Idősorelemzés gyakorlat

Biztosítási és pénzügyi matematika mesterszak
Alkalmazott matematikus mesterszak
2017/2018 tavaszi félév

Játékszabályok

- A gyakorlatokról maximum 3-szor lehet hiányozni. Aki többször hiányzik, nem kap gyakjegyet.
- 100 + x pontot lehet szerezni a félév során:
  - 20 pont: számolós beadandó feladatok (4 pont · 5)
  - 30 pont: számítógépes/szimulációs beadandó feladatok (15 pont · 2)
  - 50 pont: ZH: V.15., É 0.79 Jánossy Lajos terem
  - x pont: szorgalmi feladatok
- A ZH-n minimálisan teljesíteni kell 30 %-ot. Ha a ZH sikertelen, nem írod meg, vagy javítani szeretnél, akkor vizsgaidőszak első hetén lesz lehetőség a pótZH megírására vagy a javításra. A jobbik eredményt veszem figyelembe, azaz nem lehet rontani. Két sikertelen vagy meg nem írt ZH esetén gyakUV-t kell írnod, a beadandókért kapott pontok megmaradnak. GyakUV írása esetén legfeljebb 2-est lehet szerezni.
- A ZH-kon használható: **számológép** és egy legfeljebb A4-es méretű lapra KÉZZEL írott "puska".
- Számolós beadandók: Mindegyik maximálisan 4 pontot ér, a legjobb 5-öt veszem figyelembe. A beadandóknál több feladatot is kihirdetek, amik közül ízlés szerint válogathattok. A beadandók célja, hogy folyamatosan tanuljatok, gyakoroljatok, ezért fix határidőig lehet őket benyújtani.
- Számítógépes/szimulációs beadandók: kettőt hirdetek ki fix határidővel, mindegyikkel maximálisan 15 pontot lehet szerezni.
- A beadandóknál nem tilos, sőt, bizonyos fokig még kívánatos is a közös munka/konzultáció hallgatótársaiddal – az viszont elvárás, hogy **a gondolataidat önállóan írd le**! Amennyiben nyilvánvaló másolás gyanúja merül fel, az érintettek 0 pontban részesülnek.
- Osztályozás:

elégtelen (1)	0	–	34,99
elégséges (2)	35	–	49,99
közepes (3)	50	–	64,99
jó (4)	65	–	79,99
jeles (5)	80	–	∞^3

Infók a gyakvezetőről

Név	Varga László, *óraadó*
Munkahely	Morgan Stanley, Risk Management
Tanszék	Valószínűségelméleti és Statisztika Tanszék (ELTE TTK)
E-mail	`vargal4@cs.elte.hu`
Honlap	`vargal4.elte.hu`

Ajánlott irodalom

- Shumway-Stoffer: Time series analysis and its applications, with R examples
  Egy lebutított verziója ingyenesen letölthető: `http://www.stat.pitt.edu/stoffer/tsa4/tsaEZ.pdf`
- Brockwell-Davis: Introduction to time series and forecasting

---

Szimulációkhoz használt szoftver/programnyelv: **R**

- Statisztikai modellezésre, adatok elemzésére kiválóan alkalmas programnyelv
- Nyílt forráskódú, ma már alig van probléma, feladat, aminek a megoldására ne lenne valamilyen package – akár több is
- Jelenleg a legelterjedtebb matematikai célú programnyelv
- Letöltési helye: `https://cran.r-project.org/`
- Szövegszerkesztésre ajánlott szoftver: RStudio; letöltési helye: `https://www.rstudio.com/products/rstudio/download3/`

1.) Keressük meg "kézzel" és **R** segítségével a legjobb (legkisebb négyzetes) becslést (a, b és c paraméterek)! Ábrázoljuk az adatpontokat és az illesztett görbét!

	Adatok	Modell
a.)	x: -1, 1, 2; y: 1, 2, -1	$y = ax + b$
b.)	x: -1, 0, 1, 2; y: 8, 8, 4, 16	$y = ax^2 + bx + c$
c.)	x: 0, $\frac{1}{2}$, 1; y: 1, 3, 7	$y = a \cdot \cos(\pi x) + b \cdot \sin(\pi x)$
d.)	x: 0, -1, 1, 1; y: 1, 0, -1, 1	$a \cdot (x^2 + y^2) + b \cdot (x + y) = 1$

2.) Olvassuk be a `kerdoiv.txt` fájlt, ami egy 2017-es hallgató kérdőíves felmérés adatait tartalmazza. A következőkre válaszoltak: nem, testmagasság (cm), súly (kg), cipőméret, hányast szerzett valszámból a 2017-es vizsgán, hány percet utazik az egyetemre, szorgalmi időszakban átlagosan hány órát tanul egy héten.

a.) Nézzük meg pontdiagrammal néhány adatpár közti összefüggést (pl. magasság és súly, nem és cipőméret, stb.)!

b.) A továbbiakban célunk a testmagasság modellezése/magyarázása a többi változó segítségével. Tekintsük az alábbi regressziós modelleket:

I.) Testmagasság = Testsúly + Hiba , ami a
Testmagasság = $a_0 + a_1 \cdot$ Testsúly + Hiba modell rövidített változata

II.) Testsúly = Testmagasság + Hiba

III.) Testmagasság = Testsúly + Lábméret + Hiba

IV.) Testmagasság = Nem + Hiba

Vizsgáljuk meg a korrelációs mátrixot! Keressük meg a legjobban illeszkedő modellt!

c.) Adjunk előrejelzést a legjobbnak tűnő modell(ek) alapján egy olyan fiú hallgató testmagasságára, aki 70 kg-os, 45-ös a cipőmérete, 5-öse volt valszámból, 25 percet utazik az egyetemre és heti 12 órát tanul!

SZ1.) Ha sok magyarázóváltozónk van, akkor a "legjobb" lineáris regressziós modell megtalálása rendszerint nehéz feladat. Ennek automatizálására több eljárást kidolgoztak, ezek egyike található a szláv mitológia erdő- és vadászistenéről elnevezett `Boruta` package-ben. Használd a package `Boruta` függvényét a `kerdoiv.txt` fájlban található hallgatói kérdőív adatbázis esetén a legjobb lineáris modell megtalálására – dolgozd fel a következő weblap példáját: `https://www.analyticsvidhya.com/blog/2016/03/select-important-variables-boruta-package`. Miért egy erdőistenről nevezték el a package-et? A futtatásaid reprodukálásához szükséges fájlokat küldd el az E-mail címemre! (2p)

SZ2.) A `Crime.xlsx` fájl amerikai nagyvárosokra vonatkozóan tartalmaz bűnözési adatokat, az oszlopok fejléce (a változók pontos megnevezése) pedig a `CrimeHeader.txt` fájlban található. Olvasd be az adatokat valamilyen alkalmas módszerrel, majd az xlsx fájl 1. oszlopára (1 millió főre jutó összes jelentett bűnesetek száma) próbálj meg minél jobb regressziós modellt meghatározni a többi változó segítségével!

A beolvasáshoz követhető módszerek például:

- csv kiterjesztésű fájlt csinálsz az xlsx-ből, majd használod a `read.csv` parancsot;
- letöltesz egy olyan package-et, amiben van xlsx beolvasására alkalmas parancs.

A futtatásaid reprodukálásához szükséges fájlokat küldd el az E-mail címemre! Kommentekben vagy az E-mail szövegében értékeld a modellt józan paraszti ész alapján – ezt vártad-e; mennyire jó a modell; van-e még olyan kimaradt változó, amit magyarázóváltozóként figyelembe kellene venni? (2p)

Mostantól amennyiben az adott feladat nem jelzi külön, t mindig az egész számok halmazát futja be, azaz $t \in \mathbb{Z}$.

3.) Legyen X_t *véletlen bolyongás drift*-tel: $X_t = \delta + X_{t-1} + \varepsilon_t \quad (t = 1, 2, \ldots)$, ahol δ valós paraméter, $P(X_0 = 0) = 1$ és $\varepsilon_t \sim WN(0, \sigma^2)$.

a.) Fejezzük ki X_t-t a fehér zaj folyamat segítségével!

b.) Határozzuk meg X_t várható érték, autokovariancia és autokorreláció függvényét! Gyengén stacionárius a folyamat?

c.) Szimuláljunk egy ilyen folyamatot normális innovációk, $\delta = -1, 0, 1$, valamint $\sigma = 1, 5, 10$ esetén, majd ábrázoljuk X_t-t, a trendet és az autokorreláció függvényét!

d.) Van-e olyan $\varphi(X_t, X_{t-1}, \ldots)$ transzformáció, amivel származtatott folyamat már stacionárius?

e.) Tegyük fel, hogy a ε_t fehér zaj folyamat normális eloszlású. Határozd meg a δ paraméter és a fehér zaj σ^2 szórásnégyzet paraméterének ML-becslését az első n megfigyelés, mint n elemű minta alapján! Torzítatlan, illetve konzisztens a δ-ra adott becslés?

4.) Legyen $X_t = \frac{1}{4}\varepsilon_{t-1} + \frac{1}{2}\varepsilon_t + \frac{1}{4}\varepsilon_{t+1} \quad (t = 1, 2, \ldots)$, ahol ε_t fehér zaj.

a.) Határozzuk meg X_t várható érték, autokovariancia és autokorreláció függvényét! Gyengén stacionárius a folyamat?

b.) Szimuláljunk egy fehér zajt, készítsük el hozzá X_t-t, majd ábrázoljuk a két idősort közös diagramban! Értelmezzük a látottakat!

5.) Legyen $X_t = \alpha + \beta t + \varepsilon_t \quad (t = 1, 2, \ldots)$, ahol α, β valós paraméterek és ε_t fehér zaj.

a.) Határozzuk meg X_t várható érték, autokovariancia és autokorreláció függvényét! Gyengén stacionárius a folyamat?

b.) Van-e olyan $\varphi(X_t, X_{t-1}, \ldots)$ transzformáció, amivel származtatott folyamat már stacionárius?

6.) Legyen $(X_t)_{t\in\mathbb{R}} = U\sin(2\pi\alpha t) + V\cos(2\pi\alpha t)$, ahol α valós paraméter, U és V egymástól független, 0 várható értékű és τ szórású val. változók.

a.) Határozzuk meg X_t várható érték, autokovariancia és autokorreláció függvényét! Gyengén stacionárius a folyamat?

b.) Legyen $Y_t = X_t + \varepsilon_t$, ahol ε_t fehér zaj. Szimuláljunk egy fehér zajt $D\varepsilon_t = 1, 2, 4, 8$ esetén, készítsük el hozzá Y_t-t, ha $\tau = 2$ és $\alpha = 1$, majd ábrázoljuk X_t és Y_t idősorokat! Értelmezzük a látottakat!

7.) Legyen $X_t = \sin(2\pi U t) \quad (t = 1, 2, \ldots)$, ahol $U \sim E(0; 1)$. Mutassuk meg, hogy X_t folyamat gyengén stacionárius, de erősen *nem* stacionárius!

Útmutatás az erős stacionaritás vizsgálatához: számoljuk ki a következő két valószínűséget: $P\left(X_1 < -\frac{\sqrt{3}}{2}, X_2 < -\frac{\sqrt{3}}{2}\right)$ és $P\left(X_2 < -\frac{\sqrt{3}}{2}, X_3 < -\frac{\sqrt{3}}{2}\right)$!

B1.) **[III.6.]** Tekintsük a következő $(X_t)_{t=1,2,\ldots}$ folyamatot: $X_t = ae^{X_{t-1}} + \varepsilon_t$, ahol a valós paraméter, ε_t normális eloszlású fehér zaj 0 várható értékkel és ismeretlen σ szórással, $X_0 = x_0 \in \mathbb{R}$. Vezesd le az ismeretlen a paraméter ML-becslését!

B2.) **[III.6.]** Tekintsük a következő $(X_t)_{t\in\mathbb{Z}}$ folyamatot: $X_t = \varepsilon_t \eta_t$, ahol ε_t és η_t egymástól és időben is független folyamatok, ε_t fehér zaj 0 várható értékkel és σ szórással, $P(\eta_t = 0) = P(\eta_t = 2) = \frac{1}{2}$ $\forall t$-re. Gyengén stacionárius az X_t folyamat? Határozd meg az autokovariancia és autokorreláció függvényét!

SZ3.) Legyen $(X_t)_{t\in\mathbb{R}}$ stacionárius Gauss-folyamat (azaz minden $k \in \mathbb{Z}_+$-ra és $t_1, \ldots, t_k \in \mathbb{R}$-re $(X_{t_1}, \ldots, X_{t_k})$ együttesen normális eloszlású) és $Y_t = e^{X_t}$.

a.) Határozd meg az X_t folyamat momentumgeneráló függvényét, azaz $M_{X_t}(s) = E(e^{sX_t})$-t!

b.) Mutasd meg, hogy az Y_t folyamat gyengén stacionárius! (1+1= 2p)

SZ4.) Legyen $X_t = at + W_t \ (0 \le t \in \mathbb{R})$, ahol a ismeretlen valós paraméter, W_t Wiener-folyamat. Határozd meg az a paraméter ML-becslését egy azonos időközönként rendelkezésünkre álló n elemű minta alapján, azaz legyen $\delta > 0$ fix, $t_i - t_{i-1} = \delta \ (i = 1, 2, \ldots, n)$, $t_0 = 0$; a tapasztalati mintánk pedig $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$

– azonos időközönként "nézünk rá" a folyamatra. (2p)
Útmutatás: Könnyen látható, hogy $X_{t_j} = X_{t_{j-1}} + a\delta + W_{t_j} - W_{t_{j-1}}$ $(j = 1, 2, \ldots, n)$, ebből pedig már megállapítható az $X_{t_j} | X_{t_{j-1}}$ feltételes eloszlás.

8.) Legyenek X_t és Y_t egymással korrelálatlan stacionárius folyamatok R_X és R_Y autokovariancia függvénnyel, valamint F_X és F_Y spektrális eloszlásfüggvénnyel. Mutassuk meg, hogy ekkor a $Z_t = X_t + Y_t$ folyamat autokovariancia függvényére $R_Z(h) = R_X(h) + R_Y(h)$ és spektrális eloszlásfüggvényére $F_Z(x) = F_X(x) + F_Y(x)$ teljesül minden $h \in \mathbb{Z}$ és $x \in \mathbb{R}$ esetén.

9.) Határozzuk meg a következő folyamatok spektrális sűrűségfüggvényét, ha pedig nincsen, akkor a spektrális eloszlásfüggvényét:

a.) fehér zaj folyamat σ szórással;
b.) AR(1) folyamat: $X_t = \alpha X_{t-1} + \varepsilon_t$, ahol $|\alpha| < 1$ valós paraméter, ε_t fehér zaj σ szórással;
c.) MA(1) folyamat: $X_t = \varepsilon_t + \alpha\varepsilon_{t-1}$, ahol α valós paraméter, ε_t fehér zaj σ szórással;
d.) $X_t = U\sin(\alpha t) + V\cos(\alpha t)$, ahol α valós paraméter, U és V egymástól független, 0 várható értékű és τ szórású valószínűségi változók;
e.) $X_t = S_t + \alpha S_{t-D} + N_t$, ahol S_t és N_t stacionárius és egymástól független folyamatok 0 várható értékkel és f_S, valamint f_N spektrális sűrűségfüggvényekkel; D ismert pozitív egész szám; α ismeretlen valós paraméter.

Tanuljunk meg ábrákról olvasni - szimuláljunk az egyes folyamatokból különböző paraméterértékek esetén, ábrázoljuk a mintát, az autokorreláció függvényt, valamint a spektrális sűrűségfüggvényt (eloszlásfüggvényt)!

10.) Legyen $X_t = U\sin\left(\frac{\pi}{3}t\right) + V\cos\left(\frac{\pi}{3}t\right)$, ahol U és V egymástól független, 0 várható értékű és τ szórású valószínűségi változók; $Y_t = \varepsilon_t + 2,5\varepsilon_{t-1}$, ahol ε_t fehér zaj σ szórással; továbbá legyen $Z_t = X_t + Y_t$. Határozd meg a Z_t folyamat autokovariancia függvényét és a spektrális eloszlásfüggvényét!

11.) Döntsük el, hogy az alábbi egész számokon értelmezett függvények lehetnek-e egy stacionárius folyamat autokovariancia függvényei:

a.) $R(h) = I(h = 0)$;
b.) $R(h) = I(h = 0) - 0,5 \cdot I(|h| = 2) - 0,25 \cdot I(|h| = 3)$;

12.) Legyen egy stacionárius folyamat spektrális sűrűségfüggvénye $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in [-\pi; \pi]$. Határozd meg a folyamat autokovariancia és autokorreláció függvényét!
Útmutatás: integráljuk parciálisan az autokovariancia függvény deriváltját, majd oldjuk meg az így adódó $R'(h) = -hR(h)$ differenciálegyenletet!

B3.) [III.20.] Legyen $X_t = \sum_{i=1}^{N} a_i \cos(\omega_i t + U_i)$, ahol minden $i = 1, 2, \ldots, N$ esetén $\omega_i \neq 0$ és a_i valós konstansok, $U_i \sim E(0; 2\pi)$ függetlenek. Határozd meg az X_t folyamat spektrális eloszlásfüggvényét/sűrűségfüggvényét!

SZ5.) Az egész számokon értelmezett $R(h) = I(h = 0) + cI(|h| = 1)$ függvény a c valós paraméter mely értékei esetén lehet egy stacionárius folyamat autokovariancia függvénye? (1p)

SZ6.) Legyenek X_t és Y_t stacionárius, egymástól független, 0 várható értékű folyamatok f_X és f_Y spektrális sűrűségfüggvényekkel, továbbá legyen $Z_t = X_t \cdot Y_t$. Mutasd meg, hogy a Z_t folyamat spektrális sűrűségfüggvénye a következőképp számolható: $f_Z(x) = \int_{-\pi}^{\pi} f_X(x - y) f_Y(y)\, dy$. (2p)

SZ7.) Egy stacionárius X_t folyamat spektrális sűrűségfüggvénye a következő: $f(x) = 100 \cdot I\left(-\frac{\pi}{6} - 0,01 < x < -\frac{\pi}{6} + 0,01\right) + 100 \cdot I\left(\frac{\pi}{6} - 0,01 < x < \frac{\pi}{6} + 0,01\right)$. Határozd meg a folyamat lag 1 (egylépéses késleltetésű) autokorrelációját, azaz az $r(1) = \mathrm{cor}(X_t, X_{t+1})$ mennyiséget! (2p)

13.) Vizsgáljuk meg a különbség-előjel próbát! Legyen $X_1, \ldots, X_n$ minta egy folyamatból, és legyen Z_n annak a száma, ahányszor a folyamat értéke növekszik, azaz $Z_n = \sum_{i=2}^{n} Y_i$, ahol $Y_i = I(X_i > X_{i-1})$, $i = 2, 3, \ldots, n$.

a.) Az összefüggő valószínűségi változó sorozatokra vonatkozó CHT segítségével mutassuk meg, hogy $\frac{Z_n - \frac{n-1}{2}}{\sqrt{\frac{n-1}{12}}} \xrightarrow[n\to\infty]{d} N(0; 1)$!
b.) Generálj i.i.d. sorozatot $n = 10, 50, 100, 200, 500, 1000, 2000$ mintanagyságok esetén, majd szimulációk alapján becsüld meg az elsőfajú hiba valószínűségét!
c.) Generálj
- AR(1) folyamatból;
- driftes véletlen bolyongásból

mintákat $n = 10, 50, 100, 200, 500, 1000, 2000$ mintanagyságok esetén, majd szimulációk alapján becsüld meg a különbség-előjel próba erejét!

14.) Tegyük fel, hogy ε_t független értékű zaj 0 várható értékkel és $\sigma_\varepsilon = 1$ szórással. Bizonyítsuk be, hogy az alábbi folyamatoknak van stacionárius $MA(\infty)$ reprezentációjuk, továbbá, hogy invertálhatók (azaz létezik $AR(\infty)$ reprezentációjuk)!

a.) $X_t = 2 + 0,9X_{t-1} + \varepsilon_t$
b.) $X_t = 0,2X_{t-1} + 0,35X_{t-2} + \varepsilon_t$
c.) $X_t = 0,5X_{t-1} - 0,25X_{t-2} + \varepsilon_t$
d.) $X_t = X_{t-1} - 0,25X_{t-2} + \varepsilon_t$
e.) $X_t = \varepsilon_t + 0,5\varepsilon_{t-1}$
f.) $X_t = 0,5X_{t-1} + \varepsilon_t - 0,5\varepsilon_{t-1}$.
g.) $X_t = 0,9X_{t-1} + \varepsilon_t + 0,5\varepsilon_{t-1}$.

Határozzuk meg a következőket:

- a stacionárius megoldás várható értéke, szórása, autokorreláció-függvénye és parciális autokorreláció-függvénye;
- a folyamat $MA(\infty)$ reprezentációja;
- a folyamat spektrális sűrűségfüggvénye.

Szimuláljunk ilyen folyamatokat **R**-rel! Határozzuk meg **R** függvények segítségével is az autokorrelációkat, parciális autokorrelációkat és ábrázoljuk a spektrális sűrűségfüggvényt!

15.) Legyen $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t$, ahol ε_t véges szórású i.i.d. zaj. Bizonyítsuk be, hogy a folyamatnak akkor és csak akkor van stacionárius megoldása, ha $\alpha_1 + \alpha_2 < 1$, $\alpha_2 - \alpha_1 < 1$ és $|\alpha_2| < 1$ teljesül!

16.) Egy ARMA folyamat spektrális sűrűségfüggvénye a következő: $f(x) = \frac{2+2\cos x}{\pi(1{,}5625-2{,}5\cos x+\cos^2 x)}$. Határozd meg a folyamat rendjét és paramétereit!

17.) Generáljunk egy zérus várható értékű, $\alpha_1 = 0,5$ és $\alpha_2 = 0,3$ paraméterű AR(2) folyamatból $n = 200$ elemű mintákat.

a.) Becsüljük vissza a minták paramétereit, majd végezzünk modelldiagnosztikát – vizsgáljuk meg, mennyire jó a becslés, illetve mennyire jó a modell!

b.) Játsszunk el azzal, hogy más modelleket illesztünk (például AR(1), AR(3), ARMA(1,1)), honnan vesszük észre, hogy ezek nem lesznek jók – ha egyáltalán észrevesszük.

c.) Vizsgáljuk meg szimulációval, hogy miként teljesít az AIC és a BIC információs kritérium a helyes rend kiválasztásában! Összehasonlításként nézzük meg az AIC/BIC értékeket, ha AR(1), AR(3), AR(4), ARMA(1,1) modelleket illesztünk a fenti AR(2)-ből generált mintákra! Mi történik, ha növeljük a mintaelemszámot?

18.) Elemezzük és illesszünk alkalmas ARMA modellt a 'gnp96.txt' fájlban található idősorra, ami az USA negyedéves, szezonális hatásoktól megszabadított reál (1996-os árakon számított) GNP adatait tartalmazza 1947 és 2002 között.

a.) Ábrázoljuk az idősort, és nézzük meg az ACF és PACF függvényét! Stacionáriusnak tűnik az idősor?

b.) Alkalmas transzformációval (logdifferenciák képzése) és/vagy regressziós modell illesztésével próbáljuk meg stacionáriussá tenni az idősort!

c.) A transzformált idősor ACF és PACF ábráinak szemrevételezése után próbáljunk meg alkalmas ARMA modellt illeszteni! AIC/BIC kritériumok alapján válasszuk ki a legjobbnak tűnő ARMA modellt, majd végezzünk modelldiagnosztikát (szignifikánsak-e a becsült paraméterek, a reziduálisok származhatnak-e fehér zaj folyamatból)!

19.) Az előző példához hasonlóan elemezzük és illesszünk alkalmas ARMA modellt az 'astsa' package-ben lévő 'gtemp' idősorra, ami 1880 és 2009 között a Föld átlaghőmérsékletét tartalmazza.

B4.) **[III.27.]** Legyen $X_t = 0,3X_{t-1} + 0,04X_{t-2} + \varepsilon_t$, ahol ε_t fehér zaj 0 várható értékkel és σ_ε szórással, $0 < \mu \in \mathbb{R}$, $t \in \mathbb{Z}$.

a.) Van a folyamatnak stacionárius $MA(\infty)$ reprezentációja?

b.) Invertálható a folyamat (azaz létezik $AR(\infty)$ reprezentációja)?

c.) Állítsd elő a folyamat $MA(\infty)$ reprezentációját! (1+1+2= 4p)

B5.) **[III.27]** Egy oksági AR(2) folyamat esetén előfordulhat, hogy az autokorrelációk az $r(1) = 1/2$; $r(2) = 1/6$; $r(3) = 1/12$ értékeket veszik fel?

SZ8.) Bizonyítsd be, hogy a fordulópont próba (turning point test) próbastatisztikája eloszlásban a standard normális eloszlásoz tart, amennyiben a minta i.i.d. fehér zajból származik, azaz teljesül a nullhipotézis! (2p)

SZ9.) Legyen $X_t = 0,4X_{t-1} + 0,45X_{t-2} + \varepsilon_t + \varepsilon_{t-1} + 0,25\varepsilon_{t-2}$, ahol ε_t fehér zaj. Van stacionárius "megoldása" (MA reprezentációja) a folyamatnak? Ha van, akkor állítsd elő! (2p)

SZ10.) Mutasd meg, hogy az $X_t = \varepsilon_t + \beta\varepsilon_{t-1}$ MA(1) folyamat parciális autokorreláció függvénye $\rho_k = \frac{-(-\beta)^k}{1+\beta^2+\ldots+\beta^{2k}}$. (2p)

SZ11.) Legyen $X_1, \ldots, X_n$ fehér zajból származó minta. Mutasd meg, hogy

a.) a Bartlett-tételben (lásd elméleti összefoglaló) szereplő **W** mátrix egységmátrix!

b.) Készíts konfidenciaintervallumot a $\hat{r}(h)$, $h \in \mathbb{Z}$ empirikus autokorrelációra! (1+1= 2p)

20.) Legyen $X_1, X_2, \ldots, X_n$ minta egy stacionárius folyamatból.

a.) Torzítatlan becslése a folyamat várható értékének a $\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ mintaátlag?

b.) Határozzuk meg a mintaátlag standard hibáját, azaz $D(\overline{X}_n)$-t!

c.) Legyen $\sum_{h=-\infty}^{\infty} |R(h)| < \infty$. Hova tart $n \cdot D^2(\overline{X}_n)$, ha $n \to \infty$?

21.) Legyen $(X_n)_{n\in\mathbb{Z}}$ stacionárius folyamat μ várható értékkel és $R(h)$ autokovariancia függvénnyel. Adjunk $0 < h \in \mathbb{Z}$ lépéses legkisebb négyzetes lineáris előrejelzést X_n segítségével, azaz határozzuk meg azon a és b értékeket, amikre az $MSE(a,b) := E([X_{n+h} - (a + bX_n)]^2)$ minimális! Határozzuk meg az így adódó $\hat{X}_{n+h} = \hat{a} + \hat{b}X_n$ becslés szórását!

22.) Legyen $(X_n)_{n\in\mathbb{Z}}$ stacionárius folyamat μ várható értékkel és $R(h)$ autokovariancia függvénnyel. Célunk h lépéses legkisebb négyzetes lineáris előrejelzés meghatározása az első n megfigyelés, $X_n, \ldots, X_1$ segítségével. Jelölje a legjobb lineáris előrejelzést $\mathbb{P}_n X_{n+h} := a_0 + a_1 X_n + \ldots + a_n X_1$. Mutassuk meg, hogy

a.) a legjobb lineáris előrejelzés együtthatói kielégítik a következő egyenleteket:
$E(X_{n+h} - \mathbb{P}_n X_{n+h}) = 0$
$E\left[(X_{n+h} - \mathbb{P}_n X_{n+h})X_{n+1-j}\right] = 0 \quad j = 1, \ldots, n$

b.) a legjobb lineáris előrejelzés együtthatói kielégítik a következő egyenletrendszert: $\boxed{\mathbf{\Gamma}_n \mathbf{a}_n = \mathbf{R}_n(h)}$, ahol

- $\mathbf{\Gamma}_n = [R(i-j)]_{i,j=1}^n$
- $\mathbf{a}_n = (a_1, \ldots, a_n)^T$
- $\mathbf{R}_n(h) = (R(h), R(h+1), \ldots, R(h+n-1))^T$

c.) az előrejelzés $\mathbb{P}_n X_{n+h} = \mu + \sum_{i=1}^n a_i(X_{n+1-i} - \mu)$ alakba írható, így a feladatok elején $\mu = 0$ feltehető, majd ezzel a képlettel megkapható az előrejelzés;

d.) az előrejelzés legkisebb négyzetes hibája $R(0) - \mathbf{a}_n^T \mathbf{R}_n(h)$.

23.) Legyen X_t stacionárius AR(1) folyamat 0 várható értékkel és σ szórású innovációkkal.

a.) Határozzuk meg a folyamat egylépéses, majd h lépéses legkisebb négyzetes előrejelzését és adjuk meg az előrejelzés hibáját! Vessük össze a kapottakat a 21.) feladat eredményével!

b.) Tegyük fel, hogy ismerjük a folyamat 1. és 3. megfigyelését, de a 2. megfigyelés hiányzik. Adjunk legkisebb négyzetes lineáris becslést X_2-re X_1 és X_3 segítségével, és határozzuk meg a becslés hibáját!

24.) Legyen X_t stacionárius MA(1) folyamat 0 várható értékkel és σ szórású innovációkkal. Tegyük fel, hogy ismerjük a folyamat 1. és 2. megfigyelését, de az 3. megfigyelés hiányzik. Adjunk legkisebb négyzetes lineáris becslést X_3-ra X_1 és X_2 segítségével, és határozzuk meg a becslés hibáját!

25.) Legyen X_t stacionárius AR(p) folyamat 0 várható értékkel és σ szórású innovációkkal. Határozzuk meg a folyamat egylépéses, legkisebb négyzetes előrejelzését!

26.) Legyen X_t stacionárius AR(1) folyamat μ várható értékkel és σ szórású innovációkkal. Az előrejelzési operátorok tulajdonságait felhasználva, határozzuk meg a folyamat h lépéses, legkisebb négyzetes előrejelzését!

27.) A előrejelzést a 'gnp96.txt' fájlban található idősor következő 10 értékére a legjobbnak bizonyult ARMA modell alapján

a.) a beépített `predict` függvény segítségével;

b.) az előrejelzési operátorról tanultak alapján!

Adjunk intervallumbecslést is az előrejelzésre! Ábrázoljuk az idősort az előrejelzésekkel együtt!

28.) Legyen $X_t = \varepsilon_t - \varepsilon_{t-1}$ MA(1) folyamat σ szórású innovációkkal. Generáljunk egy 100 elemű mintát, majd adjunk legkisebb négyzetes becslést a minta 12., 13. és 14. elemére a többi mintaelem segítségével (tekintsünk úgy rájuk, mintha ezek hibásak lennének, ezért becsülni kell őket)!

B6.) **[IV.24]** Legyen $X_1, X_2, \ldots, X_n$ minta az $X_t = -\pi + \varepsilon_t + 2\varepsilon_{t-1} - \varepsilon_{t-3}$ folyamatból, ahol ε_t fehér zaj folyamat σ szórással.

a.) Adj a minta első két eleme, X_1 és X_2 segítségével h lépéses legkisebb négyzetes előrejelzést $h = 1, 2, \ldots$ esetén! Számold ki az előrejelzés hibáját!

b.) Tegyük fel, hogy $n = 1000$. Adj a minta első két eleme és utolsó két eleme segítségével becslést X_{100}-ra! (2,5+1,5= 4p)

SZ12.) Legyen X_t AR(1) folyamat α paraméterrel és 1 szórású innovációval. Mutasd meg, hogy ekkor $D^2(\overline{X}_n) = \frac{\sigma^2}{n^2(1-\alpha^2)} \frac{n - n\alpha^2 - 2\alpha + 2\alpha^{n+1}}{(1-\alpha)^2}$. (1 pont)

SZ13.) Legyen X_t stacionárius MA(1) folyamat 0 várható értékkel és σ szórású innovációkkal. Tegyük fel, hogy ismerjük a folyamat 1., 2., 4. és 5. megfigyelését, de a 3. megfigyelés hiányzik. Adjunk legkisebb négyzetes lineáris becslést X_3-ra X_1, X_2, X_4 és X_5 segítségével, és határozzuk meg a becslés hibáját! (1 pont)

29.) Legyen X_t $(t = 1, 2, \ldots)$ ARCH(1) folyamat, azaz $X_t = \sigma_t \varepsilon_t$, ahol $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2$ és ε_t gauss-i fehér zaj.

a.) Határozd meg az $X_t | \mathcal{F}_{t-1}$ feltételes eloszlást, ahol $\mathcal{F}_t = \sigma(X_s : s \leq t)$!

b.) Számold ki X_t várható értékét!

c.) Határozd meg a $\mathrm{cov}(X_{t+h}, X_t)$ mennyiséget, ha $h > 0$!

d.) Mutasd meg, hogy X_t^2 AR(1) folyamat!

e.) Határozzuk meg X_t szórásnégyzetét és 4. momentumát! Milyen paraméterértékek esetén léteznek ezek a mennyiségek?

f.) Határozzuk meg a paraméterek momentum becslését!

g.) Határozzuk meg a paraméterek ML-becslését az X_1 kezdeti érték mellett!

Szimuláljunk ARCH(1) folyamatokat **R**-rel és az egyes feladatrészek eredményeit is vizsgáljuk meg!

B7.) **[V.8]** Stacionárius ARCH(1) folyamat esetén számold ki a folyamat lapultságát, azaz a kurtosis=kurtosis$_t$ = $\frac{E(X_t - EX_t)^4}{D^4 X_t} - 3$ mennyiséget! A paraméterek mely értéke esetén lesz a folyamat csúcsosabb a normális eloszlás csúcsosságához képest?

30.) Illesszünk megfelelő GARCH modellt a(z)

a.) NYSE;

b.) GBP-CHF;

idősorra! Vizsgáljuk meg, mennyire jó a modell! Készítsünk 99%-os konfidencia intervallumot az idősor egy szakaszára!

31.) Az általános lineáris modell paraméterének legkisebb négyzetes becslését felhasználva mutasd meg, hogy az $y_i = ax_i + b, \quad i = 1, 2, \ldots, n$ egyváltozós lineáris regressziós modell esetén az együtthatók legkisebb négyzetes becslése

$$\hat{a} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\left(\sum_{i=1}^{n} x_i y_i\right) - n \cdot \overline{x} \cdot \overline{y}}{\left(\sum_{i=1}^{n} x_i^2\right) - n \cdot \overline{x}^2} \quad \text{és} \quad \hat{b} = \overline{y} - \hat{a}\overline{x}.$$

32.) Egy fagylaltárus negyedéves forgalmának alakulása (ezer gombóc):

Év	I. negyedév	II. negyedév	III. negyedév	IV. negyedév
2015	95	152	255	118
2016	102	146	248	124
2017	97	156	245	122

Jelölje az idősor elemeit $(x_t)_{t=1,2,\ldots,12}$. Néhány számítási eredmény:

$$\sum_{t=1}^{12} x_t = 1860 \qquad \sum_{t=1}^{12} t \cdot x_t = 12342 \qquad \sum_{t=1}^{12} t = 78 \qquad \sum_{t=1}^{12} t^2 = 650$$

a.) Határozd meg a lineáris trend egyenletét és értelmezd a paramétereket!

b.) Határozd meg a szezonális eltéréseket és értelmezd őket!

c.) Adj pont- és intervallumbecslést a 2018. évi eladási mennyiségekre (mind a 4 negyedévre)!

33.) Szabadítsuk meg az *astsa* package-ben lévő *jj* (Johnson&Jonhson) 1960 és 1980 közötti negyedéves adatsort a trendtől és a szezonális hatásoktól a.) hagyományos idősor-dekompozícióval;

b.) az **R** nyelv *stl* parancsa (seasonal decomposition of time series by LOESS regression) segítségével.

Illesszünk alkalmas ARMA modellt a reziduálisokra, majd adjunk előrejelzést az 1981-es év mind a 4 negyedévére!

34.) Legyen Y_t stacionárius folyamat 0 várható értékkel és legyenek a és b valós konstansok. Ha $X_t = a + bt + s_t + Y_t$, ahol s_t szezonális komponens 4 periódussal, akkor mutassuk meg, hogy $\nabla\nabla_4 X_t = (1-B)(1-B^4)X_t$ folyamat stacionárius, és fejezzük ki az autokovariancia függvényét Y_t folyamat $R_Y(h)$ autokovariancia függvényével!

35.) Tekintsük azt a lineáris szűrőt, melynek együtthatói $a_i = \frac{1}{1+2q}$, ha $-q \le i \le q$ és $a_i = 0$, ha $|i| > q$.

a.) Mutassuk meg, hogy ez a szűrő torzítás nélkül átengedi az $m_t = c_0 + c_1 t$ lineáris trendet.

b.) Amennyiben Z_t, $t \in \mathbb{Z}$ független, 0 várható értékű és σ szórású valószínűségi változók, akkor számoljuk ki az $U_t = \frac{1}{2q+1}\sum\limits_{i=-q}^{q} Z_{t-i}$ lineárisan megszűrt folyamat várható értékét és szórását! Mire következtethetünk ebből nagy q esetén?

36.) Mutassuk meg, hogy az $(a_{-2}, a_{-1}, a_0, a_1, a_2)^T = \frac{1}{9}(-1, 4, 3, 4, -1)^T$ együtthatókkal számolt lineáris szűrő torzítás nélkül átenged minden harmadfokú polinomot, továbbá eliminálja a 3 periódusú szezonális komponenseket!

B8.) [V.15] Legyen $Y_t = -6 + \varepsilon_t - 6\varepsilon_{t-1}$, ahol ε_t fehér zaj 0 várható értékkel és 2 szórással, valamint legyenek a és b valós konstansok. Ha $X_t = (a + bt)s_t + Y_t$, ahol s_t szezonális komponens 12 periódussal, akkor mutasd meg, hogy $\nabla_{12}^2 X_t$ folyamat stacionárius, és határozd meg az autokovariancia függvényét!

B9.) [V.15] Határozd meg azt az $1 + \alpha B + \beta B^2 + \gamma B^3$ alakú lineáris szűrőt – azaz keresd meg az alkalmas α, β, γ valós együtthatókat –, amely torzítás nélkül átenged bármely lineáris trendet és minden 2 periódusú szezonális komponenst eliminál!

SZ14.) Legyen $m_t = \sum\limits_{i=0}^{p} c_i t^i$ polinomiális trend, ahol $0 \ne c_i \in \mathbb{R}$. Számítsd ki a $\nabla^p m_t$ és $\nabla^{p+1} m_t$ értékeket! Mire lehet ebből következtetni? (1 pont)

SZ15.) Bizonyítsd be, hogy az $\{a_i\}_{i\in\mathbb{Z}}$ együtthatókkal számolt lineáris szűrő akkor és csak akkor enged át torzítás nélkül egy tetszőleges k fokú polinomot, ha $\sum\limits_{i=-\infty}^{\infty} a_i = 1$ és minden $r = 1, 2, \ldots, k$ esetén $\sum\limits_{i=-\infty}^{\infty} i^r a_i = 0$! (2 pont)