Estacionariedad

Gustavo A. Caffaro

Instituto Tecnológico de Santo Domingo

Nov 2021 - Ene 2022

Introducción

Contenido:

- Estacionariedad
- Consecuencias de la no estacionariedad
- Transformando una serie no estacionaria
- Diferenciación de una serie
- El operador rezago
- Suavizamiento

Estacionariedad estricta

Un proceso estocástico $\{x_t \mid t = 1, 2, \dots\}$ es **estríctamente estacionario** si, para cada colección de índices temporales $1 \leq t_1 \leq t_2 \leq \cdots \leq t_m$, la distibución conjunta de

$$(x_{t_1}, x_{t_2}, \dots, x_{t_m})$$

es idéntica a la de

$$(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_m+h}).$$

Por lo que

$$\Pr(x_{t_1} \leq c_1, \dots, x_{t_m} \leq c_k) = \Pr(x_{t_1+h} \leq c_1, \dots, x_{t_m+h} \leq c_k)$$

para todo $k \in \mathbb{N}$, todo $t_1, t_2, \dots, t_k$, todo $c_1, c_2, \dots, c_k$ y todo $h \in \mathbb{Z}$.

Estacionariedad estricta

Intuitivamente, la estacionariedad estricta nos dice que si tomamos una colección de términos en una serie, y nos desplazamos h términos, entonces la distribución de la colección original debe ser igual que la distribución de la colección desplazada.

Ej.: Si la serie x_t es estríctamente estacionaria y $h = 1$, entonces

$$\Pr(x_1 \leq a) = \Pr(x_2 \leq a) = \Pr(x_3 \leq a) = \cdots$$

$$\Pr(x_1 \leq a_1, x_{h+1} \leq a_{h+1}) = \Pr(x_t \leq a_t, x_{t+h} \leq a_{t+h})$$

Estacionariedad débil

Una serie es **(débilmente) estacionaria** sí y sólo sí

1. Tiene media constante e independiente del tiempo.
2. La función de autocovarianza, $\gamma(s, t)$ sólo depende de la diferencia entre s y t, $h = |s - t|$.

Una conclusión directa de la definición de estacionariedad débil es que $\gamma(s, t) = \gamma(s + j, t + j)$ para cualquier entero j.

En esta clase utilizaremos el término *estacionario* para referirnos a esta definición.

Autocovarianza y autocorrelación (series estacionarias)

La **función de autocovarianza de una serie estacionaria** viene dada por

$$\gamma(h) = cov(x_{t+h}, x_t) = E[(x_{t+h} - \mu)(x_t - \mu)].$$

De manera similar, la **función de autocorrelación de una serie estacionaria** viene dada por

$$\rho(h) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t+h,t+h)\gamma(t,t)}} = \frac{\gamma(h)}{\gamma(0)}$$

La desigualdad de Cauchy-Schwarz muestra que $-1 \leq \rho(h) \leq 1$ para todo h.

Autocovarianza y autocorrelación (RBG)

Ej.: Ruido Blanco Gaussiano (RBG)

Si $w_t \sim N(0, \sigma_w^2)$, entonces

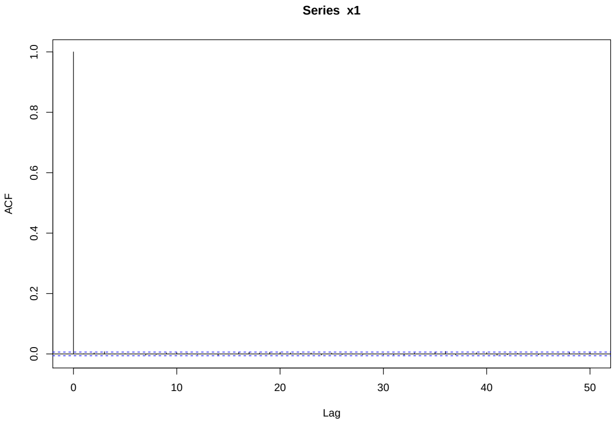
$$\gamma_w(h) = cov(w_{t+h}, w_t) = \begin{cases} \sigma_w^2, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

Como $\gamma_w(0) = \sigma_w^2$, entonces $\rho(h) = 1$ si $h = 0$ y $\rho(h) = 0$ si $h \neq 1$.

Adicionalmente, como $\mu_{wt} = \mu_w = 0$, entonces el ruido blanco gaussiano cumple con la definición de estacionariedad débil (también con la definición estricta!).

FAC (RBG)

```r
n <- 1e3
set.seed(n)
x1 <- rnorm(1e5)
acf(x1)
```



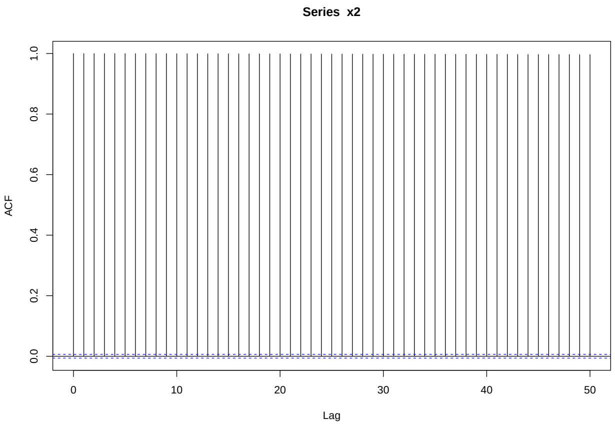
Autocovarianza y autocorrelación (RW)

Ej.: Caminata Aleatoria (RW)

Calculemos la función de covarianza, la f. de correlación y determinemos si es estacionaria.

FAC (RW)

```r
x2 <- cumsum(x1)
acf(x2)
```



Autocovarianza y autocorrelación (media móvil)

Ej.: Media móvil

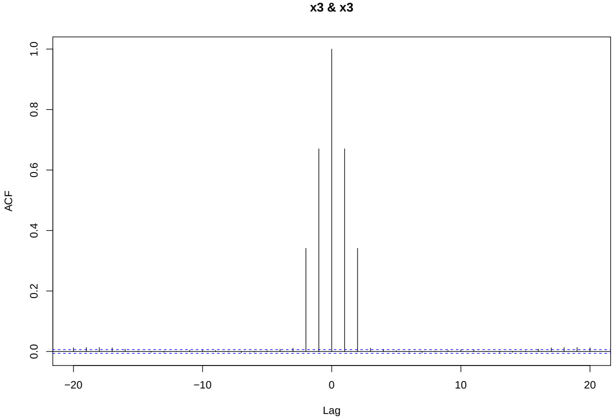
Recuerda que la autocovarianza de la media móvil de tres puntos viene dada por

$$\gamma_v(h) = \begin{cases} \frac{3}{9}\sigma_w^2, & h = 0 \\ \frac{2}{9}\sigma_w^2, & h = \pm 1 \\ \frac{1}{9}\sigma_w^2, & h = \pm 2 \\ 0, & |h| > 2. \end{cases}$$

Como la media $\mu_{vt} = \mu_v = 0$ y autocovarianza no dependen de t sino del número de pasos h, satisfacen la definición de débilmente estacionaria.

FAC (media móvil)

```
library(forecast)
x3 <- ma(x1,order=3,centre=T)
x3 <- x3[complete.cases(x3)] # ignorar NAs prod. por media movil
ccf(x3,x3,lag.max = 20) # funcion de correlacion cruzada
```



Propiedades de la función de autocovarianza

Si la serie x_t es estacionaria, entonces

1. $\gamma(h) = \gamma(-h)$
2. $|\gamma(h)| \leq \gamma(0)$

Prueba 1

$$\gamma(h) = cov(x_{t+h}, x_t) = cov(x_t, x_{t-h}) = \gamma(-h)$$

Prueba 2

Cuando $h = 0$, tenemos que $\gamma(0) = E[(x_t - \mu)^2]$.

La desigualdad de Cauchy-Schwarz implica que

$$|\gamma(h)| \leq \gamma(0)$$

Estacionariedad tendencial

Definamos una serie $x_t = \alpha + \beta t + y_t$, donde y_t es estacionaria. Entonces

$$\mu_{xt} = E(x_t) = \alpha + \beta t + \mu_y,$$

la cual no es independiente del tiempo. Por lo tanto, x_t no es estacionaria.

Sin embargo, la función de autocovarianza es independiente del tiempo, porque

$$\begin{aligned} \gamma_x(h) &= E[(x_{t+h} - \mu_{x,t+h})(x_t - \mu_{x,t})] \\ &= E[(y_{t+h} - \mu_y)(y_t - \mu_y)] \\ &= \gamma(h). \end{aligned}$$

Por lo tanto, se considera que el modeo tiene estacionariedad al rededor de su tendencia. A este comportamiento le llamamos *estacionariedad tendencial*.

Consecuencias de la no estacionariedad

Si una serie no es estacionaria

- Los choques no “mueren” (persisten para siempre)
- Las distribuciones de las pruebas estadísticas son no normales
- Sesgo en los estimadores
- Pobres proyecciones

Persistencia de los choques

Consideremos un modelo AR(1)

$$y_t = by_{t-1} + w_t,$$

el cual, si resolvemos de manera recursiva, obtenemos

$$y_t = b^t y_0 + w_t + bw_{t-1} + b^2 w_{t-2} + \cdots + b^{t-2} w_2 + b^{t-1} w_1.$$

Observa que los choques dependen en el valor de b.

Persistencia de los choques

$$y_t = b^t y_0 + w_t + bw_{t-1} + b^2 w_{t-2} + \cdots + b^{t-2} w_2 + b^{t-1} w_1.$$

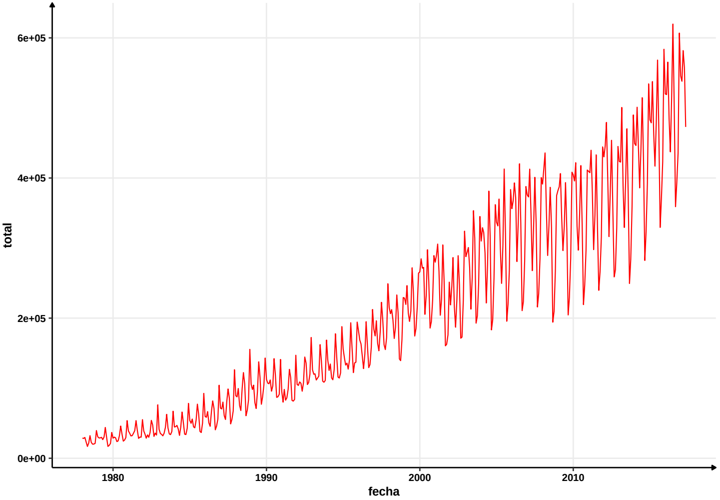
Tres casos:

1. $|b| < 1$: $b^t \to 0$ cuando $t \to \infty$, por lo que el efecto de un choque disminuye conforme el tiempo pasa.
2. $|b| = 1$: $b^t = 1, \forall t$, por lo que el efecto es persistente y su varianza crece con el tiempo (¿a qué proceso les recuerda esto?).
3. $b > 1$: los choques se vuelven más importantes con el tiempo.

Sesgos, pruebas estadísticas no normales, y malas proyecciones

- Si una serie no es estacionaria, los estimadores de los coeficientes podrían estar sesgados. Esto nos llevaría a estimaciones de relaciones **espurias** (Granger y Newbold). Un ejemplo evidente lo obtenemos asumiendo que la llegada de turistas a RD es una función del tiempo.

Llegada de turistas



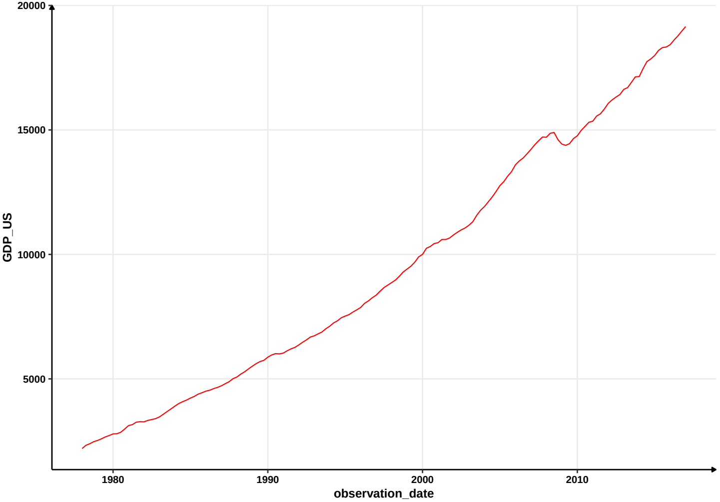
Llegada de turistas

```
x <- read.csv("../data/llegada_turistas.csv")
x$t <- 1:nrow(x)
modelo1 <- lm(total ~ t, data=x)
summary(modelo1)
```

```
##
## Call:
## lm(formula = total ~ t, data = x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -150761  -32982    -663   34586  185228
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -32647.97    4821.59  -6.771 3.81e-11 ***
## t             1008.95      17.63  57.236  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52350 on 471 degrees of freedom
## Multiple R-squared:  0.8743, Adjusted R-squared:  0.874
## F-statistic:  3276 on 1 and 471 DF,  p-value: < 2.2e-16
```

¿Observen el coeficiente de t y su significancia, pero ¿creen que el paso del tiempo es el motor de la llegada de turistas a RD?

PIB EEUU vs llegada de turistas



PIB EEUU vs llegada de turistas

```
##
## Call:
## lm(formula = GDP_US ~ turistas_rd, data = gdp_us)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2536.99  -451.37     7.65   445.72  2380.99
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.309e+03  1.332e+02   17.33   <2e-16 ***
## turistas_rd 1.163e-02  1.784e-04   65.16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 944.3 on 155 degrees of freedom
## Multiple R-squared:  0.9648, Adjusted R-squared:  0.9645
## F-statistic:  4245 on 1 and 155 DF,  p-value: < 2.2e-16
```

¿Será que el PIB de EEUU depende de los turistas que llegan a RD?

Sesgos, pruebas estadísticas no normales, y malas proyecciones

- Si una serie no es estacionaria, los estimadores de los coeficientes podrían estar sesgados. Esto nos llevaría a estimaciones de relaciones **espurias** (Granger y Newbold). Un ejemplo evidente lo obtenemos asumiendo que la llegada de turistas a RD es una función del tiempo.
- La inferencia que obtenemos de los resultados de la regresión (ej. los intervalos de confianza) no son válidos.
- Usualmente un alto R^2
- Con coeficientes sesgados tenemos proyecciones deficientes.

Pruebas estadísticas de estacionariedad

Hemos visto como, la mayoría de las veces, sólo observando la data podemos tener una buena idea de si la serie es estacionaria o no.

Sin embargo, existen algunas pruebas estadísticas que también nos ayudan a determinar la estacionariedad de una serie.

Veremos estas pruebas en la siguiente clase, cuando introduzcamos formalmente los modelos AR, MA, y ARIMA.

De no estacionariedad a estacionariedad

Vimos cómo los procesos no estacionarios no son deseables.

Sin embargo, la gran mayoría de las series que nos enfrentamos en la vida real no son estacionarias.

Dicho esto, ¿qué podemos hacer con las series no estacionarias?

¿Será que podemos transformar un proceso no estacionario a uno estacionario?

Comencemos viendo series que violan la no estacionariedad por tener **medias no constantes**.

Tipos de tendencias

- Tendencia determinística (estacionariedad alrededor de la tendencia)
- Tendencia estocástica

Tendencia determinística

Recordemos que un proceso es estacionario alrededor de su tendencia si exhibe estacionariedad al rededor de dicha tendencia. Podríamos escribir este tipo de modelos como

$$x_t = \mu_t + y_t,$$

donde x_t son las observaciones, μ_t la tendencia, y y_t es un proceso estacionario.

Tendencia determinística

Algunos tipos de tendencia determinística:

- Tendencia lineal: $\mu_t = a + bt$

$$x_t = a + bt + y_t$$

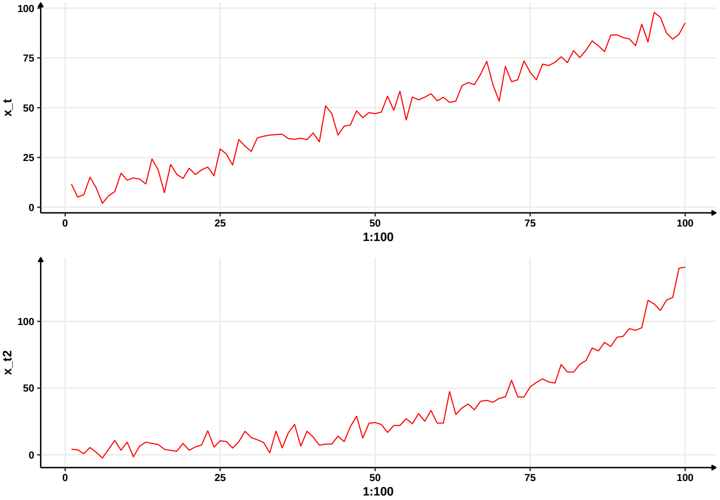
- Tendencia exponencial: $\mu_t = a + be^t$

$$x_t = a + be^t + y_t$$

En este último caso, podríamos linealizar la serie tomando $ln(x_t)$.

Tres de las técnicas más utilizadas para eliminar esta tendencia son el *detrending*, la *diferenciación*, y el *filtrado/suavizamiento*.

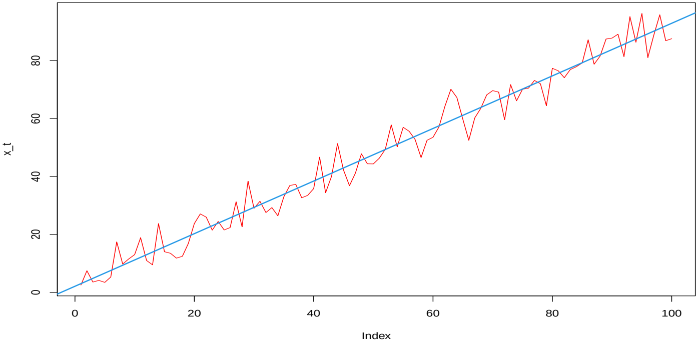
Tendencia determinística



Detrending

El *detrending* consiste en estimar la tendencia μ_t con algún método paramétrico (ej. OLS) y luego trabajar con los residuales:

$$\hat{y}_t = x_t - \hat{\mu}_t.$$



Diferenciación

La *diferenciación* consiste en restarle a cada valor de la serie su valor anterior:

$$x_t - x_{t-1} = (\mu_t + y_t) - (\mu_{t-1} + y_{t-1}) = b + y_t - y_{t-1}$$

La diferenciación juega un rol importantísimo en las series de tiempo, por lo que recibe su propia notación. La primera diferencia de la serie x_t se escribe:

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

Detrending vs. diferenciar la serie

- Una ventaja de diferenciar la serie sobre el detrending es que no se requiere estimar ningún parámetro.
- Una desventaja es que no tenemos estimaciones del componente estocástico y_t. Si quisieramos tener estimaciones de y_t, entonces es mejor estimar la tendencia de la serie.
    - Ej. ciclo del IMAE
- Otra ventaja de diferenciar la serie sobre el detrending es que no necesitamos conocer la especificación exacta de la tendencia, por lo que si el objetivo principal es convertir la data a estacionaria, diferenciar la serie no sólo es más eficiente sino que también nos evita problema de pobre especificación del modelo. En mi opinión este es el argumento con más peso.

El operador de rezago/retardo

El **operador de rezago** se define como

$$Lx_t = x_{t-1},$$

y si lo extendemos a las potencias:

$$L^2 x_t = L(Lx_t) = Lx_{t-1} = x_{t-2},$$

$$L^k x_t = x_{t-k}.$$

De igual forma, podríamos definir un operador adelantado si $L^{-1}L = 1$:

$$x_t = L^{-1}Lx_t = L^{-1}x_{t-1}$$

El operador de rezago/retardo

Podríamos reescribir la operación de diferenciación en términos del operador de rezago:

$$\nabla x_t = x_t - x_{t-1} = (1 - L)x_t.$$

Podríamos extender esta idea a diferencias de mayor orden. Por ejemplo, la segunda diferencia:

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

Diferencias de orden d

La **diferencia de orden** d se define como:

$$\nabla^d = (1 - L)^d,$$

donde podríamos expandir el operador $(1 - L)^2$ de manera algebráica para evaluar mayores valores de d.

Suavizamiento por media móvil

Anteriormente vimos cómo una media móvil centrada podía ser utilizada para suavizar el ruido blanco.

Este tipo de métodos es útil para descubrir tendencias y componentes estacionales.

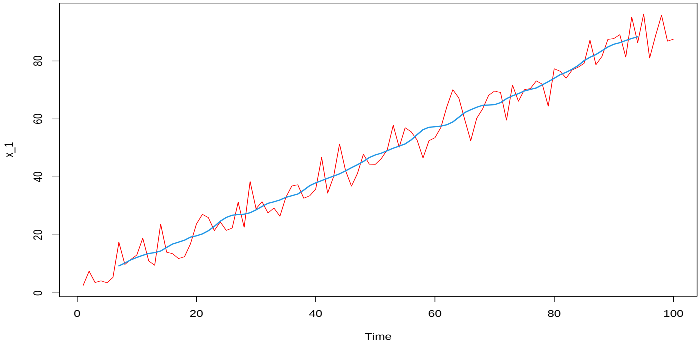
Formalmente, si x_t representan las observaciones de una serie, tenemos que

$$m_t = \sum_{j=-k}^{k} a_j x_{t-j},$$

donde $a_j = a_{-j} \geq 0$ y $\sum_{j=-k}^{k} a_j = 1$, es una media móvil centrada de dichas observaciones.

Suavizamiento por media móvil

```r
x_1 <- as.ts(x_t)
w <- c(.5, rep(1,11), .5)/12 # pesos (weights), 12 meses
x_ma <- stats::filter(x_1, sides=2, filter=w)
plot(x_1, col="red")
lines(x_ma, lwd=2, col=4)
```



Suavizamiento por kernel

El **suavizamiento por kernel** es un tipo de suavizamiento de media móvil que utiliza una función de peso llamada kernel, y se describe por

$$m_t = \sum_{i=1}^{n} w_i(t) x_i,$$

donde

$$w_i(t) = \frac{K\left(\frac{t-i}{b}\right)}{\sum_{j=1}^{n} K\left(\frac{t-j}{b}\right)},$$

son los pesos y $K(\cdot)$ es una función kernel.

El parámetro b representa el ancho de banda (bandwidth), y mientras más ancha la banda, más "suave" el suavizado.

¿Por qué?

Suavizamiento por kernel

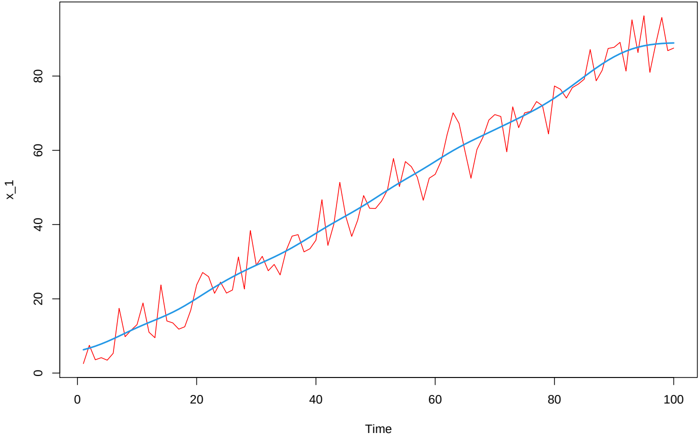
Usualmente se utiliza un kernel normal

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-z^2/2\right)$$

Este método de suavizamiento le asigna pesos elevados a observaciones cercanas al punto estimado, y van decayendo conforme se van alejando.

Suavizamiento por kernel

```
x_1_ks <- ksmooth(time(x_1), x_1, "normal", bandwidth=12) # extraer el suavizamiento por kernel:
plot(x_1, col="red")
lines(x_1_ks, lwd=2, col=4)
```



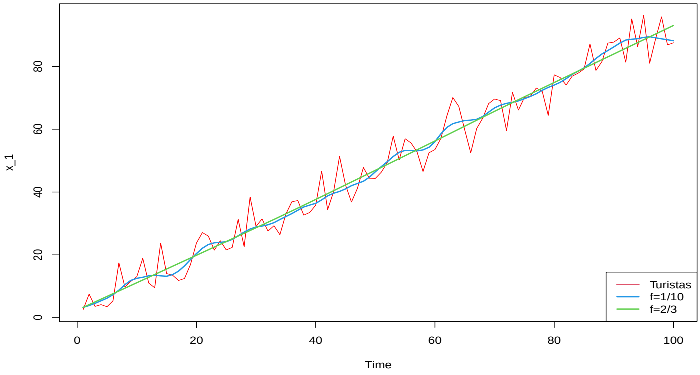
Lowess

Otra técnica de suavizamiento es correr una regresión con los k puntos más cercanos a cada observación (k-nearest neighbors). El método es intuitivo. Para cada x_t:

- Correr una regresión con los datos $\{x_{t-k/2}, \ldots, x_t, \ldots, x_{t+k/2}\}$ para predecir x_t: $m_t = \hat{x}_t$
- Formalmente, intentamos encontrar los parámetros α_i, β_i que minimizan
$$\sum_{i=1}^{T} w_i(t) \cdot (x_t - \alpha_i - \beta_i t)^2$$
- La función de peso $w_i(t)$ predeterminada en el programa R es el *tricubo*.

Lowess

```
# extraer el suavizamiento por lowess:
x_1_lowess1 <- lowess(x_1, f=1/10) # poco suavizamiento
x_1_lowess2 <- lowess(x_1, f=2/3) # mas suavizamiento
plot(x_1, col="red")
lines(x_1_lowess1, lwd=2, col=4)
lines(x_1_lowess2, lwd=2, col=3)
legend(x = "bottomright", legend=c("Turistas", "f=1/10","f=2/3"),
       col=c(2,4,3),lwd=2)
```



Regresión polinomial

Finalmente, una manera relativamente sencilla de suavizar una serie es mediante una regresión polinomial. La regresión polinomial de orden p viene dada por

$$x_t = \beta_0 + \sum_{j=1}^{p} \beta_j t^j + w_t,$$

donde los β_j se podrían estimar por mínimo cuadrados.

Ej.: regresión polinomial de 2do orden:

$$x_t = \beta_0 + \beta_1 t + \beta_2 t^2 + w_t$$

Extrayendo los errores