Análisis Estadístico de Datos Climáticos

SERIES TEMPORALES 2

2015

Contenido

- Procesos estacionarios y débilmente estacionarios

- Algunos procesos estocásticos útiles:

  Procesos puramente aleatorios (ruido blanco
  Caminatas al azar
  Procesos auto-regresivos

Media, varianza y autocovarianza de procesos (o modelos) estocásticos

Partimos de un proceso estocástico: un conjunto de variables aleatorias ordenadas en el tiempo y definidas en un conjunto de instantes, que pueden ser discretos o continuos.

Para el caso discreto, anotamos X_t , para referir a la variable aleatoria en el instante t. ($t= 0, \pm 1, \pm 2, \ldots$).

Una serie real es una realización particular del proceso estocástico. La anotamos x_t con $t = 1, 2, \ldots, N$, para el caso discreto.

Definimos (o recordamos) algunos conceptos útiles:

Media: $\mu(t) = E(X_t)$ para todo t.

Varianza: $\sigma^2(t) = Var(X_t)$ para todo t.

Función de autocovarianza (f. acv): es la covarianza de X_{t1} y X_{t2}, o sea:

$$\gamma(t_1, t_2) = E[(X_{t1} - \mu(t_1))(X_{t2} - \mu(t_2))] \text{ para todo } t_1 \text{ y } t_2.$$

(La varianza es un caso particular de la covarianza, para $t_1 = t_2$.)

Procesos (o modelos) estacionarios

Un proceso es estrictamente estacionario si la distribución conjunta de $X_{t1},\ldots.X_{tk}$ es la misma que la distribución conjunta de $X_{t1+\tau}$, $X_{tk+\tau}$, para todo $t_1,\ldots.t_k$, τ y k.

En otras palabras, trasladar el origen de tiempos en una magnitud τ no tiene efecto en las distribuciones conjuntas, las cuales dependerán solamente de los intervalos entre t_1, $t_2,\ldots.t_k$.

En particular, para k=1, la estacionariedad estricta implica que la distribución de X_t es la misma para todo t, por lo que, se cumplirá que:

$\mu(t) = \mu$
$\sigma^2(t) = \sigma^2$
para todo t (si existen).

Además, si k=2 la distribución conjunta de X_{t1} y X_{t2} depende sólo de la diferencia $(t_2 - t_1) = \tau$, que se llama lag.

Procesos (o modelos) estacionarios

Entonces, la f. acv $\gamma(t_1, t_2)$ también depende sólo de $(t_2 - t_1)$ y se puede escribir como $\gamma(\tau)$, siendo

$$\gamma(\tau) = E[(X_t - \mu)(X_{t+\tau} - \mu)] = Cov(X_t, X_{t+\tau}),$$

que recibe el nombre de coeficiente de autocovarianza en el lag τ.

Notar que $\gamma(0) = Var(X_t) = \sigma^2$

La f. acv se estandariza para obtener la función de autocorrelación (adimensionada) , que se define como:

$$\rho(\tau) = \gamma(\tau) / \gamma(0) = \gamma(\tau) / \sigma^2$$

$\rho(\tau)$ es la correlación entre X_t y $X_{t+\tau}$.
(La función de autocorrelación muestral, para $\tau = k$, es la r_k que ya hemos definimos antes.)
En lo que sigue, para los procesos discretos escribiremos $\rho(k)$ y $\gamma(k)$.

Procesos débilmente estacionarios
(o estacionarios de segundo orden)

En la práctica es útil definir un concepto de estacionariedad menos estricto.

Un proceso es débilmente estacionario si su media es constante y su f. acv depende sólo del lag, o sea:

$$E(X_t) = \mu$$

y

$$Cov(X_t, X_{t+\tau}) = \gamma(\tau)$$

(Tomando $\tau = 0$, tenemos que la varianza también es constante para un proceso débilmente estacionario.)

Un caso particular importante es cuando el proceso es gausiano, o sea cuando la distribución conjunta de $X_{t1},\ldots.X_{tk}$ es normal multivariada para todos los $t_1,\ldots.t_k$, ya que esta última queda caracterizada completamente por sus primer y segundo momentos. Entonces, para estos procesos, la estacionariedad débil implica la estacionariedad estricta.

Algunas propiedades de la acv y la ac

Sea un proceso estocástico estacionario X_t con media μ, varianza σ^2, función de autocovarianza $\gamma(\tau)$ y función de autocorrelación $\rho(\tau)$.

- **Ya vimos que** $\rho(\tau) = \gamma(\tau)/\gamma(0) = \gamma(\tau)/\sigma^2$
  Notar que $\rho(0) = 1$

- $\gamma(\tau) = \gamma(-\tau)$
  $\rho(\tau) = \rho(-\tau)$

- $| \rho(\tau) | \leq 1$

Algunos procesos (o modelos) útiles

1. Procesos puramente aleatorios ("ruido blanco")

Se llama así aun proceso en que los términos $\{Z_t\}$ son variables aleatorias mutuamente independientes e idénticamente distribuidos (iid). Habitualmente se supone además que las variables aleatorias tienen distribución normal con media cero y varianza σ_Z^2, que serán constantes para todo t.

La suposición de independencia implica que:

$$\gamma(k) = \mathrm{Cov}(Z_t, Z_{t+k}) = \sigma_Z^2 \quad \text{si} \quad k=0$$

$$= 0 \quad \text{si} \quad k = \pm 1, \pm 2, \ldots.$$

lo cual implica que:

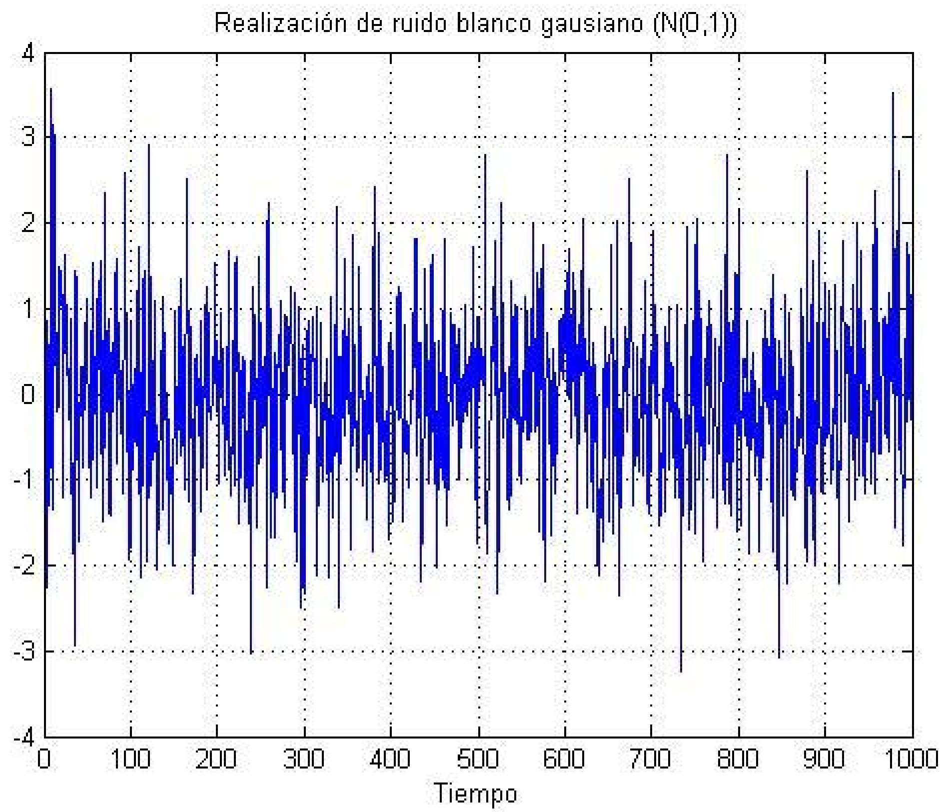
$$\rho(k) = 1 \text{ si } k = 0$$

$$= 0 \text{ si } k = \pm 1, \pm 2, \ldots.$$

Este proceso es, por definición, estrictamente estacionario.

Ejemplo: una realización de ruido blanco gausiano (N(0,1))

```
Y=randn(1,1000);
plot(Y), grid
```



El ruido blanco aparece como un componente en varios otros procesos estocásticos.

2. Caminatas al azar ("random walks")

Si $\{Z_t\}$ es un proceso puramente aleatorio con media μ y varianza σ^2, se dice que el proceso X_t es una caminata al azar si

$$X_t = X_{t-1} + Z_t$$

Suponiendo que se comienza con $X_1 = Z_1$, tenemos que

$$X_t = \sum_{i=1}^{t} Z_i$$

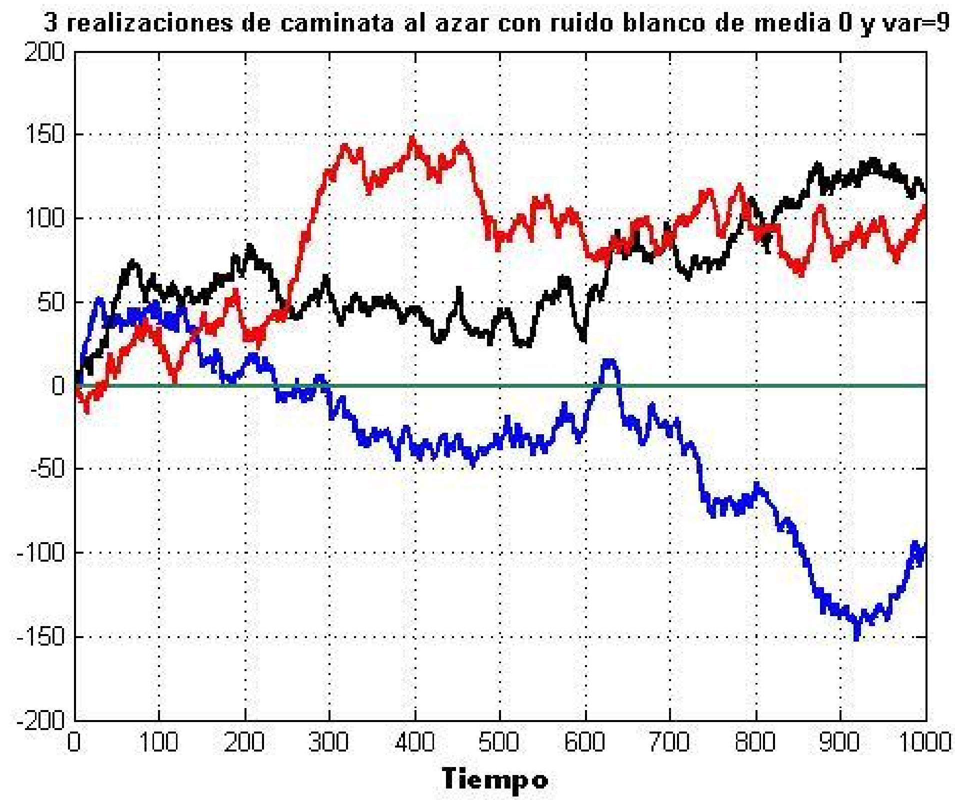
Se tiene que: $E(X_t) = t\mu$ **y** $Var(X_t) = t\sigma_Z^2$

Por lo tanto, el proceso no es estacionario.

(pero $X_t - X_{t-1} = Z_t$ es puramente aleatorio y por tanto es estacionario.)

Ejemplo: 3 realizaciones de una caminata al azar con ruido blanco gausiano de varianza = 9.

```
X=zeros(1,1000);
  for i=2:1000;
     X(i)=X(i-1)+3*randn;
   end
plot(X)
% ejecutada 3 veces
```



La media "no cambia" para valores grandes de t, pero la dispersión sí "crece" con t.

Aplicación a dispersión de contaminantes en la atmósfera (Von Storch – Zwiers, p. 202-203)

Una caminata al azar describe la trayectoria de una partícula que realiza desplazamientos aleatorios.

El movimiento de una partícula, p. ej. emitida por una chimenea o un volcán está determinada por el flujo determinístico U y muchos pequeños desplazamientos impredecibles.

Si la partícula en t está en la ubicación R(t), en t+1 estará en

$R(t+1) = R(t) + U + Z_t$ (donde Z_t es ruido blanco con media 0),

y en el instante t+k estará en:

$$R(t+k) = R(t) + kU + \sum_{i=t}^{i=t+k-1} Z_i$$

La aplicación a muchas partículas, puede permitir modelar su dispersión tridimensional.

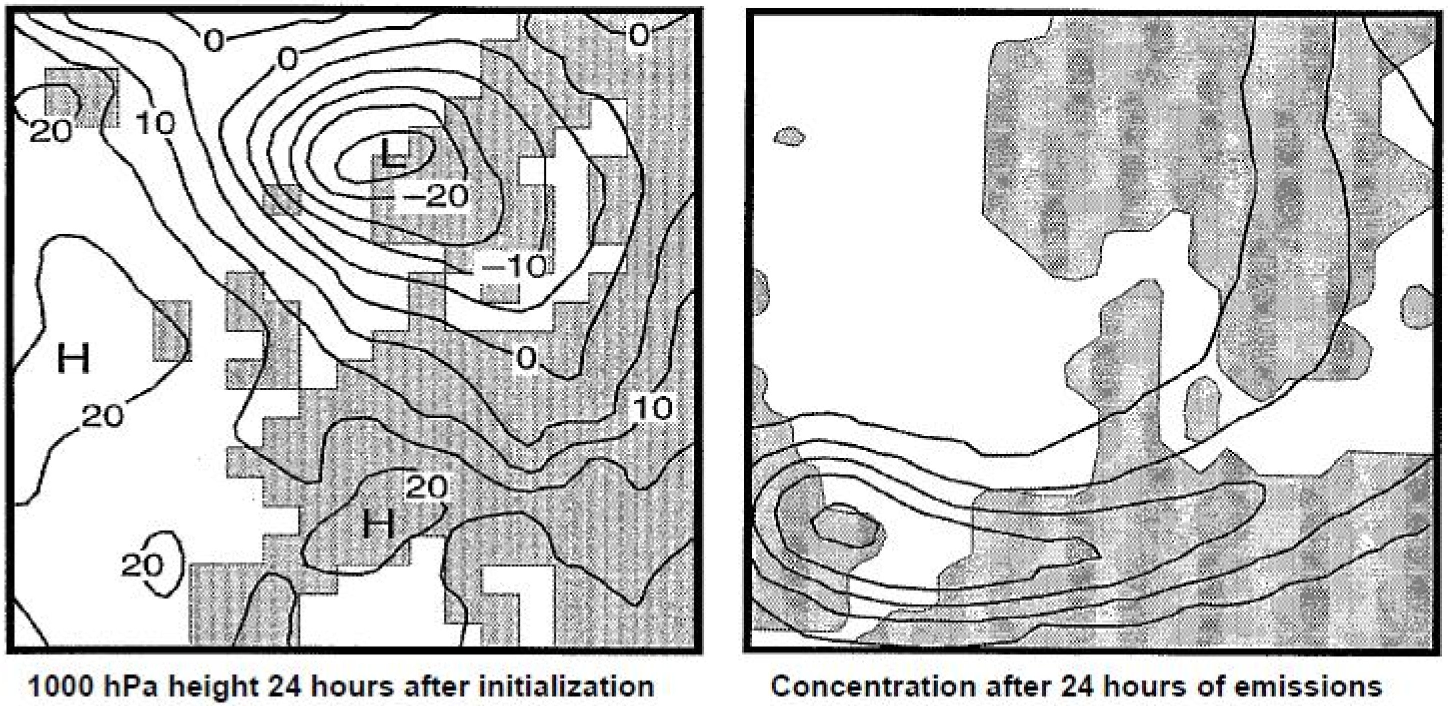


Figure 10.6: *Example of a simulation of long-range transport of air pollutants.*
Left: Simulated 1000 hPa height field 24 hours after model initialization.
Right: Distribution of pollutant continuously emitted in east England after 24 hours.
From Lehmhaus et al. [250].

A la izquierda, un pronóstico de alturas de geopotencial en 1000 hPa para Europa occidental.
Se "inyecta" SO_2 en la atmósfera simulada en un punto del este de Inglaterra.
A la derecha, la concentración de SO_2 producida por el modelo estocástico a las 24 hs.

3. Procesos auto-regresivos (AR)

Si Z_t es un proceso puramente aleatorio con media 0 y varianza σ_Z^2 (ruido blanco), se dice que el proceso $\{X_t\}$ es un proceso auto-regresivo de orden p (AR(p)) si

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + Z_t$$

En cierto sentido, se puede interpretar como un modelo de regresión lineal múltiple, pero la regresión es respecto de valores pasados de X_t, y no respecto de variables predictoras independientes.

El AR de primer orden (p=1),

$$X_t = \alpha X_{t-1} + Z_t \quad (1)$$

se llama a veces proceso de Markov o también proceso de "ruido rojo".

Este proceso tiene interés porque hay muchas variables climáticas, geofísicas e índices solares que presentan un espectro similar al de un proceso de ruido rojo (el espectro lo veremos más adelante).
(Para el caso particular α = 0, el proceso es de ruido blanco.)

Se puede demostrar que este proceso es estacionario si $|\alpha| < 1$.

En ese caso, $E(X_t)$ debe ser una constante para todo t, que se Halla tomando valores esperados en la ecuación (1), resultando $E(X_t) = 0$.

Si elevamos (1) al cuadrado y tomamos valores esperados, llegamos a que $\sigma_X^2 = \sigma_Z^2 / (1- \alpha^2)$

Si multiplicamos (1) por X_{t-k} y tomamos valores esperados, obtenemos: $\gamma(k) = \alpha\gamma(k-1)$;
luego usando que $\gamma(0) = \sigma_X^2$, llegamos a que

$$\gamma(k) = \alpha^k \sigma_X^2$$

y

$$\rho(k) = \alpha^k$$

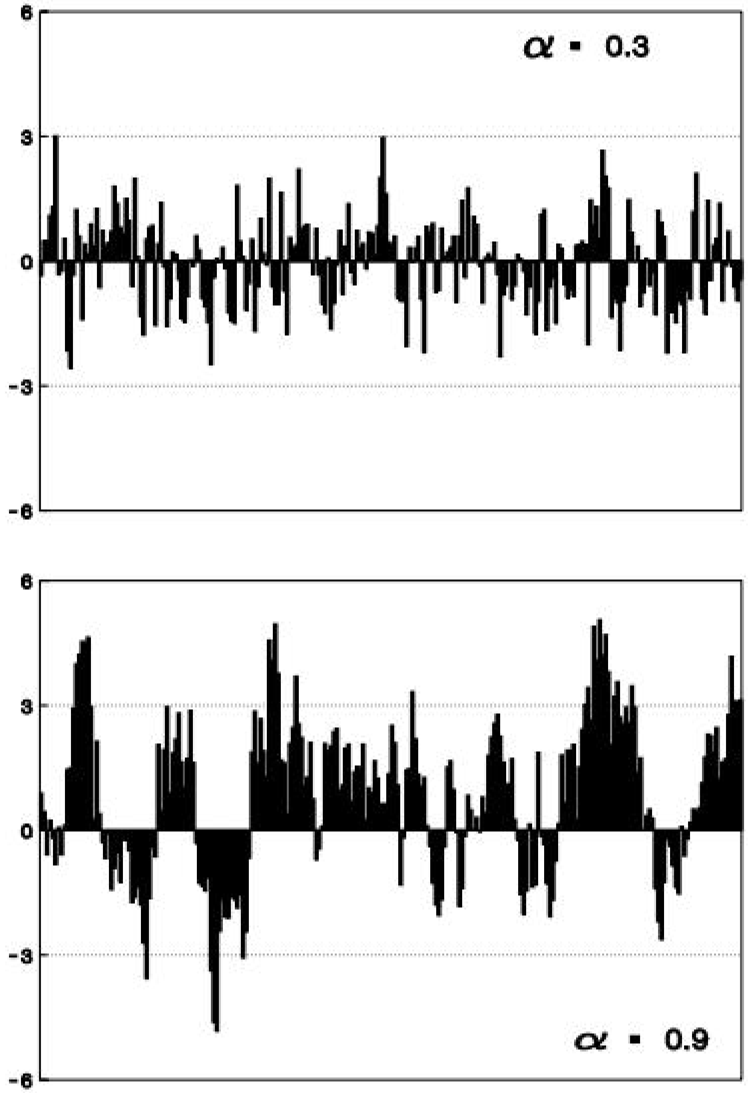


Figure 10.7: *240 time step realizations of AR(1) processes with $\alpha_1 = 0.3$ (top) and 0.9 (bottom). Both processes are forced by unit variance normally distributed white noise.*

2 realizaciones de AR(1), con $\alpha = 0.3$ y $\alpha = 0.9$.

Observar que la segunda tiene mucho más persistencia que la primera, que se parece más a la realización de un ruido blanco.

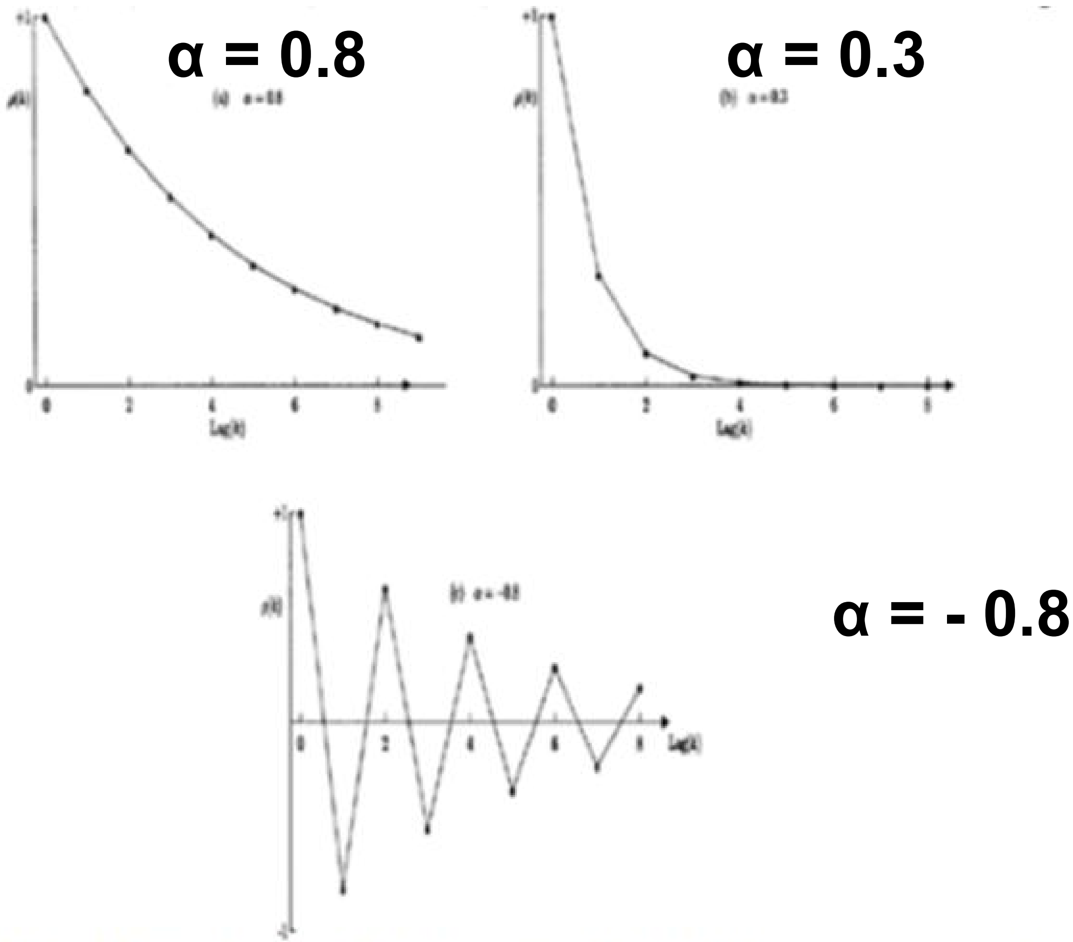


Figure 3.1 Three examples of the autocorrelation function of a first-order autoregressive process with, (a) a=0.8; (b) a=0.3; (c) a=−0.8.

Correlogramas de AR(1)

Para los AR(p) existen condiciones entre los coeficientes para que exista estacionariedad