of n independent trials such that

1. each trial has only two outcomes S (success) or F (failure).
2. the probabilities for success $p = P(\mathrm{S})$ and $q = P(\mathrm{F}) = 1 - p$ are the same for all trials.
3. the random variable X counts the number of successes in the n trials

Consider, for example $n = 2$, and $p = q = 1/2$. The possible outcomes of the experiment are $\{SS, SF, FS, FF\}$. The range of X is $\{0, 1, 2\}$ and we find the following probability distribution:

$$\begin{aligned} P(X=0) &= P(FF) = \frac{1}{4}, \\ P(X=1) &= P(FS) + P(SF) = \frac{1}{2}, \\ P(X=2) &= P(SS) = \frac{1}{4}. \end{aligned}$$

We can represent this experiment in a *tree* diagram:



In the general case, the range of X is $\{0, 1, 2, ..., n\}$. The probability of obtaining k successes during the n trials, can be computed as

$$P(X=k) = \binom{n}{k} p^k q^{n-k}. \tag{11}$$

Here, $\binom{n}{k}$ is the binomial coefficient, sometimes written as ${}_nC_k$. Hence the binomial distribution is characterized by the two parameters n and p and we write $X \sim B(n, p)$.

Proof: Consider the probability of the particular event A that represents the sequence of obtaining first k successes followed by $n-k$ failures:

$$A = \underbrace{S...S}_{k}\underbrace{F...F}_{n-k}.$$

Since the probabilities on the branches of the tree multiply due to independence, the probability of this sequence is given by

$$P(\{S...SF...F\}) = p^k \, q^{n-k}.$$

Since the random variable X only counts the number of successes, the order in which the successes appear in the sequence is irrelevant: All events with exactly k successes contribute to the total probability $P(X = k)$ and there are exactly $\binom{n}{k}$ such events.

We will later need formulas for the expectation (mean) of X and its variance. For the binomial distribution we have

$$\mathbb{E}(X) = np, \qquad \sigma_X^2 = npq\,. \tag{12}$$

Proof: First remember that, for the sum of two random variables X_1 and X_2 we have

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)\,, \tag{13}$$

hence for a sum of n random variables we have

$$\mathbb{E}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbb{E}(X_k)\,. \tag{14}$$

We can represent the binomial random variable X as a sum of Bernoulli random variables

$$X = \sum_{k=1}^{n} X_k\,,$$

where each X_k takes the value 1 in case of success, otherwise the value 0. Clearly, the expected value of X_k is

$$\mathbb{E}(X_k) = p \cdot 1 + q \cdot 0 = p$$

and, therefore, we have $\mathbb{E}(X) = np$. In order to prove the statement about the variance, remember that, for *independent* random variables, the variances add up:

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k)\,. \tag{15}$$

The variance of each X_k can be computed directly:

$$\mathrm{Var}(X_k) = \mathbb{E}(X_k^2) - \mu_k^2 = p - p^2 = p(1-p) = pq\,,$$

which yields $\mathrm{Var}(X) = npq$.

Continuous probability distributions

In the following we consider a continuous random variable X with a probability density function $p(x)$ on $\mathbb{R}$. One way to think of the probability density function is that the probability that X takes a value in the interval $[x, x + dx)$ is given by

$$P\left(x \leq X < x + dx\right) = p(x)\, dx\,.$$

For continuous probability distributions, the sums in the formulas above become integrals. For instance we have

$$1 = \int_{\mathbb{R}} p(x)\, dx\,, \tag{16}$$

$$\mu_X = \mathbb{E}(X) = \int_{\mathbb{R}} x\, p(x)\, dx\,, \tag{17}$$

$$\sigma_X^2 = \mathrm{Var}(X) = \int_{\mathbb{R}} (x - \mu_X)^2\, p(x)\, dx\,. \tag{18}$$

We also define the cumulative distribution given by $F_X(x) = P(X < x)$ that can be written as

$$F_X(x) = P(X < x) = \int_{-\infty}^{x} p(t)\, dt\,. \tag{19}$$

Normal distribution: If the random variable X has the probability density p given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\, \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} \tag{20}$$

we say that X has a normal distribution with mean μ and standard deviation σ, or $X \sim N(\mu, \sigma^2)$.

We can show by direct calculation that indeed

$$1 = \int_{\mathbb{R}} p(x)\, dx\,, \tag{21}$$

$$\mu = \mathbb{E}(X) = \int_{\mathbb{R}} x\, p(x)\, dx\,, \tag{22}$$

$$\sigma^2 = \mathrm{Var}(X) = \int_{\mathbb{R}} (x - \mu)^2\, p(x)\, dx\,. \tag{23}$$

For the proof, we need the following basic calculus result:

$$J = \int_{\mathbb{R}} \mathrm{e}^{-ax^2}\, dx = \sqrt{\frac{\pi}{a}} \tag{24}$$

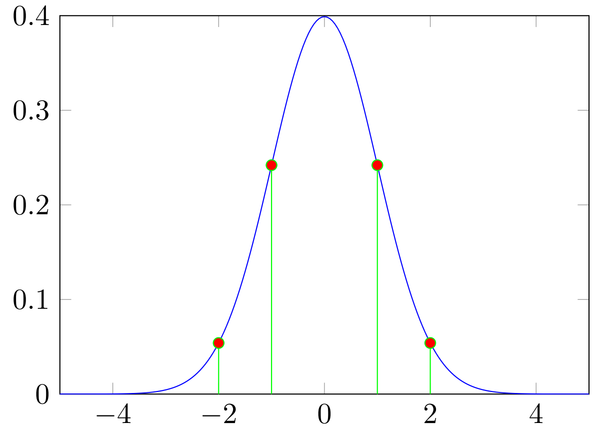


Figure 3: Standard normal distribution with $\mu = 1$, $\sigma = 1$.

Proof of this result: Consider $a = 1$ and instead of J the square J^2. Then transform to polar coordinates:

$$\begin{aligned} J^2 &= \left(\int_{\mathbb{R}} \mathrm{e}^{-x^2}\, dx\right)\left(\int_{\mathbb{R}} \mathrm{e}^{-y^2}\, dy\right) \\ &= \int_{\mathbb{R}^2} \mathrm{e}^{-(x^2+y^2)}\, dx\, dy = \int_0^\infty \int_0^{2\pi} \mathrm{e}^{-r^2} r\, d\phi\, dr \\ &= \pi \int_0^\infty 2r\, \mathrm{e}^{-r^2}\, dr = \pi\,. \end{aligned}$$

We can now proceed to prove the equations (21), (22), (23). First, we see immediately that

$$\int_{\mathbb{R}} p(x)\, dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\, dx = \frac{1}{\sqrt{2\pi}\sigma}\sqrt{2\sigma^2\pi} = 1\,.$$

In order to show (22), we first write

$$\begin{aligned} \int_{\mathbb{R}} x\, p(x)\, dx &= \int_{\mathbb{R}} (x - \mu + \mu)\, p(x)\, dx \\ &= \mu + \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} (x-\mu)\, \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\, dx\,. \end{aligned}$$

Due to symmetry (substitute $\tilde{x} = x - \mu$) the last integral is zero:

$$\int_{\mathbb{R}} (x-\mu)\, \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\, dx = \int_{\mathbb{R}} \tilde{x}\, \mathrm{e}^{-\tilde{x}^2/(2\sigma^2)}\, d\tilde{x} = 0\,.$$

For the last property (23), we need one more integral:

$$\int_{\mathbb{R}} x^2\, \mathrm{e}^{-ax^2}\, dx = \frac{1}{2a}\sqrt{\frac{\pi}{a}}\,. \tag{25}$$

This result also follows directly from what we already know: Use the basic integral (24) to define a function $J(a)$ as

$$J(a) = \int_{\mathbb{R}} \mathrm{e}^{-ax^2}\, dx = \sqrt{\frac{\pi}{a}}$$

and compute $J'(a)$ in two different ways. First, differentiate under the integral and then use the right-hand side of the above equation

$$J'(a) = -\int_{\mathbb{R}} x^2\, \mathrm{e}^{-ax^2}\, dx = \frac{d}{da}\left(\sqrt{\frac{\pi}{a}}\right) = -\frac{1}{2a}\sqrt{\frac{\pi}{a}}\,.$$

Now we can find (23) by direct computation:

$$\begin{aligned}
\int_{\mathbb{R}} (x-\mu)^2\, p(x)\, dx &= \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} (x-\mu)^2\, \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\, dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} \tilde{x}^2\, \mathrm{e}^{-\tilde{x}^2/(2\sigma^2)}\, d\tilde{x} \\
&= \frac{1}{\sqrt{2\pi}\sigma}\frac{2\sigma^2}{2}\sqrt{2\sigma^2\pi} = \sigma^2
\end{aligned}$$

Central limit theorem: The mean of a sufficiently large number of iterates of independent random variables, with a well-defined expected value and a well-defined variance, will be approximately normally distributed. In particular, for large n, the binomial distribution $B(n,p)$ becomes approximately normal with $N(np, npq)$.

Moment-generating function: For a random variable X, the moment-generating function M_X is defined as

$$M_X(s) = \mathbb{E}\left(\mathrm{e}^{sX}\right)\,. \tag{26}$$

Clearly, we have $M_X(0) = 1$. More interestingly, we find that, if we know M_X, we can compute the mean (and in fact all higher moments) of the random variable X by differentiation. For the mean, this relationship is given by

$$\mu_X = \mathbb{E}(X) = M_X'(0)\,. \tag{27}$$

Proof: From the definition of the moment-generating function, we can use the density p of X to write

$$M_X'(s) = \int_{\mathbb{R}} x\mathrm{e}^{sx} p(x)\, dx$$

If we evaluate this relationship at $s = 0$ we find immediately

$$M_X'(0) = \int_{\mathbb{R}} xp(x)\, dx = \mathbb{E}(X)$$

As an example of how to compute the moment-generating function in a concrete case, take a random variable $Z \sim N(0,1)$. We find by direct calculation:

$$\begin{aligned} M_Z(s) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{sx - x^2/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}(x^2 - 2xs)}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}(x^2 - 2xs + s^2 - s^2)}\, dx \\ &= \mathrm{e}^{s^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}(x-s)^2}\, dx = \mathrm{e}^{s^2/2} \end{aligned}$$

The trick of 'completing the square' used in the third line of the above calculation is very useful in the context of Gaussian integrals.

Simulating random numbers

In MATLAB, we can generate normally distributed random numbers (e.g. with a standard normal distribution) $Z \sim N(0,1)$ using the command `randn()`. For instance, the command `r=randn(1,1000)` creates a row vector of 1000 random numbers. In order to create histograms, one can use the command `hist()`. The following series of commands compares the histogram of the randomly generated numbers with the theoretical distribution.

```
>> [x,ps] = creategauss(10000);
>> plot(x,ps,x,1/sqrt(2*pi)*exp(-x.^2/2))
```

Here, the function `creategauss()` creates an approximation of the theoretical probability density by a sample of random numbers:

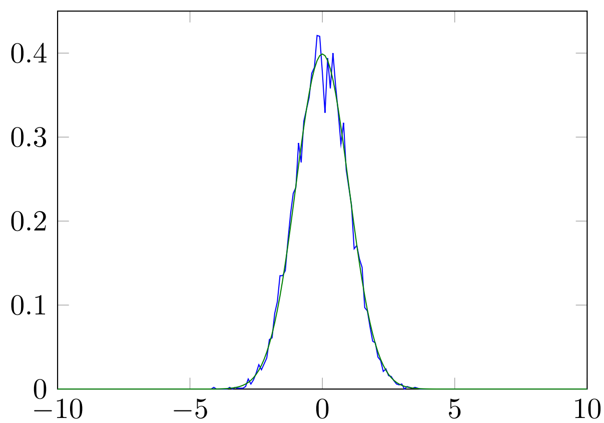


Figure 4: Standard normal distribution from 10000 samples.

```
function [x,ps] = creategauss(n)

% creategauss.m Sampling of a standard normal distribution
%   [x,ps] = creategauss(n);
%   plot(x,ps,x, 1/sqrt(2*pi)*exp(-x.^2/2))

x = [-10:0.1:10]; % create x range
dx = x(2)-x(1);   % here 0.1, of course

r = randn(1,n);   % draw n random numbers
ps = hist(r,x);   % create histogram, centers given by x
ps = ps/sum(ps)*1/dx; % normalize
end
```

(b) Brownian motion and stochastic differential equations (SDEs)

Random Walks and Brownian Motion

Simple random walk: In the following, we develop a more systematic approach to formulate a continuum limit for stochastic processes. For this purpose, imagine that we devide the interval $[0, 1]$ into n steps and define a *random walk* W_n at time steps $t_0 = 0$, $t_1 = \delta t$, $t_2 = 2\delta t$, ... with $\delta t = 1/n$ as the process that

- starts at zero, hence $W_n(0) = 0$,
- can go up or down by $1/\sqrt{n}$ with a probability of $1/2$ at each step

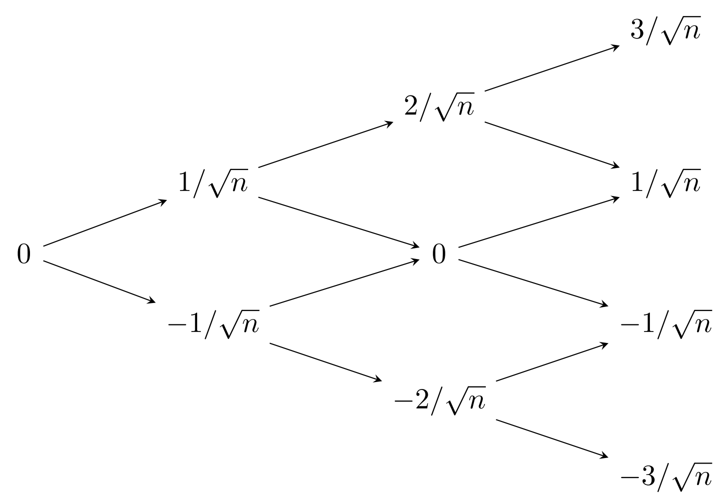


Figure 5: Simple random walk

The scaling of the 'up-' and 'down-' jumps of $1/\sqrt{n}$ is essential for the convergence of the process as we will see by the following argument: The random walk W_n can be written as a sum of independent random variables X_j, where each X_j takes the value -1 or 1 with probability 1/2, hence

$$W_n(k\delta t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{k} X_j \tag{28}$$

Clearly, the mean of W_n is zero. The variance at $k = n$ is computed as the sum of the variances of the X_j as all X_j are independent. Therefore, we find

$$\mathrm{Var}(W_n(1)) = \sum_{j=1}^{n} \left(\frac{1}{\sqrt{n}}\right)^2 \mathrm{Var}(X_j) = n \cdot 1/n = 1 \tag{29}$$

Now we see that, due to the factor $1/\sqrt{n}$, the variance of $W_n(1) = 1$ for all n, hence we can expect convergence of $W_n \to Z$ with $Z \sim N(0,1)$ by applying the central limit theorem. Often, we also write

$$W_n((k+1)\delta t) = W_n(k\delta t) + \Delta W, \qquad \Delta W = \pm\frac{1}{\sqrt{n}} = \pm\sqrt{\delta t} \tag{30}$$

where ΔW is called the *Brownian increment* which can take the values $\sqrt{\delta t}$ and $-\sqrt{\delta t}$ with probability $1/2$.
Brownian motion: We can now take the continuum limit to see that the random walk W_n converges to a continuous stochastic process W called *Brownian motion* (with respect to the measure $\mathbb{P}$. From the properties of W_n we can see that

- $W_0 = 0$
- $W_t \sim N(0, t)$
- $W_t - W_s \sim N(0, t - s)$

Moreover, due to the independence of the increments X_j, we know that all Brownian increments are independent, in particular $W_t - W_s$ is independent of the history up to the time s.

Stochastic Differential Equations

Consider the process defined by $X_t = \sigma_t W_t$. In a discrete approximation, we can compute the next value $X_{t+\Delta t}$ by drawing a random number r which is 1 or -1 with probability $1/2$ and setting

$$X_{t+\Delta t} = X_t + \sigma_t\, r\, \sqrt{\Delta t}\,.$$

A different way of writing this relationship is to consider the differential by writing

$$\Delta X_t = X_{t+\Delta t} - X_t = \sigma_t\, r\, \sqrt{\Delta t} = \sigma_t\, \Delta W$$

with the Brownian increment ΔW. The continuous version of the above equation is simply

$$dX_t = \sigma_t dW_t \tag{31}$$

and, in this form, the stochastic equation resembles a differential equation. *A (short) review of differential equations:* Let's, for a moment, put stochasticity aside and consider a basic, deterministic differential equation of the form

$$\frac{dB_t}{dt} = \mu_t$$

Writing this in a discretized version, we can write

$$\Delta B_t = B_{t+\Delta t} - B_t = \mu_t \, \Delta t$$

The continuous version of the above equation is simply

$$dB_t = \mu_t dt$$

which is very similar to (31). However, the main difference is that Δt is deterministic, whereas ΔW is stochastic. Moreover, the size of ΔW is much larger that Δt as $|\Delta W| = \sqrt{\Delta t}$.
General form of a stochastic differential equation: It is convenient to combine stochastic and deterministic contributions to the change of the stochastic process X_t in differential form by writing

$$dX_t = \mu_t dt + \sigma_t dW_t \tag{32}$$

Note that, as for ordinary differential equations the function μ_t and σ_t can depend on X_t and on t, hence

$$\mu_t = \mu(X_t, t), \qquad \sigma_t = \sigma(X_t, t)\,.$$

As for deterministic differential equations, there are no general techniques to solve any given stochastic differential equation. Many simple problems, however, can be solved by applying Ito's Lemma which we will discuss in the next section.

(c) Ito's lemma, Girsanov theorem, and martingales

Martingales

We will now analyze the mathematical structure behind the option pricing. First, we need several basic definitions, then we will see how they can be applied in a financial context.

Stochastic Process: A stochastic process Y is a collection of random variables, usually we write $Y = (Y_0, Y_1, Y_2, ...)$. We interpret the index as the number of the time-tick, and, at the moment, we only consider discrete stochastic processes. Usually, we assume that the value of Y_0 is known, and the other Y_i for $i > 0$ are random variables with several possible values. An example is the stock process S given by the above tree. Here, for example, $S_0 = 100$ and S_2 can take the values $\{60, 100, 140\}$.

Filtration: A filtration $\mathcal{F}_i$ is a history in time up to the time i. In the case of the stock process, a particular path in the tree corresponds to such a history. There are different ways to identify the history, for example by listing the corresponding nodes or stock values. Alternatively, we can also say, for instance, that we are looking at the filtration $\mathcal{F}_2 = (u, u)$ where 'u' stands for an up-movement of the stock process. We could also say $\mathcal{F}_2 = (100, 120, 140)$ meaning that $\mathcal{F}_2$ is the path of the stock where $S_0 = 100$, $S_1 = 120$, and $S_2 = 140$.

Conditional expectation: Given a filtration $\mathcal{F}_i$ up to the time i, we can compute expectations with respect to the remaining part of the tree, taking into account the nodes that are still accessible given the history $\mathcal{F}_i$. For the above history (u, u), the stock will be at the node with the value 140. Now only the two remaining values 120 and 160 are accessible. Therefore, we can compute the expectation of S_3 conditioned on the history $\mathcal{F}_2 = (u, u)$. We write $\mathbb{E}_{\mathbb{Q}}(S_3|\mathcal{F}_2 = (u, u))$. In our case, we see that

$$\mathbb{E}_{\mathbb{Q}}(S_3|\mathcal{F}_2 = (u, u)) = \frac{1}{2} \cdot 160 + \frac{1}{2} \cdot 120 = 140.$$

Martingales: It is easy to check that, for the above example, we have

$$\mathbb{E}_{\mathbb{Q}}(S_3|\mathcal{F}_2) = S_2$$

for *any* history $\mathcal{F}_2$. Or, even more general $\mathbb{E}_{\mathbb{Q}}(S_j|\mathcal{F}_i) = S_i$ for $j \geq i$. This property is called the martingale property and a stochastic process Y with the property

$$\mathbb{E}_{\mathbb{Q}}(Y_j|\mathcal{F}_i) = Y_i \tag{33}$$

for $j \geq i$ is called a $\mathbb{Q}$-martingale. Why is the stock process considered above a $\mathbb{Q}$-martingale? The answer is extraordinarily simple: We constructed the measure $\mathbb{Q}$ such that S has exactly this martingale property! Remember that we used the formula

$$q = \frac{s_{now} - s_d}{s_u - s_d}, \qquad s_{now} = q s_u + (1-q) s_d \,.$$

The latter equation ensures that

$$S_k = \mathbb{E}_{\mathbb{Q}}(S_{k+1}|\mathcal{F}_k)$$

such that the martingale property holds at each node of the tree. It is then easy to see, e.g. by induction, that the martingale property holds for the entire tree.

The claim: Let's now consider the claim X, for example the European call $X = (S_3 - K)^+$. Here things are slightly more complicated as X is, at first, only defined on the end nodes of the tree. When filling the option tree, working backwards, we constructed a new stochastic process Y, such that $Y_3 = X$ at the end nodes. At each step, we computed

$$f_{now} = q f_u + (1-q) f_d \,.$$

In other words, we used in fact the conditional expectation operator in order to construct Y, hence

$$Y_i = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_i) \tag{34}$$

For instance, for the filtration $\mathcal{F}_2 = (u, u)$, we have

$$f_{now} = \mathbb{E}_{\mathbb{Q}}\left((S_3 - K)^+ | \mathcal{F}_2 = (u,u)\right) = \frac{1}{2} \cdot 60 + \frac{1}{2} \cdot 20 = 40 \,.$$

In this way, Y is again a martingale, more precisely a $\mathbb{Q}$-martingale as we used the measure $\mathbb{Q}$ in the conditional expectation operator that defines Y.

Binomial Representation Theorem

Construction strategies: So far, the mathematical view of option pricing consists in the following steps (remember, for now, we assume that the interest rate $r = 0$):

1. For the given stock process S, construct a measure $\mathbb{Q}$, such that S is a $\mathbb{Q}$-martingale.

2. Convert a claim X, that is defined at the end nodes of the tree, into a stochastic process Y defined on the same tree as S using the conditional expectation operator

$$Y_i = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_i).$$

3. The option price is $Y_0 = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_0) = \mathbb{E}_{\mathbb{Q}}(X)$.

The final step is to clarify what the replication strategy (the stock and bond holdings in order to hedge the claim) means in mathematical terms. The appropriate interpretation is the following: We have two $\mathbb{Q}$-martingales on the same tree, and, in such a situation, it can be shown that one martingale can be constructed from the other martingale in a *previsible* way, meaning that we always know one step ahead which ϕ to choose (which is important for our hedging strategy). To formulate this more precisely, we prove the following theorem:

Theorem. (Binomial Representation Theorem) *Assume that S is a $\mathbb{Q}$-martingale and V is another $\mathbb{Q}$-martingale on the same tree. Then there exists a previsible process ϕ such that*

$$V_n = V_0 + \sum_{j=0}^{n-1} \phi_{j+1}(S_{j+1} - S_j) \tag{35}$$

Proof. Consider a step from i to $i+1$ where $S_i = s_{now}$ can go to $S_{i+1} = s_u$ or $S_{i+1} = s_d$ and $V_i = f_{now}$ can go to $V_{i+1} = f_u$ or $V_{i+1} = f_d$. Clearly, we can find ϕ_{i+1} and k_{i+1} such that

$$\begin{aligned} f_u - f_{now} &= \phi_{i+1}(s_u - s_{now}) + k_{i+1} \\ f_d - f_{now} &= \phi_{i+1}(s_d - s_{now}) + k_{i+1} \end{aligned}$$

In particular we know at time i the value of ϕ_{i+1} to be

$$\phi_{i+1} = \frac{f_u - f_d}{s_u - s_d}$$

To prove the formula

$$V_{i+1} - V_i = \phi_{i+1}(S_{i+1} - S_i)\,,$$

we need to show that $k_{i+1} = 0$ and to do so, we will make use of the assumptions that the processes S and V are $\mathbb{Q}$-martingales: Since k_{i+1} is known at time i we have

$$\begin{aligned} k_{i+1} &= \mathbb{E}_{\mathbb{Q}}(k_{i+1}|\mathcal{F}_i) \\ &= \mathbb{E}_{\mathbb{Q}}(V_{i+1} - V_i|\mathcal{F}_i) - \phi_{i+1}\mathbb{E}_{\mathbb{Q}}(S_{i+1} - S_i|\mathcal{F}_i) \\ &= 0\,. \end{aligned}$$

The theorem then follows by induction. □

Replication with non-zero interest rates: For the case $r \neq 0$, it can be shown that we need to make only minor modifications to the above application of the binomial representation theorem: Define a process Z given by $Z_i = B_i^{-1} S_i$ (the *discounted* stock process), and choose the measure $\mathbb{Q}$ such that Z is a $\mathbb{Q}$-martingale. Then define a process E using the conditional expectation operator

$$E_i = \mathbb{E}_{\mathbb{Q}}\left(B_T^{-1} X|\mathcal{F}_i\right) \tag{36}$$

and apply the binomial representation theorem such that

$$E_n = E_0 + \sum_{j=0}^{n-1} \phi_{j+1}(Z_{j+1} - Z_j) \tag{37}$$

Brownian Motion as a Martingale

In continuous time, we define a martingale in analogy to the discrete definition: A stochastic process M_t is a martingale with respect to a measure $\mathbb{P}$ (or short $\mathbb{P}$-martingale) if for all $t > s$

$$\mathbb{E}_{\mathbb{P}}(M_t|\mathcal{F}_s) = M_s \tag{38}$$

It is easy to see that a $\mathbb{P}$-Brownian motion is a $\mathbb{P}$-martingale. Intuitively, this is clear as the random walk W_n goes up and down with the same probability. Using the above properties of Brownian motion, we can easily carry out a formal proof (assuming $t > s$):

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}(W_t|\mathcal{F}_s) &= \mathbb{E}_{\mathbb{P}}(W_t - W_s + W_s|\mathcal{F}_s) \\ &= \mathbb{E}_{\mathbb{P}}(W_t - W_s|\mathcal{F}_s) + \mathbb{E}_{\mathbb{P}}(W_s|\mathcal{F}_s) \\ &= 0 + W_s = W_s \end{aligned}$$

In the last step, we used the (trivial) fact that $\mathbb{E}_{\mathbb{P}}(W_s|\mathcal{F}_s) = W_s$ and that $\mathbb{E}_{\mathbb{P}}(W_t - W_s|\mathcal{F}_s) = 0$ since we know that $W_t - W_s \sim N(0, t - s)$. We can also use the properties of Brownian motion in order to carry out more complicated calculations. Consider for example the stochastic process

$$Y_t = W_t^2 - t$$

and we can show that Y_t is also a $\mathbb{P}$-martingale in the following way: First we show that

$$\mathbb{E}_{\mathbb{P}}(W_t^2 - W_s^2 | \mathcal{F}_s) = t - s \tag{39}$$

Proof: By direct computation we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}(W_t^2 - W_s^2 | \mathcal{F}_s) &= \mathbb{E}_{\mathbb{P}}((W_t - W_s)^2 + 2W_s(W_t - W_s) | \mathcal{F}_s) \\ &= \mathbb{E}_{\mathbb{P}}((W_t - W_s)^2 | \mathcal{F}_s) + 2W_s \mathbb{E}_{\mathbb{P}}((W_t - W_s) | \mathcal{F}_s) \\ &= t - s\,. \end{aligned}$$

Again, the last line follows directly from the fact that $W_t - W_s \sim N(0, t-s)$. With the above relation (39) at hand, we find for the process Y_t:

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}(W_t^2 - t | \mathcal{F}_s) &= \mathbb{E}_{\mathbb{P}}(W_t^2 - W_s^2 + W_s^2 - t | \mathcal{F}_s) \\ &= t - s + W_s^2 - t = W_s^2 - s \end{aligned}$$

which shows that $Y_t = W_t^2 - t$ is indeed a martingale. This calculation, however, shows as well that it will be useful to find more sophisticated tests to see whether a process in continuous time is a martingale. We will see that the so-called Ito's Lemma (see below) offers for many processes a quick way to test whether they satisfy the martingale property.

Ito's Lemma

Taylor expansion of the differential: Consider a stochastic differential equation in its general form

$$dX_t = \mu_t dt + \sigma_t dW_t$$

and consider a transformation of the form

$$Y_t = f(X_t)$$

with a smooth function f. We can calculate the differential equation for Y_t by considering the following Taylor expansion:

$$\begin{aligned} \Delta Y &= f(X_t + \Delta X) - f(X_t) = f'(X_t)\Delta X + \frac{1}{2} f''(X_t)(\Delta X)^2 + ... \\ &= f'(X_t)(\mu_t \Delta t + \sigma_t \Delta W) + \frac{1}{2} f''(X_t)(\mu_t \Delta t + \sigma_t \Delta W)^2 + ... \\ &\approx f'(X_t)(\mu_t \Delta t + \sigma \Delta W) + \frac{1}{2} f''(X_t)\, \sigma_t^2 (\Delta t) \end{aligned}$$

The last approximation follows as $(\Delta W)^2 = \Delta t$ and from the fact that we consider terms up to the order Δt. Remember that $|\Delta W| = \sqrt{\Delta t}$. Terms like $(\Delta W)\Delta t$ are of the order $(\Delta t)^{3/2}$ and can be neglected. Therefore, we have the following

Lemma. (Ito's Lemma) *Assume that a stochastic process X_t is the solution of the stochastic differential equation*

$$dX_t = \mu_t dt + \sigma_t dW_t \,.$$

Then, for a transformed process $Y_t = f(X_t)$ with a deterministic, twice continuously differentiable function f, the process Y_t satisfies the stochastic differential equation given by

$$\begin{aligned} dY_t &= f'(X_t)\left(\mu_t dt + \sigma_t dW_t\right) + \frac{1}{2}\sigma_t^2 f''(X_t)dt \\ &= f'(X_t)\sigma_t dW_t + \left(f'(X_t)\mu_t + \frac{1}{2}\sigma_t^2 f''(X_t)\right) dt \,. \end{aligned}$$

Examples of the application of Ito's lemma: We can consider Ito's lemma as a generalization of the deterministic chain rule. In the following we discuss several examples and applications of Ito's lemma.

- Consider the special case, where $X_t = W_t$ and $Y_t = f(W_t)$. Then we have $\mu_t = 0$ and $\sigma_t = 1$ and we see that

$$dY_t = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt \,.$$

- Applying the above formula, we find directly relationships like

$$\begin{aligned} d(W_t^2) &= 2W_t dW_t + dt \\ d(W_t^2 - t) &= 2W_t dW_t \\ d(W_t^3) &= 3W_t^2 dW_t + 3W_t dt \\ d(W_t^6) &= 6W_t^5 dW_t + 15W_t^4 dt \end{aligned}$$

  In particular we see that the solution of the stochastic differential equation $dY_t = 2W_t dW_t$ (together with the initial condition $Y_0 = 0$) is given by $Y_t = W_t^2 - t$.

- An important example is the application to $f(x) = \mathrm{e}^x$. Then, clearly, $f(x) = f'(x) = f''(x)$. Let's assume constant volatility $\sigma_t = \sigma$ and constant drift $\mu_t = \mu$ and define

$$Y_t = f(X_t) = \mathrm{e}^{X_t} = \mathrm{e}^{\sigma W_t + \mu t} \,.$$

This process is called *exponential Brownian motion* and is of relevance in the context of the Black-Scholes model. Applying Ito's lemma, we find that

$$\begin{aligned} dY_t &= f'(X_t)dX_t + \frac{1}{2}\sigma^2 f''(X_t)dt \\ &= Y_t\left(\sigma dW_t + \left(\mu + \frac{1}{2}\sigma^2\right)dt\right) . \end{aligned}$$

In particular we see that for $\mu = -\sigma^2/2$, hence for

$$Y_t = \mathrm{e}^{\sigma W_t - \sigma^2 t/2}$$

we have

$$dY_t = \sigma Y_t dW_t$$

- For two processes X_t and Y_t satisfying the stochastic differential equations

  $$\begin{aligned} dX_t &= \mu_t dt + \sigma_t dW_t \\ dY_t &= \nu_t dt + \rho_t dW_t \end{aligned}$$

  we have for the product $Z_t = X_t Y_t$

  $$d(X_t Y_t) = X_t dY_t + Y_t dX_t + \sigma_t \rho_t dt\,.$$

  This generalizes the product rule to stochastic processes. The proof is left as an exercise. Hint: use a two-dimensional Taylor expansion of the function $f(x,y) = xy$.

Using Ito's lemma to identify martingales: Aside from solving stochastic differential equations, we can also use Ito's lemma to identify martingales. To see why, let's prove first the following (trivial) statement: Consider a stochastic process X_t given by

$$X_t = \sigma W_t + \mu t$$

with constant volatility σ and constant drift μ. Then X_t is a martingale if, and only if, the drift vanishes or $\mu = 0$. The proof is basically a repetition of the proof that Brownian motion is a martingale: Assume X_t is a martingale, then we have (for any $t > s$)

$$\mathbb{E}_{\mathbb{P}}(X_t|\mathcal{F}_s) = X_s\,.$$

As $X_s = \sigma W_s + \mu s$, and using the fact that W_t is a $\mathbb{P}$-martingale, we see that

$$\mathbb{E}_{\mathbb{P}}(X_t|\mathcal{F}_s) = \sigma \mathbb{E}_{\mathbb{P}}(W_t|\mathcal{F}_s) + \mu t = \sigma W_s + \mu t = X_s = \sigma W_s + \mu s$$

or $\mu t = \mu s$, hence $\mu = 0$. The other direction is trivial. The important interpretation of this statement is that, for *arithmetic Brownian motion* $\sigma W_t + \mu t$, we need the drift to vanish in order for the process to be a martingale. This statement is generalized by the following theorem:

Theorem. (Characterization of martingales) *Assume that a stochastic process X_t is the solution of the stochastic differential equation*

$$dX_t = \mu_t dt + \sigma_t dW_t \,.$$

and that the (technical) condition $\mathbb{E}((\int_0^T \sigma_s^2 ds)^{1/2}) < \infty$ is satisfied. Then X_t is a martingale if, and only if, X_t is driftless (hence $\mu_t = 0$).

We will not give a formal proof of this theorem, but taking our intuition from the discrete world, we know that a process on a finite tree is a martingale if the process is a martingale on each branch. Therefore, for a time-tick Δt, we can consider

$$X_{t+\Delta t} = X_t + \Delta X = \begin{cases} \mu_t \Delta t + \sigma_t \sqrt{\Delta t} \\ \mu_t \Delta t - \sigma_t \sqrt{\Delta t} \end{cases}$$

where the probability for an 'up'-jump and an 'down'-jump are 1/2. Therefore, we have

$$\mathbb{E}_{\mathbb{P}}(X_{t+\Delta t}|\mathcal{F}_t) = X_t + \mu_t \Delta t = X_t$$

if X_t is a $\mathbb{P}$-martingale. Hence $\mu_t = 0$ for that time-tick. This local feature of one 'branch' translates to the entire tree, hence $\mu_t = 0$ everywhere.

Going back to one of the examples for Ito's lemma discussed earlier, we can now see immediately that the stochastic process $Y_t = W_t^2 - t$ is a martingale as there is no drift present in $dY_t = 2W_t dW_t$. On the other hand, $Y_t = W_t^2$ is not a martingale since the stochastic differential equation is $dY_t = 2W_t dW_t + dt$ with the drift term dt present. We actually already obtained this result above directly from the properties of Brownian motion. The calculation, however, was more tedious than a simple and direct application of the above theorem in conjunction with Ito's lemma.

In the context of stock models, referring to the exponential Brownian motion we see that

$$Y_t = Y_0 e^{\sigma W_t - \sigma^2 t/2}$$

is a martingale as there is no drift term in the corresponding stochastic differential equation

$$dY_t = \sigma Y_t dW_t \,.$$

This already resembles the representation of the stock process in the martingale measure that we obtained earlier for the Black-Scholes model. The precise meaning will become clear in the next section when we discuss the change of measure and the Girsanov theorem.

The Girsanov Theorem

Radon-Nikodym derivative: Taking again our intuition from the discrete world, we know that, in the context of option pricing, we need to price the claim using the risk-neutral measure. If there are no interest rates, this measure $\mathbb{Q}$ is constructed through the requirement that the stock process S_t needs to be a $\mathbb{Q}$-martingale. This will also work in the continuous world, however, we will need a continuous model of the stock process in the first place. For the Black-Scholes model, we start from

$$S_t = S_0 \mathrm{e}^{\sigma W_t + \mu t}$$

with a $\mathbb{P}$-Brownian motion W_t. In order to define a measure $\mathbb{Q}$, such that S_t becomes a $\mathbb{Q}$-martingale, we need to know how a Brownian motion changes when the measure changes. Or, in other words, we would like to express the change of measure in terms of Brownian motions.

To prepare the change of measure in the continuous world, we go back (only for a moment) to the discrete world. Consider the same tree with two different measures $\mathbb{P}$ and $\mathbb{Q}$.

In the following, we basically only need the rule that we have to multiply probabilities on branches in order to compute the probability to get to a certain node. For instance, using the measure $\mathbb{P}$, the probability π_9 to get to the node numbered node 9 is

$$\pi_{10} = p_1 \cdot p_3 \cdot p_6$$

Under the measure $\mathbb{Q}$, the probability to reach node 9 is in general different. Let's denote it by $\tilde{\pi}_9$. Obviously, we have

$$\tilde{\pi}_{10} = q_1 \cdot q_3 \cdot q_6$$

In this way, we can express probabilities for all nodes. For node 9, we find

$$\begin{aligned} \pi_9 &= p_1 \cdot p_3 \cdot (1 - p_6) + p_1 \cdot (1 - p_3) \cdot p_5 + (1 - p_1) \cdot p_2 \cdot p_5, \\ \tilde{\pi}_9 &= q_1 \cdot q_3 \cdot (1 - q_6) + q_1 \cdot (1 - q_3) \cdot q_5 + (1 - q_1) \cdot q_2 \cdot q_5 \end{aligned}$$

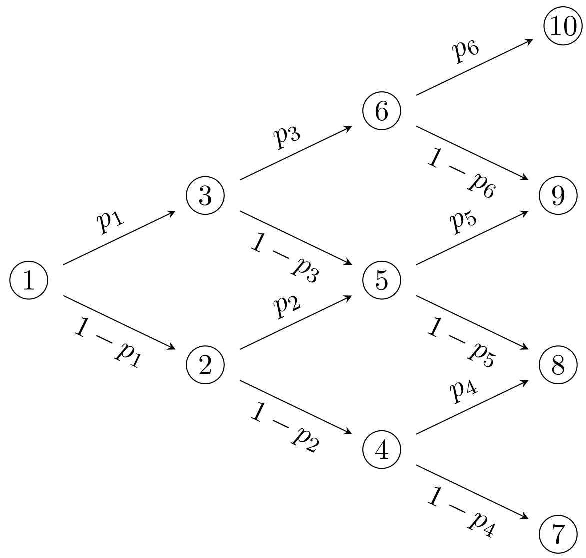


Figure 6: Binomial tree with measure $\mathbb{P}$

After these preparations, we are ready to think about the question under which circumstances it is possible to express the measure $\mathbb{Q}$ in terms of the measure $\mathbb{P}$ and vice versa. The (trivial) answer is to write simply for the probability to reach node 9

$$\tilde{\pi}_9 = \frac{\tilde{\pi}_9}{\pi_9} \cdot \pi_9,$$

in order to go from measure $\mathbb{P}$ to the measure $\mathbb{Q}$ and, on the other hand,

$$\pi_9 = \frac{\pi_9}{\tilde{\pi}_9} \cdot \tilde{\pi}_9 \,.$$

in order to go from measure $\mathbb{Q}$ to the measure $\mathbb{P}$. First, we notice that the likelihood ratio corresponds to a stochastic process. This process is called *Radon-Nikodym derivative* and denoted by $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ or $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$. Clearly, for the Radon-Nikodym derivative to be well-defined, we need to assume that nodes of the tree that are accessible under the measure $\mathbb{Q}$ are also accessible under the measure $\mathbb{P}$. In other words: we need to avoid dividing by zero when forming the likelihood ratios. The formal definition is given by the *equivalence* of the two measures: The two measures are equivalent if for each set A the statement $\mathbb{Q}(A) > 0$ is equivalent to $\mathbb{P}(A) > 0$. For the binomial

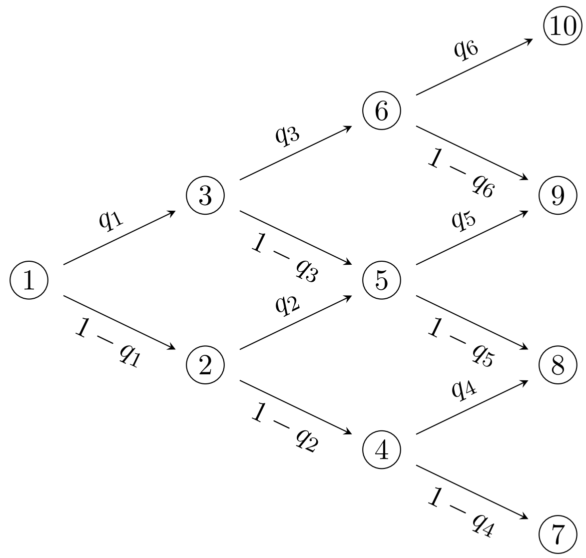


Figure 7: Binomial tree with measure $\mathbb{Q}$

tree, this reduces to the statement that, nodes that are accessible under one measure are also accessible under the other measure. By 'accessible' we mean simply that there is a non-zero probability to reach a node.

For equivalent measures, we can easily express the expectation value with respect to one measure through the expectation value taken with respect to the other measure. Consider a discrete random variable X, we find

$$\mathbb{E}_{\mathbb{P}}(X) = \sum_i x_i \pi_i = \sum_i x_i \frac{\pi_i}{\tilde{\pi}_i} \tilde{\pi}_i = \mathbb{E}_{\mathbb{Q}}\left(X \frac{d\mathbb{P}}{d\mathbb{Q}} \right) \tag{40}$$

For a normally distributed random variable, we can use this formula to characterize the change of measure. This is due to the following basic fact:

Theorem. (Characterization of Gaussian variables)
The following two statements are equivalent:

1. *A random variable X has a normal distribution $N(\mu, \sigma^2)$ under a measure $\mathbb{P}$*

2. *For all real θ, we have*

$$\mathbb{E}_{\mathbb{P}}\left(\mathrm{e}^{\theta X}\right) = \mathrm{e}^{\theta\mu + \theta^2\sigma^2/2}$$

Let's look at a simple, but very instructive example: In order to define a suitable Radon-Nikodym derivative, a simple choice is to have $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} > 0$. Let W_t be a $\mathbb{P}$-Brownian motion (hence normally distributed with zero mean and a variance t), a possible choice would be

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \mathrm{e}^{-\gamma W_t - \gamma^2 t/2} > 0\,.$$

Can we figure out what happens to W_t under the measure $\mathbb{Q}$? If we are lucky, W_t has still a normal distribution under the new measure (but maybe with a different mean and/or variance). In this case, we should be able to use the above theorem. Let's try to compute

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\left(\mathrm{e}^{\theta W_t}\right) &= \mathbb{E}_{\mathbb{P}}\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\mathrm{e}^{\theta W_t}\right) = \mathbb{E}_{\mathbb{P}}\left(\mathrm{e}^{-\gamma W_t - \gamma^2 t/2}\mathrm{e}^{\theta W_t}\right) \\
&= \mathrm{e}^{-\gamma^2 t/2}\mathbb{E}_{\mathbb{P}}\left(\mathrm{e}^{(\theta-\gamma)W_t}\right) = \mathrm{e}^{-\gamma^2 t/2}\mathrm{e}^{(\theta-\gamma)^2 t/2} \\
&= \mathrm{e}^{-\gamma\theta t + \theta^2 t/2}
\end{aligned}$$

Applying the above theorem again, we see that W_t has, under the measure $\mathbb{Q}$, again a normal distribution with mean $-\gamma t$ and variance t. In summary, we have shown that, for a $\mathbb{P}$ Brownian motion W_t, we can construct an equivalent measure $\mathbb{Q}$ such that, under the new measure $\mathbb{Q}$, the Brownian motion W_t has a mean $-\gamma t$ and a drift t. In particular $\tilde{W}_t = W_t + \gamma t$ is actually a $\mathbb{Q}$-Brownian motion, since the term γt is used to compensate for the negative drift. This is a special case of the following theorem:

Theorem. (Cameron-Martin-Girsanov theorem)
For a $\mathbb{P}$-Brownian motion W_t and a previsible process γ_t, satisfying the condition

$$\mathbb{E}_{\mathbb{P}}\left(\exp\left(\frac{1}{2}\int_0^T \gamma_t^2\,dt\right)\right) < \infty$$

there exists a measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that

$$\tilde{W}_t = W_t + \int_0^t \gamma_s\,ds$$

is a $\mathbb{Q}$-Brownian motion. The measures are related by the Radon-Nikodym derivative given by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\left(-\int_0^t \gamma_s\,dW_s - \frac{1}{2}\int_0^t \gamma_s^2\,ds\right)\,.$$

In the context of derivative pricing, we can use the Cameron-Martin-Girsanov theorem in order to construct a martingale measure. Consider for example a stochastic process $X_t = \mu t + \sigma W_t$ with a $\mathbb{P}$-Brownian motion W_t. Now, we can write

$$X_t = \sigma \left(\frac{\mu}{\sigma} t + W_t \right) = \sigma(\gamma t + W_t) = \sigma \tilde{W}_t$$

since the Cameron-Martin-Girsanov theorem guarantees the existence of an equivalent measure $\mathbb{Q}$ and the corresponding $\mathbb{Q}$-Brownian motion $\tilde{W}_t$. Clearly, X_t is not a $\mathbb{P}$-martingale, but it is obviously a $\mathbb{Q}$-martingale.

Derivation of the Black-Scholes formula using the Girsanov theorem: We can now derive again the Black-Scholes formula using the Cameron-Martin-Girsanov theorem. The main step consists in considering the Black-Scholes model with a stock and bond process given by

$$S_t = S_0 \, \mathrm{e}^{\mu t + \sigma W_t}, \qquad B_t = B_0 \mathrm{e}^{rt}$$

and forming the discounted stock process $Z_t = B_t^{-1} S_t = S_0 \, \mathrm{e}^{(\mu - r)t + \sigma W_t}$. For a European claim X with maturity T, the initial price is given by $V = \mathrm{e}^{-rT} \mathbb{E}_{\mathbb{Q}}(X)$, and Z_t is a $\mathbb{Q}$-martingale. Clearly, we want to use the Cameron-Martin-Girsanov theorem to construct $\mathbb{Q}$. Therefore, we use first Ito's lemma to find dZ_t:

$$dZ_t = Z_t \left((\mu - r + \sigma^2/2) \, dt + \sigma dW_t \right) = \sigma Z_t \left(\gamma \, dt + dW_t \right) , \qquad (41)$$

where we set $\gamma = \mu - r + \sigma^2/2$. Applying now the Cameron-Martin-Girsanov theorem, we can construct the measure $\mathbb{Q}$ such that Z_t becomes a $\mathbb{Q}$-martingale and $\tilde{W}_t = \gamma t + W_t$ is a $\mathbb{Q}$-Brownian motion. Clearly, we have then $W_t = \tilde{W}_t - \gamma t$ and we can use this in order to express the stock process S_t in terms of $\tilde{W}_t$:

$$S_t = S_0 \, \mathrm{e}^{\mu t + \sigma W_t} = S_0 \, \mathrm{e}^{\mu t + \sigma \tilde{W}_t - (\mu - r)t - \sigma^2 t/2} = S_0 \, \mathrm{e}^{\sigma \tilde{W}_t + (r - \sigma^2/2)t} \qquad (42)$$

For a European call, we have $X = (S_T - K)^+$, and thus the price of such a call is computed as

$$V = \mathrm{e}^{-rT} \mathbb{E}_{\mathbb{Q}}(X) = \mathbb{E}_{\mathbb{Q}} \left(\left(S_0 \, \mathrm{e}^{\sigma \tilde{W}_t - \sigma^2 t/2} - K \mathrm{e}^{-rT} \right)^+ \right) \qquad (43)$$

which is exactly the Black-Scholes formula.

The martingale representation theorem in the continuous world: Let's for a moment look again at the simple case of zero-interest rates. Remember that, in the discrete world, the binomial representation theorem would allow us to construct a self-financing hedging strategy to replicate our claim: First, we construct a measure $\mathbb{Q}$, such that the stock process S on the tree is a $\mathbb{Q}$-martingale. Then, in the second step, we convert the claim X into a process E, using the conditional expectation operator $E_i = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_i)$. At each step, we have

$$\Delta E_i = \phi_i \Delta S_i, \qquad E_i = E_0 + \sum_{k=1}^{i} \phi_k \Delta S_k\,.$$

In other words, at time-tick i, we need ϕ_{i+1} units of the stock S and, therefore, $\psi_{i+1} = E_i - \phi_{i+1} S_i$ units of the bond. At time zero, our portfolio has the value $\phi_1 S_0 + \psi_1 = E_0 = \mathbb{E}_{\mathbb{Q}}(X)$ which is the money we need to create it (the price of the derivative). And, trivially, at time k, our portfolio will have the value E_k such that, at maturity, it replicates the claim. For the case with interest rates, all we needed to do was to define the *discounted* stock process $Z_i = B_i^{-1} S_i$ and use this process Z to find the martingale measure $\mathbb{Q}$. And we would consider the discounted claim to define the process E via $E_i = \mathbb{E}_{\mathbb{Q}}(B_T^{-1} X|\mathcal{F}_i)$. The value of the claim X at time-tick i was then given by

$$V_i = B_i E_i = B_i \mathbb{E}_{\mathbb{Q}}(B_T^{-1} X|\mathcal{F}_i)\,. \tag{44}$$

All these ideas carry over to the continuous world. Again, let us first write down the continuous version of the martingale representation theorem:

Theorem. (Martingale representation theorem)
Given a $\mathbb{Q}$-martingale M whose volatility is always non-zero and any other $\mathbb{Q}$-martingale N, there exists a previsible process ϕ such that N can be written as

$$N_t = N_0 + \int_0^t \phi_s \, dM_s$$

Using this theorem, we can proceed in the continuous world exactly in the same way as in the discrete world: First, we find a measure $\mathbb{Q}$, such that the stock process S_t is a $\mathbb{Q}$-martingale (again, we have $r = 0$). Then, convert the claim X into a process via $E_t = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t)$. Now, apply the martingale representation theorem to construct a previsible process ϕ_t such that

$$dE_t = \phi_t \, dS_t, \qquad E_t = E_0 + \int_0^t \phi_s dS_s$$

Again, the hedging strategy will consist in holding ϕ_t units of the stock at time t and $\psi_t = E_t - \phi_t S_t$ units of the bond, such that the value of the portfolio at time t will be $V_t = \phi_t S_t + \psi_t$. What happens if we have an interest rate $r > 0$? Similar to the discrete world, we need to consider now the *discounted* stock process $Z_t = B_t^{-1} S_t$ and use this process to find the martingale measure $\mathbb{Q}$. This is exactly what was done as an example at the end of the last section for the Black-Scholes model. The value of the claim X at time t is then - compare to (44)

$$V_t = B_t E_t = B_t \mathbb{E}_{\mathbb{Q}}(B_T^{-1} X | \mathcal{F}_t) \,. \tag{45}$$

For the standard bond process $B_t = B_0 \mathrm{e}^{rt}$ we find

$$V_t = B_0 \mathrm{e}^{rt} \mathbb{E}_{\mathbb{Q}}(B_0^{-1} \mathrm{e}^{-rT} X | \mathcal{F}_t) = \mathrm{e}^{-r(T-t)} \mathbb{E}_{\mathbb{Q}}(X | \mathcal{F}_t) \tag{46}$$

The Black-Scholes Model

We already noted in the section introducing the change of measure as an example how to derive the Black-Scholes formula using the Cameron-Martin-Girsanov theorem.

$$\begin{aligned} V &= \mathrm{e}^{-rT} \mathbb{E}_{\mathbb{Q}}(X) = \mathbb{E}_{\mathbb{Q}}\left(\left(S_0 \, \mathrm{e}^{\sigma \tilde{W}_t - \sigma^2 t/2} - K \mathrm{e}^{-rT}\right)^+\right) \\ &= S_0 \Phi\left(\frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}\right) \\ &\quad - K \, \mathrm{e}^{-rT} \, \Phi\left(\frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}\right) \end{aligned}$$

Clearly, (if the claim X, as always assumed only depends on the final value of the stock S_T), for a time $0 < t < T$, we can write the value $V_t = V(S_t, t)$ for $s = S_t$ as

$$V(s,t) = \mathrm{e}^{-r(T-t)} \mathbb{E}_{\mathbb{Q}}(X | S_t = s) \tag{47}$$

and, concerning the trading strategy, the number of shares that one needs to hold to hedge the claim at time t is given by $\phi_t = \partial V / \partial s$ which is the continuous version of $\phi = (f_u - f_d)/(s_u - s_d)$ valid in the discrete world. With a little bit of algebra, one can prove that

$$\phi_t = \frac{\partial V}{\partial s}(S_t, T-t) = \Phi\left(\frac{\ln(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{(T-t)}}\right) \tag{48}$$

A second proof for the formula $\phi_t = \partial V/\partial s$ comes from the fact that the hedging strategy is self-financing, meaning that we have

$$dV_t = \phi_t dS_t + \psi_t dB_t \tag{49}$$

Under the martingale measure, the SDE for the stock process yields

$$dS_t = \sigma S_t d\tilde{W}_t + rS_t dt \tag{50}$$

and, for the bond process, we have as always $dB_t = rB_t dt$. Together, this yields the following representation of dV_t:

$$dV_t = \sigma S_t \phi_t d\tilde{W}_t + (rS_t\phi_t + r\psi_t B_t)\, dt \tag{51}$$

Now we can derive a second representation of dV_t using Ito's lemma:

$$dV_t = dV(S_t, t) = \left(\sigma S_t \frac{\partial V}{\partial s}\right) d\tilde{W}_t + \left(rS_t \frac{\partial V}{\partial s} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2} + \frac{\partial V}{\partial t}\right) dt \tag{52}$$

Comparing the two expressions, we find

$$\phi_t = \frac{\partial V}{\partial s}, \qquad rs\frac{\partial V}{\partial s} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2} + \frac{\partial V}{\partial t} = rV \tag{53}$$

The latter equation is a *partial differential equation* that can be solved (with the appropriate final condition) in order to obtain the option price.

Exercises

1. (10 points) An investor agrees to sell insurance for a portfolio of 100 identical mortgages against defaults. Assume independence and that the random variable X counting the number of defaults has a binomial distribution with $p = 0.1$. The investor agrees to pay a flat fee of \$10M if X is between 20 and 30.

   (a) Compute $P(X = 8)$, μ_X, and σ_X.

   (b) Find $P(20 \leq X \leq 30)$.

   (c) What would be the 'fair' price for the insurance contract?

   (d) How does this price change if $p = 0.11$?

2. (5 points) Show that the following formula holds for a random variable X with a Gaussian distribution (mean μ and variance σ^2), and for a real number θ:
$$\mathbb{E}\left(\mathrm{e}^{\theta X}\right) = \exp\left(\theta\mu + \frac{1}{2}\theta^2\sigma^2\right)$$

3. (5 points) Show that if B_t is a zero-volatility process and X_t is any stochastic process, then
$$d(B_t X_t) = B_t dX_t + X_t dB_t$$

4. (5 points) Use Ito's formula and the rule above to check whether the following process
$$Y_t = W_t^3 - 3tW_t$$
is a martingale.

5. (5 points) What is the solution of
$$dX_t = X_t(\sigma dW_t + \mu \sin(t) dt), \qquad X_0 = a > 0$$
where σ and μ are assumed to be constants?

6. (5 points) Consider the stochastic differential equation (SDE)
$$dX_t = \sigma dW_t + \mu dt, \qquad X_0 = a > 0$$
where σ and μ are constants.

   (a) Write down the solution X_t of this SDE.

(b) Find mean and variance of X_t.

(c) Write down the probability density $p(x,t)$ of X_t.

7. (10 points) Consider a stock $S_t = S_0 e^{\sigma W_t}$, with a $\mathbb{P}$-Brownian motion W_t.

   (a) Show that S_t is not a $\mathbb{P}$-martingale.

   (b) Use the Girsanov theorem to construct a measure $\mathbb{Q}$ such that S_t is a $\mathbb{Q}$-martingale. Express S_t in terms of a $\mathbb{Q}$-Brownian motion $\tilde{W}_t$.

   (c) Assume $\sigma = 0.2$, $S_0 = \$10$. No interest rates. What is the value of a bet that pays \$20 if the stock is worth less than \$8 after two years?

8. (10 points) Show that $s\Phi'(d_1) = k e^{-r(T-t)}\Phi'(d_2)$ when

$$\begin{aligned} d_1 &= \frac{\log(s/k) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \\ d_2 &= \frac{\log(s/k) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \end{aligned}$$

   Use this result to show that

$$\phi_t = \frac{\partial V}{\partial s}(S_t, T-t) = \Phi\left(\frac{\log(S_t/k) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right)$$