

Procesos Estocásticos y Series Temporales

Francisco Javier Mercader Martínez

Índice

1	Introducción a los procesos estocásticos	4
1.1	Concepto de proceso estocástico, trayectorias y ejemplos	4
1.2	Funciones asociadas a un proceso estocástico	6
1.3	Procesos estacionarios y débilmente estacionarios	8
1.4	Procesos gaussianos	9
1.5	Ejemplos y simulación	10
1.5.1	Ruido blanco gaussiano	10
1.5.2	Movimiento Browniano	10
1.6	Procesos gaussianos estacionarios isotrópicos	13
2	Cadenas de Markov	20
2.1	Cadenas de Markov de tiempo discreto	20
2.2	Matriz de transición de n pasos	23
2.3	Clasificación de los estados	24
2.4	Número de visitas	28
2.5	Probabilidades de absorción y tiempos medios de llegada	28
2.5.1	Probabilidades de absorción	29
2.6	Comportamiento asintótico	29
2.7	Algoritmo PageRank	33
3	Procesos de Poisson	40
3.1	Procesos de conteo	40
3.2	Proceso de Poisson	41
3.2.1	Distribución de Poisson	41
3.2.2	Definición de proceso de Poisson	42
3.3	Tiempos de ocurrencia (llegada) en un Proceso de Poisson	43
3.3.1	Distribución exponencial	44
3.3.2	Modelización de tiempos de llegada y entre llegadas consecutivas	44
3.3.3	Simulación de procesos de Poisson	46
3.4	Transformadas de procesos de Poisson	47
3.4.1	Superposición de procesos de Poisson independientes	47
3.4.2	Separación de procesos de Poisson	49
3.5	Procesos de renovación	51
4	Conceptos básicos en series temporales	57
4.1	Introducción	57
4.2	Definición de serie temporal. Gráficos.	58
4.3	Medidas para evaluación de predicciones	60

4.4	Componentes de una serie temporal y esquemas de integración	61
4.5	Diferenciación de una serie temporal y medias móviles	63
4.6	Descomposición clásica de series temporales	67
4.6.1	Método de la diferencia a la media móvil	67
4.6.2	Método de la razón a la media móvil	69
4.7	Predicciones con Análisis Clásico y limitaciones	71
4.8	La descomposición STL	73
5	Métodos de Alisado Exponencial	82
5.1	Introducción	82
5.2	Alisado Exponencial Simple	82
5.3	Alisado Exponencial Doble o método de Brown	85
5.4	Método de Holt	87
5.5	Método de Holt-Winters multiplicativo	88
5.6	Método de Holt-Winters aditivo	90
5.7	Interpretación de los parámetros de alisado	92
5.8	Taxonomía de los métodos de alisado exponencial	92
5.9	Los modelos ETS	94
5.9.1	ETS(A,N,N): Modelo de espacio de estados del alisado exponencial simple con errores aditivos	94
5.9.2	ETS(M,N,N): Modelo de espacio de estados del alisado exponencial simple con errores multiplicativos	96
5.9.3	Otros modelos ETS	98
5.9.4	Estimación, selección del modelo ETS y predicciones	99
6	Modelos ARIMA	106
6.1	Introducción	106
6.1.1	Procesos débilmente estacionarios. El correlograma.	107
6.1.2	Procesos lineales	109
6.2	Proceso de ruido blanco gaussiano	110
6.3	Modelos Autorregresivos de orden p , $AR(p)$	110
6.3.1	Determinación del orden de la autorregresión	111
6.3.2	Estimación de los parámetros del modelo	113
6.3.3	Ejemplo de modelo $AR(1)$	115
6.4	Modelos de medias móviles, $MA(q)$	116
6.4.1	Determinación del orden del modelo	116
6.4.2	Estimación de los parámetros del modelo	117
6.4.3	Ejemplo de modelo $MA(1)$	117
6.5	Modelos $ARMA(p, q)$	118
6.5.1	Estimación de los parámetros del modelo	119
6.6	Procesos lineales no estacionarios: modelo ARIMA	119
6.6.1	Paseo aleatorio	119
6.6.2	Procesos ARIMA	119
6.7	Identificación del modelo, validación y predicciones	120
6.7.1	Análisis de la estacionariedad de la serie	120
6.7.2	Determinación del orden de la parte AR y de la parte MA	121
6.7.3	Estimación de los coeficientes del modelo	121
6.7.4	Validación del modelo: análisis de los residuos	122
6.7.5	Bondad del ajuste y selección del mejor modelo	122
6.7.6	Predicciones	123
6.7.7	Un ejemplo de modelo $ARMA$ con \mathbf{R}	124
6.8	Modelos ARIMA estacionales, SARIMA	127

7	Modelos de series temporales con variables exógenas	131
7.1	Introducción	131
7.2	El modelo de Regresión Múltiple para series temporales	131
7.3	Algunos predictores útiles	132
7.4	Consideraciones sobre el mundo y sobre las predicciones	134
7.4.1	Regresión espuria	134
7.4.2	Predicciones ex-ante y ex-post. Intervalos de predicción	135
7.5	Ejemplo	136
7.6	Modelos de regresión dinámica	142

Tema 1: Introducción a los procesos estocásticos

Los procesos estocásticos modelizan cantidades numéricas que cambian con el tiempo de manera aleatoria. Ejemplos del mundo real de estos procesos incluyen:

1. Resultados sucesivos en un juego de azar.
2. Número de plazas ocupadas en un parking.
3. Porcentaje de cielo cubierto en el cielo de Madrid.
4. Evolución del precio de activos financieros: acciones, tipos de cambio de divisas o criptomonedas, materias primas, etc.
5. Indicadores económicos como la inflación, el precio de la luz, y el IBEX 35.

Los procesos estocásticos proporcionan marcos matemáticos para modelar y comprender estos fenómenos, lo que permite hacer predicciones y tomar decisiones informadas.

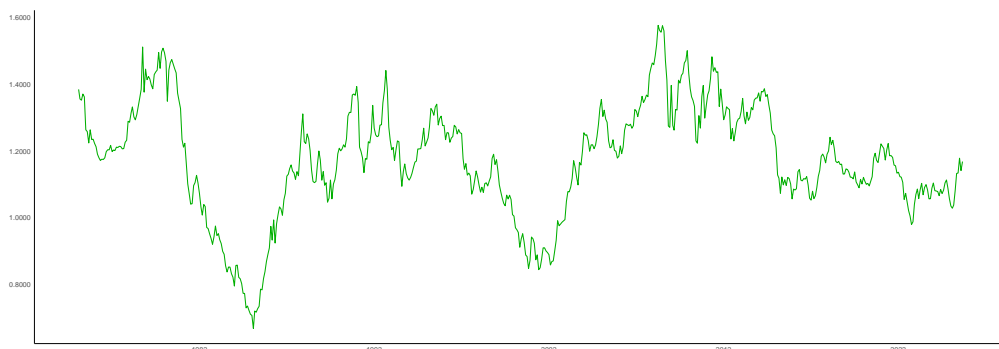


Figure 1: Evolución del tipo de cambio Euro/dólar (Fuente: Google Finance)

1.1) Concepto de proceso estocástico, trayectorias y ejemplos

A lo largo de este tema, vamos a fijar un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y un subconjunto $\mathbb{T} \subset [0, \infty)$.

Definición 1.1.1 Un **proceso estocástico** $(X_t)_{t \in \mathbb{T}}$ es una colección de variables aleatorias reales X_t definidas en el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.

Interpretamos t como el tiempo (medido en cierta unidad).

- Si \mathbb{T} es contable (por ejemplo, $\mathbb{T} = \{0 < 1 < 2 < \dots\}$), diremos que el proceso estocástico $(X_t)_{t \in \mathbb{T}}$ es de tiempo discreto.
- Si \mathbb{T} es un intervalo (por ejemplo, $\mathbb{T} = [0, T]$ o $\mathbb{T} = [0, \infty)$), diremos que el proceso estocástico es de tiempo continuo.

Para cada $t \in \mathbb{T}$ tenemos una variable aleatoria X_t . La variable aleatoria X_t tomará un valor numérico $X_t(\omega)$ para cada $\omega \in \Omega$. A los posibles valores que toma un proceso estocástico se les llama **estados**.

- Si para cada $t \in \mathbb{T}$ la variable aleatoria X_t es de tipo discreto, diremos que el proceso estocástico $(X_t)_{t \in \mathbb{T}}$ es de estado discreto.
- Si para cada $t \in \mathbb{T}$ la variable aleatoria X_t es de tipo continuo, diremos que el proceso estocástico $(X_t)_{t \in \mathbb{T}}$ es de estado continuo.

	t Discreto	t Continuo
X Discreta	Proceso de estado discreto y tiempo discreto (Unidades prod. mensualmente de un producto)	Proceso de estado discreto y tiempo continuo (Unidades producidas hasta t)
X Continua	Proceso de estado continuo y tiempo discreto (Toneladas de producción diaria de un producto)	Proceso de estado continuo y tiempo continuo (Velocidad de un vehículo en el instante t)

Figure 2: Tipos de procesos estocásticos

En lo que sigue, siempre supondremos que $0 \in \mathbb{T}$. Esta condición no es realmente restrictiva, ya que podemos desplazar el tiempo por una constante para garantizar que dicha condición se cumple.

Definición 1.1.2 Dado un proceso estocástico $(X_t)_{t \in \mathbb{T}}$, para cada realización $\omega \in \Omega$, la colección de números reales $(X_t(\omega))_{t \in \mathbb{T}}$ se llama **trayectoria** del proceso.

- Para un proceso estocástico $(X_t)_{t=0,1,2,\dots}$ de tiempo discreto, cada trayectoria define una sucesión de números reales $(x_t)_{t=0,1,2,\dots}$.
- Para un proceso estocástico $(X_t)_{t \in [0,\infty)}$ de tiempo continuo, cada trayectoria define una función real $t \mapsto x(t) : [0, \infty) \rightarrow \mathbb{R}$.

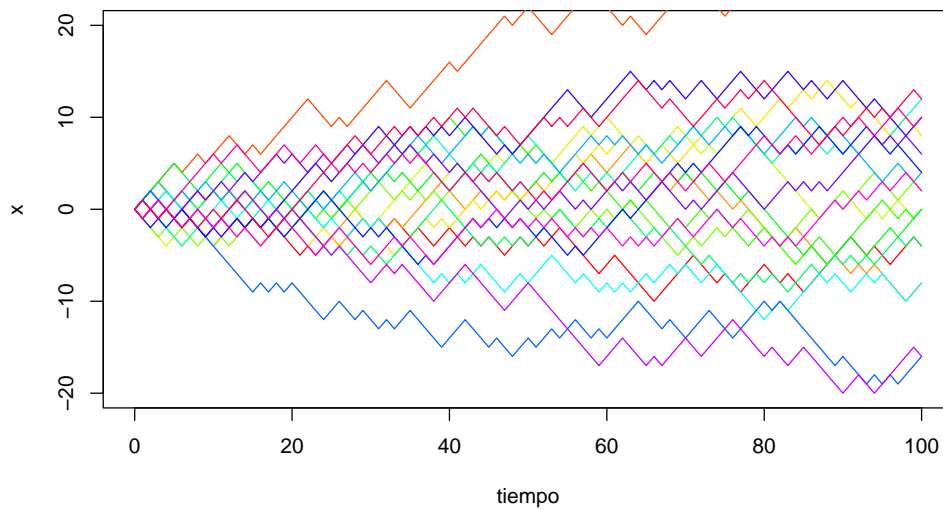
Cada trayectoria corresponde a una observación particular de la evolución del valor de un proceso estocástico en el tiempo.

Ejemplo. Supón que inviertes 100 euros en una cuenta bancaria con tipo de interés anual R , compuesto anualmente. Es decir, si X_t es la cantidad de dinero en el año t , entonces

$$X_t = 100 \cdot (1 + R)^t \text{ para } t = 0, 1, 2, \dots$$

Supongamos también que el valor de $R > 0$ es fijado cuando metes el dinero en la cuenta y sigue una distribución $\text{Exp}(1)$.

Para cada observación particular $R = r$ de la variable R tendremos una trayectoria del proceso estocástico $(X_t)_{t=0,1,2,\dots}$. Podemos simular y dibujar cinco trayectorias de este proceso estocástico mediante el siguiente código en R:



1.2) Funciones asociadas a un proceso estocástico

Dado un proceso estocástico $(X_t)_{t \in \mathbb{T}}$, cada variable aleatoria X_t tendrá su propia distribución de probabilidad la cual puede ser discreta o continua. Observar un proceso estocástico en un solo tiempo de forma aislada no es útil para describir su comportamiento como función del tiempo, ya que los valores observados en un tiempo pueden condicionar su comportamiento en tiempos distintos. Por ejemplo, en el caso del camino aleatorio simétrico en el que apostamos un euro a cara, si hemos acumulado muchas ganancias en el pasado, cual disminuirá a lo sumo en un euro en cada lanzamiento.

Por ese motivo, es necesario estudiar la distribución de probabilidad conjunta a lo largo de varios tiempos.

Definición 1.2.1 Supongamos que $(X_t)_{t \in \mathbb{T}}$ es un proceso estocástico. Dada una sucesión finita de tiempos $\{t_1 < t_2 < \dots < t_n\} \subset \mathbb{T}$, la función $F_{t_1, t_2, \dots, t_n} : \mathbb{R}^n \rightarrow [0, 1]$ definida por

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = \mathbb{P}(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n),$$

se llama **función de distribución (marginal) finito dimensional** del proceso $(X_t)_{t \in \mathbb{T}}$.

La función de distribución finito dimensional describe el comportamiento probabilístico del proceso estocástico observado en los distintos tiempo t_1, t_2, \dots, t_n . La distribución de probabilidad de un proceso está caracterizada por el conjunto de todas las distribuciones finito dimensionales. En particular, dos procesos estocásticos con las mismas funciones de distribución finito dimensionales tendrán un comportamiento probabilístico similar.

En el caso de que $(X_t)_{t \in \mathbb{T}}$ sea de estado discreto, entonces dicho comportamiento probabilístico puede ser descrito por medio la **función puntual de probabilidad finito dimensional**

$$P_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = \mathbb{P}(X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n).$$

En el caso de que $(X_t)_{t \in \mathbb{T}}$ sea de estado continuo, consideramos la **función de densidad finito dimensional** $f_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)$, la cual verifica

$$\mathbb{P}(a_1 < X_{t_1} < b_1, \dots, a_n < X_{t_n} < b_n) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) dx_1 \dots dx_n$$

para todo $-\infty \leq a_i < b_i \leq \infty$.

Definición 1.2.2 Dado un proceso estocástico $(X_t)_{t \in \mathbb{T}}$ la **función medio o función de medias** $\mu_X : \mathbb{T} \rightarrow \mathbb{R}$ se define como

$$\mu_X(t) = \mathbb{E}(X_t).$$

- Para un proceso estocástico $(X_t)_{t=0,1,2,\dots}$ de tiempo discreto, la función media define una sucesión de números reales $(\mu_X(t))_{t=0,1,2,\dots}$ de tiempo discreto, la función media define una sucesión de números reales $(\mu_X(t))_{t=0,1,2,\dots}$.
- Para un proceso estocástico $(X_t)_{t \in [0, \infty)}$ de tiempo continuo, la función media define una función real $t \mapsto \mu_X(t) : [0, \infty) \rightarrow \mathbb{R}$.

La función de medias da una idea de cómo el proceso estocástico se comporta en promedio a lo largo del tiempo.

Ejemplo. Volvamos al ejemplo donde una moneda es lanzada reiteras veces y $(X_t)_{t=0,1,2,\dots}$ son las ganancias acumuladas al apostar un euro a cara en cada lanzamiento (paseo aleatorio simétrico). En este caso, $\mu_X(0) = \mathbb{E}(X_0) = \mathbb{E}(0) = 0$ y, si $t > 0$,

$$\mathbb{E}(X_t) = \mathbb{E}(Y_1 + Y_2 + \dots + Y_t) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \dots + \mathbb{E}(Y_t) = 0,$$

donde hemos aplicado que $\mathbb{E}[Y_s] = 0$ para todo s . En definitiva, la ganancia acumulada promedio es 0 para cualquier número de lanzamientos efectuado.

Definición 1.2.3 Dado un proceso estocástico $(X_t)_{t \in \mathbb{T}}$ la **función de covarianza** $C_X : \mathbb{T} \times \mathbb{T} \rightarrow \mathbb{R}$ se define como

$$C_X(s, t) = \text{Cov}(X_s, X_t) = \mathbb{E}((X_s - \mu_X(s))(X_t - \mu_X(t))) = \mathbb{E}(X_s X_t) - \mu_X(s) \mu_X(t).$$

La **función de varianza** $\sigma_X^2 : \mathbb{T} \rightarrow \mathbb{R}$ se define como

$$\sigma_X^2(t) = \text{Var}(X_t) = C_X(t, t),$$

y la **función de correlación** $\rho_X : \mathbb{T} \times \mathbb{T} \rightarrow \mathbb{R}$ se define como

$$\rho_X(s, t) = \frac{C_X(s, t)}{\sigma_X(s)\sigma_X(t)}.$$

Ejemplo. Consideremos el proceso estocástico $(X_t)_{t \in [0, \infty)}$ definido como

$$X_t = A + Bt \quad \forall t \in [0, \infty),$$

donde A y B son variables aleatorias $N(1, 1)$ independientes.

Tenemos por un lado que la función de medias viene dada por:

$$\mu_X(t) = \mathbb{E}(X_t) = \mathbb{E}(A + Bt) = 1 + t \quad \forall t \in [0, \infty).$$

Y por otro lado la función de varianzas es:

$$C_X(s, t) = \mathbb{E}(X_s X_t) - \mu_X(s)\mu_X(t)$$

Observar que la esperanza del producto vale:

$$\begin{aligned} \mathbb{E}(X_s X_t) &= \mathbb{E}((A + Bs)(A + Bt)) \\ &= \mathbb{E}(A^2 + ABs + ABt + B^2 st) \\ &= \mathbb{E}(A^2) + \mathbb{E}(A)\mathbb{E}(B)(s + t) + \mathbb{E}(B^2)st \\ &= 2 + s + t + 2st \end{aligned}$$

Y el producto de esperanza vale:

$$\mu_X(s)\mu_X(t) = (1 + s)(1 + t) = 1 + s + t + st.$$

Por tanto, para este proceso estocástico la función de covarianzas es:

$$C_X(s, t) = (2 + s + t + 2st) - (1 + s + t + st) = 1 + st,$$

la función de varianzas:

$$\sigma_X^2(t) = \text{Var}(X_t) = C_X(t, t) = 1 + t^2,$$

y la función de correlaciones:

$$\rho_X(s, t) = \frac{1 + st}{\sqrt{(1 + t^2)(1 + s^2)}}$$

Ejemplo. Volvamos al ejemplo del paseo aleatorio simétrico $(X_t)_{t=0,1,2,\dots}$. Fijados dos números naturales n, m con $n \leq m$, tenemos

$$\begin{aligned} C_X(n, m) &= \mathbb{E}(X_n X_m) \\ &= \mathbb{E}(X_n(X_m - X_n + X_n)) \\ &= \mathbb{E}(X_n(X_m - X_n)) + \mathbb{E}(X_n^2) \\ &= \mathbb{E}(X_n)\mathbb{E}(X_m - X_n) + \mathbb{E}(X_n^2) \\ &= n = \min(n, m), \end{aligned}$$

donde hemos usado que X_n y $X_m - X_n$ son variables aleatorias independientes. Además, obsérvese que, como $X_n =$

$Y_1 + Y_2 + \dots + Y_n$ y las variables Y_1, Y_2, \dots, Y_n son independientes, se tiene que:

$$\text{Var}(X_n) = \text{Var}(Y_1^2) + \text{Var}(Y_2^2) + \dots + \text{Var}(Y_n^2) = 1 + 1 + \dots + 1 = n,$$

y por tanto:

$$\mathbb{E}(X_n^2) = \text{Var}(X_n) + (\mathbb{E}(X_n))^2 = n + 0 = n.$$

1.3) Procesos estacionarios y débilmente estacionarios

Los procesos estocásticos se pueden dividir entre estacionarios (en sentido estricto), débilmente estacionarios y no estacionarios. Intuitivamente hablando, un proceso estocástico es estacionario (en sentido estricto) si su comportamiento probabilístico o distribución no cambia con el tiempo.

Definición 1.3.1 Un proceso estocástico $(X_t)_{t \in \mathbb{T}}$ se dice que es **estacionario** si para toda sucesión finita de tiempos t_1, t_2, \dots, t_n y todo $s > 0$ se verifica que los vectores aleatorios

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \quad \text{y} \quad (X_{t_1+s}, X_{t_2+s}, \dots, X_{t_n+s})$$

tienen la misma función de distribución, es decir,

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = F_{t_1+s, t_2+s, \dots, t_n+s}(x_1, x_2, \dots, x_n)$$

para cualesquiera números reales x_1, x_2, \dots, x_n .

En particular, si un proceso es estacionario, entonces todas las variables aleatorias X_t del proceso tienen la misma distribución, es decir, están idénticamente distribuidas. Esto no implica que los vectores aleatorios que pueden formar con las variables del proceso en diferentes instantes tengan todos la misma distribución.

En la práctica es muy útil saber si un proceso estocástico es estacionario cuando es necesario predecir el comportamiento futuro de dicho proceso. Si sabemos que el proceso es estacionario, entonces observando su comportamiento en el pasado podremos inferir información de su comportamiento en el futuro.

A continuación introducimos una noción de estacionaridad menos exigente, la cual es también muy útil para predecir comportamientos futuros de procesos estocásticos reales.

Definición 1.3.2 Un proceso estocástico $(X_t)_{t \in \mathbb{T}}$ se dice que es **débilmente estacionario** si:

- 1) $\mu_X(t) = \mu_X(0)$ para todo $t \in \mathbb{T}$. Es decir, la función de medias es constante.
- 2) $C_X(s, t) = C_X(0, t - s)$ para todo $s, t \in \mathbb{T}$ con $s \leq t$. Es decir, la función de covarianzas depende sólo del salto entre tiempos.

Obsérvese que, en particular, si el proceso es débilmente estacionario tendrá función de varianzas constante.

Dado que la media $\mu_X(t)$ es determinada por la función de distribución marginal $F_t(x)$, y la covarianza $C_X(s, t)$ es determinada por la función de distribución marginal $F_{s,t}(x)$, todo proceso estacionario será también débilmente estacionario. Sin embargo, es posible que un proceso sea débilmente estacionario pero no estacionario.

Ejemplo. Consideremos el proceso estocástico $(X_t)_{t \in [0, \infty)}$ definido por

$$X_t = Y \cos(U + t),$$

donde Y y U son variables aleatorias independientes con $Y \sim N(0, 1)$ y $U \sim U[0, 2\pi]$. Entonces

$$\mu_X(t) = 0,$$

y además, para $s \leq t$,

$$\begin{aligned}
C_X(s, t) &= \mathbb{E}(X_s X_t) - \mu_X(s) \mu_X(t) \\
&= \mathbb{E}((Y \cos(U + s))(Y \cos(U + t))) \\
&= \mathbb{E}(Y^2) \mathbb{E}(\cos(U + s) \cos(U + t)) \\
&= \mathbb{E}(\cos(U + s) \cos(U + t)) \\
&= \mathbb{E}\left(\frac{1}{2} \cos(2U + s + t) + \frac{1}{2} \cos(t - s)\right) \\
&= \mathbb{E}\left(\frac{1}{2} \cos(2U + s + t)\right) + \frac{1}{2} \cos(t - s) \\
&= \frac{1}{2} \int_0^{2\pi} \left(\cos(2u + s + t) \frac{1}{2\pi}\right) du + \frac{1}{2} \cos(t - s) \\
&= 0 + \frac{1}{2} \cos(t - s) \\
&= \frac{1}{2} \cos(t - s)
\end{aligned}$$

En particular, tenemos que $C_X(s, t) = \frac{1}{2} \cos(t - s) = C_X(0, t - s)$, de donde deducimos que el proceso estocástico $(X_t)_{t \in [0, \infty)}$ es débilmente estacionario.

1.4) Procesos gaussianos

A continuación introducimos una de las familias de procesos estocásticos más importantes: los procesos gaussianos. Este tipo de procesos tienen importantes aplicaciones en Machine Learning.

Recordemos primero la noción de variable **Normal multivariante**. Una propiedad importante de las variables normales multivariadas es que su función de densidad (y, por tanto, su distribución) está completamente determinada por el vector de medias y la matriz de covarianzas.

Definición 1.4.1 Se dice que el vector aleatorio $\vec{X} = (X_1, X_2, \dots, X_n)$ sigue una distribución Normal multivariante si su función de densidad viene dada por:

$$f_X(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|C_X|}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu}_X)^\top C_X^{-1} (\vec{x} - \vec{\mu}_X) \right\} \text{ para todo } \vec{x} \in \mathbb{R}^n,$$

donde $\vec{\mu}_X$ es el vector de medias de \vec{X} y C_X es la matriz de covarianzas de \vec{X} .

Propiedad: Se puede probar, aunque no es inmediato, que una condición equivalente para el vector aleatorio (X_1, X_2, \dots, X_n) siga una distribución Normal multivariante es que, para cualesquiera números reales a_1, a_2, \dots, a_n la variable aleatoria real

$$a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

sigue una distribución Normal univariante.

Otra importante propiedad de las variables aleatorias normales multivariadas que conviene recordar es que las transformaciones lineales de estas variables también son normales multivariadas. En otras palabras, si \vec{X} es un vector aleatorio Normal multivariante de tamaño n , A es una matriz constante de tamaño $m \times n$, y \vec{b} es un vector constante de tamaño m , entonces el vector aleatorio $\vec{Y} = A\vec{X} + \vec{b}$ también es un vector aleatorio Normal multivariante.

Definición 1.4.2 Un proceso estocástico $(X_t)_{t \in \mathbb{T}}$ se dice que es **gaussiano** (o **Normal**) si para toda sucesión de tiempos $t_1, t_2, \dots, t_n \in \mathbb{T}$ el vector aleatorio

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n})$$

es una variable Normal multivariante.

Una importante propiedad de los procesos gaussianos es que las nociones de estacionaridad y estacionaridad débil son equivalentes para este tipo de procesos. En la práctica esto significa que para verificar que un proceso gaussiano es estacionario

basta analizar su media y covarianza. Concretamente, tenemos lo siguiente.

Teorema 1.4.1 Sea $(X_t)_{t \in \mathbb{T}}$ un proceso estocástico gaussiano. Si $(X_t)_{t \in \mathbb{T}}$ es débilmente estacionario, entonces es también estacionario.

1.5) Ejemplos y simulación

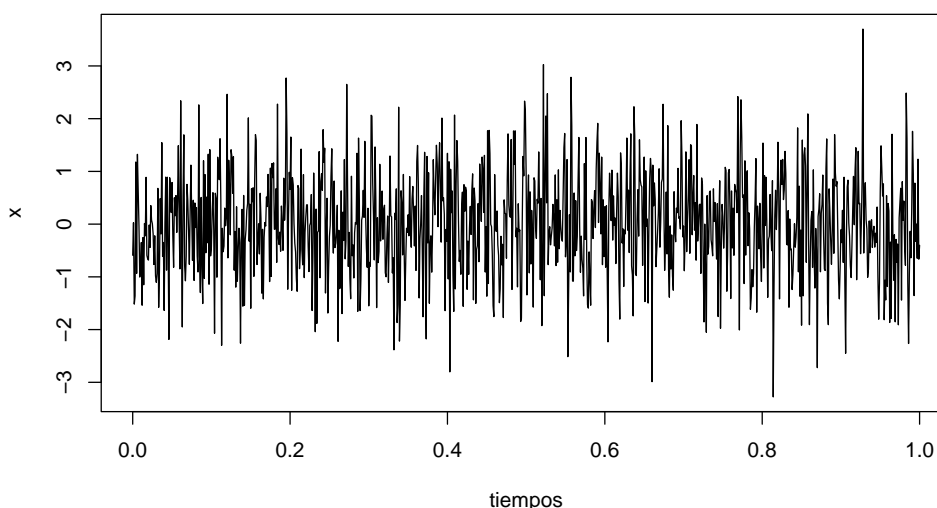
En lo que sigue vamos a estudiar y simular en R algunos ejemplos particulares de procesos estocásticos gaussianos, analizando en cada paso si se trata de procesos estacionarios o no estacionarios.

1.5.1) Ruido blanco gaussiano

Definición 1.5.1 Un proceso estocástico $(X_t)_{t \in \mathbb{T}}$ se llama **ruido blanco gaussiano** si las variables aleatorias X_t son independientes y siguen una distribución $N(0, \sigma^2)$.

Claramente el ruido blanco gaussiano es un proceso gaussiano. Además, $\mu_X(t) = 0$, $C_X(t, s) = 0$ si $t \neq s$, y $C_X(t, t) = \sigma^2$. Lo cual en particular implica que el ruido blanco gaussiano es un proceso estacionario.

Podemos simular en R una trayectoria del ruido blanco gaussiano:



Obtenemos una trayectoria extremadamente irregular. Además, observamos a simple vista un comportamiento estacionario del proceso.

1.5.2) Movimiento Browniano

Definición 1.5.2 Un proceso estocástico $(X_t)_{t \in [0, \infty)}$, de tiempo continuo se dice que es un **movimiento Browniano** o **proceso de Wiener** si cumple:

- 1) $X_0 = 0$.
- 2) Para todo par de tiempos $s \leq t$, la variable aleatoria $X_t - X_s$, sigue una distribución $N(0, t - s)$.
- 3) Para cualquier sucesión de tiempo $t_1 < t_2 < \dots < t_n$ se tiene que las variables aleatorias

$$X_{t_2} - X_{t_1}, \quad X_{t_3} - X_{t_2}, \quad \dots, \quad X_{t_n} - X_{t_{n-1}}$$

son independientes.

- 4) Las trayectorias $t \mapsto X_{t(\omega)}$ son funciones continuas.

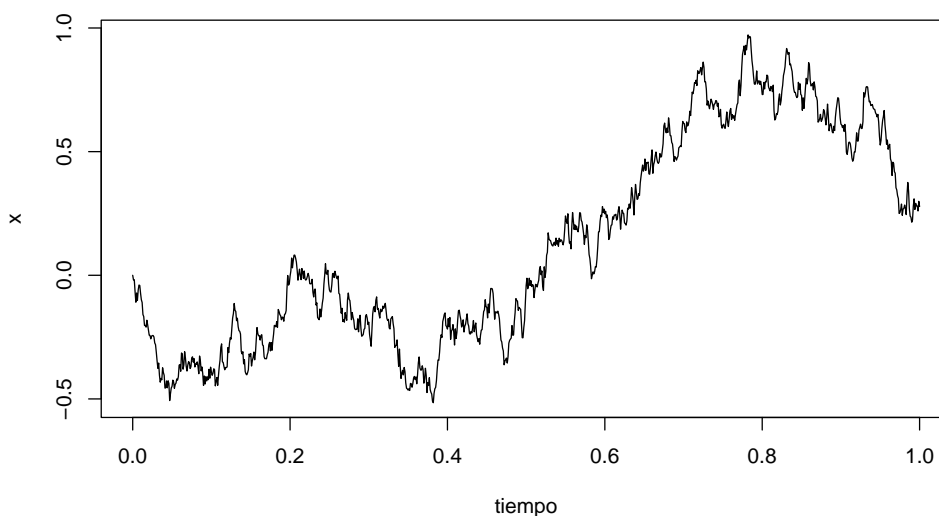
El movimiento Browniano se utiliza en finanzas para modelar la evolución del precio de activos financieros como las acciones o los tipos de cambio. En particular, se aplica a la valoración de opciones financieras mediante el bloque de Black-Scholes, y el análisis de riesgos.

A continuación vamos a simular una trayectoria del movimiento Browniano. Para ello fijamos un incremento de tiempo $\Delta t > 0$ y un número natural n . Deseamos simular una trayectoria en los instantes de tiempo $t_0 = 0, t_1 = \Delta t, \dots, t_n = n\Delta t$. En vista de la definición de movimiento Browniano, tenemos que los incrementos

$$\Delta X_1 = X_{t_1} - X_{t_0}, \quad \Delta X_2 = X_{t_2} - X_{t_1}, \quad \dots, \quad \Delta X_n = X_{t_n} - X_{t_{n-1}},$$

son variables aleatorias independientes y siguen todas ellas una distribución $N(0, \Delta t)$. De este modo, nuestra estrategia será simular los valores de los incrementos mediante la generación de una muestra aleatoria de n variables $N(0, \Delta t)$ independientes, y reconstruir la trayectoria del proceso a partir de dichos incrementos. Esto podemos lograrlo mediante el siguiente código en R:

```
set.seed(11)
n <- 1000
dt <- 1/n
tiempo <- seq(from=0, to=1, by=dt)
incrementos <- rnorm(n, 0, sqrt(dt))
x <- c(0, cumsum(incrementos))
plot(tiempo, x, type = "l", lty = 1)
```



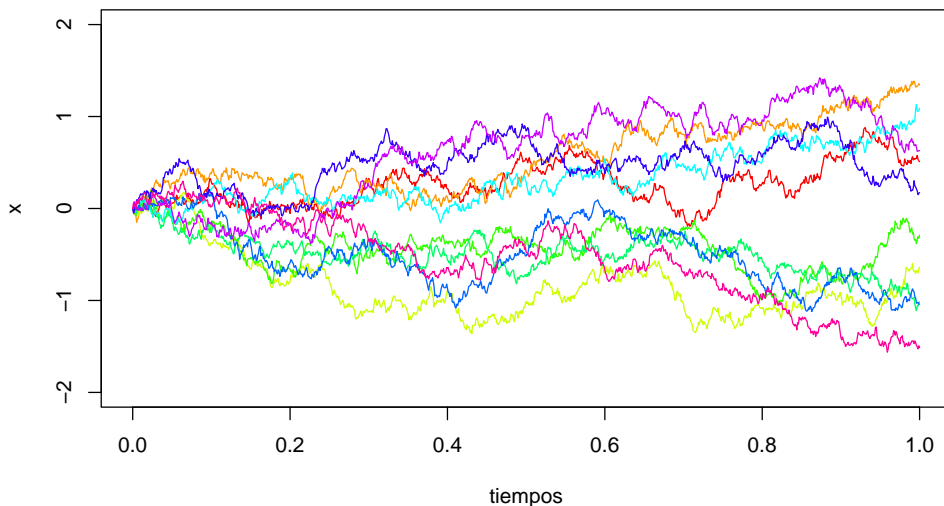
Para analizar el comportamiento probabilístico, simularemos diez trayectorias del proceso:

```
set.seed(123)
nsim <- 10
n <- 1000
dt <- 1/n
tiempos <- seq(from=0, to=1, by=dt)
incrementos <- rnorm(n, 0, sqrt(dt))
x <- c(0, cumsum(incrementos))

colores <- rainbow(nsim)

plot(tiempos, x, ylim=c(-2,2), type = "l", lty = 1, col=colores[1])
```

```
for (i in 2:nsim) {
  incrementos=rnorm(n, 0, sqrt(dt))
  x <- c(0, cumsum(incrementos))
  lines(tiempos, x, col=colores[i])
}
```



A simple vista observamos un comportamiento no estacionario del proceso. Notamos que a medida que avanza el tiempo, las trayectorias tienden a volverse más dispersas, lo que indica que el comportamiento probabilístico de las trayectorias varía con el tiempo.

El movimiento Browniano es un proceso gaussiano. De hecho, la distribución de probabilidad del vector aleatorio $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$, para $t_1 < t_2 < \dots < t_n$, es normal multivariada porque se obtiene como una transformación lineal del vector $(X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}})$, el cual es normal multivariado porque sus componentes son independientes y normales.

Además, $\mu_X(t) = \mathbb{E}(X_t) = 0$ y, para $s \leq t$,

$$\begin{aligned} C_X(s, t) &= \mathbb{E}(X_s X_t) \\ &= \mathbb{E}(X_s(X_t - X_s + X_s)) \\ &= \mathbb{E}(X_s(X_t - X_s)) + \mathbb{E}(X_s^2) \\ &= s = \min(s, t) \end{aligned}$$

En resumen, el movimiento Browniano es un proceso gaussiano con funciones de media y covarianza $\mu_X(t) = 0$, y $C_X(s, t) = \min(s, t)$, respectivamente. En particular, observamos que este proceso no es estacionario (ni siquiera en el sentido débil), ya que en general no se verifica que $C_X(s, t) = C_X(0, t - s)$. Por ejemplo, $C_X(3, 4) = \min(3, 4) = 3$; sin embargo, $C_X(0, 4 - 3) = \min(0, 1) = 0$.

A continuación vamos a detallar otra estrategia para simular el movimiento Browniano, la cual se basa en que dicho proceso es gaussiano. Aunque nos vamos a centrar en el caso particular del movimiento Browniano, este procedimiento sirve para simular cualquier proceso gaussiano conocidas sus funciones de media y covarianza. Fijados los tiempos $t_1 < \dots < t_n$ sabemos que el vector aleatorio $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ sigue una distribución multivariada de media $\mu_X = 0$ y matriz de covarianzas $C_X = (c_{i,j})_{n \times n}$ donde $c_{i,j} = \min\{t_i, t_j\}$. De esta manera, podremos simular una trayectoria mediante la obtención de una muestra del anterior vector aleatorio. Para ello utilizamos la librería `mvtnorm` de R, la cual permite simular variables aleatorias normales multivariadas.

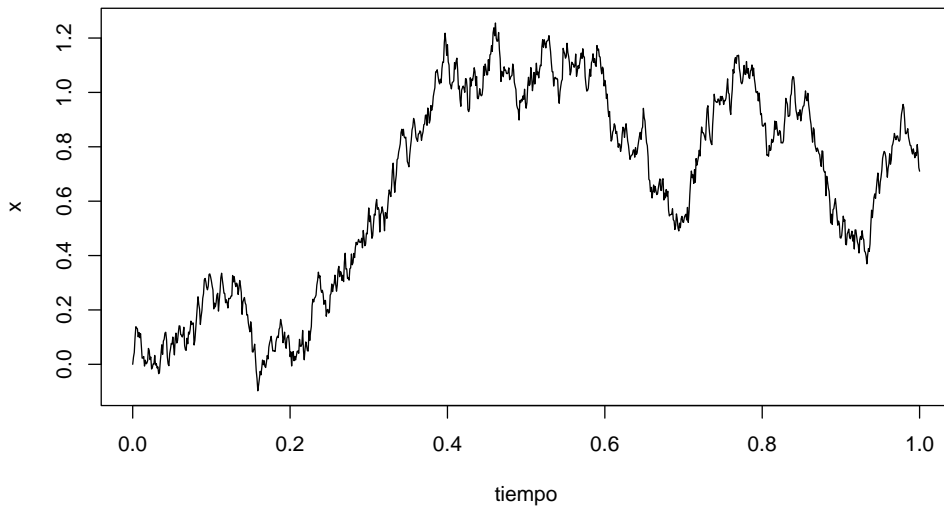
```
set.seed(23)
n <- 1000
```

```
dt <- 1/n
tiempo <- seq(from=0, to=1, by=dt)

# Definimos la matriz de covarianzas
C <- matrix(NA, nrow = n, ncol = n)
C <- dt*pmin(row(C), col(C))

# Simulamos una normal multivariada
library(mvtnorm)
x <- rmvnorm(1, sigma = C)

# Añadimos el valor 0 inicial
x <- c(0, x)
plot(tiempo, x, type = "l", lty = 1)
```



1.6) Procesos gaussianos estacionarios isotrópicos

Un tipo particular de proceso gaussiano estacionario son los procesos gaussianos estacionarios isotrópicos.

Definición 1.6.1 Decimos que un proceso gaussiano estacionario $(X_t)_{t \in \mathbb{T}}$ es **isotrópico** si su función de medias es constante, $\mu_X(t) = \mu$, y la función de covarianza depende sólo de la distancia $|t - s|$, es decir,

$$C_X(s, t) = K(|t - s|)$$

para cierta función $K : [0, \infty) \rightarrow [0, \infty)$ llamada núcleo.

Tomemos por ejemplo $\mu_X(t) = 0$ y el núcleo exponencial $K_l(x) = \exp\left(-\frac{x^2}{2l^2}\right)$, donde l es un parámetro de escala. Vamos a simular en R una trayectoria de dicho proceso:

```
set.seed(112)
n <- 1000
dt <- 1/n
tiempo <- seq(from=0, to=1, by=dt)

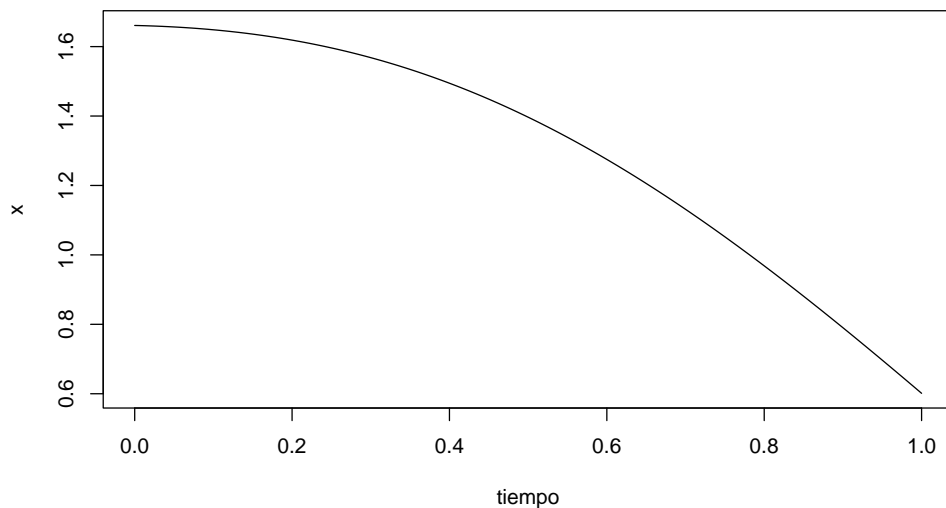
# Definimos la matriz de covarianzas
```

```

l <- 1
C <- matrix(NA, nrow = n+1, ncol = n+1)
C <- dt*abs(row(C)-col(C))
C <- exp(-C^2/(2*l^2))

# Simulamos una normal multivariada
library(mvtnorm)
x <- rmvnorm(1, sigma = C)
plot(tiempo, x, type = "l", lty = 1)

```



Para analizar el comportamiento probabilístico, podemos simular diez trayectorias del proceso:

```

set.seed(112)
nsim <- 10
n <- 1000
dt <- 1/n
tiempo <- seq(from=0, to=1, by=dt)

# Definimos la matriz de covarianzas
l <- 1
C <- matrix(NA, nrow = n+1, ncol = n+1)
C <- dt*abs(row(C)-col(C))
C <- exp(-C^2/(2*l^2))

# Simulamos una normal multivariada
library(mvtnorm)
x <- rmvnorm(1, sigma = C)

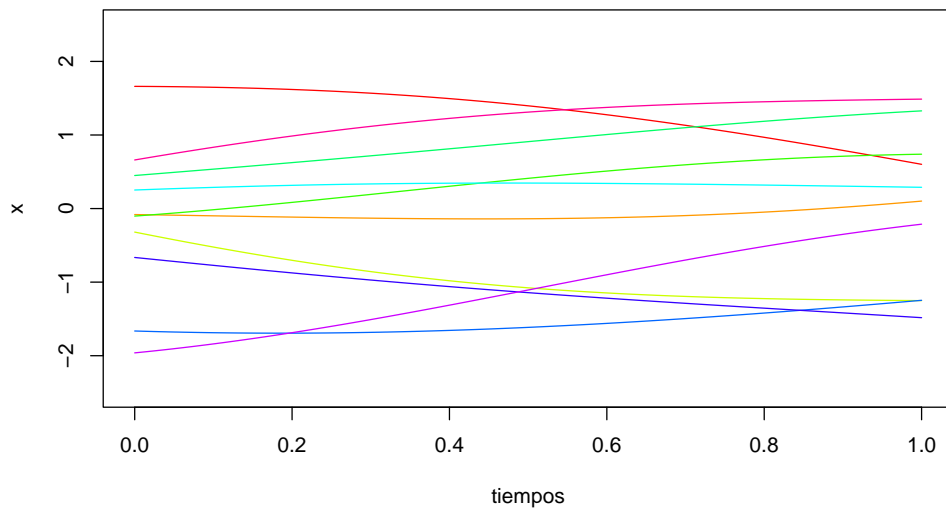
colores <- rainbow(nsim)

plot(tiempos, x, ylim=c(-2.5,2.5), type = "l", lty = 1, col=colores[1])

for (i in 2:nsim) {
  incrementos=rnorm(n, 0, sqrt(dt))

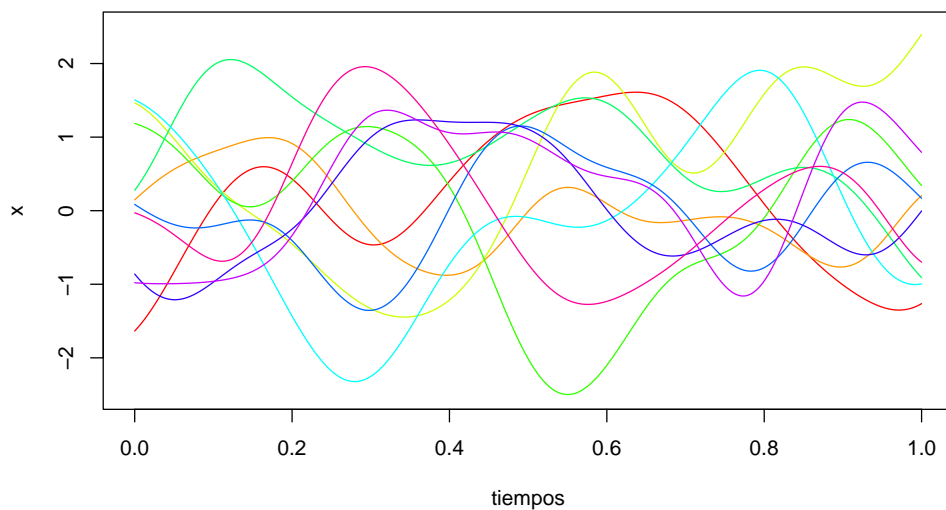
```

```
x <- rmvnorm(1, sigma = C)
lines(tiempos, x, col=colores[i])
}
```



A diferencia del movimiento Browniano, vemos que las trayectorias presentan un comportamiento estacionario, ya que su distribución es la misma a lo largo del tiempo. Otra diferencia con el movimiento Browniano es que las trayectorias tienen una apariencia suave.

Podemos cambiar el parámetro de escala, tomando $l = 0.1$. De este modo obtenemos las trayectorias:



Tomemos ahora el núcleo de Ornstein-Uhlenbeck $K - l(x) = \exp\left(-\frac{|x|}{l}\right)$.

Simulemos algunas trayectorias en R:

```
set.seed(112)
nsim <- 10
n <- 1000
dt <- 1/n
tiempo <- seq(from=0, to=1, by=dt)
```

```

# Definimos la matriz de covarianzas
l <- 1
C <- matrix(NA, nrow = n+1, ncol = n+1)
C <- dt*abs(row(C)-col(C))
C <- exp(-C/l)

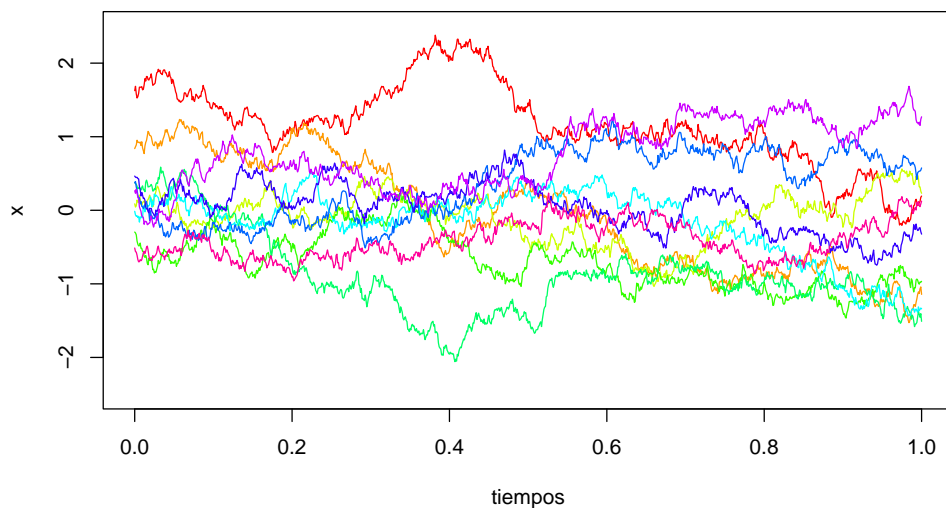
# Simulamos una normal multivariada
library(mvtnorm)
x <- rmvnorm(1, sigma = C)

colores <- rainbow(nsim)

plot(tiempos, x, ylim=c(-2.5,2.5), type = "l", lty = 1, col=colores[1])

for (i in 2:nsim) {
  x <- rmvnorm(1, sigma = C)
  lines(tiempos, x, col=colores[i])
}

```



Vemos que las trayectorias tienen una apariencia más irregular la cual nos recuerda al movimiento Browniano. Sin embargo, en este caso tenemos un proceso estacionario cuyas trayectorias se comportan de manera similar a lo largo del tiempo.

Hoja 1: Problemas de Introducción a los Procesos Estocásticos

- 1) ¿Qué es un proceso estocástico y cómo se define formalmente?
- 2) ¿Cuáles son los tipos principales de procesos estocásticos en función del tiempo?
- 3) ¿Cómo se clasifica un proceso estocástico según los estados que puede formar?
- 4) ¿Qué es una trayectoria en el contexto de un proceso estocástico?
- 5) ¿Qué es un paseo aleatorio y cómo se define en términos de variables aleatorias? Simular, usando R, 4 trayectorias de un paseo aleatorio con variables aleatorias i.i.d. dadas por la distribución $N(0, 1)$.
- 6) ¿Qué características tiene un paseo aleatorio simple y en qué se diferencia del paseo aleatorio simétrico? Simular, usando R, 6 trayectorias de un paseo aleatorio simple $p = \frac{1}{3}$.
- 7) ¿Qué es una función de distribución finito dimensional en un proceso estocástico?
- 8) ¿Cómo se define la función de medias de un proceso estocástico? ¿Y la de covarianzas y correlaciones?
- 9) ¿Qué relación existe entre la función de varianza y la función de covarianza en un proceso estocástico?
- 10) ¿Qué condiciones deben cumplirse para que un proceso sea estacionario y en qué se diferencia de la estacionariedad débil?
- 11) ¿Qué diferencias hay entre un proceso estocástico gaussiano y un ruido blanco gaussiano? Simular, usando R, 2 trayectorias de un ruido blanco gaussiano con varianza σ^2 .
- 12) ¿Cuál es la definición de un proceso de Wiener o movimiento Browniano? Simular, usando R, 3 trayectorias de un movimiento Browniano.
- 13) ¿Cuál de las siguientes es una condición necesaria para que un proceso estocástico sea débilmente estacionario?
 - (a) La varianza depende del tiempo.
 - (b) La media es constante en el tiempo.
 - (c) La covarianza es cero para todos los tiempos.
- 14) Si $(X_t)_{t \in [0, \infty)}$ es ruido blanco (gaussiano) con varianza σ^2 , ¿cuál de las siguientes afirmaciones es verdadera?
 - (a) $\text{Cov}(X_t, X_{t+h}) = \sigma^2$ para todo t y h .
 - (b) $\text{Cov}(X_t, X_{t+h}) = \sigma^2$ si $h = 0$, $\text{Cov}(X_t, X_{t+h}) = 0$ si $h \neq 0$.
 - (c) $\text{Cov}(X_t, X_{t+h}) = \frac{1}{2}\sigma^2$ para todo t y h .
- 15) ¿Cuál de las siguientes afirmaciones es verdadera para un ruido blanco?
 - (a) Tiene una función de correlación que decrece exponencialmente.
 - (b) Su media es no nula.
 - (c) Sus valores en diferentes tiempos son independientes.
- 16) ¿Qué describe la función de covarianza de un proceso estocástico?
 - (a) La relación lineal en el proceso en dos tiempos diferentes.
 - (b) La suma de todas las realizaciones del proceso.
 - (c) La frecuencia con la que el proceso cruza la media.
- 17) Una condición necesaria para que un proceso estocástico sea débilmente estacionario es que:
 - (a) La media debe ser cero.
 - (b) La covarianza depende del tiempo.

(c) La función de covarianza depende solo de la diferencia entre los tiempos.

- 18) Consideremos el proceso estocástico de tiempo discreto $(X_t)_{t=0,1,2,\dots}$, con $X_0 = a \in \mathbb{R}$ una constante cualquiera y $X_t = 3 + \varepsilon_t + 2\varepsilon_{t-1}$ para todo $t = 1, 2, \dots$, siendo $(\varepsilon_t)_{t=0,1,2,\dots}$ ruido blanco (gaussiano) de varianza σ^2 .

(a) Calcular la función de medias, covarianzas y correlaciones del proceso $(X_t)_{t=0,1,2,\dots}$

(b) Justificar si el proceso es estacionario en sentido débil.

(c) Justificar si el proceso es estacionario (en sentido fuerte).

Nota: El proceso del enunciado es un proceso denominado MA(1) que estudiará en el tema posterior.

- 19) Consideremos el proceso estocástico de tiempo discreto $(X_t)_{t=0,1,2,\dots}$, con $X_0 = a \in \mathbb{R}$ una constante cualquiera y $X_t = \alpha X_{t-1} + \varepsilon_t$, para todo $t = 1, 2, \dots$, siendo $(\varepsilon_t)_{t=1,2,\dots}$ ruido blanco (gaussiano) de varianza σ^2 . Suponiendo el proceso $(X_t)_{t=0,1,2,\dots}$ es débilmente estacionario, obtener la función de medias y correlaciones.

Nota: El proceso del enunciado es un proceso denominado AR(1) que estudiará en el tema posterior.

Supongamos que X_t es débilmente estacionario.

Nos piden función medias, covarianzas y correlaciones.

- **Función de medias** del $(X_t)_{t=0,1,2,\dots}$

$$\mathbb{E}(X_t) = \mathbb{E}(\alpha X_{t-1} + \varepsilon_t) = \underbrace{\alpha \mathbb{E}(X_{t-1})}_{\mu} + \mathbb{E}(\varepsilon_t) \xrightarrow{0} \mu$$

Esto implica que

$$\mu = \alpha \mu \begin{cases} \mu = 0 \\ \alpha = 1 \end{cases}$$

- **Función de varianzas** de $(X_t)_{t=0,1,2,\dots}$

$$\begin{aligned} \sigma_X^2 &= \text{Var}(X_t) = \text{Var}(\alpha X_{t-1} + \varepsilon_t) = \alpha^2 \underbrace{\text{Var}(X_{t-1})}_C + \underbrace{\text{Var}(\varepsilon_t)}_{\sigma^2} \\ C &\longleftrightarrow \alpha^2 C + \sigma^2 \longleftrightarrow C - \alpha^2 C = \sigma^2 \longleftrightarrow C = \frac{\sigma^2}{1 - \alpha^2} \end{aligned}$$

$C = \text{Var}(X_t) \forall t$ constante porque hemos supuesto débilmente estacionario.

Observar que para que exista la varianza C , debe cumplirse que $|\alpha| < 1$.

- **Función de correlaciones**

- 20) Sea $(X_t)_{t=0,1,2,\dots}$ un paseo aleatorio simple. Es decir, $X_0 = 0$ y $X_t = X_{t-1} + Y_t$, $(Y_t)_{t=0,1,2,\dots}$ variables aleatorias i.i.d verificando $P(Y_t = 1) = p$, $P(Y_t = -1) = 1 - p$, $p \in (0, 1)$.

(a) Calcular la función de medias, covarianzas y correlaciones del proceso $(X_t)_{t=0,1,2,\dots}$

(b) Justificar si el proceso es estacionario en sentido débil.

(c) Justificar si el proceso es estacionario (en sentido fuerte).

Vamos a calcular la media y varianza de la variable aleatoria Y_t .

$$\mathbb{E}(Y_t) = 1 \cdot p + (-1) \cdot (1 - p) = 2p - 1 \quad \forall t = 1, 2, \dots$$

$$\text{Var}(Y_t) = \mathbb{E}(Y_t^2) - (\mathbb{E}(Y_t))^2 = 1 - (4p^2 - 4p + 1) = -4p^2 + 4p = 4p(1 - p) \quad \forall t = 1, \dots, n$$

$$\mathbb{E}(Y_t^2) = 1^2 \cdot p + (-1)^2 \cdot (1 - p) = 1$$

- **Función de medias** del proceso $(X_t)_{t=0,1,2,\dots}$

$$X_0 = 0$$

$$X_1 = X_0 + Y_1 = 0 + Y_1 = Y_1$$

$$X_2 = X_1 + Y_2 = Y_1 + Y_2$$

\vdots

$$X_t = Y_1 + Y_2 + \dots + Y_t \quad \forall t = 1, 2, \dots$$

$$\mu_X(t) = \mathbb{E}(X_t) = \mathbb{E}(Y_1 + Y_2 + \dots + Y_t) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \dots + \mathbb{E}(Y_t) = t \cdot (2p - 1)$$

Observar que no es constante \implies el proceso no es estacionario ni débil ni fuerte.

- **Función de covarianzas** del proceso. Se $t \in \{0, 1, 2, \dots\}$ y $h \geq 0$:

$$\begin{aligned} C_X(t, t+h) &= \text{Cov}(X_t, X_{t+h}) \\ &= \text{Cov}(Y_1 + Y_2 + \dots + Y_t, Y_1 + Y_2 + \dots + Y_{t+h}) \\ &= \text{Cov}(Y_1, Y_1) + \text{Cov}(Y_2, Y_2) + \dots + \text{Cov}(Y_t, Y_t) \\ &= \text{Var}(Y_1) + \text{Var}(Y_2) + \dots + \text{Var}(Y_t) \end{aligned}$$

Observar que $\text{Cov}(Y_t, Y_s) = 0 \quad \forall s \neq t$ (por independencia)

- **Función de varianzas** del proceso:

$$\text{Var}(X_t) = t \cdot 4p(1-p) \text{ depende del tiempo } t.$$

- **Función de correlaciones** del proceso $(X_t)_{t=0,1,2,\dots}$

$$\rho_X(t, t+h) = \frac{C_X(t, t+h)}{\sigma_X(t)\sigma_X(t+h)} = \frac{t \cdot 4p(1-p)}{\sqrt{t4p(1-p)}\sqrt{(t+h)4p(1-p)}} = \frac{t}{\sqrt{t}\sqrt{t+h}}$$

21) Demostrar que si $(X_t)_{t \in [0, \infty)}$ es un movimiento Browniano, entonces es un proceso estocástico gaussiano no estacionario.

22) Consideremos el proceso estocástico $(X_t)_{t \in [0, \infty)}$ con $X_t = Ae^{\lambda t} + B$, con $A \sim U(0, 4)$ y $B \sim \text{Exp}(\lambda)$ siendo A y B independientes.

- Calcular la función de medias, covarianzas y correlaciones del proceso $(X_t)_{t \in [0, \infty)}$
- Justificar si el proceso es estacionario en sentido débil.
- Justificar si el proceso es estacionario (en sentido fuerte).

23) Consideremos el proceso estocástico $(X_t)_{t \in [0, \infty)}$ con $X_t = Y \sin(U + t)$, con $U \sim U(0, 2\pi)$ e $Y \sim N(0, 1)$ siendo U e Y independientes.

- Calcular la función de medias, covarianzas y correlaciones del proceso $(X_t)_{t \in [0, \infty)}$.
- Justificar si el proceso es estacionario en sentido débil.

Tema 2: Cadenas de Markov

En este tema vamos a estudiar un tipo particular de procesos estocásticos: las cadenas de Markov. Este tipo de procesos describen cambios de estado en un sistema, con la peculiaridad de que dichos cambios dependen únicamente del estado actual y no están influenciados por ningún estado que haya tomado previamente.

Por ejemplo, imaginemos un aparcamiento y consideremos la variable X_n representando el número de plazas de aparcamiento ocupadas en cada instante de tiempo $n \in \{0, 1, 2, \dots\}$. Claramente el valor de X_{n+1} dependerá de X_n ya que X_{n+1} se obtiene de X_n sumándole los coches que han aparcado y restándole los que se han ido entre los instantes n y $n + 1$. Por tanto, conocido X_n , parece razonable que la cantidad X_{n+1} no dependa de los valores previos $X_0, X_1, X_2, \dots, X_{n-1}$.

Este tipo de procesos, en los que el valor X_{n+1} depende exclusivamente de X_n y no se ve influenciado por los estados previos $X_0, X_1, X_2, \dots, x_{n-1}$ es formalizado matemáticamente mediante el concepto de cadena de Markov.

2.1) Cadenas de Markov de tiempo discreto

Definición 2.1.1 Un proceso estocástico $(X_n)_{n=0,1,2,\dots}$ de tiempo discreto se dice que es una **cadena de Markov** (de tiempo discreto) si se verifica:

- 1) Cada X_n toma valores en un conjunto numerable (es decir, finito o infinito numerable) \mathcal{S} llamado **espacio de estados**.
- 2) Se cumple que

$$\mathbb{P}(X_{n+1} = a_{n+1} | X_n = a_n, X_{n-1} = a_{n-1}, \dots, X_0 = a_0) = \mathbb{P}(X_{n+1} = a_{n+1} | X_n = a_n), \quad (1)$$

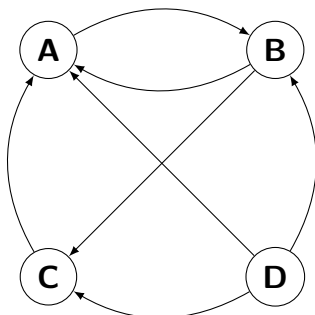
donde $a_0, a_1, \dots, a_n, a_{n+1} \in \mathcal{S}$.

La condición (1) arriba se llama **propiedad de Markov**. La interpretación es que la probabilidad de cualquier valor futuro del proceso, dado el valor actual, no está influenciada por ningún valor pasado. Se dice que las cadenas de Markov son procesos estocásticos sin memoria.

Ejemplo. (PageRank) es un algoritmo creado y desarrollado por la compañía tecnológica estadounidense Google para ordenar las apariciones de las páginas en cada búsqueda, dando preferencia a aquellas páginas que sean más "importantes" o "populares". Para medir esto, se analiza la cadena de Markov resultante de un individuo o "surfeador de la web" que va pulsando links al azar en un conjunto de páginas de internet. Por ejemplo, supongamos que el surfeador de la web navega haciendo clics al azar en las páginas A,B,C,D donde:

- A tiene enlace a B.
- B tiene enlaces a A y C.
- C tiene enlace a A.
- D tiene enlaces a las otras tres páginas.

Este proceso da lugar a una cadena de Markov con espacio de estados $\{A, B, C, D\}$, el cuál puede ser descrito mediante el siguiente grafo.



Además, tenemos las siguientes probabilidades de transición entre estados.

	A	B	C	D
A	0%	100%	0%	0%
B	50%	0%	50%	0%
C	100%	0%	0%	0%
D	33.3333%	33.3333%	33.3333%	0%

Claramente este proceso da lugar a una cadena de Markov, ya que, en cada paso, las probabilidades de visitar una página u otra, sólo depende de en qué página se encuentre el surfador, sin importar qué páginas haya visitado anteriormente.

El algoritmo PageRank trata de determinar la probabilidad con la que una página es visitada a medida que el surfador hace más y más clics, considerando como más importantes aquellas páginas para las que esta probabilidad. Éste es el criterio usado para ordenar las páginas en cada búsqueda en Google. Analizaremos este problema en detalle al final del tema.

Ejemplo. El paseo aleatorio simétrico $(X_n)_{n=0,1,2,\dots}$ es una cadena de Markov con espacio de estados infinito

$$\mathcal{S} = \mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

En este caso, para todo $i \in \mathcal{S}$

$$\begin{aligned}\mathbb{P}(X_{n+1} = i + 1 | X_n = i) &= \mathbb{P}(X_{n+1} = i - 1 | X_n = i) = \frac{1}{2}, \\ \mathbb{P}(X_{n+1} = j | X_n = i) &= 0 \text{ si } j \neq i \pm 1.\end{aligned}$$

Definición 2.1.2 Supongamos que $(X_n)_{n=0,1,2,\dots}$ es una cadena de Markov con espacio de estados \mathcal{S} . Las **probabilidades de transición** son las probabilidades

$$p_{x,y}(n) = \mathbb{P}(X_n = y | X_{n-1} = x),$$

donde $x, y \in \mathcal{S}$ y $n = 1, 2, 3, \dots$

Definición 2.1.3 Decimos que la cadena de Markov $(X_n)_{n=0,1,2,\dots}$ es **homogénea** si las probabilidades de transición no dependen del tiempo. En tal caso, definimos

$$\begin{aligned}p_{x,y} &= \mathbb{P}(X_n = y | X_{n-1} = x) \\ &= \mathbb{P}(X_1 = y | X_0 = x)\end{aligned}$$

donde hemos eliminado n de la notación.

En el resto de este tema vamos a trabajar con cadenas de Markov homogéneas con un número de estados **finito**. Por lo que a partir de ahora, cada vez que nos refiramos a una **cadena** estaremos hablando de una cadena de Markov con esas características.

En general, usaremos que $(X_n)_{n=0,1,2,\dots}$ es una cadena con espacio de estados finito que denotaremos por $\mathcal{S} = \{1, 2, 3, \dots, N\}$.

Definición 2.1.4 Definimos la **matriz de transición** de $(X_n)_{n=0,1,2,\dots}$ como

$$P = (p_{i,j})_{i,j=1,2,\dots,N} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\ \dots & \dots & \dots & \dots \\ p_{N,1} & p_{N,2} & \cdots & p_{N,N} \end{bmatrix}$$

Obsérvese que las entradas de la fila 1 describen todas las probabilidades posibles condicionadas a empezar en el estado

$i = 1$. Por tanto, la suma de todas ellas debe ser igual a 1. Obviamente, lo mismo es cierto para el resto de filas, es decir, $p_{i,1} + p_{i,2} + \dots + p_{i,N} = 1$ para cada i . Por tanto, en una matriz de transición, la suma de todos los elementos de cualquier fila debe de ser 1. Sin embargo, esta condición no tiene por qué verificarse para las columnas de una matriz de transición.

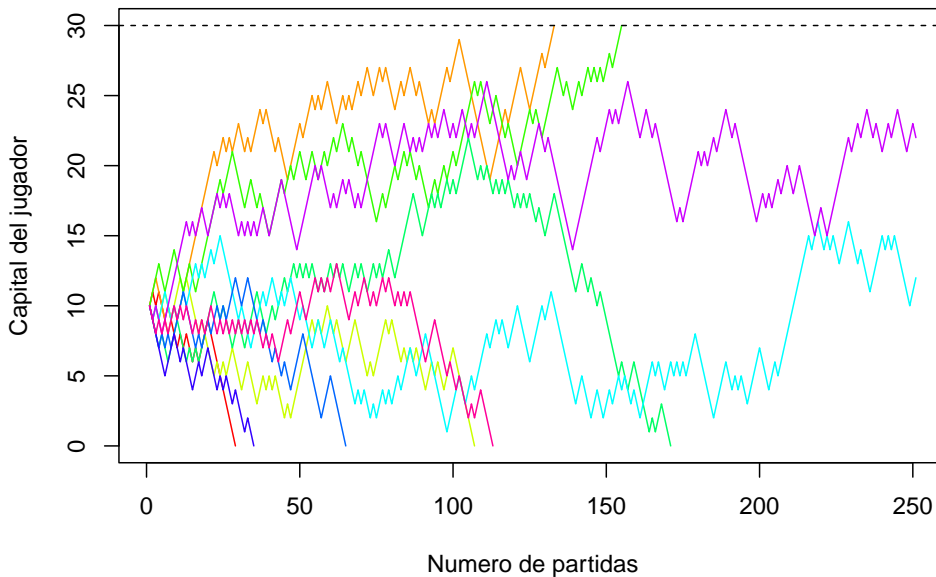
Ejemplo. (Urnas de Ehrenfest) Supongamos que tenemos dos urnas, U_1 y U_2 . En ellas están distribuidas N bolas numeradas. En cada paso, se elige un número al azar entre $\{1, 2, \dots, N\}$. A continuación se observa en qué urna está la bola con el número elegido y se cambia de urna. Denotemos por X_n el número de bolas contenidas en la urna U_1 en tiempo n . De esta manera, definimos una cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados $\mathcal{S} = \{0, 1, 2, \dots, N\}$ y probabilidades de transición

$$\begin{aligned}\mathbb{P}(X_n = i + 1 | X_{n-1} = i) &= \frac{N - i}{N}, \\ \mathbb{P}(X_n = i - 1 | X_{n-1} = i) &= \frac{i}{N}, \\ \mathbb{P}(X_n = j | X_{n-1} = i) &= 0 \quad \text{si } j \neq i \pm 1.\end{aligned}$$

Por ejemplo, si $N = 3$ tenemos espacio de estados $\{0, 1, 2, 3\}$ y una matriz de transición

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Ejemplo. (Ruina del jugador) Consideremos un individuo que juega a la ruleta, que posee una riqueza inicial de X_0 euros, y que apuesta 1 euro al rojo en cada jugada con probabilidad de ganar p . El jugador seguirá apostando hasta que, o bien alcance una riqueza objetivo M , o bien hasta que se arruine. El proceso $(X_n)_{n=0,1,2,\dots}$ de la riqueza acumulada hasta la jugada n es una cadena de Markov con espacio de estados finito $\mathcal{S} = \{0, 1, \dots, M\}$.



En este caso, el proceso de la fortuna acumulada $(X_n)_{n=0,1,2,\dots}$ es una cadena de estados $\{0, 1, 2, \dots, M\}$ y transición

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & p & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$$

2.2) Matriz de transición de n pasos

En esta sección, sea $(X_n)_{n=0,1,2,\dots}$ una cadena con espacio de estados $\mathcal{S} = \{1, 2, \dots, N\}$. En general, definimos lo siguiente.

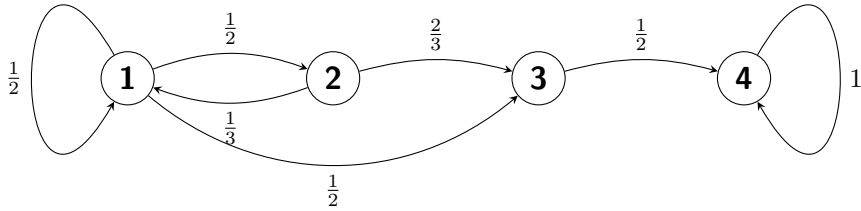
Definición 2.2.1 Dados dos estados $i, j \in \mathcal{S}$, definimos la probabilidad de transición de n pasos

$$p_{i,j}^{(n)} = \mathbb{P}(X_n = j | X_0 = i).$$

Definimos la **matriz de transición de n pasos** de $(X_n)_{n=0,1,2,\dots}$ como

$$P^{(n)} = \left(p_{i,j}^{(n)} \right)_{i,j=1,2,\dots,N} = \begin{bmatrix} p_{1,1}^{(n)} & p_{1,2}^{(n)} & \cdots & p_{1,N}^{(n)} \\ p_{2,1}^{(n)} & p_{2,2}^{(n)} & \cdots & p_{2,N}^{(n)} \\ \cdots & \cdots & \cdots & \cdots \\ p_{N,1}^{(n)} & p_{N,2}^{(n)} & \cdots & p_{N,N}^{(n)} \end{bmatrix}.$$

Para entender cómo calcular la matriz de transición de n pasos analicemos el siguiente ejemplo. Consideremos la cadena $(X_n)_{n=0,1,2,\dots}$ dada por el grafo:



Esta cadena tiene matriz de transición

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Queremos ver lo que ocurre tras dos pasos del proceso.

Por ejemplo, veamos la probabilidad de llegar al estado 1 desde el estado 1 en dos pasos

$$p_{1,1}^{(2)} = \mathbb{P}(111) + \mathbb{P}(121) = p_{1,1} \cdot p_{1,1} + p_{1,2} \cdot p_{2,1} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} = \frac{5}{12}$$

Incluyendo todos los posibles caminos, incluso aquellos que no existen en el grafo, podemos poner

$$\begin{aligned} p_{1,1}^{(2)} &= \mathbb{P}(111) + \mathbb{P}(121) + \mathbb{P}(131) + \mathbb{P}(141) \\ &= p_{1,1} \cdot p_{1,1} + p_{1,2} \cdot p_{2,1} + p_{1,3} \cdot p_{3,1} + p_{1,4} \cdot p_{4,1} \\ &= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} + 0 + 0 = \frac{5}{12}. \end{aligned}$$

Escrito de esta manera, vemos que $p_{1,1}^{(2)}$ es la primera entrada de la matriz $P^2 = P \cdot P$. Es más, tenemos que $P^2 = (p_{i,j}^2)_{i,j \in \{1,2,3,4\}}$.

En general, tenemos lo siguiente.

Teorema 2.2.1 *La matriz de transición de n pasos vendrá dada por n -ésima potencia de P . Es decir,*

$$P^n = (p_{i,j}^{(n)})_{i,j=1,2,\dots,N}.$$

Como consecuencia, se tiene la relación:

$$\pi^{(n)} = \pi^{(0)} \cdot P^n$$

donde:

$$\pi^{(0)} = (\pi_1^{(0)}, \dots, \pi_N^{(0)})$$

denota el vector de probabilidades iniciales de cada estado, es decir, $\pi_j^{(0)} = \mathbb{P}(X_0 = j)$ y

$$\pi^{(n)} = (\pi_1^{(n)}, \dots, \pi_N^{(n)})$$

denota el vector de probabilidades iniciales de cada estado en el instante n , es decir, $\pi_j^{(n)} = \mathbb{P}(X_n = j)$, para todo $j = 1, 2, \dots, N$.

Teniendo en cuenta el teorema anterior junto al hecho de que $P^{m+n} = P^m P^n$, tenemos lo siguiente.

Corolario 2.2.1 *(Ecuaciones de Chapman-Kolmogorov)*

$$p_{i,j}^{(n+m)} = \sum_{k=1}^N p_{i,k}^{(n)} p_{k,j}^{(m)}.$$

2.3) Clasificación de los estados

Sea $(X_n)_{n=0,1,2,\dots}$ una cadena con espacio de estados \mathcal{S} . Dados dos estados $x, y \in \mathcal{S}$, definimos

$$r_{x,y} = \mathbb{P}(X_n = y \text{ para algún } n \geq 1 | X_0 = x).$$

Esto es, $r_{x,y}$ es la probabilidad de que la cadena alcance el estado y (en algún tiempo futuro) si la cadena se inicia en el estado x .

Definición 2.3.1 Sean $x, y \in \mathcal{S}$ con $x \neq y$.

- 1) Decimos que y es **accesible** desde x si $r_{x,y} > 0$. En tal caso escribimos $x \rightarrow y$.
- 2) Decimos que x se **comunica** con y si son accesibles entre sí (es decir, $x \rightarrow y, y \rightarrow x$). En tal caso escribimos $x \leftrightarrow y$.

Por convenio, se considera que cualquier estado x es accesible desde sí mismo ($x \rightarrow y$), y que se comunica consigo mismo ($x \leftrightarrow x$), incluyendo el caso de que $r_{x,x} = 0$.

Cada cadena puede ser representada con un grado. Podemos ver la accesibilidad simplemente observando si en el grafo existe un camino desde x hasta y respetando la dirección de las flechas. Cuando haya caminos en ambas direcciones, significará que ambos estados se comunican.

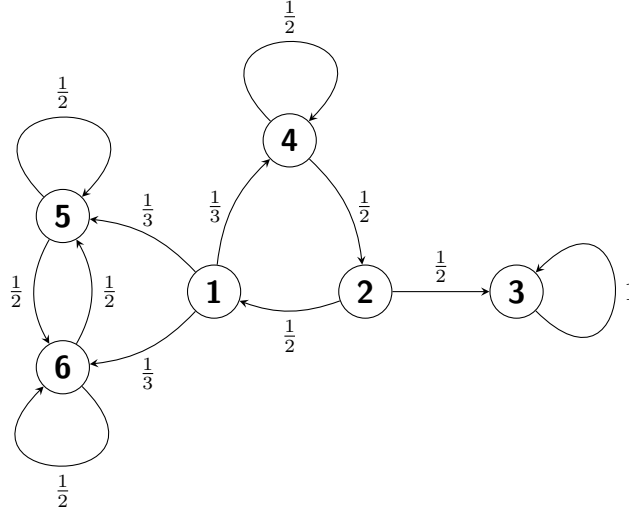


Figure 3: Observando las flechas podemos analizar la accesibilidad y la conexión entre estados

Por ejemplo, consideremos la cadena dada por el grafo de la Figura 3. Vemos que por ejemplo que $1 \rightarrow 3$ ya que podemos ir desde 1 hasta 3, siguiendo la dirección de las flechas, pasando por 4 y 2. Sin embargo, $3 \nrightarrow 1$, ya que de 3 sólo se puede volver a saltar a sí mismo. Por otro lado, $1 \leftrightarrow 2$ ya que podemos conectar ambos estados por caminos tanto empezando en 1 como empezando en 2.

En general, si un estado y es accesible desde un estado x , siempre será posible encontrar un camino desde x hasta y de modo que dicho camino no pase por un mismo estado dos veces. Para ello, basta considerar cualquier camino desde x hasta y , y si algún estado z aparece dos veces en el camino, eliminamos el tramo del camino desde la primera aparición de dicho estado hasta la última aparición del mismo, de modo que aparezca una sola vez. Dicho camino, al no pasar dos veces por un mismo estado, tendrá como mucho tantos pasos como estados tenga la cadena. En definitiva, tenemos lo siguiente.

Proposición 2.3.1 *Supongamos que el espacio de estados \mathcal{S} tiene N elementos. Entonces,*

$$x \rightarrow y \quad \text{si, y sólo si,} \quad p_{x,y}^{(n)} > 0 \quad \text{para algún } n \leq N.$$

La relación entre estados definida por la propiedad de estar comunicados es una relación de equivalencia. Es decir, tenemos lo siguiente.

Proposición 2.3.2 *Dado el espacio de estados \mathcal{S} , siempre es posible dividirlo en clases disjuntas*

$$\mathcal{S} = C_1 \cup C_2 \cup \dots \cup C_n \quad (2)$$

donde para todo $x, y \in \mathcal{S}$ se verifica

$$\begin{aligned} x \leftrightarrow y & \quad \text{si } x, y \in C_i, \\ x \nleftrightarrow y & \quad \text{si } x \in C_i, y \in C_j, i \neq j. \end{aligned}$$

Las clases en (2) se llaman *clases irreducibles*. La proposición nos dice que cada cadena admite siempre una descomposición en clases irreducibles. Una cadena se dice que es *irreducible* si sólo posee una clase irreducible, es decir, todos sus estados se comunican entre sí.

Las clases irreducibles pueden ser fácilmente identificadas mirando el grafo de la cadena. Observando de nuevo el grado de la

cadena de la Figura 3, vemos que $1 \rightarrow 5$ pero $5 \nrightarrow 1$, por lo que están en clases diferentes. Por otro lado, $1 \rightarrow 4$, $4 \rightarrow 2$, $2 \rightarrow 1$, por lo que están en la misma clase. Finalmente, 3 no está comunicado con nadie, luego forma él solo una clase. En la Figura 4, podemos ver las tres clases irreducibles de la cadena que hemos identificado.

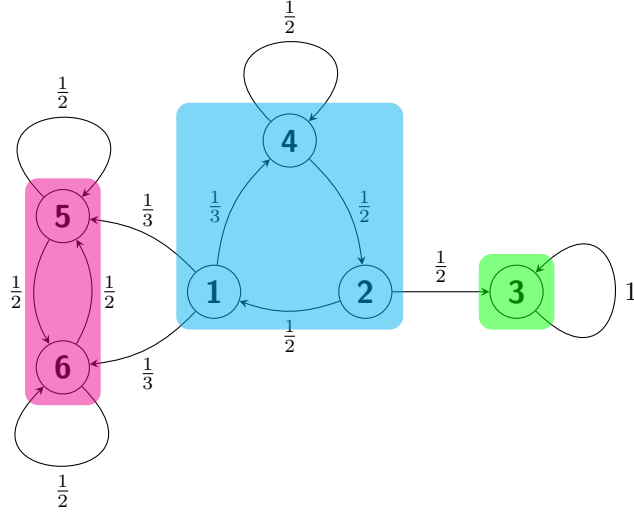


Figure 4: Tenemos en distinto color las tres clases irreducibles

A continuación vamos a introducir un criterio para clasificar los distintos estados de una cadena.

Definición 2.3.2 Sea $x \in S$.

- Decimos que x es **recurrente** si $r_{x,x} = 1$.
- Decimos que x es **transitorio** si $r_{x,x} < 1$.
- Decimos que x es **absorbente** si $p_{x,x} = 1$.

Si x es recurrente, tendremos que una vez que la cadena alcanza el estado x entonces tendremos total certeza de que volverá a alcanzar el estado x en el futuro.

Si x es transitorio, tendremos que una vez que la cadena alcanza el estado x entonces no tendremos certeza de que la cadena vuelva a alcanzar el estado x de nuevo.

Si x es absorbente, tendremos que una vez que la cadena alcanza el estado x con toda certeza permanezca en el estado x en el futuro. En particular, todo estado absorbente es también recurrente.

Mirando de nuevo al grafo de la Figura 4 arriba, vemos que el estado 3 es absorbente (y por tanto recurrente), ya que si comenzamos en él solo llegaremos al él mismo. Los estados 5 y 6 son recurrentes. Por ejemplo, vemos que si empezamos en 5 no podremos llegar a ningún otro estado excepto 6 y él mismo. Además, la única manera para que, empezando en 5, no se vuelva a visitar 5 es que la cadena se cambie al estado 6 y permanezca en ese estado en lo sucesivo. Pero la probabilidad de que eso ocurra es $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdots = 0$. Así que con probabilidad 1 se volverá en algún momento a 5, y por lo tanto es recurrente. El mismo argumento es aplicable a 6. Finalmente, los estados 1, 2 y 4 son transitorios ya que desde esos estados se accede a zonas de las que no se puede volver.

A continuación, enunciaremos, sin demostración, algunos resultados sobre el comportamiento de los estados recurrentes.

Teorema 2.3.1 Si $x \rightarrow y$ y x es recurrente, entonces y también es recurrente y, además, $r_{x,y} = 1 = r_{y,x}$.

El teorema de arriba nos dice que si y es accesible desde un estado recurrente, entonces se realizarán viajes de ida y vuelta entre x e y , una y otra vez.

Teorema 2.3.2 Toda cadena de Markov homogénea posee al menos un estado recurrente.

Como consecuencia de los dos teoremas anteriores tenemos.

Corolario 2.3.1 *Supongamos que una cadena de Markov tiene descomposición en clases irreducibles*

$$C_1 \cup C_2 \cup \dots \cup C_n.$$

Entonces si existe $x \in C_i$ el cual es recurrente (resp. transitorio), entonces todos los $y \in C_i$ son también recurrentes (resp. transitorios).

Obsérvese que, en la Figura 4 arriba, los estados 1, 2 y 4 están todos en una misma clase y son transitorios; los estados 5 y 6 están los dos en una misma clase y son recurrentes; finalmente, el estado 3 es absorbente y formará él solo una clase irreducible.

En las cadenas de Markov frecuentemente se observan patrones cíclicos, donde una sucesión de estados se recorre en un camino en forma de bucle empezando y acabando en un mismo estado. Esto da lugar al siguiente concepto.

Definición 2.3.3 *Un estado $x \in \mathcal{S}$ se dice que tiene **periodo** d si $p_{x,x}^{(n)} = 0$ para todo n que no sea divisible por d , y además d es el mayor número con esta propiedad. En el caso de que $p_{x,x}^{(n)} = 0$ para todo n (es decir, si $r_{x,x} = 0$), entonces diremos que el periodo de x es infinito. Un estado x con periodo 1 se dice que es **aperiódico**.*

En otras palabras, un estado x tiene periodo d si la cadena puede volver al estado x sólo en un número de pasos que sea múltiplo de d . En el caso de que la cadena no vuelva nunca más a x , el periodo es infinito.

Un importante propiedad es la siguiente.

Proposición 2.3.3 *Si $x \leftrightarrow y$, entonces x e y tienen el mismo periodo.*

Por definición en una cadena irreducible todos los estados se comunican, por lo que tenemos lo siguiente.

Corolario 2.3.2 *En una cadena irreducible todos los estados tienen el mismo periodo.*

Dado que todos los estados de una cadena irreducible tienen el mismo periodo, tiene sentido introducir lo siguiente.

Definición 2.3.4 *Se dice que una cadena irreducible tiene periodo d si uno de sus estados (y por tanto todos) tienen periodo d . Se dice que una cadena es aperiódica si uno de sus estados (y por tanto todos) es aperiódico.*

En el caso de cadenas no irreducibles tendremos lo siguiente.

Corolario 2.3.3 *Supongamos que una cadena de Markov tiene descomposición en clases irreducibles*

$$C_1 \cup C_2 \cup \dots \cup C_n.$$

Todos los estados pertenecientes a una misma clase C_i tienen el mismo periodo.

Veremos más adelante que conocer si una cadena irreducible es aperiódica es importante en el estudio límite de las cadenas de Markov. Para determinar si un estado es aperiódico aplicaremos un simple método que explicaremos a continuación.

Recuerda que dos números n y m son coprimos si su máximo común divisor es 1. Supongamos que podemos encontrar dos números coprimos n y m tal que $p_{x,x}^{(n)} > 0$ y $p_{x,x}^{(m)} > 0$. En este caso, por definición de periodo, d debe dividir a n y m . Como n y m son coprimos, necesariamente tendremos que $d = 1$. Por tanto, para determinar si un estado x es aperiódico, trataremos de buscar en el grafo dos caminos de x en sí mismo de modo que los números de pasos dados en sendos caminos sean coprimos.

Para ilustrar este método, volvamos al ejemplo en la Figura 4 arriba. Claramente el estado 3 es aperiódico por ser absorbente. Por otro lado, partiendo del estado 1, podremos volver a dicho estado en tres pasos siguiendo el camino $1 \rightarrow 4 \rightarrow 2 \rightarrow 1$. Pero, partiendo de 1, también podemos volver a 1 en cuatro pasos a través del camino $1 \rightarrow 4 \rightarrow 4 \rightarrow 2 \rightarrow 1$ en el que pasamos dos veces por 4. Como 3 y 4 son coprimos, necesariamente tendremos que el estado 1 es aperiódico. Además, los estados 2 y 4 también serán aperiódicos por estar en la misma clase irreducible que 1. Un argumento similar muestra que los estados 5 y 6, los cuales están en la misma clase irreducible, son aperiódicos también.

2.4) Número de visitas

En esta sección vamos a estudiar el número medio de veces que una cadena de Markov alcanza un estado determinado a lo largo del tiempo. En lo que sigue consideramos una cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados \mathcal{S} . Dado $x \in \mathcal{S}$, la probabilidad de que la cadena vuelva a x habiendo empezado en x es $r_{x,x}$. Por tanto, si empezamos en x , la probabilidad de no volver nunca más a x es $1 - r_{x,x}$. Teniendo en cuenta ambas cosas, tenemos que

$$\mathbb{P}(\text{"visitar } x \text{ exactamente } k \text{ veces"} | X_0 = x) = r_{x,x}^k (1 - r_{x,x}), \quad (3)$$

donde no estamos contando la visita a x en tiempo 0.

Comentario. Recordemos que una variable aleatoria discreta Y sigue una distribución geométrica de parámetro $p \in [0, 1]$ si

$$\mathbb{P}(Y = k) = (1 - p)^k p \quad \text{para } k = 0, 1, 2, 3, \dots$$

En este caso,

$$\mathbb{E}(Y) = \frac{1 - p}{p}.$$

Obsérvese que la anterior esperanza es infinita si $p = 0$. Para los demás valores de p tendremos valores finitos.

A partir de (3), vemos que el número de visitas a x partiendo de x sigue una distribución Geométrica de parámetro $p = 1 - r_{x,x}$. En particular, el *número esperado de visitas* a x si salimos de x es

$$\mathbb{E}(\text{"número de visitas a } x" | X_0 = x) = \frac{r_{x,x}}{1 - r_{x,x}}.$$

Además, la expresión arriba es finita si, y sólo si, $r_{x,x} < 1$, en cuyo caso x es transitorio. En caso contrario, tendremos un estado recurrente y el número esperado de visitas será infinito.

Consideremos ahora un estado y distinto de x , de modo que y es accesible desde x . Queremos calcular ahora el número medio de visitas a y empezado en x . Si partimos de x , tenemos dos posibilidades, o bien la cadena alcanza el estado y produciéndose al menos una visita, lo cual ocurre con probabilidad $r_{x,y}$; o bien no lo visita nunca, lo cual ocurre con probabilidad $1 - r_{x,y}$. De esta manera, el número esperado de visitas a y si salimos de x es

$$\begin{aligned} \mathbb{E}(\text{"número de visitas a } y" | X_0 = x) &= r_{x,y} \cdot (1 + \mathbb{E}(\text{"número de visitas a } y \text{ con } n > 1" | X_1 = y)) + (1 - r_{x,y}) \cdot 0 \\ &= r_{x,y} \left(1 + \frac{r_{y,y}}{1 - r_{y,y}} \right) \\ &= \frac{r_{x,y}}{1 - r_{y,y}}. \end{aligned}$$

De nuevo podemos observar que la esperanza anterior es finita si, y sólo si, $r_{y,y} < 1$, en cuyo caso y es transitorio.

Resumimos los resultados obtenidos en el siguiente enunciado.

Proposición 2.4.1 *Supongamos que x, y son dos estados (posiblemente iguales) de modo que $x \rightarrow y$. Entonces,*

$$\mathbb{E}(\text{"número de visitas a } y" | X_0 = x) = \frac{r_{x,y}}{1 - r_{y,y}}.$$

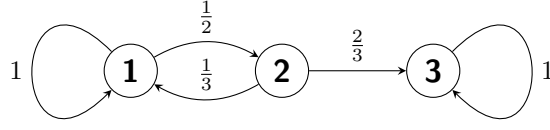
Además, la esperanza anterior es finita si y sólo si y es transitorio.

2.5) Probabilidades de absorción y tiempos medios de llegada

Cuando se estudian las cadenas de Markov, es habitual preguntarse por la probabilidad de que la cadena alcance cierto estado absorbente y también por el tiempo medio en el que alcanza dicho estado. Por ejemplo, en el caso de la ruina del jugador, podríamos plantearnos cómo de probable es alcanzar las ganancias que el jugador ha fijado antes de retirarse, cuál es la probabilidad de arruinarse, o cuánto tiempo tardarán en producirse dichos sucesos. A continuación, introduciremos los conceptos de probabilidad de absorción y tiempo medio de llegada, y veremos cómo se calculan.

2.5.1) Probabilidades de absorción

Supongamos que tenemos una cadena con todos sus estados transitorios o absorbentes. Por ejemplo, supongamos que tenemos la cadena dada el siguiente grafo:



Definimos $\tau_{1,3}$ como el número de pasos promedio necesarios para alcanzar estado absorbente 3 partiendo de 1, es decir,

$$\tau_{1,3} = \mathbb{E}[\text{"Número de pasos hasta la primera visita a 3"} | X_0 = 1].$$

Del mismo modo, podemos definir $\tau_{2,3}$ y $\tau_{3,3}$.

Obviamente $\tau_{3,3} = 0$. Para calcular $\tau_{1,3}$, tenemos que tener en cuenta que, si la cadena se encuentra en el estado 1, en el primer paso dado o bien vuelve al estado 1, o bien va al estado 2. De esa manera,

$$\tau_{1,3} = 1 + p_{1,1}\tau_{1,3} + p_{1,2}\tau_{2,3} = 1 + \frac{1}{2}\tau_{1,3} + \frac{1}{2}\tau_{2,3}.$$

Si estamos en el estado 2, daremos el primer paso o bien a 1, o bien a 3. Luego,

$$\tau_{2,3} = 1 + p_{2,1}\tau_{1,3} + p_{2,3}\tau_{3,3} = 1 + \frac{1}{3}\tau_{1,3} + \frac{2}{3}\tau_{3,3}.$$

Poniendo todo junto, obtenemos el siguiente sistema de ecuaciones

$$\begin{cases} \tau_{1,3} = 1 + \frac{1}{2}\tau_{1,3} + \frac{1}{2}\tau_{2,3}, \\ \tau_{2,3} = 1 + \frac{1}{3}\tau_{1,3} + \frac{2}{3}\tau_{3,3}, \\ \tau_{3,3} = 0 \end{cases}$$

Resolviendo el sistema anterior, obtenemos

$$\tau_{1,3} = \frac{9}{2}, \quad \tau_{2,3} = \frac{5}{2}, \quad \tau_{3,3} = 0.$$

En general, si $(X_n)_{n=0,1,2,\dots}$ una cadena con un solo estado absorbente $a \in \mathcal{S}$ y todos los demás estados transitorios, entonces los tiempos medios de llegada a $a \in \mathcal{S}$ se pueden calcular usando que:

- 1) $\tau_{a,a} = 0$.
- 2) Para cualquier estado $x \in \mathcal{S}$ distinto de a ,

$$\tau_{x,a} = 1 + \sum_{y \in \mathcal{S}} p_{x,y} \tau_{y,a}.$$

2.6) Comportamiento asintótico

Como hemos visto, en una cadena irreducible, todos los estados son recurrentes. Tenemos que es posible pasar de cualquier estado a cualquier otro. De hecho la cadena pasará infinitas veces por todos los estados con probabilidad 1. Nos preguntamos sobre cuál es la probabilidad de que a largo plazo la cadena se encuentre en un estado particular. En otras palabras, queremos saber con qué frecuencia la cadena visitará dicho estado a medida que avanzamos en el tiempo.

En lo que sigue, consideramos una cadena $(X_n)_{n=0,1,2,\dots}$ irreducible con espacio de estados $\mathcal{S} = \{1, 2, \dots, N\}$. Para cada

$n = 0, 1, 2, \dots$ y $j \in \mathcal{S}$ consideramos la probabilidad de que la cadena se encuentre en el estado j en n pasos, es decir,

$$\pi_j^{(n)} = \mathbb{P}(X_n = j).$$

De hecho, podemos considerar el vector

$$\pi^{(n)} = \left(\pi_1^{(n)}, \pi_2^{(n)}, \dots, \pi_N^{(n)} \right).$$

Al considerar la probabilidad de n pasos para todos los posibles estados, el vector $\pi^{(n)}$ verificará además que sus componentes suman 1, es decir, $\sum_{j=1}^{\infty} \pi_j^{(n)} = 1$.

Dado un estado $j \in \mathcal{S}$, teniendo en cuenta la probabilidad de empezar en cada uno de los estados, tendremos que

$$\begin{aligned} \pi_j^{(n)} &= \sum_{i=1}^N \mathbb{P}(X_0 = i) \cdot \mathbb{P}(X_n = j | X_0 = i) \\ &= \sum_{i=1}^{\infty} \pi_i^0 \cdot p_{i,j}^{(n)} \end{aligned}$$

Lo cual conduce a la siguiente ecuación matricial

$$\pi^{(n)} = \pi^{(0)} \cdot P^n. \quad (4)$$

Supongamos que existe el límite $\lim_{n \rightarrow \infty} \pi^{(n)} = \pi$. El vector $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ se llama **distribución límite** de la cadena de Markov. Puesto que $\sum_{j=1}^N \pi_j^{(n)} = 1$, la distribución límite necesariamente verificará

$$\sum_{j=1}^N \pi_j = 1. \quad (5)$$

Por otro lado, usando (4), tenemos

$$\begin{aligned} \pi &= \lim_{n \rightarrow \infty} \pi^{(n+1)} \\ &= \lim_{n \rightarrow \infty} \left(\pi^{(0)} \cdot P^{n+1} \right) \\ &= \left(\lim_{n \rightarrow \infty} \pi^{(0)} \cdot P^n \right) \cdot P \\ &= \left(\lim_{n \rightarrow \infty} \pi^{(n)} \right) \cdot P \\ &= \pi \cdot P. \end{aligned}$$

La distribución límite π cumple la ecuación matricial

$$\pi = \pi \cdot P \quad (6)$$

Podemos explicar esta ecuación de forma intuitiva como sigue. Imaginemos que para cierto tiempo n , tenemos que X_n sigue la distribución π . Entonces X_{n+1} sigue la distribución πP . Si $\pi = \pi P$, entonces X_{n+1} sigue también la distribución π . Por lo que la cadena de Markov habrá alcanzado una distribución estable en el tiempo. A una distribución π verificando (6) se llama **distribución estacionaria** de la cadena de Markov.

Hemos argumentado arriba que la distribución límite, si existe, debe cumplir las ecuaciones (5) y (6). Por otra parte, es posible demostrar que si una cadena irreducible es aperiódica entonces la distribución límite existe. Por tanto, tenemos lo siguiente.

Teorema 2.6.1 *Supongamos que $(X_n)_{n=0,1,2,\dots}$ es una cadena irreducible aperiódica. Entonces, la distribución límite existe*

y está determinada por las ecuaciones

$$\begin{cases} \pi = \pi P, \\ \sum_{j=1}^N \pi_j = 1. \end{cases}$$

De esta manera, para encontrar la distribución límite de una cadena, bastará con comprobar que la cadena es irreducible y aperiódica, lo cual garantiza la existencia de una distribución límite, y luego resolver las ecuaciones anteriores para determinarla.

Ejemplo. Consideremos una cadena de Markov con sólo dos estados $\mathcal{S} = \{1, 2\}$, y con matriz de transición

$$P = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Esta cadena es irreducible y aperiódica ya que todos los estados están comunicados en un solo paso. Por tanto, existirá la distribución límite $\pi = (\pi_1, \pi_2)$, la cual será también una distribución estacionaria. Para hallar dicha distribución, hemos de resolver la ecuación matricial

$$\begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix},$$

junto con la condición $\pi_1 + \pi_2 = 1$. Esto nos conduce al sistema de ecuaciones:

$$\begin{cases} \frac{1}{3}\pi_1 - \frac{1}{2}\pi_2 = 0, \\ -\frac{1}{3}\pi_1 + \frac{1}{2}\pi_2 = 0, \\ \pi_1 + \pi_2 = 1. \end{cases}$$

Resolviendo obtenemos la distribución límite

$$\pi_1 = \frac{3}{5} = 0.6, \quad \pi_2 = \frac{2}{5} = 0.4.$$

Dicha distribución límite es posible aproximarla por "fuerza bruta" tomando potencias sucesivas de la matriz de transición. Lo haremos con ayuda de R:

```
P <- rbind(c(2/3, 1/3),
           c(1/2, 1/2))
P2 <- P%*%P
P3 <- P2%*%P
P4 <- P3%*%P
P5 <- P4%*%P
P6 <- P5%*%P
P7 <- P6%*%P
P8 <- P7%*%P
P9 <- P8%*%P
P10 <- P9%*%P
P
##           [,1]      [,2]
## [1,] 0.6666667 0.3333333
## [2,] 0.5000000 0.5000000
P2
```

```

##          [,1]      [,2]
## [1,] 0.6111111 0.3888889
## [2,] 0.5833333 0.4166667

P3

##          [,1]      [,2]
## [1,] 0.6018519 0.3981481
## [2,] 0.5972222 0.4027778

P4

##          [,1]      [,2]
## [1,] 0.6003086 0.3996914
## [2,] 0.5995370 0.4004630

P5

##          [,1]      [,2]
## [1,] 0.6000514 0.3999486
## [2,] 0.5999228 0.4000772

P6

##          [,1]      [,2]
## [1,] 0.6000086 0.3999914
## [2,] 0.5999871 0.4000129

P7

##          [,1]      [,2]
## [1,] 0.6000014 0.3999986
## [2,] 0.5999979 0.4000021

P8

##          [,1]      [,2]
## [1,] 0.6000002 0.3999998
## [2,] 0.5999996 0.4000004

P9

##          [,1]      [,2]
## [1,] 0.6000000 0.4000000
## [2,] 0.5999999 0.4000001

P10

##          [,1] [,2]
## [1,]    0.6  0.4
## [2,]    0.6  0.4

```

Vemos que al tomar sucesivas potencias de la matriz de transición, el valor de la matriz se va estabilizando y sus filas se van aproximando a la distribución estacionaria π que hemos obtenido anteriormente. Recordemos que la potencia n -ésima de la matriz de transición corresponde a la matriz de transición de n pasos. Luego es lógico que al ir tomando potencias de la matriz de transición, obtengamos valores cada vez más cercanos a la distribución límite. Vemos también que la probabilidad límite no depende del valor inicial de la cadena, ya que todas las filas convergen a los mismos valores.

2.7) Algoritmo PageRank

Volvamos al ejemplo que vimos al principio del tema donde analizamos la cadena de Markov del algoritmo PageRank de Google. Recordemos que dicha cadena describía a un "surfeador de internet" que pulsaba al azar los links de un conjunto de páginas. El algoritmo PageRank se usa para ordenar la aparición de las páginas en cada búsqueda en Google, dando preferencia a aquellas páginas cuya probabilidad sea mayor en la distribución límite. Es decir, aquellas con mayor probabilidad serán las que sean más visitadas a largo plazo.

En nuestro ejemplo, el "surfeador de la web" navega al azar entre las páginas A, B, C y D, donde:

- A tiene enlace a B
- B tiene enlaces a A y C
- C tiene enlace a A
- D tiene enlaces a las otras tres páginas.

De esta manera, el espacio de estados es $\{A, B, C, D\}$, y la matriz de transición será

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 1 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}$$

Como hemos dicho, estamos interesados en determinar la distribución de la cadena de Markov definida por el surfeador. El problema es que dicha cadena podría no ser irreducible, por lo que la distribución límite podría no existir. De hecho, en nuestro ejemplo en particular, la cadena no es irreducible ya que, si bien todas las páginas son accesibles desde D, dicha página no es accesible desde el resto de las páginas.

Para solucionar este problema, el algoritmo de PageRank supone que, en cualquier caso, el surfeador de la web se puede "aburrir", y empezar a navegar de nuevo en cualquiera de las páginas del conjunto al azar. La probabilidad con la que, tras cada paso, el surfeador se aburre es un número fijo p . A la probabilidad complementaria $d = 1 - p$ se le llama factor de amortiguamiento (damping factor). Por regla general se suele tomar factor de amortiguamiento $d = 0.85$.

En nuestro ejemplo particular, si el surfeador se encuentra, por ejemplo, en la página B podría seguir navegando e ir a cualquiera de las páginas a las que tiene enlace, es decir, A o C, o podría aburrirse y salta a cualquiera de las páginas del conjunto A, B, C o D.

La cadena de Markov así obtenida es irreducible ya que desde cada página se puede pasar a cualquier otra página ya que existe la posibilidad de aburrirse. Además, será aperiódica ya que dichos saltos pueden darse en un solo paso. Al ser una cadena irreducible y aperiódica tendrá una distribución límite.

Desde un punto de vista más práctico, añadir la posibilidad de que el surfeador se pueda aburrir, hace que éste no pueda quedarse "atrapado" en un grupo de páginas particular conectadas entre sí, pero no a las demás, haciendo que el algoritmo desestime otras páginas que también podrían ser importantes.

En general, supongamos que el surfeador navega al azar entre n páginas con matriz de transición P . Entonces, al considerar un factor de amortiguación d obtenemos una nueva matriz de transición la cual viene dada por:

$$M = d \cdot P + (1 - d) \cdot \frac{1}{n} J_n, \quad (7)$$

donde J_n es la matriz cuadrada de tamaño n con todas sus entradas igual a 1. La nueva matriz de transición M representa que, en cada paso, con probabilidad d el surfeador actuará de acuerdo a la cadena con matriz de transición P , y con probabilidad $1 - d$ el surfeador actuará de acuerdo a la cadena consistente en elegir cualquiera de las n páginas al azar, cuya matriz de

transición es $\frac{1}{n}J_n$.

En nuestro ejemplo, tomando $d = 0.85$, calculamos con ayuda de R la matriz de transición:

```
P <- rbind(c(0, 1, 0, 0),
           c(1/2, 0, 1/2, 0),
           c(1, 0, 0, 0),
           c(1/3, 1/3, 1/3, 0))
d <- 0.85
n <- 4
M <- d*P + (1-d)*(1/n)*matrix(1, n, n)
```

Obtenemos

$$M = \begin{bmatrix} 0.0375 & 0.8875 & 0.0375 & 0.0375 \\ 0.4625 & 0.0375 & 0.4625 & 0.0375 \\ 0.8875 & 0.0375 & 0.0375 & 0.0375 \\ 0.3208333 & 0.3208333 & 0.3208333 & 0.0375 \end{bmatrix}$$

Sabemos que la cadena dada por la anterior matriz de transición posee distribución límite. Podemos calcular su valor exacto resolviendo las ecuaciones correspondientes. Sin embargo, por comodidad aproximaremos la distribución límite mediante el cálculo de potencias de la matriz de transición. Para ello podemos implementar el siguiente método iterativo en R:

```
M_nueva <- M

# Fijamos un error máximo
error_max <- 10^-15

# Inicializamos la variable de error
error <- 1

while (error > error_max) {
  M_nueva <- M_nueva%*%M
  error <- max(abs(M_nueva - M_nueva%*%M))
}

# La solución es
M_nueva[1,]

## [1] 0.3824972 0.3732476 0.2067552 0.0375000
```

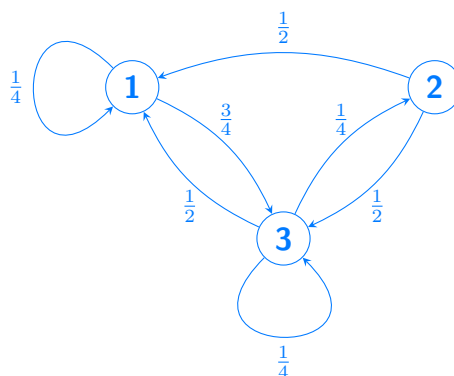
Obtenemos:

$$\pi_A = 0.3824972, \quad \pi_B = 0.3732476, \quad \pi_C = 0.2067552, \quad \pi_D = 0.0375.$$

Esto quiere decir que, a largo plazo, el surfearor gastará el 38.25% del tiempo en la página A; 37.32% del tiempo en la página B; 20.68% del tiempo en la página C; y el 3.75% del tiempo en la página D. Por tanto, el algoritmo PageRank ordenará las páginas en función de estas probabilidades, mostrando las páginas en el orden: A, B, C, D.

Hoja 2: Problemas de Cadenas de Markov

- 1) Considera la cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados $\mathcal{S} = \{1, 2, 3\}$ dada por el grafo



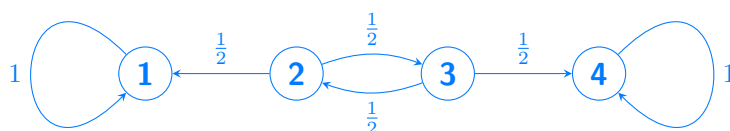
Supongamos que $\mathbb{P}(X_1 = 1) = \frac{1}{2}$ y que $\mathbb{P}(X_1 = 2) = \frac{1}{4}$.

- Encuentra la matriz de transición.
 - Determina las clases irreducibles de la cadena y clasifica los estados.
 - Determina el periodo de cada estado de la cadena.
 - Calcula $\mathbb{P}(X_1 = 3, X_2 = 2, X_3 = 1)$.
 - Calcula $\mathbb{P}(X_1 = 3, X_3 = 1)$
 - Supongamos que las probabilidades iniciales de los estados 2 y 3 son iguales, mientras que la del estado 1 es el doble de la del estado 3. Determina cuál será la probabilidad de cada estado en dos pasos, es decir, para $n = 2$.
- 2) Considera la cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados $\mathcal{S} = \{1, 2, 3\}$ y matriz de transición

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

- Dibuja el grafo de dicha cadena.
 - Determina las clases irreducibles de la cadena, clasifica los estados y calcula su periodo.
 - Sabiendo que

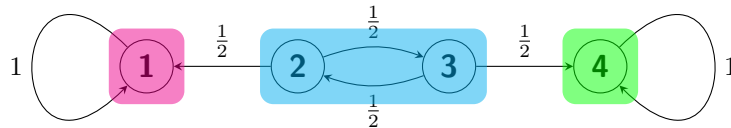
$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 2) = \frac{1}{4}$$
 encuentra $\mathbb{P}(X_1 = 3, X_2 = 2, X_3 = 1)$.
 - Sabiendo que la cadena se inició en el estado 1, determina la probabilidad de que vuelva a ese estado en 3 pasos.
- 3) Considera la cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados $\mathcal{S} = \{1, 2, 3, 4\}$ y grafo



a) Encuentra la matriz de transición.

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

b) Determina las clases irreducibles de la cadena y clasifica sus estados.



$$\underbrace{C_1 = \{1\}}_{\text{absorvente}}, \underbrace{C_2 = \{2, 3\}}_{\text{transitorios}}, \underbrace{C_3 = \{4\}}_{\text{absorvente}}$$

c) Calcula la probabilidad de ser absorbido por el estado 4 si la cadena se encuentra en el estado 2.

$$\left. \begin{aligned} r_{2,4} &= \frac{1}{2} r_{3,4} \\ r_{3,4} &= \frac{1}{2} + \frac{1}{2} r_{2,4} \end{aligned} \right\} \rightarrow r_{3,4} = \frac{1}{2} + \frac{1}{2} \left(\frac{1}{2} r_{3,4} \right) \longleftrightarrow r_{3,4} = \frac{1}{2} + \frac{1}{4} r_{3,4} \longleftrightarrow \frac{3}{4} r_{3,4} = \frac{1}{2} \longleftrightarrow r_{3,4} = \frac{2}{3}$$

$$r_{2,4} = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

d) Calcula la probabilidad de ser absorbido por el estado 4 si la cadena se encuentra en el estado 3.

$$r_{3,4} = \frac{1}{2} + \frac{1}{2} r_{2,4} = \frac{2}{3}$$

e) Calcula el número esperado de visitas al estado 2 si la cadena se encuentra en el estado 2.

$$\text{Me piden } \mathbb{E}(\text{"Nº visitas a '2'" } | X_0 = 2) = \frac{r_{2,2}}{1 - r_{2,2}} = \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3}$$

$$r_{2,2} = p_{2,1} \cdot r_{1,2} + p_{2,3} \cdot r_{3,2} = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot r_{3,2} = \frac{1}{2} r_{3,2} = \frac{1}{4}$$

$$r_{3,2} = p_{3,2} + p_{3,4} \cdot r_{4,2} = \frac{1}{2}$$

f) Calcula el número esperado de visitas al estado 2 si la cadena se encuentra en el estado 3.

$$\text{Me piden } \mathbb{E}(\text{"Nº visitas a '2'" } | X_0 = 3) = \frac{r_{3,2}}{1 - r_{2,2}} = \frac{\frac{1}{2}}{1 - \frac{1}{4}} = \frac{2}{3}$$

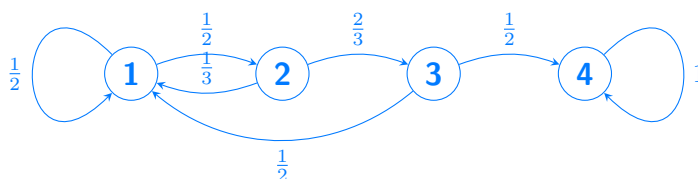
4) Considera la cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados $\mathcal{S} = \{1, 2, 3, 4, 5, 6, 7\}$ dada por la matriz de transición

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

- a) Dibuja el grafo.
 - b) Encuentra las clases irreducibles y clasifica los estados.
 - c) Observa que una vez que la cadena alcanza cualquiera de los estados del conjunto $A = \{1, 2\}$ la cadena ya no vuelve a salir más de dicho conjunto. Si la cadena se encuentra en el estado 3, encuentra la probabilidad de ser absorbido por el conjunto A .
- 5) Considera la cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados $\mathcal{S} = \{1, 2, 3, 4, 5, 6, 7\}$ dada por la matriz de transición

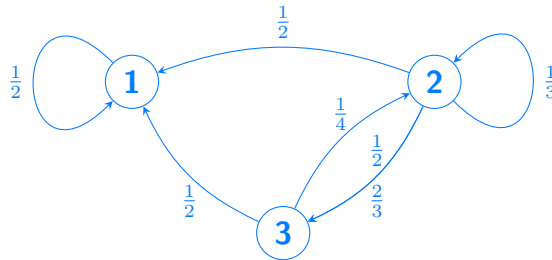
$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

- a) Dibuja el grafo.
 - b) Encuentra las clases irreducibles y clasifica los estados.
 - c) Observa que una vez que la cadena alcanza cualquiera de los estados del conjunto $A = \{1, 2\}$ la cadena ya no vuelve a salir más de dicho conjunto. Si la cadena se encuentra en el estado 4, encuentra la probabilidad de ser absorbido por el conjunto A .
- 6) Considera la cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados $\mathcal{S} = \{1, 2, 3, 4\}$



$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- a) Calcula el número esperado de visitas al estado 1 si la cadena se encuentra en el estado 2.
- b) Calcula la probabilidad de ser absorbido por estado 4 si la cadena se encuentra en estado 2.
- c) Suponiendo que la cadena se encuentra en el estado 2, calcula el número esperado de pasos para ser absorbido por el estado 4.
- 7) Considera la cadena del problema 5. Las clases irreducibles $A = \{1, 2\}$ y $B = \{5, 6, 7\}$ son cerradas. Recuerda que una clase es cerrada si una vez que la cadena visita uno de sus estados no vuelve a salir más de la clase. Suponiendo que la cadena se encuentra en el estado 4, calcula el número esperado de pasos hasta que la cadena alcance algún estado de A o B .
- 8) Considera la cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados $\mathcal{S} = \{1, 2, 3\}$ y grafo



- a) ¿Es esta cadena irreducible?

Si. Desde cualquier estado se puede llegar a cualquier otro (p.ej, $2 \rightarrow 3 \rightarrow 1, 1 \rightarrow 2 \rightarrow 3$, etc.)

- b) ¿Es esta cadena aperiódica?

Si. La cadena es irreducible y existe $P_{11} > 0$ (también $P_{22} > 0$), por lo que todos los estados tienen periodo 1.

- c) Encuentra la distribución estacionaria.

$$1) \pi = \pi \cdot P$$

$$2) \sum_{j=1}^N \pi_j = 1$$

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

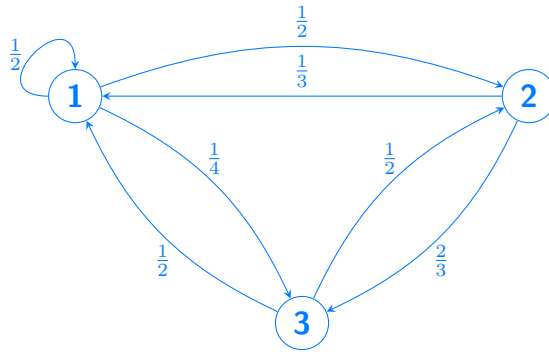
$$\pi = (\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3) \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \longleftrightarrow \pi = \left(\frac{2}{7}, \frac{3}{7}, \frac{2}{7} \right)$$

$$\begin{cases} \pi_1 = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_3 \\ \pi_2 = \frac{1}{2}\pi_1 + \frac{1}{3}\pi_2 + \frac{1}{2}\pi_3 \\ \pi_3 = \frac{2}{3}\pi_2 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \longrightarrow \begin{cases} \pi_1 - \frac{1}{2}\pi_1 = \frac{1}{2}\pi_3 \implies \frac{1}{2}\pi_1 = \frac{1}{2}\pi_3 \implies \pi_1 = \pi_3 \\ \pi_3 = \frac{2}{3}\pi_2 = \frac{2}{3} \cdot \frac{3}{7} = \frac{2}{7} \\ \pi_1 + \pi_2 + \pi_3 = 1 \rightarrow \frac{2}{3}\pi_2 + \pi_2 + \frac{2}{3}\pi_2 = \frac{7}{3}\pi_2 = 1 \rightarrow \pi_2 = \frac{3}{7} \end{cases}$$

d) ¿Es la distribución estacionaria también la distribución límite?

Si. Al ser finita, irreducible y aperiódica, existe distribución límite única y coincide con π , independientemente de la distribución inicial.

9) Considera la cadena $(X_n)_{n=0,1,2,\dots}$ con espacio de estados $\mathcal{S} = \{1, 2, 3\}$ y grafo



a) ¿Es esta cadena irreducible?

b) ¿Es esta cadena aperiódica?

c) Encuentra la distribución estacionaria.

d) ¿Es la distribución estacionaria también la distribución límite?

Tema 3: Procesos de Poisson

En este capítulo vamos a ver un tipo de procesos estocásticos que son utilizados para contar la ocurrencia en el tiempo o el espacio de cierto evento aleatorio. Se trata de los procesos de Poisson y, más en general, los procesos de conteo. Este tipo de procesos pueden usarse para contar las llegadas de clientes a una tienda, apariciones de defectos en una línea de producción, las solicitudes individuales a un servidor de internet, el paso de coches por un tramo de carretera, las llegadas de emails a la bandeja de entrada, las ocurrencias de cierto tipo de delito en un gran ciudad, etc.

3.1) Procesos de conteo

Definición 3.1.1 Un proceso estocástico en tiempo continuo $(N_t)_{t \in [0, \infty)}$ se dice que es un **proceso de conteo** si se verifica:

- 1) $N_t \geq 0$.
- 2) N_t toma valores enteros.
- 3) Si $s < t$ entonces $N_s \leq N_t$.

En particular, son procesos estocásticos de tiempo continuo y estado discreto. El valor N_t se interpreta como el número de veces que ocurre el evento aleatorio en estudio durante el intervalo de tiempo $[0, t]$. En particular, si $s < t$, entonces $N_t - N_s$ representa el número de veces que ocurre el evento aleatorio durante el intervalo de tiempo $(s, t]$. Para fijar ideas, vamos a considerar que estamos contando las llegadas de clientes a un establecimiento. Por ejemplo, si $N_1 = 4$ tendremos que han llegado 3 clientes durante el intervalo de tiempo $(1, 2]$. El tiempo podría medirse por ejemplo en horas.

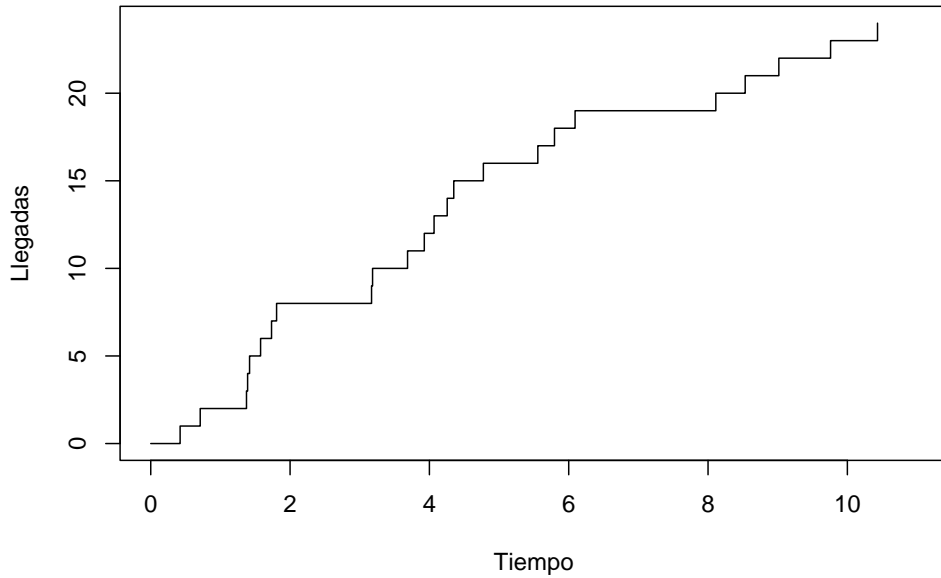


Figure 5: Trayectoria de un proceso de conteo. En cada tiempo t , N_t representa el número de llegadas ocurridas en el intervalo $[0, t]$

Definición 3.1.2 Un proceso estocástico $(X_t)_{t \in [0, \infty)}$ se dice que tiene **incrementos independientes** si para toda sucesión de tiempo $0 \leq t_1 < t_2 < \dots < t_n$ las variables aleatorias

$$X_{t_2} - X_{t_1}, \quad X_{t_3} - X_{t_2}, \quad \dots, \quad X_{t_n} - X_{t_{n-1}},$$

son independientes.

Un ejemplo de proceso con incrementos independientes es el movimiento Browniano. Para un proceso de conteo $(N_t)_{t \in [0, \infty)}$ con incrementos independientes contando las llegadas de clientes (o cualquier otro evento aleatorio), se tendrá que los números

de llegadas en intervalos disjuntos

$$(t_1, t_2], \quad (t_2, t_3], \quad \dots, \quad (t_{n-1}, t_n]$$

son independientes.

Supongamos que queremos calcular la probabilidad de que haya 2 llegadas en el intervalo $(1, 2]$ y 3 llegadas en el intervalo $(2, 3]$. Como los intervalos $(1, 2]$ y $(2, 3]$ son disjuntos, tendremos

$$\mathbb{P}(2 \text{ llegadas en } (1, 2] \text{ y } 3 \text{ llegadas en } (2, 3]) = \mathbb{P}(2 \text{ llegadas en } (1, 2]) \cdot \mathbb{P}(3 \text{ llegadas en } (2, 3])$$

3.2) Proceso de Poisson

El proceso de conteo más usado es el proceso de Poisson. En esta sección veremos su definición y principales propiedades. Antes de introducir este concepto, necesitamos recordar la distribución de Poisson.

3.2.1) Distribución de Poisson

Definición 3.2.1 Una variable aleatoria discreta X se dice que tiene distribución de **Poisson** de parámetro $\lambda > 0$ si toma valores en $\{0, 1, 2, 3, \dots\}$ y tiene función puntual de probabilidad

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{para todo } k \in \{0, 1, 2, \dots\}.$$

En este caso escribimos $X \sim \text{Poisson}(\lambda)$.

Algunas propiedades de la distribución de Poisson que conviene recordar son las siguientes.

Proposición 3.2.1 Supongamos que $X \sim \text{Poisson}(\lambda)$. Entonces $\mathbb{E}(X) = \lambda$ y $\text{Var}(X) = \lambda$.

La suma de variables Poisson independientes es Poisson. Más concretamente, tenemos lo siguiente.

Proposición 3.2.2 Si X_1, X_2, \dots, X_n son variables independientes tal que $X_i \sim \text{Poisson}(\lambda_i)$. Entonces,

$$X_1 + X_2 + \dots + X_n \sim \text{Poisson}(\lambda_1 + \lambda_2 + \dots + \lambda_n).$$

Las variables Poisson pueden ser interpretadas como el límite en distribución de variables aleatorias binomiales. Más específicamente tenemos lo siguiente.

Proposición 3.2.3 Consideremos una sucesión de variables aleatorias Y_1, Y_2, \dots , donde $Y_n \sim \text{Binomial}(n, p_n)$ con $p_n = \frac{\lambda}{n}$ para cierto $\lambda > 0$. Entonces la función puntual de probabilidad de Y_n converge a la función puntual de probabilidad de una variable Poisson(λ), es decir,

$$\lim_{n \rightarrow \infty} \mathbb{P}(Y_n = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{para todo } k \in \{0, 1, 2, \dots\}.$$

Demostración. Fijado $k \in \{0, 1, 2, \dots\}$, como $Y_n \sim \text{Binomial}(n, p_n)$, tenemos

$$\begin{aligned} \mathbb{P}(Y_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{n!}{(n-k)!n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}. \end{aligned} \tag{8}$$

Por otro lado, observamos que

$$\begin{aligned}\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda}, \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} &= 1, \\ \lim_{n \rightarrow \infty} \frac{n!}{(n-k)!n^k} &= \lim_{n \rightarrow \infty} \left(\frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \dots \cdot \frac{n-(k-1)}{n}\right) \\ &= \lim_{n \rightarrow \infty} \left(1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right)\right) = 1,\end{aligned}$$

donde para calcular el último límites hemos usado que tenemos un número constante k de factores que convergen a 1.

Tomando límites cuando $n \rightarrow \infty$ en (8) teniendo en cuenta las anteriores igualdades, obtenemos el resultado deseado.

3.2.2) Definición de proceso de Poisson

El proceso de Poisson se usa para contar la ocurrencia de un evento aleatorio cuyas observaciones se dan en el tiempo con una frecuencia promedio constante λ . Por ejemplo, supongamos que mediante la observación se llega a la conclusión de que a un establecimiento llegan, de media, 5 clientes por hora. Puesto que las llegadas son aleatorias, cada hora llegará un número de clientes que no tiene por qué ser exactamente igual a 5; sin embargo, dicha cifra se verificará en promedio al avanzar el tiempo.

Definición 3.2.2 Un proceso de conteo $(N_t)_{t \in [0, \infty)}$ se dice que es un **proceso de Poisson** con tasa $\lambda > 0$ si se verifican las siguientes condiciones:

- 1) $N_0 = 0$.
- 2) N_t tiene incrementos independientes.
- 3) Para cada $t, s \geq 0$ se verifica que $N_{t+s} - N_s \sim \text{Poisson}(\lambda t)$.

Para justificar la definición anterior pensemos en los clientes que llegan al establecimiento. Supongamos que llegan de media λ clientes por hora. Sea N_t el número de clientes que han llegado a la tienda tras t horas. Queremos encontrar qué distribución debería seguir la variable aleatoria N_t .

Para simplificar el problema, vamos a suponer en primer lugar que, dado $n \in \mathbb{N}$, los clientes van llegando en instantes de tiempo equiespaciados

$$0 < \frac{t}{n} < \frac{2t}{n} < \frac{3t}{n} < \dots < t,$$

de forma que al final de cada intervalo $\left(\frac{kt}{n}, \frac{(k+1)t}{n}\right]$ podrá llegar un solo cliente con probabilidad p , o no llegará ninguno con probabilidad $1 - p$. Además, consideramos que la llegada o no de un cliente en cada intervalo ocurre de forma independiente de lo que haya pasado en los intervalos anteriores.

Bajo estos supuestos, tenemos que el número total N_t de clientes llegados en el intervalo de tiempo $[0, t]$ sigue una distribución Binomial(n, p). En particular, como llegan de media λ clientes por hora, tendremos que

$$\lambda t = \mathbb{E}(N_t) = np.$$

Necesariamente, se tiene que verificar que $p = \frac{\lambda t}{n}$. De esta manera obtenemos la probabilidad de llegada de un cliente en cada intervalo, la cual depende del número de subdivisiones n en el que dividimos el intervalo de tiempo.

Para pasar a tiempo continuo, tomemos límites cuando $n \rightarrow \infty$. De este modo, como consecuencia de la Proposición 3.2.3, obtenemos que

$$\mathbb{P}(N_t = k) = \frac{(t + \lambda)^k e^{-\lambda t}}{k!}, \quad k \in \{0, 1, 2, \dots\}$$

Concluimos que N_t sigue una distribución Poisson(λt), lo cual justifica el punto 3 de la definición de Proceso de Poisson.

Es lógico considerar que los clientes que llegan en dos intervalos de tiempo disjuntos sean independientes, por lo que el proceso de conteo correspondiente debería tener incrementos independientes. Por lo cual, la condición 2 de la definición anterior es natural.

Como hemos observado arriba tenemos que $\mathbb{E}(N_t) = \lambda t$, lo cual es también consecuencia de la condición 3. En particular, la tasa λ de un proceso de Poisson nos informa del número medio de llegadas por unidad de tiempo.

Ejemplo. A una gasolinera llegan de media 10 coches por hora. Supongamos que la llegada de coches a la gasolinera sigue un proceso de Poisson. Calcula

- 1) Probabilidad de que lleguen 2 coches entre las 9:00 y las 9:20.
- 2) Probabilidad de que lleguen 3 coches entre las 9:00 y las 9:20 y 7 coches entre las 10:20 y las 11:00.

Solución. Consideramos el proceso $(N_t)_{t \geq 0}$ de conteo de coches tras t horas. Como llegan una media de 10 coches por hora, tendremos que $\lambda = 10$.

- 1) Sea X = "Num. de coches entre las 9:00 y las 9:20". Debido a la condición 3 de la definición de proceso de Poisson, tenemos que $X \sim \text{Poisson}(10/3)$ ya que entre las 9:00 y las 9:20 ha pasado $\frac{1}{3}$ de hora. Por tanto,

$$\mathbb{P}(X = 2) = \frac{e^{-\frac{10}{3}} \cdot \left(\frac{10}{3}\right)^2}{2!} = 0.1982.$$

Alternativamente, podemos usar el siguiente código en R para calcular el resultado anterior mediante la función puntual de probabilidad de la Poisson:

```
t <- 1/3
lambda <- 10
# Evaluamos la función puntual de probabilidad de la Poisson
dpois(2, t*lambda)
## [1] 0.1981889
```

- 2) Sea Y = "Núm. de coches entre las 10:20 y las 11:00". Tenemos que $Y \sim \text{Poisson}(20/3)$ y, además, X, Y son independientes. Entonces,

$$\begin{aligned} \mathbb{P}(X = 3, Y = 7) &= \mathbb{P}(X = 3)\mathbb{P}(Y = 7) \\ &= \frac{e^{-\frac{10}{3}} \cdot \left(\frac{10}{3}\right)^3}{3!} \frac{e^{-\frac{20}{3}} \cdot \left(\frac{20}{3}\right)^7}{7!} = 0.0325 \end{aligned}$$

Alternativamente, podemos usar el siguiente código de R para calcular el resultado anterior:

```
t <- 1/3
s <- 2/3
lambda <- 10
# Evaluamos la función puntual de probabilidad de la Poisson
dpois(3, t*lambda)*dpois(7, s*lambda)
## [1] 0.03254399
```

3.3) Tiempos de ocurrencia (llegada) en un Proceso de Poisson

Hemos visto que los procesos de Poisson surgen de manera natural cuando se trata de contar las llegadas de clientes a un establecimiento. En esta sección vamos a estudiar qué distribución de probabilidad sigue el tiempo aleatorio que pasa entre la llegada de dos clientes consecutivos.

3.3.1) Distribución exponencial

Recordemos que una variable aleatoria continua τ tiene distribución exponencial de parámetro $\lambda > 0$, en cuyo caso escribimos $\tau \sim \text{Exp}(\lambda)$, si toma valores reales positivos y tiene función de densidad

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0.$$

Recordemos además que si $\tau \sim \text{Exp}(\lambda)$, entonces la función de distribución de τ es $F(t) = 1 - e^{-\lambda t}$. En particular,

$$\mathbb{P}(\tau > t) = 1 - F(t) = e^{-\lambda t}.$$

Las variables exponenciales suelen interpretarse como el tiempo de espera para la observación de un evento aleatorio. Por ejemplo, la llegada del primer cliente a una tienda, el tiempo de fallo de un sistema, el tiempo de quiebra de una entidad financiera, tiempo en el que una acción alcanza un determinado nivel, etc.

Proposición 3.3.1 Si $\tau \sim \text{Exp}(\lambda)$, entonces

$$\mathbb{E}(\tau) = \frac{1}{\lambda}, \quad \text{Var}(\tau) = \frac{1}{\lambda^2}.$$

Si pensamos en la llegada del primer cliente a un establecimiento, la proposición anterior nos dice que el tiempo medio que esperaremos en observar llegar dicho cliente será $\frac{1}{\lambda}$.

Por otro lado, tendremos que $\mathbb{P}(\tau \leq t)$ es la probabilidad que el primer cliente llegue en algún instante del intervalo de tiempo $[0, t]$. Por tanto, $\mathbb{P}(\tau > t)$ es la probabilidad de que ningún cliente llegue durante el intervalo de tiempo $[0, t]$.

Una importante propiedad de las variables exponenciales es que no tienen memoria, es decir, tenemos lo siguiente.

Proposición 3.3.2 Si $\tau \sim \text{Exp}(\lambda)$, entonces

$$\mathbb{P}(\tau > s) = \mathbb{P}(\tau > t + s | \tau > t)$$

para todo $t, s > 0$.

La anterior propiedad, conocida como **falta de memoria** de la distribución exponencial, quiere decir que la probabilidad de que en el intervalo de tiempo $[0, s]$ no haya ninguna llegada, es igual a la probabilidad de que en el intervalo de tiempo $[t, t + s]$ de igual longitud no haya ninguna llegada si en el intervalo de tiempo previo $[0, t]$ no ha habido llegadas, es decir,

$$\mathbb{P}(\text{no hay llegadas en } [0, s]) = \mathbb{P}(\text{no hay llegadas en } [t, t + s] | \text{no hay llegadas en } [0, t]).$$

Por ejemplo, si en 30 minutos no ha llegado ningún cliente a la tienda, la probabilidad de esperar 10 minutos más y no ver ningún cliente es la misma que si no hubiéramos esperado los 30 minutos iniciales.

3.3.2) Modelización de tiempos de llegada y entre llegadas consecutivas

El siguiente resultado, para el cual no damos la demostración, caracteriza a los procesos de Poisson en términos de tiempos de espera exponenciales.

Teorema 3.3.1 Un proceso estocástico $(N_t)_{t \in [0, \infty)}$ es un proceso de Poisson con tasa λ si y sólo si existe una sucesión $\tau_1, \tau_2, \tau_3, \dots$ de variables aleatorias independientes con distribución $\text{Exp}(\lambda)$ tal que

$$N_t = \max\{n : \tau_1 + \tau_2 + \dots + \tau_n \leq t\}. \quad (9)$$

Podemos pensar en τ_i como el tiempo de espera entre las llegadas, es decir, τ_1 es el tiempo transcurrido hasta la primera llegada, τ_2 es el tiempo transcurrido entre la primera y la segunda llegada, τ_3 es el tiempo entre la segunda y la tercera llegada, etc. De esta manera, $T_n = \tau_1 + \tau_2 + \dots + \tau_n$ es el tiempo transcurrido hasta la n -ésima llegada. En particular, la

condición (9) puede ser reescrita como

$$N_t = \max\{n : T_n \leq t\}.$$

Si, por ejemplo, $N_t = 4$, la condición anterior nos dice que

$$T_4 \leq t < T_5.$$

En términos de llegadas, tenemos que en el instante t habrán llegado los cuatro primeros clientes, pero no el quinto.

Recordemos que λ era el promedio de llegadas por unidad de tiempo. Por otro lado, como $\mathbb{E}(\tau) = \frac{1}{\lambda}$, tendremos que esperar de media $\frac{1}{\lambda}$ unidades de tiempo entre cada dos llegadas sucesivas.

Ejemplo. A la bandeja de entrada de mi email llegan de media 2 emails por hora (o un email cada media hora). Supongamos que abro el correo a las 8:00 am. Calcula:

- 1) La probabilidad de que el primer email llegue más tarde de las 8:40.
- 2) La probabilidad de que el primer email llegue más tarde de las 11:00 si no ha llegado ningún correo hasta las 9:00.

Solución: Consideremos los tiempos $\tau_1, \tau_2, \tau_3, \dots$ entre llegadas de emails, medidos en horas.

- 1) Tenemos que $\tau_1 \sim \text{Exp}(2)$, luego

$$\mathbb{P}(\tau_1 > 0.5) = e^{-2 \cdot 0.5} = 0.3679.$$

Alternativamente, podemos usar R para calcular la probabilidad anterior mediante la función de distribución de la exponencial:

```
lambda <- 2
# Evaluamos la función de distribución de la exponencial
1 - pexp(0.5, lambda)
## [1] 0.3678794
```

- 2) Usando que τ_1 no tiene memoria

$$\mathbb{P}(\tau_1 > 3 | \tau_1 > 1) = P(\tau_1 > 2) = e^{-2 \cdot 2} = 0.0183.$$

Alternativamente, podemos usar R para calcular la probabilidad anterior.

```
lambda <- 2
# Evaluamos la función de distribución de la exponencial
1 - pexp(2, lambda)
## [1] 0.01831564
```

Recordemos que una variable aleatoria continua T tiene distribución gamma, en cuyo caso escribimos $T \sim \Gamma(k, \lambda)$, si toma valores positivos y tiene función de densidad

$$f(t) = \frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}, \quad t > 0,$$

donde $\Gamma(k)$ es la función Gamma, la cual viene dada por $\Gamma(k) = (k-1)!$ cuando k es entero.

La distribución gamma se obtiene cuando sumamos variables independientes con distribución exponencial.

Proposición 3.3.3 *Supongamos que $\tau_1, \tau_2, \tau_3, \dots$ son variables independientes con distribución $\text{Exp}(\lambda)$. Entonces se verifica que*

$$T_n = \tau_1 + \tau_2 + \dots + \tau_n \sim \Gamma(n, \lambda).$$

La anterior propiedad es muy útil para calcular probabilidades relativas a tiempos de llegada, ya que nos da la distribución del tiempo de llegada T_n del n -ésimo cliente.

Ejemplo. Volviendo al ejemplo anterior. Calcula:

- 1) La probabilidad de que el cuarto email llegue pasadas las 10:00 am.
- 2) La probabilidad de que el cuarto email llegue más de 3 horas después del segundo email si el segundo email llega después de las 9:00 am.

Solución:

- 1) Como $T_4 = \tau_1 + \tau_2 + \tau_3 + \tau_4 \sim \Gamma(n = 4, \lambda = 2)$, tenemos

$$\mathbb{P}(T_4 > 2) = \int_2^\infty \frac{2^4 t^3 e^{-2t}}{3!} dt = \frac{8}{3} \int_2^\infty t^3 e^{-2t} dt = 0.4435,$$

donde la anterior integral puede ser calculada a mano mediante integración por partes, o con ayuda del ordenador.

Alternativamente, usando R tenemos.

```
n <- 4
lambda <- 2
# Evaluamos la función de distribución de la distribución gamma
1 - pgamma(2, n, lambda)
## [1] 0.4334701
```

- 2) Como τ_3, τ_4 son independientes de τ_2, τ_1 , por ser las τ_i independientes entre sí, tenemos que

$$\begin{aligned} \mathbb{P}(\tau_3 + \tau_4 > 3 | \tau_1 + \tau_2 > 1) &= \mathbb{P}(\tau_3 + \tau_4 > 3) \\ &= \int_3^\infty 4te^{-2t} dt = 0.0173, \end{aligned}$$

donde hemos usado que $\tau_3 + \tau_4 \sim \Gamma(n = 2, \lambda = 2)$. Además, la anterior integral puede ser calculada a mano mediante integración por partes, o con ayuda del ordenador.

Alternativamente, usando R tenemos.

```
n <- 2
lambda <- 2
1 - pgamma(3, n, lambda)
## [1] 0.01735127
```

3.3.3) Simulación de procesos de Poisson

Simular proceso de Poisson es sencillo usando los tiempos entre llegadas. Para ello basta simular los tiempos entre llegadas como variables aleatorias exponenciales independientes y contabilizar el número de llegadas ocurridas.

Veamos un ejemplo en R:

```
# Valores para simulación
lambda <- 2 # tasa (llegadas por unidad de tiempo)
T <- 10 # Intervalo de tiempo en el que observamos el proceso

# Simulación de los tiempos entre llegadas
set.seed(123)
tau <- NULL
i <- 1
```

```

while (sum(tau) < T) {
  tau[i] <- rexp(n = 1, rate = lambda)
  i <- i+1
}

# Calculamos los valores del proceso de Poisson
tiempos <- c(0, cumsum(tau))
N <- 0:length(tau)

# Dibujamos la trayectoria
plot(tiempos, N, type = "s",
      xlab = "Tiempo", ylab = "Llegadas", xlim = c(0, 10))

```

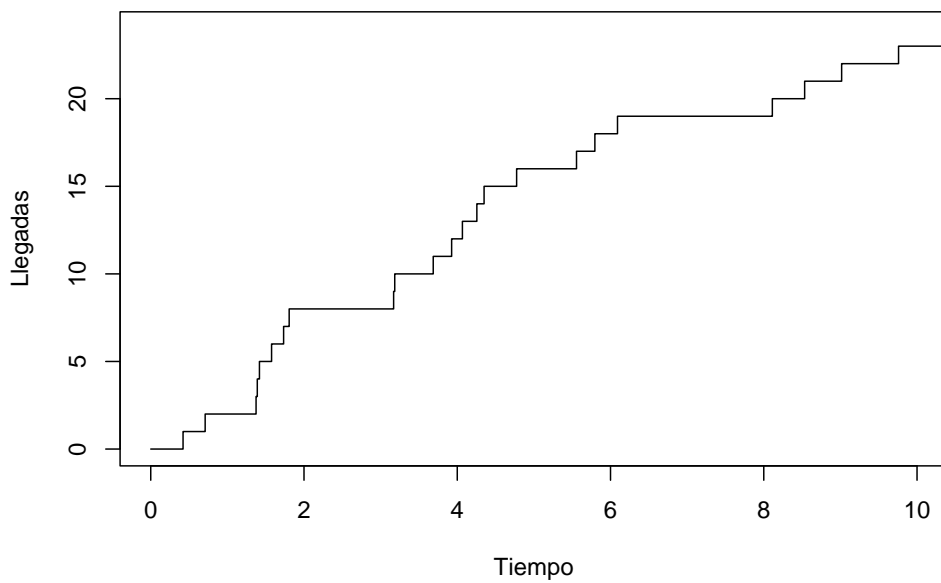


Figure 6: Simulación de la trayectoria de un proceso de Poisson

3.4) Transformadas de procesos de Poisson

En esta sección vamos a ver dos tipos de transformaciones de procesos de Poisson, la superposición y la separación.

3.4.1) Superposición de procesos de Poisson independientes

Consideremos dos procesos de Poisson $(N_t^1)_{t \in [0, \infty)}$ y $(N_t^2)_{t \in [0, \infty)}$ independientes con tasas λ_1 y λ_2 , respectivamente. Consideremos el proceso $(N_t)_{t \in [0, \infty)}$ definido como

$$N_t = N_t^1 + N_t^2 \text{ para todo } t \in [0, \infty).$$

Dados $t, s \geq 0$, se tiene que los incrementos

$$N_{t+s}^1 - N_s^1 \sim \text{Poisson}(\lambda_1 t), \quad N_{t+s}^2 - N_s^2 \sim \text{Poisson}(\lambda_2 t),$$

y son variables aleatorias independientes. Por tanto, debido a la Proposición 3.2.2 tenemos que

$$N_{t+s} - N_t = (N_{t+s}^1 - N_t^1) + (N_{t+s}^2 - N_t^2)$$

sigue una distribución Poisson($(\lambda_1 + \lambda_2)t$). Es más, $N_0 = N_0^1 + N_0^2 = 0$, y el proceso N_t tiene incrementos independientes (por tenerlos N^1 y N^2). Dado que se verifican todas las condiciones de la definición de proceso de Poisson, podemos concluir que el proceso $(N_t)_{t \in [0, \infty)}$ es de Poisson con tasa $\lambda_1 + \lambda_2$.

En general, tendremos lo siguiente.

Teorema 3.4.1 (Superposición) Sean $(N_t^1), (N_t^2), \dots, (N_t^m)$ procesos de Poisson independientes con tasas $\lambda_1, \lambda_2, \dots, \lambda_m$ respectivamente. Sea además

$$N_t = N_t^1 + N_t^2 + \dots + N_t^m \text{ para todo } t \in [0, \infty).$$

Entonces el proceso $(N_t)_{t \in [0, \infty)}$ es un proceso de Poisson con tasa $\lambda_1 + \lambda_2 + \dots + \lambda_m$.

A continuación simulamos en R la superposición de dos procesos de Poisson independientes.

```
# Valores para simulación de los procesos
lambda1 <- 2 # tasa del primer proceso de Poisson
lambda2 <- 0.5 # tasa del segundo proceso de Poisson
T <- 10 # Intervalo de tiempo en el que observamos el proceso

# Simulación de los tiempos entre llegadas
set.seed(123)
tau1 <- NULL
tau2 <- NULL
i <- 1
while(sum(tau1) < T){
  tau1[i] <- rexp(n = 1, rate = lambda1)
  i <- i + 1
}

i <- 1
while (sum(tau2) < T) {
  tau2[i] <- rexp(n = 1, rate = lambda2)
  i <- i + 1
}

# Definimos los tiempos y los valores correspondientes del proceso de Poisson 1
tiempos1 <- c(0, cumsum(tau1))
N1 <- 0:length(tau1)

# Definimos los tiempos y los valores correspondientes del proceso de Poisson 2
tiempos2 <- c(0, cumsum(tau2))
N2 <- 0:length(tau2)

# Definimos los tiempos y los valores correspondientes del proceso de superposición
tiempos <- union(tiempos1, tiempos2)
tiempos <- sort(tiempos)
N <- 0:(length(tiempos) - 1)

# Dibujamos la trayectoria del proceso 1
plot(tiempos1, N1, type = "s",
     xlab = "Tiempo", ylab = "Llegadas", xlim = c(0, 10), ylim = c(0, max(N)))
```



```
# Dibujamos la trayectoria del proceso 2
lines(tiempos2, N2, type = "s", col="blue")

# Dibujamos la trayectoria del proceso superposición
lines(tiempos, N, type = "s", col="red")

# Añadimos una leyenda
legend(0, 25, legend = c("Proceso 1", "Proceso 2", "Superposición"),
      col=c("black", "blue", "red"), lty=c(1, 1, 1))
```

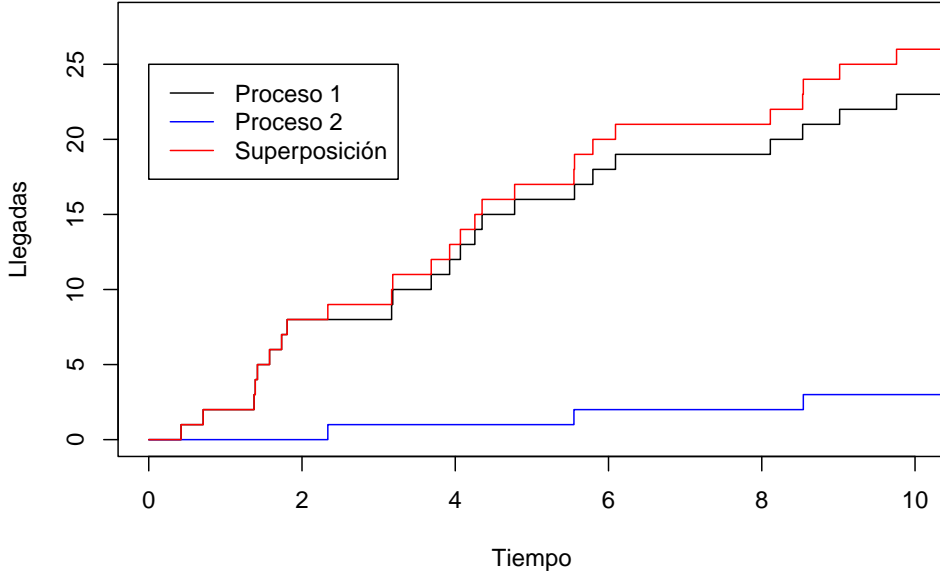


Figure 7: Superposición de dos procesos de Poisson

3.4.2) Separación de procesos de Poisson

Imaginemos una cafetería en los Pirineos donde los clientes, de España y de Francia, van llegando de acuerdo a un proceso de Poisson $(N_t)_{t \in [0, \infty)}$ con tasa de 10 clientes por hora. Supongamos además que el 60% de los clientes son españoles y el 40% de los clientes son franceses. De esta manera, los clientes españoles tendrán su propio proceso de conteo $(N_t^1)_{t \in [0, \infty)}$ y, de la misma manera, los clientes franceses tendrán otro proceso de conteo $(N_t^2)_{t \in [0, \infty)}$. Nos preguntamos qué tipo de procesos son $(N_t^1)_{t \in [0, \infty)}$ y $(N_t^2)_{t \in [0, \infty)}$.

En general, un proceso de Poisson $(N_t)_{t \in [0, \infty)}$ puede ser separado en dos procesos $(N_t^1)_{t \in [0, \infty)}$ y $(N_t^2)_{t \in [0, \infty)}$ contando llegadas de "tipo 1" y "tipo 2", de acuerdo a una sucesión Y_1, Y_2, \dots de variables aleatorias independientes entre sí, e independientes del proceso $(N_t)_{t \in [0, \infty)}$, tal que

$$\mathbb{P}(Y_i = 1) = p \text{ y } \mathbb{P}(Y_i = 2) = 1 - p.$$

De esta manera, si $T_n = \tau_1 + \tau_2 + \dots + \tau_n$ es el tiempo transcurrido hasta la n -ésima llegada, entonces los procesos $(N_t^1)_{t \in [0, \infty)}$ contando las llegadas de tipo 2 hasta tiempo t vendrán respectivamente dados por

$$N_t^1 = \#\{n : Y_n = 1, T_n \leq t\}, \quad N_t^2 = \#\{n : Y_n = 2, T_n \leq t\}, \quad (10)$$

donde $\#$ denota el cardinal de un conjunto.

Tenemos el siguiente resultado.

Teorema 3.4.2 (Separación) Sea $(N_t)_{t \in [0, \infty)}$ un proceso de Poisson con tasa λ . Consideremos que cada llegada es de tipo

1 con probabilidad p o de tipo 2 con probabilidad $1 - p$, independientemente de las demás llegadas y de $(N_t)_{t \in [0, \infty)}$. Entonces, los procesos $(N_t^1)_{t \in [0, \infty)}$ y $(N_t^2)_{t \in [0, \infty)}$ contando las llegadas de tipo 1 y las llegadas de tipo 2, respectivamente, verifican:

- 1) $(N_t^1)_{t \in [0, \infty)}$ es un proceso de Poisson con tasa λp .
- 2) $(N_t^2)_{t \in [0, \infty)}$ es un proceso de Poisson con tasa $\lambda(1 - p)$.
- 3) $(N_t^1)_{t \in [0, \infty)}$ y $(N_t^2)_{t \in [0, \infty)}$ son procesos de Poisson independientes.

Volvamos al ejemplo de la cafetería en los Pirineos. Dijimos que llegan una media de 10 clientes por hora, por lo que el correspondiente proceso de Poisson $(N_t)_{t \in [0, \infty)}$ tendrá tasa $\lambda = 10$. Al ser el 60% de los clientes que llegan españoles y el 40% franceses, tendremos que los clientes españoles llegarán a la cafetería siguiendo un proceso de Poisson con tasa $10 \cdot 0.6 = 6$, y los clientes franceses llegarán a la cafetería siguiendo un proceso de Poisson con tasa $10 \cdot 0.4 = 4$.

A continuación simulamos en R la separación del proceso de clientes en españoles y franceses:

```
# Valores para simulación de los procesos
lambda <- 10 # tasa del proceso de Poisson
p <- 0.6 # tasa del segundo proceso de Poisson
T <- 10 # Intervalo de tiempo en el que observamos el proceso

# Simulación de los tiempos entre llegadas y tipos de llegadas
set.seed(123)
tau <- NULL
y <- NULL

i <- 1
while (sum(tau) < T) {
  tau[i] <- rexp(n = 1, rate = lambda)
  y[i] = sample(c(1, 2), size = 1)
  i <- i+1
}

# Definimos los tiempos y los valores de N
tiempos <- c(0, cumsum(tau))
N <- 0:length(tau)

# Definimos los valores del proceso de Poisson 1
N1 <- NULL
N1[y==1] <- 1
N1[y==2] <- 0
N1 <- c(0, cumsum(N1))

# Definimos los valores del proceso de Poisson 2
N2 <- NULL
N2[y==2] <- 1
N2[y==1] <- 0
N2 <- c(0, cumsum(N2))

# Dibujamos la trayectoria del proceso 1
plot(tiempos, N1, type = "s",
     xlab = "Tiempo", ylab = "Llegadas", xlim = c(0, 10), ylim = c(0, max(N)))
```

```
# Dibujamos la trayectoria del proceso 2
lines(tiempos, N2, type = "s", col = "blue")

# Dibujamos la trayectoria del proceso 2
lines(tiempos, N, type = "s", col = "red")

# Añadimos una leyenda
legend(0, 18, legend = c("Españoles", "Franceses", "Totales"),
      col = c("black", "blue", "red"), lty = c(1, 1, 1))
```

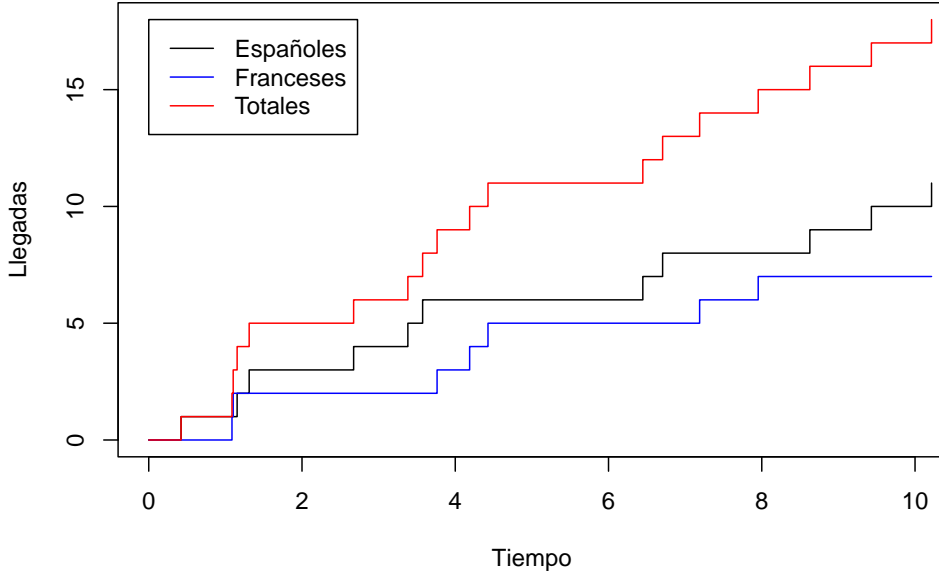


Figure 8: Separación del proceso de Poisson

3.5) Procesos de renovación

Hemos visto anteriormente que los procesos de Poisson son caracterizados como el proceso de conteo cuando los tiempos entre llegadas dados por una sucesión $\tau_1, \tau_2, \tau_3, \dots$ de variables independientes y exponenciales con la misma tasa λ . Una generalización de los procesos de Poisson son los procesos de renovación donde los tiempos entre llegadas son independientes e idénticamente distribuidos, pero no necesariamente siguen una distribución exponencial. Dicho concepto es definido formalmente como sigue.

Definición 3.5.1 Sean $\tau_1, \tau_2, \tau_3, \dots$ variables aleatorias no negativas (es decir, $\tau_n \geq 0$), independientes, e idénticamente distribuidas, de modo que $\mathbb{P}(\tau_n = 0) < 1$ para todo n . Llamamos **proceso de renovación** con tiempos entre llegadas $\tau_1, \tau_2, \tau_3, \dots$ al proceso estocástico $(N_t)_{t \in [0, \infty)}$ verificando $N_0 = 0$ y

$$N_t = \max\{n : T_n \leq t\} \quad (11)$$

donde $T_n = \tau_1 + \tau_2 + \dots + \tau_n$.

Al igual que con los proceso de Poisson interpretamos $\tau_1, \tau_2, \tau_3, \dots$ como los tiempos entre llegadas, por lo que $T_n = \tau_1 + \tau_2 + \dots + \tau_n$ es el tiempo de la n -ésima llegada. La condición (11) implica que si $N_t = n$, entonces

$$T_n \leq t < T_{n+1}.$$

Ejemplo: Imaginemos una máquina que funciona durante un tiempo α y, tras fallar, tarda un tiempo β en ser separada. Si la máquina falla y es reparada reiteradas veces, obtendremos variables aleatorias independientes $\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3, \dots$ con los correspondientes tiempos de fallo y reparación. El tiempo total de cada ciclo donde la máquina empieza a funcionar, falla y es reparada viene dado por $\tau_i = \alpha_i + \beta_i$. Las variables $\tau_1, \tau_2, \tau_3, \dots$ son i.i.d. Podemos, por tanto, considerar el proceso de renovación $(N_t)_{t \in [0, \infty)}$ contando el número total de ciclos hasta tiempo t .

Una importante propiedad de los procesos de renovación es la siguiente versión de la ley de los grandes números.

Teorema 3.5.1 (Ley de los grandes números para procesos de renovación) Sea $(N_t)_{t \in [0, \infty)}$ un proceso de renovación con tiempos entre llegadas $\tau_1, \tau_2, \tau_3, \dots$. Si $\mathbb{E}(\tau_i) = \mu$, entonces con probabilidad 1 se verifica que

$$\frac{N_t}{t} \rightarrow \frac{1}{\mu} \text{ cuando } t \rightarrow \infty.$$

La interpretación del anterior resultado es sencilla. Si el tiempo medio entre llegadas (ciclos consecutivos) de un proceso de renovación es de μ horas, entonces en el largo plazo se producirán $\frac{1}{\mu}$ llegadas (ciclos) por hora. En el caso de un proceso de Poisson con tasa λ se tendrá que

$$\frac{N_t}{t} \rightarrow \lambda \text{ cuando } t \rightarrow \infty,$$

ya que los tiempos entre llegadas tienen media $\frac{1}{\lambda}$.

La demostración del anterior resultado se basa en la ley fuerte de los grandes números la cual recordamos a continuación.

Lema 3.5.1 (Ley fuerte de los grandes números) Sean $\tau_1, \tau_2, \tau_3, \dots$ variables aleatorias i.i.d. con $\mathbb{E}(\tau_i) = \mu$. Entonces con probabilidad 1 se verifica que

$$\frac{T_n}{n} \rightarrow \mu \text{ cuando } n \rightarrow \infty,$$

donde $T_n = \tau_1 + \tau_2 + \dots + \tau_n$ para cada $n \in \mathbb{N}$.

A continuación damos la demostración de la ley de los grandes números para procesos de renovación.

Demostración del Teorema 3.5.1 Para cada $n \in \mathbb{N}$, sea $T_n = \tau_1 + \tau_2 + \dots + \tau_n$. Como las variables $\tau_1, \tau_2, \tau_3, \dots$ son i.i.d. con media μ , la ley de los grandes números nos dice que, con probabilidad 1, se verifica

$$\frac{T_n}{n} \rightarrow \mu \text{ cuando } n \rightarrow \infty. \quad (12)$$

Ahora, por definición de proceso de renovación tenemos que

$$T_{N_t} \leq t < T_{N_t+1}.$$

Dividiendo entre N_t tenemos que

$$\frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{T_{N_t+1}}{N_t}.$$

De ahí se sigue que

$$\frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{T_{N_t+1}}{N_t+1} \frac{N_t+1}{N_t}. \quad (13)$$

Puesto que las trayectorias de N_t son crecientes, dando infinitos saltos de longitud 1, tendremos que $\lim_{t \rightarrow \infty} N_t = \infty$. Como consecuencia de (12) se verifica, con probabilidad 1, que

$$\lim_{t \rightarrow \infty} \frac{T_{N_t}}{N_t} = \mu.$$

Por las mismas razones también se cumplirá, con probabilidad 1, que

$$\lim_{t \rightarrow \infty} \frac{T_{N_t+1}}{N_t+1} \frac{N_t+1}{N_t} = \mu.$$

Finalmente, tomando límite en (13) cuando $t \rightarrow \infty$, y teniendo en cuenta lo anterior, llegamos a la conclusión deseada, ya

que $\frac{t}{N_t}$ se encuentra, con probabilidad 1, entre dos sucesiones las cuales convergen a μ .

Volvamos al ejemplo donde una máquina se estropea y es reparada reiteradas veces. Supongamos que, de media, la máquina tarda una hora en estropearse, y quince minutos en ser reparada. Supongamos, además, que cada tiempo de fallo α_i y cada tiempo en reparación β_i siguen distribuciones exponenciales. De este modo, $\alpha_i \sim \text{Exp}(1)$ y $\beta_i \sim \text{Exp}(4)$. El proceso de renovación resultante puede ser simulado fácilmente en R.

```
# Dijemos los valores de los parámetros
lambda_1 <- 1 # tasa de fallo
lambda_2 <- 4 # tasa de reparación
T <- 10      # Intervalo de tiempo en el que observamos el proceso

# Simulación de los tiempos entre llegadas
set.seed(13)
tau <- NULL
i <- 1
while (sum(tau) < T) {
  alpha <- rexp(n = 1, rate = lambda_1)
  beta <- rexp(n = 1, rate = lambda_2)
  tau[i] <- alpha + beta
  i <- i + 1
}

# Calculamos los valores del proceso de renovación
tiempos <- c(0, cumsum(tau))
N <- 0:length(tiempos)

# Dibujamos la trayectoria
plot(tiempos, N, type = "s",
      xlab = "Tiempo", ylab = "Ciclos", xlim = c(0, 10))
```

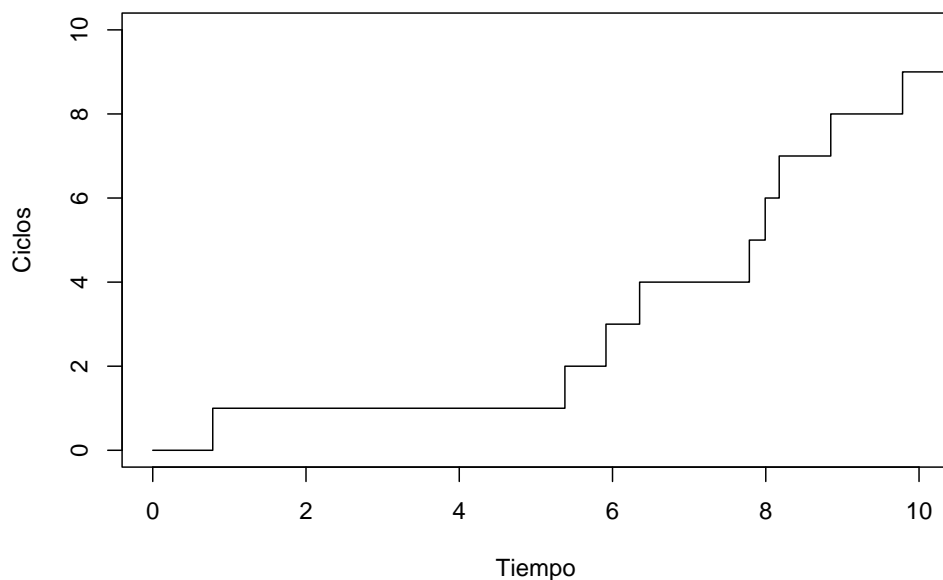


Figure 9: Simulación del proceso de renovación donde una máquina se estropea y repara reiteradamente

El tiempo medio de cada ciclo es $\mathbb{E}(\tau_i) = \mathbb{E}(\alpha_i) + \mathbb{E}(\beta_i) = 1 + \frac{1}{4} = \frac{5}{4}$. Por tanto, la ley de los grandes números para procesos de renovación nos dice que

$$\frac{N_t}{t} \rightarrow \frac{4}{5} \text{ cuando } t \rightarrow \infty.$$

Vamos a comprobar dicho resultado simulando una trayectoria a un plazo elevado de tiempo (por ejemplo, $T = 1000$), y dibujando la gráfica del cociente $\frac{N_t}{t}$.

```

# Dijemos los valores de los parámetros
lambda_1 <- 1 # tasa de fallo
lambda_2 <- 4 # tasa de reparación
T <- 1000     # Intervalo de tiempo en el que observamos el proceso

# Simulación de los tiempos entre llegadas
set.seed(13)
tau <- NULL
i <- 1
while (sum(tau) < T) {
  alpha <- rexp(n = 1, rate = lambda_1)
  beta <- rexp(n = 1, rate = lambda_2)
  tau[i] <- alpha + beta
  i <- i + 1
}

# Calculamos los valores del proceso de renovación
tiempos <- c(0, cumsum(tau))
N <- 0:length(tiempos)
cociente <- N / tiempos

# Dibujamos la trayectoria
plot(tiempos, cociente, type = "S",
     xlab = "Tiempo", ylab = "N_t/t", xlim = c(0, 1000))

# Dibujamos una línea horizontal a la altura de la inversa de la media
abline(h=4/5, col="blue")

```

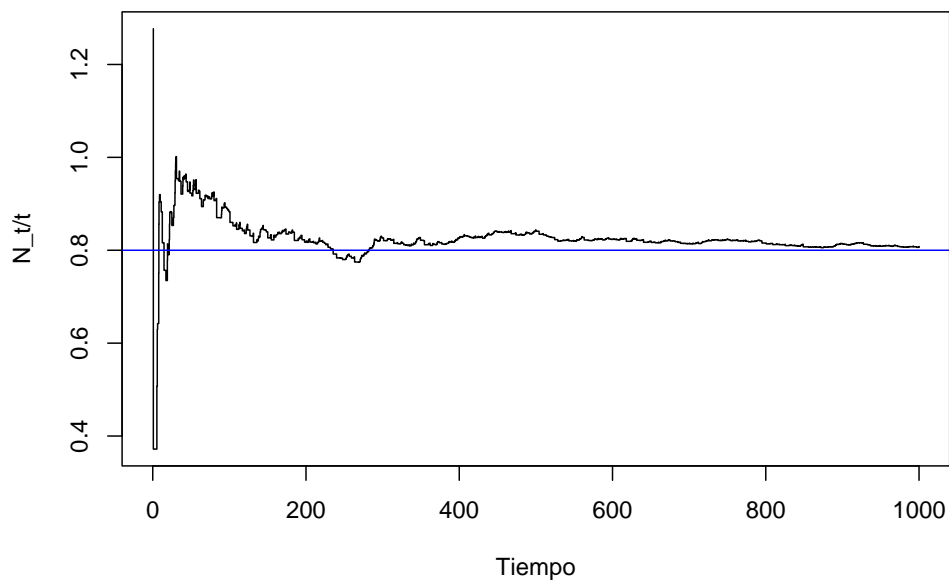


Figure 10: El cociente entre el proceso de renovación y el tiempo se va aproximando a una constante, debido a la ley de los grandes números

Hoja 3: Problemas de Procesos de Poisson y de conteo

- 1) Sea $(N_t)_{t \in [0, \infty)}$ un proceso de Poisson con tasa $\lambda = 2$, y sean $\tau_1, \tau_2, \tau_3, \dots$ los tiempos entre llegadas.
- a) Encuentra la probabilidad de que la primera llegada ocurra después de $t = 0.5$.
 - b) Encuentra la probabilidad de que la primera llegada ocurra antes de $t = 3$ si no ha habido ninguna llegada antes de $t = 1$.
 - c) Encuentra la probabilidad de que haya cuatro llegadas después de $t = 4$ si la tercera llegada ocurrió exactamente en $t = 2$.
 - d) Supongamos que empezamos observando el proceso en tiempo $t = 10$. Sea T el tiempo que tarda en ver la primera llegada. Encuentra $\mathbb{E}(T)$ y $\text{Var}(T)$.
- 2) Sea $(N_t)_{t \in [0, \infty)}$ un proceso de Poisson con tasa $\lambda = 1$.
- a) Calcula la probabilidad de que haya dos llegadas en $(0, 2]$ y tres llegadas en $(2, 4]$.
 - b) Encuentra la probabilidad de que haya dos llegadas en $(0, 2]$ y tres llegadas en $(1, 4]$.
- 3) Sea $(N_t)_{t \in [0, \infty)}$ un proceso de Poisson con tasa λ .
- a) Calcula la función de medias, covarianzas y correlaciones para el proceso $(N_t)_{t \in [0, \infty)}$.
 - b) ¿Es un proceso débilmente estacionario?
 - c) ¿Es un proceso estacionario (en sentido fuerte)?
- 4) Sea $(N_t)_{t \in [0, \infty)}$ un proceso de Poisson con tasa λ .
- a) Considera $0 < s < t$. Prueba que si $N_t = n$, entonces N_s sigue una distribución binomial $\mathcal{B}\left(n, \frac{s}{t}\right)$, es decir,

$$P(N_s = k | N_t = n) = \binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k} \text{ para } k \in \{0, 1, 2, \dots, n\}.$$

- b) Sea τ_1 el tiempo de la primera llegada. Prueba que si $N_t = 1$, entonces τ_1 sigue una distribución uniforme en $(0, t)$, es decir,

$$P(\tau_1 \leq x | N_t = 1) = \frac{x}{t} \text{ para } 0 \leq x \leq t.$$

- 5) En un aeropuerto los aviones llegan a la pista de aterrizaje siguiendo un proceso de Poisson $\{N_t\}_{t \geq 0}$. Sabiendo que llegan de media $\lambda = \frac{1}{2}$ aviones por hora. Si empezamos a observar los aviones que llegan a partir de las 8 de la mañana, calcule:
- a) La probabilidad de que no lleguen aviones entre las 8 y las 10 de la mañana.
 - b) La probabilidad de que el primer, segundo, tercer y cuarto avión lleguen respectivamente entre las 8 y las 9, entre las 9 y las 10, entre las 11 y las 12, y entre las 12 y las 13.
 - c) El cuarto avión de la mañana aterrice más de dos horas más tarde que el tercero.
 - d) El quinto avión de la mañana aterrice más de tres horas más tarde que el tercero.
- 6) A la pista de aterrizaje de otro aeropuerto llegan vuelos nacionales e internacionales. Cada hora llegan de media 4 vuelos nacionales y 2 internacionales. Suponemos que ambos tipos de vuelos llegan a la pista de aterrizaje siguiendo procesos de Poisson independientes, y que empezamos a observar los aviones a las 8.00 de la mañana.
- a) Calcula la probabilidad de que el sexto avión del día aterrice más de media hora más tarde que el quinto avión del día.

- b) Calcula la probabilidad de que no llegue ningún avión durante la primera media hora y que el primer vuelo internacional del día llegue antes de una hora.
- c) Calcula la probabilidad de que durante la primera hora hayan aterrizado exactamente 3 internacionales si han aterrizado un total de 7 aviones durante el mismo periodo de una hora.
- 7) A una cafetería en la frontera entre Suiza y Alemania llegan de media 3 clientes cada hora, siguiendo un proceso de Poisson. Se sabe que $\frac{2}{3}$ de los clientes que llegan son suizos y $\frac{1}{3}$ son alemanes, no habiendo nunca clientes de otras nacionalidades.
- a) Calcular la probabilidad de que en una hora hayan llegado exactamente dos clientes y en dos horas exactamente 5 clientes.
- b) Supongamos que en una hora llegan dos clientes. Calcular la probabilidad de que exactamente uno de ellos sea suizo.
- 8) Dos equipos, A y B, juegan un partido de fútbol. El número de goles anotados por el Equipo A sigue un proceso de Poisson N_t^1 con tasa $\lambda_1 = 0.02$ goles por minuto, y el número de goles anotados por el Equipo B sigue un proceso de Poisson N_t^2 con tasa $\lambda_2 = 0.03$ goles por minuto. Se asume que los dos procesos son independientes. Sea N_t el número total de goles hasta el instante t , incluyendo dicho instante. El partido dura 90 minutos.
- a) Probabilidad de que el partido acabe en 0-0.
- b) Probabilidad de que al menos haya un total de dos goles en el partido.
- c) Probabilidad de que el partido acabe en 1-2.
- d) Probabilidad de que el partido acabe en empate.
- 9) Sean (N_t^1) y (N_t^2) dos procesos de Poisson independientes con tasas λ_1 y λ_2 , respectivamente. Consideremos el proceso $(N_t)_t$ donde $N_t = N_t^1 + N_t^2$. Demuestra que si $N_t = n$, entonces N_t^1 sigue una distribución binomial $\mathcal{B}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$, es decir,
- $$P(N_t^1 = k | N_t = n) = \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(1 - \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{n-k} \text{ para } k \in \{0, 1, 2, \dots, n\}.$$
- 10) En el problema 8, encuentra la probabilidad de que el equipo B meta gol el primero. Esto es, encuentra la probabilidad de que al menos se anote un gol en el partido y que el primero en marcar sea el equipo B.
- 11) Una fábrica opera con una máquina que falla de manera recurrente y se repara inmediatamente. El tiempo (en horas) de funcionamiento de la máquina hasta el fallo sigue una distribución $\text{Exp}\left(\frac{1}{5}\right)$. Cada vez que la máquina falla, se repara. El tiempo de reparación sigue una distribución exponencial de media 10 minutos.
- a) Calcula la tasa de fallos por hora en el largo plazo.
- b) Simula 10 trayectorias de dicho proceso en R.
- 12) En una carrera de Fórmula 1 un coche entra en boxes de manera recurrente. El tiempo (en horas) que circula hasta entrar en boxes sigue una distribución $\text{Exp}(0.5)$. El proceso de puesta a punto consta de las siguientes fases:
- Repostaje de combustible: Esta fase toma un tiempo fijo de 2 minutos.
 - Cambio de neumáticos: El tiempo empleado en esta fase sigue una distribución exponencial con media de 3 minutos.
 - Ajustes mecánicos: El tiempo (en minutos) empleado en esta sigue una distribución uniforme en $(0, 8)$.

Responde a las siguientes preguntas:

- a) Calcula la tasa de entradas a boxes por hora en el largo plazo.
- b) Simula 10 trayectorias de dicho proceso en R.

Tema 4: Conceptos básicos en series temporales

4.1) Introducción

El estudio de series temporales es fundamental para el proceso de predicción (Forecasting), que consiste en anticipar eventos futuros basándose en datos históricos y patrones identificables. Esta disciplina es esencial en diversos campos, desde la economía hasta la ingeniería, y su aplicación es vital para la toma de decisiones informadas en el ámbito empresarial y gubernamental.

¿Qué se puede predecir?

La predicción se utiliza en múltiples situaciones de la vida cotidiana y la industria, como son la gestión de intervalos, la planificación de la producción, la gestión energética o la asignación de personal. La capacidad de predecir con precisión depende de varios factores, entre ellos la comprensión de las causas subyacentes, la disponibilidad de datos históricos, la consistencia del comportamiento pasado y futuro, y el impacto de las predicciones en los propios eventos. Por ejemplo, las predicciones a corto plazo de la demanda eléctrica residencial suelen ser precisas debido a la estabilidad de los factores que la afectan y la abundancia de datos históricos relevantes.

Diferencias entre predicciones, objetivos y planificación

Es crucial diferenciar entre predicciones, objetivos y planificación. Predecir implica anticipar el futuro con la mayor precisión posible mediante el análisis históricos y el conocimiento de eventos futuros. Los objetivos son metas específicas que deben que deben alinearse con las predicciones y servir como guía para la planificación. La planificación involucra diseñar estrategias y acciones basadas en las predicciones para alcanzar los objetivos establecidos.

Determinación de qué predecir

Decidir qué predecir es un paso crucial. Puede involucrar observaciones individuales o datos agregados. Es importante definir el horizonte de predicción, que puede ser a corto plazo (una hora, o unos días), a medio plazo (por ejemplo, un mes) o a largo plazo (por ejemplo, un año). Además, se debe determinar la frecuencia de las predicciones, que puede ser diaria, semanal, mensual o anual, según las necesidades de los usuarios de las predicciones. La recolección y organización de datos relevantes y precisos es esencial para asegurar la calidad de las predicciones.

Datos y métodos de predicción

Los métodos cuantitativos de predicción son útiles cuando hay datos numéricos disponibles y se espera que los patrones pasados continúen en el futuro. Los datos de series temporales, como las ventas trimestrales o la demanda horaria de electricidad, son fundamentales en estos análisis. La descomposición de datos permite estudiar tendencias y patrones estacionales. Entre los métodos más utilizados se encuentran el análisis clásico de series, el suavizamiento exponencial y los modelos ARIMA, efectivos para capturar y predecir patrones en los datos históricos. Los modelos explicativos, que emplean variables predictoras y los modelos mixtos, que combinan características de varios enfoques, pueden mejorar la precisión y la utilidad de las predicciones.

Pasos básicos en Forecasting

Las etapas a tener en cuenta en un problema de predicciones son:

- 1) Definición del problema: Entender el uso y los usuarios de las predicciones para asegurar que se adapten a sus necesidades específicas.
- 2) Recolección de información: Obtener datos estadísticos relevantes y utilizar la experiencia de expertos en el campo.
- 3) Análisis preliminar: Graficar y analizar datos para identificar patrones, tendencias y estacionalidad, lo cual ayuda en la selección del modelo de predicción adecuado.
- 4) Elección y ajuste de modelos: Seleccionar y ajustar el método de predicción más apropiado basado en los datos disponibles y la situación específica.

- 5) Uso de las predicciones: Aplicar los resultados en la planificación y toma de decisiones, alineando las acciones con las expectativas futuras y los objetivos de la organización o empresa.

4.2) Definición de serie temporal. Gráficos.

Una de las técnicas para hacer inferencias sobre el futuro con base en lo ocurrido en el pasado, es el análisis de series de tiempo.

Dada una serie temporal, nuestros objetivos principales serán: describir el comportamiento de la serie, investigar el mecanismo generador de la serie temporal y buscar posibles patrones temporales que permitan sobrepasar la incertidumbre del futuro.

Se suele hablar estudio clásico o descriptivo de las series temporales para referirse a la metodología que se ha venido empleando desde la segunda mitad del siglo XIX. A pesar de sus limitaciones, los métodos clásicos se siguen usando por su sencillez y simplificación. En la década de los 60 se plantearon métodos de previsión alternativos a los clásicos que suplen varias de sus limitaciones, como son los métodos de alisado exponencial, y a principios de los 70 aparece un nuevo enfoque, debido a los estadísticos Box y Jenkins, para los modelos univariantes de series temporales.

Definición 4.2.1 Llamamos *serie temporal* a un conjunto de mediciones de cierto fenómeno o experimento registradas secuencialmente en el tiempo. Estas observaciones suelen denotarse por:

$$\{x_{t_1}, x_{t_2}, \dots, x_{t_n}\} \text{ con } \{t_1 < t_2 < \dots < t_n\} = T \subset \mathbb{R}$$

donde x_t , representa el valor, en el instante t_i , de una variable aleatoria que evoluciona con el tiempo.

Observación 4.2.1 Teniendo en cuenta la definición anterior, una serie temporal puede verse también como una realización (o trayectoria) de un proceso estocástico $(X_t)_{t \in T}$.

En muchas áreas de conocimiento, las observaciones de interés son obtenidas en instantes del tiempo. Por ejemplo, cada hora, cada día, datos mensuales, trimestrales, semestrales o bien registradas por algún aparato de medición en forma continua.

- De una manera sencilla, podemos decir que si $T \subset \mathbb{Z}$ es un conjunto contable, se dice que la serie es de *tiempo discreto*, mientras que si $T \subset \mathbb{R}$ es un intervalo real, se dice que la serie es de *tiempo continuo*.
- Cuando $t_{i+1} - t_i = k$, para todo $i = 1, \dots, n - 1$, se dice que la serie es *equiespaciada*; en caso contrario será no equiespaciada.

En adelante, trabajaremos por comodidad con series de tiempo equiespaciadas. Por tanto, se denotarán de la siguiente forma:

$$\{x_{t_1}, x_{t_2}, \dots, x_{t_n}\} = \{x_1, x_2, \dots, x_n\}$$

El primer paso en el análisis de cualquier serie de tiempo consiste en representarla gráficamente (**gráfico temporal**). En la representación gráfica de las series temporales se utilizan los ejes cartesianos. En el eje de abscisas se representa el tiempo t , y en el eje de ordenadas, los valores de la magnitud observada x_t . Se obtiene así una nube de puntos (t, x_t) que unidos por segmentos, proporcionan una visión dinámica de la evolución de la variable a lo largo del tiempo.

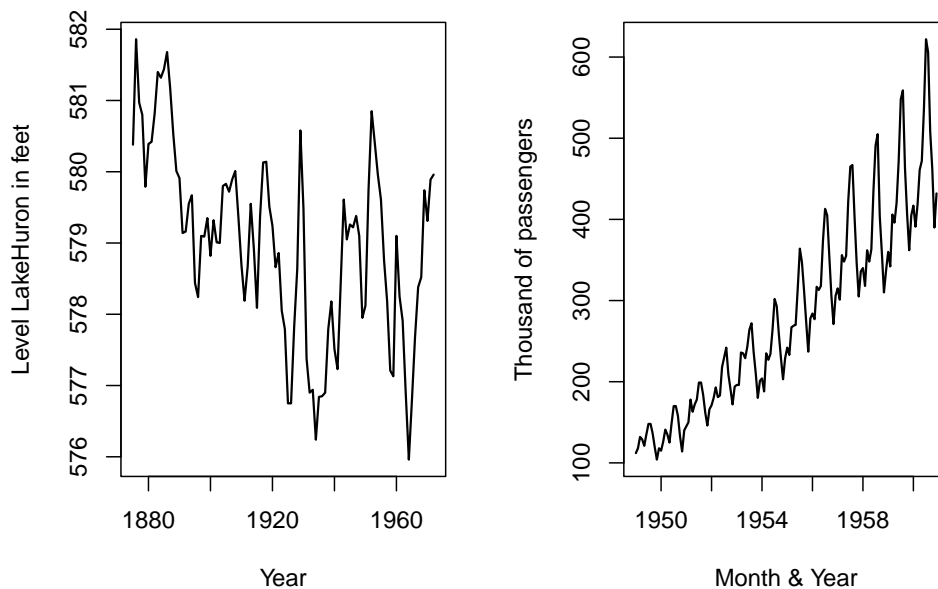
Veamos algunos ejemplos de series de datos incluidos en R. La primera, LakeHuron, contiene las medidas anuales del nivel del lago Huron (en pies), en el tramo 1875-1972. Observamos una tendencia decreciente que parece estabilizarse en el tramo final. La segunda serie de datos, AirPassengers, contiene los viajeros mensuales en aerolíneas internacionales, en el tramo 1949-1960. Observamos un patrón periódico que se repite a lo largo de la serie, así como tendencia creciente. La tercera, EuStockMarkets, contiene los precios diarios al cierre de 4 índices stock europeos (DAX, SMI, CAC y FTSE), en el tramo 1991-1998. Observamos que las cuatro series presentan un patrón de comportamiento similar, con tendencia creciente y picos de crecimiento en los mismo tramos temporales.

```
par(mfrow=c(1,2))
ts.plot(LakeHuron,
        gpars = list(xlab = "Year",
```

```

        ylab = "Level LakeHuron in feet",
        lwd = 1.5))
ts.plot(AirPassengers,
        gpars = list(xlab = "Month & Year",
        ylab = "Thousand of passengers",
        lwd = 1.5))

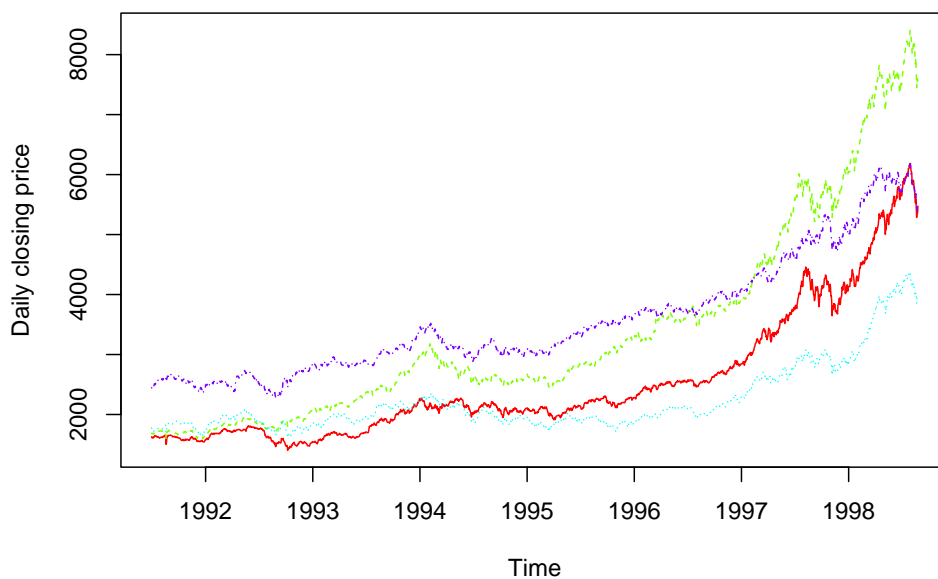
```



```

ts.plot(EuStockMarkets,
        gpars = list(xlab = "Time",
        ylab = "Daily closing price",
        col = rainbow(4),
        lty = c(1:4)))

```



El gráfico de la serie permitirá:

- a) *Detectar outliers (datos atípicos)*: se refiere a puntos de la serie que se escapan de lo "normal". Un outlier puede corresponderse con un error de medición, o bien con un valor real del fenómeno, el cual a su vez puede tener o no influencia destacada en el modelo propuesto. Conviene identificar los outliers para determinar si deben eliminarse, reemplazarse por otro valor, o bien mantenerlos en el análisis.
- b) *Detectar estacionalidad*: la variación estacional representa un movimiento periódico de la serie de tiempo. Cada estación puede corresponder con un trimestre, un mes, un día, etc.
- c) *Detectar tendencias*: la tendencia representa el comportamiento predominante de la serie una vez eliminada la estacionalidad. Esta puede ser definida vagamente como la curva maestra que muestra el cambio de la media a lo largo del tiempo.

4.3) Medidas para evaluación de predicciones

En esta sección introduciremos conceptos y alguna notación referidos al contexto de predicción con series temporales.

En primer lugar, debemos destacar que para que tenga sentido la realización de predicciones se supondrá que los datos recopilados a lo largo del tiempo se han tomado siempre en las mismas magnitudes y que la serie presenta cierta estabilidad en la estructura del fenómeno estudiado.

En el análisis clásico, las predicciones futuras se realizan teniendo en cuenta únicamente el valor de la serie para instantes pasados. Si disponemos de T observaciones de la serie en estudio, la información disponible viene dada por:

$$x_1, x_2, \dots, x_T$$

Dada esta información, la **predicción** de la variable en estudio en el instante siguiente $T + 1$ se denotará por:

$$\hat{x}_{T+1}$$

y en un instante posterior cualquiera $T + m$ será:

$$\hat{x}_{T+m}$$

Así, el **error de predicción** vendrá dado por la diferencia entre el valor observado en la serie y la predicción:

$$e_{T+1} = x_{T+1} - \hat{x}_{T+1}$$

Comentar que, al igual que sucede en muchos procesos de inferencia, se pueden realizar predicciones puntuales y predicciones por intervalos de confianza para las observaciones futuras. En la mayoría de los casos, nos limitaremos a proporcionar estimaciones puntuales.

Para poder comparar diversos procedimientos de predicción, es decir, para responder a la pregunta ¿qué método es mejor para predecir esta serie?, necesitamos disponer de algún criterio que nos permita evaluar las predicciones realizadas.

Como primer paso, podemos comentar la existencia de algunos métodos extremadamente básicos, denominados métodos ingenuos, que sirven de referencia en la comparación de predicciones. Es decir, las predicciones que se obtienen con cualquier procedimiento suelen compararse con las predicciones que se obtendrían con algún método ingenuo. Si el método ingenuo proporciona mejores predicciones (menores errores de predicción), esto será indicativo de que el método usado no es el adecuado y conviene probar con otro.

Dos ejemplos de métodos ingenuos que se utilizan como referencia son:

- *Método ingenuo I*: La predicción para el próximo instante es igual a la observación actual (último valor observado).

$$\hat{x}_{t+1} = x_t$$

- *Método ingenuo II*: La predicción para el próximo instante es igual a la observación actual más el último incremento

observado.

$$\hat{x}_{t+1} = x_t + (x_t - x_{t-1})$$

Como segunda opción para medir la capacidad predictiva de un modelo, se suelen calcular las siguientes medidas. Si se dispone de T observaciones y se han realizado predicciones desde el instante 1, las medidas son:

- *RMSE (la raíz del error cuadrático medio):*

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^T e_t^2}{T}}$$

- *MAE (el error absoluto medio):*

$$\text{MAE} = \frac{\sum_{t=1}^T |e_t|}{T}$$

Evidentemente, cuanto menores sean estas medidas mayor será la capacidad predictiva del modelo. Conviene indicar la existencia de otras medidas para evaluar predicciones, además de las vistas aquí.

Por último destacaremos que para que un modelo predictivo sea adecuado no basta con que los errores de predicción sean pequeños (valores pequeños de RMSE y de MAE), sino que además pretendemos describir completamente el comportamiento sistemático de la serie (la parte determinista), y de manera que la parte que queda sin describir (los errores de predicción) deberían ser un ruido blanco.

4.4) Componentes de una serie temporal y esquemas de integración

El principal fin del estudio de series temporales es realizar predicciones. Un punto muy importante a tener en cuenta en este contexto es la cantidad de información en una serie temporal.

Por ejemplo, podemos considerar la serie temporal correspondiente a la hora de salida del sol (medida todos los días durante 10 años). Con la serie de observaciones pasadas podemos realizar una buena predicción para días futuros, es decir, la serie contiene mucha información relevante para predecir el futuro. Consideremos por otra parte la serie temporal correspondiente al número premiado del sorteo de la lotería (medida todos los sábados durante 20 años). En este caso las observaciones pasadas no nos permiten realizar una buena predicción del fenómeno en el futuro, es decir, la serie contiene poca información relevante para predecir el futuro.

Podríamos decir que la primera serie es prácticamente determinista mientras que la segunda es totalmente aleatoria. En general, las series de tiempo contienen una parte determinista (permiten realizar predicciones) y otra parte aleatoria (perturbación impredecible).

En el **Análisis Clásico** se supone que una serie temporal, $\{x_t\}_{t=1,\dots,n}$, se puede descomponer en todas o algunas de las siguientes componentes:

- Componente de Tendencia (T_t)
- Componente de Ciclo (C_t)
- Componente de Estacionalidad (S_t)
- Componente Irregular (I_t)

y que la serie de tiempo se obtiene como una función de sus componentes:

$$x_t = f(T_t, C_t, S_t, I_t)$$

Vamos a continuación el significado de cada una de las componentes, aunque resulta complicado dar una definición exacta de cada una de ellas.

- a) **Componente de Tendencia (Trend):** Es la componente que representa la evolución a lo largo de la serie, es decir, la curva maestra de evolución del fenómeno en estudio. Las variaciones de la tendencia de una serie pueden deberse a diversos motivos.
- b) **Componente de Ciclo (Cycle):** El factor cíclico refleja movimientos oscilatorios por encima y por debajo de la tendencia a largo plazo. La duración de un ciclo se suele medir desde un pico al siguiente pico de la serie suavizada (o desde un valle al siguiente valle). Se entiende que el periodo de cada ciclo es siempre superior al año. A veces se confunde con la componente estacional. La diferencia principal es que las fluctuaciones de los ciclos no tienen frecuencia fija, y suelen asociarse a aspectos económicos, mientras que las fluctuaciones de la componente estacional tienen frecuencia fija y se asocian con aspectos del calendario.

Observación: En general, las componentes de tendencia y ciclo son muy difíciles de separar, de manera que suelen fusionarse en una componente denominada Tendencia-Ciclo. En adelante, nosotros trabajaremos con ambas componentes fusionadas y usaremos la notación T_t y hablaremos de Tendencia para referirnos a la componente Tendencia-Ciclo.

- c) **Componente de Estacionalidad (Seasonal):** La componente estacional recoge el comportamiento periódico (repetitivo) de la serie. Las razones de la estacionalidad suelen ser tipo físico-natural (tiempo meteorológico, ciclos biológicos, etc.) y de tipo institucional (vacaciones, horarios comerciales, etc.). Es decir, aparece debido al efecto del calendario.

En general, la estacionalidad se refiere a las oscilaciones de una serie temporal que se completan dentro de un año y que se repiten en años sucesivos. Por tanto, el periodo de esta componente suele ser menor o igual a un año. Por ejemplo, si disponemos de datos mensuales podemos tener estacionalidad con periodo 12 meses (un año), si disponemos de datos trimestrales podemos tener estacionalidad con periodo 4 trimestres (un año). Si disponemos de datos diarios, podemos tener estacionalidad con periodo 7 días (una semana), si disponemos de datos cada hora, podemos tener estacionalidad con periodo 24 horas (un día), etc. El **periodo** de la componente estacional hace referencia al número de datos que conforman la parte repetitiva de la serie.

En Análisis Clásico, se supone que la componente estacional (patrón periódico) se repite de forma fija a lo largo de toda la serie. Por tanto, si el periodo es L , diremos que la componente estacional tiene L estaciones que podemos denotar por $\{h_1, h_2, \dots, h_L\}$ y buscamos estimar un valor correspondiente a cada estación.

Por ejemplo, si se trata de datos mensuales, probablemente observamos en la serie temporal un patrón repetitivo (cada 12 datos). Tendremos entonces que el periodo es $L = 12$, cada mes del año sería una estación ($h_1 = \text{enero}, h_2 = \text{febrero}, \dots, h_{12} = \text{diciembre}$) y por tanto estimaremos un valor que se repetirá para todos los enero, un valor que se repetirá para todos los febrero, ..., un valor que se repetirá para todos los diciembre. Si se trata de datos diarios (por ejemplo, ingresos por ventas de unos grandes almacenes), observaremos en la serie temporal un patrón repetitivo cada semana (cada 7 datos). Tendremos entonces que el periodo es $L = 7$, cada día sería una estación ($h_1 = \text{lunes}, h_2 = \text{martes}, \dots, h_7 = \text{domingo}$) y por tanto estimaremos un valor que se repetirá para todos los lunes, un valor que se repetirá para todos los martes, ..., un valor que se repetirá para todos los domingos.

- d) **Componente Irregular (Reminder):** Esta componente viene dada por las variaciones de la serie que no están recogidas en las demás componentes. La idea es que recoja la perturbación aleatoria pura de la serie, es decir, la parte impredecible de la serie. Lo ideal es que se corresponda con un *ruido blanco*, de manera que toda la parte determinista de la serie quede recogida en las componentes Tendencia-Ciclo y Estacional.

Como ya hemos indicado, en el análisis clásico se supone que el valor que toma la serie $\{x_t\}_t$ en cada instante se puede expresar como una función del valor correspondientes a sus componentes en dicho instante:

$$x_t = f(T_t, S_t, I_t),$$

donde T_t representa la Tendencia-Ciclo. Conviene señalar que una serie temporal no tienen por qué estar presentes todas las componentes.

Los esquemas más utilizados para describir la forma en que las componentes se integran para dar lugar a la serie son el esquema *aditivo* y el *multiplicativo*.

- **Esquema aditivo:** Según este esquema, el valor de la serie en cualquier instante se obtiene como *suma* de los valores correspondientes a sus componentes en dicho instante:

$$x_t = T_t + S_t + I_t$$

- **Esquema multiplicativo:** Según este esquema, el valor de la serie en cualquier instante se obtiene como *producto* de los valores correspondientes a sus componentes en dicho instante:

$$x_t = T_t \times S_t \times I_t$$

Obsérvese que al tomar logaritmo neperianos, el esquema multiplicativo se transforma en aditivo.

Aunque por su sencillez los esquemas más utilizados son los dos anteriores, también suelen utilizarse los esquemas mixtos, donde algunas componentes se integran de forma multiplicativa y otras de forma aditiva. Un ejemplo de **esquema mixto** sería:

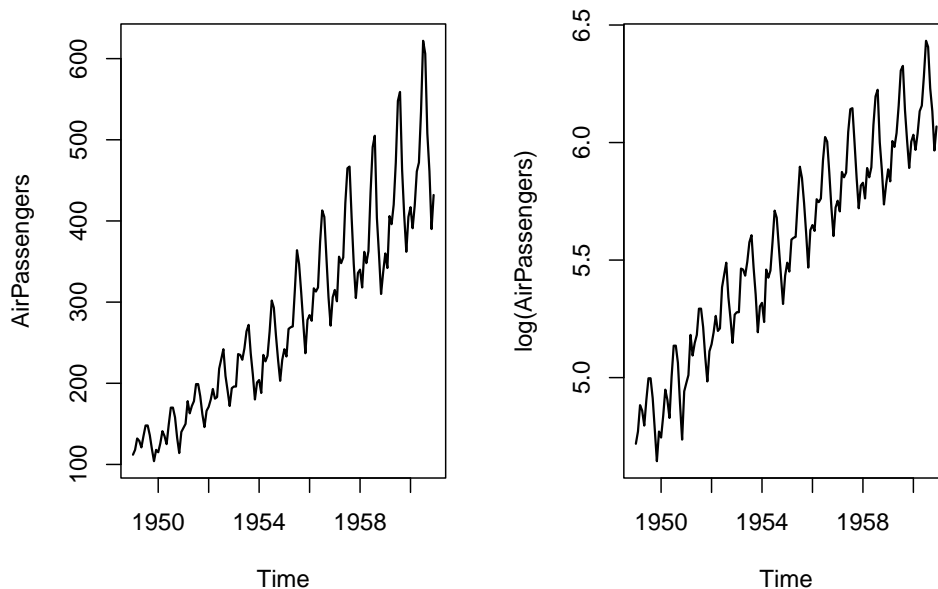
$$x_t = T_t \times S_t + I_t$$

¿Esquema aditivo o multiplicativo? Para determinar si las componentes de la serie se combinan de forma aditiva o multiplicativa, se pueden usar varias herramientas:

- 1) Inspección visual: En un esquema aditivo, la representación gráfica de la serie mostrará una componente estacional aproximadamente estable en todo momento, independientemente de los valores que tome la tendencia (nivel de la serie). Sin embargo, en un esquema multiplicativo, las desviaciones típicas aumentan cuando crece la media.

A continuación mostramos un ejemplo de esquema multiplicativo (figura de la izquierda) y un esquema aditivo (figura de la derecha).

```
par(mfrow=c(1,2))
ts.plot(AirPassengers, lwd = 1.5)
ts.plot(log(AirPassengers), lwd = 1.5)
```



4.5) Diferenciación de una serie temporal y medias móviles

La representación gráfica de la serie temporal nos puede ayudar a decidir el tipo de tendencia de la serie, pero a veces no es tan sencillo. Un método que permite determinar el tipo de tendencia de una serie y a su vez eliminarla consiste en *tomar*

diferencias.

Diferenciación de una serie

Definición 4.5.1 Dada una serie temporal $\{x_t\}_{t=1,\dots,n}$, se define su **diferencia de primer orden** como la serie que resulta de restar a cada observación, la observación anterior:

$$\Delta x_t = x_t - x_{t-1}$$

De manera análoga, se puede definir la *diferencia de orden 2* de la serie como:

$$\Delta^2 x_t = \Delta(x_t - x_{t-1}) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2}$$

así como la *diferencia de orden k*, en general.

Observación 4.5.1 Al tomar diferencias de orden uno, la serie resultante tiene un dato menos (se pierde el primer dato). Si se toman diferencias de orden dos, la serie resultante tiene dos datos menos (se pierden los dos primeros). En general, al tomar diferencias de orden k , la serie resultante tiene k datos menos (se pierden los k primeros datos).

- Si la tendencia es de tipo lineal, ésta se puede eliminar tomando diferencias de orden uno:

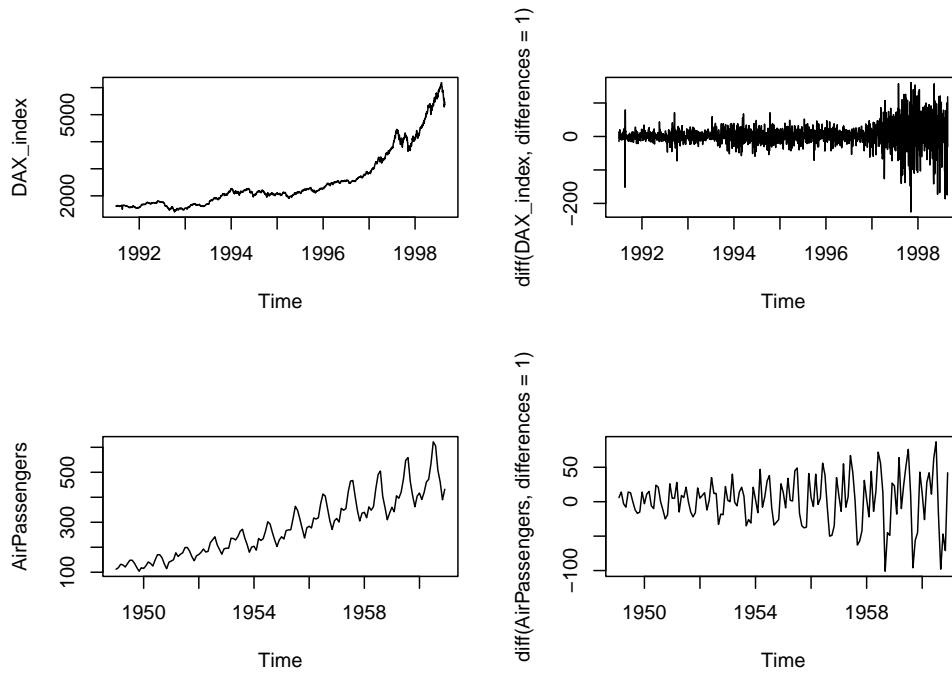
$$\begin{aligned}x_t &= b_0 + b_1 \cdot t + \varepsilon_t \text{ con } \varepsilon_t \text{ ruido blanco} \\ \Rightarrow \Delta x_t &= x_t - x_{t-1} = (b_0 + b_1 \cdot t + \varepsilon_t) - (b_0 + b_1 \cdot (t-1) + \varepsilon_{t-1}) \\ \Rightarrow \Delta x_t &= b_1 + (\varepsilon_t - \varepsilon_{t-1}) = b_1 + \tilde{\varepsilon}_t\end{aligned}$$

es decir, la serie obtenida al tomar diferencias de orden uno en una constante (b_1) más un ruido blanco ($\tilde{\varepsilon}_t$).

- En general, si la tendencia es de tipo polinómico de orden k , ésta se puede eliminar tomando diferencias de orden k .

Veamos un ejemplo del efecto de tomar diferencias de orden 1 sobre la serie de precios del índice DAX sobre la serie de pasajeros:

```
DAX_index <- EuStockMarkets[ , 1]
par(mfrow = c(2,2))
ts.plot(DAX_index)
ts.plot(diff(DAX_index, differences = 1))
ts.plot(AirPassengers)
ts.plot(diff(AirPassengers, differences = 1))
```

- Si la tendencia es de tipo exponencial, ésta se puede eliminar tomando diferencias en el logaritmo de la serie:

$$x_t = e^{b_0 + b_1 \cdot t} \times e^{\varepsilon_t} \Rightarrow \ln(x_t) = b_0 + b_1 \cdot t + \varepsilon_t$$

$$\Rightarrow \Delta \ln(x_t) = \ln(x_t) - \ln(x_{t-1}) = b_1 + \tilde{\varepsilon}_t$$

Por otra parte, la diferencia de logaritmos coincide con el logaritmo del cociente, así que:

$$\Rightarrow \Delta \ln(x_t) = \ln(x_t) - \ln(x_{t-1}) = \ln\left(\frac{x_t}{x_{t-1}}\right)$$

y usando el desarrollo de Taylor para el logaritmo se tiene que:

$$\Delta \ln(x_t) = \ln\left(\frac{x_t}{x_{t-1}}\right) \simeq \frac{x_t - x_{t-1}}{x_{t-1}} \text{ (Tasa de variación relativo)}$$

Por tanto, si la tendencia es de tipo exponencial, las tasas de variación relativa oscilarán alrededor de un valor constante.

Medias Móviles

Otro método para ayudar a identificar la tendencia de una serie (análisis descriptivo de la serie), son las **medias móviles**, aunque también puede usarse para realizar predicciones y para eliminar la componente estacional de una serie.

Este método consiste en realizar, para cada instante de tiempo t , la media de unas cuantas observaciones. Necesitamos determinar cuántas observaciones se usarán para el cálculo de la media móvil (longitud u orden de la media móvil).

Las medias móviles eliminan las irregularidades de la serie observada, es decir, se produce un suavizado de la serie eliminando o atenuando el efecto de la componente irregular, consiguiendo así identificar la trayectoria que sigue la tendencia. Además, en cuanto mayor sea el orden de la media móvil no es tan inmediato porque, si es demasiado pequeño, puede que no elimine las irregularidades, y si es demasiado grande, podemos perder información sobre cambio de la serie (en el caso extremo tendríamos siempre una tendencia constante).

Un aspecto importante de este método frente al ajuste de la tendencia por mínimos cuadrados es que en este caso el análisis de la tendencia se realiza desde un enfoque local, es decir, la trayectoria de la tendencia se obtiene localmente utilizando algunas observaciones de la serie (no todas). En este sentido, este método permitirá obtener mejores predicciones a corto plazo que si ajustamos la tendencia por mínimos cuadrados. Por contra, no dispondremos de un modelo matemático que nos permita describir la serie y realizar predicciones a largo plazo.

Podemos distinguir dos tipos: medias móviles centradas y medias móviles asimétricas.

Las **medidas móviles centradas** revisten mayor interés, se suelen utilizar para describir la tendencia de la serie de forma más flexible y adecuada usando mínimos cuadrados.

Definición 4.5.2 Dada una serie temporal $\{x_t\}_{t=1,\dots,n}$, se define su **media móvil centrada de orden** $(2p + 1)$ como la serie obtenida mediante la siguiente expresión:

$$MM(2p + 1)_t = \frac{x_{t-p} + x_{t-p+1} + \dots + x_t + \dots + x_{t+p-1} + x_{t+p}}{2p + 1}$$

es decir, en cada instante t , se calcula la media usando las p observaciones anteriores, las p observaciones posteriores y la propia observación en el instante t .

Si el orden de la media móvil es par, entonces debemos realizar dos veces medias móviles para conseguir que sea centrada.

Definición 4.5.3 Dada una serie temporal $\{x_t\}_{t=1,\dots,n}$, se define su **media móvil centrada de orden** $(2p)$ como la serie obtenida mediante la siguiente expresión:

$$MM(2p)_t = \frac{MM(2p)_{t-0.5} + MM(2p)_{t+0.5}}{2}$$

donde:

$$MM(2p)_{t-0.5} = \frac{x_p + \dots + x_{t-1} + x_t + \dots + x_{t+p-1}}{2p}$$

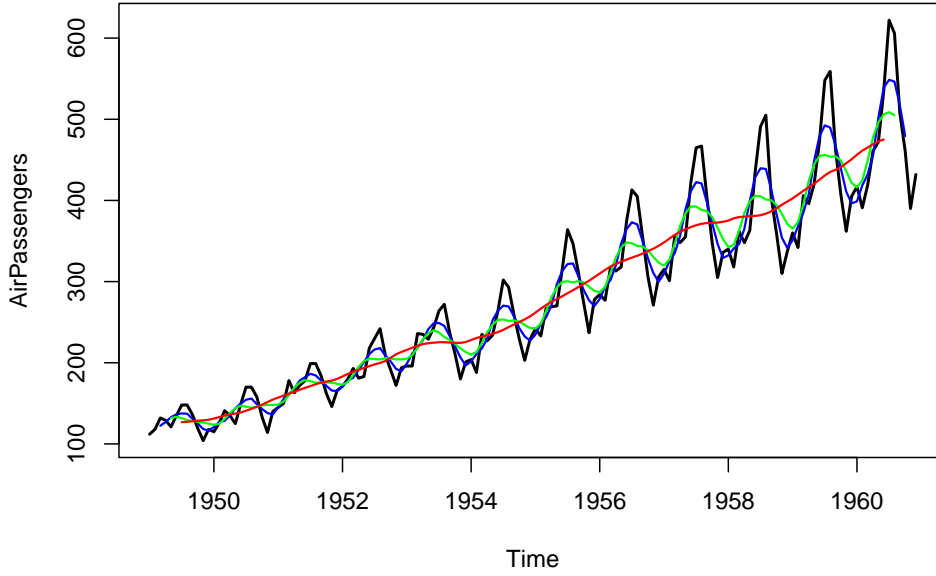
y

$$MM(2p)_{t+0.5} = \frac{x_{t-p+1} + \dots + x_t + x_{t+1} + \dots + x_{t+p}}{2p}$$

Observación 4.5.2 La serie obtenida por medias móviles centradas (ya sea de orden impar $2p + 1$ o de orden par $2p$) pierde los p primeros valores y los p últimos. Es decir, la serie suavizada tendrá $2p$ datos menos que la original, así que no conviene tomar el orden de la media móvil demasiado alto porque se perderían muchos datos.

La siguiente figura muestra cómo aumenta el suavizado de una serie (eliminación de las irregularidades) conforme aumentamos el orden de la media móvil (en negro, la serie original; en azul, medias móviles de orden 5; en verde medias móviles de orden 8; en rojo, medias móviles de orden 12):

```
library(forecast)
ts.plot(AirPassengers, lwd = 2)
lines(ma(AirPassengers, order = 5),
      col = "blue", lwd = 1.5)
lines(ma(AirPassengers, order = 8, centre = TRUE),
      col = "green", lwd = 1.5)
lines(ma(AirPassengers, order = 12),
      col = "red", lwd = 1.5)
```



Definición 4.5.4 Dada una serie temporal $\{x_t\}_{t=1,\dots,n}$, la **media móvil asimétrica de orden p** se define de la siguiente forma:

$$MMA(p)_t = \frac{x_{t-p+1} + \dots + x_{t-1} + x_t}{p}$$

es decir, consiste en realizar la media aritmética de las p últimas observaciones de la serie.

Esta variante de las medias móviles se puede usar para realizar predicciones en un modelo con tendencia aproximadamente constante localmente. Suele ser más adecuado para las predicciones a corto plazo que un ajuste por mínimos cuadrados, debido a su enfoque local para la estimación de la tendencia. Es decir, si disponemos de las observaciones hasta el instante T , la predicción para el instante siguiente viene dada por la media de las p últimas observaciones:

$$\hat{x}_{T+1} = \frac{x_{T-p+1} + \dots + x_{T-1} + x_T}{p}$$

En esta expresión se aprecia que todas las observaciones que intervienen en la media tienen el mismo peso. En el tema siguiente veremos que los métodos de alisado exponencial suelen dar mejores resultados porque se les da más peso a las observaciones más recientes que a las más antiguas.

4.6) Descomposición clásica de series temporales

El objetivo de esta sección es mostrar las técnicas clásicas para descomponer una serie temporal en sus 3 componentes fundamentales (Tendencia-Cielo, Estacional e Irregular). Para ello, es necesario determinar previamente si el esquema de integración de dichas componentes se considera aditivo, o multiplicativo.

Comenzamos con el **enfoque clásico** para extraer las componentes en un esquema aditivo, conocido como el método de la diferencia a la media móvil.

4.6.1) Método de la diferencia a la media móvil

El método de la diferencia a la media móvil es el método que se usa para desestacionalizar una serie que presenta un esquema aditivo:

$$x_t = T_t + S_t + I_t$$

Consideramos una serie temporal con el esquema anterior, presentando componente estacional, y llamaremos L al periodo de la componentes estacional. En muchas ocasiones L será un número par, por ejemplo $L = 12$ para datos mensuales, $L = 4$ para datos trimestrales, etc.

La idea fundamental para intentar eliminar la componentes estacional consiste en tomar medias móviles de orden L (periodo de la componente estacional). Realizamos los siguientes pasos:

- 1) *Calcular la media móvil centrada de orden L para la serie y determinar la componente Tendencia-Ciclo.*

Calculando la media móvil centrada de orden L , donde L representa el periodo de la componente estacional, se consigue eliminar en gran parte el efecto de la componente estacional así como de la componente irregular.

Por tanto, la serie obtenida por medias móviles servirá como estimación de la componente Tendencia-Ciclo (T_t):

$$T_t = MM(L)_t$$

Recordar que si L es par, la media móvil centrada se obtiene realizando el promedio de dos series de medias móviles. Además, si L es par, la serie obtenida por medias móviles tiene L datos menos que la serie original. Por ejemplo, si la serie original se refiere a datos mensuales, tendremos un datos menos cada mes. Y si L es impar, la serie obtenida por medias móviles tiene $L - 1$ datos menos que la serie original.

- 2) *Obtener los índices brutos de variación estacional (IBVE).*

Como el esquema es aditivo:

$$x_t = T_t + S_t + I_t$$

una estimación de las componentes Estacional e Irregular conjuntamente viene dada por:

$$\widehat{S_t + I_t} = x_t - T_t = x_t - MM(L)_t$$

es decir, hay que restar a la serie original la serie obtenida mediante medias móviles de orden L (de ahí el nombre de método de la diferencia a la media móvil).

Cada uno de los valores obtenidos mediante la diferencia anterior se denomina índice bruto de variación estacional en el instante t :

$$(IBVE)_t = \widehat{S_t + I_t} = x_t - MM(L)_t$$

donde el calificativo de "bruto" se debe a que el índice estacional está contaminado por la componente irregular.

- 3) *Obtener los índices de variación estacional no normalizados (IVENN).*

Con el fin de eliminar la componente irregular de los índices anteriores, calcularemos la media de los índices para cada estación. Si disponemos de información correspondiente a K periodos completos y un total de N observaciones, entonces se cumple que:

$$N = L \cdot K$$

Por ejemplo, pensemos en datos mensuales y un total de 5 años completos, es decir, $12 \cdot 5 = 60$ datos. Como ya hemos indicado, al realizar medias móviles se pierden L datos (uno por cada estación, si L es par), de manera que para cada estación tendríamos de exactamente $K - 1$ datos (en lugar de K).

Llamaremos índice de variación estacional no normalizado al promedio de los IBVE para cada estación h :

$$(IVENN)_h = \frac{\sum_{\{t:h(t)=h\}} (IBVE)_t}{K - 1} \quad h = h_1, h_2, \dots, h_L$$

donde $\{t : h(t) = h\}$ denota el conjunto de instantes de tiempo que se corresponden con la estación h . Si K no es muy pequeño, con el promedio conseguimos atenuar enormemente el efecto de la componente irregular, de manera que estos nuevos índices contienen información de la componente estacional pero no de la componente irregular.

Podemos poner una expresión más general para obtener los índices de variación estacional no normalizados, que contemple la posibilidad de que la serie tenga periodos incompletos o bien que L sea impar, en cuyo caso, el número de datos $(IBVE)_t$ disponible para cada estación puede ser diferente. Denotemos por K_h al número de datos $(IBVE)_t$

disponible para cada estación h , entonces:

$$(IVENN)_h = \frac{\sum_{\{t:h(t)=h\}} (IBVE)_t}{K_h} \quad h = h_1, h_2, \dots, h_L$$

4) *Obtener los índices de variación estacional normalizados (IVE) y determinar la componente Estacional*

Si una serie no tiene componente estacional, en un modelo aditivo se tendría que $S_t \equiv 0$. Por otra parte, es razonable suponer que la componente estacional no debe afectar al nivel de la serie, de manera que en el modelo aditivo la media de la componente estacional debería ser 0.

Por tanto, definimos los índices de variación estacional normalizados para cada estación h como

$$(IVE)_h = (IVENN)_h = \frac{\sum_{h \in \{h_1, \dots, h_L\}} (IVENN)_h}{L}$$

aunque existen otras formas de normalizar los índices de variación estacional.

Estos índices suponen una **estimación de la componente estacional** de la serie para cada estación h . Es decir, en este paso determinamos la componente estacional para cada instante de tiempo t , haciéndola coincidir con el índice de variación estacional para la estación, $h(t)$, correspondiente al instante t :

$$S_t = (IVE)_{h(t)}$$

5) *Desestacionalizar la serie.*

En el modelo aditivo, la desestacionalización de la serie se consigue restando a la serie original los índices de variación estacional:

$$D_t = x_t - S_t = x_t - (IVE)_{h(t)} \quad (\text{serie desestacionalizada})$$

6) *Determinar la componente Irregular.*

Una vez extraídas la componente Estacional (S_t) y la Tendencia-Ciclo (T_t), la componente Irregular (I_t) se obtiene de forma inmediata mediante:

$$I - t = x_t - T_t - S_t$$

La situación ideal sería que dicha componente irregular se comportara como un ruido blanco.

A continuación, mostramos el **enfoque clásico** para extraer las componentes de una serie temporal si el esquema es multiplicativo.

4.6.2) Método de la razón a la media móvil

El método de la razón a la media móvil es el método que se usa para desestacionalizar una serie que presenta un esquema multiplicativo:

$$x_t = T_t \times S_t \times I_t$$

Los pasos para desestacionalizar la serie y extraer las componentes son similares al caso aditivo, pero realizando cocientes en lugar de diferencias:

1) *Calcular la media móvil centrada de orden L para la serie y determinar la componente Tendencia-Ciclo.*

$$T_t = MM(L)_t$$

2) *Obtener los índices brutos de variación estacional (IBVE).*

Como el esquema es multiplicativo, una estimación de las componentes Estacional e Irregular conjuntamente viene

dada por:

$$\widehat{S_t \times I_t} = \frac{x_t}{\widehat{T_t}} = \frac{x_t}{MM(L)_t}.$$

3) *Obtener los índices de variación estacional no normalizados (IVENN).*

Llamaremos índice de variación estacional no normalizado al promedio de los *IBVE* para cada estación h . Denotemos por K_h al número de datos $(IBVE)_t$ disponible para cada estación h , entonces:

$$(IVENN)_h = \frac{\sum_{\{t:h(t)=h\}} (IBVE)_t}{K_h} \quad h = h_1, h_2, \dots, h_L$$

4) *Obtener los índices de variación estacional normalizados (IVE) y determinar la componente Estacional*

Si una serie no tiene componente estacional, en un modelo multiplicativo se tendría que $S_t \equiv 1$. Por otra parte, es razonable suponer que la componente estacional no debe afectar al nivel de la serie, de manera que en el modelo multiplicativo la media de la componente estacional debería ser 1.

Sin embargo, los índices calculados en el apartado anterior pueden no tener media uno. Por tanto, definimos los índices de variación estacional normalizados para cada estación h como:

$$(IVE)_h = \frac{(IVENN)_h}{\frac{\sum_{h \in \{h_1, \dots, h_L\}} (IVENN)_h}{L}}$$

los cuales suponen una **estimación de la componente estacional** de la serie para cada estación h . Es decir, en este paso determinamos la componente estacional para cada instante de tiempo t , haciéndola coincidir con el índice de variación estacional para la estación, $h(t)$, correspondiente al instante t :

$$S_t = (IVE)_{h(t)}$$

5) *Desestacionalizar la serie.*

En el modelo multiplicativo, la desestacionalización de la serie se consigue dividiendo la serie por los índices de variación estacional:

$$D_t = \frac{x_t}{(IVE)_{h(t)}}$$

6) *Determinar la componente Irregular.*

Una vez extraídas la componente estacional S_t , y la Tendencia T_t , la componente Irregular o restante se obtiene de forma inmediata mediante:

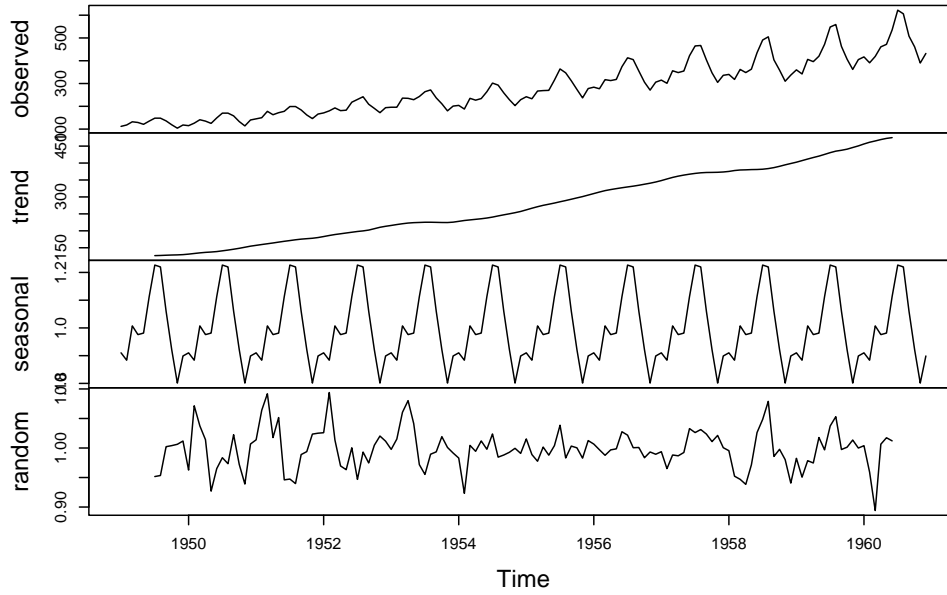
$$I_t = \frac{x_t}{T_t \times S_t}$$

La situación ideal sería que dicha componente se comportara como un ruido gaussiano de media 1, o bien como la exponencial de un ruido gaussiano de media cero.

A modo de ejemplo, a continuación mostramos una serie temporal y las componentes que resultan con la **descomposición clásica**:

```
plot(decompose(AirPassengers, type = 'multiplicative'))
```

Decomposition of multiplicative time series



Nota: Podemos comprobar que la componente estacional se mantiene estable a lo largo del tiempo. Es decir, todos los meses de enero toman el mismo valor independientemente del año. Lo mismo sucede con los febreros, marzos, ..., diciembres. Esta situación es característica para la descomposición clásica.

4.7) Predicciones con Análisis Clásico y limitaciones

En el enfoque clásico, la forma de realizar predicciones se basa en la descomposición de la serie en sus componentes fundamentales (tendencia, estacionalidad e irregular). Extraídas las componentes, se predice cada una de ellas y la predicción de la serie original se obtiene combinando dichas predicciones según se trate de un esquema aditivo o multiplicativo.

¿Cómo realizar predicciones de la componente irregular I_t ?

Como se trata de la parte aleatoria de la serie, las predicciones de la componente irregular valdrán cero en el esquema aditivo y uno en el multiplicativo.

Es decir, para todo instante t , tendremos que la predicción de la componente irregular en dicho instante es:

$$\hat{I}_t = 0 \text{ (esquema aditivo)}, \hat{I}_t = 1 \text{ (esquema multiplicativo)}$$

¿Cómo realizar predicciones de la Componente Estacional S_t ?

Bastará con repetir el valor obtenido para cada estación. Por ejemplo, si tenemos datos mensuales y la componente estacional extraída asigna un valor de 50 a los enero, entonces la predicción de la componente estacional será siempre de 50 para los futuros enero. De la misma forma se realizaría la predicción para el resto de meses, pues el enfoque clásico considera componente estacional estable.

Si la serie original tiene periodo L y disponemos de observaciones hasta el instante T , entonces la predicción de la componente estacional en el instante $T + m$ es:

$$\hat{S}_{T+m} = S_{T+m-L} \quad m = 1, 2, \dots, L$$

donde S_{T+m-L} representa el valor de la componente estacional en el instante $T + m - L$, es decir, el valor del IVE en la estación que corresponde a dicho instante.

¿Cómo realizar predicciones de la componente Tendencia-Ciclo T_t ?

La componente tendencia puede modelizarse con cualquier método para serie sin estacionalidad. Por ejemplo, realizando un ajuste por mínimos cuadrados como predicto (predictores) el tiempo o potencias del mismo. El modelo obtenido para

la tendencia usando mínimos cuadrados es adecuado para representar el comportamiento de la serie y captar los aspectos más permanentes del fenómeno estudiado (predicciones a medio o largo plazo). Sin embargo, no es aconsejable para las predicciones a corto plazo.

Por ejemplo, si la Tendencia es lineal respecto al tiempo:

$$T_t = b_0 + b_1 \cdot t,$$

podemos estimar la ecuación de dicha recta usando regresión lineal simple. Por tanto, en este caso las predicciones de la tendencia en cualquier instante t , se obtendrán mediante:

$$\hat{T}_t = \hat{b}_0 + \hat{b}_1 \cdot t$$

Y si la Tendencia es polinómica:

$$\hat{T}_t = \hat{b}_0 + \hat{b}_1 \cdot t + \hat{b}_2 \cdot t^2 + \dots + \hat{b}_k \cdot t^k$$

también podemos estimar los coeficientes del modelo usando regresión lineal simple.

Obsérvese que se trata de un *enfoque global* del análisis de la tendencia, el cual permite obtener un modelo matemático que modeliza la tendencia de la serie usando todas las observaciones de la serie, y además todas con el mismo peso.

Los procedimientos adecuados para obtener la tendencia desde un *enfoque local* que permita realizar predicciones adecuadas a corto plazo son, por ejemplo, el método de las medias móviles asimétricas o los métodos de alisado exponencial, que veremos en un tema posterior.

¿Cómo realizar predicciones de la serie original?

Si disponemos de T observaciones de la serie y queremos realizar predicciones para los L instantes siguiente (L es el periodo), en el caso de esquema aditivo tendremos:

$$\hat{x}_{T+m} = \hat{T}_{T+m} + \hat{S}_{T+m} \quad m = 1, 2, \dots, L$$

y en el caso de esquema multiplicativo tendremos:

$$\hat{x}_{T+m} = \hat{T}_{T+m} \times \hat{S}_{T+m} \quad m = 1, 2, \dots, L$$

Limitaciones del Análisis Clásico

El análisis clásico de series temporales tiene dos usos principalmente:

- *Describir el comportamiento de la serie temporal* (análisis descriptivo), haciendo uso por ejemplo de la descomposición de la serie en las componentes Tendencia, Estacionalidad e Irregular.
- *Realizar predicciones a medio o largo plazo.*

Como método descriptivo sigue siendo interesante su uso gracias a la sencillez de los métodos expuestos. Del mismo modo, cuando sólo nos interesa estudiar el comportamiento futuro de la serie a largo plazo, es decir, estudiar cómo evolucionaría la serie a grandes rasgos, también interesa emplear un análisis clásico.

Sin embargo, el enfoque global de los métodos clásicos impide realizar predicciones adecuadas a corto plazo, siendo ésta la principal limitación de un análisis clásico. Además, los métodos clásicos proporcionan demasiada rigidez al modelo, perdiendo en ocasiones información relevante.

Para suplir las carencias del método clásico, se propusieron nuevos métodos como la descomposición STL, que veremos a continuación, o los métodos de alisado exponencial, que veremos en el tema posterior.

4.8) La descomposición STL

El método STL (*Seasonal and Trend decomposition using LOESS*) es una técnica alternativa a la descomposición clásica utilizada para descomponer una serie temporal en sus componentes básicas: tendencia, estacionalidad e irregular. Se basa en el suavizado de datos mediante LOESS (*locally estimated scatterplot smoothing*), que permite la obtención de las componentes forma más versátil y flexible.

Mientras que la descomposición clásica considera que la componente estacional es completamente estable, la descomposición STL permite que la componente estacional varíe a lo largo del tiempo. La diferencia fundamental radica en el proceso de desestacionalización de la serie: mientras que el enfoque clásico usa medias móviles, el STL usa el suavizado local LOESS.

Importante: El método STL considera que las componentes de la serie se combinan siguiendo un **esquema aditivo**. En el caso de esquema multiplicativo, primero debe realizarse una transformación logarítmica para convertirlo en esquema aditivo.

¿En qué consiste el método LOESS?

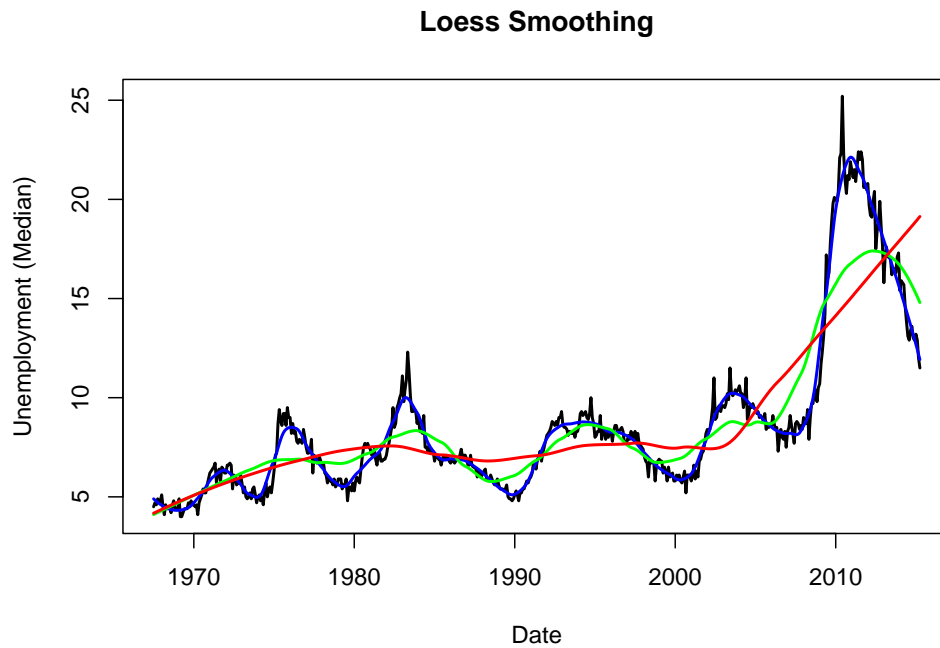
LOESS realiza agresiones locales a lo largo de los datos. En cada punto x_i de la serie, se ajusta un polinomio (generalmente de primer o segundo grado) utilizando solo los datos cercanos a ese punto. Los puntos cercanos se definen por una ventana de vecindad, cuyo tamaño está determinado por un parámetro de suavizado. Este parámetro puede ser una proporción del total de puntos (ancho de banda o bandwidth). Los puntos dentro de la ventana de vecindad se ponderan de acuerdo con su distancia al punto x_i . Los puntos más cercanos tienen mayor peso en la estimación del polinomio local que los puntos más lejanos. Para cada punto x_i de la serie, se calcula un valor ajustado \hat{x} mediante la regresión local ponderada. Este proceso se repite para cada punto de la serie, resultando en una curva suavizada.

LOESS es una técnica eficaz para suavizar los datos y detectar tendencias locales sin imponer una estructura global rígida. Es especialmente útil para datos que no siguen una forma predefinida y permite capturar patrones locales mediante ajustes flexibles y ponderados.

A continuación mostramos un ejemplo del método LOESS, así como el efecto del parámetro de suavizado. Para ello usaremos el conjunto de datos "economics" del paquete `ggplot2`, y suavizaremos la serie correspondiente al tiempo de desempleo (variable `uempmed`, que representa la mediana del número de días sin empleo). Observamos que la serie original (en negro) contiene numerosas irregularidades, que se suavizan al aplicar LOESS con parámetros de suavizado 10% (en azul), 30% (en verde) y 50% (en rojo). Indicar que si la serie original tiene por ejemplo 100 observaciones y se usa un suavizado del 50% (parámetro `span=0.5`), entonces para cada punto x_i de la serie, la ventana de vecindad se construye con 25 datos a la izquierda de x_i y otros 25 datos a la derecha de x_i .

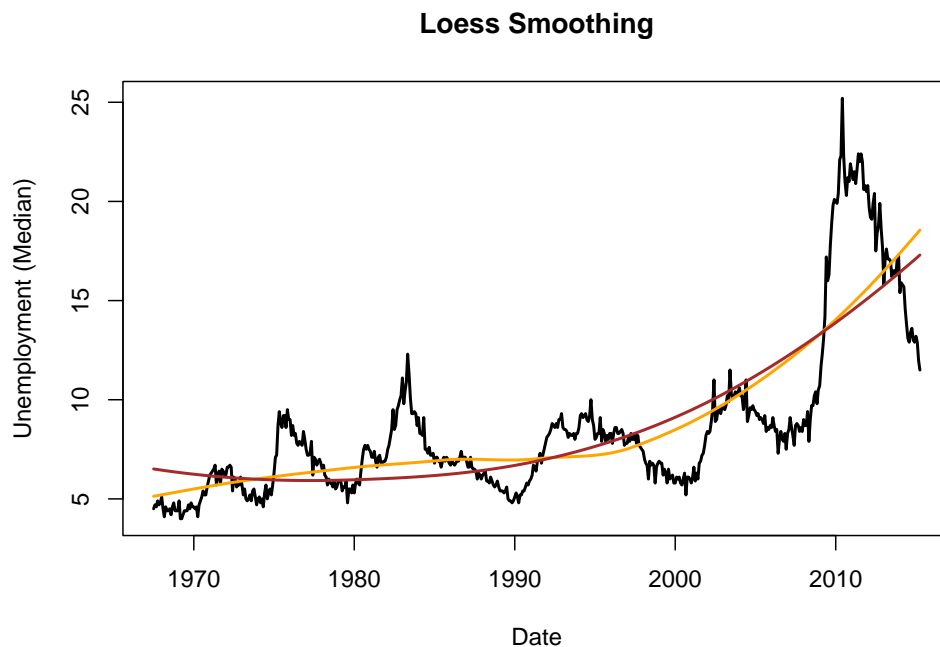
```
library(ggplot2)
datos <- economics
datos$index <- 1:nrow(datos)
loess10 <- loess(uempmed ~ index, data=datos, span = 0.10)
loess30 <- loess(uempmed ~ index, data=datos, span = 0.30)
loess50 <- loess(uempmed ~ index, data=datos, span = 0.50)

plot(datos$uempmed, x=datos$date, type="l", main="Loess Smoothing",
      xlab="Date", ylab="Unemployment (Median)", lwd = 2)
lines(predict(loess10), x=datos$date, col="blue", lwd = 2)
lines(predict(loess30), x=datos$date, col="green", lwd = 2)
lines(predict(loess50), x=datos$date, col="red", lwd = 2)
```



Y para valores altos del parámetro de suavizado tendremos un suavizado global en lugar de local (obsérvese que la curva marrón se corresponde con la ecuación de una parábola que ajusta globalmente los datos de la serie):

```
loess100 <- loess(uempmed ~ index, data = datos, span = 1)
loess200 <- loess(uempmed ~ index, data = datos, span = 2)
plot(datos$uempmed, x=datos$date, type="l", main="Loess Smoothing",
      xlab="Date", ylab="Unemployment (Median)", lwd = 2)
lines(predict(loess100), x=datos$date, col="orange", lwd = 2)
lines(predict(loess200), x=datos$date, col="brown", lwd = 2)
```



Resumen de las etapas en la descomposición STL

- 1) Realizar un suavizado de la serie original mediante LOESS con un ancho de banda grande, con el fin de captar las variaciones a largo plazo (eliminando así las fluctuaciones debidas a la componente estacional). El resultado de este suavizado nos dará una primera estimación de la Tendencia-Ciclo (T_t).

En R, se puede fijar el parámetro de ancho de banda correspondiente a la tendencia (`t.window`) o dejar que tome un valor por defecto. En R, dicho parámetro además debe tomar un valor impar y representa el número de observaciones consecutivas usadas para estimar la tendencia. Valores pequeños permiten cambios más rápidos.

- 2) Restar a la serie original la Tendencia obtenida en la etapa anterior. Es decir, hacemos $x_t - T_t$, para conseguir una serie temporal con tendencia constante.
- 3) Dividir la serie de la etapa anterior (sin tendencia) en subseries: Supongamos que la serie original tiene estacionalidad de periodo L , entonces hay que dividir la serie de la etapa anterior en L subseries, una para cada estación. Por ejemplo, para datos mensuales durante varios años, tendremos que $L = 12$. Esta etapa implica considerar una serie para cada estación, es decir, una serie con los datos de enero, otra con los datos de febrero, ..., y otra con los datos de diciembre.
- 4) Obtener una estimación de la componente Estacional (S_t) mediante suavizado LOESS: En esta etapa se fija el valor del parámetro `s.window` correspondiente al ancho de banda del suavizado para la estacionalidad. En R, dicho parámetro debe tomar un valor impar y no tiene valor por defecto. Para cada subserie de la etapa anterior, se aplica el suavizado LOESS y `s.window` representa el número de periodos (años) consecutivos usados en la estimación de cada valor de la componentes estacional. Valores pequeños permiten cambios más rápidos. Y en el caso de usar todos los periodos se considera estacionalidad estable (igual que en un enfoque clásico). Al igual que en la descomposición clásica, las estimaciones resultantes se normalizan para conseguir que la media de cada periodo (año) completo sea cero (recordar que se supone esquema aditivo).
- 5) Obtener la componente Irregular: Para ello bastará restar a la serie original, las estimaciones obtenidas para la estacionalidad y la tendencia ($I_t = x_t - T_t - S_t$).

Nota: La extracción de las componentes puede ser mejorada siguiendo un ciclo iterativo como el siguiente.

Con la tendencia inicial T_t y la estacionalidad inicial S_t estimadas, se sigue un ciclo iterativo para ajustar y mejorar las estimaciones:

- a) Ajuste de la Tendencia: eliminar el efecto de la estacionalidad de la serie original (serie desestacionalizada)

$$x_t - S_t$$

y obtener una nueva estimación de la componente Tendencia-Ciclo (nueva T_t) mediante un nuevo suavizado LOESS sobre la serie desestacionalizada. En este proceso se requiere nuevamente del parámetro `t.window` correspondiente al ancho de banda del suavizado para capturar la tendencia subyacente.

- b) Ajuste de la Estacionalidad: Se elimina la tendencia ajustada de la serie

$$x_t - T_t$$

Las subseries estacionales se suavizan de nuevo para recalcular S_t , teniendo en cuenta nuevamente un parámetro del ancho de banda para la estacionalidad (`s.window`).

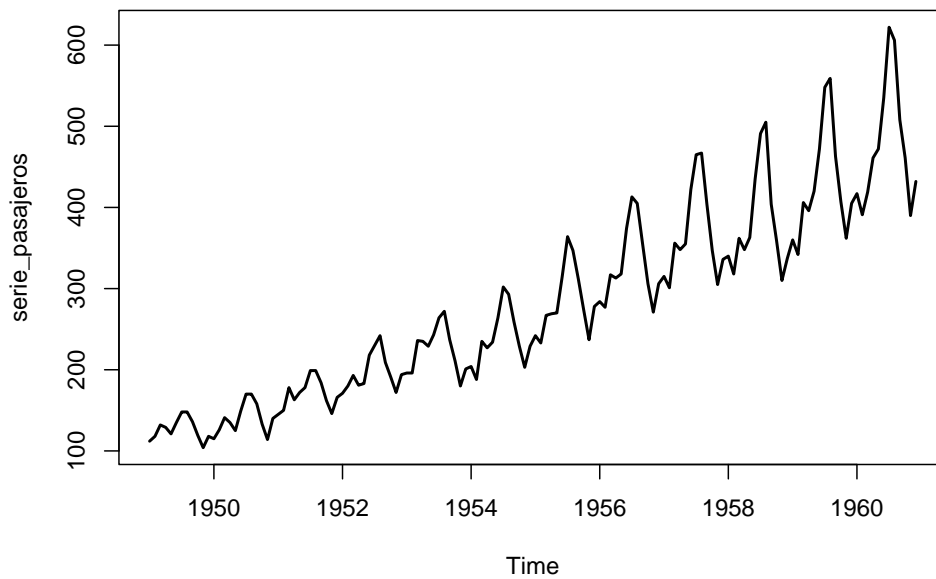
- c) Ajuste de la componente Irregular: Se recalcula como

$$I_t = x_t - T_t - S_t$$

Este ciclo se repite varias veces para mejorar la precisión de las componentes.

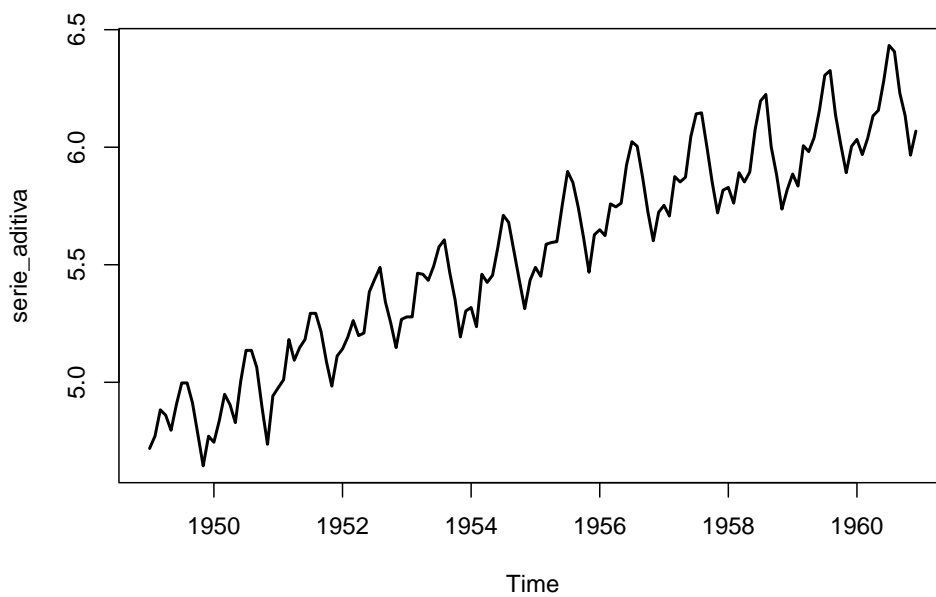
Veamos como ejemplo la descomposición STL sobre la serie `AirPassengers` incluida en R.

```
serie_pasajeros <- AirPassengers
ts.plot(serie_pasajeros, lwd = 2)
```



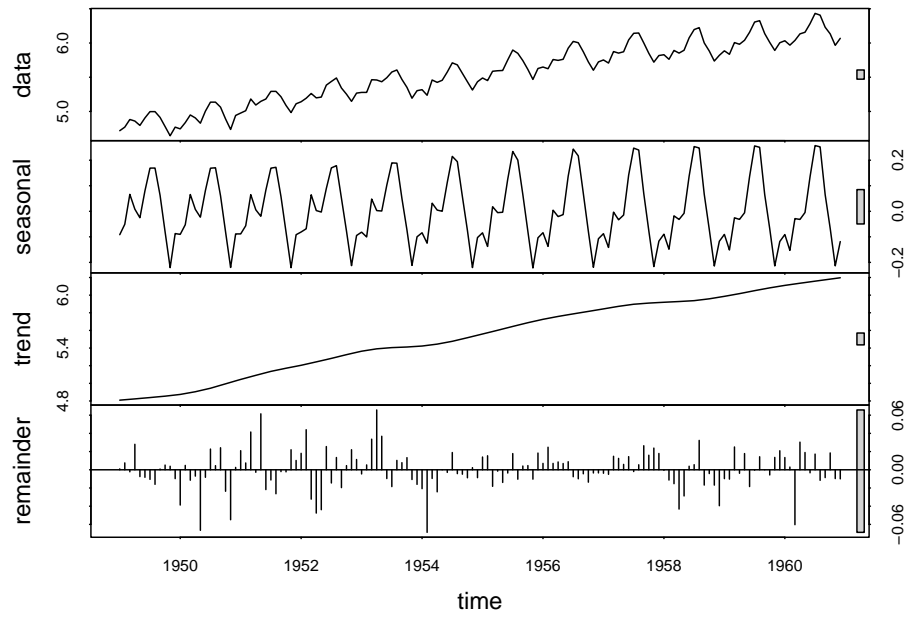
Observamos esquema multiplicativo, así que transformamos los datos para conseguir esquema aditivo, necesario para descomposición STL.

```
serie_aditiva <- log(serie_pasajeros)
ts.plot(serie_aditiva, lwd = 2)
```



Realizando ahora la descomposición STL sobre la serie aditiva, fijando `s.window` (ancho de banda de la estacionalidad):

```
modelo_STL <- stl(serie_aditiva, s.window = 7)
plot(modelo_STL)
```



Nota: Obsérvese que el aspecto de la componente estacional varía a lo largo del tiempo, a diferencia de lo que sucede en el enfoque clásico.

Hoja 4: Conceptos Básicos de Series Temporales

- 1) Los siguientes datos se corresponden con las ventas cuatrimestrales (en miles de unidades) de una empresa que elabora latas de conserva.

Año:	2021	2021	2021	2022	2022	2022	2023	2023	2023
Cuatrimestre:	C1	C2	C3	C1	C2	C3	C1	C2	C3
Ventas (miles):	23	30	14	31	37	20	34	41	26

Se pretende realizar una descomposición clásica de la serie y se pide:

- Representar la serie en estudio y describir sus aspectos más relevantes. ¿Presenta componente estacional? ¿Cuál es el periodo L ?
- Justificar si, para la serie en estudio, es más adecuado un esquema aditivo o bien uno multiplicativo.
- Completar la siguiente tabla para obtener las componentes Tendencia, Estacionalidad e Irregular por el método de la diferencia a la media móvil o por el método de la razón a la media móvil, según se considere.

Año:	2021	2021	2021	2022	2022	2022	2023	2023	2023
Cuatrimestre:	C1	C2	C3	C1	C2	C3	C1	C2	C3
Ventas (miles):	23	30	14	31	37	20	34	41	26
Tendencia (T_t): Medias Móviles Orden = ¿ ?									
IBVE:									
IVE no normalizados:									
Comp. Estacional: $S_h = IVE_h$									
Serie Desestacionalizada (D_t):									
Comp. Irregular (I_t):									

- Explica cómo podrías predecir las ventas futuras. Supongamos que la tendencia se ajusta mediante la ecuación de la recta siguiente: $T(t) = 19.5 + 1.8t$, $t = 1, 2, \dots$
 - Obtener las predicciones de los cuatrimestres C1, C2 y C3 del año 2024.
 - Obtener los valores ajustados de la serie, es decir, los valores predichos en el rango de datos observado.
 - Olvidemos la descomposición estacional realizada y consideremos el Método Ingenuo I. Calcula las predicciones que se obtendrían con dicho método sólo para los datos del 2023, así como las medidas de error MAE y RMSE correspondientes.
- 2) Repetir el problema anterior suponiendo que los datos de ventas fueron los siguientes:

Año:	2021	2021	2021	2022	2022	2022	2023	2023	2023
Cuatrimestre:	C1	C2	C3	C1	C2	C3	C1	C2	C3
Ventas (miles):	10	21	4	21	39	8	30	61	12

Y para el apartado (4), supondremos que la tendencia se ajusta mediante la ecuación de la recta siguiente: $T(t) = 4.2 + 3.85 \cdot t$, $t = 1, 2, \dots$

Año:	2021	2021	2021	2022	2022	2022	2023	2023	2023
Cuatrimestre:	C1	C2	C3	C1	C2	C3	C1	C2	C3
Ventas (miles):	10	21	4	21	39	8	30	61	12
Tendencia (T_t): Medias Móviles Orden = ¿ ?									
IBVE:									
IVE no normalizados:									
Comp. Estacional: $S_h = IVE_h$									
Serie Desestacionalizada (D_t):									
Comp. Irregular (I_t):									

- 3) Los siguientes datos se corresponden con las pernoctaciones trimestrales (en miles de personas) de una ciudad turística.

Año	2021	2021	2021	2021	2022	2022	2022	2022	2023	2023	2023	2023
Trimestre:	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Pernoctaciones (miles):	10	21	4	25	21	39	8	51	30	61	12	73

Se pretende realizar una descomposición clásica de la serie y se pide:

- 1) Representar la serie en estudio y describir sus aspectos más relevantes. ¿Presenta componente estacional? ¿Cuál sería el periodo L ?
- 2) Justificar si, para la serie en estudio, es más adecuado un esquema aditivo o bien uno multiplicativo.
- 3) Completar la siguiente tabla para obtener las componentes Tendencia, Estacionalidad e Irregular por el método de la diferencia a la media móvil o por el método de la razón a la media móvil, según se considere.

Año	2021	2021	2021	2021	2022	2022	2022	2022	2023	2023	2023	2023
Trimestre:	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Pernoctaciones (miles):	10	21	4	25	21	39	8	51	30	61	12	73
IBVE:												
IVE no normalizados:												
Comp. Estacional: $S_h = IVE_h$												
Serie Desestacionalizada (D_t):												
Comp. Irregular (I_t):												

4) Explica cómo podrías predecir las ventas de los trimestre T1, T2, T3 y T4 del año 2024.

4) Repetir el problema anterior suponiendo que los datos de pernoctaciones fueron los siguientes:

Año	2021	2021	2021	2021	2022	2022	2022	2022	2023	2023	2023	2023
Trimestre:	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Pernoctaciones (miles):	23	30	14	32	31	37	20	39	34	41	26	43

Año	2021	2021	2021	2021	2022	2022	2022	2022	2023	2023	2023	2023
Trimestre:	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Pernoctaciones (miles):	23	30	14	32	31	37	20	39	34	41	26	43
IBVE:												
IVE no normalizados:												
Comp. Estacional: $S_h = IVE_h$												
Serie Desestacionalizada (D_t):												
Comp. Irregular (I_t):												

5) Resolver los problemas anteriores usando R para comprobar si los resultados obtenidos a mano son correctos. Obsérvese que, en lugar de usar la función "decompose()" de R, la descomposición estacional debe implementarse paso a paso para poder completar las tablas. Posteriormente puede usarse la función "decompose()" de R para comprobar resultados.

6) Responder a las siguientes cuestiones:

a) ¿Qué es una serie temporal? ¿Cuál es su relación con los procesos estocásticos?

b) Explicar en qué consiste el proceso de diferenciación (tomar diferencias) en una serie temporal y para qué sirve.

- c) Explicar en qué consiste el proceso de realizar medias móviles en una serie temporal y para qué sirve.
- d) ¿Cuáles son las componentes de una serie temporal según el enfoque clásico? ¿Cómo suelen combinarse las componentes para dar lugar a la serie? Es decir, ¿qué esquemas son los habituales en el enfoque clásico?
- e) Utilidad y limitaciones del análisis clásico de series temporales.
- f) Similitudes y diferencias entre la descomposición clásica de series y la descomposición STL.

Tema 5: Métodos de Alisado Exponencial

5.1) Introducción

Una de las críticas que se les hace a estos métodos clásicos es que no se adaptan a lo largo del tiempo de forma natural: en los métodos clásicos, la tendencia y la estacionalidad se estiman una sola vez, usando todas las observaciones y todas con el mismo peso. Sin embargo, parece más lógico que las observaciones más recientes tengan un mayor peso en las predicciones futuras y que las estimaciones de la tendencia y estacionalidad se actualicen conforme se obtienen nuevas observaciones.

Una familia de modelos que aparecen en la década de los años 60, intenta solucionar este problema. Se les conoce como **técnicas de alisado exponencial**, y se trata de técnicas bastante sencillas. Por este motivo, y a pesar de los años transcurridos, siguen siendo utilizadas en ciertas actividades de pronóstico, principalmente para realizar **predicciones a corto plazo**. Aplicar técnicas más sofisticadas no siempre es justa.

El estudio de una serie temporal viene marcado por las diferentes metodologías empleadas en su tratamiento. La metodología que presentamos en este tema trata de explicar la trayectoria de una serie observada a través de la información contenida en los datos históricos, es decir, intenta capturar el comportamiento sistemático que muestra el pasado de la serie y en base a ello realizar predicciones respecto al futuro.

Los métodos de alisado exponencial se clasifican dependiendo de si la serie en estudio presenta o no estacionalidad. En este tema veremos los siguientes:

1) Si la serie no presenta estacionalidad:

- a) Método de alisado exponencial simple: se usa cuando la tendencia se considera constante (localmente).
- b) Método de alisado exponencial doble: se usa cuando la tendencia se considera lineal (localmente).
- c) Método de Holt: se usa cuando la tendencia se considera lineal (localmente).

2) Si la serie presenta estacionalidad:

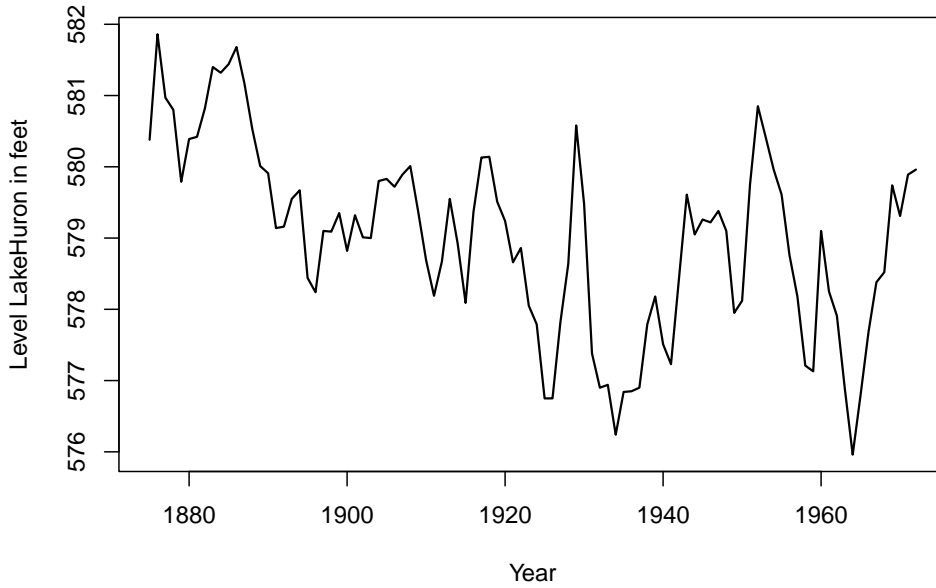
- a) Método de Holt-Winters para modelos multiplicativos: se usa cuando la serie presenta estacionalidad, la tendencia se considera lineal localmente y ambas componentes se integran con esquema multiplicativo.
- b) Método de Holt-Winters para modelos aditivos: se usa cuando la serie presenta estacionalidad, la tendencia se considera lineal localmente y ambas componentes se integran con esquema aditivo.

5.2) Alisado Exponencial Simple

Es el método más sencillo de alisado exponencial. Se aplica cuando la serie en estudio no presenta estacionalidad y la tendencia de la serie se considera constante (localmente). Es decir, no se advierte un patrón claro de tendencia ni estacionalidad.

Un ejemplo ilustrativo sería la serie con el siguiente gráfico:

```
ts.plot(LakeHuron,  
        gpars = list(xlab = "Year",  
                     ylab = "Level LakeHuron in feet",  
                     lwd = 1.5))
```



En este caso suponemos que la serie en estudio se expresa de la siguiente forma en cada instante:

$$\begin{cases} x_t = T_t + I_t = a_t + I_t & \text{modelo aditivo} \\ x_t = T_t \times I_t = a_t \times I_t & \text{modelo multiplicativo} \end{cases}$$

donde hemos considerado que la componente de tendencia es constante en cada instante t (constante localmente): $T_t = a_t$, con $a_t \in \mathbb{R}$.

Obsérvese que a_t representa la *altura* de la serie en cada instante t cuando realizamos su representación gráfica. Por este motivo se dice que a_t representa el **nivel de la serie**.

Al no disponer de componente estacional y suponer tendencia constante, nuestro objetivo se reduce a estimar el nivel de la serie en cada instante.

Como el resto de componente estacional y suponer tendencia constante, nuestro objetivo se reduce a estimar el nivel de la serie en cada instante.

Como el resto de métodos de alisado exponencial, el alisado exponencial simple se caracteriza por dar más peso a las observaciones más recientes que a las observaciones más antiguas a la hora de realizar predicciones.

Dada una serie de observaciones, $\{x_t\}_{t=1,2,\dots,n}$, en el alisado exponencial simple se obtiene otra serie denominada serie alisada (smoothing series), que denotaremos por $\{A_t\}_{t=1,2,\dots,n}$, y que representa la estimación del nivel de la serie (a_t) en cada instante.

La serie alisada en cada instante t se obtiene a partir de la serie original mediante la siguiente expresión:

$$A_t = \hat{a}_t = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \dots + \alpha(1 - \alpha)^{t-1} x_1 \quad (14)$$

donde $\alpha \in (0, 1)$ se conoce como *parámetro de alisado*.

Es decir, la serie alisada se obtiene, en cada instante, como una media ponderada de las observaciones actual y anteriores de la serie original, donde las observaciones más recientes poseen más peso que las observaciones más antiguas. De hecho, el nombre de alisado exponencial se debe a que el peso de las observaciones decrece exponencialmente a medida que nos separamos del último instante observado (los valores de los pesos $\alpha, \alpha(1 - \alpha), \alpha(1 - \alpha)^2, \dots, \alpha(1 - \alpha)^{t-1}$ decrecen exponencialmente al ser $\alpha \in (0, 1)$).

Otra de las características de los métodos de alisado exponencial es que las series alisadas se pueden obtener de manera

recurrente, pudiendo actualizarse cada vez que disponemos de una nueva observación.

Teniendo en cuenta la expresión (14), la serie alisada en el instante $t - 1$, viene dada por:

$$A_{t-1} = \alpha x_{t-1} + \alpha(1 - \alpha)x_{t-2} + \alpha(1 - \alpha)^2 x_{t-3} + \cdots + \alpha(1 - \alpha)^{t-2} x_1$$

de manera que:

$$(1 - \alpha)A_{t-1} = \alpha(1 - \alpha) + \alpha(1 - \alpha)^2 x_{t-2} + \alpha(1 - \alpha)^3 x_{t-3} + \cdots + \alpha(1 - \alpha)^{t-1} x_1$$

y teniendo en cuenta nuevamente (14), se obtiene la siguiente expresión que permite calcular la serie alisada de forma recurrente:

$$\boxed{A_t = \alpha x_t + (1 - \alpha)A_{t-1} \quad \forall t, \text{ con } \alpha \in [0, 1]} \quad (15)$$

Es decir, la estimación del nivel en cada instante se puede obtener de forma recurrente:

$$\boxed{a_t = \alpha x_t + (1 - \alpha)a_{t-1} \quad \forall t, \text{ con } \alpha \in [0, 1]}$$

Nota: haciendo abuso de la notación, hemos usado a_t en la serie alisada para denotar a la estimación del nivel. Además, consideramos $\alpha \in [0, 1]$, permitiendo que α tome los valores 0 y 1 en las fórmulas recurrentes para facilitar la interpretación del parámetro de alisado.

Para aplicar la fórmula recursiva anterior necesitamos disponer, por una parte, del valor del parámetro de alisado α , y por otra parte, del valor inicial de la serie alisada A_0 (nivel inicial de la serie a_0).

Como valor inicial del nivel se suele usar el primer valor de la serie original, es decir:

$$a_0 = A_0 = x_1$$

o bien la media de las primeras observaciones, es decir:

$$a_0 = A_0 = \frac{x_1 + \cdots + x_k}{k} \text{ con } k \text{ pequeño}$$

Con respecto al parámetro de alisado α , comentar que si se utiliza un valor próximo a uno ($\alpha \simeq 1$) se da mucho peso a la última observación de la serie y muy poco peso a las pasadas. Por el contrario, si se utiliza un parámetro de alisado próximo a cero ($\alpha \simeq 0$), se le da mucho peso al conjunto de las observaciones pasadas y poco a la última observación (véase (15)).

Se ha comprobado empíricamente que un valor de $\alpha = 0.2$ suele dar buenos resultados. Sin embargo, la mayoría del software permite seleccionar el valor de *alpha* que minimiza la suma de cuadrados de los errores de estimación. En R podremos seleccionar como parámetro de suavizado el valor de α óptimo según dicho criterio (minimiza la suma de cuadrados de los errores de predicción para el instante siguiente).

Nuestro fin último es realizar predicciones de la serie para instantes futuros. Si disponemos de información de la serie hasta el instante T , $\{x_t\}_{t=1,2,\dots,T}$, las predicciones de la serie usando el método de alisado exponencial simple vienen dadas por:

$$\boxed{\begin{aligned} \hat{x}_{T+1/T} &= a_T = A_T \\ \hat{x}_{T+2/T} &= a_T = A_T \\ &\vdots \\ \hat{x}_{T+m/T} &= a_T = A_T \quad \forall m = 1, 2, 3, \dots \end{aligned}}$$

donde $\hat{x}_{T+m/T}$ denota el valor de la predicción en el instante $T + m$ cuando disponemos de observaciones de la serie hasta el instante T . Es decir, la predicción de la serie en el instante siguiente viene dada por la estimación del nivel en el último instante observado, o lo que es lo mismo, el valor de la serie alisada en el último instante observado. Y ésta también sería la predicción para instantes posteriores.

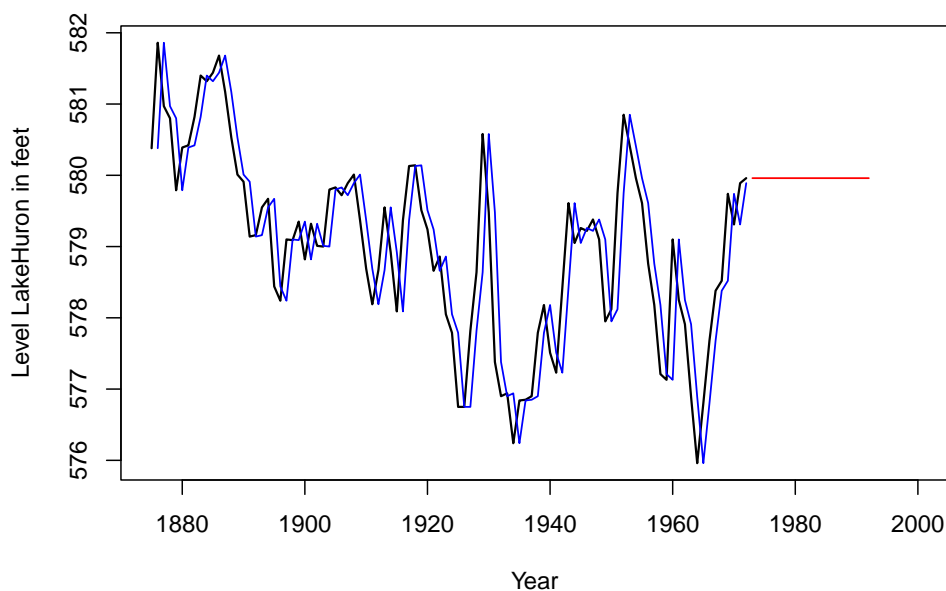
Sin embargo, a medida que disponemos de información nueva (nuevas observaciones), se calculan los valores de la serie alisada para los nuevos instantes y por lo tanto se revisan las predicciones de manera automática.

Podemos decir que el alisado exponencial simple estima en cada instante t el *nivel de la serie* (altura de la serie), y al suponer tendencia constante las predicciones vienen dadas por el último nivel estimado. En el siguiente gráfico se muestra la serie original (negro), la serie ajustada (azul) usando Alisado Exponencial Simple y las predicciones (rojo) para futuros instantes.

```
SES <- HoltWinters(LakeHuron, beta = FALSE, gamma = FALSE)

ts.plot(LakeHuron, xlim = c(1875, 2000),
        gpars = list(xlab = "Year",
                     ylab = "Level LakeHuron in feet",
                     lwd = 1.5))

lines(SSES$fitted[, 1], col = "blue", lwd = 1.2)
lines(predict(SSES, n.ahead = 20), col = "red", lwd = 1.2)
```

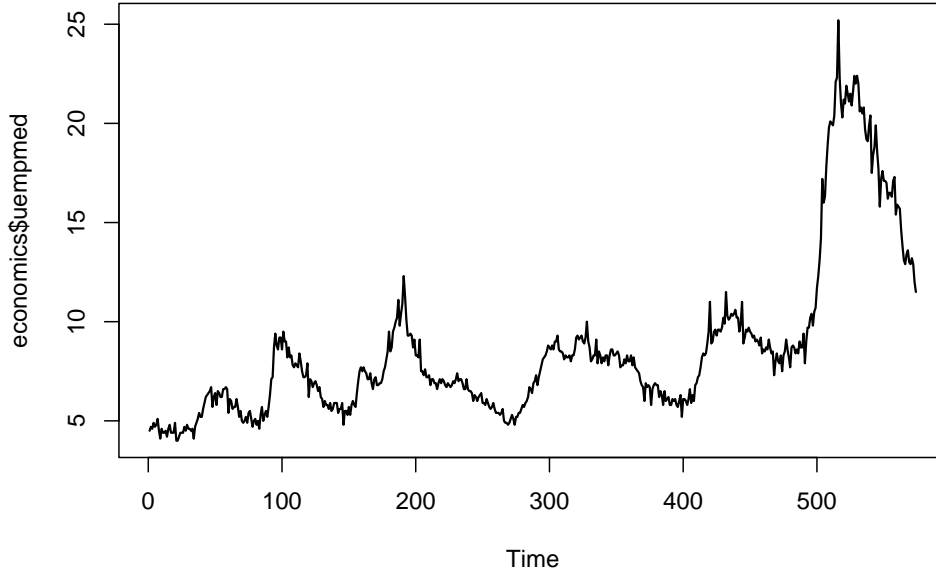


5.3) Alisado Exponencial Doble o método de Brown

Este método de alisado se aplica cuando la serie en estudio no presenta estacionalidad y la tendencia de la serie se considera lineal (localmente).

Un ejemplo ilustrativo sería la serie con siguiente gráfico:

```
library(ggplot2)
ts.plot(economics$uempmed, lwd = 1.5)
```



Esto es, suponemos que la serie en estudio se expresa de la siguiente forma en cada instante:

$$\begin{cases} x_t = T_t + I_t = (a_t + b_t \cdot t) + I_t & \text{modelo aditivo} \\ x_t = T_t \times I_t = (a_t + b_t \cdot t) \times I_t & \text{modelo multiplicativo} \end{cases}$$

donde hemos considerado que la componente de tendencia es lineal en cada instante t (lineal localmente): $T_t = a_t + b_t \cdot t$, con $a_t, b_t \in \mathbb{R}$.

Obsérvese que a_t representa el **nivel** de la serie en cada instante t , mientras que b_t representa la **pendiente** de la tendencia lineal en dicho instante.

Al no disponer de componente estacional y suponer tendencia localmente lineal, nuestro objetivo se reduce a estimar el nivel de la serie y la pendiente de la tendencia en cada instante.

El alisado exponencial doble consiste en realizar dos veces el proceso de alisado sobre la serie observada. Es decir, primero se aplica el proceso de alisado simple a la serie original y luego se aplica el proceso de alisado simple sobre la serie alisada previamente. Además, se utiliza el mismo parámetro de alisado en las dos etapas.

Dada una serie de observaciones, $\{x_t\}_{t=1,2,\dots,n}$, en el alisado exponencial doble se obtienen de forma recursiva las dos series alisadas siguientes:

$$\begin{aligned} A_t &= \alpha x_t + (1 - \alpha)A_{t-1} \\ A_t^{(2)} &= \alpha A_t + (1 - \alpha)A_{t-1}^{(2)} \end{aligned}$$

donde $\alpha \in [0, 1]$ es el *parámetro de alisado*.

Las dos series obtenidas por alisado exponencial doble sirven para estimar, en cada instante t , el nivel de la serie (a_t) y la pendiente de la tendencia (b_t). Se puede comprobar, con un desarrollo bastante laborioso, que estimaciones del nivel y de la pendiente vienen dadas por:

$$\begin{aligned} a_t &= 2A_t - A_t^{(2)} \\ b_t &= \frac{\alpha}{(1 - \alpha)} (A_t - A_t^{(2)}) \end{aligned}$$

Nota: haciendo abuso de la notación, hemos usado a_t y b_t en las series alisadas para denotar a las estimaciones del nivel y de la tendencia.

Para aplicar la fórmula recursiva anterior necesitamos disponer, por una parte, del valor del parámetro de alisado α , y por otra parte, de los valores iniciales de las series alisadas A_0 y $A_0^{(2)}$ (o equivalentemente de los valores iniciales del nivel a_0 y pendiente b_0). Como valores iniciales se suelen tomar la ordenada en el origen ($a_0 = cte$) y la pendiente ($b_0 = pendiente$) de la recta que ajusta a la serie por mínimos cuadrados usando todos los datos.

Con respecto al parámetro de alisado α , nuevamente podemos seleccionar aquel que minimice la suma de cuadrados de los errores de estimación.

Nuestro fin último es realizar predicciones de la serie para instantes futuros. Si disponemos de información de la serie hasta el instante T , $\{c_t\}_{t=1,2,\dots,T}$, las predicciones de la serie usando el método de alisado exponencial doble vienen dadas por:

$$\begin{aligned}\hat{x}_{t+1/T} &= a_T + b_T \cdot 1 \\ \hat{x}_{t+2/T} &= a_T + b_T \cdot 2 \\ &\vdots \\ \hat{x}_{t+m/T} &= a_T + b_T \cdot m \quad \forall m = 1, 2, 3, \dots\end{aligned}$$

Es decir, la predicción de la serie en el instante siguiente viene dada por la estimación del nivel en el último instante observado más la estimación de la pendiente en el último instante observado.

Igual que en el alisado simple, a medida que disponemos de información nueva (nuevas observaciones), se calculan los valores de las series alisadas para los nuevos instantes y por tanto se revisan las predicciones de manera automática.

5.4) Método de Holt

Este método de alisado se aplica en las mismas situaciones que el alisado exponencial doble, es decir, cuando la serie en estudio no presenta estacionalidad y la tendencia de la serie se considera lineal localmente.

Nuevamente, la serie en estudio se expresa de la siguiente forma en cada instante:

$$\begin{cases} x_t = T_t + I_t = (a_t + b_t \cdot t) + I_t & \text{modelo aditivo} \\ x_t = T_t \times I_t = (a_t + b_t \cdot t) \times I_t & \text{modelo multiplicativo} \end{cases}$$

donde a_t representa el **nivel** de la serie en cada instante t , mientras que b_t represente la **pendiente** de la tendencia lineal en dicho instante.

La ideal del método de Holt es similar al alisado exponencial doble: en cada instante t , queremos estimar el nivel de la serie (a_t) y la pendiente de la tendencia (b_t). A diferencia del alisado exponencial doble, en el método de Holt se usan dos parámetros de alisado que denotaremos por $\alpha \in [0, 1]$ y $\beta \in [0, 1]$.

En el método de Holt, la estimación del nivel en el instante t viene dada por la siguiente serie alisada:

$$a_t = \alpha x_t + (1 - \alpha)(a_{t-1} + b_{t-1})$$

es decir, se obtiene como media ponderada de la última observación y (nivel+pendiente) en el instante anterior.

Por otra parte, **la estimación de la pendiente en el instante t viene dada por la siguiente serie alisada:**

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

es decir, se obtiene como media ponderada del último incremento de nivel y la pendiente en el instante anterior.

Para aplicar la fórmula recursiva anterior necesitamos disponer de los parámetros de alisado $\alpha \in [0, 1]$ y $\beta \in [0, 1]$, y de los valores iniciales del nivel (a_0) y pendiente (b_0). Como valores iniciales se suelen tomar la ordenada en el origen ($a_0 = cte$) y la pendiente ($b_0 = pendiente$) de la recta que ajusta a la serie por mínimos cuadrados usando todos los datos.

Con respecto a los parámetros de alisado, podemos seleccionar aquellos que minimicen la suma de cuadrados de los errores de estimación.

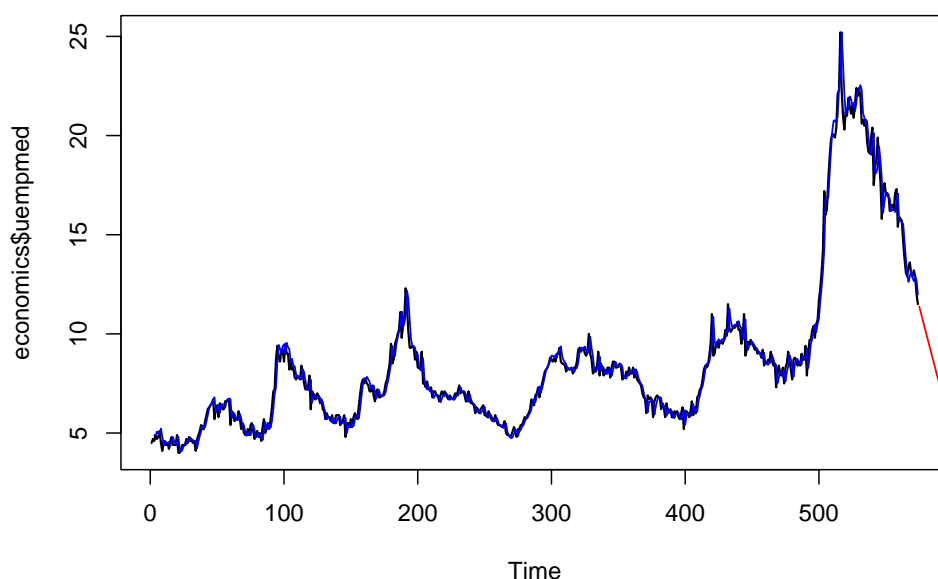
Las predicciones con el método de Holt se obtienen de forma análoga al alisado doble: si disponemos de información de la serie hasta el instante T , $\{x_t\}_{t=1,2,\dots,T}$, las predicciones vienen dadas por

$$\begin{aligned}\hat{x}_{t+1/T} &= a_T + b_T \cdot 1 \\ \hat{x}_{t+2/T} &= a_T + b_T \cdot 2 \\ &\vdots \\ \hat{x}_{t+m/T} &= a_T + b_T \cdot m \quad \forall m = 1, 2, 3, \dots\end{aligned}$$

Como siempre, a medida que disponemos de nuevas observaciones, se calculan los valores de las series alisadas para los nuevos instantes y por tanto se revisan las predicciones de manera automática.

En el siguiente gráfico se muestra la serie original (negro), la serie ajustada (azul) usando el método de Holt y las predicciones (rojo) para futuros instantes.

```
Holt <- HoltWinters(economics$uempmed, gamma = FALSE)
ts.plot(economics$uempmed, lwd = 1.5)
lines(Holt$fitted[, 1], col = "blue", lwd = 1.2)
lines(predict(Holt, n.ahead = 20), col = "red", lwd = 1.2)
```



5.5) Método de Holt-Winters multiplicativo

Este método de alisado es una variación del método de Holt que sirve para realizar predicciones en series que presentan estacionalidad. Además de la estacionalidad, se supone tendencia lineal (localmente) y que ambas componentes (tendencia y estacionalidad) se integran en la serie de forma multiplicativa.

Por tanto, la serie en estudio se expresa de la siguiente forma en cada instante:

$$x_t = T_t \times S_t \times I_t = (a_t + b_t \cdot t) \times S_t \times I_t \quad \text{modelo multiplicativo}$$

$$x_t = T_t \times S_t + I_t = (a_t + b_t \cdot t) \times S_t + I_t \quad \text{modelo multiplicativo (mixto)}$$

donde a_t representa el **nivel** de la serie en cada instante t , b_t representa la **pendiente** de la tendencia lineal en dicho instante y S_t la componente **estacional** de la serie que supondremos de período L .

Siguiendo la idea del método de Holt, en cada instante t necesitamos estimar el nivel de la serie (a_t), la pendiente de la tendencia (b_t) y el factor estacional (S_t). En este caso se usarán tres series aisladas y tres parámetros de alisado que denotaremos por $\alpha \in [0, 1]$, $\beta \in [0, 1]$ y $\gamma \in [0, 1]$.

En el método de Holt-Winters multiplicativo, la estimación del nivel en el instante t viene dada por la siguiente serie alisada:

$$a_t = \alpha \frac{x_t}{S_{t-L}} + (1 - \alpha)(a_{t-1} + b_{t-1})$$

es decir, se obtiene como media ponderada de la última observación "desestacionalizada" y (nivel+pendiente) en el instante anterior.

Por otra parte, **la estimación de la pendiente en el instante t viene dada por la siguiente serie alisada:**

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

es decir, se obtiene como media ponderada del último incremento de nivel y la pendiente en el instante anterior.

Finalmente, **la estimación del factor estacional en el instante t viene dada por la siguiente serie aislada:**

$$S_t = \gamma \cdot \frac{x_t}{a_t} + (1 - \gamma)S_{t-L}$$

Para aplicar la fórmula recursiva anterior necesitamos disponer de los parámetros de alisado $\alpha \in [0, 1]$, $\beta \in [0, 1]$ y $\gamma \in [0, 1]$, y de los valores iniciales de nivel (a_0), pendiente (b_0), y factores estacionales ($S_0, S_{-1}, \dots, S_{-L+1}$). Como valores iniciales para el nivel y la pendiente (a_0 y b_0) se suelen tomar la ordenada en el origen ($a_0 = cte$) y la pendiente ($b_0 = pendiente$) de la recta que ajusta a la serie desestacionalizada por mínimos cuadrados usando todos los datos, y como valores iniciales de los factores estacionales se suelen tomar los índices de variación estacional (IVE) obtenidos al desestacionalizar la serie (visto en análisis clásico).

Con respecto a los parámetros de alisado, podemos seleccionar aquellos que minimicen la suma de cuadrados de los errores de estimación.

Las predicciones con el método de Holt-Winters multiplicativo se obtienen de la siguiente forma: si disponemos de información de la serie hasta el instante T , $\{x_t\}_{t=1,2,\dots,T}$, las predicciones vienen dadas por

$$\begin{aligned} \hat{x}_{T+1/T} &= (a_T + b_T \cdot 1) \times S_{T+1-L} \\ \hat{x}_{T+2/T} &= (a_T + b_T \cdot 2) \times S_{T+2-L} \\ &\vdots \\ \hat{x}_{T+m/T} &= (a_T + b_T \cdot m) \times S_{T+m-L} \quad \forall m = 1, 2, \dots, L \end{aligned}$$

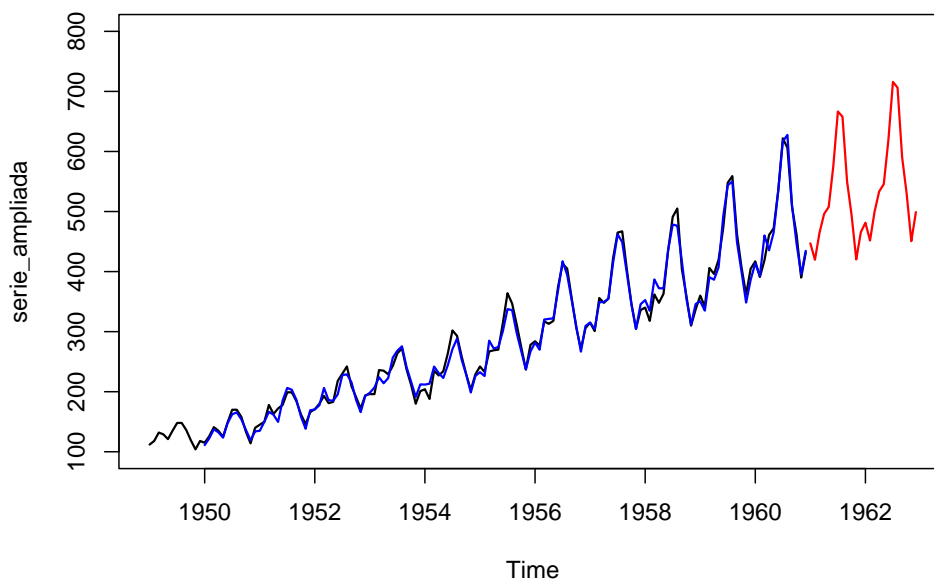
Como siempre, a medida que disponemos de nuevas observaciones, se calculan los valores de las series alisadas para los nuevos instantes y por tanto se revisan las predicciones de manera automática.

En el siguiente gráfico se muestra la serie original (negro), la serie ajustada (azul) usando Holt-Winters multiplicativo y las predicciones (rojo) para futuros instantes.

```
library(forecast)
HWm <- HoltWinters(AirPassengers, seasonal = "multiplicative")
pred_HWm <- predict(HWm, n.ahead = 24)
pred_HWm.ts <- ts(pred_HWm,
                  start = c(1961, 1), frequency = 12)

serie_ampliada <- ts(c(AirPassengers, rep("NA", 24)),
                    start = start(AirPassengers),
                    frequency = frequency(AirPassengers))

ts.plot(serie_ampliada, lwd = 1.5, ylim = c(100, 800))
lines(HWm$fitted[, 1], col = "blue", lwd = 1.5)
lines(pred_HWm.ts, col = "red", lwd = 1.5)
```



5.6) Método de Holt-Winters aditivo

Este método de alisado se aplica en los mismo casos que el método anterior, salvo que los componentes de estacionalidad y tendencia se integran de forma aditiva. Por tanto, la serie en estudio se expresa de la siguiente forma en cada instante:

$$x_t = T_t + S_t + I_t = (a_t + b_t \cdot t) + S_t + I_t \quad \text{modelo aditivo}$$

donde a_t representa el **nivel** de la serie en cada instante t , b_t representa la **pendiente** de la tendencia lineal en dicho instante y S_t la componente **estacional** de la serie que supondremos de período L .

Siguiendo la idea del método anterior, en cada instante t necesitamos estimar el nivel de la serie (a_t), la pendiente de la tendencia (b_t) y el factor estacional (S_t). En este caso se usarán tres series alisadas y tres parámetros de alisado que denotaremos por $\alpha \in [0, 1]$, $\beta \in [0, 1]$ y $\gamma \in [0, 1]$.

En el método de Holt-Winters aditivo, la estimación del nivel en el instante t viene dada por la siguiente serie alisada:

$$a_t = \alpha(x_t - S_{t-L}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

es decir, se obtiene como media ponderada de la última observación "desestacionalizada" y (nivel+pendiente) en el instante

anterior.

Por otra parte, la estimación de la pendiente en el instante t viene dada por la siguiente serie alisada:

$$b_t = \beta(a_t - a_{t-1} + (1 - \beta)b_{t-1})$$

es decir, se obtiene como media ponderada del último incremento de nivel y la pendiente en el instante anterior.

Finalmente, la estimación del factor estacional en el instante t viene dada por la siguiente serie alisada:

$$S_t = \gamma \cdot (x_t - a_t) + (1 - \gamma)S_{t-L}$$

Para aplicar la fórmula recursiva anterior necesitamos disponer de los parámetros de alisado $\alpha \in [0, 1]$, $\beta \in [0, 1]$ y $\gamma \in [0, 1]$, y de los valores iniciales de nivel (a_0), pendiente (b_0), y factores estacionales ($s_0, S_{-1}, \dots, S_{-L+1}$). Como valores iniciales para el nivel y la pendiente (a_0 y b_0) se suelen tomar la ordenada en el origen ($a_0 = cte$) y la pendiente ($b_0 = pendiente$) de la recta que ajusta a la serie desestacionalizada por mínimos cuadrados usando todos los datos, y como valores iniciales de los factores estacionales se suelen tomar los índices de variación estándar (IVE) obtenidos al desestacionalizar la serie (visto en análisis clásico).

Con respecto a los parámetros de alisado, podemos seleccionar aquellos que minimicen la suma de cuadrados de los errores de estimación.

Las predicciones con el método de Holt-Winters aditivo se obtienen de la siguiente forma: si disponemos de información de la serie hasta el instante T , $\{x_t\}_{t=1,2,\dots,T}$, las predicciones vienen dadas por

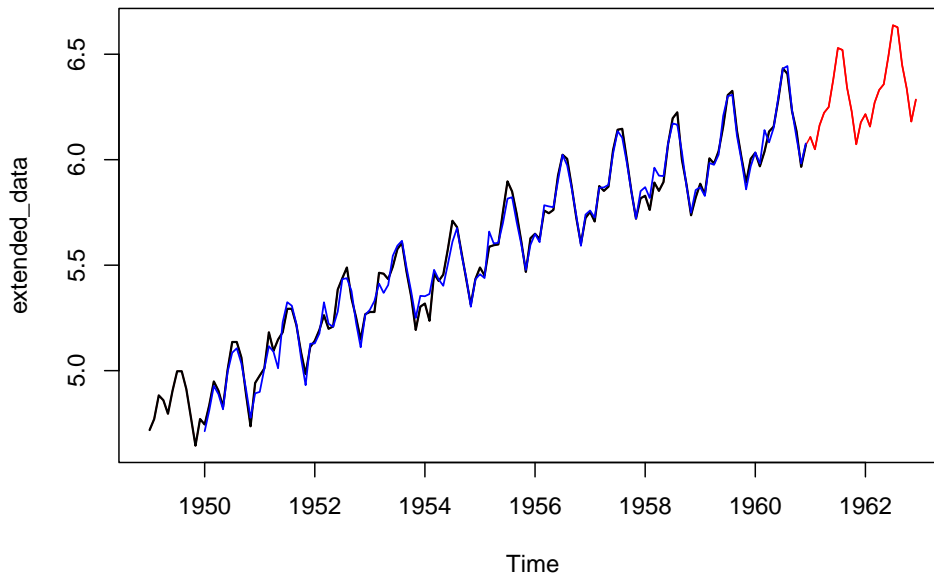
$$\begin{aligned}\hat{x}_{T+1/T} &= (a_T + b_T \cdot 1) + S_{T+1-L} \\ \hat{x}_{T+2/T} &= (a_T + b_T \cdot 2) + S_{T+2-L} \\ &\vdots \\ \hat{x}_{T+m/T} &= (a_T + b_T \cdot m) + S_{T+m-L} \quad \forall m = 1, 2, \dots, L\end{aligned}$$

Como siempre, a medida que disponemos de nuevas observaciones, se calculan los valores de las series alisadas para los nuevos instantes y por tanto se revisan las predicciones de manera automática.

En el siguiente gráfico se muestra la serie original (negro), la serie ajustada (azul) usando Holt-Winters y las predicciones (rojo) para futuros instantes.

```
data <- log(AirPassengers)
Holt_Winters_aditivo <- HoltWinters(data, seasonal = "additive")
forecasted_values <- predict(Holt_Winters_aditivo, n.ahead = 24)
extended_data <- ts(c(data, forecasted_values), start = start(data), frequency = frequency(data))

ts.plot(extended_data, col = "red")
lines(log(AirPassengers), lwd = 1.5)
lines(Holt_Winters_aditivo$fitted[, 1], col = "blue", lwd = 1.2)
lines(predict(Holt_Winters_aditivo, n.ahead = 24), col = "red", lwd = 1.2)
```



5.7) Interpretación de los parámetros de alisado

Para clarificar un poco el papel que tiene cada uno de los parámetros de alisado en los métodos vistos, podemos proporcionar la siguiente interpretación aproximada de los parámetros de alisado α , β y γ :

- Alfa (α): Parámetro de suavizado exponencial que controla el peso relativo dado a las observaciones más recientes a la hora de estimar el nivel de la serie, en contraposición a la media global de la serie. Cuando alfa toma el valor 1, se utiliza exclusivamente la única observación más reciente; cuando alfa toma el valor 0, las observaciones antiguas cuentan tanto peso como las recientes. Alfa se utiliza en todos los métodos de alisado.
- Beta (β): Parámetro de suavizado exponencial que controla el peso relativo dado a las observaciones recientes a la hora de estimar la tendencia de la serie en el presente. Toma valores de 0 a 1. Los valores próximos a 1 indican un mayor peso para los valores más recientes. Beta se utiliza sólo en los métodos de suavizado exponencial con una tendencia lineal localmente. No se utiliza en el alisado exponencial simple ni doble.
- Gamma (γ): Parámetro de suavizado exponencial que controla el peso relativo dado a las observaciones recientes al estimar la estacionalidad del presente. Toma valores de 0 a 1. Los valores próximos a 1 corresponden a un mayor peso para las observaciones recientes. La gamma se utiliza en todos los métodos de suavizado exponencial con componente estacional. No se utiliza en los métodos de alisado simple, doble y de Holt.

5.8) Taxonomía de los métodos de alisado exponencial

Los métodos de alisado desarrollados en las secciones anteriores suponen una parte del total de métodos disponibles con ese mismo enfoque.

Las formas más usuales de variar y combinar las componentes de tendencia y estacionalidad se muestran en la siguiente tabla, dando lugar a un total de 9 métodos de alisado exponencial (fuente: Hyndman, R.J. and Athanasopoulos, G., 2021):

Trend Component	Seasonal Componentes		
	N	A	M
	(None)	(Additive)	(Multiplicative)
N (None)	(N,N)	(N, A)	(N, M)
A (Additive)	(A, N)	(A, A)	(A, M)
A_d (Additive damped)	(A _d , N)	(A _d , A)	(A _d , M)

Table 1: Tabla 8.5 extraída de Hyndman and Athanasopoulos (2021)

La mayoría de dichos métodos son los desarrollados anteriormente en este tema, incluyendo como novedad los casos de tendencia amortiguada (con parámetro de alisado adicional denotado por ϕ), en los que la pendiente de la tendencia se ve amortiguada conforme aumenta el horizonte de predicción (h):

Short hand	Method
(N,N)	Simple exponential smoothing
(A,N)	Holt's linear method
(A _d ,N)	Additive damped trend method
(A,A)	Additive Holt-Winter's method
(A,M)	Multiplicative Holt-Winter's method
(A _d ,M)	Holt-Winter's damped method

Table 2: Tabla extraída de Hyndman and Athanasopoulos (2021)

La siguiente tabla muestra las fórmulas recurrentes para aplicar cada uno de los nueve métodos de alisado mencionados arriba, así como la expresión para obtener las **predicciones puntuales** con horizonte h (fuente Hyndman, R.J. and Athanasopoulos, G., 2021). Indicar que la notación empleada para dicha tabla difiere de la usada en nuestro caso.

Cabe destacar, a modo resumen, que los métodos de alisado exponencial permiten obtener **predicciones puntuales** de la serie para instantes futuros. Aunque no se establece ninguna regla para el horizonte de predicción permitido, estos métodos tienen sentido para **predicciones a corto plazo**, siendo recomendable actualizar las fórmulas recurrentes conforme se dispone de nuevas observaciones de la serie.

En la próxima sección se introducen los modelos ETS, que suponen una extensión de los métodos de alisado exponencial.

Trend	Seasonal		
	N	A	M
N	$\hat{y}_{t+h t} = \ell_t$ $\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$	$\hat{y}_{t+h t} = \ell_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1 - \gamma)s_{t-m}$	$\hat{y}_{t+h t} = \ell_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1 - \gamma)s_{t-m}$
A	$\hat{y}_{t+h t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$\hat{y}_{t+h t} = \ell_t + hb_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$	$\hat{y}_{t+h t} = \ell_t + hb_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} - b_{t-1})) + (1 - \gamma)s_{t-m}$
A_d	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-m}$	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} - \phi b_{t-1})) + (1 - \gamma)s_{t-m}$

Table 3: Tabla 8.6 extraída de Hyndman and Athanasopoulos (2021)

5.9) Los modelos ETS

Los modelos ETS, cuyas siglas se corresponden con (Error, Trend, Seasonal), representan los modelos estadísticos que subyacen bajo los métodos de alisado exponencial, dando lugar a las mismas estimaciones puntuales y permitiendo la obtención de intervalos de predicción.

Cada modelo ETS consiste en una ecuación que describe al proceso estocástico generador de la serie observada y, por otro lado, una o varias ecuaciones que describen cómo varían con el tiempo los *estados* (nivel, pendiente de la tendencia y estacionalidad). Por eso estos modelos se conocen como **modelos de espacio de estados** (state space models).

Por facilitar la distinción, hablaremos de *métodos* de alisado exponencial y *modelos* de espacio de estados. Para cada método de alisado existen dos modelos: uno con errores aditivos y otro con errores multiplicativos. Ambos proporcionan las mismas predicciones puntuales pero diferentes intervalos de predicción.

Notación: Usaremos la nomenclatura **ETS (Error, Trend, Seasonal)**, donde:

- *Error* = $\{A, M\}$ según sean errores aditivos (A) o multiplicativos (M).
- *Trend* = $\{N, A, A_d\}$ según sea tendencia (lineal localmente) aditiva (A), amortiguada (A_d), o bien no haya tendencia (N).
- *Seasonal* = $\{N, A, M\}$ según sea esquema aditivo (A) o multiplicativo (M) para combinar la estacionalidad con la tendencia, o bien no se contemple componente estacional (N).

Por ejemplo, ETS(A,N,N) representaría el modelo de espacio de estados del alisado exponencial simple con errores aditivos.

Veremos a continuación alguno de estos modelos en detalle.

5.9.1) ETS(A,N,N): Modelo de espacio de estados del alisado exponencial simple con errores aditivos

Recordemos las ecuaciones vistas para el método de alisado simple:

$$a_t = \alpha x_t + (1 - \alpha)a_{t-1} \quad (\text{ecuación suavizado})$$

$$\hat{x}_{t+1/t} = a_t \quad (\text{ecuación predicción})$$

Obsérvese que la ecuación del suavizado también puede reescribirse como sigue:

$$\begin{aligned} a_t &= a_{t-1} + \alpha(x_t - a_{t-1}) \\ &= a_{t-1} + \alpha(x_t - \hat{x}_{t/t} - 1) \\ &= a_{t-1} + \alpha e_t \end{aligned}$$

donde $e_t = x_t - \hat{x}_{t/t} = x_t - a_{t-1}$ es el residuo o error de predicción en el instante t .

Por encanto, tendremos las dos siguientes ecuaciones. La primera describe a la serie observada y la segunda cómo varían con el tiempo los estados (en este caso, el nivel de la serie):

$$x_t = a_{t-1} + e_t$$

$$a_t = a_{t-1} + \alpha e_t$$

A partir de las ecuaciones anteriores, el modelo estadístico ETS(A,N,N) se construye considerando una distribución probabilística para los errores. En este caso, y por analogía con otras metodologías estadísticas, supondremos que los residuos provienen de distribuciones $N(0, \sigma^2)$ independientes.

Por tanto, el modelo ETS(A,N,N) viene dado por las ecuaciones:

$$X_t = a_{t-1} + \epsilon_t$$

$$a_t = a_{t-1} + \alpha \epsilon_t$$

con (ϵ_t) un ruido blanco gaussiano, es decir $\epsilon_t \sim N(0, \sigma^2)$ independientes para todo t , donde X_t denota el proceso estocástico generador de la serie.

La interpretación del parámetro de alisado α es el mismo que vimos. Es decir, si $\alpha = 0$, el nivel de la serie no cambia con el tiempo, y el modelo se reduce a:

$$X_t = a_0 + \epsilon_t$$

Y si $\alpha = 1$, para cada instante t se tiene que $X_t = a_t$, y el modelo se reduce a un paseo aleatorio:

$$X_t = a_t = a_{t-1} + \epsilon_t = X_{t-1} + \epsilon_t$$

Podemos simular varias trayectorias de este proceso en R mediante:

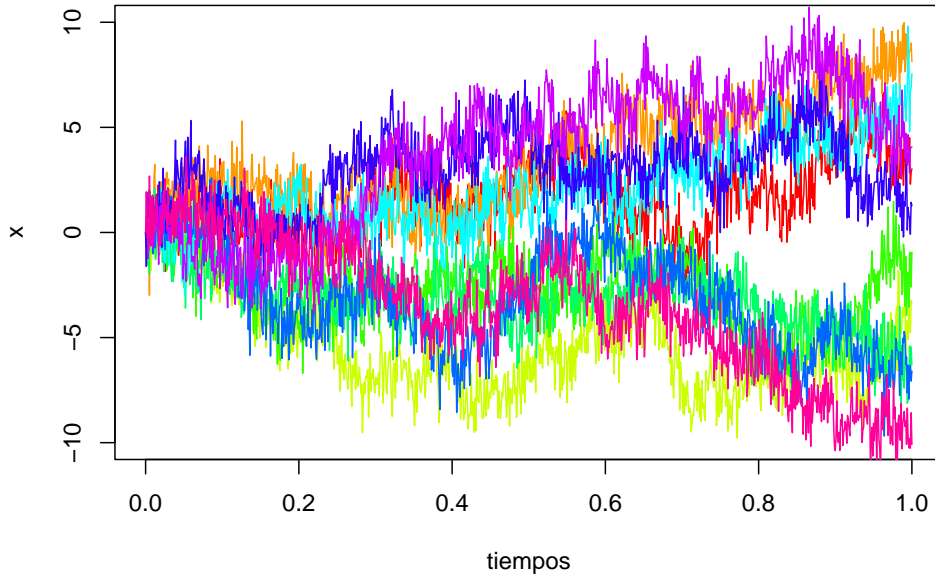
```
set.seed(123)

# parámetros
a0 <- 0
alpha <- 0.2
x0 <- a0

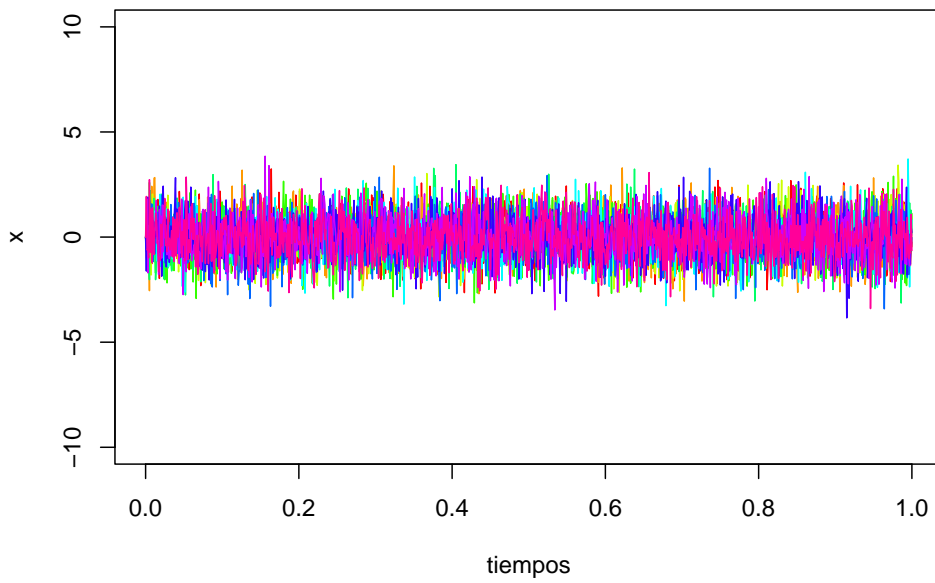
nsim <- 10
n <- 1000

# Proceso
e <- rnorm(n)
a <- a0 + c(0, cumsum(alpha * e))
x <- c(x0, a[1:n]+e)

dt <- 1/n
tiempos <- seq(from=0, to=1, by=dt)
colores <- rainbow(nsim)
plot(tiempos, x, ylim = c(-10, 10), type = "l", lty = 1, col = colores[1])
for (i in 2:nsim) {
  e <- rnorm(n)
  a <- a0 + c(0, cumsum(alpha * e))
  x <- c(x0, a[1:n]+e)
  lines(tiempos, x, col = colores[i])
}
```



Para $\alpha = 0$, obtenemos un ruido blanco gaussiano:



5.9.2) ETS(M,N,N): Modelo de espacio de estados del alisado exponencial simple con errores multiplicativos

De forma análoga a lo visto anteriormente, podemos considerar el modelo con errores multiplicativos, definiendo los errores en términos relativos:

$$e_t^* = \frac{x_t - \hat{x}_{t/t-1}}{\hat{x}_{t/t-1}}$$

Y sustituyendo $\hat{x}_{t/t-1} = a_{t-1}$ en la expresión anterior, tenemos:

$$x_t = a_{t-1} + a_{t-1}e_t^* = a_{t-1}(1 + e_t^*)$$

Obsérvese que los residuos usuales ($e_t = x_t - \hat{x}_{t/t-1}$) y los relativos (e_t^*), se relacionan mediante:

$$e_t = \hat{x}_{t/t-1} e_t^* = a_{t-1} e_t^*$$

Por tanto, la ecuación de suavizado se reescribe:

$$a_t = a_{t-1} + \alpha e_t = a_{t-1} + \alpha a_{t-1} e_t^* = a_{t-1} (1 + \alpha e_t^*)$$

A partir de las ecuaciones anteriores, el modelo estadístico ETS(M,N,N) se construye considerando una distribución probabilística para los errores relativos. En este caso, supondremos que provienen de distribuciones $N(0, \sigma^2)$ independientes.

Por tanto, el modelo ETS(M,N,N) viene dado por las ecuaciones:

$$X_t = a_{t-1} (1 + \epsilon_t) a_t = a_{t-1} (1 + \alpha \epsilon_t)$$

con (ϵ_t) un ruido blanco gaussiano, es decir, $\epsilon_t \sim N(0, \sigma^2)$ independientes para todo t , donde X_t denota el proceso estocástico generador de la serie. Podemos simular varias trayectorias de este proceso en R mediante:

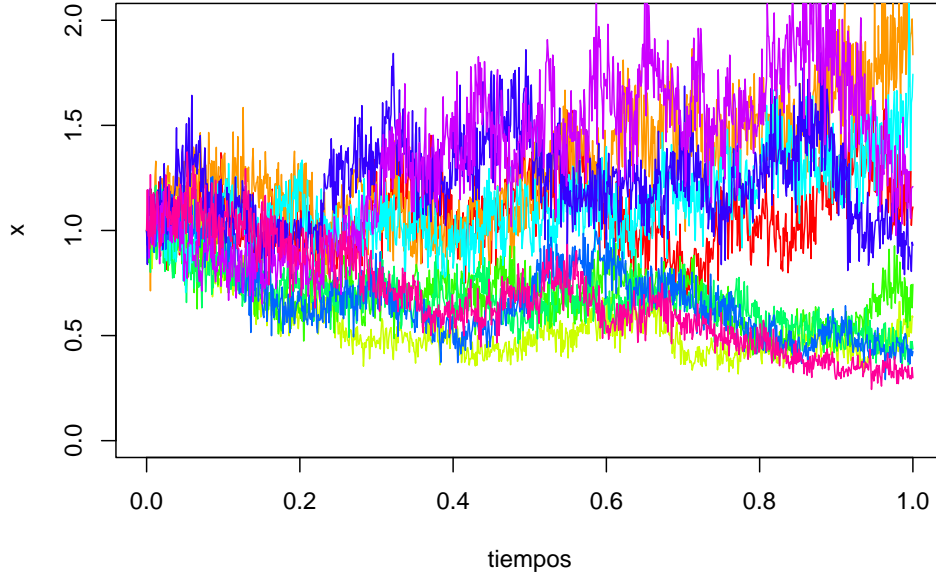
```
set.seed(123)

# parámetros
a0 <- 1
alpha <- 0.2
x0 <- a0

nsim <- 10
n <- 1000

# Proceso
e <- rnorm(n, mean=0, sd=0.1)
a <- a0*c(1, cumprod(1 + alpha*e))
x <- c(x0, a[1:n]*(1+e))

dt <- 1/n
tiempos <- seq(from=0, to=1, by=dt)
colores <- rainbow(nsim)
plot(tiempos, x, ylim=c(0, 2), type = "l", lty = 1, col=colores[1])
for(i in 2:nsim){
  e <- rnorm(n, mean=0, sd=0.1)
  a <- c(a0, a0*cumprod(1 + alpha*e))
  x <- c(x0, a[1:n]*(1+e))
  lines(tiempos, x, col=colores[i])
}
```



5.9.3) Otros modelos ETS

De forma análoga, se pueden desarrollar las ecuaciones que definen los diferentes modelos ETS según variemos el tipo de error, la tendencia y estacionalidad. Los resultados aparecen recogidos en las siguientes tablas (fuente: Hyndman, R.J. and Athanasopoulos, G., 2021):

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_t + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
A_d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$

Table 4: Tabla 8.7 extraída de Hyndman and Athanasopoulos (2021)

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_t(1 + \varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$
	$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$	$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$
		$s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$
	$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$
	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$
A _d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$
	$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$
	$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$
		$s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Table 5: Tabla 8.7 extraída de Hyndman and Athanasopoulos (2021)

5.9.4) Estimación, selección del modelo ETS y predicciones

Para la **estimación de los parámetros** (parámetros de suavizado y estados iniciales) en los modelos ETS, se utiliza el método de máxima verosimilitud en lugar de minimizar la suma de los errores al cuadrado. Recordemos que la verosimilitud representa la probabilidad de que los datos se generen a partir del modelo especificado, y una verosimilitud alta indica un buen ajuste del modelo.

Destacar que, en el caso de modelos con errores aditivos y asumiendo distribución Normal para los errores, el método de máxima verosimilitud proporciona los mismos resultados que minimicen la suma de cuadrados de los errores. Sin embargo, para modelos con errores multiplicativos no es cierto.

El empleo de la estimación máximo-verosímil en los modelos ETS, permite el uso de criterios de información para la **selección del modelo**, como el AIC (Criterio de Información de Akaike) y el BIC (Criterio de Información Bayesiano). Estos criterios ayudan a determinar qué modelo ETS es el más apropiado para una serie temporal dada. Recordemos las definiciones del AIC y BIC:

$$AIC = 2 \cdot k - 2 \cdot \ln(Lik) \quad BIC = k \cdot \ln(n) - 2 \cdot \ln(Lik)$$

donde Lik es la verosimilitud del modelo, k es el número de parámetros estimados y n el número de datos. El modelo con el menor valor de AIC o BIC se considera el mejor ajuste.

Si disponemos de una serie $\{x_t\}_{t=1,2,\dots,T}$, las **predicciones** con modelos ETS se obtienen iterando las ecuaciones del modelo para los instantes $t = T + 1, \dots, T + h$, siendo h el horizonte de predicción deseado, y asumiendo que $\varepsilon_t = 0$ para $t > T$.

Nota: Las predicciones puntuales obtenidas con los métodos de alisado exponencial coinciden con las obtenidas mediante el correspondiente modelo ETS (tanto en el caso de modelo con errores aditivos como multiplicativos), siempre que se usen los mismos parámetros de alisado. Además, la predicción puntual coincide con la media de la distribución para las predicciones, excepto en el caso de modelos con estacionalidad multiplicativa.

La ventaja principal de los modelos ETS frente a los métodos de alisado exponencial es que permiten la obtención de intervalos de predicción, puesto que se dispone de una expresión para las variables del proceso estocástico, $\{X_t\}_{t=1,2,\dots,T}$, y se asume distribución Normal para los errores.

De forma resumida, para la mayoría de los modelos ETS, un intervalo de predicción tendrá la siguiente forma:

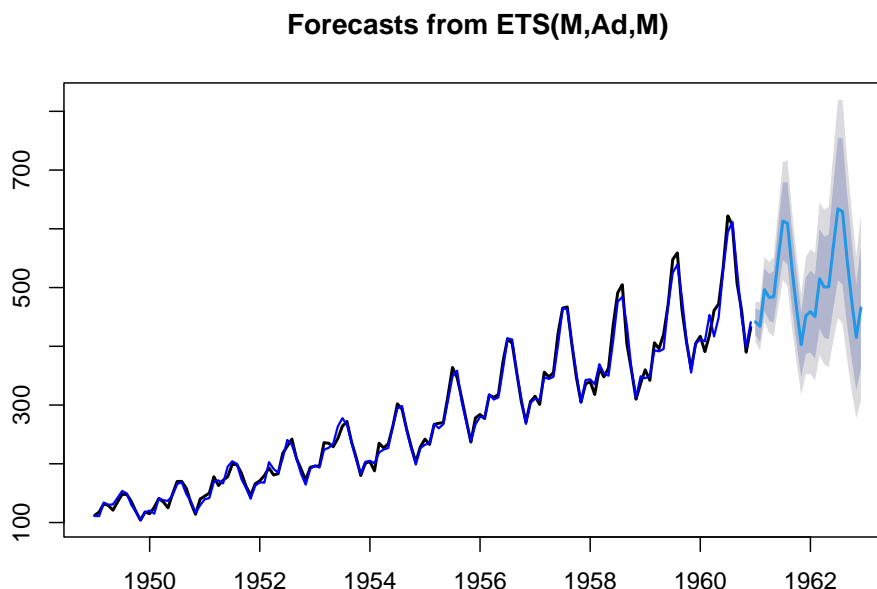
$$(\text{Estimación puntual} \pm \text{Error de predicción}) = (\hat{x}_{T+h/T} \pm c \cdot \sigma_h)$$

donde c es el cuantil de la distribución $N(0,1)$ necesario según el nivel de confianza prefijado y σ_h^2 es la varianza de la predicción con horizonte h . Indicar que la expresión analítica de σ_h^2 es complicada de obtener para algunos modelos ETS.

Veamos un ejemplo de aplicación de los modelos ETS para la serie "AirPassengers".

En el siguiente gráfico se muestra la serie original (negro), la serie ajustada (azul oscuro) usando el modelo ETS(M,A,M) y los intervalos de predicción para los dos años siguientes (sombras grises). En azul claro se muestra la media de la distribución para las predicciones, es decir, el punto medio del intervalo de predicción.

```
library(forecast)
ETS_model <- ets(AirPassengers, model = "MAM")
plot(forecast(ETS_model, h = 24), lwd = 2)
lines(ETS_model$fitted, col = "blue", lwd = 1.5)
```



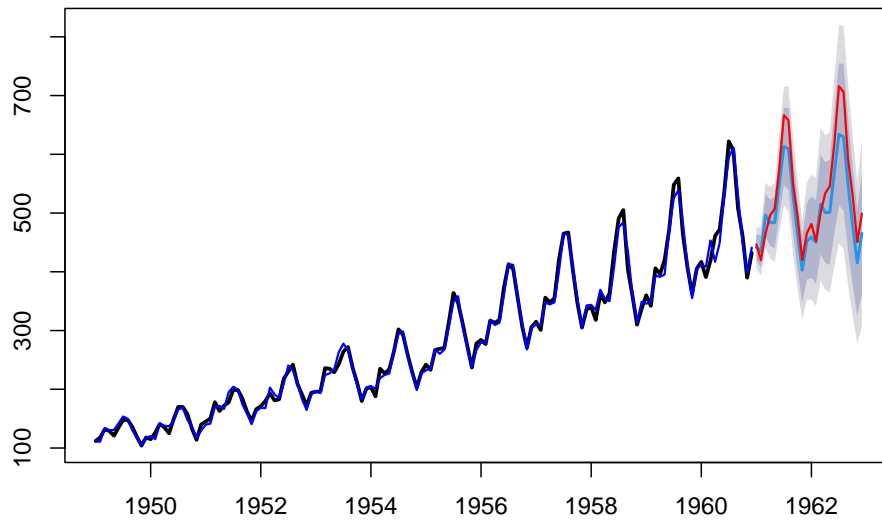
Nota: En este caso la predicción puntual no coincide con la media de la distribución para las predicciones, pues se trata de un modelo ETS con estacionalidad multiplicativa.

Podemos ver la diferencia entre la predicción puntual (en rojo) con la media de la distribución para las predicciones (en azul claro):

```
library(forecast)
HWm <- HoltWinters(AirPassengers, seasonal = "multiplicative")
pred_HWm <- predict(HWm, n.ahead = 24)
pred_HWm.ts <- ts(pred_HWm,
                  start = c(1961, 1), frequency = 12)

ETS_model <- ets(AirPassengers, model = "MAM")
predicciones <- forecast(ETS_model, h = 24)
plot(forecast(ETS_model, h = 24), lwd = 2.5)
lines(predicciones$fitted, col = "blue", lwd = 1.5)
lines(pred_HWm.ts, col = "red", lwd = 1.5)
```

Forecasts from ETS(M,Ad,M)



Hoja 5: Métodos de Alisado Exponencial

1) Responder a las siguientes preguntas

- a) ¿Qué son los métodos de alisado exponencial y para qué sirven?
- b) Clasificación de los métodos de alisado exponencial.
- c) ¿Qué son los parámetros de alisado y dónde intervienen? ¿Cuántos parámetros de alisado deben emplearse?
- d) ¿Cómo se obtiene los parámetros de alisado óptimos?
- e) Supongamos que se está analizando una serie temporal que no tiene estacionalidad y cuyas observaciones se recopilan cada minuto. El analista ha decidido emplear el método de Holt para predecir los próximos 30 minutos, y transcurrido ese tiempo, volverá a aplicar el método de Holt con las nuevas observaciones reales para predecir los siguientes 30 minutos. ¿Te parece adecuado?
- f) Explica cómo se interpretaría que en el método de Holt-Winters (aditivo o multiplicativo), se obtuviera un parámetro de alisado óptimo $\gamma=0$.
- g) Los métodos de alisado exponencial, ¿permiten obtener intervalos de predicción o sólo predicciones puntuales?
- h) ¿Qué son los modelos ETS y qué significan sus siglas? ¿Qué relación tiene con los métodos de alisado exponencial?
- i) ¿Los modelos ETS, ¿permiten obtener intervalos de predicción o sólo predicciones puntuales? ¿Se necesita algún supuesto en términos de distribuciones probabilísticas?
- j) ¿Qué relación existe entre las predicciones puntuales de los métodos de alisado exponencial y las obtenidas con modelos ETS?

2) Responder a las siguientes cuestiones:

- a) Para analizar una serie temporal con 100 observaciones, se usó alisado exponencial simple (AES). Se sabe que la última observación vale $x(100) = 28$, que el parámetro de alisado óptimo fue $\alpha = 0.8$ y que la estimación del nivel de la serie en el instante 99 fue $a(99) = 35$. Calcular las predicciones para los próximos 5 instantes.

$$a_t = \alpha x_t + (1 - \alpha)a_{t-1} \implies a_{100} = 0.8 \cdot x_{100} + (1 - 0.8) \cdot a_{99} = 0.8 \cdot 28 + 0.2 \cdot 35 = \boxed{29.4}$$

$$\hat{x}_{101|100} = \dots = \hat{x}_{105|100} = a_{100} = 29.4$$

- b) Repetir el apartado (a) suponiendo que se usó el método de Holt y que se tienen los siguientes datos: $x(100) = 28, \alpha = 0.8, \beta = 0.7, a(99) = 35, b(99) = 1$.

$$\bullet a_{100} = \alpha x_{100} + (1 - \alpha)(a_{99} + b_{99}) = 0.8 \cdot 28 + 0.2 \cdot (35 + 1) = 19.6$$

$$\bullet b_{100} = \beta(a_{100} - a_{99}) + (1 - \beta)b_{99} = 0.7 \cdot (29.6 - 35) + 0.3 \cdot 1 = -3.48$$

$$\hat{x}_{100+h|100} = a_{100} + h \cdot b_{100} = 29.6 + h \cdot (-3.48) \quad h = 1, 2, 3, 4, 5.$$

$$h = 1 \implies \hat{x}_{101|100} = 26.12$$

$$h = 2 \implies \hat{x}_{102|100} = 22.64$$

$$h = 3 \implies \hat{x}_{103|100} = 19.16$$

$$h = 4 \implies \hat{x}_{104|100} = 15.68$$

$$h = 5 \implies \hat{x}_{105|100} = 12.20$$

- 3) Para una serie de datos con 30 observaciones recopiladas por cuatrimestres, comenzando en el primer cuatrimestre de 2014, se observó componente estacional y se aplicó Holt-Winters aditivo. Se pide calcular las predicciones para los años

2024 y 2025 sabiendo que se obtuvieron los siguientes datos:

$$x(30) = 250, \alpha = 0.8, \beta = 0.9, \gamma = 0, a(29) = 200, \\ b(29) = -2, S(1) = 25, S(2) = -15, S(3) = -10$$

Observar que $\gamma = 0$, por tanto:

$$S_t = S_{t-L} \forall t \implies S_{28} = S_{31} = S_1 = 25 \\ S_{29} = S_{32} = S_2 = -15 \\ S_{30} = S_{33} = S_3 = -10$$

Como la serie empieza en el 1^{er} cuatrimestre de 2014 ($t = 1$) entonces termina en 3^{er} cuatrimestre de 2023 ($t = 30$). Me piden predicciones para los cuatrimestre de 2024 y 2025, es decir, predicciones en horizontes $h = 1, 2, 3, 4, 5, 6$.

$$a_{30} = \alpha(x_{30} - S_{30-L}) + (1 - \alpha)(a_{29} + b_{29}) = 0.8 \cdot (250 - (-10)) + 0.2 \cdot (200 + (-2)) = 247.6$$

$$b_{30} = \beta(a_{30} - a_{29}) + (1 - \beta)b_{29} = 0.9 \cdot (247.6 - 250) + 0.1 \cdot (-2) = 42.64$$

$$S_{30} = S_3 = -10$$

$$h = 1 \quad \hat{x}_{31|30} = (a_{30} + 1 \cdot b_{30}) + S_{28} = 247.6 - 42.64 + 25 = 315.24$$

$$h = 2 \quad \hat{x}_{32|30} = (a_{30} + 2 \cdot b_{30}) + S_{29} = 247.6 + 2 \cdot 42.64 - 15 = 317.88$$

$$h = 3 \quad \hat{x}_{33|30} = (a_{30} + 3 \cdot b_{30}) + S_{30} =$$

$$h = 4 \quad \hat{x}_{34|30} = (a_{30} + 4 \cdot b_{30}) + S_{31} =$$

$$h = 5 \quad \hat{x}_{35|30} = (a_{30} + 5 \cdot b_{30}) + S_{32} =$$

$$h = 6 \quad \hat{x}_{36|30} = (a_{30} + 6 \cdot b_{30}) + S_{33} =$$

4) Repetir el Problema 3 suponiendo esquema multiplicativo y que

$$S(1) = 15, S(2) = 0.6, S(3) = 0.9$$

5) En la siguiente tabla se muestra una serie temporal con datos trimestrales a la que se le ha aplicado el método de Holt-Winters. Responde a las siguientes cuestiones.

- 1) Representar la serie, y justificar si, para la serie en estudio, es más adecuado un esquema aditivo o bien uno multiplicativo.
- 2) De acuerdo al apartado anterior, completa los huecos de la tabla, sabiendo que:

$$\alpha = 0.2, \beta = 0.3, \gamma = 0.5,$$

$$a(0) = 619.31, b(0) = 4.06,$$

$$S(-3) = 21.09, S(-2) = 80.22, S(-1) = -55.66, S(0) = -45.66$$

Para $t = 1$: Datos trimestrales, $L = 4$

$$a_1 = \alpha(x_1 - S_{1-4}) + (1 - \alpha)(a_0 + b_0) = 0.2 \cdot (653 - 21.09) + (1 - 0.2) \cdot (619.31 + 4.06) = 625.078$$

$$b_1 = \beta(a_1 - a_0) + (1 - \beta)b_0 = 0.3 \cdot (625.078) + (1 - 0.3) \cdot 4.06 = 4.5724$$

$$S_1 = \gamma(x_1 - a_1) + (1 - \gamma)S_{1-L} = 0.5 \cdot (653 - 625.078) + 0.5 \cdot 21.09 = 24.506$$

Para $t = 4$:

$$a_4 = \alpha(x_4 - S_{4-4}) + (1 - \alpha)(a_3 + b_3) = 0.2 \cdot (598 - (-45.66)) + 0.8 \cdot (636.39 + 5.11) = 641.932$$

$$b_4 = \beta(a_4 - a_3) + (1 - \beta)b_3 = 0.3 \cdot (641.932 - 636.39) + 0.7 \cdot 5.11 = 5.2396$$

$$S_4 = \gamma(x_4 - a_4) + (1 - \gamma)S_{4-L} = 0.5 \cdot (598 - 641.932) + 0.5 \cdot (-45.66) = -44.796$$

$$\hat{X}_{4|3} = (a_3 + 1 \cdot b_3) + S_0 = (636.391 + 5.11) + (-45.66) = 595.84$$

3) Calcular las predicciones para el año 2024.

Año	Trimestre	Serie	a	b	S	Serie ajustada
2021	T1	653				644.47
2021	T2	716	630.88	4.94	82.67	709.87
2021	T3	583	636.39	5.11	-54.52	580.17
2021	T4	598				
2022	T1	688	650.44	6.22	31.03	671.68
2022	T2	736	655.99	6.02	81.34	739.33
2022	T3	604	661.32	5.81	-55.92	607.49
2022	T4	634	669.46	6.51	-40.13	622.33
2023	T1	709	676.37	6.63	31.83	707.01
2023	T2	756				
2023	T3	615	684.16	5.14	-62.54	631.55
2023	T4	635	686.46	4.29	-45.79	649.17

6) En la siguiente tabla se muestra una serie de temporal con datos trimestrales a la que se le ha aplicado el método de Holt-Winters. Responde a las siguientes cuestiones:

- 1) Representar la serie, y justificar si, para la serie en estudio, es más adecuado un esquema aditivo o bien uno multiplicativo.
- 2) De acuerdo al apartado anterior, completa los huecos de la tabla, sabiendo que:

$$\alpha = 0.2, \beta = 0.3, \gamma = 0.5,$$

$$a(0) = 560.94, b(0) = 28.96$$

$$S(-3) = 1.04, S(-2) = 1.92, S(-1) = 0.50, S(0) = 0.54$$

3) Calcular las predicciones para el año 2024.

Año	Trimestre	Serie	a	b	S	Serie ajustada
2021	T1	685.40				614.66
2021	T2	1286.91	643.42	35.11	1.96	1220.67
2021	T3	340.36				
2021	T4	395.44	717.74	36.30	0.55	385.66
2022	T1	972.63	781.88	44.66	1.17	821.04
2022	T2	1571.84	821.71	43.21	1.94	1619.13
2022	T3	419.89	859.63	41.62	0.49	433.15
2022	T4	566.80	928.81	49.89	0.58	491.63
2023	T1	1199.91				
2023	T2	1919.85	1031.61	49.90	1.90	2016.41
2023	T3	468.72	1054.73	41.86	0.47	534.94
2023	T4	572.49	1075.41	35.51	0.56	633.69

Tema 6: Modelos ARIMA

6.1) Introducción

Tanto en los métodos de análisis clásico de series como en los métodos de alisado exponencial, partimos de un esquema establecido a priori: descomposición de la serie en las componentes tendencia-ciclo, estacionalidad e irregular. Sin embargo, a principios de 1970 aparece un nuevo enfoque en el estudio de series temporales univariantes (debido a los estadísticos Box y Jenkins) y se basa en estudiar la correlación de los datos. *Este nuevo enfoque consiste en considerar que la serie temporal en estudio ha sido generada por un proceso estocástico. El objetivo en este caso es identificar el proceso estocástico que ha generado la serie, para posteriormente poder realizar predicciones.*

Por tanto, se pretende construir un modelo que nos permita explicar la estructura y preveer la evolución, a corto y medio plazo, de una serie temporal. La variable observada puede ser económica (I.P.C., demanda de un producto, existencias en un determinado almacén, etc.), física (temperatura de un proceso, velocidad del viento en una central eólica, concentración en la atmósfera de un contaminante, etc.) o social (número de nacimientos, votos de un determinado partido, etc.).

Recordemos que la definición de serie temporal (una sucesión de valores de una variable obtenidos de manera secuencial en el tiempo) coincide con el concepto de realización de un proceso estocástico. Es decir, los datos x_1, x_2, \dots, x_n de una serie temporal observados en n instantes de tiempo pueden interpretarse como una trayectoria o realización de particular de un proceso estocástico $(X_t)_{t=1,2,\dots,n}$. Teniendo en cuenta esta interpretación, la teoría de los procesos estocásticos será aplicable al estudio de series temporales.

Si dispusiéramos de muchas realizaciones de un mismo proceso estocástico, es decir, de muchas series temporales generadas por un mismo proceso, podríamos intentar obtener la función de distribución de cada variable X_i del proceso, aunque no sería sencillo. En general hay que contentarse con conocer algunas características del proceso como la función de medias, la función de varianzas, etc. Supongamos, por ejemplo, que disponemos de las siguientes series que han sido generadas por un mismo proceso:

Serie 1: $\{x_{11}, x_{12}, \dots, x_{1n}\}$

\vdots

Serie k: $\{x_{k1}, x_{k2}, \dots, x_{kn}\}$

entonces podemos estimar la media de cada variable X_i del proceso mediante:

$$\begin{aligned}\hat{\mu}_1 &= \frac{x_{11} + x_{21} + \dots + x_{k1}}{k} \\ &\vdots \\ \hat{\mu}_n &= \frac{x_{1n} + x_{2n} + \dots + x_{kn}}{k}\end{aligned}$$

No obstante, nos encontraremos con una importante restricción al trabajar con series temporales: en muchos casos, la serie observada es la única realización accesible del proceso estocástico que la ha generado. Por ejemplo, en la serie de turistas que visitan España mes a mes o en la cantidad de unidades producidas diariamente en una fábrica, solo disponemos de una única trayectoria concreta del fenómeno, sin acceso a múltiples realizaciones independientes del mismo proceso.

Este problema, que a simple vista parece insalvable, requiere que apliquemos ciertas restricciones e hipótesis al tipo de proceso estocástico que genera la serie en estudio. Específicamente, necesitaremos que el proceso estocástico sea *estacionario* (al menos en sentido débil) y *ergódico*. Estas condiciones garantizarán que los datos observados a lo largo de un período de tiempo suficientemente amplio sean representativos del comportamiento probabilístico del proceso subyacente, y que el conocimiento obtenido de los datos actuales sea útil para comprender su comportamiento en momentos futuros.

6.1.1) Procesos débilmente estacionarios. El correlograma.

Tal como estudiamos en el Tema 1, un proceso estocástico $\{X_t\}_{t=1,2,\dots}$ se dice **estacionario en sentido débil** (o débilmente estacionario) si cumple que su función de medias es constante, y su función de covarianzas sólo depende del retardo o salto temporal, es decir:

- $\mu_X(t) = \mu$ para cierta constante μ .
- $C_X(t, t+k) = \gamma_k$, para cierta cantidad γ_k que sólo depende de k (y no del instante t).

Cuando el proceso sea débilmente estacionario, se cumplirá además que la función de varianzas es constante en el tiempo, y que la función de correlaciones sólo depende del retardo o salto temporal, es decir:

- $\sigma_X(t) = \sigma$ para cierta constante σ .
- $C_X(t, t+k) = \rho_k$, para cierta cantidad ρ_k que sólo depende de k (y no del instante t).

Observación 6.1.1 *La estacionariedad en sentido débil no garantiza la estacionariedad en sentido estricto, excepto en el caso de normalidad de las variables del proceso. No obstante, la estacionariedad en sentido débil garantizará que algunas características del proceso estocástico tenga un comportamiento estable a lo largo del tiempo.*

En el estudio de procesos estocásticos estacionarios, la función de correlaciones (o autocorrelaciones), denotada como ρ_X , es de especial importancia. Como mencionamos, cuando el proceso es débilmente estacionario, esta función depende únicamente del retardo k , tomando un valor ρ_k que varía sólo en función de este desfase temporal y no del instante específico. La representación gráfica de la función ρ_k en relación con el retardo k se denomina correlograma (o autocorrelograma).

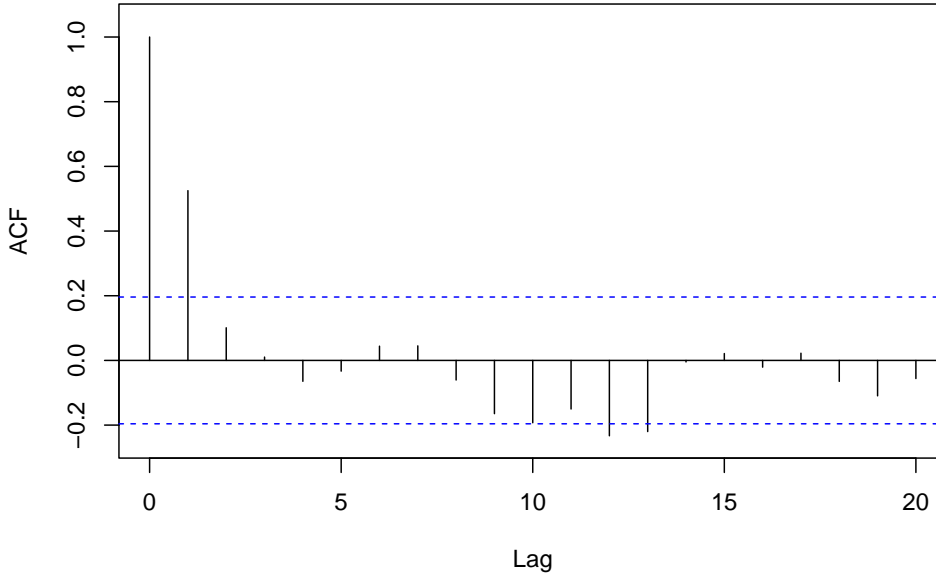


Figure 11: Ejemplo de correlograma

Observar que el correlograma, tal y como se ha definido, sólo tiene sentido para procesos débilmente estacionarios.

En el contexto de procesos estocásticos débilmente estacionarios, aunque dispongamos de una única realización (una única serie temporal), podemos obtener una estimación de las características del proceso del siguiente modo. Si denotamos por $\{x_1, x_2, \dots, x_n\}$ a la única serie temporal observada del proceso, se tiene:

- Estimación de la función de medias constante:

$$\hat{\mu} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- Estimación de la función de varianzas constante:

$$\widehat{\sigma^2} = s_x^2 = \frac{\sum_{t=1}^n (x_t - \bar{x})^2}{n}$$

- Estimación de la función de covarianzas:

$$\hat{\gamma}_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{n}$$

- Estimación de la función de correlaciones:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

Si buscamos el comportamiento de estacionariedad para las series temporales, necesitaremos ver gráficas que se mantienen en un nivel constante con unas pautas estables de oscilación. En la figura se muestra un ejemplo de serie estacionaria, realización de un proceso estocástico estacionario.

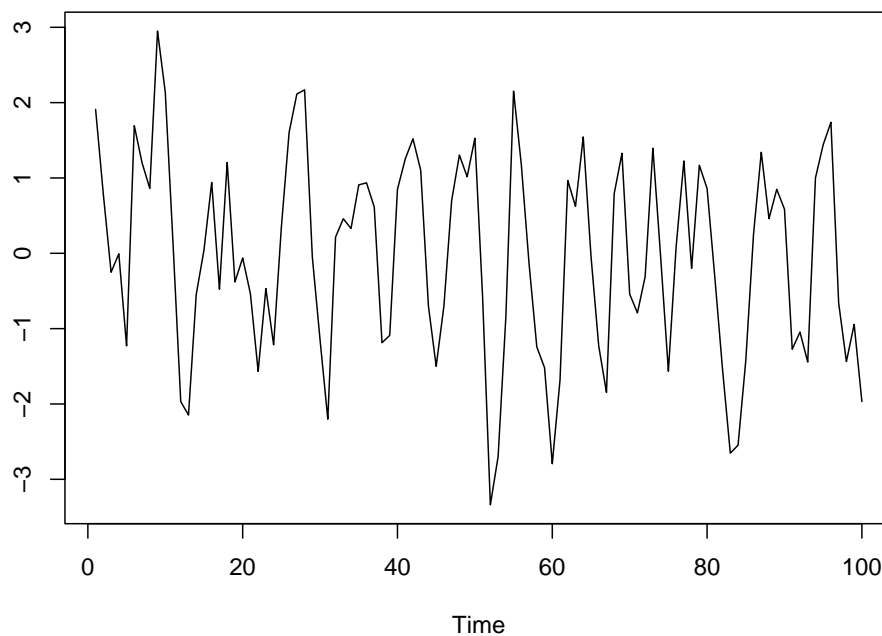


Figure 12: Ejemplo de serie estacionaria (en sentido débil)

En la práctica del análisis de series encontraremos series con problemas de estacionariedad que afectan a cualquiera de sus parámetros básicos, siendo los más frecuentes las inconstancias en media y varianza. En la figura se muestran dos ejemplos de series no estacionarias, la primera en media y la segunda en varianza.

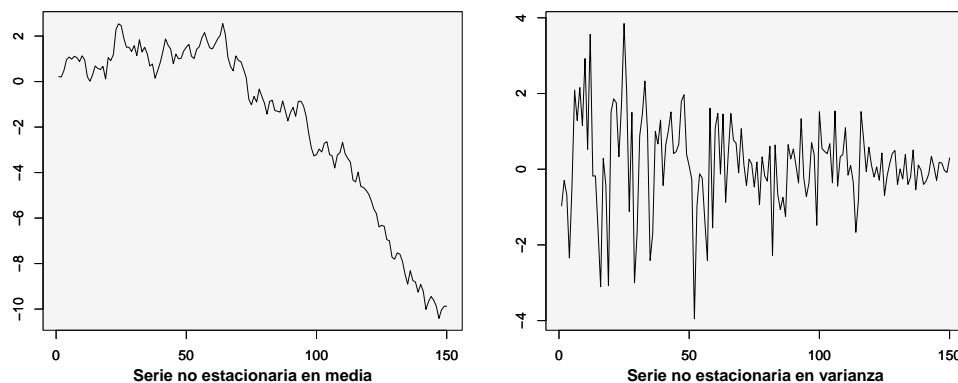


Figure 13: Ejemplo de series no estacionarias

Además de ser estacionario, el proceso estocástico ha de ser *ergódico*. Este concepto es algo más complejo y sólo indicaremos que una condición necesaria para que un proceso sea ergódico es que $\lim_{k \rightarrow \infty} \rho_k = 0$, es decir, que las autocorrelaciones sean nulas para retardos altos. Esto quiere decir, que para valores altos del retardo habrá poca dependencia entre las observaciones. En caso contrario los valores de la serie alejados en el tiempo estarían altamente correlados y por tanto no se podrían obtener estimaciones consistentes de la función de medias, varianzas, etc.

En adelante, supondremos siempre que trabajamos con procesos ergódicos y nos centramos en estudiar si el proceso que ha generado la serie es o no estacionario. Llegados a este punto, cabría preguntarse si la estacionariedad resulta una condición muy restrictiva, es decir, si en la práctica existen muchas series que proceden de procesos estocásticos no estacionarios. En este sentido podemos decir que, aunque trabajemos con series no estacionarias, en general se podrá conseguir la estacionariedad mediante una transformación sencilla en los datos.

De hecho, las dos transformaciones más usuales para conseguir la estacionariedad de una serie son:

- Realizar una transformación de Box-Cox (cuando la serie no es constante en varianza), siendo la más frecuente el tomar logaritmos neperianos en los datos.
- Tomar diferencias en la serie (cuando la serie no es constante en media): si la tendencia es lineal se tomarán diferencias de orden 1, si la tendencia es cuadrática se tomarán diferencias de orden 2, etc. Los procesos que no son estacionarios, pero que se convierten en estacionarios al tomar diferencias, se denominan *procesos integrados*.

Finalizamos la sección indicando una propiedad para procesos estacionarios fácil de demostrar.

Proposición 6.1.1 Si $\{X_t\}_t$ es un proceso estacionario, entonces el proceso primera diferencia

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

también sigue siendo estacionario.

6.1.2) Procesos lineales

Los procesos lineales son una clase particular de procesos estocásticos que incluyen a los siguientes tipos de procesos, que estudiaremos con más detalle en las próximas secciones:

- Proceso puramente aleatorios o ruido blanco gaussiano (también suele denominarse simplemente ruido blanco).
- Procesos autorregresivos, $AR(p)$.
- Procesos de medias móviles, $MA(q)$.
- Procesos autorregresivos y de medias móviles, $ARMA(p, q)$.
- Procesos autorregresivos y de medias móviles no estacionarios, $ARIM(p, d, q)$.

6.2) Proceso de ruido blanco gaussiano

Recordemos que un ruido blanco gaussiano es un proceso estocástico $(\varepsilon_t)_{t=1,2,\dots,n}$ verificando que $\varepsilon_t \sim N(0, \sigma^2)$ para todo $t = 1, 2, \dots, n$ e independientes entre sí.

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

En el caso del proceso de ruido blanco gaussiano, los correlogramas simple y parcial no presentarán ninguna correlación significativa (salvo para el retardo 0, donde la correlación es de 1). A continuación se muestra un ejemplo de ruido blanco gaussiano y su correlograma simple.

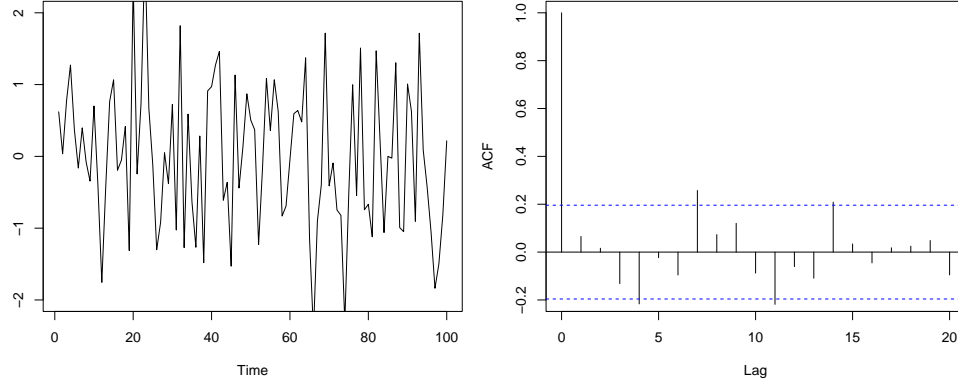


Figure 14: Ejemplo de ruido blanco gaussiano y su correlograma

6.3) Modelos Autorregresivos de orden p , $AR(p)$

Al representar la influencia de hechos pasados sobre el presente (y en consecuencia sobre el futuro) de un proceso estocástico, podemos considerar diferentes expresiones. Una de ellas consiste en colocar el valor actual del proceso dependiente linealmente de valores pasados del propio proceso, más una perturbación aleatoria que se comporta como un ruido blanco gaussiano:

$$X_t = \delta + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t \quad (16)$$

donde δ representa una constante y ε_t es un ruido blanco, es decir, las variables ε_t son i.i.d. y todas ellas tienen distribución $N(0, \sigma^2)$.

Esta formulación se denomina autorregresiva porque en cierto modo es un modelo de regresión del proceso sobre sí mismo.

Observando el modelo propuesto en (16), si lo consideramos estacionario, con $E(X_t) = \mu$, se tiene que:

$$\mu = \delta + a_1 \mu + a_2 \mu + \dots + a_p \mu \implies \mu = \frac{\delta}{1 - a_1 - a_2 - \dots - a_p}$$

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

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

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

Un elemento que se suele utilizar para expresar la formulación de los procesos lineales es el llamado operador de retardo B . Tal operador actúa sobre un término de un proceso estocástico reduciendo el índice temporal en una unidad:

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

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

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

o equivalentemente:

$$[a_p(B)] X_t = \varepsilon_t$$

donde

$$a_p(x) = 1 - a_1 \cdot x - a_2 \cdot x^2 - \dots - a_p \cdot x^p$$

se denomina *polinomio característico* del proceso autorregresivo.

Como propiedad, destacaremos que la condición necesaria y suficiente para que un proceso $AR(p)$ sea estacionario es que las raíces de su polinomio característico estén fuera del círculo unidad del plano complejo.

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

- Determinación del orden p del modelo autorregresivo.
- Una vez fijado éste, determinar los parámetros a_i del modelo.

6.3.1) Determinación del orden de la autorregresión

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

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

$$\rho(X_t, X_{t-2} | X_{t-1}) = 0,$$

donde la notación anterior se interpreta como la correlación entre las variables X_t y X_{t-2} eliminando el efecto de X_{t-1} .

Esta dependencia puede ilustrarse con el esquema siguiente:

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

donde las flechas muestran una relación de dependencia directa.

Sin embargo, en un $AR(2)$ además del efecto de X_{t-2} que se transmite a X_t a través de X_{t-1} , existe un efecto directo de X_{t-2} sobre X_t , por lo que, en general,

$$\rho(X_t, X_{t-2} | X_{t-1}) \neq 0.$$

Por otro lado, conocidos X_{t-1} y X_{t-2} , el valor de X_{t-3} es irrelevante para predecir X_t , es decir,

$$\rho(X_t, X_{t-3} | X_{t-1}, X_{t-2}) = 0.$$

En este caso, podemos escribir:

$$\begin{array}{ccccccc}
 & & \uparrow & & - \longrightarrow - & & \downarrow \\
 AR(2) : & X_{t-3} & & \rightarrow X_{t-2} & & \rightarrow X_{t-1} & \rightarrow X_t \\
 & & \downarrow & & - \longrightarrow - & & \uparrow
 \end{array}$$

La función de autocorrelación simple tiene sólo en cuenta que X_t y X_{t-2} están relacionadas en ambos casos, pero si medimos la relación directa entre X_t y X_{t-2} , esto es, eliminando el efecto debido a X_{t-1} , encontraremos que para un $AR(1)$ este efecto es nulo y para un $AR(2)$ no.

En general, un $AR(p)$ presenta efectos directos de observaciones separadas por $1, 2, \dots, p$ retardos y los efectos directos para retardos superiores son nulos, es decir,

$$\rho(X_t, X_{t-k} | X_{t-1}, X_{t-2}, \dots, X_{t-k+1}) \neq 0 \quad \text{si } k \leq p,$$

$$\rho(X_t, X_{t-k} | X_{t-1}, X_{t-2}, \dots, X_{t-k+1}) = 0 \quad \text{si } k > p.$$

Esta idea es la clave para la utilización de la función de autocorrelación parcial, entendiendo el coeficiente de autocorrelación parcial de orden k como una medida de la relación lineal entre observaciones separadas k períodos con independencia de los valores intermedios. De este concepto se deduce que un proceso $AR(p)$ tendrá los p primeros coeficientes de autocorrelación parcial distintos de cero. Llamaremos autocorrelograma parcial a la representación de los coeficientes de correlación parcial en función del retardo. *Por tanto, para determinar el orden de un modelo autorregresivo nos fijaremos en el correlograma parcial: el número de coeficientes que sean 'significativamente' distintos de cero indica el orden del proceso de AR.*

A continuación mostramos cómo serían los autocorrelogramas simples y parciales **teóricos** de modelos $AR(1)$ y $AR(2)$.

1) Modelos $AR(1)$:

a) Con el parámetro $a_1 > 0$:

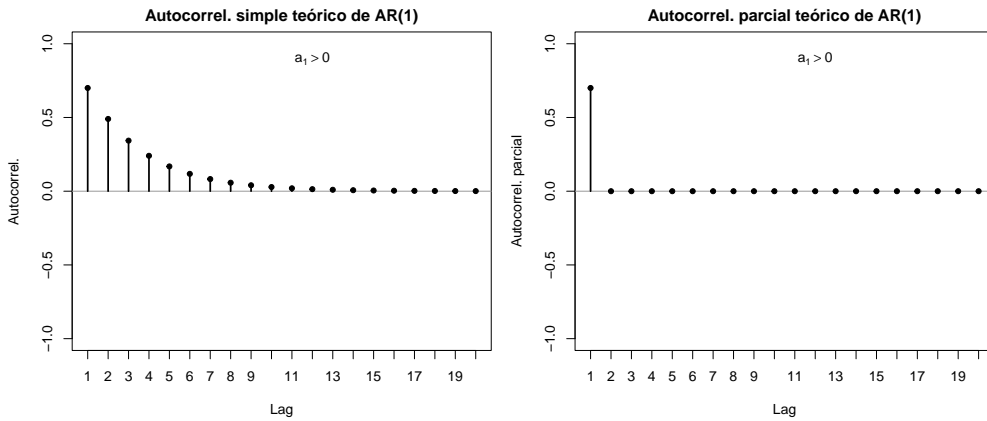


Figure 15: Correlogramas teóricos $AR(1)$

b) Con el parámetro $a_1 < 0$:

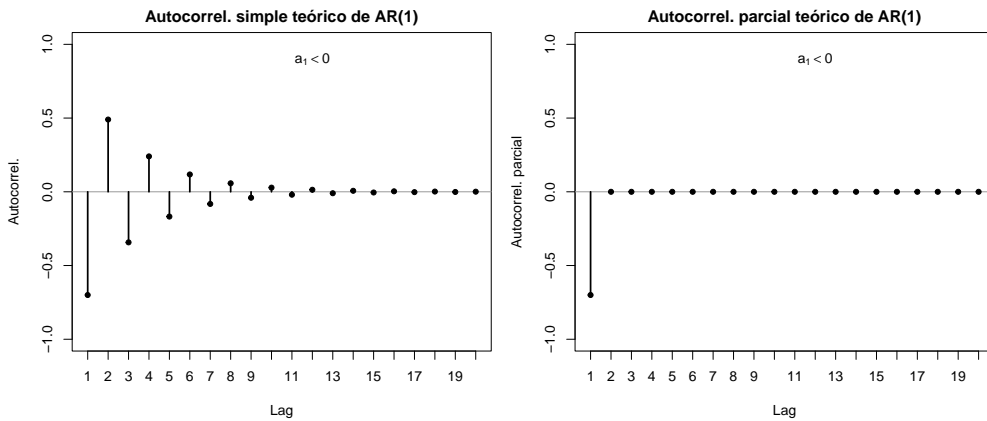


Figure 16: Correlogramas teóricos $AR(1)$

2) Modelos $AR(2)$:

a) Con los parámetros $a_1 > 0, a_2 > 0$:

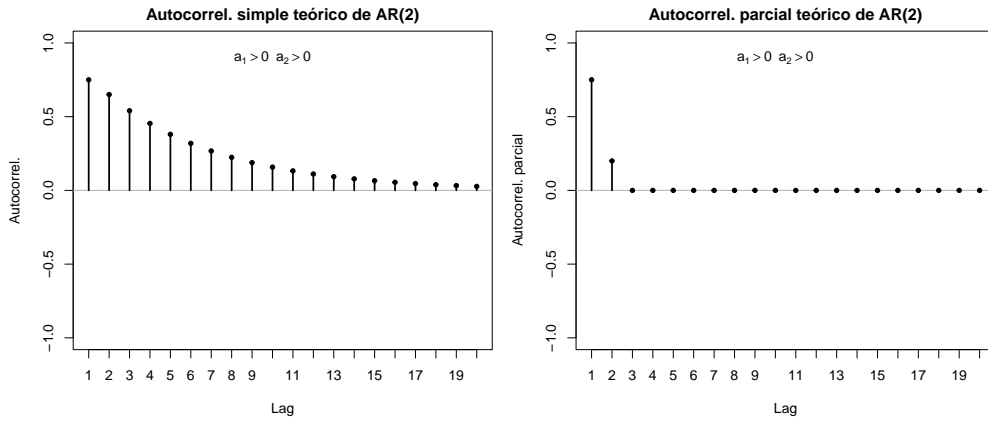


Figure 17: Correlogramas teóricos AR(2)

b) Con los parámetros $a_1 < 0$, $a_2 > 0$:

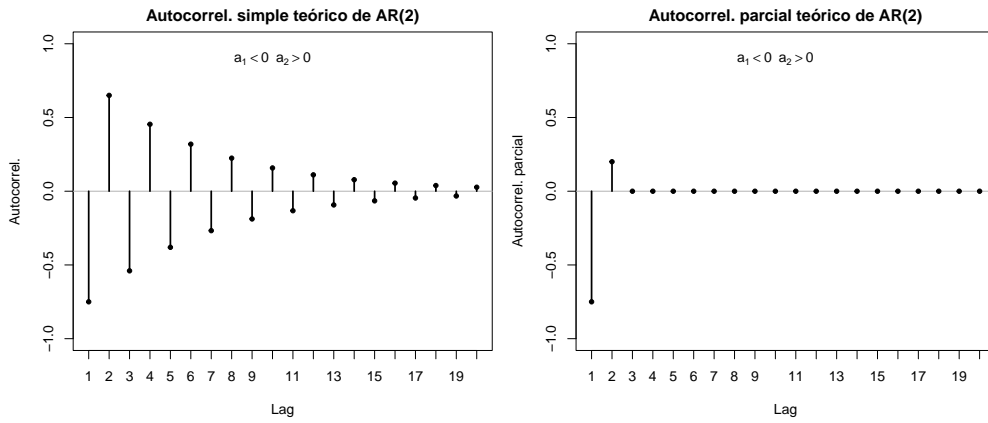


Figure 18: Correlogramas teóricos AR(2)

c) Con los parámetros $a_1 > 0$, $a_2 < 0$:

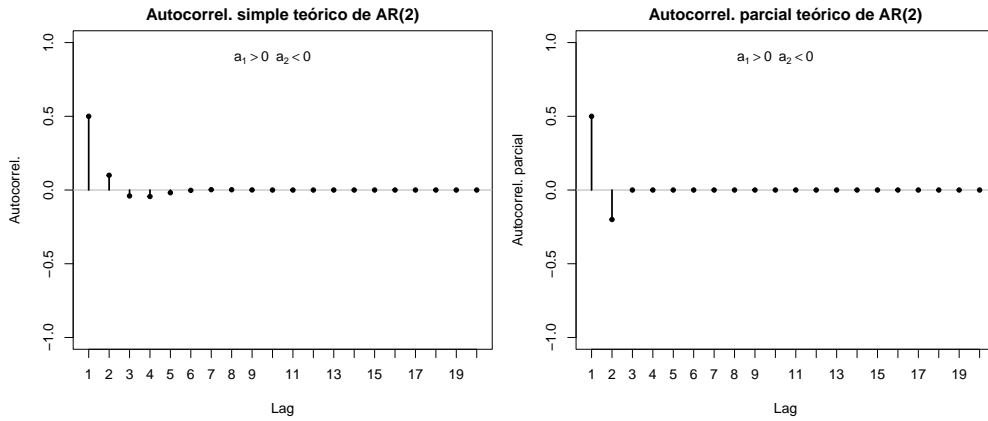


Figure 19: Correlogramas teóricos AR(2)

d) Con los parámetros $a_1 < 0$, $a_2 < 0$:

6.3.2) Estimación de los parámetros del modelo

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

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} + E(\varepsilon_t X_{t-j}) \quad (17)$$

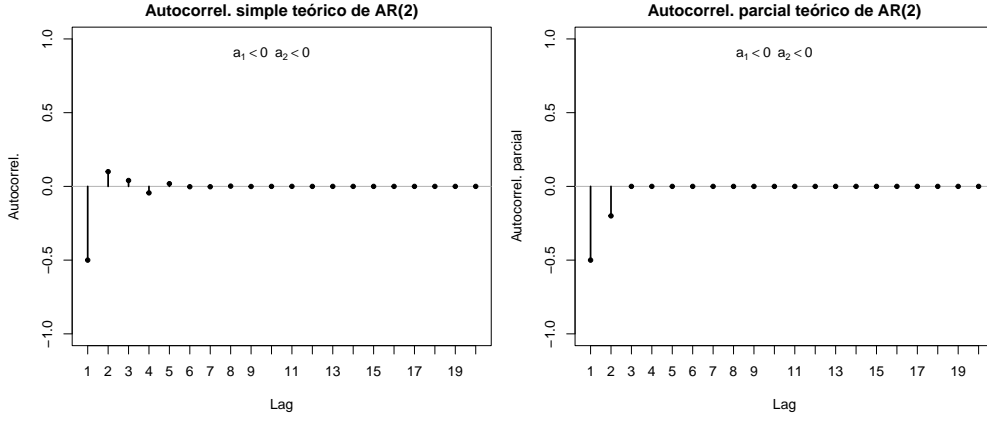


Figure 20: Correlogramas teóricos AR(2)

multiplicando por X_{t-j} y tomando esperanzas

$$E(X_t X_{t-1}) = a_1 E(X_{t-1} X_{t-j}) + a_2 E(X_{t-2} X_{t-j}) + \cdots + a_p E(X_{t-p} X_{t-j}) + E(\varepsilon_t X_{t-j}) \quad (18)$$

Obsérvese que al ser $E(\varepsilon_t \varepsilon_{t-j}) = 0$ para $j > 0$, se tiene que $E(\varepsilon_t X_{t-j}) = 0$ para $j > 0$ y $E(\varepsilon_t X_t) = \sigma_\varepsilon^2$, donde σ_ε^2 denota la varianza común de las perturbaciones ε_t . Por tanto, reescribiendo (18) en términos de covarianzas (hemos supuesto proceso de media cero), tendremos las **ecuaciones de Yule-Walker** para el proceso $AR(p)$ usando covarianzas:

$$\begin{aligned} \gamma_0 &= a_1 \gamma_1 + a_2 \gamma_2 + \cdots + a_p \gamma_p + \sigma_\varepsilon^2 \\ \gamma_j &= a_1 \gamma_{j-1} + a_2 \gamma_{j-2} + \cdots + a_p \gamma_{j-p} \quad j > 0 \end{aligned}$$

Por otro lado, dividiendo en las ecuaciones anteriores por γ_0 (varianza común del proceso), tendremos las **ecuaciones de Yule-Walker** para un proceso $AR(p)$ usando correlaciones:

$$\begin{aligned} \rho_0 &= 1 \\ \rho_j &= a_1 \rho_{j-1} + a_2 \rho_{j-2} + \cdots + a_p \rho_{j-p} \quad j > 0 \end{aligned}$$

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

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

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

$$\rho_p = a_1 \rho_{p-1} + a_2 \rho_{p-2} + \cdots + a_p$$

Llamando:

$$\begin{aligned} \mathbf{a}' &= (a_1, a_2, \dots, a_p) & \rho' &= (\rho_1, \rho_2, \dots, \rho_p) \\ \mathbf{R} &= \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{pmatrix} \end{aligned}$$

el sistema anterior se escribe matricialmente:

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

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

Además, las ecuaciones de Yule-Walker coinciden con el criterio de mínimos cuadrados para los residuos. Así, si consideramos un proceso $AR(p)$:

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

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

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

donde \hat{x}_t denotamos el valor estimado de la serie. Los parámetros a_1, a_2, \dots, a_p que minimizan la suma de cuadrados de los residuos coinciden con la solución a las ecuaciones de Yule-Walker.

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

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

Si suponemos que la media del proceso es cero y su varianza es constante igual a la unidad, denotando por:

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

el sistema anterior se transforma en:

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

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

6.3.3) Ejemplo de modelo $AR(1)$

Consideremos el modelo Autorregresivo de orden 1 expresado mediante:

$$X_t = 0.9X_{t-1} + \varepsilon_t$$

Una serie temporal generada por el proceso anterior, viene dada en el gráfico de la siguiente figura.

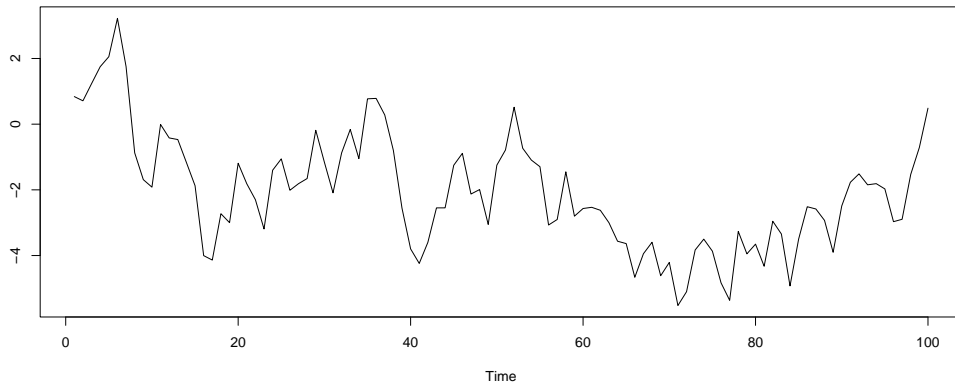


Figure 21: Ejemplo de serie $AR(1)$

A continuación se muestran las correlaciones simples y parciales estimadas a partir de la serie $AR(1)$ simulada.

Como se observa, el correlograma simple representa varias correlaciones significativas decreciendo de forma sinusoidal, pero

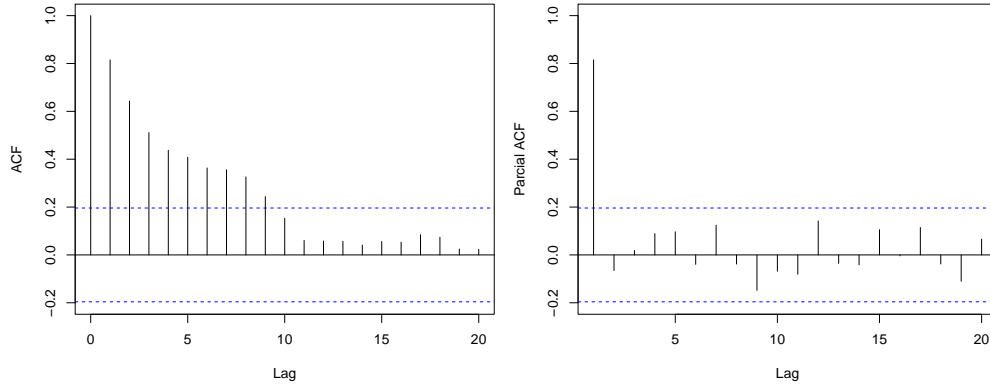


Figure 22: Correlogramas simple y parcial serie AR(1)

sólo es significativa la correlación parcial correspondiente al retardo 1, lo que indica que se trata de un modelo AR(1).

6.4) Modelos de medias móviles, $MA(q)$

Otra alternativa de representación de la dependencia respecto al pasado consiste en considerar el valor actual como el resultado de la combinación de q factores aleatorios independientes entre si más una perturbación aleatoria contemporánea al modelo:

$$X_t = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \cdots + b_q\varepsilon_{t-q}$$

donde ε_t es un ruido blanco gaussiano. Al modelo anterior se le denomina proceso de medias móviles.

Nota: Dependiendo del software usado, la notación usada para representar los modelos $MA(q)$ puede variar considerando los coeficientes b_j cambiados de signo. En nuestro caso, usaremos la notación usual del software R.

Obviamente, se tiene:

$$\begin{aligned} E(X_t) &= 0 \\ \text{Var}(X_t) &= \sigma_\varepsilon^2(1 + b_1^2 + b_2^2 + \cdots + b_q^2) \end{aligned}$$

Usando el operador de retardos tendremos:

$$X_t = (1 + b_1B + b_2B^2 + \cdots + b_qB^q)\varepsilon_t$$

o equivalentemente:

$$X_t = [b_q(B)]\varepsilon_t$$

donde el polinomio $b_q(x) = (1 + b_1x + b_2x^2 + \cdots + b_qx^q)$ recibe el nombre del polinomio característico del proceso de medias móviles.

Estos procesos siempre son estacionarios (no necesitan condición sobre los parámetros b_i). Además, un proceso $MA(q)$ se puede ver como un $AR(\infty)$ siempre y cuando el proceso sea invertible (las raíces del polinomio característico deben estar fuera del círculo unidad).

$$X_t = [b_q(B)]\varepsilon_t \implies [b_q(B)]^{-1}X_t = \varepsilon_t$$

6.4.1) Determinación del orden del modelo

Consideremos el proceso $MA(q)$ de media nula:

$$X_t = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \cdots + b_q\varepsilon_{t-q}$$

Recordemos que la función de covarianzas viene dada por:

$$\gamma_k = \text{Cov}(X_t, X_{t-k}) = E(X_t X_{t-k}) - E(X_t)E(X_{t-k})$$

Entonces, teniendo en cuenta que la función de medias es nula y que $(\varepsilon_t)_t$ es un proceso de ruido blanco gaussiano, la función de covarianzas para los proceso $MA(q)$ quedaría:

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 b_k + \sigma_\varepsilon^2 \sum_{j=k+1}^q b_{j-k} b_j & \text{si } k = 1, 2, \dots, q \\ 0 & \text{si } k > q \end{cases}$$

de manera que las correlaciones serán nulas para retardos mayores de q :

$$\rho_k = 0 \text{ si } k > q$$

Este último resultado tendrá una gran importancia práctica porque nos permitirá identificar el orden del proceso MA al que se ajusta una serie temporal dada. *Para ello, observaremos el autocorrelograma de la serie: el número de coeficientes que sean 'significativamente' distintos de cero indican el orden del proceso MA .*

Veamos en la siguiente tabla un resumen de las características básicas de los procesos AR y MA

	AR(p)	MA(q)
Expresión	$\varepsilon_t = (1 - a_1 B - \dots - a_p B^p) X_t$	$X_t = (1 + b_1 B + \dots + b_q B^q) \varepsilon_t$
Estacionario	Raíces del polinomio característico fuera del círculo unidad	Siempre
Correlograma	Infinitos valores no nulos decreciendo de manera amortiguada	Valores no nulos hasta un retardo q , el resto nulos
Correlograma parcial	Valores no nulos hasta un retardo p , el resto nulos	Infinitos valores no nulos decreciendo de manera amortiguada

6.4.2) Estimación de los parámetros del modelo

La estimación de los parámetros en modelos MA resulta más complicada que en modelos AR puesto que las ecuaciones son no lineales en los parámetros y para resolverlas es necesario recurrir a procedimientos iterativos.

Además, rara vez se trabaja en la práctica con un sistema MA puro, sino que se utiliza una combinación de modelo AR y MA dando lugar a los llamados modelos $ARMA$ que trataremos en la siguiente sección.

6.4.3) Ejemplo de modelo $MA(1)$

Consideremos el modelo de Medias Móviles de orden 1 expresado mediante:

$$X_t = \varepsilon_t + 0.8\varepsilon_{t-1}$$

Una serie temporal generada por el proceso anterior, viene dada en el gráfico de la siguiente figura.

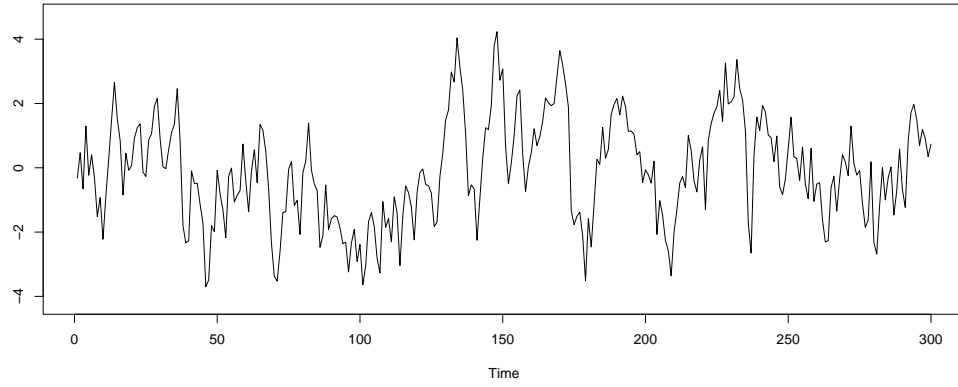


Figure 23: Ejemplo de serie MA(1)

A continuación se muestran las correlaciones simples y parciales estimadas a partir de la serie $MA(1)$ simulada.

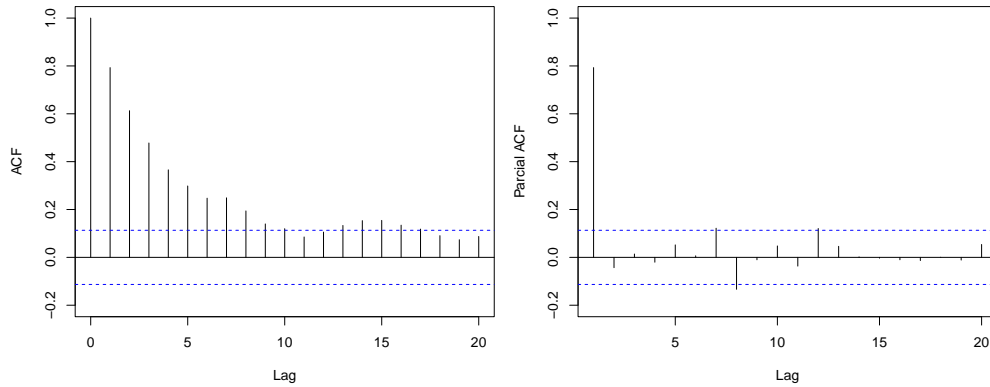


Figure 24: Correlogramas simple y parcial serie $MA(1)$

Como se observa, el correlograma simple presenta sólo una correlación significativa, mientras que el correlograma parcial presenta varias correlaciones parciales significativas, lo que indica que se podría tratar de un modelo $MA(1)$.

6.5) Modelos $ARMA(p, q)$

Existen procesos que encuentra su representación óptima mediante una combinación de los dos modelos anteriores. Tales modelos reciben el nombre de procesos autorregresivos de medias móviles ($ARMA$). Su expresión vendrá dada por:

$$X_t - a_1 X_{t-1} - a_2 X_{t-2} - \dots - a_p X_{t-p} = \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q}$$

y usando el operador de retardos:

$$(1 - a_1 B - a_2 B^2 - \dots - a_p B^p) X_t = (1 + b_1 B + b_2 B^2 + \dots + b_q B^q) \varepsilon_t$$

o equivalentemente:

$$[a_p(B)] X_t = [b_q(B)] \varepsilon_t$$

donde $a_p(x)$ y $b_q(x)$ son los polinomios característicos.

El proceso $ARMA$ es estacionario si las raíces del polinomio $a_p(x)$ están fuera del círculo unidad, y será invertible si las raíces del polinomio $b_q(x)$ están fuera del círculo unidad.

Como se observa, el modelo contiene p retardos del autorregresivo y q medias móviles, por consiguiente se representará por $ARMA(p, q)$, es decir, el primer índice indicará el orden de la autorregresión y el segundo el de las medias móviles.

Observación 6.5.1 Los modelos AR y MA se pueden obtener como caso particular del modelo $ARMA$ haciendo $q = 0$ o bien $p = 0$.

Si se cumplen las condiciones para que el modelo sea considerado estacionario, todo modelo $ARMA(p, q)$ se puede considerar como un $MA(\infty)$ de la forma:

$$X_t = \frac{(1 + b_1B + b_2B^2 + \dots + b_qB^q)}{(1 - a_1B - a_2B^2 - \dots - a_pB^p)} \varepsilon_t$$

6.5.1) Estimación de los parámetros del modelo

El problema de estimación de los parámetros $ARMA$ a partir de n valores observados de la serie, x_1, x_2, \dots, x_n , resulta mucho más complicado que en modelos AR puesto que como veremos seguidamente las ecuaciones del sistema que se obtienen al aplicar mínimos cuadrados no son lineales. Veamos el caso más sencillo, un proceso $ARMA(1, 1)$. Este proceso viene caracterizado por:

$$X_t = a_1X_{t-1} + b_1\varepsilon_{t-1} + \varepsilon_t$$

los residuos ε_t vendrán dados por:

$$\begin{aligned} e_1 &= x_1 - \hat{x}_1 = x_1 \\ e_2 &= x_2 - \hat{x}_2 = x_2 - a_1x_1 - b_1e_1 \\ &\dots \\ e_n &= x_n - \hat{x}_n = x_n - a_1x_{n-1} - b_1e_{n-1} \end{aligned}$$

Obviamente e_3 depende de e_2 y este a su vez depende de e_1 . En general, los residuos contendrán potencias y productos cruzados de los parámetros a_1 y b_1 y por tanto no se tratará de un sistema lineal.

6.6) Procesos lineales no estacionarios: modelo ARIMA

6.6.1) Paseo aleatorio

Hemos visto en las secciones anteriores que los procesos MA finitos son siempre estacionarios y que los AR lo son si las raíces del polinomio característico están fuera del círculo unidad. Consideremos un proceso $AR(1)$:

$$X_t = a_1X_{t-1} + \varepsilon_t$$

Si $|a_1| > 1$, el proceso resulta "explosivo" (crece rápidamente), si $|a_1| < 1$, el proceso es estacionario, mientras que si $|a_1| = 1$, no es ni explosivo ni estacionario. Recordemos que este tipo de procesos reciben el nombre de **paseos aleatorios**. Como veremos en el siguiente apartado, se trata de un proceso integrado de orden 1 (puesto que su primera diferencia $X_t - X_{t-1} = \varepsilon_t$ es estacionaria).

Observar que si denotamos al operador diferencia mediante:

$$\nabla X_t = X_t - X_{t-1}$$

el paseo aleatorio se escribe como:

$$\nabla X_t = \varepsilon_t$$

6.6.2) Procesos ARIMA

Como acabamos de mostrar, un paseo aleatorio es un proceso $AR(1)$ que no es estacionario, pues la raíz de su polinomio característico es unitaria. Pero si tomamos diferencias de orden 1, conseguimos que el proceso resultante sea estacionario. Esta idea puede generalizarse para cualquier proceso $ARMA$, dando lugar a lo que se conocen como *procesos autorregresivos de medias móviles (ARIMA)*.

En el ejemplo del paseo antes mencionado se ha obtenido un proceso estacionario aplicando el operador diferencia una vez.

Definición 6.6.1 Diremos que un proceso sigue un modelo $ARIMA(p, d, q)$ si al aplicar d veces el operador diferencia se obtiene un proceso estacionario $ARMA(p, q)$.

Así, si denotamos por:

$$Y_t = \nabla^d X_t = (1 - B)^d X_t$$

a la serie obtenida al aplicar d veces el operador diferencia, si la serie original X_t seguía un modelo $ARIMA(p, d, q)$ se puede expresar mediante:

$$(1 - a_1 B - a_2 B^2 - \dots - a_p B^p)(1 - B)^d X_t = (1 + b_1 B + b_2 B^2 + \dots + b_q B^q) \varepsilon_t$$

Observación 6.6.1 Los modelos AR , MA y $ARIMA$ se obtienen como caso particular de los modelos $ARIMA$, tomando como parámetro $d = 0$.

6.7) Identificación del modelo, validación y predicciones

Una vez descritas algunas de las propiedades más importantes de los modelos AR , MA , $ARMA$ y $ARIMA$, vamos a estudiar cómo identificar el proceso estocástico del que procede la serie en estudio.

Podemos decir que el objetivo concreto perseguido a lo largo de este tema es intentar identificar el proceso $ARIMA(p, d, q)$ que probablemente haya generado nuestra serie. Como es habitual, debemos comenzar realizando un análisis descriptivo previo de la serie que incluya detección de outliers. En el proceso de identificación del proceso generador de la serie seguiremos las siguientes etapas:

1. Analizar la estacionariedad de la serie y determinar el parámetro d .
2. Determinar el orden de la parte autorregresiva (parámetro p) y de la parte media móvil (parámetro q).
3. Estimar los coeficientes del modelo: parámetros a_i, b_j y *cte*.
4. Determinar los residuos $e_t = \hat{\varepsilon}_t$ y validar el modelo.
5. Realizar predicciones.

6.7.1) Análisis de la estacionariedad de la serie

La primera etapa en la identificación del proceso generador de la serie consiste en determinar si la serie (proceso generador) es estacionaria. Como ya adelantamos en un apartado anterior, los incumplimientos de la estacionariedad suelen deberse a que la función de medias o la función de varianzas no resulten constantes.

Como herramienta para verificar la estacionariedad en varianza, podemos usar la representación gráfica de la serie: si observamos que las fluctuaciones de la serie se amplifican con el tiempo o con el nivel de la serie, será indicativo de que la función de varianzas no permanece constante.

En el caso de que la serie no sea estacionaria en varianza, se suele realizar una transformación de Box-Cox, siendo la más frecuente tomar logaritmos neperianos en los datos para conseguir varianza constante.

Una vez conseguida la estacionariedad en varianza, analizaremos si es estacionaria en media. Como herramienta para verificar la estacionariedad en media, podemos usar la representación gráfica de la serie: si la trayectoria de la serie oscila aleatoriamente alrededor de un valor constante, será indicativo de que la función de medias es constante, pero si observamos que el nivel de la serie varía a lo largo del tiempo, será indicativo de no estacionariedad en media.

En el caso de que la serie no sea estacionaria en media, en general se conseguirá que se convierta en estacionaria tomando diferencias de orden d . Por ejemplo, si observamos tendencia lineal en la serie, tomando diferencias de orden 1 conseguiremos un nivel constante de la serie, y si observamos tendencia cuadrática, tomando diferencias de orden 2 se conseguirá la estacionariedad en media.

Por otra parte, también podemos observar el autocorrelograma de la serie: si las correlaciones estimadas no decrecen rápidamente con el retardo, podría indicarnos que el proceso generador tiene una raíz del polinomio característico igual a

uno (paseo aleatorio) y por tanto no es estacionario. En estos casos se suelen tomar diferencias de orden uno y se vuelve a observar el autocorrelograma. En general basta con tomar diferencias de orden uno o dos para lograr la estacionariedad.

6.7.2) Determinación del orden de la parte *AR* y de la parte *MA*

Como herramientas básicas para determinar los órdenes p y q del modelo *ARMA* se suelen utilizar tanto el correlograma simple de la serie como el correlograma parcial.

Para establecer algunas pautas a la hora de determinar los valores de p y q , debemos recordar las propiedades de los modelos *AR* y *MA* estudiadas en apartados anteriores.

- En un modelo *AR*(p), las correlaciones parciales teóricas son nulas para retardos mayores de p . Sin embargo, las correlaciones teóricas nunca se hacen cero, pero decaen rápidamente a partir del retardo p .

Por tanto, si se trata de un modelo autorregresivo, el orden p se puede determinar a partir del correlograma parcial identificando los valores significativos del mismo.

- En un modelo *MA*(q), las correlaciones teóricas son nulas para retardos mayores de q . Sin embargo, las correlaciones parciales teóricas nunca se hacen cero, pero decaen rápidamente a partir del retardo q .

Por tanto, si se trata de un modelo de medias móviles, el orden q se puede determinar a partir del correlograma simple identificando los valores significativos del mismo.

- En un modelo *ARMA*(p, q) tanto las correlaciones como las correlaciones parciales teóricas nunca se hacen cero.

Por tanto, si se trata de un modelo *ARMA*, será difícil identificar los órdenes p y q , puesto que no se produce un corte ni en el autocorrelograma simple ni en el parcial. Podremos al menos proponer algunos modelos como candidatos de partida y posteriormente valorar si son reducibles.

Evidentemente, si trabajáramos con las *correlaciones y correlaciones parciales teóricas* el proceso de identificación se simplificaría bastante. Sin embargo, en la práctica siempre trabajaremos con las *correlaciones y correlaciones parciales estimadas* a partir de la serie en estudio, de manera que el problema de identificación del modelo resulta todavía más complejo. De hecho, en muchas ocasiones el autocorrelograma estimado a partir de la serie suele presentar ciertas oscilaciones que no se corresponden con el modelo teórico.

Otro factor importante a la hora de determinar los órdenes de la parte autorregresiva y de medias móviles es el *tamaño de la serie*. La identificación será más fácil cuanto mayor sea el tamaño de la serie. En este sentido se han realizado simulaciones que muestran una gran diferencia entre las correlaciones (parciales) teóricas y las estimadas para series de tamaño pequeño (inferior a 60 observaciones).

6.7.3) Estimación de los coeficientes del modelo

Una vez determinados los órdenes p y q del modelo *ARMA*, necesitamos estimar los coeficientes a_i y b_j del modelo:

$$X_t - a_1X_{t-1} - a_2X_{t-2} - \cdots - a_pX_{t-p} = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \cdots + b_q\varepsilon_{t-q}$$

Además, debemos estimar la constante del modelo, en caso de que éste la incluya.

La estimación de los parámetros se puede realizar minimizando la suma de cuadrados residual o bien maximizando la verosimilitud, dando ambos procedimientos resultados similares (aunque no iguales). Como hemos comentado en secciones anteriores, si el modelo contempla parte *MA* será necesario resolver sistemas de ecuaciones no lineales. Por ese motivo, diferentes softwares pueden mostrar soluciones distintas para una misma serie dependiendo del algoritmo de resolución empleado.

En nuestro caso, realizaremos la estimación de los coeficientes del modelo haciendo uso del software *R*, que además suele incluir (como en el caso de la regresión lineal múltiple) contrastes sobre la significación de cada uno de los parámetros a_i, b_j , así como de la constante. Estos contrastes nos será de gran utilidad a la hora de identificar el modelo generador de la serie.

En este contexto también haremos uso del *principio de parsimonia*, según el cual debemos seleccionar como modelo óptimo aquel que contenga menor número de parámetros entre todos los modelos considerados adecuados.

6.7.4) Validación del modelo: análisis de los residuos

Para que el modelo propuesto en la etapa anterior sea adecuado, es necesario que los residuos (diferencia entre los valores observados de la serie y los ajustados por el modelo propuesto) se comporten como un ruido blanco gaussiano. Por tanto, la validez del modelo propuesto para por verificar las siguientes hipótesis sobre los residuos:

- Los residuos se comportan como una distribución normal (hipótesis de normalidad).
- Los residuos tienen varianza constante (hipótesis de homocedasticidad).
- Los residuos son independientes (hipótesis de independencia).

Esta validación coincide con la realizada en el caso de los modelos de regresión, por lo que se pueden usar procedimientos similares.

Por ejemplo, la *normalidad* se puede verificar mediante contrastes no paramétricos como los de Kolmogorov-Smirnov o de Shapiro-Wilks, o bien a través de métodos gráficos.

Por otra parte, podemos contrastar la *homocedasticidad* observando el gráfico de dispersión de residuos frente a los valores ajustados o bien mediante el gráfico de los residuos frente al tiempo: estos gráficos deben mostrar que los residuos se sitúan aleatoriamente alrededor del cero, dentro de una banda de amplitud constante.

Con respecto a la *independencia* de los residuos, haremos uso del autocorrelograma: éste no debe presentar ninguna correlación significativa, han de ser todas prácticamente nulas para que los residuos se supongan independientes.

6.7.5) Bondad del ajuste y selección del mejor modelo

Una medida usual para cuantificar la bondad de un ajuste consiste en calcular la suma de cuadrados residual del modelo ajustado. Por ejemplo, en análisis de regresión múltiple, se puede utilizar esta medida para determinar lo "bueno" que es el ajuste realizado.

Por tanto, un primer criterio para determinar el "mejor" modelo consistirá en seleccionar, dentro de los modelos considerados válidos por verificar las hipótesis de los residuos, aquel modelo que presente una menor suma de cuadrados residual. Sin embargo, este criterio no tiene en cuenta ni el número de parámetros del modelo (órdenes p y q), ni la magnitud de las observaciones. Es decir, si realizamos transformaciones en los datos, el modelo resultante no es comparable con el original a través de los residuos (por ejemplo, una transformación logarítmica siempre proporcionará residuos de menor magnitud que para los datos originales).

Existen otros criterios para la selección de modelos, entre los que destacaremos los criterios de información AIC, AIC corregido y BIC, basados en la verosimilitud y que tienen en cuenta el número de parámetros a estimar.

Es importante destacar que estos criterios de información no son adecuados para seleccionar el orden la diferenciación (parámetro d), pues los valores de la verosimilitud no son comparables para modelos con diferentes órdenes de diferenciación. Por tanto, es conveniente que la selección del orden d se realice de otra forma (por ejemplo, manualmente), y posteriormente se pueden usar los criterios AIC y BIC para determinar los órdenes p y q .

La **selección del modelo** se puede realizar de manera "manual" (realizando los pasos indicados al principio de la sección), o bien utilizando funciones que realizan la selección de forma automática. En la siguiente figura, extraída de Hyndman and Athanasopoulos (2021), se muestran los pasos a seguir en las dos alternativas:

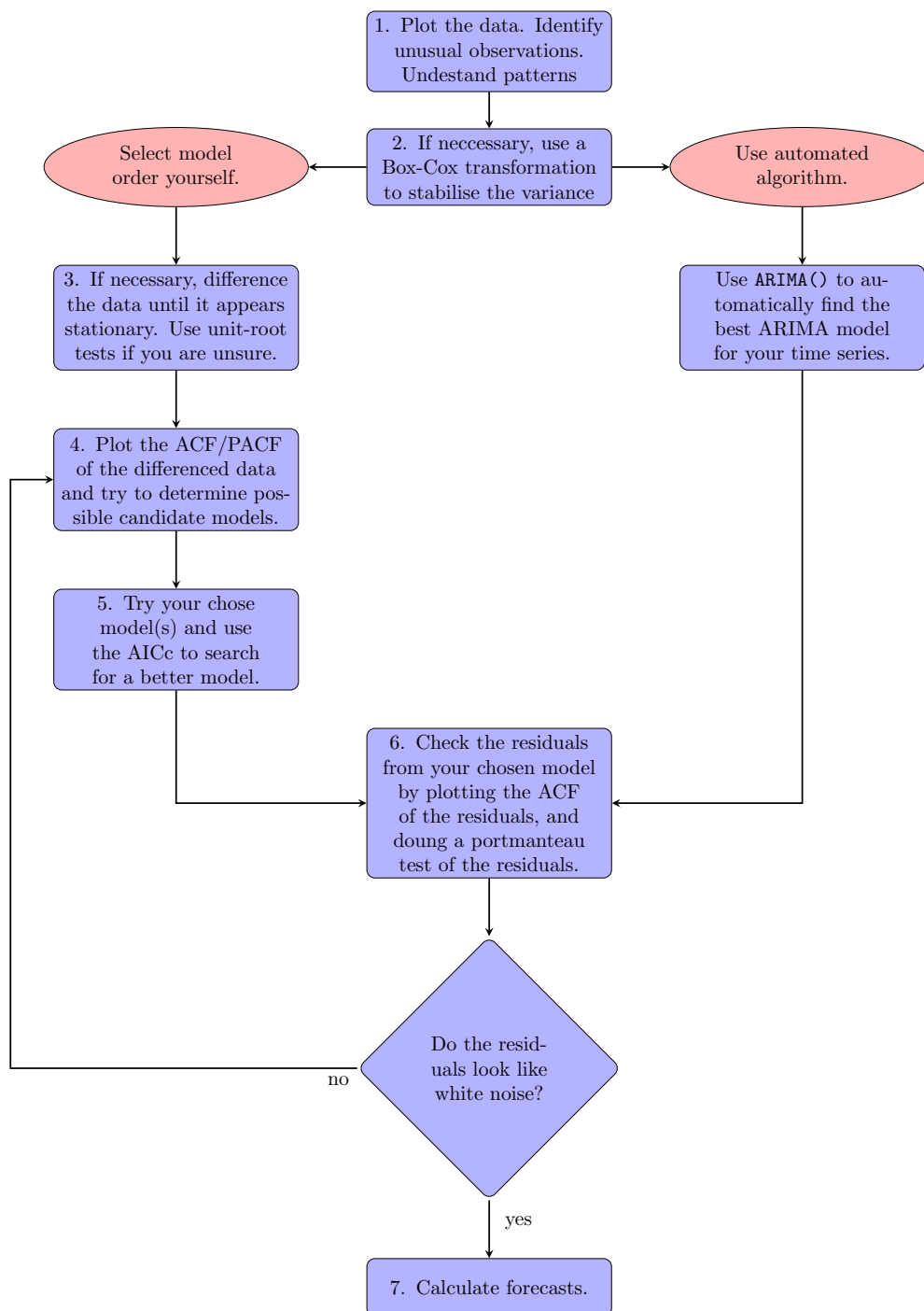


Figure 25: Pasos en metodología ARIMA

6.7.6) Predicciones

Una vez validado el modelo, el siguiente paso y fin último de nuestro estudio es realizar predicciones para valores futuros de la variable. Para obtener esas previsiones haremos uso del modelo teórico que se ha identificado como adecuado, proyectándolo hacia el futuro.

Además, las predicciones sirven para contrastar la adecuación de nuestro modelo: las discrepancias sistemáticas entre predicciones y valores observados cuestionarán la validez de nuestro modelo.

Supongamos que hemos observado la serie hasta un instante T , y denotemos por \hat{x}_{T+h} a la predicción de la serie en el instante $T+h$ a partir del modelo ajustado. La forma de proceder será la siguiente:

- La estimación del término (ruido) para instantes anteriores a T se corresponde con el error de previsión a un periodo vista, es decir:

$$\hat{\varepsilon}_t = x_t - \hat{x}_t \quad t \leq T$$

- Los términos de ruido posteriores al instante T se considerarán nulos:

$$\hat{\varepsilon}_t = 0 \quad t > T$$

- Los valores de la serie para instantes posteriores a T que se requieran en el cálculo de nuevas predicciones, se sustituyen por sus propias predicciones:

$$x_{T+h} = \hat{x}_{T+h} \quad h = 1, 2, \dots$$

donde el parámetro h representa el *horizonte de predicción*.

Veámos a modo de ejemplo cómo proceder a la hora de realizar predicciones con dos modelos concretos:

Ejemplo 6.7.1 Consideremos el modelo ARIMA $(1,1,0)$:

$$(1 - 0.6B)(1 - B)X_t = \varepsilon_t$$

Supongamos que conocemos el valor de la serie hasta un instante t_0 y pretendemos realizar previsiones más allá de ese instante. A partir del modelo teórico, tenemos que la expresión explícita del modelo de previsión será:

$$\hat{x}_{t_0+h} = 1.6\hat{x}_{t_0+h-1} - 0.6\hat{x}_{t_0+h-2} + \hat{\varepsilon}_{t_0+h} - 0.3\hat{\varepsilon}_{t_0+h-1} - 0.3\hat{\varepsilon}_{t_0+h-2} \quad \text{para } h = 1, 2, 3, \dots$$

y por consiguiente:

$$\begin{aligned} \hat{x}_{t_0+1} &= 1.6x_{t_0} - 0.6x_{t_0-1} - 0.5\hat{\varepsilon}_{t_0} - 0.3\hat{\varepsilon}_{t_0-1} \\ \hat{x}_{t_0+2} &= 1.6\hat{x}_{t_0+1} - 0.6x_{t_0} - 0.3\hat{\varepsilon}_{t_0} \\ \hat{x}_{t_0+3} &= 1.6\hat{x}_{t_0+2} - 0.6\hat{x}_{t_0+1} \\ &\vdots \\ \hat{x}_{t_0+h} &= 1.6\hat{x}_{t_0+h-1} - 0.6\hat{x}_{t_0+h-2}. \end{aligned}$$

Observa que para obtener las predicciones \hat{x}_{t_0+1} y \hat{x}_{t_0+2} necesitamos $\hat{\varepsilon}_{t_0-1}$ y $\hat{\varepsilon}_{t_0}$.

Dichos valores los podemos sacar de la fórmula recurrente:

$$\hat{\varepsilon}_{t_0-h} = \hat{x}_{t_0-h} - 1.6\hat{x}_{t_0-h-1} + 0.6\hat{x}_{t_0-h-2} + 0.5\hat{\varepsilon}_{t_0-h-1} + 0.3\hat{\varepsilon}_{t_0-h-2} \quad \text{para } h = 0, 1, 2, \dots$$

la cual permite obtener, a partir de los valores observados de la serie, todos los residuos de los tiempos observados $\hat{\varepsilon}_{t_0}, \hat{\varepsilon}_{t_0-1}, \hat{\varepsilon}_{t_0-2}, \dots$

Como se observa, si una serie incorpora medias móviles, la influencia directa del ruido se transmite tantos periodos hacia adelante como orden del proceso MA. A partir de este instante, la influencia de estos términos es indirecta a través de los valores obtenidos.

Nota: Las expresiones para obtener las predicciones puntuales permiten observar que, al aumentar el horizonte de predicción, se recurre a valores estimados en lugar de datos reales, lo que se traduce en mayores errores de predicción. La metodología ARIMA permite, además de obtener predicciones puntuales como hemos descrito arriba, obtener intervalos de predicción para instantes futuros.

6.7.7) Un ejemplo de modelo ARMA con R

Para ilustrar algunos de los conceptos vistos en este tema, primero vamos a generar una serie temporal a partir de un modelo ARMA, proporcionando nosotros mismo tanto los órdenes p y q , como el valor de los coeficientes del modelo. Posteriormente, analizaremos la serie simulada para determinar el modelo generador comprobando que se asemeja al modelo ARMA fijado en la simulación. Evidentemente, cuanto mayor sea el tamaño de la serie simulada, los resultados del ajuste serán más próximos al modelo generador real.

Concretamente, vamos a generar una serie a partir del siguiente modelo $ARMA(p = 2, q = 1)$:

$$X_t = 0.9X_{t-1} - 0.5X_{t-2} + 0.7\varepsilon_{t-1} + \varepsilon_t$$

donde ε_t representa un ruido blanco gaussiano.

```
set.seed(1234)
serie_simulada <- arima.sim(n = 1000,
                           list(ar = c(0.9, -0.5), ma = c(0.7)),
                           sd = 1)
```

```
ts.plot(serie_simulada)
```

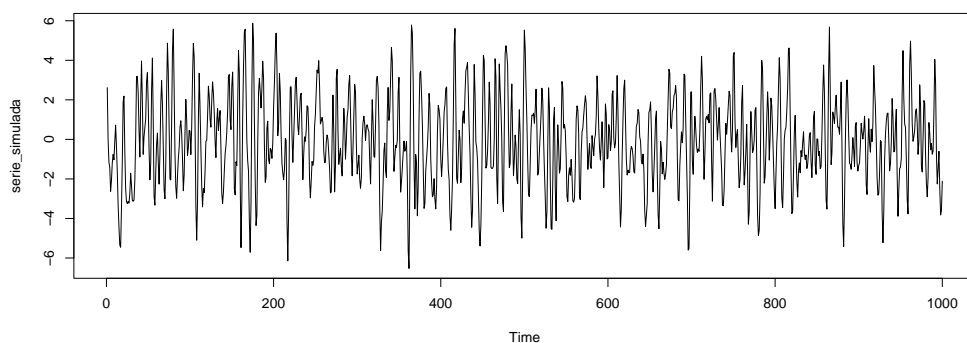


Figure 26: Serie simulada

```
acf(serie_simulada)
pacf(serie_simulada)
```

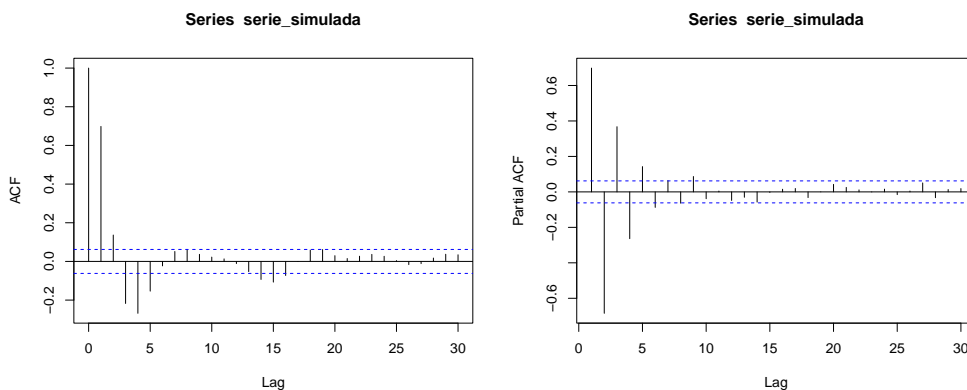


Figure 27: Correlogramas simple y parcial

Observamos que la serie es estacionaria (en media y varianza), además centrada (media cero). En los correlogramas vemos bastantes correlaciones simples y parciales significativas, lo que nos sugiere que el proceso generador probablemente contenga tanto parte AR (autorregresiva) como parte MA (medias móviles). Nótese que es especialmente significativo el retardo 1 en el correlograma simple y los retaros 1 y 2 en el correlograma parcial, lo que nos sugiere un modelo $ARMA(2, 1)$, es decir, un $ARIMA(2, 0, 1)$.

```
modelo_ARIMA201 <- arima(serie_simulada,
                        order = c(2, 0, 1),
                        include.mean = FALSE)
```

```
modelo_ARIMA201
```

```
##
```

```
## Call:
## arima(x = serie_simulada, order = c(2, 0, 1), include.mean = FALSE)
##
## Coefficients:
##          ar1          ar2          ma1
##      0.8900  -0.4853  0.7366
## s.e.  0.0311   0.0307  0.0248
##
## sigma^2 estimated as 0.9965:  log likelihood = -1418.74,  aic = 2845.49
```

Obtenemos estimaciones de los coeficientes muy parecidos a los coeficientes reales usados en la simulación. Podemos probar cómo sería el ajuste si consideramos algún orden superior en la parte *AR* y/o en la *MA*.

```
modelo_ARIMA301 <- arima(serie_simulada,
                        order = c(3, 0, 1),
                        include.mean = FALSE)

modelo_ARIMA301

##
## Call:
## arima(x = serie_simulada, order = c(3, 0, 1), include.mean = FALSE)
##
## Coefficients:
##          ar1          ar2          ar3          ma1
##      0.8938  -0.4910  0.0048  0.7340
## s.e.  0.0467   0.0599  0.0437  0.0345
##
## sigma^2 estimated as 0.9965:  log likelihood = -1418.74,  aic = 2847.47
```

```
modelo_ARIMA202 <- arima(serie_simulada,
                        order = c(2, 0, 2),
                        include.mean = FALSE)

modelo_ARIMA202

##
## Call:
## arima(x = serie_simulada, order = c(2, 0, 2), include.mean = FALSE)
##
## Coefficients:
##          ar1          ar2          ma1          ma2
##      0.8833  -0.4819  0.7446  0.0080
## s.e.  0.0657   0.0431  0.0737  0.0693
##
## sigma^2 estimated as 0.9965:  log likelihood = -1418.74,  aic = 2847.47
```

```
modelo_ARIMA302 <- arima(serie_simulada,
                        order = c(3, 0, 2),
                        include.mean = FALSE)

modelo_ARIMA302

##
## Call:
```

```
## arima(x = serie_simulada, order = c(3, 0, 2), include.mean = FALSE)
##
## Coefficients:
##          ar1          ar2          ar3          ma1          ma2
##          0.7350   -0.3505   -0.0712   0.8932   0.1180
## s.e.    1.3579    1.2106    0.6614   1.3573   0.9989
##
## sigma^2 estimated as 0.9965:  log likelihood = -1418.73,  aic = 2849.46
```

Obsérvese que al aumentar el orden de la parte *AR* y/o la *MA*, obtenemos valores de AIC mayores (peores modelos), indicativo de que el modelo óptimo es el *ARMA*(2,1).

6.8) Modelos ARIMA estacionales, SARIMA

El modelo SARIMA (Seasonal Autoregressive Integrated Moving Average) es una extensión de los modelos ARIMA que incorpora componente estacional. Es decir, variaciones que se repiten en intervalos regulares de tiempo (por ejemplo, mensuales, trimestrales o anuales).

El modelo se denota mediante $SARIMA(p, d, q)(P, D, Q)_L$, donde (p, d, q) representan los órdenes de la parte ARIMA no estacional y (P, D, Q) representan los mismos conceptos pero para la parte estacional. El parámetro L representa la periodicidad estacional (por ejemplo, 12 para datos mensuales con un patrón anual). El modelo completo combina tanto los componentes estacionales como no estacionales para capturar la estructura completa de la serie temporal.

La estimación de un modelo SARIMA sigue los mismos principios que los modelos ARIMA, pero con la adición de los componentes estacionales. Los pasos típicos incluyen:

- Utilizar gráficos de la serie temporal, de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), para determinar los posibles valores de los órdenes (p, d, q) y (P, D, Q) .
- Estimar los parámetros del modelo utilizando métodos como la máxima verosimilitud.
- Analizar los residuos del modelo asegurarse de que no presentan autocorrelación significativa y que son aproximadamente normales.

Como es habitual, para seleccionar el modelo SARIMA más adecuado, se pueden utilizar criterios de información como el AIC, el AIC corregido y el BIC. Una vez estimado y validado el modelo SARIMA, se pueden realizar predicciones futuras.

Nota: La comparación entre modelos ARIMA (o bien SARIMA) y ETS no debe realizarse usando los criterios de información AIC o BIC, porque son modelos de distinta clase y la verosimilitud se calcula de forma diferente en cada caso. En su lugar, conviene compararlos realizando validación cruzada y calculando medias de bondad de ajuste como RMSE, MAE o MAPE.

Hoja 6: Modelos ARIMA

Cuestiones Teóricas

- 1) ¿Qué es la estacionariedad en una serie temporal?
- 2) ¿Por qué es importante que una serie sea estacionaria para la modelización ARIMA?
- 3) ¿Qué es la diferencia en el contexto de las series temporales y cómo ayuda a estabilizar una serie?
- 4) Explica la diferencia entre diferenciación de primer orden, diferenciación de segundo orden y diferencia estacional.
- 5) ¿Cuál es la expresión formal para un modelo $AR(p)$?
- 6) ¿Qué diferencia hay entre correlaciones imples y correlaciones parciales?
- 7) ¿Cómo se identifica un ruido blanco gaussiano?
- 8) ¿Cómo se identifica un modelo $AR(p)$?
- 9) Para un modelo $AR(p)$, ¿cuándo tendremos que se trata de un proceso estacionario?
- 10) ¿Qué es un modelo de media móvil $MA(q)$?
- 11) ¿Cómo se identifica un modelo $MA(q)$? Demostrar que en un $MA(q)$ las correlaciones simples son nulas para retardos superiores a q .
- 12) Para un modelo $MA(q)$, ¿cuándo tendremos que se trata de un proceso estacionario?
- 13) Define formalmente un modelo $ARMA(p, q)$.
- 14) ¿Cómo se seleccionan los órdenes p y q en un modelo $ARMA$?
- 15) Indica la fórmula general de un modelo $ARIMA(p, d, q)$ usando el operador de retardo.
- 16) ¿Cómo se estiman los coeficientes de un modelo $ARIMA$? Deducir las ecuaciones de Yule-Walker en términos de correlaciones para un proceso $AR(p)$.
- 17) ¿Cuál es la utilidad de los criterios de información de Akaike (AIC) y de Bayes (BIC) en el contexto de metodología $ARIMA$?
- 18) ¿Cuál es la importancia de verificar que los residuos de un modelo $ARIMA$ se asemejen a ruido blanco?
- 19) Resume los pasos para analizar una serie temporal usando modelos $ARIMA$.
- 20) ¿Qué efectos tiene la transformación logarítmica en una serie temporal y cuándo se utiliza?

Problemas

- 21) En la siguiente tabla se muestra una serie temporal con datos trimestrales. Para modelar dicha serie se propuso el modelo $ARMA$ dado por

$$(1 - 0.9B + 0.5B^2)X_t = (1 + 0.7B + B^2)\varepsilon_t$$

- a) Escribe órdenes y polinomios característicos
- b) Prueba que el modelo es estacionario
- c) Completa los huecos de la tabla tomando como valores iniciales $X(-1) = 0.2$, $X(0) = 0.5$, $e(-1) = 0$, $e(0) = 0$.
- d) Calcula las predicciones para el año 2024.

t	Año	Trimestre	Serie	residuos
1	2020	T1	2.45	2.10
2	2020	T2	4.01	
3	2020	T3		-0.18
4	2020	T4	2.73	0.04
5	2021	T1	-1.03	-0.98
6	2021	T2	-1.74	1.20
7	2021	T3	-2.45	-1.26
8	2021	T4	0.03	1.05
9	2022	T1		-0.62
10	2022	T2	0.20	-0.50
11	2022	T3	-1.01	-0.16
12	2022	T4	-2.51	-0.89
13	2023	T1	-3.33	-0.79
14	2023	T2	-2.88	0.30
15	2023	T3	-0.85	
16	2023	T4	1.22	

22) En la siguiente tabla se muestra una serie temporal con datos trimestrales. Para modelar dicha serie se propuso el modelo *ARIMA* dado por

$$\left(1 - \frac{1}{2}B + \frac{1}{8}B^2\right)(1 - B)X_t = (1 + 0.4B - 0.2B^2 + 0.1B^3)\varepsilon_t$$

- a)** Escribe órdenes y polinomios característicos.
- b)** Comprueba que la serie obtenida al aplicar el operador diferencia es estacionaria.
- c)** Completa los huecos de la tabla tomando como valores iniciales $X(-2) = 0.1$, $X(-1) = 0.2$, $X(0) = 0.5$, $e(-2) = 0$, $e(-1) = 0$, $e(0) = 0$.
- d)** Calcula las predicciones para el año 2024.

t	Año	Trimestre	Serie	residuos
1	2020	T1	0.99	0.35
2	2020	T2		0.56
3	2020	T3	0.17	
4	2020	T4	-1.84	-0.05
5	2021	T1	-2.84	-0.70
6	2021	T2	-3.68	
7	2021	T3	-3.95	-0.08
8	2021	T4	-4.21	-0.14
9	2022	T1	-4.03	0.32
10	2022	T2	-3.94	-0.18
11	2022	T3	-3.94	0.13
12	2022	T4	-2.60	1.23
13	2023	T1	-1.71	-0.23
14	2023	T2	-1.17	0.59
15	2023	T3	-0.59	0.01
16	2023	T4		-0.70

Tema 7: Modelos de series temporales con variables exógenas

7.1) Introducción

En los temas anteriores, hemos visto técnicas para analizar series temporales donde se extrae información exclusivamente de la propia serie en estudio, ya sea descomponiendo la serie en varias componentes, o bien identificando el modelo generador a partir de las autocorrelaciones de los datos.

En este tema consideraremos el uso de otras variables diferentes a la variable interés para explicar y predecir la serie en estudio. Por ejemplo, consideremos el consumo eléctrico horario de un hogar. Por una parte, presenta varias estacionalidades: periodo 24 horas (estacionalidad diaria) y periodo 168 horas (estacionalidad semanal). Además, presenta correlación significativa para el retardo de 1 hora. Por tanto, podríamos plantear un modelo SARIMA para este tipo de series. Sin embargo, también parece claro que existen otros factores un modelo SARIMA para este tipo de series. Sin embargo, también parece claro que existen otros factores externos que influyen en el consumo eléctrico horario de un hogar, como son la temperatura (debido al uso de climatización eléctrica), el tipo de día (diferenciando entre laborables y festivos) o el precio de la energía (si el usuario tiene contrato indexado al precio de mercado). A estas otras variables, distintas a la variable objeto de estudio y con influencia sobre ésta, se les denominan **variables exógenas**, que actuarán como predictores.

Notación: En adelante, usaremos la notación y_t para referirnos al valor de la serie en estudio en el instante t , y usaremos $x_{i,t}$ con $i = 1, 2, \dots, k$ para referirnos a la variable exógena predictora x_i en el instante t . Esto supone un cambio respecto a la notación usada en los temas anteriores (donde la serie en estudio se denotaba por x_t), siendo ahora necesario distinguir entre variable de interés a predecir y variables predictoras.

7.2) El modelo de Regresión Múltiple para series temporales

Por analogía al modelo de Regresión Lineal Múltiple (RLM) visto en asignaturas anteriores, en el contexto de series temporales el modelo RLM se suele expresar de la siguiente forma:

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t \text{ con } \varepsilon_t \sim N(0, \sigma^2) \text{ independientes, } t \in \mathbb{T}.$$

En este caso, Y_t representa el proceso estocástico generador de la serie en estudio y_t , y asumimos que los predictores x_i no son variables aleatorias sino que son controlables, o al menos serán variables observables.

Recordemos que los errores deben tener media nula para proporcionar predicciones sesgadas. Además deben estar incorrelados entre sí y con los predictores, para garantizar que se ha extraído toda la información posible de los datos (parte determinista). La hipótesis de Normalidad y varianza constante (Homocedasticidad) se utiliza para la obtención de intervalos de predicción.

A continuación comentamos algunas etapas habituales en los análisis RLM (visto en asignaturas anteriores) que son de aplicación a los modelos de regresión de series temporales.

- La estimación de los coeficientes β_j , se puede realizar mediante el criterio de mínimos cuadrados o mediante estimación máximo verosímil. Recordemos que la función `lm()` de R permite la estimación de los coeficientes en un modelo RLM, pero también se pueden usar otras funciones más específicas para series temporales como `tslm()` del paquete `forecast` o la función `TSLM()` del paquete `fable`.
- La selección del modelo se puede realizar usando los métodos iterativos backward, forward y stepwise. Otra opción es buscar el mejor conjunto de regresores (predictores) atendiendo al cálculo de medidas de bondad del ajuste que permitan la comparativa entre modelos con distinto número de parámetros: R cuadrado ajustado (R^2_{adjusted}), validación cruzada (CV), criterios AIC , $AIC_{\text{corrected}}$ y BIC .
- La validación del modelo se realiza a través del análisis de los residuos (hipótesis de Normalidad, Homocedasticidad e Independencia), así como estudiando la multicolinealidad y la existencia de valores influyentes.
- Una vez seleccionado y validado el modelo, se pueden calcular tanto predicciones puntuales como intervalos de predicción para instantes futuros.

Indicar que también se pueden proponer modelos de regresión múltiple no lineales, por ejemplo considerando potencias de las variables exógenas o interacciones entre ellas.

7.3) Algunos predictores útiles

Existen predictores de especial utilidad en el contexto de modelos de regresión para series temporales, muchos de ellos debidos al "efecto calendario" sobre la serie en estudio. A continuación describimos con detalle algunos de ellos.

El tiempo como predictor de las Tendencias

Es común que los datos de la serie temporal muestren una tendencia. Por ejemplo, una tendencia lineal se puede modelar utilizando como predictor el tiempo t :

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t \text{ con } \varepsilon_t \sim N(0, \sigma^2) \text{ independientes, } t \in \mathbb{T}.$$

Y una tendencia cuadrática se puede modelar utilizando como predictores el tiempo t y su cuadrado t^2 :

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t \text{ con } \varepsilon_t \sim N(0, \sigma^2) \text{ independientes, } t \in \mathbb{T}.$$

Sin embargo, en general no se recomienda la modelización de la tendencia con potencias de t^2 o superiores, pues no suele reflejar de forma realista la evolución futura de la serie. En su lugar, es preferible el uso de tendencias lineales a trozos.

La creación del predictor tiempo puede realizarse manualmente o bien usando funciones especiales, como es el caso de `trend` dentro de la función `tslm()` o el caso de `trend()` dentro de la función `TSLM()`.

Variables dummy

En general, los predictores toman valores numéricos. Sin embargo, podemos disponer de predictores categóricos que tengan una influencia significativa en la serie analizada. Por ejemplo, en una serie de consumo eléctrico horario, el tipo de día (laborable o festivo) tiene influencia relevante. Esta situación se puede tratar dentro del marco de los modelos de regresión múltiple creando una "variable dummy" que toma el valor 1 para la categoría "día festivo" y el valor 0 para "día laborable".

Definición 7.3.1 Una variable dummy también conocida como "variable indicadora", es una variable binaria que sólo toma los valores 0 y 1.

Las variables dummy también se utilizan como predictor para distinguir periodos vacacionales especiales como la Semana Santa, que no tienen un calendario fijo: para los instantes correspondientes a Semana Santa, la variable dummy toma el valor 1, y para el resto de instantes toma el valor 0.

Si hay más de dos categorías, entonces la variable puede codificarse usando varias variables dummy. Concretamente, se necesitan tantas variables dummy como el número total de categorías menos uno. Algunas funciones de R (como por ejemplo `TSLM()`) tratan automáticamente esta situación si se especifica que la variable categórica es de tipo factor, de manera que no es necesario crear manualmente las variables dummy correspondientes.

Por ejemplo, si consideramos el predictor "día de la semana", entonces se crearán 6 variables dummy para identificar si se trata de un lunes, martes, ..., o domingo. En el modelo de regresión, la interpretación de cada uno de los coeficientes asociados a las variables dummy es una medida del efecto de esa categoría en relación con la categoría omitida. Para el predictor categórico "día de la semana" del ejemplo, si consideramos el "domingo" como categoría omitida, tendremos 6 variables dummy (la primera para indicar los lunes, la segunda para los martes, ..., la sexta para los sábados); el coeficiente de regresión de la variable dummy de los lunes se interpreta como el efecto de los lunes en relación a los domingos, efecto que se refleja en un aumento o disminución de la constante del modelo (intercept).

Observación 7.3.1 En particular, las variables dummy se pueden usar como predictores de la Estacionalidad de una serie (variables dummy estacionales). Si la serie presenta Estacionalidad de periodo L , consideraremos $L - 1$ variables dummy como indicadoras de cada estación.

Una variable dummy también se puede usar para tener en cuenta un valor atípico en los datos. En lugar de omitir el valor atípico, la variable dummy elimina su efecto. En este caso, la variable dummy toma el valor 1 para la observación atípica y 0 en todos los demás casos.

Variables de Intervención como variables dummy

A menudo es necesario modelar "intervenciones" que pueden afectar a la serie temporal en estudio, entendiendo como tal la ocurrencia de eventos especiales que no vienen recogidos en otras variables exógenas. Por ejemplo, la serie temporal del consumo eléctrico horario de un hogar, vendrá influenciada por eventos del tipo: (1) reemplazamiento de la iluminaria y electrodomésticos del hogar por dispositivos de muy bajo consumo; (2) corte puntual de una hora del suministro eléctrico.

En el caso (1), introduciremos una variable dummy que tome el valor 0 antes del evento y el valor 1 desde el instante del evento en adelante. Se trata de una variable que produce un escalón en el nivel de la serie (step variable), pues el evento produce un efecto permanente en la serie en estudio. Este tipo de variables también pueden reflejar el efecto de un evento en un periodo temporal concreto, por ejemplo, una subida considerable del precio de la electricidad durante varias horas podría cambiar temporalmente el comportamiento del consumidor por temas económicos.

En el caso (2), introduciremos una variable dummy que tome el valor 1 en el instante sin suministro eléctrico y que tome el valor 0 en el resto de instantes. Se trata de variable pico (spike variable), pues el evento produce un efecto sólo en un instante. Este escenario es equivalente a la gestión de atípicos (outliers) usando variables dummy.

Número de días laborables

En el caso de analizar series de datos mensuales, otro predictor que suele ser útil es el número de días laborables del mes, ya que puede variar sustancialmente de un mes a otro con efecto sobre la serie en estudio.

Por ejemplo, si consideramos la serie de consumo eléctrico mensual de una industria, puede haber meses con menos días laborales que otros afectando al consumo eléctrico de dicha industria. Una opción sería considerar un predictor que represente el número de días laborables de cada mes. Otra opción sería definir 7 predictores de la siguiente forma:

x_1 = número de lunes laborables del mes

x_2 = número de martes laborables del mes

...

x_7 = número de domingos laborables del mes

Series de Fourier como predictores de la Estacionalidad

Una alternativa al uso de variables dummy estacionales consiste en emplear las series de Fourier, especialmente en el caso de series con periodos estacionales largos. Tiene como base que cualquier función periódica puede aproximarse por una combinación de términos de senos y cosenos.

Si L es el periodo estacional, entonces los primeros términos de Fourier vienen dados por:

$$x_{1,t} = \sin\left(\frac{2\pi t}{L}\right); \quad x_{2,t} = \cos\left(\frac{2\pi t}{L}\right); \quad x_{3,t} = \sin\left(\frac{4\pi t}{L}\right); \quad x_{4,t} = \cos\left(\frac{4\pi t}{L}\right);$$

y en general, para frecuencia $k = 1, 2, 3, \dots$:

$$x_{2k-1,t} = \sin\left(\frac{2k\pi t}{L}\right); \quad x_{2k,t} = \cos\left(\frac{2k\pi t}{L}\right)$$

El número máximo de pares de términos de Fourier ("seno" y "coseno") que deben considerarse es $K = \left\lceil \frac{L}{2} \right\rceil$, donde $\lceil \cdot \rceil$ denota la parte entera. Si L es par, se puede comprobar que no se requieren más términos porque al aumentar k se obtienen series redundantes, bien idénticas o cambiadas de signo. Si L es impar, las frecuencias $k > \frac{L}{2}$ no dan lugar a términos redundantes, pero su contribución para describir la estacionalidad suele ser marginal.

Por ejemplo, para datos mensuales donde $L = 12$, tendremos que $K = 6$ y por tanto hay un total de 12 variables predictoras (6 senos y 6 cosenos). En realidad, se puede ver que el predictor número 11 se anula para todo $t = 1, 2, 3, \dots$:

$$x_{11,t} = \sin\left(\frac{12\pi t}{12}\right) = 0$$

En resumen, al usar términos de Fourier se necesitan menos predictores que con variables dummy, especialmente cuando el periodo L es grande. Por ejemplo, para datos semanales donde $L = 52$. Cuando el periodo es pequeño, por ejemplo para datos cuatrimestrales, no hay ventaja en usar términos de Fourier frente a dummies.

7.4) Consideraciones sobre el mundo y sobre las predicciones

En esta sección vamos a considerar algunos aspectos a tener en cuenta cuando se propone un modelo de Regresión Múltiple para una serie temporal, así como a la hora de realizar predicciones.

7.4.1) Regresión espuria

La regresión espuria en series temporales ocurre cuando se observa una relación significativa entre dos series temporales pueden estar influenciadas por tendencias o patrones a largo plazo similares, lo que da la apariencia de una relación significativa cuando, en realidad, no hay una relación causal directa entre las variables.

Las series temporales no estacionarias a menudo presentan una alta correlación simplemente debido a sus tendencias comunes a largo plazo, no porque estén relacionadas causalmente. Por ejemplo, consideremos las series de la siguiente figura, una representando los pasajeros en Australia y la otra la producción de arroz en Guinea.

Ambas parecen estar relacionadas simplemente porque tienen tendencias similares. Si realizamos una regresión lineal entre ambas, podríamos obtener un valor alto de R^2 y coeficientes significativos, sugiriendo una relación entre las dos series. Sin embargo, nada tienen que ver.

Indicar que en ocasiones, la regresión espuria puede proporcionar predicciones aceptables a muy corto plazo, pero en general no serán adecuadas cuando se avance en el tiempo.



Figure 28: Regresión espuria. Fuente: Hyndman and Athanasopoulos (2021)

Algunos signos de la regresión espuria son un R^2 elevado y alta autocorrelación de los residuos. Si las series no son estacionarias, una solución común es tomar diferencias para transformar las series en estacionarias.

7.4.2) Predicciones ex-ante y ex-post. Intervalos de predicción

Las predicciones ex-ante y ex-post son dos enfoques utilizados en la evaluación y comparación de modelos de pronóstico en el análisis de series temporales. A continuación se describe la diferencia entre ambos tipos de predicciones.

Predicciones ex-ante

Las predicciones ex-ante son aquellas que se realizan en el futuro, utilizando únicamente la información disponible hasta el momento presente. Esto supone tener que realizar predicciones tanto para los predictores (variables exógenas) como para la variable de interés, tal como lo haríamos en una situación real.

En este caso tenemos dos fuentes de incertidumbre: la propia del modelo ajustado para explicar nuestra variable de interés y la incertidumbre debida a la predicción de los predictores.

Por ejemplo, en una serie de consumo eléctrico horario donde contamos con la Temperatura ambiente como predictor, para realizar predicciones futuras "ex-ante" necesitaremos predecir primero la Temperatura y así poder predecir el consumo eléctrico. Las predicciones de Temperatura las podemos realizar nosotros mismos o bien obtenerlas de agencias como AEMET.

Predicciones ex-post

Las predicciones ex-post son aquellas que se realizan utilizando observaciones reales de los predictores en el periodo para el cual se está haciendo la predicción. Por tanto, no se trata de una situación realista, puesto que se está asumiendo que los predictores se pueden predecir de forma perfecta.

En este caso tenemos una única fuente de incertidumbre: la del modelo ajustado que explica nuestra variable de interés en función de los predictores. En esencia, permite una evaluación más clara del rendimiento del modelo en estudio, separando otras fuentes de incertidumbre ajenas a dicho modelo.

Por ejemplo, en la serie de consumo eléctrico horario, podemos plantear el cálculo de medidas de bondad del ajuste usando predicciones ex-ante y ex-post, con el fin de evaluar el rendimiento del modelo por sí solo y conjuntamente con la fuente de incertidumbre que supone predecir los predictores.

Cuando resulta complicado predecir los predictores, otra opción que se puede plantear consiste en utilizar los predictores retardados en el tiempo. Lo que se persigue es que las predicciones futuras de la variable de interés se puedan realizar sin necesidad de predecir los predictores.

Si queremos realizar predicciones con horizonte h , el modelo a considerar sería:

$$Y_{t+h} = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t \text{ con } \varepsilon_t \sim N(0, \sigma^2) \text{ independientes, } t \in \mathbb{T}.$$

Esta formulación hace que el modelo sea operativo en un escenario realista, además de que permite considerar el efecto retardado que podrían tener algunos predictores sobre la variable de interés.

Intervalos de predicción

En la mayoría de las situaciones, cuando se habla de predicciones, nos referimos a predicciones puntuales. Es decir, un valor concreto que sirve de estimación de la futura observación para la variable de interés. Sin embargo, no proporcionamos información de la incertidumbre o error asociado a dicha estimación puntual.

Una manera de cuantificar el error asociado a la predicción puntual se consigue proporcionando un intervalo de predicción, que es un intervalo de la recta real que contendrá con probabilidad alta a la observación futura que pretendemos predecir. Los niveles de confianza que se fijan para obtener intervalos de predicción suelen ser del 90% al 99%. La forma común de obtener un intervalo de predicción es:

$$\text{Intervalo Predicción} = (\text{Predicción Puntual} \pm \text{Margen Error})$$

donde la predicción puntual sabemos que se obtiene sustituyendo los valores de los predictores en el modelo ajustado y el margen de error va a depender del nivel de confianza fijado y la distribución de los errores del modelo. Por tanto, para

obtener un intervalo de predicción, es necesario conocer el modelo de distribución de los errores (de ahí la hipótesis que necesitamos verificar en el modelo de regresión).

Por ejemplo, en el caso de una serie con T observaciones modelizada mediante regresión lineal simple:

$$Y_t = \beta_0 + b_1 x_t + \varepsilon_t \text{ con } \varepsilon_t \sim N(0, \sigma^2) \text{ independientes, } t \in \mathbb{T}$$

el intervalo de predicción aproximado, al 95% viene dado por:

$$\hat{y} = 1.96 \hat{\sigma}_e \sqrt{1 + \frac{1}{T} + \frac{(x - \hat{x})^2}{(T-1)s_x^2}}$$

donde la predicción puntual es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

y la estimación de error estándar de la regresión simple es:

$$\hat{\sigma}_e = \sqrt{\frac{1}{T-2} \sum_{t=1, \dots, T} e_t^2}$$

Nota: El margen de error crece conforme el valor del predictor x se aleja de la media.

Si la serie se modeliza mediante regresión lineal múltiple:

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t \text{ con } \varepsilon_t \sim N(0, \sigma^2) \text{ independientes, } t \in \mathbb{T}$$

un intervalo de predicción aproximado, al $100 \cdot (1 - \alpha)\%$ de confianza, viene dado por:

$$\hat{y} \pm z_{1-\frac{\alpha}{2}} \hat{\sigma}_e \sqrt{1 + \mathbf{x}^* (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{x}^*)'}$$

donde α es el nivel de significación, $z_{1-\frac{\alpha}{2}}$ es el cuantil de la $N(0, 1)$, \mathbf{x}^* denota el vector fila con los nuevos valores de los predictores, X es una matriz de diseño del modelo RLM, \hat{y} es la predicción puntual dada por:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k$$

y $\hat{\sigma}_e$ es la desviación estándar residual dada por:

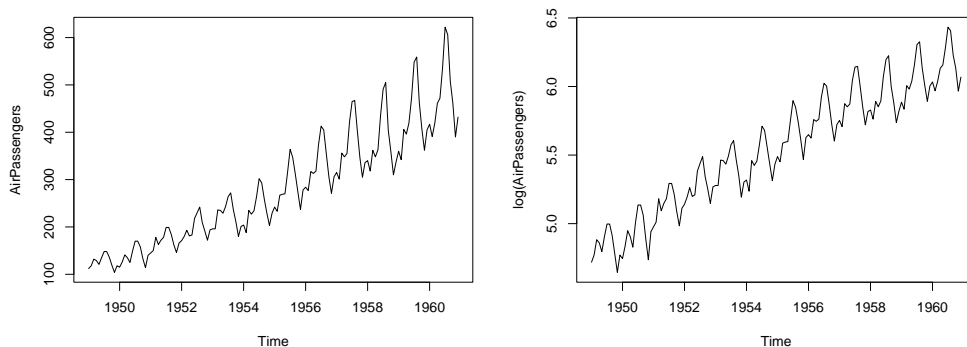
$$\hat{\sigma}_e = \sqrt{\frac{1}{T-k-1} \sum_{t=1, \dots, T} e_t^2}$$

7.5) Ejemplo

Consideremos la serie "AirPassengers" incluida en R, cuyos datos se corresponden con el número de pasajeros mensuales en aerolíneas norteamericanas desde enero de 1949 hasta diciembre de 1960. Obsérvese que es un objeto de tipo "ts" (time series) y que su representación refleja un esquema multiplicativo entre tendencia y estacionalidad.

Con el fin de estimar un modelo lineal (en lugar de no-lineal), trabajaremos con la serie transformada mediante logaritmos neperianos, dando lugar a un esquema aditivo entre tendencia y estacionalidad.

```
par(mfrow = c(1,2))
plot(AirPassengers)
plot(log(AirPassengers))
```

Parece claro que la serie `log(AirPassengers)` presenta estacionalidad y tendencia lineal en función del tiempo. Vamos a estimar un modelo de regresión lineal múltiple para explicar el comportamiento de la serie. Crearemos manualmente los predictores "tiempo" y "dummy estacionales".

```
t <- 1:144 # predictor tiempo
library(forecast)
estacional.dummy <- seasonaldummy(log(AirPassengers)) # predictores dummy estacionales

lm.model <- lm(log(AirPassengers) ~ t + estacional.dummy)
summary(lm.model)
```

```
##
## Call:
## lm(formula = log(AirPassengers) ~ t + estacional.dummy)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.156370	-0.041016	0.003677	0.044069	0.132324

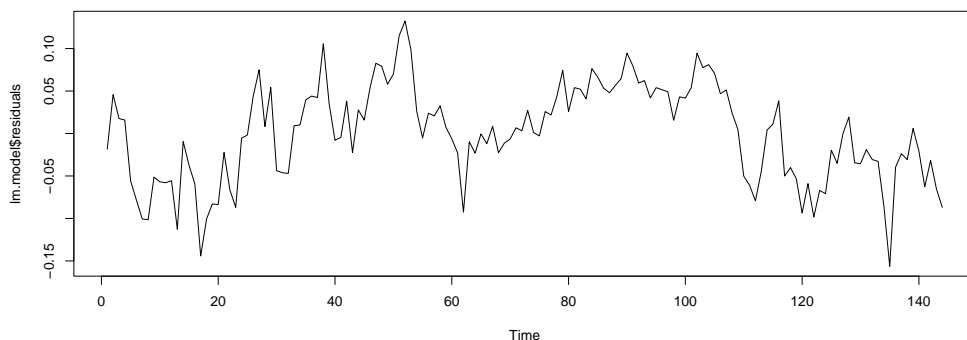
```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.7054593	0.0194850	241.491	< 2e-16 ***
t	0.0100688	0.0001193	84.399	< 2e-16 ***
estacional.dummyJan	0.0213211	0.0242461	0.879	0.380816
estacional.dummyFeb	-0.0007338	0.0242400	-0.030	0.975897
estacional.dummyMar	0.1294934	0.0242344	5.343	3.92e-07 ***
estacional.dummyApr	0.0982245	0.0242294	4.054	8.59e-05 ***
estacional.dummyMay	0.0958519	0.0242250	3.957	0.000124 ***
estacional.dummyJun	0.2179981	0.0242212	9.000	2.25e-15 ***
estacional.dummyJul	0.3219404	0.0242179	13.293	< 2e-16 ***
estacional.dummyAug	0.3126456	0.0242153	12.911	< 2e-16 ***
estacional.dummySep	0.1680110	0.0242132	6.939	1.64e-10 ***
estacional.dummyOct	0.0298527	0.0242118	1.233	0.219790
estacional.dummyNov	-0.1138650	0.0242109	-4.703	6.41e-06 ***

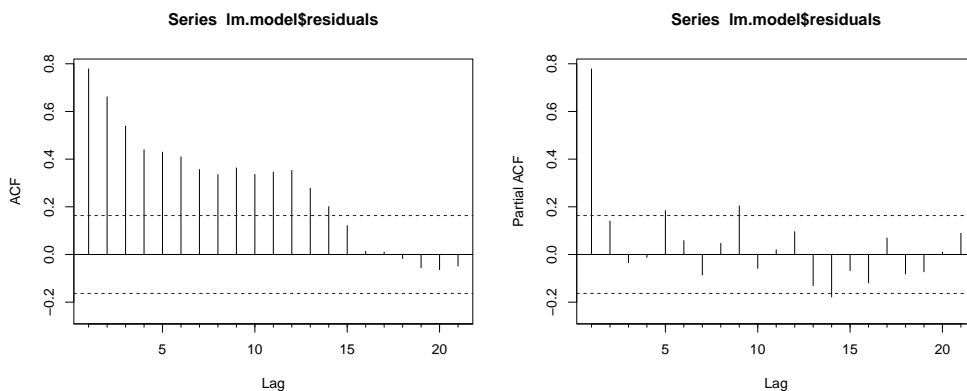
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0593 on 131 degrees of freedom
## Multiple R-squared:  0.9835, Adjusted R-squared:  0.982
## F-statistic: 649.4 on 12 and 131 DF, p-value: < 2.2e-16
```

Obtenemos una bondad del ajuste alta (Adjusted R-squared: 0.9764). Pero es importante verificar si los errores se comportan como un ruido blanco. Con los correlogramas simple y parcial concluimos que no es el caso. Por tanto, dichos errores deberían modelizarse con un ARIMA (véase la sección de regresión dinámica).

```
ts.plot(lm.model$residuals)
```



```
par(mfrow = c(1,2))
Acf(lm.model$residuals)
Pacf(lm.model$residuals)
```



Veamos cómo sería el ajuste usando la función `tslm()` del paquete `forecast`, que permite crear automáticamente los predictores de la tendencia lineal y las dummy estacionales:

```
tslm.model <- tslm(log(AirPassengers) ~ trend + season)
summary(tslm.model)
```

```
##
## Call:
## tslm(formula = log(AirPassengers) ~ trend + season)
##
## Residuals:
```

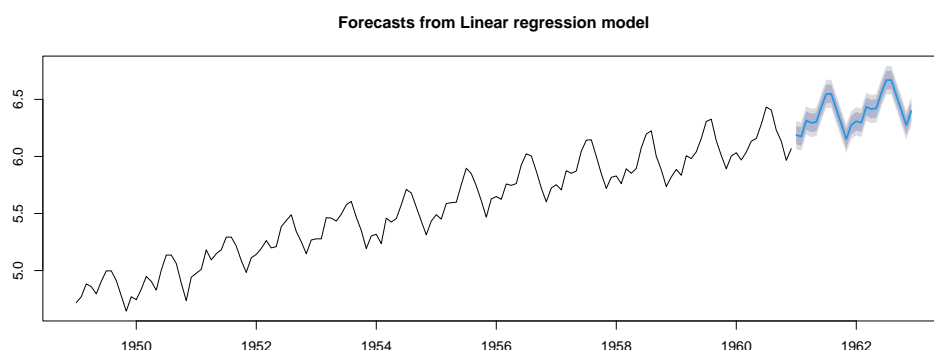
	Min	1Q	Median	3Q	Max
	-0.156370	-0.041016	0.003677	0.044069	0.132324

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.7267804	0.0188935	250.180	< 2e-16 ***
trend	0.0100688	0.0001193	84.399	< 2e-16 ***
season2	-0.0220548	0.0242109	-0.911	0.36400
season3	0.1081723	0.0242118	4.468	1.69e-05 ***
season4	0.0769034	0.0242132	3.176	0.00186 **

```
## season5      0.0745308  0.0242153   3.078  0.00254 **
## season6      0.1966770  0.0242179   8.121  2.98e-13 ***
## season7      0.3006193  0.0242212  12.411  < 2e-16 ***
## season8      0.2913245  0.0242250  12.026  < 2e-16 ***
## season9      0.1466899  0.0242294   6.054  1.39e-08 ***
## season10     0.0085316  0.0242344   0.352  0.72537
## season11     -0.1351861  0.0242400  -5.577  1.34e-07 ***
## season12     -0.0213211  0.0242461  -0.879  0.38082
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0593 on 131 degrees of freedom
## Multiple R-squared:  0.9835, Adjusted R-squared:  0.982
## F-statistic: 649.4 on 12 and 131 DF,  p-value: < 2.2e-16

plot(forecast(tslm.model, h = 24))
```



¿Y si usamos términos de Fourier en lugar de dummy estacionales?

Veamos cómo afecta K (número de términos de Fourier a incluir) a la modelización de la componente estacional:

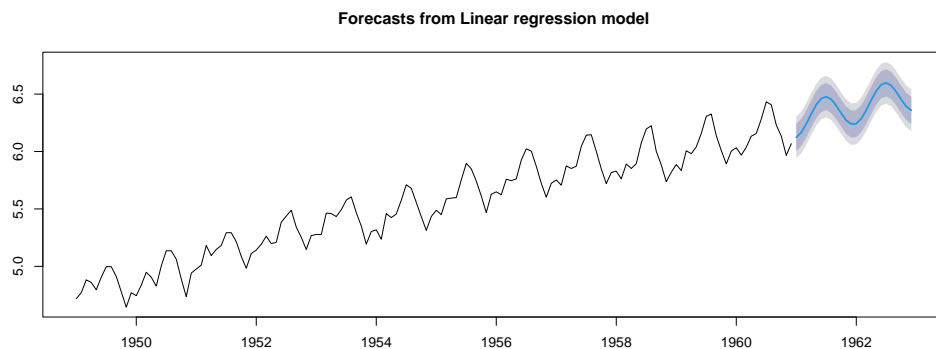
```
tslm.model_Fourier <- tslm(log(AirPassengers) ~
                             trend + fourier(log(AirPassengers), K = 1))

summary(tslm.model_Fourier)

##
## Call:
## tslm(formula = log(AirPassengers) ~ trend + fourier(log(AirPassengers),
##      K = 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.254905 -0.060940  0.004394  0.069431  0.186910
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.8145681   0.0150203  320.537  < 2e-16
## trend           0.0100360   0.0001798   55.809  < 2e-16
## fourier(log(AirPassengers), K = 1)S1-12 -0.0494811   0.0105698  -4.681  6.66e-06
## fourier(log(AirPassengers), K = 1)C1-12 -0.1417735   0.0105500 -13.438  < 2e-16
```

```
##
## (Intercept) ***
## trend ***
## fourier(log(AirPassengers), K = 1)S1-12 ***
## fourier(log(AirPassengers), K = 1)C1-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08951 on 140 degrees of freedom
## Multiple R-squared:  0.9598, Adjusted R-squared:  0.9589
## F-statistic: 1113 on 3 and 140 DF,  p-value: < 2.2e-16

plot(forecast(tslm.model_Fourier,
             data.frame(fourier(log(AirPassengers), K = 1, h = 24))))
```



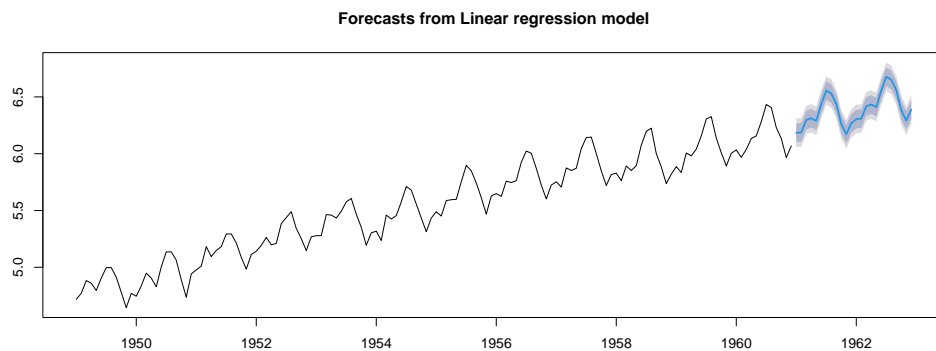
```
tslm.model_Fourier <- tslm(log(AirPassengers) ~
                           trend + fourier(log(AirPassengers), K = 4))

summary(tslm.model_Fourier)

##
## Call:
## tslm(formula = log(AirPassengers) ~ trend + fourier(log(AirPassengers),
##      K = 4))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.137989 -0.041654  0.000409  0.044630  0.133818
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.8121301   0.0102284  470.467 < 2e-16
## trend          0.0100696   0.0001225   82.208 < 2e-16
## fourier(log(AirPassengers), K = 4)S1-12 -0.0493556   0.0071920  -6.863 2.28e-10
## fourier(log(AirPassengers), K = 4)C1-12 -0.1418071   0.0071785 -19.754 < 2e-16
## fourier(log(AirPassengers), K = 4)S2-12  0.0786811   0.0071806  10.957 < 2e-16
## fourier(log(AirPassengers), K = 4)C2-12 -0.0228136   0.0071785  -3.178 0.001841
## fourier(log(AirPassengers), K = 4)S3-12 -0.0087300   0.0071785  -1.216 0.226072
## fourier(log(AirPassengers), K = 4)C3-12  0.0272914   0.0071785   3.802 0.000217
## fourier(log(AirPassengers), K = 4)S4-12  0.0256117   0.0071778   3.568 0.000499
```

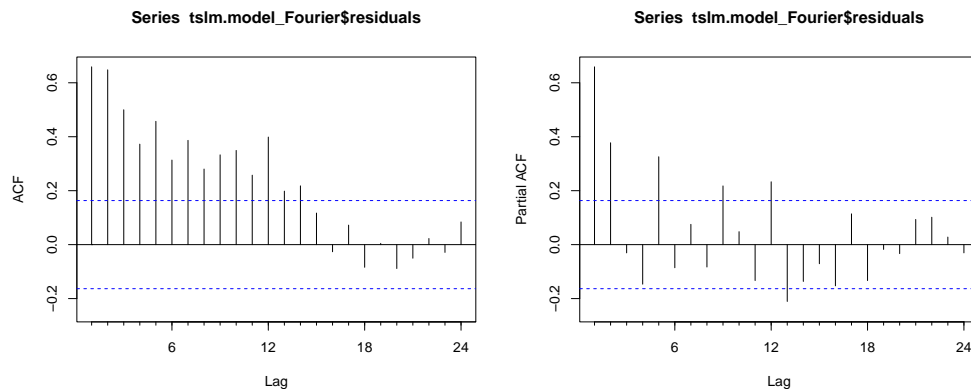
```
## fourier(log(AirPassengers), K = 4)C4-12 0.0221465 0.0071785 3.085 0.002472
##
## (Intercept) ***
## trend ***
## fourier(log(AirPassengers), K = 4)S1-12 ***
## fourier(log(AirPassengers), K = 4)C1-12 ***
## fourier(log(AirPassengers), K = 4)S2-12 ***
## fourier(log(AirPassengers), K = 4)C2-12 **
## fourier(log(AirPassengers), K = 4)S3-12
## fourier(log(AirPassengers), K = 4)C3-12 ***
## fourier(log(AirPassengers), K = 4)S4-12 ***
## fourier(log(AirPassengers), K = 4)C4-12 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0609 on 134 degrees of freedom
## Multiple R-squared: 0.9822, Adjusted R-squared: 0.981
## F-statistic: 819.9 on 9 and 134 DF, p-value: < 2.2e-16

plot(forecast(tslm.model_Fourier,
             data.frame(fourier(log(AirPassengers), K = 4, h = 24))))
```



Observamos que al aumentar K , modelizamos la componente estacional de forma más precisa. En cualquier caso, los errores asociados al modelo estimado no son un ruido blanco, sino que están correlados. Por tanto deberían modelizarse con ARIMA (regresión dinámica).

```
par(mfrow = c(1, 2))
Acf(tslm.model_Fourier$residuals)
Pacf(tslm.model_Fourier$residuals)
```



7.6) Modelos de regresión dinámica

Los modelos de regresión dinámica son una extensión de los modelos de regresión para series temporales permitiendo la inclusión de efectos autorregresivos y de media móvil en los residuos del modelo. Esto es especialmente útil cuando las observaciones sucesivas en una serie temporal están correlacionadas. Por tanto, estos modelos incluyen tanto predictores exógenos como componentes ARIMA para los errores, lo que permite capturar la estructura de correlación en los residuos y mejorar la precisión de las predicciones.

El modelo de regresión dinámica suele expresarse de la siguiente forma:

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \eta_t$$

donde los errores η_t no son un ruido blanco sino que se modelizan con un $ARIMA(p, d, q)$ mediante:

$$(1 - a_1 B - a_2 B^2 - \cdots - a_p B^p)(1 - B)^d \eta_t = (1 + b_1 B + b_2 B^2 + \cdots + b_q B^q) \varepsilon_t$$

con $\varepsilon_t \sim N(0, \sigma^2)$ independientes, $t \in \mathbb{T}$, es decir, ruido blanco. En realidad, los errores η_t pueden corresponder con un modelo ARIMA estacional (SARIMA) o no estacional (simplemente ARIMA).

Obsérvese que el modelo contempla dos fuentes de error: la perturbación η_t del modelo de regresión y el término de error ε_t del modelo ARIMA.

Para estimar los coeficientes del modelo de regresión dinámica se debe minimizar la suma de cuadrados de los errores ε_t , pues si se plantea la minimización de los errores η_t , la estimación de los coeficientes no sería apropiada y tampoco las inferencias sobre ellos. También se puede recurrir a la estimación por máxima verosimilitud.

Importante: Para estimar adecuadamente los coeficientes del modelo de regresión dinámica, es conveniente que todas las series que intervienen sean estacionarias (tanto la variable de interés como los predictores, a excepción de los debidos a efectos calendario).

En general, si alguna de las series no es estacionaria, se podrán diferenciar en cada una de ellas (no sólo en la que no es estacionaria), consiguiendo así que todas sean estacionarias y se mantenga el tipo de relación inicial.

Como es habitual en este contexto, tras el análisis descriptivo previo de la serie y la realización de las transformaciones necesarias, se procede a la **selección y validación del modelo**, etapas que se realizan con procedimientos similares a los vistos con anterioridad, incluyendo la consideración de predictores útiles como variables de tipo calendario o variables retardadas en el tiempo. Y finalmente, para las **predicciones**, se tienen en cuenta tanto la parte del modelo de regresión como la parte ARIMA de los errores. También en este caso se podrán obtener predicciones puntuales e intervalos de predicción.

Regresión dinámica con presencia de estacionalidad

Como vimos en los modelos de regresión múltiple para series, la estacionalidad puede modelizarse de dos formas: mediante el uso de variables dummy o incluyendo términos de Fourier, siendo preferible el uso de los términos de Fourier en el caso de valores altos de la periodicidad. Por otro lado, los modelos ETS y SARIMA están diseñados para periodicidades bajas, como máximo de 24.

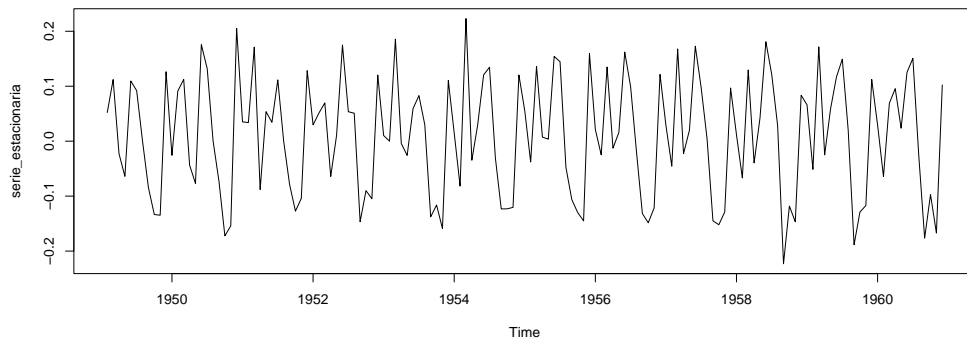
Para las series con periodicidad alta (superiores a 12 o 24) y para las series que presentan diferentes valores de periodicidad (por ejemplo, datos horarios con periodicidades de un día, $L_1 = 24$, y de una semana, $L_2 = 168$), se recomienda el uso de las series de Fourier para modelizar la componente estacional. Con el número de términos de Fourier a incluir en el modelo (parámetro denotado por K), se controla el suavizado de la componente estacional. Otra ventaja de usar los términos de Fourier para modelizar la estacionalidad de la serie es que, en general, se consigue que los errores η_t puedan modelarse con un modelo ARIMA no estacional, aunque no siempre se consigue.

Ejemplo

Seguiremos con el ejemplo de la serie AirPassengers de la sección anterior, donde vimos que tras realizar el ajuste del modelo de regresión múltiple, los errores estaban correlados.

Para el ajuste del modelo de regresión dinámica (ajuste conjunto del modelo de regresión y del modelo ARIMA), conviene que todas las series que intervienen (exceptuando las de tipo calendario) sean estacionarias. Por tanto, trabajaremos con la serie transformada:

```
serie_estacionaria <- diff(log(AirPassengers), 1)
plot(serie_estacionaria)
```



Con la función `auto.arima()` del paquete `forecast`, realizaremos la selección automática de los órdenes $(p, d, q)(P, D, Q)$ del modelo SARIMA para los errores, usando como predictores del modelo de regresión las variables exógenas "dummy estacionales".

```
estacional.dummy <- seasonaldummy(serie_estacionaria)

dyreg.model_auto <- auto.arima(serie_estacionaria,
                               xreg = as.matrix(estacional.dummy))

summary(dyreg.model_auto)
```

```
## Series: serie_estacionaria
## Regression with ARIMA(1,0,1)(1,0,0)[12] errors
##
## Coefficients:
##      ar1      ma1      sar1  intercept      Jan      Feb      Mar      Apr
##      0.6011 -0.8724  0.2290      0.1240 -0.0978 -0.1350  0.0144 -0.1422
## s.e.  0.2141  0.1537  0.0876      0.0132  0.0206  0.0196  0.0192  0.0191
##      May      Jun      Jul      Aug      Sep      Oct      Nov
##      -0.1175  0.0076 -0.0096 -0.1238 -0.2583 -0.2513 -0.2582
## s.e.  0.0189  0.0189  0.0189  0.0191  0.0193  0.0197  0.0203
##
## sigma^2 = 0.001306:  log likelihood = 279.36
## AIC=-526.72  AICc=-522.41  BIC=-479.32
##
## Training set error measures:
```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
## Training set	-7.885777e-05	0.03418937	0.02630177	NaN	Inf	0.7618447	-0.06587283

Observamos que, a pesar de haber usado variables dummy para modelizar la estacionalidad, los residuos requieren e un ARIMA estacional.

Veamos que ahora los residuos sí se pueden asumir independientes

```
par(mfrow = c(1, 2))
Acf(dyreg.model_auto$residuals)
```

```
Pacf(dyreg.model_auto$residuals)
```

