Introduction to Stochastic Processes

Kazuhisa Matsuda

Department of Economics
The Graduate Center, The City University of New York,
365 Fifth Avenue, New York, NY 10016-4309
Email: maxmatsuda@maxmatsuda.com
http://www.maxmatsuda.com/

January 2005

Abstract

This paper presents the introductory knowledge of stochastic processes for finance majors in mind. Main topics are sample paths properties of stochastic processes such as continuous paths or discontinuous paths. Martingale property and Markov property are introduced.

[1] Continuous Function and Discontinuous Function

[1.1] Function

Definition 1.1 Function A function $f : a \rightarrow f(a)$ or $f : A \rightarrow B$ on $\mathbb{R}$ uniquely maps (relates) a set of input values $a \in A$ to a set of output values $f(a) \in B$. The domain of a function is the set A on which a function is defined and the set of all actual outputs $f(a) \in B$ is called the range of a function.

A function is a many-to-one mapping (i.e. not one-to-many mapping). For example, a function $f(a) = a$ is a one-to-one mapping, $f(a) = -a^2$ is a two-to-one mapping except for $a = 0$, and $f(a) = \sin(2\pi a)$ is a many-to-one mapping.

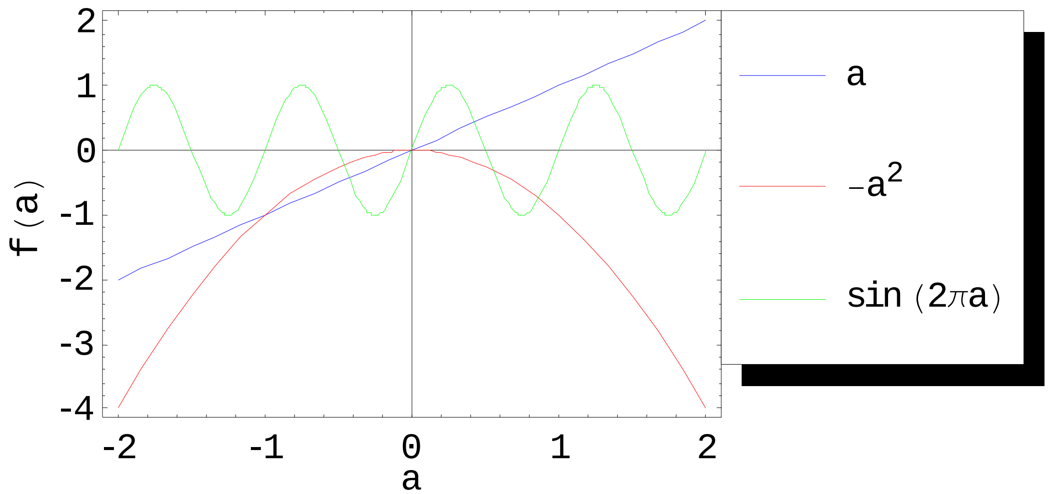


Figure 1.1: Examples of a function $f : a \rightarrow f(a)$.

[1.2] Left Limit and Right Limit of a Function

Definition 1.2 Left limit and Right limit of a function A function $f : a \rightarrow f(a)$ on $\mathbb{R}$ has a left limit $f(b-)$ at a point $a = b$ if $f(a)$ approaches $f(b-)$ when a approaches b from the below (the left-hand side):

$$\lim_{a \rightarrow b-} f(a) = f(b-).$$

A function $f : a \rightarrow f(a)$ on $\mathbb{R}$ has a right limit $f(b+)$ at a point $a = b$ if $f(a)$ approaches $f(b+)$ when a approaches b from the above (right-hand side):

$$\lim_{a \rightarrow b+} f(a) = f(b+).$$

[1.3] Right Continuous Function and Right Continuous with Left Limit (RCLL) Function

Definition 1.3 Right continuous function A function f on $\mathbb{R}$ is said to be right continuous at a point $a = b$ if it satisfies the following conditions:

(1) $f(b)$ is defined. In other words, a point b is in the domain of a function f.
(2) Right limit of the function as a approaches b from the above (right hand side) exists, i.e. $\lim_{a \to b+} f(a) = f(b+)$.
(3) $f(b+) = f(b)$.

Definition 1.4 Right continuous with left limit (rcll) function A function f on $\mathbb{R}$ is said to be right continuous with left limit at a point $a = b$ if it satisfies the following conditions:

(1) $f(b)$ is defined. In other words, a point b is in the domain of a function f.
(2) Right limit of the function as a approaches b from the above (right hand side) exists, i.e. $\lim_{a \to b+} f(a) = f(b+)$. Left limit of the function as a approaches b from the below (left hand side) exists, i.e. $\lim_{a \to b-} f(a) = f(b-)$.
(3) $f(b+) = f(b)$.

The above definitions imply that a rcll function is right continuous, but the reverse is not true. In other words, a rcll function is more restrictive than a right continuous function because a rcll function needs left limit. This point is illustrated in Figure 1.2.

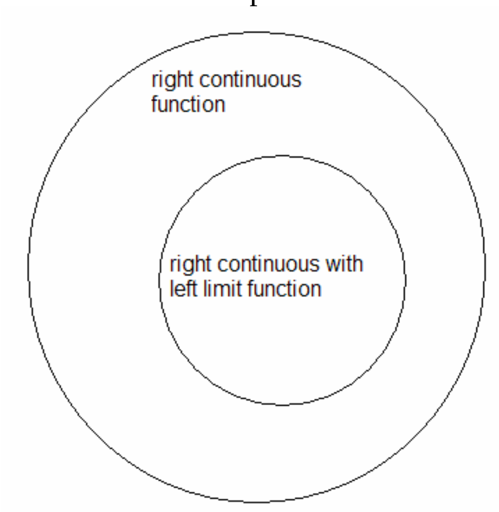


Figure 1.2: Relationship between rc function and rcll function.

Consider a piecewise constant function defined as (illustrated in Figure 1.3):

$$f(a)=\begin{cases}0 & \text{if} \quad a<1 \\ 1 & \text{if} \quad 1\le a<2 \\ 2 & \text{if} \quad 2\le a<3\end{cases}. \tag{1.1}$$

The right limit at a point $a=1$ is equal to the actual value of the function at a point $a=1$:

$$f(1+)=f(1)=1,$$

this means f is right continuous at a point $a=1$. But the left limit at a point $a=1$ is not equal to the actual value of the function at a point $a=1$:

$$f(1-)=0\neq f(1)=1,$$

this means f is not left continuous at a point $a=1$. Therefore, this function is right continuous with left limit. And the jump size is:

$$f(1+)-f(1-)=1-0=1.$$

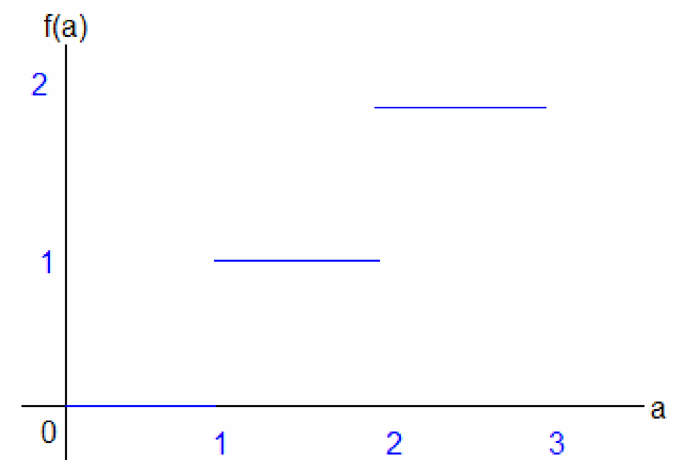


Figure 1.3: Right continuous with left limit (rcll) function.

[1.4] Left Continuous Function and Left Continuous with Right Limit (LCRL) Function

Definition 1.5 Left continuous function A function f on $\mathbb{R}$ is said to be left continuous at a point $a=b$ if it satisfies the following conditions:

(1) $f(b)$ is defined. In other words, a point b is in the domain of a function f.
(2) Left limit of the function as a approaches b from the below (left hand side) exists, i.e. $\lim_{a\to b-} f(a)=f(b-)$.
(3) $f(b-)=f(b)$.

Definition 1.6 Left continuous with right limit (lcrl) function A function f on $\mathbb{R}$ is said to be left continuous with right limit at a point $a = b$ if it satisfies the following conditions:

(1) $f(b)$ is defined. In other words, a point b is in the domain of a function f.
(2) Right limit of the function as a approaches b from the above (right hand side) exists, i.e. $\lim_{a \to b+} f(a) = f(b+)$. Left limit of the function as a approaches b from the below (left hand side) exists, i.e. $\lim_{a \to b-} f(a) = f(b-)$.
(3) $f(b-) = f(b)$.

Consider a piecewise constant function defined as:

$$f(a) = \begin{cases} 0 & \text{if} \quad a \le 1 \\ 1 & \text{if} \quad 1 < a \le 2 \\ 2 & \text{if} \quad 2 < a \le 3 \end{cases}. \tag{1.2}$$

The left limit at a point $a = 1$ is equal to the actual value of the function at a point $a = 1$:

$$f(1-) = f(1) = 0,$$

this means f is left continuous at a point $a = 1$. But the right limit at a point $a = 1$ is not equal to the actual value of the function at a point $a = 1$:

$$f(1+) = 1 \ne f(1) = 0,$$

this means f is not right continuous at a point $a = 1$. Therefore, this function is left continuous with right limit. And the jump size is:

$$f(1+) - f(1-) = 1 - 0 = 1.$$

[1.5] Continuous Function

Definition 1.7 Continuous function A function $f : a \to f(a)$ on $\mathbb{R}$ is said to be continuous at a point $a = b$ if it satisfies the following conditions:

(1) $f(b)$ is defined. In other words, a point b is in the domain of a function f.
(2) Right limit of the function as a approaches b from the above (right hand side) exists, i.e. $\lim_{a \to b+} f(a) = f(b+)$. Left limit of the function as a approaches b from the below (left hand side) exists, i.e. $\lim_{a \to b-} f(a) = f(b-)$.
(3) $f(b+) = f(b-) = f(b)$.

In other words, a continuous function is a left and right continuous function which in turn means that a continuous function is the most restrictive among rc, rcll, and continuous functions. All the functions in Figure 1.1 are continuous.

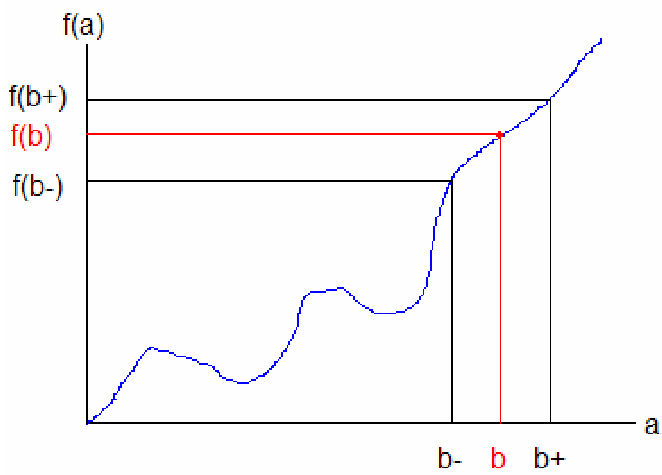


Figure 1.4: Illustration of a continuous function.

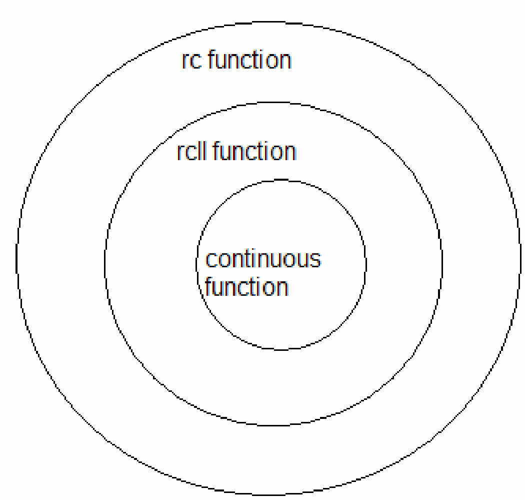


Figure 1.5: Relationship between rc, rcll, and continuous functions.

[1.6] Discontinuous Function

Definition 1.8 Discontinuous function A function $f : a \rightarrow f(a)$ on $\mathbb{R}$ is said to be discontinuous at a point $a = b$ (called a point of discontinuity) if it fails to satisfy being a continuous function.

There are three different categories of points of discontinuities.

Definition 1.9 A function with removable discontinuity (singularity) A function $f : a \rightarrow f(a)$ on $\mathbb{R}$ is said to have a removable discontinuity at a point $a = b$ if it satisfies the following conditions:

(1) $f(b)$ is defined or $f(b)$ is not defined.
(2) Left limit $\lim_{a\to b-} f(a) = f(b-)$ exists. Right limit $\lim_{a\to b+} f(a) = f(b+)$ exists.
(3) $f(b-) = f(b+) \neq f(b)$.

This means that a removable discontinuity at a point $a = b$ looks like a dislocated point as shown by Figure 1.6 where the example is a function:

$$f(a) = \begin{cases} -a+5 & \text{if } \ a \neq 3 \\ 5 \ \text{ if } \ a = 3 \end{cases} .$$

This function has a left limit 2 which is equal to the right limit at a point $a = 3$:

$$f(3-) = f(3+) = 2,$$

but these limits are not equal to the actual value that this function takes at a point $a = 3$:

$$f(3-) = f(3+) = 2 \neq f(3) = 5,$$

which indicates that f is discontinuous at a point $a = b$.

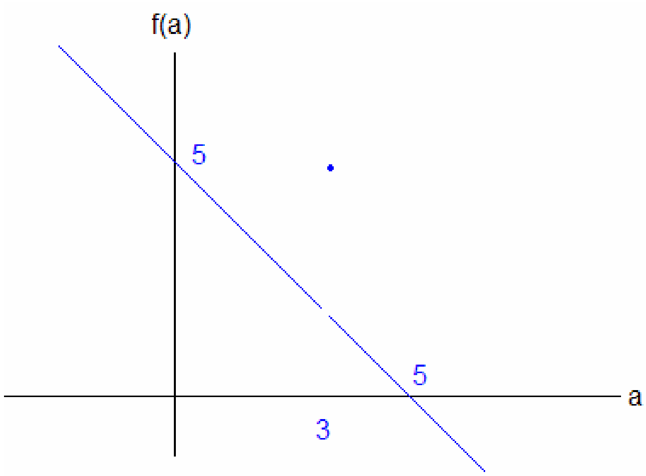


Figure 1.6: Example of a removable discontinuity with the defined discontinuity point $f(3) = 5$.

Or, consider a function:

$$f(a) = \frac{a^2 - 25}{a - 5},$$

which is undefined at a point $a = 5$. But its left limit and right limit exist and they are equal:

$$f(5-) = f(5+) = 10.$$

Therefore, it is a function with removable discontinuity, too.

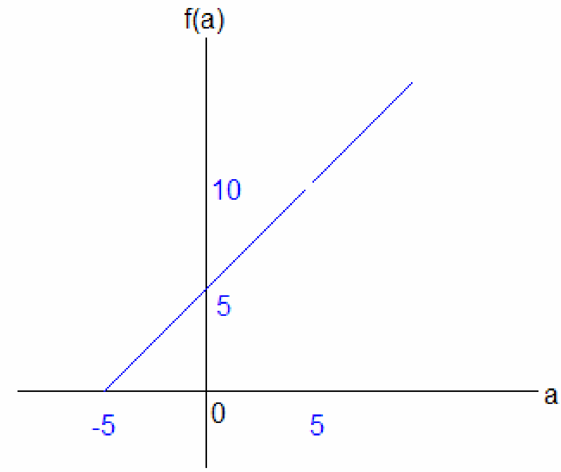


Figure 1.7: Example of a removable discontinuity with the undefined discontinuity point $f(5)$.

Definition 1.10 A function with discontinuity of the first kind (jump discontinuity)
A function $f : a \to f(a)$ on $\mathbb{R}$ is said to have a jump discontinuity at a point $a = b$ if it satisfies the following conditions:

(1) $f(b)$ is defined. In other words, a point b is in the domain of a function f.
(2) Left limit $f(b-)$ exists. Right limit $f(b+)$ exists.
(3) $f(b-) \neq f(b+)$.

Then, the jump is defined by the amount $f(b+) - f(b-)$.

Consider a function:

$$f(a) = \begin{cases} 1 & \text{if} \quad a > 1 \\ 0 & \text{if} \quad a = 1 \\ -1 & \text{if} \quad a < 1 \end{cases}. \tag{1.3}$$

This function has a left limit -1 which is not equal to the right limit 1 at a point $a = 1$:

$$f(1-) = -1 \neq f(1+) = 1,$$

and the jump size is:

$$f(1+) - f(1-) = 1 - (-1) = 2.$$

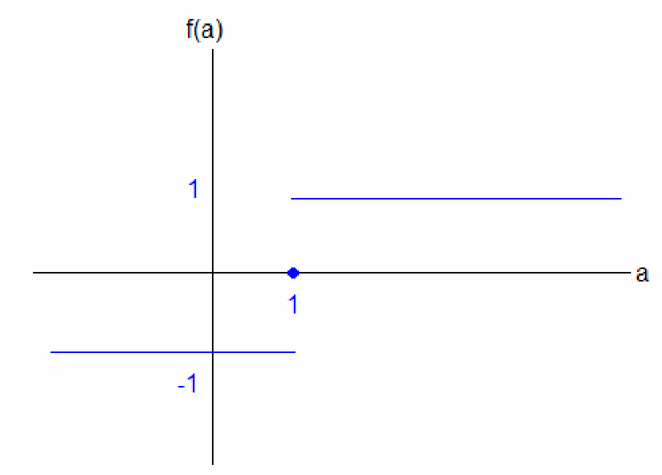


Figure 1.8: Example of a jump discontinuity.

Definition 1.11 A function with discontinuity of the second kind (essential discontinuity) A function $f : a \to f(a)$ on $\mathbb{R}$ is said to have an essential discontinuity at a point $a = b$ if either (or both) of left limit $f(b-)$ or right limit $f(b+)$ does not exist.

The typical example of an essential discontinuity given in most textbooks is the function:

$$f(a) = \begin{cases} \sin(1/a) & \text{if} \quad a \neq 0 \\ 0 \quad \text{if} \quad a = 0 \end{cases}, \tag{1.4}$$

which does not have both left limit $f(b-)$ and right limit $f(b+)$ as shown by Figure 1.9.

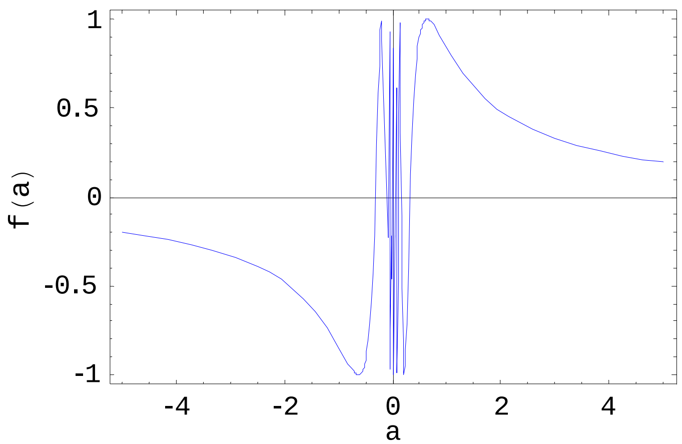


Figure 1.9: Example of an essential discontinuity.

Figure 1.10 illustrates the relationship between rcll, continuous, and discontinuous functions.

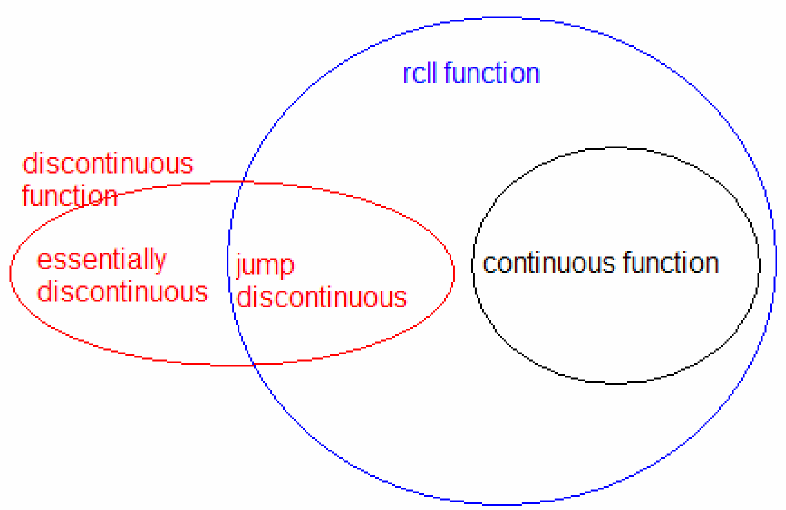


Figure 1.10: Relationship between rcll, continuous, and discontinuous functions.

[2] Stochastic Process

A stochastic process is a collection of random variables:

$$(X_{t\in[0,T]}),$$

where the index denotes time. Note that we are interested in the continuous time stochastic process where the time index takes any value in the interval $t\in[0,T]$ (or it could be an infinite horizon). Discrete time stochastic process can be defined using a countable index set $t\in\mathbb{N}$:

$$(X_{t\in\mathbb{N}}).$$

A stochastic process is defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$ (see Appendix 1 and 2 for details) where Ω is an arbitrary set and $\mathbb{P}$ is a probability measure on $\mathcal{F}_{t\in[0,T]}$. $\mathcal{F}_{t\in[0,T]}$ is called a filtration which is an increasing family of σ-algebras of a subset of Ω which satisfy for $\forall 0\le s\le t$:

$$\mathcal{F}_s\subseteq\mathcal{F}_t.$$

Intuitively speaking, a filtration is an increasing information flow about $(X_{t\in[0,T]})$ as time progresses.

We can alternatively state that a continuous time stochastic process is a random function:

$$X:[0,T]\times\Omega\to\mathbb{R}.$$

After the realization of the randomness ω, a sample path of $(X_{t\in[0,T]})$ is defined as:

$$X(\omega): t \to \mathbb{R} \text{ or } X(\omega): t \to X_t(\omega).$$

A stochastic process $(X_{t\in[0,T]})$ is said to be nonanticipating with respect to the filtration $\mathcal{F}_t$ or $\mathcal{F}_t$-adapted if the value of X_t is revealed at time t for each $t \in [0, T]$.

[2.1] Convergence of Random Variables

Definition 2.1 Pointwise convergence Let $\left(X_{n\in\mathbb{N}}(\omega)\right)$ be a sequence of real valued random variables on a space $(\Omega, \mathcal{F}, \mathbb{P})$ under a scenario (i.e. event or randomness) $\omega \in \Omega$. A sequence $\left(X_{n\in\mathbb{N}}(\omega)\right)$ is said to converge pintwisely to a random variable X if:

$$\lim_{n\to\infty} X_n(\omega) = X .$$

Pointwise convergence is the strongest notion of convergence because it requires convergence to a random variable X for all scenarios (samples) $\omega \in \Omega$, i.e. even for those scenarios with zero probability.

Definition 2.2 Almost sure convergence Let $\left(X_{n\in\mathbb{N}}(\omega)\right)$ be a sequence of real valued random variables on a space $(\Omega, \mathcal{F}, \mathbb{P})$ under a scenario $\omega \in \Omega$. A sequence $\left(X_{n\in\mathbb{N}}(\omega)\right)$ is said to converge almost surely to a random variable X if:

$$\mathbb{P}\left(\lim_{n\to\infty} X_n(\omega) = X\right) = 1.$$

Almost sure convergence is weaker than pointwise convergence since those samples $\omega \in \Omega$ with non convergence $\lim_{n\to\infty} X_n(\omega) \neq X$ have zero probability:

$$\mathbb{P}\left(\lim_{n\to\infty} X_n(\omega) = X\right) + \mathbb{P}\left(\lim_{n\to\infty} X_n(\omega) \neq X\right) = 1 + 0 = 1.$$

Almost sure convergence is used in the strong law of large numbers. Almost sure convergence implies convergence in probability which in turn implies convergence in distribution.

Definition 2.3 Convergence in probability Let $\left(X_{n\in\mathbb{N}}\right)$ be a sequence of real valued random variables on a space $(\Omega, \mathcal{F}, \mathbb{P})$. A sequence $\left(X_{n\in\mathbb{N}}\right)$ is said to converge in probability to a random variable X if for every $\varepsilon \in \mathbb{R}^+$:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|X_n - X\right| > \varepsilon\right) = 0,$$

or equivalently:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|X_n - X\right| \le \varepsilon\right) = 1.$$

Intuitively speaking, convergence in probability means that the probability of X_n getting closer to X rises (and eventually converges to 1) as we take n larger and larger. Convergence in probability is used in the weak law of large numbers. Convergence in probability implies convergence in distribution.

Definition 2.4 Convergence in mean square Let $\left(X_{n\in\mathbb{N}}\right)$ be a sequence of real valued random variables on a space $(\Omega, \mathcal{F}, \mathbb{P})$. A sequence $\left(X_{n\in\mathbb{N}}\right)$ is said to converge in mean square to a random variable X if for every $\varepsilon \in \mathbb{R}^+$:

$$\lim_{n\to\infty} E\left(\left|X_n - X\right|^2\right) = 0.$$

Convergence in mean square implies convergence in probability following Chebyshev's inequality.

Definition 2.5 Convergence in distribution (Weak convergence) Let $\left(X_{n\in\mathbb{N}}\right)$ be a sequence of real valued random variables on a space $(\Omega, \mathcal{F}, \mathbb{P})$. A sequence $\left(X_{n\in\mathbb{N}}\right)$ is said to converge in distribution to a random variable X if for $z \in \mathbb{R}$:

$$\lim_{n\to\infty} \mathbb{P}\left(X_n \le z\right) = \mathbb{P}\left(X \le z\right).$$

Loosely speaking, convergence in distribution means that only when we take sufficiently large n, the probability that X_n is in the interval $[a,b]$ approaches the probability that X is in the interval $[a,b]$. Convergence in distribution is the weakest definition of convergence in the sense that it does not imply any other convergence but implied by all other notions of convergence listed above.

[2.2] Law of Large Numbers and Central Limit Theorem

Definition 2.6 Weak law of large numbers Let $X_1, X_2, X_3 \ldots$ be *i.i.d* random variables from a distribution with mean μ and variance $\sigma^2 < \infty$. Define its sample mean as:

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Then, the sample mean $\overline{X}_n$ converges in probability to the (population) mean μ:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) = 0,$$

for any $\varepsilon \in \mathbb{R}^+$. Or equivalently:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = 1.$$

Definition 2.7 Strong law of large numbers Let $X_1, X_2, X_3 \ldots$ be *i.i.d* random variables from a distribution with mean $\mu < \infty$. Define its sample mean as:

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Then, the sample mean $\overline{X}_n$ converges almost surely to the (population) mean μ:

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1,$$

for any $\varepsilon \in \mathbb{R}^+$.

Definition 2.8 Central limit theorem Let $X_1, X_2, X_3 \ldots$ be *i.i.d* random variables from a distribution with mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Define the sum of a sequence of random variables as:

$$S_n = X_1 + X_2 + ... + X_n.$$

We know the followings:

$$E\left[S_n\right] = E\left[X_1\right] + \left[X_2\right] + ...\left[X_n\right] = n\mu,$$
$$Var\left[S_n\right] = Var\left[X_1\right] + Var\left[X_2\right] + ... + Var\left[X_n\right] = n\sigma^2.$$

Then, informally, the sum S_n converges in distribution to a normal distribution with mean $n\mu$ and variance $n\sigma^2$ as $n \to \infty$:

$$\lim_{n\to\infty} \mathbb{P}\left(S_n \le b\right) = \mathbb{P}\left(Y \le b\right)$$
$$= \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left\{-\frac{1}{2}\frac{\left(Y - n\mu\right)^2}{n\sigma^2}\right\} dY,$$

where Y is a normal random variable, i.e. $Y \sim N\left(n\mu, n\sigma^2\right)$.

Formal central limit theorem is a standardization of the above informal one. Define a random variable Z_n as:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Then, Z_n converges in distribution to the standard normal distribution as $n \to \infty$:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left(Z_n \le b\right) &= \mathbb{P}\left(Z \le b\right) \\ &= \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}Z^2\right) dZ,\end{aligned}$$

where Z is the standard normal random variable, i.e. $Z \sim N\left(0,1\right)$.

[2.3] Inequalities

Definition 2.9 Markov's inequality Let X be a nonnegative random variable. Then, for any $b \in \mathbb{R}^+$:

$$\mathbb{P}\left(X \ge b\right) \le \frac{E\left[X\right]}{b}.$$

Proof

$$E\left[X\right] = \int_0^\infty X d\mathbb{P}\left(X\right) = \int_0^b X d\mathbb{P}\left(X\right) + \int_b^\infty X d\mathbb{P}\left(X\right).$$

This means:

$$E\left[X\right] \ge \int_b^\infty X d\mathbb{P}\left(X\right) \ge \int_b^\infty b d\mathbb{P}\left(X\right) = b\int_b^\infty d\mathbb{P}\left(X\right) = b\mathbb{P}\left(X \ge b\right).$$

Thus:

$$\frac{E\left[X\right]}{b} \ge \mathbb{P}\left(X \ge b\right).$$

□

Markov's inequality provides an upper bound of the probability that a nonnegative random variable is greater than an arbitrary positive constant b by relating a probability to an expectation. A variant of Markov's inequality is called Chebyshev's inequality.

Definition 2.10 Chebyshev's inequality Let X be a random variable on $\mathbb{R}$ (i.e. both $\mathbb{R}^+$ and $\mathbb{R}^-$) with mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Then, for any $k \in \mathbb{R}^+$:

$$\mathbb{P}\left(\left|X - \mu\right| \geq k\right) \leq \frac{\sigma^2}{k^2}.$$

Proof

Start with Markov's inequality:

$$\mathbb{P}\left(X \geq b\right) \leq \frac{E\left[X\right]}{b}.$$

Replace a random variable X with a random variable $(X-\mu)^2$ and b with k^2:

$$\mathbb{P}\left((X-\mu)^2 \geq k^2\right) \leq \frac{E\left[(X-\mu)^2\right]}{k^2} = \frac{\sigma^2}{k^2},$$

which in turn indicates:

$$\mathbb{P}\left(\left|X - \mu\right| \geq k\right) \leq \frac{\sigma^2}{k^2},$$
$$\mathbb{P}\left(\left|X - \mu\right| \geq k\sigma\right) \leq \frac{1}{k^2}.$$

□

Chebyshev's inequality provides bounds of random variables from any distributions as long as their means and variances are known. For example, when $k = 2$:

$$\mathbb{P}\left(\left|X - \mu\right| \geq 2\sigma\right) \leq \frac{1}{4}$$
$$\mathbb{P}\left(-X + \mu \geq 2\sigma, X - \mu \geq 2\sigma\right) \leq \frac{1}{4}$$
$$\mathbb{P}\left(X \leq \mu - 2\sigma, X \geq \mu + 2\sigma\right) \leq \frac{1}{4}$$
$$\mathbb{P}\left(\mu - 2\sigma \leq X \leq \mu + 2\sigma\right) \geq \frac{3}{4}.$$

This tells us that the probability that any random variable lies within two standard deviations is at least .75.

Definition 2.11 Cauchy-Schwarz's inequality Let X and Y be jointly distributed random variables on $\mathbb{R}$ with each having finite variance. Then:

$$\left(E[XY]\right)^2 \leq E[X^2]E[Y^2].$$

Proof

Omitted.

[3] Putting Structure on Stochastic Processes

The purpose of any mathematical (statistical) modeling regardless of the field is to fit less complicated models to the highly complicated real world phenomena as accurate as possible. Mathematical models are less complicated in the sense that they make some simplifying assumptions or put some simplifying structures (restrictions) on the real world phenomena for the purpose of gaining tractability. There are some popular dependence structures put on stochastic processes which mathematicians have developed and used for years.

[3.1] Process with Independent and Stationary Increments: Imposing Structure on a Probability Measure $\mathbb{P}$

Before giving the definition of processes with independent and stationary increments, we must know the basics.

Definition 3.1 Conditional probability The conditional probability of an arbitrary event A given an event with positive probability B is:

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

When $\mathbb{P}(B) = 0$, $\mathbb{P}(A|B)$ is undefined.

Definition 3.2 Statistical (Stochastic) independence Two arbitrary events A and B are said to be independent, if and only if:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

This definition of independence has two advantages. Firstly, it is symmetric in A and B. In other words, an event A's independence of an event B implies an event B's independence of an event A. Secondly, this definition holds even when an event B has zero probability, i.e. $\mathbb{P}(B) = 0$.

When two arbitrary events A and B are independent, from the definition of a conditional probability:

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

It is important to note that this is a result of statistical independence and not the definition. This is because the above equation is not true (i.e. undefined) when $\mathbb{P}(B) = 0$ and it is not symmetric in that $\mathbb{P}(A|B) = \mathbb{P}(A)$ does not necessarily imply $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Definition 3.3 Mutual statistical independence Arbitrary events A_1, A_2, …, A_n are said to be mutually independent, if and only if:

$$\mathbb{P}(A_1 \cap A_2 \cap ... \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_n).$$

Definition 3.4 Processes with Independent and Stationary Increments A stochastic process $\left(X_{t\in[0,T]}\right)$ with values in $\mathbb{R}$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be a process with independent and stationary increments if it satisfies the following conditions:

(1) Its increments are independent. In other words, for $t_1 < t_2 < ... < t_n$:

$$\mathbb{P}(X_{t_2} - X_{t_1} \cap X_{t_3} - X_{t_2} \cap ... \cap X_{t_n} - X_{t_{n-1}}) = \mathbb{P}(X_{t_2} - X_{t_1})\mathbb{P}(X_{t_3} - X_{t_2})...\mathbb{P}(X_{t_n} - X_{t_{n-1}}).$$

(2) Its increments are stationary: i.e. for $\forall h \in \mathbb{R}^+$, $X_{t+h} - X_t$ has the same distribution as X_h. In other words, the distribution of increments does not depend on t (i.e. temporal homogeneity).

Consider an increasing sequence of time $0 < t_1 < t_2 < ... < t_n < t < u < \infty$ where t is the present. As a result of independent increments condition:

$$\begin{aligned}
&\mathbb{P}(X_u - X_t \,|\, X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n}) \\
&= \frac{\mathbb{P}(X_u - X_t \cap X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n})}{\mathbb{P}(X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n})} \\
&= \frac{\mathbb{P}(X_u - X_t)\mathbb{P}(X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n})}{\mathbb{P}(X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n})} \\
&= \mathbb{P}(X_u - X_t),
\end{aligned}$$

which means that there is no correlation (probabilistic dependence structure) on the increments among the past, the present, and the future.

For example, independent increments condition means that when modeling a log stock price $\ln S_t$ as an independent increment process, the probability distribution of a log stock price in year 2005 – 2006 is independent of the way the log stock price increment has evolved over the years (i.e. stock price dynamics), i.e. it doesn't matter if this stock crushes or soars in year 2004 – 2005):

$$\mathbb{P}(\ln S_{2006} - \ln S_{2005} \,|\, ..., \ln S_{2003} - \ln S_{2002}, \ln S_{2004} - \ln S_{2003}, \ln S_{2005} - \ln S_{2004})$$
$$= \mathbb{P}(\ln S_{2006} - \ln S_{2005}) .$$

Using the simple relationship $X_u \equiv (X_u - X_t) + X_t$ for an increasing sequence of time $0 < t_1 < t_2 < ... < t_n < t < u < \infty$:

$$\mathbb{P}(X_u \,|\, X_0, X_{t_1}, X_{t_2}, ..., X_{t_n}, X_t) = \mathbb{P}((X_u - X_t) + X_t \,|\, X_0, X_{t_1}, X_{t_2}, ..., X_{t_n}, X_t)$$
$$= \mathbb{P}(X_u \,|\, X_t) ,$$

which holds because an increment $(X_u - X_t)$ is independent of X_t by definition and the value of X_t depends on its realization $X_t(\omega)$. This is a strong probabilistic structure imposed on a stochastic process because this means that the conditional probability of the future value X_u depends only on the previous realization $X_t(\omega)$ and not on the entire past history of realizations $X_0, X_{t_1}, X_{t_2}, ..., X_{t_n}, X_t$ (i.e. called Markov property which is discussed soon).

Although this condition seems too strong, it imposes a very tractable property on the process. Because if two variables X and Y are independent:

$$E[XY] = E[X]E[Y] ,$$
$$Var[X+Y] = Var[X] + Var[Y] ,$$
$$Cov[X,Y] = 0 \text{ (i.e. } Corr[X,Y] = 0 \text{)}.$$

Stationary increments condition means that the distributions of increments $X_{t+h} - X_t$ do not depend on the time t, but they depend on the time-distance h of two observations (i.e. interval of time). In other words, the probability density function of increments does not change over time. For example, if you model a log stock price $\ln S_t$ as a process with stationary increments, the distribution of increment in year 2005 – 2006 is the same as that in year 2050 – 2051:

$$\ln S_{2006} - \ln S_{2005} \overset{d}{=} \ln S_{2051} - \ln S_{2050} .$$

There is no doubt that the above independent and stationary increments conditions impose a strong structure on a stochastic process $\left(X_{t\in[0,T]}\right)$, as a result of these restrictions, the mean and variance of the process is tractable:

$$E[X_t] = \mu_0 + \mu_1 t,$$
$$Var[X_t] = \sigma_0^{\ 2} + \sigma_1^{\ 2} t,$$

where $\mu_0 = E[X_0]$, $\mu_1 = E[X_1] - \mu_0$, $\sigma_0^{\ 2} = E[(X_0 - \mu_0)^2]$, and $\sigma_1^{\ 2} = E[(X_1 - \mu_1)^2] - \sigma_0^{\ 2}$.

[3.2] Martingale: Structure on Conditional Expectation

[3.2.1] Definition of Martingale

Originally, the word 'martingale' comes from a French acronym of a gambling strategy. Imagine a coin flip gamble in which you win if a head turns up and you lose if a tail turns up. Martingale strategy requires a gambler to double his bet after every loss. Following martingale strategy, a gambler can recover all the losses he made and end up with an initial amount of his wealth plus an initial bet. Table 3.1 gives a sample path of a martingale strategy in which a gambler initially owns \$200 of wealth, start betting with a stake of \$2, and due to his bad luck his first win comes at the seventh trial. As you can see, he basically ends up where he started, i.e. his initial wealth of \$200 (plus an initial bet of \$2). Thus, a martingale strategy tells that after gambling many hours a gambler gains nothing (loses nothing) and his wealth remains constant on average.

Table 3.1 Martingale Gambling Strategy

Trial	0	1	2	3	4	5	6	7
Result		Loss	Loss	Loss	Loss	Loss	Loss	Win
Bet		\$2	\$4	\$8	\$16	\$32	\$64	\$128
Net Gain		-\$2	-\$4	-\$8	-\$16	-\$32	-\$64	+\$128
Wealth	\$200	\$198	\$194	\$186	\$170	\$138	\$74	+\$202

In probability theory, a stochastic process is said to be a martingale if its sample path has no trend. Formally, a martingale is defined as the follows.

Definition 3.5 Martingale Consider a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$. A rcll stochastic process $(X_t)_{t\in[0,\,T]}$ is said to be a martingale with respect to the filtration $\mathcal{F}_t$ and under the probability measure $\mathbb{P}$ if it satisfies the following conditions:

(1) X_t is nonanticipating.

(2) $E[|X_t|] < \infty$ for $\forall t \in [0,\ T]$. Finite mean condition.

(3) $E[X_u | \mathcal{F}_t] = X_t$ for $\forall u > t$.

In other words, if a stochastic process is a martingale, then, the best prediction of its future value is its present value. Note that the definition of martingale makes sense only when the underlying probability measure P and the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ have been specified.

The fundamental property of a martingale process is that its future variations are completely unpredictable with the filtration $\mathcal{F}_t$:

$$\forall u > 0,\ E[x_{t+u} - x_t \,|\mathcal{F}_t] = E[x_{t+u}\,|\mathcal{F}_t] - E[x_t\,|\mathcal{F}_t] = x_t - x_t = 0 .$$

Finite mean condition is necessary to ensure the existence of the conditional expectation.

[3.2.2] An Example of Continuous Martingale: A Standard Brownian Motion

Let $(B_{t\in[0,\infty)})$ be a standard Brownian motion process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, $(B_{t\in[0,\infty)})$ is a continuous martingale with respect to the filtration $\mathcal{F}_{t\in[0,\infty)}$ and the probability measure $\mathbb{P}$.

Proof

By definition, $(B_{t\in[0,\infty)})$ is a nonanticipating process (i.e. $\mathcal{F}_{t\in[0,\infty)}$ - adapted process) with the finite mean $E[|B_t|] = 0 < \infty$ for $\forall t \in [0,\infty)$. For $\forall 0 \le t \le u < \infty$:

$$B_u = B_t + \int_t^u dB_v .$$

Using the fact that a Brownian motion is a nonanticipating process, i.e. $E[B_t\,|\mathcal{F}_t] = B_t$:

$$\begin{aligned}
E[B_u - B_t\,|\mathcal{F}_t] &= E[B_u\,|\mathcal{F}_t] - E[B_t\,|\mathcal{F}_t] = E[B_t + \int_t^u dB_v\,|\mathcal{F}_t] - B_t \\
E[B_u - B_t\,|\mathcal{F}_t] &= E[B_t\,|\mathcal{F}_t] + E[\int_t^u dB_v\,|\mathcal{F}_t] - B_t \\
E[B_u - B_t\,|\mathcal{F}_t] &= B_t + 0 - B_t = 0 ,
\end{aligned}$$

or in other words:

$$\begin{aligned}
E[B_u\,|\mathcal{F}_t] &= E[B_t + \int_t^u dB_v\,|\mathcal{F}_t] = E[B_t\,|\mathcal{F}_t] + E[\int_t^u dB_v\,|\mathcal{F}_t] = B_t + 0 \\
E[B_u\,|\mathcal{F}_t] &= B_t ,
\end{aligned}$$

which is a martingale condition.

□

Let $(B_{t\in[0,\infty)})$ be a standard Brownian motion process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, a Brownian motion with drift $(X_{t\in[0,\infty)}) \equiv (\mu t + \sigma B_{t\in[0,\infty)})$ is not a continuous martingale with respect to the filtration $\mathcal{F}_{t\in[0,\infty)}$ and the probability measure $\mathbb{P}$.

Proof

By definition, $(X_{t\in[0,\infty)})$ is a nonanticipating process (i.e. $\mathcal{F}_{t\in[0,\infty)}$- adapted process) with the finite mean $E[X_t] = E[\mu t + \sigma B_t] = \mu t < \infty$ for $\forall t \in [0,\infty)$ and $\mu \in \mathbb{R}$. For $\forall 0 \le t \le u < \infty$:

$$X_u = X_t + \int_t^u dX_v \,.$$

Using the fact that a Brownian motion with drift is a nonanticipating process, i.e. $E[X_t | \mathcal{F}_t] = X_t$:

$$E[X_u | \mathcal{F}_t] = E[X_t + \int_t^u dX_v | \mathcal{F}_t] = E[X_t | \mathcal{F}_t] + E[\int_t^u dX_v | \mathcal{F}_t]$$
$$E[X_u | \mathcal{F}_t] = X_t + \mu(u-t),$$

which violates a martingale condition.

□

But one way to transform nonmartingales into martingales is to make the process driftless. In other words, eliminating the trend of the process which is sometimes called a detrending. Consider the following example.

A detrended Brownian motion with drift defined as:

$$(X_{t\in[0,\infty)} - \mu t) \equiv (\mu t + \sigma B_{t\in[0,\infty)} - \mu t) \equiv (\sigma B_{t\in[0,\infty)}),$$

is a continuous martingale with respect to the filtration $\mathcal{F}_{t\in[0,\infty)}$ and the probability measure $\mathbb{P}$.

Proof

For $\forall 0 \le t \le u < \infty$:

$$E[X_u - \mu u | \mathcal{F}_t] = E[(X_t - \mu t) + (\int_t^u dX_v - \mu \int_t^u dv) | \mathcal{F}_t]$$

$$E[X_u - \mu u | \mathcal{F}_t] = E[(X_t - \mu t) | \mathcal{F}_t] + E[(\int_t^u dX_v - \mu \int_t^u dv) | \mathcal{F}_t]$$
$$E[X_u - \mu u | \mathcal{F}_t] = X_t - \mu t + \mu(u - t) - \mu(u - t)$$
$$E[X_u - \mu u | \mathcal{F}_t] = X_t - \mu t \,,$$

which satisfies a martingale condition.

□

[3.2.3] Martingale Asset Pricing

Most of financial asset prices are not martingales because they are not completely unpredictable and most financial time series have trends. Consider a stock price process $\{S_t; 0 \le t \le T\}$ on a filtered probability space $(\Omega, \mathcal{F}_{t \in [0,T]}, \mathbb{P})$ and let r be the risk-free interest rate. In a small time interval Δ, risk-averse investors expect S_t to grow at some positive rate. This can be written as under actual probability measure $\mathbb{P}$:

$$E^{\mathbb{P}}[S_{t+\Delta} | \mathcal{F}_t] > S_t \,.$$

This means that a stock price S_t is not martingale under $\mathbb{P}$ and with respect to $\mathcal{F}_t$. To be more precise, risk-averse investors expect S_t to grow at a rate greater than r because a stock is risky:

$$E^{\mathbb{P}}[e^{-r\Delta} S_{t+\Delta} | \mathcal{F}_t] > S_t \,.$$

The stock price discounted by the risk-free interest rate $e^{-r\Delta} S_{t+\Delta}$ is not martingale under $\mathbb{P}$ and with respect to $\mathcal{F}_t$.

How can we convert a discounted stock price $e^{-r\Delta} S_{t+\Delta}$ into a martingale? First approach is to eliminate the trend. The trend in this case is the risk premium π which risk-averse investors demand for bearing extra amount of risk. If we can estimate π correctly, a discounted stock price $e^{-r\Delta} S_{t+\Delta}$ can be converted into a martingale by detrending:

$$E^{\mathbb{P}}[e^{-\pi\Delta} e^{-r\Delta} S_{t+\Delta} | \mathcal{F}_t] = E^{\mathbb{P}}[e^{-(r+\pi)\Delta} S_{t+\Delta} | \mathcal{F}_t] = S_t \,.$$

But this approach involves the rather difficult job of estimating the risk premium π and is not used in quantitative finance. Martingale asset pricing uses the second approach to convert non-martingales into martingales by changing the probability measure. We will try to find an equivalent probability measure $\mathbb{Q}$ (called risk-neutral measure) under which a discounted stock price becomes martingale:

$$E^{\mathbb{Q}}[e^{-r\Delta} S_{t+\Delta} | \mathcal{F}_t] = S_t \,.$$

[3.2.4] Submartingales and Supermartingales

Definition 3.6 Submartingale Consider a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$. A rcll stochastic process $(X_t)_{t\in[0,\ T]}$ is said to be a submartingale with respect to the filtration $\mathcal{F}_t$ and under the probability measure $\mathbb{P}$ if it satisfies the following conditions:

(1) X_t is nonanticipating.
(2) $E[|X_t|] < \infty$ for $\forall t \in [0,\ T]$. Finite mean condition.
(3) $E[X_u | \mathcal{F}_t] \geq X_t$ for $\forall u > t$.

Intuitively, a submartingale is a stochastic process with a positive (upward) trend. A submartingale gains or grows on average as time progresses.

Definition 3.7 Supermartingale Consider a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$. A rcll stochastic process $(X_t)_{t\in[0,\ T]}$ is said to be a supermartingale with respect to the filtration $\mathcal{F}_t$ and under the probability measure $\mathbb{P}$ if it satisfies the following conditions:

(1) X_t is nonanticipating.
(2) $E[|X_t|] < \infty$ for $\forall t \in [0,\ T]$. Finite mean condition.
(3) $E[X_u | \mathcal{F}_t] \leq X_t$ for $\forall u > t$.

Intuitively, a supermartingale is a stochastic process with a negative (downward) trend. A supermartingale loses or declines on average as time progresses.

By definition, any martingale is a submartingale and a supermartingale.

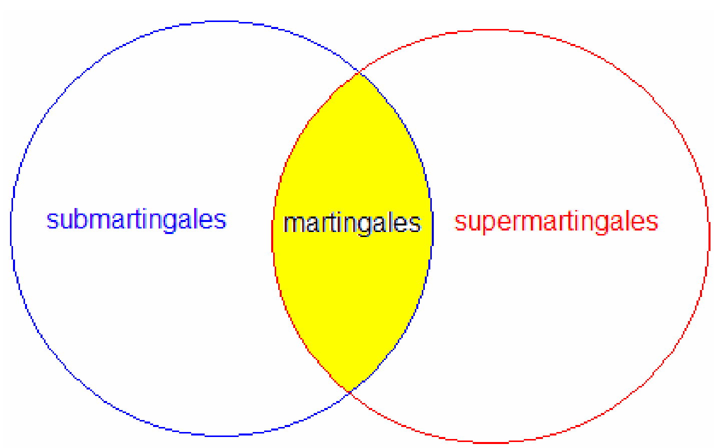


Figure 3.1: Relationship among martingales, submartingales, and supermartingales.

[3.3] Markov Processes: Structure on Conditional Probability

This section gives a brief introduction to a class of stochastic processes called Markov processes which impose a restriction on the conditional probabilities. This differs from martingales which impose a structure on conditional expectations.

[3.3.1] Discrete Time Markov Chains

Definition 3.8 Discrete time Markov chain Consider a discrete time stochastic process $(X_n)_{n\in\mathbb{N}}$ (i.e. $n = 0,1,2,...$) defined on a filtered probability space $(\Omega, \mathcal{F}_{n\in\mathbb{N}}, \mathbb{P})$ which takes values in a countable or a finite set E called a state space of the process. A realization X_n is said to be in state $i \in E$ at time n if $X_n = i$. An E-valued discrete time Markov chain is a stochastic process which satisfies for $\forall n \in \mathbb{N}$ and $\forall i, j \in E$:

$$\mathbb{P}(X_{n+1} = j \mid X_0, X_1, X_2, ..., X_n = i) = \mathbb{P}(X_{n+1} = j \mid X_n = i).$$

This is called a Markov property. Markov property means that the probability of a random variable X_{n+1} at time $n+1$ (tomorrow) being in a state j conditional on the entire history of the stochastic process $(X_0, X_1, X_2, ..., X_n)$ is equal to the probability of a random variable X_{n+1} at time $n+1$ (tomorrow) being in a state j conditional only on the value of a random variable at time n (today). In other words, the history (sample path) of the stochastic process $(X_0, X_1, X_2, ..., X_n)$ is of no importance in that the way this stochastic process evolved or the dynamics $(X_1 - X_0, X_2 - X_1, ...)$ does not mean a thing in terms of the conditional probability of the process. The only factor which influences the conditional probability of a random variable X_{n+1} at time $n+1$ (tomorrow) is the state of a random variable at time n (today).

The probability $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ which is a conditional probability of moving from a state i at time n to a state j at time $n+1$ is called a one step transition probability. In the general cases, transition probabilities are dependent on the states and time such that $\forall m \neq n \in \mathbb{N}$:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) \neq \mathbb{P}(X_{m+1} = j \mid X_m = i).$$

When transition probabilities are independent of time n, discrete time Markov chains are said to be time homogeneous or stationary.

Definition 3.9 Time homogeneous (stationary) discrete time Markov chain Consider a discrete time stochastic process $(X_n)_{n\in\mathbb{N}}$ (i.e. $n = 0,1,2,...$) defined on a filtered probability space $(\Omega, \mathcal{F}_{n\in\mathbb{N}}, \mathbb{P})$ which takes values in a countable or a finite set E called a state space of the process. A realization X_n is said to be in state $i \in E$ at time n

if $X_n = i$. An E-valued time homogeneous discrete time Markov chain is a stochastic process which satisfies for $\forall n \in \mathbb{N}$ and $\forall i, j \in E$:

$$\begin{aligned}\mathbb{P}(X_{n+1} = j \mid X_0, X_1, X_2, ..., X_n = i) &= \mathbb{P}(X_{n+1} = j \mid X_n = i) \\ &= \mathbb{P}(X_1 = j \mid X_0 = i) \\ &= \mathbb{P}(j \mid i).\end{aligned}$$

In other words, transition probabilities do not depend on time n and only depend on transition states from i to j. A matrix of transition probabilities $\mathbb{P} = \left\| \mathbb{P}(j \mid i) \right\|_{i, j \in E}$ is called a transition probability matrix:

$$\left\| \mathbb{P}(j \mid i) \right\|_{i, j \in E} = \left\| \begin{matrix} \mathbb{P}(0|0) & \mathbb{P}(1|0) & \mathbb{P}(2|0) & \mathbb{P}(3|0) & \cdots \\ \mathbb{P}(0|1) & \mathbb{P}(1|1) & \mathbb{P}(2|1) & \mathbb{P}(3|1) & \cdots \\ \mathbb{P}(0|2) & \mathbb{P}(1|2) & \mathbb{P}(2|2) & \mathbb{P}(3|2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ \mathbb{P}(0|i) & \mathbb{P}(1|i) & \mathbb{P}(2|i) & \mathbb{P}(3|i) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{matrix} \right\|.$$

Transition probabilities $\mathbb{P}(j \mid i)$ satisfy the following conditions:

(1) $\mathbb{P}(j \mid i) \geq 0$ for $\forall i, j \in E$.
(2) $\sum_{j \in E} \mathbb{P}(j \mid i) = 1$ for $\forall i \in E$.

Condition (2) guarantees the occurrence of a transition including a case in which the state remains unchanged.

Proposition 3.1 Defining a discrete time Markov chain An E-valued general discrete time Markov chain $(X_n)_{n \in \mathbb{N}}$ is completely defined if it satisfies the following conditions:

(1) All transition probabilities $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n)$ are known.
(2) The probability distribution of the initial (i.e. time 0) state of the Markov chain $\mathbb{P}(X_0 = i_0) = \mathbb{P}_0$ is known.

Proof

Consider obtaining the joint probability distribution of an E-valued general discrete time Markov chain $(X_n)_{n \in \mathbb{N}}$. From the definition of a conditional probability:

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_n = i_n)$$
$$= \mathbb{P}(X_n = i_n \mid X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-1} = i_{n-1}) \mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-1} = i_{n-1}).$$

Since $(X_n)_{n\in\mathbb{N}}$ is a Markov chain:

$$\mathbb{P}(X_n = i_n \mid X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-1} = i_{n-1}) = \mathbb{P}(X_n = i_n \mid X_{n-1} = i_{n-1}) = \mathbb{P}(i_n \mid i_{n-1}).$$

Joint probability can be calculated as:

$$\begin{aligned}
&\mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_n = i_n) \\
&= \mathbb{P}(i_n \mid i_{n-1}) \mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-1} = i_{n-1}) \\
&= \mathbb{P}(i_n \mid i_{n-1}) \mathbb{P}(X_{n-1} = i_{n-1} \mid X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-2} = i_{n-2}) \\
&\quad \times \mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-2} = i_{n-2}) \\
&= \mathbb{P}(i_n \mid i_{n-1}) \mathbb{P}(i_{n-1} \mid i_{n-2}) \mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-2} = i_{n-2}) \\
&= \mathbb{P}(i_n \mid i_{n-1}) \mathbb{P}(i_{n-1} \mid i_{n-2}) ... \mathbb{P}(i_2 \mid i_1) \mathbb{P}(i_1 \mid i_0) \mathbb{P}_0
\end{aligned}$$

□

Consider a transition probability of a time homogeneous discrete time Markov chain $(X_n)_{n\in\mathbb{N}}$ from a state i at time k (i.e. $X_k = i$) to a state j at time $k+n$. This is called a n-step transition probability and expressed as:

$$\mathbb{P}(X_{k+n} = j \mid X_k = i) = \mathbb{P}(X_n = j \mid X_0 = i) = \mathbb{P}^{(n)}(j \mid i).$$

Proposition 3.2 n step transition probability matrix (a special case of Chapman-Kolmogorov equation) Consider a time homogeneous discrete time Markov chain $(X_n)_{n\in\mathbb{N}}$ defined on a filtered probability space $(\Omega, \mathcal{F}_{n\in\mathbb{N}}, \mathbb{P})$ which takes values in a countable or a finite set E called a state space of the process. Then, its n-step transition probability matrix from a state i at time k (i.e. $X_k = i$) to a state j at time $k+n$ is given by for $\forall k, n \in \mathbb{N}$ and $\forall i, j \in E$:

$$\mathbb{P}^{(n)}(j \mid i) = \mathbb{P}^n(j \mid i) = \sum_{v\in E} \mathbb{P}^{(y)}(v \mid i) \mathbb{P}^{(z)}(j \mid v) = \sum_{v\in E} \mathbb{P}^y(v \mid i) \mathbb{P}^z(j \mid v).$$

where $y + z = n$ and $\mathbb{P}^{(0)}(j \mid i)$ is defined as:

$$\mathbb{P}^{(0)}(j \mid i) = \begin{cases} 1 & \text{for} \quad i = j \\ 0 & \text{for} \quad i \neq j \end{cases}.$$

Proof

When $n = 1$:

$$\mathbb{P}^{(1)}(j|i) = \mathbb{P}(j|i).$$

When $n = 2$:

$$\mathbb{P}^{(2)}(j|i) = \sum_{v \in E} \mathbb{P}(v|i)\mathbb{P}(j|v).$$

By induction:

$$\mathbb{P}^{(n+1)}(j|i) = \sum_{v \in E} \mathbb{P}(v|i)\mathbb{P}^{n}(j|v).$$

□

One interesting topic about this n step transition probability matrix is its asymptotic behavior as $n \to \infty$. As n becomes larger, the initial state i becomes less important and in the limit as $n \to \infty$, $\mathbb{P}^{n}(j|i)$ is independent of i. We recommend Karlin and Taylor (1975) for more details.

[3.3.2] Markov Processes

Definition 3.10 Markov Processes (Continuous time Markov chains) Consider a continuous time stochastic process $\left(X_{t\in[0,T]}\right)$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ which takes values in $\mathbb{N}$ (for simplicity) called a state space of the process. $\left(X_{t\in[0,T]}\right)$ is said to be a time homogeneous Markov process if for $\forall h \in \mathbb{R}^{+}$ and $\forall i, j \in \mathbb{N}$:

$$\mathbb{P}_h(j|i) = \mathbb{P}(X_{t+h} = j | \mathcal{F}_t) = \mathbb{P}(X_{t+h} = j | X_t = i).$$

Markov property means that the probability of a random variable X_{t+h} at time $t+h$ (tomorrow) being in a state j conditional on the entire history of the stochastic process $\mathcal{F}_{[0,t]} \equiv X_{[0,t]}$ is equal to the probability of a random variable X_{t+h} at time $t+h$ (tomorrow) being in a state j conditional only on the value of a random variable at time t (today). In other words, the history (sample path) of the stochastic process $\mathcal{F}_{[0,t]}$ is of no importance in that the way this stochastic process evolved or the dynamics does not mean a thing in terms of the conditional probability of the process.

[4] Sample Path Properties of Stochastic Processes

[4.1] Continuous Stochastic Process

In this section we give the formal definition of the continuity of a sample path of a stochastic process which we believe is underrated in the literature. Note that the continuity of path and continuity of time are different subjects. In other words, a continuous time stochastic process does not imply continuous stochastic process.

There are different notions of continuity of sample paths which use different notions of convergence of random variables we saw in section 1.

Definition 4.1 Continuous in mean square A real valued stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be continuous in mean square if for $\forall t \in [0, T]$:

$$\lim_{s\to t} E[\left|X_s - X_t\right|^2] = 0 .$$

Continuity in mean square implies continuity in probability following Chebyshev's inequality.

Definition 4.2 Continuous in probability A real valued stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be continuous in probability if for $\forall t \in [0, T]$ and every $\varepsilon \in \mathbb{R}^+$:

$$\lim_{s\to t} \mathbb{P}\left(\left|X_s - X_t\right| > \varepsilon\right) = 0 ,$$

or equivalently:

$$\lim_{s\to t} \mathbb{P}\left(\left|X_s - X_t\right| \le \varepsilon\right) = 1 .$$

Intuitively speaking, continuity in probability means that the probability of X_s getting closer to X_t rises (and eventually converges to 1) as s approaches t.

For example, a Brownian motion process is continuous in mean square and continuous in probability although its proof is not that easy (refer to any stochastic process textbook for this). But it turns out that the above definitions of continuity are too loose because a Poisson process can be proven to be continuous in mean square and probability (proof is omitted). Therefore, a more strict definition of continuity is used for the definition of a continuity of a sample path of a stochastic process.

Definition 4.3 Continuous stochastic process A real valued nonanticipating stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be (almost surely) continuous if a sample path of the process $(X_{t\in[0,T]}(\omega))$ is almost surely a continuous function for $\forall t \in [0,\ T]$. In other words, a sample path of the process $(X_{t\in[0,T]}(\omega))$ satisfies for $\forall t \in [0,\ T]$:

(1) Right limit of the process as s approaches t from the above (right hand side) exists, i.e. $\lim_{s\to t, s>t} X_s = X_{t+}$. Left limit of the process as s approaches t from the below (left hand side) exists, i.e. $\lim_{s\to t, s<t} X_s = X_{t-}$.

(2) $X_{t+} = X_{t-} = X_t$.

This means that a continuous stochastic process is a right continuous and left continuous stochastic process.

[4.2] Right Continuous with Left Limit (RCLL) Stochastic Process

Definition 4.4 Right continuous with left limit (rcll) stochastic process A real valued nonanticipating stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be a rcll stochastic process if for $\forall t \in [0,\ T]$:

(1) Right limit of the process as s approaches t from the above (right hand side) exists, i.e. $\lim_{s\to t, s>t} X_s = X_{t+}$. Left limit of the process as s approaches t from the below (left hand side) exists, i.e. $\lim_{s\to t, s<t} X_s = X_{t-}$.

(2) $X_{t+} = X_t$.

In other words, only the right continuity is needed (this allows jumps). Apparently, a continuous stochastic process implies a rcll stochastic process (but the reverse is not true). What we encounter in finance literatures are all rcll stochastic processes (for the modeling of stock price dynamics. Rcll processes include jump discontinuous process such as Poisson processes and infinite activity Lévy processes. Essentially discontinuous processes are useless in finance because they don’t have either (or both) of the left limit X_{t-} or the right limit X_{t+}.

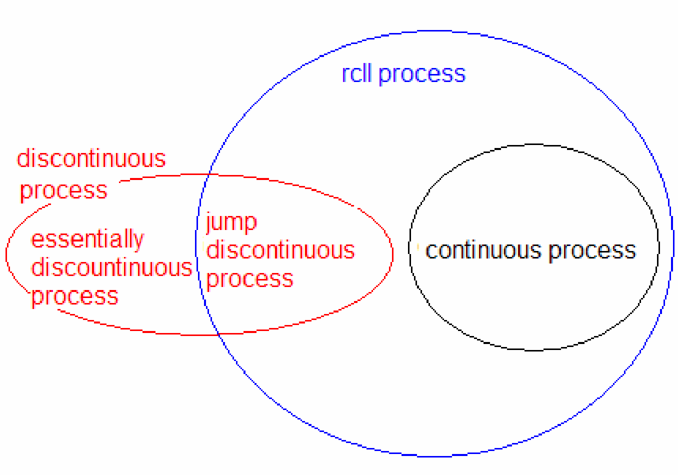
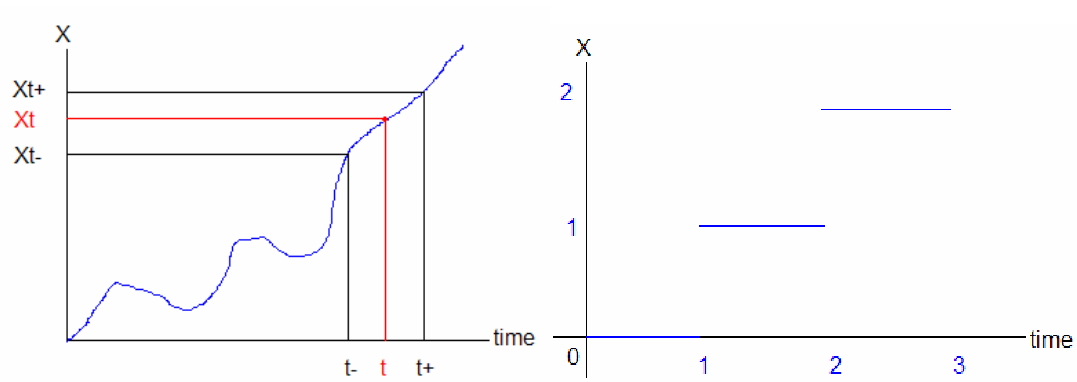


Figure 4.1: Relationship between rcll, continuous, and jump discontinuous processes.



A) A continuous stochastic process. B) A jump discontinuous stochastic process.
Figure 4.2: Examples of rcll stochastic processes.

Suppose t is a discontinuity point. The jump of the stochastic process at t is defined as:

$$\Delta X_t = X_t - X_{t-} .$$

A rcll nonanticipating stochastic process $(X_{t \in [0,T]})$ can have a finite number of large jumps and countable number (possibly infinite) of small jumps.

[4.3] Total Variation

Definition 4.5 Total variation of a function Let $f(x)$ be a bounded function defined in the interval $[a,b]$:

$$f(x) : [a,b] \to \mathbb{R} .$$

The interval can be infinite, i.e. $[-\infty, \infty]$. Consider partitioning the interval $[a,b]$ with the points:

$$a = x_0 < x_1 < x_2 ... x_{n-1} < x_n = b .$$

Then, the total variation of a function $f(x)$ is defined by:

$$T(f) = \sup \sum_{i=1}^{n} \left| f(x_i) - f(x_{i-1}) \right| ,$$

where sup indicates a supremum (least upper bound).

Definition 4.6 Function of finite variation A function $f(x)$ on the interval $[a,b]$ is said to be a function of finite variation, if its total variation on the interval $[a,b]$ is finite:

$$T(f) = \sup \sum_{i=1}^{n} \left| f(x_i) - f(x_{i-1}) \right| < \infty .$$

Proposition 4.1 Every bounded increasing or decreasing function is of finite variation on the interval $[a,b]$.

Proof

Consider an increasing function $f(x)$ on the interval $[a,b]$. By its definition, for $\forall i$:

$$f(x_i) - f(x_{i-1}) \geq 0 .$$

And:

$$T(f) = \sup \left\{ f(x_n) - f(x_{n-1}) + f(x_{n-1}) - f(x_{n-2}) + ... + f(x_2) - f(x_1) + f(x_1) - f(x_0) \right\}$$
$$T(f) = \sup \left\{ f(x_n) - f(x_0) \right\}$$
$$T(f) = \sup \left\{ f(b) - f(a) \right\} ,$$

which is finite because $f(x)$ is bounded:

$$-\infty < f(a), f(b) < \infty .$$

Definition 4.7 Total variation of a stochastic process Consider a real valued stochastic process $(X_{t \in [0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t \in [0,T]}, \mathbb{P})$. Partition the time interval $[0,T]$ with the points:

$$0 = t_0 < t_1 < t_2 ... < t_{n-1} < t_n = T .$$

Then, the total variation of a stochastic process $(X_{t\in[0,T]})$ on the time interval $[0,T]$ is defined by:

$$T(X) = \sup \sum_{i=1}^{n} \left| X(t_i) - X(t_{i-1}) \right|,$$

where sup indicates a supremum (least upper bound).

Definition 4.8 Stochastic process of finite variation A real valued stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ on the interval $[0,T]$ is said to be a stochastic process of finite variation, if the total variation on the interval $[0,T]$ of a sample path of the process is finite with probability 1:

$$\mathbb{P}(T(X) = \sup \sum_{i=1}^{n} \left| X(t_i) - X(t_{i-1}) \right| < \infty) = 1.$$

References

Brzezniak, Z., and Zastawniak, T., 1999, Basic Stochastic Processes, Springer.

Cont, R., and Tankov, P., 2004, Financial Modelling with Jump Processes, Chapman & Hall/CRC Financial Mathematics Series.

Feller, W., 1968, An Introduction to Probability Theory and Its Applications, John Wiley & Sons.

Karlin, S., and Taylor, H., 1975, A First Course in Stochastic Processes, Academic Press.

Oksendal, B., 2003, Stochastic Differential Equations: An Introduction with Applications, Springer.

Rogers, L.C.G., and Williams, D., 2000, Diffusions, Markov Processes and Martingales, Cambridge University Press.