Idősorok és többdimenziós statisztika gyakorlat

Matematika alapszak, matematikai elemző szakirány
2023/2024 őszi félév

Játékszabályok

- A gyakorlatokról maximum 3-szor lehet hiányozni. Aki többször hiányzik, nem kap gyakjegyet.
- 100 + x pontot lehet szerezni a félév során:
  - 28 pont: **B** jelű, "kis" beadandó feladatok (4 pont · 7)
  - 22 pont: "nagy" szimulációs/számítógépes beadandó feladat
  - 50 pont: ZH: dec. 13., hagyományos kézzel írós
  - 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 Canvas-on keresztül. 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.
- A ZH-n minden offline/online anyag használható, a kommunikáció bármely formája más személyekkel viszont tilos.
- **B** jelű beadandók: Mindegyik maximálisan 4 pontot ér, a legjobb 7-et veszem figyelembe. Több feladatot is kihirdetek, amik közül ízlés szerint válogathattok. A beadandók célja, hogy folyamatosan tanuljatok, gyakoroljatok.
- "Nagy" beadandó: nagyobb lélegzetvételű, számítógép segítségével megoldandó feladat.
- Aki a beadandókból összesen nem éri el a 30%-ot (15 pont), nem kap gyakjegyet.
- Minden beadandónál elvárás az önálló munkavégzés. 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	Citi, MQA
Tanszék	Valószínűségelméleti és Statisztika Tanszék (ELTE TTK)
E-mail	`vargala4@gmail.com`
Honlap	`vargal4.elte.hu`

Ajánlott irodalom (példatárak)

APZ	Arató–Prokaj–Zempléni: Bevezetés a valószínűségszámításba és alkalmazásaiba: példákkal, szimulációkkal.
BP	Barczy–Pap: Sztochasztikus folyamatok példatár, 1. rész.
BP2	Barczy–Pap: Sztochasztikus folyamatok példatár 2. rész.
VL	Varga László: bevezető valószínűségszámítás kurzus példáinak megoldásai.
VL2	További példák megoldásokkal.

A félév során modellezésre/szimulációkhoz használt 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
- **R** nyelv/szoftver letöltési helye
- Szövegszerkesztésre ajánlott szoftver: RStudio; letöltési helye
- R MarkDown: egyszerű szövegszerkesztő és kódot is futtató package, rövid bevezető

Ismétlésre ajánlott: VL-ből 81–85

1.) Legyenek $Y_1, Y_2, \ldots$ i.i.d. valószínűségi 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, vizsgáljuk meg, homogének-e, valamint írjuk fel az átmenetvalószínűség-mátrixukat!

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

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

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

2.) Tönkremenési feladat. 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.) Ha nem halaszt évet, akkor milyen valószínűséggel lesz egy most

b1.) másodéves hallgató 4 év múlva harmadéves?

b2.) elsőéves hallgató még 4 év múlva is egyetemista?

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

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

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

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

g.) 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?

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

i.) 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 500 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.) [09.27.] Egy urnában öt cédula van, rajtuk a 0, 1, 1, 2 és 2 számokkal. Egymás után húzunk cédulákat az urnából visszatevéssel, legyen X_n az összes addig kihúzott cédulán lévő szám szorzata modulo 3. Például ha az első négy húzás 2, 1, 2, 2, akkor $2 \cdot 1 \cdot 2 \cdot 2 = 8 \equiv 2 \; mod(3)$, ezért $X_4 = 2$. Mutasd meg, hogy $X_n \quad n = 1, 2, \ldots$ Markov-lánc, rajzold fel a gráfját és határozd meg az átmenetvalószínűségi mátrixot!

B2.) [10.04.] Egy részvény évente 2-szer fizethet osztalékot attól függően, hogy az igazgatótanács megszavazza-e az osztalékfizetést. Egy részvény névértéke 10 ezer Ft, aktuális árfolyama 20 ezer Ft. Rendelkezel 200 ezer Ft-tal, most úgy döntesz, hogy veszel 10 darabot a részvényből. Tegyük fel, hogy a részvényenként kifizetett osztalék normális eloszlású 200 Ft várható értékkel és 100 Ft szórással. Az osztalékot névérték alapján fizetik. Sok-sok év adatai alapján megfigyeltük, hogy amennyiben a részvény adott időpontban fizet osztalékot, akkor 60% eséllyel fél év múlva is fizetni fog; míg amennyiben nem fizet osztalékot, akkor fél év múlva 80% eséllyel nem fog fizetni. Ha hosszú távon gondolkozol a részvénybe való befektetéssel, akkor a következő 15 év során várhatóan mennyi bevételed fog származni osztalékokból? Szerinted a mostani befektetési lehetőségeket figyelembe véve, megéri ebbe a részvénybe fektetni a pénzt? Indokolj!

B3.) **[10.11.]** Tekintsük a 4. feladat történetét. Az egyetem átszervezi a képzést, ennek következtében a másodév lényegesen könnyebben teljesíthetővé válik: a másodévesek 15%-a fog várhatóan évet ismételni és csak 5% hagyja ott végleg a szakot. Ebben az új helyzetben válaszolj a következőkre:

a.) Mennyivel nő a régi képzéshez képest annak az esélye, hogy egy gólya el fogja végezni a szakot?

b.) Minden egyéb feltevés változatlansága mellett a matematika szak mennyi pénzébe kerül 1 félév során az államnak? Vesd össze a korábbival!

SZ1.) **[10.11.]** Tekintsük a bolyongást a Petersen-gráfon. Tegyük fel, hogy az éleken oda-vissza lehet közlekedni és egy adott pontból bármely szomszédos pontba azonos valószínűséggel lehet átmenni. Add meg az átmenetvalószínűségi mátrixot és osztályozd az állapotokat! Határozd meg a stacionárius eloszlást! (1p)

SZ2.) **[10.11.]** Legyen $\{X_k,\ k \geq 0\}$ homogén Markov-lánc $S = \{0, 1, 2\}$ állapottérrel, $X_0 = 1$, az átmenetvalószínűség mátrix $\begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0.2 & 0.8 \end{pmatrix}$. Mutasd meg, hogy az $Y_k = I(X_k \leq 1)$ folyamat nem Markov-lánc! (2p)

SZ3.) **[10.11.]** 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 150 ezer Ft. Ha a partnernek az elmúlt két év mindegyikében volt balesete, akkor évi 300 ezer Ft-ot fizet. Amennyiben az ügyfélnek az elmúlt két év egyikében volt balesete, akkor a biztosítási díj 200 ezer Ft. Ha a partnernek balesete volt az elmúlt évben, akkor 80% a valószínűsége, hogy az aktuális évben nem lesz balesete. Ha az ügyfélnek nem volt balesete az elmúlt évben, akkor 95% eséllyel nem 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)

Ismétlésre ajánlott: VL-ből 54–74

7.) Az előző két hónap során megfigyeltem, hogy a sarki zöldségesben a nektarin kilóját a minőségétől függően változó áron árulják. A megfigyelt értékesítési árak: 800, 1400, 1000, 1200, 1600, 1400, 1000, 1600, 1600, 1500, 1200, 1300, 1400, 1300. Az teljesen kiszámíthatatlan, hogy amikor betérek a zöldségesbe, épp milyen minőségű nektarint árulnak. Összesen 2000 forintot szánok nektarinra, jelölje Y azt a valószínűségi változót, hogy hány kiló nektarint tudok vásárolni.

a.) Hogyan modellezzük a nektarin árát? Alkalmas statisztikai próba végrehajtásával is mutasd meg, hogy a választott eloszlás megfelelő!

b.) Várhatóan hány kiló nektarint fogok tudni venni? Adj 95%-os konfidenciaintervallum-becslést is erre a mennyiségre!

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

Ott-hon	$\xrightarrow{\text{gyalog}}$ 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	**Egye-tem**

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!

9.) 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.

10.) 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!

11.) 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 kovarianciamátrixot és a korrelációs mátrixot!

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!

12.) 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.) $E(Y|X)$ =?

13.) 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, a 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!

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

B4.) [10.25.] Jolán a városi piacon értékesíti almáját. Mind az értékesítési mennyiség, mint az ár véletlennek tekinthető (lehet vele alkudozni). Az egy nap alatt eladott mennyiség (kg) egyenletes eloszlású 20 várható értékkel és 4 szórással, míg az eladott almák ára (Ft/kg) egyenletes eloszlású 450 és 550 paraméterekkel. Tegyük fel, hogy a mennyiség és az ár függetlenek egymástól. Költségei naponta 8 ezer forintot tesznek ki (benzinköltség).

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

b.) Határozd meg annak a valószínűségét (szimulációval is jó), hogy Jolán napi profitja meghaladja az 3 ezer forintot! (1+3= 4p)

B5.) [11.08.] Legyen X és Y együttes sűrűségfüggvénye $f_{X,Y}(x,y) = \frac{1}{16\pi} I(x^2 + y^2 < 16)$, valamint definiáljuk az $U = \operatorname{arctg}\left(\frac{Y}{X}\right)$, $V = \sqrt{X^2+Y^2}$ valószínűségi változókat. Határozd meg U és V együttes sűrűségfüggvényét! Független egymástól U és V?

Útmutatás: használjuk az $\frac{1}{\sqrt{\operatorname{tg}^2 x + 1}} = \cos x$ szögfüggvény-összefüggést!

B6.) [11.08.] Legyen X és Y együttes sűrűségfüggvénye $f_{X,Y}(x,y) = \frac{3}{4} \cdot I(0 < y < 1 - x^2 < 1)$. Határozd meg a $D^2(X|Y)$ feltételes szórásnégyzetet!

SZ4.) [11.08.] Legyenek X és Y egymástól független valószínűségi változók. Mutasd meg, hogy $U = XY$ sűrűségfüggvénye $f_U(u) = \int\limits_{-\infty}^{\infty} \frac{1}{|y|} f_X\left(\frac{u}{y}\right) f_Y(y)\, dy.$ (2p)

SZ5.) [11.08.] 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)

SZ6.) [11.08.] 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)

SZ7.) [11.08.] 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)

SZ8.) [11.08.] 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! (2p)

14.) Egy adott típusú LED izzó várható élettartama a műszaki leírása alapján 10 év. Az izzó ára 1000 Ft. A gyártónak az izzóra valamennyi garanciaidőt vállalni kell, ami azt jelenti, hogy amennyiben egy izzó ezen belül hibásodik meg, akkor köteles a vevő számára egy ugyanolyan típusú új izzót adni. Hány év legyen a garanciaidő, ha a gyártó azzal számol, 1 millió izzót fog egy év során eladni, illetve azt szeretné, hogy ez idő alatt 100 ezer Ft-nál kevesebb pénzt kelljen költenie meghibásodott garanciális izzók pótlására? (Keressük meg azt a garanciaidőt, amire a meghibásodásra költött pénz összessége nagy, például legalább 99%-os eséllyel 100 ezer Ft alatt marad.)

15.) 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 $P(U < V)$ és $P(-1 < U < 1, 0 < V < 3)$ valószínűségeket!

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

17.) 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.) Mi az esélye, hogy egy férfi 180 cm magas és 80 kiló?

b.) Mi az esélye, hogy egy férfi magasabb 180 cm-nél és 80 kilogramm a testsúlya?

c.) Mi az esélye, hogy egy férfi magasabb 180 cm-nél és könnyebb 80 kg-nál?

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

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

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

B7.) [11.15.] Legyen $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3\left(\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}\right)$.

a.) Milyen eloszlású az $\begin{pmatrix} X_3 \\ X_1 \end{pmatrix}$ valószínűségi vektorváltozó?

b.) Határozd meg $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ eloszlását, ahol $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}$ és $\mathbf{b} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

c.) $P(X_1 > 2,\ 0 < X_2 < 1,\ X_3 > 0) = ?$

d.) $E(X_1 | X_3 = -1) = ?$

B8.) [11.15.] 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 $-1 \leq \rho \leq 1$ valós szám? Az eredményt számold ki számítógép segítségével, készíts belőle alkalmas grafikont (x tengely: ρ; y tengely: a valószínűség)! Értékeld a látottakat – erre számítottál?

SZ9.) **[11.15.]** Legyen $(X,Y,Z)^T$ együttes sűrűségfüggvénye $f_{X,Y,Z}(x,y,z) = c\cdot\exp\left\{-\frac{1}{8}x^2-\frac{1}{2}y^2-\frac{1}{2}z^2+\frac{1}{4}x+\frac{1}{2}yz-\frac{1}{8}\right\}$. Határozd meg a c értékét és számítsd ki az $E(Y|X=0,Z=0)$ feltételes várható értéket! (2p)

SZ10.) **[11.15.]** Oldd meg a 16 feladatot a normális korreláció tételének felhasználása nélkül! (2p)

SZ11.) **[11.15.]** 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 46,86 mp-es világcsúcsát! (2p)

18.) Generálj 10^4 darab véletlen számot az

a.) inverziós módszerrel az $f(x)=\frac{3x^2}{56}I(2<x<4)$ sűrűségfüggvényű eloszlásból;

b.) elutasításos módszerrel az $f(x)=\frac{3}{8}\left(1+\frac{x^2}{4}\right)^{-5/2}$, $x\in\mathbb{R}$ sűrűségfüggvényű eloszlásból.

Ábrázold a véletlen számokat pontdiagramon és készíts szintvonalas ábrá(ka)t!

19.) Tekintsük az alábbi kétdimenziós sűrűségfüggvényt:

$$f_{X,Y}(x,y)=\begin{cases}\frac{10}{49}y & \text{ha } 1<x^2<y<4,\ x>0\\ 0 & \text{egyébként}\end{cases}$$

Generáljunk véletlen vektorokat ebből az eloszlásból, majd készítsünk többdimenziós sűrűséghisztogramot, valamint ábrázoljuk a magfüggvényes becslést és a szintvonalakat!

20.) Generálj

a.) $\mathrm{N}(1,2^2)$ és $\mathrm{N}(4,3^2)$ eloszlású valószínűségi változókat 0,9 korrelációval;

b.) $\mathrm{N}(1,3^2)$, $\mathrm{N}(2,2^2)$, $\mathrm{N}(3,1^2)$ eloszlású valószínűségi változókat úgy, hogy az első kettő közti korreláció 0,8, a második és a harmadik közti 0,5, valamint az első és a harmadik közötti 0,2.

A véletlen számokból készíts többdimenziós sűrűséghisztogramot, valamint ábrázold a magfüggvényes becslést és a szintvonalakat!

B9.) **[11.29.]** Tekintsük a következő sűrűségfüggvényt:

$f(x)=\frac{1}{2x\sqrt{\pi}}\exp\left\{-\frac{(\log x-m)^2}{4}\right\}I(x\geq 0)$, ahol m valós paraméter.

a.) Vezesd le n elemű minta esetén az m paraméter maximum likelihood becslését!

b.) Szimuláció segítségével készíts 95%-os konfidenciaintervallumot a 100 elemű mintából készített ML-becslésre! A véletlen számok generálását alkalmas tanult módszerrel végezd el!

Útmutatás: konfidenciaintervallum például úgy készíthető, ha sokszor, mondjuk 1000-szer állítasz elő 100 elemű mintát, amiből 1000 becslés készíthető az m paraméterre, ezekből pedig a megfelelő tapasztalati kvantiliseket kiszámolva kapható intervallumbecslés.

B10.) **[11.29.]** Töltsd le a honlapomról a B10.r fájlt és futtasd le R/RStudio segítségével.

a.) Írd le nagy vonalakban, mi történik a kódban! Értékeld ki az eredményeket!

b.) Módosítsd a `konturdiag` és `hist_2d` függvényeket úgy, hogy a `pontdiag` függvényhez hasonlóan az ábráknak legyen címe!

SZ12.) **[11.29.]** Tekintsük az alábbi eloszlásfüggvényt, ahol R valós paraméter:

$$F(x)=\begin{cases}0 & \text{ha } x\leq -R\\ \frac{3}{4}\frac{x}{R}-\frac{1}{4}\left(\frac{x}{R}\right)^3+\frac{1}{2} & \text{ha } -R<x\leq R\\ 1 & \text{ha } x>R\end{cases}$$

Vezesd le n elemű minta esetén R paraméter momentum becslését és szimuláció segítségével készíts 95%-os konfidenciaintervallumot a 100 elemű mintából készített momentum becslésre! A véletlen számok generálását alkalmas tanult módszerrel végezd el! (2p)

21.) Állítsuk elő a következő többdimenziós eloszlások kopuláját és kopula-sűrűségfüggvényét, illetve ábrázoljuk a kopula-sűrűségfüggvény szintvonalait.

a.) $F_{X,Y}(x,y)=\begin{cases}\frac{(x+1)(e^y-1)}{x+2e^y-1} & \text{ha } (x,y)\in[-1;1]\times(0,\infty)\\ 1-e^{-y} & \text{ha } (x,y)\in(1;\infty)\times(0,\infty)\\ 0 & \text{egyébként}\end{cases}$

b.) $f_{X,Y}(x,y)=I(0<y<2x<2)$

22.) Generáljunk egy 1000 elemű kétdimenziós mintát úgy, hogy a peremek standard normálisak legyenek, a köztük lévő korreláció pedig 0.3 legyen. Vegyünk be a mintába néhány kiugró (outlier) értéket, majd nézzük meg, ennek hatására mennyire változtak meg a Pearson-, Kendall- és Spearman-féle korrelációk!

Magyarázzuk az egyik dimenziót a másik dimenzió segítségével lineáris modellel és itt is vizsgáljuk meg az outlierek hatását!

23.) Tekintsük azt az Archimédészi kopulát, aminek kopulageneráló függvénye $\varphi(u)=(-\log u)^\theta$, ahol $\theta\geq 1$. Állítsd elő a kopulát és számold ki a Kendall-féle τ értékét.

24.) Generáljunk 1000 elemű kétdimenziós mintákat Gauss, 3 szabadságfokú Student t, Gumbel, Clayton kopulákból

a.) különböző paraméterek esetén;

b.) úgy, hogy megegyezik a Kendall-féle τ korrelációjuk!

Vessük őket össze a minták pontdiagramját!

25.) Generáljunk 1000 elemű kétdimenziós mintákat úgy, hogy a peremek eloszlása 1 paraméterű exponenciális és 5 szabadságfokú t-eloszlású legyen, a kopulájuk pedig θ paraméterű Clayton-kopula legyen. Ábrázoljuk a minták pontdiagramját különböző θ paraméterértékek esetén!

26.) Legyen X és Y lognormális eloszlású 2 és 0,6 paraméterekkel, kopulájuk pedig legyen Gauss, 3 szabadságfokú Student t, Gumbel, Clayton kopulákból úgy, hogy megegyezik a Kendall-féle τ korrelációjuk! Számítsuk ki mindegyik esetben a $P(X<4,\ Y<4)$, $P(X>13,\ Y>13)$ és $P(X<4,\ Y>13)$ valószínűségeket!

27.) Generáljunk 1000 elemű mintát 2 szabadságfokú Student t kopulából $0,7$ korreláció esetén. Illesszünk Gauss, Student t, Gumbel és Clayton kopulákat a mintára (becsüljük meg a kopula paramétereit), majd végezzünk illeszkedésvizsgálatot!

28.) Legyen X és Y független t_d eloszlású, $r = -0,8$. Legyen $Z = rX + \sqrt{1-r^2}Y$. Teszteljük különböző d értékek esetén alkalmas statisztikai próbával, hogy X és Z kopulája d szabadságfokú t-kopula-e.

29.) Tekintsük a MagyarReszvenyek.xlsx fájlban lévő mintát, amely Budapesti Értéktőzsdén nyíltan kereskedett néhány kiemelt részvény árfolyamadatait tartalmazza 2017-től kezdődően.

a.) Olvassuk be az adatokat, hozzuk őket az elemzéshez alkalmas formátumokra!

b.) Tekintsük az OTP és a MOL adatokat 2017-től 2022-ig (5 évnyi adatmennyiség).

I.) Vizsgáljuk meg az adatokat a tanult vizuális módszerekkel és mutatókkal egyaránt, szükség esetén végezzünk alkalmas transzformációt az eredeti adatokon!

II.) Illesszünk különböző kopulákat, vizsgáljuk meg a paraméterbecslés stabilitását tolóablakos módszerrel! Adjunk intervallumbecslést is a paraméter(ek)re!
Tolóablakos módszer: paramétert becslünk egy bizonyos intervallumra, majd ezt az intervallumot kicsit eltoljuk előre az időben és újra paramétert becslünk, majd újra megismételjük. A végén a becsült paraméterek stabilitását megvizsgáljuk, ábrázoljuk.

III.) Végezzünk illeszkedésvizsgálatot – melyik kopula illeszkedik elfogadhatóan az adatainkra?

IV.) Illesszünk megfelelően illeszkedő eloszlásokat a peremekre! Próbálkozzunk például átskálázott t-eloszlással, azaz ha X t-eloszlást követ d szabadságfokkal, legyen $Y = \sigma \cdot X$, $\sigma > 0$, ezt az Y-t nevezzük skálázott t-eloszlásnak (paraméterei: σ, d). A becslést például elvégezhetjük ML-becsléssel, a loglikelihood-függvény (-1)-szeresét minimalizálhatjuk az R-ben beépített `nlm` rutinnal.

V.) Becsüljük meg annak a valószínűségét, hogy egy nap alatt mind az OTP, mind a MOL részvények l %-ot esnek, ahol $l = 1, 2, \ldots, 10$. Vessük össze az értékeket azzal, amikor (tévesen) normális kopulával és normális peremekkel dolgozunk!

c.) Oldjuk meg az előző pontban kérdéseit az OTP, MOL, Richter adatokra együttesen!

B11.) **[12.06.]** Legyen $\theta \in [-1; 1]$ valós paraméter. Mutassuk meg, hogy az $F_{X,Y}(x, y) = (1 - e^{-x} - e^{-y} + (1-\theta)e^{-x-y})^{-1}$ együttes eloszlásfüggvénnyel definiált eloszlás kopulája $C(u, v) = \frac{uv}{1-\theta(1-u)(1-v)}$! Van olyan θ, amelyre X és Y független egymástól?

B12.) **[12.13.]** Legyen X és Y normális eloszlású 10 és 2 paraméterekkel, kopulájuk pedig legyen Clayton kopula úgy, hogy a Kendall-féle τ korreláció értéke $0,5$ legyen! Számítsd ki a $P(X < 7,\ Y < 7)$, $P(X < 13,\ Y > 7)$ és $P(X > 12,\ Y > 13)$ valószínűségeket!

SZ13.) **[12.13.]** Legyen X_1 és X_2 független azonos, abszolút folytonos eloszlású F eloszlásfüggvénnyel. Mutassuk meg, hogy $(X_1^*, X_2^*)^T$ kopulája az alábbi függvény:

$$C(u_1, u_2) = 2\sqrt{u_1}F(\min(F^{-1}(1-\sqrt{1-u_1}), F^{-1}(\sqrt{u_2}))) - \\ - [F(\min(F^{-1}(1-\sqrt{1-u_1}), F^{-1}(\sqrt{u_2})))]^2,$$

ahol $X_1^* = \min(X_1, X_2)$, $X_2^* = \max(X_1, X_2)$. (3p)

SZ14.) **[12.13.]** Határozd meg $(Y,\ 1-Y)^T$ kopuláját, ahol $Y \sim E(0; 10)$. (2p)

SZ15.) **[12.13.]** Legyenek X, Y és Z normális eloszlásúak 9 és 3 paraméterekkel, kopulájuk pedig legyen 5 szabadságfokú t kopula úgy, hogy a Kendall-féle τ korreláció értéke $0,7$ legyen! Számítsd ki a $P(X < 5,\ Y < 5,\ Z < 5)$, $P(X < 5,\ 5 < Y < 13,\ Z > 13)$ és $P(X > 13,\ Y > 13,\ Z > 13)$ valószínűségeket! (2p)

30.) 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!

d.) Szignifikáns hatással van a hallgatók neme a testmagasságukra?

e.) Szignifikáns hatással van a hallgatók tavalyi vizsgajegye a testmagasságukra?

31.) Egy ügyvédi vállalkozásnak két irodája van: egy Budapesten és egy Miskolcon. 2023-ban Budapesten 4 dolgozó bruttó fizetése 1000 e Ft, 1200 e Ft, 1100 e Ft, 1020 e Ft volt. A miskolci telephelyen dolgozók közül 3-an bruttó 900 e, 700 e, 800 e Ft-ot vittek haza.

a.) A telephely hány %-ban magyarázza a fizetések változékonyságát?

b.) A telephely szignifikáns hatással van a fizetésekre?

c.) Hogyan változik az előző két kérdésre adott válasz, ha Győrben is van egy telephely és két ott dolgozó fizetése 900 e Ft és 1100 e Ft?

32.) A távolsági autóbusz-közlekedés néhány jellemző adata egy év során:

Járat	Szállított utasok száma (fő)	Utazási távolság (km/fő) átlaga	szórása
Menetrend szerinti	450	17	5
Szerződéses	30	23	15
Külön	3	151	70
Összesen/együtt	...	...	...

a.) Számítsd ki a táblázat kipontozott celláit!

b.) A járat fajtája szignifikáns hatással van az utazási távolságra?

33.) A következő táblázat a 2016. 1. félévben tanár szakos BSc-s hallgatóknak tartott 4 valszám gyakorlat év végi, 100-ra skálázott végső pontszámait tartalmazza:

Gyakvezér	Pontszámok										
Cs. V.	98	87	102	92	52	46	95	60	81	55	60	94
	81	58	80	93	70	66	49	94	50	88	74	
W. G.	77	46	54	57	50	45	39	63	26	107	75	
	66	52	109	91	35	65						
B. Á.	86	94	54	61	42	59	88	81	81	80	102	72
	88	96	58	90	110	58	80	90	84	80	94	
V. L.	66	60	72	49	52	54	80	56	36	91	68	
	60	51	40	38	54	62						

Vizsgáljuk meg, az év végi pontszám függ-e attól, hogy a hallgató melyik csoportba jár! Hány %-ban magyarázza a pontszámok változékonyságát az, hogy a hallgatók melyik csoportba járnak?

34.) Tekintsük a `https://stats.idre.ucla.edu/stat/data/binary.csv` linken elérhető amerikai egyetemi felvételi adatokat és próbáljuk meg modellezni a felvétel eredményét (valószínűségét) a rendelkezésre álló magyarázó változókkal! Értelmezzük a kapott eredményeket!
A modell alapján milyen eséllyel vesznek fel egy diákot, aki egy 1-es rangú iskolából jön, a jegyeinek átlaga 3,8 és a felvételin 790 pontot ért el?