Análisis de Series de Tiempo

Carrera de Especialización en Inteligencia Artificial

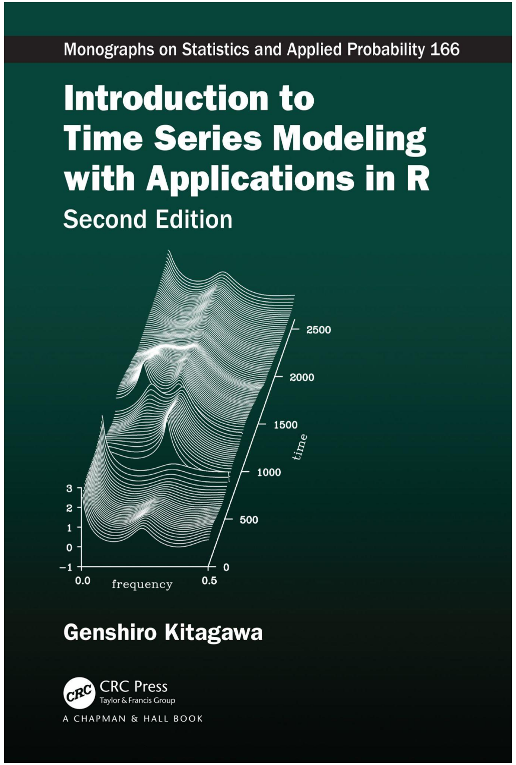
Clase 1

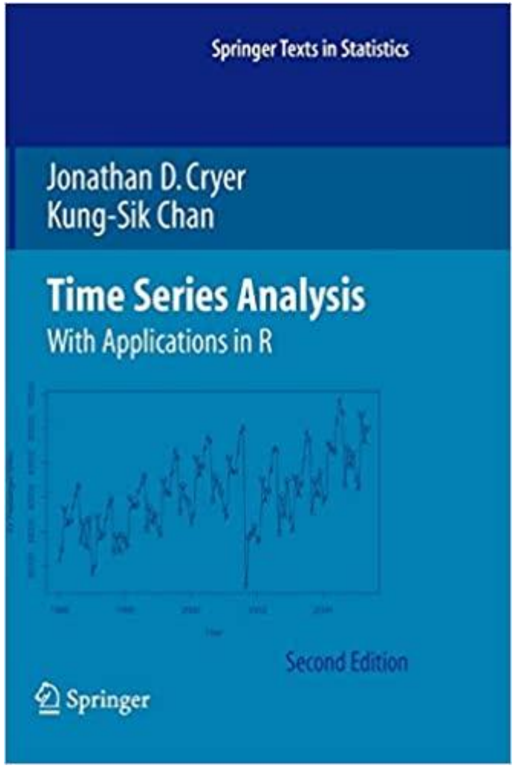
Ing. Magdalena Bouza, Ing. Carlos German Carreño Romano

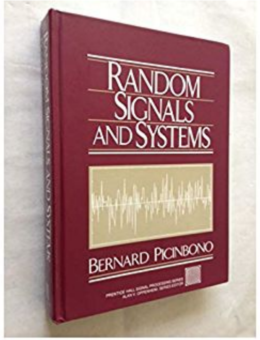
Cronograma

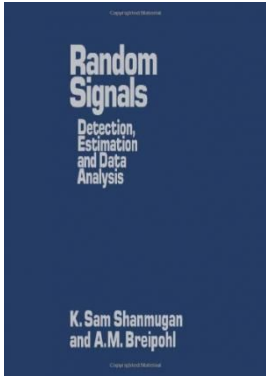
Clase 1	Introducción, nociones básicas de modelado
Clase 2	Modelos estacionarios
Clase 3	Aplicaciones, presentación de TP
Clase 4	Tendencia y estacionalidad
Clase 5	Predicciones
Clase 6	Aplicaciones. Análisis espectral
Clase 7	Heterocedasticidad. Markov.
Clase 8	Evaluación final

Bibliografía



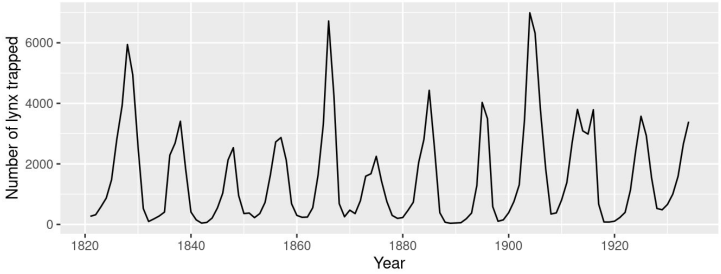


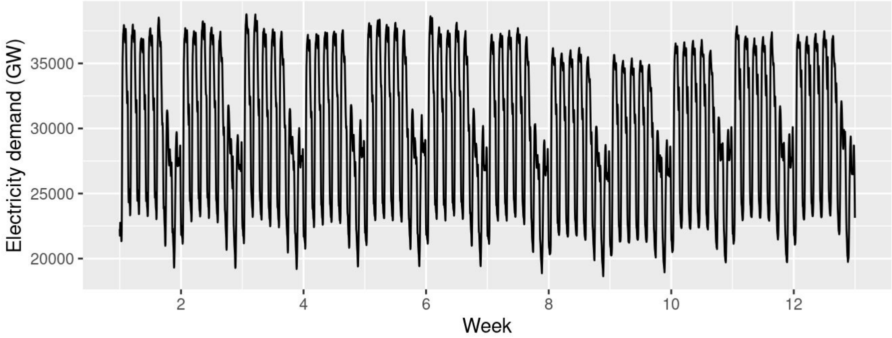


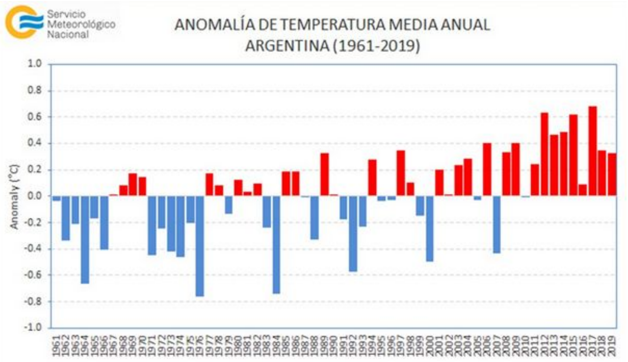


¿Qué es una serie de tiempo?

Un registro de un fenómeno que varía en el tiempo de forma irregular es una serie de tiempo.







¿Para qué estudiar Series de Tiempo?

En general no podemos asumir que la **observación** surge independientemente de una población común. Estudiar modelos que incorporen **dependencia** es la clave en TSA.

- **Descripción**: se desea resumir las características de una serie de tiempo.
- **Modelado**: se busca identificar un modelo aproximado del problema. Es necesario identificar un modelo correcto, así como los parámetros asociados al mismo.
- **Predicción**: el objetivo es conocer, o predecir, el comportamiento futuro del sistema.
- **Extracción de señales**: de todo el modelo, se busca extraer ciertas señales que resultan de interés para el problema en cuestión

Clasificación de Series de Tiempo

- **Continuas o discretas**: En general se consideran series de tiempo discretas, ya que los datos suelen medirse en intervalos de tiempo
- Univariadas o **multivariadas**:
- **Estacionarias** o **no estacionarias**: cuando la serie de tiempo estudiada proviene de realizaciones de un proceso estocástico con una estructura invariante en el tiempo se lo llama estacionario.
- **Gaussianas** o no gaussianas: Si la serie de tiempo sigue una distribución Gaussiana o no,
- **Lineales** o no lineales: se dice que una serie de tiempos es lineal si puede modelarse como la salida de un sistema lineal
- Observaciones faltantes y outliers:

Series de tiempo y Procesos estocásticos

Procesos estocásticos

Las series de tiempo forman parte de lo que se conoce como procesos estocásticos.
Así como las variables aleatorias mapean los posibles resultados de un experimento aleatorio a un número (real), los procesos estocásticos mapean los resultados de un experimento aleatorio a un conjunto de funciones en el tiempo.

- recordar media, var, cov, autocorrelacion, PSD
- definir estacionariedad

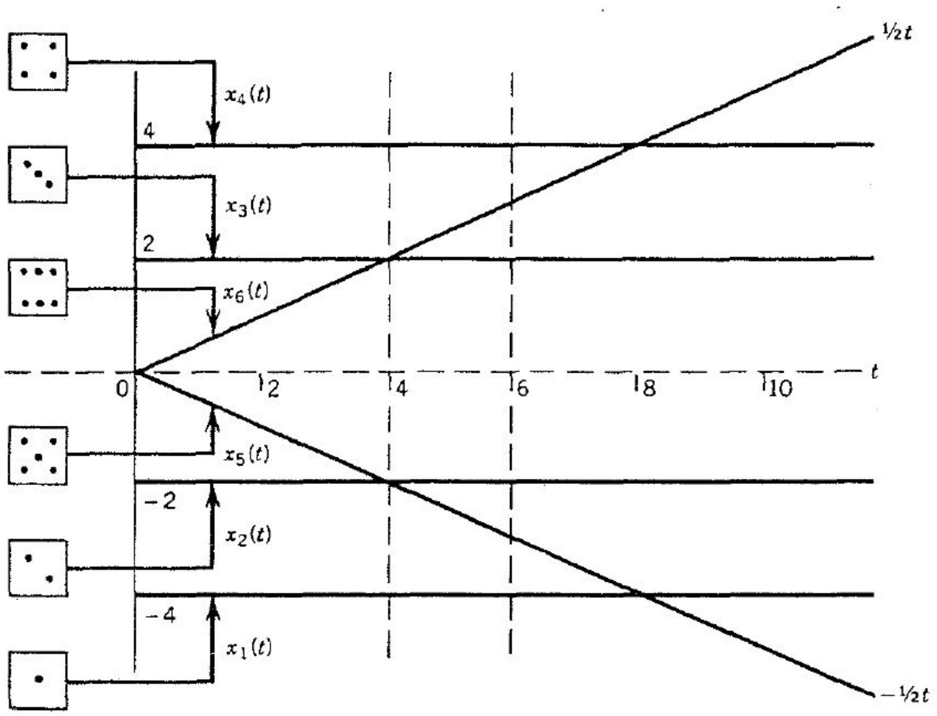


Figure 3.2 Example of a random process.

Ref.: Random Signals: Detection, Estimation and Data Analysis, *Shanmugan, Breipohl*

Algunos momentos

Una serie de tiempo $Y(t)$, puede entonces modelarse a través de la secuencia de variables $\{Y(t), t = t_k, t_{k+1}, \ldots, t_{k+n}\}$ que definen un proceso estocástico. Gran parte de la información de la distribución de estos procesos está contenida en los momentos de dicha distribución: medias, varianzas y covarianzas. A diferencia de las variables aleatorias, estos momentos deberán definirse como función del tiempo.

- **Media** (o esperanza): $\mu_t = \mathbb{E}[Y_t]$
- **Función de autocovarianza**: $C_{t,s} = Cov(Y_t, Y_s) = \mathbb{E}\left[(Y_t - \mu_t)(Y_s - \mu_s)\right]$
- **Función de autocorrelación**: $R_{t,s} = Corr(Y_t, Y_s) = \frac{Cov(Y_t, Y_s)}{\sqrt{var(Y_t)var(Y_s)}}$

Algunos momentos (caso multivariado)

Análogamente, si se tiene una serie de tiempo multivariada, donde $Y_t = [Y_t^{(1)}, \ldots, Y_t^{(l)}]$, definimos

- **Media** (o esperanza): $\boldsymbol{\mu}_t = [\mu_t^{(1)}, \ldots, \mu_t^{(l)}] = [\mathbb{E}[Y_t^{(1)}], \ldots, \mathbb{E}[Y_t^{(l)}]]$
- **Matriz de cross-covarianza**: $C_{t,s} = \begin{bmatrix} C_{t,s}^{(1,1)} & \ldots & C_{t,s}^{(1,l)} \\ \vdots & \ddots & \vdots \\ C_{t,s}^{(l,1)} & \ldots & C_{t,s}^{(l,l)} \end{bmatrix}$, donde

$$C_{t,s}^{(i,j)} = Cov(Y_t^{(i)}, Y_s^{(j)}) = \mathbb{E}[(Y_t^{(i)} - \mu_t^{(i)})(Y_s^{(j)} - \mu_s^{(j)})]$$

- **Matriz de cross-correlación**: $R_{t,s} = \begin{bmatrix} R_{t,s}^{(1,1)} & \ldots & R_{t,s}^{(1,l)} \\ \vdots & \ddots & \vdots \\ R_{t,s}^{(l,1)} & \ldots & R_{t,s}^{(l,l)} \end{bmatrix}$

Estacionariedad

Una suposición que suele hacerse a la hora de modelar series de tiempo es la estacionariedad.

Se dice que un proceso es **estacionario** (*stationary*) si la distribución del proceso no cambia a lo largo del tiempo. Matemáticamente, quiere decir que

$$f_{Y_{t_1},\ldots,Y_{t_n}} = f_{Y_{t_{1+\delta}},\ldots,Y_{t_{n+\delta}}}, \quad \forall t_1,\ldots t_n, \forall\delta$$

Una conclusión que se desprende de esta definición es que su los procesos son estacionarios, su media es constante en el tiempo y las funciones de autocorrelación y autocovarianza dependen sólo de la diferencia de tiempos:

$$\mu_t = \mu_s = \mu \quad \forall t, s$$

$$C_{t,s} = Cov(Y_t, Y_s) = Cov(Y_{t+\delta}, Y_{s+\delta}) = C_{s-t}, \quad \forall\, s, t, \delta$$

$$R_{t,s} = Corr(Y_t, Y_s) = Corr(Y_{t+\delta}, Y_{s+\delta}) = R_{s-t}, \quad \forall\, s, t, \delta$$

Estacionariedad débil

En muchos casos, si bien no se cumple el supuesto de estacionariedad, vale lo que se conoce como estacionariedad débil.

Se dice que un proceso es **débilmente estacionario (DE)** si sólo se cumple

$$\mu_t = \mu_s = \mu \quad \forall t, s$$

$$C_{t,s} = Cov(Y_t, Y_s) = Cov(Y_{t+\delta}, Y_{s+\delta}) = C_{s-t}, \quad \forall\, s, t, \delta$$

$$R_{t,s} = Corr(Y_t, Y_s) = Corr(Y_{t+\delta}, Y_{s+\delta}) = R_{s-t}, \quad \forall\, s, t, \delta$$

Es decir que sólo se pide que sean estacionarios los momentos hasta de segundo orden.

En general cuando hablemos de series de tiempo estacionarias nos estaremos refiriendo a este tipo de estacionariedad.

Estimación de momentos

Dadas N muestras de una serie de tiempo DE $\{y_1, \ldots, y_N\}$, entonces podemos estimar los momentos como:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\hat{C}_k = \frac{1}{N}\sum_{n=k+1}^{N} (y_n - \hat{\mu})(y_{n-k} - \hat{\mu})$$

$$\hat{R}_k = \frac{\hat{C}_k}{\hat{C}_0}$$

Preprocesamiento

¿ Cuándo aplicar un preprocesamiento?

Cuando la distribución de la serie de tiempo cambia a lo largo del tiempo se dice que es **no estacionaria**.
En estos casos, pueden aplicarse transformaciones para que la serie de tiempo resultante sea aproximadamente estacionaria.

Algunos métodos:
1. Transformación de variables
2. Diferenciación
3. Promedio móvil

1. Transformación de variables

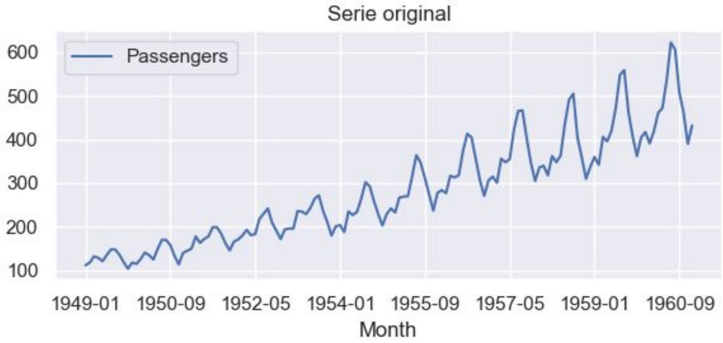
Muchas series de tiempo tienen la característica de que sus varianzas van aumentando a medida que avanza el tiempo.

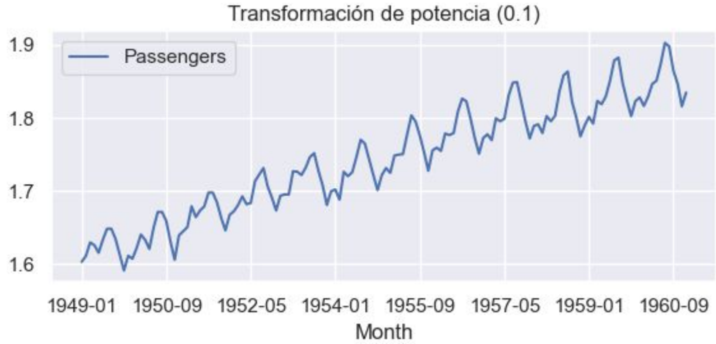
Se pueden aplicar transformaciones a los puntos de la serie de tiempo.

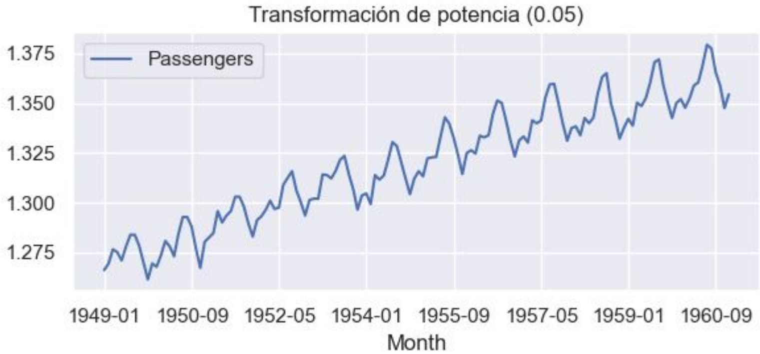
- Transformación Box-Cox

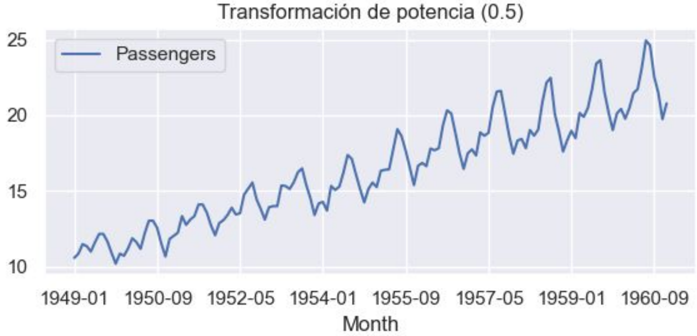
$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^{\lambda}-1}{\lambda} & si\ \lambda \neq 0, \\ \ln y_i & si\ \lambda = 0 \end{cases}$$

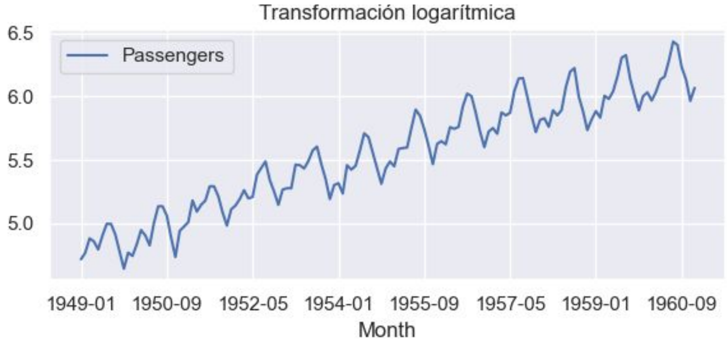
1. Ejemplos


Serie original
Passengers
600
500
400
300
200
100
1949-01
1950-09
1952-05
1954-01
1955-09
1957-05
1959-01
1960-09
Month


Transformación de potencia (0.1)
Passengers
1.9
1.8
1.7
1.6
1949-01
1950-09
1952-05
1954-01
1955-09
1957-05
1959-01
1960-09
Month


Transformación de potencia (0.05)
Passengers
1.375
1.350
1.325
1.300
1.275
1949-01
1950-09
1952-05
1954-01
1955-09
1957-05
1959-01
1960-09
Month


Transformación de potencia (0.5)
Passengers
25
20
15
10
1949-01
1950-09
1952-05
1954-01
1955-09
1957-05
1959-01
1960-09
Month


Transformación logarítmica
Passengers
6.5
6.0
5.5
5.0
1949-01
1950-09
1952-05
1954-01
1955-09
1957-05
1959-01
1960-09
Month

2. Diferenciación

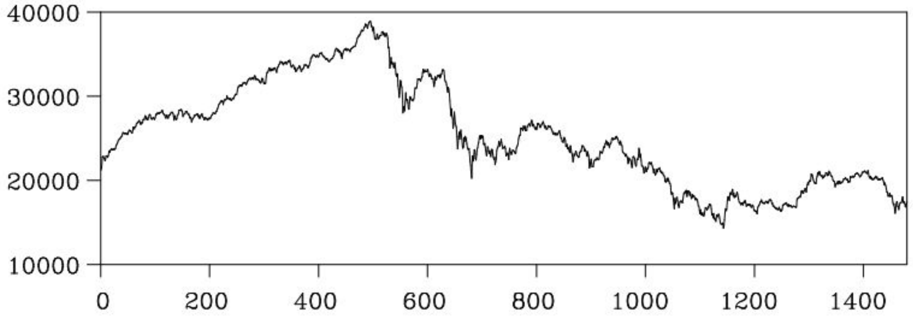
Si la serie presenta un tendencia podemos analizar en su lugar la serie diferenciada. Dada una serie de tiempo y_n, $n = 0, 1, \ldots$ definimos

$$z_n = \Delta y_n = y_n - y_{n-1}$$

Motivación: si $y_n = an + b$, al diferenciar obtendríamos una constante.

Observación: En general, si y_n se corresponde con un polinomio de grado n, diferenciando n veces recuperamos una constante.

Ejemplos



(g) Stock price index data

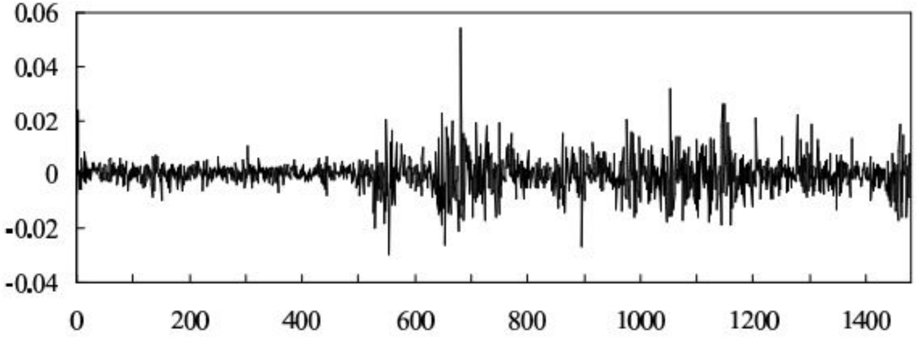


Figure 1.2: *Difference of the logarithm of the* Nikkei 225 *data.*

3. Promedio móvil

Para una serie de tiempo y_n, el promedio móvil de $(2k+1)$ términos está dado por

$$T_n = \frac{1}{2k+1} \sum_{j=-k}^{k} y_{n+j}$$

Si modificamos la definición intercambiando la media por la mediana obtenemos la mediana móvil, definida como

$$T_n = \text{mediana}\{y_{-k}, \ldots, y_k\}$$

En general, la mediana móvil puede capturar cambios en la tendencia más rápido que el promedio móvil.

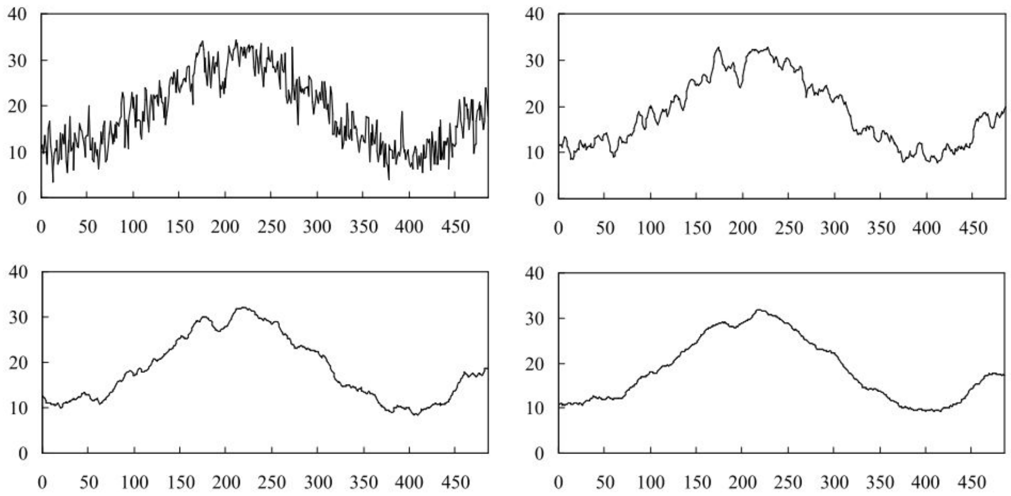


Figure 1.4 *Maximum temperature data and its moving average. Top left: original data, top right: moving average with $k = 5$, bottom left: $k = 17$, bottom right: $k = 29$.*

Ref.: Introduction to Time Series Modeling, *Genshiro Kiagawa*