Újfajta okság-felfogás az idősorok „nyoma” alapján

Dr. Rappai Gábor[2], Abaligeti Gallusz[1], Kehl Dániel[2]

[1]BCE-KTK,
[2]PTE-KTK

2012. június 4.

Tartalomból

Idősorok nyoma

Alapötlet és definíció

Adott egy idősor, ennek a nyomát keressük a következőképpen:

1. illesztünk egy jól illeszkedő idősori modellt (δ_t maradékkal),
2. a várható érték egyenlet ismeretében hozzárendelünk az idősorhoz egy $(-1, 0, +1)$ sorozatot, melynek modellbe vonásával eltérés-négyzetösszege (SSE) tovább csökken

Definíció (idősor nyoma)

Legyenek $D^+ : \{1 \dots T\} \to \{0, +1\}$ és $D^- : \{1 \dots T\} \to \{-1, 0\}$ azok a sorozatok, amik $\alpha, \beta > 0$ feltétel mellett a

$$\sum_{t=1}^{T} (\delta_t - \alpha D_t^- - \beta D_t^+)^2 \to \min$$

megoldják, ekkor a $D_t := D_t^- + D_t^+$-t az idősor nyomának hívjuk.

A várható érték egyenlet maradékait rendezzük **növekvő sorrendbe** és jelöljük ϵ_t-vel a t-ik legnagyobb reziduumot. Belátható hogy, ha

$$k^* = \arg\max_{k \in \{1 \ldots T\}} \frac{1}{k} \left(\sum_{t=1}^{k} \epsilon_t \right)^2 \quad \text{és} \quad K^* = \arg\max_{K \in \{1 \ldots T\}} \frac{1}{T - K + 1} \left(\sum_{t=K}^{T} \epsilon_t \right)^2,$$

akkor az idősor nyoma a következő – három értékű – változó sorozat lesz:

$$D_t = \begin{cases} +1, & \text{ha } \epsilon_{K^*} \leq \delta_t \\ 0, & \text{ha } \epsilon_{k^*} < \delta_t < \epsilon_{K^*} \\ -1, & \text{ha } \delta_t \leq \epsilon_{k^*}. \end{cases}$$

Megjegyzés: a K^*-ra vonatkozó feladat visszavezethető a k^*-ra vonatkozó feladat alakjára (de nem ugyanarra).

Idősorok nyoma

A nyom kiszámítása

Most tegyük fel, hogy $\delta_1, \ldots, \delta_T$ FAE valószínűségi változók, ha növekvő sorba vannak rendezve, akkor elég nagy T mellett $\epsilon_t \approx F^{-1}(t/T)$ ahol $t \in \{1 \ldots T\}$.

Áttérve az $x = t/T$ változóra a maximum probléma a következő lesz

$$\max_{x \in (0,1]} \frac{1}{x} \left(\int_0^x F^{-1}(y)\,\mathrm{d}y \right)^2.$$

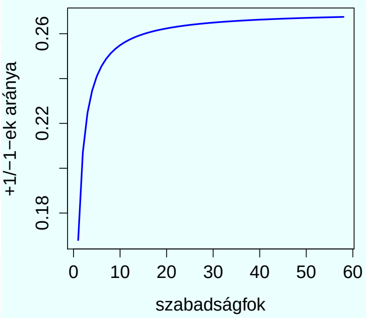
Ezt a kifejezést x szerint deriválva majd véve a $z = F^{-1}(x)$ helyettesítést, első rendű feltételhez jutunk

$$2F(z)z = \int_{-\infty}^{z} u\,\mathrm{d}F(u),$$

ebből $x^* = F(z^*)$ lesz az elméleti kvantilis. Véges, rendezett mintában pedig nagyjából a $k^* = \lfloor Tx^* \rfloor$-ik elemnél lesz szélsőérték.

Idősorok nyoma

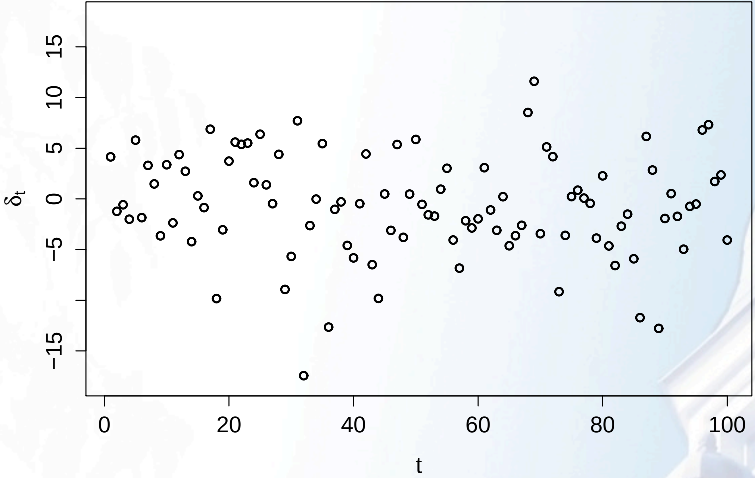
Néhány nevezetes, szimmetrikus eloszlás nyoma



Eloszlás	-1	**0**	$+1$
Egyenletes	0.333	0.333	0.333
Normális	0.270	0.460	0.270
Student (5)	0.241	0.518	0.241
Student (10)	0.255	0.490	0.255
Student (50)	0.267	0.476	0.267
Student (100)	0.268	0.474	0.268
Student (1000)	0.270	0.460	0.270

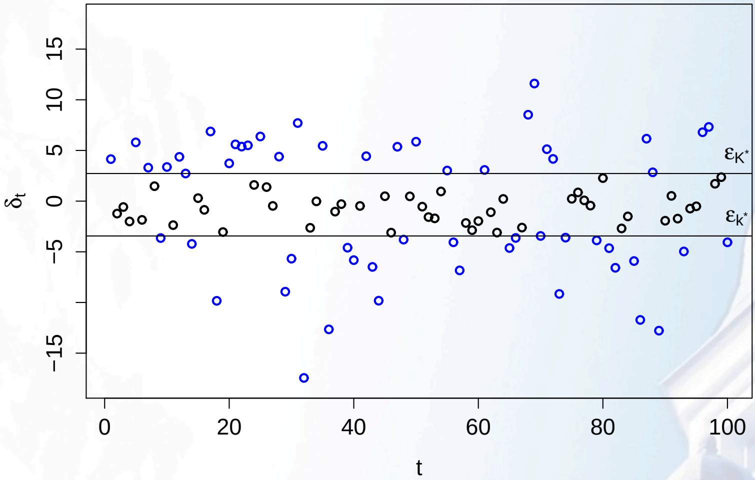
Idősorok nyoma

A nyom kiszámítása empirikus idősoroknál: normális reziduumok $\mathcal{N}(0, \sigma^2)$



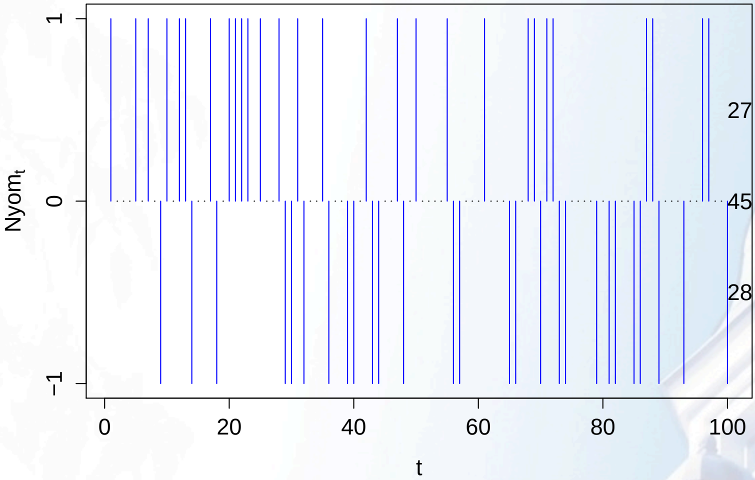
Idősorok nyoma

A nyom kiszámítása empirikus idősoroknál: normális reziduumok $\mathcal{N}(0, \sigma^2)$



Idősorok nyoma

A nyom kiszámítása empirikus idősoroknál: normális reziduumok $\mathcal{N}(0, \sigma^2)$



Definíció (Erős Granger-okság)

*Az X_t erősen **nem** Granger-oka Y_t-nek pontosan akkor, ha*

$$\mathbf{P}(y_t \mid y_{t-1}, \ldots, y_1) = \mathbf{P}(y_t \mid y_{t-1}, \ldots, y_1, x_{t-1}, \ldots x_1),$$

minden $t = 1, \ldots, T$ és minden realizáció esetében.

Megjegyzés: ezt elég nehéz tesztelni empirikus idősorok esetében. A következőkben csupán elsőrendű okságot fogunk vizsgálni, ekkor elegendő a következőt ellenőrizni:

$$\mathbf{P}(y_t \mid y_{t-1}) = \mathbf{P}(y_t \mid y_{t-1}, x_{t-1}),$$

minden $t = 1, \ldots, T$ és minden y_t, y_{t-1} és x_{t-1} realizáció esetében.

Jelenségek közti okság

Okság a nyom alapján

A problémához felírható logisztikus regresszió a következő:

$$\mathbf{P}(y_t = -1 \mid y_{t-1}, x_{t-1}) = \frac{\exp\left\{\alpha_0 + \beta_0 y_{t-1} + \gamma_0 x_{t-1}\right\}}{1 + \sum_{k=0}^{1} \exp\left\{\alpha_k + \beta_k y_{t-1} + \gamma_k x_{t-1}\right\}}$$

$$\mathbf{P}(y_t = 0 \mid y_{t-1}, x_{t-1}) = \frac{\exp\left\{\alpha_1 + \beta_1 y_{t-1} + \gamma_1 x_{t-1}\right\}}{1 + \sum_{k=0}^{1} \exp\left\{\alpha_k + \beta_k y_{t-1} + \gamma_k x_{t-1}\right\}}$$

$$\mathbf{P}(y_t = +1 \mid y_{t-1}, x_{t-1}) = \frac{1}{1 + \sum_{k=0}^{1} \exp\left\{\alpha_k + \beta_k y_{t-1} + \gamma_k x_{t-1}\right\}}$$

Amennyiben a $\gamma_0 = \gamma_1 = 0$ hipotézist el tudjuk elvetni, akkor X nyom értelemben oka Y-nak.

Jelenségek közti okság

Egy egyszerű példa

Tekintsük a következő VAR modellt:

$$Y_t = \alpha Y_{t-1} + \beta X_{t-1} + \epsilon_t^Y$$
$$X_t = \gamma X_{t-1} + \epsilon_t^X,$$

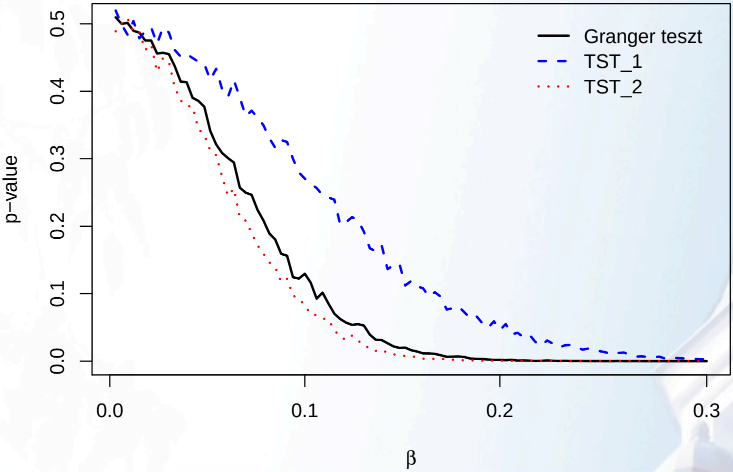
ahol $|\alpha|, |\beta|, |\gamma| < 1$ valamint $\epsilon_t^X, \epsilon_t^Y \sim \mathcal{N}(0, 1)$. A két idősor közti okságot a β paraméter határozza meg, ha

- $\beta = 0$, akkor nincs okság,
- $\beta \neq 0$, akkor van okság.

A $\beta \in [0, 0.3]$ tartományt 101 részve osztva, minden értékhez szimuláltunk idősorokat, majd a Granger-teszt és a logisztikus regresszió paramétertesztjeinek a *p*-értékeit vizsgáltuk.

Jelenségek közti okság

Egy egyszerű példa



Köszönjük a figyelmet!

A nyommal kapcsolatos további lehetséges kutatási irányok:

- extrém érték elmélet: optimális treshold kiválasztása
- outlier szűrés

Potenciális előnyök, hátrányok:

+ tetszőleges modellre elvégezhető
+ erős okságot tesztelhető
- információ vesztés
- kétszer annyi teszt mint a Granger-nél

Köszönjük a figyelmet!

Felhasznált irodalom:

- Gary Chamberlain :
  The general equivalence of granger and sims causality.
  In *Econometrica*, 50. évf. (1982) 3. sz., 569–581. p.
- J. P. Florens, M. Mouchart :
  A note on noncausality.
  In *Econometrica*, 50. évf. (1982) 3. sz., 583–591. p.
- Rocco Mosconi, Raffaello Seri :
  Non-causality in bivariate binary time series.
  In *Journal of Econometrics*, 132. évf. (2006) 2. sz., 379–407. p.
- Monica Billio, Silvestro Di Sanzo :
  Granger-causality in Markov Switching Models.
  http://merlin.fae.ua.es/activities/pdf/SDiSanzo.pdf