

▼ 13. Nagy számok törvénye és a centrális határeloszlás-tétel. Alkalmazások és szemléltetések.

Valószínűségszámítás és statisztika számítógép algebrai támogatással	Készült a PTE PMMIK-n a TÁMOP -4.1.2.A/1-11/1-2011-0098 támogatásával
Klincsik Mihály	

▼ 13.1. Bevezetés

Ha egy X véletlen változó eloszlása ismeretlen, akkor is tudunk becslést adni az $\{|X - a| > \epsilon\}$ alakú esemény valószínűségére, amennyiben ismert a várható érték és esetleg a szórása. Ezek a becslések a **Markov-egyenlőtlenség** illetve a **Csebisev-egyenlőtlenség**. A kapott egyenlőtlenségek hozzájárulnak olyan nevezetes eredmények megalapozásához, amelyek az esemény **relatív gyakoriságáról** szólnak, miszerint **konvergálnak** az esemény valószínűségéhez. A konvergencia alatt **valószínűségi mértékben** való konvergenciát értünk és nem sorozat konvergenciát. A **nagy számok törvényét** alkalmazzuk a **Monte Carlo-módszerrel** történő integrálás során és a **minőségbiztosításban** a mintavételezés becslésére!

A **centrális határeloszlás-tétel** olyan valószínűségi változók határértékével összefüggő eredmény, amely szerint ha n független és azonos eloszlású valószínűségi változónak létezik a közös várható értéke és szórása, akkor az összeg **standardizáltjának eloszlás függvénye** konvergál a standard $N(0, 1)$ normál eloszlás **eloszlásfüggvényéhez**, n tart végtelen esetén. A centrális határeloszlás tétel az összeg változó eseményeinek becslésére hatékonyan alkalmazható.

Célunk az alábbi fogalmak és összefüggések megismerése, megértése	
Markov-egyenlőtlenség bizonyítása és alkalmazása	A nagy számok törvényének és a centrális határeloszlás-tétel kapcsolata
Csebisev-egyenlőtlenség bizonyítása és alkalmazása	Valószínűségi változók konvergenciája valószínűségi mértékben és eloszlásban
Nagy számok (gyenge) törvényének Csebisev-alakja és alkalmazása	Monte Carlo módszer alkalmazása integrálok közelítő számítására
Nagy számok (gyenge) törvényének Bernoulli-alakja és alkalmazása	Egyenletek gyökének közelítő meghatározására a Maple <code>fsolve(egyenlet, x = a .. b)</code> eljárását használhatjuk.
Centrális határeloszlás tétel és alkalmazásai	

▼ 13.2. Markov egyenlőtlenség

13.2.1. TÉTEL. Markov - egyenlőtlenség (A.A. MARKOV, 1856-1922, orosz matematikus)

Ha az $Y \geq 0$ **nem** – **negatív** valószínűségi változónak létezik az $E(Y)$ várható értéke és $a > 0$ tetszőleges, akkor

$$P(Y \geq a) \leq \frac{E(Y)}{a}$$

egyenlőtlenség teljesül.

BIZONYÍTÁS

Válasszuk külön a diszkrét és folytonos eset bizonyítását!

(a) Folytonos eset

Legyen az Y **folytonos valószínűségi változó** sűrűség függvénye $y = f(x)$. Az $Y \geq 0$ feltétel miatt $f(x) = 0$, ha $x \leq 0$ negatív! A várható érték számításából kiindulva csökkenteni fogjuk az integrál értékét

$$E(Y) = \int_0^{\infty} x f(x) dx = \int_0^a x f(x) dx + \int_a^{\infty} x f(x) dx \geq \int_a^{\infty} x f(x) dx \geq \int_a^{\infty} a f(x) dx = a P(Y \geq a)$$

Az egyenlőtlenséget osztva $0 < a$ -val kapjuk a tétel állítását!

b) Legyen Y **diszkrét valószínűségi változó**, melynek $y_k \geq 0$ értékei nem-negatívak az $Y \geq 0$ feltétel miatt. Ekkor ugyanúgy, mint az előzőekben a várható érték számításából kiindulva csökkentjük az összeget

$$E(Y) = \sum_k y_k p_k = \sum_{y_k < a} y_k p_k + \sum_{y_k \geq a} y_k p_k \geq \sum_{y_k \geq a} y_k p_k \geq \sum_{y_k \geq a} a p_k = a \cdot \sum_{y_k \geq a} p_k = a \cdot P(Y \geq a)$$

Az egyenlőtlenséget $a > 0$ -val osztva kapjuk a tétel állítását!

Q.e.d.

Megjegyzések

1. A tétel használható eredményt akkor ad, ha $\frac{E(Y)}{a} < 1$ teljesül.
2. A tétel állítása szerint, ha 'a' értéke nagy, akkor a valószínűségi változó 'a'-nál nagyobb értéket "kis" valószínűséggel vesz fel.

13.2.2. PÉLDA.

A PTE PMMIK számítógépes laborjaiban a számítógépek nagy igénybevétele következtében a gépek átlagos élettartama (javítástól javításig) 2 hónap. A számítógépek legfeljebb hány százaléka bírja ki 3 hónapig javítás nélkül?

Megoldás

Legyen T egy számítógép élettartama hónapokban. A Markov-egyenlőtlenség $T \geq 0$ feltétele teljesül, mert negatív élettartam nincs. A várható érték $E(T) = 2$ hónap adott. A Markov-egyenlőtlenséget közvetlenül alkalmazhatjuk $a = 3$ hónap értékkel

$$P(T > 3) \leq \frac{E(Y)}{a} = \frac{2}{3} \approx 0.6666$$

Ez azt jelenti, hogy maximum 66,7% az esély arra, hogy **egy számítógép** élettartama 3 hónapnál hosszabb. Így az összes számítógép legfeljebb 66.7%-át nem kell javítani 3 hónapon belül.

▼ 13.3. Csebisev-egyenlőtlenség

Ha a Markov egyenlőtlenségben elvégezzük az $Y = (X - E(X))^2$ helyettesítést, akkor olyan becsléshez jutunk, amely

- (i) egyrészt tetszőleges előjelű valószínűségi változóra alkalmazható lesz
- (ii) másrészt X szórásának létezni kell!

13.3.1. TÉTEL. CSEBISEV –EGYENLŐTLENSÉG (P.L. CSEBISEV, 1821 - 1894, orosz matematikus)

Legyen adott az X valószínűségi változó, melynek $\mu = E(X)$ a várható értéke, valamint a

$\sigma^2 = \text{Var}(X)$ varianciája véges. Ha ε tetszőleges pozitív szám, akkor fenáll a

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

egyenlőtlenség.

Amennyiben ε értékét a σ standard szórás k -szorosának választjuk, azaz $\varepsilon = k \sigma$, akkor a Csebisev-egyenlőtlenség alakja

$$P(|X - \mu| \geq k \cdot \sigma) \leq \frac{1}{k^2}$$

BIZONYÍTÁS.

Írjuk fel a Markov-egyenlőtlenséget az $Y = (X - \mu)^2 \geq 0$ valószínűségi változóval valamint az $a = \varepsilon^2$ konstanssal!

$$P((X - \mu)^2 \geq \varepsilon^2) \leq \frac{E((X - \mu)^2)}{\varepsilon^2}$$

Mivel jobb oldalon az $E((X - \mu)^2) = \sigma^2$ az X valószínűségi változó szórás négyzete és az $\varepsilon^2 \leq (X - \mu)^2$ egyenlőtlenség ekvivalens az $\varepsilon \leq |X - \mu|$ egyenlőtlenséggel ($\varepsilon > 0$), ezért

$$P(\varepsilon \leq |X - \mu|) \leq \frac{\sigma^2}{\varepsilon^2},$$

amely a bizonyítandó Csebisev-egyenlőtlenség.

A Csebisev egyenlőtlenség második alakját úgy kapjuk, ha az elsőben alkalmazzuk az

$$\varepsilon = k \sigma$$

helyettesítést!

Q.e.d.

13.3.2. PÉLDA. Szimmetrikus becslés a várható érték körül

Szalagokat gyártó cég által készített 3 méter hosszú szalag hossza változó a gyártásból és a mérésből eredő pontatlanság következtében. A 3 méteres szalag pontatlanságának szórása nagyszámú statisztikára alapozva 0.03 [m] = 3 [cm]. Mekkora valószínűséggel esik egy 3 méteres darab hossza a [2.94, 3.06] intervallumba?

MEGOLDÁS

Legyen X a legyártott szalag hossza, melynek várható értéke 3 méter $E(X)=3$. A feladat alapján X szórása $\sigma(X)=0.03$.

Írjuk fel a Csebisev-egyenlőtlenség második alakját!

A feltételek alapján a $k = 2$ esettel van dolgunk, mert a 3 várható értéktől az eltérések lefelé és fölfelé is $0.06 = 3.06 - 3 = 3 - 2.94$ értékkel térnek el és ez a szórás kétszerese

$$0.06 = 2 \cdot 0.03 = 2 \cdot \sigma.$$

A Csebisev-egyenlőtlenség alapján

$$P(|X - 3| \geq 2 \cdot 0.03) \leq \frac{1}{2^2} = 0.25$$

teljesül. Ahonnan a kiegészítő esemény valószínűségére az alábbi adódik

$$P(|X - 3| < 2 \cdot 0.03) = 1 - P(|X - 3| \geq 2 \cdot 0.03) \geq 1 - 0.25 = 0.75.$$

Tehát legalább 75%-os biztonsággal állíthatjuk, hogy szalag hossza [2.94, 3.06] közé esik!

▼ 13.4. Nagy számok (gyenge) törvényének Csebisev-alakja. Monte Carlo

integrálás.

A Csebisev- egyenlőtlenség alapján igazolni tudjuk a nagy számok törvényének Csebisev-alakját.

13.4.1. TÉTEL. NAGY SZÁMOK (GYENGE) TÖRVÉNYÉNEK CSEBISEV-ALAKJA.

Legyenek $X_1, X_2, \dots, X_n, \dots$ azonos eloszlású, független valószínűségi változók (diszkrét vagy folytonos) végtelen sorozata és legyen $\mu = E(X_i)$ a közös várható érték és $\sigma^2 = Var(X_i)$ a variancia.

Ha

$$S_n = X_1 + X_2 + \dots + X_n$$

a változók összege, akkor

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}$$

ahonnan

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

BIZONYÍTÁS

Az $\bar{X}_n = \frac{S_n}{n}$ átlag valószínűségi változó várható értéke ugyancsak μ , mert az E várható érték lineáris és mindegyik i -re $E(X_i) = \mu$

$$E\left(\frac{S_n}{n}\right) = \frac{1}{n} \cdot E(X_1 + X_2 + \dots + X_n) = \frac{n \cdot \mu}{n} = \mu.$$

Mivel $X_1, X_2, \dots, X_n$ valószínűségi változók kölcsönösen függetlenek és mindegyik szórása σ , ezért

$$\sigma_{S_n} = \sqrt{Var(S_n)} = \sqrt{\sigma^2 + \sigma^2 + \dots + \sigma^2} = \sqrt{n} \sigma.$$

Ahonann a számtani átlag szórására $\sigma_{\bar{X}_n} = \frac{\sqrt{n} \sigma}{n} = \frac{\sigma}{\sqrt{n}}$ adódik. Alkalmazzuk a

Csebisev-egyenlőtlenséget X helyett $\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ valószínűségi változóra és kapjuk a tételben szereplő egyenlőtlenséget

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{\sigma^2}{n \varepsilon^2}.$$

A határérték reláció következménye a $\lim_{n \rightarrow \infty} \frac{\sigma^2}{n \varepsilon^2} = 0$ egyenlőségnek. Q.e.d. ♣

13.4.2. PÉLDA. SZERENCSE JÁTÉKOS CSÓDJE

A rulett játékban a veszítés valószínűsége mindig egy kicsit nagyobb, mint a nyerésé. Az egyszerűség kedvéért tegyük fel, hogy 1 euro nyerési esélye $p = \frac{49}{100}$ és a veszítés esélye

$q = 1 - p = \frac{51}{100}$. Az egymást követő nyerések és veszítések modellezhetők az $X_1, X_2, \dots$ független véletlen változók olyan sorozatával, melyben

$$P(X_i = 1) = \frac{49}{100}, P(X_i = -1) = \frac{51}{100}.$$

Az X_i várható értéke

$$E(X_i) = 1 \cdot \frac{49}{100} + (-1) \cdot \frac{51}{100} = \frac{49 - 51}{100} = -\frac{2}{100} = -\frac{1}{50}.$$

A játékos tőkéje n játék után az

$$S_n = X_1 + X_2 + \dots + X_n$$

összeg. A nagyszámok törvénye azt mondja ki, hogy

$$P\left(\left|\frac{S_n}{n} + \frac{1}{50}\right| > \varepsilon\right) \rightarrow 0,$$

ha $n \rightarrow \infty$. Ezért $\frac{S_n}{n}$ nagy valószínűséggel közel kerül $-\frac{1}{50}$ értékhez, ahonnan S_n közel kerül

$\left(-\frac{n}{50}\right)$ -hez. Tehát a játékos **szükségszerűen veszít**.

Választ kaphatunk a nagy számok törvénye kapcsán arra is, hogyan alakul sok játék után a játékos $\{S_n > 0\}$ nyerési eseményének esélye. Ugyanis n játék után a játékos nyerési esélyét becsülhetjük a

$$P(S_n > 0) = P\left(\frac{S_n}{n} > 0\right) \leq P\left(\left|\frac{S_n}{n} + \frac{1}{50}\right| > \frac{1}{50}\right),$$

valószínűséggel, amely a nagy számok törvénye szerint tart 0-hoz, ha $n \rightarrow \infty$. Tehát a játékos nyerési esélye nullához tart, ha a játékok száma növekszik.

13.4.3. PÉLDA. MONTE CARLO (MC) INTEGRÁLÁS

Közelítsük az $\int_a^b f(x) dx$ integrál értékét véletlen értékek generálásával olyan esetben, amikor a primitív függvény létezik, de nem elemi függvény. Ilyen esetekben numerikus integrálással tudjuk csak az integrál közelítő értékét számolni.

A közelítés első lépéseként vegyük az $X_1, X_2, \dots, X_n, \dots$ $[a, b]$ intervallumon egyenletes eloszlású véletlen változók

$$f(X_1), f(X_2), f(X_3), \dots, f(X_n), \dots$$

$y = f(x)$ függvény melletti képeinek átlagát

$$\langle f \rangle_N = \frac{1}{N} \sum_{k=1}^N f(X_k).$$

13.4.4. TÉTEL. Monte Carlo integrálközelítő összegek várható értéke az integrál

Mutassuk meg, hogy az

$$\langle I \rangle_N = (b - a) \cdot \langle f \rangle_N = \frac{b - a}{N} \sum_{k=1}^N f(X_k)$$

valószínűségi változó várható értéke az

$$E(\langle I \rangle_N) = \int_a^b f(x) dx$$

integrál.

Bizonyítás

A várható érték lineáris tulajdonsága miatt

$$E(\langle I \rangle_N) = E\left(\frac{b-a}{N} \sum_{k=1}^N f(X_k)\right) = \frac{b-a}{N} \cdot \sum_{k=1}^N E(f(X_k)) = \frac{b-a}{N} \cdot \sum_{k=1}^N \int_a^b f(x) \cdot p(x) dx$$

ahol $p(x) = \frac{1}{b-a}$ az $[a, b]$ intervallumon egyenletes eloszlású $Uniform(a, b)$ sűrűség függvénye. Így

$$\frac{b-a}{N} \cdot \sum_{k=1}^N \int_a^b f(x) \cdot p(x) dx = \frac{b-a}{N} \cdot \frac{1}{b-a} \cdot \sum_{k=1}^N \int_a^b f(x) dx = \frac{1}{N} \sum_{k=1}^N \int_a^b f(x) dx = \int_a^b f(x) dx$$

mert az integrál értékét adjuk össze önmagával N-szer.

A nagyszámok törvénye szerint tudjuk, hogy az

$$\langle I \rangle_N = (b-a) \cdot \langle f \rangle_N$$

valószínűségi mértékben konvergál a várható értékhez, mely az $\int_a^b f(x) dx$ integrál. Q.e.d. ♣

A hiba becsléshez használjuk a statisztikai szórás

$$Standard\ hiba = (b-a) \sqrt{\frac{\langle f^2 \rangle_N - \langle f \rangle_N^2}{N}}$$

képletét, ahol $\langle f^2 \rangle_N = \frac{1}{N} \sum_{k=1}^N f^2(X_k)$ a függvényértékek négyzetének átlaga.

Ekkor az **integrál közelítő értéke** a Monte Carlo módszer alapján

$$\langle I \rangle_N \pm Standard\ hiba = (b-a) \cdot \langle f \rangle_N \pm (b-a) \sqrt{\frac{\langle f^2 \rangle_N - \langle f \rangle_N^2}{N}}.$$

Monte Carlo módszerrel határozzuk meg az

$$\int_0^2 \frac{1}{\sqrt{1 - \frac{1}{2} \sin(\pi x)^2}} dx$$

integrál értékét.

A Maple megadja az integrál értékét a beépített **elliptikus függvény** segítségével, melynek kerekített értékét elfogadjuk az integrál **pontos értékének**.

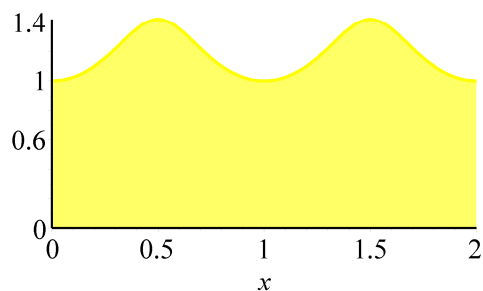
A Monte Carlo integráláshoz 1000 véletlen számot generálunk a $[0,2]$ intervallumban egyenletes

Maple kód a grafikon rajzolásához és az integrál pontos értékének kiszámításához.

```
> restart; a, b := 0, 2;
f := x -> 1 / (sqrt(1 - sin(Pi * x)^2 / 2)) :
plot(f(x), x = a..b, filled = true, color = yellow);
pontos := evalf(int(f(x), x = a..b));
```

$a, b := 0, 2$

eloszlással! Kiszámoljuk az integrál közelítő értékét és a standard hibát. Felrajzoljuk a közelítő érték megfelelő környezetét, melynek sugara a standard hiba és ebbe berajzoljuk pontos értéket és a közelítő értéket!

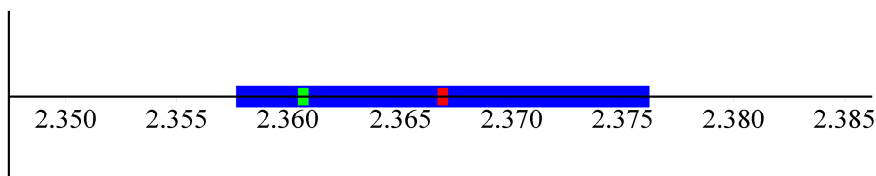


pontos := 2.360681197

(1.4.1)

Maple kód az integrál értékének kiszámításához Monte Carlo módszerrel

```
> with(Statistics) : with(StringTools) : N := 1000 :
> randomize( ) : randomize( ) :
  V := Sample( Uniform(a, b), N ) :
  T := evalf( 1/N · add( f( Vi ), i = 1..N ) ) :
  közelítés := (b - a) · T :
  S2 := evalf( 1/N · add( f( Vi )2, i = 1..N ) ) :
> hiba := (b - a) · √( (S2 - T2) / N ) ; Pontos érték = pontos
> Join( [ "Az integrál értékének becslése", convert( közelítés, string ), "±", convert( hiba, string ) ],
  "" )
> alsó := közelítés - hiba : felső := közelítés + hiba :
> plot( [ [ [ alsó, 0 ], [ felső, 0 ] ], [ [ pontos, -0.01 ], [ pontos, 0.01 ] ], [ [ közelítés, -0.01 ], [ közelítés,
  0.01 ] ] ], thickness = [ 10, 5, 5 ], color = [ blue, green, red ], view = [ alsó - 0.01, felső + 0.01, -0.1
  ..0.1 ], tickmarks = [ 10, 0 ] )
  hiba := 0.009275733718
  Pontos érték = 2.3606811970
  "Az integrál értékének becslése 2.366939074 ± .9275733718e-2"
```



Látható, hogy a megadott intervallumba esik a pontos érték! Kísérletezzünk más N értékkel!

▼ 13.5. Nagy számok (gyenge) törvényének Bernoulli - alakja

13.5.1. TÉTEL. (Jacob D. Bernoulli, 1659 – 1705) NAGY SZÁMOK TÖRVÉNYE.

Egy $p > 0$ valószínűségű A eseményre végezzünk n független kísérletet és jelölje $k_n(A)$ az n

kísérletből az A esemény bekövetkezési számát. Ekkor az

$$X_{\text{átlag}} = \frac{k_n(A)}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

relatív gyakoriság valószínűségi változóra az alábbi egyenlőtlenség teljesül

$$P\left(\left|\frac{k_n(A)}{n} - p\right| \geq \varepsilon\right) \leq \frac{p \cdot (1-p)}{n \cdot \varepsilon^2} \leq \frac{1}{4 \cdot n \cdot \varepsilon^2}.$$

A képletekben X_i jelöli az i-edik kísérlet **indikátor** vagy **Bernoulli**- változóját, amely

$$X_i = \begin{cases} 1, & \text{ha } A \text{ bekövetkezik} \\ 0, & \text{ha } A \text{ nem következik be} \end{cases}.$$

BIZONYÍTÁS

Az X_k változók független Bernoulli-típusú véletlen változók azonos p -paraméterrel. Ezért $k_n(A)$ összeg változó **(n,p)** paraméterű binomiális - eloszlású valószínűségi változó. Így **várható értéke** $E(k_n(A)) = n p$ és **szórása** $\sigma(k_n(A)) = \sqrt{n p (1-p)}$ (lásd a nevezetes diszkrét valószínűségi változóknál). A várható érték és szórás tulajdonságait felhasználva kapjuk a következőket

$$E\left(\frac{k_n(A)}{n}\right) = \frac{E(k_n(A))}{n} = \frac{n p}{n} = p \text{ és}$$

$$\sigma\left(\frac{k_n(A)}{n}\right) = \frac{\sigma(k_n(A))}{n} = \frac{\sqrt{n p (1-p)}}{n} = \sqrt{\frac{p (1-p)}{n}}$$

A Csebisev – egyenlőtlenségben szereplő X helyébe helyettesítsük be a $\frac{k_n(A)}{n}$ változót. Ekkor a jobb oldali felső korlát a következő lesz

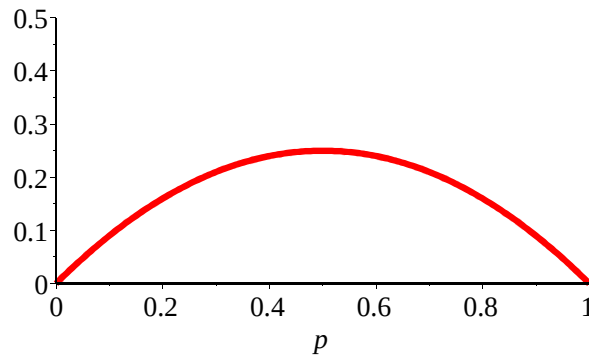
$$\frac{\sigma^2\left(\frac{k_n(A)}{n}\right)}{\varepsilon^2} = \frac{p \cdot (1-p)}{n \varepsilon^2}.$$

Ezzel megkaptuk a tétel állításában szereplő egyenlőtlenség első részét. A második egyenlőtlenség következik a $p \cdot (1-p) \leq \frac{1}{4}$ egyenlőtlenségből (lásd az ábrát). Q.e.d.

> restart :

plot($p \cdot (1-p)$, $p = 0..1$, color = red, thickness = 3, scaling = constrained, view = [0..1, 0..0.5], title = $A \cdot p \cdot (1-p)$ függvény)

A $p*(1-p)$ függvény



A tételben szereplő esemény komplementerére teljesül a

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{k_n(A)}{n} - p\right| < \varepsilon\right) = 1$$

reláció.

13.5.2. DEFINÍCIÓ. KONVERGENCIA VALÓSZÍNŰSÉGI MÉRTÉKBEN

Tekintsük az $X_1, X_2, \dots$ valószínűségi változók végtelen sorozatát. Azt mondjuk, hogy az X_n sorozat konvergál az Y valószínűségi változóhoz **valószínűségi mértékben**, ha minden $\varepsilon > 0$ esetén

$$\lim_{n \rightarrow \infty} P(|X_n - Y| \geq \varepsilon) = 0$$

vagy ezzel ekvivalens módon

$$\lim_{n \rightarrow \infty} P(|X_n - Y| < \varepsilon) = 1.$$

Jelölése $\lim_{n \rightarrow \infty} X_n \stackrel{Pr}{=} Y$ vagy $X_n \xrightarrow{Pr} Y$.

A fenti definíció alapján a nagy számok törvénye azt mondja ki, hogy $\frac{k_n(A)}{n} \xrightarrow{Pr} p$, vagyis a relatív gyakoriságok **valószínűségi mértékben konvergálnak** az esemény p valószínűségéhez! Minden korábbi szemléltetés, amikor számítógépes szimulációval relatív gyakoriságokat rajzoltunk és megmutattuk, hogy nagy n esetén közel kerülnek a valószínűséghez, akkor a nagyszámok törvényének Bernoulli-alakja működött a háttérben.

13.5.3. PÉLDA. MINŐSÉGBIZTOSÍTÁS A MINTA MINIMÁLIS ELEMSZÁMÁRA

Gyártó soron készült termékek átvételkor a gyártó azt állítja, hogy a termékek $p = 0.05$ része hibás.

Minimálisan hány elemű mintát kell kivenni, ha legalább 90%-os biztonsággal akarunk meggyőződni arról, hogy a megadott p aránytól való eltérés legfeljebb 0.02?

MEGOLDÁS

(a) eset. Ha felhasználjuk, hogy a $p = 0.05$ selejtarány ismert.

Jelölje A azt az eseményt, hogy hibás alkatrészt veszünk ki. Az n független kísérletből jelölje $k_n(A)$ az A esemény bekövetkezési számát, vagyis gyakoriságát! A feladat a következő egyenlőtlenség megoldását kérdezi n -re

$$P\left(\left|\frac{k_n(A)}{n} - 0.05\right| < 0.02\right) \geq 0.9$$

A nagyszámok Bernoulli-alakjában az $\left\{\left|\frac{k_n(A)}{n} - p\right| > \varepsilon\right\}$ esemény szerepel. A kérdés fordított irányú egyenlőtlenségre vonatkozik. Így első lépésben alkalmazzuk a tagadás valószínűségéről szóló tételt, majd a nagyszámok törvényének Bernoulli-alakját

$$\begin{aligned} P\left(\left|\frac{k_n(A)}{n} - 0.05\right| < 0.02\right) &= 1 - P\left(\left|\frac{k_n(A)}{n} - 0.05\right| \geq 0.02\right) \geq \\ &\geq 1 - \frac{p \cdot (1-p)}{n \cdot \varepsilon^2} = 1 - \frac{0.05 \cdot 0.95}{n \cdot 0.02^2} \geq 0.9. \end{aligned}$$

Maple-ben felírjuk az utóbbi egyenlőtlenséget, hogy ne kelljen kézzel megoldani n -re!

> restart :

> p := 0.05;

$$e := 0.9 \leq 1 - \frac{p(1-p)}{n \cdot 0.02^2}$$

> solve(e, n)

$$p := 0.05$$

$$e := 0. \leq 0.1 - \frac{118.7500000}{n}$$

$$\text{RealRange}(-\infty, \text{Open}(0.)), \text{RealRange}(1187.500000, \infty)$$

(1.5.1)

Tehát, ha $n > 1187.5$, akkor teljesülnek a feltételek.

(b) eset. Ha p értéke ismeretlen lenne.

Ha **nem ismernénk** a $p = 0.05$ **valószínűséget**, akkor a tétel utolsó egyenlőtlenségét használni tudjuk a becsléshez. Ebben az esetben még nagyobb értéket kapunk n -re! Ugyanis az alábbi egyenlőtlenség megoldásából adódik n értéke.

$$> 1 - \frac{1}{4 \cdot n \cdot 0.02^2} \geq 0.9; \text{solve}(\%, n);$$

$$0. \leq 0.1 - \frac{625.0000000}{n}$$

$$\text{RealRange}(-\infty, \text{Open}(0.)), \text{RealRange}(6250., \infty)$$

(1.5.2)

Tehát p ismerete nélkül legalább 6250 elemű mintát kellene kivenni, szemben az (a) esettel, amikor legalább 1188 terméket kell kiválasztani.

(c) eset. Ha ismerjük a selejt eloszlását.

A nagy elemszámú minta kiválasztása és átvizsgálása **csökkenthető**, ha ismerjük a selejteket eloszlását! Amennyiben feltesszük, hogy a sokaság elemszáma **jóval nagyobb**, mint amennyit kiveszünk belőle, akkor a mintavétel tekinthető **viSSzatevéseSnek**. Ekkor a $\text{Binomial}(n, p)$ eloszlással dolgozhatunk. A két paraméter közül ismerjük $p = 0.05$ értékét, viszont n értékét a megadott feltételből kell számolni. Tehát $k_n(A) = B_{n,p}$ binomiális eloszlású változó és az eseményre vonatkozó egyenlőtlenséget rendezzük a $k_n(A)$ gyakoriságra

$$\left|\frac{k_n(A)}{n} - p\right| < \varepsilon$$

$$\begin{aligned}
 -\varepsilon &< \frac{k_n(A)}{n} - p < \varepsilon \\
 p - \varepsilon &< \frac{k_n(A)}{n} < p + \varepsilon \\
 n \cdot (p - \varepsilon) &< k_n(A) < n \cdot (p + \varepsilon).
 \end{aligned}$$

Így a keresett feltételt

$$P\left(\left|\frac{k_n(A)}{n} - p\right| < \varepsilon\right) = \sum_{n \cdot (p - \varepsilon) < k < n \cdot (p + \varepsilon)} \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \geq 0.9$$

képlettel adhatjuk meg, ahol $p = 0.05$ és $\varepsilon = 0.02$ adottak. Ebből n értékét kell számolni. Ez egy nagyon bonyolult nem-lineáris egyenlőtlenség. A baloldal kiszámolható az

$$F(x) = P(B_{n,p} < x) = CDF(B_{n,p}, x)$$

eloszlásfüggvény segítségével a

$$F(n \cdot (p + \varepsilon)) - F(n \cdot (p - \varepsilon))$$

képletet.

A megoldás menete tehát az, hogy a Maple beépített $CDF(B_{n,p})$ eloszlás függvénye segítségével kiszámoljuk n különböző értékeire a fenti eltéréseket, majd felrajzoljuk az eltérést n függvényében.

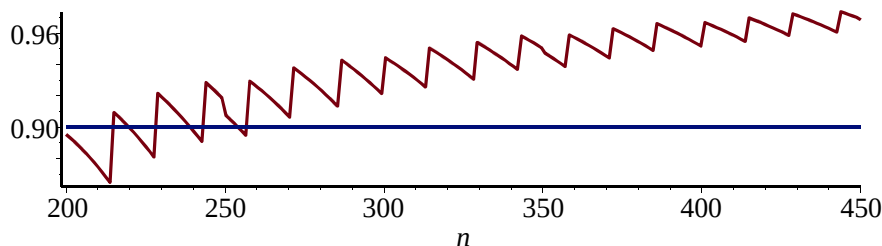
> restart : with(Statistics) :

> n := 'n': X := RandomVariable(Binomial(n, 0.05)) :

> FX := unapply(CDF(X, 0.07·n) - CDF(X, 0.02·n), n)

> plot([FX(n), 0.9], n = 200..450)

$FX := n \rightarrow -\frac{1}{20.^n} (0.05263157895 \text{ binomial}(n, \text{floor}(0.07 n))$
 $+ 1.) \text{ hypergeom}([1.000000000, -1. n + 1. \text{floor}(0.07 n) + 1.],$
 $[1.000000000 \text{ floor}(0.07000000000 n) + 2.000000000], -0.05263157895)$
 $19.^{n - 1. \text{floor}(0.07 n)}) + \frac{1}{20.^n} (0.05263157895 \text{ binomial}(n, \text{floor}(0.02 n))$
 $+ 1.) \text{ hypergeom}([1.000000000, -1. n + 1. \text{floor}(0.02 n) + 1.], [2.000000000$
 $+ 1.000000000 \text{ floor}(0.02000000000 n)], -0.05263157895) 19.^{n - 1. \text{floor}(0.02 n)})$



Valószínű, ilyen **cikk-cakk függvényt** kedves olvasónk még valószínűséggel kapcsolatban nem látott. Most tudjuk igazán értékelni a Bernoulli-tételben szereplő egyenlőtlenséget és a nagyszámok törvényét! Hol ugrik a cikk-cakk függvény hirtelen felfelé és meddig megy lefelé? Ennek

kiderítését az olvasóra bízjuk. Minden esetre van egy növekvő trend a grafikonban és $n = 200 \dots 260$ között a 0.90 valószínűségi szint alatt még vannak pontok, majd onnan már nem csökken akkorát, hogy a 0.90 szint alá csökkenne vissza.

Az ábra alapján azt mondhatjuk, hogy kb. elegendő kiválasztani $n = 260$ elemű mintát és megvizsgálni. Ha ebben a mintában kevesebb a selejt, mint a 260 elem 5%-a, ami 13 darab, akkor vegyük át a teljes szállítmányt. Egyébként utasítsuk vissza a teljes rakományt. Ekkor döntésünk 90%-ban garantálja, hogy a selejt 7% alatt lesz az egészben. Persze az esetek 10%-ában hibázhatunk. A nagy számok törvénye ilyen módon garanciát ad a minőségre, hogy a vevő is és az eladó is az elvárásoknak megfelelően elégedett legyen a piacon!

▼ 13.6. Nagy számok törvényének igazolása a szabályos pénzérme kísérletére

A nagy számok törvényének **történetéhez** tartozik, hogy először Jacob Bernoulli, Ars Conjectandi (A sejtés művészete) című könyvében jelent meg 1713-ban, a szerző halála után 8 évvel, amelyet unokaöccse Nikolaus Bernoulli jelentetett meg. Bernoulli a binomiális eloszlás ügyes matematikai becslésével bizonyította a tételt, amely akkoriban nagy teljesítmény volt, hiszen a Csebisev-egyenlőtlenség még nem volt ismert. Ezután a tétel, mint Bernoulli-tétel néven vált ismertté. Poisson bizonyított egy általánosabb tételt és ő nevezte el a tételt először nagy számok törvényének. Alekszander Hincsin orosz matematikus 1929-ben bizonyította a nagy számok gyenge törvényének általános alakját. Sok más verzió is ismert, amelyekben nem szükséges feltenni olyan szigorú feltételeket, mint az eloszlások egyenlőségét vagy a varianciák végeességét.

Érdeemes megismerni a nagyszámok törvényének **bizonyításával** az egyik legegyszerűbb esetben, amikor **szabályos pénzérmét dobálunk**. Ehhez meg kell mutatni, hogy a szabályos pénzérme ismételt és független feldobásakor a fej dobások aránya durván $\frac{1}{2}$. Vajon milyen

matematikai analízisbeli eszközöket kell felvonultatni minimálisan ahhoz, hogy ezt az egyszerű helyzetet igazoljuk. A binomiális eloszlásnál láttuk, hogy az

$$A_{n,k} = \{n \text{ dobásból } k \text{ fejet kapunk}\}$$

esemény valószínűsége $p = 1 - p = \frac{1}{2}$ szabályos érme esetén

$$P(A_{k,n}) = \binom{n}{k} \cdot \left(\frac{1}{2}\right)^n.$$

Ha k jelöli a fej dobások kedvező eseteinek számát, akkor a relatív gyakoriság a $\frac{k}{n}$ arány. A

relatív gyakoriság értéke akkor esik az $\frac{1}{2}$ érték $\varepsilon > 0$ sugarú környezetébe, ha

$$\frac{1}{2} - \varepsilon < \frac{k}{n} < \frac{1}{2} + \varepsilon$$

$$n \cdot \left(\frac{1}{2} - \varepsilon\right) < k < n \cdot \left(\frac{1}{2} + \varepsilon\right)$$

teljesül. Rögzített n kísérletszám mellett a

$$B = \bigcup_{n \cdot \left(\frac{1}{2} - \varepsilon\right) < k < n \cdot \left(\frac{1}{2} + \varepsilon\right)} A_{k,n}$$

összeg esemény jelöli pontosan azt, hogy a fej dobások **relatív gyakorisága** beesik az $\left(\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon\right)$ intervallumba.

13.6.1. TÉTEL. NAGYSZÁMOK TÖRVÉNYE SZABÁLYOS ÉRME KÍSÉRLETÉRE

Tetszőleges $\varepsilon > 0$ esetén

$$P\left(\bigcup_{n\left(\frac{1}{2}-\varepsilon\right) < k < n\left(\frac{1}{2}+\varepsilon\right)} A_{k,n}\right) \rightarrow 1,$$

ha $n \rightarrow \infty$.

BIZONYÍTÁS

Elegendő bizonyítani a

$$(*) \quad P\left(\bigcup_{n\left(\frac{1}{2}+\varepsilon\right) < k \leq n} A_{k,n}\right) \rightarrow 0$$

és

$$(**) \quad P\left(\bigcup_{0 \leq k < n\left(\frac{1}{2}-\varepsilon\right)} A_{k,n}\right) \rightarrow 0$$

határértékeket, ha $n \rightarrow \infty$. Ugyanis $P\left(\bigcup_{k=0}^n A_{k,n}\right) = 1$ teljesül minden n értékre és így a (*) és (**) esetekből kimaradó k indexekre összegezve az eseményeket a valószínűség határértéke 1 lesz.

Most csak a (*) képletet igazoljuk és az olvasóra bízunk a (**) bizonyítását, amely hasonlóan megy. Használjuk a **binomiális eloszlás** képletét

$$P\left(\bigcup_{n\left(\frac{1}{2}+\varepsilon\right) < k \leq n} A_{k,n}\right) = \sum_{k > n\left(\frac{1}{2}+\varepsilon\right)} P(A_{k,n}) = \sum_{k > n\left(\frac{1}{2}+\varepsilon\right)} \binom{n}{k} \cdot 2^{-k}$$

Az alábbi **algebrai átalakítások** jellemzőek a valószínűségelmélet bizonyítása során. Legyen $\lambda > 0$ tetszőleges, ekkor

$$\begin{aligned} \sum_{k > n\left(\frac{1}{2}+\varepsilon\right)} \binom{n}{k} \cdot 2^{-k} &\leq \sum_{k > n\left(\frac{1}{2}+\varepsilon\right)} e^{\lambda\left(k-n\left(\frac{1}{2}+\varepsilon\right)\right)} \binom{n}{k} \cdot 2^{-k} = \\ &= e^{-\lambda \cdot n \cdot \varepsilon} \sum_{k > n\left(\frac{1}{2}+\varepsilon\right)} \binom{n}{k} \cdot \left(\frac{1}{2} \cdot e^{\frac{\lambda}{2}}\right)^k \cdot \left(\frac{1}{2} \cdot e^{-\frac{\lambda}{2}}\right)^{n-k} = e^{-\lambda \cdot n \cdot \varepsilon} \sum_{k=0}^n \binom{n}{k} \cdot \left(\frac{1}{2} \cdot e^{\frac{\lambda}{2}}\right)^k \cdot \left(\frac{1}{2} \cdot e^{-\frac{\lambda}{2}}\right)^{n-k} \end{aligned}$$

Használjuk a binomiális-tételt $a = \frac{1}{2} \cdot e^{\frac{\lambda}{2}}$ és $b = \frac{1}{2} \cdot e^{-\frac{\lambda}{2}}$ értékekre. Így az egyenlőtlenség jobb oldala

$$e^{-\lambda \cdot n \cdot \varepsilon} \cdot \left(\frac{1}{2} \cdot e^{\frac{\lambda}{2}} + \frac{1}{2} \cdot e^{-\frac{\lambda}{2}}\right)^n$$

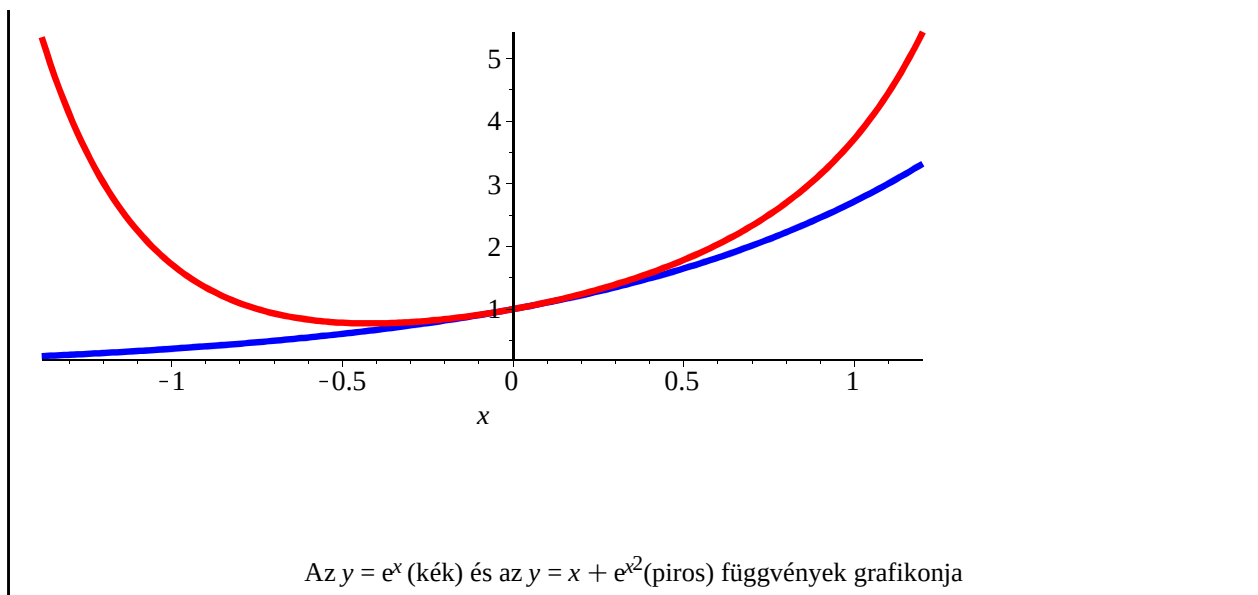
Alkalmazzuk az

$$e^x \leq x + e^{x^2}$$

egyenlőtlenséget, amely igaz pozitív és negatív x értékekre egyaránt.

Ezt szemlélteti az alábbi ábra.

> `plot([e^x, x + e^x^2], x=-1.38..1.2, color=[blue, red], thickness=3)`



Az egyenlőtlenséget használva kétszer

$$e^{-\lambda \cdot n \cdot \varepsilon} \cdot \left(\frac{1}{2} \cdot e^{\frac{\lambda}{2}} + \frac{1}{2} \cdot e^{-\frac{\lambda}{2}} \right)^n \leq e^{-\lambda \cdot n \cdot \varepsilon} \cdot \left(\frac{1}{2} \cdot \left(\frac{\lambda}{2} + e^{\left(\frac{\lambda}{2} \right)^2} \right) + \frac{1}{2} \cdot \left(-\frac{\lambda}{2} + e^{\left(-\frac{\lambda}{2} \right)^2} \right) \right)^n = e^{n \cdot \left(\frac{\lambda^2}{4} - \lambda \cdot \varepsilon \right)}.$$

Mivel tetszőleges $\lambda > 0$ esetén teljesülnek a fenti becslések, ezért választhatjuk λ értékét úgy, hogy a kitevőben szereplő $\frac{\lambda^2}{4} - \lambda \cdot \varepsilon$ másodfokú kifejezésnek λ -ban minimuma legyen. Ez a $\lambda = 2 \cdot \varepsilon$ mellett teljesül és így a keresett valószínűség felső becslése

$$P \left(\bigcup_{n \left(\frac{1}{2} + \varepsilon \right) < k \leq n} A_{k,n} \right) \leq e^{-n \cdot \varepsilon^2},$$

ahol a jobb oldal nullához tart, ha n tart végtelenhez. Ezzel a tételt igazoltuk. ♣

A bizonyításból kiderül, hogy egy általános tétel speciális esetének bizonyítása sok meglepetést tartogat.

▼ 13.7. Centrális határeloszlás-tétel szemléltetése és alkalmazásai

A nagy számok törvénye kimondja, hogy független azonos eloszlású valószínűségi változók átlaga "valószínűségi értelemben" közel kerül tetszőleges változó $E(X_i) = \mu$ várható értékéhez, ha n értéke nagy. A közelséget n -értékére pontosítani kellene. Ha például sorozatokat nézünk, akkor az $S_n = \sqrt{n}$ vagy $S_n = \ln n$ mind olyan sorozat, amelyekre $\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0$. Mégis óriási különbség van a két nullához konvergálási rendben. Azt a nullához konvergálást, amit a nagy számok törvényében kaptunk **pontosítja a centrális határeloszlás tétel**.

13.7.1. KÍSÉRLET. A binomiális eloszlás standardizáltja hogyan közelíti az $N(0, 1)$ standard normál eloszlás sűrűségfüggvényét?

Standardizáljuk a $B \sim \text{Binomiális}(n, p)$ binomiális eloszlást. Mivel B várható értéke

$$E(B) = n \cdot p$$

szórása

$$\sigma(B) = \sqrt{n \cdot p \cdot (1 - p)},$$

Maple kód

```
> restart : with(Statistics) : with(plots) :
> B := RandomVariable(Binomial(n, p)) :
  n := 100; p := 0.3;
> X := [seq( (k - n*p) / sqrt(n*p*(1-p)), k = 0..n) ] :
```

ezért standardizáltja

$$X = \frac{B - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}$$

Ezt a transzformációt végrehajtjuk B értékkészletén, vagyis a

$$k = 0, 1, 2, \dots, n$$

értékeken.

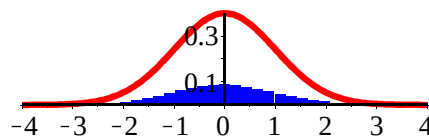
A standardizált X^* binomiális eloszlású változót a

$$P\left(X^* = \frac{k - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}\right) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, \quad k = 0, 1, 2, 3, \dots, n$$

képlet adja meg! Rajzoljuk fel ennek tüske vagy pálcika diagramját a standard $N(0, 1)$ normál eloszlás sűrűség függvényével együtt!

A rajzolás érdekében az $n = 100$ és $p = 0.3$ értékeket választottuk!

```
> standard := [seq([ [X[k+1], 0], [X[k+1],
    Probability(B = k) ]], k = 0..n) ] :
> tüskerajz := plot(standard, -4..4, color = blue,
    thickness = 5) :
> sűrűség := DensityPlot(Normal(0, 1), range =
    -4..4, color = red, thickness = 3) :
> display([ tüskerajz, sűrűség ], caption
    = Binomiális eloszlás standardizáltja és
    az N(0,1) sűrűsége)
n := 100
p := 0.3
```



Binomiális eloszlás standardizáltja és az $N(0,1)$ sűrűsége

Az illeszkedés **borzasztó!** Első gondolatunk az volt, hogy a kettő eloszlás jól fog illeszkedni egymáshoz, mert mindkettő **standardizált** értékek. A rosszban az a jó, hogy eloszlások **jellege** azonos, ami alatt a terjedelmet és az alakot értjük. Úgy lehetne egymáshoz igazítani a kettőt, ha a tüskéket **függőleges irányban megnövelnénk** annyival, hogy elérjék a harang görbét! Pontosan ezt tesszük!

Az **eltérés** abból adódik, hogy a binomiális eloszlás **diszkrét**, míg a normál eloszlás **folytonos**. Ezért, ha a pálcikák hosszait összeadjuk, akkor 1-t kapunk, míg ha a harang görbe alatti területet számoljuk, akkor is 1-t kapunk. A **hosszúság** 1 dimenziós mennyiség, míg a **terület** 2 dimenziós mennyiség. Az illeszkedési eltérés onnan ered, hogy két különböző dimenziót hasonlítottunk össze, ami képtelenség. Hogyan lehetne összehasonlítani azonos dimenziókat? Úgy, ha a tüskék végeit összekötjük egyenes szakaszokkal és a keletkezett **trapézok** (illetve két háromszög) területeinek összegét vesszük. Ennek a területnek kellene 1-t adni!

Ha valaki emlékszik a **trapéz-formulára**, akkor tudja, hogy a trapézok területének összege megegyezik a függőleges pálcikák összege, kivéve az első és utolsót, mert azok felét kell hozzáadni és ezt az összeget szorozni kell a pálcikák közötti távolsággal. Mivel "nagy" n esetén az első és utolsó érték közel nulla és a pálcikák összege 1, ezért a trapézok területének összege számértékileg egyenlő pálcikák közötti távolsággal. A k -ik pálcika az x -tengelyen az $x_k = \frac{k - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}$ értéknél

van, ezért

$$\Delta x = x_k - x_{k-1} = \frac{1}{\sqrt{n \cdot p \cdot (1 - p)}}$$

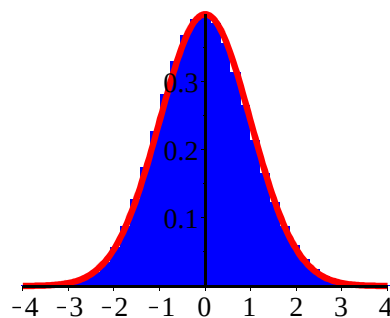
Nézzük mit kapunk a Maple-vel!

```
> faktor := sqrt(n*p*(1-p)) :
> standard_modos := [seq([ [X[k+1], 0],
    [X[k+1], faktor*Probability(B = k) ]], k = 0..n) ] :
> tüskerajz_modos := plot(standard_modos,
    -4..4, color = blue, thickness = 5) :
display([ tüskerajz_modos, sűrűség ])
```

a pálcikák közötti távolság. Tehát a trapéz területek összege

$$T_{\text{trapéz}} = \frac{1}{\sqrt{n \cdot p \cdot (1-p)}}.$$

Ahhoz, hogy a trapéz területek összege 1 legyen, ahhoz meg kell szorozni a pálcikák magasságait $\sqrt{n \cdot p \cdot (1-p)}$ -vel, vagyis a binomiális eloszlás szórásával.



Ennél pontosabb illeszkedésről álmodni se mertünk volna.

13.7.2. A Centrális Határeloszlás Tétel

Az előző részben a standard normál eloszlás **sűrűség** függvényét közelítettük a standardizált binomiális **eloszlással**. Láttuk, hogy az eltérések miatt be kellett építeni egy **szorzó faktort**! A centrális határeloszlás-tétel a standard normál eloszlás **eloszlásfüggvényének** egyre jobb közelítéséről szól standardizált összegek eloszlásfüggvényeivel, ahol **nem kell beépíteni** szorzó faktort.

13.7.2. TÉTEL. Centrális Határeloszlás Tétel (CHT)

Legyenek $X_1, X_2, \dots, X_n, \dots$ független és azonos eloszlású valószínűségi változók végtelen sorozata, amelyek közös $\mu = E(X_i)$ várható értéke és $\sigma = \sigma(X_i) > 0$ szórása létezik. Ekkor az

$$S_n^* = \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{(X_k - \mu)}{\sigma} = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}$$

valószínűségi változók sorozatára fennáll a

$$\lim_{n \rightarrow \infty} P(S_n^* < x) = \Phi(x)$$

határérték reláció, ahol $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ a standard normál eloszlás

eloszlásfüggvénye és $S_n = X_1 + X_2 + \dots + X_n$ az összeg változó.

CHT története.

A Centrális Határeloszlás tétel első változatát A. de Moivre (1667-1754) francia matematikus bizonyította 1738-ban

$X_i \sim \text{Binomial}\left(\frac{1}{2}\right)$ esetben. P-S. Laplace (1749-1827) francia matematikus újra felfedezte a tételt és 1812-ben

általánosította azt $X_i \sim \text{Binomial}(p)$ esetre, amikor $p \neq \frac{1}{2}$. A tételt C.F. Gauss (1777-1855) német matematikus és

csillagász bizonyította általános formában. Az első matematikailag precíz bizonyítást A.M. Ljapunov (1857-1918) orosz matematikus adta meg 1901-1902-ben megszabadulva az azonos eloszlás feltételétől. A tételt centrális határeloszlás tételnek először Gy. Pólya (1887-1985) magyar matematikus nevezte 1920-ban. A tételt gyengébb feltétel mellett és erősebb következtetéssel J.W. Lindeberg (1876-1932) finn matematikus 1922-ben bizonyította.

Az alkalmazások előtt nézzünk néhány szemléltetést, amelyek az **eloszlásokban** térnek el

egymástól.

▼ 13.7.3. Szemléltetés Bernoulli-eloszlással

Legyenek az $X_i \sim \text{Bernoulli}(p)$ valószínűségi változók független Bernoulli-eloszlásúak közös $0 < p < 1$ paraméterrel. Ekkor $E(X_i) = p$ a várható érték, $\sigma(X_i) = \sqrt{p \cdot (1 - p)}$ a szórás és az összeg változó $S_n = X_1 + X_2 + \dots + X_n \sim \text{Binomial}(n, p)$ binomiális eloszlású.

A szemléltetés a központi határeloszlás tétel de Moivre-Laplace-által felismert verzióját adja.

13.7.3. TÉTEL. (de Moivre, 1738–Laplace, 1812)

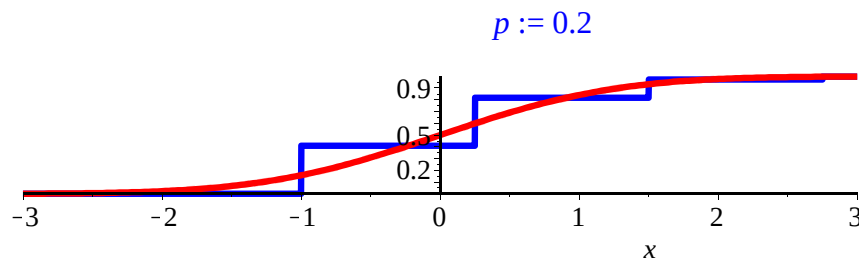
Legyen $a, b \in \mathbb{R} \cup \{\pm \infty\}$ és $a < b$. Ekkor

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

ahol $S_n = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$ binomiális eloszlás és a konvergencia $[a, b]$ – ben egyenletes.

Nézzük a $p = 0.2$ paraméter esetét és változzon n értéke 4 és 50 között. Animációval felrajzoljuk a standardizált binomiális eloszlás **eloszlásfüggvényét** különböző n értékre és a normál eloszlás eloszlásfüggvényét!

```
> restart :
> with(Statistics) : with(plots) :
> p := 0.2
> Gauss := CDF(Normal(0, 1), x) :
> rajzGauss := plot(Gauss, x = -3..3, color = [red, green], thickness = 3) :
> N := 50 :
> rajzok := NULL :
  for n from 4 to N do
    S := (RandomVariable(Binomial(n, p)) - n * p) :
          sqrt(p * (1 - p))
    rajzok := rajzok, display(
      plot(
        CDF(
          S / sqrt(n), x
        ), x = -3..3, color = blue, thickness = 3
      ),
      rajzGauss
    ) :
  end do:
  rajzok := [rajzok] :
> display(rajzok, insequence = true)
```



A diszkrét eloszlás lépcsős függvényei n növekedésével egyre jobban megközelítik a normál eloszlás folytonos eloszlásfüggvényét.

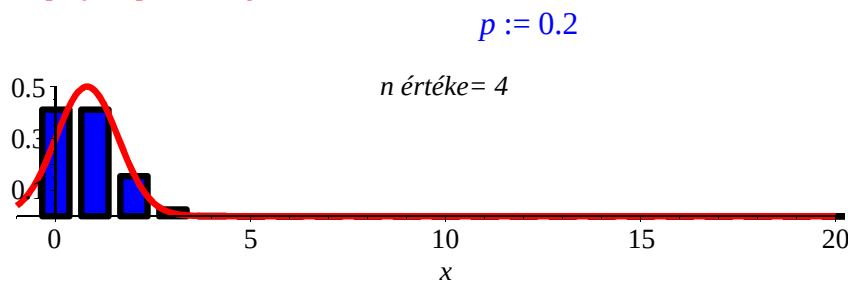
Javasoljuk az olvasónak, hogy változtassa N és p paraméterek értékét és kísérletezzon!

Egy **másik szemléltetési** lehetőség, ha rögzített p valószínűség mellett a

$\text{binomial}(n, p)$ eloszlás oszlop diagramjához megrajzoljuk a $\text{Normal}(n \cdot p, \sqrt{n \cdot p \cdot (1 - p)})$ normál eloszlás sűrűség függvényét n értékének növekedésével!

A rajzoljuk fel az ábrákat animációval, ha $p = 0.2$ és n értéke növekszik 4-től egyesével 50-ig!

```
> restart : with(Statistics) : with(plots) :
> p := 0.2
> kepek := NULL :
  for n from 4 to 50 do
>   kep1 := ColumnGraph( [ seq(ProbabilityFunction(Binomial(n, p), k), k = 0..20) ], color
      = blue, thickness = 3, width = 0.7, distance = 0.3, offset = -0.35) :
>   kep2 := plot( PDF( Normal( n*p, sqrt( n*p*(1-p) ) ), x ), x = -1..20, color = red, thickness = 3) :
>   num := convert(n, string) : szoveg := textplot( [ 10, 0.5, cat(n értéke = , num) ], color
      = black) :
>   kepek := kepek, display( [ kep1, kep2, szoveg ] ) :
  end do:
> display(kepek, insequence = true)
```



Látható, hogy a normál eloszlás sűrűség függvénye és a binomiális eloszlás egyre jobban megközelítik egymást n növekedésével.

▼ 13.7.4.Szemléltetés Poisson-eloszlással

Legyenek az $X_i \sim \text{Poisson}(\lambda)$ valószínűségi változók független Poisson-eloszlásúak közös

$0 < \lambda$ paraméterrel. Ekkor $E(X_i) = \lambda$ a várható érték, $\sigma(X_i) = \sqrt{\lambda}$ a szórás és az összeg

változó $S_n = X_1 + X_2 + \dots + X_n \sim \text{Poisson}(n \cdot \lambda)$ is Poisson-eloszlású. Nézzük a $\lambda = 0.5$

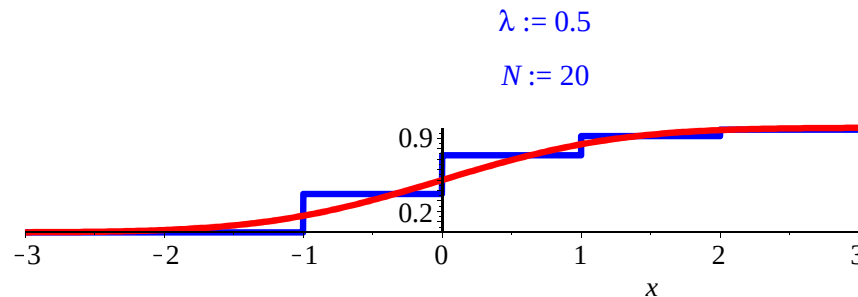
paraméter esetét és változzon n értéke 2 és 20 között. Animációval felrajzoljuk a Poisson eloszlás eloszlásfüggvényét különböző n értékre és a normál eloszlás eloszlásfüggvényét!

```
> restart :
> with(Statistics) : with(plots) :
> lambda := 0.5
> Gauss := CDF(Normal(0, 1), x) :
> rajzGauss := plot( Gauss, x = -3..3, color = [ red, green ], thickness = 3) :
> N := 20
> rajzok := NULL :
```

```

for n from 2 to N do
  S := RandomVariable( Poisson( n·λ ) ) :
  rajzok := rajzok, display( [ plot( CDF( S, x·√λ·√n + n·λ ), x=-3..3, color = blue, thickness
    = 3 ), rajzGauss ] ) :
end do:
rajzok := [ rajzok ] :
> display(rajzok, insequence = true)

```



Mivel a *Statistics* csomagba nincs beépítve a diszkrét változók transzformációja úgy, mint

például az $S_n^* = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}$, ezért a transzformációt a *CDF* eloszlásfüggvénybe kellett

beépítenünk! Az összeg változót $S_n \sim \text{Poisson}(n \cdot \lambda)$ formában definiáljuk és ennek eloszlásfüggvényével felépítjük a standardizált összeg eloszlásfüggvényét a

$$P(S_n^* < x) = P\left(\frac{S_n - n \cdot \lambda}{\sqrt{n} \cdot \sqrt{\lambda}} < x\right) = P\left(S_n < n \cdot \lambda + x \cdot \sqrt{n} \cdot \sqrt{\lambda}\right) = CDF\left(S_n, n \cdot \lambda + x \cdot \sqrt{n} \cdot \sqrt{\lambda}\right)$$

képlet alapján.

Javasoljuk az olvasónak, hogy változtassa *N* és *λ* paraméterek értékét és kísérletezzon!

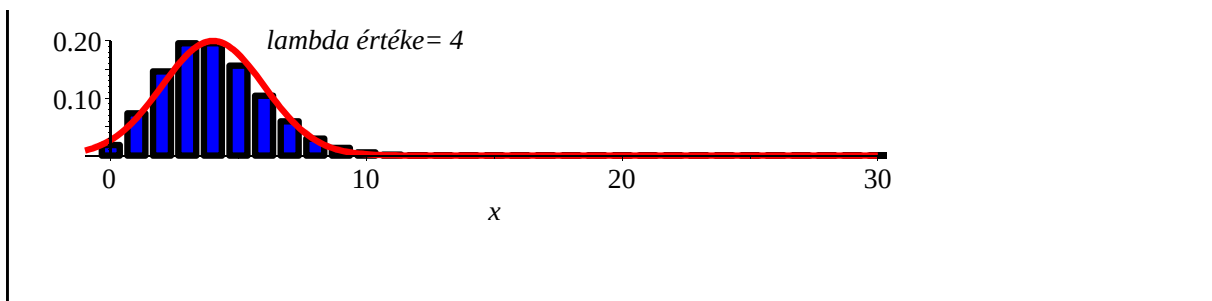
Egy **másik szemléltetési** lehetőséget kapunk, ha a *Poisson*(*λ*) eloszlás oszlop diagramjához megrajzoljuk a *Normal*(*λ*, $\sqrt{\lambda}$) normál eloszlás sűrűség függvényét *λ* értékének növekedésével!

A rajzoljuk fel az ábrákat animációval, ha *λ* értéke növekszik 4-től egyesével 15-ig!

```

> restart : with(Statistics) : with(plots) :
> kepek := NULL :
  for λ from 4 to 15 do
> kep1 := ColumnGraph( [ seq( ProbabilityFunction( Poisson( λ ), k ), k = 0..30 ) ], color
  = blue, thickness = 3, width = 0.7, distance = 0.3, offset = -0.35 ) :
> kep2 := plot( PDF( Normal( λ, √λ ), x ), x = -1..30, color = red, thickness = 3 ) :
> num := convert( λ, string ) : szöveg := textplot( [ 10, 0.2, cat( lambda értéke = , num ) ], color
  = black ) :
> kepek := kepek, display( [ kep1, kep2, szöveg ] ) :
  end do:
> display(kepek, insequence = true)

```



Látható, hogy a normál eloszlás sűrűség függvénye és a Poisson- eloszlás egyre jobban megközelítik egymást λ növekedésével.

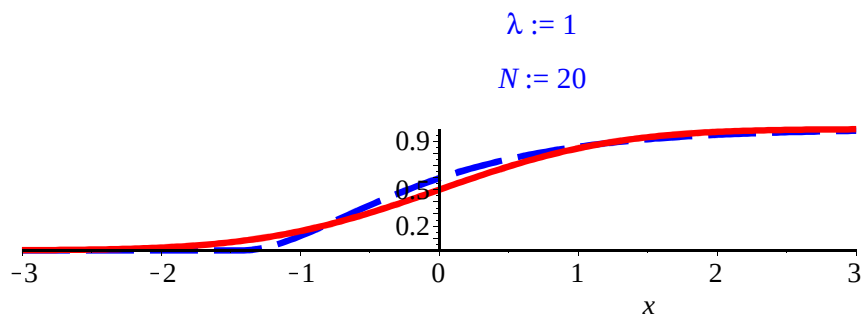
▼ 13.7.5.Szemléltetés exponenciális-eloszlással

Legyenek az $X_i \sim \text{Exponential}(\lambda)$ valószínűségi változók független exponenciális-eloszlásúak közös $0 < \lambda$ paraméterrel. Ekkor $E(X_i) = \lambda$ a várható érték, $\sigma(X_i) = \lambda$ a szórás és az összeg változó

$S_n = X_1 + X_2 + \dots + X_n \sim \text{Erlang}(\lambda, n)$ Erlang-eloszlású λ és n paraméterekkel. Nézzük a $\lambda = 1$ paraméter esetét és változzon n értéke 2 és 20 között.

Animációval felrajzoljuk az $\text{Erlang}(\lambda, n)$ eloszlás eloszlásfüggvényét $\lambda = 1$ és különböző n értékre, valamint a normál eloszlás eloszlásfüggvényét!

```
> restart :
> with(Statistics) : with(plots) :
  λ := 1;
> Gauss := CDF(Normal(0, 1), x) :
> rajzGauss := plot(Gauss, x = -3..3, color = [red, green], thickness = 3) :
> N := 20
> rajzok := NULL :
  for n from 2 to N do
    rajzok := rajzok, display([plot(CDF(RandomVariable(Erlang(λ, n))), λ · (n + x · √n), x
      = -3..3, color = blue, thickness = 3, linestyle = 3), rajzGauss)]) :
  end do;
  rajzok := [rajzok] :
> display(rajzok, insequence = true)
```



A Statistics csomagban a folytonos változók olyan transzformációja beépített úgy, mint például

az $S_n^* = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}$. Azonban a transzformáció után a CDF eloszlásfüggvényt számoló eljárás

rossz értékeket adott vissza, ha λ értéke kicsi (≤ 0.2) és $x \leq -3$! Ezért helyette az $S_n \sim \text{Erlang}(\lambda, n)$ összeg változót definiáltuk és ennek eloszlásfüggvényével a standardizált összeg eloszlásfüggvényét

$$P(S_n^* < x) = P\left(\frac{S_n - n \cdot \lambda}{\sqrt{n} \cdot \lambda} < x\right) = P(S_n < n \cdot \lambda + x \cdot \sqrt{n} \cdot \lambda) = CDF(S_n, n \cdot \lambda + x \cdot \sqrt{n} \cdot \lambda)$$

képlettel számoltuk.

Javasoljuk az olvasónak, hogy változtassa N és λ paraméterek értékét és kísérletezzon!

13.7.6. DEFINÍCIÓ. KONVERGENCIA ELOSZLÁSBAN

Tekintsük az $X_1, X_2, \dots, X_n, \dots$ valószínűségi változók végtelen sorozatát. Legyen $F_{X_n}(t)$ az X_n valószínűségi változó eloszlás függvénye és $F_Y(t)$ az Y változó eloszlás függvénye. Azt mondjuk, hogy az X_n sorozat **konvergál** az Y valószínűségi változóhoz **eloszlásban**, ha

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F_Y(t)$$

teljesül minden olyan $t > 0$ helyen, ahol az $F_Y(t)$ folytonos.

Jelölése $\lim_{n \rightarrow \infty} X_n \stackrel{d}{=} Y$ vagy $X_n \xrightarrow{d} Y$, ahol d = **d**istribution vagy eloszlás kezdő betűje.

Ezen definíció alapján a határeloszlás tétel a megadott feltételek mellett azt állítja, hogy

$$\lim_{n \rightarrow \infty} \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \stackrel{d}{=} Z,$$

ahol $Z \sim N(0, 1)$ standard normál eloszlású változó.

A tárgyalt kétféle konvergencia között a kapcsolatot tisztázza az alábbi tétel.

13.7.7. TÉTEL.

A valószínűségi mértékben való konvergenciából következik az eloszlásban való konvergencia

Tegyük fel, hogy $X_n \xrightarrow{Pr} Y$ teljesül. Ekkor igaz lesz az $X_n \xrightarrow{d} Y$ reláció is.

Bizonyítás

Tekintsük az alábbi átalakításokat

$$\begin{aligned} P(X_n \leq t) &= P(\{X_n \leq t\} \cap \{|X_n - Y| \leq \varepsilon\}) + P(\{X_n \leq t\} \cap \{|X_n - Y| > \varepsilon\}) \leq \\ &\leq P(Y \leq t + \varepsilon) + P(|X_n - Y| > \varepsilon). \end{aligned}$$

Legyen $\eta > 0$ tetszőleges. Mivel $F_Y(t)$ folytonos a t pontban, ezért ha $\varepsilon > 0$ elegendően kicsiny, akkor

$$F_Y(t + \varepsilon) = P(Y \leq t + \varepsilon) < P(Y \leq t) + \frac{\eta}{2}.$$

Másrészt az $X_n \xrightarrow{Pr} Y$ feltétel miatt $P(|X_n - Y| > \varepsilon) \leq \frac{\eta}{2}$, ha n elég nagy. A fenti egyenlőtlenségek alapján

$$F_n(t) = P(X_n \leq t) < F(t) + \eta.$$

Hasonló módon kapjuk az alsó becslést is, ha n értéke elegendően nagy, akkor

$$F(t) - \eta < F_n(t).$$

Tehát azt kaptuk, hogy $\lim_{n \rightarrow \infty} F_n(t) = F(t) = P(Y < t)$ minden olyan t pontban, ahol $F(t)$ folytonos.

Vagyis $X_n \xrightarrow{d} Y$ teljesül. ♣

Megjegyzés.

A 13.7.7. tétel megfordítása általában nem igaz. Igaz viszont a megfordítás, ha $Y = c$ konstans valószínűségi változó.

13.7.8. LEMMA.

Tegyük fel, hogy $X_n \xrightarrow{d} c$ teljesül valamilyen c konstanssal. Ekkor igaz a fordított $X_n \xrightarrow{Pr} c$ konvergencia.

Tehát, ha a határérték **konstans**, akkor a valószínűségi mértékben való konvergencia **ekvivalens** az eloszlásban való konvergenciával.

13.7.9. PÉLDA. FÖLDMÉRÉSEK ÁTLAGÁNAK HIBÁJA

Egy földmérő **ismert távolságot**, mondjuk 1 kilométert szeretne megmérni. Tudja, hogy a pontos mérést torzítják a légköri hatások és egyéb emberi hibák. Ezért több mérést végez ugyanarra a távolságra és a mérési **adatok átlagát** veszi. Feltételezi, hogy az egyes mérések független valószínűségi változók közös eloszlással, amelyek várható értéke $\mu = 1$ [km] és standard szórása $\sigma = 0.0002$ [km].

Hányszor mérje meg a távolságot a földmérő, hogy 95%-os megbízhatósággal a mérések átlaga legfeljebb 0.0001 [km] távolsággal térjen el a valódi távolságtól?

Megoldás

Jelölje S_n az n mérés összegét és $\frac{S_n}{n}$ a mérések átlagát, melynek várható értéke $\mu = 1$ és szórása

$\sigma\left(\frac{S_n}{n}\right) = \frac{0.0002}{\sqrt{n}}$. A kérdés első megválaszolásához használjuk a Csebisev-egyenlőtlenséget $\varepsilon = 0.0001$ paraméterrel

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq 0.0001\right) \leq \frac{0.0002^2}{n \cdot 0.0001^2} = \frac{4}{n}$$

A feladatban a **tagadás** esemény valószínűségét kell becsülni, ezért

$$P\left(\left|\frac{S_n}{n} - \mu\right| \leq 0.0001\right) = 1 - P\left(\left|\frac{S_n}{n} - \mu\right| \geq 0.0001\right) \geq 1 - \frac{4}{n} \geq 0.95.$$

Az utolsó egyenlőtlenség megoldása $n \geq 80$.

Tehát a Csebisev-egyenlőtlenség szerint, ha 80 vagy ennél több mérést végzünk, akkor legalább 0.95 annak a valószínűsége, hogy a mérések átlagának hibája kisebb, mint 10 cm!

Korábban láttuk, hogy a Csebisev-egyenlőtlenség túl nagy n érték szolgáltat! Azt reméljük, hogy a centrális határeloszlással pontosabb becslést kapunk n értékére. Tegyük fel, hogy n értéke elegendően nagy ahhoz, hogy S_n közelítőleg normál eloszlású legyen. Ekkor

$$P\left(\left|\frac{S_n}{n} - \mu\right| \leq 0.0001\right) = P\left(-0.5 \cdot \sqrt{n} < \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} < 0.5 \cdot \sqrt{n}\right) = P\left(-0.5 \cdot \sqrt{n} < S_n^* < 0.5 \cdot \sqrt{n}\right) \approx$$

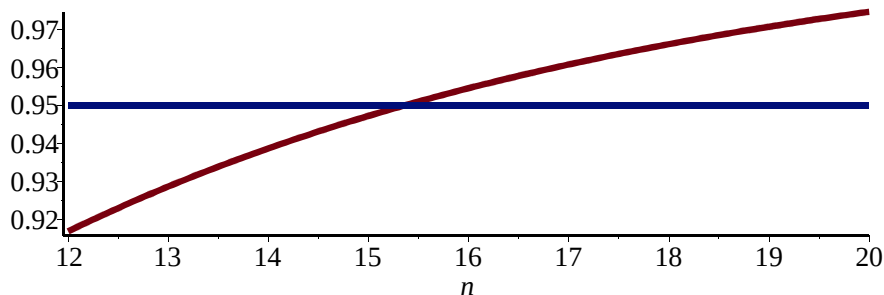
$$\approx \frac{1}{\sqrt{2\pi}} \int_{-0.5\sqrt{n}}^{+0.5\sqrt{n}} e^{-\frac{t^2}{2}} dt = \Phi(0.5\sqrt{n}) - \Phi(-0.5\sqrt{n}) \geq 0.95$$

> restart : with(Statistics) :

> e := CDF(Normal(0, 1), 0.5*sqrt(n)) - CDF(Normal(0, 1), -0.5*sqrt(n)) = 0.95

> plot([lhs(e), rhs(e)], n = 12..20, thickness = 3)

$$e := \operatorname{erf}\left(0.2500000000 \sqrt{n} \sqrt{2}\right) = 0.95$$



> n = fsolve(e, n = 14..18)

$$n = 15.3658352800$$

(1.7.1)

Tehát a centrális határeloszlás-tétel segítségével $n \geq 16$ (jóval kisebb) értéket kaptunk, mint a Csebisev-egyenlőtlenséggel. Maradt azonban az a kérdés, hogy $n = 16$ elegendően nagy-e ahhoz, hogy S_n^* és S_n eloszlását normál eloszlással közelíthessük?

13.7.10. PÉLDA. A 13.7.9. PÉLDA FOLYTATÁSA

Tegyük fel, hogy a földmérő egy **ismeretlen hosszúságú** távolságot mér meg ugyanazzal az eszközzel, ugyanolyan körülmények között, mint az előző feladatban. Ehhez $n = 36$ alkalommal mér, majd ezeket átlagolja. Milyen biztos lehet abban, hogy az átlag a valódi érték 0.0002 [km] = 20 [cm] sugarú környezetébe esik?

Megoldás

Használjuk a normál közelítést a becsléshez

$$P\left(\left|\frac{S_n}{n} - \mu\right| < 0.0002\right) = P(|S_n^*| < 0.5\sqrt{n}) = P(|S_n^*| < 3) \approx \frac{1}{\sqrt{2\pi}} \int_{-3}^3 e^{-\frac{t^2}{2}} dt = 0.997$$

> with(Statistics) :

evalf(CDF(Normal(0, 1), 3) - CDF(Normal(0, 1), -3))

$$0.9973002039$$

(1.7.2)

Tehát a földmérő 99.7%-ban biztos lehet abban, hogy az átlag érték a valódi távolság 0.0002 [km] sugarú környezetébe esik.

13.7.11. PÉLDA. Véletlenszerű számlák kerekítéseinek összegéről

Jelenleg a legkisebb fizető eszközünk az 5 Forint. Ezért, ha készpénzzel fizetünk, akkor a

pénztárgépek a fizetendő összeget automatikusan kerekítik a legközelebbi 5 forinthez vagy lefelé vagy fölfelé. Így például a 32 Ft is és a 28 Ft is kerekítés után 30 Ft lesz. A kerekítés lefelé is és felfelé is maximálisan 2.5 Ft lehet. Tegyük fel, hogy a kerekítések folytonosan változhatnak -2.5 és $+2.5$ között.

13.7.11. PÉLDA.

Azt szeretnénk tanulmányozni, hogy 100 független és véletlenszerű **kerekítés összegének abszolútértéke** mekkora valószínűséggel haladja meg a 10 Ft-ot?

- (i) Mennyi az összegzett kerekítések várható értéke és szórása?
- (ii) Készítsünk szimulációs programot, amely kiszámolja 100 kerekített érték összegét ismételten 200-szor, majd felrajzolja a kérdéses esemény relatív gyakoriságát!
- (iii) A centrális határeloszlás-tétel segítségével számoljuk ki az esemény valószínűségét! Egyezik-e a relatív gyakoriságokból adódó értékkel?

Megoldás

(i) A kerekítési értékek modellezhetők $X_1, X_2, X_3, \dots, X_{100}$ független valószínűségi változókkal, ahol mindegyik változó folytonos egyenletes eloszlású a $[-2.5, 2.5]$ intervallumon

$$X_i \sim \text{Uniform}(-2.5, 2.5).$$

A 100 kerekítés halmozott összege az

$$S_{100} = X_1 + X_2 + \dots + X_{100}$$

valószínűségi változó.

Mivel minden $i = 1, 2, 3, \dots, 100$ esetén

$$E(X_i) = 0, \text{Var}(X_i) = \frac{(2.5 - (-2.5))^2}{12} = \frac{25}{12} \text{ és } \sigma(X_i) = \frac{5}{2\sqrt{3}} \approx 1.443$$

ezért

$$E(S_{100}) = 0, \text{Var}(S_{100}) = 100 \cdot \frac{25}{12} = \frac{25^2}{3}, \sigma(S_{100}) = \frac{25}{\sqrt{3}} \approx 14.43.$$

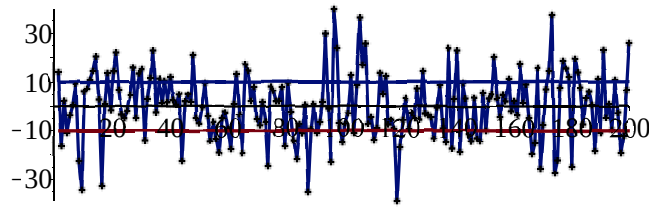
Tehát hosszú távon a felfelé és lefelé kerekítések kiegyenlítik egymást és a várható érték 0 Ft.

(ii) Felveszünk egy kerekítési változót és kiszámoljuk várható értékét és szórását!

Az alábbi program összead 100 véletlen kerekítési értéket és ezt megismétli 200-szor. A kapott $S_1, S_2, \dots, S_{200}$ véletlen összegeket felrajzolja vonalas diagrammal és hozzá az $y = 10$ és $y = -10$ szinteket!

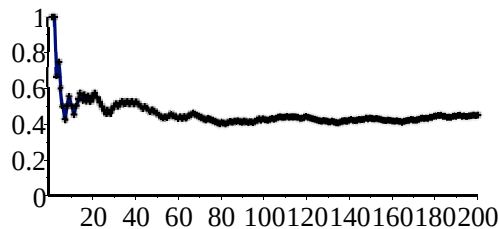
```
> restart : with(Statistics) : with(plots) :
> X := RandomVariable(Uniform(-2.5, 2.5)) :
> μ := Mean(X); σ := StandardDeviation(X);
> N := 200 : n := 100 :
  S := NULL :
  for i from 1 to N do
    M := Sample(X, n)
  > S := S, add(M[k], k = 1..n) :
  end do:
  S := [S] :
> vonalak := LineChart(S) :
  határok := plot([-10, 10], 1..N) :
  display([vonalak, határok])

μ := 0.
σ := 1.443375673
```

Az alábbi ciklus megvizsgálja, hogy mely S_k összegre teljesül az $\{|S_k| > 10\} = \{S_k < -10\} \cup \{S_k > 10\}$ feltétel a $k = 1, 2, 3, \dots, 200$ közül. Ezután felrajzolja ezek relatív gyakoriságát.

```
> rgy := NULL : gyak := 0 :
for k from 1 to N do
  if abs(S[k]) > 10 then gyak := gyak + 1 end if:
  rgy := rgy,  $\frac{gyak}{k}$  :
end do:
rgy := [rgy] :
rajz1 := LineChart(rgy, view = [1..N, 0..1]) : rajz1
```



(iii) A centrális határeloszlás-tétellel számoljuk ki az $|X_1 + X_2 + \dots + X_{100}| > 10$ esemény valószínűségét!

A relatív gyakoriságok azt **sejtetik**, hogy a keresett

$$p = P(|X_1 + X_2 + \dots + X_{100}| > 10)$$

valószínűség kb. 0.5. Pontosabb értéket közelítő számítással tudunk mondani.

Egy lehetséges felső becslést ad p -re a Csebisev-egyenlőtlenség, melyben

$$X_1 + X_2 + \dots + X_{100} = S_{100}, E(S_{100}) = 0 \text{ és } \sigma^2(S_{100}) = \frac{25^2}{3}.$$

Ezért

$$P(|X_1 + X_2 + \dots + X_{100}| > 10) \leq \frac{\sigma^2(S_{100})}{\epsilon^2} = \frac{\frac{25^2}{3}}{10^2} \approx 2.083.$$

Mivel minden valószínűség 0 és 1 közé esik, ezért a Csebisev-egyenlőtlenség nem ad használható eredményt.

Bontsuk két részre az abszolútértéket a centrális határeloszlás -tétellel való számításához

$$P(|X_1 + X_2 + \dots + X_{100}| > 10) = P(X_1 + X_2 + \dots + X_{100} > 10) + P(X_1 + X_2 + \dots + X_{100} < -10).$$

Standardizáljuk az összeget és használjuk a normál eloszlás eloszlásfüggvényével való közelítést

$$\begin{aligned} P(X_1 + X_2 + \dots + X_{100} > 10) &= P(X_1 + X_2 + \dots + X_{100} - n \cdot \mu > 10 - n \cdot \mu) = \\ &= P\left(\frac{X_1 + X_2 + \dots + X_{100} - n \cdot \mu}{\sigma \cdot \sqrt{n}} > \frac{10 - n \cdot \mu}{\sigma \cdot \sqrt{n}}\right) = P\left(Z_n > \frac{10 - n \cdot \mu}{\sigma \cdot \sqrt{n}}\right) = P\left(Z_n > \frac{10 - 100 \cdot 0}{1.443337 \cdot \sqrt{100}}\right) \end{aligned}$$

$P(X_1 + X_2 + \dots + X_{100} > 10) = P(Z_n > 0.692823) = 1 - \Phi(0.692823) \approx 0.24421$,
ahol Z_n az összeg standardizáltja 0 várható értékkel és 1 szórással, $\Phi(x)$ a standard normál eloszlás eloszlásfüggvénye.

Hasonlóan $P(X_1 + X_2 + \dots + X_{100} < -10) = P(Z_n < -0.692823) = 1 - \Phi(0.692823) \approx 0.24421$.

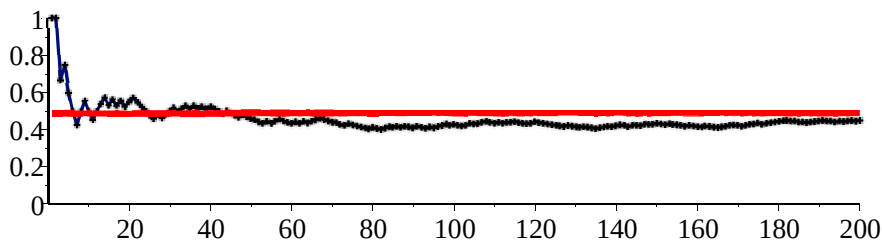
Így a keresett p valószínűsége

$$p = 0.24421 + 0.24421 = 0.48842$$

Ugyanez a számítás Maple-segítségével!

```
>  $\Phi := x \rightarrow \text{CDF}(\text{Normal}(0, 1), x) :$ 
>  $p_1 := 1 - \Phi\left(\frac{(10 - \mu \cdot n)}{\sigma \cdot \sqrt{n}}\right) : p_2 := \Phi\left(-\frac{(10 - \mu \cdot n)}{\sigma \cdot \sqrt{n}}\right) :$ 
>  $p := p_1 + p_2$ 
>  $\text{rajz2} := \text{plot}(p, 1..N, \text{thickness} = 3, \text{color} = \text{red}) :$ 
>  $\text{display}([ \text{rajz1}, \text{rajz2} ])$ 
```

$p := 0.488422316639885$



Láthatóan a kiszámolt valószínűséghez konvergálnak a relatív gyakoriságok.

A **közelítő formula** jelentőségét szeretnénk hangsúlyozni azzal, hogy írtunk egy olyan Maple programot, amely összead 100 darab egyenletes eloszlású változót.

A $P(S_{100} > 10)$ egyenlőtlenséget az eloszlásfüggvénnyel $(1 - \text{CDF}(S_{100}, 10))$ módon tudjuk számolni!

```
>  $S := \text{RandomVariable}(\text{Uniform}(-2.5, 2.5)) :$ 
> for  $k$  from 2 to 10 do
>    $S := S + \text{RandomVariable}(\text{Uniform}(-2.5, 2.5))$ 
> end do;
```

A Maple olyan hosszadalmasan számol, hogy ezt csak gyors géppel rendelkezők és türelmes olvasók indítsák el!

```
>  $\# 1 - \text{CDF}(S, 10)$ 
```

13.7.12. PÉLDA. A 13.5.3. példa megoldása centrális határeloszlás tétellel

Visszatérünk a minimális mintvételi elemszám meghatározásához, amikor legalább 90%-os biztonsággal szeretnénk meggyőződni arról, hogy a $p = 0.05$ selejt aránytól való eltérés legfeljebb 0.02?

A becsléshez az egyenlőtlenséget hozzuk olyan alakra, amely alkalmas a centrális határeloszlás tétel alkalmazására!

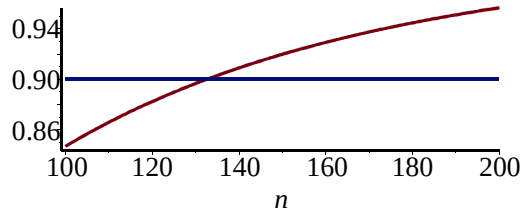
$$P\left(\left|\frac{k_n(A)}{n} - 0.05\right| < 0.02\right) = P\left(0.03 < \frac{k_n(A)}{n} < 0.07\right) = P(0.03 \cdot n < k_n(A) < 0.07 \cdot n)$$

$$P(0.03 \cdot n < k_n(A) < 0.07 \cdot n) \approx \frac{1}{\sqrt{2\pi}} \int_{a_n^*}^{b_n^*} e^{-\frac{x^2}{2}} dx = \Phi(b_n^*) - \Phi(a_n^*)$$

ahol

$$a_n^* = \frac{0.03 \cdot n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} = -\frac{0.02}{\sqrt{0.02 \cdot 0.98}} \sqrt{n} = -\sqrt{\frac{0.02}{0.98}} \cdot \sqrt{n}$$

$$\text{és } b_n^* = \frac{0.07 \cdot n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} = \frac{0.02}{\sqrt{0.02 \cdot 0.98}} \sqrt{n} = \sqrt{\frac{0.02}{0.98}} \cdot \sqrt{n}$$

A további számításokhoz használjuk a Maple beépített eljárásait! Rajzoljuk fel n változtatásával a normál eloszlással kifejezett valószínűséget! Az ábra szerint a 0.9 valószínűségi szintet a görbe az $n=130$ és $n=140$ között biztosan metszi. Ezt a metszéspontot megkeresi a Maple <i>fsolve</i> eljárása, amelyet egyenletek numerikus megoldásának közelítő számítására fejlesztettek ki.	<pre>> restart : with(Statistics) : with(plots) : > F := unapply(CDF(Normal(0, 1), $\frac{\sqrt{0.02}}{\sqrt{0.98}} \cdot \sqrt{n}$) - CDF(Normal(0, 1), $-\frac{\sqrt{0.02}}{\sqrt{0.98}} \cdot \sqrt{n}$), n) : > plot([F(n), 0.9], n = 100..200)</pre>  <pre>> n = fsolve(F(n) = 0.9, n = 120..140) n = 132.5716294000</pre> (1.7.3)
--	---

Tehát a centrális határeloszlás tétel alapján kapott legkisebb elemszám $n = 133$. Ez azonban túl kevésnek tűnik, hiszen láttuk a 16.3.3 feladat (c) pontjában, hogy binomiális eloszlást feltételezve, $n < 260$ esetben lesznek olyan n értékek, melyekre a valószínűség 0.9 alatt lesz. Például az $n = 180$ minta elemszámra a megbízhatóság csak 0.86 körül lesz.

<pre>> n := 180 > X := RandomVariable(Binomial(n, 0.05)) : > p = CDF(X, 0.07 \cdot n) - CDF(X, 0.02 \cdot n)</pre> $n := 180$ $p = 0.862385591605296$ (1.7.4)

Pontosabb becslést akkor tudnánk mondani, ha ismernénk a sokaság N elemszámát. Mert ilyen esetben hipergeometrikus eloszlással számolhatunk!

▼ 13.8. Ellenőrző kérdések

13.8.1. Mit mond ki a **Markov-egyenlőtlenség**?

13.8.2. Milyen feltételek mellett nem ad használható eredményt a Markov-egyenlőtlenség?

13.8.3. Milyen helyettesítéssel kapjuk a **Csebisev-egyenlőtlenséget** a Markov-egyenlőtlenségből?

13.8.4. Mit mond ki a **nagy számok törvényének Csebisev-alakja**?

13.8.5. **Hogyan kapjuk** a Csebisev-egyenlőtlenségből a nagy számok törvényének Csebisev-alakját?

13.8.6. Mit mond ki a **nagy számok törvényének Bernoulli-alakja**?

13.8.7. **Hogyan kapjuk** a Csebisev-egyenlőtlenségből a nagy számok törvényének Bernoulli-alakját?

13.8.8. Mikor mondjuk, hogy az $X_1, X_2, \dots, X_n, \dots$ valószínűségi változók sorozata **valószínűségi mértékben** konvergál az Y valószínűségi változóhoz?

13.8.9. Mit mond ki a **centrális határeloszlás- tétel**?

13.8.10. Mit mond ki a **deMoivre - Laplace-tétel**?

13.8.11. Mikor mondjuk, hogy az $X_1, X_2, \dots, X_n, \dots$ valószínűségi változók sorozata **eloszlásban** konvergál az Y valószínűségi változóhoz?

13.8.12. Mi a kapcsolat a valószínűségi mértékben való konvergencia és az eloszlásban való konvergencia között?

13.8.13. Hogyan **közelítjük** a binomiális eloszlást normál eloszlással?

13.8.14. Hogyan **közelítjük** a Poisson eloszlást normál eloszlással?

13.8.15. Milyen módon közelítünk egy integrált **Monte Carlo módszerrel**?

▼ 13.9. Gyakorló feladatok

13.9.1. feladat

Legyen $X \sim \text{Bernoulli}(p)$ valószínűségi változó, amelyre $P(X = 1) = p$ és $P(X = 0) = 1 - p$, ahol $0 < p < 1$. Hasonlítsuk össze a $P(X \geq 1)$ valószínűséget a Markov-egyenlőtlenséggel kapott becsléssel!

13.9.2. feladat

Legyen $X \sim \text{Exponential}(\lambda)$ exponenciális eloszlású valószínűségi változó $E(X) = 1$ várható értékkel. Számoljuk ki rendre az $x = 0.5, 1$ és 2 értékekre a $P(X \geq x)$ valószínűségek pontos értékeit és hasonlítsuk össze a Markov-egyenlőtlenségben kapott felső korlátokkal!

13.9.3. feladat

Legyenek az $X_1, X_2, \dots, X_{10}$ független valószínűségi változók Poisson eloszlásúak $E(X_i) = 1$ közös várható értékkel. Számoljuk ki a $P(X_1 + X_2 + \dots + X_{10} > 15)$ valószínűséget a Markov-egyenlőtlenség használatával! Majd számoljuk a $P(X_1 + X_2 + \dots + X_{10} > 15)$ valószínűséget, használva a független Poisson-eloszlások összege Poisson eloszlás tétel! Hasonlítsuk össze a két eredményt!

13.9.4. feladat

Legyen X folytonos valószínűségi változó, melynek várható értéke és varianciája is 20, azaz
(*) $E(X) = \text{Var}(X) = 20$

(i) Mit mondhatunk a **Csebisev-egyenlőtlenség** alapján a $P(0 \leq X \leq 40)$ valószínűségről?

(ii) Készítsünk olyan $X \sim \text{Normal}(\mu, \sigma)$ normál eloszlású változót, melyre teljesül a (*) feltétel! Mit mondhatunk ekkor a $P(0 \leq X \leq 40)$ valószínűségről? Hasonlítsuk össze az (i) pontban kapott értékkel!

(iii) Készítsünk olyan $X \sim \text{Uniform}(a, b)$ egyenletes eloszlású változót, melyre teljesül a (*) feltétel! Mit mondhatunk ekkor a $P(0 \leq X \leq 40)$ valószínűségről? Hasonlítsuk össze az (i) pontban kapott értékkel!

13.9.5. feladat

A **centrális határeloszlás tétel** kimondja, hogy független valószínűségi változók összege tart a normál eloszláshoz, tekintet nélkül arra, hogy mennyire furcsa egyedi változóból indultunk ki.

Teszteljük ezt számítógépes szimulációval! Válasszunk 25 független értéket a $[0, 1]$ intervallumban az alábbiakban megadott eloszlások sűrűség függvényével és számoljuk ki az S_{25} összegüket! Ismételjük meg a számítást $N=1000$ -szer és rajzoljuk fel a kapott összegek oszlop diagramját! Rajzoljuk be az ábrába a $Normal(E(S_{25}), \sigma(S_{25}))$ normál eloszlás sűrűség függvényét! Milyen az illeszkedés a normál sűrűség függvény és az oszlop diagram között?

(i) $f(x) = 1$

(ii) $f(x) = 2 \cdot x$

(iii) $f(x) = 3 \cdot x^2$

(iv) $f(x) = 4 \cdot \left| x - \frac{1}{2} \right|$

(v) $f(x) = 2 - 4 \cdot \left| x - \frac{1}{2} \right|$

13.9.6. feladat

Milyen nagy legyen n értéke, hogy az $S_n = X_1 + X_2 + \dots + X_n$ közelítőleg normál eloszlású legyen?

Ez a szám néha meglepően kicsi. Adjuk meg a választ számítógépes szimulációval! Válasszunk n darab számot a $[0, 1]$ intervallumból az $f(x)$ sűrűséggel, ahol $n = 3, 6, 12, 20$ és $f(x)$ a 13.10.5. feladatban adott sűrűség függvény. Számoljuk ki az S_n összeget, majd ismételjük meg 1000-szer a

kísérletet! Rajzoljuk fel az oszlop diagramot és a normál eloszlást!

Mennyi legyen n értéke, hogy az illeszkedés megfelelő legyen?

13.9.7. feladat

Legyen $X \sim \text{Binomial}(n, p)$ binomiális eloszlású valószínűségi változó.

(i) Ha $n = 100$ és $p = \frac{1}{4}$, akkor pontos számítással kapjuk a $P(X \leq 25) = 0.5534708238$

valószínűséget. Használjuk a centrális határeloszlás tétel de Moivre-Laplace verzióját és közelítsük a $P(X \leq 25)$ és $P(X < 26)$ valószínűségeket!

(ii) Ha $n = 100$ és $p = \frac{1}{4}$, akkor pontos számítással kapjuk a $P(X \leq 2) = 1.874075144 \cdot 10^{-10}$

valószínűséget. Használjuk a centrális határeloszlás tétel de Moivre-Laplace verzióját és közelítsük a $P(X \leq 2)$ valószínűséget!

13.9.8. feladat

Dobjunk fel egy szabályos pénzérmét 1000-szer egymás után. Mekkora a valószínűsége, hogy fej dobások száma 450 és 550 közé esik.

(a) Számoljunk Csebisev-egyenlőtlenséggel!

(b) Számoljunk centrális határeloszlás tétellel!

13.9.9. feladat

Legyenek az $X_1, X_2, \dots, X_{625}$ független azonos eloszlású valószínűségi változók sorozata, amelyek közös sűrűség függvénye az

$$f(x) = \begin{cases} 3 \cdot (1-x)^2, & \text{ha } 0 \leq x \leq 1 \\ 0, & \text{egyébként} \end{cases}$$

függvény. Alkalmazzuk a centrális határeloszlás tételt a $P(X_1 + X_2 + \dots + X_{625} < 170)$ valószínűség közelítésére!

13.9.10. feladat

Legyenek az $X_1, X_2, \dots, X_{144}$ független azonos eloszlású valószínűségi változók sorozata, amelyek

várható értéke $\mu = E(X_i) = 2$ és varianciája $\sigma^2 = \text{Var}(X_i) = 4$. Közelítsük a

$P(X_1 + X_2 + \dots + X_{144} > 144)$ valószínűséget a centrális határeloszlás tétellel!

13.9.11. feladat

Legyenek az $X_1, X_2, \dots, X_n, \dots$ független $N(0, 1)$ eloszlású valószínűségi változók sorozata.
Tekintsük az

$$Y_n = X_1^2 + X_2^2 + \dots + X_n^2$$

négyzetösszeget $n = 1, 2, 3, \dots$ esetén!

(i) Mutassuk meg, hogy $E(X_i) = 1$

(ii) Parciális integrálás segítségével igazoljuk, hogy $E(X_i^4) = 3$. Ebből következik

hogy $\text{Var}(X_i^2) = 2$

(iii) A centrális határeloszlás tétel alapján adjunk közelítést a $P(Y_{100} > 110)$

valószínűségre!

13.9.12. feladat

Legyenek az $X_1, X_2, \dots, X_n, \dots$ független egyenletes eloszlásúak a $[-1, 1]$ intervallumban, azaz $X_i \sim U(-1, 1)$. Tekintsük a

$$T_n = \frac{1}{n} \sum_{i=1}^n X_i^2$$

a négyzetösszegek átlagát.

(a) Teljesülhet-e a

$$\lim_{n \rightarrow \infty} P(|T_n - a| > \varepsilon) = 0$$

valamilyen a értékkel és tetszőleges $\varepsilon > 0$ -val? Hogyan lehet ez igaz?

(b) Határozzuk meg a értékét!

(c) Készítsünk számítógépes szimulációt!

13.9.13. feladat

A Monte Carlo integráláshoz legyen U_1 és U_2 független valószínűségi változók eloszlása egyenletes a $[-1, 1]$ intervallumban: $U_1, U_2 \sim \text{Uniform}(0, 1)$ és legyen

$$X = \begin{cases} 1 & , \text{ ha } U_1^2 + U_2^2 < 1, \\ 0 & , \text{ egyébként} \end{cases}$$

Bernoulli-változó. Képezzünk $n = 1000$ véletlen értéket az U_1 és U_2 változókra, majd az X változóra is az

$$X_n = \begin{cases} 1 & , \text{ ha } U_{1,n}^2 + U_{2,n}^2 < 1, \\ 0 & , \text{ egyébként} \end{cases} \quad (n = 1, 2, \dots, 1000)$$

feltétel vizsgálattal. Ekkor $E(X) = \frac{\pi}{4}$. A **nagy számok törvénye** alapján

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X_k}{n} = \frac{\pi}{4}$$

valószínűségi mértékben teljesül. Készítsünk számítógépes szimulációt ez alapján a π értékének közelítő számítására!

13.9.14. feladat

Legyen az X diszkrét valószínűségi változó táblázata a következő

$$X \quad \left| \quad X=0 \quad \right| \quad X$$

$$\begin{array}{c|c|c}
 & & = t \\
 \hline
 P & 1 & - \frac{1}{t} \frac{1}{t}
 \end{array}$$

ahol $t > 1$. Tehát $P(X=0) = 1 - \frac{1}{t}$ és $P(X=t) = \frac{1}{t}$.

Ezért $E(X) = 1$ és $Var(X) = t - 1$.

Tekintsük a $P(|X - 1| > a)$ valószínűséget!

(a) Ellenőrizzük, hogy $t = 10$ és $a = 8$ esetén $P(|X - 1| > a) = \frac{1}{10}$ és a Csebisev-egyenlőtlenség

felső korlátja ugyanerre a valószínűségre $\frac{9}{64}$. A különbség $\frac{9}{64} - \frac{1}{10} \approx 0.04$.

Azt mondjuk, hogy $t = 10$ mellett a **Csebisev rés** az X változóra $a = 8$ esetén 0.04.

(b) Határozzuk meg a Csebisev rés értékét $t=10$ mellett $a = 5$ - re és $a = 10$ - re!

(c) Vajon tudunk-e olyan t és a értéket megadni, hogy a Csebisev rés kisebb legyen, mint 0.01, 0.0001, 0.00001,...stb.

(d) Vajon lehet-e javítani a Csebisev-egyenlőtlenséget olyan értelemben, hogy a felső becslés közelebb legyen a tényleges valószínűséghez?

13.9.15. feladat

Önnek egy újonnan induló vállalkozása van, melynek neve CyberHírek KFT. és profilja a számítógépes játékok eladása Interneten. Statisztikai vizsgálatok alapján egy internetes eladással foglalkozó cég - mint amilyen az Öné- várható T élettartama $\mu = 30$ hónap.

(a) Legfeljebb mekkora az esélye, hogy a CyberHírek tovább él, mint $t_0 = 38$ hónap? A becsléshez használja a Markov-egyenlőtlenséget!

(b) Mennyit változik a fenti esély, ha statisztikai adatokból egyrészt az ismert, hogy egy internetes eladással foglalkozó cég élettartamának szórása $\sigma = 4$ hónap, másrészt a T élettartam eloszlása az átlagra nézve szimmetrikus? A becsléshez használja a Csebisev-egyenlőtlenséget!

(c) Mennyit változik az (a) pontban kért esély, ha tudjuk, hogy $T \sim N(30, 8)$, vagyis az élettartam normál eloszlású?

13.9.16. feladat

Egy repülőtérre érkező repülőgépek X száma egy adott t időintervallumon belül átlagosan $\mu = 100$ repülő. (a) Legfeljebb mekkora az esélye, hogy az adott repülőtérre a t idő alatt több, mint $N = 120$ repülő érkezik? A becsléshez használja a Markov-egyenlőtlenséget!

(b) Mennyit változik a fenti esély, ha statisztikai adatokból ismert egyrészt, hogy az érkező repülőgépek X számának szórása $\sigma = 10$ repülő, másrészt az eloszlás jó közelítéssel szimmetrikus az átlagra? A becsléshez használja a Csebisev-egyenlőtlenséget!

(c) Mennyit változik az (a) pontban kért esély, ha tudjuk, hogy $X \sim \text{Poisson}(100)$, vagyis az érkező repülőgépek száma Poisson - eloszlású?

13.9.17. feladat

A gyártmányok 5%-a hibás. A minőségellenőrzés csak akkor találja elfogadhatónak a tételt, ha abban legfeljebb 7% hibás van. Mekkora legyen a tételben levő gyártmányok n darabszáma, hogy a tétel fogadható legyen legalább 0.95 valószínűséggel?

Tipp: Használja a nagy számok J. Bernoulli-féle törvényét! $P\left(\varepsilon \leq \left| \frac{k_A}{n} - p \right| \right) \leq \frac{p(1-p)}{n\varepsilon^2}$