Villamosmérnök A4 — 11. hét
Kétdimenziós normális eloszlás, cht - Megoldások

Kétdimenziós normális összefoglalás

Egy kétdimenziós (X,Y) valószínűségi változó **kovariancia mátrixa**:

$$\Sigma = \begin{pmatrix} \mathbb{V}\mathrm{ar}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathbb{V}\mathrm{ar}(Y) \end{pmatrix}$$

Korrelációs együttható: $\mathrm{r}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathbb{SD}(X)\mathbb{SD}(Y)}$.

Kétdimenziós normális eloszlás: Standard kétdimenziós normális eloszlású egy (U,V) pár, ha sűrűségfüggvénye

$$\varphi(u,v) := \frac{1}{2\pi}\exp\left(-\frac{u^2+v^2}{2}\right) \qquad (u,v\in\mathbb{R}).$$

(U,V) zérus várható érték vektorú, egységmátrix kovarianciájú pár. Azt mondjuk, hogy az $(X,Y) = (U,V)\mathbf{A}+\mu$ pár kétdimenziós normális eloszlású $\mu\in\mathbb{R}^2$ várható érték vektorral és $\Sigma\in\mathbb{R}^{2\times 2}$ (invertálható) kovariancia mátrixszal, ha (U,V) standard normális pár, és $\mathbf{A}^\top\mathbf{A} = \Sigma$, ahol $\mathbf{A}\in\mathbb{R}^{2\times 2}$.
Egy kétdimenziós normális (X,Y) pár sűrűségfüggvénye:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2r\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\},$$

ahol $\mathbb{E}X = \mu_X$, $\mathbb{E}Y = \mu_Y$, $\mathbb{SD}(X) = \sigma_X$, $\mathbb{SD}(Y) = \sigma_Y$, és $r = r(X,Y)$ X és Y korrelációs egyyüthatója. Kétdimenziós normális eloszlás esetében a perem- illetve feltételes eloszlások is normálisok: a marginálisok eloszlása $N(\mu_X,\sigma_X)$ és $N(\mu_Y,\sigma_Y)$, a feltételes eloszlások pedig $N(\mu_{X|Y=y},\sigma_{X|Y=y})$, ahol

$$\mu_{X|Y=y} = \mu_X + (y-\mu_Y)\,r\,\frac{\sigma_X}{\sigma_Y} \qquad \text{és} \qquad \sigma_{X|Y=y} = \sqrt{1-r^2}\,\sigma_X\,.$$

Megjegyzés: az itt használt jelölés kicsit eltér a korábbiaktól. A szórás angolul standard deviation, ezért a megszokott $D(X)$ jelölés mellett $\mathbb{SD}(X)$ jelölést is használjuk. A szórásnégyzet angolul variance, ezért a megszokott $D^2(X)$ jelölés mellett a $\mathbb{V}\mathrm{ar}(X)$ jelölés is használjuk. Továbbá X és Y korrelációs együtthatójára a $\mathrm{r}(X,Y)$ jelölést használjuk.

A normális eloszlással számolható vagy közelíthető feladatok esetén elfogadható a standard normális eloszlásfüggvényével és excel függvénnyel való válasz is. A megoldásokban önkényesen váltogatjuk, hogy melyik formát választjuk, az is előfordul, hogy mindkét formát.

Kétdimenziós normális feladatok

1. Tegyük fel, hogy egy jólmenő étterem heti összbevétele normális eloszlást követ 1 millió forint várható haszonnal, és 700000 forint szórással. Mi annak a valószínűsége, hogy kevesebb, mint 1.5 millió forint a bevétele két, egymást követő héten? Itt tegyünk még fel függetlenséget! Majd nézzük meg, hogyan változik annak a valószínűsége, hogy a második héten több mint 2 millió forint a bevétel feltéve, hogy az első héten 1.5 millió forint volt a bevétel és a korreláció -0.5! Mi a második hét várható bevétele ugyanezen feltétel mellett? Mennyi a két hét várható összbevétele? Mi a szórás?

   Megoldás: Rögzítsük a jelöléseket: $\mu := 1$, $\sigma := 0.7$ millió forintban számolva. (X,Y) pár pedig az első, illetve második hét (véletlen) összbevétele (millióban), mindkettő tehát normális (μ,σ) paraméterrel. Függetlenséget feltéve (azaz $r = 0$):

$$\begin{aligned}\mathbb{P}(X<1.5, Y<1.5) &= \mathbb{P}(X<1.5)\mathbb{P}(Y<1.5) = \mathbb{P}\left(\frac{X-1}{0.7}<\frac{5}{7}\right)\mathbb{P}\left(\frac{Y-1}{0.7}<\frac{5}{7}\right) = \\ &= \Phi(5/7)\Phi(5/7)\approx 0.762^2 = 0.581\end{aligned}$$

   Namost tegyük fel, hogy $r = -0.5$. Az (X,Y) pár kétdimenziós (nem-standard) normális, peremeloszlásai is normálisak: $Y|X = x$ $(x = 1.5)$ várható értéke: $\mu_Y + r(x-\mu_X)\frac{\sigma_Y}{\sigma_X} = 1-0.5(1.5-1) = 3/4 = 0.75$ (ez a feltételes várható érték), szórása $\sigma_Y\sqrt{1-r^2} = 0.7\sqrt{3}/2\approx 0.606$, azaz behelyettesítve és standardizálva kapjuk, hogy

$$\mathbb{P}(Y>2\,|\,X=1.5) = \mathbb{P}\left(\frac{Y-0.75}{0.606} > 2.06\,|\,X=1.5\right) = 1-\Phi(2.06)\approx 0.02.$$

   Standardizálás helyett elfogadható a következő megoldás is:

$$\mathbb{P}(Y>2\,|\,X=1.5) = 1 - NORMDIST(2; 0.75; 0.606; TRUE).$$

Most koncentráljunk az összbevételre. A feladat nem kérdezi, de az összbevétel egydimenziós normális eloszlású, mert kétdimenziós normális eloszlás minden lineáris függvénye is normális eloszlású. Az összbevétel várható értéke könnyen adódik: $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 2$. A szórásnégyzet kiszámolásához jó felidézni, hogy ha átrendezzük a korreláció definícióját, akkor $\mathrm{Cov}(X,Y) = \mathrm{r}(X,Y)D(X)D(Y)$ összefüggést kapjuk. Ezt és az összeg szórásnégyzetére (más néven varianciájára) vonatkozó összefüggésből az összbevétel szórásnégyzete (szórás gyökvonással adódik): $\mathbb{V}\mathrm{ar}(X+Y) = \mathbb{V}\mathrm{ar}(X)+\mathbb{V}\mathrm{ar}(Y)+2\mathrm{Cov}(X,Y) = 2\times 0.7^2+2\times(-0.5)0.7^2 = 0.49$.

2. Budapesten májusban az átlagos hőmérséklet 25 °C, 7 °C szórással, valamint az átlagnyomás 10^5 Pa, 2×10^4 Pa szórással. A hőmérséklet/nyomás változása szoros összhangban van, köztük lévő korreláció 0.7. Írjuk fel a kovariancia mátrixot majd határozzuk meg a következőket:

   (a) Mi a valószínűsége annak, hogy egy nap melegebb lesz, mint 40 °C? És, hogy alacsonyabb a nyomás 6×10^4 Pa-nál?

   (b) Egy nap 20 °C-ot mértünk. Mi annak a valószínűsége, hogy a légnyomás 1.2×10^5 Pa fölött járt? Átlagosan mekkora volt a légnyomás? Mekkora a szórás?

   (c) Feltéve, hogy egy nap 10^5 Pa volt a légnyomás, mi annak a valószínűsége, hogy melegebb volt, mint 35 °C? Átlagosan hány fok volt aznap? Mekkora a szórás?

   Megoldás: Rögzítsük az adatokat: $\mu_X := 25$, $\sigma_X := 7$, $\mu_Y := 10^5$, $\sigma_Y := 2\times 10^4$ és $r := 0.7$, X jelenti az átlag hőmérsékletet, Y pedig a hozzátartozó nyomást. (X,Y) kétdimenziós normális, a kovariancia mátrix (a kovariancia mátrix definíciója felül olvasható, az egyetlen bonyodalom az, hogy ki kell számolni a kovarianciát a $\mathrm{Cov}(X,Y) = \mathrm{r}(X,Y)D(X)D(Y)$ összefüggéssel:

   $$\Sigma = \begin{pmatrix} 49 & 9.28\times 10^4 \\ 9.28\times 10^4 & 4\times 10^8 \end{pmatrix}.$$

   (a) $\mathbb{P}(X>40) = \mathbb{P}\left(\frac{X-25}{7} > \frac{15}{7}\right) = 1-\Phi(15/7) \approx 0.016$,
   vagy másképpen $\mathbb{P}(X>40) = 1 - NORMDIST(40;25;7;TRUE)$;
   $\mathbb{P}(Y<6\times 10^4) = \mathbb{P}\left(\frac{X-10^5}{2\times 10^4} < -2\right) = 1-\Phi(2) \approx 0.023$,
   vagy másképpen $\mathbb{P}(Y<6\times 10^4) = NORMDIST(6\times 10^4;10^5;2\times 10^4;TRUE)$;

   (b) $x=20$, $Y\,|\,X=x$ normális eloszlású $\left(\mu_Y + \frac{\sigma_Y}{\sigma_X}r(x-\mu_X), \sigma_Y\sqrt{1-r^2}\right)$ paraméterekkel. Azaz ennek a várható értéke: $10^5 + \frac{2\times 10^4}{7}(20-25)0.7 = 90000$, és szórása: $2\times 10^4\sqrt{1-0.7^2} \approx 14283$. A keresett valószínűség pedig:

   $$\mathbb{P}(Y>1.2\times 10^5\,|\,X=20) = \mathbb{P}\left(\frac{Y-90000}{14283} > 2.1\,|\,X=20\right) = 1-\Phi(2.1) \approx 0.018,$$

   vagy másképpen $\mathbb{P}(Y>1.2\times 10^5\,|\,X=20) = 1 - NORMDIST(120000;90000;14283;TRUE)$.

   (c) $y=10^5$, $X\,|\,Y=y$ normális eloszlású $\left(\mu_X + \frac{\sigma_X}{\sigma_Y}r(y-\mu_Y), \sigma_X\sqrt{1-r^2}\right)$ paraméterekkel. Azaz ennek a várható értéke: $25+\frac{7}{2\times 10^4}(10^5-10^5)0.7 = 25$, és szórása: $7\sqrt{1-0.7^2} \approx 5$. A keresett valószínűség pedig:

   $$\mathbb{P}(X>35\,|\,Y=10^5) = \mathbb{P}\left(\frac{X-25}{5} > 2\,|\,Y=10^5\right) = 1-\Phi(2) \approx 0.023,$$

   vagy másképpen $\mathbb{P}(X>35\,|\,Y=10^5) = 1 - NORMDIST(35;25;5;TRUE)$.

3. Magyarországon a felnőtt férfiak testmagassága átlagosan 178 cm, 9 cm szórással, míg testsúlyuk 85 kg, 10 kg szórással. A korrelációs együttható 0.7, azaz minél magasabb valaki, annál súlyosabb is. Írjuk fel a kovariancia mátrixot!

   (a) Mi a valószínűsége annak, hogy egy férfi magasabb 2 méternél? És, hogy nehezebb 100 kg-nál?

   (b) Feltéve, hogy egy férfi 80 kg, mi annak a valószínűsége, hogy magasabb, mint 180 cm? Várhatóan hány cm magas egy ilyen férfi? Mekkora a szórás?

   (c) Átlagosan mekkora súlyú egy 190 cm magas férfi?

   (d) Átlagosan milyen magas egy 94.3 kg-os férfi?

   (e) Hasonlítsuk össze az utolsó két eredményt.

   Megoldás: Rögzítsük az adatokat: $\mu_X := 178$, $\sigma_X := 9$, $\mu_Y := 85$, $\sigma_Y := 10$ és $r := 0.7$, X jelenti a magasságot, Y pedig hozzátartozó testsúlyt. (X,Y) kétdimenziós normális, a kovariancia mátrix:

   $$\Sigma = \begin{pmatrix} 81 & 63 \\ 63 & 100 \end{pmatrix}.$$

   (a) $\mathbb{P}(X>200) = \mathbb{P}\left(\frac{X-178}{9} > \frac{22}{9}\right) = 1-\Phi(22/9) \approx 0.0073$; $\mathbb{P}(Y>100) = \mathbb{P}\left(\frac{X-85}{10} > 1.5\right) = 1-\Phi(1.5) \approx 0.067$;

(b) $y = 80$, $X \mid Y = y$ normális eloszlású $\left(\mu_X + \frac{\sigma_X}{\sigma_Y} r(y-\mu_Y), \sigma_X\sqrt{1-r^2}\right)$ paraméterekkel. Azaz ennek a várható értéke: $178 + (9/10)(80-85) \times 0.7 = 174.85$, és szórása: $9\sqrt{1-0.7^2} \approx 6.43$. A keresett valószínűség pedig:

$$\mathbb{P}(X > 180 \mid Y = 80) = \mathbb{P}\left(\frac{X - 174.85}{6.43} > \frac{180-174.85}{6.43} \mid Y = 80\right) = 1 - \Phi(0.817) \approx 0.206.$$

Érdekességképpen: feltétel nélkül $\mathbb{P}(X > 180) = 1 - \Phi(2/9) \approx 0.412$.

(c) $x = 190$, $Y \mid X = x$ normális eloszlású $\left(\mu_Y + \frac{\sigma_Y}{\sigma_X} r(x-\mu_X), \sigma_Y\sqrt{1-r^2}\right)$ paraméterekkel. Azaz ennek a várható értéke: $85 + (10/9)(190-178) \times 0.7 \approx 94.3$.

(d) A korábbiak mintájára a 94.3 súlyú férfiák általában $178 + (9/10)(94.3-85) \times 0.7 \approx 183.8$ magasak.

(e) Általános megfogalmazásban a (c) és (d) feladatokban a regressziós egyeneseket (ami normális eloszlás esetén egybeesik a feltéles vérható értékkel) értékeltük ki megadott helyeken. A két alfeladat demonstrálja, hogy a két regressziós egyenes nem egymás inverze.

4. Az X áramerősség normális $N(230, 6)$ eloszlású, a mérőkészülék Z hibája ettől független $N(0, 8)$ eloszlású, mi az $Y = X + Z$ értéket mérjük. Mi a valószínűsége, hogy $Z > X/20$?

   Megoldás: A feladatban az Y képletét semmire nem használjuk. Először is a kérdés átírható: $\mathbb{P}(Z > X/20) = \mathbb{P}(20Z - X > 0)$. Vegyük észre, hogy mivel kétdimenziós normális eloszlás lineáris függvénye, ezért $20Z - X$ normális eloszlású, $\mathbb{E}(20Z - X) = -230$ várható értékkel és $\mathbb{V}\mathrm{ar}(20Z - X) = \mathbb{V}\mathrm{ar}(20Z) + \mathbb{V}\mathrm{ar}(-X) + 2\mathrm{Cov}(20Z, -X) = 400\mathbb{V}\mathrm{ar}(Z) + \mathbb{V}\mathrm{ar}(X) - 40\mathrm{Cov}(Z, X) = 400 \cdot 64 + 36 - 40 \cdot 0$ varianciával vagyis $\mathbb{SD}(20Z - X) = 160,112$ szórással. Ezt használva $\mathbb{P}(20Z - X > 0) = 1 - NORMDIST(0; -230; 160,112; TRUE) \approx 0,075$.

5. Legyen a (X, Y) pár kétdimenziós normális eloszlású, r korrelációval. Mi az eloszlása $U = X + Y$-nak és $V = X - Y$-nak? Független-e U a V-től? Számoljuk ki a várható értékeket és szórásokat is!

   Megoldás: Az (U, V) kétdimenziós normális eloszlású pár, mert az (X, Y) kétdimenziós normális eloszlású valószínűségi változó pár lineáris függvénye. Emiatt U és V pontosan akkor függetlenek, ha korrelálatlanok: $\mathrm{Cov}(U, V) = \mathrm{Cov}(X + Y, X - Y) = \mathrm{Cov}(X, X) - \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, X) - \mathrm{Cov}(Y, Y)$, a két középső tag kiejti egymást, így használva, hogy a variancia az az önmagától való kovariancia adódik $\mathrm{Cov}(U, V) = \mathbb{V}\mathrm{ar}(X) - \mathbb{V}\mathrm{ar}(Y)$, azaz, ha $\sigma_X = \sigma_Y$, akkor függetlenek, egyébként függőek. Valamint $\mathbb{E}(X \pm Y) = \mu_X \pm \mu_Y$; $\mathbb{V}\mathrm{ar}(X \pm Y) = \sigma_X + \sigma_Y \pm 2r\sigma_X\sigma_Y$.

6. Hogyan generálna le kétdimenziós normális eloszlású véletlen pontokat a síkon, melyek várható értéke (μ_1, μ_2), szórása (σ_1, σ_2), korrelációs együtthatója pedig r. Függetlenek-e a koordináták, ha $r = 0$?

   Megoldás: Vegyünk egy (U, V) standard normális véletlen párt (azaz U, V standard normális véletlen szám, egymástól függetlenül - nem azonosak az előző feladatban szereplőkkel). Adottak a $\mu_1, \mu_2 \in \mathbb{R}$; $\sigma_1, \sigma_2 \in \mathbb{R}^+$ és $-1 < r < 1$ számok. Kell: egy (X, Y) pár, hogy $\mathbb{E}X = \mu_1$, $\mathbb{E}Y = \mu_2$, $\mathbb{SD}(X) = \sigma_1$, $\mathbb{SD}(Y) = \sigma_2$, illetve $\mathrm{Cov}(X, Y) = r\sigma_1\sigma_2$, és persze (X, Y) kétdimenziós normális. Ezt lineáris transzformációval érjük el a standard (U, V) párból. Legyen $X = \sigma_1 U + \mu_1$, ezzel az egyik marginális már megvan, a másikat $Y := aU + bV + \mu_2$ alakban keressük. Y várható értéke be van lőve, már csak $a, b \in \mathbb{R}$ – egyelőre ismeretlen – paramétereket kell úgy megválasztani, hogy Y szórása, illetve X, Y kovarianciája megfelelő legyen. Erre pedig a következő adódik: a függetlenség miatt $\mathbb{V}\mathrm{ar}(Y) = \mathbb{V}\mathrm{ar}(aU + bV) = a^2 + b^2 = \sigma_2^2$, illetve $\mathrm{Cov}(X, Y) = \mathrm{Cov}(\sigma_1 U + \mu_1, aU + bV + \mu_2) = \sigma_1 a = r\sigma_1\sigma_2$, azaz $a = r\sigma_2$, illetve $b = \sigma_2\sqrt{1-r^2}$. Ezzel $(X = \sigma_1 U + \mu_1;\ Y = r\sigma_2 U + \sigma_2\sqrt{1-r^2}V + \mu_2)$ kétdimenziós normális a megfelelő paraméterekkel. U, V-t a megszokott módon így generálunk az (A1,B1) cella párban: A1=Norm.Inv(Rand();0;1), B1=Norm.Inv(Rand();0;1). Ebből pedig az előbbi transzformációval kapunk kétdimenziós normálisat. Ha $r = 0$, a koordináták függetlenek, mert normális esetben korrelálatlanságból következik a függetlenség, és ez konkrétan meg is jelenik, ugyanis a fenti transzformációban: $r = 0$ esetén Y csak V-től függ, U-tól nem.

CHT

Centrális határeloszlás-tétel (CHT): legyenek $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók, $\mu := \mathbb{E}X_i \in \mathbb{R}$ és $\sigma := \mathbb{SD}(X_i) \in \mathbb{R}^+$. Ekkor

$$\mathbb{P}\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} < x\right) \longrightarrow \Phi(x), \text{ amint } n \to +\infty, \ (x \in \mathbb{R}).$$

Mindez szóban: elég nagy n esetén a – FAE – valószínűségi változók standardizált összege közelítőleg (standard) normális eloszlású. Speciálisan: ha az X_i változókat azonos, p paraméterű Bernoulli változóknak választjuk, akkor jutunk el a de Moivre–Laplace tételhez (avagy binomiális CHT).

7. Mennyi annak a valószínűsége, hogy 12 000 kockadobás során előforduló hatosok száma 1900 és 2150 közé esik?

   Megoldás: Legyen S_{12000} a hatosok száma, ekkor S_{12000} egy $n = 12000$, $p = \frac{1}{6}$ paraméterű binomiális val. változó, ami köztudottan felírható független, azonos eloszlású valószínűségi változók összegeként: $I_1 + I_2 + \cdots + I_{12000}$ alakban, ahol az I_k indikátor 1 ha a k. dobásunk 6-os, egyébként 0. Vagy fejből vagy az előbbi indikátoros felírásból adódik, hogy $\mathbb{E}S_{12000} = np = 2000$ és $\mathbb{SD}(S_{12000}) = \sqrt{npq} = \frac{100}{\sqrt{6}}$. Az indikátoros felírásból adódóan S_{12000} közelíthető $N(2000, \frac{100}{\sqrt{6}})$ eloszlással vagyis

$$\mathbb{P}(S_{12000} \in [1900, 2150]) = \mathbb{P}(S_{12000} - 2000 \in [-100, 150]) = \mathbb{P}(\frac{S_{12000} - 2000}{\frac{100}{\sqrt{6}}} \in [-\sqrt{6}, \frac{3}{2}\sqrt{6}]) \approx \Phi(\frac{3}{2}\sqrt{6}) - \Phi(-\sqrt{6}).$$

A válasz természetesen a $NORMDIST(2150; 2000; \frac{100}{\sqrt{6}}, TRUE) - NORMDIST(1900; 2000; \frac{100}{\sqrt{6}}, TRUE)$ alakban is elfogadható. Megjegyezzük, hogy mivel ismerjük S_{12000} pontos eloszlását (binomiális a megadott paraméterekkel), ezért a választ pontosan is kiszámolhattuk volna. A fenti közelítő megoldás jóval kevesebb számolást igényel.

8. Egy gyár két fajta érmét gyárt: egy igazságosat, és egy hamisat ami 55% eséllyel mutat fejet. Van egy ilyen érménk, de nem tudjuk igazságos-e vagy pedig hamis. Ennek eldöntésére a következő statisztikai tesztet hajtjuk végre: Feldobjuk az érmét 1000-szer. Ha legalább 525-ször fejet mutat, akkor hamisnak nyilvánítjuk, ha 525-nél kevesebb fej lesz a dobások között, akkor az érmét igazságosnak tekintjük. Mi a valószínűsége, hogy a tesztünk téved abban az esetben, ha az érme igazságos volt? És ha hamis volt?

   Megoldás:

   Kétféle hibával kell számolnunk:

   - Igazságos az érme, de mi hamisnak mondjuk. Ennek valószínűsége $\mathbb{P}(S_{1000}^{\text{igaz}} \geq 525)$, ahol S_{1000}^{igaz} binomiális eloszlású ($n = 1000, p = 0,5$) paraméterekkel (hiszen feltesszük, hogy az igazságos érmével dobtunk). A cht miatt S_{1000}^{igaz} közelíthető egy $1000 \cdot 0,5 = 500$ várható értékű és $\sqrt{1000 \cdot 0,5 \cdot 0,5} \approx 15,81$ szórású normális eloszlással, ezért $\mathbb{P}(S_{1000}^{\text{igaz}} \geq 525)$ körülbelül $1 - NORMDIST(525; 500; 15,81, TRUE) \approx 0,057$.
   - Hamis az érme, de mi igazságosnak mondjuk. Ennek valószínűsége $\mathbb{P}(S_{1000}^{\text{hamis}} < 525)$, ahol S_{1000}^{hamis} binomiális eloszlású ($n = 1000, p = 0,55$) paraméterekkel. A cht miatt S_{1000}^{hamis} közelíthető egy $1000 \cdot 0,55 = 550$ várható értékű és $\sqrt{1000 \cdot 0,55 \cdot 0,45} \approx 15,73$ szórású normális eloszlással, ezért $\mathbb{P}(S_{1000}^{\text{hamis}} < 525)$ körülbelül $NORMDIST(525; 500; 15,73, TRUE) \approx 0,055$.

   Megjegyezzük, hogy ez tekinthető egy bevezető feladatnak a matematikai statisztikába.

9. Határozzuk meg azt a k egész számot, amelyre igaz, hogy annak a valószínűsége, hogy 1000 érmedobás során a fejek száma $500 - k$ és $500 + k$ közé esik, kb. 0.9.

   Megoldás: S_{1000} binomiális eloszlású ($n = 1000, p = 0,5$) paraméterekkel. A cht miatt S_{1000} közelíthető egy $1000 \cdot 0,5 = 500$ várható értékű és $\sqrt{1000 \cdot 0,5 \cdot 0,5} \approx 15,81$ szórású normális eloszlással. Itt kényelmesebb a standard normális eloszlásra való visszavezetéssel dolgozni, mert a $\Phi(-x) = 1 - \Phi(x)$ szimmetriából adódó azonossághoz jobban hozzászoktunk.

$$0,9 \approx \mathbb{P}(500 - k < S_{1000} < 500 + k) = \mathbb{P}(< \frac{-k}{15,81} < \frac{S_{1000}}{15,81} < \frac{k}{15,81}) = \Phi(\frac{k}{15,81}) - \Phi(\frac{-k}{15,81}) = 2\Phi(\frac{k}{15,81}) - 1.$$

   A fentiből $\Phi(\frac{k}{15,81}) \approx \frac{0,9+1}{2} = 0,95$ adódik, amiből

$$\frac{k}{15,81} \approx \Phi^{-1}(0,95) = NORM.INV(0,9; 0; 1) \approx 1,64.$$

   Így $k \approx 1,64 \cdot 15,81 \approx 26$.

10. Mennyi a valószínűsége annak, hogy 50 darab független és azonos eloszlású valószínűségi változó összege a [0, 30] intervallumba esik, ha egy ilyen változó eloszlása a [0, 1] intervallumon

    (a) egyenletes;
    Megoldás: A bevezetésben szereplő jelöléssel: $m = \frac{1}{2}$, $\sigma = \frac{1}{\sqrt{12}}$, és $n = 50$. Behelyettesítve:

$$P(S_{50} < 30) = P(\frac{S_{50} - 25}{\sqrt{\frac{50}{12}}} < \frac{30 - 25}{\sqrt{\frac{50}{12}}}) \approx \Phi(\sqrt{6}) \approx 0.993$$

    (b) $f(x) = 2x$ sűrűségfüggvény szerint alakul?
    Megoldás: A bevezetésben szereplő jelöléssel: $m = \int_0^1 2x \cdot x dx = \frac{2}{3}$, $\sigma^2 = \int_0^1 2x \cdot x^2 dx - m^2 = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18}$, és $n = 50$. Behelyettesítve:

$$P(S_{50} < 30) = P(\frac{S_{50} - 50 \cdot \frac{2}{3}}{\sqrt{\frac{50}{18}}} < \frac{30 - 50 \cdot \frac{2}{3}}{\sqrt{\frac{50}{18}}}) \approx \Phi(-1.63) \approx 0.0516$$

11. Becsüljük meg annak valószínűségét, hogy 10 000 kockadobás összege 34 800 és 35 200 közé esik.

    Megoldás: Legyen $S_{10000} = X_1 + X_2 + \dots X_{10000}$ 10000 darab független kockadobás összege. Egy darab kockadaobás várható értéke 3,5, szórása pedig $\sqrt{\frac{7 \cdot 5}{12}} \approx 1,7$ (lásd VII/6-os feladat). Így a CHT miatt S_{10000} közelíthető egy $10000 \cdot 3,5 = 35000$ várható értékű $\sqrt{10000} \cdot 1,7 = 170$ szórású normális eloszlással. Így a válasz közelítőleg $NORMDIST(35200; 35000; 170; TRUE) - NORMDIST(34800; 35000; 170; TRUE) \approx 0,76$.

12. Egy kockát folyamatosan feldobunk addig, amíg a dobások összege meghaladja a 300-at. Becsüljük meg annak valószínűségét, hogy legalább 80 dobásra van ehhez szükség.

    Megoldás: Akkor van szükség legalább 80 dobásra, ha az első 79 dobás összege nem haladja meg a 300-t. Az előző feladat mintájára a CHT-t használva a következő adódik:

    $$\mathbb{P}(S_{79} \leq 300) = \mathbb{P}(\frac{S_{79} - 79 \cdot 3,5}{\sqrt{79} \cdot 1,7} \leq \frac{300 - 79 \cdot 3,5}{\sqrt{79} \cdot 1,7}) \approx \Phi(1,55) \approx 0,94.$$

13. Adott 100 égőnk, melyek élettartama egymástól független exponenciális eloszlású, 5 óra várható értékkel. Tegyük fel, hogy az égőket egymás után használjuk, azonnal kicserélve azt, amelyik kiégett. Becsüljük meg annak valószínűségét, hogy 525 óra után még van működő égőnk.

    Megoldás: $S_{100} = X_1 + X_2 + \cdots X_{100}$, ahol $X_1, X_2, \ldots, X_{100}$ független azonososan $\frac{1}{5}$ paraméterű exponenciális eloszlású valószínűségi változók. $\mathbb{E}(S_{100}) = 100 \cdot \mathbb{E}(X) = 100 \cdot 5 = 500$. $\mathbb{SD}(S_{100}) = \sqrt{100} \cdot \mathbb{SD}(X) = 10 \cdot 5 = 50$. Az exponenciális eloszlás szórása ha nem tudjátok fejből, akkor az $\mathbb{SD}(X) = \sqrt{\mathbb{E}(X^2) - (\mathbb{E}(X))^2}$ és az $\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^{\infty} x^2 \frac{1}{5} e^{-\frac{1}{5}x} dx$ összefüggésekből számolható.

    A CHT alapján S_{100} közelítőleg $N(500; 50)$ eloszlású, így a kérdésre a következő közelítő válasz adható:

    $$\mathbb{P}(S_{100} > 525) \approx 1 - NORMDIST(525; 500; 50; TRUE) \approx 0,31, \text{ vagy másképpen}$$
    $$\mathbb{P}(S_{100} > 525) = \mathbb{P}(\frac{S_{100} - 500}{50} > \frac{525 - 500}{50}) \approx 1 - \Phi(\frac{1}{2}) \approx 0,31.$$

14. A jegyiroda előtt a fiatalok hosszú sorban állnak egy koncertjegyért. Ebben a pillanatban éppen 18-an állnak az egyik pénztár előtt. Megfigyeltem, hogy egy vásárló kiszolgálási ideje memória nélküli (más néven örökifjú) valószínűségi változó 3 perc átlaggal és a kiszolgálási idők függetlenek. Becsülje meg annak a valószínűségét, hogy a most utolsóként álló fiatal több mint 60 percet fog a pénztár előtt eltölteni!

    Megoldás: A most utolsó fiatalnak meg kell várnia a többiek kiszolgálását és a sajátját is, így a pénztár előtt eltöltött ideje így irható: $S_{18} = X_1 + X_2 + \cdots X_{18}$, ahol $X_1, X_2, \ldots, X_{18}$ független azonososan $\frac{1}{3}$ paraméterű exponenciális eloszlású valószínűségi változók (örökifjúság miatt). Az előző feladattal analóg módon, így a $\mathbb{P}(S_{18}) > 60$ kérdésre körülbelül válaszolhatunk a $1 - NORMDIST(60; 18 \cdot 3; \sqrt{18} \cdot 3; TRUE)$ formulával.

15. Egy szabályos érmét 40-szer feldobunk, és X-szel jelöljük a kapott fejek számát. Határozzuk meg annak valószínűségét, hogy $X = 20$

    - a binomiális eloszlás segítségével,

      Megoldás: Binomiális eloszlással kell számolni, így a választ a $BINOMDIST(20; 40; 0,5; FALSE)$ függvény szolgáltatja, ami körülbelül $0,1253$. Itt azért kellett FALSE-t írnunk az Excel függvénybe, mert a binomiális eloszlás súlyfüggvényét akartuk használni, ami hasonló a folytonos eloszlások sűrűségfüggvényéhez (persze utóbbi nem a valószínűséget adja meg az előbbivel ellentétben).

    - a deMoivre–Laplace-tételt használva. Ez utóbbihoz segítség: $\mathbb{P}\{X = 20\} = \mathbb{P}\{19.5 \leq X < 20.5\}$, ami persze nem számít amíg X-et binomiálisnak (azaz egész értékűnek) tekintjük, de fontos lesz a deMoivre–Laplace-tétel alkalmazásánál.

      Megoldás: Az X binomiális, mint tudjuk felírható független indikátorok összegeként, ezért a cht használva kozelíthető egy $40 \cdot 0,5 = 20$ várható értékű és $\sqrt{40 \cdot 0,5 \cdot 0,5} = 0,5\sqrt{40}$ szórású normális eloszlással, így

      $$\mathbb{P}(19,5 \leq X < 20,5) \approx NORMDIST(20,5; 20; 0,5\sqrt{40}, TRUE) - NORMDIST(19,5; 20; 0,5\sqrt{40}, TRUE),$$

      ami körülbelül $0,1256$. Látható, hogy meglehetősen jó közelítést kaptunk.