Ch. 12 Stochastic Process

1 Introduction

The analysis of experimental data that have been observed at different points in time leads to new and unique problems in statistical modeling and inference. The obvious correlation introduced by the sampling of adjacent points in time can severely restrict the applicability of the many conventional statistical method traditionally dependent on the assumption that these adjacent observations are independent and identically distributed. The systematic approach by which one goes about answering the mathematical and statistical questions posed by these time **correlation** is commonly referred as **time series analysis**. Time series analysis is the theory of stochastic processes dealing with system which develop in time in accordance with probabilistic laws.

A particularly important aspect of real observable phenomena, which the random variables concept cannot accommodate, is their *time dimension*; the concept of random variable is essential static. A number of economic phenomena for which we need to formulate probability models come in the form of dynamic processes for which we have discrete sequence of observations in time. Observed data referring to economic variables such as inflation, national income, money stock, represent examples where the **time dependency** might be very important. The problem we have to face is extend the simple probability model,

$$\mathbf{\Phi} = \{f(x;\theta), \theta \in \mathbf{\Theta}\},$$

to one which enables us to model *dynamic phenomena*. We have already moved in this direction by proposing the random vector probability model

$$\mathbf{\Phi} = \{f(x_1, x_2, ..., x_T;\theta), \theta \in \mathbf{\Theta}\}.$$

The way we viewed this model so far has been as representing different characteristics of the phenomenon in question in the form of the jointly distributed r.v.'s $X_1, X_2, ..., X_T$. If we reinterpret this model as representing the same characteristic but at successive points in time then this can be viewed as a dynamic probability model. With this as a starting point let us consider the dynamic probability model in the context of $(\mathcal{S}, \mathcal{F}, \mathcal{P})$.

2 The Concept of a Stochastic Process

The natural way to make the concept of a random variable dynamic is to extend its domain by attaching a date to the elements of the sample space $\mathcal{S}$.

Definition:
Let $(\mathcal{S}, \mathcal{F}, \mathcal{P})$ be a probability space, let $\mathcal{T}$ be an index set of real numbers and let $\mathbb{R}^{\mathcal{T}}$ be the product space generated by taking a copy of $\mathbb{R}$ for each element of $\mathcal{T}$. Then, a stochastic process is a measurable mapping $\mathbf{x} : \mathcal{S} \longmapsto \mathbb{R}^{\mathcal{T}}$, where

$$\mathbf{x}(s) = \{X_t(s), t \in \mathcal{T}\}.$$

$\mathcal{T}$ is called the index set, the the random variable $X_t(s)$ is called a coordinate of the process. A stochastic can also be characterized as a mapping from $\mathcal{S} \times \mathcal{T} \to \mathbb{R}$. The ordered sequence of random variables $\{X_t(s),\ t \in \mathcal{T}\}$ is called a stochastic process.[1]

Three main elements of a stochastic process $\{X_t(s),\ t \in \mathcal{T}\}$ are:
(a). its range space (sometimes called the state space),[2] usually $\mathbb{R}$;
(b). the index $\mathcal{T}$, usually one of $\mathbb{R}, \mathbb{R}_+ = [0, \infty)$, $\mathcal{Z} = \{..., 0, 1, 2, ...\}$ and
(c). the dependence structure of the r.v.'s $\{X_t(s), t \in \mathcal{T}\}$ itself (level) or even the second moment of $X_t(s)$ (variation).

In what follows a stochastic process will be denoted by $\{X_t, t \in \mathcal{T}\}$ (s is dropped and and we are concerning exclusively on *discrete stochastic process*. That is, the index set $\mathcal{T}$ is a countable set such as $\mathcal{T} = \{...., -2, -1, 0, 1, 2, ...\}$.

[1] Since a stochastic process determine the joint density of $X_1, X_2, ..., X_T$, the marginal density of $X_t, t = 1, 2, ..., T$ is not necessary to be identical.

[2] In the function $y = f(x)$, x is referred to as the *argument* of the function, and y is called the *value* of the function. We shall also alternatively refer x as the *independent variable* and y as the *dependent variable*. The set of all permissible value that x can take in a given context is known as the *domain* of the function. The value into which an x value is mapped is called the *image* of that x value. The set of all images is called the *range* of the function, which is the set of all values that y variable will take.

2.1 The Joint Distribution of a Stochastic Process in a Finite Time Horizon

The dependence structure of $\{X_t, t \in \mathcal{T}\}$, in direct analogy with the case of a random vector, should be determined by the joint distribution of the process. The question arises, however, since $\mathcal{T}$ is commonly an infinite set, do we need an infinite dimensional distribution to define the structure of the process ?

This question was tackled by Kolmogorov (1933) who showed that when the stochastic process satisfies certain regularity conditions the answer is definitely 'no'. In particular, if we define the 'tentative' joint distribution of the process for the subset $(t_1 < t_2 < ... < t_T)$ of $\mathcal{T}$ by $F(x_{t_1}, x_{t_2}, ..., x_{t_T}) = Pr(X_{t_1} \leq x_1, X_{t_2} \leq x_2, ..., X_{t_T} \leq x_T)$, then if the stochastic process $\{X_t, t \in \mathcal{T}\}$ satisfies the conditions:

(a). **symmetry**: $F(x_{t_1}, x_{t_2}, ..., x_{t_T}) = F(x_{t_{j1}}, x_{t_{j2}}, ..., x_{t_{jT}})$ where $j1, j2, ..., jT$ is any permutation of the indices $1, 2, ..., T$ (i.e. reshuffling the ordering of the index does not change the distribution).
(b). **compatibility**: $\lim_{x_T \to \infty} F(x_{t_1}, x_{t_2}, ..., x_{t_T}) = F(x_{t_1}, x_{t_2}, ..., x_{t_{T-1}})$ (i.e. the dimensionality of the joint distribution can be reduced by marginalization);

there exist a probability space $(\mathcal{S}, \mathcal{F}, \mathcal{P})$ and a stochastic process $\{X_t, t \in \mathcal{T}\}$ defined on it whose finite dimensional distribution is the distribution $F(x_{t_1}, x_{t_2}, ..., x_{t_T})$ as defined above. That is, the probability structure of the stochastic process $\{X_t, t \in \mathcal{T}\}$ is completely specified by the joint distribution of $F(x_{t_1}, x_{t_2}, ..., x_{t_T})$ for all values of T (a positive integer) and any subset $(t_1, t_2, ..., t_T)$ of $\mathcal{T}$.

2.2 The First and Second Moment of a Stochastic Process

Given that, for a specific t, X_t is a random variable, we can denote its distribution and density function by $F(x_t)$ and $f(x_t)$ respectively. Moreover the mean, variance and higher moments of X_t (as a r.v.) can be defined as standard form

as:

$$\begin{aligned} E(X_t) &= \int_{x_t} x_t f(x_t) dx_t = \mu_t, \\ E(X_t - \mu_t)^2 &= \int_{x_t} (x_t - \mu_t)^2 f(x_t) dx_t = \gamma_t^2, \text{ and} \\ E(X_t)^r &= \mu_{rt}, \quad r \geq 1, \end{aligned}$$

for all $t \in \mathcal{T}$.

Definition (Autocovariance Function):
The linear dependence measures between X_i and X_j

$$\gamma_{i,j} = E[(X_i - \mu_i)(X_j - \mu_j)], \quad i, j \in \mathcal{T},$$

is now called the **autocovariance function**. In standardized form

$$r_{i,j} = \frac{\gamma_{i,j}}{\gamma_i \gamma_j}, \quad i, j \in \mathcal{T},$$

is called is **autocorrelation function**.

These numerical characteristics of the stochastic process $\{X_t, t \in \mathcal{T}\}$ play an important role in the analysis of the process and its application to modeling real observable phenomena. We say that $\{X_t, t \in \mathcal{T}\}$ is an uncorrelated process if $r_{i,j} = 0$ for any $i, j \in \mathcal{T}, i \neq j$.

Example:
One of the most important example of a stochastic process is the normal process. The stochastic process $\{X_t, t \in \mathcal{T}\}$ is said to be normal (or Gaussian) if any finite subset of $\mathcal{T}$, say $t_1, t_2, ..., t_T$, $(X_{t_1}, X_{t_2}, ..., X_{t_T}) \equiv \mathbf{x}_T'$ has a multivariate normal distribution, i.e.

$$f(x_{t_1}, x_{t_2}, ..., x_{t_T}) = (2\pi)^{-T/2} |\mathbf{V}_T|^{-1/2} \exp[-\frac{1}{2}(\mathbf{x}_T - \boldsymbol{\mu}_T)' \mathbf{V}_T^{-1} (\mathbf{x}_T - \boldsymbol{\mu}_T)],$$

where

$$\boldsymbol{\mu}_T = E(\mathbf{x}_T) = \begin{bmatrix} E(X_{t_1}) \\ E(X_{t_1}) \\ \cdot \\ \cdot \\ \cdot \\ E(X_{t_T}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_T \end{bmatrix}, \ and$$

$$\mathbf{V}_T = E(\mathbf{x}_T - \boldsymbol{\mu}_T)(\mathbf{x}_T - \boldsymbol{\mu}_T)' = \begin{bmatrix} \gamma_{t_1}^2 & \gamma_{t_1,t_2} & \cdot & \cdot & \cdot & \gamma_{t_1,t_T} \\ \gamma_{t_2,t_1} & \gamma_{t_2}^2 & \cdot & \cdot & \cdot & \gamma_{t_2,t_T} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \gamma_{t_T,t_1} & \cdot & \cdot & \cdot & \cdot & \gamma_{t_T}^2 \end{bmatrix}.$$

As in the case of a normal random variable, the distribution of a normal stochastic process is characterized by the first two moment but now they are function of t.

One problem so far in the definition of a stochastic process given above is much too general to enable us to obtain a operational probability model. In the analysis of stochastic process we only have a single realization of the process and we will have to deduce the value of μ_t and γ_t^2 with the help of a single observation. (which is impossible !)

The main purpose of the next three sections is to consider various special forms of stochastic process where we can construct probability models which are manageable in the context of statistical inference. Such manageability is achieved by imposing certain restrictions which enable us to reduce the number of unknown parameters involved in order to be able to deduce their value from a single realization. These restrictions come in two forms:

(a). restriction on the **time-heterogeneity** of the process; and
(b). restriction on the **memory** of the process.

3 Model's Restictions in a Stochastic Process

3.1 Restricting the Time-Heterogeneity of a Stochastic Process

For an arbitrary stochastic process $\{X_t, t \in \mathcal{T}\}$ the distribution function $F(x_t; \theta_t)$ depends on t with the parameter θ_t characterizing it being function of t as well. That is, a stochastic process is time-heterogeneous in general. This, however, raises very difficult issues in modeling real phenomena because usually we only have one observation for each t. Hence in practice we will have to estimate θ_t on the basis of a single observation, which is impossible. For this reason we are going to consider an important class of **stationary** process which exhibit considerable time-homogeneity and can be used to model phenomena approaching their *equilibrium steady − state*, but continuously undergoing 'random' functions. This is the class of stationary stochastic processes.

𝔇efinition (Strongly Stationary):
A stochastic process $\{X_t, t \in \mathcal{T}\}$ is said to be **(strictly) stationary** if any subset $(t_1, t_2, ..., t_T)$ of $\mathcal{T}$ and any τ,

$$F(x_{t_1}, ..., x_{t_T}) = F(x_{t_1+\tau}, ..., x_{t_T+\tau}).$$

That is, the distribution of the process remains unchanged when shifted in time by an arbitrary value τ. In terms of the marginal distributions, (strictly) stationarity implies that

$$F(X_t) = F(X_{t+\tau}), \;\; t \in \mathcal{T},$$

and hence $F(x_{t_1}) = F(x_{t_2}) = ... = F(x_{t_T})$. That is stationarity implies that $X_{t_1}, X_{t_2}, ..., X_{t_T}$ are (individually) identically distributed.

The concept of stationarity, although very useful in the context of probability theory, is very difficult to verify in practice because it is defined in terms of distribution function. For this reason the concept of the **second** order stationarity, defined in terms of the first two moments, is commonly preferred.

Definition (Weakly Stationary):
A stochastic process $\{X_t, t \in \mathcal{T}\}$ is said to be (**weakly) stationary** if

$$\begin{aligned} E(X_t) &= \mu \ \ for\ all\ t; \\ \gamma_{t_i,t_j} &= E[(X_{t_i} - \mu)(X_{t_j} - \mu)] = \gamma_{|t_j - t_i|}, \ \ t_i, t_j \in \mathcal{T}. \end{aligned}$$

These suggest that weakly stationarity for $\{X_t, t \in \mathcal{T}\}$ implies that its mean and variance $\gamma_{t_i}^2 = \gamma_0$ are constant and free of t and its autocovariance depends on the interval $|t_j - t_i|$; not t_i and t_j. Therefore, $\gamma_k = \gamma_{-k}$.

Example:
Consider the normal stochastic process in the above example. With the weakly stationarity assumption, now

$$\boldsymbol{\mu}_T = E(\mathbf{X}_T) = \begin{bmatrix} \mu \\ \mu \\ . \\ . \\ . \\ \mu \end{bmatrix} \mathbf{V}_T = \begin{bmatrix} \gamma_0 & \gamma_1 & . & . & . & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & . & . & . & \gamma_{T-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \gamma_{T-1} & . & . & . & . & \gamma_0 \end{bmatrix},$$

a sizeable reduction in the number of unknown parameters from $T + [T(T+1)/2]$ to $(T+1)$. It is important, however, to note that even in the case of stationarity the number of parameters increase with the size of the subset $(t_1, ..., t_T)$ although the parameters do not depend on $t \in \mathcal{T}$.[3] This is because time-homogeneity does not restrict the 'memory' of the process. In the next section we are going to consider 'memory' restrictions in an obvious attempt to 'solve' the problem of the parameters increasing with the size of the subset $(t_1, t_2, ..., t_T)$ of $\mathcal{T}$.

[3]To see this, consider $\mathbf{V}_{T+1} = \begin{bmatrix} \gamma_0 & \gamma_1 & . & . & . & \gamma_{T-1} & \gamma_T \\ \gamma_1 & \gamma_0 & . & . & . & \gamma_{T-2} & \gamma_{T-1} \\ . & . & . & . & . & . & \\ . & . & . & . & . & . & \\ . & . & . & . & . & . & \\ \gamma_{T-1} & . & . & . & . & \gamma_0 & \gamma_1 \\ \gamma_T & . & . & . & . & \gamma_1 & \gamma_0 \end{bmatrix}$. With an additional observation $X_{t_{T+1}}$, we have an additional autocovariance γ_T.

3.2 Restricting the Memory of a Stochastic Process

3.2.1 Asymptotic Independence

In the case of a typical economic times series, viewed as a particular realization of a stochastic process $\{X_t,\ t \in \mathcal{T}\}$ one would expect that the dependence between X_{t_i} and X_{t_j} would tend to weaken as the distance $(t_j - t_i)$ increase. Formally, this dependence can be described in terms of the joint distribution $F(x_{t_1}, x_{t_2}, ..., x_{t_T})$ as follows:

Definition (Asymptotically Independent):
A stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be **asymptotically independent** if for any subset $(t_1, t_2, ..., t_T)$ of $\mathcal{T}$ and any τ, $\beta(\tau)$ defined by

$$|F(x_{t_1}, x_{t_2}, ..., x_{t_T}, x_{t_{1+\tau}}, ..., x_{t_{T+\tau}}) - F(x_{t_1}, x_{t_2}, ..., x_{t_T})F(x_{t_{1+\tau}}, ..., x_{t_{T+\tau}})| \leq \beta(\tau) \quad \textit{goes to zero as } \tau \to \infty.$$

That is if $\beta(\tau) \to 0$ as $\tau \to \infty$ the two subsets $(X_{t_1}, X_{t_2}, ..., X_{t_T})$ and $(X_{t_{1+\tau}}, ..., X_{t_{T+\tau}})$ become independent.

A particular case of asymptotic independence is that of $m-$*dependence* which restricts $\beta(\tau)$ to be zero for all $\tau > m$. That is, X_{t_1} and X_{t_2} are independent for $|t_1 - t_2| > m$.

3.2.2 Asymptotic Un-correlation

An alternative way to express the weakening of the dependence between X_{t_i} and X_{t_j} as $|t_j - t_i|$ increases in terms of the autocorrelation function which is a measure of linear dependence.

Definition (Asymptotically Uncorrelated):
A stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be **asymptotically uncorrelated** if for there exists a sequence of constants $\{\rho(\tau), \tau \geq 1\}$ defined by

$$\left|\frac{\gamma_{t,t+\tau}}{\gamma_t \gamma_{t+\tau}}\right| \leq \rho(\tau), \quad \textit{for all } t \in \mathcal{T},$$

such that

$$0 \leq \rho(\tau) \leq 1 \quad and \quad \sum_{\tau=0}^{\infty} \rho(\tau) < \infty.$$

As we can see, the sequence of constants $\{\rho(\tau), \tau \geq 1\}$ defines an upper bound for the sequence of autocorrelation coefficients $r(t, t+\tau)$. Moreover, given that $\rho(\tau) \to 0$ as $\tau \to \infty$ is a necessary and $\rho(\tau) < \tau^{-(1+\delta)}$ for $\delta > 0$, a sufficient condition for $\sum_{\tau=0}^{\infty} \rho(\tau) < \infty$, the intuition underlying the above definition is obvious.

At this stage it is important to note that the above concept of asymptotic independence and uncorrelatedness which restrict the memory of a stochastic process are **not** defined in terms of a stationary stochastic process but a general time-heterogeneous process. This is the reason why $\beta(\tau)$ and $\rho(\tau)$ for $\tau \geq 1$ define only upper bounds for the two measures of dependence given that when equality is used in their definition they will depend on $(t_1, t_2, ..., t_T)$ as well as τ.

3.2.3 Mixing Process

A more general formulation of asymptotic independence can be achieved using the concept of a σ-field generated by a random vector. Let $\mathcal{F}_1^t$ denote the σ-field generated by $X_1, X_2, ..., X_T$ where $\{X_t, t \in \mathcal{T}\}$ is a stochastic process. A measure of the dependence among the elements of the stochastic process can be defined in terms of the events $B \in \mathcal{F}_{-\infty}^t$ and $A \in \mathcal{F}_{t+\tau}^{\infty}$ by

$$\alpha(\tau) = \sup_{\tau} |P(A \cap B) - P(A)P(B)| .$$

Definition (Strongly Mixing Process):
A stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be **strongly mixing** $(\alpha - mixing)$ if $\alpha(\tau) \to 0$ as $\tau \to \infty$.

As we can see, this is a direct generalization of the asymptotic independence which is defined in terms of particular events A and B related to the definition of the joint distribution function. In the case where $\{X_t, t \in \mathcal{T}\}$ is an independent process $\alpha(\tau) = 0$ for $\tau \geq 1$.

Another interesting special case defined above of a mixing process is the $m-dependent$ process where $\alpha(\tau)=0$ for $\tau>m$. In this sense an independent process is a zero-dependent process. The usefulness of the concept of an $m-dependent$ process stems from the fact that commonly in practice any asymptotically independent (or mixing) process can be approximated by such a process for 'large enough' m.[4]

A stronger form of mixing, sometimes called $uniform\ mixing$, can be defined in terms of the following measure of dependence:

$$\varphi(\tau)=\sup_{\tau}|P(A|B)-P(A)|,\quad P(B)>0.$$

Definition (Uniformly Mixing Process):
A stochastic process $\{X_t,\ t\in\mathcal{T}\}$ is said to be **uniformly mixing** $(\varphi-mixing)$ if $\varphi(\tau)\to 0$ as $\tau\to\infty$.

Looking at the two definitions of mixing we can see that $\alpha(\tau)$ and $\varphi(\tau)$ define absolute and relative measures of temporal dependence, respectively. The former is based on the definition of dependence between two events A and B separated by τ periods using the absolute measure

$$[P(A\cap B)-P(A)\cdot P(B)]\geq 0$$

and the latter the relative measure

$$[P(A|B)-P(A)]\geq 0.$$

Because $\varphi(\tau)\geq\alpha(\tau)$ (why ?),[5] $\varphi-mixing$ implies $\alpha-mixing$.

3.2.4 Ergodicity

In the context of weakly-stationary stochastic process, asymptotic uncorrelatedness can be defined more intuitively in terms of the temporal covariance as

[4]For example, the Augmented Dickey-Fuller test in an $ARIMA(p,1,q)$ model. See Chapter 21 for detail.

[5]By definition, $\varphi(\tau)=[P(A|B)-P(A)]=\frac{P(A\cap B)}{P(B)}-P(A)$, i.e. $P(B)\varphi(\tau)=P(A\cap B)-P(A)\cdot P(B)]=\alpha(\tau)$. Since $0<P(B)<1$, we have $\varphi(\tau)\geq\alpha(\tau)$. That is if $\varphi(\tau)=0$, then $\alpha(\tau)$ must be zero.

follows:

$$Cov(X_t, X_{t+\tau}) = \gamma_\tau \to 0 \quad as\ \tau \to \infty.$$

A stronger form of such memory restriction is so called ergodicity property. Ergodicity can be viewed as a condition which ensures that the memory of the process as measured by γ_τ "weakens by averaging overtime".

Definition:
A weakly-stationary stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be **ergodic** if [6]

$$\lim_{T\to\infty}\left(\frac{1}{T}\sum_{\tau=0}^{T}\gamma_\tau\right) = 0.$$

3.3 Some Equvalent Implications

Definition:
A *necessary* condition is in the nature of a prerequisite: suppose that a statement p is true *only if* another statement q is true; then q constitutes a *necessary* condition of p. Symbolically, we express this as follows:

$$p \Longrightarrow q$$

which is read:
1. "p *only if* q," or alternative
2. "*if* p, *then* q". It is also logically correct to mean
3. "p *implies* q", and
4. "p *is a stronger condition than* q " and
5. $p \subset q$.

[6] which imply $\sum_{\tau=0}^{\infty}|\gamma_\tau| < \infty$ and this further implies that $\gamma_\tau \to 0$.

4 Some Special Stochastic Process

We will consider briefly several special stochastic process which play an important role in econometric modeling. These stochastic processes will be divided into **parametric** and **non-parametric** process. The non-parametric process are defined in terms of their joint distribution function or the first few joint moments. On the other hand, parametric process are defined in terms of a generating mechanism which is commonly a functional form based on a non-parametric process.

4.1 Non-Parametric process

4.1.1 White Noise Process

𝔇efinition (White Noise):
A stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be a **white-noise** process if

$$\begin{aligned} (a).\ E(X_t) &= 0; \\ (b).\ E(X_t X_\tau) &= \begin{cases} \sigma^2 & if\ t = \tau \\ 0 & if\ t \neq \tau. \end{cases} \end{aligned}$$

Hence, a white-noise process is both time-homogeneous, in view of the fact that it is a weakly-stationary process, and has no memory. In the case where $\{X_t,\ t \in \mathcal{T}\}$ is also assumed to be normal the process is also strictly stationary.

Despite its simplicity (or because of it) the concept of a white-noise process plays a very important role in the context of parametric time-series models to be considered next, as a basic building block.

4.1.2 Martingales Process

Let $\{X_t, t \in \mathcal{T}\}$ be a stochastic process defined on $(\mathcal{S}, \mathcal{F}, P(\cdot))$ and let $\{\mathcal{F}_t\}$ be a sequence of $\sigma - fields\ \mathcal{F}_t \subset \mathcal{F}$ for all t (i.e. $\{\mathcal{F}_t\}$ is an increasing sequence of $\sigma - fields$) satisfying the following conditions:
(a). X_t is a random variable relatives to $\{\mathcal{F}_t\}$ for all $t \in \mathcal{T}$.
(b). $E(|X_t|) < \infty$ for all $t \in \mathcal{T}$.

(c). $E(X_t|\mathcal{F}_{t-1}) = X_{t-1}$, for all $t \in \mathcal{T}$.

Then $\{X_t, t \in \mathcal{T}\}$ is said to be a **martingale** with respect to $\{\mathcal{F}_t, t \in \mathcal{T}\}$.

4.1.3 Markov Process

An important class of stochastic process is that of Markov process. These process are based on so- called Markov property that 'the future' of the process, given the 'present', is independent of the 'past'.

𝔇efinition (Markov Process):
A stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be a **Markov** process if for every Borel function $h(X_t) \in \mathcal{B}_t^{\infty}$ ('the future') such that

$$\begin{aligned} E|h(X_t)| &< \infty, \\ E(h(X_t)|\mathcal{B}^t_{-\infty}) &= E(h(X_t|\mathcal{F}_{t-1}), \end{aligned}$$

where $\mathcal{B}^b_\alpha = \{\mathcal{F}_t, \alpha < t < b\}$.

4.1.4 Brownian Motion

A particular form of a Markov process with a long history in physics is the so-called Brownian motion (or Wiener) process.

𝔇efinition (Brownian Motion):
Let $(\mathcal{S}, \mathcal{F}, \mathcal{P})$ be a complete probability space. Then $W : \mathcal{S} \times [0,1] \to \mathbb{R}^1$ is a standard Wiener process if each of $r \in [0,1]$, $W(\cdot, r)$ is $\mathcal{F}$-measurable, and in addition,
(a). The process starts at zero: $\mathcal{P}[W(\cdot, 0) = 0] = 1$.
(b). The increments are independent: if $0 \le a_0 \le a_1 ... \le a_k \le 1$, then $W(\cdot, a_i) - W(\cdot, a_{i-1})$ is independent of $W(\cdot, a_j) - W(\cdot, a_{j-1})$, $j = 1, .., k$, $j \ne i$ for all $i = 1, ..., k$.
(c). The increments are normally distributed: For $0 \le a \le b \le 1$, the increment $W(\cdot, b) - W(\cdot, a)$ is distributed as $N(0, b-a)$.

4.2 Parametric Stochastic Processes

The main difference between the type of stochastic process considered so far and the one to be considered in this section is that the latter are defined in terms of a generating mechanism; they are 'derived' stochastic processes.

4.2.1 (Weakly) Stationary Process

Definition ($AR(1)$ Process) 13:
A stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be a **autoregressive of order one** (AR(1)) if it satisfies the stochastic difference equation,

$$X_t = \phi X_{t-1} + u_t$$

where ϕ is a constant and u_t is a white-noise process.

Definition ($AR(p)$ Process):
A stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be a **autoregressive of order** p ($AR(p)$) if it satisfies the stochastic difference equation,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + u_t,$$

where $\phi_1, \phi_2, ..., \phi_p$ are constants and u_t is a white-noise process.

Definition ($MA(q)$ Process):
A stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be a **moving average process of order** q ($MA(q)$) if it can be expressed in the form

$$X_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ... + \theta_q u_{t-q},$$

where $\theta_1, \theta_2, ..., \theta_q$ are constants and u_t is a white-noise process.

That is, the white-noise process is used to build the process $\{X_t,\ t \in \mathcal{T}\}$, being a linear combination of the last q u_{t-i}'s.

Definition ($ARMA(p,q)$ Process):
A stochastic process $\{X_t,\ t \in \mathcal{T}\}$ is said to be an **autoregressive moving average process of order** p, q ($ARMA(p,q)$) if it can be expressed in the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ... + \theta_q u_{t-q},$$

where $\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q$ are constants and u_t is a white-noise process.

Definition ($ARFIMA(p,d,q)$ Process):
A stochastic process $\{Y_t,\ t \in \mathcal{T}\}$ is said to be an **fractionally autoregressive integrated moving average process of order** p, d, q ($ARFIMA(p,d,q)$) if it can be expressed as a stationary $ARMA(p,q)$ process after fractionally-differenced "d" times:

$$(1-L)^d Y_t = X_t,$$

and

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ... + \theta_q u_{t-q},$$

where $\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q$ are constants, $|d| < 0.5$ and u_t is a white-noise process.

4.2.2 Non-Stationary Process

Definition (Unit Root Process):
A stochastic process $\{Y_t,\ t \in \mathcal{T}\}$ is said to be an **autoregressive integrated moving average process of order** p, q ($ARIMA(p,1,q)$) if it can be expressed as a stationary $ARMA(p,q)$ process after first-differenced

$$(1-L)^1 Y_t = X_t,$$

and

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ... + \theta_q u_{t-q},$$

where $\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q$ are constants and u_t is a white-noise process.