Časové řady II

Ondřej Vencálek
Univerzita Palackého v Olomouci
ondrej.vencalek@upol.cz

seminář pro VŠB-TUO
2015-03-20, Ostrava

evropský sociální fond v ČR | EVROPSKÁ UNIE | MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY | OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Nové kreativní týmy v prioritách vědeckého bádání

CZ.1.07/2.3.00/30.0055

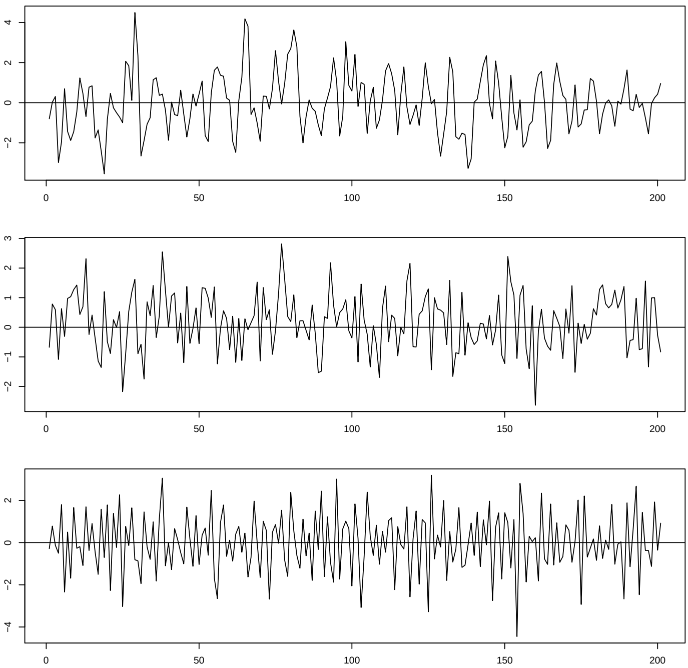
Tento projekt je spolufinancován z ESF a státního rozpočtu ČR.

Písmenkové zmatky

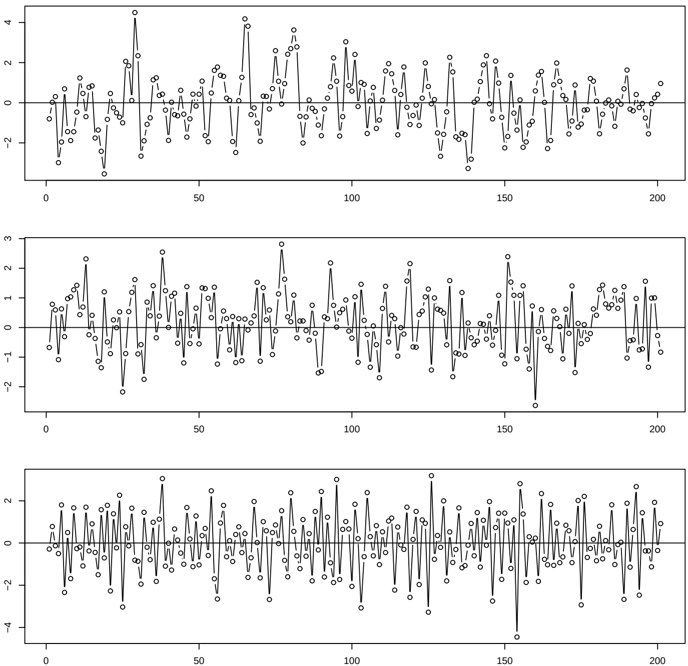
Upozornění týkající se používání velkých a malých písmen v odborné literatuře:

- statistici občas až pedantsky rozlišují symboly Y (náhodná veličina) a y (realizace náhodné veličiny; reálné číslo),
- „časovkáři“ (zvláštní druh statistiků) dosti často používají symbolu y pro náhodnou veličinu

Omluva: v této prezentaci občas použiji pro náhodnou veličinu velké a občas malé písmeno.

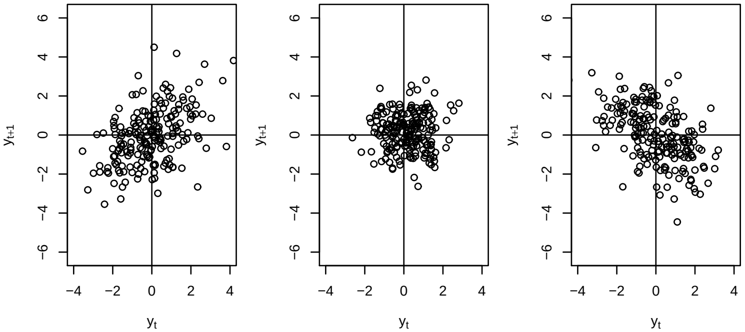


4
2
0
−2
0
50
100
150
200
3
2
1
0
−1
−2
0
50
100
150
200
2
0
−2
−4
0
50
100
150
200



4
2
0
−2
0
50
100
150
200
3
2
1
0
−1
−2
0
50
100
150
200
2
0
−2
−4
0
50
100
150
200

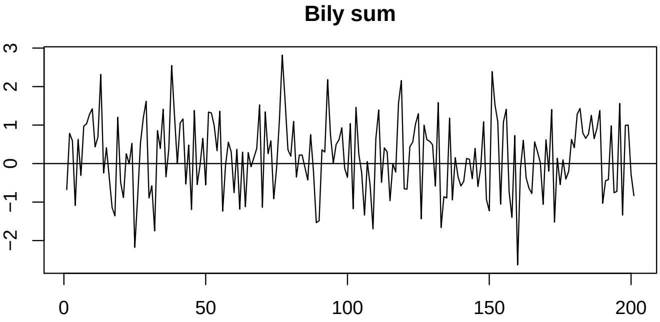
Korelace po sobě jdoucích hodnot



Bílý šum (white noise)

$$\epsilon_t, \quad t = \dots, -1, 0, 1, \dots \text{i.i.d. } \epsilon_t \sim N(0, \sigma^2),$$

tj. **nezávislé**, stejně rozdělené s nulovou stř. hodnotou.



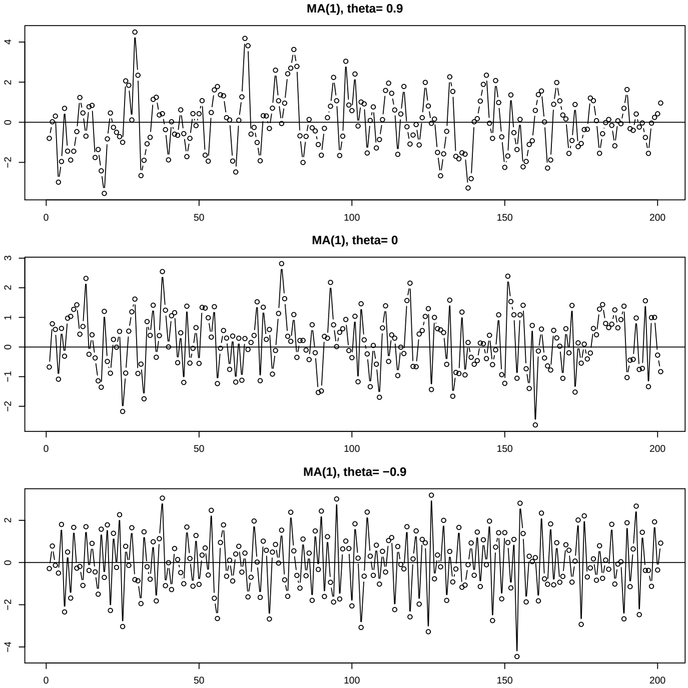
http://en.wikipedia.org/wiki/White_noise

Procesy MA(1)

Posloupností klouzavých součtů řádu 1 (Moving Average – MA) rozumíme posloupnost náhodných veličin danou předpisem

$$y_t = \epsilon_t + \theta \epsilon_{t-1},$$

kde $\{\epsilon_t\}$ je bílý šum ($\epsilon_t \sim N(0, \sigma^2)$),
$\theta \in \mathbb{R}$ je parametr procesu.


MA(1), theta= 0.9
MA(1), theta= 0
MA(1), theta= −0.9
0
50
100
150
200

Procesy MA(1): střední hodnota a rozptyl

$$\begin{aligned}
E(Y_t) &= E(\epsilon_t + \theta\epsilon_{t-1}) = E(\epsilon_t) + \theta E(\epsilon_{t-1}) = 0 + \theta \cdot 0 = 0 \\
var(Y_t) &= var(\epsilon_t + \theta\epsilon_{t-1}) = var(\epsilon_t) + var(\theta\epsilon_{t-1}) + 2cov(\epsilon_t, \theta\epsilon_{t-1}) = \\
&= var(\epsilon_t) + \theta^2 var(\epsilon_{t-1}) = \sigma^2 + \theta^2\sigma^2 = \sigma^2(1 + \theta^2)
\end{aligned}$$

Všimněte si: střední hodnota a rozptyl jsou v čase konstantní.

Procesy MA(1): korelace

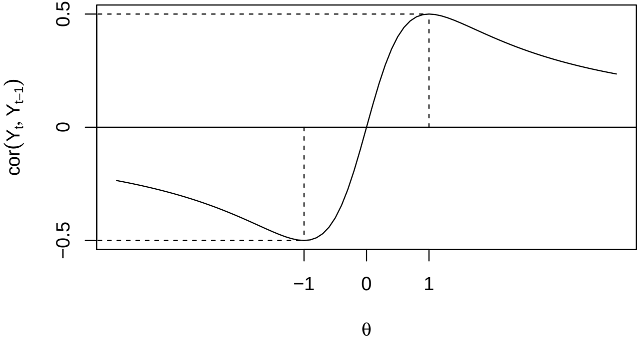
$$
\begin{aligned}
cov(Y_t, Y_{t-1}) &= cov(\epsilon_t + \theta\epsilon_{t-1}, \epsilon_{t-1} + \theta\epsilon_{t-2}) = \\
&= cov(\epsilon_t, \epsilon_{t-1}) + \theta cov(\epsilon_t, \epsilon_{t-2}) + \\
&\quad + \theta cov(\epsilon_{t-1}, \epsilon_{t-1}) + \theta^2 cov(\epsilon_{t-1}, \epsilon_{t-2}) = \\
&= 0 + 0 + \theta var(\epsilon_{t-1}) + 0 = \theta\sigma^2 \\
cov(Y_t, Y_{t-k}) &= cov(\epsilon_t + \theta\epsilon_{t-1}, \epsilon_{t-k} + \theta\epsilon_{t-1-k}) = \\
&= cov(\epsilon_t, \epsilon_{t-k}) + \theta cov(\epsilon_t, \epsilon_{t-1-k}) + \\
&\quad + \theta cov(\epsilon_{t-1}, \epsilon_{t-k}) + \theta^2 cov(\epsilon_{t-1}, \epsilon_{t-1-k}) = \\
&= 0 + 0 + 0 + 0 = 0 \text{ pro } k > 1
\end{aligned}
$$

Připomeňme, že $var(Y_t) = var(Y_{t-1}) = \sigma^2(1 + \theta^2)$, tudíž

$$
\begin{aligned}
cor(Y_t, Y_{t-1}) &= \frac{\theta}{1 + \theta^2} \\
cor(Y_t, Y_{t-k}) &= 0 \text{ pro } k > 1
\end{aligned}
$$

Všimněte si: korelace dvou po sobě jdoucích (náhodných) hodnot procesu je v čase konstantní.

Procesy MA(1): korelace



$$|cor(Y_t, Y_{t-1})| \leq 1/2$$

STACIONARITA

U procesů MA(1) jsme ukázali, že mají následující vlastnosti:

- střední hodnota je konstantní v čase, tj.
  $E(Y_t) = \mu$ pro všechna $t = \dots, -1, 0, 1, \dots$
- rozptyl je konstantní v čase, tj.
  $var(Y_t) = \sigma^2$ pro všechna $t = \dots, -1, 0, 1, \dots$
- korelace dvou hodnot náhodného procesu závisí jen na jejich vzdálenosti, nikoliv na čase, tj.
  $cor(Y_t, Y_{t+h}) = cor(Y_s, Y_{s+h})$ pro libovolné t, s a časovou vzdálenost h.

Procesy, které tyto vlastnosti mají, se nazývají slabě stacionární.

Důležitost stacionarity popisuje *Cipra* slovy: „V Box-Jenkinsově metodologii lze modelovat pouze stacionární časové řady, přičemž ovšem pomocí různých transformací (nejčastěji pomocí diferencování) je možné mnoho nestacionárních časových řad z praxe převést na stacionární.“

AUTOKORELAČNÍ FUNKCE

Autokorelační funkcí stacionární časové řady $\{Y_t, t \in \mathbb{Z}\}$ rozumíme funkci

$$\begin{array}{rcl} \rho : \mathbb{Z} & \to & [-1, 1] \\ k & \mapsto & cor(Y_t, Y_{t+k}) \end{array}$$

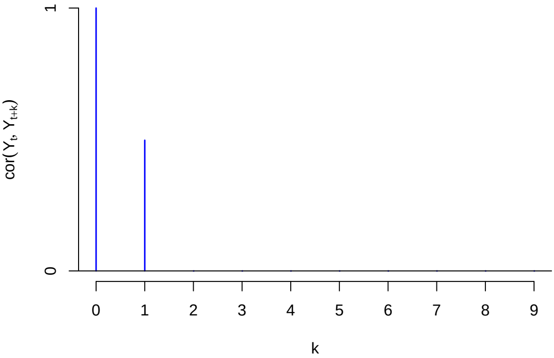
Poznámky:

- rozmyslete si, že bez předpokladu stacionarity procesu nemá autokorelační funkce smysl
- jde o sudou funkci, tj. $\rho(k) = \rho(-k)$,
  neboť $cor(Y_t, Y_{t+k}) = cor(Y_t, Y_{t-k})$ (podmínka stacionarity)
  proto stačí zkoumat autokorel. funkci jen pro $k \geq 0$
- $\rho(0) = 1$, neboť $\rho(0) = cor(Y_t, Y_t) = 1$

Autokorelační funkce procesů MA(1)

Ukázali jsme, že

- $\rho(1) = cor(Y_t, Y_{t-1}) = \frac{\theta}{1+\theta^2}$
- $\rho(k) = cor(Y_t, Y_{t-k}) = 0$ pro $k > 1$



Graf: autokorelační funkce procesu MA(1) s parametrem $\theta = 0,9$.

Procesy MA(q)

Posloupností klouzavých součtů řádu q (Moving Average – MA) rozumíme posloupnost náhodných veličin danou předpisem

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q},$$

kde $\{\epsilon_t\}$ je bílý šum ($\epsilon_t \sim N(0, \sigma^2)$),
$\theta_1 \in \mathbb{R}, \ldots, \theta_q \in \mathbb{R}$ jsou parametry procesu.

Procesy MA(q): stř. hodnota, rozptyl, autokorelační funkce

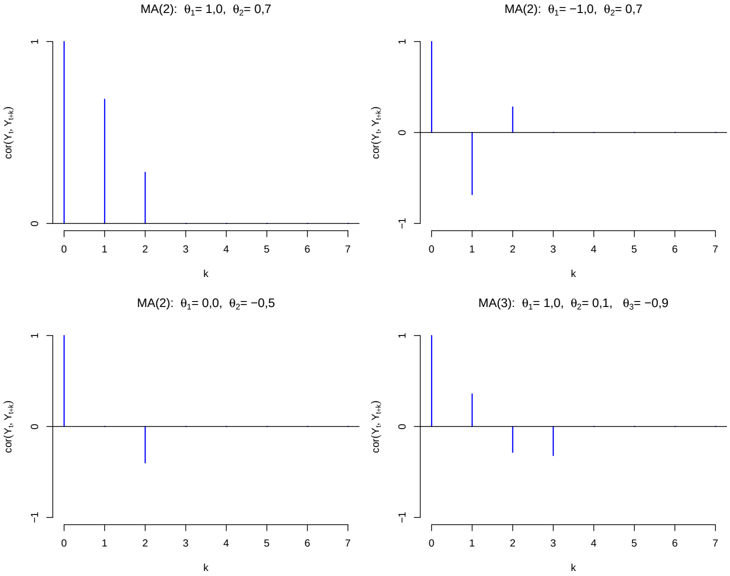
$$
\begin{aligned}
E(Y_t) &= 0 \\
var(Y_t) &= (1 + \theta_1^2 + \ldots + \theta_q^2) \cdot \sigma^2
\end{aligned}
$$

Autokorelační funkce:

$$
\begin{aligned}
\rho_k = cor(Y_t, Y_{t+k}) &= \frac{\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \ldots + \theta_q^2} \quad \text{pro } k \leq q \\
\rho_k = cor(Y_t, Y_{t+k}) &= 0 \quad \text{pro } k > q
\end{aligned}
$$

Opět vidíme, že procesy MA(q) jsou stacionární.
Odvození je jednoduché cvičeníčko – zkuste si.

Procesy MA(q): příklady autokorelačních funkcí



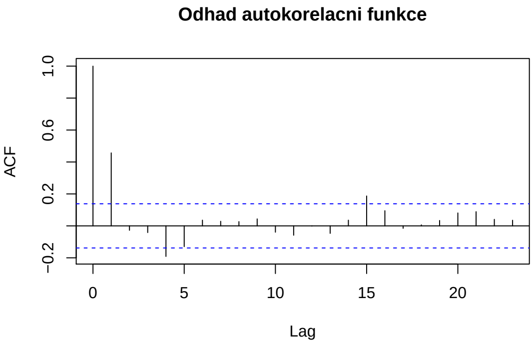
Proč je pro nás autokorel. funkce důležitá?

- Korelační koeficient umíme z dat odhadnout,
- Tedy umíme odhadnout, jak vypadá autokorelační funkce,
- Podle tvaru odhadnuté autokorel. funkce se můžeme rozhodnout, jaký model bude vhodný pro naše data.
- Hodnotu odhadu (auto)korelace můžeme využít k odhadu parametrů (θ) daného procesu.

Příklad interpretace odhadu autokorel. funkce

- Na začátku jsme viděli simulovaných 200 hodnot procesu MA(1) s parametrem $\theta = 0,9$.
- Jeho autokorelační funkce má hodnotu $90/181 \doteq 0,497$ pro $k = 1$ a je nulová pro $k > 1$.
- Představme si, že nevíme, že jde o proces typu MA(1) a neznáme hodnotu θ.

Odhadnutá autokorelační funkce:



Odhad autokorel. funkce

- Odhad autokovarianční funkce:
  $c_k = \frac{1}{n} \sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})$
- Odhad autokorelační funkce:
  $r_k = \frac{c_k}{c_0}$

Odhad autokorel. funkce – Bartlettova aproximace

Je-li $\rho_k = 0$ pro $k > k_0$, pak

$$var(r_k) \approx \frac{1}{n}\left(1 + 2\sum_{j=1}^{k_0} r_j^2\right) \quad \text{pro} \quad k > k_0$$

Využití odhadu autokorel. funkce pro odhad parametru θ

Víme

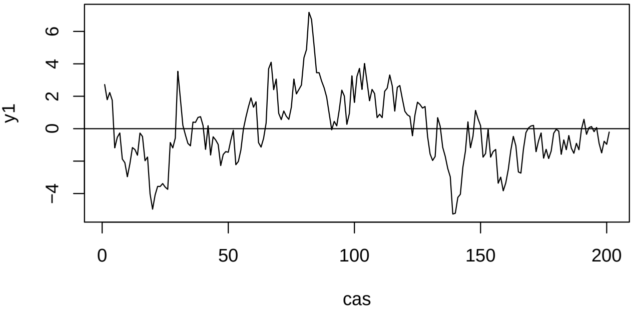
$$\rho(1) = \frac{\theta}{1+\theta^2}$$

a umíme hodnotu autokorelace odhadnout.

$$\begin{aligned} \widehat{\rho}(1) \cdot \widehat{\theta}^2 - \widehat{\theta} + \widehat{\rho}(1) &= 0 \\ \widehat{\theta} &= \frac{1 - \sqrt{1 - 4\widehat{\rho}(1)^2}}{2\widehat{\rho}(1)} \end{aligned}$$

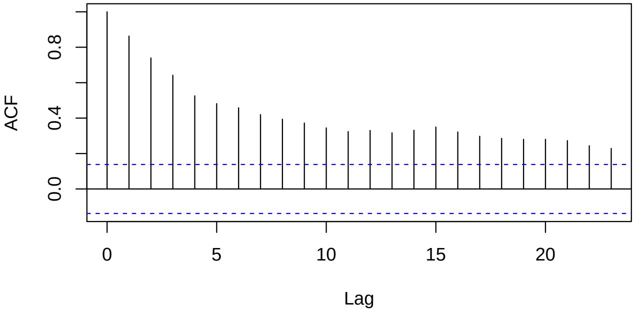
Druhý kořen odpovídá procesu, který není *invertibilní* (vysvětlíme později).

Jiný typ modelu



Autokorelační funkce předešlého procesu

(odhad)



Procesy AR(1)

Autoregresním procesem řádu 1 rozumíme posloupnost náhodných veličin $\{y_t\}$ danou předpisem

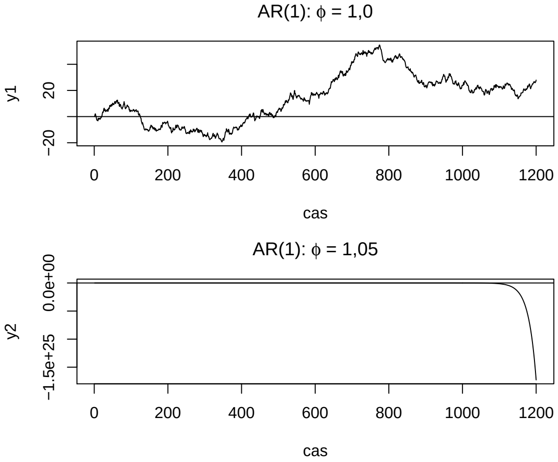
$$y_t = \phi y_{t-1} + \epsilon_t,$$

kde $\{\epsilon_t\}$ je bílý šum ($\epsilon_t \sim N(0, \sigma^2)$),
$\phi \in \mathbb{R}$ je parametr procesu.

Satcionarita procesů AR(1)

$$y_t = \phi y_{t-1} + \epsilon_t,$$

Proces AR(1) nemusí být stacionární – záleží na hodnotě parametru ϕ.



Satcionarita procesů AR(1) – neformální odvození

$$\begin{aligned}
y_t &= \phi y_{t-1} + \epsilon_t \\
y_t &= \phi(\phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
y_t &= \phi^2 y_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \\
y_t &= \phi^3 y_{t-3} + \phi^2\epsilon_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \\
&\vdots \\
y_t &= \phi^k y_{t-k} + \sum_{j=0}^{k-1} \phi^j \epsilon_{t-j}, \qquad k \in \mathbb{N}
\end{aligned}$$

Pro $|\phi| < 1$ je „modrý“ člen zanedbatelný, zbytek je proces MA, o němž víme, že je stacionární.

Procesy AR(1) jsou stacionární, pokud $|\phi| < 1$.

Stacionární procesy AR(1): střední hodnota

$$\begin{aligned}
E(y_t) &= E(\phi y_{t-1} + \epsilon_t) = \\
&= \phi E(y_{t-1}) + E(\epsilon_t) = \\
&= \phi E(y_t) + 0 \\
E(y_t)(1-\phi) &= 0 \\
E(y_t) &= 0
\end{aligned}$$

Využili jsme předpoklad stacionarity, nejprve v rovnosti $E(y_{t-1}) = E(y_t)$, poté v posledním kroku, kde $|\phi| < 1$.

Stacionární procesy AR(1): rozptyl

$$\begin{aligned}
var(y_t) &= var(\phi y_{t-1} + \epsilon_t) = \\
&= \phi^2 var(y_{t-1}) + var(\epsilon_t) + 2\phi cov(y_{t-1}, \epsilon_t) = \\
&= \phi^2 var(y_t) + var(\epsilon_t) + 0 = \\
&= \phi^2 var(y_t) + \sigma^2 \\
var(y_t)(1 - \phi^2) &= \sigma^2 \\
var(y_t) &= \frac{\sigma^2}{1 - \phi^2}
\end{aligned}$$

Pozornost si zaslouží rovnost $cov(y_{t-1}, \epsilon_t) = 0$,
uvědomme si, že y_t je definováno pomocí *minulých* hodnot procesu $\{y_t\}$ a *současné* hodnoty bílého šumu ϵ_t. Hodnota v čase $t - 1 (y_{t-1})$, tedy musí být nezávislá s hodnotou bílého šumu v čase t, který je z pohledu času $t - 1$ *budoucností*.

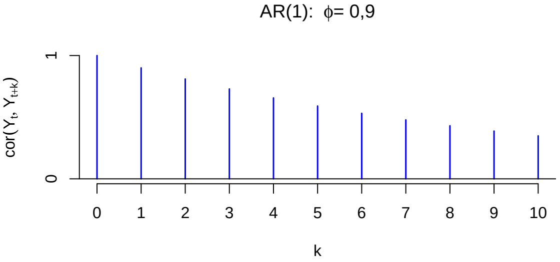
Stacionární procesy AR(1): korelace

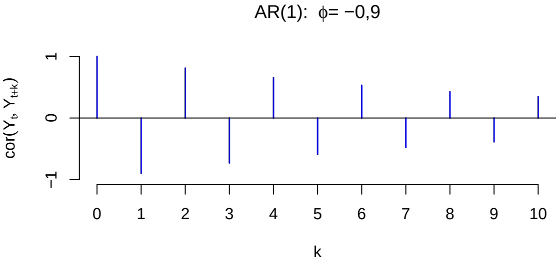
Pro výpočet použijeme vztah, který jsme odvodili o 3 slajdy dříve:

$$y_t = \phi^k y_{t-k} + \sum_{j=0}^{k-1} \phi^j \epsilon_{t-j}, \qquad k \in \mathbb{N}$$

$$\begin{aligned}
cov(y_t, y_{t-k}) &= cov(\phi^k y_{t-k} + \sum_{j=0}^{k-1} \phi^j \epsilon_{t-j}, y_{t-k}) = \\
&= \phi^k cov(y_{t-k}, y_{t-k}) + \sum_{j=0}^{k-1} \phi^j cov(\epsilon_{t-j}, y_{t-k}) = \\
&= \phi^k cov(y_{t-k}, y_{t-k}) + 0 = \\
&= \phi^k \sigma^2 \\
\rho(k) = cor(y_t, y_{t-k}) &= \phi^k
\end{aligned}$$

Stacionární procesy AR(1): autokorelační funkce





Potřeba „přímé“ korelace

$$\begin{aligned} y_t &= \phi y_{t-1} + \epsilon_t \\ y_{t-1} &= \phi y_{t-2} + \epsilon_{t-1} \end{aligned}$$

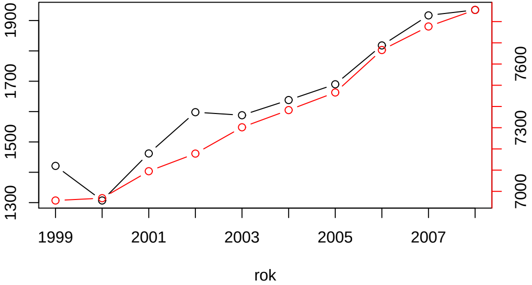
$$\begin{array}{ccc} y_t & \longleftarrow & \epsilon_t \\ \uparrow & & \\ y_{t-1} & \longleftarrow & \epsilon_{t-1} \\ \uparrow & & \\ y_{t-2} & & \end{array}$$

- Korelace mezi y_t a y_{t-2} je „přes y_{t-1}“.
- Existuje však mezi nimi „přímá“ korelace?
- KONCEPT PARCIÁLNÍ KORELACE

Připomenutí jednoho příkladu z minula

- X: počet právníků ve státě Kansas
- Y: počet úmrtí v důsledku pádu ze schodů v USA

máme k dispozici údaje za období 1999 až 2008 (10 let)
$cor(X, Y) = 0,98$



Připomenutí jednoho příkladu z minula

- X: počet právníků ve státě Kansas
- Y: počet úmrtí v důsledku pádu ze schodů v USA
- Z: počet obyvatel USA

$$\begin{array}{ccccc} X & & \overset{???}{\longleftrightarrow} & & Y \\ & \nwarrow & & \nearrow & \\ & & Z & & \end{array}$$

Je mezi X a Y nějaký jiný vztah, než že obě přímo úměrně rostou s rostoucím Z?

Odpověď na předešlou otázku

- Odhadněme (lineární) závislost X na Z
  $X = \beta_0 + \beta_1 Z + \epsilon_X$
  $\widehat{X} = \widehat{\beta_0} + \widehat{\beta_1} Z$
- Odhadněme (lineární) závislost Y na Z
  $Y = \beta_0^* + \beta_1^* Z + \epsilon_Y$
  $\widehat{Y} = \widehat{\beta_0}^* + \widehat{\beta_1}^* Z$

Studujme korelaci veličin $X - \widehat{X}$ a $Y - \widehat{Y}$, tj. zabývejme se tou částí vztahu X a Y, kterou nelze vysvětlit pomocí Z.

$$cor(X - \widehat{X}, Y - \widehat{Y}) =: \rho_{X,Y.Z}$$

Parciální korelační koeficient X a Y při daném Z.

Otázka pro Vás!

Jaký byste očekávali paricální korelační koeficient X a Y při daném Z?

- X: počet právníků ve státě Kansas
- Y: počet úmrtí v důsledku pádu ze schodů v USA
- Z: počet obyvatel USA

Jde o korelační koeficient

- rozdílu skutečného počtu právníků v Kansasu a očekávaného (dle velikosti Americké populace)
- rozdílu skutečného počtu úmrtí ... a očekávaného (dle velikosti Americké populace)

Já bych očekával hodnotu blízkou nule.

Otázka pro Vás!

Jaký byste očekávali paricální korelační koeficient X a Y při daném Z?

- X: počet právníků ve státě Kansas
- Y: počet úmrtí v důsledku pádu ze schodů v USA
- Z: počet obyvatel USA

Jde o korelační koeficient

- rozdílu skutečného počtu právníků v Kansasu a očekávaného (dle velikosti Americké populace)
- rozdílu skutečného počtu úmrtí ... a očekávaného (dle velikosti Americké populace)

Já bych očekával hodnotu blízkou nule.

Překvapení

$$\rho_{X,Y.Z} = 0,68$$

Tak že by přece jen ti právníci ...

Možné vysvětlení

- Kansas nebyl v příkladu zvolen úplně náhodně, byl to stát (z 51) s druhou nejsilnější korelací mezi X a Y.
- Korelační koeficient se pohyboval kolem hodnoty 0,91 (medián).
- Parciální korelační koeficient se pohyboval kolem hodnoty 0,19 (medián).

(Ale pořád je tu možnost, že „přece jen ti právníci ..." – že existuje kauzální vztah mezi počtem právníků v Kansasu a počtem úmrtí při pádu ze schodů v USA)

PARCIÁLNÍ AUTOKORELAČNÍ FUNKCE

Parciální autokorelační funkcí stacionární časové řady $\{Y_t, t \in \mathbb{Z}\}$ rozumíme funkci

$$\begin{aligned} \rho : \mathbb{Z} &\rightarrow [-1, 1] \\ k &\mapsto pcor(Y_t, Y_{t+k}), \end{aligned}$$

kde $pcor(Y_t, Y_{t+k})$ označuje parciální korelační koeficient veličin Y_t a Y_{t+k} při daných $Y_{t+1}, \ldots, Y_{t+k-1}$.

Parciální autokorelační funkce v procesech AR(1)

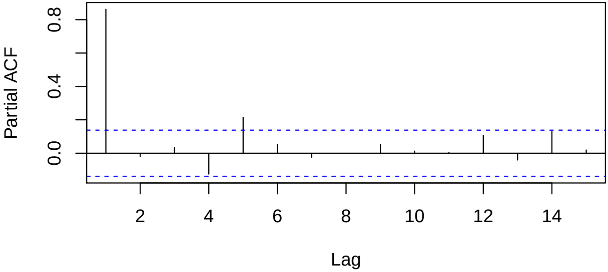
Spočtěme např.$pcor(Y_t, Y_{t-2})$

$$\begin{aligned} y_t &= \phi y_{t-1} + \epsilon_t \\ \widehat{y}_t &= \phi y_{t-1} \\ y_t - \widehat{y}_t &= \epsilon_t \end{aligned}$$

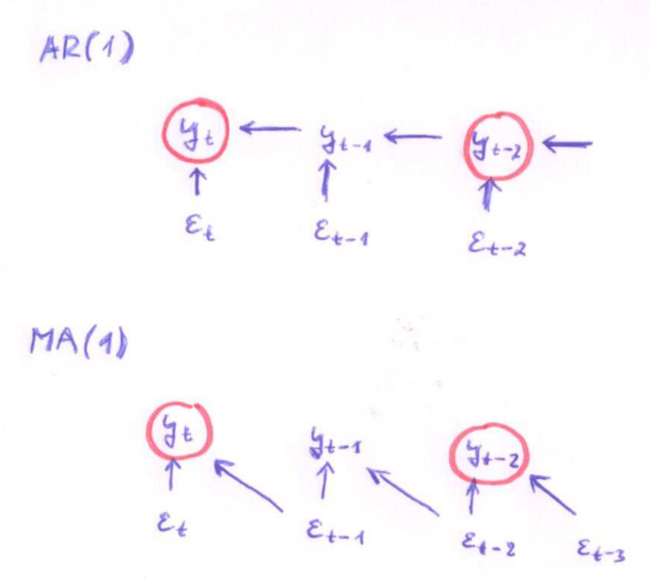
$$\begin{aligned} y_{t-1} &= \phi y_{t-2} + \epsilon_{t-1} \\ \widehat{y}_{t-2} &= \frac{1}{\phi} y_{t-1} \\ y_{t-2} - \widehat{y}_{t-2} &= -\frac{1}{\phi}\epsilon_{t-1} \end{aligned}$$

$$pcor(Y_t, Y_{t-2}) = cor(y_t - \widehat{y}_t, y_{t-2} - \widehat{y}_{t-2}) = cor(\epsilon_t, -\frac{1}{\phi}\epsilon_{t-1}) = 0$$

Parciální autokorelační funkce v procesech AR(1)



Parciální autokorelační funkce v procesech AR(1) a MA(1)



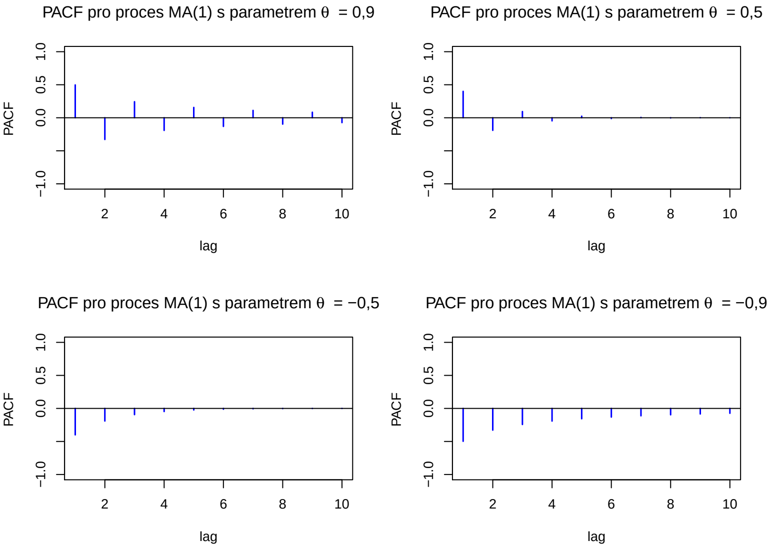
Parciální autokorelační funkce v procesech MA(1)

$$pcor(y_t, t_{t-k}) = \frac{(-1)^{k-1}\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}} \quad k = 2, 3, \dots$$

Je omezena geometricky klesající posloupností

$$|pcor(y_t, t_{t-k})| < |\theta|^k$$

Parciální autokorelační funkce v procesech MA(1)



Procesy AR(p)

Autoregresním procesem řádu p rozumíme posloupnost náhodných veličin $\{y_t\}$ danou předpisem

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t,$$

kde $\{\epsilon_t\}$ je bílý šum ($\epsilon_t \sim N(0, \sigma^2)$),
$\phi_1 \in \mathbb{R}, \ldots, \phi_p \in \mathbb{R}$ jsou parametry procesu.

Procesy AR(p): podmínka stacionarity, stř. hodnota, rozptyl

$$
\begin{aligned}
E(Y_t) &= 0 \\
var(Y_t) &= \frac{\sigma^2}{1 - \phi_1\rho_1 - \ldots - \phi_p\rho_p}
\end{aligned}
$$

Podmínka stacionarity: všechna řešení rovnice

$$1 - \sum_{j=1}^{p} \phi_j x^j = 0$$

musí ležet vně jednotkového kruhu.
(pro $p = 1$ má rovnice tvar $1 - \phi_1 x = 0$, kořen je $1/\phi_1$, podmínka je tedy $|1/\phi_1| > 1$, tj/ $|\phi_1| < 1$, viz dříve)

Procesy AR(p): autokorelační funkce

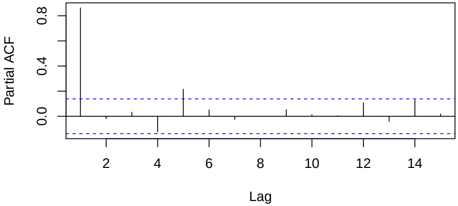
$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t \\
y_t y_{t-k} &= \phi_1 y_{t-1} y_{t-k} + \phi_2 y_{t-2} y_{t-k} + \ldots + \phi_p y_{t-p} y_{t-k} + \epsilon_t y_{t-k} \\
E y_t y_{t-k} &= \phi_1 E y_{t-1} y_{t-k} + \phi_2 E y_{t-2} y_{t-k} + \ldots + \phi_p E y_{t-p} y_{t-k} + E\epsilon_t y_{t-} \\
\rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}
\end{aligned}
$$

pro $k \in \mathbb{N}$
Soustava prvních p těchto rovnic ($k = 1, \ldots, p$) se nazývá *Yuleova-Walkerova soustava rovnic*, jejím řešením získáme hodnoty parametrů $\phi_1, \ldots, \phi_p$ vyjádřené pomocí hodnot autokorelační funkce $\rho_1, \ldots, \rho_p$, které umíme odhadnout.

Procesy AR(p): parciální autokorelační funkce

$$\rho_{kk} = 0 \qquad \text{pro } k > p$$

Odhad PACF pro data simulovaná z procesu AR(1) s parametrem $\phi = 0,9$:



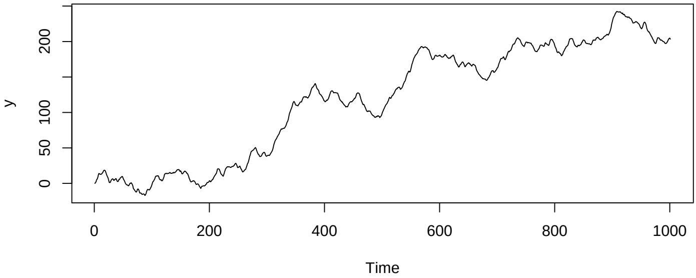
Procesy ARMA(p, q)

Smíšeným procesem řádu p a q rozumíme posloupnost náhodných veličin $\{y_t\}$ danou předpisem

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q},$$

kde $\{\epsilon_t\}$ je bílý šum ($\epsilon_t \sim N(0, \sigma^2)$),
$\phi_1 \in \mathbb{R}, \ldots, \phi_p \in \mathbb{R}$ a $\theta_1 \in \mathbb{R}, \ldots, \theta_q \in \mathbb{R}$ jsou parametry procesu.

Box, Jenkins a nestacionarita



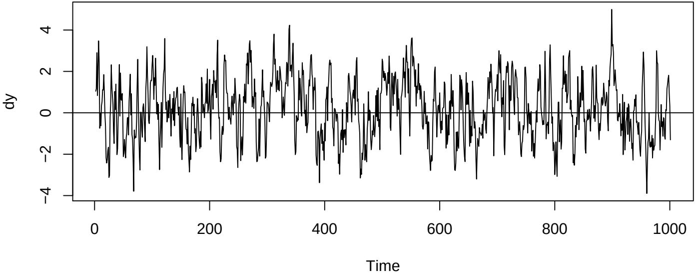
Procesy ARIMA(p, d, q)

diferencování procesu:

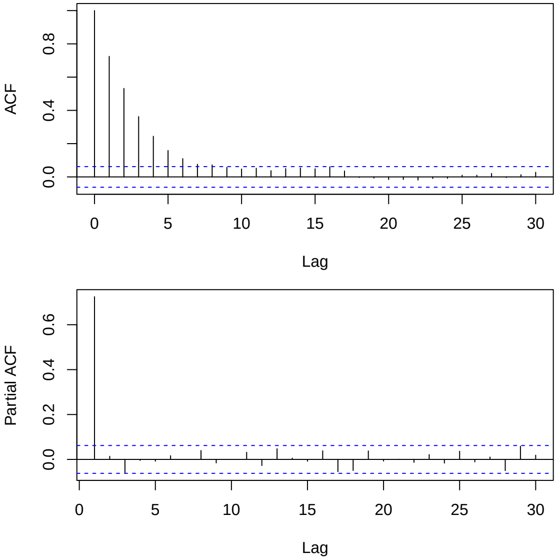
$$
\begin{aligned}
\Delta y_t &= y_t - y_{t-1} \\
\Delta^2 y_t &= \Delta(\Delta y_t) = \Delta y_t - \Delta y_{t-1} = \\
&= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \\
\dots
\end{aligned}
$$

Proces $\{y_t\}$, jehož d-tá diference je smíšeným procesem ARMA(p, q) se označuje jako proces ARIMA(p, d, q).

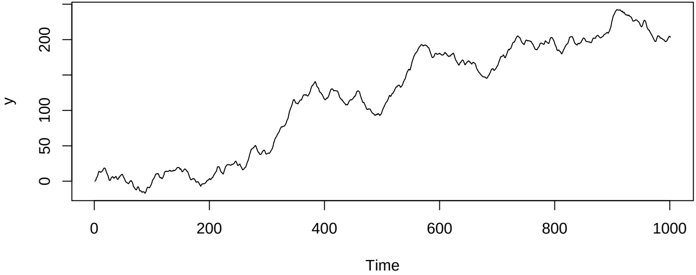
Po 1. diferencování



ACF a PACF diferencované řady



Tajenka



Řada byla generována pomocí generátoru náhodných čísel jako proces ARIMA(1,1,0), kde parametr $\phi = 0,7$.

Box, Jenkins a sezónnost

$$\begin{array}{cccc} y_1 & y_{L+1} & y_{2L+1} & \cdots \\ y_2 & y_{L+2} & y_{2L+2} & \cdots \\ \vdots & \vdots & \vdots & \\ y_L & y_{2L} & y_{3L} & \cdots \end{array}$$

- Dlouhodobý aspekt
  $y_t = \phi_1 y_{t-L} + \phi_2 y_{t-2L} + \ldots + \phi_P y_{t-PL} + \eta_t + \Theta_1 \eta_{t-L} + \ldots + \Theta_Q \eta_{t-QL}$
- Krátkodobý aspekt
  $\eta_t = \varphi_1 \eta_{t-1} + \varphi_2 \eta_{t-2} + \ldots + \varphi_p \eta_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}$

$\{\epsilon_t\}$ je bílý šum

SARIMA $(p, 0, q) \times (P, 0, Q)_L$
obecněji SARIMA $(p, d, q) \times (P, D, Q)_L$