

## Capítulo 3

# Procesos estocásticos estacionarios

**Maurice. S. Bartlett** (1910-2002)

Estadístico británico. Fue profesor en Manchester, donde fundó el Departamento de Estadística, en Londres, University College y finalmente en Oxford, hasta su jubilación. Hizo contribuciones fundamentales a los procesos estocásticos, especialmente a la inferencia de las autocorrelaciones y el espectro. Fue elegido en 1961 miembro de la Royal Society.

### 3.1 Introducción

Los fenómenos dinámicos que observamos mediante series temporales pueden clasificarse en dos clases. Los primeros son los que toman valores estables en el tiempo alrededor de un nivel constante, sin mostrar tendencia a crecer o a decrecer a largo plazo, como la cantidad de lluvias anuales en una región, la temperatura media del año o la proporción de nacimientos que corresponde a varones. Estos procesos se denominan estacionarios y en este capítulo comenzamos su estudio, que continuará en los siguientes. Una segunda clase de procesos que estudiaremos en el capítulo 6 son los procesos no estacionarios, que son aquellos que pueden mostrar tendencia, estacionalidad y otros efectos evolutivos en el tiempo. Por ejemplo la serie de la renta anual en un país, las ventas de una empresa o la demanda de energía son series que evolucionan en el tiempo con tendencias más o menos estables. La clasificación de una serie como estacionaria o no depende del periodo de observación, ya que la serie puede ser estable en un periodo dado pero no estacionaria en un periodo mayor. Por ejemplo, la temperatura en un punto de la tierra puede parecer estable en un intervalo de 20 años, y mostrar una tendencia creciente en un periodo de miles de años. Como regla general cuanto más amplio sea el periodo de observación más difícil es que el proceso sea estable.

Las series temporales evolucionan normalmente con cierta inercia, que se manifiesta en cierta dependencia entre sus valores presentes y sus valores previos. Para medir esta la dependencia lineal entre el presente y el pasado se utilizan las funciones de autocovarianzas y autocorrelaciones, que definiremos en este capítulo, y que generalizan la idea de covarianza y correlación entre dos variables. El análisis que presentamos en este libro se basa principalmente en la dependencia temporal, y es el más utilizado en la práctica. Un análisis alternativo es el análisis frecuencial, que se utiliza para describir series que son el resultado de la mezcla de muchas ondas cíclicas. Este segundo tipo de análisis, que tiene especialmente aplicaciones en física e ingeniería se presenta brevemente al final del capítulo.

### 3.2 El concepto de proceso estocástico

Un proceso estocástico es un conjunto de variables aleatorias  $\{z_t\}$  donde el índice  $t$  toma valores en un cierto conjunto  $C$ . En nuestro caso, este conjunto es ordenado y corresponde a los instantes temporales (días, meses, años, etc.). Para cada valor  $t$  del conjunto  $C$  (para cada instante temporal) esta definida una variable aleatoria,  $z_t$ , y los valores observados de las variables aleatorias en distintos instantes forman una serie temporal. Es decir, una serie de  $T$  datos,  $(z_1, \dots, z_t, \dots, z_T)$ , es una muestra de tamaño uno del vector

de  $T$  variables aleatorias ordenadas en el tiempo correspondientes a los momentos  $t = 1, \dots, T$ , y la serie observada se considera una realización o trayectoria del proceso estocástico.

El proceso queda caracterizado si definimos la distribución de probabilidad conjunta de las variables aleatorias  $(z_1, \dots, z_t, \dots, z_T)$ , para cualquier valor de  $T$ . Estas distribuciones se denominan las distribuciones finito-dimensionales del proceso. Diremos que conocemos la estructura probabilística de un proceso estocástico cuando se conozcan estas distribuciones, que determinan la distribución de cualquier subconjunto de variables y, en particular, las distribuciones marginales de cada variable.

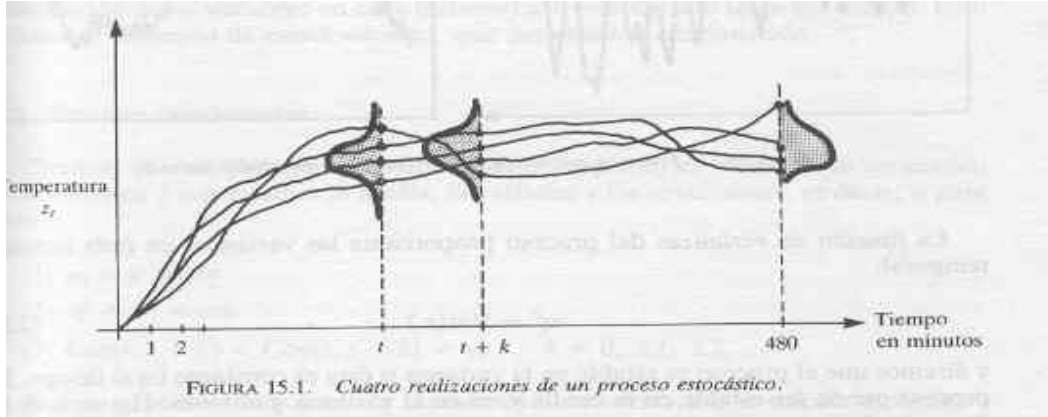


Figura 3.1: Cuatro realizaciones de un proceso estocástico

Por ejemplo, supongamos que en una empresa industrial se mide la temperatura de un horno cada minuto desde que comienza a funcionar, a las 9 horas, hasta que se apaga, a las 17 horas. Cada día se obtienen 480 observaciones que corresponden a la temperatura del horno en los instantes 9h 1m, 9h 2m, ..., 16h 59m, 17h, que asociaremos a los valores  $t = 1, \dots, 480$ . El conjunto de las mediciones de la temperatura una día cualquiera constituyen una realización del proceso  $\{z_t\}$ ,  $t = 1, \dots, 480$ , siendo  $z_t$ , la variable aleatoria: temperatura del horno en el momento  $t$ . Si disponemos de muchas realizaciones del proceso (datos de muchos días) podemos obtener la distribución de probabilidad de las variables que lo componen. Para ello, hay que suponer que la temperatura del horno en un instante  $t_0$  fijo, sigue en distintos días una distribución de probabilidad definida. Esto implica que al observar  $N$  días distintos la frecuencia relativa con la que  $z_{t_0}$  pertenece a un intervalo fijo  $(a, b)$  converge, al aumentar el número de días, a una constante, que es la probabilidad del intervalo  $(a, b)$  en la distribución de probabilidad de la variable  $z_{t_0}$ . Si esta hipótesis es cierta, lo que implica que la situación del horno es la misma en los distintos días, podemos obtener la distribución de cada variable  $z_t$ , estudiando la distribución de la  $t$ -ésima observación los distintos días. Análogamente, podemos estudiar la distribución conjunta de dos variables consecutivas,  $(z_{t_0}, z_{t_0+1})$ , tomando los pares de valores observados en los instantes  $t_0$  y  $t_0 + 1$  en distintos días. Por el mismo procedimiento podemos obtener la distribución conjunta de cualquier grupo de las variables del proceso. La figura 3.1 presenta cuatro realizaciones de este proceso.

Como segundo ejemplo, consideremos una variable climática (como la temperatura, la lluvia, la humedad, la contaminación) a las 12 horas de cada día del año en una ciudad concreta. Las 365 observaciones recogidas cada año para esta variable constituyen una realización del proceso estocástico. Al comenzar otro año se obtiene una nueva realización del proceso y así sucesivamente. Si estudiamos el valor el 9 de Junio de los distintos años tendremos la distribución de probabilidad de una variable aleatoria. Por ejemplo, la figura 3.2 presenta 12 realizaciones de la serie de lluvia en Madrid en los 12 meses del año. Tenemos por tanto 12 variables aleatorias, una para cada mes y la trayectoria de los 12 valores en un año dado representa una realización del proceso estocástico. Si consideramos un mes cualquiera, los 12 valores disponibles forman una muestra de tamaño 12 de esa variable aleatoria.

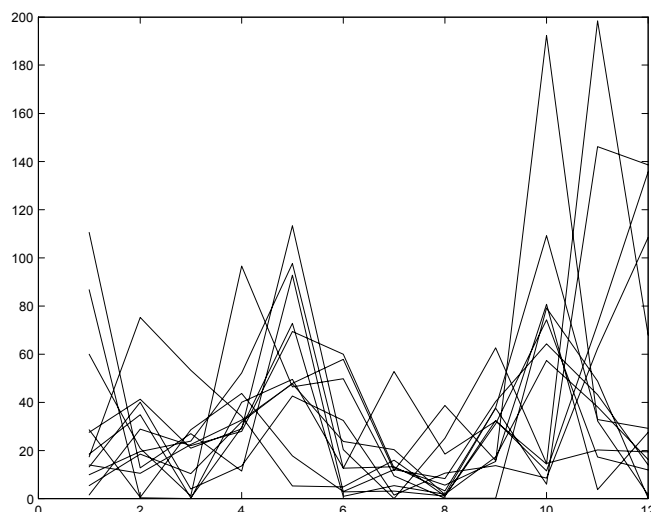


Figura 3.2: Series mensuales de la cantidad de lluvia en Madrid en los años 1988 a 1999

Podemos observar varias realizaciones de un proceso estocástico en fenómenos estables repetitivos a lo largo del tiempo. Por ejemplo, con las variables climáticas, astronómicas o ambientales que siguen cada año un patrón similar como consecuencia de la rotación de la tierra alrededor del sol. También, en variables que se repiten en las mismas condiciones, como el número de clientes en un puesto de servicio (supermercado, gasolinera, etc) en la primera semana del mes en distintos meses que podemos suponer similares, o el ritmo cardiovascular de una persona a lo largo del día (en días similares) , o el rendimiento obtenido por una acción en distintos días, etc.

Determinar la distribución conjunta del proceso requiere observar un gran número de realizaciones, para estimar la probabilidad de los distintos intervalos. Esta tarea se simplifica mucho cuando podemos suponer que la distribución conjunta es normal multivariante, ya que, entonces, quedará determinada por el vector de medias y la matriz de varianzas y covarianzas entre las variables.

Llamaremos *función de medias* del proceso a una función del tiempo que proporciona las esperanzas de las distribuciones marginales  $z_t$  para cada instante:

$$E[z_t] = \mu_t \quad (3.1)$$

Un caso particular importante, por su simplicidad, aparece cuando todas las variables tiene la misma media y la función de medias es una constante. Entonces, las realizaciones del proceso no mostrarán ninguna tendencia y diremos que el proceso es estable en la media. Si, por el contrario, las medias cambian con el tiempo, las observaciones de distintos momentos mostrarán dicho cambio. La figura 3.3 presenta una serie temporal que muestra, aproximadamente, una media constante a lo largo del tiempo. Por el contrario la serie de la figura 3.4 muestra claramente una media no estable. Finalmente, la serie de lluvias de la figura 3.2 tampoco tiene media constante, ya que la media de lluvia es distinta en los diferentes meses del año.

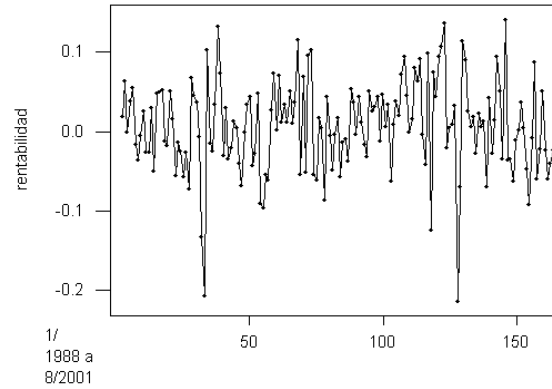


Figura 3.3: Rentabilidad mensual de la bolsa de Madrid en el periodo enero de 1988 hasta Septiembre del 2001

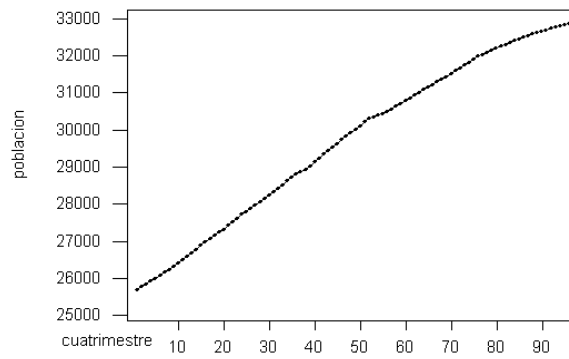


Figura 3.4: Serie de población mayor de 16 años en España en el periodo primer cuatrimestre de 1977 al cuarto del 2000

Se denomina *función de varianzas del proceso* a la que proporciona las varianzas en cada instante temporal:

$$\text{Var}(z_t) = \sigma_t^2 \quad (3.2)$$

y diremos que el proceso es estable en la varianza si ésta es constante en el tiempo. Un proceso puede ser estable en la media y no en la varianza y al revés. Por ejemplo, la serie de la figura 3.3 parece tener media constante pero parece haber mayor variabilidad en ciertos tramos que en otros con lo que la varianza de la serie no sería constante.

La estructura de dependencia lineal entre las variables aleatorias del proceso se representa por las funciones de covarianza y correlación. Llamaremos *función de autocovarianzas* del proceso a la función de dos

argumentos que describe las covarianzas entre dos variables del proceso en dos instantes cualesquiera:

$$\gamma(t, t+j) = \text{Cov}(z_t, z_{t+j}) = E[(z_t - \mu_t)(z_{t+j} - \mu_{t+j})] \quad (3.3)$$

En particular, tenemos que

$$\gamma(t, t) = \text{Var}(z_t, z) = \sigma_t^2$$

La función de medias y la de autocovarianzas cumplen en un proceso estocástico el mismo papel que la media y la varianza para una variable escalar.

Las autocovarianzas tienen dimensiones, las del cuadrado de la serie, por lo que no son convenientes para comparar series medidas en unidades distintas. Podemos obtener una medida adimensional de la dependencia lineal generalizando la idea del coeficiente de correlación lineal entre dos variables. Llamaremos coeficiente de autocorrelación de orden  $(t, t+j)$  al coeficiente de correlación entre las variables  $z_t, z_{t+j}$  y *función de autocorrelación* a la función de dos argumentos que describe estos coeficientes para dos valores cualesquiera de las variables. Esta función será

$$\rho(t, t+j) = \frac{\text{Cov}(t, t+j)}{\sigma_t \sigma_{t+j}} = \frac{\gamma(t, t+j)}{\gamma^{1/2}(t, t) \gamma^{1/2}(t+j, t+j)} \quad (3.4)$$

### 3.3 Procesos estacionarios

#### 3.3.1 Definición

La obtención de las distribuciones de probabilidad del proceso es posible en ciertas situaciones, por ejemplo con variables climáticas, donde podemos suponer que cada año se observa una realización el mismo proceso, o técnicas, que pueden generarse en un laboratorio. Sin embargo, en muchas situaciones de interés, como ocurre con variables económicas o sociales, sólo podemos observar una realización del proceso. Por ejemplo, si observamos la serie de crecimientos anuales de la riqueza de un país no es posible volver atrás en el tiempo para generar otra realización. El proceso estocástico existe conceptualmente, pero no es posible obtener muestras sucesivas o realizaciones independientes del mismo. Para poder estimar las características «transversales» del proceso (medias, varianzas, etc.) a partir de su evolución «longitudinal» es necesario suponer que las propiedades «transversales» (distribución de las variables en cada instante) son estables a lo largo del tiempo. Esto conduce al concepto de estacionaridad, que definimos a continuación.

Diremos que un proceso estocástico (serie temporal) es *estacionario en sentido estricto* si:

- (1) las distribuciones marginales de todas las variables son idénticas;
- (2) las distribuciones finito-dimensionales de cualquier conjunto de variables sólo dependen de los retardos entre ellas.

La primera condición establece que, en particular, la media y la varianza de todas las variables son las mismas. También lo son los coeficientes de asimetría y kurtosis de las distribuciones marginales, ya que estas distribuciones son las mismas para todos los retardos. La segunda condición impone que la dependencia entre las variables además sólo depende de sus retardos, es decir la misma dependencia existe entre las variables  $z_t, z_{t+j}, z_{t+j+h}$  que entre las variables  $z_{t+k}, z_{t+k-j}, z_{t+k-j-h}$ . Estas dos condiciones pueden resumirse estableciendo que la distribución conjunta de cualquier conjunto de variables no se modifica si trasladamos las variables en el tiempo, es decir:

$$F(z_i, z_j, \dots, z_k) = F(z_{i+h}, z_{j+h}, \dots, z_{k+h}).$$

La estacionaridad estricta es una condición muy fuerte, ya que para contrastarla es necesario disponer de las distribuciones conjuntas para cualquier selección de variables del proceso. Una propiedad más débil, pero más fácil de contrastar en la práctica, es la *estacionaridad en sentido débil*, que implica la estabilidad de la media, la varianza y la estructura de covarianzas a lo largo del tiempo. Un proceso es estacionario en sentido débil si, para todo  $t$ :

1.  $\mu_t = \mu = cte$
2.  $\sigma_t^2 = \sigma^2 = cte$
3.  $\gamma(t, t-k) = E[(z_t - \mu)(z_{t-k} - \mu)] = \gamma_k \quad k = 0, \pm 1, \pm 2$

Las dos primeras condiciones indican que la media y la varianza son constantes. La tercera que la covarianza entre dos variables depende sólo de su separación. En un proceso estacionario las autocovarianzas y autocorrelaciones sólo dependen del retardo entre las observaciones y en particular, la relación entre  $z_t$  y  $z_{t-k}$ , es siempre igual a la relación entre  $z_t$  y  $z_{t+k}$ . Podemos escribir:

$$Cov(z_t, z_{t+k}) = Cov(z_{t+j}, z_{t+j+k}) = \gamma_k, \quad j = 0 \pm 1, \pm 2, \dots$$

y también, para las autocorrelaciones:

$$\rho(k) = \frac{\gamma_k}{\gamma_0}.$$

Se verifica:

1.  $\gamma_k = \gamma_{-k}$ .
2.  $\gamma_0 = \sigma^2$
3.  $\rho_k = \gamma_k / \gamma_0 = \rho_{-k}$

Se denomina *función de autocorrelación simple (fas)* a la representación de los coeficientes de autocorrelación del proceso en función del retardo.

La estacionaridad débil no garantiza la estabilidad completa del proceso. Por ejemplo, la distribución de las variables  $z_t$  puede estar cambiando en el tiempo. Sin embargo, si suponemos que estas variables tienen *conjuntamente* una distribución normal  $n$ -dimensional, como ésta queda determinada por las medias, las varianzas y las covarianzas, todas las distribuciones marginales serán idénticas, y también lo serán las distribuciones de subconjuntos de hasta  $n-1$  variables del tipo  $(z_{t_1}, z_{t_1+j}, z_{t_1+k})$  y  $(z_{t_2}, z_{t_2+j}, z_{t_2+k})$  para cada  $t$ . Por tanto, para procesos con distribuciones conjuntamente normal la estacionaridad débil coincide con la estricta.

### 3.3.2 Combinaciones de procesos estacionarios

Una propiedad importante de los procesos estacionarios es que son estables ante combinaciones lineales, es decir, los procesos obtenidos mediante combinaciones lineales de procesos estacionarios son también estacionarios. En particular, la combinación lineal formada por los incrementos de un proceso estacionario conduce también a una serie estacionaria. Es decir, si  $z_t$  es estacionario, el proceso  $w_t$  definido por:

$$w_t = z_t - z_{t-1}$$

es también estacionario. En efecto, la esperanza de  $w$ , es siempre cero, y su varianza es constante, ya que, llamando  $\gamma_1$  a la covarianza entre observaciones contiguas:

$$Var(w_t) = Var(z_t) + Var(z_{t-1}) - 2Cov(z_t, z_{t-1}) = 2(\sigma^2 - \gamma_1),$$

que no depende de  $t$ . Además su función de autocovarianza es:

$$Cov(w_t, w_{t+k}) = E[(z_t - z_{t-1})(z_{t+k} - z_{t+k-1})] = 2\gamma_k - \gamma_{k+1} - \gamma_{k-1}$$

y depende sólo del retardo  $k$  y no del instante  $t$ .

Vamos a demostrar en general que toda combinación lineal de procesos estacionarios conduce a un nuevo proceso estacionario. Sea  $\mathbf{z}_t = (z_{1t}, \dots, z_{kt})$  un vector de  $k$  procesos estacionarios donde suponemos que las autocovarianzas sólo dependen del retardo y las covarianzas entre dos componentes en dos instantes de tiempo sólo dependen de los dos componentes considerados y del retardo entre los instantes temporales. En estas condiciones el vector de series es estacionario. Consideremos el proceso escalar definido por el vector de constantes  $\mathbf{c}' = (c_1, \dots, c_k)$ :

$$y_t = \mathbf{c}'\mathbf{z}_t = c_1 z_{1t} + \dots + c_k z_{kt}$$

que será una combinación lineal de los componentes del vector  $\mathbf{z}_t$ . La esperanza de ese proceso es

$$E(y_t) = c_1 E(z_{1t}) + \dots + c_k E(z_{kt}) = \mathbf{c}' \boldsymbol{\mu}$$

donde  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$  es el vector de las medias de los componentes. Como las medias son constantes también lo será  $E(y_t)$ . La varianza del proceso  $y_t$  será:

$$\text{var}(y_t) = E(\mathbf{c}'(\mathbf{z}_t - \boldsymbol{\mu})(\mathbf{z}_t - \boldsymbol{\mu})\mathbf{c}) = \mathbf{c}'\Gamma_z\mathbf{c}$$

donde  $\Gamma_z$  es la matriz de covarianzas entre los componentes del vector en el mismo instante. Como los componentes son estacionarios, la matriz de covarianzas entre ellos es también constante. Análogamente, se comprueba que

$$\text{cov}(y_t y_{t+k}) = \mathbf{c}'\Gamma_z(k)\mathbf{c} = \gamma_k$$

donde  $\Gamma_z(k)$  contiene las covarianzas entre los componentes en distintos instantes que, por hipótesis, sólo dependen del retardo. Por tanto el proceso  $y_t$  es estacionario.

Una consecuencia de este resultado es que las autocovarianzas de un proceso estacionario no pueden ser cualesquiera y deben verificar que la matriz de autocovarianzas de orden  $k$  del proceso definida por

$$\Gamma_k = \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{k-2} \\ \cdot & \cdot & \dots & \cdot \\ \gamma_{k-1} & \gamma_{k-2} & \dots & \gamma_0 \end{bmatrix}$$

sea definida no negativa, es decir si  $\mathbf{c}$  es un vector arbitrario debe verificarse que  $\mathbf{c}'\Gamma_k\mathbf{c} \geq 0$ . En efecto, si definimos el proceso

$$y_t = c_1 z_t + c_2 z_{t-1} + \dots + c_k z_{t-k}$$

su varianza es  $\mathbf{c}'\Gamma_k\mathbf{c}$ , que debe ser no negativa.

Una conclusión importante de este resultado es que para que una secuencia de números pueda ser la función de autocovarianzas de un proceso esta secuencia debe verificar que la matriz de autocovarianzas asociada de orden  $k$ , para  $k = 1, 2, 3, \dots$  debe ser definida no negativa. Si consideramos un proceso de varianza unidad las autocovarianzas coinciden con las autocorrelaciones por lo que la misma propiedad se aplica a la matriz de autocorrelaciones

$$R_k = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \cdot & \cdot & \dots & \cdot \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{bmatrix}$$

que debe ser también definida no negativa.

### 3.4 Proceso de ruido blanco

Un proceso estacionario muy importante es el definido por:

1.  $E[z_t] = 0$
2.  $\text{Var}(z_t) = \sigma^2$
3.  $\text{Cov}(z_t, z_{t-k}) = 0, \quad k = \pm 1, \pm 2, \dots$

Este proceso se denomina *proceso de ruido blanco*. Por ejemplo, si tiramos una moneda en instantes  $t = 1, 2, \dots$  y definimos  $z_t = -1$  si se obtiene cara y  $z_t = +1$  si se obtiene cruz, se obtiene un proceso de ruido: la esperanza es cero, la varianza es constante (igual a la unidad) y las covarianzas son nulas. Como segundo ejemplo, si apuntamos el número que se obtiene en la extracción de un bombo que contiene el mismo número de dígitos entre el cero y el nueve y consideramos la secuencia de valores  $z_t - 4.5$ , llamando  $z_t$  al valor

observado en la  $t$ -ésima extracción,  $t = 1, 2, \dots$  se obtiene un proceso de ruido. En estos procesos conocer los valores pasados no proporciona ninguna información sobre el futuro ya que el proceso «no tiene memoria».

Si suponemos que en un proceso de ruido todas las variables tienen distribución normal, la incorrelación garantiza la independencia y llamaremos al proceso resultante *proceso de ruido blanco normal o de variables independientes*.

### 3.5 Estimación de los momentos de procesos estacionarios

Supongamos un proceso estacionario con media  $\mu = E[z_t]$ , varianza  $\sigma^2 = \gamma_0 = \text{Var}(z_t)$  y covarianzas  $\gamma_k = \text{Cov}(z_t, z_{t-k})$  del que se observa una realización  $(z_1, \dots, z_T)$ . Vamos a estudiar como estimar la media, la varianza, las covarianzas y las autocorrelaciones del proceso a partir de esta única realización disponible.

#### 3.5.1 Estimación de la media

Un estimador centrado de la media poblacional es la media muestral. En efecto, llamando  $\bar{z}$  la media muestral:

$$\bar{z} = \frac{\sum_{t=1}^T z_t}{T},$$

se verifica que

$$E(\bar{z}) = \frac{\sum_{t=1}^T E(z_t)}{T} = \mu.$$

Para datos independientes la varianza de la media muestral como estimador de la media poblacional es  $\sigma^2/T$ . En consecuencia, al aumentar el tamaño muestral el error cuadrático medio de estimación dado por

$$E(\bar{z} - \mu)^2$$

y que como el estimado es centrado coincide con la varianza, tiende a cero. En un proceso estocástico estacionario esta propiedad no es necesariamente cierta, y es posible que al aumentar el tamaño muestral no disminuya constantemente la varianza de la estimación. Por ejemplo, supongamos un proceso que comienza en el instante  $t = 1$ , donde la primera variable  $z_1$ , es un valor al azar de una variable aleatoria de media cero y varianza  $\sigma^2$ . A continuación los valores posteriores del proceso son iguales al valor inicial de manera que  $z_t = z_{t-1} = \dots = z_1$ . En este proceso estacionario la esperanza de cada variable es cero, ya que en distintas realizaciones la variable inicial irá tomando distintos valores, pero si disponemos de una realización única del proceso siempre  $\bar{z} = z_1$  y la varianza de esta estimación es  $\sigma^2$ , sea cual sea el tamaño de la realización. En adelante supondremos que este tipo de situaciones no se da, de manera que las nuevas observaciones no tienen correlación uno con las anteriores y, en consecuencia, proporcionan información. En este caso decimos que el proceso es *ergódico* para la estimación de la media y entonces se verifica que:

$$\lim_{T \rightarrow \infty} E(\bar{z} - \mu)^2 \rightarrow 0$$

Para identificar las condiciones para que un proceso sea ergódico para la estimación de la media, vamos a calcular el error cuadrático medio de la media muestral como estimador de la media poblacional para procesos estacionarios. Como el estimador  $\bar{z}$  es centrado para  $\mu$ , el error cuadrático medio coincide con la varianza del estimador y tenemos:

$$\text{var}(\bar{z}) = E(\bar{z} - \mu)^2 = \frac{1}{T^2} E\left(\sum_{t=1}^T (z_t - \mu)\right)^2$$

que podemos escribir como:

$$\text{var}(\bar{z}) = \frac{1}{T^2} \left[ \sum_{t=1}^T E(z_t - \mu)^2 + 2 \sum_{i=1}^T \sum_{j=i+1}^T E((z_i - \mu)(z_j - \mu)) \right]$$



El primer sumatorio dentro de los corchetes es  $T\sigma^2$ . El segundo sumatorio doble contiene  $T - 1$  veces las covarianzas de orden uno, es decir, entre  $z_t$  y  $z_{t+1}$ ,  $T - 2$  veces las covarianzas de orden 2 y, en general,  $T - i$  veces las de orden  $i$ . En consecuencia, la varianza de la media muestral puede escribirse como:

$$\text{var}(\bar{z}) = \frac{1}{T} \left[ \sigma^2 + 2 \sum_{i=1}^{T-1} \left(1 - \frac{i}{T}\right) \gamma_i \right] \quad (3.5)$$

Observemos en primer lugar que si todas las covarianzas son cero, es decir, tenemos un proceso de ruido blanco, la varianza de la media muestral es  $\sigma^2/T$ . Sin embargo, cuando las covarianzas no sean cero, la varianza de la media muestral para procesos estacionarios puede ser considerablemente mayor que para observaciones independientes. En efecto, si las  $\gamma_i$  son positivas el sumatorio puede ser muy grande. La condición para que la  $\text{var}(\bar{z})$  tienda a cero al aumentar  $T$  es que el sumatorio converja a una constante al aumentar  $T$ . Una condición necesaria (aunque no suficiente, véase el apéndice 3.1) de que la suma converja es:

$$\lim_{i \rightarrow \infty} \gamma_i \rightarrow 0$$

lo que supone que la dependencia entre observaciones tiende a cero al aumentar el retardo. Por ejemplo, una serie constante tiene correlación uno entre las observaciones y no es ergódica. Tampoco lo será una serie con componentes deterministas periódicos, como la serie

$$z_t = A \cos wt + a_t$$

donde  $A$  y  $w$  son constantes para toda la realización. Las covarianzas entre las observaciones nunca van a cero al aumentar el retardo, por el carácter periódico de la serie y aumentar el tamaño muestral no ayuda a mejorar la estimación de la media poblacional.

La ecuación (3.5) puede escribirse utilizando las autocorrelaciones como:

$$\text{var}(\bar{z}) = \frac{\gamma_0}{T} \left[ 1 + 2 \sum_{i=1}^{T-1} \left(1 - \frac{i}{T}\right) \rho_i \right]$$

donde  $\rho_i = \gamma_i/\gamma_0$  son las autocorrelaciones del proceso. Observemos que, como ocurría con las autocovarianzas, una condición suficiente para que el sumatorio converja y el proceso sea ergódico en la estimación de la media es que  $\lim_{i \rightarrow \infty} \rho_i \rightarrow 0$ .

En resumen, la propiedad de ergodicidad, que puede aplicarse a la estimación de cualquier parámetro, se satisface si las nuevas observaciones del proceso proporcionan información adicional sobre el parámetro, de manera que al aumentar el tamaño muestral disminuye contantemente su error de estimación. En la estimación de la media esto no ocurrirá si existe una dependencia tan fuerte que las nuevas observaciones son previsible desde el pasado y no proporcionan información adicional para estimar la media. La figura 3.5 presenta las autocorrelaciones de dos procesos estacionarios. El primero es ergódico, ya que las autocorrelaciones tienden a cero con el retardo, mientras que el segundo no lo es. En adelante supondremos que el proceso es ergódico, lo que supone en la práctica que hemos eliminado posibles términos sinusoidales deterministas de la forma  $A \cos(\omega t + \delta)$  que pudieran existir.

### 3.5.2 Estimación de las autocovarianzas y autocorrelaciones

Si la media del proceso fuese conocida, el estimador de las autocovarianzas de orden  $k$  es:

$$\tilde{\gamma}_k = \frac{1}{T-k} \sum_{t=k+1}^T (z_t - \mu)(z_{t-k} - \mu) \quad (3.6)$$

y es fácil comprobar al tomar esperanzas que este estimador es centrado para estimar  $\gamma_k = E((z_t - \mu)(z_{t-k} - \mu))$ . Sin embargo, cuando  $\mu$  es desconocida y la sustituimos por su estimador,  $\bar{z}$ , se demuestra que el estimador no es centrado. Un estimador alternativo de  $\gamma_k$  que tiene mejores propiedades cuando  $\mu$  es desconocida, es:

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^T (z_t - \bar{z})(z_{t-k} - \bar{z}) \quad (3.7)$$

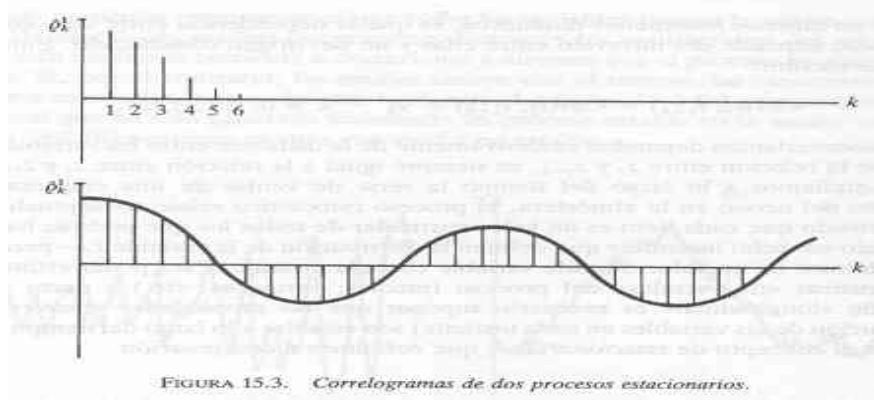


Figura 3.5: Correlograma de dos procesos estacionarios

que es también un estimador sesgado de la autocovarianza poblacional pero tiene menor error cuadrático de estimación que el anterior. En particular la varianza del proceso se estima mediante  $\hat{\gamma}_0$ .

Una ventaja adicional de utilizar el estimador (3.7) es que la matriz de autocovarianzas estimada:

$$\hat{\Gamma}_k = \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \dots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \dots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & \dots & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \dots & \hat{\gamma}_0 \end{bmatrix}$$

es siempre definida no negativa. Esto puede no ocurrir si en lugar de dividir por el tamaño muestral lo hacemos por el número de términos en la suma, como en el estimador (3.6). Esta es una propiedad importante porque en otro caso la secuencia estimada puede no corresponder a ningún proceso estacionario.

Las autocorrelaciones se estiman por

$$r_k = \hat{\gamma}_k / \hat{\gamma}_0$$

y el vector de autocorrelaciones estimadas,  $\mathbf{r} = (r_1, \dots, r_k)'$  se distribuye aproximadamente para  $T$  grande de forma normal con media  $\boldsymbol{\rho}$ , el vector de autocorrelaciones teóricas y matriz de covarianzas  $\mathbf{V}_\rho/T$  donde los términos de la matriz  $\mathbf{V}_\rho$  vienen dados por las fórmulas de Bartlett (véase Fuller, 1996 para una prueba de este resultado y el apéndice 3.2 de este capítulo para las expresiones). En particular, si sólo los primeros  $q$  coeficientes de autocorrelación son distintos de cero, las varianzas de las autocorrelaciones estimadas se aproximan por :

$$\text{var}(r_k) = \frac{T-k}{T(T+2)} \left(1 + 2 \sum_{j=1}^q \rho_j^2\right), \quad k > q \quad (3.8)$$

Por tanto, si todos las autocorrelaciones del proceso son nulas, la varianza asintótica de los coeficientes de autocorrelación estimados es

$$\text{var}(r_k) = \frac{T-k}{T(T+2)}$$

que puede aproximarse, para  $T$  grande, por  $1/T$ . En la práctica, con series reales cuando  $T$  no es muy grande podemos tener pocos términos para calcular la autocorrelación muestral de orden  $k$  y se recomienda que el orden máximo para el que se calculen autocorrelaciones muestrales sea  $k_{\max} \leq T/4$ .

En la figura 3.8 se han representado dos líneas paralelas al eje x a altura  $\pm 2/\sqrt{T}$ . Estas líneas proporcionan aproximadamente un intervalo del 95% donde deberían estar los coeficientes de autocorrelación estimados si la serie se ha generado como un proceso de ruido blanco.

Señalaremos por último que las estimaciones de los coeficientes de autocorrelación y de las covarianzas están a su vez correladas entre sí, y tanto más cuanto mayor sean los coeficientes teóricos que estamos estimando (véase el apéndice 3.2). Sólo cuando todos los coeficientes de autocorrelación teóricos son cero no existe correlación entre las estimaciones.

**Ejemplo 3.1**

Vamos a calcular la función de autocorrelación muestral para los datos de las leguas recorridas por Colón. Comenzaremos calculando la media de la serie:

$$\bar{z} = \frac{9 + 45 + \dots + 59 + 49}{34} = 31.83$$

y la varianza de los datos:

$$s_z^2 = \frac{(9 - 31.83)^2 + (45 - 31.83)^2 + \dots + (49 - 31.83)^2}{34} = 266.97$$

El coeficiente de autocorrelación de primer orden se calcula como

$$r_1 = \frac{[(9 - 31.83)(45 - 31.83) + \dots + (59 - 31.83)(49 - 31.83)] / 34}{266.97} = .52$$

y el de orden dos

$$r_2 = \frac{[(9 - 31.83)(60 - 31.83) + \dots + (31 - 31.83)(49 - 31.83)] / 34}{266.97} = -.002$$

y así sucesivamente. La figura 3.6 presenta la función de autocorrelación obtenida. El coeficiente de correlación de orden cero, de la variable con sí misma, es siempre uno y no se ha presentado en el gráfico que incluye las bandas de confianza a distancia  $2/\sqrt{T}$ . Únicamente el primer coeficiente de autocorrelación parece ser significativo, indicando que las leguas recorridas un día determinado dependen de las recorridas el día anterior, pero no de las de otros días previos.

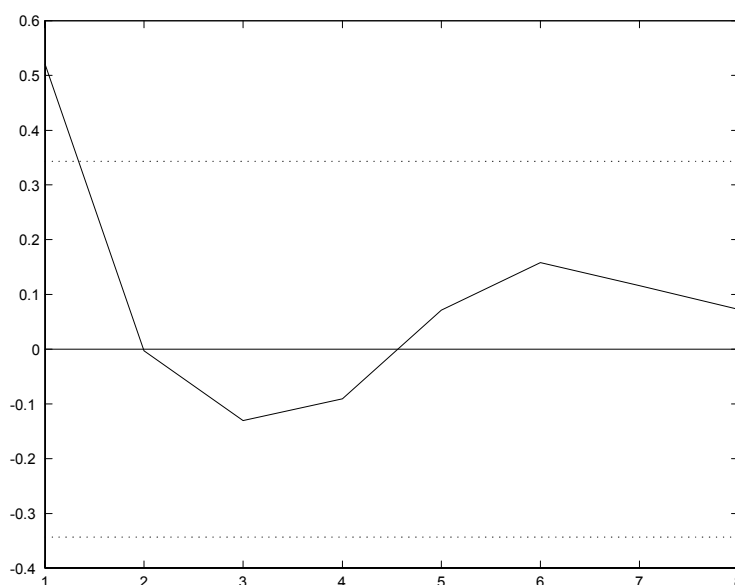


Figura 3.6: Función de autocorrelación para los datos de Colón.

**Ejemplo 3.2**

La figura 3.7 presenta una simulación de un proceso de ruido blanco y la figura 3.8 los coeficientes de autocorrelación muestrales para retardos mayores que cero de esta serie. Se han dibujado en el gráfico dos bandas de confianza del 95% para las autocorrelaciones estimadas a una distancia  $2/\sqrt{T}$ , que representan el intervalo de confianza aproximado al 95% donde deben estar los coeficientes estimados si la serie se ha generado por un proceso de ruido blanco. Observemos que un intervalo del 95% implica que una de cada 20

autocorrelaciones es esperable que salga fuera de las bandas en promedio cuando el proceso es verdaderamente ruido blanco, y eso es lo que ocurre en la figura 3.8, donde uno de los coeficientes calculados sale fuera de las bandas.

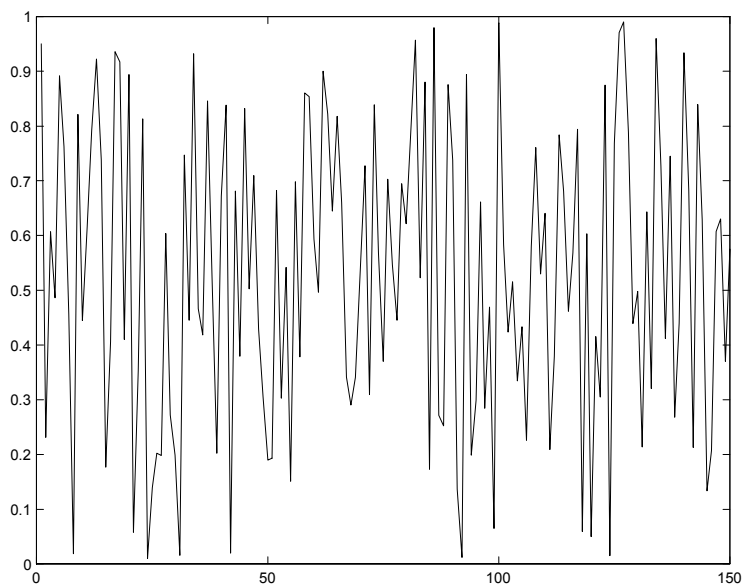


Figura 3.7: Simulación de 150 observaciones de un proceso de ruido blanco

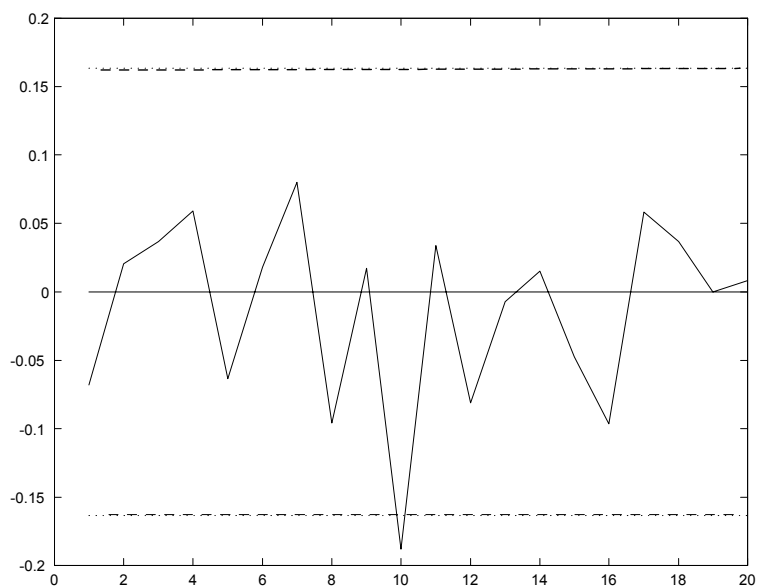


Figura 3.8: Función de autocorrelación para un proceso de ruido blanco

### Ejemplo 3.3

La figura 3.9 presenta la función de autocorrelación para la serie de rendimientos en la bolsa de Madrid. Se observa que todos los coeficientes de autocorrelación son pequeños, y están dentro de las bandas de  $2/\sqrt{T}$ . Esto sugiere que el proceso es ruido blanco, y los rendimientos pasados no nos dan información para prever

los rendimientos futuros. El lector puede pensar que la figura muestra una apariencia cíclica, pero si se generan 30 números aleatorios normales con un ordenador y realiza su gráfico comprobará que tienen una forma parecida.

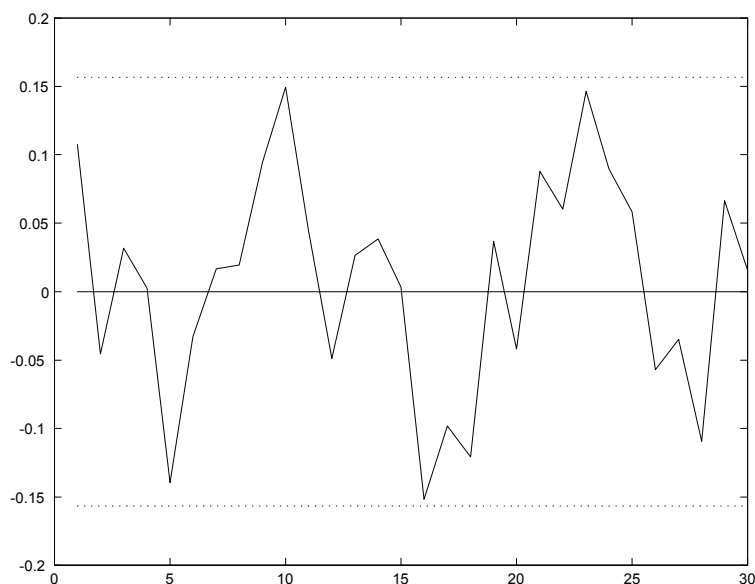


Figura 3.9: Función de autocorrelación de los datos de los rendimientos con el índice general de la bolsa de Madrid, 1988 a 2001.

### 3.6 El espectro de un proceso estacionario(\*)

Esta sección puede omitirse en primera lectura ya que no se utiliza en el resto del libro. Se introduce para facilitar al lector la comprensión de la literatura de series que utiliza el espectro. En el Capítulo 2 se introdujo el *periodograma*, como una función que nos informa sobre la importancia relativa, medida por la variabilidad explicada, de las posibles periodicidades en la serie temporal. Vimos que dada una serie temporal que se ha generado como suma de ondas armónicas deterministas, es posible decomponer su variabilidad en términos asociados a cada onda y que el periodograma para una frecuencia  $f_j$  era una medida de la varianza dedida al componente de esta frecuencia. Se demuestra en el apéndice 3.3 que una expresión alternativa del periodograma de una serie es:

$$I(f_j) = 2(\hat{\gamma}_0 + 2 \sum_{k=1}^{T-1} \hat{\gamma}_k \cos 2\pi f_j k) = 2\hat{\gamma}_0(1 + 2 \sum_{k=1}^{T-1} r_j \cos 2\pi f_j k)$$

que pone de manifiesto que el periodograma es una combinación lineal de las autocovarianzas o autocorrelaciones muestrales calculadas a partir de la serie observada. El periodograma es una forma alternativa de representar la estructura de dependencia lineal observada en una serie. Conocidas las autocovarianzas muestrales podemos obtener el periodograma según la fórmula anterior, y se demuestra que, análogamente, dado el periodograma  $I(f_j)$  podemos calcular las autocovarianzas de la serie.

Vimos también en el capítulo 2 que una forma simple de suavizar el periodograma es representarlo como un histograma, con longitudes de la base  $1/T$  y alturas de los rectángulos iguales a  $I(f_j)$ . Entonces el área encerrada por este periodograma suavizado es la varianza de la serie. Supongamos ahora que la longitud de la serie observada tiende a infinito. Entonces las propiedades de la serie se irán aproximando a las propiedades del proceso estocástico que ha generado la realización observada. Las frecuencias básicas  $1/2 \geq f_j \geq 1/T$  tenderán entonces a cubrir todo el intervalo  $1/2 \geq f \geq 0$ , y el periodograma suavizado tenderá a una curva

suave, que dependerá del proceso generador de la serie temporal. Si el proceso que ha generado los datos es estacionario, las autocovarianzas estimadas  $\hat{\gamma}_k$  tenderán a las teóricas del proceso estacionario,  $\gamma_k$ , y se define el espectro de este proceso mediante la expresión:

$$S(f) = 2(\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos 2\pi f k) \quad (3.9)$$

Esta función está definida entre  $0 \leq f \leq 1/2$  y existe para todo proceso estacionario y ergódico. Se demuestra en el apéndice 3.3 que el espectro es una transformación matemática de la función de autocovarianzas del proceso y que ambas recogen la misma información. Es decir, conocidas las autocovarianzas puede calcularse el espectro del proceso y al contrario. Los análisis basados en las covarianzas suelen denominarse *análisis en el dominio temporales*, ya que las covarianzas definen la relación dinámica temporal entre las observaciones del proceso. Los análisis basados en el espectro se suelen denominar *análisis en el dominio frecuencial*, ya el espectro tiene en cuenta la contribución de los distintos armónicos a la variabilidad de la serie. La utilidad del espectro es que, como ocurre con el periodograma, podemos detectar los ciclos principales que forman el proceso. La relación entre el periodograma y el espectro es similar a la de histograma y función de densidad, la primera corresponde a una muestra y la segunda a la población. Por analogía con el periodograma, el área bajo el espectro en un intervalo de frecuencias representa la varianza que explican las funciones sinusoidales con frecuencia en ese intervalo. Observemos que un proceso de variables incorreladas tiene  $\gamma_k = 0$  para  $k \geq 1$ , con lo que su espectro será constante, y dado por

$$S(f) = 2\sigma^2$$

siendo  $\sigma^2$  la varianza del proceso. Esto justifica la denominación de ruido blanco para estos procesos, por analogía con el espectro físico de la luz blanca que es constante. (En física el espectro representa la intensidad de los componentes de una señal en función de la longitud de onda, y es de allí donde proviene la representación estadística). Se demuestra que el área bajo el espectro proporciona siempre la varianza del proceso estacionario:

$$\gamma_0 = \int_0^{1/2} S(f) df$$

resultado que es esperable por la intuición del espectro como límite del periodograma suavizado.

Se define la función de densidad espectral como:

$$s(f) = \frac{S(f)}{\gamma_0} = 2(1 + 2 \sum_{k=1}^{\infty} \rho_k \cos 2\pi f k).$$

El nombre de densidad proviene de que tiene las mismas propiedades que una función de densidad de probabilidad: es no negativa,  $s(f) \geq 0$ , e integra a uno en su dominio de definición  $(0, 1/2)$ . Representa la contribución relativa de los armónicos en cada intervalo a la variabilidad total del proceso.

El análisis temporal suele ser más útil en series económicas, sociales o biológicas, donde no son esperables ciclos deterministas importantes. El análisis frecuencial se ha utilizado mucho en ingeniería eléctrica y de telecomunicaciones, cuando las series de interés están formadas por combinaciones lineales de señales periódicas contaminadas con un ruido aditivo. El espectro del proceso permite entonces, como vimos con el periodograma, obtener la importancia relativa de los distintos armónicos. El espectro tiene además un interés teórico, ya que puede demostrarse que podemos construir una representación de cualquier proceso estacionario como una suma infinita de funciones sinusoidales para todas las frecuencias en el intervalo  $[0, 1/2]$  con amplitudes que son variables aleatorias independientes. Esta caracterización de los procesos estacionarios permite que muchas propiedades de estos procesos pueden estudiarse mejor en el dominio frecuencial que en el temporal.

### Ejercicios

1. El gráfico adjunto indica la lluvia mensual caída en Santiago de Compostela en los años 1988-1997. Justificar las hipótesis necesarias para considerar que estas diez series son realizaciones de un proceso estocástico y comentar las propiedades del proceso. ¿Es estacionario?.

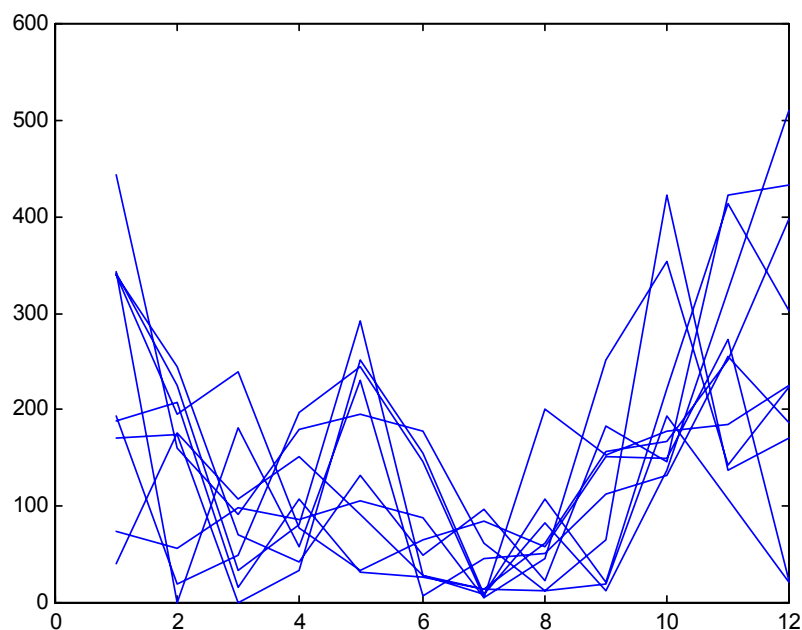


Figura 3.10: Precipitación mensual en Santiago de Compostela en el periodo 1988-1997.

3.2 Consideremos el proceso  $z_t = .5z_{t-1} + a_t$  donde  $a_t$  es un proceso de ruido. Calcular su media y sus autocovarianzas ¿Es estacionario?

3.3 Consideremos el proceso  $z_t = .5a_{t-1} + a_t$ . Calcular su media y sus autocovarianzas ¿Es estacionario?

3.4 Dada la serie estacionaria,  $z_1, \dots, z_T$ , demostrar que si las autocovarianzas son todas positivas la media del proceso se estimará con mayor varianza que en el caso en que todas las autocovarianzas son nulas.

3.5. Supongamos un proceso donde sólo la autocorrelación de primer orden es distinta de cero y su valor es negativo. Demostrar utilizando los resultados del capítulo que para que la varianza del proceso sea positiva el mínimo valor de la autocorrelación es -.5. Comprobarlo como caso particular con el proceso del ejercicio 3.4.

3.6 Demostrar que en el caso anterior la media del proceso se estima con menor variabilidad que con datos independientes

3.7 Justificar utilizando los resultados del capítulo que si un proceso es de variables independientes la varianza de las autocorrelaciones es  $1/T$ .

### Apéndice 3.1 Condición de convergencia

Dada una secuencia de números  $a_i$  una condición necesaria y suficiente para que la serie  $\sum a_i$  converja es que para todo  $\varepsilon > 0$  exista un  $N(\varepsilon)$  para el cual

$$|a_{n+1} + a_{n+2} + \dots + a_m| < \varepsilon, \quad \text{si } m > n > N(\varepsilon)$$

Se comprueba que si  $\sum a_i$  converge entonces siempre lo hace  $\sum |a_i|$ .

### Apéndice 3.2. La formulas de Barlett para las autocovarianzas estimadas

La varianza asintótica del coeficiente de autorrelación de orden  $k$  es

$$\text{var}(r_k) = \frac{1}{T} \sum_{-\infty}^{\infty} (\rho_i^2 + \rho_{i-k}\rho_{i+k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_i^2\rho_k^2)$$

mientras que las covarianzas vienen dadas por

$$\text{cov}(r_k, r_{k+h}) = \frac{1}{T} \sum_{-\infty}^{\infty} (\rho_i \rho_{i-h} + \rho_{i-k} \rho_{i+k+h} - 2\rho_{k+h} \rho_i \rho_{i-h} - 2\rho_k \rho_i \rho_{i-k-h} + 2\rho_i^2 \rho_k \rho_{k+h})$$

Si suponemos que sólo las primeras  $q$  autocorrelaciones son cero, entonces para  $k > h$  las correlaciones entre los coeficientes estimados son:

$$\text{corr}(r_k, r_{k+h}) = \frac{2 \sum_{i=0}^{\infty} \rho_i \rho_{i+k}}{1 + 2 \sum_{j=1}^q \rho_j^2}$$

y vemos que sólo cuando todos los coeficientes de autocorrelación teóricos son cero no existe correlación entre las estimaciones.

### Apendice 3.3 El espectro y las autocovarianzas

Supongamos, sin pérdida de generalidad, una serie temporal de media muestra cero para simplificar la exposición. Vimos en el Capítulo 2 que el periodograma viene dado por

$$I(f_j) = \frac{T \widehat{R}_j^2}{2} = \frac{T}{2} \left[ \left( \frac{2}{T} \sum_{t=1}^T z_t \text{sen}(wt) \right)^2 + \left( \frac{2}{T} \sum_{t=1}^T z_t \cos(wt) \right)^2 \right]$$

que puede escribirse

$$I(f_j) = \frac{2}{T} \left[ \sum_{t=1}^T \sum_{v=1}^T z_t \text{sen}(wt) z_v \text{sen}(wv) + \sum_{t=1}^T \sum_{v=1}^T z_t \cos(wt) z_v \cos(wv) \right]$$

y utilizando que  $\cos(a-b) = \cos(a)\cos(b) + \text{sen}(a)\text{sen}(b)$ , podemos escribir

$$I(f_j) = \frac{2}{T} \sum_{t=1}^T \sum_{v=1}^T z_t z_v \cos(w(t-v)) = 2 \sum_{t=1}^T \sum_{v=1}^T \left( \frac{1}{T} z_t z_v \right) \cos(w(t-v)).$$

En esta suma vemos que aparecerán las sumas de términos  $(z_t z_v / T)$ , que son las autocovarianzas muestrales. Vamos a ver los coeficientes con los que aparecerán las autocovarianzas de distintos órdenes. La covarianza  $\widehat{\gamma}_0$  vendrá dada por los términos con  $t = v$ , entonces  $\cos(w(t-v)) = 1$ . Para  $\widehat{\gamma}_1$  necesitamos  $v = t-1$  o también  $v = t+1$ , por lo que tendremos dos veces la suma  $\sum z_t z_{t-1} / T$  y el coeficiente del coseno será  $\cos w$ . Análogamente, para  $\widehat{\gamma}_k$  tendremos dos veces la suma  $\sum z_t z_{t-k} / T$  y el coeficiente del coseno será  $\cos wk$ , y podemos escribir

$$I(f_j) = 2(\widehat{\gamma}_0 + 2 \sum_{k=1}^T \widehat{\gamma}_k \cos 2\pi f k)$$

Cuando  $T \rightarrow \infty$ , si suponemos que la serie es una realización de un proceso estocástico estacionario, las autocovarianzas muestrales convergen a las del proceso y tendremos la expresión límite del espectro

$$S(f) = 2(\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos 2\pi f k) \quad (3.10)$$

Una expresión alternativa del espectro es

$$S(f) = 2 \sum_{k=-\infty}^{\infty} \gamma_k e^{-i2\pi f k},$$

ya que  $\gamma_k = \gamma_{-k}$ ,  $\cos(x) = \cos(-x)$  y  $\text{sen}(x) = -\text{sen}(-x)$ , al desarrollar  $e^{-i2\pi f k}$  como  $\cos 2\pi f k + i \text{sen} 2\pi f k$ , la suma de los senos se anulará y la suma de los cosenos será igual para  $k$  negativo y positivo, con lo que resulta 3.10.



El espectro se expresa a veces en función de la frecuencia angular  $w = 2\pi f$ , y, entonces, en lugar de representarse en el intervalo  $(0, .5)$  se representa en el intervalo  $(0, \pi)$ . Para que se mantenga la propiedad de que la varianza del proceso es igual al area bajo el espectro, entonces el espectro suele definirse como

$$S(w) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i w k},$$

Algunos autores prefieren definir el espectro en el intervalo  $(-\pi, \pi)$ , con  $S^*(-w) = S^*(w)$ , su expresión entonces es:

$$S^*(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i w k}$$

Estas expresiones ponen de manifiesto que, matemáticamente, el espectro es la transformada de Fourier de la función de autocovarianzas. Por ejemplo, utilizando la representación con frecuencias angulares,  $S(w)$ , Puede demostrarse que existe la transformación inversa, dada por

$$\gamma_k = \int_{-\pi}^{\pi} e^{i w k} S(w) dw$$

En resumen, el espectro y la función de covarianzas contienen la misma información y podemos pasar de una a la otra mediante las ecuaciones anteriores.