

# TEMA 5. PROCESOS ESTOCÁSTICOS

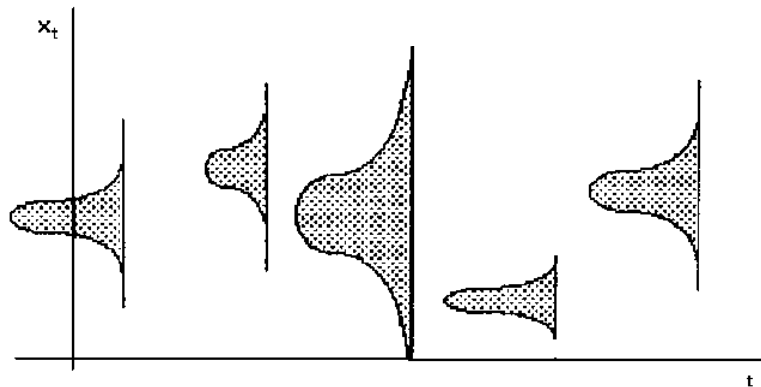
En el estudio de las variables aleatorias realizado hasta ahora se han explorado las características aleatorias del fenómeno pero se ha mantenido una premisa por defecto, que esas características aleatorias permanecen constantes a través del tiempo. Al incluir en el estudio la presencia de la variable determinística **tiempo** se está considerando que, de alguna forma, la variable aleatoria depende del tiempo. En otras palabras, la variable aleatoria dependerá del fenómeno probabilístico y del tiempo. En consecuencia, cualquier función que se establezca en términos de la variable aleatoria, como lo son la función de distribución o la función de densidad, serán también dependientes del tiempo.

Uno de los objetivos de este capítulo es construir un modelo que nos permita explicar la estructura y prever la evolución, al menos a corto plazo, de una variable que observamos a lo largo del tiempo. La variable observada puede ser económica (I.P.C., demanda de un producto, existencias en un determinado almacén, etc...), física (temperatura de un proceso, velocidad del viento en una central eólica, concentración en la atmósfera de un contaminante, etc..) o social (número de nacimientos, votos de un determinado partido, etc..). Supondremos a lo largo del tema que los datos se obtienen en intervalos regulares de tiempo (horas, días, años,..) y el objetivo es utilizar la posible "inercia" en el comportamiento de la serie con el fin prever su evolución futura. Así, una serie temporal será una sucesión de valores de una variable obtenidos de manera secuencial durante el tiempo.

## 5.1. Concepto de proceso estocástico.

**Definición** Un proceso estocástico es una colección o familia de variables aleatorias  $\{X_t, t \in T\}$ , ordenadas según el subíndice  $t$  que en general se suele identificar con el tiempo.

Por tanto, para cada instante  $t$  tendremos una variable aleatoria distinta representada por  $X_t$ , con lo que un proceso estocástico puede interpretarse como una sucesión de variables aleatorias cuyas características pueden variar a lo largo del tiempo. Por ejemplo, si observamos sólo unos pocos valores de  $t$ , tendríamos una imagen similar a la de la figura siguiente:



en la que se representa para cada  $t$  la función de densidad correspondiente a  $X_t$ . Aunque en la figura se han representado unas funciones de densidad variables, un proceso estocástico no tiene por que presentar esas diferencias en la función de densidad a lo largo del tiempo. Como más adelante se comentará presentan un especial interés aquellos procesos cuyo comportamiento se mantiene constante a lo largo de  $t$ .

A los posibles valores que puede tomar la variable aleatoria se le denominaran **estados**, por lo que se puede tener un espacio de estados discreto y un espacio de estados continuo. Por otro lado, la variable tiempo puede ser de tipo discreto o de tipo continuo. En el caso del tiempo discreto se podría tomar como ejemplo que los cambios de estado ocurran cada día, cada mes, cada año, etc.. En el caso del tiempo continuo, los cambios de estado se podrían realizar en cualquier instante.

Por tanto, dependiendo de cómo sea el conjunto de subíndices  $T$  y el tipo de variable aleatoria dado por  $X_t$  se puede establecer la siguiente clasificación de los procesos estocásticos:

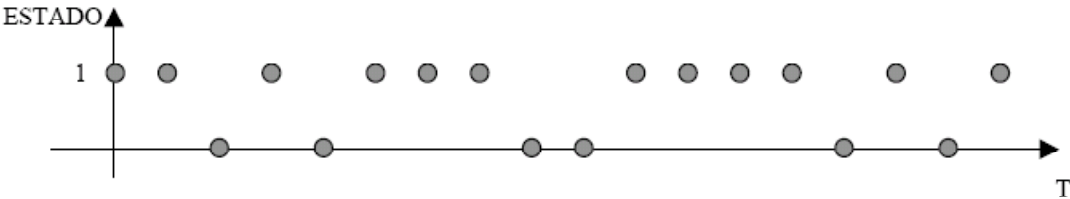
- Si el conjunto  $T$  es continuo, por ejemplo  $\mathbb{R}^+$ , diremos que  $X_t$  es un proceso estocástico de parámetro continuo.
- Si por el contrario  $T$  es discreto, por ejemplo  $\mathbb{N}$ , diremos que nos encontramos frente a un proceso estocástico de parámetro discreto.
- Si para cada instante  $t$  la variable aleatoria  $X_t$  es de tipo continuo, diremos que el proceso estocástico es de estado continuo.
- Si para cada instante  $t$  la variable aleatoria  $X_t$  es de tipo discreto, diremos que el proceso estocástico es de estado discreto.

	$t$ <b>Discreto</b>	$t$ <b>Continuo</b>
$X$ <b>Discreta</b>	<i>Proceso de estado discreto y tiempo discreto (<b>Cadena</b>)</i> (Unidades producidas mensualmente de un producto)	<i>Proceso de estado discreto y tiempo continuo (<b>Proc. Saltos Puros</b>)</i> (Unidades producidas hasta el instante $t$ )
$X$ <b>Continua</b>	<i>Proceso de estado continuo y tiempo discreto</i> (Toneladas de producción diaria de un producto)	<i>Proceso de estado continuo y tiempo continuo (<b>Proceso Continuo</b>)</i> (Velocidad de un vehículo en el instante $t$ )

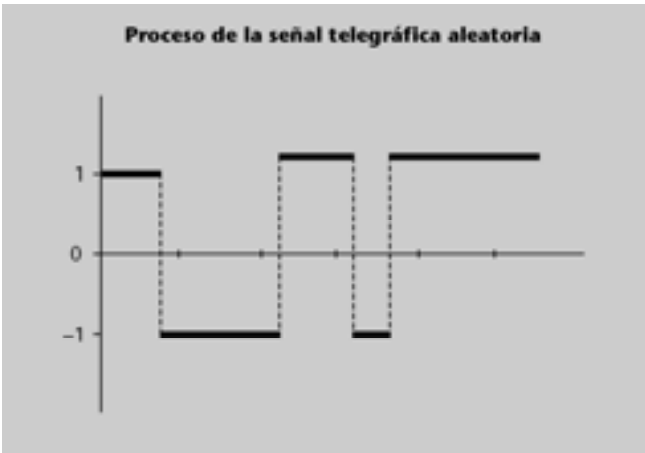
Una **Cadena** es un proceso estocástico en el cual el tiempo se mueve en forma discreta y la variable aleatoria sólo toma valores discretos en el espacio de estados. Un **Proceso de Saltos Puros** es un proceso estocástico en el cual los cambios de estados ocurren en forma aislada y aleatoria pero la variable aleatoria sólo toma valores discretos en el espacio de estados. En un **Proceso Continuo** los cambios de estado se producen en cualquier instante y hacia cualquier estado dentro de un espacio continuo de estados.

Como un ejemplo de una **Cadena**, considere una máquina dentro de una fábrica. Los posibles estados para la máquina son que esté operando o que esté fuera de funcionamiento y la verificación de esta característica se realizará al principio de cada día de trabajo. Si hacemos corresponder el estado 'fuera de funcionamiento' con el valor 0 y el estado 'en

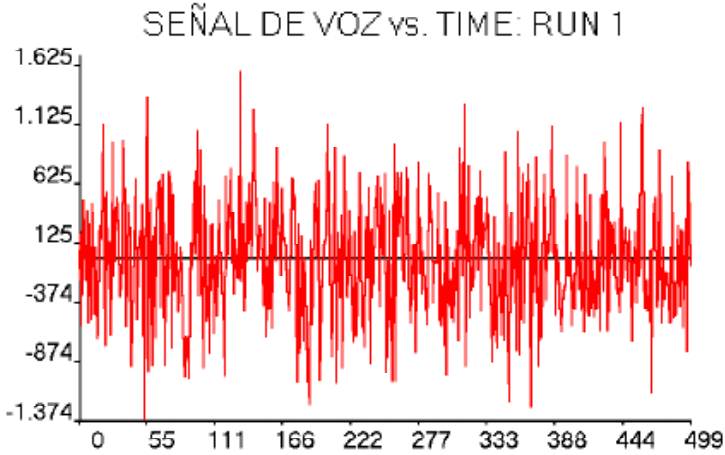
operación' con el valor 1, la siguiente figura muestra una posible secuencia de cambios de estado a través del tiempo para esa máquina.



Para el caso de los **Procesos de Saltos Puros** se puede considerar como un ejemplo una **señal telegráfica**. Sólo hay dos posibles estados (por ejemplo 1 y -1) pero la oportunidad del cambio de estado se da en cualquier instante en el tiempo, es decir, el instante del cambio de estado es aleatorio. La siguiente figura muestra una señal telegráfica.



Como un ejemplo de un **Proceso Continuo**, se puede mencionar la señal de voz vista en la pantalla de un osciloscopio. Esta señal acústica es transformada en una señal eléctrica analógica que puede tomar cualquier valor en un intervalo continuo de estados. La figura siguiente muestra una señal de voz la cual está modulada en amplitud.



## 5.2. Procesos de Estado Discreto.

En el caso de procesos estocásticos con espacio de estados discreto, una secuencia de variables que indique el valor del proceso en instantes sucesivos suele representarse de la siguiente manera:

$$\{X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n\}$$

en la que cada variable  $X_i$ ,  $i = 0, \dots, n$ , tiene una distribución de probabilidades que, en general, es distinta de las otras variables aunque podrían tener características comunes.

El principal interés del estudio a realizar en el caso discreto es el cálculo de probabilidades de ocupación de cada estado a partir de las probabilidades de cambio de estado. Si en el instante  $n - 1$  se está en el estado  $x_{n-1}$ , ¿con qué probabilidad se estará en el estado  $x_n$  en el instante siguiente  $n$ ? Esta probabilidad se denotará como:

$$P(X_n = x_n / X_{n-1} = x_{n-1})$$

A este tipo de probabilidad condicionada se le denomina **probabilidad de transición o de cambio de estado**. A las probabilidades del tipo  $P(X_n = x_n)$  se les denomina **probabilidades de ocupación de estado**.

Otro tipo de probabilidad de interés es la de ocupar un cierto estado en un instante  $n$ , dado que en todos los instantes anteriores, desde  $n = 0$  hasta  $n - 1$ , se conoce en qué estados estuvo el proceso. Esto se puede escribir como:

$$P(X_n = x_n / X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1})$$

Nótese que esta probabilidad depende de toda la historia pasada del proceso, mientras que la probabilidad de transición depende únicamente del estado actual que ocupe el proceso.

**Propiedad de Markov:** Se dice que un proceso cumple la propiedad de Markov cuando toda la historia pasada del proceso se puede resumir en la posición actual que ocupa el proceso para poder calcular la probabilidad de cambiar a otro estado, es decir, se cumple la propiedad siguiente:

$$P(X_n = x_n / X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = P(X_n = x_n / X_{n-1} = x_{n-1})$$

Aquellas Cadenas que cumplen la propiedad de Markov se llaman **Cadenas de Markov**. Otra manera de denotar a las probabilidades de transición es de la forma siguiente:

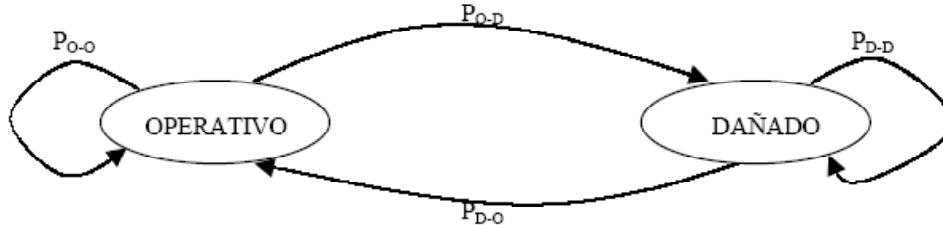
$$P(X_n = j / X_{n-1} = i) = p_{ij}(n)$$

Una propiedad interesante que puede tener una Cadena es que los valores  $p_{ij}(n)$  no dependan del valor de  $n$ . Es decir, las probabilidades de cambiar de estado son las mismas en cualquier instante. Esta propiedad indica que las probabilidades de transición son estacionarias.

### Cadenas de Markov

Considere un equipo de computación el cual será revisado al inicio de cada día para verificar si está operativo o si está dañado. El diagrama de estados de la siguiente figura muestra los dos posibles estados del proceso y las relaciones que se deben cumplir para pasar de un

estado al otro. Así, con probabilidad PO-D se pasa del estado operativo al estado dañado; ésta es una probabilidad de transición o de cambio de estado. En el caso de que el valor de la probabilidad anterior no cambie a través del tiempo y, esto se repite para todas las probabilidades involucradas en la figura, se dice que la Cadena es estacionaria.



Si en un día cualquiera el ordenador está dañado al comenzar el día, sólo puede ocurrir una de dos cosas para el inicio del siguiente día: que siga dañado (con probabilidad PD-D) o que haya sido reparado en ese día (con probabilidad PD-O). Por otro lado, si en un día cualquiera el ordenador está funcionando correctamente al comenzar el día, sólo puede ocurrir una de dos cosas para el inicio del siguiente día: que se dañe (con probabilidad PO-D) o que se mantenga operativo (con probabilidad PO-O). Estos son los únicos eventos que pueden ocurrir.

Si se define una variable aleatoria  $X_n$  como el estado del ordenador al inicio del día  $n$ , se podrían asignar los valores 0 y 1 a los posibles estados 'dañado' u 'operativo', respectivamente.

Otra pregunta de interés es, ¿con qué probabilidad el estado del computador en el día  $n$  es 'operativo'?; ésta es una probabilidad de ocupación de estado. Esa probabilidad se denota como  $P(X_n = 1)$  o también  $\pi_n(1)$ . De igual forma, la probabilidad, por ejemplo, de que el computador esté dañado en la primera oportunidad de ser observado será  $\pi_0(0)$ .

Utilizando la nomenclatura descrita anteriormente, se puede escribir:

$$\begin{aligned} P(X_{n+1} = 1/X_n = 0) &= p \Rightarrow P(X_{n+1} = 0/X_n = 0) = 1 - p \\ P(X_{n+1} = 0/X_n = 1) &= q \Rightarrow P(X_{n+1} = 1/X_n = 1) = 1 - q \end{aligned}$$

Estos datos se podrían resumir en una matriz que se llama matriz de probabilidades de transición o matriz de transición de estados.

$$P_n = \begin{pmatrix} P(X_{n+1} = 0/X_n = 0) & P(X_{n+1} = 1/X_n = 0) \\ P(X_{n+1} = 0/X_n = 1) & P(X_{n+1} = 1/X_n = 1) \end{pmatrix} = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix}$$

En general, consideremos una cadena de Markov con  $k$  estados posibles  $s_1, s_2, \dots, s_k$  y probabilidades estacionarias. Para  $i = 1, 2, 3, \dots, k$  y  $j = 1, 2, 3, \dots, k$ , denotaremos por  $p_{ij}$  a la probabilidad condicionada de que el proceso esté en el estado  $s_j$  en un determinado momento si estuvo en el estado  $s_i$  en el momento inmediatamente anterior. Entonces **la matriz de transición de la cadena de Markov** se define como una matriz cuadrada  $k \times k$  de la siguiente forma:

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}$$

El elemento en la fila  $i$ , columna  $j$ ,  $p_{ij} = P(X_n = s_j / X_{n-1} = s_i)$ , representa la probabilidad de transición de un paso.

El usar esta representación en forma matricial nos facilita el cómputo de las probabilidades de transición en más de un paso. En general  $\mathbf{P} \times \mathbf{P} = \mathbf{P}^2$  corresponde a las probabilidades de transición en dos pasos, y  $\mathbf{P}^3, \mathbf{P}^4, \dots, \mathbf{P}^m, \dots$  corresponden a las probabilidades de transición en 3, 4, ...,  $m$  pasos respectivamente. De hecho, la matriz  $\mathbf{P}^m$  se conoce como la **matriz de transición en  $m$  pasos** de la cadena de Markov.

#### Distribución Inicial de la Cadena.

En ocasiones no se conoce a ciencia cierta la ubicación del proceso en el instante inicial. Esta ocupación podría presentarse en forma probabilística como un vector de ocupación inicial  $\boldsymbol{\pi}_0$ . Este vector tiene como componentes la probabilidad de ocupar cada uno de los estados en el instante inicial, es decir, los términos  $\pi_0(j)$ , para cada estado  $j = 1, 2, \dots, k$ .

$$\boldsymbol{\pi}_0 = (\pi_0(1), \pi_0(2), \pi_0(3), \dots, \pi_0(k))$$

con:

$$\pi_0(j) \geq 0, \quad \sum_{j=1}^k \pi_0(j) = 1$$

#### Distribución de las probabilidades de ocupación de estados de la Cadena en un instante cualquiera.

Al igual que para el instante inicial, se puede definir un vector de ocupación de estado para cada instante  $n$ . Se le llamará  $\boldsymbol{\pi}_n$ . Este vector tiene como componentes la probabilidad de ocupar cada uno de los estados en el instante  $n$ :

$$\boldsymbol{\pi}_n = (\pi_n(1), \pi_n(2), \pi_n(3), \dots, \pi_n(k))$$

con:

$$\pi_n(j) \geq 0, \quad \sum_{j=1}^k \pi_n(j) = 1$$

En general se tiene la siguiente relación:

$$\pi_n(j) = P(X_n = j) = \sum_{i=1}^k P(X_0 = i) \cdot P(X_n = j / X_0 = i) = \sum_{i=1}^k \pi_0(i) \cdot P(X_n = j / X_0 = i)$$

que matricialmente se escribe:

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 \cdot \mathbf{P}^n$$

**Ejercicio** Después de mucho estudio sobre el clima, hemos visto que si un día está soleado, en el 70% de los casos el día siguiente continua soleado y en el 30% se pone nublado. También nos fijamos en que si un día está nublado, la probabilidad de que esté soleado el día siguiente es 0.6 y la probabilidad de que se ponga nublado es 0.4. Además, si un día está soleado la probabilidad de que continúe soleado el día siguiente es 0.7 y la probabilidad de que al día siguiente esté nublado es 0.3. Entonces:

- Construir la matriz de transición de la cadena.
- Obtener la matriz de transición tras dos días.
- Si hoy está nublado, ¿cuál es la probabilidad de que esté nublado tres días después?.

## Procesos de Saltos Puros

En este caso, el proceso sigue siendo discreto en estados pero la gran diferencia es que los cambios de estado ocurren en cualquier instante en el tiempo (tiempo continuo). Hemos apuntado anteriormente que un ejemplo típico de procesos de saltos puros es la señal telegráfica.

Otros ejemplos de procesos de saltos puros son los siguientes:

**Definición** Un proceso estocástico en tiempo continuo  $\{N(t), t \geq 0\}$  se dice que es un **Proceso de Conteo** si representa el número de veces que ocurre un suceso hasta el instante de tiempo  $t$ .

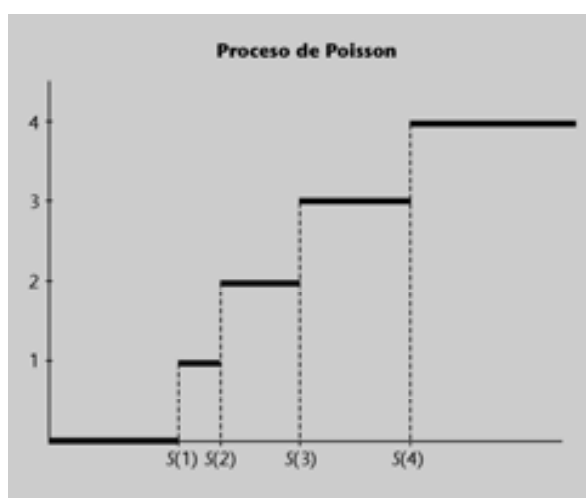
En particular, se tiene que  $N(t) \in \mathbb{N}$ , y  $N(s) \leq N(t) \forall s < t$ .

**Definición** Un proceso de conteo se dice que es un **Proceso de Poisson** homogéneo de tasa  $\lambda > 0$ , si:

1.  $N(0) = 0$
2.  $N(t_1) - N(t_0), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1})$  son variables aleatorias independientes (proceso de incrementos independientes).
3.  $N(t + s) - N(s)$  ("Nº de sucesos que ocurren entre el instante  $s$  y el instante  $t + s$ ") sigue una distribución de Poisson de parámetro  $\lambda t$ .

El proceso de Poisson se utiliza básicamente para modelar los llamados procesos de colas. En ellos se pueden incluir muchos procesos: coches que llegan al peaje de una autopista, clientes que llegan a un banco, peticiones que llegan a un servidor de Internet, llamadas que pasan por una centralita, etc.

Consideremos, por ejemplo, el caso de un servidor de Internet. Queremos estudiar las peticiones que va recibiendo este servidor. Podemos suponer que las peticiones llegan de una en una. Si llamamos  $N(t)$  al número de peticiones que ha recibido el servidor hasta el instante  $t$ , encontramos que una posible realización de este proceso será del tipo:



donde  $S(1)$  indica el instante en que llega la primera petición,  $S(2)$  indica el instante en que llega la segunda y, en general,  $S(i)$  indica el instante en que llega la  $i$ -ésima. Éste es un ejemplo de un proceso que se puede representar utilizando el proceso de Poisson.

Tiene algunas características fundamentales:

- Para cada instante  $t$ ,  $N(t)$  seguirá una distribución de Poisson de parámetro  $\lambda t$ .
- Las diferencias entre los tiempos de llegadas consecutivas siguen una distribución exponencial de parámetro  $\lambda$ , es decir,  $S(i+1) - S(i)$  siguen una exponencial de parámetro  $\lambda$ .

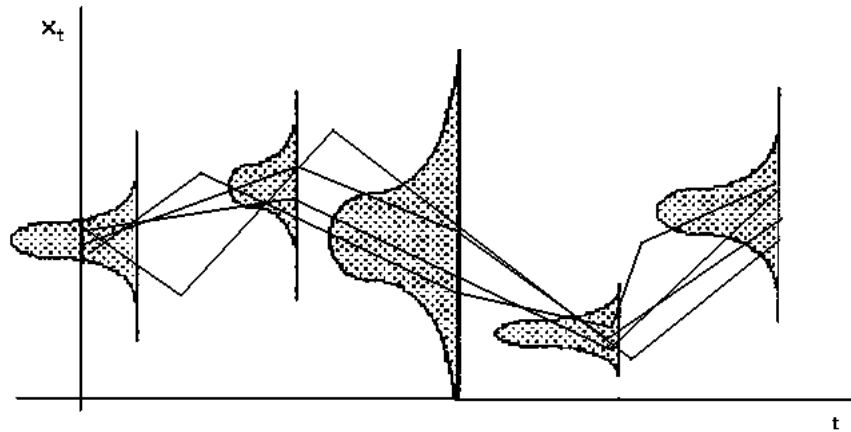
### 5.3. Procesos de Estado Continuo.

#### Realización de un Proceso Estocástico: Series temporales.

Un concepto que debe ser definido es el de realización: una realización de una experiencia aleatoria es el resultado de una repetición de esa experiencia. Así, en la experiencia aleatoria "lanzar una vez un dado" una realización sería el cinco que nos ha salido en ese único lanzamiento. En ese caso, la realización se reduce a un único número  $\{X\}$ . Si repetimos la experiencia, obtendremos otra realización distinta. En una experiencia bidimensional, por ejemplo la identificación de unos individuos a través de su peso y su estatura, una realización sería el resultado obtenido al tomar un individuo al azar y medir en él ambos parámetros. Ahora la realización es una pareja de números  $\{X_1, X_2\}$ . En una experiencia  $n$ -dimensional tendríamos análogamente un grupo de  $n$  valores  $\{X_1, \dots, X_n\}$  y, siguiendo con la analogía, en un proceso estocástico tendríamos un grupo de infinitos valores  $\{X_t\}$ , con  $t \in \Theta$ .

Veamos un ejemplo de realización de un proceso estocástico. Consideremos que estamos interesados en estudiar cómo evoluciona la temperatura en un punto  $P$  de un motor durante una hora de funcionamiento del mismo a régimen constante. Podemos considerar que la variable "temperatura instantánea en ese punto  $P$ " es un proceso estocástico continuo de parámetro continuo. Si realizamos la experiencia una vez obtendremos una gráfica temporal de la evolución de la temperatura: esa gráfica (o los valores numéricos asociados) es una realización del proceso estocástico. Si repetimos la experiencia en las mismas condiciones obtendremos otra realización, representada por otra gráfica o por sus correspondientes resultados numéricos.





Por tanto, una realización de un proceso estocástico es una sucesión de infinitos valores de una cierta variable a lo largo del tiempo. Si  $t$  tiene una consideración continua obtendremos una representación continua, mientras que si  $t$  es discreto obtendremos una sucesión de puntos.

Si recordamos la definición de serie temporal que dimos en la introducción, "una serie temporal será una sucesión de valores de una variable obtenidos de manera secuencial durante el tiempo", la única diferencia que hay entre ellas radica en que la realización consta de infinitos elementos y la serie temporal de un número limitado. Así, desde el punto de vista de los procesos estocásticos, diremos que

**Definición:** Una *serie temporal* es una realización parcial de un proceso estocástico de parámetro tiempo discreto.

Por tanto, la teoría de los procesos estocásticos será de aplicación a las series temporales. No obstante, nos encontraremos con una fuerte restricción que radica en el hecho de que en muchas series temporales, ellas son la única realización observable del proceso estocástico que las ha generado (por ejemplo, la serie de turistas que visitan España mes a mes, o la unidades producidas diariamente en una factoría, etc.). Esto, en general nos colocará, cuando queramos deducir las características del proceso a partir de las de la serie, en una situación hasta cierto punto similar a la que tendríamos al intentar describir la composición de una urna en base a una única bola extraída. Tal problema, que parece insalvable, requerirá que apliquemos una serie de restricciones e hipótesis al tipo de proceso que queramos analizar.

## Características de un proceso estocástico.

Del mismo modo que en una variable unidimensional  $X$ , podemos calcular su media, su varianza y otras características, y en variables  $n$ -dimensionales obtenemos un vector de medias, matriz de covarianzas, etc., en un proceso estocástico podemos obtener algunas características que describen su comportamiento: medias, varianzas y covarianzas. Puesto que las características del proceso pueden variar a lo largo de  $t$  estas características no serán parámetros sino que serán funciones de  $t$ . Así:

**Definición:** Llamaremos *función de medias* del proceso a una función de  $t$  que proporciona las medias de las distribuciones marginales para cada instante  $t$

$$\mu_t = E(X_t)$$

**Definición:** Llamaremos *función de varianzas* del proceso a una función de  $t$  que proporciona las varianzas de las distribuciones marginales para cada instante  $t$

$$\sigma_t^2 = Var(X_t)$$

**Definición:** Llamaremos *función de autocovarianzas* del proceso a la función que proporciona la covarianza existente entre dos instante de tiempo cualesquiera:

$$Cov(t, s) = Cov(s, t) = Cov(X_t, X_s)$$

y por último,

**Definición:** Llamaremos *función de autocorrelación* a la estandarización de la función de covarianzas:

$$\rho_{t,s} = \frac{Cov(t, s)}{\sigma_t \sigma_s}$$

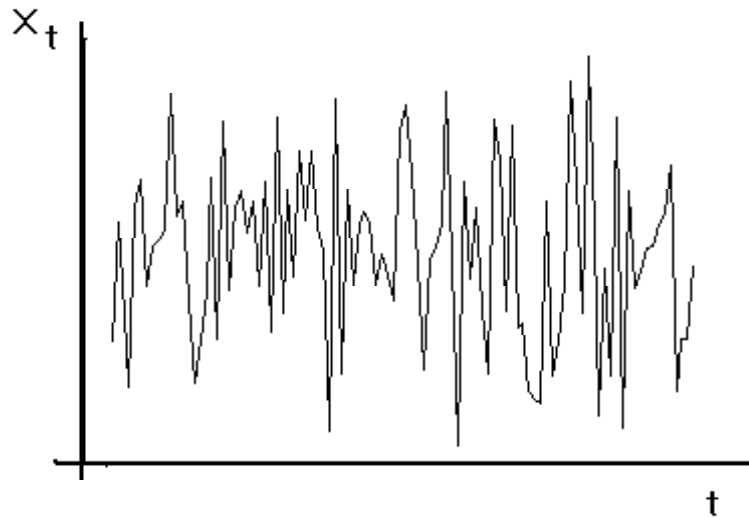
En general, estas dos últimas funciones dependen de dos parámetros (dos instantes). Una condición de estabilidad que aparece en muchos fenómenos es que la dependencia sólo dependa, valga la redundancia, de la "distancia" entre ellos y no del instante considerado. En estos casos tendremos:

$$Cov(t, t + j) = Cov(s, s + j) = \gamma_j \quad j = 0, \pm 1, \pm 2, \dots$$

Por otro lado, si estudiamos casos concretos como la evolución de las ventas de una empresa o la concentración de un contaminante, sólo disponemos de una única realización y aunque el proceso estocástico exista, al menos conceptualmente, para poder estimar las características "transversales" del proceso (medias, varianzas, etc..) a partir de la serie es necesario suponer que estas permanecen "estables" a lo largo de  $t$ . Esta idea nos conduce a lo que se entiende por condiciones de estacionariedad de un proceso estocástico (o de una serie temporal).

## Procesos estocásticos estacionarios.

En una primera aproximación, llamaremos estacionarios a aquellos procesos estocásticos que tengan un comportamiento constante a lo largo del tiempo. Si buscamos el correspondiente comportamiento de las series temporales asociadas a esos procesos, veremos gráficas que se mantienen en un nivel constante con unas pautas estables de oscilación. La siguiente figura muestra un ejemplo de serie estacionaria, realización de un proceso estocástico estacionario.



En la práctica del análisis de series encontraremos series con problemas de estacionariedad que afecten a cualquiera de sus parámetros básicos, siendo los que más suelen afectar al proceso de análisis las inconstancias en media y varianza. En las siguientes figuras se muestran dos ejemplos de series no estacionarias, la primera en media y la segunda en varianza.

**Definición:** Diremos que un proceso es estacionario en sentido estricto si al realizar un mismo desplazamiento en el tiempo de todas las variables de cualquier distribución conjunta finita, resulta que esta distribución no varía, es decir:

$$F(X_{i_1}, X_{i_2}, \dots, X_{i_r}) = F(X_{i_1+j}, X_{i_2+j}, \dots, X_{i_r+j})$$

para todo conjunto de índices  $(i_1, i_2, \dots, i_r)$  y todo  $j$ .

Esta condición resulta bastante restrictiva y por consiguiente se adoptan otras un poco más "débiles"

**Definición:** Diremos que un proceso estocástico es *estacionario en sentido débil* si mantiene constantes todas sus características lo largo del tiempo, es decir, si para todo  $t$  :

1.  $\mu_t = \mu$
2.  $\sigma_t^2 = \sigma^2$
3.  $Cov(t, t + j) = Cov(s, s + j) = \gamma_j \quad j = 0, \pm 1, \pm 2, \dots$

**Nota.-**

*En algunos libros este tipo de estacionariedad recibe el nombre de estacionariedad en sentido amplio o estacionariedad de segundo orden. Por otro lado si sólo exigimos que la función de medias sea constante se dirá que el proceso es estacionario de primer orden o en media. Por último, indicar que la estacionariedad en sentido débil no garantiza la estacionariedad, ahora bien bajo la suposición de normalidad de las variables si se verifica esta igualdad.*

Es inmediato probar que si un proceso es estacionario en sentido débil la función de autocorrelación viene dada por:

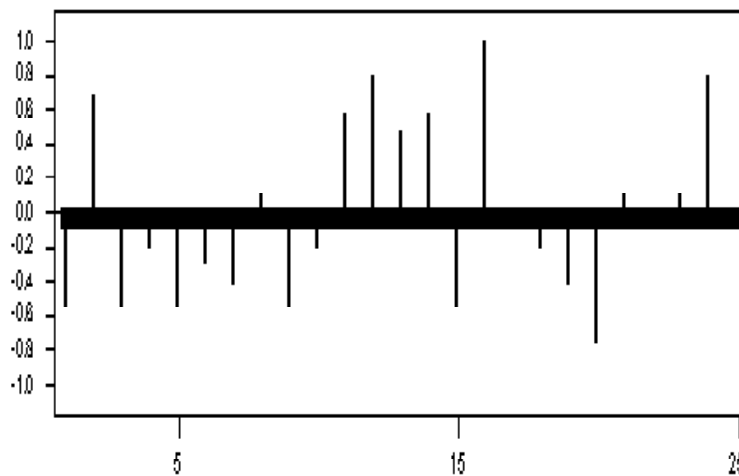
$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{\gamma_j}{\sigma^2}$$

Obviamente esta función es simétrica ( $\rho_j = \rho_{-j}$ ) y sólo dependerá de la distancia o retardo entre los instantes de tiempo o también llamado "decalaje" . Así:

**Definición:** Se denomina *función de autocorrelación simple (acf)* a dicha función:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{\gamma_j}{\sigma^2}$$

y la representación gráfica en forma de diagrama de barras de sus valores recibe el nombre de *correlograma*.



Observar que el correlograma, tal y como se ha definido, sólo tiene sentido para series estacionarias.

## Proceso de ruido blanco.

Un proceso estocástico que emplearemos con frecuencia en los apartados siguientes es el llamado proceso de ruido blanco. Se trata de un proceso estacionario  $\varepsilon_t$  que cumple las siguientes características:

- $E(\varepsilon_t) = 0$
- $Var(\varepsilon_t) = \sigma^2$
- $Cov(\varepsilon_t, \varepsilon_s) = 0$  si  $t \neq s$

Podemos interpretar un ruido blanco como una sucesión de valores sin relación alguna entre ellos, oscilando en torno al cero dentro de un margen constante. En este tipo de procesos, conocer valores pasados no proporciona ninguna información sobre el futuro ya que el proceso es "puramente aleatorio" y por consiguiente "carece de memoria"

Una de las características principales de un proceso estocástico es su correlograma y en el caso del proceso de ruido blanco no puede ser más sencillo: puesto que las variables correspondientes a distintos instantes están incorreladas, todos los coeficientes de correlación serán 0 salvo el correspondiente a un retardo 0 que será 1.

## Modelos Auto-Regresivos de orden $p$ , $AR(p)$ .

Al tratar de representar la influencia que hechos pasados tienen sobre el presente (y en consecuencia sobre el futuro) de un proceso estocástico, podemos considerar diferentes expresiones alternativas. Una de ellas consiste en colocar el valor actual del proceso como dependiente de modo lineal de valores pasados del propio proceso, más una perturbación aleatoria, **que supondremos normalmente distribuida**, que evite que el modelo sea determinista:

$$X_t = \delta + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t \quad (1)$$

donde  $\delta$  representa una constante.

A esta formulación la llamaremos autorregresiva, pues en la misma expresión se aprecia el porque de la denominación: de algún modo es un modelo de regresión del proceso sobre si mismo.

Si observamos el modelo propuesto en (1), si lo consideramos estacionario, con  $E(X_t) = \mu$ , se tiene que:

$$\mu = \delta(1 + a_1\mu + a_2\mu + \dots + a_p\mu) \Rightarrow \mu = \frac{\delta}{1 - a_1 - a_2 - \dots - a_p}$$

por consiguiente, para que exista la media, necesitamos que:

$$a_1 + a_2 + \dots + a_p \neq 1$$

Sin perder generalidad, en el desarrollo del apartado supondremos que el proceso está centrado, esto es,  $\delta = 0$ .

Un elemento que resulta de gran interés en estudio de estos modelos procesos lineales es el llamado operador retardo  $B$ . Tal operador actúa sobre un término de un proceso estocástico reduciendo el índice temporal en una unidad:

$$B X_t = X_{t-1} \Rightarrow B^k X_t = X_{t-k}$$

y por tanto, un proceso autorregresivo puede expresarse en la forma:

$$\varepsilon_t = (1 - a_1 B - a_2 B^2 - \dots - a_p B^p) X_t$$

Consideremos la expresión autorregresiva antes definida. En principio, no hay nada que nos limite el número de periodos del pasado que influyen en el valor actual del proceso, sin embargo tal planteamiento (probablemente cierto en términos teóricos) provoca problemas no justificables por las ventajas prácticas obtenidas: el hecho de tener que estimar infinitos coeficientes de regresión no se justifica cuando en la práctica se observa que sólo los periodos más recientes tienen influencia significativa en el valor actual del proceso. Con ello, la expresión anterior se trunca, dejando el valor actual dependiente de los últimos  $p$  valores del proceso, más una perturbación aleatoria:

$$\varepsilon_t = (1 - a_1 B - a_2 B^2 - \dots - a_p B^p) X_t$$

Un proceso de estas características se denomina *proceso autorregresivo de orden  $p$   $AR(p)$* . Si denotamos por  $a_p(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p$ , al llamado polinomio característico del proceso obtenemos, el siguiente resultado que nos garantiza la estacionariedad del proceso.

**Proposición:** La condición necesaria y suficiente para que un proceso  $AR(p)$  sea estacionario es que las raíces de su polinomio característico estén fuera del círculo unidad del plano complejo.

### Determinación del orden y de los parámetros del modelo $AR$ .

Los dos problemas fundamentales que nos presentan los procesos autorregresivos son:

- Determinación del grado  $p$  del polinomio autorregresivo.
- Una vez fijado este, determinar los parámetros  $a_i$  del modelo.

### Determinación de los parámetros del modelo

En esta sección nos centraremos en la segunda de las cuestiones. Así, si consideramos un proceso  $AR(p)$ :

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t \quad (2)$$

multiplicando por  $X_{t-j}$  y tomado esperanzas

$$E(X_t X_{t-j}) = a_1 E(X_{t-1} X_{t-j}) + a_2 E(X_{t-2} X_{t-j}) + \dots + a_p E(X_{t-p} X_{t-j}) + E(\varepsilon_t X_{t-j}) \quad (3)$$

para  $j = 0$  al ser la función de autocovarianzas simétrica se tiene:

$$\gamma_0 = a_1 \gamma_1 + a_2 \gamma_2 + \dots + a_p \gamma_p + \sigma \varepsilon^2$$

puesto que al ser  $E(\varepsilon_t \varepsilon_{t-j}) = 0$  para  $j > 0$  se tiene (sin mas que aplicar de manera recursiva (2))  $E(\varepsilon_t X_{t-j}) = \sigma \varepsilon^2$ . Por otro lado, dividiendo en (3) por  $\gamma_0$  es fácil probar que la función de autocorrelación asociada al proceso debe verificar la misma ecuación en diferencias:

$$\rho_j = a_1 \rho_{j-1} + a_2 \rho_{j-2} + \dots + a_p \rho_{j-p} \quad j > 0$$

Particularizando la ecuación anterior para  $j = 1, 2, \dots, p$  se obtiene un sistema de ecuaciones que relaciona las  $p$  primeras autocorrelaciones con los parámetros del proceso. Llamaremos ecuaciones de Yule-Walker al sistema:

$$\rho_1 = a_1 + a_2 \rho_1 + \dots + a_p \rho_{p-1}$$

$$\rho_j = a_1 \rho_1 + a_2 \rho_2 \dots + a_p \rho_{p-2}$$

$$\rho_j = a_1 \rho_{p-1} + a_2 \rho_{p-2} + \dots + a_p$$

llamando:

$$\mathbf{a}' = (a_1, a_2, \dots, a_p) \quad \boldsymbol{\rho}' = (\rho_1, \rho_2, \dots, \rho_p)$$

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{pmatrix}$$

el sistema anterior se escribe matricialmente:

$$\boldsymbol{\rho} = \mathbf{R} \cdot \mathbf{a} \Rightarrow \mathbf{a} = \mathbf{R}^{-1} \cdot \boldsymbol{\rho}$$

por consiguiente, los valores de los parámetros  $\mathbf{a}$  se pueden obtener una vez estimada la matriz de autocorrelaciones de orden  $p$ .

Veamos que el modelo de Yule-Walker nos da estimadores óptimos según mínimos cuadrados de los coeficientes  $a_k$  a partir de un conjunto términos observados de la serie. Así, si consideramos un proceso  $AR(p)$ :

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t$$

a partir de  $n$  valores observados de la serie,  $x_1, x_2, \dots, x_n$ , los residuos  $e_t$  vendrán dados por:

$$e_t = x_t - (a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_p x_{t-p}) = x_t - \hat{x}_t$$

donde por  $\hat{x}_t$  denotamos el valor estimado de la serie.

Con el fin obtener los estimadores óptimos de los parámetros según mínimos cuadrados, calcularemos las derivadas parciales respecto de  $a_k$ . Igualando estas parciales a cero, obtenemos las llamadas ecuaciones normales del modelo:

$$\sum_{k=1}^p a_k \sum_{i=1}^n x_{i-k} x_{i-j} = \sum_{i=1}^n x_i x_{i-j} \quad j = 1, 2, \dots, p$$

Si denotamos por:

$$\hat{\rho}_{k-j} = \hat{\rho}_{j-k} = \sum_{i=1}^n x_{i-k} x_{i-j}$$

el sistema anterior se transforma en:

$$\sum_{k=1}^p a_k \cdot \hat{\rho}_{k-j} = \hat{\rho}_j \quad j = 1, 2, \dots, p$$

que se corresponden con las ecuaciones de Yule-Walker donde se ha sustituido las correlaciones teóricas por sus estimaciones.

### Determinación del orden de la autorregresión.

Determinar el orden de un proceso autorregresivo a partir de su función de autocorrelación es difícil. En general esta función es una mezcla de decrecimientos exponenciales y sinusoidales, que se amortiguan al avanzar el retardo, y no presenta rasgos fácilmente identificables con el orden del proceso. Para resolver este problema se introduce la función de autocorrelación parcial.

Si comparamos un  $AR(1)$  con un  $AR(2)$  vemos que aunque en ambos cada observación está relacionada con las anteriores, el tipo de relación entre observaciones separadas dos períodos, es distinto. En el  $AR(1)$  el efecto de  $X_{t-2}$  sobre  $X_t$ , es siempre a través de  $X_{t-1}$ , y no existe efecto directo entre ambas. Conocido  $X_{t-1}$ , el valor de  $X_{t-2}$  es irrelevante para prever  $X_t$ , Esta dependencia puede ilustrarse:

$$AR(1) : X_{t-3} \rightarrow X_{t-2} \rightarrow X_{t-1} \rightarrow X_t$$

