Valószínűségszámítás és statisztika
1.b. zárthelyi megoldása

1) A szimmetrikus differencia definíció szering a következő: $A \circ B = \bar{A}B + A\bar{B}$. Mutassuk meg, hogy

$$(A + \bar{B}) \circ (\bar{A} + B) = A \circ B$$

(5 pont)

Megoldás: A következő azonos átalakításokkal kapjuk, hogy

$(A + \bar{B}) \circ (\bar{A} + B) =$	
$=\overline{A + \bar{B}}(\bar{A} + B) + (A + \bar{B})\overline{\bar{A} + B} =$	A $\circ$ definíciója alapján
$=(\bar{A}B)(\bar{A} + B) + (A + \bar{B})(A\bar{B}) =$	De-Morgan azonosság
$=(\bar{A}B)\bar{A} + (\bar{A}B)B + A(A\bar{B}) + \bar{B}(A\bar{B}) =$	Disztributivitás
$=\bar{A}\bar{A}B + \bar{A}BB + AA\bar{B} + A\bar{B}\bar{B} =$	Szorzat asszociativitása és kommutativitása miatt
$=\bar{A}B + \bar{A}B + A\bar{B} + A\bar{B} =$	$XX = X$ miatt
$=\bar{A}B + A\bar{B} =$	$X + X = X$ miatt
$=A \circ B$	A $\circ$ definíciója alapján

2) A $(0, 1)$ intervallumot 2 véletlenszerűen rádobott ponttal felosztjuk 3 szakaszra. Mennyi annak a valószínűsége, hogy mindegyik szakasz hossza kisebb, mint 1/2?

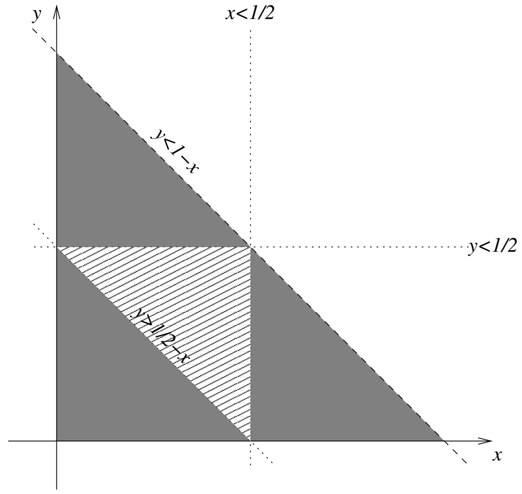
(10 pont)

Megoldás: Jelölje A azt az eseményt, hogy mindhárom három szakasz hossza kisebb 1/2-nél. A $P(A)$ valószínűséget geometriai módszerrel határozhatjuk meg. Legyen a $(0, 1)$ intervallumra dobott két pont koordinátája $x \in (0, 1)$ és $y \in (0, 1)$, és tegyük fel, hogy $x \leq y$. Ekkor a lehetséges elemi események Ω halmaza az x–y síkon egy háromszögnek felel meg, melynek területe $\mu(\Omega) = 1/2$.

Az A esemény bekövetkezéséhez a következő feltételeknek kell egyszerre teljesülni:

$$\begin{aligned} x &< 1/2 \\ y - x &< 1/2 \iff y < 1/2 + x \\ 1 - y &< 1/2 \iff y < 1/2 \end{aligned}$$

Ezek a feltételek szintén egy háromszöget határoznak meg, melynek területe $\mu(A) = 1/8$ (lásd a 1. ábrát). Ebből következik, hogy az A esemény bekövetkezésének valószínűsége $P(A) = \frac{\mu(A)}{\mu(\Omega)} = 1/4$.



1. ábra. Az Ω elemi események halmazát a szürke háromszög, a kedvező esetek halmazát pedig a satírozott háromszög jelöli.

3) Egy városban ugyanannyi férfi van, mint nő. Minden 100 férfi közül 5, és minden 10000 nő közül 25 színvak. Mennyi a valószínűsége, hogy a színvakokról vezetett nyilvántartásból egy találomra kiválasztott karton egy férfi adatait tartalmazza?

(15 pont)

Megoldás: Jelölje A azt az eseményt, hogy egy véletlenszerűen kiválasztott városlakó férfi, B pedig azt, hogy színvak. Ekkor a feladat szerint

$$P(A) = 1/2, \qquad P(B \mid A) = \frac{5}{100} = \frac{1}{20}, \tag{1}$$

$$P(\bar{A}) = 1/2 \qquad P(B \mid \bar{A}) = \frac{25}{10000} = \frac{1}{400}. \tag{2}$$

A kérdés az, hogy mennyi a $P(A \mid B)$ valószíneség. Mivel A és $\bar{A}$ teljes eseményrendszert alkot, ezért a Bayes-tétel alapján azt kaphatjuk, hogy

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \bar{A})P(\bar{A})} \tag{3}$$

$$= \frac{\frac{1}{20}\frac{1}{2}}{\frac{1}{20}\frac{1}{2} + \frac{1}{400}\frac{1}{2}} = \frac{20}{21} \approx 95.2\% \tag{4}$$

4) Egy gyümölcsösben az egyes gyümölcsöket megtámadó férgek száma egymástól független, Poisson-eloszlású valószínűségi változó, gyümölcsönként λ várható értékkel. A gyümölcsöst rovarirtóval kezelik, és ez p valószínűséggel öli meg a férgeket. Ebből a gyümölcsösből n gyümölcsöt vásárolunk.

(a) Mennyi a vásárolt gyümölcsökben lévő férgek számának eloszlása, várható értéke és szórása?

(b) Mennyi a férges gyümölcsök számának eloszlása, várható értéke és szórása?

(20 pont)

Megoldás: A permetezés előtt az i-edik gyümölcsben lévő férgek számát jelöljük ξ_i-vel, a permetezés után pedig η_i-vel. A permetezés előtt a férgek száma λ paraméterű Poisson-eloszlást követ, tehát

$$P(\xi_i = k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

Ha az i-edik gyümölcsben a permetezés előtt pontosan k féreg volt, akkor a permetezés után a férgek száma binomiális eloszlást követ $k(1-p)$ várható értékkel és $\sqrt{kp(1-p)}$ szórással:

$$P(\eta_i = l \mid \xi_i = k) = \binom{k}{l}(1-p)^l p^{k-l}, \quad 0 \le l \le k$$

Mivel $\xi_i = k$ teljes eseményrendszert alkot, ezért a teljes valószínűség tétele alapján a permetezés után az i-edik gyümölcsben lévő férgek számának eloszlása

$$P(\eta_i = l) = \sum_{k=l}^{\infty} P(\eta_i = l \mid \xi_i = k)P(\xi_i = k)$$

Határozzuk meg ennek a generátor-függvényét!

$$\begin{aligned}
G_{\eta_i}(z) &= \sum_{l=0}^{\infty} P(\eta_i = l)z^l = \sum_{l=0}^{\infty}\sum_{k=l}^{\infty} P(\eta_i = l \mid \xi_i = k)P(\xi_i = k)z^l \\
&= \sum_{k=0}^{\infty} P(\xi_i = k)\sum_{l=0}^{k} P(\eta_i = l \mid \xi_i = k)z^l \\
&= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda}\sum_{l=0}^{k}\binom{k}{l}(1-p)^l p^{k-l} z^l = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}((1-p)z+p)^k \\
&= e^{\lambda((1-p)z+p-1)} = e^{\lambda(1-p)(z-1)}
\end{aligned}$$

Ez éppen egy $\lambda(1-p)$ várható értékű Poisson-eloszlású valószínűségi változó generátorfüggvénye. n gyümölcs vásárlása esetén a vásárolt gyümölcsökben lévő férgek számát jelölje $\zeta_n = \xi_1 + \xi_2 + \cdots + \xi_n$. Mivel az egyes gyümölcsökben lévő férgek száma független, azért ζ_n generátorfüggvénye

$$G_{\zeta_n}(z) = [G_{\xi_i}(z)]^n = e^{n\lambda(1-p)(z-1)}$$

Ez pedig nyilvánvalóan egy $n\lambda(1-p)$ paraméterű Poisson-eloszlású változó generátor-fügyvénye, ezért

$$\begin{aligned}\mathbf{M}(\zeta_n) &= n\lambda(1-p),\\ \mathbf{D}(\zeta_n) &= \sqrt{n\lambda(1-p)}.\end{aligned}$$

Jelölje az n vásárolt gyümölcs közötti férges gyümölcsök számát ν_n. Annak a valószínűsége, hogy az i-edik gyümölcsben a permetezés után nincsen féreg: $P(\eta_i = 0) = G_{\eta_i}(0) = e^{-\lambda(1-p)}$. Mivel minden gyümölcs egymástól függetlenül férges vagy nem férges, ezért ν_n eloszlása binomiális:

$$P(\nu_n = k) = \binom{n}{k} e^{-\lambda(1-p)k}(1-e^{-\lambda(1-p)})^{n-k},$$

melynek várható értéke és szórása:

$$\begin{aligned}\mathbf{M}(\nu_n) &= ne^{-\lambda(1-p)},\\ \mathbf{D}(\nu_n) &= \sqrt{ne^{-\lambda(1-p)}(1-e^{-\lambda(1-p)})}.\end{aligned}$$

5) Az X_1 és X_2 valószínűségi változók eloszlásának együttes sűrűségfüggvénye

$$h(x,y) = \frac{1}{2\pi}\left[\left(\sqrt{2}e^{-\frac{x^2}{2}} - e^{-x^2}\right)e^{-y^2} + \left(\sqrt{2}e^{-\frac{y^2}{2}} - e^{-y^2}\right)e^{-x^2}\right]$$

Határozzuk meg a perem-eloszlásokat, valamint a korrelációs együtthatót!

(25 pont)

Megoldás: Az X_1 peremeloszlása (perem-sűrűségfüggvény) $f(x) = \int_{-\infty}^{\infty} h(x,y)dy$. Ehhez $\int_{-\infty}^{\infty} e^{-\alpha u^2}du$ alakú integrálokat kell kiszámítani. Mivel a normális eloszlás normált, ezért tudjuk, hogy

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}}du = 1.$$

Ebből $\alpha = \frac{1}{2\sigma^2}$ helyettesítéssel és átrendezéssel kapjuk, hogy

$$\int_{-\infty}^{\infty} e^{-\alpha u^2}du = \sqrt{\frac{\pi}{\alpha}}$$

Az X_1 peremeloszlása tehát

$$f(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi} \left[\left(\sqrt{2} e^{-\frac{x^2}{2}} - e^{-x^2} \right) e^{-y^2} + \left(\sqrt{2} e^{-\frac{y^2}{2}} - e^{-y^2} \right) e^{-x^2} \right] dy$$
$$= \frac{1}{2\pi} \left[\left(\sqrt{2} e^{-\frac{x^2}{2}} - e^{-x^2} \right) \int_{-\infty}^{\infty} e^{-y^2} dy \right.$$
$$\left. + e^{-x^2} \left(\sqrt{2} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy - \int_{-\infty}^{\infty} e^{-y^2} dy \right) \right]$$
$$= \frac{1}{2\pi} \left[\left(\sqrt{2} e^{-\frac{x^2}{2}} - e^{-x^2} \right) \sqrt{\pi} + e^{-x^2} \left(2\sqrt{\pi} - \sqrt{\pi} \right) \right] = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Mivel $h(x, y)$ szimmetrikus x-ben és y-ban, ezért X_2 peremeloszlása nyilvánvalóan $g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$.

A peremerloszlások tehát standard normális eloszlásúak, tehát $\mathbf{M}(X_1) = \mathbf{M}(X_2) = 0$, és $\mathbf{D}(X_1) = \mathbf{D}(X_2) = 1$. A korrelációs együttható ezért

$$R(X_1, X_2) = \frac{\mathbf{M}(X_1 X_2) - \mathbf{M}(X_1)\mathbf{M}(X_2)}{\mathbf{D}(X_1)\mathbf{D}(X_2)} = \mathbf{M}(X_1 X_2)$$

kifejezésre egyszerűsüdik. Az $\mathbf{M}(X_1 X_2)$ várható értéket az együttes eloszlás segítségével a következőképpen számíthatjuk ki: $\mathbf{M}(X_1 X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, h(x, y) dx\, dy$. Ez azonban szétesik

$$\int_{-\infty}^{\infty} u e^{-\alpha u^2} du$$

alakú integrálokra, amelyek egy páros és egy páratlan tényező szorzata. Ezeknek az integrálokan az értéke ezért 0. Ebből következik, hogy $\mathbf{M}(X_1 X_2) = 0$, azaz $R(X_1, X_2) = 0$.

Vegyük észre, hogy habár X_1 és X_2 peremeloszlásai standard normális eloszlásúak, az együttes eloszlásuk azonban *nem* normális eloszlású. Ha az együttes eloszlás normális eloszlás lenne, akkor a korrelálatlanságból következne a függetlenség is. Mint láttuk, X_1 és X_2 korrelálatlanok, de mivel $h(x, y) \neq f(x)g(y)$, ezért X_1 és X_2 *nem* függetlenek.