Statisztika gyakorlat

Alkalmazott matematikus szakirány

Játékszabályok

- Az órákon részt kell venni, maximum 3-szor lehet hiányozni. Aki többször hiányzik, nem kap gyakjegyet.
- 100 + x pontot lehet szerezni a félév során:
  · 40 pont: 1. ZH a félév közepén
  · 40 pont: 2. ZH a félév végén
  · 20 pont: beadandó feladatokkal (2 pont · 10)
  · x pont: szorgalmi feladatokkal
- Mindkét ZH-n minimálisan teljesíteni kell a 30 %-ot, azaz a 12 pontot.
- Ha egy ZH sikertelen, nem írod meg, vagy javítani szeretnél, akkor vizsgaidőszak első hetén lesz lehetőség a pótZH megírására vagy a javításra. Csak az egyik ZH anyagából javíthatsz, és a jobbik eredményt veszem figyelembe, azaz nem lehet rontani. Két sikertelen vagy meg nem írt ZH esetén gyakUV-t írsz, és maximum kettest kaphatsz.
- A ZH-kon a kiosztott táblázatokon kívül használni lehet egy A4-es lapra (akár mindkét oldalára) KÉZZEL írott "puskát", valamint számológépet.
- Beadandók: Mindegyik maximálisan 2 pontot ér, a **legjobb 10-et** 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.
- Osztályozás:

1	0	–	34,99
2	35	–	49,99
3	50	–	64,99
4	65	–	79,99
5	80	–	1000

Személyes adatok

Név	Varga László
Tanszék	Valószínűségelméleti és Statisztika Tanszék (ELTE TTK)
Szoba	D 3-309
E-mail	`vargal4@cs.elte.hu`
Honlap	`www.cs.elte.hu/~vargal4`

Ajánlott irodalom

- Bolla–Krámli: Statisztikai következtetések elmélete
- Móri–Szeidl–Zempléni: Matematikai statisztikai feladatok

1.) Jelölje F^{-1} az egydimenziós eloszlásfüggvény általánosított inverzét. Mutasd meg, hogy

a.) minden $0 < y < 1$ esetén $F\left(F^{-1}(y)\right) \le y \le F(F^{-1}(y)+0)$;

b.) ha $U \sim E(0,1)$, akkor $F^{-1}(U)$ eloszlásfüggvénye éppen F;

c.) ha az X valószínűségi változó F eloszlásfüggvénye folytonos, akkor $F(X)$ eloszlása a $(0,1)$ intervallumon egyenletes.

2.) Legyen $Y_1,\dots,Y_n$ a $(0,1)$ intervallumon egyenletes eloszlásból vett minta.

a.) Milyen eloszlású Y_k^*?

b.) Határozd meg a rendezett minta várható értékét!

c.) Legyen $Z_1 = \frac{Y_1^*}{Y_2^*}, Z_2 = \left(\frac{Y_2^*}{Y_3^*}\right)^2, \dots, Z_{n-1} = \left(\frac{Y_{n-1}^*}{Y_n^*}\right)^{n-1}, Z_n = (Y_n^*)^n$. Mutasd meg, hogy Z_1, $Z_2,\dots,Z_n$ eloszlása megegyezik Y_1, $Y_2,\dots,Y_n$ eloszlásával!

3.) Legyen $X_1^* < X_2^* < \dots < X_n^*$ az f sűrűségfüggvényű és F eloszlásfüggvényű abszolút folytonos eloszlásból vett rendezett minta.

a.) Határozd meg X_k^* eloszlás- és sűrűségfüggvényét!

b.) Határozd meg $X_1^*, X_2^*, \dots, X_n^*$ együttes sűrűségfüggvényét!

4.) Legyen $X_1,\dots,X_n$ exponenciális eloszlásból vett minta, és $Y_1 = nX_1^*, Y_2 = (n-1)(X_2^*-X_1^*), \dots, Y_{n-1} = 2(X_{n-1}^* - X_{n-2}^*), Y_n = X_n^* - X_{n-1}^*$.

a.) Mutasd meg, hogy Y_1, $Y_2,\dots,Y_n$ eloszlása megegyezik X_1, $X_2,\dots,X_n$ eloszlásával!

b.) Határozd meg az X_1^*, $X_2^*,\dots,X_n^*$ rendezett minta várható értékét és kovarianciamátrixát!

5.) Legyen $X \sim$Exp(λ). Határozd meg X móduszát és tetszőleges kvantilisét! Hasonlítsd össze a mediánt és a várható értéket!

6.) Legyen $X \sim$Ind(p). Határozd meg X móduszát és kvantilisfüggvényét!

B1.) **[II.25.]** Legyen $Y_1,\dots,Y_n$ a $(0,1)$ intervallumon egyenletes eloszlásból vett minta. Határozd meg Y_k^* $(1 \le i \le n)$ tetszőleges momentumát és szórását!

B2.) **[II.25.]** Legyen $X \sim$Cauchy$(0,1)$. Határozd meg X móduszát, várható értékét és kvantilisfüggvényét! Hasonlítsd össze a mediánt és a várható értéket!

SZ1.) Határozd meg a $\mathrm{Bin}(3,p)$ eloszlás móduszát és kvantilisfüggvényét! (1 pont)

SZ2.) Határozd meg a lognormális eloszlás móduszát és kvantilisfüggvényét! (1 pont)

SZ3.) Legyen $X_1^* < X_2^* < \cdots < X_n^*$ az f sűrűségfüggvényű és F eloszlásfüggvényű abszolút folytonos eloszlásból vett rendezett minta. Mutasd meg, hogy az $X_i^* = t$ feltétel mellett $X_1^*, \ldots, X_{i-1}^*$ és $X_{i+1}^*, \ldots, X_n^*$ független, az előbbi eloszlása olyan, mint egy $i-1$ elemű rendezett mintáé az $f_1(x) = \frac{f(x)}{F(t)}\chi_{(x<t)}$ $(x \in \mathbb{R})$ sűrűségfüggvényű eloszlásból, míg az utóbbi eloszlása olyan, mint egy $n-i$ elemű rendezett mintáé az $f_2(x) = \frac{f(x)}{1-F(t)}\chi_{(x<t)}$ $(x \in \mathbb{R})$ sűrűségfüggvényű eloszlásból. (2 pont)

SZ4.) Definiáljuk egy X valószínűségi vektorváltozónak az $\mathcal{F}$ σ-algebrára vonatkozó feltételes kovarianciamátrixát a
$\Sigma(X|\mathcal{F}) = E\left\{[X - E(X|\mathcal{F})][X - E(X|\mathcal{F})]^\top \middle| \mathcal{F}\right\}$ formulával. (Könnyen látható, hogy $\Sigma(X|\mathcal{F}) = E(XX^\top|\mathcal{F}) - E(X|\mathcal{F})E(X|\mathcal{F})^\top$.) Bizonyítsuk be a teljes szórásnégyzet tételét: $\Sigma(X) = E\left(\Sigma(X|\mathcal{F})\right) + \Sigma\left(E(X|\mathcal{F})\right)$! (2 pont)

SZ5.) Legyen $X_1^* < X_2^* < \cdots < X_n^*$ rendezett minta abszolút folytonos eloszlásból. Számítsuk ki $i < j$ esetén az (X_i^*, X_j^*) együttes sűrűségfüggvényét. Tegyük fel, hogy egyenletes mintából indultunk ki, számítsuk ki $\mathrm{Cov}(X_i^*, X_j^*)$-t! (3 pont)

7.) Keressünk elégséges statisztikát a következő eloszláscsaládokból vett n elemű minta esetén, és ahol tudjuk, írjuk fel az kapott elégséges statisztika eloszlását is.

a.) $\mathrm{Ind}(p)$, $0 < p < 1$ paraméter,
b.) $\mathrm{Bin}(r,p)$, $r \geq 1$ egész ismert, $0 < p < 1$ paraméter,
c.) $\mathrm{Geo}(p)$, $0 < p < 1$ paraméter,
d.) $\mathrm{NegBin}(r,p)$, $r \geq 1$ egész ismert, $0 < p < 1$ paraméter,
e.) diszkrét egyenletes az $\{1, 2, \ldots, N\}$ halmazon, $N \geq 1$ egész paraméter,
f.) $\mathrm{Exp}(\lambda)$, $\lambda > 0$ paraméter,
g.) $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ és $\sigma > 0$ paraméterek,
h.) $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ ismert, $\sigma > 0$ paraméter,
i.) $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ paraméter, $\sigma > 0$ ismert,
j.) $E(a,b)$, $-\infty < a < b < \infty$ paraméterek,
k.) $E(-\vartheta, \vartheta)$, $\vartheta > 0$ paraméter,
l.) $E(\vartheta, 2\vartheta)$, $\vartheta > 0$ paraméter,
m.) $E(-\vartheta, 2\vartheta)$, $\vartheta > 0$ paraméter.
n.) $\mathrm{Béta}(a,b)$, $a, b > 0$ paraméterek,
o.) $\mathrm{Béta}(a,b)$, $a > 0$ ismert, $b > 0$ paraméter,
p.) $\mathrm{Béta}(a,b)$, $a > 0$ paraméter, $b > 0$ ismert,
q.) $\mathrm{Cauchy}(x_0)$, $x_0 \in \mathbb{R}$ paraméter.

8.) Legyen $X_1, ..., X_{20}$ i.i.d. minta $N(m, 1^2)$ eloszlásból. Célunk az ismeretlen m paraméter becslése. Tekintsük az alábbi három statisztikát:

- $T_1(\mathbf{X}) = X_8$,
- $T_2(\mathbf{X}) = \frac{X_3 + X_7}{2}$,
- $T_3(\mathbf{X}) = \frac{X_9 + X_{19}}{8}$.

a.) A fenti statisztikák közül melyek torzítatlanok? Amelyik nem torzítatlan, hogyan tudnánk torzítatlanná tenni?
b.) Vizsgáljuk meg a fenti statisztikák közül a torzítatlanokat hatásosság szempontjából!

9.) Legyen $X_1, ..., X_n$ i.i.d. minta ismeretlen eloszlásból.

a.) Torzítatlan becslés-e a várható értékre nézve az átlag?
b.) Torzítatlan becslés-e a szórásnégyzetre nézve a tapasztalati szórásnégyzet? Amennyiben nem az, hogyan tudnánk torzítatlanná tenni?
c.) Mikor konzisztens becslése a várható értéknek az átlag?
d.) Adjunk torzítatlan és konzisztens becslést az eloszlásfüggvényre!

10.) n elemű λ-paraméterű exponenciális minta esetén adjunk torzítatlan becslést $e^{-3\lambda}$-ra és $\frac{1}{\lambda}$-ra!

11.) n elemű λ-paraméterű Poisson minta esetén adjunk torzítatlan becslést $e^{-\lambda}$-ra és λ^2-re!

12.) Adjunk torzítatlan becslést a $[0,\theta]$ intervallumon egyenletes eloszlás paraméterére

a.) a mintaátlag
b.) a maximum
c.) a minimum

segítségével. Hasonlítsuk őket össze hatásosság szempontjából! Melyik becslés konzisztens?

13.) Mutassuk meg, hogy exponenciális eloszlású minta esetén $T(\mathbf{X}) = n \cdot \min(X_1, ..., X_n)$ statisztika torzítatlan, de nem konzisztens becslése a várható értéknek!

14.) Adjunk torzítatlan becslést $\mathrm{Bin}(2,p)$ eloszlásból származó i.i.d. minta esetén $\frac{1}{p}$-re!

15.) Legyen $X_1, ..., X_n$ i.i.d. minta valamely véges szórású eloszlásból, és tekintsük a $\mathrm{T}(\mathbf{X}) = a_1X_1 + ... + a_nX_n$ alakú lineáris becsléseket, ahol $a_1, ..., a_n \in \mathbb{R}$. Feltéve, hogy $\mathrm{T}(\mathbf{X})$ a várható érték torzítatlan becslése, mely $a_1, ..., a_n$ számokra lesz minimális a $D^2(T(\mathbf{X}))$?

B3.) [III.11.] Tegyük fel, hogy a valószínűségszámítás vizsgán 6-an vettek részt. Az első 5 vizsgázó jegyei: 2, 3, 4, 4, 5.

a.) Adj torzítatlan becslést az 5 megfigyelés alapján a kapott jegy szórásnégyzetére!

b.) A hatodik vizsgázó mely érdemjegyére lesz a 6 jegy korrigált tapasztalati szórásnégyzete a legnagyobb, illetve a legkisebb?

B4.) [III.11.] Keressünk minimális elégséges statisztikát a következő eloszláscsaládokból vett n elemű minta esetén, és ahol tudjuk, írjuk fel az kapott elégséges statisztika eloszlását is.

a.) $\mathrm{Poi}(\lambda)$, $\lambda > 0$ paraméter,

b.) $\mathrm{Gamma}(\alpha, \lambda)$, $\alpha > 0$ ismert, $\lambda > 0$ paraméter!

B5.) [III.11.] Torzítatlan becslés-e a mintaátlag reciproka az exponenciális eloszlás paraméterére? Hogyan tudnánk torzítatlanná tenni? Konzisztens a becslés? (2 pont)

SZ6.) 8 véletlen számot jegyeztünk fel: 10,22,39,66,52,99,17,75. Ha tudjuk, hogy ezek az $\{1, 2, ..., N\}$ halmazból vett véletlen minta elemei, akkor hogyan becsülnéd az N paramétert? (1 pont)

SZ7.) A Négyfülű Fülemüle párt kíváncsi rá, hogy milyen a támogatottsága a lakosság körében, ezért megbíz egy közvélemény-kutató vállalatot, hogy kérdezzen meg 1000 embert. A közvéleménykutató azt találja, hogy a megkérdezettek közül 200-an szavaznának a Négyfülű Fülemülére. Becsüld meg a párt támogatottságát és annak szórását? Adj felső becslést a szórásra! (2 pont)

SZ8.) Legyen $X_1, ..., X_n$ i.i.d. minta Bin(k,p)-ből, $Y_1, ..., Y_n$ i.i.d. minta Bin(l,p)-ből, és tegyük fel, hogy a két minta egymástól is független. Milyen (a, b) számpárokra lesz $a\overline{X} + b\overline{Y}$ a p paraméter torzítatlan becslése? Ezen számpárok közül melyikre lesz a becslés szórása minimális? (2 pont)

16.) Határozzuk meg az ismeretlen paraméter(ek) ML becslését, ha a minta

a.) $\mathrm{Exp}(\lambda)$ eloszlású;

b.) $\mathrm{Poi}(\lambda)$ eloszlású.

c.) $\mathrm{E}(a, b)$ eloszlású, ahol $a < b$, mindkettő paraméter;

d.) $\mathrm{N}(m, \sigma^2)$ eloszlású, m és σ paraméterek.

Torzítatlan a becslés? Ha nem az, próbáljuk torzítatlanná tenni! Konzisztens a becslés?

17.) Legyen $X_1, ..., X_n$ Pascal-eloszlású (geometriai eloszlású) minta p paraméterrel.

a.) Adjunk meg X_3 függvényeként torzítatlan becslést $p(1-p)^4$-re!

b.) Adjunk maximum likelihood becslést $p(1-p)$-re!

18.) Legyenek X_1,...,X_n és Y_1,...,Y_m egymástól független λ illetve $\frac{1}{\lambda}$ paraméterű exponenciális eloszlású minták. Határozzuk meg az ismeretlen paraméter együttes ML becslését!

19.) Tegyük fel, hogy a minta kétparaméteres eloszláscsaládból származik, a paraméterek a és b.

Ekkor mutassuk meg, hogy az $\begin{cases} E_{a,b}X &= m_1 \\ E_{a,b}X^2 &= m_2 \end{cases}$ egyenletrendszer megoldása megegyezik az $\begin{cases} E_{a,b}X &= m_1 \\ D^2_{a,b}X &= s_n^2 \end{cases}$ egyenletrendszer megoldásával.

20.) Becsüld a paramétert momentum-módszerrel, ha a minta eloszlása:

a.) $\mathrm{Exp}(\lambda)$;

b.) $\mathrm{Poi}(\lambda)$;

c.) $\mathrm{E}(a, b)$;

d.) $\mathrm{E}(-a, a)$;

e.) $\mathrm{N}(2m + 5, (\frac{1}{d})^2)$.

21.) Legyen az $X_1, \ldots, X_n$ minta a következő diszkrét eloszlásból: $P(X_1 = 1) = c$, $P(X_1 = 2) = 3c$, $P(X_1 = 3) = 1 - 4c$ (c az ismeretlen paraméter). Tegyük fel, hogy az n mintaelemből y_i darab veszi fel az i értéket (i=1,2,3).

a.) Határozzuk meg c momentum-becslését!

b.) Határozzuk meg c ML-becslését!

22.) Adott egy n elemu minta az $\mathrm{E}(0, b)$ eloszlásból.

a.) Adjunk maximum likelihood becslést b-re!

b.) Tegyük ezt a becslést torzítatlanná!

c.) Mutassuk meg, hogy konzisztens becsléshez jutottunk!

23.) Egy gyárban a termékek minőségét úgy ellenőrzik, hogy minden nap n terméket vizsgálnak meg. Az adott napi gyártmányokat (nevezzük ezek összességét tételnek) akkor fogadják el, ha minden egyes megvizsgált gyártmány jó. m nap után azt tapasztalták, hogy x tételt fogadtak el. Adjunk maximum likelihood becslést annak p valószínuségére, hogy egy termék selejtes (tegyük fel, hogy minden termékre ugyanez a valószínűség, és hogy az egyes termékek minősége független egymástól).

24.) Egy halastóból kifogtunk n db halat, megjelöltük és visszadobtuk őket. Ezután visszatevés nélkül kifogunk m db halat, melyek közül x jelölt és $m-x$ jelöletlen. Adjunk maximum likelihood becslést a halastóban található jelöletlen halak N számára!

B6.) [III.31.] Adjunk a Béta$(\vartheta, 1)$ eloszlású n elemű minta esetén becslést a ϑ ismeretlen paraméterre momentum-, és ML-módszerrel!

SZ9.) Legyenek $X_{i,j}$-k ($i = 1, \dots, n$ és $j = 1, \dots, r > 1$) független $N(\mu_i, \sigma^2)$ eloszlásúak. Határozd meg $\vartheta = (\mu_1, \dots, \mu_n, \sigma^2)^T$ ML-becslését! (1 pont)

SZ10.) Rendelkezésünkre áll a következő minta Cauchy(a, b) eloszlásból, ahol a és b ismeretlen paraméterek: 0,89; 2,52; -0,55; -6.73; 0,85; 198,4; 0.03; 1.84; 1,14; -30,18. Add meg az ismeretlen paraméterek ML-becslését, az egyenletrendszer megoldásához használj alkalmas szoftvert és küldd el a forráskódot E-mail-ben! (2 pont)

SZ11.) Legyen $(X_1, Y_1), \dots, (X_n, Y_n)$ i.i.d. minta kétdimenziós normális eloszlásból, melynek peremei standard normálisak, az ismeretlen korrelációt jelölje $\rho \in (-1, 1)$. Adjuk meg ρ ML-becslését! Mutassuk meg, hogy az ML-becslés 1-hez tartó valószínűséggel egyértelmű! (2 pont)

25.) Számítsuk ki a Fisher-információt a következő eloszláscsaládokból vett n elemű minta esetén:

a.) Poi(λ), $\lambda > 0$ paraméter,
b.) Bin(r, p), $r \geq 1$ ismert, $0 < p < 1$ paraméter,
c.) Exp(λ), $\lambda > 0$ paraméter,
d.) $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ és $\sigma > 0$ paraméterek,
e.) $N(\vartheta, \vartheta^2)$, $\vartheta \in \mathbb{R}$ paraméter,
f.) $E(0, \vartheta)$, $\vartheta > 0$ paraméter,
g.) folytonosan differenciálható f sűrűségfüggvénnyel képzett eltolás- és skálaparaméteres család: $f_{\mu,\sigma}(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$ $(x \in \mathbb{R})$, ahol $\mu \in \mathbb{R}$ eltolásparaméter és $\sigma > 0$ skálaparaméter.

26.) Egy szövőszéken n orsó forog, a szálszakadásig eltelő idő eloszlása mindegyiken ugyanolyan exponenciális eloszlású. Az eloszlás paraméterét szeretnénk becsülni, ezért a $t > 0$ időpontban megfigyeljük, hány orsón nem szakadt még el a szál.

a.) Melyik t maximalizálja a megfigyelés Fisher-információját?
b.) Ha a t időpontban még X orsón nem volt szakadás és még egy megfigyelésre van lehetőségünk, azt mikor végezzük?

27.) n elemű mintát veszünk a $0 < p < 1$ paraméterű indikátoreloszlásból.

a.) Adjunk $g(p) = p^k$-ra egyszerű torzítatlan becslést.
b.) Keressünk teljes elégséges statisztikát.
c.) Blackwellizálással adjunk hatásos becslést a paraméter minden olyan függvényére, amely torzítatlanul becsülhető.

28.) $\lambda > 0$ paraméterű Poisson-eloszlásból vett n elemű minta esetén adjunk hatásos becslést a $g(\lambda) = e^{-\lambda}$ mennyiségre.

29.) A $(0, \vartheta)$ intervallumon egyenletes eloszlásból vett n elemű minta esetén blackwellizáljuk a $\vartheta > 0$ paraméter $T(X) = (n+1)X_1^*$ becslését. Hatásos?

30.) Legyenek $X \sim \Gamma(\alpha, \lambda)$, $Y \sim \Gamma(\beta, \lambda)$ függetlenek! Mutassuk meg, hogy $U = X + Y$ és $V = \frac{Y}{X+Y}$ függetlenek, $U \sim \Gamma(\alpha + \beta, \lambda)$ és $V \sim$ Beta(α, β)!

31.) Exponenciális eloszlásból vett n elemű minta esetén

a.) teljes-e a mintaelemek összege? (Útmutatás: egy $X \geq 0$ valószínűségi változó *Laplace-transzformáltja* a $[0, \infty) \ni t \mapsto Ee^{-tX} \in (0, \infty)$ függvény. Megmutatható, hogy a Laplace-transzformált egyértelműen meghatározza az eloszlást.)
b.) Mekkora a paraméterre vonatkozó információs határ?
c.) A $T(X) = \frac{n-1}{X_1 + \cdots + X_n}$ torzítatlan becslés eléri-e az információs határt?
d.) Lehet-e egyetlen mintaelemből torzítatlanul becsülni a paramétert?
e.) Számítsuk ki a $T_c(X_1) = \frac{1}{c} I(X_1 < c)$ becslés blackwellizáltjának a határértékét $c \searrow 0$ esetén!

32.) $\lambda > 0$ paraméterű exponenciális eloszlásból vett minta esetén blackwellizálással adjunk jó minőségű torzítatlan becslést a $g(\lambda) = e^{-c\lambda}$ függvényre ($c > 0$ konstans).

B7.) [III.24.] Számítsuk ki a Fisher-információt a Cauchy$(\theta, 1)$ eloszlású n elemű minta esetén!

B8.) [III.24.] Geometriai eloszlás $0 < p < 1$ paraméterére adjunk hatásos becslést n elemű mintából blackwellizálás segítségével.

SZ12.) Számítsuk ki az n elemű mintában rejlő Fisher-információt, ha a mintaelemek közös sűrűségfüggvénye a következő:
$f_\vartheta(x) = \frac{x}{\vartheta^2} e^{-\frac{x^2}{2\vartheta^2}} I(x > 0)$, ahol $\vartheta > 0$ valós paraméter! (1 pont)

SZ13.) Legyen $X_1, \dots, X_n$ a $(0, \vartheta)$ intervallumon egyenletes eloszlású minta. Számítsuk a ϑ ismeretlen paraméter $T(\mathbf{X}) = 2X_1$ becslésének blackwellizáltját! (1 pont)

SZ14.) Legyen $X_1, \dots, X_n$ a $(\vartheta, 1 + \vartheta)$ intervallumon egyenletes eloszlású minta. Teljes az $S(\mathbf{X}) = (X_1^*, X_n^*)$ statisztika? (2 pont)

SZ15.) Legyenek $X_1, \dots, X_n$ i.i.d. minta a diszkrét $\{1, 2, \dots, N\}$ halmazon.

Adj torzítatlan becslést N-re egy mintaelem (például X_1) segítségével, majd blackwellizálással készíts hatásos becslést! (3 pont)

33.) Legyen $X_1, \ldots, X_n$ a λ paraméterű exponenciális eloszlásból származó minta $(0 < \lambda < 1)$. Számítsuk ki $g(\lambda)$ Bayes-becslését gamma apriori eloszlás esetén, ha
a.) $g(\lambda) = \lambda$;
b.) $g(\lambda) = \lambda^2$.
c.) $g(\lambda) = e^{2\lambda}$.

34.) Legyen $X_1, \ldots, X_n$ a ϑ paraméterű geometriai eloszlásból származó minta $(0 < \vartheta < 1)$. Adjunk Bayes-becslést ϑ-ra, ha az apriori eloszlás egyenletes a paramétertéren.

35.) Béta$(\vartheta, 1)$ eloszlás esetén becsüljük a paramétert
a.) momentum-módszerrel,
b.) maximum likelihood módszerrel,
c.) Bayes-módszerrel, ha az apriori eloszlás gamma!

36.) Valaki feldobott egy marék aprót és a fejre esett érméket nekünk adta. Összesen 13 pénzdarabot kaptunk. Adjunk Bayes-becslést a feldobott érmék számára, ha az apriori eloszlás
a.) Poi(6),
b.) Geo$\left(\frac{1}{6}\right)$.

B9.) [IV.14.] Legyen $X_1, \ldots, X_n$ a $\left(\vartheta - \frac{1}{2}, \vartheta + \frac{1}{2}\right)$ intervallumon egyenletes eloszlású minta. Határozd meg a ϑ paraméter Bayes-becslését a $(13, 10, 7, 19, 21, 17)$ minta alapján, ha ϑ apriori eloszlása E$(10, 20)$.

B10.) [IV.14.] Az előző évszázad 80-as, 90-es éveiben az adattárolás eszközeként szolgáltak a hajlékonylemezek (floppy-k), melyek egyik hátránya az volt, hogy hamar meghibásodtak. Egy év használat után megfigyeltek 5 kislemezt, amelyeken 3, 1, 4, 6 és 2 szektorhibát találtak. Az IBM mérnökei úgy találták, hogy a szektorhibák száma 1 év használat után Poisson eloszlást követ, jelölje az ismeretlen paramétert λ. Adjuk meg a paraméter Bayes-becslését, amennyiben az apriori eloszlás a következő:
$P(\lambda = 1) = 0,4 \quad P(\lambda = 1,5) = 0,6$.

SZ16.) Legyen $X_1, \ldots, X_n$ a $\mathrm{N}(m, \sigma^2)$ eloszlású minta, ahol m ismert, σ pedig ismeretlen paraméterek. Határozd meg az m paraméter Bayes-becslését, ha m apriori eloszlása $\mathrm{N}(\mu, \tau^2)$! (1 pont)

SZ17.) Legyen $X_1, \ldots, X_n$ Cauchy$(\vartheta, 1)$ eloszlású minta. Határozd meg a ϑ paraméter Bayes-becslését a $(-3, 4, 0, 15, -7, 10)$ minta alapján, ha ϑ apriori eloszlása $\mathrm{N}(10, 1^2)$. Az integrálok kiszámításához használj alkalmas szoftvert és küldd el a forráskódot E-mail-ben! (2 pont)

37.) Valaki azt állítja, hogy a klíma változik, és ezt azzal véli bizonyítottnak, hogy az elmúlt 10 évben 2-szer is volt jégeső, pedig korábban az egyes évekre a jégeső valószínűsége a hivatalos adatok alapján csupán $p = 0,1$ volt. Írjuk fel a hipotéziseket, a próbát és állapítsuk meg az első-, másodfajú hibák valószínűségét, valamint az erőfüggvényt a $p = 0,2$ pontban!

38.) Az X valószínűségi változó egyenletes eloszlású a $(-b; 1 + 2b)$ intervallumon. A H_0: $b = 0$ hipotézist szeretnénk ellenőrizni a H_1: $b > 0$ hipotézis ellenében, e célból a következő próbát alkalmazzuk: egy megfigyelést végzünk és ha ez a $(0,1; 0,85)$ intervallumba esik, elfogadjuk H_0-t, különben elvetjük.
a.) Írjuk fel a próba erőfüggvényét! Mekkora a próba terjedelme?
b.) Valaki egy másik próbát javasol: két független megfigyelést végezzünk, és akkor fogadjuk el H_0-t, ha ezek összege a $(0,1; 1,7)$ intervallumba esik. Számítsuk ki itt is a terjedelmet és az erőfüggvényt!

39.) Az X valószínűségi változó eloszlása kétféle lehet:

	k	0	1	2	3	4	5
H_0	$P(X = k)$	0.32	0.17	0.05	0.09	0.14	0.23
H_1	$P(X = k)$	0.05	0.12	0.21	0.39	0.19	0.04

Egyetlen megfigyelést végzünk (jelölje ezt X), ennek alapján kétféle próbát javaslunk:
1. próba: $2 \le X \le 4$, akkor elvetjük H_0-t, különben elfogadjuk.
2. próba: ha $X \le 2$, akkor elfogadjuk H_0-t, különben elvetjük.
a.) Melyiket tartjuk jobb próbának?
b.) Adjuk meg a legerősebb $0,1$ terjedelmű próbát!
c.) Határozzuk meg a legerősebb $0,05$ terjedelmű próbát 3 elemű minta esetén!

40.) Legyen két megfigyelésünk a $(3; p)$ paraméterű binomiális eloszlásból. Adjuk meg a legjobb olyan próbát az alábbi hipotézisekre, melynek elsőfajú hiba valószínűsége $0,04$:
$H_0: \; p \ge \frac{1}{2}$
$H_1: \; p < \frac{1}{2}$

41.) Keressünk n elemű $N(m, \sigma^2)$ független minta esetén egyenletesen legerősebb α terjedelmű próbát a

$H_0 : m = m_0$

$H_1 : m > m_0$

hipotézisvizsgálati feladatra, ha σ ismert!

42.) Legyen $X_1, ..., X_n$ minta az

$$f(x) = \begin{cases} \frac{2x}{a^2} & \text{ha } 0 \leq x \leq a \\ 0 & \text{különben} \end{cases}$$

sűrűségfüggvényű eloszlásból. Tekintsük a következő hipotéziseket:

H_0: $a = 1$

H_1: $0 < a < 1$

Adjunk meg α terjedelemhez egyenletesen legerősebb próbát!

43.) A $(0, 1)$ intervallumon egyenletes eloszlásból 10 megfigyelést végeztek és az egyiket megmondták nekünk, jelölje ezt X, állítólag a legnagyobbikat (H_0). El szeretnénk dönteni, hogy ez igaz-e. Adjuk meg az egyenletesen legerősebb $0,95$ szintű valószínűséghányados próbát a kérdés eldöntésére.

B11.) [IV.28.] Az (A) gép által termelt termékek egy bizonyos jellemzője $N(11, 1^2)$, míg a (B) gépen termelt termékeké $N(13, 4^2)$ eloszlású. Legyenek

H_0 : a mintánk az (A) gépen készült

H_1 : a mintánk a (B) gépen készült

Ha egy 16 elemű minta átlaga legfeljebb 12, akkor elfogadjuk H_0-t, különben elvetjük.

a.) Mekkora az elsőfajú és a másodfajú hiba valószínűsége?

b.) Milyen c értéket adjunk meg a 12 helyett ahhoz, hogy legfeljebb $0,05$ legyen a próba terjedelme? Ekkor mennyi a másodfajú hiba valószínűsége?

B12.) [IV.28.] Legyen öt megfigyelésünk a λ paraméterű Poisson-eloszlásból. Adjuk meg a legjobb olyan próbát az alábbi hipotézisekre, melynek elsőfajú hiba valószínűsége $0,02$:

$H_0 : \ \lambda \geq 1$

$H_1 : \ \lambda < 1$

SZ18.) A Politikatudományi Kar HÖK elnöke nagyon fontosnak tartja népszerűségét. Amennyiben a hallgatók legfeljebb 70%-a utálja, az számára elfogadható (H_0 hipotézis). Az ennél nagyobb arány esetén (H_1 hipotézis) lemond. Minden negyedév végén 10 hallgatót kérdez meg (közvéleménykutatást tart). Az elnök akkor mond le, ha a tízből legalább 8 diák utálja.

a.) Mekkora a próba terjedelme?

b.) Várhatóan hány negyedévet fog tevékenykedni az elnök, ha stabilan a diákok 65%-a utálja? (1+1 pont)

SZ19.) Legyen X_1 minta az $f(x)$ sűrűségfüggvényű eloszlásból. Tekintsük a következő hipotéziseket:

H_0: $f(x) = f_0(x) = 2(1 - x) \cdot I(0 < x < 1)$

H_1: $f(x) = f_1(x) = 2x \cdot I(0 < x < 1)$

Adjunk meg α terjedelemhez egyenletesen legerősebb próbát! (1 pont)

SZ20.) Legyen X_1 minta $\mathrm{Bin}(n, p)$ eloszlásból, ahol n ismert. Tekintsük a következő hipotéziseket:

H_0: $p \sim \text{Béta}(\beta, \beta)$

H_1: $p \sim E(0, 1)$

Adjunk meg α terjedelemhez egyenletesen legerősebb próbát! (3 pont)

44.) Az alábbi minta 4 év október 18-án Budapesten mért napi középhőmérséklet adatait tartalmazza.

Középhőm. (C fok) adatok:

14,8	12,2	16,8	11,1

a.) Ellenőrizzük a H_0: $m = 15$ hipotézist $\alpha = 0,05$ elsőfajú hibavalószínűség mellett *értelmes* alternatív hipotézissel szemben. A korábbi tapasztalatok alapján tekintsük az értékek szórását 2-nek.

b.) Adjuk meg a p-értéket is!

c.) Ne használjunk a szórásra vonatkozóan előzetes információt!

45.) Tegyük fel, hogy az emberi magasság normális eloszlású.

a.) Végezzünk statisztikai próbát arra vonatkozóan, hogy a gyakorlaton lévő lányok átlagmagassága 170 cm!

b.) Végezzünk statisztikai próbát arra vonatkozóan, hogy a gyakorlaton lévő fiúk átlagmagassága 180 cm!

46.) A Természettudományi Kar II. évfolyamán az egyik gyakorlati csoportban 10-en írtak statisztika zárthelyit. Két feladatsor volt, mindkettőben 30 pontot lehetett elérni. Tegyük fel, hogy az elért pontszámok normális eloszlásúak. A pontszámokat tartalmazza az alábbi táblázat:

1. feladatsor	12	11	8	14	10
2. feladatsor	15	14	9	16	11

a.) Vajon az első feladatsor nehezebb volt?

b.) Mennyiben változik a helyzet, ha nem 10 diákról, hanem csak 5-ről van szó, és a 2. feladatsor a pótZH eredménye?

47.) Egy matematikaversenyen 18 hallgató vett részt, akik különféle magyarországi egyetemekre járnak. A hallgatók pontszámai (50 a maximum):

Hol tanul	Pontszámok					
BCE	26					
BME	35	23	40	27		
DTE	40	18				
ELTE	34	44	34	35	48	43
NYME	23					
SZIE	35					
SZTE	42	21	32			

Állíthatjuk-e, hogy az ELTE hallgatói jobbak a BME-seknél?

48.) Az alábbi két minta 10 egyforma képességűnek feltételezett sportoló súlylökésben elért eredményeit tartalmazza. A sportolók két ötfős csoportban készültek az edzőtáborban. Edzéstervük ugyanaz volt, de az első csoportban készülők minden reggel fejenként 10 tojást és 25 túró rudit ettek meg. A második csoportban készülőknek reggel és este 1-1 kg szalonnát és 1-1 kg madártejet kellett megenni. 2 hét felkészülés után értékelték az eredményeket. Tételezzük fel, hogy normális eloszlásból származnak a minták és a terjedelem 5%.

1. csoport	15,8	15,2	16,3	17,1	16,1
2. csoport	19,0	12,1	17,2	14,7	21,0

a.) Melyik diéta volt jobb, ha a dobások szórását 2-nek tekintjük?

b.) Állíthatjuk-e, hogy a második csoportban nagyobb változékonyságot mutat a sportolók teljesítménye?

c.) Ha nem ismerjük a szórást, akkor tekinthetjük-e valamelyik diétát jobbnak?

$F_{4,4;0,95} = 4,4$ $F_{5,5;0,95} = 5,05$ $F_{4,4;0,975} = 9,6$ $F_{5,5;0,975} = 7,15$

B13.) [V.12.] Az alábbi két minta néhány azonos típusú autó fogyasztási adatait tartalmazza. A gyári adatok szerint az autók fogyasztása normális eloszlású. Az első sorban a szerviz előtti, a másodikban a szerviz utáni értékek találhatók. Mit gondolunk, vajon a szervizelés csökkentette a fogyasztást?

Szerviz előtt ($\frac{l}{100\,km}$)	$7,9$	$8,1$	$8,8$	$7,2$	$6,0$	$6,6$	$6,0$
Szerviz után ($\frac{l}{100\,km}$)	$7,5$	$7,5$	$8,1$	$7,2$	$5,7$	$6,0$	$5,6$

B14.) [V.12.] Két eltérő márkájú instant kávé oldódási idejét tesztelték, melyekből minden alkalommal azonos mennyiséget tettek 1 dl forrásban lévő vízbe. A kísérletek eredményeit a következő táblázat tartalmazza:

Kávémárka	Oldódási idő (mp)							
Mokka Makka	8	5	6	6	5	7	6	7
Coffee Pronto	5	4	3	3	6	4		

Tegyük fel, hogy az oldódási idők normális eloszlást követnek. Állíthatjuk-e, hogy a két kávéfajta ugyanannyi idő alatt oldódik fel?

SZ21.) A Hurka húsgyárban minden szállítás előtt megvizsgálják a kolbászok szulfáttartalmát. 2015. április 28-án a még megengedett szint %-ban a mérések a következők voltak: 98,5; 101,4; 99,5; 100.9 és 100,7. A korábbi tapasztalatok alapján az ellenőr az eredményekről feltételezi, hogy 1 szórásúak.

a.) Elfogadható-e a $H_0 : m = 100$ nullhipotézis α=0,05 elsőfajú hibavalószínűség mellett?

b.) Mennyi a próba erőfüggvényének az értéke az $m = 102$ pontban? Hány elemű mintára van szükség, ha azt szeretnénk, hogy ez az érték legalább $0,99$ legyen? (0,5+1=1,5 pont)

SZ22.) A 47.) feladatra vonatkozó kérdés:
Állíthatjuk-e, hogy az ELTE hallgatói jobbak voltak a versenyen az összes többi egyetem hallgatóinál? (1 pont)

SZ23.) A gyári adatok szerint a TT220V típusú izzók élettartama exponenciális eloszlást követ 1000 óra várható értékkel. Egy doboznyi ilyen fajtájú izzót vásároltunk tesztelési céllal, és arra a következő idők alatt égtek ki (óra): 1010, 1043, 930, 850, 1300, 930. Hajtsunk végre alkalmas próbát annak eldöntésére, hogy a tesztünk alapján lehet-e az izzók várható élettartama 1000 óra! (2 pont)

49.) Az Informatikai Kar III. évfolyamán 300-an tanulnak. Megszámolták, hogy a legutóbbi vizsgaidőszakban hányszor buktak az egyes hallgatók. Az eredményeket tartalmazza az alábbi táblázat:

Bukások száma	0	1	2	3	4
Hallgatók száma	80	113	77	27	3

a.) Elfogadhatjuk-e azt a hipotézist, hogy egy hallgató bukásszáma Bin(4; 0,25) eloszlású?

b.) és azt, hogy Bin(4;p) eloszlású?

50.) Az alábbi táblázatban adatok találhatók azon személyek számáról, akik lórúgás következtében haltak meg 10 porosz hadtestben 20 év alatt (1875–

1894) (összesen $10 \cdot 20 = 200$ adat):

halálesetek száma	0	1	2	3	4
gyakoriság	109	65	22	3	1

Ellenőrizzük azt a hipotézist, hogy a halálesetek száma egy hadtestben egy év alatt Poisson-eloszlású!

51.) Személysérüléses közúti közlekedési balesetek az előidéző okok szerint két kiemelt évben(forrás: nagyrészt `http://www.ksh.hu/docs/hun/xstadat/xstadat_eves/i_ods003.html`):

Baleset oka	2003-ban	2013-ban
A járművezetők hibája	17769	14356
A gyalogosok hibája	1885	981
Az utasok hibája	49	63
A járművek műszaki hibája	105	81
Pályahiba	162	203
Egyéb okok	6	7
Összesen	19976	15691

Vizsgáljuk meg, hogy a személysérüléses közúti közlekedési balesetek 2003-as megoszlása (eloszlása) megegyezik-e a 2013-as megoszlással (eloszlással)!

52.) Az alábbi kontingencia-táblázat mutatja, hogy 100 évben a csapadék mennyisége és az átlaghőmérséklet hogyan alakult:

Csapadék / Hőmérséklet	Kevés	Átlagos	Sok
Hűvös	15	10	5
Átlagos	10	10	20
Meleg	5	20	5

A cellákban az egyes esetek gyakoriságai találhatóak. Tekinthető-e a csapadékmennyiség és a hőmérséklet függetlennek?

53.) Rendelkezésünkre áll a következő minta: 0,55; 0,59; 0,34; 0,69; 0,95; 0,34; 0,53; 0,54; 0,03; 0,11.

a.) Elfogadhatjuk-e azt a hipotézist, hogy a minta $E(0,2)$ eloszlású?

b.) Elfogadhatjuk-e azt a hipotézist, hogy a minta $E(0,1)$ eloszlású?

c.) Elfogadhatjuk-e azt a hipotézist, hogy a minta $Exp(2)$ eloszlású?

B15.) [V.12.] Egy szabálytalannak gondolt érmét addig dobálunk fel, míg második alkalommal fejet nem kapunk. Ezt a kísérletet 100-szor megismételtük, a fejet másodszorra a következő dobásokra kaptuk meg: $2, 2, 3, 6, 3, 2, 3, 4, \ldots, 2$. A könnyebb áttekinthetőség kedvéért ezeket táblázatba foglalhatjuk:

Hányadik dobásra kaptuk a 2. fejet	2	3	4	5	6	7	8
Gyakoriságok	35	28	16	11	7	2	1

Mit gondolsz, az érme valóban szabálytalan? Vizsgáld meg, hogy a dobások száma a 2. fejig negatív binomiális eloszlást követ-e (a rend a feladat szövege alapján ismert, azaz: ...)!

B16.) [V.12.] Algebrából évfolyamZH-t írtak a hallgatók. Két csoport volt, a csoportok és a megszerzett pontok (100 volt a maximum) szerinti gyakoriságokat tartalmazza a következő táblázat:

Pontszám / Csoport	0-20	21-40	41-60	61-80	81-100
A	10	10	15	14	4
B	6	10	20	16	9

Mit gondolsz, a megszerzett pontok függetlenek attól, hogy ki melyik csoportot írta?

SZ24.) A honlapomon az alkmatos statisztika résznél találhatsz egy adatsort, ami a napfoltok számát tartalmazza 1700-tól 2013-ig. Elérhetősége: `www.cs.elte.hu/~vargal4/napfolt.txt`. Vajon a napfoltok száma milyen eloszlást követhet? Állításodat támaszd alá valamilyen tanult statisztikai módszerrel! (2 pont)

SZ25.) 100 napon keresztül feljegyezték egy város energiafogyasztását. A lenti táblázat azt tartalmazza, hogy az egyes intervallumokba hány megfigyelés esett, valamint azt is, hogy az adott intervallumba eső értékeknek mennyi az átlaga. Tekinthető-e az energiafogyasztás normális eloszlásúnak? (1 pont)

Intervallumok	< 5000	$5000 - 6000$	$6000 - 7000$	> 7000
Gyakoriságok	20	31	28	21
Átlagok	3875	5700	6500	7800