

1. Alapfeladatok

1. Legyenek X és Y független kockadobások eredményei. Mennyi $X + Y$ és $X - Y$ kovarianciája? Függetlenek-e?

Megoldás. A kovarianca bilinearitását használva látjuk, hogy $\text{cov}(X + Y, X - Y) = 0$. (És nem is kell használnunk X és Y függetlenességét.) Viszont $X + Y$ és $X - Y$ nem függetlenek, $X + Y = 12 \Rightarrow X - Y = 0$.

2. Legalább mekkora valószínűséggel állíthatjuk, hogy egy szabályos érmével végzett 100-as dobássorozatokban a fejdobások száma legalább 44 és legfeljebb 56? Mit mondhatunk 1000-es dobássorozat esetén arról, hogy a fejdobások száma legalább 440 és legfeljebb 560?

Megoldás. Alkalmazzuk a Csebisev-egyenlőtlenséget az fejek számára, vagyis az X valószínűségi változóra, melynek eloszlása binomiális eloszlás: és a várható értéke 50:

$$\mathbb{P}(44 \leq X \leq 56) = 1 - \mathbb{P}(|X - 50| \geq 7) \geq 1 - \frac{D^2(X)}{7^2} = 1 - \frac{100 \cdot 0,5 \cdot 0,5}{7^2} = 1 - \frac{25}{49} = 49\%.$$

Hasonlóképpen, ha Y a fejek száma 1000 dobásból, akkor $n = 1000$ és $\mathbb{E}(Y) = 500$, így

$$\mathbb{P}(440 \leq Y \leq 560) = 1 - \mathbb{P}(|Y - 500| \geq 60) \geq 1 - \frac{D^2(Y)}{60^2} = 1 - \frac{1000 \cdot 0,5 \cdot 0,5}{60^2} = 93,3\%.$$

Vagyis ugyanolyan arányok közé esés valószínűségére jobb becslést tudunk mondani, ha több dobás van.

Az Azuma-egyenlőtlenségből az első esetben 24%, második esetben 99%-os alsó becslés adódik. Ha a globális de Moivre-Laplace-szal közelítünk, akkor 74% és 99.99% értéket kapunk.

3. Legyen $n \geq 1$ esetén $\xi_n \sim \text{Poisson}(n)$ és $x \in \mathbb{R}$. Számoljuk ki a $\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\xi_n - n}{\sqrt{n}} < x\right)$ határértéket.

Megoldás. Tudjuk korábban, hogy ha X és Y függetlenek, Poisson-eloszlásúak, akkor $X + Y$ is Poisson-eloszlású, és paramétere a paraméterek összege. Ezért ha $Y_1, Y_2, \dots$ független 1 paraméterű Poisson-eloszlásúak, ξ_n azonos eloszlású az $Y_1 + \dots + Y_n$ összeggel. Azt is tudjuk továbbá, hogy Y_1 várható értéke és szórása is 1. Vagyis a kérdéses valószínűség így írható fel:

$$\mathbb{P}\left(\frac{\sum_{j=1}^n Y_j - n\mathbb{E}(Y_1)}{D(Y_1)\sqrt{n}} < x\right).$$

Mivel $Y_1, Y_2, \dots$ függetlenek, azonos eloszlásúak, véges szórásúak, teljesítik a centrális határeloszlástétel feltételeit. A tétel alapján ennek a valószínűségnek a limesze $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx$.

4. Tegyük fel, hogy az X valószínűségi változó exponenciális eloszlású λ paraméterrel, azaz sűrűségfüggvénye $\lambda e^{-\lambda x} \mathbb{I}(x \geq 0)$, azaz $\lambda e^{-\lambda x}$, ha $x > 0$, és 0 különben. Számítsuk ki és hasonlítsuk össze a $\mathbb{P}(X \geq t)$ és $\mathbb{P}(X \geq s + t | X \geq s)$ valószínűségeket, ahol s, t pozitív számok.

Megoldás. A feltételes valószínűség definícióját és az exponenciális eloszlás eloszlásfüggvényének alakját felhasználva

$$\begin{aligned} \mathbb{P}(X \geq s + t | X \geq s) &= \frac{\mathbb{P}(\{X \geq s + t\} \cap \{X \geq s\})}{\mathbb{P}(X \geq s)} = \frac{1 - \mathbb{P}(X < s + t)}{1 - \mathbb{P}(X < s)} = \\ &= \frac{1 - F(s + t)}{1 - F(s)} = \frac{1 - (1 - e^{-\lambda(s+t)})}{1 - (1 - e^{-\lambda s})} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \\ &= e^{-\lambda t} = 1 - (1 - e^{-\lambda t}) = 1 - F(t) = \mathbb{P}(X \geq t). \end{aligned}$$

Ezt nevezik az exponenciális eloszlás **örökifjú tulajdonságának**.

5. Arizona egy kis falujában 201-en szavaztak az amerikai elnökválasztáson. Feltételezzük, hogy minden szavazat azonos valószínűséggel érkezik az egyes jelöltekre. A szavazatokat egyenként, sorban egymás után számolták meg. 103-an szavaztak Trumpra, és 98-as Harrisre. Mennyi a valószínűsége, hogy az első három szavazat mindegyikét Trump kapta?

Megoldás

Tekintsük ezt egy egyszerű szimmetrikus bolyongásnak, ahol Trump szavazatai a felfelé, Harris szavazatai a lefelé lépések. Tudjuk, hogy a bolyongás 201 lépés után az 5-ben ér véget. Az ilyen utak száma $\binom{201}{103}$.

Ezek között kell meghatároznunk a három felfelé lépéssel kezdődő utak arányát, hiszen, mivel minden utat egyformán valószínűnek tételezünk fel, ez adja meg a feltételes valószínűséget. Ilyenkor a 3-ból kell 198 lépés alatt az 5-be jutni, az ilyen utak száma $\binom{198}{100}$ (további 100 felfelé lépés kell). Vagyis a kérdéses valószínűség:

$$\frac{\binom{198}{100}}{\binom{201}{103}} = \frac{103 \cdot 102 \cdot 101}{201 \cdot 200 \cdot 199},$$

ami azt is mutatja, hogy számolhatunk úgy is, hogy a 201 beérkezett szavazatot, amiből 103 Trumpra lett leadva, véletlen sorrendben dolgozzák fel.

6. Legyen $\nu_r = \inf\{n : S_n = r\}$ az első olyan időpont, amikor a bolyongás megérkezik az $r > 0$ szintre, M_n pedig az első n lépés során elért legnagyobb szint. Ekkor $n \geq 1$, $n \equiv r \pmod{2}$ esetén

$$\mathbb{P}(\nu_r \leq n) = \mathbb{P}(M_n \geq r),$$

továbbá

$$\mathbb{P}(\nu_r = n) = \frac{r}{n} \mathbb{P}(S_n = r).$$

Megoldás

Az állítás első része könnyen látható: az r szintet pontosan akkor értük el n lépésen belül, ha az első n lépés során a maximum legalább r .

A második rész bizonyításához: az r szint első elérése csak úgy következhet be az n . lépésben, ha $n - 1$ lépés után a maximum $r - 1$, a bolyongás is éppen itt tartózkodik, és felfelé lépünk. Ezt átalakíthatjuk úgy, hogy az $(r - 1)$ -ben végződő sétákból kivonjuk azokat, amik $(r - 1)$ -ben végződnek, de a maximum nagyobb, mint $r - 1$. Ez utóbbinak az átlakításához használhatjuk a tükrözési elvet:

$$\begin{aligned} \mathbb{P}(\nu_r = n) &= \frac{1}{2} \cdot \mathbb{P}(M_{n-1} = r - 1, S_{n-1} = r - 1) = \\ &= \frac{1}{2} \cdot (\mathbb{P}(S_{n-1} = r - 1) - \mathbb{P}(S_{n-1} = r - 1, M_{n-1} \geq r)) = \\ &= \frac{1}{2} \cdot (\mathbb{P}(S_{n-1} = r - 1) - \mathbb{P}(S_{n-1} = r + 1)) \stackrel{(*)}{=} \frac{1}{2^n} \left(\binom{n-1}{\frac{n-1}{2}} - \binom{n-1}{\frac{n-1}{2} - 1} \right) = \\ &= \frac{1}{2^n} \cdot \binom{n}{\frac{n-r}{2}} \cdot \left(\frac{n+r}{2n} - \frac{n-r}{2n} \right) = \frac{1}{2^n} \cdot \binom{n}{\frac{n-r}{2}} \cdot \frac{r}{n} \stackrel{(*)}{=} \frac{r}{n} \mathbb{P}(S_n = r). \end{aligned}$$

A $(*)$ lépésekben az S_n eloszlását használtuk (adott számú lépésből a felfelé menőket választjuk ki, innen adódik a binomiális együttható).

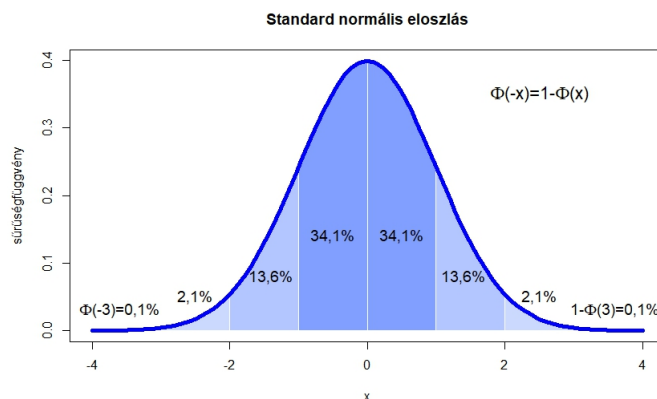
2. Gondolkodtató feladatok

1. Szeretnénk megállapítani, hogy hány dohányos él Budapesten. Ezért megkérdezzük n véletlenszerűen kiválasztott budapesti lakost arról, hogy dohányoznak-e (visszatevéses mintavétellel). A Csebisev-egyenlőtlenség alapján milyen nagyra kell n -et választani, ha azt szeretnénk, hogy a kapott relatív gyakoriság legfeljebb 1 százalékot tévedjen (additív értelemben) legalább 95% valószínűséggel? És ha a de Moivre–Laplace-tétel *globális alakját* (illetve a centrális határeloszlástételt) használjuk? (A megoldásban használhatjuk a közelítést a határérték alapján, a hibataggal nem kell számolni.)

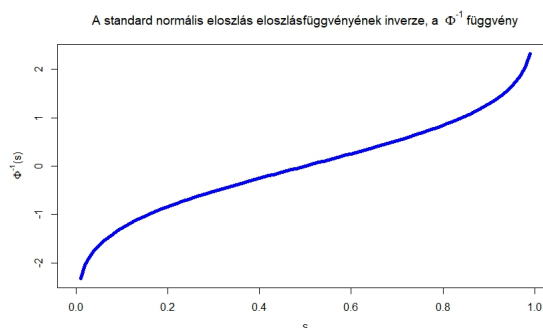
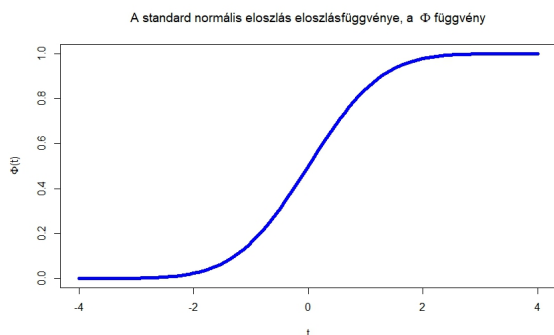
Megoldás. Tegyük fel, hogy minden megkérdezett egymástól függetlenül p valószínűséggel dohányzik ($p \in [0, 1]$ ismeretlen). A megkérdezettek száma legyen n (ez ismert, sőt szabadon megválasztható). Itt feltettük, hogy mindenki válaszol, és igazat mond (vagyis nem megbízható válaszok esetén az eredmény sem megbízható), és mindenkit azonos valószínűséggel találunk meg. A függetlenséget viszonylag könnyű biztosítani, ha függetlenül választunk, és a válaszadók nem látják egymás választát.

A relatív gyakoriság: Y_n/n , azaz Y_n ember mondta, hogy dohányzik. Itt Y_n binomiális eloszlású, az Y_n/n várható értéke p . Vagyis a feltétel, amit teljesíteni kell: minden $p \in [0, 1]$ -re

$$\mathbb{P}\left(\left|\frac{Y_n}{n} - p\right| < 0,01\right) \geq 0,95 \quad \Leftrightarrow \quad \mathbb{P}\left(\left|\frac{\sum_{i=1}^n X_i}{n} - p\right| \geq 0,01\right) \leq 0,05. \quad (1)$$



1. ábra. A φ függvény



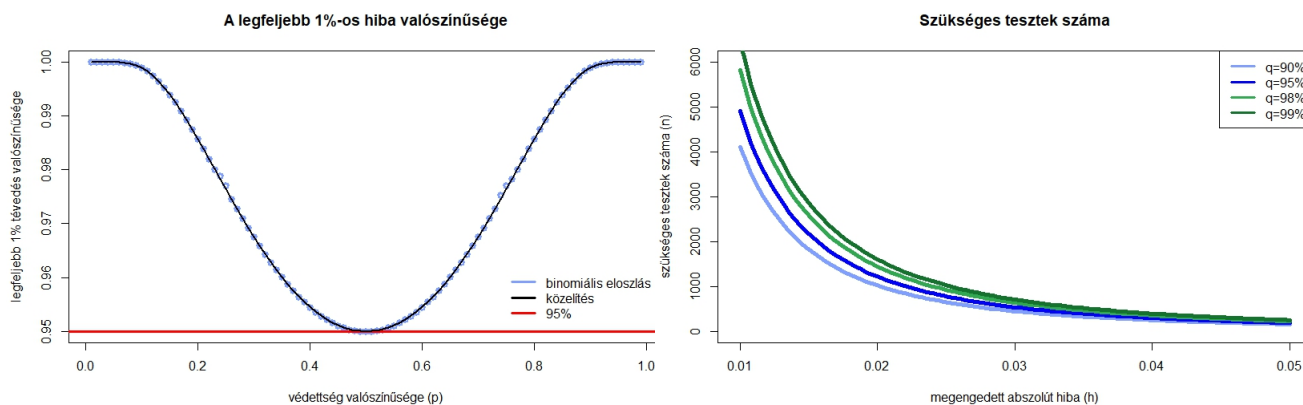
2. ábra. A standard normális eloszlásfüggvény, azaz $\Phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ és Φ^{-1}

A globális de Moivre–Laplace-tételből a következőt tudjuk (hiszen $\sum_{j=1}^n X_j$ eloszlása binomális eloszlás n renddel és p paraméterrel):

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(a \leq \frac{\sum_{j=1}^n X_j - np}{\sqrt{np(1-p)}} \leq b\right) = \Phi(b) - \Phi(a) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx.$$

Alakítsuk át a kérdésünket úgy, hogy ehhez hasonló alakot kapjunk.

$$\begin{aligned} \mathbb{P}\left(\left|\frac{Y_n}{n} - p\right| < 0,01\right) &= \mathbb{P}\left(-0,01 \leq \frac{Y_n - np}{n} \leq 0,01\right) = \\ &= \mathbb{P}\left(-\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}} \leq \frac{Y_n - np}{\sqrt{np(1-p)}} \leq \frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right). \end{aligned}$$



3. ábra. A szükséges tesztek száma normális közelítéssel néhány megbízhatósági szint mellett a megengedett abszolút hiba függvényében (balra), illetve annak valószínűsége, hogy legfeljebb 1%-ot téved a becslés, a valós p véletlenségi arány függvényében, $n = 9604$ teszt esetén

Itt megjelent Y_n standardizáltja, használhatjuk a globális de Moivre–Laplace-tételt a binomiális eloszlású Y_n -re (bár a tételben a és b rögzített volt, itt pedig a határok is végtelenhez tartanak; a Berry–Esséen-tétel akkor lenne jól használható, ha például $0,1 < p < 0,9$ -et tudnánk)

$$\mathbb{P}\left(\left|\frac{\sum_{i=1}^n X_i}{n} - p\right| < 0,01\right) \approx \Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) = 2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1,$$

a $\Phi(-a) = 1 - \Phi(a)$ azonosság alapján. Tehát az kell, hogy

$$2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 0,95$$

$$\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 0,975$$

$$\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}} \geq \Phi^{-1}(0,975) = 1,96$$

Itt használtuk, hogy a Φ monoton növekvő. Ebből

$$n \geq 196^2 \cdot p(1-p) = 38416 \cdot p(1-p)$$

Ehhez pedig $p(1-p) \leq 1/4$ alapján elég, hogy

$$n \geq \left(\frac{1,96 \cdot 0,5}{0,01}\right)^2 = 9604.$$

Vagyis $n = 9604$ embert elég megkérdezni. Számítógéppel ellenőrizve ez lényegében a pontos érték (3. ábra), bár talán 9602 is elég.

A másik módszerrel:

Csebisev-egyenlőtlenség véges szórású Y -ra és $t > 0$ -ra:

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq t) \leq \frac{D^2(Y)}{t^2}.$$

Itt $Y = \sum X_i$ binomiális eloszlású, várható értéke pn . Tehát a Csebisev-egyenlőtlenség alapján

$$\mathbb{P}\left(\left|\frac{\sum_{i=1}^n X_i}{n} - p\right| \geq 0,01\right) = \mathbb{P}\left(\left|\sum_{i=1}^n X_i - pn\right| \geq 0,01n\right) \leq \frac{D^2(\sum X_i)}{0,01^2 n^2} = \frac{np(1-p)}{0,01^2 n^2} = \frac{p(1-p)}{0,01^2 n}.$$

Ezért elég, hogy minden p -re

$$\frac{p(1-p)}{n \cdot 0,01^2} \leq 0,05.$$

Mivel $p(1-p) \leq 1/4$ minden p -re, ehhez elég, hogy

$$n \geq \frac{1}{4 \cdot 0,01^2 \cdot 0,05} = 50000.$$

Ennek a példának egy részletesebb kifejtése:

<https://ematlap.hu/tudomany-tortenet-2020-12/992-mennyit-teszteljunk-2-v3>

<https://ematlap.hu/tudomany-tortenet-2020-12/992-mennyit-teszteljunk-2-v3>

2. Arizona egy kis falujában 201-en szavaztak az amerikai elnökválasztáson. Feltételezzük, hogy minden szavazat azonos valószínűséggel érkezik az egyes jelöltekre. A szavazatokat egyenként, sorban egymás után számolták meg. 103-an szavaztak Trumpra, és 98-as Harrisre. Mennyi a valószínűsége, hogy a szavazatszámolásnál végig Trump vezetett? Egyenlőséget sem engedünk meg menet közben.

Megoldás

A folyamathoz rendelünk egy egyszerű szimmetrikus bolyongást, mely felfelé lép, ha megszámolnak egy Trump-szavazatot, lefelé, ha megszámolnak egy Harris-szavazatot. Mivel előzetesen azt feltételeztük, hogy mindenki $1/2$ valószínűséggel szavaz a két jelöltre, valóban egy egyszerű szimmetrikus bolyongást kapunk. A kérdés így fogalmazható meg: feltéve, hogy a bolyongás $103 - 98 = 5$ -ben ér véget, mennyi a valószínűsége, hogy végig pozitív? Vagyis, a $0 \rightarrow 5$ utak között hány olyan van, ami végig pozitív?

Ezt pedig átfogalmazhatjuk úgy, hogy az előző feladathoz hasonlót kapjunk. $X_0 = 0, X_1, X_2, \dots, X_{201}$ helyett tekintsük az $5 - X_{201}, 5 - X_{200}, \dots, 5 - X_1, 5 - X_0$ sorozatot. Ez 0-ból indul, 5-ben ér véget, és ha az X -ek mind pozitívak, akkor ez azt jelenti, hogy ez a folyamat először a 201. lépésben éri el az 5 szintet. Könnyen látható, hogy a módosított folyamat is egyszerű szimmetrikus bolyongás. Ezért a szavazásra vonatkozó kérdés pontos megfelelője ez lesz: feltéve, hogy egy egyszerű szimmetrikus bolyongás 201 lépés után az 5-ben ér véget, mennyi a valószínűsége, hogy korábban nem járt az 5-ben?

Az előző feladatból rögtön következik, hogy ez a feltételes valószínűség $5/201$. Mivel pedig a kérdés ekvivalens az eredetivel, $5/201$ annak valószínűsége, hogy Trump végig vezetett a szavazatszámolás során.

3. Gyakorló feladatok

1. Legyen $0 < q < 1$ esetén $\xi_q \sim \text{Geo}(q)$ és $x > 0$. Számoljuk ki a $\lim_{q \rightarrow 0} \mathbb{P}(q\xi_q < x)$ limeszt. Vagyis: ha független kísérleteket végzünk, mindegyik q valószínűséggel sikerül, akkor ξ_q az első sikeres kísérlet bekövetkezésének sorszámát jelöli. Bizonyítsuk be, hogy az így kapott

$$x \mapsto \lim_{q \rightarrow 0} \mathbb{P}(q\xi_q < x)$$

függvény deriváltja sűrűségfüggvény (azaz nemnegatív és a valós egyenesen vett integrálja 1).

Megoldás. Mivel ξ_q csak egész értékeket vehet fel, és azt is tudjuk, hogy akkor legalább k az értéke, ha az első $k - 1$ kísérlet nem sikerült:

$$\mathbb{P}(q\xi_q < x) = 1 - \mathbb{P}\left(\xi_q \geq \left\lceil \frac{x}{q} \right\rceil\right) = 1 - (1 - q)^{\lceil x/q \rceil - 1}.$$

Itt

$$(1 - q)^{x/q - 1} \leq (1 - q)^{\lceil x/q \rceil - 1} \leq (1 - q)^{x/q}.$$

Ha $q \rightarrow 0$, akkor az alsó és a felső becslés is e^{-x} -hez tart.

Hasonlóképpen a Taylor-sorfejtés alapján

$$(1 - q)^{\lceil x/q \rceil - 1} \geq \exp(-q(\lceil x/q \rceil - 1)) \rightarrow e^{-x} \quad (q \rightarrow 0).$$

A negatív x -re nyilvánvaló, hogy a $\mathbb{P}(q\xi_q < x)$ valószínűség limesze 0.

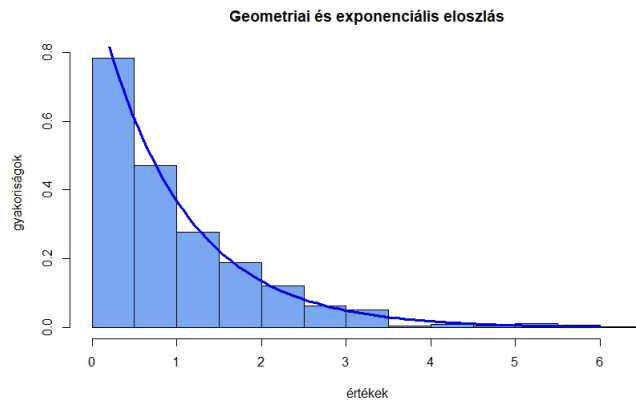
Tehát a limesz pozitív x -re $1 - e^{-x}$ és különben 0.

$\lim_{q \rightarrow 0} \mathbb{P}(q\xi_q < x) = \mathbb{P}(Y < x) = \int_0^x e^{-s} ds$, ahol Y exponenciális eloszlású.

Ennek deriváltja 0, ha $x < 0$, és e^{-x} különben. Ezért azt kell ellenőriznünk, hogy a függvény majdnem mindenütt nemnegatív, és

$$\int_0^\infty e^{-x} dx = 1.$$

Mivel az integrandus primitív függvénye $-e^{-x}$, ez könnyen látható.



4. ábra. A $q\xi_q$ hisztogramja $q = 0,001$ esetén 1000 elemű mintából és az e^{-x} függvény

2. Legyen $n \geq 1$ esetén $\xi_n \sim \text{NegBin}(n, \frac{1}{2})$ és $x \in \mathbb{R}$. Számoljuk ki a $\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\xi_n - 2n}{\sqrt{2n}} < x\right)$ limeszt. Vagyis: ha független kísérleteket végzünk, mindegyik $1/2$ valószínűséggel sikerül, akkor ξ_n az n . sikeres kísérlet bekövetkezésének sorszámát jelöli.

Megoldás. Legyen Y_j az, hogy a $j - 1$. sikeres kísérlet után hányat kell várni a j -re. Ekkor $\xi_n = \sum_{j=1}^n Y_j$.

Ha $Y_1, Y_2, \dots$ függetlenek, geometriai eloszlásúak $1/2$ paraméterrel, akkor ξ_n eloszlása megegyezik $\sum_{j=1}^n Y_j$ eloszlásával, hiszen Y_j jelentheti azt, hogy a j . sikeres kísérletre mennyit kell várni a $j - 1$. után. Azt is tudjuk, hogy $\mathbb{E}(Y_j) = 2$ és $D(Y_j) = \sqrt{(1-p)/p^2} = \sqrt{2}$. Ezért a kérdéses valószínűség így írható fel:

$$\mathbb{P}\left(\frac{\sum_{j=1}^n Y_j - n\mathbb{E}(Y_1)}{D(Y_1)\sqrt{n}} < x\right).$$

Mivel $Y_1, Y_2, \dots$ függetlenek, azonos eloszlásúak, véges szórásúak, teljesítik a centrális határeloszlástétel feltételeit. A tétel alapján ennek a valószínűségnek a limesze $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx$.