

Model ARIMA

Zastosowanie do analizy niestacjonarnych szeregów czasowych

Dawid Andrys

9 grudnia 2024 r.

Plan prezentacji

- 1 Wprowadzenie do teorii szeregów czasowych
- 2 Objaśnienie metodologii modeli ARIMA
- 3 Przykład z wykorzystaniem rzeczywistych danych przestępczości w Chinach

- Szereg czasowy zmiennej Y

$$\{Y_t : t = 0, 1, \dots\}$$

$$\mu_Y = \mathbb{E}[Y_t], \quad \sigma_Y^2 = \mathbb{E}[(Y_t - \mu_Y)^2]$$

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$$

- Autokorelacja i autokowariancja

$$\rho_k = \frac{\mathbb{E}[(Y_t - \mu_Y)(Y_{t+k} - \mu_Y)]}{\mathbb{E}[(Y_t - \mu_Y)^2]\mathbb{E}[(Y_{t+k} - \mu_Y)^2]}$$

$$\widehat{\rho}_k = \frac{\sum_{t=1}^{n-k} [(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})]}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

$$\gamma_k = \mathbb{E}[(Y_t - \mu_Y)(Y_{t+k} - \mu_Y)]$$

$$\widehat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} [(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})]$$

$$k = |t - s|$$

- ACF i PACF

$$ACF : \{Y_t, Y_{t+k}\} \rightarrow [-1, 1],$$

$$ACF(Y_t, Y_{t+k}) = \mathbb{E}[Y_t \overline{Y_{t+k}}]$$

$$PACF : \{Y_t, Y_{t+k}\} \rightarrow [-1, 1],$$

$$PACF(Y_t, Y_{t+k}) = ACF(Y_{t+k} - \widetilde{Y_{t+k}}, Y_t - \widetilde{Y_t})$$

Uwaga

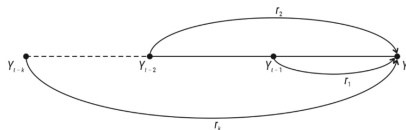
Dla $k=1$ zachodzi równość:

$$ACF(Y_t, Y_{t+k}) = PACF(Y_t, Y_{t+k})$$

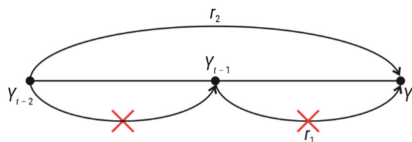
Wprowadzenie do teorii szeregów czasowych

Uwaga

Ilustracje zjawisk autokorelacji i autokowariancji w szeregu czasowym



Rys. 1.1. Ilustracja zjawiska autokorelacji w szeregu czasowym



Rys. 1.2. Ilustracja zjawiska autokowariancji w szeregu czasowym

[Nazarko, Chodakowska, 2022]

- Autoregresja

$$Y_t = \sum_{j=1}^p \phi_j Y_{t-j} + e_t$$

Przykład 1. Biały szum

Szereg niezależnych zmiennych losowych o postaci:

$$\{e_t : t = 0, 1, \dots\}$$

o jednakowym rozkładzie ze skończoną wartością oczekiwaną i skończoną wariancją. Kiedy zachodzi:

$$X_t \sim N(0, \sigma_e^2)$$

szereg nazywamy gaussowskim białym szumem.

$$\rho_k = \begin{cases} 1 & : k = 0 \\ 0 & : wpp \end{cases} \quad \gamma_k = \begin{cases} \sigma_e^2 & : k = 0 \\ 0 & : wpp \end{cases}$$

Przykład 2. Błądzenie przypadkowe

Proces błądzenia przypadkowego możemy opisać następującym równaniem:

$$\begin{cases} Y_t = Y_{t-1} + e_t \\ Y_1 = e_1 \end{cases}$$

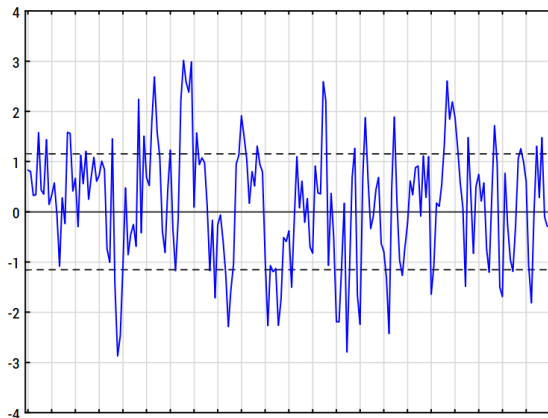
$$\gamma_k = t\sigma_e^2 \quad \text{dla } t \text{ i } s \text{ takich, że } 1 \leq t \leq s$$

$$\rho_k = \sqrt{\frac{t}{s}}$$

Objaśnienie metodologii modeli ARIMA

- 1 Proces stochastyczny stacjonarny
- 2 Proces stochastyczny niestacjonarny

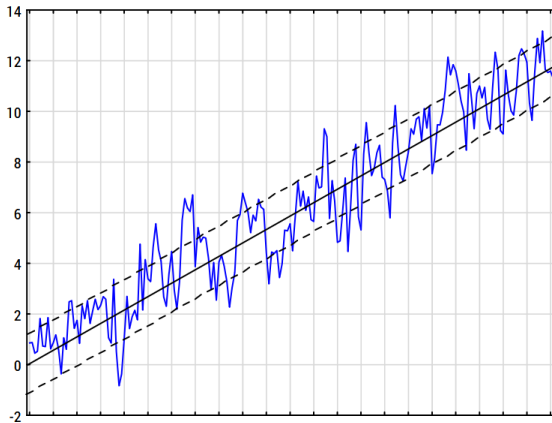
Objaśnienie metodologii modeli ARIMA



Rys. 2.1. Przykład szeregu stacjonarnego ze stałą średnią i wariancją

[Nazarko, Chodakowska, 2022]

Objaśnienie metodologii modeli ARIMA



Rys. 2.2. Przykład szeregu niestacjonarnego ze względu na średnią (z trendem)

[Nazarko, Chodakowska, 2022]

Objaśnienie metodologii modeli ARIMA

Aby przybliżyć metodologię modeli ARIMA, wprowadzimy następujące operatory:

- operator przesunięcia wstecz B

$$BY_t = Y_{t-1}$$

- operator różnicowy:

$$\nabla Y_t = Y_t - Y_{t-1} = (1 - B)Y_t$$

Zachodzą następujące własności:

$$B(BY_t) = BY_{t-1} = Y_{t-2}$$

$$B^k Y_t = Y_{t-k}$$

$$\nabla^d = (1 - B)^d$$

Objaśnienie metodologii modeli ARIMA

Operatory wykorzystujemy do:

- oceny stacjonarności procesu stochastycznego zgodnie z **testami pierwiastków jednostkowych**:

$$Y_t = \frac{1}{2} Y_{t-1} + e_t \Leftrightarrow (1 - \frac{1}{2}B) Y_t = e_t$$

$\Rightarrow B = 2 \Leftrightarrow$ Proces jest stacjonarny

$$Y_t = -\frac{7}{5} Y_{t-1} - \frac{2}{5} Y_{t-2} + e_t$$

$$\Leftrightarrow (1 + \frac{2}{5}B)(1 + B) Y_t = e_t$$

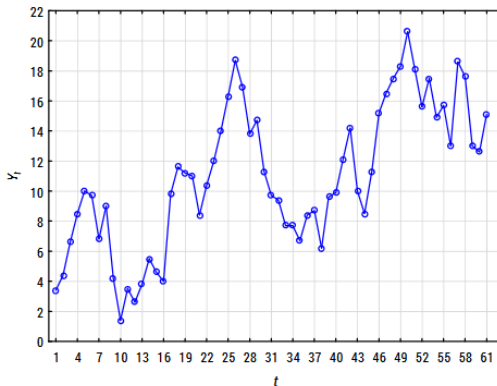
$\Rightarrow B = -\frac{1}{4} \vee B = -1 \Leftrightarrow$ Proces jest niestacjonarny

- sprowadzenia procesu niestacjonarnego do postaci stacjonarnej poprzez **różnicowanie z rzędem d**

$$\hat{Y}_t = \nabla^d(Y_t)$$

- Proces stochastyczny niestacjonarny, którego można sprowadzić do stacjonarnego poprzez różnicowanie z rzędem d , nazywamy **zintegrowanym w stopniu d**

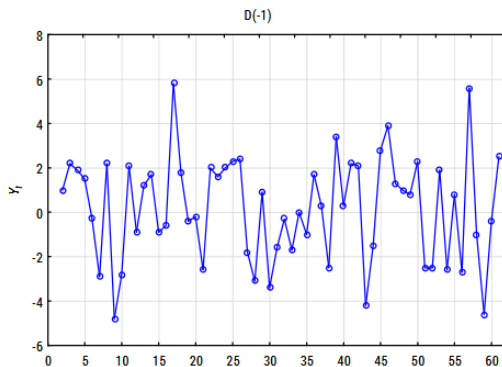
Objaśnienie metodologii modeli ARIMA



Rys. 2.3. Szereg średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r

[Nazarko, Chodakowska, 2022]

Objaśnienie metodologii modeli ARIMA



Rys. 2.4. Szereg średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r. po operacji różnicowania $d = 1$

[Nazarko, Chodakowska, 2022]

Objaśnienie metodologii modeli ARIMA

Modele ARIMA (ang. *Autoregressive Integrated Moving Average*, wywodzące się z prac G.E.P. Boxa i G.M. Jenkinsa, są jedną z najpopularniejszych uniwersalnych metod analizy szeregów czasowych, których różnice są zintegrowane w stopniu d .

W modelach ARIMA bieżąca wartość zmiennej Y jest interpretowana jako funkcja jej przeszłych wartości i nieskorelowanych wahań:

$$Y_t = f(Y_{t-k}, e_{t-k}) + e_t$$

Objaśnienie metodologii modeli ARIMA

Wyróżniamy następujące typy modeli, które wykorzystujemy później w konstrukcji modelu ARIMA:

- AR(p): model autoregresji rzędu p, w którym obserwacja jest skorelowana z p poprzedzającymi obserwacjami:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t = (1 - \phi_1 B - \dots - \phi_p B^p) Y_t = \phi(B) Y_t = e_t$$

gdzie „wyraz wolny” nazywamy resztą dla momentu, okresu t

- MA(q): model średniej ruchomej rzędu q, w którym każda obserwacja jest skorelowana z odpowiadającą resztą q-tej obserwacji i nieskorelowana z poprzednimi:

$$Y_t = e_t - \theta_1 Y_{t-1} - \dots - \theta_q Y_{t-q} = (1 - \theta_1 B - \dots - \theta_q B^q) Y_t = \theta(B) e_t$$

Objaśnienie metodologii modeli ARIMA

W modelu ARMA(p,q) wykorzystujemy kombinację AR(p) i MA(q):

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

$$\Leftrightarrow \phi(B) Y_t = \theta(B) e_t$$

Z kolei model ARIMA(p,d,q) wprowadza do modelu ARMA(p,q) różnicowanie z rzędem d w celu usunięcia niestacjonarności. W innych słowach, **model ARIMA(p,d,q) pozwala na d-krotne zróżnicowanie procesu niestacjonarnego i zastosowanie na tak uzyskanym procesie standardowym modelu ARMA(p,q)**

Objaśnienie metodologii modeli ARIMA

Przykłady

$$\text{ARIMA}(2,0,0) \equiv \text{AR}(2)$$

$$\text{ARIMA}(1,0,1) \equiv \text{ARMA}(1,1)$$

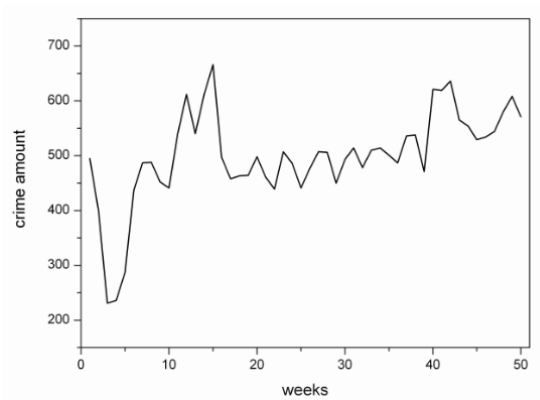
$$\phi(B)(1 - B)Y_t = e_t \rightarrow \text{ARIMA}(1,1,0)$$

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)Y_t = e_t \rightarrow \text{ARIMA}(3,1,0)$$

Wzór ogólny

$$\phi(B)(1 - B)^d Y_t = \theta(B)e_t \Leftrightarrow \phi(B) \nabla^d Y_t = \theta(B)e_t \rightarrow \text{ARIMA}(p,d,q)$$

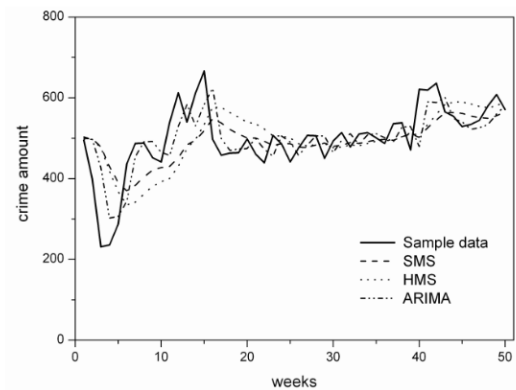
Przykład z wykorzystaniem rzeczywistych danych przestępczości w Chinach



Rys. 3.1. Szereg czasowy danych dotyczących przestępczości

[Chen, Yuan, Shu, 2008]

Przykład z wykorzystaniem rzeczywistych danych przestępczości w Chinach



Rys. 3.2. Szereg czasowy po zróżnicowaniu z wykorzystaniem modelu ARIMA (w porównaniu z rezultatami dla dwóch modeli wygładzania wykładniczego)

[Chen, Yuan, Shu, 2008]

 Joanicjusz Nazarko, Ewa Chodakowska (2022)

Prognostowanie w zarządzaniu przedsiębiorstwem. Cz. V. Prognostowanie na podstawie modeli ARIMA

Oficyna Wydawnicza Politechniki Białostockiej.

 Peng Chen, Hongyong Yuan, Xueming Shu (2008)

Forecasting Crime Using the ARIMA Model

IEEE (Fifth International Conference on Fuzzy Systems and Knowledge Discovery).