Idősorok és többdimenziós statisztika gyakorlat

Matematika alapszak, matematikai elemző szakirány
2019/2020 őszi félév

Játékszabályok

- A gyakorlatokról maximum 3-szor lehet hiányozni. Aki többször hiányzik, nem kap gyakjegyet.
- 100 + x pontot lehet szerezni a félév során:
  - · 32 pont: B jelű beadandó feladatok (4 pont · 8)
  - · 30 pont: számítógépes/szimulációs beadandó feladatok (15 pont · 2)
  - · 38 pont: ZH: XII.11., 0–823 Kitaibel Pál terem
  - · x pont: SZ jelű 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. Sikertelen vagy meg nem írt ZH esetén gyakUV-t kell írnod, a beadandókért kapott pontok megmaradnak. GyakUV írása esetén legfeljebb 2-est lehet szerezni, bármennyi is az összpontszám.
- A ZH-kon használható: **számológép** és egy legfeljebb A4-es méretű lapra KÉZZEL írott "puska". A lap mindkét oldalára írhatsz.
- B jelű beadandók: Mindegyik maximálisan 4 pontot ér, a legjobb 8-at veszem figyelembe. 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	–	∞^π

Infók a gyakvezetőről

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

Ajánlott irodalom (példatárak)

- APZ Arató–Prokaj–Zempléni: Bevezetés a valószínűségszámításba és alkalmazásaiba: példákkal, szimulációkkal. Elérési helye: `http://www.math.u-szeged.hu/~barczy/sztoc_foly_gyak1.pdf`
- BP Barczy–Pap: Sztochasztikus folyamatok példatár, 1. rész. Elérési helye: `http://www.math.u-szeged.hu/~barczy/sztoc_foly_gyak1.pdf`
- BP2 Barczy–Pap: Sztochasztikus folyamatok példatár 2. rész. Elérési helye: `http://www.math.u-szeged.hu/~barczy/sztoc_foly_gyak2.pdf`
- VL Varga László: bevezető valószínűségszámítás kurzus példáinak megoldásai. Elérési helye: `http://vargal4.elte.hu/Vsz1_FeladatokMego.pdf`
- VL2 További példák megoldásokkal. Elérési helye: `http://vargal4.elte.hu/PluszPeldak.pdf`

Ajánlott irodalom (bevezető idősorelmélet)

- Shumway-Stoffer: Time series analysis and its applications, with R examples Elérési helye: `https://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf`
- Brockwell-Davis: Introduction to time series and forecasting

---

Ismétlésre ajánlott: VL-ből 81–85

1.) Legyen $Y_0 = 1$, $Y_1, Y_2, \ldots$ i.i.d. val. változók, $P(Y_1 = 1) = 1 - P(Y_1 = -1) = p \in (0,1)$. Vizsgáljuk meg, hogy a következő sorozatok Markov-láncot alkotnak-e, és ha igen, írjuk fel az átmenetvalószínűség-mátrixukat!

a.) $X_n = Y_1 + \ldots + Y_n \quad n = 1, 2, \ldots$

b.) $X_n = Y_n \cdot Y_{n+1} \quad n = 1, 2, \ldots$

c.) $X_n = Y_1 \cdot \ldots \cdot Y_n \quad n = 1, 2, \ldots$

2.) A 0 időpontban 2000 Ft-om van. Az $1, 2, \ldots$ időpontokban 1000 Ft értékű fogadásokat kötök. A fogadást p valószínűséggel megnyerem, $1 - p$ valószínűséggel pedig elveszítem. Célom az, hogy a tőkémet 4000 Ft-ra növeljem, amint ez teljesül, befejezem a játékot. A játéknak akkor is vége van, ha tőkém 0 Ft-ra esik. Legyen X_t a t-edik fogadás után a tőkém nagysága.

a.) Mutasd meg, hogy X_t Markov-lánc, határozd meg az állapotteret és az átmenetvalószínűségi mátrixot! Ábrázold a Markov-lánc gráfját és osztályozd az állapotokat!

b.) Mi az esélye, hogy 4000 Ft-tal / 0 Ft-tal fejezem be a játékot, ha $p = 0,6$?

c.) Várhatóan hány lépésben fejeződik be a játék, ha $p = 0,6$?

d.) Nézzük meg, konvergál-e az átmenetmátrix n-edik hatványa $n \to \infty$ esetén!

3.) Tegyük fel, hogy a kóla iparágban kétféle termék van: Coke és Poke. Ha egy személy legutóbb Coke-ot vett, akkor 90% eséllyel legközelebb is Coke-ot vesz. Amennyiben legutóbb Poke-ot ivott, akkor 80% valószínűséggel a következő vásárlás során is Poke-ot fog venni. Tegyük fel, hogy kezdetben a Coke gyártója a piac 60%-át uralja.

a.) Ábrázold a Markov-lánc gráfját, írd fel az átmenetvalószínűség-mátrixot és osztályozd az állapotokat!

b.) Ha egy vásárló Poke-ot iszik, akkor mennyi a valószínűsége, hogy a mostantól számított 2. vásárláskor Coke-ot fog inni?

c.) Ha egy vevő Coke-ot iszik, akkor mennyi a valószínűsége, hogy a mostantól számított 3. vásárláskor is Coke-ot fog inni?

d.) Számítsuk ki a két márka piaci részesedését (a fogyasztók hány %-a fogyasztja az egyiket és a másikat) a 3. időpont után!

e.) Vizsgáljuk meg szimulációval az n lépéses átmenetvalószínűség mátrixot és a két márka piaci részesedését $n \to \infty$ esetén! Nézzük meg a piaci részesedés konvergenciáját akkor, ha kezdetben a Poke márka teljesen uralja a piacot!

f.) Számítsd ki a stacionárius valószínűségeket!

g.) Tegyük fel, hogy egy fél literes kóla ára 300 Ft, előállítási költsége pedig 200 Ft. Felmérésekből kiderítették, hogy kólát rendszeresen 120 ezer ember fogyaszt, ők hetente átlagosan 1-et isznak meg. Számoljunk 52 héttel egy évben. Számítsd ki, hogy a Coke és a Poke gyártója (hosszú távon) mekkora profitra tesznek szert 1 év alatt!

h.) Ha Peti ezen a héten Coke-ot fogyaszt, akkor várhatóan hány hét múlva fogyasztja el a következő Coke-ot?

4.) Modellezzük egy egyetem matematika BSc szakos hallgatóit Markov-lánccal! Az elsőévesek 10%-a évet ismétel, 10% hagyja ott végleg a matek alapszakot, a többiek a következő évtől másodévesek lesznek. A másodév a legnehezebb, a másodévesek 20%-a ismétel évet és 15%-uk adja fel végleg a szakját. A harmadévesek 10%-a ismétel évet, és 5%-uk dönt az alapszak elhagyása mellett, a többiek elvégzik a szakot.

a.) Határozd meg a Markov-lánc állapotterét és az átmenetvalószínűségi mátrixot! Ábrázold a Markov-lánc gráfját és osztályozd az állapotokat!

b.) Milyen valószínűséggel lesz egy elsőéves hallgató 2 év múlva harmadéves?

c.) Milyen valószínűséggel lesz egy elsőéves hallgató 3 év múlva harmadéves?

d.) Milyen valószínűséggel hagyja el egy elsőéves hallgató végzettség nélkül a matematika alapszakot?

e.) Mi az esélye annak, hogy egy másodéves hallgató elvégzi a szakot?

f.) Várhatóan hány évig lesz egy másodéves hallgató másodéves?

g.) Várhatóan hány évet tölt egy elsőéves egyetemista az egyetemen?

h.) A visszajelzések szerint a munkaerőpiac évente 50 matematika alapszakos hallgatót képes felszívni. Ha a hallgatóknak várhatóan a 40%-a tovább szeretne tanulni valamilyen mesterszakra, akkor az egyetemnek hány gólyát érdemes felvennie a szakra?

i.) Nézzük meg, konvergál-e az átmenetmátrix n-edik hatványa $n \to \infty$ esetén!

j.) Tegyük fel, hogy "sok" éven keresztül, minden évben 60 hallgatót vesznek fel matematika alapszakra és más egyetemekről nem jönnek át másod-, illetve harmadévesek erre a szakra. "Sok-sok" év múlva várhatóan hány hallgató fog tanulni az egyes évfolyamokon? Ha egy hallgató képzése egy félévre 200 ezer Ft, akkor a szak összesen mennyi pénzébe kerül 1 félév során az államnak?

5.) Egy országban a családok három csoportba oszthatók: városban, városok vonzáskörzetében vagy vidéken élők. Egy évben a városlakó családok 15%-a átköltözik a vonzáskörzetbe és 5%-a vidékre költözik; ugyanakkor a vonzáskörzetben élők 6%-a városba és 4%-a vidékre költözik; végül a vidéki lakosság 4%-a a városokba és 6%-a a városok vonzáskörzetébe települ át. Jelenleg a családok 40%-a városokban, 35%-a pedig a városok vonzáskörzetében él.

a.) Ha egy család városban lakik, akkor mi a valószínűsége, hogy 2 év múlva is városban fog lakni?

b.) Két év múlva a családok hány %-a fog városban élni?

c.) Ha ez a modell hosszú távon jól írja le a lakosok költözési viselkedését, akkor hogyan fog alakulni az ország lakóhely szerinti megoszlása?

6.) A Google által használt **PageRank** algoritmus a keresőben talált oldalakat sorrendbe állítja, egyszerűsített változata a következőképpen írható le.
Tekintsünk egy Markov-láncot, melynek gráfját jelölje $G = (V, E)$, ennek csúcshalmaza az adott adatbázisban található oldalak, továbbá $(u, v) \in E$ ha az u oldalról mutat link a v oldalra, az élre pedig az $\frac{1}{d_{ki}(u)}$ számot írjuk, ahol $d_{ki}(u)$ az u csúcsból más csúcsokba mutató élek száma (ki-fokszám). Amennyiben egy oldal semelyik másik oldalra se mutat, akkor egyenletes eloszlás szerint véletlenszerűen választunk egy oldalt és oda ugrunk, azaz az összes többi csúcshoz berajzolunk élet, és az azonos $\frac{1}{|V|-1}$ számokat írunk rájuk. A Markov-lánc stacionárius eloszlása adja meg az oldalak rangszámait, minél nagyobb a szám, annál "fontosabb" az adott oldal.
Tekintsünk egy olyan esetet, amikor összesen az (a, b, c, d) négy oldal van egy adatbázisban a következő linkekkel: $a \to b$, $b \to c$, $d \to c$, $b \to d$.

a.) Írjuk fel a PageRank által meghatározott bolyongás állapotmátrixát és osztályozzuk az állapotokat!

b.) Határozzuk meg a stacionárius eloszlást, majd adjuk meg az oldalak sorrendjét!

Gyakorlásra ajánlott: BP2-ből 1.10–1.15, 1.16, 1.20, 1.21, 1.25, 2.1, 2.2, 2.16–2.19, 2.23, 2.27, 4.16–4.22, 4.25–4.31, 5.6-5.8, 5.13, 5.15, 5.17, 9.1

B1.) [X.2.] Egy urna a kiinduló helyzetben két festetlen golyót tartalmaz. Véletlenszerűen kiválasztunk egy golyót és feldobunk egy pénzérmét. Ha a választott golyó festetlen és fejet dobunk, akkor a kiválasztott festetlen golyót pirosra festjük; ha a választott golyó festetlen és írást dobunk, akkor a kiválasztott festetlen golyót feketére festjük. Ha a golyó már korábban meg lett festve, akkor átfestjük a másik színre. Jelöljük az urna állapotát az alábbi számhármassal: (i, j, k), ahol i az urnában lévő festetlen, j a piros, k pedig a fekete golyók száma. Legyen X_t az urna állapota a t-edik lépés után. Mutasd meg, hogy X_t Markov-lánc és határozd meg az átmenetvalószínűségi mátrixot!
Segítség: a Markov-láncnak 6 állapota van, a kezdeti állapot a $(2, 0, 0)$ számhármassal írható le.

B2.) [X.2.] Tekintsük a 3. feladat történetét. Egy reklámügynökség megkeresi a Coke tulajdonosát egy olyan ajánlattal, hogy a márkát rendkívül menővé fogják tenni vagány TV-reklámokkal, ezáltal annak a valószínűsége, hogy egy fogyasztó a következő héten a másik márkát fogyasztja, 10%-ról 5%-ra fog lecsökkenni. A reklámkampányra 50 millió Ft-os ajánlatot adnak. Számításokkal alátámasztva adj tanácsot, a Coke tulajdonosának érdemes-e belekezdeni a reklámkampányba!

B3.) **[X.2.]** Egy vállalat telefonon értékesíti termékét. Tekintsünk egy potenciális vevőt, akit még nem hívott fel a cég képviselője – az 1. hívás után 60% az esélye, hogy az illető lanyha érdeklődést fog tanúsítani a termék iránt, 30% eséllyel komolyan fog érdeklődni és 10% eséllyel a vállalat törölni fogja a potenciális vevők köréből. Az a vevő, aki korábban lanyha érdeklődést mutatott, a következő hívás során 20% eséllyel továbbra is alacsony érdeklődést fog mutatni, 30% valószínűséggel komoly érdeklődő lesz, 30% eséllyel megveszi a terméket és 20% eséllyel törölni kell a listából. A korábban komolyan érdeklődő potenciális ügyfél a következő hívás során 10% eséllyel alacsony érdeklődést fog mutatni, 40% valószínűséggel továbbra is komoly érdeklődő marad, míg 50% eséllyel megvásárolja a terméket.

a.) Mi az esélye, hogy egy potenciális vevő megveszi a terméket?

b.) A vállalat átlagosan hányszor fog felhívni egy potenciális vevőt?

SZ1.) Tekintsük a szimmetrikus bolyongást az $n \geq 5$ hosszú körön. Tegyük fel, hogy az éleken oda-vissza lehet közlekedni. Add meg az átmenetmátrixot és osztályozd az állapotokat! Határozd meg a stacionárius eloszlást! (1p)

SZ2.) Legyen $\{X_k,\ k \geq 0\}$ homogén Markov-lánc $S = \{0, 1, 2\}$ állapottérrel, $X_0 = 1$, az átmenetvalószínűség mátrix $\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0.1 & 0.9 \end{pmatrix}$. Mutasd meg, hogy az $Y_k = I(X_k \geq 1)$ folyamat nem Markov-lánc! (2p)

SZ3.) Egy gépjármű-biztosítással foglalkozó vállalat ügyfeleinek az alapján határozza meg a biztosítási díjat, hogy volt-e balesetük az előző két évben. Ha egy ügyfélnek az elmúlt két évben nem volt balesete, akkor a biztosítási díj évi 100 ezer Ft. Ha a partnernek az elmúlt két év mindegyikében volt balesete, akkor évi 200 ezer Ft-ot fizet. Amennyiben az ügyfélnek az elmúlt két év egyikében volt balesete, akkor a biztosítási díj 150 ezer Ft. Ha a partnernek balesete volt az elmúlt évben, akkor 10% a valószínűsége, hogy az aktuális évben is balesete lesz. Ha az ügyfélnek nem volt balesete az elmúlt évben, akkor csak 3% eséllyel lesz balesete az aktuális évben. Tegyük fel, hogy a fenti számok hosszú távon stabilnak tekinthetők. A biztosítónak mennyi az összes ügyfélre vonatkozó átlagos biztosítási díja? (2p)

SZ4.) Tekintsük a 4. feladat történetét. Milyen valószínűséggel lesz egy elsőéves hallgatóból harmadéves? (1p)

SZ5.) Legyenek X_n és Y_n független Markov láncok a $\{0; 1; 2\}$ állapottéren. Mindkét lánc átmenetmátrixa $\begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix}$. Tegyük fel, hogy $X_0 = 0$ és $Y_0 = 2$, valamint legyen $T = \inf\{n \geq 0 : X_n = Y_n\}$. Határozd meg T várható értékét! (3p)

7.) Legyen $(X_t)_{t\in\mathbb{R}}$ Poisson-folyamat λ intenzitással. Mutasd meg a Poisson-folyamat alábbi tulajdonságait:

a.) az autokovariancia függvénye $\mathrm{Cov}(X_t, X_s) = \lambda \cdot \min(t, s)$;

b.) ha $t > s$, akkor $X_s|X_t \sim \mathrm{Bin}\big(X_t, \frac{s}{t}\big)$;

c.) ha $t < s$, akkor $X_s|X_t \sim\ X_t{+}\mathrm{Poi}\big(\lambda(s-t)\big)$.

8.) Tegyük fel, hogy a tűzoltóság telefonközpontjába a hívások Poisson-folyamat szerint érkeznek az egész nap során, óránként átlagosan 4 hívás. Tekintsük a folyamat kezdő, 0. időpontjának éjfélt.

a.) Mennyi a valószínűsége, hogy az első órában kettőnél kevesebb hívás érkezik?

b.) Mennyi annak az együttes bekövetkezési valószínűsége, hogy 6 és 7 óra között 3, 6 és 9 óra között 13, valamint 6 és 11 óra között 20 hívás érkezik?

c.) Mennyi annak az együttes bekövetkezési valószínűsége, hogy 6 és 8 óra között 2, valamint 7 és 10 óra között 4 hívás érkezik?

d.) Átlagosan mennyi idő telik el két hívás között?

e.) Átlagosan/várhatóan mennyi hívás érkezik be 10 és 12 óra között? Mennyire változékony ez az átlagos érték?

f.) Feltéve, hogy az első órában hat hívás érkezett, mennyi a valószínűsége, hogy a második órában legalább két hívás jön?

g.) Várhatóan mikor érkezik a 15. hívás?

h.) Feltéve, hogy a 1 óra előtt nem érkezett be hívás, milyen eséllyel érkezik be az 1. hívás 3 óra után?

i.) Feltéve, hogy a 3. hívás 2 órakor érkezett be, milyen eséllyel érkezik be a 4. hívás 4 óra után?

j.) Tegyük fel, hogy 3 óráig 10 hívás érkezett be. Ekkor mi a valószínűsége, hogy 1-ig 4 hívás érkezett be? Mennyi az 1 óráig beérkező hívások várható értéke és szórása?

k.) Tegyük fel, hogy a 3 óráig 10 hívás érkezett be. Ekkor mi a valószínűsége, hogy 5-ig 15 hívás érkezik be? Mennyi az 5-ig beérkező hívások várható értéke és szórása?

9.) Egy banki ügyintézőhöz a pénzt felvenni szándékozók $\lambda = 8$ fő/óra intenzitású Poisson-folyamat szerint érkeznek. A pénzt befizetni szándékozók ettől függetlenül, $\mu = 2$ fő/óra intenzitású Poisson-folyamat szerint érkeznek.

a.) Mennyi az esélye, hogy fél óra alatt legfeljebb ketten érkeznek a banki ügyintézőhöz?

b.) Feltéve, hogy 1 óra alatt 8 ügyfél érkezik, mennyi az esélye, hogy ebből hárman szeretnének pénzt befizetni?

c.) Mi a valószínűsége, hogy nyitást követően előbb érkezik 5 ügyfél pénzfelvételi célból, mielőtt egy pénzbefizető ügyfél betérne?

d.) Mi a valószínűsége, hogy nyitást követően előbb érkezik 5 ügyfél pénzfelvételi célból, mielőtt 2 pénzbefizető ügyfél betérne?

10.) Barnus és Marci egypetéjű ikertestvérek, akik közösen létrehoznak egy társkereső profilt az egyik legnagyobb társkereső honlapon. A srácok megállapodnak benne, hogy az első lánnyal, aki hajlandó a profil alanyával randevúra menni, Barnus fog találkozni. Tegyük fel, hogy fél nap alatt átlagosan 10 lány tetszik nekik azok közül, akiket korábban még nem láttak – ezekre a lányokra rá is írnak. Ha egy lányt online megszólítanak, akkor $\frac{1}{10}$ eséllyel lesz a chat-elésből személyes randevú. Mi az esélye, hogy Marcinak több mint 3 napot kell várnia az első randira?

Gyakorlásra ajánlott: BP-ből 2.4.2-2.4.4, 2.4.7, 2.4.9

B4.) **[X.9.]** Egy biztosítóhoz a károkat Poisson-folyamat szerint jelentik be 0,5 kár/hét intenzitással.

a.) Mi az esélye, hogy a 6. káresetet 3 héttel az 5. káreset után jelentik be?

b.) Mi annak az együttes valószínűsége, hogy az első három héten 2, a 4-7. hetekben pedig összesen 3 kárt jelentettek be?

c.) Határozd meg a 13. bejelentett kár időpontjának várható értékét és szórását!

d.) Ha az első 5 héten 4 káreset következett be, akkor milyen eséllyel következik be 3 káreset az első 2 héten?

B5.) **[X.16.]** Egy délelőtt 8 órakor nyitó fehérneműboltba óránként átlagosan 10 vevő tér be, egy potenciális vásárló $\frac{2}{3}$ eséllyel nő, $\frac{1}{3}$ valószínűséggel férfi. Tekintsük a vevők számát Poisson-folyamatnak. *Tegyük fel*, hogy délelőtt 8 és 9 között 10 nő ment be az üzletbe. Ezzel a feltétellel, számold ki a következő események valószínűségét:

a.) 8 és 9 között pontosan 10 férfi is betért a boltba;

b.) 8 és 9 között legalább 20 vevő ment be az üzletbe;

c.) nyitás után előbb lép be a boltba a 2. férfi, mint a 3. női látogató. (1+1+2=4p)

SZ6.) Legyen X_t Poisson-folyamat, τ_i az i-edik esemény bekövetkezési időpontja. Mutasd meg, hogy $\tau_1|X_t = 1 \sim E(0;t)$, azaz ha tudjuk, hogy a t időpontig pontosan 1 esemény következett be, akkor ennek az eseménynek a bekövetkezési időpontja egyenletes eloszlású a $(0;t)$ intervallumon. (1p)

SZ7.) Négy bolha oda-vissza ugrál Vili a vizsla és Leó a labrador között. Az ugrások időpontjait független, $\lambda = 3$ ugrás/perc intenzitású Poisson-folyamatok adják meg. Kezdetben az összes bolha Vilin tanyázik. Adjuk meg 5 perc múlva a Vilin lévő bolhák számának eloszlását! (2p)

SZ8.) Tegyük fel, hogy a tűzoltóság telefonközpontjába a hívások Poisson-folyamat szerint érkeznek, óránként átlagosan 4 hívás. Feltéve, hogy az első 3 órában 11 hívás érkezett, várhatóan mikor érkezett ezek közül az utolsó (tehát a 11.)? (2p)

Ismétlésre ajánlott: VL-ből 54–74

11.) Legyenek $X, Y \sim E(0;1)$ függetlenek. Határozd meg a következő transzformáltak sűrűségfüggvényét: a.) $X+Y$ b.) $X-Y$ c.) XY

12.) Legyenek $X, Y \sim E(0;1)$ függetlenek. Határozzuk meg az $E\left(\frac{X}{Y}\right)$ értéket!

13.) Szandi minden nap villamossal és busszal közlekedik, hogy eljusson az egyetemre. Az alábbi ábra tartalmazza a közlekedési időket:

Otthon	$\xrightarrow{\text{gyalog}}$ 5 p	Villamos-megálló	$\xrightarrow{\text{villamossal}}$ 10 p	$\xrightarrow{\text{gyalog}}$ 2 p	Busz-megálló	$\xrightarrow{\text{busszal}}$ 8 p	$\xrightarrow{\text{gyalog}}$ 5 p	**Egyetem**

Tapasztalatai alapján átlagosan 2 percet vár a villamos megállójában és 4 percet a busz megállójában. Szandi ma reggel később ébredt fel, ezért csak 7:43-kor indult el otthonról, az első órája 8:15-kor kezdődik. Becsüljük különböző, értelmes valószínűségi modellek segítségével annak a valószínűségét, hogy el fog késni!

14.) Mutassuk meg, hogy $U = X + Y$ és $V = \frac{Y}{X+Y}$ függetlenek, és határozzuk meg az eloszlásukat, amennyiben

a.) $X \sim \mathrm{Exp}(\lambda)$, $Y \sim \mathrm{Exp}(\lambda)$ függetlenek;

b.) $X \sim \Gamma(\alpha, \lambda)$, $Y \sim \Gamma(\beta, \lambda)$ függetlenek.

15.) Legyen X és Y együttes sűrűségfüggvénye $f_{X,Y}(x,y) = \frac{e^{-\frac{x}{y}-y}}{y} I(x>0, y>0)$.

a.) X és Y független egymástól?

b.) $P(X < 5Y)$=?

c.) Milyen eloszlású X az $Y = y$ feltétel mellett? Számítsuk ki az $E(X|Y)$ feltételes várható értéket!

16.) Mely c valós paraméter esetén lesz kétdimenziós sűrűségfüggvény

$$f_{X,Y}(x,y) = \begin{cases} c & \text{ha } 0 < y < 2x < 2 \\ 0 & \text{egyébként} \end{cases} ?$$

a.) Adjuk meg a peremsűrűségfüggvényeket és a kovarianciamátrixot!

b.) Generáljunk véletlen számokat ebből az eloszlásból, majd becsüljük meg a $P\left(X > \frac{1}{2},\ Y < 1\right)$ valószínűséget! Számoljuk is ki a pontos értéket!

c.) Számítsuk ki és becsüljük szimulációval az $E(X|Y)$ feltételes várható értéket és a $D^2(X|Y)$ =? feltételes szórásnégyzetet!

d.) Legyen $U = X + Y$ és $V = X - Y$. Határozd meg az (U, V) valószínűségi vektorváltozó (együttes) sűrűségfüggvényét!

17.) Legyen c alkalmas valós szám, X és Y együttes sűrűségfüggvénye $f_{X,Y}(x,y) = (cx+1) \cdot I(x>0,\ y>0,\ x+y<1)$.

a.) c =? b.) $P(Y < 2X^2)$ =? c.) $E(Y|X)$ =?

18.) Három nyugdíjas barátnő, Kató néni, Luca néni és Margó néni együtt mennek a városi nagypostára, hogy feladják a gázszámlájukat. Mivel délelőtt fél 10-kor, munkaidőben mennek, az 6 pultból mindössze 2-nél van ügyfél, ezért mindannyian egyszerre oda tudnak állni egy-egy postai ügyintézőhöz feladni a csekket. Tegyük fel, hogy a csekkfeladás (kiszolgálási) ideje exponenciális eloszlású 2 perc várható értékkel. Határozd meg annak a valószínűségét, hogy

a.) Kató néni előbb végez az ügyintézéssel, mint Luca néni;

b.) Kató néni végez az ügyintézéssel elsőként, Luca néni másodikként, Margó néni pedig utolsóként!

19.) Legyen X és Y független standard normális eloszlású, $\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 2X+Y \\ -X+Y \end{pmatrix}$.

a.) Milyen eloszlást követ a V valószínűségi változó?

b.) Határozzuk meg U és V együttes eloszlását! Mi a kovarianciamátrix és az együttes sűrűségfüggvény? Rajzoljuk ki az együttes sűrűségfüggvény szintvonalait!

c.) Határozzuk meg a $P(U < V)$ valószínűséget!

20.) Legyen (X, Y) kétdimenziós normális eloszlású. Határozzuk meg az $X|Y$ feltételes eloszlást!

21.) Tegyük fel, hogy a magyar férfiak magassága és testsúlya kétdimenziós normális eloszlású. A férfiak átlagmagassága 178 cm, 9 cm szórással; átlagos testsúlyuk pedig 85 kg, 10 kg szórással. A magasság és a testtömeg közötti korreláció 0,7. Számoljuk ki pontosan és R-es szimulációval egyaránt az alábbi mennyiségeket.

a.) Feltéve, hogy egy férfi 80 kg, mi a valószínűsége, hogy magasabb 180 cm-nél?

b.) Átlagosan mekkora súlyú egy 190 cm magas férfi?

c.) Átlagosan milyen magas egy 94,33 kg-os férfi?

Gyakorlásra ajánlott: VL2-ből 1–4, VL-ből 99–102, APZ-ből 6.28, 6.29

B6.) [X.16.] Juliska néni a városi piacon értékesíti almáját. Mind az értékesítési mennyiség, mint az ár véletlennek tekinthető (lehet vele alkudozni). Az egy nap alatt eladott mennyiség (kg) Pareto-eloszlású $\frac{5}{4}$ és 10 paraméterekkel, míg az eladott almák ára (Ft/kg) egyenletes eloszlású 180 és 220 paraméterekkel. Tegyük fel, hogy a mennyiség és az ár függetlenek egymástól. Költségei naponta 4000 forintot tesznek ki (benzinköltség).

a.) Várhatóan mennyi profitra fog szert tenni egy nap alatt?

b.) Határozd meg annak a valószínűségét, hogy Juliska néni napi profitja meghaladja az 5000 forintot! (1+3= 4p)

B7.) [XI.6.] Legyen (X, Y) sűrűségfüggvénye $f(x, y) = 2\,I(0 < x < y < 1)$. Határozd meg $U = \frac{X}{Y}$ eloszlását!

Útmutatás: ehhez tekintsük a $V = Y$ valószínűségi változót, majd határozd meg U és V együttes sűrűségfüggvényét!

B8.) [XI.6.] Legyen (X, Y) sűrűségfüggvénye $f(x, y) = e^{-y}\,I(0 \le x \le y)$. Számítsd ki az $E(Y|X)$ feltételes várható értéket és a $D(Y|X)$ feltételes szórást!

B9.) [XI.20.] Az emberek IQ-ja 100 pont várható értékű 15 szórással. A Mensa nevű egyesületbe bárki jelentkezhet, aki egy hivatalos IQ-teszten magas, 130 feletti pontot ér el. Zsuzsi és Dóri testvérek. Ha tudjuk, hogy Zsuzsi Mensa-tag lett 135 ponttal, akkor mi az esélye, hogy Dóri is sikerrel jelentkezhet a Mensába, amennyiben a két testvér intelligenciahányadosa közti korreláció $-1 \le \rho \le 1$ valós szám? Az eredményt számold ki számítógép segítségével, készíts belőle alkalmas grafikont (x tengely: ρ, y tengely: a valószínűség) és küldd el az Email címemre! Értékeld a látottakat – erre számítottál?

SZ9.) Legyenek X és Y egymástól független valószínűségi változók. Bizonyítsd be, hogy $U = XY$ sűrűségfüggvénye $f_U(u) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_X\left(\frac{u}{y}\right) f_Y(y)\,dy.$ (2p)

SZ10.) Legyenek $X, Y \sim N(0, 1)$ eloszlású, egymástól független valószínűségi változók. Határozd meg $U = \frac{X}{Y}$ sűrűségfüggvényét és várható értékét! Milyen nevezetes eloszlást követ U? (2p)

SZ11.) Az Északi és a Déli tömb közötti távolságot egymástól függetlenül 4-en lemérik. Tegyük fel, hogy az egyes mérések eredményei közös, a valódi távolsággal megegyező várható értékű és véges szórású normális eloszlást követnek. Mi az esélye, hogy az első 3 mérés átlaga pontosabban közelíti a valódi távolságot, mint a 4. mérés? (2 pont)

SZ12.) Box-Müller transzformáció. Legyenek $X_1, X_2 \sim E(0, 1)$ függetlenek. Legyenek $Y_1 = \sqrt{-2 \log X_1} \cos(2\pi X_2)$, $Y_2 = \sqrt{-2 \log X_1} \sin(2\pi X_2)$. Mutassuk meg, hogy $Y_1, Y_2 \sim N(0, 1)$ függetlenek! (2p)

SZ13.) Legyen $f_{X,Y}(x, y) = \begin{cases} c & \text{ha } 0 < x < 2 \text{ és } -2 + x < y < 2 - x \\ 0 & \text{különben} \end{cases}$

Határozd meg a c értékét, majd az $E(X|Y)$ feltételes várható értékeket! (1p)

SZ14.) Tegyük fel, hogy egy gyorsúszó alvásideje normális eloszlásúnak tekinthető 8 óra várható értékkel és 1 óra szórással. Tegyük fel továbbá, hogy amennyiben x órát alszik, akkor a 100 méter gyorsúszáson elért ideje ugyancsak normális eloszlású $58 - x$ mp várható értékkel és 1 mp szórással. Számold ki annak a valószínűségét, hogy a gyorsúszó megdönti a 100 méter gyorsúszás 46,91 mp-es világcsúcsát! (2p)

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

22.) Legyen X_t *véletlen bolyongás drift*-tel: $X_t = \delta + X_{t-1} + \varepsilon_t$ $(t = 1, 2, \ldots)$, 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?

23.) 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!

24.) 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 függvény
- a folyamat $MA(\infty)$ reprezentációja.

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

25.) Generáljunk egy 1000 elemű mintát az $X_t = \omega + \alpha \cdot X_{t-1} + \varepsilon_t$ folyamatból, ahol $\omega = 2$, $\alpha = 0,6$ és $\varepsilon_t \sim WN(0;1)$. Becsüljük vissza az ω és α paramétereket

- az `arima` függvénnyel különböző 'method'-ok esetén (ML-becslés, legkisebb négyzetes becslés);
- az `lm` függvénnyel, azaz lineáris regresszióként tekintve a feladatra.

Hasonlítsuk össze a becsléseket és értelmezzük a kapottakat! Próbáljunk magyarázatot találni a kapott furcsaságra!

26.) 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.) Illesszünk más modelleket (például AR(1), AR(3), ARMA(1,1)), majd próbáljuk meg észrevenni, hogy ezek miért nem lesznek jók!
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?

27.) 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)!
d.) Adjunk előrejelzést az idősor következő 10 értékére a legjobbnak bizonyult ARMA modell alapján a beépített `predict` függvény segítségével! Adjunk intervallumbecslést is az előrejelzésre! Ábrázoljuk az idősort az előrejelzésekkel együtt!

28.) 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.

29.) 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

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

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

Év	I. negyedév	II. negyedév	III. negyedév	IV. negyedév
2016	95	152	255	118
2017	102	146	248	124
2018	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! Használjuk az előző feladat eredményét megfelelő szereposztással!
b.) Határozd meg a szezonális eltéréseket és értelmezd őket!
c.) Adj pontbecslést a 2019. évi eladási mennyiségekre (mind a 4 negyedévre)!

31.) Szabadítsuk meg az *astsa* package-ben lévő *jj* (Johnson&Jonhson) 1960 és 1980 közötti negyedéves adatsort a trendtől és a szezonális hatásoktól hagyományos idősor-dekompozícióval! Illesszünk alkalmas ARMA modellt a reziduálisokra, majd adjunk előrejelzést az 1981-es év mind a 4 negyedévére!

B10.) **[XI.27.]** Legyen $X_t = \frac{1}{2}\varepsilon_t - \varepsilon_{t-1} + \frac{1}{4}\varepsilon_{t-2} \quad (t \in \mathbb{Z})$, ahol ε_t fehér zaj. Határozzuk meg X_t várható érték, autokovariancia és autokorreláció függvényét! Gyengén stacionárius a folyamat?

B11.) **[XII.11]** Egy stacionárius AR(2) folyamat esetén előfordulhat, hogy az autokorrelációk az $r(1) = 1/2$; $r(2) = 1/6$; $r(3) = 1/12$ értékeket veszik fel?

B12.) **[XII.11.]** A Siófoki Aranyparton egy lángosárus forgalmának alakulása (bevétel, ezer Ft) havonta:

Év	május	június	július	augusztus	szeptember
2015	513	1052	2555	2118	634
2016	631	1246	2848	2249	831
2017	459	968	2648	2224	348
2018	642	1157	2245	1822	730

Jelölje az idősor elemeit $(x_t)_{t=1,2,\ldots,20}$. Néhány számítási eredmény:

$$\sum_{t=1}^{20} x_t = 27920 \qquad \sum_{t=1}^{20} t \cdot x_t = 292781 \qquad \sum_{t=1}^{20} t = 210 \qquad \sum_{t=1}^{20} t^2 = 2870$$

a.) Határozd meg a lineáris trend egyenletét!
b.) Határozd meg a szezonális eltéréseket!
c.) Adj pontbecslést a 2019. évi eladási mennyiségekre (mind az 5 hónapra)!

SZ15.) 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)

SZ16.) Legyen $X_t = \sin(2\pi U t)$ $(t = 1, 2, \ldots)$, ahol $U \sim E(0; 1)$. Mutassuk meg, hogy X_t folyamat gyengén stacionárius, de erősen *nem* stacionárius! (2p)
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 $P\left(X_2 < -\frac{\sqrt{3}}{2}, X_3 < -\frac{\sqrt{3}}{2}\right)$!

SZ17.) 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)