

# Capítulo 1

## Introducción a los Procesos Aleatorios

### 1.1. Procesos Aleatorios y el Teorema de Kolmogorov

**Definición 1.1** Un *Proceso Aleatorio* es una colección de variables aleatorias  $X(t)$  donde  $t$  varía en un espacio de parámetros  $T$ .

En general,  $T$  será el conjunto de los números reales o algún subconjunto como  $\mathbb{N}$ ,  $\mathbb{Z}$ ,  $[0, 1]$  o  $[0, \infty)$  y con frecuencia lo interpretaremos como el tiempo.

Como vemos, un proceso aleatorio  $X$  es una función de dos variables:  $X : \Omega \times T \rightarrow \mathbb{R}$ . Para cada valor fijo de  $t \in T$ ,  $X(\cdot, t)$  es una función medible de  $\Omega$  en  $\mathbb{R}$ , mientras que para cada valor fijo de  $\omega \in \Omega$  tenemos una función  $X(\omega, \cdot)$  de  $T$  en  $\mathbb{R}$ , que representa una *trayectoria* del proceso. Una vez que se selecciona  $\omega \in \Omega$  la función que obtenemos es determinística y podemos pensar que la medida de probabilidad  $P$  sobre el espacio  $\Omega$  representa el mecanismo aleatorio que selecciona la trayectoria. Esta probabilidad  $P$  se conoce como la *distribución* o *ley* del proceso.

Cualquier colección finita de instantes  $\{t_1, \dots, t_n\} \subset T$  define un vector  $(X(t_1), \dots, X(t_n))$  al cual asociamos una medida de probabilidad en  $(\mathbb{R}^n, \mathcal{B}^n)$  que está determinada por la distribución  $P$  del proceso: para cualquier boreliano  $B$  en  $\mathcal{B}^n$ ,

$$P_{t_1, \dots, t_n}(B) = P((X(t_1), \dots, X(t_n)) \in B).$$

En particular si  $B$  es el producto de borelianos en  $\mathbb{R}$ :  $B = B_1 \times \dots \times B_n$ ,

$$P_{t_1, \dots, t_n}(B_1 \times \dots \times B_n) = P(X(t_1) \in B_1, \dots, X(t_n) \in B_n).$$

Estas medidas se conocen como las distribuciones finito-dimensionales del proceso  $X$  y tienen dos propiedades sencillas de demostrar pero importantes. En primer lugar,

$$P_{t_1, \dots, t_n, t_{n+1}}(B_1 \times \dots \times B_n \times \mathbb{R}) = P_{t_1, \dots, t_n}(B_1 \times \dots \times B_n) \quad (1.1)$$

es decir, si añadimos una coordenada sobre la cual no hay ninguna restricción real, la probabilidad no cambia. En segundo lugar, si  $i_1, \dots, i_n$  es una permutación de  $1, \dots, n$ , entonces

$$P_{t_1, \dots, t_n}(B_1 \times \dots \times B_n) = P_{t_{i_1}, \dots, t_{i_n}}(B_{t_1} \times \dots \times B_{t_n}) \quad (1.2)$$

La demostración de ambas propiedades es inmediata.

Lo que resulta más interesante, y menos evidente, es que estas dos sencillas propiedades son suficientes para garantizar la existencia de un proceso aleatorio, como lo demuestra el siguiente teorema debido a Kolmogorov.

**Teorema 1.1 (Kolmogorov)** *Supongamos que para cualesquiera  $t_1, \dots, t_n \in T$ , las medidas de probabilidad  $P_{t_1, \dots, t_n}$  están definidas sobre  $(\mathbb{R}^n, \mathcal{B}^n)$ . Para que exista un proceso aleatorio cuyas distribuciones finito-dimensionales sean*

$$P_{t_1, \dots, t_n}(B), \quad B \in \mathcal{B}^n$$

*es necesario y suficiente que se cumplan las condiciones (1.1) y (1.2).*

Veamos primero algunas ideas previas a la demostración del teorema. Sea  $T$  un conjunto cualquiera de índices y sea  $\mathbb{R}^T$  el conjunto de las funciones de  $T$  en  $\mathbb{R}$ . Así, un elemento  $x \in \mathbb{R}^T$  es una función  $x: T \rightarrow \mathbb{R}$  que toma el valor  $x(t)$  en el punto  $t \in T$ .

Para cada  $t \in T$  definimos la *proyección* o la *función coordenada*  $\Pi_t: \mathbb{R}^T \rightarrow \mathbb{R}$  por

$$\Pi_t(x) = x(t)$$

Un *cilindro* o *cilindro finito-dimensional* es un conjunto definido por restricciones sobre una cantidad finita de coordenadas: Si  $k$  es un entero y  $H \in \mathcal{B}^k$  es un boreliano de  $\mathbb{R}^k$ , un cilindro es un conjunto de la forma

$$A = \{x \in \mathbb{R}^T : (x(t_1), \dots, x(t_k)) \in H\} \quad (1.3)$$

La colección de los cilindros, que denotaremos por  $\mathcal{B}_0^T$ , es un álgebra. Para ver esto observemos, en primer lugar, que  $\mathbb{R}^T$  es un cilindro (basta tomar  $H = \mathbb{R}^k$ ). En segundo lugar, si  $A$  está dado por (1.3), su complemento es

$$A^c = \{x \in \mathbb{R}^T : (x(t_1), \dots, x(t_k)) \in \mathbb{R}^k - H\}$$

que también es un cilindro.

Supongamos ahora que  $A$  está dado por (1.3) y  $B$  por

$$B = \{x \in \mathbb{R}^T : (x(s_1), \dots, x(s_j)) \in I\}$$

donde  $I \in \mathcal{B}^j$  y el conjunto de índices  $\{s_1, \dots, s_j\}$  es disjunto de  $\{t_1, \dots, t_k\}$ . Sea  $m = k + j$ , pongamos  $(u_1, \dots, u_m) = (t_1, \dots, t_k, s_1, \dots, s_j)$  y sea  $H' = H \times \mathbb{R}^j$ ,  $I' = \mathbb{R}^k \times I$ . Entonces

$$A = \{x \in \mathbb{R}^T : (x(u_1), \dots, x(u_m)) \in H'\} \quad \text{y} \quad B = \{x \in \mathbb{R}^T : (x(u_1), \dots, x(u_m)) \in I'\} \quad (1.4)$$

y tenemos que

$$A \cup B = \{x \in \mathbb{R}^T : (x(u_1), \dots, x(u_m)) \in H' \cup I'\} \quad (1.5)$$

Como  $H' \cup I' \in \mathcal{B}^m$ ,  $A \cup B$  es un cilindro. Si los conjuntos de índices no son disjuntos, tomamos un vector  $(u_1, \dots, u_m)$  formado por ambos conjuntos de índices, sin repeticiones ( $m < k + j$ ). Es posible ver que existen conjuntos  $H'$  e  $I'$  en  $\mathcal{B}^m$  tales que las representaciones (1.4) son válidas y por lo tanto, también lo es (1.5). Esto muestra que  $\mathcal{B}_0^T$  es cerrado bajo uniones y complementos y por lo tanto es un álgebra.

La  $\sigma$ -álgebra generada por los cilindros, que denotaremos por  $\mathcal{B}^T$ , es la  $\sigma$ -álgebra generada por las proyecciones:  $\mathcal{B}^T = \sigma\{\Pi_t : t \in T\}$ . En consecuencia, las proyecciones son funciones medibles en el espacio  $(\mathbb{R}^T, \mathcal{B}^T)$ . Si  $P$  es una medida de probabilidad sobre este espacio medible,  $\{\Pi_t : t \in T\}$  es un proceso aleatorio en  $(\mathbb{R}^T, \mathcal{B}^T, P)$ , que se conoce como el proceso canónico. Por lo tanto, para demostrar el Teorema de Kolmogorov basta probar que existe una medida  $P$  sobre  $(\mathbb{R}^T, \mathcal{B}^T)$  que tiene las distribuciones finito-dimensionales adecuadas. Esta medida se conoce como la *ley* o la *distribución* del proceso.

Veamos ahora cual es la idea de la demostración del teorema. Si  $A$  es el cilindro descrito por (1.3) definimos

$$P(A) = P_{t_1, \dots, t_k}(H)$$

En primer lugar hay que demostrar que esta definición es consistente: Si  $A$  tiene otra representación el resultado de aplicar la definición es el mismo. Para esto es necesario usar las condiciones de consistencia.

Consideremos ahora dos cilindros disjuntos  $A$  y  $B$ . Por el procedimiento que explicamos anteriormente podemos suponer que los conjuntos de índices son idénticos. Supongamos que sus representaciones están dadas por (1.4) de modo que (1.5) es válida. Si  $H' \cap I'$  fuese vacío,  $A \cap B$  también lo sería, de modo que

$$\begin{aligned} P(A \cup B) &= P_{u_1, \dots, u_n}(H' \cup I') \\ &= P_{u_1, \dots, u_n}(H') + P_{u_1, \dots, u_n}(I') \\ &= P(A) + P(B) \end{aligned}$$

y vemos que  $P$  es aditiva en  $\mathcal{B}_0^T$ . Además  $P(\mathbb{R}^T) = 1$ .

Si podemos mostrar que  $P$  es  $\sigma$ -aditiva, el Teorema de Extensión de Caratheodory permite extender  $P$  a una medida de probabilidad sobre  $\mathcal{B}^T$ , y por la forma como definimos  $P$  sobre  $\mathcal{B}_0^T$

$$P(x \in \mathbb{R}^T : (\Pi_{t_1}(x), \dots, \Pi_{t_n}(x)) \in H) = P_{t_1, \dots, t_n}(H)$$

y el proceso  $\{\Pi_t : t \in T\}$  tiene las distribuciones finito-dimensionales requeridas.

La demostración de la  $\sigma$ -aditividad de  $P$  sobre  $\mathcal{B}_0^T$  no es difícil pero es notacionalmente complicada y la omitiremos. El lector interesado en este y otros detalles de la demostración puede consultar el libro de P. Billingsley [Billingsley 95].

**Ejemplo 1.1 (Una Sucesión de Variables Independientes.)**

Sea  $F_n(x)$ ,  $n \geq 1$  una sucesión de funciones de distribución cualesquiera. Para cualquier colección de enteros  $n_1, n_2, \dots, n_k$  y de números reales  $x_1, x_2, \dots, x_k$  definimos las funciones de distribución

$$F_{n_1, \dots, n_k}(x_1, \dots, x_k) = F_{n_1}(x_1) \cdots F_{n_k}(x_k)$$

Es inmediato ver que estas funciones de distribución satisfacen las condiciones del Teorema de Kolmogorov y por lo tanto existe un proceso a tiempo discreto  $X_j$ , para  $j \in \mathbb{Z}$  o  $j \in \mathbb{N}$ . Las variables de cualquier subconjunto finito también son independientes. Este proceso es una sucesión de variables independientes. Si además  $F_n = F$  para todo  $n$  se trata de una sucesión de variables independientes idénticamente distribuidas.

**Definición 1.2** Sea  $\{X(t), t \in T\}$  un proceso aleatorio. Si  $E|X(t)| < \infty$  para todo  $t \in T$ , la *función de media* del proceso es  $m(t) = EX(t)$ . Si  $EX^2(t) < \infty$  para todo  $t \in T$ , decimos que el proceso está en  $L^2$  y definimos su *función de covarianza*  $R(s, t)$  por

$$R(s, t) = \text{Cov}(X(s), X(t)) = E[(X(s) - m(s))(X(t) - m(t))]$$

## 1.2. Separabilidad

Ahora que tenemos una manera de definir un proceso aleatorio es natural preguntarnos, por ejemplo, bajo que condiciones las trayectorias del proceso son continuas, o son diferenciables. Lamentablemente, el conjunto de las funciones continuas o el de las funciones diferenciables, no están en la  $\sigma$ -álgebra  $\mathcal{B}^T$  y por lo tanto no pueden ser medidos con la probabilidad  $P$ . Dicho de otra manera, la construcción de Kolmogorov, basada en los cilindros, no da suficiente información sobre conjuntos complejos como el de las funciones continuas, cuya definición depende de **todas** las coordenadas del proceso y no sólo de una cantidad numerable. Veamos un ejemplo.

**Ejemplo 1.2**

Sobre el espacio de probabilidad  $([0, 1], \mathcal{B}[0, 1], \lambda)$  donde  $\lambda$  es la medida de Lebesgue, definimos los procesos  $X(t, \omega) = 0$  para todo  $\omega$  y todo  $t \in [0, 1]$  y

$$Y(t, \omega) = \begin{cases} 0 & \text{si } t \neq \omega \\ 1 & \text{si } t = \omega \end{cases}$$

Entonces, para cada  $t$ ,  $P(\omega : X(t, \omega) = Y(t, \omega)) = 1$ , y en consecuencia ambos procesos tienen las mismas distribuciones finito-dimensionales. Sin embargo, todas las trayectorias de  $X$  son continuas mientras que ninguna de las trayectorias de  $Y$  lo es. Más aun,

$$P(\sup_{[0,1]} X(t, \omega) = 0) = 1, \quad P(\sup_{[0,1]} Y(t, \omega) = 1) = 1.$$

**Definición 1.3** Diremos que el proceso  $\{Y(t), t \in T\}$  es una *versión* del proceso  $\{X(t), t \in T\}$  si  $P(X(t) = Y(t)) = 1$  para cada  $t \in T$ .

En el ejemplo anterior,  $Y$  es una versión de  $X$ . Es fácil ver que si un proceso es versión de otro, ambos tienen las mismas distribuciones finito-dimensionales, y por lo tanto la misma ley.

Un camino para poder considerar propiedades de las trayectorias como continuidad o diferenciability, o poder hablar del supremo de las trayectorias, es la idea de separabilidad, que permite obtener una versión del proceso cuyas trayectorias están esencialmente determinadas por sus valores en un conjunto numerable de parámetros. Consideramos únicamente procesos a valores reales cuyo conjunto de parámetros sea un subconjunto de  $\mathbb{R}$ .

**Definición 1.4** Sea  $\{X(t), t \in T\}$  un proceso aleatorio. Diremos que es *separable* si existe un conjunto numerable denso  $T_0 \subset T$ , llamado el conjunto separante, y un conjunto  $\Lambda$  de probabilidad cero tal que, si  $\omega \notin \Lambda$  y  $t \in T$ , existe una sucesión  $t_n \in T_0$ ,  $t_n \rightarrow t$  con  $X(t_n, \omega) \rightarrow X(t, \omega)$ .

Por lo tanto el comportamiento del proceso está determinado por su comportamiento en  $T_0$ . Es inmediato que si un proceso tiene trayectorias continuas con probabilidad 1 entonces es separable y cualquier conjunto denso de  $T$  es separante.

La demostración del siguiente resultado, que no incluiremos, puede hallarse en [Ash & Gardner 75] o en [Billingsley 95].

**Teorema 1.2 (Separabilidad)** Sea  $\{X(t), t \in T\}$  un proceso aleatorio a valores reales, donde  $T \subset \mathbb{R}$ , entonces existe una versión separable de este proceso.

## 1.3. Algunas Clases Importantes de Procesos

### 1.3.1. Procesos Estacionarios

(a) *Procesos Débilmente Estacionarios o Estacionarios en Sentido Amplio.*

**Definición 1.5** Sea  $\{X(t), t \in T \subset \mathbb{R}\}$  un proceso en  $L^2$ . Decimos que  $X(t)$  es *estacionario en sentido amplio* o *débilmente estacionario* si la función de media es constante:

$$m(t) = \mathbb{E} X(t) = m, \quad \text{para todo } t \in T,$$

y la función de covarianza  $R(t, s)$  sólo depende de la diferencia  $t - s$ :

$$R(t, s) = R(t - s)$$

Como  $R(s, t) = \overline{R(t, s)}$  tenemos en el caso estacionario  $R(t - s) = \overline{R(s - t)}$ . Si el proceso es real, la función de covarianza también lo es y en consecuencia  $R(h) = R(-h)$ , es decir, la función de covarianza es par.

#### Ejemplos 1.3

1. Sea  $X_n, n \geq 1$  una sucesión de variables independientes centradas con  $\mathbb{E} X_n^2 = \sigma_n^2$  y  $\sum \sigma_n^2 < \infty$ . Sea  $\lambda_n, n \geq 1$  una sucesión de números reales. El proceso  $Y(t) = \sum_{n=1}^{\infty} e^{i\lambda_n t} X_n$  es estacionario en sentido amplio puesto que  $\mathbb{E} Y(t) = 0$  y

$$R(t, s) = \sum_{n=1}^{\infty} e^{i\lambda_n(t-s)} \sigma_n^2 = R(t - s)$$

2. Sea  $X(t) = \cos(t\eta + \theta)$  donde  $\theta$  tiene distribución uniforme en  $[0, 2\pi]$  mientras que  $\eta$  es independiente de  $\theta$  y tiene función de distribución  $G(x)$ . El proceso  $X(t)$  es estacionario en sentido amplio puesto que  $\mathbb{E} X(t) = 0$  para todo  $t$  y

$$R(s, t) = \frac{1}{2} \int_{-\infty}^{\infty} \cos[(t-s)x] dG(x) = R(t - s)$$

3. *Proceso de Promedio Móvil.* Sea  $\{X_n, n \in \mathbb{Z}\}$  una sucesión de v.a.i.i.d. centradas y con varianza 1. Definimos

$$Y_n = \sum_{j=0}^k a_j X_{n-j}, \quad n \in \mathbb{Z}$$

donde los coeficientes  $a_j$  son constantes dadas.  $Y_n$  se conoce como un proceso de promedio móvil y es estacionario en sentido amplio:  $\mathbb{E} Y_n = 0$  y

$$\begin{aligned} R(n, n+m) &= \mathbb{E}[Y_n Y_{n+m}] = \mathbb{E}\left[\left(\sum_{j=0}^k a_j X_{n-j}\right) \left(\sum_{p=0}^k a_p X_{n+m-p}\right)\right] \\ &= \begin{cases} a_k a_{k-m} + \cdots + a_{m+1} a_1 & \text{si } m \leq k \\ 0 & \text{si } m > k \end{cases} \\ &= R(m) \end{aligned}$$

Sea  $\{X(t), t \in T\}$  un proceso estacionario en sentido amplio. A partir del criterio para continuidad en media cuadrática (Teorema ??) vemos que  $X$  es continuo en media cuadrática para todo  $t \in T$  si y sólo si la función de covarianza  $R(h)$  es continua en  $h = 0$ . En cuanto a la diferenciabilidad en media cuadrática,  $X$  es diferenciable en media cuadrática si y sólo si la función de covarianza  $R(h)$  es dos veces diferenciable en  $h = 0$ . Si el proceso es diferenciable tenemos

$$\text{Cov}(X(s), X'(t)) = R'(s-t), \quad \text{Cov}(X'(s), X'(t)) = R''(s-t) \quad (1.6)$$

(b) *Procesos Estacionarios en Sentido Estricto.*

**Definición 1.6** Un proceso  $\{X(t), t \in T \subset \mathbb{R}\}$  es estacionario en sentido estricto si sus distribuciones finito-dimensionales son invariantes bajo traslaciones. Más precisamente, sea  $t_1, t_2, \dots, t_n$  un conjunto de índices y  $h > 0$ , entonces las funciones de distribución de  $(X(t_1), X(t_2), \dots, X(t_n))$  y de  $(X(t_1+h), X(t_2+h), \dots, X(t_n+h))$  coinciden.

Un proceso estacionario en sentido estricto que está en  $L^2$  también es estacionario en sentido amplio. El recíproco no es cierto en general.

#### Ejemplos 1.4

1. Una sucesión de variables aleatorias independientes idénticamente distribuidas es un proceso estacionario en sentido estricto.
2. Sea  $X_n, n \geq 1$  una sucesión de v.a.i.i.d. y supongamos que  $a_n, n \geq 1$  es una sucesión de números reales tales que la serie

$$\sum_{n=1}^{\infty} a_n X_{m+n}$$

converge en probabilidad. El proceso  $Y_m = \sum_{n=1}^{\infty} a_n X_{m+n}$  es estacionario en sentido estricto.

La función  $R$  es definida positiva: para cualquier  $n$ , cualesquiera  $t_1, \dots, t_n \in T$  y cualesquiera números reales  $c_1, \dots, c_n$  se tiene

$$\begin{aligned} \sum_{ij} R(t_i - t_j) c_i c_j &= \sum_{ij} \text{E}[(X(t_i) - m(t_i))(X(t_j) - m(t_j))] c_i c_j \\ &= \text{E} \left[ \sum_i (X(t_i) - m(t_i)) c_i \sum_j (X(t_j) - m(t_j)) c_j \right] \\ &= \text{E} \left[ \sum_i (X(t_i) - m(t_i)) c_i \right]^2 \geq 0 \end{aligned}$$

Si la función  $R$  es continua en 0 (lo que equivale a que el proceso sea continuo en probabilidad), el Teorema de Bochner nos dice que puede ser representada de la siguiente manera:

$$R(h) = \int_{\mathbb{R}} e^{ihx} d\mu(x) \quad (1.7)$$

donde  $\mu$  es una medida de Borel sobre la recta con masa total igual a  $R(0) = \sigma^2$ , que se conoce como la *medida espectral* del proceso. Si la varianza del proceso vale 1,  $\mu$  es una medida de probabilidad. La función de distribución  $F$  asociada a la medida  $\mu$ :  $F(x) = \mu(-\infty, x]$ , es la *función de distribución espectral* del proceso. Si  $\mu$  es absolutamente continua respecto a la medida de Lebesgue, su densidad se conoce como la *densidad espectral*.

Si el proceso es real entonces  $R(h) = R(-h) = R(|h|)$  y la representación espectral es

$$\begin{aligned} R(h) &= \int_{\mathbb{R}} e^{ihx} d\mu(x) = \int_{\mathbb{R}} [\cos(hx) + i \text{sen}(hx)] d\mu(x) \\ &= \int_{\mathbb{R}} \cos(hx) d\mu(x) \end{aligned} \quad (1.8)$$

y la medida espectral es simétrica respecto a  $x = 0$ .

La función de distribución espectral se puede obtener a partir de la función de covarianza. Si  $a$  y  $b$  son puntos de continuidad de  $F$ ,

$$F(b) - F(a) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-ita} - e^{-itb}}{it} R(t) dt$$

Si la función de covarianza es absolutamente integrable:  $\int_{\mathbb{R}} |R(t)| dt < \infty$ , entonces existe una densidad espectral que se puede obtener por la ecuación:

$$f(u) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iut} R(t) dt$$

es decir, la densidad es la transformada inversa de la función de correlación.

### 1.3.2. Procesos con Incrementos Estacionarios

**Definición 1.7** Un proceso aleatorio  $\{X(t), t \in T\}$  tiene *incrementos estacionarios* si para todo  $h$  fijo

$$Y(t) = X(t+h) - X(t)$$

es un proceso estacionario.

Al igual que en la sección anterior, hay dos posibilidades: que los incrementos sean estacionarios en sentido débil o en sentido estricto.

### 1.3.3. Procesos con Incrementos Independientes

**Definición 1.8** Un proceso  $\{X(t), t \in T \subset \mathbb{R}\}$  tiene *incrementos independientes* si para cualesquiera  $t_0 < t_1 < \dots < t_n$  las variables aleatorias

$$X(t_0), X(t_1) - X(t_0), \dots, X(t_n) - X(t_{n-1})$$

son independientes.

Para estos procesos, las distribuciones finito-dimensionales están determinadas por la distribuciones de las variables  $X(t)$ ,  $t \in T$  y  $X(t) - X(s)$ ,  $s, t \in T$ : Si  $F_t(x) = P(X(t) \leq x)$  y  $G_{s,t}(x) = P(X(t) - X(s) \leq x)$  entonces, para  $t_1 < t_2 < \dots < t_n$

$$\begin{aligned} F_{t_1, \dots, t_n}(A_1, \dots, A_n) &= P(X(t_1) \in A_1, \dots, X(t_n) \in A_n) \\ &= \int \dots \int \mathbf{1}_{A_1}(x_1) \mathbf{1}_{A_2}(x_1 + x_2) \dots \mathbf{1}_{A_n}(x_1 + \dots + x_n) \\ &\quad \times F_{t_1}(dx_1) G_{t_1, t_2}(dx_2) \dots G_{t_n, t_{n-1}}(dx_n) \end{aligned}$$

Un caso particular importante es el de los *procesos con incrementos independientes y estacionarios*. Supondremos que  $X(0) = 0$ . Observamos que

$$X(t) = (X(t) - X(t-h)) + (X(t-h) - X(t-2h)) + \dots + (X(h) - X(0))$$

donde  $h = t/n$ , o sea

$$X(t) = X_1^{(n)} + \dots + X_n^{(n)},$$

donde las variables  $X_j^{(n)}$ ,  $j = 1, \dots, n$  son i.i.d.

**Definición 1.9** La variable  $X$  tiene una distribución infinitamente divisible si para todo  $n$  existen v.a.i.i.d.  $X_1^{(n)}, \dots, X_n^{(n)}$  tales que la distribución de  $X$  coincide con la distribución de  $X_1^{(n)} + \dots + X_n^{(n)}$ .

En consecuencia la ley de  $X(t)$  es infinitamente divisible. Estas leyes tienen gran importancia en el estudio del teorema central del límite.

Recordemos que la función característica de una suma de variables independientes es el producto de las funciones características de los sumandos. Esto implica que  $X$  tiene distribución infinitamente divisible si y sólo si su función característica  $\phi_X$  se puede escribir como

$$\phi_X(t) = (\phi_n(t))^n, \quad \text{para todo } t$$

donde  $\phi_n$  es una función característica.

**Ejemplo 1.5**

Las distribuciones estables simétricas en  $\mathbb{R}$  son aquellas cuya función característica es  $e^{-c|t|^\alpha}$  para  $\alpha \in (0, 2]$  y algún  $c > 0$ . Para  $\alpha = 2$  tenemos la ley gaussiana centrada de varianza  $c$ . Las leyes estables son infinitamente divisibles, para ver esto basta tomar  $\phi_n(t) = e^{-c|t|^\alpha/n}$

Si  $\{X(t), t \in T\}$  tiene incrementos estacionarios e independientes y  $T = [0, \infty)$  o  $T = 0, 1, 2, \dots$  entonces

$$E(X(t)) = m_0 + mt, \quad \text{Var}(X(t)) = \sigma_0^2 + \sigma^2 t$$

donde  $m_0 = E(X(0))$  y  $\sigma_0^2 = \text{Var}(X(0))$ . Veamos esto para la media del proceso. Llamemos  $f(t) = E(X(t)) - E(X(0))$ , entonces

$$\begin{aligned} f(s+t) &= E[X(s+t) - X(0)] = E[X(s+t) - X(s) + X(s) - X(0)] \\ &= E[X(s+t) - X(s)] + E[X(s) - X(0)] = E[X(t) - X(0)] + E[X(s) - X(0)] \\ &= f(t) + f(s) \end{aligned}$$

y bajo condiciones débiles de regularidad, la solución de esta ecuación funcional es  $f(t) = f(1)t$ . En consecuencia

$$E(X(t)) = E(X(0)) + [E(X(1)) - E(X(0))]t = m_0 + mt$$

**1.3.4. Procesos de Markov**

Un proceso de Markov es un proceso con la propiedad de que si conocemos el valor de  $X(t)$ , el valor de  $X(s)$ ,  $s > t$  no depende del valor de  $X(u)$ ,  $u < t$ . Dicho de otra manera, si conocemos el presente, el futuro es independiente del pasado. La definición formal es la siguiente.

**Definición 1.10** Un proceso  $\{X(t), t \in T\}$  es un proceso de Markov si para cualquier colección finita  $t_1 < t_2 < \dots < t_n$  de puntos en  $T$  se tiene que para cualquier  $x$

$$P(X(t_n) \leq x | X(t_1), \dots, X(t_{n-1})) = P(X(t_n) \leq x | X(t_{n-1})) \quad c. p. 1. \quad (1.9)$$

Para  $A$  un intervalo de  $\mathbb{R}$  la función

$$P(x, s; t, A) = P(X(t) \in A | X(s) = x), \quad t > s,$$

se conoce como la función de transición y es fundamental en el estudio de la estructura de los procesos de Markov. Si esta probabilidad es invariante bajo traslaciones en el tiempo, es decir, si

$$P(x, s; t, A) = P(x, s+h; t+h, A),$$

para cualquier  $h$ , decimos que el proceso es homogéneo.

Es posible demostrar que la distribución de probabilidad de

$$(X(t_1), X(t_2), \dots, X(t_n))$$

puede calcularse en términos de las funciones de transición y de la distribución inicial de  $X(t)$ .

Si  $X(t+h) - X(t)$  es independiente de  $X(s)$ ,  $s \leq t$ , entonces  $X$  es un proceso de Markov

$$\begin{aligned} P(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1) \\ &= P(X(t_n) - X(t_{n-1}) \leq x_n - x_{n-1} | X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1) \\ &= P(X(t_n) - X(t_{n-1}) \leq x_n - x_{n-1} | X(t_{n-1}) = x_{n-1}) \\ &= P(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}) \end{aligned}$$

Un proceso de Markov con espacio de estado finito o numerable se llama una *Cadena de Markov*. Un proceso de Markov con trayectorias continuas es una *difusión*.

**Ejemplos 1.6**

1. Un proceso de Poisson es una cadena de Markov con parámetro continuo.
2. Sea  $X_1, X_2, \dots$  una sucesión de v.a.i. entonces  $S_n = X_1 + \dots + X_n$  es un proceso de Markov.

### 1.3.5. Martingalas

**Definición 1.11** Un proceso aleatorio  $\{X(t), t \in T\}$  es una *martingala* si  $E|X(t)| < \infty$  para todo  $t$  y, para cualquier  $n > 1$  y  $t_1 < t_2 < \dots < t_n < t_{n+1}$ ,

$$E(X(t_{n+1})|X(t_1), \dots, X(t_n)) = X(t_n) \quad c.p. 1 \quad (1.10)$$

En particular, si  $T = \mathbb{N}$  vemos que la sucesión de variables aleatorias  $X_n, n \geq 1$  es una martingala si y sólo si  $E|X_n| < \infty$  para todo  $n$  y

$$E(X_{n+1}|X_1, \dots, X_n) = X_n \quad c.p. 1$$

#### Ejemplos 1.7

1. Sea  $Y, X_1, X_2, \dots$  variables aleatorias con  $E|Y| < \infty$  y definimos

$$Z_n = E(Y|X_1, \dots, X_n).$$

Este proceso es una martingala: Observemos primero que

$$\begin{aligned} E(Z_{n+1}|X_1, \dots, X_n) &= E(E(Y|X_1, \dots, X_{n+1})|X_1, \dots, X_n) \\ &= E(Y|X_1, \dots, X_n) = Z_n \end{aligned}$$

con probabilidad 1. Como  $Z_1, \dots, Z_n$  son variables aleatorias en el mismo espacio que  $X_1, \dots, X_n$ ,

$$E(Z_{n+1}|Z_1, \dots, Z_n, X_1, \dots, X_n) = E(Z_{n+1}|X_1, \dots, X_n) = Z_n.$$

Tomando esperanza condicional dadas  $Z_1, \dots, Z_n$  en ambos lados de la ecuación anterior obtenemos

$$E(Z_{n+1}|Z_1, \dots, Z_n) = Z_n.$$

2. Sea  $X_1, X_2, \dots$  una sucesión de v.a.i. centradas entonces  $S_n = X_1 + \dots + X_n$  es una martingala:

$$\begin{aligned} E(S_{n+1}|S_1, \dots, S_n) &= E(S_n + X_{n+1}|S_1, \dots, S_n) \\ &= S_n + E(X_{n+1}|S_1, \dots, S_n) \\ &= S_n + E(X_{n+1}) = S_n \end{aligned}$$

Si en lugar de (1.10) se tiene

$$E(X(t_{n+1})|X(t_1), \dots, X(t_n)) \leq (\geq) X(t_n) \quad c.p. 1 \quad (1.11)$$

tenemos una supermartingala (submartingala). Observamos que si  $X$  es una supermartingala,  $-X$  es una submartingala. Además, la media de una martingala es constante, la de una supermartingala es creciente. Una supermartingala  $X(t), 0 \leq t$  es una martingala si y sólo si  $E(X(t)) = E(X(0))$ .

**Teorema 1.3 (Convergencia de Martingalas)** Si  $\{X(t), t \geq 0\}$  es una martingala integrable, es decir si  $\sup_{t \geq 0} E|X(t)| < \infty$  entonces existe un límite c.s.  $\lim_{t \rightarrow \infty} X(t) = Y$  e  $Y$  es una variable aleatoria integrable.

### 1.3.6. Procesos Gaussianos

#### La Distribución Gaussiana

Una variable aleatoria  $X$  tiene *distribución Gaussiana* o *normal* con parámetros  $m$  y  $\sigma^2$  si su transformada de Fourier es

$$\phi_X(t) \equiv E(e^{itX}) = \exp\{imt - \frac{1}{2}\sigma^2 t^2\}$$

y en este caso usaremos la notación  $X \sim \mathcal{N}(m, \sigma^2)$ . Cuando  $m = 0$  y  $\sigma^2 = 1$  decimos que  $X$  tiene una distribución Gaussiana *típica* o *estándar* o que esta variable está *normalizada*. En este caso su densidad y su función de distribución son

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad y \quad \Phi(x) = \int_{-\infty}^x \varphi(y) dy$$



Es sencillo verificar que si  $X$  tiene una distribución estándar entonces  $\sigma X + m$  tiene distribución  $\mathcal{N}(m, \sigma^2)$ .

En el caso de un vector aleatorio  $\mathbf{X} = (X_1, \dots, X_n)$  decimos que tiene una *distribución Gaussiana* de parámetros  $\mathbf{m}$  y  $\Sigma$ , donde  $\mathbf{m}$  es un vector  $n$ -dimensional y  $\Sigma$  es una matriz simétrica, definida positiva en sentido amplio y de dimensión  $n \times n$ , si su transformada de Fourier esta dada por

$$\phi(\mathbf{t}) \equiv E(e^{i\langle \mathbf{t}, \mathbf{X} \rangle}) = \exp\{i\langle \mathbf{m}, \mathbf{t} \rangle - \frac{1}{2}\langle \mathbf{t}, \Sigma \mathbf{t} \rangle\} \quad \mathbf{t} \in \mathbb{R}^n \quad (1.12)$$

Si  $\mathbf{m} = \mathbf{0}$  y  $\Sigma = I$ , la matriz identidad, decimos que la distribución de  $\mathbf{X}$  es Gaussiana *típica* o *estándar* en  $\mathbb{R}^n$ . Si la matriz  $\Sigma$  no es singular se puede verificar que la distribución tiene densidad respecto a la medida de Lebesgue en  $\mathbb{R}^n$  dada por

$$\begin{aligned} \varphi_n(\mathbf{x}) &= \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m})^t \Sigma^{-1} (\mathbf{x} - \mathbf{m})\right\} \\ &= \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2|\Sigma|} \sum_{j,k} \Sigma_{jk} (x_j - m_j)(x_k - m_k)\right\} \end{aligned} \quad (1.13)$$

donde  $\mathbf{x}^t$  denota el vector traspuesto de  $\mathbf{x}$ ,  $|\Sigma|$  es el determinante de  $\Sigma$  y  $\Sigma_{jk}$  son los cofactores de  $\Sigma$ .

Es posible verificar que

$$\mathbf{m} = E(\mathbf{X}), \quad \Sigma = \text{Var}(\mathbf{X}) = E[(\mathbf{X} - \mathbf{m})(\mathbf{X} - \mathbf{m})^t]$$

$\mathbf{m}$  se llama la *media* o el *vector de medias* y  $\Sigma$  es la *varianza* de  $\mathbf{X}$ .

En el caso de un vector bidimensional con distribución Gaussiana típica la densidad se escribe

$$\varphi(x, y; \rho) = \frac{1}{2\pi(1 - \rho^2)^{1/2}} \exp\left\{-\frac{x^2 + y^2 - 2\rho xy}{2(1 - \rho^2)}\right\}$$

donde  $\rho = E(XY)$  es la correlación entre  $X$  y  $Y$ . Es inmediato ver que si la correlación es nula, la densidad se factoriza en el producto de dos densidades normales típicas y por lo tanto las variables son independientes. Como el recíproco es siempre cierto (dos variables independientes tienen correlación nula) tenemos el siguiente resultado.

**Proposición 1.1** *Dos variables aleatorias Gaussianas son independientes si y sólo si su correlación vale 0.*

Este resultado se extiende a más de dos variables de modo que  $X_1, X_2, \dots, X_n$  con distribución Gaussiana son independientes si y sólo si la matriz  $\Sigma$  es diagonal. En consecuencia, un vector normal típico tiene componentes normales típicas independientes.

**Observación 1.1** En la Proposición 1.1 es fundamental que la distribución *conjunta* de los dos vectores sea Gaussiana: no basta con que las marginales lo sean. En efecto, supongamos que  $X$  es una variable normal típica y sea  $B$  una variable de Bernoulli con probabilidad de éxito 0.5. Definimos  $Y = X$  si  $B = 1$  e  $Y = -X$  si  $B = 0$ . Entonces  $Y$  también tiene distribución normal típica y la correlación entre ambas variables es

$$E(XY) = E[E(XY|B)] = \frac{E(X^2)}{2} + \frac{E(-X^2)}{2} = 0$$

Sin embargo,  $X$  e  $Y$  no son independientes porque, por ejemplo,

$$P(|X| > 1, |Y| < 1) = 0 \neq P(|X| > 1)P(|Y| < 1).$$

**Observación 1.2** La importancia de la distribución normal en Probabilidades surge, principalmente, del Teorema Central de Límite, que afirma que bajo condiciones generales, la suma de variables independientes normalizadas converge c.s. a una distribución normal típica.

Es inmediato ver que el producto de dos funciones de la forma (1.12) es otra función de la misma forma, de modo que la suma de vectores Gaussianos independientes es otro vector Gaussiano cuya media es la suma de las medias y cuya varianza es la suma de las varianzas.

Hay un recíproco parcial del resultado anterior. Si  $X$  e  $Y$  son independientes y su suma es Gaussiana entonces ambas son Gaussianas. Este resultado se debe a P. Lévy y H. Cramér. La demostración de este resultado se puede encontrar en [Stromberg 94]

Con mayor generalidad, sea  $\mathbf{X} = (X_1, \dots, X_n)$  un vector normal típico y sea  $A$  una matriz  $k \times n$ . Veamos que el vector  $\mathbf{Y} = A\mathbf{X}$  también es un vector Gaussiano centrado. Comenzamos calculando

$$\langle \mathbf{t}, \mathbf{Y} \rangle = \sum_{j=1}^k t_j Y_j = \sum_{j=1}^k \sum_{\ell=1}^n t_j a_{j\ell} X_\ell$$

y como las  $X_\ell$  son normales típicas independientes

$$\begin{aligned} E(e^{i\langle \mathbf{t}, \mathbf{Y} \rangle}) &= E \exp\{i \sum_{\ell} X_\ell (\sum_j t_j a_{j\ell})\} \\ &= \exp\{-\frac{1}{2} \sum_{\ell} (\sum_j t_j a_{j\ell})^2\} \\ &= \exp\{-\frac{1}{2} \mathbf{t}^t A A^t \mathbf{t}\} \end{aligned}$$

y a partir de la definición vemos que  $\mathbf{Y}$  tiene distribución Gaussiana centrada con matriz de covarianza  $\Gamma = A A^t$ .

El recíproco también es cierto: Supongamos que  $\mathbf{Y}$  es un vector Gaussiano centrado con matriz de covarianza  $\Gamma$ . Hemos visto que esta matriz es simétrica y definida positiva en sentido amplio, por lo tanto existe una matriz ortogonal  $O$  que la diagonaliza:

$$O\Gamma O^t = D$$

donde  $D$  es diagonal y sus entradas son positivas.

Consideremos el vector  $\mathbf{Z} = O\mathbf{Y}$ . Este vector es Gaussiano y tiene covarianza

$$E(\mathbf{Z}\mathbf{Z}^t) = E(O\mathbf{Y}\mathbf{Y}^t O^t) = O\Gamma O^t = D$$

luego

$$E(Z_i Z_j) = \begin{cases} 0, & i \neq j \\ d_{ii}, & i = j \end{cases}$$

y en consecuencia las variables  $Z_i$  son independientes  $\mathcal{N}(0, d_{ii})$ . Si  $d_{ii} = 0$  definimos  $X_i \equiv 0$  y para el resto  $X_i = Z_i / \sqrt{d_{ii}}$ , entonces las variables  $X_i$  distintas de 0 son independientes  $\mathcal{N}(0, 1)$  y hay una matriz  $A$  tal que  $\mathbf{Y} = A\mathbf{X}$ .

Si  $\Gamma$  tiene rango máximo decimos que la distribución de  $\mathbf{Y}$  no es degenerada y en este caso los elementos de la diagonal de  $D$  son estrictamente positivos, es decir, todas las componentes de  $\mathbf{X}$  son distintas de cero. De ahora en adelante, las distribuciones que consideraremos serán de este tipo, a menos que se especifique lo contrario.

Si hacemos la transformación  $\mathbf{Y} = A\mathbf{X} + \mathbf{m}$ , donde  $\mathbf{X}$  de nuevo es un vector normal típico y  $\mathbf{m} = (m_1, \dots, m_n)$ , el vector  $\mathbf{Y}$  tiene distribución  $\mathcal{N}(\mathbf{m}, \Sigma)$ . Recíprocamente, si  $\mathbf{Y}$  tiene distribución  $\mathcal{N}(\mathbf{m}, \Sigma)$ , existe un vector normal típico  $\mathbf{X}$  tal que  $\mathbf{Y} = A\mathbf{X} + \mathbf{m}$ .

**Observación 1.3** Al igual que en la observación 1.1, la distribución *conjunta* de las variables debe ser Gaussiana para que cualquier combinación lineal de ellas también lo sea. El mismo ejemplo anterior muestra que se pueden tener variables  $X$  e  $Y$  con distribución individual normal y tales que  $X + Y$  no es normal.

### La Representación Polar.

Sean  $X$  e  $Y$  dos variables gaussianas típicas independientes y consideremos el par  $(X, Y)$ , que representa un punto (aleatorio) en el plano. Este punto tiene una representación en coordenadas polares:

$$X = A \cos \Gamma, \quad Y = A \sen \Gamma, \quad (1.14)$$

donde  $A = \sqrt{X^2 + Y^2}$  y  $\tan \Gamma = Y/X$ . El jacobiano de la transformación (1.14) es

$$\frac{\partial(X, Y)}{\partial(A, \Gamma)} = A$$

y en consecuencia  $A$  y  $\Gamma$  tienen densidad conjunta

$$f_{A, \Gamma}(a, \gamma) = \frac{1}{2\pi} a \exp\left(-\frac{a^2}{2}\right)$$

para  $a \geq 0$ ,  $0 \leq \gamma < 2\pi$ . Esto muestra que  $A$  y  $\Gamma$  son independientes,  $A$  tiene distribución de Rayleigh:

$$f(a) = a \exp\left(-\frac{a^2}{2}\right)$$

mientras que  $\Gamma$  tiene distribución uniforme en  $[0, 2\pi)$ .

### Distribuciones Condicionales

Sea  $\mathbf{X} \sim \mathcal{N}(\mathbf{m}, \Sigma)$ . Para  $k < n$  consideremos las siguientes particiones de  $\mathbf{X}$ ,  $\mathbf{m}$  y  $\Sigma$ :

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \quad \mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

donde

$$\begin{aligned} \mathbf{X}_1 &= (X_1, \dots, X_k)^t, & \mathbf{X}_2 &= (X_{k+1}, \dots, X_n)^t, \\ \mathbf{m}_1 &= (m_1, \dots, m_k)^t, & \mathbf{m}_2 &= (m_{k+1}, \dots, m_n)^t \end{aligned}$$

$\Sigma_{ii}$  es la matriz de covarianza de  $\mathbf{X}_i$ ,  $i = 1, 2$ , y  $\Sigma_{12} = (\sigma_{ij})$  donde  $\sigma_{ij} = \text{Cov}(X_i, X_j)$  para  $1 \leq i < j \leq n$ .

Si  $\mathbf{X}$  tiene distribución  $\mathcal{N}(\mathbf{m}, \Sigma)$  donde  $\Sigma$  tiene rango máximo, la distribución condicional de  $\mathbf{X}_1$  dado que  $\mathbf{X}_2 = \mathbf{x}_2$ , es  $\mathcal{N}(\mathbf{m}_{1.2}, \Sigma_{11.2})$  donde

$$\begin{aligned} \Sigma_{11.2} &= \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \\ \mathbf{m}_{1.2} &= \mathbf{m}_1 + \Sigma_{12} \Sigma_{22}^{-1} (\mathbf{x}_2 - \mathbf{m}_2) \end{aligned}$$

Observamos que  $\mathbf{m}_{1.2}$  es una función lineal de  $\mathbf{x}_2$  y  $\Sigma_{11.2}$  no depende de  $\mathbf{x}_2$ . La matriz  $\Sigma_{12} \Sigma_{22}^{-1}$  se conoce como la matriz de regresión de  $\mathbf{X}_2$  sobre  $\mathbf{X}_1$ .

### Procesos Gaussianos

**Definición 1.12** Un proceso  $\{X(t), t \in T\}$  es un *proceso gaussiano* o *normal* si para cualquier colección finita  $t_1, \dots, t_n$  en  $T$ , el vector  $(X(t_1), \dots, X(t_n))$  tiene una distribución gaussiana

La definición anterior quiere decir que para cualquier vector  $\mathbf{t} = (t_1, \dots, t_n)$  de elementos de  $T$ , existen un vector  $n$ -dimensional, que llamaremos  $\mathbf{m} = \mathbf{m}(\mathbf{t}) = (m(t_1), \dots, m(t_n))$ , y una matriz definida positiva y simétrica de tamaño  $n \times n$ , que llamaremos  $\Sigma = \Sigma(\mathbf{t}) = (R(t_i, t_j))$  donde  $R(t_i, t_j) = \text{Cov}(X(t_i), X(t_j))$  tales que la función característica del vector  $(X(t_1), \dots, X(t_n))$  es

$$E(e^{i\langle \mathbf{t}, \mathbf{X} \rangle}) = \exp\left\{i\langle \mathbf{m}, \mathbf{t} \rangle - \frac{1}{2}\langle \mathbf{t}, \Sigma \mathbf{t} \rangle\right\}.$$

Por los resultados del Capítulo 1, una definición equivalente es la siguiente:

**Definición 1.13** Un proceso  $\{X(t), t \in T\}$  es un *proceso gaussiano* o *normal* si para cualquier colección finita  $t_1, \dots, t_n$  en  $T$  y cualesquiera números reales  $a_1, \dots, a_n$ , la variable  $\sum_j a_j X(t_j)$  tiene distribución normal.

Nuestra primera observación es que siempre hay dos funciones asociadas a un proceso gaussiano. En primer lugar la función de media:  $m(t) = \mathbb{E}X(t)$ , para  $t \in T$  y en segundo, la función de covarianza:

$$R(s, t) = \text{Cov}(X(s), X(t)) = \mathbb{E}[(X(s) - m(s))(X(t) - m(t))] \quad \text{para } s, t \in T.$$

Esta última es simétrica:  $R(s, t) = R(t, s)$  y definida positiva: para cualesquiera  $t_1, \dots, t_n$  y cualesquiera números complejos  $c_1, \dots, c_n$  se tiene que

$$\sum_{ij} R(t_i, t_j) c_i \bar{c}_j \geq 0$$

Más interesante aun es que basta con tener estas dos funciones para definir un proceso gaussiano.

**Teorema 1.4** *Sea  $m(t), t \in T$  una función y  $R(s, t)$  una función simétrica definida positiva, entonces existe un proceso Gaussiano  $\{X(t), t \in T\}$  tal que  $m$  y  $R$  son, respectivamente, su función de media y de covarianza. Además, la distribución de  $X$  es única.*

*Idea de la Demostración.* Para cualesquiera  $t_1, \dots, t_n$  en  $T$ , distintos dos a dos, sea  $P_{t_1, \dots, t_n}$  la medida gaussiana en  $\mathbb{R}^n$  con media

$$\mathbf{m}(t_1, \dots, t_n) = (m(t_1), \dots, m(t_n))^t$$

y matriz de covarianza

$$\Sigma_{t_1, \dots, t_n} = (R(t_i, t_j)) \quad \text{para } i, j = 1, \dots, n.$$

Es fácil verificar que las distribuciones  $\{P_{t_1, \dots, t_n}\}$  verifican las condiciones de consistencia del Teorema de Kolmogorov. En consecuencia, hay una única medida de probabilidad  $P$  sobre el espacio medible  $(\mathbb{R}^T, \mathcal{B}^T)$  cuya restricción a los cilindros que dependen de las coordenadas  $t_1, \dots, t_n$  es  $P_{t_1, \dots, t_n}$ . Esto es válido para cualquier conjunto finito de parámetros  $\{t_1, \dots, t_n\}$  distintos dos a dos. La medida  $P$  es única y se conoce como la medida gaussiana generada por  $(m, R)$ . ■

### Ejemplo 1.8

Si  $T = \mathbb{N}$  tenemos una sucesión de variables gaussianas. En la próxima sección estudiaremos algunas propiedades relacionadas con la convergencia de estas sucesiones.

### Ejemplo 1.9 (El proceso coseno)

Un proceso gaussiano muy sencillo se obtiene de la siguiente manera: Sean  $\eta$  y  $\zeta$  dos variables gaussianas estándar independientes y  $\alpha$  una constante positiva fija. El proceso  $X(t)$  se define como

$$X(t) = \eta \cos \alpha t + \zeta \sin \alpha t$$

Es inmediato que este proceso es gaussiano y que  $\mathbb{E}X(t) = 0$ . También es fácil calcular la covarianza,

$$\begin{aligned} R(t, t+h) &= \mathbb{E}[(\eta \cos \alpha t + \zeta \sin \alpha t)(\eta \cos \alpha(t+h) + \zeta \sin \alpha(t+h))] \\ &= \cos \alpha t \cos \alpha(t+h) + \sin \alpha t \sin \alpha(t+h) \\ &= \cos \alpha h \end{aligned}$$

Si usamos la representación polar:  $\eta = A \cos \phi$  y  $\zeta = A \sin \phi$ , que discutimos en el capítulo 1, podemos escribir el proceso  $X$  como

$$X(t) = A \cos(\alpha t - \phi)$$

que es la forma usual de representar el proceso.

Si consideramos un proceso gaussiano centrado  $\{X(t), t \in T\}$ , su distribución está determinada por la función de covarianza  $R(s, t)$  y, en principio, deberíamos poder determinar las propiedades de sus trayectorias a partir de  $R$ .

Una función asociada a la covarianza de un proceso gaussiano que juega un papel de gran importancia en el problema de determinar condiciones necesarias y suficientes para que las trayectorias sean continuas es la *varianza incremental*, también conocida como *variograma*, que se define como

$$d(s, t) = [\mathbb{E}(X(s) - X(t))^2]^{1/2} = [R(s, s) - 2R(s, t) + R(t, t)]^{1/2}.$$

Observamos que si el proceso es estacionario y tiene varianza  $\sigma^2$

$$d^2(s, t) = 2[\sigma^2 - R(|t - s|)].$$

De acuerdo a la definición 2.5 vemos que el proceso  $X$  es continuo en media cuadrática en  $t$  si y sólo si

$$\lim_{s \rightarrow t} d(s, t) = 0$$

Esta función es una pseudo-métrica sobre el espacio de parámetros  $T$ :  $d(s, t)$  puede valer 0 sin que  $t = s$ . Si el espacio  $T$  tiene una métrica  $\tau$  y  $(T, \tau)$  es compacto entonces, suponiendo que el proceso es continuo en media cuadrática, es posible probar que la continuidad del proceso respecto a  $\tau$  equivale a la continuidad respecto a  $d$  (ver [Adler (1990)]). Por lo tanto, es razonable usar la métrica  $d$ , que viene dada por el proceso, para el estudio del problema de continuidad.

### Sucesiones Gaussianas

**Teorema 1.5** Sea  $\mathbf{X}_n$  una sucesión de vectores gaussianos  $n$ -dimensionales con parámetros  $(\mathbf{m}_n, R_n)$ . La sucesión  $\mathbf{X}_n$  converge débilmente si y sólo si

$$\mathbf{m}_n \rightarrow \mathbf{m}, \quad R_n \rightarrow R. \quad (1.15)$$

En este caso la distribución límite también es gaussiana con parámetros  $\mathbf{m}$  y  $R$ .

*Demostración.* Para que la sucesión de vectores  $\mathbf{X}_n$  converja débilmente a un límite es necesario y suficiente que la sucesión de funciones características  $\phi_n(\mathbf{t})$  converja a una función continua en 0. Consideremos la sucesión  $\{\log \phi_n(\mathbf{t})\}$  cerca de  $\mathbf{t} = \mathbf{0}$  donde

$$\log \phi_n(\mathbf{t}) = i\langle \mathbf{m}_n, \mathbf{t} \rangle - \frac{1}{2}\langle \mathbf{t}, R_n \mathbf{t} \rangle.$$

Para que esta sucesión converja es necesario y suficiente que (1.15) sea válida, en cuyo caso

$$\phi_n(\mathbf{t}) \rightarrow \phi(\mathbf{t}) = \exp\{i\langle \mathbf{m}, \mathbf{t} \rangle - \frac{1}{2}\langle \mathbf{t}, R \mathbf{t} \rangle\}$$

que es la función característica de una distribución gaussiana de parámetros  $(\mathbf{m}, R)$ . ■

**Teorema 1.6** Para una sucesión gaussiana  $\{X_n, n \geq 1\}$ , convergencia en probabilidad es equivalente a convergencia en media cuadrática.

*Demostración.* Convergencia en media cuadrática implica convergencia en probabilidad, así que sólo tenemos que probar el recíproco. Usamos la notación

$$E(X_j - X_k) = m_{j,k}, \quad \text{Var}(X_j - X_k) = \sigma_{j,k}^2.$$

Para que  $(X_n)$  converja en media cuadrática es necesario y suficiente que sea de Cauchy, es decir, que

$$E[(X_j - X_k)^2] = \sigma_{j,k}^2 + m_{j,k}^2 \rightarrow 0, \quad \text{cuando } j, k \rightarrow \infty.$$

Por lo tanto, si  $(X_n)$  no converge en media cuadrática, tenemos que

$$\limsup_{j,k \rightarrow \infty} (\sigma_{j,k}^2 + m_{j,k}^2) > 0$$

pero

$$P(|X_j - X_k| > \varepsilon) = \int_{|x| > \varepsilon} \frac{1}{\sqrt{2\pi}\sigma_{j,k}} \exp\left\{-\frac{(x - m_{j,k})^2}{2\sigma_{j,k}^2}\right\} dx$$

y como  $\sigma_{j,k}^2$  y  $m_{j,k}$  no tienden simultáneamente a 0, para  $\varepsilon > 0$  suficientemente pequeño tenemos

$$\limsup_{j,k \rightarrow \infty} P(|X_j - X_k| > \varepsilon) \geq \frac{1}{2}$$

pero esto implica que  $(X_n)$  no converge en probabilidad, lo cual es una contradicción. ■

**Corolario 1.1** Para una sucesión gaussiana  $\{X_n, n \geq 1\}$ , convergencia c.s. implica convergencia en media cuadrática.

### Procesos Gaussianos Estacionarios

Recordemos que en el Capítulo 2 vimos la definición de proceso estacionario, tanto en sentido débil o amplio como en sentido estricto. Para los procesos gaussianos ambas definiciones coinciden porque, como hemos visto, las distribuciones finito-dimensionales están totalmente determinadas por la función de medias  $m(t)$  y la función de covarianzas  $R(s, t)$ . Por lo tanto un proceso gaussiano es estacionario si su función de media es constante y su función de covarianza satisface

$$R(t, s) = R(t - s) \quad \text{para } s, t \in T.$$

Un proceso estacionario tiene varianza constante:  $\text{Var}(X(t)) = R(0) = \sigma^2$ . En general, a menos que se diga lo contrario, supondremos que los procesos estacionarios están centrados, de modo que  $m(t) = 0$  para todo  $t$ , y tienen varianza 1.

Como ejemplo tenemos el proceso coseno, que definimos en la sección 3.1, cuya función de covarianza está dada por  $R(h) = \cos \alpha h$ .

Si la función  $R$  es continua en 0, el Teorema de Bochner nos dice que puede ser representada de la siguiente manera:

$$R(h) = \int_{\mathbb{R}} e^{ihx} d\mu(x) \quad (1.16)$$

donde  $\mu$  es una medida de Borel sobre la recta con masa total igual a  $R(0) = \sigma^2$ , que se conoce como la *medida espectral* del proceso. Si la varianza del proceso vale 1,  $\mu$  es una medida de probabilidad y si  $\mu$  es absolutamente continua respecto a la medida de Lebesgue, su densidad se conoce como la *densidad espectral* del proceso.

Si el proceso es real entonces  $R(h) = R(-h) = R(|h|)$  y la representación espectral es

$$\begin{aligned} R(h) &= \int_{\mathbb{R}} e^{ihx} d\mu(x) = \int_{\mathbb{R}} [\cos(hx) + i \text{sen}(hx)] d\mu(x) \\ &= \int_{\mathbb{R}} \cos(hx) d\mu(x) \end{aligned} \quad (1.17)$$

La medida espectral es simétrica respecto a  $x = 0$ .

El *momento espectral* de orden  $k$  se define como

$$\lambda_k = \int_{\mathbb{R}} x^k d\mu(x)$$

siempre que la integral exista. La existencia de estos momentos está asociada a la regularidad de las trayectorias del proceso. Por la simetría de  $\mu$ , los momentos de orden impar son todos nulos.

**Lema 1.1** *Supongamos que  $R(0) = 1$ ,*

$$i) \lim_{h \rightarrow 0} \frac{2(1 - R(h))}{h^2} = \lambda_2 = \int_{\mathbb{R}} \lambda^2 d\mu(x).$$

ii) *La derivada segunda  $R''(h)$  existe y es finita en  $h = 0$  si y sólo si  $\lambda_2 < \infty$ , en cuyo caso  $R''(0) = -\lambda_2$  y  $R''(h)$  existe para todo  $h$ .*

*Demostración.* i) Si  $\lambda_2 < \infty$  entonces, teniendo en cuenta que  $R(0) = 1$ , tenemos que

$$\frac{2(1 - R(h))}{h^2} = \int_{\mathbb{R}} \lambda^2 \frac{1 - \cos h\lambda}{h^2 \lambda^2 / 2} d\mu(\lambda)$$

y por el teorema de convergencia dominada obtenemos el resultado. Si  $\lambda_2 = \infty$  el resultado sigue del lema de Fatou.

ii) Si  $\lambda_2 < \infty$ , derivando dos veces la ecuación (1.8) dentro de la integral obtenemos que  $R''(h)$  existe y

$$-R''(h) = \int_{\mathbb{R}} \lambda^2 \cos \lambda h d\mu(\lambda)$$

de donde obtenemos que  $R''(0) = -\lambda_2$ . Por otro lado, si  $R(h)$  tiene segunda derivada finita en 0, ponemos  $S(x) = R(x) - R(x-t)$  para obtener

$$\frac{-2(1-R(t))}{t^2} = \frac{S(t) - S(0)}{t^2} = \frac{S'(\theta t)}{t}, \quad 0 < \theta = \theta(t) < 1.$$

En consecuencia

$$\begin{aligned} \frac{-2(1-R(t))}{t^2} &= \frac{R'(\theta t) - R'((\theta-1)t)}{t^2} \\ &= \theta[R''(0) + o(1)] - (\theta-1)[R''(0) + o(1)] \\ &= R''(0) + o(1) \quad \text{cuando } t \rightarrow 0 \end{aligned}$$

y a partir de i) obtenemos  $\lambda_2 < \infty$ . ■





# Bibliografía

- [Adler (1990)] Adler, R.J., *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, Institute of Mathematical Statistics Lecture Notes Vol. 12, 1990.
- [Anderson (1984)] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis, 2nd. Ed.*, Wiley, 1984.
- [Ash & Gardner 75] Ash, R.B. & M. F. Gardner, *Topics in Stochastic Processes*, Academic Press, 1975
- [Belyaev 61] Belyaev, Yu. K. *Continuity and Hölder conditions for sample functions of stationary Gaussian processes*, Proc. Fourth Berkeley Symp. Math. Statist. Prob. **2** (1961) 23-33.
- [Berman 64] Berman, S.M., *Limit Theorems for the Maximum Term in Stationary Sequences*, Ann. Math. Statist. **35** (1964) 502-516.
- [Berman 87] Berman, S.M., *An Extension of Plackett's differential Equation for the Multivariate Normal Density*, SIAM J. Alg. Disc. Meth. **8** (1987) 196-197.
- [Billingsley 95] Billingsley, P., *Probability and Measure, 3rd. Ed.*, Wiley, 1995.
- [Breiman 68] Breiman, L., *Probability*, Addison-Wesley, 1968.
- [Cramér & Leadbetter 67] Cramér, H. & M.R. Leadbetter, *Stationary and Related Sochastic Processes*, Wiley, 1967.
- [Csörgő & Révész 81] Csörgő, M. & P. Révész, *Strong Approximations in Probability and Statistics*, Academic Press, 1981.
- [Csörgő 79] Csörgő, M., *Brownian Motion - Wiener Process*, Canad. Math. Bull. **22**(1979) 257-279.
- [Doob 53] Doob, J.L. *Stochastic Processes*, Wiley, 1953.
- [Dudley 67] Dudley, R.M., *The Sizes of Compact Subsets of Hilbert Space and Continuity of Gaussian Processes*, J. Functional Analysis **1** (1967) 290-330.
- [Dudley 73] Dudley, R.M., *Sample Functions of the Gaussian Process*, Ann. Probab. **1** (1973) 66-103.
- [Dvoretzki et al. 61] Dvoretzki, A., P. Erdős & S. Kakutani, *Nonincrease everywhere of the Brownian motion process*, Proc. 4th Berkeley Symp. on Math. Stat. Prob. Vol II (1961) 103-116
- [Fernandez de la Vega 73] Fernandez de la Vega, W., *On Almost Sure Convergence of Quadratic Variation*, Ann. Probab. **2** (1973) 551-552.
- [Freedman 71] Freedman, D., *Brownian Motion and Diffusion*, Holden-Day, 1971.
- [Gikhman & Skorokhod 69] Gikhman, I.I. & A.V. Skorokhod, *Introduction to the Theory of Random Processes*, W.B. Saunders Co., 1969.

- [Hida 80] Hida, T., *Brownian Motion*, Springer, 1980.
- [Hida & Hitsuda 93] Hida, T. & M. Hitsuda, *Gaussian Processes*, American Math. Soc. 1993.
- [Ibragimov & Rozanov 78] Ibragimov, I.A. & Y.A. Rozanov, *Gaussian Random Processes*, Springer 1978.
- [Jain & Marcus 78] Jain, N.C. & M.B. Marcus, *Continuity of Subgaussian Processes*, en: Probability on Banach Spaces, Advances in Probability and Related Topics, Vol. 4, J. Kuelbs (ed.), 81-196, Marcel Dekker, 1978.
- [Kahane 85] Kahane, J.-P. *Some Random Series of Functions*, Cambridge Univ. Press, 1985.
- [Kannan 79] Kannan, D., *An Introduction to Stochastic Processes*, North-Holland, 1979.
- [Kolmogorov 56] Kolmogorov, A.N., *Foundations of Probability*, Chelsea, 1956.
- [Koroliuk 81] Koroliuk, V.S., *Manual de la Teoría de Probabilidades y Estadística Matemática*, Mir, 1981.
- [Marcus & Shepp 72] Marcus, M.B. & L.A. Shepp, *Sample Behavior of Gaussian Processes*, en: Proc. 6th. Berkeley Symp. Math. Statist. Probab., L.M. Le Cam, J. Neyman & E. L. Scott (eds.), 423-441, Univ. California Press, 1972.
- [Mathai & Pederzoli 77] Mathai, A.M. & G. Pederzoli, *Characterizations of the Normal Probability Law*, Wiley, 1977.
- [Ortega 82] Ortega, J., *Asymptotic Behaviour of Gaussian Random Fields*, Z. Wahr. verw. Gebiete 59, 169-177 (1982).
- [Pickands 69a] Pickands, J. III *Upcrossing Probabilities for Stationary Gaussian Processes*, Trans. Amer. Math. Soc. **145** (1969) 51-73.
- [Pickands 69b] Pickands, J. III *Asymptotic Properties of the Maximum in a Stationary Gaussian Process*, Trans. Amer. Math. Soc. **145** (1969) 75-86.
- [Plackett 54] Plackett, R.L. *A Reduction Formula for Normal Multivariate Integrals*, Biometrika **41** (1954) 351-360.
- [Samorodnitsky & Taqqu 94] Samorodnitsky, G. & Taqqu, M.S. *Stable Non-Gaussian Random Processes*, Chapman & Hall, 1994.
- [Slepian 62] Slepian, D., *The One-sided Barrier Problem for Gaussian Noise*, Bell System Tech. J. **41** (1962) 463-501.
- [Stromberg 94] Stromberg, K., *Probability for Analysts*, Chapman & Hall, 1994.
- [Tong 90] Tong, Y.L., *The Multivariate Normal Distribution*, Springer, 1990.
- [Wentzell 81] Wentzell, A.D., *A Course in the Theory of Stochastic Processes*, McGraw-Hill, 1981.
- [Wróbel 82] Wróbel, A. *On the Almost Sure Convergence of the Square Variation of the Brownian Motion*, Probab. Math. Statist. **3** (1982) 97-101.
- [Wschebor 01] Wschebor, M. *One-Parameter Gaussian Processes: Lectures on the Distribution of the Maximum*, XIV Escuela Venezolana de Matemáticas, Asociación Matemática Venezolana, 2001.