Time-Series Analysis: Stochastic Models

Evgeny Burnaev, Alexey Zaytsev

Skoltech, Moscow, Russia

All models are wrong, but some are useful

- We'll consider some models for time series.
- They have different ideas behind them and can be arbitrary complex.
- But they are able to provide reasonable predictions and are interpretable.

Outline

1. Auto Regression (AR)
2. Moving average
3. Autocovariance function
4. Auto Regression (ARMA) process

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

- If $\varphi = 1$, we get nonstationary Random walk process.

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

- Can we make y_t stationary (i.e. $|\varphi| < 1$)?

$$\mathbb{E}[y_t] = \varphi \mathbb{E}[y_{t-1}] = 0 \ \ \text{(from stationarity)}$$

$$\mathbb{E}[y_t^2] = \varphi^2 \mathbb{E}[y_{t-1}^2] + \sigma^2 = \frac{\sigma^2}{1 - \varphi^2} \ \ \text{(from stationarity)}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

- Can we make y_t stationary (i.e. $|\varphi| < 1$)?

$$\mathbb{E}[y_t] = \varphi \mathbb{E}[y_{t-1}] = 0 \ \ \text{(from stationarity)}$$

$$\mathbb{E}[y_t^2] = \varphi^2 \mathbb{E}[y_{t-1}^2] + \sigma^2 = \frac{\sigma^2}{1 - \varphi^2} \ \ \text{(from stationarity)}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

- Can we make y_t stationary (i.e. $|\varphi| < 1$)?

$$\mathbb{E}[y_t] = \varphi \mathbb{E}[y_{t-1}] = 0 \ \text{(from stationarity)}$$

$$\mathbb{E}[y_t^2] = \varphi^2 \mathbb{E}[y_{t-1}^2] + \sigma^2 = \frac{\sigma^2}{1 - \varphi^2} \ \text{(from stationarity)}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

- Can we make y_t stationary (i.e. $|\varphi| < 1$)?

$$\mathbb{E}[y_t] = \varphi \mathbb{E}[y_{t-1}] = 0 \ \text{(from stationarity)}$$

$$\mathbb{E}[y_t^2] = \varphi^2 \mathbb{E}[y_{t-1}^2] + \sigma^2 = \frac{\sigma^2}{1 - \varphi^2} \ \text{(from stationarity)}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t,\ \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\mathbb{E}[y_t] = 0,\ \ \mathbb{E}[y_t^2] = \frac{\sigma^2}{1 - \varphi^2}$$

$$\begin{aligned}
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|}\gamma_y(0)\ \ \text{(check for } h > 0 \text{ and } h < 0\text{)} \\
&= \frac{\varphi^{|h|}\sigma^2}{1 - \varphi^2}
\end{aligned}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\mathbb{E}[y_t] = 0, \ \ \mathbb{E}[y_t^2] = \frac{\sigma^2}{1 - \varphi^2}$$

$$\begin{aligned}
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|} \gamma_y(0) \ \ \text{(check for } h > 0 \text{ and } h < 0) \\
&= \frac{\varphi^{|h|} \sigma^2}{1 - \varphi^2}
\end{aligned}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\mathbb{E}[y_t] = 0, \ \mathbb{E}[y_t^2] = \frac{\sigma^2}{1 - \varphi^2}$$

$$\begin{aligned}
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|} \gamma_y(0) \ \ \text{(check for } h > 0 \text{ and } h < 0\text{)} \\
&= \frac{\varphi^{|h|} \sigma^2}{1 - \varphi^2}
\end{aligned}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \;\; \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\mathbb{E}[y_t] = 0, \;\; \mathbb{E}[y_t^2] = \frac{\sigma^2}{1 - \varphi^2}$$

$$\begin{aligned}
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|} \gamma_y(0) \;\; \text{(check for } h > 0 \text{ and } h < 0\text{)} \\
&= \frac{\varphi^{|h|} \sigma^2}{1 - \varphi^2}
\end{aligned}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\begin{aligned}
\mathbb{E}[y_t] &= 0, \ \mathbb{E}[y_t^2] = \frac{\sigma^2}{1 - \varphi^2} \\
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|} \gamma_y(0) \ \ (\text{check for } \ h > 0 \ \text{ and } \ h < 0) \\
&= \frac{\varphi^{|h|} \sigma^2}{1 - \varphi^2}
\end{aligned}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \;\; \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\begin{aligned}
\mathbb{E}[y_t] &= 0, \;\; \mathbb{E}[y_t^2] = \frac{\sigma^2}{1-\varphi^2} \\
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|}\gamma_y(0) \;\; (\text{check for } \; h > 0 \; \text{ and } \; h < 0) \\
&= \frac{\varphi^{|h|}\sigma^2}{1-\varphi^2}
\end{aligned}$$

- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \; \{\varepsilon_t\}_{t\geq 1} \sim WN(0, \sigma^2)$$

Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\begin{aligned}
\mathbb{E}[y_t] &= 0, \; \mathbb{E}[y_t^2] = \frac{\sigma^2}{1-\varphi^2} \\
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|}\gamma_y(0) \; \text{(check for } h > 0 \text{ and } h < 0) \\
&= \frac{\varphi^{|h|}\sigma^2}{1-\varphi^2}
\end{aligned}$$

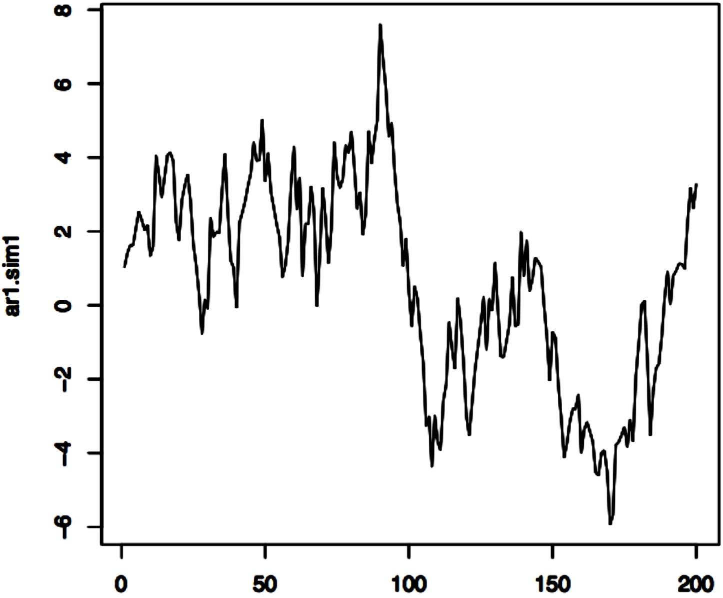
- **Example**: AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

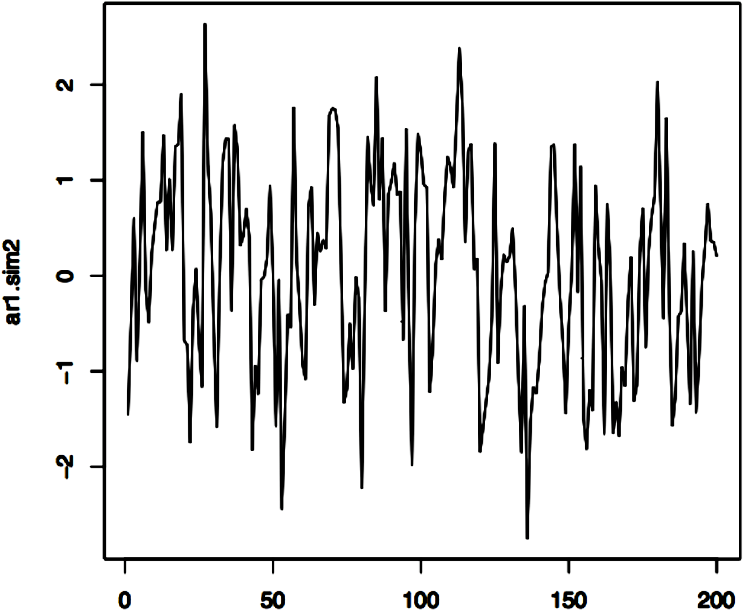
Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\begin{aligned}
\mathbb{E}[y_t] &= 0, \ \ \mathbb{E}[y_t^2] = \frac{\sigma^2}{1-\varphi^2} \\
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|}\gamma_y(0) \ \ \text{(check for } h > 0 \text{ and } h < 0\text{)} \\
&= \frac{\varphi^{|h|}\sigma^2}{1-\varphi^2}
\end{aligned}$$

AR(1): $\varphi = 0.95$



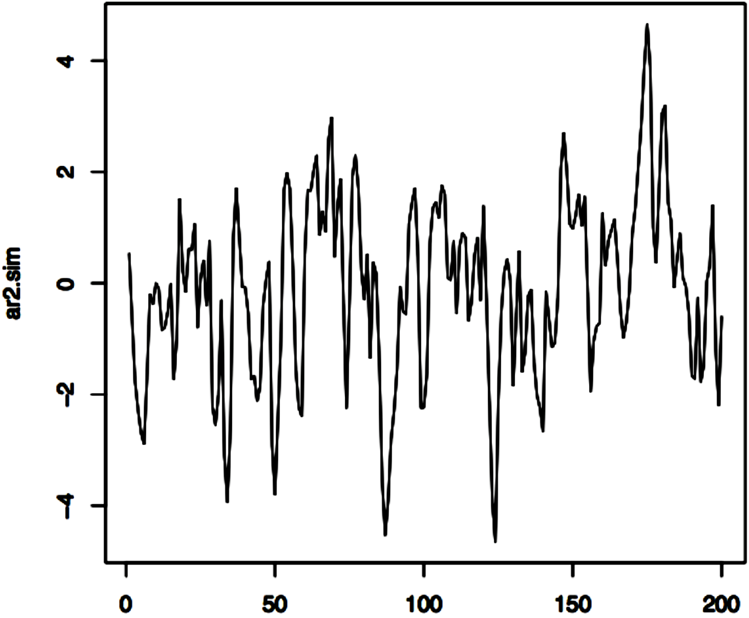
AR(1): $\varphi = 0.5$



- AR(2) process (**AutoRegression**)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t, \;\; \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

AR(2): $\varphi_1 = 0.9$, $\varphi_2 = 0.2$



- AR(p) process (**AutoRegression**)

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t, \ \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

- Stationarity condition: all roots of the polynomial $\Phi(z)$ lie inside the unit circle, $|z_i| < 1$,

$$\Phi(z) = 1 - \sum_{i=1}^{p} \varphi_i z^{p-i}.$$

- AR(p) process (**AutoRegression**)

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t, \ \ \{\varepsilon_t\}_{t\geq 1} \sim WN(0, \sigma^2)$$

- Stationarity condition: all roots of the polynomial $\Phi(z)$ lie inside the unit circle, $|z_i| < 1$,

$$\Phi(z) = 1 - \sum_{i=1}^{p} \varphi_i z^{p-i}.$$


AR(0)
AR(1)
AR(2)
φ = 0.3
φ = 0.9
φ1 = 0.3
φ2 = 0.3
φ1 = 0.9
φ2 = -0.8

- **Example**: MA(1) process (**Moving Average**)

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1},$$
$$\{\varepsilon_t\}_{t\geq 1} \sim WN(0, \sigma^2),$$

$\{\varepsilon_t\}$ is a white noise process. We have $\mathbb{E}[y_t] = 0$ and

$$\begin{aligned}\gamma_y(t+h, t) &= \mathbb{E}(y_{t+h}y_t) \\ &= \mathbb{E}[(\varepsilon_{t+h} + \theta\varepsilon_{t+h-1})(\varepsilon_t + \theta\varepsilon_{t-1})] \\ &= \begin{cases}\sigma^2(1+\theta^2) \text{ if } h = 0, \\ \sigma^2\theta \text{ if } h = \pm 1, \\ 0 \text{ otherwise}\end{cases}\end{aligned}$$

- Thus $\{y_t\}_{t\geq 1}$ is stationary

- **Example**: MA(1) process (**Moving Average**)

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1},$$
$$\{\varepsilon_t\}_{t\geq 1} \sim WN(0, \sigma^2),$$

$\{\varepsilon_t\}$ is a white noise process. We have $\mathbb{E}[y_t] = 0$ and

$$\begin{aligned}\gamma_y(t+h, t) &= \mathbb{E}(y_{t+h}y_t)\\ &= \mathbb{E}[(\varepsilon_{t+h} + \theta\varepsilon_{t+h-1})(\varepsilon_t + \theta\varepsilon_{t-1})]\\ &= \begin{cases}\sigma^2(1+\theta^2) & \text{if } h = 0,\\ \sigma^2\theta & \text{if } h = \pm 1,\\ 0 & \text{otherwise}\end{cases}\end{aligned}$$

- Thus $\{y_t\}_{t\geq 1}$ is stationary

- **Example**: MA(1) process (**Moving Average**)

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1},$$
$$\{\varepsilon_t\}_{t\geq 1} \sim WN(0, \sigma^2),$$

$\{\varepsilon_t\}$ is a white noise process. We have $\mathbb{E}[y_t] = 0$ and

$$\begin{aligned}\gamma_y(t+h,t) &= \mathbb{E}(y_{t+h}y_t) \\ &= \mathbb{E}[(\varepsilon_{t+h} + \theta\varepsilon_{t+h-1})(\varepsilon_t + \theta\varepsilon_{t-1})] \\ &= \begin{cases} \sigma^2(1+\theta^2) \text{ if } h = 0, \\ \sigma^2\theta \text{ if } h = \pm 1, \\ 0 \text{ otherwise} \end{cases}\end{aligned}$$

- Thus $\{y_t\}_{t\geq 1}$ is stationary

- **Example**: MA(1) process (**Moving Average**)

$$
\begin{aligned}
y_t &= \varepsilon_t + \theta \varepsilon_{t-1}, \\
\{\varepsilon_t\}_{t \geq 1} &\sim WN(0, \sigma^2),
\end{aligned}
$$

$\{\varepsilon_t\}$ is a white noise process. We have $\mathbb{E}[y_t] = 0$ and

$$
\begin{aligned}
\gamma_y(t+h, t) &= \mathbb{E}(y_{t+h} y_t) \\
&= \mathbb{E}[(\varepsilon_{t+h} + \theta \varepsilon_{t+h-1})(\varepsilon_t + \theta \varepsilon_{t-1})] \\
&= \begin{cases} \sigma^2(1+\theta^2) \text{ if } h = 0, \\ \sigma^2 \theta \text{ if } h = \pm 1, \\ 0 \text{ otherwise} \end{cases}
\end{aligned}
$$

- Thus $\{y_t\}_{t \geq 1}$ is stationary

- **Example**: MA(1) process (**Moving Average**)

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1},$$
$$\{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2),$$

$\{\varepsilon_t\}$ is a white noise process. We have $\mathbb{E}[y_t] = 0$ and

$$\begin{aligned}\gamma_y(t+h, t) &= \mathbb{E}(y_{t+h} y_t) \\ &= \mathbb{E}[(\varepsilon_{t+h} + \theta \varepsilon_{t+h-1})(\varepsilon_t + \theta \varepsilon_{t-1})] \\ &= \begin{cases} \sigma^2(1+\theta^2) \text{ if } h = 0, \\ \sigma^2 \theta \text{ if } h = \pm 1, \\ 0 \text{ otherwise} \end{cases}\end{aligned}$$

- Thus $\{y_t\}_{t \geq 1}$ is stationary

- **Example**: MA(1) process (**Moving Average**)

$$
\begin{aligned}
y_t &= \varepsilon_t + \theta \varepsilon_{t-1}, \\
\{\varepsilon_t\}_{t\geq 1} &\sim WN(0, \sigma^2),
\end{aligned}
$$

$\{\varepsilon_t\}$ is a white noise process. We have $\mathbb{E}[y_t] = 0$ and

$$
\begin{aligned}
\gamma_y(t+h, t) &= \mathbb{E}(y_{t+h} y_t) \\
&= \mathbb{E}[(\varepsilon_{t+h} + \theta \varepsilon_{t+h-1})(\varepsilon_t + \theta \varepsilon_{t-1})] \\
&= \begin{cases} \sigma^2(1+\theta^2) \text{ if } h = 0, \\ \sigma^2\theta \text{ if } h = \pm 1, \\ 0 \text{ otherwise} \end{cases}
\end{aligned}
$$

- Thus $\{y_t\}_{t\geq 1}$ is stationary

- MA(q) process (**Moving Average**)

$$y_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$
$$\{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2),$$

  $\{\varepsilon_t\}$ is a white noise process.
- For MA we have limit range of covariances between y_t.

- Sample autocovariance function

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (y_{t+|h|} - \overline{y})(y_t - \overline{y})$$

- $\approx$ sample covariance of $(y_1, y_{h+1}), \ldots, (y_{n-h}, y_n)$, except that
  - we normalize by n instead of $n-h$, and
  - we subtract the full sample mean
- We estimate sample variance and obtain sample autocorrelation function

- Sample autocovariance function

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (y_{t+|h|} - \overline{y})(y_t - \overline{y})$$

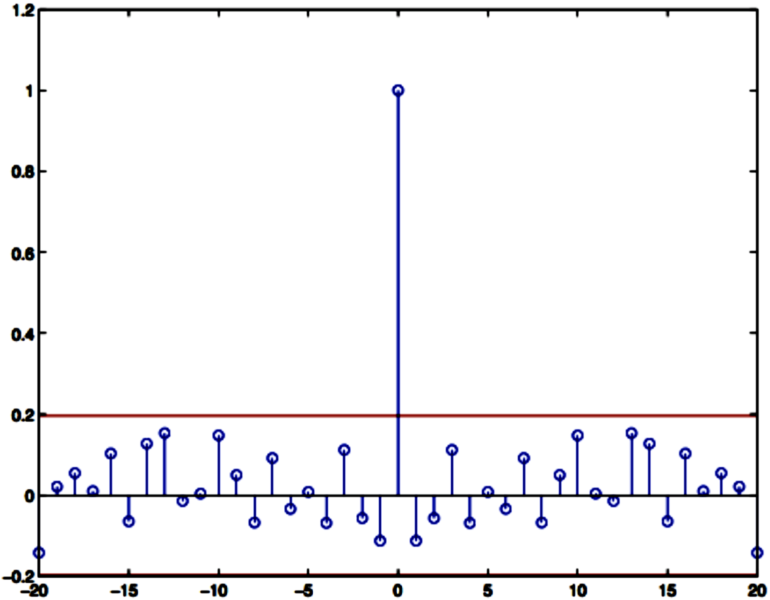
- $\approx$ sample covariance of $(y_1, y_{h+1}), \ldots, (y_{n-h}, y_n)$, except that
  - we normalize by n instead of $n - h$, and
  - we subtract the full sample mean
- We estimate sample variance and obtain sample autocorrelation function

- Sample autocovariance function

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (y_{t+|h|} - \overline{y})(y_t - \overline{y})$$

- $\approx$ sample covariance of $(y_1, y_{h+1}), \ldots, (y_{n-h}, y_n)$, except that
  - we normalize by n instead of $n - h$, and
  - we subtract the full sample mean
- We estimate sample variance and obtain sample autocorrelation function

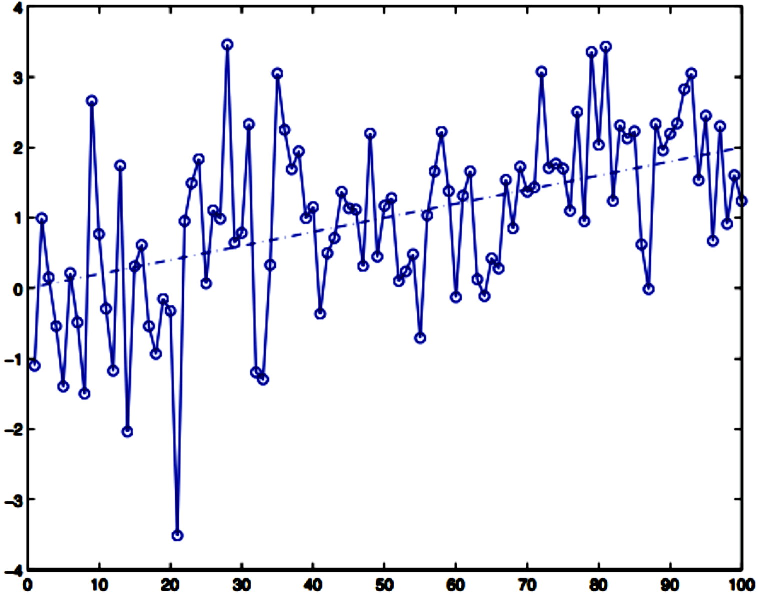
Sample ACF for Gaussian noise



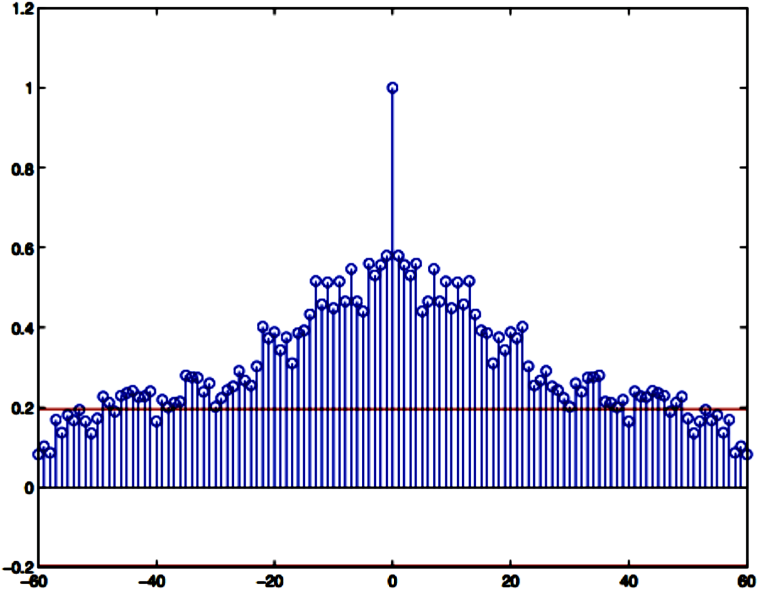
We can recognize the sample autocorrelation functions of many non-white (even non-stationary) time-series

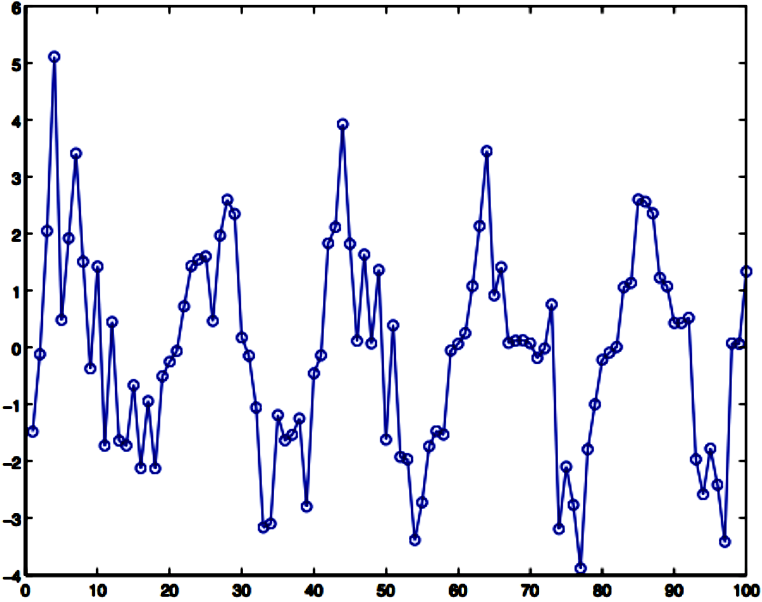
Time series:	**Sample ACF:**
White	zero
Trend	Slow decay
Periodic	Periodic
MA(q)	Zero for $\lvert h \rvert > q$
AR(p)	Decays to zero exponentially

Trend

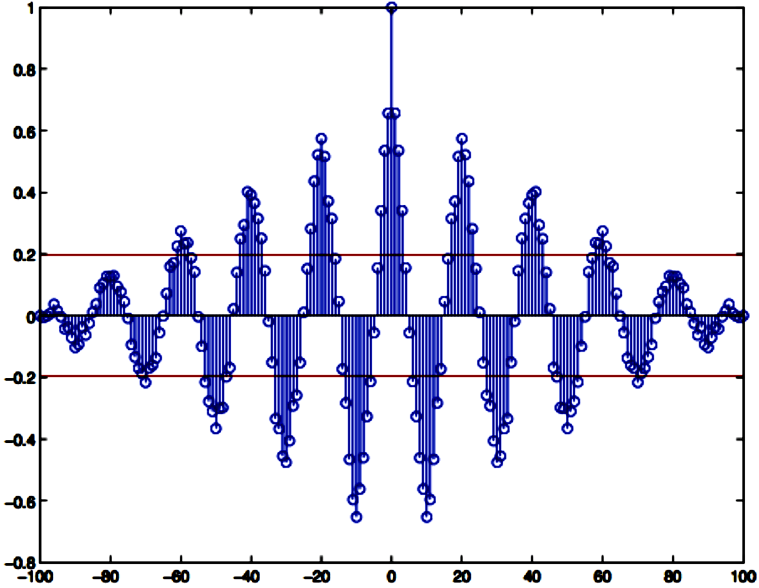


Sample ACF: Trend

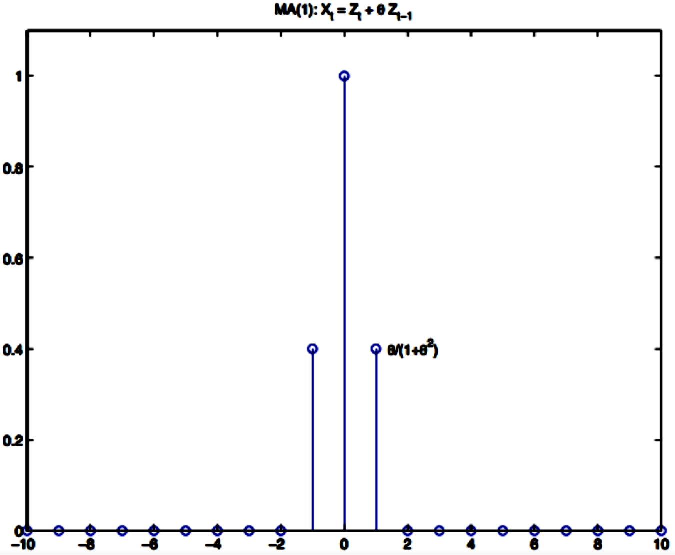



6
5
4
3
2
1
0
−1
−2
−3
−4
0
10
20
30
40
50
60
70
80
90
100

Sample ACF: Trend

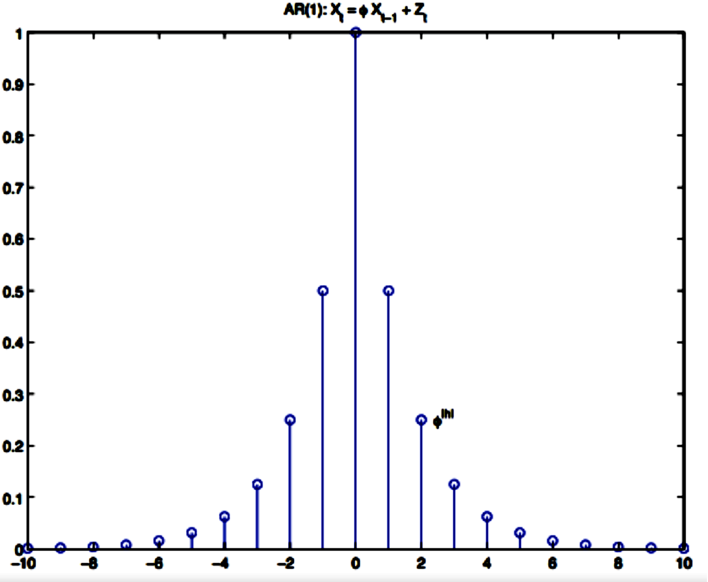


ACF: MA(1)



MA(1): $X_t = Z_t + \theta Z_{t-1}$
$\theta/(1+\theta^2)$
1
0.8
0.6
0.4
0.2
0
−10
−8
−6
−4
−2
0
2
4
6
8
10

ACF: AR



Outline

1. Auto Regression (AR)
2. Moving average
3. Autocovariance function
4. Auto Regression (ARMA) process

- An **ARMA(p,q) process** $\{y_t\}_{t\geq 1}$ is a stationary process that satisfies

$$y_t - \varphi_1 y_{t-1} - \ldots - \varphi_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots \theta_q \varepsilon_{t-q},$$

where $\{\varepsilon_t\}_{t\geq 1} \sim WN(0, \sigma^2)$

- Given n observations, in case of **AR(p) process** the parameters can be estimated by least-squares

$$\widehat{\varphi} = \arg\min_{\varphi} \sum_{t=p+1}^{n} [y_t - \varphi_1 y_{t-1} - \ldots - \varphi_p y_{t-p}]^2$$

- In matrix form for

$$\mathbf{X} = \begin{bmatrix} y_{n-1} & y_{n-2} & \cdots & y_{n-p-1} \\ y_{n-2} & y_{n-3} & \cdots & y_{n-p-2} \\ \vdots & \vdots & \vdots & \vdots \\ y_p & y_{p-1} & \cdots & y_1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_{p+1} \end{bmatrix}$$

then

$$\widehat{\varphi} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

- An **ARMA(p,q) process** $\{y_t\}_{t\geq 1}$ is a stationary process that satisfies

$$y_t - \varphi_1 y_{t-1} - \ldots - \varphi_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots \theta_q \varepsilon_{t-q},$$

where $\{\varepsilon_t\}_{t\geq 1} \sim WN(0, \sigma^2)$

- Given n observations, in case of **AR(p) process** the parameters can be estimated by least-squares

$$\widehat{\varphi} = \arg\min_{\varphi} \sum_{t=p+1}^{n} [y_t - \varphi_1 y_{t-1} - \ldots - \varphi_p y_{t-p}]^2$$

- In matrix form for

$$\mathbf{X} = \begin{bmatrix} y_{n-1} & y_{n-2} & \cdots & y_{n-p-1} \\ y_{n-2} & y_{n-3} & \cdots & y_{n-p-2} \\ \vdots & \vdots & \vdots & \vdots \\ y_p & y_{p-1} & \cdots & y_1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_{p+1} \end{bmatrix}$$

then

$$\widehat{\varphi} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

- An **ARMA(p,q) process** $\{y_t\}_{t\geq 1}$ is a stationary process that satisfies

$$y_t - \varphi_1 y_{t-1} - \ldots - \varphi_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots \theta_q \varepsilon_{t-q},$$

  where $\{\varepsilon_t\}_{t\geq 1} \sim WN(0, \sigma^2)$
- Given n observations, in case of **AR(p) process** the parameters can be estimated by least-squares

$$\widehat{\varphi} = \arg\min_{\varphi} \sum_{t=p+1}^{n} \left[y_t - \varphi_1 y_{t-1} - \ldots - \varphi_p y_{t-p}\right]^2$$

- In matrix form for

$$\mathbf{X} = \begin{bmatrix} y_{n-1} & y_{n-2} & \cdots & y_{n-p-1} \\ y_{n-2} & y_{n-3} & \cdots & y_{n-p-2} \\ \vdots & \vdots & \vdots & \vdots \\ y_p & y_{p-1} & \cdots & y_1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_{p+1} \end{bmatrix}$$

  then

$$\widehat{\varphi} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

- AR models assume that the relation between past and future is linear
- Nonlinear Auto Regressive (NAR) formulation

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + \varepsilon_t,$$

where the missing information is lumped into a noise term ε_t

- We will consider this relationship as a particular instance of a dependence

$$y_t = f(\mathbf{x}_t) + \varepsilon, \;\; y_t \in \mathbb{R}^1, \, \mathbf{x}_t \in \mathbb{R}^p,$$

where

$$\mathbf{x}_t = [y_{t-1}, y_{t-2}, \dots, y_{t-p}]$$

- We train $f(\cdot)$ using ML regression algorithm and a sample $(\mathbf{X}, \mathbf{y})$, where

$$\mathbf{X} = \{\mathbf{x}_{p+1}, \mathbf{x}_{p+2}, \dots, \mathbf{x}_T\}$$

and

$$\mathbf{y} = \{y_{p+1}, y_{p+2}, \dots, y_T\}$$

- AR models assume that the relation between past and future is linear
- Nonlinear Auto Regressive (NAR) formulation

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + \varepsilon_t,$$

  where the missing information is lumped into a noise term ε_t

- We will consider this relationship as a particular instance of a dependence

$$y_t = f(\mathbf{x}_t) + \varepsilon, \;\; y_t \in \mathbb{R}^1, \mathbf{x}_t \in \mathbb{R}^p,$$

  where

$$\mathbf{x}_t = [y_{t-1}, y_{t-2}, \dots, y_{t-p}]$$

- We train $f(\cdot)$ using ML regression algorithm and a sample $(\mathbf{X}, \mathbf{y})$, where

$$\mathbf{X} = \{\mathbf{x}_{p+1}, \mathbf{x}_{p+2}, \dots, \mathbf{x}_T\}$$

  and

$$\mathbf{y} = \{y_{p+1}, y_{p+2}, \dots, y_T\}$$

- AR models assume that the relation between past and future is linear
- Nonlinear Auto Regressive (NAR) formulation

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}) + \varepsilon_t,$$

where the missing information is lumped into a noise term ε_t
- We will consider this relationship as a particular instance of a dependence

$$y_t = f(\mathbf{x}_t) + \varepsilon, \;\; y_t \in \mathbb{R}^1, \, \mathbf{x}_t \in \mathbb{R}^p,$$

where

$$\mathbf{x}_t = [y_{t-1}, y_{t-2}, \ldots, y_{t-p}]$$

- We train $f(\cdot)$ using ML regression algorithm and a sample $(\mathbf{X}, \mathbf{y})$, where

$$\mathbf{X} = \{\mathbf{x}_{p+1}, \mathbf{x}_{p+2}, \ldots, \mathbf{x}_T\}$$

and

$$\mathbf{y} = \{y_{p+1}, y_{p+2}, \ldots, y_T\}$$

- AR models assume that the relation between past and future is linear
- Nonlinear Auto Regressive (NAR) formulation

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + \varepsilon_t,$$

  where the missing information is lumped into a noise term ε_t
- We will consider this relationship as a particular instance of a dependence

$$y_t = f(\mathbf{x}_t) + \varepsilon, \ \ y_t \in \mathbb{R}^1, \, \mathbf{x}_t \in \mathbb{R}^p,$$

  where

$$\mathbf{x}_t = [y_{t-1}, y_{t-2}, \dots, y_{t-p}]$$

- We train $f(\cdot)$ using ML regression algorithm and a sample $(\mathbf{X}, \mathbf{y})$, where

$$\mathbf{X} = \{\mathbf{x}_{p+1}, \mathbf{x}_{p+2}, \dots, \mathbf{x}_T\}$$

  and

$$\mathbf{y} = \{y_{p+1}, y_{p+2}, \dots, y_T\}$$

There is a zoo of related models

- ARIMA — model differences $y_t - y_{t-1}$ instead of y_t
- SARIMA — takes into account seasonality
- ARCH - autoregression with conditional heterscedasticity
- GARCH - generalized ARCH

Conclusions

- Autoregression (AR) and Moving average (MA) models describe different types of relations between neighbour observations of time series
- For autoregression autocorrelations are non-zero for all differences
- For Moving average autocorrelations are zero (and we specify the order to define the range of dependence)
- ARIMA model unites AR and MA models
- Using Autocorrelation function we can identify the right model ARMA(p, q)