Szeregi czasowe

Grzegorz Górzny

19 May 2021

Czym jest szereg czasowy

Szereg czasowy - realizacja procesu stochastycznego, którego dziadzina jest czas (zazwyczaj z regularnymi odstepami). przykłady to: wartość akcji na giełdzie, liczba chorych, zanieczyszczenie powietrza, poziom CO_2 w powietrzu itp.

Składowe szeregu

W szeregu czasowym wyróżniamy kilka składników: trend, wahania sezonowe i bład losowy.

- Model addytywny: $X_t = m_t + s_t + e_t$
- Model multiplikatywny: $X_t = m_t * s_t * e_t$

gdzie m_t - funkcja trendu, s_t - funkcja sezonowości, e_t - losowe odchylenie.

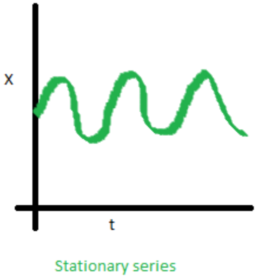
Proces stacjonarny

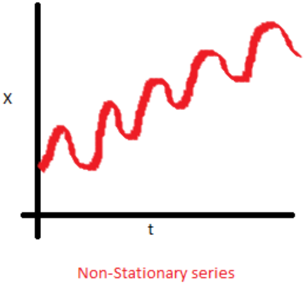
Zakładamy, że losowe odchylenie (e_t) jest stacjonarny. Tzn.

- $E[e_t] = \mu$, dla każdego t
- $Var(e_t) = \sigma^2 < \infty$
- $Cov(e_t, e_s) = Cov(e_{t+h}, e_{s+h})$ dla każdego h całkowitego.

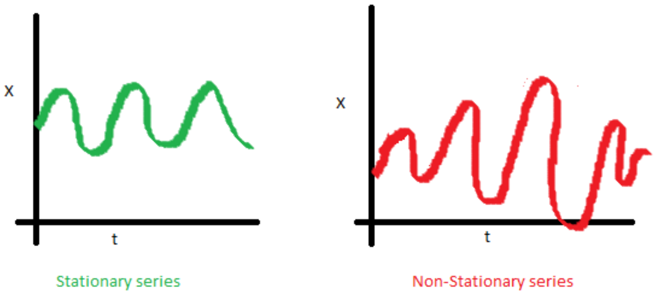
Dla uproszczenia bedziemy zapisywać $\gamma(h) = Cov(e_0, e_h)$, a $\gamma(h)$ nazwiemy funkcja autokowariancji.

Proces stacjonarny c.d.

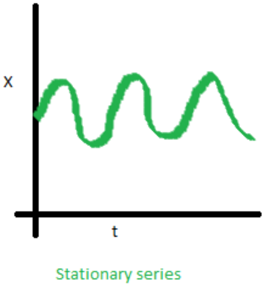


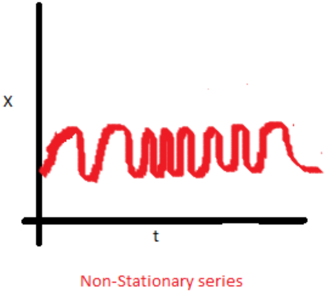


Proces stacjonarny c.d.



Proces stacjonarny c.d.





Funkcja autokowaraincji

Własności funkcji autokowariancji $\gamma(h)$:

- $\gamma(0) \geq 0$
- $\gamma(0) \geq \gamma(h)$
- $\gamma(h) = \gamma(-h)$

Funkcja autokorelacji

Jeśli funkcje $\gamma(h)$ podzielimy przez $\gamma(0)$, otrzymamy funkcje $\rho(h)$ - zwana funkcja autokorelacji. Własności funkcji autokorelacji $\gamma(h)$:

- $\rho(0) = 1$
- $1 \geq |\rho(h)|$
- $\rho(h) = \rho(-h)$

Proces średniej ruchomej

Procesem średniej ruchomej rzedu q (ozn. $MA(q)$) bedziemy nazywać proces postaci:

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

gdzie $Z_t, ...Z_{t-q}$ sa nieskorelowanymi zmiennymi losowymi o tej samej średniej równej 0 i wariancji σ^2 (zmienne te nazywane sa białym szumem)
Proces średniej ruchomej jest procesem stacjonarnym dla skończonych q. Funkcja aukokorelacji wynosi 0 dla $h > q$.

Proces autoregresji

Procesem autoregresji rzedu p (ozn. $AR(p)$) bedziemy nazywać proces postaci:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

Proces autoregresji nie musi być stacjonarny. Jeśli mamy proces $AR(1)$ postaci $X_t = X_{t-1} + Z_t$ (tzn. $\phi_1 = 1$) mamy bładzenie losowe. Łatwo zauważyć, że wariancja bedzie zależeć od czasu - $X_t = X_0 + \sum_{i=0}^{t} Z_i$ i $Var(X_t) = Var(X_0) + t\sigma^2$

Proces arma

Proces Arma rzedu p i q (ozn. $ARMA(p,q)$) nazywać proces postaci:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

Wielomian Autoregresji jest postaci:

$$\phi(s) = 1 - \phi_1 s - \phi_2 s^2 - \cdots - \phi_p s^p$$

Proces $ARMA(p,q)$ jest stacjonarny, gdy pierwiastki wielomianu autoregresji spełniaja zależność $|s| > 1$

Proces arima

Proces ARIMA rzedu p, d i q (ozn. $AIRMA(p, d, q)$) nazwiemy proces, który po d-krotnym zróżnicowaniu jest procesem $ARMA(p, q)$. Zróżnicowanie szeregu, czyli wykonaniu operacji $X_{t+1} - X_t$ d razy. Tak powstały nowy szereg czasowy bedzie szeregiem $ARMA(p, q)$

Funkcja autokorelacji czastkowej

Funkcje autokorelacji $\alpha(n)$ zdefiniujemy nastepujaco: $\alpha(0) = 1$ oraz $\alpha(n) = \phi_{nn}$, gdzie ϕ_{nn} - n-ta współrzedna wektora ϕ_n, który spełnia równanie $R_n\phi_n = \rho_n$.
R_n - macierz autokowariancji (rzedu n), $\rho_n = (\rho_1, \rho_2, ..., \rho_n)$
Dla procesu $AR(p)$, $\alpha(n) = 0$ dla $n > p$

Operacje na szeregach czasowych

- Operacja przesuniecia (lag) - przesuniecie o d to wykonanie operacji $X_{t+d} - X_t$
- Różnicowanie d-krotne - d-krotne złożenie przesuniecia o 1
- Transformacja Boxa-Coxa
- Wygładzanie wykładnicze

Pierwsze dwie można wykonać w R za pomoca funkcji *diff()*, odpowiednio z parametrem lag dla przesuniecie, differences dla różnicowania.

Wygładzanie wykładnicze

Wygładzanie wykładnicze na szeregu czasowym $\{X_t\}$ tworzy nowy szereg $\{Y_t\}$, spełniajacy poniższe założenia:
1) $Y_0 = X_0$
2) $Y_t = a \cdot X_t + (1 - a) \cdot X_{t-1}$
3) $a \in [0, 1]$

Test pierwiastków wielomianu AR(p)

Rozszerzony test Dickeya-Fullera - służy do testowania nastepujacej hipotezy

H_0 : istnieje pierwiastek s_i dla którego zachodzi $|s_i| \leq 1$

W praktyce można stosować do sprawdzenia, czy proces ARIMA jest stacjonarny. Test ten może (ale nie musi) działać dla danych sezonowych - na takich szeregach trzeba wykonać operacje usuwajace sezonowość (np. opóźnienie).

Test można wykonać w R za pomoca funkcji *adf.test()* z biblioteki tseries

SARIMA

Z szeregiem *SARIMA* mamy do czynienia gdy w szeregu czasowym mamy paramerty jak w szeregu $ARIMA(p, d, q)$, wystepuja czynniki okresowe. Sa to:

- P - rzad sezonowej autoregresji
- D - rzad różnicowania okresowego
- Q - rzad sezonowej średniej ruchomej
- m - liczba obserwacji w danym okresie

Pierwsze trzy sa odpowiednikami sezonowymi dla parametrów szeregu *ARIMA*. Szereg ten zapisujemy tak: $SARIMA(p, d, q)(P, D, Q)[m]$.

Źródło

- Obrazki ze strony *https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/*
- *https://www.aptech.com/blog/introduction-to-the-fundamentals-of-time-series-data-and-analysis/*
- *https://www.pluralsight.com/guides/time-series-forecasting-using-r*
- notatki własne