Mathematics of Financial Derivatives
Part II

Bernd Pape
Department of Mathematics and Statistics
University of Vaasa
e-mail: bepa@uwasa.fi
url: www.uwasa.fi/∼bepa/

0. A Short Review of Probability Theory

Probability Spaces

The sample space $\Omega = \{\omega_1, \ldots, \omega_k\}$ *is the set of all possible outcomes* ω_i, $i = 1, 2, ..., k$ *of an experiment. Subsets of* Ω *are called events.*

A function $P : \Omega \to \mathbb{R}$ *is then defined as a probability measure (for short probability), if the following conditions hold:*

(P1) $P(A) \geq 0$ *for all* $A \subset \Omega$

(P2) $P(\Omega) = 1$

(P3) $A, B \subset \Omega,\ A \cap B = \emptyset$ *then*
$$P(A \cup B) = P(A) + P(B)$$

The following properties of P *follow then by application of elementary set theory (exercise):*
a) $P(A^C) = 1 - P(A)$,
b) $P(\emptyset) = 0$,
c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Note that in order for such relations to be well defined, we must be sure that the <u>*empty set*</u> $\emptyset$*, the* <u>*complement*</u> A^C*, the* <u>*union*</u> $A \cup B$ *and the* <u>*intersection*</u> $A \cap B$ *are subsets of the sample space* Ω *as well. In other words, the collection of events must be closed under set operations. A collection of subsets of* Ω *that is closed under set operations is called an algebra. Formally we define it as follows:*

A collection $\mathcal{F} = F_1, F_2, \ldots, F_n$ *of subsets of* $\Omega = \{\omega_1, \ldots, \omega_k\}$ *is called an* <u>*algebra*</u> *on* Ω *iff:*

(A1) $\Omega \in \mathcal{F}$

(A2) $F \in \mathcal{F} \;\Rightarrow F^C \in \mathcal{F}$

(A3) $F_i, F_j \in \mathcal{F} \Rightarrow F_i \cup F_j \in \mathcal{F}$

The empty set $\emptyset$ *and the intersection* $F_i \cap F_j$ *are then also included in* $\mathcal{F}$ *(exercise). Probability is then defined as a function* $P : \mathcal{F} \to \mathbb{R}$ *having properties (P1)-(P3) with* $A, B \subset \Omega$ *replaced by* $A, B \subset \mathcal{F}$*. The triple* $(\Omega, \mathcal{F}, P)$ *is called a* <u>*probability space*</u>*.*

Sigma-fields

If Ω is a general continuous space consisting of infinitely many events, then the definitions of an algebra and a probability measure need to be generalized as follows:

A collection $\mathcal{F} = F_1, F_2, \ldots$ of subsets of Ω is called a σ-algebra or σ-field on Ω iff:

(S1) $\quad \Omega \in \mathcal{F}$

(S2) $\quad F \in \mathcal{F} \;\Rightarrow F^C \in \mathcal{F}$

(S3) $\quad F_1, F_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} \in \mathcal{F}$

Probability is then defined as a function $P : \mathcal{F} \rightarrow \mathbb{R}$ satisfying:

(P1) $\quad P(A) \geq 0$ *for all* $A \in \mathcal{F}$

(P2) $\quad P(\Omega) = 1$

(P3) $\quad A_1, A_2, ... \in \mathcal{F},\; A_i \cap A_j = \emptyset \;\forall i \neq j$

$$\Rightarrow P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Conditional Probability

When modelling asset prices over several periods, we are often in a situation that the sample space varies through time. For example, if we want to model a stock price a certain time period ahead, then the possible range of prices that we may reasonably expect the stock to attain, is certainly much larger within a years time than within the next few seconds. The basic building block for time-varying sample spaces is conditional probability defined as:

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \quad \textit{assuming } P(B) \neq 0$$

The conditional probability $P(A|B)$ may be regarded as an ordinary probability with B as new sample space. This may be seen as *follows:*

(P1) $P(A|B) \geq 0,$ *because*
$P(A \cap B) \geq 0$ *and* $P(B) > 0$

(P2) $$P(B|B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

(P3) *Let* A_1, A_2 *be mutually exlusive events, that is* $A_1, A_2 \in \mathcal{F}$ *and* $A_1 \cap A_2 = \emptyset$, *then:*

$$\begin{aligned}
P(A_1 \cup A_2|B) &= \frac{P((A_1 \cup A_2) \cap B)}{P(B)} \\
&= \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} \\
&= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} \\
&= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} \\
&= P(A_1|B) + P(A_2|B)
\end{aligned}$$

Events A *and* B *are called* *<u>independent</u>*, *iff* $P(A \cap B) = P(A)P(B)$. *This implies that the conditional probabilities* $P(A|B)$ *and* $P(B|A)$

reduce to their corresponding unconditional probabilities, as then, for example:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

Suppose that $\mathcal{P} = \{A_1, A_2, \ldots, A_n\}$ forms a partition of the sample space Ω, that is:

$A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$, *and* $A_i \cap A_j = \emptyset \; \forall \; i \neq j$.

Then the law of total probability states that for any $B \subset \Omega$:

$$P(B) = P(A_1)P(B|A_1) + \ldots + P(A_n)P(B|A_n)$$

This may be seen as follows:

$$\begin{aligned}
&P(A_1)P(B|A_1) + \ldots + P(A_n)P(B|A_n) \\
\stackrel{\text{(Def)}}{=} &P(A_1)\frac{P(B \cap A_1)}{P(A_1)} + \ldots + P(A_n)\frac{P(B \cap A_n)}{P(A_n)} \\
\stackrel{\text{(P3)}}{=} &P((B \cap A_1) \cup \ldots \cup (B \cap A_n)) \\
= &P(B \cap (A_1 \cup \ldots \cup A_n)) \\
\stackrel{\text{(Def)}}{=} &P(B \cap \Omega) = P(B)
\end{aligned}$$

Information Structure

Time varying partitions are a convenient tool to describe how we learn more about the outcome of an experiment over time. The partition $\mathcal{P}_t$ at time point t splits then the sample space Ω according to the outcomes that are still possible given the information gathered so far.

Example 1

Suppose a stock price S_t starts at $S_0 = 100$ and may in the next two periods $t = 1, 2$, either fall or rise by 50 currency units. Denote for convenience:

$$\begin{aligned}
\omega_1 &= \{S_0 = 100, S_1 = 50, S_2 = 0\} \\
\omega_2 &= \{S_0 = 100, S_1 = 50, S_2 = 100\} \\
\omega_3 &= \{S_0 = 100, S_1 = 150, S_2 = 100\} \\
\omega_4 &= \{S_0 = 100, S_1 = 150, S_2 = 200\}
\end{aligned}$$

The sample space is then $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$.

As an information tree the structure is

$$\begin{tikzcd}[row sep=small]
 & & S_2 = 200 \quad \omega_4 \\
 & S_1 = 150 \arrow[ur] \arrow[dr] & \\
 & & S_2 = 100 \quad \omega_3 \\
S_0 = 100 \arrow[uur] \arrow[ddr] & & \\
 & & S_2 = 100 \quad \omega_2 \\
 & S_1 = 50 \arrow[ur] \arrow[dr] & \\
 & & S_2 = 0 \quad \omega_1 \\
t = 0 & t = 1 & t = 2
\end{tikzcd}$$

Our information at each time point $t = 1, 2$ *is reflected by the partitions:*

$$\mathcal{P}_0 = \{S_0 = 100\} = \{\omega_1, \omega_2, \omega_3, \omega_4\} = \Omega$$
$$\mathcal{P}_1 = \{\{S_1 = 50\}, \{S_1 = 150\}\} = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$$
$$\mathcal{P}_2 = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$$

Note that the partition at each time point is obtained as a further subdivision of the preceeding partition. This must be so for a consistent description of a learning process.

For suppose that some subset of the partition would increase over time. Then, depending upon wheter such an increased subset is consistent with the corresponding state of the experiment, this would correspond to either forgetting a part of the history or even an inconsistent description of the experiment.

Example 1(continued)

Consider: $\mathcal{P}_2' = \{\{\omega_1\}, \{\omega_2, \omega_3\}, \{\omega_4\}\}$.
This would enable us to correctly infer the possible end values $S_2 = 0$, $S_2 = 100$, *and* $S_2 = 200$. *But with knowledge of the partition alone (that is without checking the individual elements inside its subsets) we could not tell whether the final state* $S_2 = 100$ *has been obtained via* $S_1 = 50$ *or via* $S_1 = 150$.

Consider now: $\mathcal{P}_2'' = \{\{\omega_1, \omega_4\}, \{\omega_2\}, \{\omega_3\}\}$.
This is obviously an inconsistent description of the experiment, because based upon the outcome of S_1 *we can already rule out one of*

the states ω_1 *and* ω_4*, and in this case we can even seperate them based upon the value of* S_2*, since* $S_2(\{\omega_1\}) = 0$ *and* $S_2(\{\omega_4\}) = 200$*.*

If the conditional probabilities given the events in the preceeding partition are known, we may use the law of total probability in order to obtain the (unconditional) probability of an event at each time step.

<u>Example 1</u>(continued)

Suppose the stock price S_t *is equally likely to rise or to fall at each time step, that is:*

$$
\begin{aligned}
&P(\{S_1 = 50\}|\{S_0 = 100\}) \\
=&P(\{S_1 = 150\}|\{S_0 = 100\}) \\
=&P(\{S_2 = 0\}|\{S_1 = 50\}) = \ldots = \frac{1}{2}
\end{aligned}
$$

Then, by the law of total probability:

$$\begin{aligned}
&P(\{S_1 = 50\})\\
=&P(\{S_1 = 50\}|\{S_0 = 100\})P(\{S_0 = 100\})\\
=&\frac{1}{2} \cdot 1 = \frac{1}{2},
\end{aligned}$$

$$\begin{aligned}
&P(\{S_1 = 150\})\\
=&P(\{S_1 = 150\}|\{S_0 = 100\})P(\{S_0 = 100\})\\
=&\frac{1}{2} \cdot 1 = \frac{1}{2},
\end{aligned}$$

$$\begin{aligned}
&P(\{S_2 = 0\})\\
=&P(\{S_2 = 0\}|\{S_1 = 50\})P(\{S_1 = 50\})\\
+&P(\{S_2 = 0\}|\{S_1 = 150\})P(\{S_1 = 150\})\\
=&\frac{1}{2} \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{4},
\end{aligned}$$

$$\begin{aligned}
&P(\{S_2 = 100\})\\
=&P(\{S_2 = 100\}|\{S_1 = 50\})P(\{S_1 = 50\})\\
+&P(\{S_2 = 100\}|\{S_1 = 150\})P(\{S_1 = 150\})\\
=&\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2},
\end{aligned}$$

$$P(\{S_2 = 200\}) = P(\{S_2 = 0\}) = \frac{1}{4}.$$

Note that to every partition there exists a unique algebra consisting of all unions of its sets and their complements. We say then that a partition $\mathcal{P}$ generates the algebra $\mathcal{F}$. Conversely, we may obtain the generating partition of $\mathcal{F}$ as the collection of the smallest nonempty sets in $\mathcal{F}$, that is for finite spaces:

$$\mathcal{P} = \{A \in \mathcal{F} : A \neq \emptyset,\ A \cap B = A \text{ or } \emptyset \quad \forall B \in \mathcal{F}\}$$

Example 1(continued)

The algebras generated by the partitions $\mathcal{P}_0$, $\mathcal{P}_1$, and $\mathcal{P}_2$ are:

$$\begin{aligned}
\mathcal{F}_0 &= \{\emptyset, \Omega\}, \\
\mathcal{F}_1 &= \{\emptyset, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega\}, \\
\mathcal{F}_2 &= \{\emptyset, \{\omega_1\}, \{\omega_2, \omega_3, \omega_4\}, \{\omega_2\}, \{\omega_1, \omega_3, \omega_4\}, \\
&\qquad \{\omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_4\}, \{\omega_1, \omega_2, \omega_3\}, \\
&\qquad \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_4\} \\
&\qquad \{\omega_2, \omega_3\}, \{\omega_1, \omega_4\}, \Omega\}
\end{aligned}$$

Note in the above example that every set contained in an algebra $\mathcal{F}_t$ is also contained *in the following algebra $\mathcal{F}_{t+1}$, since each partition $\mathcal{P}_t$ has been obtained by a further subdivision of the preceeding partition $\mathcal{P}_{t-1}$. So each algebra is a subset of the following algebra (written as $\mathcal{F}_{t-1} \subset \mathcal{F}_t$).*

A sequence of algebras $\mathbb{F} = \{\mathcal{F}_t : t = 0, 1, \dots T\}$ having the property $\mathcal{F}_{t-1} \subset \mathcal{F}_t$ for all $t = 0, 1, \dots T$ is called a <u>filtration</u>. Obviously, filtrations are good candidates for describing the learning process of investors regarding the evolution of security prices over time.

Random Variables

A random variable X is often loosely defined as any function from the sample space Ω to the real line $\mathbb{R}$ ($X : \Omega \to \mathbb{R}$). However, if we wish to associate probabilities with the values of a random variable, say $P(X = x) := P(\{X = x\})$, we must first make sure that the event $\{X = x\}$ is actually contained in some algebra $\mathcal{F}$ for all values x the random variable X may attain.

For that reason we define a random variable X on the probability space $(\Omega, \mathcal{F}, P)$ as an invertible function $X : \Omega \to \mathbb{R}$ with the important property that it is measurable with respect to algebra $\mathcal{F}$ (for short $\mathcal{F}$-measurable, denoted as $X \in \mathcal{F}$), that is:

$$\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}, \text{ for all } x \in \mathbb{R}.$$

This means that whichever value x the random variable X attains, the corresponding *event ω giving rise to this value through the transformation $X(w) = x$, is contained in the algebra $\mathcal{F}$; thereby allowing us to perform all the set operations we need in order to have a well defined probability.*

We may then define an <u>induced probability</u> for any range of values $X' \subset \mathbb{R}$ the random variable may obtain as:

$$P(X \in X') := P(\{\omega \in \Omega : X(\omega) \in X'\}.$$

<u>Example 1</u>(continued)

The stock prices $S_t,\ t = 0, 1, 2$ may all be regarded as random variables with induced probabilities:

$$P(S_0 = 100) = P(\{S_0 = 100\}) = 1$$
$$P(S_1 = 50) = P(\{S_1 = 50\}) = \frac{1}{2}$$

...and so on.

The probability density $f_X(x)$ denotes the probability $P(X = x)$ for discrete random variables (that is, for r.v.'s attaining only a countable number of values), while for continuous random variables (that is, continuous functions $X : \Omega \to I\!R$), it denotes the limit

$$f_X(x) = \lim_{h \to 0} P(X < x + h) - P(X \leq x - h) = \frac{dF_X(x)}{dx},$$

where $F_X(x) := P(X \leq x)$ denotes the cumulative distribution function of X.

The expected value of X, denoted $\mu_X = \mathsf{E}(X)$, is defined as

$$\mu_X = \begin{cases} \int_{-\infty}^{\infty} x f_X(x)\,dx & \text{for } X \text{ continous}, \\ \sum_{x:X(\omega)=x} xP(X = x) & \text{for } X \text{ discrete.} \end{cases}$$

The expected value has the following properties:

(E1) $\quad E(c) = c,\quad$ *if c is a constant,*

(E2) $\quad E(aX + bY) = aE(X) + bE(Y),$ *if a, b are constants and X, Y are random variables.*

Two discrete random variables X and Y are called <u>*independent*</u> *if for all $x, y \in \mathbb{R}$:*

$$P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\} P(\{Y = y\})$$

In that case we have additionally:

(E3) $E(XY) = E(X)E(Y)$ *for X, Y indep.*

The <u>*variance*</u> *of X is defined as*

$$\sigma_X^2 = \mathsf{V}(X) := \mathsf{E}(X - \mu_X)^2$$

It has the following properties:

(V1) $\mathsf{V}(X) = \mathsf{E}(X^2) - \mathsf{E}(X)^2$

(V2) $\mathsf{V}(c) = 0$, *if c is a constant*

(V3) $\mathsf{V}(aX + bY) = \ldots$

$= a^2\mathsf{V}(X) + b^2\mathsf{V}(Y) + 2ab\mathsf{Cov}(X, Y)$

for a, b constants and X, Y r.v.'s

with <u>*Covariance*</u>

$\mathsf{Cov}(X, Y) := \mathsf{E}[(X - \mu_X)(Y - \mu_Y)]$

$\ldots = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$

The Set Indicator Function

A particularly easy and useful random variable for our purpose is the indicator function of set A defined as:

$$\mathbb{I}_A := 1 \text{ if } \omega \in A, \text{ and } 0 \text{ otherwise.}$$

It is important for us because we can use it to describe, to which member of a generating partition the true state belongs.

The indicator function has the following important property:

$$\begin{aligned} E(\mathbb{I}_A) &= \sum_{x:\mathbb{I}_A(\omega)=x} xP(\mathbb{I}_A = x) \\ &= 0 \cdot P(\mathbb{I}_A = 0) + 1 \cdot P(\mathbb{I}_A = 1) \\ &= P(\{\omega \in A\}) = P(A) \end{aligned}$$

1. Stochastic Processes and Martingales

Stochastic Processes

A sequence of random variables $X_1, X_2, \ldots, X_T$ *defined on the same sample space is called a stochastic process. In other words, a stochastic process* $S(t) := S(t, \omega)$ *may be defined as a real valued function*

$$S(t, \omega) : \{0, 1, \ldots, T\} \times \Omega \to \mathbb{R}.$$

Keeping t *fixed yields the random variable*

$$S_t(\omega) : \Omega \to \mathbb{R}.$$

Keeping ω *fixed yields the sample path*

$$S_\omega(t) : \{0, 1, \ldots, T\} \to \mathbb{R}.$$

A stochastic process $S = \{S(t) : t = 0, 1, \ldots, T\}$ *is said to be adapted to the filtration* $\mathbb{F} = \{\mathcal{F}_t : t = 0, 1, \ldots, T\}$*, if the random variable* $S_t(\omega)$ *is* $\mathcal{F}_t$*-measurable for every* $t = 0, 1, \ldots, T$*.*

Example 1(continued)

The sequence of stock prices S_0, S_1, S_2 may be regarded as realizations of a stochastic process $S = \{S(t) : t = 0, 1, 2\}$. S is adapted to the filtration $I\!\!F = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\}$, since all stock prices S_0, S_1, S_2 are measurable with respect to their corresponding algebras, for example S_1:

$$\{\omega \in \Omega : S_1(\omega) = 50\} = \{\omega_1, \omega_2\} \in \mathcal{F}_1$$
$$\{\omega \in \Omega : S_1(\omega) = 150\} = \{\omega_3, \omega_4\} \in \mathcal{F}_1$$

Note, that because $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$, S_1 is also $\mathcal{F}_2$-measurable, but not $\mathcal{F}_0$-measurable. That is

$$\{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\} \subset \mathcal{F}_2, \text{ but}$$
$$\{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\} \not\subseteq \mathcal{F}_0,$$

because when we get to know S_1 at time point $t = 1$, then we know it still at $t = 2$, but do not know it yet at $t = 0$. This is why we may call a stochstic process, which is adapted to $I\!\!F = \{\mathcal{F}_t\}$, observable with respect to information set $\mathcal{F}_t$ or $\{\mathcal{F}_t\}$.

Conditional Expectations

In analogy to the unconditional expectation as the sum

$$E(X) = \sum_{x:X(\omega)=x} xP(X = x)$$

we may define the conditional expectation of X given the event A (in the discrete case) as:

$$E(X|A) = \sum_{x:X(\omega)=x} xP(X = x|A)$$

As discussed earlier, conditional probabilities may be regarded as ordinary probabilities on the reduced sample space consisting of the event which it is conditioned upon. This implies, that the conditional expectation $E(X|A)$ may be regarded as an ordinary expectation on the reduced sample space A.

In particular, it has the properties (E1) and (E2) of an ordinary expected value:

(E1) $E(c|A) = c,$ *if c is a constant,*
(E2) $E(aX + bY|A) = aE(X|A) + bE(Y|A),$
if a, b are constants and X, Y
are random variables.

As a special case we may choose the expectation of a random variable X conditional upon the event that another random variable Y attains a certain value y, that is:

$$E(X|\{Y = y\}) = \sum_{x:X(\omega)=x} xP(X = x|\{Y = y\})$$

Doing this for all values y which the random variable Y may attain, we may extend the definition of the conditional expectation of X given the event $\{Y = y\}$ to the <u>conditional expectation of X given the random variable Y</u> as

$$E(X|Y)(\omega) := E(X|\{Y(\omega) = y\}),\ \forall y : Y(\omega) = y$$

Note that conditioning on a random variable rather than on a single event implies that $E(X|Y)(\omega)$ *(or* $E(X|Y)$*, for short), unlike* $E(X|A)$*, is a random variable itself, as it depends on* ω *just like* $Y(\omega)$ *does. This implies that the induced probability of* $E(X|Y)$ *is*

$$P(E(X|Y)(\omega)) = P(Y = y), \quad \forall y : Y(\omega) = y$$

In our attempt to model stock prices and what investors think about them, it is of particular interest to calculate the expected value of a stock price given its own history, that is conditional upon the particular set of the generating partition $\mathcal{P}_t$*, that we experience at the moment. In other words, denoting the sets of the generating partition* $\mathcal{P}_t$ *as* $A_{t,1}, \ldots, A_{t,n}$ *and the corresponding indicator functions* $\mathbb{I}_{A_{t,1}}, \ldots, \mathbb{I}_{A_{t,n}}$ *we are interested in conditional expectation of the type* $E(X|\mathbb{I}_{A_{t,i}}) = E(X|A_{t,i})$ *if* $A_{t,i}$ *is the true state.*

We may do this simulstaneously for all sets $A_{t,i},\ i = 1,\dots n$ of the generating partition $\mathcal{P}_t$ by introducing a random variable

$$I_t := i,\ i = 1,\dots n, \quad \text{for } \mathbb{I}_{A_{t,i}} = 1$$

and defining the conditional expectation of X given information set $\mathcal{F}_t$ as

$$E_t(X)(\omega) = E(X|\mathcal{F}_t)(\omega) := E(X|I_t)(\omega),$$

where I_t has been defined above and $\mathcal{F}_t$ is the algebra generated by $\mathcal{P}_t$.

The definition above for discrete random variables can be shown to be consistent with the following general definition of $E(X|\mathcal{F}_t)$ as the unique random variable satisfying:
a) $E(X|\mathcal{F}_t) \in \mathcal{F}_t, \quad (E_t(X)$ is $\mathcal{F}_t$-measurable)
b) $E(E(X|\mathcal{F}_t)\mathbb{I}_{A_t}) = E(X\mathbb{I}_{A_t}), \quad \forall A_t \in \mathcal{F}_t$.

Example 1(continued)

We shall now calculate the probability distributions of $E(S_2|\mathcal{F}_i)(\omega)$ for $i = 0, 1, 2$.

The only partitioning set of $\mathcal{F}_0$ *is* $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$*. Therefore:*

$$E(S_2|\mathcal{F}_0) = E(S_2|\Omega) = \sum_x xP(\{S_2(\omega) = x\}|\Omega)$$
$$= \sum_x x\frac{P(\{S_2(\omega) = x\} \cap \Omega)}{P(\Omega)} = \sum_x x\frac{P(\{S_2(\omega) = x\})}{1}$$
$$= \sum_x xP(S_2 = x) = \mathrm{E}(S_2) = 0 \cdot \frac{1}{4} + 100 \cdot \frac{1}{2} + 200 \cdot \frac{1}{4}$$
$$=100,$$
$$P(E(S_2|\Omega)) = P(\Omega) = 1.$$

$\mathcal{P}_1 = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\},$ *therefore:*

$$E(S_2|\{\omega_1, \omega_2\}) = E(S_2|S_1 = 50)$$
$$=0 \cdot P(S_2 = 0|S_1 = 50) + 100 \cdot P(S_2 = 100|S_1 = 50)$$
$$=0 \cdot \frac{1}{2} + 100 \cdot \frac{1}{2} = 50,$$
$$P(E(S_2|\{\omega_1, \omega_2\})) = P(\{\omega_1, \omega_2\}) = P(S_1 = 50) = \frac{1}{2}.$$

Similarly:

$$E(S_2|S_1 = 150) = 150,\ P(E(S_2|S_1 = 150)) = \frac{1}{2}.$$

For calculating the probability distributions of $E(S_2|\mathcal{F}_2)(\omega)$, note that for $i = 1, 2, 3, 4$:

$$\begin{aligned} E(S_2|\omega_i) &= xP(\{S_2(\omega_i) = x\}|\omega_i) \\ =&S_2(\omega_i)P(\omega_i|\omega_i) = S_2(\omega_i), \end{aligned}$$

and:

$$P(E(S_2|\omega_i)) = P(\omega_i) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Therefore:

$$\begin{aligned} &E(S_2|\{\omega_1\}) = 0, && P(E(S_2|\{\omega_1\}) = \frac{1}{4}; \\ &E(S_2|\{\omega_2\}) = 100, && P(E(S_2|\{\omega_2\}) = \frac{1}{4}; \\ &E(S_2|\{\omega_3\}) = 100, && P(E(S_2|\{\omega_3\}) = \frac{1}{4}; \\ &E(S_2|\{\omega_4\}) = 200, && P(E(S_2|\{\omega_4\}) = \frac{1}{4}. \end{aligned}$$

We may deduce from the example above, that the conditional expectation with respect to the samllest possible algebra $\mathcal{F}_0 = \{\emptyset, \Omega\}$ always equals the unconditional expectation, that is $E_0(S_t) = \mathsf{E}(S_t)$.

Furthermore we deduce that the conditional expectation with respect to the largest possible algebra at endpoint T equals the random variable itself, that is $E_T(S_T) = S_T$.

Finally, it appears from the example above, that it doesn't matter whether we forecast S_2 directly by calculating it its expected value $E_0(S_2) = \mathsf{E}(S_2)$, or whether we try to forecast it indirectly by forecasting its conditional expectation at $t = 1$, since:

$$E_0(E_1(S_2)) = \frac{1}{2} \cdot 50 + \frac{1}{2} \cdot 150 = 100 = E_0(S_2).$$

This result holds in general, since by the law of total probability we have that:

$$
\begin{aligned}
E_0(X) = \mathsf{E}(X) &= \sum_{x:X(\omega)=x} xP(X=x) \\
&= \sum_{x:X(\omega)=x} x \left(\sum_{i=1}^{n} P(X=x|I_t=i)P(I_t=i) \right) \\
&= \sum_i \left(\sum_x xP(X=x|I_t=i) \right) \cdot P(I_t=i) \\
&= \sum_i E(X|I_t=i) \cdot P(I_t=i) \\
&= \mathsf{E}(E(X|I_t)) = E_0(E_t(X))
\end{aligned}
$$

This result may be generalized further by recalling that a conditional expectation given some event A may be always thought of as an unconditional expectation with A representing the whole sample space. Reverting this logic, we may then regard the unconditional expectation in some sample space Ω as a conditional expectation in an even larger sample space Ω' (such that $\Omega \subset \Omega'$), given the event Ω. In other words, we may generalize the result obtained above:

$$
\mathsf{E}(X) = E(X|\mathcal{F}_0) = E(E(X|\mathcal{F}_t)|\mathcal{F}_0) = E_0(E_t(X))
$$

for $\mathcal{F}_0 \subset \mathcal{F}_1$ into:

$$E_1(X) = E(X|\mathcal{F}_1) = E(E(X|\mathcal{F}_2)|\mathcal{F}_1) = E_1(E_2(X)).$$

for $\mathcal{F}_1 \subset \mathcal{F}_2$. This important result is called law of iterated expectations. It can be shown, that the order of indexes in the iterated conditional expectation does not matter, that is:

$$E(E(X|\mathcal{F}_1)|\mathcal{F}_2) = E(E(X|\mathcal{F}_2)|\mathcal{F}_1) = E(X|\mathcal{F}_1)$$

Example 1(continued)

We verify the law of iterated expectations for S_2 conditional on $\mathcal{F}_1$ and $\mathcal{F}_2$:

$$E(E(S_2|\mathcal{F}_1)|\mathcal{F}_2) = E(E(S_2|\mathcal{F}_2)|\mathcal{F}_1) = E(S_2|\mathcal{F}_1)$$

The probability distribution of $E(S_2|\mathcal{F}_1)$ has already been calculated above as:

$E(S_2\|\mathcal{F}_1)(\omega)$	50	150
$P(E(S_2\|\mathcal{F}_1))$	1/2	1/2

Let us now calculate the probability distribution of $E(E(S_2|\mathcal{F}_2)|\mathcal{F}_1)$*:*

$$\begin{aligned}
&E(E(S_2|\mathcal{F}_2)|\{\omega_1,\omega_2\})\\
=&0\cdot P(S_2=0|S_1=50)+100\cdot P(S_2=100|S_1=50)\\
=&0\cdot\frac{1}{2}+100\cdot\frac{1}{2}=50,\ \textit{and similarly:}\\
&E(E(S_2|\mathcal{F}_2)|\{\omega_3,\omega_4\})=100\cdot\frac{1}{2}+200\cdot\frac{1}{2}=150
\end{aligned}$$

The probability distribution of $E(E(S_2|\mathcal{F}_2)|\mathcal{F}_1)$ *is therefore:*

$E(E(S_2\|\mathcal{F}_2)\|\mathcal{F}_1)(\omega)$	*50*	*150*
$P(E(E(S_2\|\mathcal{F}_2)\|\mathcal{F}_1))$	*1/2*	*1/2*

For calculating the probability distribution of $E(E(S_2|\mathcal{F}_1)|\mathcal{F}_2)(\omega)$*, note that for any* $B\subset A$*:*

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(B)}{P(B)}=1$$

We have therefore:

$$E(E(S_2|\mathcal{F}_1)|\{\omega_1\}) = 50 \cdot P(\{\omega_1, \omega_2\}|\{\omega_1\}) = 50$$
$$E(E(S_2|\mathcal{F}_1)|\{\omega_2\}) = 50 \cdot P(\{\omega_1, \omega_2\}|\{\omega_2\}) = 50$$
$$E(E(S_2|\mathcal{F}_1)|\{\omega_3\}) = 150 \cdot P(\{\omega_3, \omega_4\}|\{\omega_3\}) = 150$$
$$E(E(S_2|\mathcal{F}_1)|\{\omega_4\}) = 150 \cdot P(\{\omega_3, \omega_4\}|\{\omega_4\}) = 150$$

all with probability $P(\{\omega_i\}) = \frac{1}{4}$, *such that:*

$E(E(S_2\vert\mathcal{F}_1)\vert\mathcal{F}_2)(\omega)$	*50*	*150*
$P(E(E(S_2\vert\mathcal{F}_1)\vert\mathcal{F}_2))$	*1/2*	*1/2*

Other properties of $E(X|\mathcal{F}_t)$

The following properties of the conditional expectation with respect to an algebra $\mathcal{F}_t$ ensure that we may take random variables out of the conditional expectation, if they are known at time t (that is $\mathcal{F}_t$-measurable).

Specifically, let X_1, X_2 and Y_1, Y_2 be random variables, of which X_1 and X_2 are $\mathcal{F}_t$-measurable, that is known at time t. We have then:

$$E(X_1Y_1+X_2Y_2|\mathcal{F}_t) = X_1E(Y_1|\mathcal{F}_t)+X_2E(Y_2|\mathcal{F}_t)$$

Furthermore, if $X \in \mathcal{F}_t$, then $E(X|\mathcal{F}_t) = X$.

Finally, if X is independent of X, that is:
$P(\{X = x\}|F_t) = P(X = x)\forall x \in \mathbb{R},\ F_t \in \mathcal{F}_t$,
then: $E(X|\mathcal{F}_t) = \mathsf{E}(X)$.
This means that we may replace the conditional with the unconditional expectation, if X is unrelated to the information structure generated by the filtration $\mathbb{F} = \{\mathcal{F}_0, \mathcal{F}_1, \ldots \mathcal{F}_T\}$.

Martingales

Let $S = \{S(t) : t = 0, 1, \ldots, T\}$ *be a stochastic process with sample space* Ω *and probability measure* P*, adapted to the filtration* $\mathbb{F} = \{\mathcal{F}_t : t = 0, 1, \ldots, T\}$.

Then S *is called a:*
–*martingale* if $E(S_\omega(t+1)|\mathcal{F}_t) = S_\omega(t)$,
–*supermartingale* if $E(S_\omega(t+1)|\mathcal{F}_t) \leq S_\omega(t)$,
–*submartingale* if $E(S_\omega(t+1)|\mathcal{F}_t) \geq S_\omega(t)$,
for all $\omega \in \Omega$ *and for all* $t < T$.

Note that if S *is a martingale:* $E_t(S_{t+1}) = S_t$, *we have by the law of iterated expectations:*

$$S_{t-1} = E_{t-1}(S_t) = E_{t-1}(E_t(S_{t+1})) = E_{t-1}(S_{t+1})$$

Therefore, by applying this procedure s *times:*

$$E(S_\omega(t+s)|\mathcal{F}_t) = S_\omega(t) \text{ for all } s \geq 0,\ \omega \in \Omega$$

That is, the current value of a martingale is the best forecast of all future values.

Similarly we have for supermartingales

$$E(S_\omega(t+s)|\mathcal{F}_t) \le S_\omega(t) \text{ for all } s \ge 0,\ \omega \in \Omega,$$

and for submartingales

$$E(S_\omega(t+s)|\mathcal{F}_t) \ge S_\omega(t) \text{ for all } s \ge 0,\ \omega \in \Omega.$$

Inserting the definition of a martingale into the unconditional expectation and applying the law of iterated expectations yields:

$$E(S_t) = E(E_{t-1}(S_t)) = E(S_{t-1}) = \ldots = E(S_0)) = S_0,$$

so a martingale is "constant on average".

Similarly we obtain for supermartingales

$$E(S_t) = E(E_{t-1}(S_t)) \le E(S_{t-1}) = \ldots \le E(S_0) = S_0,$$

and for submartingales

$$E(S_t) = E(E_{t-1}(S_t)) \ge E(S_{t-1}) = \ldots \ge E(S_0) = S_0.$$

So supermartingales "decline on average" and submartingales "increase on average".

If S is a martingale, then the conditional expectation of the martingale difference, $\Delta S_{t+1} := S(t+1) - S(t)$, *is zero, because:*

$$E_t(\Delta S_{t+1}) = E_t(S_{t+1}) - E_t(S_t) = S_t - S_t = 0$$

Example 1(continued)

$S = \{S(t) : t = 0, 1, 2\}$ *is a martingale, since:*

$$E_0(S_1) = \frac{1}{2} \cdot 50 + \frac{1}{2} \cdot 150 = 100 = S_0, \text{ and}$$

$$E_1(S_2|S_1 = 50) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 100 = 50 = S_1,$$

$$E_1(S_2|S_1 = 150) = \frac{1}{2} \cdot 100 + \frac{1}{2} \cdot 200 = 150 = S_1.$$

Accordingly: $\mathsf{E}(S_0) = \mathsf{E}(S_1) = \mathsf{E}(S_2) = 100$.

Example 2

The driftless random walk,

$$S_t = S_{t-1} + \epsilon_t, \; E_t(\epsilon_{t+1}) = 0 \; \forall \omega \in \Omega, t > 0$$

is a martingale, because:

$$E_t(S_{t+1}) = E_t(S_t) + E_t(\epsilon_{t+1}) = S_t.$$

Example 3

The random walk with drift,

$S_t = \mu + S_{t-1} + \epsilon_t,\ E_t(\epsilon_{t+1}) = 0\ \forall \omega \in \Omega,\ t > 0$

is a submartingale (respectively supermartingale) for $\mu \geq 0$ (respectively $\mu \leq 0$):

$$E_t(S_{t+1}) = \mu + S_t \gtreqless S_t, \quad \textit{for } \mu \gtreqless 0.$$

Example 4

If $S = \{S(t) : t = 0, 1, \ldots, T\}$ is adapted to the filtration $\mathbb{F} = \{\mathcal{F}_t : t = 0, 1, \ldots, T\}$, then the stochastic process defined by the sequence of conditional expectations $M_t := E(S_T|\mathcal{F}_t) = E_t(S_T)$ is a martingale, beause by the law of iterated expectations:

$$E_t(M_{t+1}) = E_t E_{t+1}(S_T) = E_t(S_T) = M_t.$$

Note the dependence of the martingale property upon the probability measure P in the probability spaces $(\Omega, \mathcal{F}_t, P)$, $t = 0, 1, \ldots T$. This will be illustrated in the following:

<u>Example 5</u>

Let $S_t = \begin{cases} S_{t-1} + 1 \text{ with probability } p; \\ S_{t-1} - 1 \text{ with probability } (1-p). \end{cases}$

$$\begin{aligned} \Rightarrow \quad E_t(S_{t+1}) &= (S_t + 1) \cdot p + (S_t - 1) \cdot (1 - p) \\ &= S_t \cdot (p + 1 - p) + 1 \cdot p - 1 \cdot (1 - p) \\ &= S_t + (2p - 1) \end{aligned}$$

So S_t will be a martingale for $p = 1/2$, a submartingale for $p \geq 1/2$, and a supermartingale for $p \leq 1/2$.

Continuous-Time Martingales

The concept of a martingale may be extended into continuous time as follows:

Let $\mathcal{F}$ be a σ-algebra and $\{\mathcal{I}_t \subset \mathcal{F}, t \geq 0\}$ a continuous time filtration, that is an increasing family of sub-sigma-fields $\mathcal{I}_t \subset \mathcal{I}_{t+h}$, for all $t \geq 0, h > 0$. A stochastic process $\{S_t, t \geq 0\}$ is a martingale with respect to the family of information sets $\mathcal{I}_t$ and probability measure P, if for all $t > 0$,

(M1) S_t *is $\mathcal{I}_t$-adapted (known, given $\mathcal{I}_t$),*

(M2) $\mathrm{E}(|S_t|) < \infty$ *(forecasts are finite),*

(M3) $E_t(S_{t+u}) := E(S_{t+u}|\mathcal{I}_t) = S_t \; \forall u > 0$
(latest observation is best forecast of all future observations);

where all expectations are assumed to be taken with respect to the probability measure P.

Super- and submartingales are analogously defined by replacing (M3) with the condition $E_t(S_{t+u}) \leq S_t$ for supermartingales and $E_t(S_{t+u}) \geq S_t$ for submartingales.

We have then for martingale differences, like in the discrete case ($\Delta S_{t+u} := S_{t+u} - S_t$):

$$E_t(\Delta S_{t+u}) = E_t(S_{t+u}) - E_t(S_t) = S_t - S_t = 0,$$

that is, future changes in S_t are unpredictable.

Example 6

We shall soon introduce (generalized) Brownian motion as a continuous time process, starting at 0, with normally distributed and uncorrelated increments, that is $X_0 = 0$,

$$\Delta X_t := X_t - X_{t-\Delta t} \sim N(\mu\Delta t, \sigma^2\Delta t) \ \forall t, \Delta t,$$

and $\mathsf{E}[(\Delta X_u - \mu\Delta t)(\Delta X_s - \mu\Delta t)] = 0 \ \forall \Delta t,\ u \neq s.$

We have then:

X_t is known at time t ($\mathcal{I}_t$-adapted). (M1)
$\mathsf{E}(|X_t|) = \mu t < \infty$ *for any finite t. (M2)*

$$
\begin{aligned}
\text{(M3)} \qquad E_t(X_{t+u}) &= E_t(X_t + (X_{t+u} - X_t)) \\
&= E_t(X_t) + E_t(\Delta X_{t+u}) \\
&= X_t + \mu u, \ \textit{ for all } u > 0.
\end{aligned}
$$

So X_t is a martingale for $\mu = 0$ and a sub/ (super)-martingale for $\mu \geq 0$ ($\mu \leq 0$).

Note that even when $\mu \neq 0$, it is easy to transform X_t into a martingale by subtracting the deterministic function μt, that is after defining a new process $Z_t := X_t - \mu t$ we get:

$$
\begin{aligned}
E_t(Z_{t+u}) &= E_t(X_{t+u} - \mu(t+u)) \\
&= E_t(X_{t+u}) - \mu(t+u) \\
&= X_t + \mu u - \mu t - \mu u \\
&= X_t - \mu t = Z_t
\end{aligned}
$$

Doop-Meyer Decomposition

The above example was our first illustration of the Doop-Meyer Decomposition, that is the decomposition of a submartingale (for $\mu > 0$) into a deterministic trend and a martingale component.

In order to understand the conditions under which this can be done, we need to shortly discuss the properties of the sample paths (trajectories) $S_\omega(t)$ of a stochastic process S.

We say that a martingale S_t is continuous, if its increments $\Delta S_t := S_{t+u} - S_t$ satisfy for $u \to 0$:

$$P(\Delta S_t > \epsilon) \to 0, \text{ for all } \epsilon > 0,$$

meaning that its sample paths $S_\omega(t)$ are continuous (with probability one).

If the martingale contains jumps which are not too concentrated, then it may still be *written as a right-continuous martingale, that is a martingale whose increments ΔS_t obey the same condition*

$$P(\Delta S_t > \epsilon) \to 0, \text{ for all } \epsilon > 0,$$

where the limiting sequence $u \to 0$ is confined to $u > 0$ (also written as $u \downarrow 0$ or $u \to 0+$), that is, the sample paths are continuous from the right, but not necessarily from the left.

So every continuous martingale is obviously also right-continuous, whereas the converse does not hold.

We are now in the position to state a set of sufficient conditions under which we may split up a submartingale into a deterministic trend and a martingale component as follows:

Theorem. Let X_t be a right-continuous submartingale w.r.t filtration $\{\mathcal{I}_t\}$, and $\mathsf{E}[X_t] < \infty$ for all t, then X_t admits the decomposition

$$X_t = M_t + A_t,$$

where M_t is a right-continuous martingale and A_t is an increasing $\mathcal{I}_t$-measurable process.

Note that the theorem stated above implies that every discrete process $X'_{t_i} \in \{I_{t_i}\}$ with $E(X'_{t_i}) < \infty$ may be decomposed into a deterministic trend and a martingale component, since it may always be regarded as a continuous process that remains constant until immediately before the next timestep, that is:

$$X(t_i + u_i) = X'_{t_i}, \text{ for } \; t_i \le t_i + u_i < t_{i+1}.$$

Example 5 *(continued)*

Let $S_t = \begin{cases} S_{t-1} + 1 \text{ with probability } p; \\ S_{t-1} - 1 \text{ with probability } (1-p). \end{cases}$

Earlier we found that

$$E_t(\Delta S_{t+1}) = E_t(S_{t+1} - S_t) = 2p - 1.$$

$$\Rightarrow E_t(\Delta S_{t+u}) = E_t(S_{t+u} - S_t) = E_t\left(\sum_{t'=t+1}^{t+u} \Delta S_{t'+1}\right)$$

$$= \sum_{t'=t+1}^{t+u} E_t(\Delta S_{t'+1}) = \sum_{t'=t+1}^{t+u} (2p-1) = (2p-1)u$$

Therefore: $E_t(S_{t+u}) = S_t + (2p-1)u$.

Define a new process: $Z_t := S_t + (1-2p)t$

$$\begin{aligned}\Rightarrow\ E_t(Z_{t+u}) &= E_t(S_{t+u} + (1-2p)(t+u)) \\ &= E_t(S_{t+u}) + (1-2p)(t+u) \\ &= S_t + (2p-1)u + (1-2p)t + (1-2p)u \\ &= S_t + (1-2p)t = Z_t\end{aligned}$$

Example 7

Assume generalized Brownian motion without drift, implying:

$$\Delta X_{t+u} := X_{t+u} - X_t \sim N(0, \sigma^2 u)\ \forall\ t, u \geq 0$$

and consider the squared process $S_t := X_t^2$.

$$\begin{aligned}E_t(\Delta S_{t+u}) &= E_t(X_{t+u}^2 - X_t^2) = E_t((X_t + \Delta X_{t+u})^2 - X_t^2) \\ &= E_t(X_t^2 + 2X_t \Delta X_{t+u} + (\Delta X_{t+u})^2 - X_t^2) \\ &= 2X_t E_t(\Delta X_{t+u}) + E_t(\Delta X_{t+u}^2) = \sigma^2 u \neq 0\end{aligned}$$

Define a detrended process: $Z_t := S_t - \sigma^2 t$

$$\begin{aligned}\Rightarrow\ E_t(Z_{t+u}) &= E_t(S_{t+u} - \sigma^2(t+u)) \\ &= E_t(S_t + \Delta S_{t+u} - \sigma^2 t - \sigma^2 u) \\ &= S_t + \sigma^2 u - \sigma^2 t - \sigma^2 u = Z_t.\end{aligned}$$

2. Modelling Randomness in Asset Prices

Suppose a stock price $S(t \in [0,T])$ is observed at n equidistant time points $0 = t_0 < t_1 < \ldots < t_n = T$ of length $h = T/n$, that is $t_k - t_{k-1} = h,\ k = 1,...,n$, such that $S_k := S(t_k) = S(kh)$.

Denote the price innovation ΔW_k as the unpredictable component of the price change $\Delta S_k = S(kh) - S((k-1)h)$ that is:

$$\Delta W_k := (S_k - S_{k-1}) - E_{k-1}(S_k - S_{k-1}),$$

where $E_{k-1}(\cdot)$ denotes the conditional expectation with respect to information set $\mathcal{I}_{k-1}$.

Note the following properties of ΔW_k:

1. ΔW_k is unpredictable given I_{k-1}, because:

$$\begin{aligned} E_{k-1}(\Delta W_k) &= E_{k-1}(S_k - S_{k-1}) - E_{k-1}(S_k - S_{k-1}) \\ &= 0, \end{aligned}$$

2. ΔW_k is a <u>martingale difference</u>, that is the accumulated innovation process $W_k := \sum_{i=1}^{k} \Delta W_i$ is a martingale:

$$\begin{aligned} E_{k-1}(W_k) &= E_{k-1}(\Delta W_1 + \ldots + \Delta W_{k-1} + \Delta W_k) \\ &= E_{k-1}(\Delta W_1 + \ldots + \Delta W_{k-1}) + E_{k-1}(\Delta W_k) \\ &= \Delta W_1 + \ldots + \Delta W_{k-1} + 0 = W_{k-1}. \end{aligned}$$

3. As a consequence, price innovations are <u>uncorrelated</u>:

$$\begin{aligned} &\mathsf{Cov}(\Delta W_i, \Delta W_{i+j}) = \mathsf{E}(\Delta W_i \Delta W_{i+j}) - \mathsf{E}(\Delta W_i)\mathsf{E}(\Delta W_{i+j}) \\ =&\mathsf{E}E_{i+j-1}(\Delta W_i \Delta W_{i+j}) = \mathsf{E}(\Delta W_i E_{i+j-1}(\Delta W_{i+j})) \\ =&\mathsf{E}(\Delta W_i \cdot 0) = 0 \end{aligned}$$

The variance V_k of the price innovations ΔW_k is $V_k := \mathsf{E}(\Delta W_k^2)$, because $E(\Delta W_k) = 0$. The variance of the accumulated innovation process $W_k = \sum_{i=1}^{k} \Delta W_i$ at time $t_k = kh$ is

$$V := \mathsf{V}\left(\sum_{i=1}^{k} \Delta W_i\right) = \sum_{i=1}^{k} \mathsf{V}(\Delta W_i) = \sum_{i=1}^{k} V_i,$$

because the increments of ΔW_i are uncorrelated (see (V3)).

Merton(1990) showed that V_k is proportional to h under the following mild assumptions:

(A1): *$V > A_1 > 0$ with A_1 independent of n, meaning that there will always be some variation or randomness involved, no matter at how small intervals we observe the price process (volatility is bounded from below by A_1).*

(A2): *$V < A_2 < \infty$ with A_2 independent of n, meaning that observing the price process more and more often does not increase volatility to such an extend that it becomes infinitely large (volatility is bounded from above by A_2).*

$$\text{(A3)}: \; V_k > A_3 \cdot V_{max}, \; 0 < A_3 < 1, \; \textit{and}$$
$$V_{max} = \max\{V_1, \ldots, V_n\} \; \textit{indep. of } n,$$

meaning that volatility is not concentrated in such a way that any of the subintervals would be left without volatility.

Under these mild assumptions, the variance V_k of the price innovations ΔW_k will be proportional to the length of the observation interval h, where the constant of proportionality may depend upon the subinterval k, that is, $V_k = E(\Delta W_k^2) = \sigma_k^2 h$, which we now show:

Combining (A2) and (A3) yields:

$$A_2 \overset{(A2)}{>} V \overset{(Def)}{=} \sum_{k=1}^{n} V_k \overset{(A3)}{>} nA_3 V_{max} \overset{(Def)}{>} nA_3 V_k$$

$$\Rightarrow \qquad V_k < \frac{A_2}{nA_3} = \frac{A_2 h}{A_3 T},$$

that is V_k is bounded from above by a line proportional to h.

On the other hand, we know from (A1):

$$nV_{max} \overset{(Def)}{>} V \overset{(Def)}{=} \sum_{k=1}^{n} V_k \overset{(A1)}{>} A_1 \Rightarrow V_{max} > \frac{A_1}{n}$$

This may be combined with (A3) as follows:

$$V_k > A_3 \cdot V_{max} \Rightarrow V_k > \frac{A_3 A_1}{n} = \frac{A_3 A_1}{T} \cdot h$$

So V_k is also bounded from below by a line proportional to h, that is:

$$\frac{A_3 A_1}{T} \cdot h < V_k < \frac{A_2}{A_3 T} \cdot h.$$

This implies that for each k we can find a constant $\sigma_k^2 \in \left[\frac{A_3 A_1}{T}, \frac{A_2}{A_3 T}\right]$, such that:

$$V_k = E(\Delta W_k^2) = \sigma_k^2 h.$$

The Wiener Process

We saw in Example 2, that the symmetric random walk $S_t = S_{t-1} + \epsilon_t, \quad E_t(\epsilon_t) = 0$ *is a martingale. We can extend this into a continuous-time process with the required properties of price innovations as follows.*

Suppose that accumulated price changes $W(t \in [0, T])$, *to be observed at* n *equidistant time points* $0 = t_0 < t_1 < \ldots < t_n = T$, *of length* $h = T/n$ *(that is,* $t_k - t_{k-1} = h$, $k = 1, \ldots, n$*), are modelled as the sum of* k *independent price innovations* ΔW_i*, that is,*

$$W_k := W(t_k) = W(kh) = \sum_{i=1}^{k} \Delta W_i,$$

where each of the price innovations ΔW_i *has the following probability distribution,*

$$\Delta W_i = \pm\sqrt{h} \quad \text{with probability} \quad \frac{1}{2}.$$

The expected value and variance of the price innovations are:

$$\mathsf{E}(\mathit{\Delta} W_i) = \frac{1}{2}\cdot\sqrt{h} + \frac{1}{2}\cdot(-\sqrt{h}) = 0,$$

$$\mathsf{V}(\mathit{\Delta} W_i) = \mathsf{E}(\mathit{\Delta} {W_i}^2) = \frac{(\sqrt{h})^2}{2} + \frac{(-\sqrt{h})^2}{2} = h.$$

The expected value and variance of the accumulated price innovations at time t_k are then:

$$\mathsf{E}(W(t_k)) = \sum_{i=1}^{k} \mathsf{E}(\mathit{\Delta} W_i) = 0,$$

$$\mathsf{V}(W(t_k)) = \sum_{i=1}^{k} \mathsf{V}(\mathit{\Delta} W_i) = kh.$$

Now, in order to obtain a process that is continuous in time, divide each interval into finer and finer subintervals, that is, let $h \to 0$, or $n \to \infty$ for the whole time span T.

We have then, by virtue of the central limit theorem:

$$W(T) = \sum_{i=1}^{n} \mathit{\Delta} W_i \quad \xrightarrow{d} \quad N(0, nh) = N(0, T)$$

where $\xrightarrow{d}$ *denotes convergence in distribution for* $n \to \infty$*, also called weak convergence.*

Since we could have used any time interval $[t_1, t_2] \subset [0, T]$ *to apply the same procedure, we have also that:*

$$W(t_2) - W(t_1) \xrightarrow{d} N(0, t_2 - t_1), \quad 0 \le t_1 \le t_2 \le T.$$

To summarize, we have constructed a process with the following properties:

(B1)*:* $W(0) = 0$*,*
(B2)*:* $W(t)$ *is continuous,*
(B3)*: Increments of* $W(t)$ *are independent, that is, if* $0 \le t_0 < \ldots < t_n$*, then* $W(t_1) - W(t_0), \ldots, W(t_n) - W(t_{n-1})$ *are independent;*
(B4)*: If* $0 \le t_1 \le t_2$*,*
then: $W(t_2) - W(t_1) \sim N(0, t_2 - t_1)$*.*

A stochastic process with properties (B1)–(B4) *is called (standard)* *Brownian motion.*

(Standard) Brownian motion turns out to be equivalent to the following definition of a Wiener process.

A Wiener process W_t relative to a family of information sets $\{\mathcal{I}_t\}$ (filtration) is a stochastic process satisfying:

(W1)*:* $W_0 = 0$,
(W2)*:* W_t *is continuous,*
(W3)*:* W_t *is adapted to the filtration* $\{\mathcal{I}_t\}$,
(W4)*: For* $s \leq t$, $W_t - W_s$ *is independent of* $\mathcal{I}_s$, *with* $\mathsf{E}(W_t - W_s) = 0$ *and* $\mathsf{V}(W_t - W_s) = t - s$.

Note that although the definition of a Wiener process makes no explicit statement about the distribution of its increments $W_t - W_s$, *we know that they are normally distributed with mean 0 and variance* $t - s$ *from the equivalence of Brownian motion and the Wiener process. In the following we shall use the terms Brownian motion and Wiener process interchangeably.*

The Wiener process W_t has the following important properties:

a) *W_t is a martingale w.r.t. filtration $\{\mathcal{I}_t\}$, because by application of* (W4) *for $s \leq t$:*
$E_s(W_t - W_s) = \mathsf{E}(W_t - W_s) = 0$
$\Rightarrow \; E_s(W_t) = E_s(W_s) = W_s$;

b) *W_t has independent increments by property* (B3) *of Brownian Motion;*

c) $\mathsf{E}(W_t) \stackrel{(W1)}{=} \mathsf{E}(W_t - W_0) \stackrel{(W4)}{=} 0$;

d) $\mathsf{V}(W_t) \stackrel{(W1)}{=} \mathsf{V}(W_t - W_0) \stackrel{(W4)}{=} t$;

e) *Law of iterated logarithms:*

$$\limsup_{t \to \infty} \frac{W_t}{\sqrt{2t \ln(\ln t)}} = 1$$

means that for sufficiently large t:

$$W_t \le \sqrt{2t \ln(\ln t)};$$

f)

$$\lim_{t\to\infty} W_t/t = 0$$

(follows from law of iterated logarithms);

g) ${W_t}^2 - t$ *is an* $\mathcal{I}_t$ *martingale (shown in Example 7);*

h) Geometric Brownian Motion:
$\exp\{\sigma W_t - (\sigma^2/2)t\}$ *is an* $\mathcal{I}_t$ *martingale (Exercise).*

i) Generalized Brownian Motion:
$\mu t + \sigma W_t \sim N(\mu t, \sigma^2 t)$, *because:*

$$\mathsf{E}(\mu t + \sigma W_t) \overset{(\mathsf{E1,E2})}{=} \mu t + \sigma \mathsf{E}(W_t) \overset{c)}{=} \mu t,$$
$$\mathsf{V}(\mu t + \sigma W_t) \overset{(\mathsf{V2,V3})}{=} \sigma^2 \mathsf{V}(W_t) \overset{d)}{=} \sigma^2 t.$$

Note that this implies also $\mu\Delta t + \sigma\Delta W_t \sim N(\mu\Delta t, \sigma^2\Delta t)$ *with* $\Delta W_t := W_{t+\Delta t} - W_t$*, because:*

$$\mu\Delta t + \sigma\Delta W_t = \mu(t+\Delta t) + \sigma W_{t+\Delta t} - (\mu t + \sigma W_t)$$

and

$$\mu(t+\Delta t) + \sigma W_{t+\Delta t} \sim N(\mu(t+\Delta t), \sigma^2(t+\Delta t)).$$

Therefore by independence (B3),(W4)*:*

$$\mu\Delta t + \sigma\Delta W_t \sim N(\mu(t+\Delta t) - \mu t, \sigma^2(t+\Delta t) - \sigma^2 t)$$

j) Scaling Property of Brownian Motion:

$$t' = at \quad \Longrightarrow \quad W_{t'} \sim \sqrt{a}W_t$$

This may be seen as follows:

$$W_t \sim N(0, t) \quad \Rightarrow \quad W_{t'=at} \sim N(0, at)$$

But we have also by application of i):

$$\sqrt{a}W_t \sim N(0, at)$$

Therefore:

$$W_{t'=at} \sim \sqrt{a}W_t$$

Jump Processes

In constructing Brownian motion from the random walk we achieved the required proportionality of the innovations variance to the observation interval h, i.e. $\mathsf{V}(\Delta W_i) = h$, by choosing the increments ΔW_i as continuous functions of h ($\Delta W_i = \pm\sqrt{h}$) with constant probabilities $p = \pm 1/2$.

Choosing ΔW_i as a continuous function of h was necessary in order to make the sample paths of W_t continuous. That is, in order to get

$$P(W(t+h) - W(t) > \epsilon) \stackrel{h \to 0}{\longrightarrow} 0 \ \forall \epsilon > 0$$

we needed $\Delta W_i \to 0$ for $h \to 0$.

Brownian motion meets this requirement, since

$$\mathsf{E}(|\Delta W_i|) = \frac{1}{2}\left|\sqrt{h}\right| + \frac{1}{2}\left|-\sqrt{h}\right| = \sqrt{h},$$

which is a continuous function of h.

Now, if we don't insist on price innovations to be continuous functions of time, we may also go the opposite way to generate the required proportionality of innovation variance to the length of the observation interval by postulating price innovations of fixed size (say 1, for simplicity), but probability of occurrence that depends upon the length of the observation interval. We shall do this below.

The result is a process that allows for discontinuities in the sample paths, no matter how small the length of the observation interval is choosen. Such processes are called jump processes.

<u>The Poisson Process</u>

Merton(1976) suggested to model stock prices as a combination of a continuous time process based upon Brownian motion to describe ordinary price movements and a jump process based upon Poisson processes in order to describe the impact of extraordinary news (rare events) upon the stock price. We shall now discuss the jump component of the process.

A stochastic process $\{N_t\}_{t\geq 0}$ *is called a <u>Poisson process</u> with parameter* λ *if it has the following properties:*

(Po1)*:* $N_0 = 0$,
(Po2)*: Increments of* N_t *are independent, that is, if* $0 \leq t_0 < \ldots < t_n$*, then we have* $N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}$ *are independent;*
(Po3)*:* $\Delta N_{t+h} := N_{t+h} - N_t \sim \text{Poisson}(\lambda h)\ \forall t, h$, *that is:*

$$P(\Delta N_{t+h} = k) = \frac{(\lambda h)^k}{k!} e^{-\lambda h},\ k = 0, 1, \ldots$$

The Poisson process may be regarded as a counting process with the random variable N_t representing the number of events (price innovations) that occur during the time interval $[0, t]$.

The Poisson process has the following important properties:

a) *During a small interval h, at most one event occurs with probability ≈ 1.*
   To see this, expand $P(\Delta N_{t+h} = k)$ in a Taylor series around $h = 0$ and keep only up to linear terms in h:

$$P(\Delta N_{t+h} = 0) = \frac{(\lambda h)^0}{0!} e^{-\lambda h} = e^{-\lambda h} \approx 1 - \lambda h$$
$$P(\Delta N_{t+h} = 1) = \frac{(\lambda h)^1}{1!} e^{-\lambda h} = (\lambda h) e^{-\lambda h} \approx \lambda h$$
$$P(\Delta N_{t+h} = (k > 1)) = \frac{(\lambda h)^k}{k!} e^{-\lambda h} \approx \frac{(\lambda h)^k}{k!} \approx 0$$

b) *The information up to time t does not help to predict the occurence of an event in the next instance; that is we have that*

$E_t(\Delta N_{t+h}) = \mathsf{E}(\Delta N_{t+h})$, *because* ΔN_{t+h} *is by* (Po2) *independent of* $\mathcal{I}_t$, *generated by the outcomes of* $N_t - N_0 = N_t$.

c) *Events occur at a constant rate* λ, *that is:* $\mathsf{E}(\Delta N_{t+h}) = \lambda h$ *(exercise).*

d) $\mathsf{E}(N_t) = \mathsf{E}(N_t - N_0) = \mathsf{E}(\Delta N_{0+t}) = \lambda t$.

e) $\mathsf{V}(\Delta N_{t+h}) = \lambda h$, *that is, the variance increases proportional to the length of the observation interval* h *with rate* λ *(exercise).*

f) $\mathsf{V}(N_t) = \mathsf{V}(N_t - N_0) = \mathsf{V}(\Delta N_{0+t}) = \lambda t$.

Note that N_t *is not a martingale, since*

$$E_t(N_{t+u}) = E_t(N_t + \Delta N_{t+u}) = N_t + \lambda u \neq N_t.$$

However, the <u>*compensated Poisson process*</u> $J_t := N_t - \lambda t$ *is a martingale, which may be seen as follows:*

$$
\begin{aligned}
E_t(J_{t+u}) &= E_t(N_t + \Delta N_{t+u} - \lambda(t+u)) \\
&= N_t + \lambda u - \lambda t - \lambda u \\
&= N_t - \lambda t = J_t.
\end{aligned}
$$

Furthermore, $\mathsf{V}(J_t) = \mathsf{V}(N_t) = \lambda t$*.*

That is, ΔJ_{t+h} *is a martingale difference with variance proportional to the observation interval* h*, as we required for reasonable price innovation processes.*

Simulating Price Processes

Recall that we introduced the price innovation ΔW_k as the difference between price change $\Delta S_k = S(kh) - S((k-1)h)$ and its conditional expectation at time point $(k-1)h$, that is:

$$\Delta W_k = \Delta S_k - E_{k-1}(\Delta S_k),\ k = 1, \ldots, n, \quad n = \frac{T}{h}.$$

Now that we identified the increments of Brownian motion and compensated Poisson processes as possible candidates for the innovation terms ΔW_k, we may turn the logic around and simulate sample paths of the price process S_t at any observation interval h we want, given the particular model we choose for ΔW_k.

Specifically, if we define a_k as the conditional expected price change per unit time at time point $t_k = kh$, that is:

$$a_k := E_{k-1}(S(kh) - S((k-1)h))/h = E_{k-1}(\Delta S_k)/h,$$

we get for the resulting price change ΔS_k:

$$\Delta S_k = E_{k-1}(\Delta S_k) + \Delta W_k = a_k h + \Delta W_k.$$

Example 8

Suppose we observe a stock price S at monthly intervals. The current stock price is S_0=100 and we expect it to appreciate within T =12 months to $S_{12} = 220$ without any seasonal patterns (that is $a := a_1 = \ldots = a_{12}$):

$$a = \frac{\mathsf{E}(S_T) - S_0}{T} = \frac{220 - 100}{12} = 10.$$

Assume furthermore that we expect price innovations to be governed by Brownian motion with an annual variance $\mathsf{V}(W_T) = 48$. The monthly variance is then by independence of the increments ΔW_t (B3):

$$\sigma^2 := \mathsf{V}(W_t) = \mathsf{V}(W_{T/12}) = \frac{1}{12}\mathsf{V}(W_T) = \frac{48}{12} = 4.$$

Therefore, we may simulate sample paths for the stock price as:

$$\Delta S_k = at + \sigma \Delta W_k = 10 \cdot t + 2 \cdot \Delta W(1).$$

In practice this amounts to starting the process at $S_0 = 100$ *and adding for each month the expected appreciation* $a = 10$ *and 2 times a* $N(0,1)$*-distributed random variable, since* $\Delta W(1) \sim N(0,1)$.

The above approach may be generalized to

$$\Delta S_t = a(S_t, t)\Delta t + \sigma(S_t, t)\Delta W_t,$$

since we may decide for new expected one-period-returns and volatilities of the price innovations both for each price S_t *and at each time point* t.

3. Stochastic Calculus

<u>*Motivation*</u>

Suppose we want to price a derivative, whose value F is related to the stock price S_t through some relation $F = F(S_t, t)$. Then, if we could apply the standard calculus of deterministic functions, we would relate instant changes in the value of the derivative dF to instant changes of the stock price dS_t and the passage of time through the total derivative

$$dF = \frac{\partial F}{\partial S_t} dS_t + \frac{\partial F}{\partial t} dt.$$

So we need an infinitesimal version of the approximation $\Delta S_t = a\Delta t + \sigma \Delta W_t$, where for simplicity we have assumed the drift parameter $a := a(S_t, t)$ and the diffusion parameter $\sigma := \sigma(S_t, t)$ to be constants. In a deterministic setup the infinitesimal version of ΔS_t would be $dS_t = a\, dt + \sigma\, dW_t$, where dW_t would again be obtained as $dW_t = \frac{\partial W_t}{\partial t} dt$.

Differentiation in Stochastic Environments

Unfortunately the requirement for the variance of the innovation process $\varDelta W_k$,

$$\mathsf{V}(\varDelta W_k) = \mathsf{E}(\varDelta {W_k}^2) = {\sigma_k}^2 h,$$

implies that the derivative of $W(t)$ does not exist, even if it is a continuous Wiener process.

In order to see this, try to take the limit

$$\lim_{h \to 0} \frac{\varDelta W_{t+h}}{h} = \lim_{h \to 0} \frac{W(t+h) - W(t)}{h}.$$

As $W(t)$ is a random process, this requires us to specify which kind of convergence we are looking for.

We will take the limit in the sense of mean square convergence, since it turns out to be the only limit, with respect to which the integral of a Wiener process can be defined.

We are thus looking for a process A_t satisfying

$$\lim_{h\to 0} \mathsf{E}\left(\frac{\Delta W_{t+h}}{h} - A_t\right)^2 = 0,$$

where $\lim_{h\to 0}$ *now denotes the usual limit of deterministic calculus.*

Now, since

$$\begin{aligned} &\mathsf{E}\left(\frac{\Delta W_{t+h}}{h} - A_t\right)^2 \\ =&\mathsf{E}\left[\left(\frac{\Delta W_{t+h}}{h}\right)^2\right] - 2\mathsf{E}\left(\frac{\Delta W_{t+h} \cdot A_t}{h}\right) + \mathsf{E}(A_t)^2, \end{aligned}$$

this requires among others the existence of

$$\mathsf{E}\left[\left(\frac{\Delta W_{t+h}}{h}\right)^2\right].$$

But $\mathsf{E}(\Delta W_{t+h}{}^2) \xrightarrow{h\to 0} \sigma_t{}^2 h$*, such that:*

$$\mathsf{E}\left[\left(\frac{\Delta W_{t+h}}{h}\right)^2\right] = \frac{\mathsf{E}(\Delta W_{t+h}{}^2)}{h^2} = \frac{\sigma_t{}^2 h}{h^2} = \frac{\sigma_t{}^2}{h} \xrightarrow{h\to 0} \infty.$$

So there is no such thing as the derivative of $W(t)$, even when $W(t)$ is continuous Brownian motion.

Integration in Stochastic Environments

We are still trying to give a meaning to the stochastic differential equation (SDE)

$$dS_t = a(S_t, t)dt + \sigma(S_t, t)dW_t.$$

So far, it doesn't look good because we have just learnt that a derivative $\frac{\partial W_t}{\partial t}$ cannot be defined, impeding our attempt to define dW_t.

But recall what total differentials like dS_t are meant for. In the end, we are not interested in the instantaneous change of a stock price dS_t corresponding to some instantaneous random event dW_t after an infinitesimal small time interval dt. In all practical situations we will rather be interested in a finite stock price change ΔS_t due to the accumulation of random events ΔW_t during the finite time interval Δt.

If dW_t could be meaningful defined, we would calculate $\Delta S_t = S(t) - S(t - \Delta t)$ as

$$\Delta S_t = \int_{t-\Delta t}^{t} a(S_u, u)du + \int_{t-\Delta t}^{t} \sigma(S_u, u)dW_u.$$

We shall see in the following that both integrals on the right hand side can be meaningfully defined in stochastic environments. This allows for the meaningful interpretation of stochastic differential equations as shorthand notations for the corresponding integral equations.

Riemann Integrals on Stochstic Processes

Recall the definition of the Riemann integral in ordinary calculus:

$$\int_a^b f(t)dt = \lim_{n\to\infty} \sum_{i=1}^{n} f(\tau_i)\Delta t_i$$

(if the limit exists), where $\Delta t_i = t_i - t_{i-1}$, $a = t_0 < t_1 < \cdots < t_n = b$ *and* $\max|t_i - t_{i-1}| \to 0$ *as* $n \to \infty$, *and finally* $t_{i-1} \le \tau_i \le t_i$.

In exactly the same way we can define pathwise the Riemann integral on the stochastic process X_t *as:*

$$I = \int_a^b X_t \, dt = \lim_{n\to\infty} \sum_{i=1}^{n} X_{\tau_i}\Delta t_i,$$

provided the limit exists and $\mathsf{E}(I)^2 < \infty$. *This definition may be used to deal with integrals of the form*

$$\int_{t-\Delta t}^{t} a(S_u, u)du.$$

Note that the Riemann integral on a stochastic process, unlike the usual Riemann integral, is a random variable, not a number.

The Ito Integral

We shall now deal with integrals of the form

$$\int_{t-\Delta t}^{t} \sigma(S_u, u) dW_u.$$

Recall for that purpose the definition of the Riemann-Stieltjes integral: Let f and g be bounded function on $[a, b]$, then with the notations above the Riemann-Stieltjes integral of f with respect to g is defined as

$$\int_a^b f(x) dg(x) = \lim_{n\to\infty} \sum_{i=1}^{n} f(\tau_i) \Delta g_i$$

(if the limit exists) with $\Delta g_i = g(t_i) - g(t_{i-1})$, $\max|\Delta g_i| \to 0$ *as $n \to \infty$, and $t_{i-1} \le \tau_i \le t_i$.*

Note. If g is differentiable, then

$$\int_a^b f(t)\, dg(t) = \int_a^b f(t) g'(t)\, dt.$$

Suppose next that $f(t)$ is a stochastic process and W_t the Wiener process. Then the stochastic integral of $f(t)$ w.r.t W_t is a random *variable defined as*

$$I = \int_a^b f(t)\, dW_t = \lim \sum_{i=1}^{n} f(t_{i-1}) \Delta W_{t_i},$$

where $\Delta W_{t_i} = W_{t_i} - W_{t_{i-1}}$*, and* "lim" *is the limit in quadratic mean, i.e.,*

$$\lim_{n\to\infty} \mathsf{E} \left[\sum_{i=1}^{n} f(t_{i-1}) \Delta W_{t_i} - I \right]^2 = 0$$

Note that this time it makes a difference how τ_i *in* $f(\tau_i)$ *is selected from the sub-interval* $[t_{i-1}, t_i]$*. Here we have selected* $\tau_i = t_{i-1}$*, the lower bound, corresponding to the Itó integral to be defined shortly. Different selections of* τ_i *result in different values of the stochastic integral. For example, the Stratonovich integral, defined as*

$$I_{\frac{1}{2}} = \lim \sum_{i=1}^{n} \frac{1}{2} \left(f(t_i) + f(t_{i-1}) \right) \left(W_{t_i} - W_{t_{i-1}} \right)$$

is another stochastic integral with values that are in general different from those of the Itó integral, as we shall see soon.

We are now in a position to define the Itó integral as follows:

Consider the stochastic difference equation:

$\Delta S_k = a(S_{k-1}, k)h + \sigma(S_{k-1}, k)\Delta W_k,\ k = 1, \ldots, n$

where $\Delta W_k = W_k - W_{k-1}$ *is a Wiener process with zero mean and variance* h *and let*
1. $\sigma(S_t, t)$ *be* *non-anticipating*, *that is both* $\mathcal{I}_t$*-adapted and independent of all future increments* $\Delta W_{t'}$ *with* $t' > t$, *and*
2. $\sigma(S_t, t)$ *be* *non-explosive*, *that is:*

$$\mathsf{E}\left[\int_0^T \sigma(S_t, t)^2\, dt\right] < \infty.$$

Then the *Itó integral*

$$I = \int_0^T \sigma(S_t, t) dW_t$$

is defined as the mean square limit:

$$\lim_{n\to\infty} \mathsf{E}\left[\sum_{k=1}^{n} \sigma(S_{k-1}, k)\Delta W_k - I\right]^2 = 0.$$

Properties of the Ito Integral

Let $\mathcal{L}$ denote the set of Itó integrable stochastic processes, i.e., if $f \in \mathcal{L}$ then $\int_0^t f(u)\, dW_u$ exists.

(1) Linearity:

If $f, g \in \mathcal{L}$ then

$$\int_0^t (af(u) + bg(u))\, dW_u = a \int_0^t f(u) dW_u + b \int_0^t g(u)\, dW_u,$$

where $a, b \in \mathbb{R}$ are constants.

(2) Addidivity with respect to subintervals:

If $0 \leq s \leq t$ then

$$\int_0^t f(u)\, dW_u = \int_0^s f(u)\, dW_u + \int_s^t f(u)\, dW_u$$

The following properties of the Itó integral require use of the fact that $\mathsf{E}(I_n - I)^2 \stackrel{n\to\infty}{\longrightarrow} 0$ implies $F(I_n) \stackrel{n\to\infty}{\longrightarrow} F(I)$ which again implies $\mathsf{E}({I_n}^p) \stackrel{n\to\infty}{\longrightarrow} \mathsf{E}(I^p)$ for any $p = 1, 2, \ldots$.

(3) Zero mean: $\mathsf{E}\left[\int_0^t f(u)\,dW_u\right] = 0$.

To see this, recall that independence of $f(u)$ from all future increments $\Delta W_{u'>u}$ implies:

$\mathsf{E}[f(t_{i-1})(W_{t_i} - W_{t_{i-1}})] = \mathsf{E}\,[f(t_{i-1})]\,\mathsf{E}\,[(W_{t_i} - W_{t_{i-1}})] = 0.$

for all partitions $0 = t_0 < t_1 < \cdots < t_n = t$.

Therefore,

$$\mathsf{E}\left[\sum_{i=1}^{n} f(t_{i-1})(W_{t_i} - W_{t_{i-1}})\right] = \sum_{i=1}^{n} \mathsf{E}[f(t_{i-1})]\mathsf{E}[(W_{t_i} - W_{t_{i-1}})] = 0,$$

Now, by definition of the Itó integral:

$$\lim_{n\to\infty} \mathsf{E}\left(I_n - \int_0^t f(u)\,dW_u\right)^2 = 0$$

$$\text{with } I_n = \sum_{i=1}^{n} f(t_{i-1})(W_{t_i} - W_{t_{i-1}})$$

and we have just shown that $\mathsf{E}(I_n) = 0\ \forall n$ *implying that* $\lim_{n\to\infty}\mathsf{E}(I_n) = 0$.

Therefore:

$$\mathsf{E}\left[\int_0^t f(u)\,dW_u\right] = \lim_{n\to\infty} \mathsf{E}(I_n) = 0.$$

(4) Variance: $\mathsf{V}\left[\int_0^t f(u)\,dW_u\right] = \int_0^t E[f(u)^2]du.$

Denote $f_i = f(t_i)$. *Then*

$$\mathsf{V}\left[\sum_{i=1}^{n} f_{i-1}(W_{t_i} - W_{t_{i-1}})\right] = \mathsf{E}\left[\sum_{i=1}^{n} f_{i-1}(W_{t_i} - W_{t_{i-1}})\right]^2$$

$$= \sum_{i=1}^{n} \mathsf{E}\left[f_{i-1}(W_{t_i} - W_{t_{i-1}})\right]^2$$

$$+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathsf{E}\left[f_{i-1}f_{j-1}(W_{t_i} - W_{t_{i-1}})(W_{t_j} - W_{t_{j-1}})\right]$$

$$= \sum_{i=1}^{n} \mathsf{E}[f_{i-1}^2]\mathsf{E}\left[(W_{t_i} - W_{t_{i-1}})^2\right]$$

$$+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathsf{E}\left[f_{i-1}f_{j-1}(W_{t_i} - W_{t_{i-1}})\right]\mathsf{E}\left[(W_{t_j} - W_{t_{j-1}})\right]$$

$$= \sum_{i=1}^{n} \mathsf{E}\left[f_{i-1}\right]^2 \mathsf{E}\left[(W_{t_i} - W_{t_{i-1}})^2\right] = \sum_{i=1}^{n} \mathsf{E}\left[f_{i-1}\right]^2 \Delta t_i$$

Thus, by definition of the Riemann integral

$$V\left[\int_0^t f(u)\,dW_u\right] = E\left[\left(\int_0^t f(u)\,dW_u\right)^2\right]$$

$$= \lim_{n\to\infty} \mathsf{E}[I_n(t)^2] = \lim_{n\to\infty}\sum_{i=1}^{n} \mathsf{E}[f(t_{i-1})^2]\Delta t_i$$

$$= \int_0^t \mathsf{E}[f(u)^2]\,du.$$

(5) *Covariance: Using the same approach as above, we can show (exercise):*

$$
\begin{aligned}
&\mathsf{Cov}\left(\int_0^t f(u)\,dW_u,\ \int_0^t g(u)\,dW_u\right)\\
&= \mathsf{E}\left[\left(\int_0^t f(u)\,dW_u\right)\left(\int_0^t g(u)\,dW_u\right)\right]\\
&= \int_0^t \mathsf{E}[f(u)g(u)]du
\end{aligned}
$$

(6) $\int_0^t f(u)dW_u$ *is a martingale.*

This is seen as follows:

$$
\begin{aligned}
E_t\left[\int_0^{t+\Delta t} f(u)\,dW_u\right] &\overset{(2)}{=} E_t\left[\int_0^t f(u)\,dW_u + \int_t^{t+\Delta t} f(u)\,dW_u\right]\\
&= E_t\left[\int_0^t f(u)\,dW_u\right] + E_t\left[\int_t^{t+\Delta t} f(u)\,dW_u\right]\\
&= \int_0^t f(u)\,dW_u
\end{aligned}
$$

because $\int_0^t f(u)\,dW_u$ *is* $\mathcal{I}_t$*-measurable, and by independence (of the increments of* W_t*) and property (3)*

$$
E_t\left[\int_t^{t+\Delta t} f(u)\,dW_u\right] = \mathsf{E}\left[\int_t^{t+\Delta t} f(u)\,dW_u\right] = 0
$$

"Magnitudes" of dt and dW_t

From the Riemann integral we find that

$$\int_0^t du = t$$

Consider next the "integral"

$$\int_0^t (du)^2$$

Using an analogy with the definition of the Riemann integral, let $0 = t_0 < t_1 < \cdots t_n = t$ *be a partition of the interval* $[0, t]$ *such that* $t_i - t_{i-1} = t/n$. *Then*

$$\begin{aligned}
\int_0^t (du)^2 &= \lim_{n\to\infty} \sum_{i=1}^{n} (t_i - t_{i-1})^2 \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{t}{n}\right)^2 \\
&= \lim_{n\to\infty} \frac{t^2}{n} \\
&= 0.
\end{aligned}$$

These results suggest that for the infinitesimal *increment dt it is reasonable set* $(dt)^2 = 0$. *In fact for any $a > 1$, using the above integral approach, we have* $(dt)^a = 0$.

Let us next consider dW_t.

By definition of the Itó integral

$$\int_0^t dW_u = W_t,$$

which suggests that dW_t is a random variable.

From the properties (3) and (4), we can deduce $\mathsf{E}[dW_t] = 0$ and

$$\mathsf{V}\left[\int_0^t dW_u\right] = \mathsf{V}[W_t] = t = \int_0^t du,$$

which suggests that $\mathsf{V}[dW_t] = dt$.

Finally, because the increments of the Wiener process are normally distributed, we can (symbolically) denote

$$dW_t \sim N(0, dt).$$

I.e., dW_t can be considered as a normally distributed random variable with mean zero and variance dt (standard deviation $\sqrt{dt}$).

Consider next $(dW_t)^2$. *Above we observed that* $(dt)^2 = 0$.

Interpreting $(dW_t)^2$ *through the Itó integral, consider again*

$$I_n(t) = \sum_{i=1}^{n} (W_{t_i} - W_{t_{i-1}})^2.$$

The expected value of $I_n(t)$ *is* t. *It is a good candidate for the Itó integral. The quadratic mean becomes*

$$\begin{aligned} \mathsf{E}\left[I_n(t) - t\right]^2 &= \mathsf{E}[I_n(t)]^2 - 2t\mathsf{E}[I_n(t)] + t^2 \\ &= \mathsf{E}[I_n(t)]^2 - t^2, \end{aligned}$$

because $\mathsf{E}[I_n(t)] = \sum_{i=1}^n \mathsf{E}(W_{t_i} - W_{t_{i-1}})^2 = \sum_{i=1}^n (t_i - t_{i-1}) = t$.

Now (assume again that $t_i - t_{i-1} = t/n$*)*

$$\begin{aligned} \mathsf{E}[I_n(t)]^2 &= \mathsf{E}\left[\sum_{i=1}^{n} (W_{t_i} - W_{t_{i-1}})^2\right]^2 \\ &= \sum_{i=1}^{n} \mathsf{E}(W_{t_i} - W_{t_{i-1}})^4 \\ &\quad +2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathsf{E}\left[(W_{t_i} - W_{t_{i-1}})^2 (W_{t_j} - W_{t_{j-1}})^2\right] \end{aligned}$$

Now, using $\mathsf{E}(X^4) = 3\sigma^4$ *for any* $X \sim N(0, \sigma^2)$ *and the properties of the Wiener process:*

$$\begin{aligned}
\mathsf{E}[I_n(t)]^2 &= \sum_{i=1}^{n} 3(t_i - t_{i-1})^2 \\
&\quad + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(t_i - t_{i-1})(t_j - t_{j-1}) \\
&= 3n\left(\frac{t}{n}\right)^2 + 2\frac{n(n-1)}{2}\left(\frac{t}{n}\right)^2 \\
&= 3\frac{t^2}{n} + \frac{n-1}{n}t^2.
\end{aligned}$$

Thus in all

$$\mathsf{E}[I_n(t) - t]^2 = 3\frac{t^2}{n} + \frac{n-1}{n}t^2 - t^2 \to 0, \text{ as } n \to \infty,$$

i.e.,

$$\int_0^t (dW_u)^2 = t = \int_0^t du.$$

Thus we have an important additional result:

$$(dW_t)^2 = dt.$$

How about $dt\, dW_t$*?*

Using again the integral approach, consider

$$\sum_{i=1}^{n} \Delta t_i \Delta W_{t_i} := \sum_{i=1}^{n} (t_i - t_{i-1})(W_{t_i} - W_{t_{i-1}}).$$

The increments ΔW_{t_i} *are independent, so*

$$\begin{aligned}
&\mathsf{E}\left(\sum_{i=1}^{n} \Delta t_i \Delta W_{t_i}\right)^2 \\
=&\mathsf{E}\left(\sum_{i=1}^{n} \Delta t_i{}^2 \Delta W_{t_i}{}^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \Delta t_i \Delta t_j \Delta W_{t_i} \Delta W_{t_j}\right) \\
=&\sum_{i=1}^{n} \Delta t_i{}^2 \mathsf{E}(\Delta W_{t_i}{}^2) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \Delta t_i \Delta t_j \mathsf{E}(\Delta W_{t_i}) \mathsf{E}(\Delta W_{t_j}) \\
=&\sum_{i=1}^{n} \Delta t_i{}^3 = \sum_{i=1}^{n} \left(\frac{t}{n}\right)^3 = \frac{t^3}{n^2} \to 0 \textit{ as } n \to \infty.
\end{aligned}$$

Thus we can state

$$\int_0^t du\, dW_u = 0.$$

so that we get

$$dt\, dW_t = 0.$$

We summarize our results in the following multiplication table for infinitesimal increments:

$\times$	dt	dW_t
dt	$dt\,dt = 0$	$dt\,dW_t = 0$
dW_t	$dW_t dt = 0$	$dW_t dW_t = dt$

Note that $(dW_t)^2$ is not the same as dW_t^2!
As a matter of fact, we have not even yet defined the meaning of dW_t^2. Returning to our original goal of relating changes in asset prices to increments of Brownian motion by means of stochastic differential equations a meaningful interpretation of $df(W_t)$ may be found by means of the stochastic integral equation

$$\int_0^t df(W_u) = \int_0^t \frac{df}{dW_u} dW_u.$$

So in order to find $\int_0^t dW_u{}^2$ we need to calculate

$$\int_0^t \frac{dW_u{}^2}{dW_u} dW_u = \int_0^t 2W_u\,dW_u = 2\int_0^t W_u\,dW_u.$$

In evaluating $I_n^{(0)} = \sum_{i=1}^n W_{t_{i-1}}(W_{t_i} - W_{t_{i-1}})$ *we use the following trick:*

$$W_{t_{i-1}} = \frac{1}{2}(W_{t_i} + W_{t_{i-1}}) - \frac{1}{2}(W_{t_i} - W_{t_{i-1}}),$$

$$(W_{t_i} + W_{t_{i-1}})(W_{t_i} - W_{t_{i-1}}) = {W_{t_i}}^2 - {W_{t_{i-1}}}^2.$$

Therefore, using $t_n = t$ *and* $W_{t_0} = W_0 = 0$*:*

$$\begin{aligned}\sum_{i=1}^n W_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) &= \frac{1}{2}\sum_{i=1}^n (W_{t_i}^2 - W_{t_{i-1}}^2) - \frac{1}{2}\sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 \\ &= \frac{1}{2}W_t^2 - \frac{1}{2}\sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2.\end{aligned}$$

When discussing the meaning of $(dW_t)^2$*, we had found that the sum on the right hand side converges in mean square to* t*. Thus:*

$$I_0(t) := \int_0^t W_s dW_s = \frac{1}{2}(W_t^2 - t).$$

Note: This is not an application of a relation similar to deterministic calculus

(Wrong!) $\lim(X + Y) = \lim(X) + \lim(Y)$,

*where "*lim*" denotes the limit in quadratic mean, because in general, this identity does not hold for convergence in quadratic mean!*

Instead, we deduce mean square convergence of $I_n^{(0)}$ to $I_0(t)$ from mean square convergence of $X_n = \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2$ to t as follows:

$$\mathsf{E}\left(\frac{1}{2}(W_t^2 - X_n) - \frac{1}{2}(W_t^2 - t)\right)^2 = \left(-\frac{1}{2}\right)^2 \mathsf{E}(X_n - t)^2 \xrightarrow{n\to\infty} 0.$$

Similarly, if we select the end points W_{t_i} of the subintervals, we get (exercise):

$$I_1(t) := \lim \sum_{i=1}^n W_{t_i}(W_{t_i} - W_{t_{i-1}}) = \frac{1}{2}(W_t^2 + t).$$

Recall that I_0 is a martingale whereas I_1 is not, none of which equals the integral function in deterministic calculus $\int x\,dx = \frac{1}{2}x^2$.

The Stratonovich integral is

$$\begin{aligned} I_{\frac{1}{2}}(t) &= \lim \sum_{i=1}^n \frac{1}{2}(W_{t_i} + W_{t_{i-1}})(W_{t_i} - W_{t_{i-1}}) \\ &= \frac{1}{2} \lim \sum_{i=1}^n (W_{t_i}^2 - W_{t_{i-1}}^2) = \frac{1}{2} W_t^2 = \frac{1}{2}(I_1 + I_0). \end{aligned}$$

i.e., the same as the Riemann integral.

Ito's Lemma

Let us return to our original goal to relate the instantaneous change in the price of some derivative $F = F(S_t, t)$ to instant changes in the price S_t of the underlying asset, whose increments at discrete observation intervals are given by the stochastic difference equation:

$$\Delta S_k = a_k h + \sigma_k \Delta W_k, \quad k = 1, \dots, n;$$

where we have written $\Delta S_k = S_{t_k} - S_{t_{k-1}}$, $a_k = a(S_{t_{k-1}}, t_k)$, $\sigma_k = \sigma(S_{k-1}, t_k)$, $\Delta W_k = W_{t_k} - W_{t_{k-1}}$, and we have partitioned the observation interval $[0, T]$ of S_t as before into $0 = t_0 < t_1 < \cdots < t_n = T$ with

$$h = t_k - t_{k-1} = \frac{T}{n}, \quad k = 1, \dots, n.$$

We have meanwhile learnt that the only meaningful interpretation of the infinitesimal version of ΔS_k: $dS_t = a(S_t, t)\, dt + \sigma(S_t, t)\, dW_t$, (called Itó process) is to take it as a shorthand notation for the integral equation

$$\int_0^t dS_u = \int_0^t a(S_u, u)\, du + \int_0^t \sigma(S_u, u)\, dW_u.$$

Our goal is to relate $dF(S_t, t)$ to dS_t.

The standard approach of using the total derivative $dF = F_s\, dS_t + F_t\, dt$ with F_s and F_t denoting the partial derivatives

$$F_s = \frac{\partial F(S_t, t)}{\partial S_t}, \quad F_t = \frac{\partial F(S_t, t)}{\partial t}$$

does not work in stochastic environments as such, but needs to be modified as follows.

Consider the Taylor series expansion to second order of an infinitely differentiable function $f : x \to \mathbb{R}$ around $x_0 \in \mathbb{R}$:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + R,$$

where the remainder R does not exceed the last explicitely calculated addend of the sum (here: $|R| \leq |\frac{1}{2} f''(x_0)(x - x_0)^2|$).

In our case of two variables the Taylor expansion to second order is:

$$\begin{aligned} F(S_{t_k}, t_k) &= F(S_{t_{k-1}}, t_{k-1}) + F_s \Delta S_k + F_t h + \tfrac{1}{2} F_{ss} (\Delta S_k)^2 \\ &\quad + \tfrac{1}{2} F_{tt}\, h^2 + F_{st}\, h\, \Delta S_k + R. \end{aligned}$$

With $\Delta F_k := F(S_{t_k}, t_k) - F(S_{t_{k-1}}, t_{k-1})$*, we get by subtracting* $F(S_{t_{k-1}}, t_{k-1})$ *on both sides:*

$$\begin{aligned} \Delta F_k &= F_s \Delta S_k + F_t h + \tfrac{1}{2} F_{ss} (\Delta S_k)^2 \\ &\quad + \tfrac{1}{2} F_{tt}\, h^2 + F_{st}\, h\, \Delta S_k + R. \end{aligned}$$

As $n \to \infty$, $h = t_k - t_{k-1} \to dt$, $\Delta S_k \to dS_t$, *and* $\Delta F_k \to dF$, *and* $R \to 0$ *because it consists of terms* $(\Delta t_k)^m$ *and* $(\Delta W_k)^m$ *with* $m \geq 3$. *So we get*

$$\begin{aligned} dF(S_t, t) &= F_s\, dS_t + F_t\, dt + \tfrac{1}{2} F_{ss} (dS_t)^2 \\ &\quad + \tfrac{1}{2} F_{tt} (dt)^2 + F_{st}\, dt\, dS_t. \end{aligned}$$

Using $(dt)^2 = dt\, dW_t = 0$ *and* $(dW_t)^2 = dt$ *from the calculation rules for differentials, we obtain:*

$$\begin{aligned} (dS_t)^2 &= (a(S_t,t)\, dt + \sigma(S_t,t)\, dW_t)^2 \\ &= a^2(S_t,t)(dt)^2 + 2a(S_t,t)\sigma(S_t,t) dt\, dW_t + \sigma^2(S_t,t)(dW_t)^2 \\ &= \sigma^2(S_t,t)\, dt. \end{aligned}$$

Similarly we obtain:

$$\begin{aligned} dt\, dS_t &= dt\,(a(S_t,t)dt + \sigma(S_t,t)\,dW_t) \\ &= a(S_t,t)(dt)^2 + \sigma(S_t,t)\,dt\,dW_t \\ &= 0. \end{aligned}$$

After inserting these expressions for $(dS_t)^2$ *and* $dt\,dS_t$ *we get:*

$$dF = F_s\,dS_t + F_t\,dt + \frac{1}{2}F_{ss}\sigma^2(S_t,t)\,dt$$

Replacing dS_t *with its Itó representation yields*

$$\begin{aligned} dF(S_t,t) &= F_s\,[a(S_t,t)dt + \sigma(S_t,t)dW_t] \\ &\quad + F_t\,dt + \tfrac{1}{2}F_{ss}\sigma^2(S_t,t)\,dt, \end{aligned}$$

which yields *<u>Ito's differential equation</u>:*

$$dF = \left[F_t + a(S_t,t)F_s + \frac{1}{2}\sigma^2(S_t,t)F_{ss}\right]dt + \sigma(S_t,t)\,dW_t$$

The result is summarized as Itó's Lemma:

Itó's Lemma: *Let $F(S_t, t)$ be a twice differentiable function of t and of the random process S_t with Itó differential equation*

$$dS_t = a_t\,dt + \sigma_t\,dW_t, \quad t \geq 0,$$

with $a_t = a(S_t, t)$ and $\sigma_t = \sigma(S_t, t)$ continuously twice-differentiable (real valued) functions. Then

$$dF = F_s\,dS_t + F_t\,dt + \frac{1}{2}F_{ss}\sigma_t^2\,dt,$$

or, after substituting for the right hand side of dS_t above

$$dF = \left(F_s\,a_t + F_t + \frac{1}{2}F_{ss}\sigma^2\right)dt + F_s\sigma_t\,dW_t,$$

where

$$F_s = \frac{\partial F}{\partial S_t},\ F_t = \frac{\partial F}{\partial t},\ \text{and}\ F_{ss} = \frac{\partial^2 F}{\partial S_t^2}.$$

The major usage of the Itó formula in finance is to find the (Itó) stochastic differential equation (SDE) for the financial derivative once the (Itó) SDE of the underlying asset is given.

Example 9

Let $F(W_t, t) = W_t^2$.

We may use $dF = F_s\, dS_t + F_t\, dt + \frac{1}{2} F_{ss} \sigma_t^2\, dt$ with $a_t = 0$ and $\sigma_t = 1$, such that $S_t = W_t$. Therefore:

$$dF(W_t, t) = F_w\, dW_t + F_t\, dt + \frac{1}{2} F_{ww}\, dt$$

Now:

$$F_w = \partial W^2 / \partial W = 2W$$
$$F_t = \partial W^2 / \partial t = 0$$
$$F_{ww} = \partial^2 F / \partial W^2 = 2$$

Therefore:

$$dF(W_t, t) = 2W_t\, dW_t + dt.$$

Thus the drift parameter of F is $a(F, t) = 1$ and the diffusion parameter is $\sigma(F, t) = 2W_t$.

Example 10.

$$F(W_t, t) = 3 + t + e^{W_t}.$$

$$\begin{aligned}
\Rightarrow dF(W_t, t) &= F_t\, dt + F_w\, dW_t + \tfrac{1}{2} F_{ww}\, dt \\
&= dt + e^{W_t}\, dW_t + \tfrac{1}{2} e^{W_t}\, dt \\
&= \left(1 + \tfrac{1}{2} e^{W_t}\right)\, dt + e^{W_t}\, dW_t.
\end{aligned}$$

Example 11. Consider geometric Brownian motion

$$S_t = S_0 e^{\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\}},$$

where S_0 is a constant. Using Itó with $\sigma(W_t, t) = 1$:

$$\begin{aligned}
dS_t &= \tfrac{\partial S_t}{\partial W_t}\, dW_t + \tfrac{\partial S_t}{\partial t}\, dt + \tfrac{1}{2} \tfrac{\partial^2 S_t}{\partial W_t^2}\, dt \\
&= S_0 \sigma e^{\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\}}\, dW_t + (\mu - \tfrac{1}{2}\sigma^2) S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}\, dt \\
&\quad + \tfrac{1}{2}\sigma^2 S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}\, dt \\
&= S_t \sigma\, dW_t + (\mu - \tfrac{1}{2}\sigma^2) S_t\, dt + \tfrac{1}{2}\sigma^2 S_t\, dt \\
&= S_t(\mu\, dt + \sigma\, dW_t),
\end{aligned}$$

or

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma dW_t,$$

or

$$dS_t = \mu S_t dt + \sigma S_t dW_t.$$

Comparing to the general form $dS_t = a(S_t, t)dt + \sigma(S_t, t)dW_t$ we find that $a(S_t, t) = \mu\, S_t$, and $\sigma(S_t, t) = \sigma\, S_t$.

Itó's formula as an integration tool

Suppose our task is to evaluate

$$\int_0^t W_s \, dW_s.$$

Make a guess

$$F(W_t, t) = \frac{1}{2} W_t^2.$$

Then using Itó

$$dF(W_t, t) = W_t \, dW_t + \frac{1}{2} \, dt.$$

The integral form is

$$\frac{1}{2} W_t^2 = F(W_t, t) = \int_0^t dF(W_s, s) = \int_0^t W_s \, dW_s + \frac{1}{2} \int_0^t ds.$$

So

$$\int_0^t W_s \, dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t.$$

The start off point here is to make a "good guess".

Example 12 (Integration by parts)

Consider the Itó integral

$$\int_0^t s\,dW_s.$$

Make a start off guess

$$F(W_t, t) = tW_t.$$

Then

$$dF(W_t, t) = W_t\,dt + t\,dW_t,$$

and

$$tW_t = \int_0^t dF(W_s, s) = \int_0^t W_s\,ds + \int_0^t s\,dW_s.$$

So

$$\int_0^t s\,dW_s = tW_t - \int_0^t W_s\,ds.$$

Example 13 (*Geometric Brownian motion*)

Consider

$$\frac{dS_t}{S_t} = \mu\,dt + \sigma\,dW_t \quad \left(\Rightarrow \ \left(\frac{dS_t}{S_t}\right)^2 = \sigma^2\,dt\right).$$

Let

$$F(S_t, t) = \ln S_t.$$

Then

$$\begin{aligned} dF(S_t, t) &= F_t\,dt + F_s\,dS_t + \tfrac{1}{2}F_{ss}(dS_t)^2 \\ &= \tfrac{1}{S_t}dS_t - \tfrac{1}{2}\tfrac{1}{S_t^2}(dS_t)^2 \\ &= \mu\,dt + \sigma\,dW_t - \tfrac{1}{2}\sigma^2\,dt \\ &= (\mu - \tfrac{1}{2}\sigma^2)\,dt + \sigma\,dW_t. \end{aligned}$$

We get

$$\begin{aligned} \ln S_t - \ln S_0 &= \textstyle\int_0^t dF(S_u, u) \\ &= \textstyle\int_0^t(\mu - \frac{1}{2}\sigma^2)du + \int_0^t \sigma\,dW_u \\ &= (\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t. \end{aligned}$$

So adding $\ln S_0$ *on both sides and taking the exponential yields:*

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}.$$

Integral form of Itós Lemma

Consider again Itós formula:

$$dF = F_s \, dS_t + F_t \, dt + \frac{1}{2} F_{ss} \sigma_t^2 \, dt.$$

We know that this is shorthand notation for the integral equation:

$$\int_0^t dF(S_u, u) = F(S_t, t) - F(S_0, 0)$$
$$= \int_0^t F_s \, dS_u + \int_0^t \left[F_u(S_u, u) + \frac{1}{2} F_{ss}(S_u, u) \sigma_u^2 \right] du$$

Rearranging terms in this integral form yields:

$$\begin{aligned} \int_0^t F_s dS_u &= [F(S_t, t) - F(S_0, 0)] \\ &\quad - \int_0^t \left[F_u(S_u, u) + \frac{1}{2} F_{ss}(S_u, u) \sigma_u^2 \right] du, \end{aligned}$$

which represents the stochastic integral $\int F_s dS$ *as a function of integrals with respect to time.*

Multivariate Itó formula

Consider a 2×1 *vector of stochastic processes obeying the following SDE:*

$$\begin{pmatrix} dS_1(t) \\ dS_2(t) \end{pmatrix} = \begin{pmatrix} a_1(t) \\ a_2(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}(t) & \sigma_{12}(t) \\ \sigma_{21}(t) & \sigma_{22}(t) \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}$$

or:

$$\begin{aligned} dS_1(t) &= a_1(t)\, dt + \sigma_{11}(t)\, dW_1(t) + \sigma_{12}(t)\, dW_2(t) \\ dS_2(t) &= a_2(t)\, dt + \sigma_{21}(t)\, dW_1(t) + \sigma_{22}(t)\, dW_2(t), \end{aligned}$$

where it is assumed that the Wiener processes $W_1(t)$ *and* $W_2(t)$ *are independent.*

Suppose $F(S_1(t), S_2(t), t)$ *is a twice differentiable real valued function.*

Using Taylor expansion and taking limit in the same manner as in the univariate case yields (with $(dt)^2 = 0$*,* $dt\, dS_1 = 0$*, and* $dt\, dS_2 = 0$*)*

$$\begin{aligned} dF &= F_t\, dt + F_{s_1}\, dS_1 + F_{s_2}\, dS_2 \\ &\quad + \frac{1}{2}\left[F_{s_1,s_1}(dS_1)^2 + F_{s_2,s_2}(dS_2)^2 + 2F_{s_1,s_2} dS_1 dS_2\right]. \end{aligned}$$

The independence of W_1 and W_2 implies that $dW_1\,dW_2 = 0$ (exercise).

Therefore

$$(dS_1)^2 = \sigma_{11}^2\,dW_1(t)^2 + \sigma_{12}^2\,dW_2(t)^2 = (\sigma_{11}^2 + \sigma_{12}^2)\,dt,$$

$$(dS_2)^2 = \sigma_{21}^2\,dW_1(t)^2 + \sigma_{22}^2\,dW_2(t)^2 = (\sigma_{21}^2 + \sigma_{22}^2)\,dt,$$

and

$$\begin{aligned} dS_1\,dS_2 &= \sigma_{11}\sigma_{21}\,dW_1(t)^2 + \sigma_{12}\sigma_{22}\,dW_2(t)^2 \\ &= (\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22})dt. \end{aligned}$$

Using these in the multivariate Itó formula

$$\begin{aligned} dF &= F_t\,dt + F_{s_1}\,dS_1 + F_{s_2}\,dS_2 \\ &\quad + \tfrac{1}{2}\left[F_{s_1,s_1}(dS_1)^2 + F_{s_2,s_2}(dS_2)^2 + 2F_{s_1,s_2}dS_1dS_2\right]. \end{aligned}$$

gives dF as a function of dW_1 and dW_2.

Example 14

Consider interest rate derivatives. Assume that the yield curve depends on two state variables, a short rate r_t, and a long rate R_t. Denote the price of the derivative as $F(r_t, R_t, t)$. Assume

$$dr_t = a_1(t)\,dt + \sigma_{11}(t)\,dW_1(t) + \sigma_{12}(t)\,dW_2(t),$$

and

$$dR_t = a_2(t)\,dt + \sigma_{21}(t)\,dW_1(t) + \sigma_{22}(t)dW_2(t).$$

Straightforward application of the Itó formula gives

$$\begin{aligned} dF &= F_t\,dt + F_r\,dr_t + F_R\,dR_t \\ &\quad + \tfrac{1}{2}\left[F_{rr}(\sigma_{11}^2 + \sigma_{12}^2) + F_{RR}(\sigma_{21}^2 + \sigma_{22}^2) + 2F_{rR}(\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22})\right] dt, \end{aligned}$$

Note that shocks to the short and the long rate are correlated in this model, because:

$$\begin{aligned} \mathrm{Cov}(dr_t, dR_t) &= \mathrm{Cov}(\sigma_{11}dW_1 + \sigma_{12}dW_2, \sigma_{21}dW_1 + \sigma_{22}dW_2) \\ &= \mathrm{E}\left[(\sigma_{11}dW_1 + \sigma_{12}dW_2)(\sigma_{21}dW_1 + \sigma_{22}dW_2)\right] \\ &= \sigma_{11}\sigma_{21}\mathrm{E}(dW_1)^2 + \sigma_{12}\sigma_{22}\mathrm{E}(dW_2)^2 \\ &= [\sigma_{11}(t)\sigma_{21}(t) + \sigma_{12}(t)\sigma_{22}(t)]\,dt. \end{aligned}$$

4. Pricing Derivatives with PDE's

Partial Differential Equations

We shall in the following sketch the line of thoughts that led to the invention of the well known Black-Scholes option pricing formula. In the preceeding chapter we have learnt that we may use Itós lemma in order to transform the stochastic differential equation (SDE) for the underlying assets price S_t into another SDE for the price of the corresponding derivative contract $F_t(S_t, t)$, where in both SDE's randomness enters through the same Wiener process W_t. We may then construct a replicating portfolio (which is in mathematical terms just a linear combination of the two SDE's), in order to eliminate all randomness in the value of the replicating portfolio, which should then yield the return of a riskfree investment by the no-arbitrage assumption. Mathematically this corresponds to reducing the SDE into a non-stochstic partial differential differential equation, PDE for short.

So our strategy in pricing F_t follows the following scheme:

SDE for $S_t \stackrel{Ito}{\longrightarrow}$ SDE for $F_t \stackrel{noarbitrage}{\longrightarrow}$ PDE for F_t

The PDE for F_t is then solved by using the boundary conditions of the derivative contract, which impose certain values upon F_t at specific timepoints t (for example expiration date T) and/or specific prices S (for example strike price or exercise price X).

Example 15

Boundary conditions for a call option with exercise price X

$$F(S_T, T) = \max(S_T - X, 0).$$

Generally $F(S_t, t)$ is unknown.

Other examples for boundary conditions are:

$$S_t, F_t \geq 0 \quad \forall\ 0 \leq t \leq T, \qquad F(S, t) = G(S, T),$$

where G could be any known function.

Construction of risk-free portfolios

We consider now the reduction of two stochastic differential equations (SDE's) for the underlyings price S_t and the derivative price F_t into a non-stochastic partial differential equation for F_t.

Construct a portfolio P_t with θ_1 units invested in F and θ_2 units invested in S:

$$P_t = \theta_1 F(S_t, t) + \theta_2 S_t.$$

Following Itós lemma it evolves as:

$$\begin{aligned} dP &= \frac{\partial P}{\partial t} dt + \frac{\partial P}{\partial F} dF + \frac{\partial P}{\partial S} dS \\ &\quad + \frac{1}{2}\frac{\partial^2 P}{\partial F^2}(dF)^2 + \frac{1}{2}\frac{\partial^2 P}{\partial S^2}(dS)^2 + \frac{\partial^2 P}{\partial F \partial S} dF\, dS \end{aligned}$$

Next we assume (erroneously) that θ_1 and θ_2 are constants, that is, both independent of t and S_t. Such an assumption holds only over infinitesimal time intervals, which makes everything what follows only approximate.

If we nevertheless assume θ_1 and θ_2 to be constants, then all partial derivatives above equal zero, except:

$$\frac{dP}{dF} = \theta_1 \quad \text{and} \quad \frac{dP}{dS} = \theta_2.$$

so

$$dP = \theta_1 dF + \theta_2 dS.$$

Now let S_t follow an Itó process:

$$dS = a(S_t, t)\, dt + \sigma(S_t, t) dW_t,$$

then using Itós lemma:

$$dF = (F_t + \frac{1}{2}\sigma_t^2 F_{ss}) dt + F_s\, dS,$$

and $dP = \theta_1\, dF + \theta_2\, dS$ becomes

$$dP = \theta_1 (F_t + \frac{1}{2}\sigma_t^2 F_{ss})\, dt + (\theta_1 F_s + \theta_2) dS.$$

Note that we can eliminate the term proportional to dS, by making the following special choice for the porftolio weights θ_1 and θ_2. Set:

$$\theta_1 = 1 \quad \text{and} \quad \theta_2 = -F_s.$$

Then:

$$dP = (F_t + \frac{1}{2}\sigma_t^2 F_{ss})\, dt,$$

that is, the hedge portfolio $P = F - F_s S$ is <u>risk-free</u>! But in an arbitrage-free market a riskless investment should earn the riskfree interest rate (henceforth denoted as r), that is, under continuous compounding:

$$P_t = P_0 \exp(rt),$$

the total differential of which is (standard calculus applies!):

$$dP = \frac{\partial P}{\partial t} dt = rP_0 e^{rt} dt = rP\, dt.$$

Equating the formulas for dP we get

$$r\, P\, dt = (F_t + \frac{1}{2}\sigma_t^2 F_{ss})\, dt,$$

implying:

$$rP = F_t + \frac{1}{2}\sigma_t^2 F_{ss},$$

or, after inserting $P = F - F_s S$:

$$r(F - F_s S) = F_t + \frac{1}{2}\sigma_t^2 F_{ss}.$$

Arranging the terms gives a standard partial differential equation:

$$-r\,F + r\,F_s S + F_t + \frac{1}{2}\sigma_t^2 F_{ss} = 0$$

to be solved. The solution depends on the diffusion term $\sigma_t := \sigma(S_t, t)$ *of the Itó process* S_t *and on the boundary conditions. Note for later reference, that the solution does not depend upon the drift term* $a(S_t, t)$ *of the Itó process* $dS = a(S_t, t)dt + \sigma(S_t, t)dW_t$.

For example, geometric Brownian motion with the boundary conditions of the European option leads to the Black-Scholes formula. But American options have different solutions, for which no closed form is known (unless it is an American call for which the boundary conditions coincide with those of the European option, because early exercise is suboptimal).

Example 16 Black-Scholes call option formula

Assume the boundary condition of an European call option with strike price X, that is:

$$F(S_T, T) = \max(S_T - X, 0),$$

and geometric Brownian motion (GBM):

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t.$$

This implies

$$a(S_t, t) = \mu S_t \quad \text{and} \quad \sigma_t := \sigma(S_t, t) = \sigma S_t,$$

such that we need to solve the so called Black-Scholes PDE,

$$-r\,F + r\,F_s S + F_t + \frac{1}{2}\sigma^2 F_{ss} S^2 = 0,$$

which we obtain by inserting $\sigma_t = \sigma S_t$ into

$$-r\,F + r\,F_s S + F_t + \frac{1}{2}\sigma_t^2 F_{ss} = 0.$$

(note once again that we did not need to make use of our knowledge about the drift term, that is, $a(S_t, t) = \mu S_t$.)

The Black-Scholes PDE

$$-r\,F + r\,F_s S + F_t + \frac{1}{2}\sigma^2 F_{ss} S^2 = 0$$

is solved by the famous Black-Scholes call option formula (exercise):

$$F(S_t, t) = S_t N(d_1) - X e^{-r\tau} N(d_2),$$

where

$$d_1 = \frac{\log(S_t/X) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}},$$

$$d_2 = \frac{\log(S_t/X) + (r - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} = d_1 - \sigma\sqrt{\tau},$$

$$N(d_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_i} e^{-\frac{1}{2}x^2}\, dx$$

are the values of the standard normal cdf at d_i, $i = 1, 2$, *and* $\tau = T - t$ *is the time to maturity.*

Furthermore, it satisfies the relevant boundary condition $F(S_T, T) = \max(S_T - X, 0)$, as may be seen by considering the limiting behaviour of F_t as $\tau \to 0$, for $S > X$ and $S < X$ seperately is follows:

$$
\begin{array}{llcl}
S > X : & d_1,\, d_2 & \overset{\tau\to 0}{\longrightarrow} & \infty \\
& \Rightarrow N(d_1),\, N(d_2) & \overset{\tau\to 0}{\longrightarrow} & 1 \\
& \Rightarrow \quad F(S_t, t) & \overset{\tau\to 0}{\longrightarrow} & S_T - X, \\
S < X : & d_1,\, d_2 & \overset{\tau\to 0}{\longrightarrow} & -\infty \\
& \Rightarrow N(d_1),\, N(d_2) & \overset{\tau\to 0}{\longrightarrow} & 0 \\
& \Rightarrow \quad F(S_t, t) & \overset{\tau\to 0}{\longrightarrow} & 0,
\end{array}
$$

which may be summarized as:

$$F(S_T, T) = \max(S_T - X, 0).$$

Note that the Black-Scholes formula holds only for this particular boundary condition (with a similar formular for the corresponding put option with a boundary condition of the form $F(S_T, T) = \max(X - S_T, 0)$). In general it is not possible to find analytical expressions for the solutions of the Black-Scholes PDE under arbitrary boundary conditions.

The role of dividends

If S_t pays constant dividends δ per time unit, then the return of the hedge portfolio $P = F - F_s S$ entails price appreciation of both the stock and dividends, which by the no-arbitrage assumption has to equal the risk-free rate times portfolio value. That is, the return of the hedge portfolio is

$$dP + \delta\, dt = rPdt$$

and the PDE becomes

$$-rF + rF_s S + \delta + F_t + \frac{1}{2} F_{ss} \sigma^2 = 0.$$

Thus the constant dividends cause no extra problems. However, if dividends follow a random process themselves, then the PDE approach does not work in the general case, unless one assumes dividends and stock price to follow the same random process, because a linear combination of two SDE's cannot eliminate more than one random process.

Exotic Options

There are lots of different kinds of options, like lookback options, Asian, Knock-In, Knock-out, ladder, multiasset, etc.

In each of these the PDE is basically the same, but boundary conditions change, which lead to different solutions of the PDE. For example:

Floating lookback: The payoff is $S_T - S_{\min}$, *if positive, where* $S_{\min} = \min_{t\in[0,T]}\{S_t\}$.

Fixed lookback: The payoff is $S_{\max} - X$, *if positive, where* $S_{\max} = \max_{t\in[0,T]}\{S_t\}$ *and* X *is the strike price.*

Ladder options: Several thresholds, such that if the underlying price reaches these thresholds, the return of the option is "locked in".

Trigger options: option comes into life first after the underlyings price has crossed a certain threshold.

Knock-out Options: Option expires as soon as the underlyings price has crossed a certain threshold.

Asian options: The payoff depends upon the average price of the underlying asset.

Solving PDE's in practice

Consider again the PDE in the framework of Black-Scholes (S_t is assumed to be geometric Brownian motion):

$$F_t + rF_sS + \frac{1}{2}F_{ss}\sigma^2S^2 = rF, \ S \geq 0, \ 0 \leq t \leq T.$$

Closed form solutions, like the B-S call price, cannot often be found, or the solution is difficult.

Numerical solutions give values for $F(S,t)$ for (discrete) combinations (S_i, t_j) of the stock price S and time variable t.

Finite difference methods approximate the relevant PDE by a set of difference equations and solve these equations iteratively from the boundary conditions.

Finite Difference Methods

To solve the PDE numerically, one approximates the PDE with finite *increments* ΔS *and* Δt*. Partitions for the range of* S *and* t *are needed:*

1. *Select grid size for* ΔS *and* Δt*.*
2. *Select an appropriate range for* S*,* $S_{\text{min}} \leq S \leq S_{\text{max}}$*.*
3. *Determine the boundary conditions.*
4. *Determine the value of* $F(S,t)$ *at the grid points.*

Suppose that the life of the option is T*. Divide it into* n *subintervals* $\Delta t = T/n$*. In the same manner divide the "reasonable" range of the stock price into* m *subintervals* $\Delta S = S_{\text{max}}/m$*. Then we have a grid* (t_j, S_i) *where* $t_j = j\Delta t$ *and* $S_i = i\Delta S$*,* $i = 0, 1, \ldots, m$*,* $j = 1, \ldots, n$*, i.e., time is running as*

$$0, \Delta t, 2\Delta t, \ldots, (n-1)\Delta t, T$$

and S *as*

$$0, \Delta S, 2\Delta S, \ldots, (m-1)\Delta S, S_{\text{max}}.$$

The Implicit Finite Difference Method

The idea is to replace the partial derivatives in the relevant PDE with their discrete approximations at points $F_{ij} = F(S_i, t_j)$. For the first order partial derivatives the choices are

Backward difference:

$$\frac{\partial F}{\partial S} \approx \frac{F_{i,j} - F_{i-1,j}}{\Delta S}.$$

Forward difference:

$$\frac{\partial F}{\partial S} \approx \frac{F_{i+1,j} - F_{i,j}}{\Delta S}.$$

Central difference:

$$\frac{\partial F}{\partial S} \approx \frac{F_{i+1,j} - F_{i-1,j}}{2\Delta S},$$

which is the average of the previous two.

The second derivative is approximated by

$$\frac{\partial^2 F}{\partial S^2} \approx \left[\frac{F_{i+1,j} - F_{i,j}}{\Delta S} - \frac{F_{i,j} - F_{i-1,j}}{\Delta S}\right]\frac{1}{\Delta S}$$
$$= \frac{F_{i+1,j} - 2F_{i,j} + F_{i-1,j}}{(\Delta S)^2}.$$

In the imlicite finite difference method we use the forward difference for $\partial F/\partial t$ and the central difference for $\partial F/\partial S$. Inserting these into the Black-Scholes PDE

$$F_t + r\,F_s S + \frac{1}{2}\sigma^2 F_{ss} S^2 = r\,F$$

with $S_i = i\Delta S$ we arrive at

$$\frac{F_{i,j+1}-F_{i,j}}{\Delta t} + r\,i\Delta S \frac{F_{i+1,j}-F_{i-1,j}}{2\,\Delta S}$$

$$+\tfrac{1}{2}\sigma^2 (i\,\Delta S)^2 \frac{F_{i+1,j}-2F_{i,j}+F_{i-1,j}}{(\Delta S)^2} = rF_{i,j}$$

for $i = 1, 2, \ldots, m-1$ and $j = 0, 1, \ldots, n-1$. Rearranging terms and noting that the ΔS terms cancel out, we get

$$a_i F_{i-1,j} + b_i F_{i,j} + c_i F_{i+1,j} = F_{i,j+1}$$

where

$$\begin{aligned} a_i &= \tfrac{1}{2} i \left(r - i\sigma^2\right) \Delta t, \\ b_i &= 1 + \left(r + i^2\sigma^2\right) \Delta t, \\ c_i &= -\tfrac{1}{2}\, i \left(r + i\sigma^2\right) \Delta t. \end{aligned}$$

This relates F for any gridpoint to its value at the same and the two neighbouring stock-prices from the preceeding timepoint.

To illustrate the method consider an American put on a non-dividend paying stock.

The value of the put at time T $(= t_n)$ is $\max(K - S_T, 0)$, *where S_T is the stock price at time T and K is the strike price. Thus*

(T) $F_{i,n} = F(i\Delta S, t_n) = \max(K - i\Delta S, 0)$,
$i = 0, 1, \ldots, m.$

The value of the put for $S = 0$ is K. So

(0) $$F_{0,j} = K, \quad j = 0, 1, \ldots, n.$$

When the stock price goes to infinity, the value of the put approaches zero. Thus we approximate

(max) $$F_{m,j} = 0, \quad j = 0, 1, \ldots, n.$$

Equations (0), (max) and (T) define the values of the put option along the three edges $S = 0$, $S = S_{\max}$, and $t = T$ of the grid (t_i, S_j), $i = 0, 1, \ldots, m$, $j = 0, 1, \ldots, n$.

To fill the remaining nodes of the grid, we start from the points $T - \Delta t$ with $j = n - 1$, where the difference equations for the price F of the derivative read

$$a_i F_{i-1,n-1} + b_i F_{i,n-1} + c_i F_{i+1,n-1} = F_{i,n}$$

for $i = 1, 2, \ldots, m - 1$. The right-hand sides of these are known from the boundary conditions (T). Furthermore, from (0) and (max):

$$\begin{aligned} F_{0,n-1} &= K \\ F_{m,n-1} &= 0. \end{aligned}$$

So we end up with $m-1$ simultaneous equations with $m - 1$ unknowns:

$$F_{1,n-1}, F_{2,n-1}, \ldots, F_{m-1,n-1}.$$

Given the solutions for $F_{i,n-1}$, if $F_{i,n-1} < K - i\Delta S$, then early exercise at time $T - \Delta t$ is optimal and $F_{i,n-1}$ is set equal to $K - i\Delta S$. The nodes with $T - i\Delta t$, $i = 2, 3, \ldots, m$, are handled in a similar way, eventually giving $F_{0,1}, F_{0,2}, \ldots, F_{0,n-1}$, one of which is the option price of interest.

The relationship of the neighboring option prices in the implicit method is

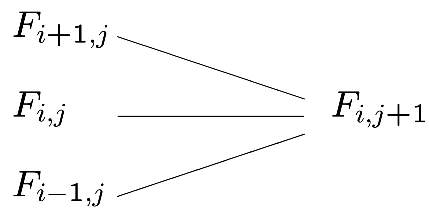

Figure. Relationship of the derivative price at time $t + \Delta t$ to three values of derivative at time t in the implicit method.

Thus in the implicit method the value at time t depends directly on its two neighbors, and hence indirectly (implicitly) on all option values at that time step.

The advantages of the implicit method are that it is robust, and it always converges to the solution of the differential equation as ΔS and Δt approach to zero. The disadvantage is that it is pretty tedious to program, because $m-1$ simultaneous equations need to be solved. A considerable simplification is reached in the explicit finite difference method, which relates F to its value at the same and the two neighbouring stockprices at the following time point.

Explicit Finite Difference Method

To simplify the calculations of the implicit method, let us assume additionally that the $\partial F/\partial S$ and $\partial^2 F/\partial S^2$ are the same at grid points (i,j) and $(i,j+1)$, such that we may replace our earlier approximations

$$\frac{\partial F}{\partial S} \approx \frac{F_{i+1,j} - F_{i-1,j}}{2\Delta S} \text{ and}$$
$$\frac{\partial^2 F}{\partial S^2} \approx \frac{F_{i+1,j} - 2F_{i,j} + F_{i-1,j}}{(\Delta S)^2}$$

by

$$\frac{\partial F}{\partial S} \approx \frac{F_{i+1,j+1} - F_{i-1,j+1}}{2\Delta S} \text{ and}$$
$$\frac{\partial^2 F}{\partial S^2} \approx \frac{F_{i+1,j+1} - 2F_{i,j+1} + F_{i-1,j+1}}{(\Delta S)^2}.$$

Inserting these into the Black-Scholes PDE

$$F_t + r\,F_s S + \frac{1}{2}\sigma^2 F_{ss} S^2 = r\,F$$

we get in a similar manner as for the implicit finite difference method:

$$F_{i,j} = a_i^* F_{i-1,j+1} + b_i^* F_{i,j+1} + c_i^* F_{i+1,j+1}$$

with

$$\begin{aligned} a_i^* &= -\frac{i\,\Delta t}{2(1+r\,\Delta t)}\left(r-\sigma^2 i\right) \\ b_i^* &= \frac{1}{1+r\,\Delta t}\left(1-\sigma^2 i^2\,\Delta t\right) \\ c_i^* &= \frac{i\,\Delta t}{2(1+r\,\Delta t)}\left(r+\sigma^2 i\right) \end{aligned}$$

Example 17. Consider an American put option with $T = 1/3$ years, i.e., 4 months, $K = 21$, $r = 10\%$ p.a., and

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t$$

where $\sigma = 30\%$ p.a. Then we know

$$F(S_T, T) = \max(21 - S_T, 0)$$

$$F(S_t, t) \to 0 \text{ as } S_t \text{ gets large.}$$

and

$$F(0, t) = 21$$

for all t.

4. Equivalent Martingale Measures

Motivation

Recall from the discrete one-period model that an arbitrage-free market ensures the existence of a risk-neutral probability measure $Q(\omega)$, *that is a function* $Q : \Omega \to \mathbb{R}$ *with*

$$Q(\omega) > 0 \,\forall\, \omega \in \Omega, \ \sum_{\omega \in \Omega} Q(\omega) = 1, \ \mathsf{E}^Q(\Delta S^*) = 0,$$

where ΔS^* *denotes the discounted price difference* $S(t_1)/(1 + r_f) - S(t_0)$, *and* $\mathsf{E}^Q(\cdot)$ *denotes the expectation operator with the true probabilities* $P(\omega)$ *replaced by the risk-neutral probabilities* $Q(\omega)$.

This led to the risk-neutral valuation principle, which we could use to price any attainable contingent claim X *as the expected value of the riskfree discounted payoff* X^*, *where the expectation had to be taken again with respect to the risk-neutral probability measure* Q, *that is:* $X_0 = \mathsf{E}^Q(X^*)$.

Probability Measures

In order to extend the risk-neutral valuation principle into the multiperiod framework with continuous time and continuous prices, let us first take a closer look at probability measures of continuous random variables.

As an illustration consider the probability density $f(z)$ of a standard normal distribution,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

Then the probability of finding the random variable Z near a specific value $\bar{z}$ is

$$P\left(\bar{z} - \frac{1}{2}\Delta < Z < \bar{z} + \frac{1}{2}\Delta\right) = \int_{\bar{z}-\frac{1}{2}\Delta}^{\bar{z}+\frac{1}{2}\Delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.$$

which is a real number (between zero and one). Thus the probability associates a real number (in this case between zero and one) to intervals on real line, or more generally to (Borel) sets. Such functions are called measures in mathematics or measure functions.

Because Δ *is small*

$$\begin{aligned}\int_{\bar{z}-\frac{1}{2}\Delta}^{\bar{z}+\frac{1}{2}\Delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz &\approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\bar{z}^2} \int_{\bar{z}-\frac{1}{2}\Delta}^{\bar{z}+\frac{1}{2}\Delta} dz \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\bar{z}^2} \Delta.\end{aligned}$$

For infinitesimal Δ, *denoted as* dz *we denote the associated measure by* $dP(z)$ *or simply* dP. *Thus in the above case, we have*

$$\begin{aligned}dP(z) &= P\left(\bar{z} - \frac{1}{2}dz < Z < \bar{z} + \frac{1}{2}dz\right) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz\end{aligned}$$

Generally, if P *is a probability measure, we have*

$$\int_{-\infty}^{\infty} dP = 1.$$

With these notations, e.g.,

$$\mathsf{E}[X] = \int_{-\infty}^{\infty} x\, dP(x).$$

So the expected value is mathematically an integral with respect to probability measure.

The Fundamental Theorem of Asset Pricing

We consider next the generalization of the risk neutral valuation principle to continuous random variables in continuous time.

A probability measure $\tilde{P}$ is called a martingale measure for the discounted price process $\tilde{S}_t = e^{-rt}S_t$, if $\tilde{S}_t$ is a martingale under $\tilde{P}$, that is:

$$\mathsf{E}_t^{\tilde{P}}[e^{-r(t+u)}S_t] = e^{-rt}S_t, \text{ whenever } u \geq 0$$

where r is the riskfree rate, and $E_t^{\tilde{P}}$ indicates that the expectation is taken with respect to probability measure $\tilde{P}$.

The fundamental theorem of asset pricing establishes then the equivalence of the absence of arbitrage opportunities and the existence of a martingale measure for the discounted price process, just like the existence of a risk-neutral probability measure was equivalent to the absence of arbitrage opportunities in the discrete one-period model.

The absence of arbitrage opportunities under presence of a martingale measure $\tilde{P}$ for the discounted price process $\tilde{S}_t$ leads then to the <u>risk-neutral valuation principle</u> in very much the same way as in the discrete one-period model, that is:

$$X_t = \mathsf{E}_t^{\tilde{P}}\left[e^{-r\tau} X_T\right] \quad \textit{with} \quad \tau = T - t$$

is the arbitrage-free price at time t of any random payoff X_T at expiration date T.

Thus a martingale measure can be viewed as a representation of the market's current opinion on the evolution of values of underlying assets and the prices of all derivatives contingent to them. Consequently the knowledge of the martingale measure is all that is needed, in principle, to value whatever derivative securities.

Then given a stock price process S_t with probability measure P the goal is to find the martingale measure $\tilde{P}$.

The Radon-Nikodym Derivative

Before considering how to change the probability measure of a stochastic process (in order to transform it into a martingale and apply the risk-neutral valuation principle), let us first figure out how to transform probability measures of continuous random variables.

As an example, consider again a $N(0,1)$ distributed random variable Z with probability measure

$$dP(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz,$$

and define a transformation function

$$\xi(z) = e^{z\mu - \frac{1}{2}\mu^2}.$$

Then $d\tilde{P}(z) = \xi(z)dP(z)$ is again a probability measure, because:

$$d\tilde{P}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2 + z\mu - \frac{1}{2}\mu^2} dz = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-\mu)^2} dz,$$

which is the probability measure of a $N(\mu, 1)$ distributed random variable.

Furthermore, we can recover the original probability measure $dP(z)$ by multiplying $d\tilde{P}(z)$ by $\frac{1}{\xi(z)} = e^{-z\mu+\frac{1}{2}\mu^2}$:

$$\begin{aligned}\frac{d\tilde{P}(z)}{\xi(z)} &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2+z\mu-\frac{1}{2}\mu^2}dz\, e^{-z\mu+\frac{1}{2}\mu^2}\\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz = dP(z)\end{aligned}$$

So we have just shown, that there exists a function $\xi(z)$, such that if we multiply the probability measure of a standard normal random variable with it, we obtain a new probability. The transformed random variable is again normal but has a different mean, and the transformation is reversible.

Before stating the general condition, under which such a transformation can be made, let us first introduce some notation. For that purpose, recall that by the fundamental theorem of calculus:

$$F(z) = \int_a^z f(x)dx \quad \Rightarrow \quad \frac{dF(z)}{dz} = f(z).$$

We apply this to the transformed probability measure $\tilde{P}$:

$$\tilde{P}(z) = \int_{-\infty}^{z} d\tilde{P}(x) = \int_{-\infty}^{z} \xi(x) dP(x) \ \Rightarrow \ \frac{d\tilde{P}(z)}{dP(z)} = \xi(z).$$

This justifies calling the transformation function ξ a derivative of $\tilde{P}$ with respect to P. In mathematical measure theory this is known as the <u>Radon-Nikodym derivative</u>. Similarly, we may express the inverse transformation as the derivative of P with respect to $\tilde{P}$:

$$\frac{dP(z)}{d\tilde{P}(z)} = \frac{1}{\xi(z)}.$$

Obviously, in order to find a transformation from P to $\tilde{P}$ and vice versa we need:

$$\tilde{P}(dz) > 0 \textit{ if and only if } P(dz) > 0$$

(because otherwise we would have to divide by zero in at least one transformation).

The Radon-Nikodym Theorem states that the condition

$$\tilde{P}(dz) > 0 \text{ if and only if } P(dz) > 0$$

is not only necessary but also sufficient to grant the existence of the Radon-Nikodym derivative ξ (and its inverse $1/\xi$).

If P and $\tilde{P}$ satisfy the above mentioned condition of the theorem, they are called equivalent probability measures, in the sense that it is always possible to transform P into $\tilde{P}$ and vice versa by:

$$d\tilde{P}(z) = \xi(z)dP(z), \quad dP(z) = \frac{1}{\xi(z)}d\tilde{P}(z).$$

Note for later reference, that an equivalent way of writing down this transformation is:

$$\begin{aligned}\tilde{P}(A) &= \int_A d\tilde{P}(z) = \int_{-\infty}^{\infty} \mathbb{I}_A d\tilde{P}(z) \\ &= \int_{-\infty}^{\infty} \mathbb{I}_A \xi(z) dP(z) = E^P(\mathbb{I}_A \xi) \ \forall A \in \mathcal{F},\end{aligned}$$

where $\mathbb{I}_A$ is the indicator function of event A and $\mathcal{F}$ is the (sigma-)algebra, upon which Z is defined.

The Girsanov Theorem

The following theorem is essential for changing the probability measure if continuous stochastic processes.

Theorem (Girsanov) *Let* $W(t), 0 \le t \le T$ *be a standard Wiener process on a probability space* $(\Omega, \mathcal{F}, P)$. *Let* $\mathcal{F}(t), 0 \le t \le T$ *be the accompanying filtration, and let* $X(t), 0 \le t \le T$ *be a stochastic process adapted to this filtration. For* $0 \le t \le T$, *define*

$$\tilde{W}(t) = W(t) - \int_0^t X(u)\, du,$$
$$\xi(t) = \exp\left(\int_0^t X(u)\, dW_u - \frac{1}{2}\int_0^t X(u)^2\, du\right),$$

and define a new probability measure

$$\tilde{P}(A) = \int_A \xi(T) dP(z) = E^P(I\!I_A \xi(T)) \;\forall A \in \mathcal{F}.$$

Then $\tilde{W}_t$ *is a standard Wiener process under the new probability measure* $\tilde{P}$, *provided that*

$$\mathsf{E}\left[e^{\int_0^t X(u)\, du}\right] < \infty,\; t \in [0, T].$$

Note that the transformation process ξ *is a martingale with* $\mathsf{E}(\xi(t)) = 1 \ \forall \ t \in [0, T]$. *In order to see this, note first that*

$$\int_0^0 X(u)\, dW_u = \int_0^0 X(u)^2\, du = 0.$$

Therefore

$$\xi(0) = \exp\left(\int_0^0 X(u)\, dW_u - \frac{1}{2}\int_0^0 X(u)^2\, du\right) = e^0 = 1.$$

In order to show that $\mathsf{E}(\xi(t)) = 1 \ \forall \ t \in [0, T]$ *it suffices then to show that* ξ *is a martingale. To see this, define*

$$A(t) = \int_0^t X(u)^2\, du, \quad B(t, W_t) = \int_0^t X(u)\, dW_u$$

implying

$$dA(t) = X(t)^2\, dt, \quad dB(t, W_t) = X(t)\, dW_t,$$

and apply Itós lemma

$$d\xi = \frac{\partial \xi}{\partial t} dt + \frac{\partial \xi}{\partial B} dB + \frac{1}{2}\frac{\partial^2 \xi}{\partial B^2}(dB)^2.$$

Using $\frac{\partial \xi}{\partial t} = \frac{\partial \xi}{\partial A}\frac{\partial A}{\partial t}$ *we obtain*

$$
\begin{aligned}
d\xi(t) &= -\frac{1}{2}\xi(t)X^2(t)dt + \xi(t)X(t)dW_t + \frac{1}{2}\xi(t)X^2(t)dt \\
&= \xi(t)X(t)\,dW_t,
\end{aligned}
$$

such that

$$\xi(t) - \xi(0) = \int_0^t d\xi(s) = \int_0^t \xi(s)X(s)\,dW_s.$$

Therefore

$$\xi(t) = 1 + \int_0^t \xi(s)X(s)\,dW_s.$$

Recall now, that $\int \xi(s)X(s)\,dW_s$*, like any other Itó-integral, is a martingale, that is:*

$$E_t\left(\int_0^{t+u} \xi(s)X(s)\,dW_s\right) = \int_0^t \xi(s)X(s)\,dW_s.$$

Therefore:

$$
\begin{aligned}
E_t[\xi(t+u)] &= E_t\left(1 + \int_0^{t+u} \xi(s)X(s)\,dW_s\right) \\
&= 1 + \int_0^t \xi(s)X(s)\,dW_s = \xi(t),
\end{aligned}
$$

that is, $\xi(t)$ *is a martingale.*

In order to gain some intuition for the meaning of the process $X(t)$ in the Girsanov theorem, let us assume that X is a constant, say $X = \mu$. We obtain then by evaluating the integrals in $\xi(t)$:

$$\xi(t) = \exp\left[\mu W(t) - \frac{1}{2}\mu^2 t\right],$$

in particular for $t = 1$, where $W(t) = W(1) \sim N(0,1)$:

$$\xi(1) = \exp\left[\mu W(1) - \frac{1}{2}\mu^2\right],$$

which is identical to the Radon-Nikodym derivative we used in order to transform the $N(0,1)$-distributed r.v. Z into a $N(\mu,1)$-distributed r.v. $\tilde{Z}$. Furthermore, going back to the Girsanov theorem for $X = \mu$:

$$\tilde{W}(t) = W(t) - \mu t.$$

We observe that $X(t)$ introduces a general drift in the Girsanov theorem, just like μ shifted the mean in the Radon-Nikodym derivative. However, $X(t)$ is an $\mathcal{I}_t$-adapted process, allowing for fairly complicated drifts.

Although $X(t)$ introduces a drift into $\tilde{W}(t)$, according to the Girsanov theorem, this drift is removed under the new probability measure $\tilde{P}$. We may verify this for our special case $X = \mu$ by calculating the expected value of $\tilde{W}(T)$ under the new probability measure $\tilde{P}$:

$$\mathsf{E}^{\tilde{P}}\left(\tilde{W}(T)\right) = \int_{-\infty}^{\infty} \tilde{W}(T)\, d\tilde{P} = \int_{-\infty}^{\infty} (W(T) - \mu T)\xi(T)\, dP$$

$$= \int_{-\infty}^{\infty} (w - \mu T) \exp\left(\mu w - \frac{1}{2}\mu^2 T\right) \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{w^2}{2T}\right) dw$$

$$= \frac{1}{\sqrt{2\pi T}} \int_{-\infty}^{\infty} (w - \mu T) \exp\left(-\frac{(w-\mu T)^2}{2T}\right) dw$$

$$= \frac{1}{\sqrt{2\pi T}} \int_{-\infty}^{\infty} y \exp\left(-\frac{y^2}{2}\right) dy = 0,$$

where we have used that the probability measure of the $N(0,T)$-distributed random variable $W(T)$ is

$$dP(W(T)) = \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{w^2}{2T}\right) dw,$$

and we have substituted $y(w) = w - \mu T$.

So there is indeed no drift in the process $\tilde{W}(t) = W(t) - \mu t$ under the new probability measure $\tilde{P}$.

Application: Geometric Brownian Motion

Let

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

Then

$$Z_t := \ln\left(\frac{S_t}{S_0}\right) = (\mu - \frac{\sigma^2}{2})t + \sigma W_t$$

with probability measure

$$dP(z) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{[z-(\mu-\frac{\sigma^2}{2})t]^2}{2\sigma^2 t}\right\} dz = f(z)dz,$$

because $Z_t \sim N((\mu - \frac{\sigma^2}{2})t), \sigma^2 t)$.

By substituting $z(y) = \ln y$ we find for the probability measure of $Y_t = e^{Z_t} = S_t/S_0$:

$$dP(z(y)) = \frac{dP}{dz}\frac{dz}{dy}\,dy = f(z(y)) \cdot \frac{1}{y}\,dy$$
$$= \frac{1}{\sqrt{2\pi\sigma^2 t y}} \exp\left\{-\frac{1}{2\sigma^2 t}\left[\ln y - (\mu - \frac{\sigma^2}{2})t\right]^2\right\} dy$$
$$= f(y)\,dy$$

where $f(y)$ *denotes the probability density function of a lognormally distributed random variable* $Y_t \sim \Lambda((\mu - \frac{\sigma^2}{2})t), \sigma^2 t)$.

Substituting $y(s) = s/S_0$ *yields for the probability measure of* $S_t = S_0 Y_t$:

$$dP(y(s)) = \frac{dP}{dy}\frac{dy}{ds}\,ds = f(y(s)) \cdot \frac{1}{S_0}\,ds$$
$$= \frac{1}{\sqrt{2\pi\sigma^2 t s}} \exp\left\{-\frac{1}{2\sigma^2 t}\left[\ln\left(\frac{s}{S_0}\right) - (\mu - \frac{\sigma^2}{2})t\right]^2\right\} ds$$
$$= f(s)\,ds$$

S_t *is not a martingale under this measure, because recalling that* $\exp(\sigma W_t - \frac{\sigma^2}{2}t)$ *is a martingale with respect to* P *we find:*

$$\begin{aligned} E_t^P(S_{t+u}) &= E_t^P\left(S_0 e^{\mu(t+u)} e^{-\frac{\sigma^2}{2}(t+u)+\sigma W_{t+u}}\right) \\ &= S_0 e^{\mu(t+u)} E_t^P\left(e^{-\frac{\sigma^2}{2}(t+u)+\sigma W_{t+u}}\right) \\ &= S_0 e^{\mu(t+u)} e^{-\frac{\sigma^2}{2}t+\sigma W_t} \\ &= e^{\mu u} S_0 e^{\left(\mu-\frac{\sigma^2}{2}\right)t+\sigma W_t} \\ &= e^{\mu u} S_t. \end{aligned}$$

But the discounted process $\tilde{S}_t = e^{-rt}S_t$ would be a martingale if $\mu = r$, because then:

$$E_t^P\left(e^{-r(t+\mu)}S_{t+u}\right) = e^{-r(t+\mu)}E_t^P(S_{t+u})$$
$$= e^{-r(t+\mu)}e^{\mu u}S_t = e^{-rt}e^{-(\mu-r)u}S_t = e^{-rt}S_t.$$

However, in the real world we have usually $\mu > r$, because investors require a positive risk premium for risky investments. Luckily the Girsanov theorem tells us, that we can find a probability measure $\tilde{P}$, that transforms the drift term in such a way that $\tilde{S}_t$ becomes a martingale.

For that purpose, rewrite $Z_t = (\mu - \frac{\sigma^2}{2})t + \sigma W_t$ *in the following way:*

$$\begin{aligned} Z_t &= (r - \frac{\sigma^2}{2})t + \sigma W_t + (\mu - r)t \\ &= (r - \frac{\sigma^2}{2})t + \sigma \tilde{W}_t \end{aligned}$$

with

$$\tilde{W}_t = W_t - \frac{1}{\sigma}(r - \mu)t,$$

that is,

$$\tilde{W}_t = W_t - \int_0^t X(u)\, du$$

with

$$X(u) = \frac{1}{\sigma}(r - \mu).$$

The Girsanov theorem tells us then that $\tilde{W}_t$ *is a standard Wiener process under the new probability measure*

$$d\tilde{P}(\tilde{w}) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{\tilde{w}^2}{2t}\right) d\tilde{w},$$

implying that $\exp(\sigma\tilde{W}_t - \frac{\sigma^2}{2}t)$ *is a martingale with respect to* $\tilde{P}$.

We have therefore, referring to our earlier calculation of $E_t^P(S_{t+u})$*:*

$$
\begin{aligned}
&E_t^{\tilde{P}}(S_{t+u}) = E_t^{\tilde{P}}\left(S_0 e^{Z_{t+u}}\right) \\
=&E_t^{\tilde{P}}\left(S_0 e^{(r-\frac{\sigma^2}{2})(t+u)+\sigma\tilde{W}_{t+u}}\right) \\
=&e^{ru}S_t,
\end{aligned}
$$

such that

$$
\begin{aligned}
E_t^{\tilde{P}}(\tilde{S}_{t+u}) &= E_t^{\tilde{P}}\left(e^{-r(t+u)}S_{t+u}\right) \\
&= e^{-r(t+u)}E_t^{\tilde{P}}(S_{t+u}) \\
&= e^{-r(t+u)}e^{ru}S_t \\
&= e^{-rt}S_t = \tilde{S}_t,
\end{aligned}
$$

implying that the discounted process $\tilde{S}_t$ *is a martingale under the new probability measure* $\tilde{P}$.

Risk-neutral valuation in practice

In practice, this calls for the following strategy in pricing derivatives, where the underlying may be assumed to follow geometric Brownian motion:

1. Simulate GBM with drift μ replaced by the riskfree rate r.
2. Using many simulations, take the average of the derivatives payoff X_T at expiration T as an approximation for $\mathsf{E}^{\tilde{P}}(X_T)$.
3. Discount the result with the risk-free rate to get an approximation for $X_0 = e^{-rt}\mathsf{E}^{\tilde{P}}(X_T)$.

Example 17

Risk-neutral valuation by Monte-Carlo simulation of a European call with 6 months to maturity, strike-price $X = 50$, stock-price $S = 47$, risk-free rate $r = 5\%$, and volatility $\sigma = 30\%$.