1. Legyen $\underline{X} = X_1, \ldots, X_{200}$ független minta b paraméterű Poisson-eloszlásból ($b > 0$). Legyen $T_1(\underline{X}) = \frac{X_1+X_2+\ldots+X_{100}}{100}$, $T_2(\underline{X}) = \frac{X_1+X_2+\ldots+X_{200}}{200}$.
   a) Milyen a számra igaz, hogy $T_2^2(\underline{X}) - aT_1(\underline{X})$ torzítatlan becslése b-nek? (10 pont)
   b) Igaz-e, hogy $\frac{1}{3}T_1(\underline{X}) + \frac{2}{3}T_2(\underline{X})$ torzítatlan becslése b-nek? (4 pont)
   c) $T_1(\underline{X})$ és $T_2(\underline{X})$ közül melyiknek kisebb a szórása? (4 pont)

2. Exponenciális eloszlásból származó minta esetén igaz-e, hogy a mintaátlag konzisztens becslése a paraméter reciprokának? (8 pont)

3. *a)* p paraméterű geometriai eloszlású mintából adjunk torzítatlan becslést a paraméterre! (8 pont)
   b) Az alábbi minta geometriai eloszlásból származik: 2, 1, 2, 7, 3, 3, 6, 2, 1, 1, 3, 1, 6.

   Mennyi az *a)* feladatban adott becslés értéke a mintán? (4 pont)

4. 50 gyerek magasságát mérjük meg centiméterben. Feltételezzük, hogy mindegyikük magassága egymástól független, $N(130, 25)$ eloszlású valószínűségi változó. Jelölje X a mért magasságok átlagát. a) Milyen eloszlású X? (8 pont) b) Mennyi $P(X < 125)$? (4 pont)

---

1. *a)* Tudjuk, hogy független, n elemű Z mintából

$$\mathbb{E}\left(\overline{Z}\right) = \mathbb{E}Z_1; \qquad D^2\left(\overline{Z}\right) = \frac{D^2(Z_1)}{n}.$$

Ezt $n = 100$-ra és $n = 200$-ra alkalmazva, továbbá felhasználva, hogy X_1 eloszlása b paraméterű Poisson-eloszlás, azt kapjuk, hogy

$$\mathbb{E}_b(T_1) = \mathbb{E}X_1 = b; \quad \mathbb{E}_b(T_2) = \mathbb{E}X_1 = b; \quad D_b^2(T_1) = \frac{\mathbb{E}X_1}{100} = \frac{b}{100}; \quad D_b^2(T_2) = \frac{\mathbb{E}X_1}{200} = \frac{b}{200}.$$

Az utóbbi két összefüggésből a szórásnégyzet definícióját felhasználva adódik, hogy

$$\mathbb{E}_b\left(T_2^2\right) = D^2(T_2) + (\mathbb{E}_b(T_2))^2 = \frac{b}{100} + b^2.$$

Tehát a megadott becslés várható értéke:

$$\mathbb{E}_b\left(T_2^2\right) - aT_1 = \frac{b}{100} + b^2 - ab = b\left(1 + \frac{b}{100} - a\right).$$

A becslés akkor torzítatlan becslése b-nek, ha várható értéke minden b-re b-vel egyezik meg. Látható, hogy ez egyetlen a-ra sem teljesül az összes b-vel egyszerre, tehát nincs a feladat feltételeinek megfelelő a szám.

b) Az *a)* feladat eredményeit és a várható érték tulajdonságait felhasználva:

$$\mathbb{E}_b\left(\frac{1}{3}T_1 + \frac{2}{3}T_2\right) = \frac{1}{3}\mathbb{E}_b(T_1) + \frac{2}{3}\mathbb{E}_b(T_2) = \frac{1}{3}b + \frac{2}{3}b = b$$

teljesül minden $b > 0$-ra. Tehát a megadott becslés torzítatlan becslése b-nek.

c) Az *a)* feladat szerint a $T_2(\underline{X})$ becslés szórása kisebb.

2. A nagy számok erős törvénye szerint ha $X_1, X_2, \ldots$ független valószínűségi változók, azonos eloszlásúak, és várható értékük létezik, akkor $(X_1 + \ldots + X_n)/n \to \mathbb{E}X_1$ 1 valószínűséggel $n \to \infty$ esetén. Ha most $X_1, X_2, \ldots$ független, $\exp(a)$ paraméterű exponenciális eloszlású valószínűségi változók, akkor egyrészt a tétel alkalmazható, másrészt $\mathbb{E}_a(X_1) = \frac{1}{a}$. Tehát

$$P_a\left(\frac{X_1 + \ldots + X_n}{n} \to \frac{1}{a}\ (n \to \infty)\right) = 1$$

minden $a > 0$-ra teljesül, és ez definíció szerint azt jelenti, hogy a mintaátlag konzisztens becslése a paraméter reciprokának.

3. a) Ha $X_1, \ldots, X_n$ p paraméterű geometriai eloszlású minta, akkor $P(X_i = 1) = p$ a geometriai eloszlás definíciója szerint. Legyen $I_i = 1$, ha $X_i = 1$, és 0 különben, $i = 1, \ldots, n$-re. Ez tehát annak indikátora, hogy az i. mintaelem 1, várható értéke az $X_i = 1$ esemény valószínűsége, azaz p. Legyen most $T(X_1, \ldots, X_n) = \frac{I_1 + \ldots + I_n}{n}$. Ekkor

$$E_p(T(X_1, \ldots, X_n)) = \mathbb{E}_p\left(\frac{I_1 + \ldots + I_n}{n}\right) = \frac{n \cdot p}{n} = p.$$

Ez teljesül minden $p \in (0,1)$-re, azaz a paraméter minden lehetséges értékére. Ez azt jelenti, hogy T, azaz az 1 mintaelemek relatív gyakorisága torzítatlan becslés b-re.

b) Most $n = 13$, és az *a)* rész jelöléseivel

$$T(X_1, \ldots, X_{13}) = \frac{4}{13},$$

hiszen a mintaelemek között 4 egyes van.

4. a) Tudjuk, hogy független, n elemű Z mintából

$$\mathbb{E}\left(\overline{Z}\right) = \mathbb{E}Z_1; \qquad D^2\left(\overline{Z}\right) = \frac{D^2(Z_1)}{n}.$$

Ugyanakkor független normális eloszlású valószínűségi változók összege is normális eloszlású, ezt rögzített számmal elosztva is normális eloszlású valószínűségi változót kapunk. Tehát ha $Z_1, \ldots, Z_{50}$ az egyes gyerekek testmagassága, akkor X is normális eloszlású valószínűségi változó. X várható értéke Z_1 várható értéke, azaz 130, szórásnégyzete pedig $\frac{25}{50} = \frac{1}{2}$. Tehát $X \sim N\left(130, \frac{1}{2}\right)$.

b) Mivel $X \sim N\left(130, \frac{1}{2}\right)$, ismert állítás szerint

$$P(X < 125) = \Phi\left(\frac{125 - 130}{\sqrt{1/2}}\right) \approx \Phi(-7.07) \approx 1 - \Phi(7.07) \approx 1.$$

1. Egy pénzérmével fejet és írást lehet dobni. A fej dobásának valószínűségét, p-t nem ismerjük ($p \in (0,1)$). Adjunk maximumlikelihood-becslést p-re, ha a minta a következő: 100 független dobásból 62 volt fej. (12 pont)

2. $X_1, \ldots, X_n$ független minta $N(m, 100)$ eloszlásból.
a) Mennyi a mintaközép várható értéke? (4 pont)
b) Adjunk torzítatlan becslést m-re! (8 pont)
c) Mennyi a megadott torzítatlan becslés értéke az alábbi mintán? (4 pont)

162,1	172,9	172,6	168,9	159,1	182,7

3. A $(0,a)$ intervallumon egyenletes eloszlásból származó független $X_1, X_2, \ldots$ minta esetén milyen c rögzített számra lesz
a) a tapasztalati szórásnégyzet c-szerese a^2-re torzítatlan becslés, ha $n = 1000$? (8 pont)
b) $\left(c \cdot \left(\frac{X_1+\ldots+X_n}{n}\right)^2\right)$ konzisztens becsléssorozat a^2-re? (8 pont)

4. Független, b paraméterű Poisson-eloszlású mintából adjunk momentummódszeres becslést a paraméterre! (6 pont)

A megoldásokat indokolni kell, a teljes pontszámhoz jó végeredmény és helyes indoklás szükséges. Összesen 50 pontot lehet elérni, a várható ponthatárok: 40, 61, 72, 83.

Az elégséges határa 20 pont.

Az eredmények az ETR infosheet rovatában lesznek elérhetők, a megoldások pedig itt:
http://www.cs.elte.hu/∼agnes/gyak

Statisztika 1. zárthelyi dolgozat 2011. március 24.

1. Egy pénzérmével fejet és írást lehet dobni. A fej dobásának valószínűségét, p-t nem ismerjük ($p \in (0,1)$). Adjunk maximumlikelihood-becslést p-re, ha a minta a következő: 100 független dobásból 62 volt fej. (12 pont)

2. $X_1, \ldots, X_n$ független minta $N(m, 100)$ eloszlásból.
a) Mennyi a mintaközép várható értéke? (4 pont)
b) Adjunk torzítatlan becslést $10m$-re! (8 pont)
c) Mennyi a megadott torzítatlan becslés értéke az alábbi mintán? (4 pont)

162,1	172,9	172,6	168,9	159,1	182,7

3. A $(0,a)$ intervallumon egyenletes eloszlásból származó független $X_1, X_2, \ldots$ minta esetén milyen c rögzített számra lesz
a) a tapasztalati szórásnégyzet c-szerese a^2-re torzítatlan becslés, ha $n = 1000$? (8 pont)
b) $\left(c \cdot \left(\frac{X_1+\ldots+X_n}{n}\right)^2\right)$ konzisztens becsléssorozat a^2-re? (8 pont)

4. Független, b paraméterű Poisson-eloszlású mintából adjunk momentum-módszeres becslést a paraméterre! (6 pont)

A megoldásokat indokolni kell, a teljes pontszámhoz jó végeredmény és helyes indoklás szükséges. Összesen 50 pontot lehet elérni, a várható ponthatárok: 40, 61, 72, 83.

Az elégséges határa 20 pont.

Az eredmények az ETR infosheet rovatában lesznek elérhetők, a megoldások pedig itt:
http://www.cs.elte.hu/∼agnes/gyak

1. Legyen $I_i = 1$, ha az i. dobás fej, 0 különben. $I_1, \ldots, I_{100}$ független minta a p paraméterű indikátoreloszlásból.

   Maximumlikelihood-módszerrel felírjuk a likelihood-függvényt:

$$f_p(I_1 = k_1, \ldots, I_{100} = k_{100}) = p^{\sum_{i=1}^{100} k_i} \cdot (1-p)^{100-\sum_{i=1}^{100} k_i};$$
$$f(p) = p^{\sum_{i=1}^{100} I_i} \cdot (1-p)^{100-\sum_{i=1}^{100} I_i}.$$

   Azt a p-t kell tehát választanunk, amelyre f maximális. (f-ben csak a mintaelemek összege szerepel, tehát a mintaelemek összege elégséges statisztika.) A logaritmusfüggvény szigorúan monoton növő, tehát ehelyett a loglikelihood-függvény maximumhelyét is elég megkeresni:

$$l(p) = \log f(p) = \sum_{i=1}^{100} I_i \cdot \log p + \left(100 - \sum_{i=1}^{100} I_i\right) \cdot \log(1-p).$$

   Ennek deriváltja:

$$l'(p) = \frac{\sum_{i=1}^{100} I_i}{p} - \frac{100 - \sum_{i=1}^{100} I_i}{1-p}.$$

   Ennek nullhelye:

$$l'(p) = 0 \Leftrightarrow \left(\sum_{i=1}^{100} I_i\right)(1-p) = \left(100 - \sum_{i=1}^{100} I_i\right) p \Leftrightarrow \sum_{i=1}^{100} I_i = 100p \Leftrightarrow p = \frac{\sum_{i=1}^{100} I_i}{100}.$$

   Ugyanakkor l folytonosan differenciálható, $\lim_{p\to 0} l(p) = \lim_{p\to 1} l(p) = -\infty$. Tehát g-nek globális maximuma van a deriváltjának nullhelyében: $\hat{p} = \frac{\sum_{i=1}^{100} I_i}{100} = \frac{62}{100} = 0,62$, beírva, hogy a 100 dobás közül 62 volt fej. Ez tehát a maximumlikelihood-becslés.

2. *a)* A mintaközép várható értéke független minta esetén a mintaelemek közös várható értéke, azaz az $N(m, 100)$ eloszlás várható értéke: m.

   b) Mivel minden m-re $\mathbb{E}\left(10 \cdot \overline{X}\right) = 10\mathbb{E}\left(\overline{X}\right) = 10m$ az *a)* feladat szerint, $10 \cdot \overline{X}$ torzítatlan becslés $10m$-re.

   c) A megadott mintán $10 \cdot \overline{X} \approx 1697,17$.

3. *a)* A kérdés, hogy milyen c számra lesz

$$\mathbb{E}_a\left(cs_n^2\right) = a^2 \quad \forall a > 0.$$

   Mivel a minta független, $U(0, a)$ eloszlású:

$$\mathbb{E}_a\left(cs_n^2\right) = c\mathbb{E}_a\left(s_n^2\right) = c \cdot \frac{n-1}{n} \cdot D^2(X_1) = c\frac{n-1}{n} \cdot \frac{a^2}{12}.$$

   Tehát $c = \frac{12n}{n-1} = \frac{12000}{999} \approx 12,01$ jó választás.

   b) A nagy számok erős törvénye szerint a mintaközepekből álló sorozat 1 valószínűséggel konvergál a várható értékhez (független, azonos eloszlású, véges várható értékű valószínűségi változókról van szó). Tehát a mintaközepekből álló sorozat minden a-ra 1 valószínűséggel konvergál $a/2$-höz. Ezért a mintaközepek négyzetéből álló sorozat minden a-ra 1 valószínűséggel konvergál $(a/2)^2$-hez. Ebből következik, hogy a mintaközepek négyzetének négyszereséből álló sorozat minden a-ra 1 valószínűséggel konvergál a^2-hez, azaz konzisztens becslést ad a^2-re. Tehát a 4 jó, és látható, hogy más megoldás nincs.

4. A b paraméterű Poisson-eloszlás várható értéke b. Ezért azt a b-t kell választanunk, amelyre a mintaközép (a tapasztalati eloszlás első momentuma) megegyezik b-vel: $\overline{X} = b$, b momentummódszeres becslése a mintaátlag.

Statisztika 1. zárthelyi dolgozat 2011. március 24.

1. Legyen $X_1, \ldots, X_n$ független, $\exp(a)$ eloszlásból származó minta. Számítsuk ki a mintaközép várható értékét és szórását! (8 pont)
2. $X_1, \ldots, X_n$ független minta $N(m, 100)$ eloszlásból.
   a) Mennyi a mintaközép várható értéke? (4 pont)
   b) Adjunk torzítatlan becslést m-re! (8 pont)
   c) Mennyi a megadott torzítatlan becslés értéke az alábbi mintán? (4 pont)

   162,1 172,9 172,6 168,9 159,1 182,7
3. A $(0, a)$ intervallumon egyenletes eloszlásból származó független $X_1, X_2, \ldots$ minta esetén milyen c rögzített számra lesz
   a) a tapasztalati szórásnégyzet c-szerese a^2-re torzítatlan becslés, ha $n = 1000$? (8 pont)
   b) $\left(c \cdot \left(\frac{X_1+\ldots+X_n}{n}\right)^2\right)$ konzisztens becsléssorozat a^2-re? (8 pont)
4. Független, b paraméterű Poisson-eloszlású mintából adjunk torzítatlan becslést be^{-b}-re! (10 pont)

A megoldásokat indokolni kell, a teljes pontszámhoz jó végeredmény és helyes indoklás szükséges. Összesen 50 pontot lehet elérni, a várható ponthatárok: 40, 61, 72, 83.

Az elégséges határa 20 pont.

Az eredmények az ETR infosheet rovatában lesznek elérhetők, a megoldások pedig itt: http://www.cs.elte.hu/∼agnes/gyak

Statisztika 1. zárthelyi dolgozat 2011. március 24.

1. Legyen $X_1, \ldots, X_n$ független, $\exp(a)$ eloszlásból származó minta. Számítsuk ki a mintaközép várható értékét és szórását! (8 pont)
2. $X_1, \ldots, X_n$ független minta $N(m, 100)$ eloszlásból.
   a) Mennyi a mintaközép várható értéke? (4 pont)
   b) Adjunk torzítatlan becslést m-re! (8 pont)
   c) Mennyi a megadott torzítatlan becslés értéke az alábbi mintán? (4 pont)

   162,1 172,9 172,6 168,9 159,1 182,7
3. A $(0, a)$ intervallumon egyenletes eloszlásból származó független $X_1, X_2, \ldots$ minta esetén milyen c rögzített számra lesz
   a) a tapasztalati szórásnégyzet c-szerese a^2-re torzítatlan becslés, ha $n = 1000$? (8 pont)
   b) $\left(c \cdot \left(\frac{X_1+\ldots+X_n}{n}\right)^2\right)$ konzisztens becsléssorozat a^2-re? (8 pont)
4. Független, b paraméterű Poisson-eloszlású mintából adjunk torzítatlan becslést be^{-b}-re! (10 pont)

A megoldásokat indokolni kell, a teljes pontszámhoz jó végeredmény és helyes indoklás szükséges. Összesen 50 pontot lehet elérni, a várható ponthatárok: 40, 61, 72, 83.

Az elégséges határa 20 pont.

Az eredmények az ETR infosheet rovatában lesznek elérhetők, a megoldások pedig itt: http://www.cs.elte.hu/∼agnes/gyak