Vizsgadolgozat

Megoldás

Tanszéki általános alapelvek

A pontozási útmutató célja, hogy a javítók a dolgozatokat egységesen értékeljék. Ezért az útmutató minden feladat (legalább egy lehetséges) megoldásának főbb gondolatait, és az ezekhez rendelt részpontszámokat közli. Az útmutatónak nem célja a feladatok teljes értékű megoldásának részletes leírása; a leírt lépések egy maximális pontszámot érő megoldás vázlatának tekinthetők.

Az útmutatóban feltüntetett részpontszámok csak akkor járnak a megoldónak, ha a kapcsolódó gondolat egy áttekinthető, világosan leírt és megindokolt megoldás egy lépéseként szerepel a dolgozatban. Így például az anyagban szereplő ismeretek, definíciók, tételek puszta leírása azok alkalmazása nélkül nem ér pontot (még akkor sem, ha egyébként valamelyik leírt tény a megoldásban valóban szerephez jut). Annak mérlegelése, hogy az útmutatóban feltüntetett pontszám a fentiek figyelembevételével a megoldónak (részben vagy egészében) jár-e, teljes mértékben a javító hatásköre.

Részpontszám jár minden olyan ötletért, részmegoldásért, amelyből a dolgozatban leírt gondolatmenet alkalmas kiegészítésével a feladat hibátlan megoldása volna kapható. Ha egy megoldó egy feladatra több, egymástól lényegesen különböző megoldást is elkezd, akkor legfeljebb az egyikre adható pontszám. Ha mindegyik leírt megoldás vagy megoldásrészlet helyes vagy helyessé kiegészíthető, akkor a legtöbb részpontot érő megoldáskezdeményt értékeljük. Ha azonban több megoldási kísérlet között van helyes és (lényeges) hibát tartalmazó is, továbbá a dolgozatból nem derül ki, hogy a megoldó melyiket tartotta helyesnek, akkor a kevesebb pontot érő megoldáskezdeményt értékeljük (akkor is, ha ez a pontszám 0). Az útmutatóban szereplő részpontszámok szükség esetén tovább is oszthatók. Az útmutatóban leírttól eltérő jó megoldás természetesen maximális pontot ér.

Aritmetikai hiba esetén elszámolásonként 1-1 pont vonandó le a feladatokból. Ez alól kivétel, ha az elszámolás lényegesen egyszerűsíti vagy módosítja a feladat felépítését. Ilyen esetekben azon feladatrészekért, amik az elszámolás okán fel sem merültek, nem jár pont.

1. Írjuk fel az alábbi definíciót, illetve állítást:

   (a) Hogyan definiáljuk egy valószínűségi változó szórásnégyzetét?

   (b) Milyen feltétel esetén, és hogyan fejezhető ki az X és Y valószínűségi változók szorzatának várható értéke $\mathbb{E}(X)$ és $\mathbb{E}(Y)$ segítségével, az előadáson elhangzott állítás szerint?

   Megoldás:
   (10 pont) Egy valószínűségi változó szórásnégyzete:

   $$\mathbb{D}^2(X) = \mathbb{E}(X - \mathbb{E}(X))^2$$

   (Ha csak az szerepel, hogy saját magával vett kovarianciája, akkor 3 pont.)
   (Ha a $\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ formulát adja meg (hibátlanul), akkor 3 pont.)
   (jegyzet: 6.4.1)
   (4 pont) Ha X és Y független valószínűségi változók, és
   (2 pont) $\mathbb{E}(XY)$, $\mathbb{E}(X)$ és $\mathbb{E}(Y)$ létezik,
   (4 pont) akkor

   $$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

   (jegyzet: 6.1.4) (A feltételek felírásáért akkor jár a pont, ha az állítás is szerepel.)

2. Összekeverünk 4 kártyalapot, melyek közül pontosan az egyik piros, majd (lefordíva és) egymásra téve őket magunk elé helyezzük a lapokat. Ezek után elkezdünk dobálni egy szabályos érmét. Legfeljebb 4-szer dobunk, de amennyiben egy dobásnál írás az eredmény, megállunk.

Végül annyi kártyát húzunk a megkevert lapokból (mindig a felsőt kihúzva), ahány fejet dobtunk.
Mennyi a valószínűsége annak, hogy 4 darab fejet dobtunk, ha tudjuk, hogy a húzott lapok között van piros?

Megoldás:
Tekintsük a következő eseményeket:
(0 pont) Pi=a húzott lapok között szerepel a piros, F_i=i darab fejet dobunk, pontosabban az első i dobás fej, és $i < 4$ esetén az $(i+1)$-edik írás $(i = 0, 1, 2, 3, 4)$.
(1 pont) A kérdezett valószínűség $\mathbb{P}(F_4|Pi)$
(1 pont) az egyszerű Bayes-tételből
(2 pont) $\mathbb{P}(F_4|Pi) = \dfrac{\mathbb{P}(Pi|F_4)\,\mathbb{P}(F_4)}{\mathbb{P}(Pi)}$
(1 pont) A teljes valószínűség tételéből
(2 pont)

$$\mathbb{P}(Pi) = \sum_{i=0}^{4} \mathbb{P}(Pi \mid F_i)\,\mathbb{P}(F_i)$$

(Ha nem két lépésben írja le, hanem azonnal a Bayes-tételt helyesen és arra hivatkozik, akkor is jár a pont.)
(1 pont) Ha $i < 4$, akkor $\mathbb{P}(F_i) = \dfrac{1}{2^{i+1}}$
(1 pont) $\mathbb{P}(F_4) = \dfrac{1}{2^4}$
(1 pont) Ha nem dobunk fejet, akkor persze nem húzhatunk pirosat, így $\mathbb{P}(Pi \mid F_0) = 0$
(1 pont) Ha 4 fejet dobunk, akkor biztosan kihúzzuk a piros lapot, azaz $\mathbb{P}(Pi \mid F_4) = 1$
(1 pont) $\mathbb{P}(Pi \mid F_1) = \dfrac{1}{4}$
(2-2 pont) valamint (számításba véve, hogy hanyadikra húznuk pirosat)

$$\mathbb{P}(Pi \mid F_2) = \frac{1}{4} + \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{2} \qquad \mathbb{P}(Pi \mid F_3) = \frac{1}{4} + \frac{3}{4} \cdot \frac{1}{3} + \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{3}{4}$$

(vagy ha a húzás sorrendjét nem vesszük figyelembe:

$$\mathbb{P}(Pi \mid F_2) = \frac{\binom{3}{1}}{\binom{4}{2}} = \frac{3}{6} = \frac{1}{2} \qquad \mathbb{P}(Pi \mid F_3) = \frac{\binom{3}{2}}{\binom{4}{3}} = \frac{3}{4}$$

így kiszámolva is nyilván ugyanannyi pontot ér)
(2 pont) Behelyettesítve

$$\mathbb{P}(Pi) = \frac{1}{4} \cdot \frac{1}{2^2} + \frac{1}{2} \cdot \frac{1}{2^3} + \frac{3}{4} \cdot \frac{1}{2^4} + 1 \cdot \frac{1}{2^4} = \frac{4+4+3+4}{64} = \frac{15}{64}.$$

(2 pont) Tehát $\mathbb{P}(F_4|Pi) = \dfrac{1 \cdot \frac{1}{2^4}}{\frac{15}{64}} = \dfrac{4}{15}$

3. Egy program egyenletesen véletlenszerűen generál egy egész számot az $[1; 5]$ zárt intervallumban (tehát az intervallumban minden egész szám $1/5$ valószínűséggel adódik kimenetként). Tegyük fel, hogy ezzel a programmal egymástól függetlenül 1000 értéket generálunk, végül pedig összeadjuk a kapott eredményeket.
Közelítőleg mennyi a valószínűsége annak, hogy az összeg 2900 és 3100 közé esik?

Megoldás:
(0 pont) Jelölje X_i az i-edik generált számot

(1 pont) $\mathbb{P}(2900 < \sum_{i=1}^{1000} X_i < 3100) =?$
(2 pont) Sztenderdizálunk:

$$\mathbb{P}\left(\frac{2900 - \mathbb{E}(\sum_{i=1}^{1000} X_i)}{\mathbb{D}(\sum_{i=1}^{1000} X_i)} < \frac{\sum_{i=1}^{1000} X_i - \mathbb{E}(\sum_{i=1}^{1000} X_i)}{\mathbb{D}(\sum_{i=1}^{1000} X_i)} < \frac{3100 - \mathbb{E}(\sum_{i=1}^{1000} X_i)}{\mathbb{D}(\sum_{i=1}^{1000} X_i)}\right) =?$$

(1 pont) $X_1, \ldots, X_{1000}$ azonos eloszlású, együttesen független valószínűségi változók
(1 pont) ezért $\mathbb{E}(\sum_{i=1}^{1000} X_i) = 1000 \cdot \mathbb{E}(X_1)$ és $\mathbb{D}(\sum_{i=1}^{1000} X_i) = \sqrt{1000} \cdot \mathbb{D}(X_1)$
(1 pont) X_1 értékkészlete $\{1, 2, 3, 4, 5\}$, és mindegyik értéket $1/5$ valószínűséggel veszi fel
(1 pont) $\mathbb{E}(X_1) = \frac{1+2+3+4+5}{5} = 3$
(2 pont) $\mathbb{D}^2(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2$
(1 pont) $\mathbb{E}(X_1^2) = \frac{1^2+2^2+3^2+4^2+5^2}{5} = \frac{1+4+9+16+25}{5} = \dfrac{55}{5} = 11$
(1 pont) $\mathbb{D}^2(X_1) = 11 - 9 = 2$
(1 pont) tehát a kérdés

$$\mathbb{P}\left(\frac{-100}{\sqrt{2000}} < \frac{\sum_{i=1}^{1000} X_i - 1000 \cdot 3}{\sqrt{1000} \cdot \sqrt{2}} < \frac{100}{\sqrt{2000}}\right) \approx \mathbb{P}\left(-2{,}24 < \frac{\sum_{i=1}^{1000} X_i - 1000 \cdot 3}{\sqrt{1000} \cdot \sqrt{2}} < 2{,}24\right)$$

(1 pont) a centrális határeloszlás-tétel miatt
(2 pont) $\frac{\sum_{i=1}^{1000} X_i - 1000 \cdot 3}{\sqrt{1000} \cdot \sqrt{2}}$ közelítőleg standard normális eloszlású.
(2 pont) Tehát a keresett mennyiség: $\Phi(2{,}24) - \Phi(-2{,}24)$
(2 pont) Mivel $\Phi(-2{,}24) = 1 - \Phi(2{,}24)$
(1 pont) ezért a keresett mennyiség $2 \cdot \Phi(2{,}24) - 1 \approx 2 \cdot 0{,}9875 - 1 = 0{,}975$.

4. Legyen $X \sim B(10; 0{,}2)$ binomiális eloszlású valószínűségi változó, továbbá legyen $Y = 3X + 2$ és $Z = 5 - 2X$.

   (a) Számoljuk ki Y és Z korrelációs együtthatóját.

   (b) Adjuk meg $Y + Z$ és $Y - Z$ kovarianciamátrixát.

Megoldás:
(a)
(4 pont) Y a Z-nek is lineáris transzformáltja: $Y = -\frac{3}{2}Z + \frac{19}{2}$
(3 pont) (mindkét változó szórása pozitív és véges), így a tanult tétel alapján $\mathrm{corr}(Y, Z) = \pm 1$
(3 pont) az előjelet a Z együtthatójának előjele határozza meg
(1 pont) tehát $\mathrm{corr}(Y, Z) = -1$.

(*második megoldás*: A korreláció a definíció alapján is kiszámítható
(2 pont) $\mathrm{corr}(Y, Z) = \dfrac{\mathrm{cov}(Y, Z)}{\mathbb{D}(Y)\,\mathbb{D}(Z)}$
(1 pont) $\mathbb{D}(X) = \sqrt{10 \cdot 0{,}2 \cdot 0{,}8} = \sqrt{1{,}6}$
(1 pont) $\mathbb{D}(Y) = \mathbb{D}(3X + 2) = \mathbb{D}(3X) = 3\,\mathbb{D}(X) = 3\sqrt{1{,}6}$
(1 pont) $\mathbb{D}(Z) = \mathbb{D}(5 - 2X) = \mathbb{D}(-2X) = 2\,\mathbb{D}(X) = 2\sqrt{1{,}6}$
(2 pont) $\mathrm{cov}(Y, Z) = \mathrm{cov}(3X + 2, 5 - 2X) = -6 \cdot \mathrm{cov}(X, X)$
(2 pont) Továbbá $\mathrm{cov}(X, X) = \mathbb{D}^2(X) = 1{,}6$
(1 pont) tehát $\mathrm{corr}(Y, Z) = \dfrac{-6 \cdot 1{,}6}{3 \cdot \sqrt{1{,}6} \cdot 2 \cdot \sqrt{1{,}6}} = -1$)

(b)
(2 pont) tudja, hogy mik a kovariancia mátrix elemei
(2 pont) $Y + Z = X + 7$, $Y - Z = 5X - 3$

(2 pont) $\mathbb{D}^2(Y+Z) = \mathbb{D}^2(X) = 1{,}6$ $\qquad \mathbb{D}^2(Y-Z) = 25\,\mathbb{D}^2(X) = 25\cdot 1{,}6 = 40$
(3 pont) $\operatorname{cov}(Y+Z, Y-Z) = \operatorname{cov}(X+7, 5X-3) = 5\cdot\operatorname{cov}(X,X) = 5\cdot\mathbb{D}^2(X) = 8$
(1 pont) a kovariancia mátrix tehát

$$\begin{pmatrix} 1,6 & 8 \\ 8 & 40 \end{pmatrix}.$$

5. Legyen (X,Y) folytonos valószínűségi vektorváltozó, aminek együttes sűrűségfüggvénye:

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

Határozzuk meg X sűrűségfüggvényét, az $\mathbb{E}(Y\mid X)$ regressziót, illetve az $\mathbb{E}(Y)$ várható értéket.

Megoldás:
(2 pont) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y$
(Ha a fenti egyenlet implicit jelenik meg az f_X kiszámolásában, akkor is jár a pont.)
(1+1 pont) $= \int_0^x (10x^2y)\,\mathrm{d}y$ (az integrálási határért jár az egyik pont)
(1 pont) $= 5x^4$
(1 pont) ha $0<x<1$, különben 0.
(3 pont) $\mathbb{E}(Y\mid X=x) = \int_{-\infty}^{\infty} y\cdot f_{Y|X}(y\mid x)\,\mathrm{d}y$
(2 pont) $f_{Y|X}(y\mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ (ahol $f_X(x)\neq 0$)
(1 pont) $= \frac{10x^2y}{5x^4} = \frac{2y}{x^2}$
(1 pont) Tehát $\mathbb{E}(Y\mid X=x) = \int_0^x y\cdot\frac{2y}{x^2}\,\mathrm{d}y$
(1 pont) $= \frac{2}{3}x$, (ha $0<x<1$ és 0 egyébként)
(1 pont) A val. változó X-et visszahelyettesítve: $\mathbb{E}(Y\mid X) = \frac{2}{3}X$
(2 pont) $\mathbb{E}(Y) = \int_{-\infty}^{\infty}\mathbb{E}(Y\mid X=x)f_X(x)\,\mathrm{d}x$
(vagy $\mathbb{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\cdot f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$)
(Ha a fenti egyenlet implicit jelenik meg az f_X kiszámolásában, akkor is jár a pont.)
(további 3 pont arányosan)
$\mathbb{E}(Y) = \frac{5}{9} = 0.5555$
(Ha csak hányados, vagy csak tizedestört alakban van megadva a megoldás, akkor is jár a pont.)

6.* Egy fontos csomagot várunk, melyet a LINK csomagküldő szolgálat kézbesít. A cég késve kézbesíti a rábízott csomagokat, a késésük órában mérve egyenletes eloszlású a $(0;1)$ intervallumon. Ha nem érkezik meg időben a csomag, kártérítést kapunk, melynek összege x késés esetén tízezer forintban számolva egyenletes eloszlású az $(x-x^2;x)$ intervallumon.
Mennyi a valószínűsége annak, hogy 2500 Ft-nál kevesebb kártérítést kapunk?

Megoldás:
(0 pont) Jelölje X az órában mért késést, Y a kártérítés összegét tízezer forintban mérve
(1 pont) $\mathbb{P}(Y<0.25) = ?$
(5 pont) a teljes valószínűség tétele alapján

$$\mathbb{P}(Y<0.25) = \int \mathbb{P}(Y<0.25|X=x)\cdot f_X(x)\,\mathrm{d}x$$

(2 pont) $\mathbb{P}(Y<0.25|X=x) = 1$, ha $x<0.25$
(2 pont) $\mathbb{P}(Y<0.25|X=x) = \int_{x-x^2}^{0.25} f_{Y|X}(y|x)\,\mathrm{d}y$, ha $x\geq 0.25$
(2 pont)

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{x^2} & \text{ha } x-x^2<y<x \\ 0 & \text{egyébként.} \end{cases}$$

(3 pont) $\mathbb{P}(Y<0.25|X=x) = \int_{x-x^2}^{0.25}\frac{1}{x^2}\,\mathrm{d}y = \frac{x^2-x+0.25}{x^2}$, ha $x\geq 0.25$
(3 pont) $\mathbb{P}(Y<0.25) = \int_0^{0.25} 1\cdot 1\,\mathrm{d}x + \int_{0.25}^{1}\frac{x^2-x+0.25}{x^2}\cdot 1\,\mathrm{d}x = 1.75 + \ln 0.25 =$
(2 pont) $= 0,3637$