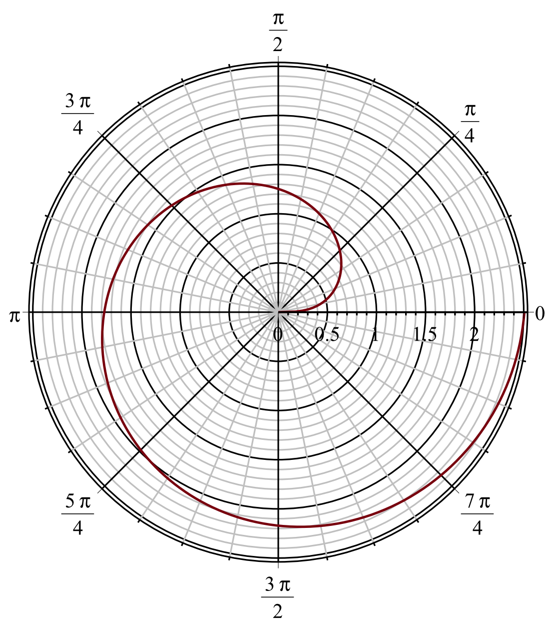
9. Gyakorlat

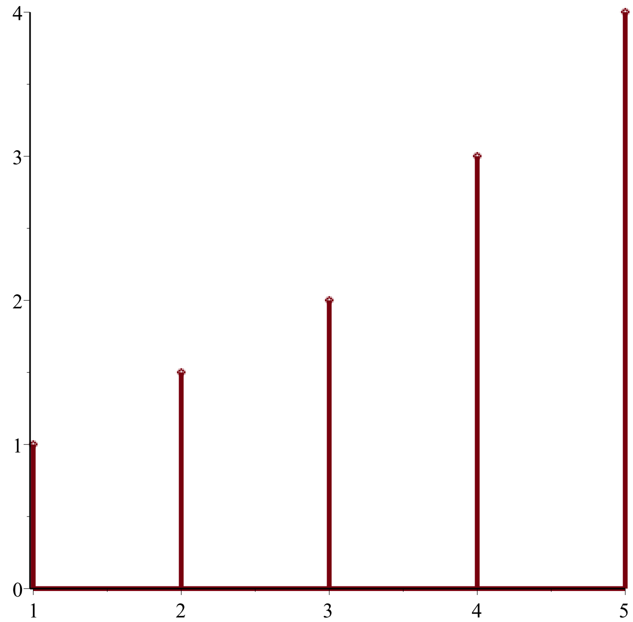
Maple ismerkedés

Grafikonok, függvényábrázolás II.:
Egyéb ábrázoló eljárások: plots[pointplot], plots[polarplot], DynamicSystems[DiscretePlot], Statistics[LineChart], Statistics[Histogram]

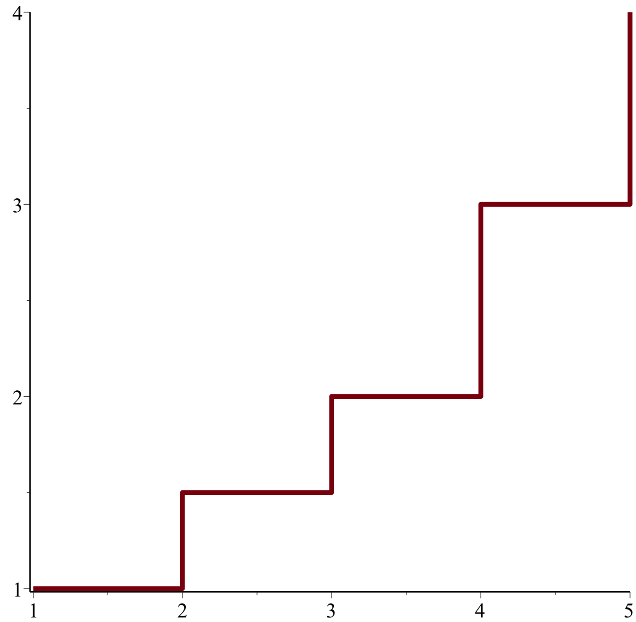
```
> restart;
> plots[polarplot](sqrt(t), t = 0..2*Pi);
```



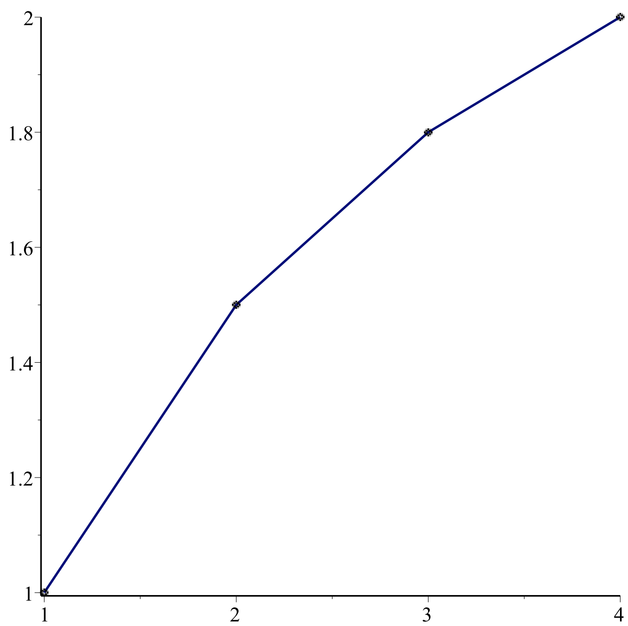

```
> DynamicSystems[DiscretePlot]([1, 2, 3, 4, 5], [1, 1.5, 2, 3, 4], style = stem, thickness = 3);
```



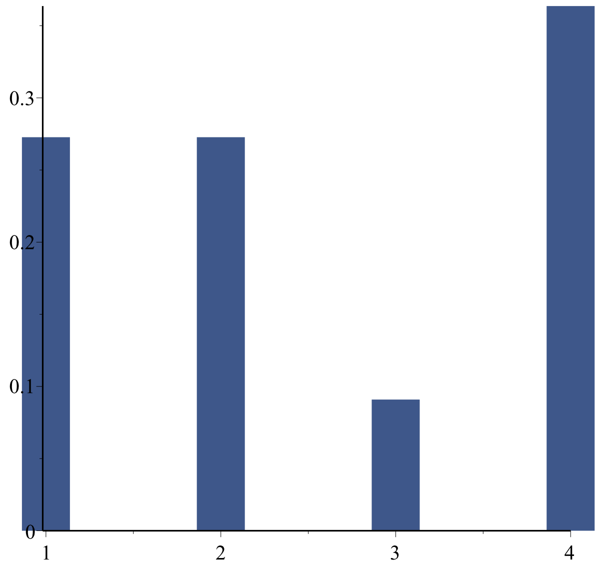

```
> DynamicSystems[DiscretePlot]([1, 2, 3, 4, 5], [1, 1.5, 2, 3, 4], style = stair, thickness = 3);
```



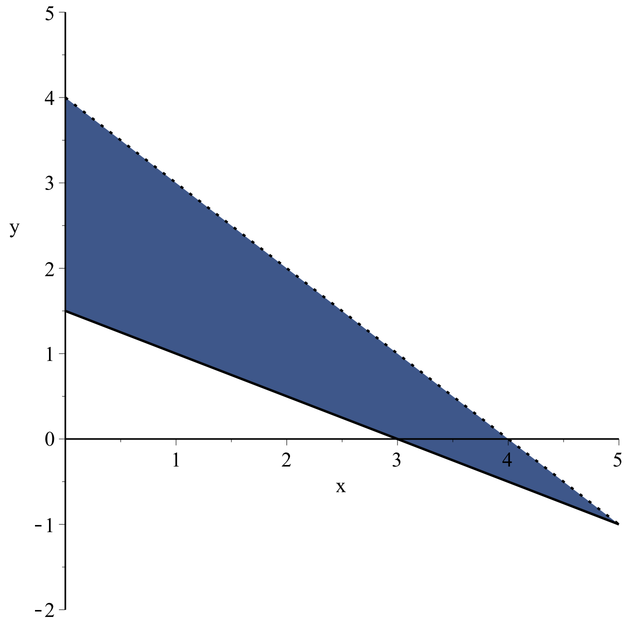

```
> Statistics[LineChart](<1,1.5,1.8,2>, style = line);
```




```
> Statistics[Histogram]([4, 1, 2, 4, 3, 1, 4, 2, 2, 1, 4],
  discrete = true, thickness = 30);
```

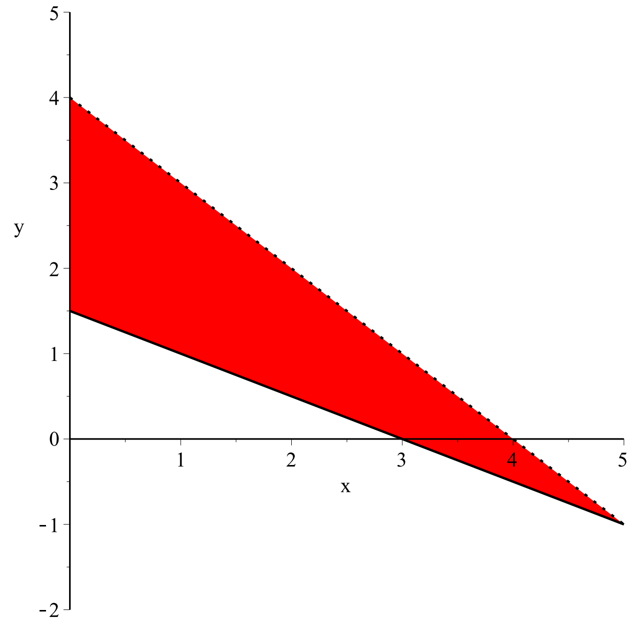



```
> plots[inequal]({x + y < 4, x + 2*y >= 3}, x = 0..5, y = -2..5);
```

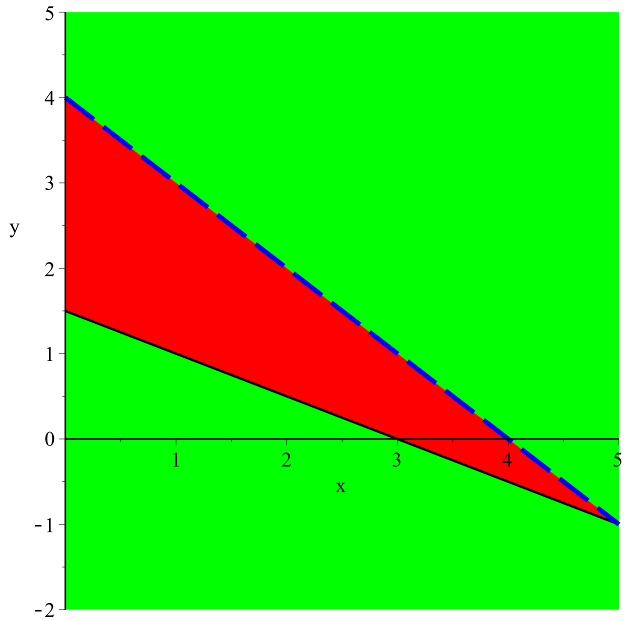



```
> P1 := plots[inequal]({x + y < 4, x + 2*y >= 3}, x = 0..5, y =
  -2..5, optionsfeasible = [color = red, transparency = 0.5]);
  P1;
```

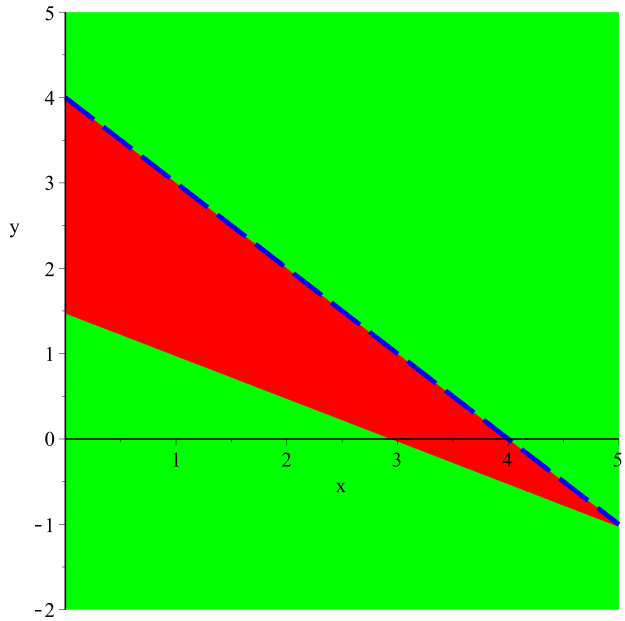
$$P1 := PLOT(\ldots)$$



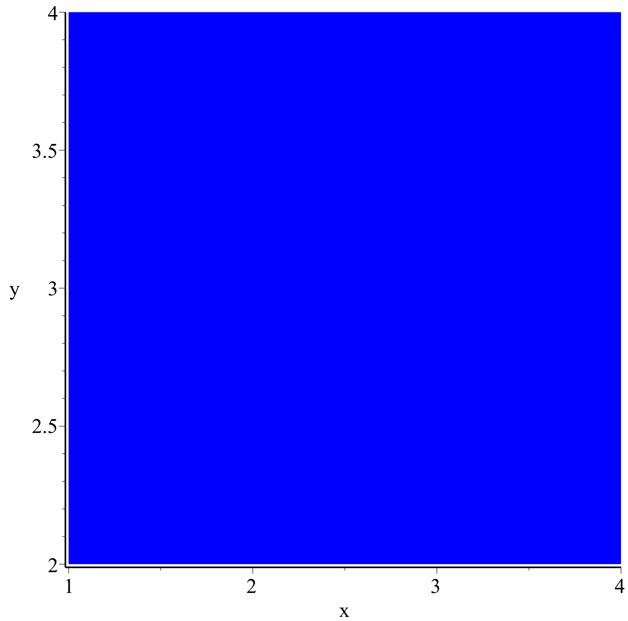

```
> plots[inequal]({x + y < 4, x + 2*y >= 3}, x = 0..5, y = -2..5,
  optionsfeasible = [color = red, transparency = 0.5],
  optionsexcluded = [color = green, transparency = 0.5],
  optionsopen = [color = blue, thickness = 3, linestyle = dash]);
```



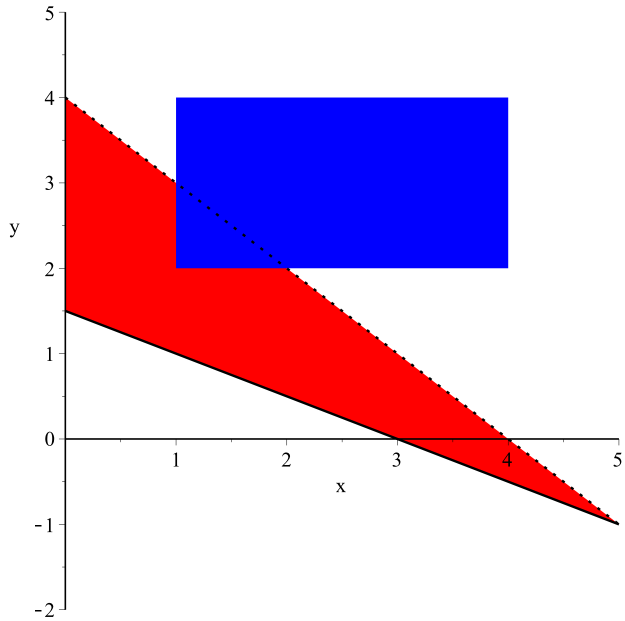

```
> plots[inequal]({x + y < 4, x + 2*y >= 3}, x = 0..5, y = -2..5,
  optionsfeasible = [color = red, transparency = 0.5],
  optionsexcluded = [color = green, transparency = 0.5],
  optionsopen = [color = blue, thickness = 3, linestyle = dash],
  optionsclosed = [color = red, thickness = 3]);
```




```
> P2 := plots[inequal]({}, x = 1..4, y = 2..4, optionsfeasible =
  [color = blue, transparency = 0.5]): P2;
```




```
> plots[display]([P2, P1]);
```



Várható érték, szórás

Elméleti összefoglaló

A valószínségi változóval számszersített kísérleti eredmény számértékként jelenik meg a számegyenesen. Sokszori elvégzéskor sok pont –egy ponthalmaz– keletkezik. Ennek szerkezetét teljes részletességgel a valószínségi változó eloszlása írja le, ám ez adott esetben akár sok (esetleg végtelen sok) értékbl állhat. Ugyanakkor bármilyen nagy adathalmaz (valószínségi eloszlás) alapvet jellege elég jól megragadható két egyszer mérszámmal: a várható értékkel és a szórással.
Els megközelítésben fontosnak és szemléletesnek tn adat, hogy mi a ponthalmaz "közepe", vagyis az átlagos érték, ill. hogy ezen érték körül mennyire "terül szét", vagyis mennyire szóródik a többi pont. Ez a két adat a várható érték és a szórás; mindkett egy-egy egyszer mérszámmal jellemezhet. Ahhoz képest, hogy a háttérben igen változatos kísérleti konstrukciók és valószínségi modellek állhatnak, gyakorlati szempontból ez a két egyszer szám els közelítésben meglepen jól jellemzi az eloszlás szerkezetét.
Rögtön szögezzük le, hogy a várható érték nem a *legvalószínbb* értéket jelenti, inkább az "átlag" fogalmának a valószínségelméleti megfelelje! Nem kizárt esetenként, hogy az átlagos és a legvalószínbb érték egybeesik, de gyakran nem így van, st, diszkrét változónál az is gyakori,

hogy az átlagos érték el sem állhat: nincs a változó értékkészletében.
Egy kísérletsorozatban a valószínségi változó átlagos értéke az összes (akár ismétlden) elforduló érték átlaga. Könnyen megmutatható, hogy ez a lehetséges értékeknek a relatív gyakoriságokkal - elméleti határesetben pedig az azok határértékeként adódó valószínségekkel súlyozott átlaga. Ha ugyanis az összeget $x_1, x_2 \ldots$ értékenként csoportosítva képezzük, akkor a teljes összeg $x_1 \cdot k_1 + x_2 \cdot k_2 + \ldots$, ahol $k_1, k_2 \ldots$ az egyes értékek elfordulási darabszámai. Az összeget elosztva az összes érték darabszámával kapjuk az átlagot:
$\overline{X} = \dfrac{x_1 \cdot k_1 + x_2 \cdot k_2 + \ldots}{n} = x_1 \cdot \dfrac{k_1}{n} + x_2 \cdot \dfrac{k_2}{n} + \ldots$, a $\dfrac{k_i}{n}$ értékek pedig elméleti határesetben az egyes értékek elfordulási valószínségeihez tartanak: $\mathrm{E}(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots$
Ugyanaz az átlag különféleképp szóródó értékekbl is kijöhet. Ha egy céltáblának átellenes oldalait találjuk el lövésekkel, akkor is megvan átlagban a közepe, mégis jobb teljesítménynek számít, ha inkább mindig a közepét találjuk el. Az átlagtól való eltérés mérszáma a szórás. Ez definíció szerint a várható értéktl való eltérések négyzete várható értékének négyzetgyöke. (Hogy miért ez a célszer definíció, és miért nem elegend mondjuk egyszeren a különbségek abszolútértékének átlagát venni, az mélyebb meggondolásokat igényel, ezért itt nem részletezzük. Azt is szoktuk mondani, hogy elször *szórásnégyzetet* vagy más szóval *varianciát* számolunk:
$\mathrm{Var}(X) = \sigma_X^2 = \mathrm{E}\big((X - \mathrm{E}(X))^2\big) = \big(x_1 - \mathrm{E}(X)\big)^2 \cdot p_1 + \big(x_2 - \mathrm{E}(X)\big)^2 \cdot p_2 + \ldots$
A hisztogram→sűrűségfüggvény átmenet alapján megtehető –itt most nem részletezendő– meggondolások alapján a várható értéknek a diszkrét valószínségi változóknál még szummával felírt definíciója –a szummázás határértékeként– folytonos valószínűségi változóknál integrálba megy át:

$$\mathrm{E}(X) = \sum_{i=1}^{n} x_i \cdot p_i \rightarrow E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, \mathrm{d}x$$

A várható értékhez hasonlóan a variancia diszkrét változóknál még szummaként felírt definíciója integrállá alakul:

$$\mathrm{Var}(X) = \sigma_X^2 = \sum_{i=1}^{n} \big(x_i - \mathrm{E}(X)\big)^2 \cdot p_i \rightarrow Var(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^2 \cdot f(x)\, \mathrm{d}x$$

A szórás ennek négyzetgyöke: $\sigma_X = \sqrt{\mathrm{Var}(x)}$.

Kidolgozott feladatok

A *palcikadiagram* eljárás várható érték és szórás szemléltetésére

Ha adott egy Y diszkrét valószínségi változó y értékekkel és a hozzájuk tartozó py valószínségekkel, akkor az alábbi eljárás pálcikadiagramon ábrázolja az eloszlást, valamint szemlélteti a várható értéket és a szórást.

```
> palcikadiagram := proc (y, py, mu, sigma)
    local db, palcikaplot, varhatopont, szoraspont,
  varhatoplot, szorasplot;
    db := nops(y);
    palcikaplot := DynamicSystems[DiscretePlot](y, py, style
  = stem, color = blue, thickness = 3);
    varhatopont := [[mu, 0]];
    szoraspont := [[mu-sigma, 0], [mu+sigma, 0]];
```

```
  varhatoplot := plot(varhatopont, style = point, symbol =
soliddiamond, symbolsize = 20, color = red);
  szorasplot := plot(szoraspont, style = point, symbol =
diamond, symbolsize = 20, color = red);
  plots[display](palcikaplot, varhatoplot, szorasplot);
end proc:
```

1. Feladat (Kockadobás / 2.)*

Dobjunk egy (6 oldalú) dobókockával, jelölje X a dobott számot. Az elz feladatsorban felírtuk az X valószínségi változó eloszlását és szemléltettük azt pálcikadiagramon.
d) Határozzuk meg X várható értékét, varianciáját és szórását. Szemléltessük ezeket az a) pontban kapott pálcikadiagramon!
e) Írjunk Maple-eljárást a várható érték, variancia és szórás kiszámítására!
f) Számítsuk ki, hogy mekkora valószínséggel esik X a várható érték szórás sugarú körzetébe.

Megoldás

```
> restart;
```

A kockadobás kísérlet közvetlenül egy 1-6 értékek közti számszer eredményt ad. Az X valószínségi változó lehetséges értékeit egy listában rögzítjük:

```
> x := [1, 2, 3, 4, 5, 6];
```

$$x := [1, 2, 3, 4, 5, 6] \qquad \textbf{(2.2.2.1.1)}$$

Az egyes értékek bekövetkezési valószínségeit szintén listában rögzíthetjük.

```
> px := [1/6, 1/6, 1/6, 1/6, 1/6, 1/6];
```

$$px := \left[\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right] \qquad \textbf{(2.2.2.1.2)}$$

Az X diszkrét valószínségi változó eloszlása a lehetséges x_i értékekhez azok bekövetkezési valószínségét ($p_i = P(X = x_i)$) rendel függvény, melyet az alábbi táblázattal szemléltethetünk:

```
> eloszlas_X = linalg[augment](matrix(2, 1, ['x_i', 'p_i']), matrix(2, 6, [x, px])); # a linalg csomag augment eljárása két mátrix összeragasztására használható
```

$$eloszlas\_X = \begin{bmatrix} x\_i & 1 & 2 & 3 & 4 & 5 & 6 \\ p\_i & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \qquad \textbf{(2.2.2.1.3)}$$

d) Határozzuk meg X várható értékét, varianciáját és szórását. Szemléltessük ezeket az a) pontban kapott pálcikadiagramon!

Az X valószínségi változó várható értéke az x_i értékek p_i valószínségekkel súlyozott átlaga:

```
> db := nops(x);
  EX := sum(x[i]*px[i], i = 1..db); evalf(%);
```

$$db := 6$$

$$EX := \frac{7}{2}$$

$$3.500000000 \qquad \textbf{(2.2.2.1.4)}$$

A kockadobás várható értéke tehát 3.5, ami az egyenl valószínséggel bekövetkez

1, 2, 3, 4, 5, 6 értékek átlaga, épp félúton az 1 és 6 értékek között. Érdekes módon ez az átlag nincs a változó lehetséges értékei között, vagyis 0 a gyakorisága!
Következre számoljuk ki a varianciát!

```
> VarX := sum((x[i] - EX)^2*px[i], i = 1..db); evalf(%);
```

$$VarX := \frac{35}{12}$$

$$2.916666667 \quad (2.2.2.1.5)$$

Ennek négyzetgyöke a szórás:

```
> sigmaX := sqrt(VarX); evalf(%);
```

$$sigmaX := \frac{1}{6}\sqrt{105}$$

$$1.707825129 \quad (2.2.2.1.6)$$

A variancia meghatározható a tanult $Var(X) = E(X^2) - E(X)^2$ képlettel is, mely a gyakorlatban sokszor jobbal használható, mint a definíció. Ehhez elbb meg kell határozni $E(X^2)$-et!

```
> EX2 := sum(x[i]^2*px[i], i = 1..db); evalf(%);
```

$$EX2 := \frac{91}{6}$$

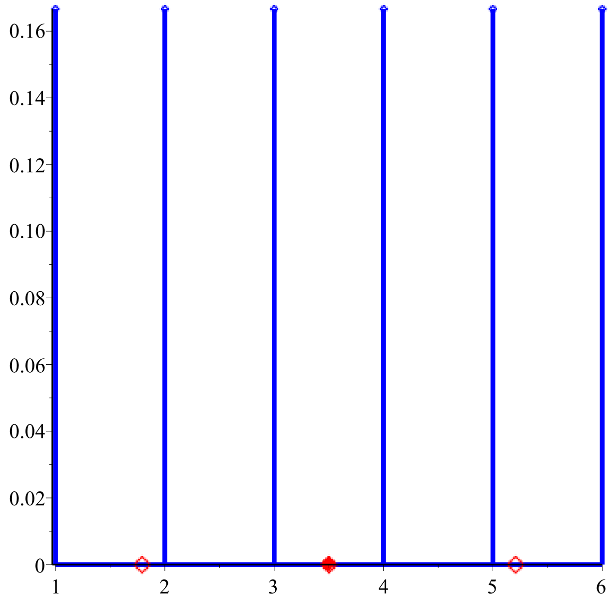
$$15.16666667 \quad (2.2.2.1.7)$$

```
> VarX = EX2-EX^2;
```

$$\frac{35}{12} = \frac{35}{12} \quad (2.2.2.1.8)$$

Az a) pontban készített pálcikadiagramot kiegészíthetjük úgy, hogy a várható értéket és a szórást is megjelenítjük. Ehhez a munkalap elején definiált *palcikadiagram* eljárást használhatjuk!

```
> palcikadiagram(x, px, EX, sigmaX);
```



e) Írjunk Maple-eljárást a várható érték, variancia és szórás kiszámítására!

```
> varhato_ertek := proc(y, py)
    local db;
    db := nops(y);
    return sum(y[i]*py[i], i = 1..db);
  end proc:
> variancia := proc(y, py)
    local varhato, db;
    db := nops(y);
    varhato := varhato_ertek(y, py);
    return sum((y[i] - varhato)^2*py[i], i = 1..db);
  end proc:
> szoras := proc(y, py)
    return sqrt(variancia(y, py));
  end proc:
```

Próbáljuk is ki ezeket!

```
> varhato_ertek(x, px);
  variancia(x, px);
  szoras(x, px);
```

$$\frac{7}{2}$$

$$\frac{35}{12}$$

$$\frac{1}{6}\sqrt{105} \qquad \textbf{(2.2.2.1.9)}$$

f) Számítsuk ki, hogy mekkora valószínséggel esik X a várható érték szórás sugarú körzetébe.

A kérdés a $P\big(E(X)-\sigma(X)\le X\le E(X)+\sigma(X)\big)$ valószínségre vonatkozik. Ez a fenti diagramon a két üres belsej jelöl közötti oszlopok magasságainak összege.

```
> also := evalf(EX - sigmaX); # itt lebegpontosan ki kell értékelni az eredményt, különben nem számként lesz kezelve
  felso := evalf(EX + sigmaX);
```

$$also := 1.792174871$$

$$felso := 5.207825129 \qquad \textbf{(2.2.2.1.10)}$$

```
> osszeg := 0:
  for i from 1 to db do
    if x[i] >= also and x[i] <= felso then
      osszeg := osszeg + px[i];
    end if;
  end do;
  `P(EX - sigmaX <= X <= EX + sigmaX)` := osszeg;
```

$$P(EX - sigmaX <= X <= EX + sigmaX) := \frac{2}{3} \qquad \textbf{(2.2.2.1.11)}$$

Tehát $\frac{2}{3}$ az esélye annak, hogy X a várható érték szórás sugarú körzetébe esik, $\frac{1}{3}$ eséllyel viszont a szórásnál nagyobb mértékben tér el a várható értéktl!

Ezt a Statistics csomag segítségével is megkaphatjuk, ha definiáljuk az X-nek megfelel, diszkrét egyenletes eloszlású valószínségi változót.

```
> with(Statistics):
> X := RandomVariable(DiscreteUniform(1, 6));
```

$$X := \_R \qquad \textbf{(2.2.2.1.12)}$$

```
> Probability({EX - sigmaX <= X, X <= EX + sigmaX});
```

$$\frac{2}{3} \qquad \textbf{(2.2.2.1.13)}$$

Az eloszlásfüggvénybl is kiszámolható a keresett valószínség:

```
> F := x -> CDF(X, x); # X eloszlásfüggvénye
```

$$F := x \rightarrow Statistics\text{:-}CDF(X, x) \qquad \textbf{(2.2.2.1.14)}$$

```
> F(EX + sigmaX) - limit(F(y), y = EX - sigmaX, left);
```

$$\frac{2}{3} \qquad \textbf{(2.2.2.1.15)}$$

Megjegyzés: a Maple az eloszlásfüggvényt az általunk tanult $F(x)=P(X<x)$ képlettl eltéren az $F(x)=P(X\le x)$ képlettel definiálja, ezért kellett bal oldali határértéket számítani a kivonandónál:

$$P\big(E(X)-\sigma(X)\le X\le E(X)+\sigma(X)\big)=P\big(X\le E(X)+\sigma(X)\big)-P\big(X<E(X)-\sigma(X)\big)=F\big(E(X)+\sigma(X)\big)-\lim_{y\to (E(X)-\sigma(X))^-}F(y)$$

.

```
> Mean(X);
  ExpectedValue(X);
```

$$\frac{7}{2}$$

$$\frac{7}{2} \tag{2.2.2.1.16}$$

```
> Variance(X);
  StandardDeviation(X);
```

$$\frac{35}{12}$$

$$\frac{1}{6}\sqrt{105} \tag{2.2.2.1.17}$$

2. Feladat (Dobás két kockával / 2.)

Dobjunk két dobókockával egyszerre! Az *Y* valószínségi változóban jegyezzük fel a két dobott érték maximumát!
e) Számítsuk ki a várható értéket, a varianciát és a szórást! Jelöljük be az eloszlást ábrázoló pálcikadiagramon a várható értéket és a szórást!
f) Számítsuk ki, hogy mekkora valószínséggel lesz *Y* kisebb a várható értékénél!

Megoldás

```
> restart;
```

A változó értékkészlete 1-t1 6-ig terjed - a feljegyzett maximum 1 is lehet, ha mindkét dobás 1-es, de persze ez viszonylag ritka érték lesz.

```
> y := [1, 2, 3, 4, 5, 6];
```

$$y := [1, 2, 3, 4, 5, 6] \tag{2.2.3.1.1}$$

A valószínségi változó értékeihez tartozó valószínségeket meghatároztuk az elz feladatsorban:

```
> py := [1/36, 3/36, 5/36, 7/36, 9/36, 11/36];
```

$$py := \left[\frac{1}{36}, \frac{1}{12}, \frac{5}{36}, \frac{7}{36}, \frac{1}{4}, \frac{11}{36}\right] \tag{2.2.3.1.2}$$

e) Számítsuk ki a várható értéket, a varianciát és a szórást! Jelöljük be az eloszlást ábrázoló pálcikadiagramon a várható értéket és a szórást!

```
> db := nops(y);
  EY := sum(y[i]*py[i], i = 1..db); evalf(%);
```

$$db := 6$$

$$EY := \frac{161}{36}$$

$$4.472222222 \tag{2.2.3.1.3}$$

Látható, hogy a maximumok átlagos értéke az értéktartomány fels széléhez van közelebb. Ez annak tudható be, hogy a magasabb maximum-értékek elfordulási valószínsége nagyobb, mivel többféle módon tudnak elállni.
Számítsuk ki a varianciát és a szórást is!

```
> VarY := sum((y[i] - EY)^2*py[i], i = 1..db); evalf(%);
```

$$VarY := \frac{2555}{1296}$$

$$1.971450617 \tag{2.2.3.1.4}$$

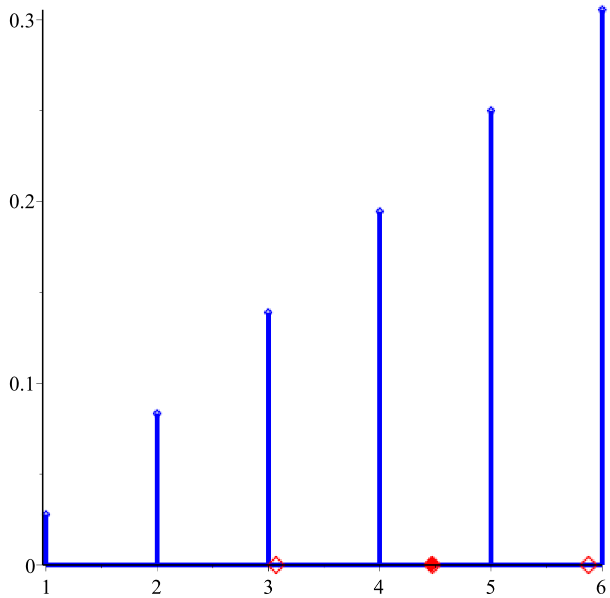
```
> sigmaY := sqrt(VarY); evalf(%);
```

$$sigmaY := \frac{1}{36}\sqrt{2555}$$

$$1.404083551 \qquad \textbf{(2.2.3.1.5)}$$

Ábrázoljuk pálcikadiagramon az eloszlást, bejelölve a várható értéket és a szórást!

```
> palcikadiagram(y, py, EY, sigmaY);
```



Szépen láthatók az egyenletesen növekv valószínségek és a várható érték "jobbra húzása".

f) Számítsuk ki, hogy mekkora valószínséggel lesz *X* kisebb a várható értékénél!

A $P(X < E(X))$ valószínség a kérdés. Ez a fenti diagramon a tömött piros jelöltl balra lév oszlopok magasságainak összege.

```
> osszeg := 0;
  for i from 1 to db do
    if y[i] < EY then
      osszeg := osszeg + py[i];
    end if;
  end do;
  `P(X < EX)` := osszeg;
```

$$osszeg := 0$$

$$P(X < EX) := \frac{4}{9} \qquad \textbf{(2.2.3.1.6)}$$

Tehát $\frac{4}{9}$ az esélye annak, hogy X kisebb a várható értékénél. Ez is mutatja, hogy a várható érték (általában) nem fele-fele arányban osztja fel a valószínűségeket – ezt a szerepet egy másik statisztikai mutató, a *medián* tölti be. A kett eltérése pedig az eloszlás *ferdeség*ére utal, ami a pálcikadiagramon itt jól látszik.

3. Feladat (Nyeremény a csokiban)

Egy akció keretében azt hirdetik, hogy minden negyedik csokiban nyerkódot találunk. Veszünk tíz darab csokit.
Legyen Y az a valószínségi változó, mely azt mutatja, hogy a vásárolt csokik közül hányban van nyerkód.
a) Adjuk meg a valószínségi változó értékkészletét és az értékekhez tartozó valószínségeket!
b) Számítsuk ki a várható értéket és a szórást!
c) Ábrázoljuk az eloszlást az Y-p síkban pálcikadiagramon, szemléltessük a várható értéket és a szórást!
d) Végezzünk szimulációt 1000 kísérlettel! Készítsük el a kapott minta hisztogramját és vessük össze a c) részben kapott pálcikadiagrammal!
e) Rajzoljuk meg az eloszlásfüggvényt!

Megoldás

```
> restart;
```

a) Adjuk meg a valószínségi változó értékkészletét és az értékekhez tartozó valószínségeket!
A 10 vásárolt csoki közül elvileg 0 és 10 között akárhány nyer lehet. A változó értékkészletét kézzel beírva is megadhatnánk, de nagyobb elemszámok esetére érdemes sorozatként generálni:

```
> n := 10;
```

$$n := 10 \tag{2.2.4.1.1}$$

```
> y := [seq(i, i = 0..n)];
```

$$y := [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10] \tag{2.2.4.1.2}$$

A binomiális eloszlás tipikus esetével van dolgunk: minden csoki kibontása független, kétállapotú, adott valószínséggel sikerrel járó elemi kísérlet, és adott számú sikeres kísérlet valószínségét kérdezzük. Ez egy $n = 10$, $p = \frac{1}{4}$ paraméter binomiális eloszlásra vezet (a binomiális eloszlás ismeretét feltételezve annak számítását itt nem részletezzük).

```
> p := 1/4;
```

$$p := \frac{1}{4} \tag{2.2.4.1.3}$$

```
> py := [seq(binomial(n, i)*p^i*(1-p)^(n-i), i = 0..n)];
  evalf(%);
```

$$py := \left[\frac{59049}{1048576}, \frac{98415}{524288}, \frac{295245}{1048576}, \frac{32805}{131072}, \frac{76545}{524288}, \frac{15309}{262144}, \frac{8505}{524288}, \frac{405}{131072}, \frac{405}{1048576}, \frac{15}{524288}, \frac{1}{1048576}\right]$$

$$[0.05631351471, 0.1877117157, 0.2815675735, 0.2502822876, 0.1459980011, 0.05839920044, 0.01622200012, 0.003089904785, \tag{2.2.4.1.4}$$

0.0003862380981, 0.00002861022949, 9.536743164 10^{-7}]

Táblázatosan összefoglalva:

```
> linalg[augment](matrix(2, 1, ['x_i', 'p_i']), matrix(2, 11, [y, evalf(py, 3)])); # a linalg csomag augment eljárása két mátrix összeragasztására használható
```

[[x_i, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10], **(2.2.4.1.5)**

[p_i, 0.0563, 0.188, 0.282, 0.250, 0.146, 0.0584, 0.0162, 0.00309, 0.000386, 0.0000286, 9.54 10^{-7}]]

b) Számítsuk ki a várható értéket és a szórást!

```
> db := 11;
EY := sum(y[i]*py[i], i = 1..db); evalf(%);
```

$$db := 11$$

$$EY := \frac{5}{2}$$

$$2.500000000 \qquad \textbf{(2.2.4.1.6)}$$

Vegyük észre, hogy ez az eredmény teljesen egybevág a szemléletünkkel: ha átlagosan 4 csokiból 1 nyer, akkor arányosan 10-b1 2.5 nyer van, mivel a 10-ben 2.5-ször van meg a 4.

A szórást közvetlenül is számolhatjuk, a variancia külön meghatározása nélkül:

```
> sigmaY := sqrt(sum((y[i] - EY)^2*py[i], i = 1..db));
evalf(%);
```

$$sigmaY := \frac{1}{4}\sqrt{30}$$

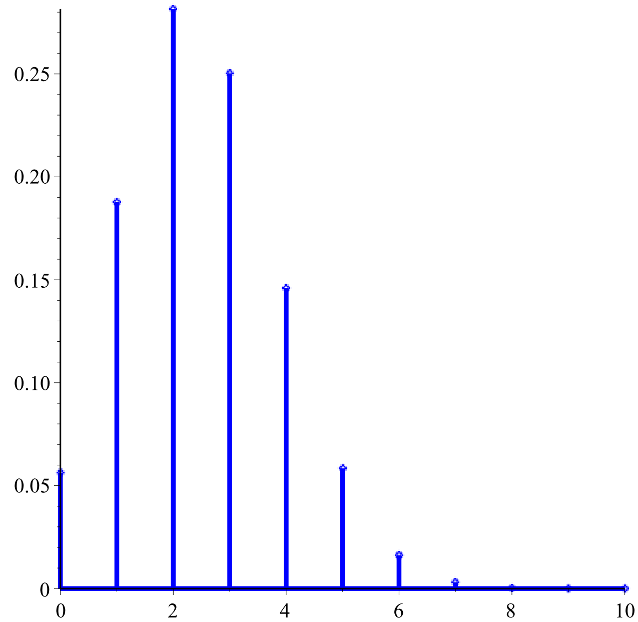
$$1.369306394 \qquad \textbf{(2.2.4.1.7)}$$

Ez nagyjából azt jelenti, hogy a 2.5-ös átlagtól átlagosan ~1.4-et tér el a tíz csoki között a nyerk száma.

c) Ábrázoljuk az eloszlást az Y - p síkban pálcikadiagramon, szemléltessük a várható értéket és a szórást!

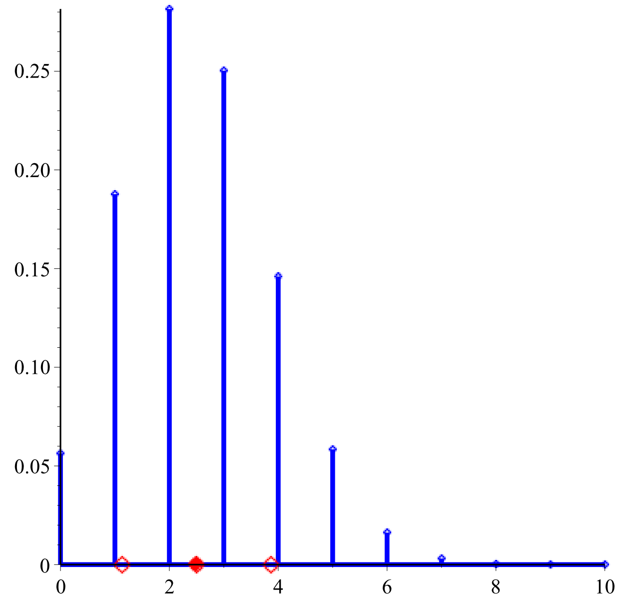
A pálcikadiagram:

```
> with(DynamicSystems):
P1 := DiscretePlot(y, py, style = stem, color = blue, thickness = 3): P1;
```



Ugyanez várható értékkel és szórással:

```
> palcikadiagram(y, py, EY, sigmaY);
```



Az látszik, hogy a nyer csokik darabszáma túlnyomó valószínséggel az 1, 2, 3 vagy 4 számok közül kerül ki. Ezen tartományon kívüli érték viszonylag ritkán fordul el.
Mivel nevezetes eloszlásról (binomiális) van szó, a várható értéket és a szórást egyszeren meghatározhatjuk a *Statistics* csomag segítségével is!

```
> with(Statistics):
  randomize():
```

```
> Y := RandomVariable(Binomial(10, 1/4));
```

$$Y := \_R \tag{2.2.4.1.8}$$

```
> ExpectedValue(Y); # Mean(Y) is mködik
```

$$\frac{5}{2} \tag{2.2.4.1.9}$$

```
> StandardDeviation(Y);
```

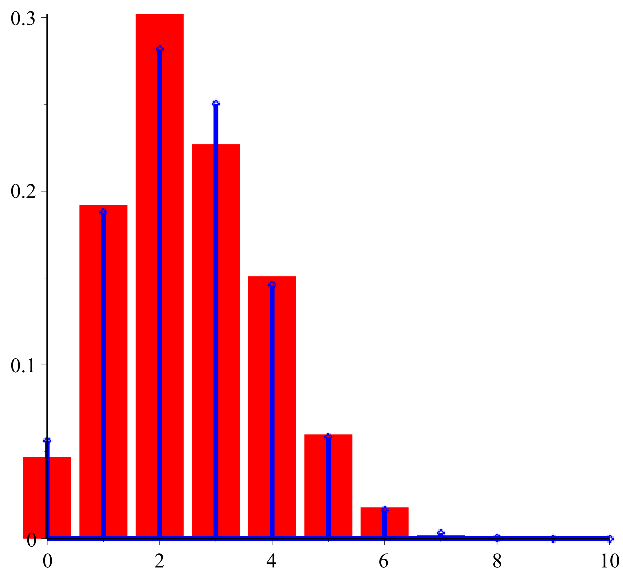
$$\frac{1}{4}\sqrt{30} \tag{2.2.4.1.10}$$

d) Végezzünk szimulációt 1000 kísérlettel! Készítsük el a kapott minta hisztogramját és vessük össze a b) részben kapott pálcikadiagrammal!
Használjuk az elbb definiált binomiális eloszlású *Y* valószínségi változót!
Vegyünk 1000 elem mintát *Y*-ból és ábrázoljuk hisztogramon.

```
> minta := Sample(Y, 1000);
```

$$minta := \begin{bmatrix} \text{1 .. 1000 } Vector_{row} \\ \text{Data Type: } float_8 \\ \text{Storage: rectangular} \\ \text{Order: Fortran\_order} \end{bmatrix} \quad \textbf{(2.2.4.1.11)}$$

```
> H := Histogram(minta, discrete=true, thickness=30,
  color=red): # diszkrét eloszlás esetén fontos a
  "discrete" opciót "true"-ra állítani!
  plots[display](H, P1); # piros a hisztogram, kék az
  elméleti eloszlás
```



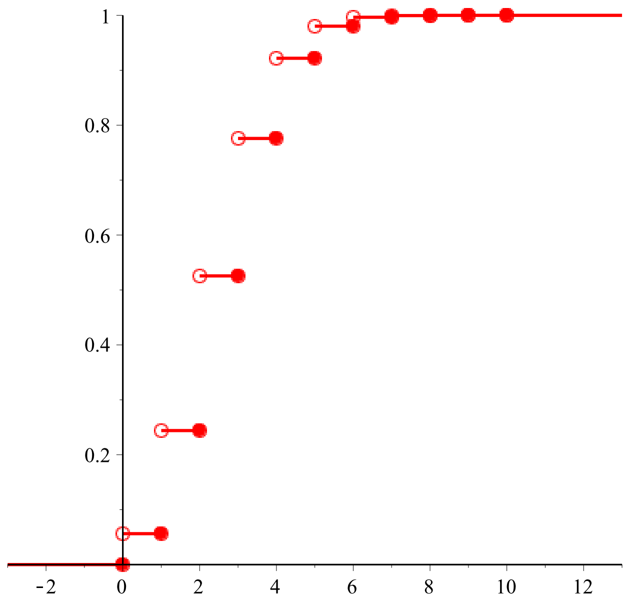
Láthatjuk, hogy –bár kisebb kilengések megfigyelhetők– a minta eloszlása összességében jól illeszkedik az elméleti eloszlásra. A mintaelemszám növelésével az illeszkedés is egyre jobb lesz.

e) Rajzoljuk meg az eloszlásfüggvényt!

Egy adott valószínségi változó eloszlásfüggvénye az a teljes valós számegyenesen értelmezett $F(x)$ függvény, melynek értéke minden x helyen azt a valószínséget mutatja, mellyel a valószínségi változó értéke kisebb x-nél, azaz $F(x) = P(X < x)$. Ez diszkrét esetben mindig egy balról folytonos lépcss függvény lesz.

Használjuk újra a múlt heti feladatsorban definiált eljárást!

```
> diszkret_eloszlasfuggveny := proc (x, px)
    local db, szintek, szakaszok, szakaszok_plot,
  elso_szakasz, elso_szakasz_plot, utolso_szakasz,
  utolso_szakasz_plot, jobb_vegpontok,
  jobb_vegpontok_plot, bal_vegpontok, bal_vegpontok_plot;
    db := nops(x);
    szintek := [seq(sum(px[i], i = 1..k), k = 1..db)];
    szakaszok := [seq([[x[i], szintek[i]], [x[i + 1],
  szintek[i]]], i = 1..db - 1)];
    szakaszok_plot := plot(szakaszok, x[1] - 3..x[db] + 3,
  color = red, thickness = 2);
    elso_szakasz := [[x[1] - 3, 0], [x[1], 0]];
    elso_szakasz_plot := plot(elso_szakasz, x[1] - 3..x
  [db] + 3, color = red, thickness = 2);
    utolso_szakasz := [[x[db], 1], [x[db] + 3, 1]];
    utolso_szakasz_plot := plot(utolso_szakasz, x[1] - 3..
  x[db] + 3, color = red, thickness = 2);
    jobb_vegpontok := [[x[1], 0], seq([x[i + 1], szintek
  [i]], i = 1..db - 1)];
    jobb_vegpontok_plot := plots[pointplot]
  (jobb_vegpontok, symbol = solidcircle, symbolsize = 18,
  color = red); # tegyünk minden szakasz jobb végére
  tömött karikát
    bal_vegpontok := [seq([x[i], szintek[i]], i = 1..db)];
    bal_vegpontok_plot := plots[pointplot](bal_vegpontok,
  symbol = circle, symbolsize = 18, color = red); #
  tegyünk minden szakasz bal végére üres karikát
    plots[display](szakaszok_plot, elso_szakasz_plot,
  utolso_szakasz_plot, jobb_vegpontok_plot,
  bal_vegpontok_plot);
  end proc:
> diszkret_eloszlasfuggveny(y, py);
```

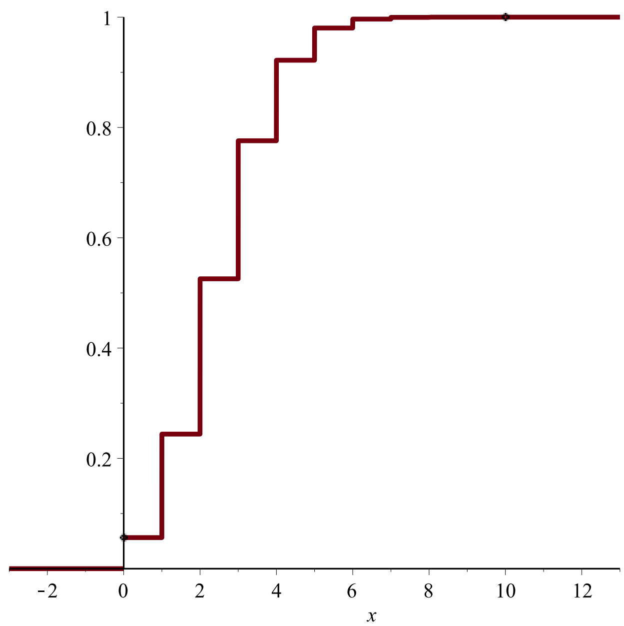


Mint minden eloszlásfüggvény, ez is 0-ról 1-re kúszik fel monoton módon, és a diszkrét eloszlásokra jellemz lépcszetességet mutatja. A lépcsk ott a legmagasabbak, ahol legnagyobbak a valószínségek (legmagasabbak a pálcikák a fenti pálcikadiagramon).
Az eloszlásfüggvényt a Statistics csomaggal is elkészíthetjük, amihez a korábban definiált Y valószínségi változót használhatjuk fel.

```
> cdf_Y := CDF(Y, x);
```

$$cdf\_Y := \begin{cases} 0 & x < 0 \\ 1 & 10 \leq x \\ \sum_{i=0}^{\text{floor}(x)} \text{binomial}(10, i) \left(\frac{1}{4}\right)^{i} \left(\frac{3}{4}\right)^{10-i} & \mathit{otherwise} \end{cases} \quad \textbf{(2.2.4.1.12)}$$

```
> plot(cdf_Y, x=-3..13, thickness = 3, discont = true);
```



A harmadik lehetség, ha elkészítjük a kumulált valószínségek listáját, majd a *DynamicSystems DiscretePlot* eljárásával ábrázoljuk az eloszlásfüggvényt:

```
> N := nops(y);
  Fy := [seq(sum(py[i], i = 1..n), n = 1..N)];
```

$$N := 11$$

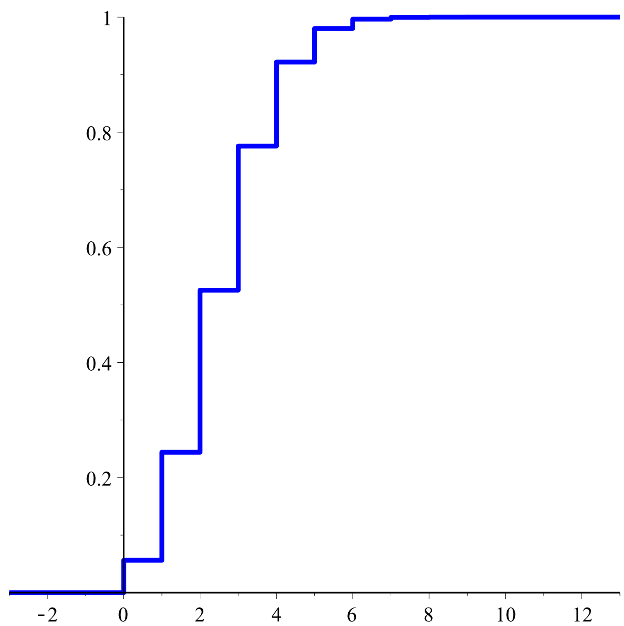
$$Fy := \left[\frac{59049}{1048576}, \frac{255879}{1048576}, \frac{137781}{262144}, \frac{203391}{262144}, \frac{483327}{524288}, \frac{513945}{524288}, \frac{261225}{262144}, \frac{262035}{262144}, \frac{1048545}{1048576}, \frac{1048575}{1048576}, 1\right] \quad \textbf{(2.2.4.1.13)}$$

```
> yy := [y[1]-3, op(y), y[N]+3];
  Fyy := [0, op(Fy), 1]; # az op() paranccsal sorozattá
  alakíthatunk egy listát, ami ezután könnyen
  beilleszthet egy vesszkkel elválasztott felsorolásba
```

$$yy := [-3, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13]$$

$$Fyy := \left[0, \frac{59049}{1048576}, \frac{255879}{1048576}, \frac{137781}{262144}, \frac{203391}{262144}, \frac{483327}{524288}, \frac{513945}{524288}, \frac{261225}{262144}, \frac{262035}{262144}, \frac{1048545}{1048576}, \frac{1048575}{1048576}, 1, 1\right] \quad \textbf{(2.2.4.1.14)}$$

```
> DiscretePlot(yy, Fyy, style = stair, color = blue,
  thickness = 3);
```



4. Feladat (Egyenletes eloszlás / 2.)

Válasszunk ki véletlenszeren (geometriai valószínség szerint) egy *Y* számot a $[0, 1]$ intervallumból!
f) Számítsuk ki *Y* várható értékét, varianciáját és szórását!
g) Mekkora valószínséggel lesz *Y* várható értéktl való eltérése nagyobb, mint a szórás?

Megoldás

```
> restart;
```

A véletlenszer választás azt jelenti, hogy Y a $[0, 1]$ intervallumon egyenletes eloszlású valószínségi változó.
Hozzuk létre a követelménynek megfelel véletlen változót!

```
> with(Statistics):
  randomize():
> Y := RandomVariable(Uniform(0, 1));
```

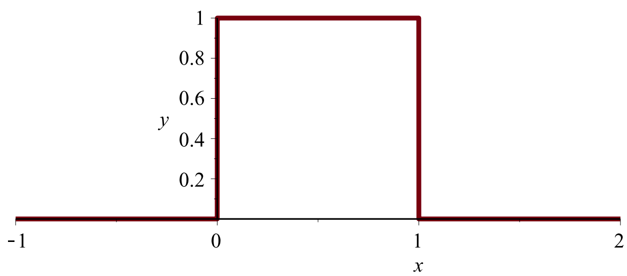
$$Y := \_R \quad \textbf{(2.2.5.1.1)}$$

Ennek srségfüggvényét az elz feladatsorban meghatároztuk:

```
> f := x -> piecewise(x <= 0, 0, x <= 1, 1, 0);
  plot(f(x), x = -1..2, y = 0..1, scaling = constrained,
```

```
thickness = 3);
```

$$f := x \rightarrow piecewise(x \leq 0, 0, x \leq 1, 1, 0)$$



f) Számítsuk ki *Y* várható értékét, varianciáját és szórását!

A *hisztogram→srségfüggvény* átmenet alapján megtehet, itt most nem részletezend meggondolások alapján a várható értéknek a diszkrét változóknál még szummával felírt definíciója –a szummázás határértékeként– folytonos változóknál integrálba megy át:

$$E(Y) = \sum_{i=1}^{n} y_i \cdot p_i \rightarrow E(Y) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Az integrál kiszámítása a konkrét esetre Maple segítségével:

```
> EY := int(f(x)*x, x = 0..1); evalf(%);
```

$$EY := \frac{1}{2}$$

$$0.5000000000 \qquad \textbf{(2.2.5.1.2)}$$

Szimmetriai okokból érthet, hogy a várható érték az intervallum felezpontja. Ez egyenletes eloszlásnál mindig így van, de általánosságban nem feltétlenül igaz!
A várható értékhez hasonlóan a variancia diszkrét változóknál még szummaként felírt definíciója integrállá alakul:

$$Var(Y) = \sigma^2(Y) = \sum_{i=1}^{n} p_i \cdot \left(y_i - E(Y)\right)^2 \rightarrow Var(Y) = \sigma^2(Y) = \int_{-\infty}^{\infty} f(x) \cdot (x - E(Y))^2 \, dx$$

A szórás ennek négyzetgyöke. A konkrét esetre Maple segítségével:

```
> VarY := int(f(x)*(x-EY)^2, x = 0..1); evalf(%);
```

$$VarY := \frac{1}{12}$$

$$0.08333333333 \qquad (2.2.5.1.3)$$

```
> sigmaY := sqrt(VarY); evalf(%);
```

$$sigmaY := \frac{1}{6}\sqrt{3}$$

$$0.2886751347 \qquad (2.2.5.1.4)$$

Mivel nevezetes eloszlásról (egyenletes) van szó, a várható értéket, a varianciát és a szórást egyszeren meghatározhatjuk a *Statistics* csomag segítségével is! Ehhez a már korábban definiált Y valószínségi változót használjuk.

```
> ExpectedValue(Y);
```

$$\frac{1}{2} \qquad (2.2.5.1.5)$$

```
> Variance(Y);
```

$$\frac{1}{12} \qquad (2.2.5.1.6)$$

```
> StandardDeviation(Y);
```

$$\frac{1}{6}\sqrt{3} \qquad (2.2.5.1.7)$$

g) Mekkora valószínséggel lesz Y várható értéktl való eltérése nagyobb, mint a szórás?

A kérdés a $P\big(Y \le E(Y) - \sigma(Y) \text{ vagy } Y \ge E(Y) + \sigma(Y)\big)$ valószínségre vonatkozik. Egyszerbb azonban az ellentét esemény valószínségét kiszámolni a srségfüggvény integrálásával, majd 1-bl kivonni:

$$1 - P\big(E(Y) - \sigma(Y) \le Y \le E(Y) + \sigma(Y)\big) = 1 - \int_{E(Y)-\sigma(Y)}^{E(Y)+\sigma(Y)} f(x)\, dx$$

```
> also := EY - sigmaY;
  felso := EY + sigmaY;
```

$$also := \frac{1}{2} - \frac{1}{6}\sqrt{3}$$

$$felso := \frac{1}{2} + \frac{1}{6}\sqrt{3} \qquad (2.2.5.1.8)$$

```
> P := 1 - int(f(x), x = also..felso); evalf(%);
```

$$P := 1 - \frac{1}{3}\sqrt{3}$$

$$0.4226497307 \qquad (2.2.5.1.9)$$

Tehát 42.26 % az esélye annak, hogy Y a várható értékétl a szórásnál nagyobb mértékben tér el.

Ez a valószínség a Statistics csomag Probability függvényével is kiszámolható:

```
> Probability(Y <= also) + Probability(Y >= felso); # itt
  kihasználtuk a valószínség additivitását
```

$$1-\frac{1}{3}\sqrt{3} \qquad (2.2.5.1.10)$$

5. Feladat (Nem egyenletes eloszlás / 2.)*

Válasszunk ki véletlenszeren (geometriai valószínség szerint) egy Y számot a $[0, 1]$ intervallumból és emeljük négyzetre: $Z = Y^2$!
a) Ábrázoljuk Z eloszlását pontdiagramon illetve hisztogramon egy 10000 elem minta alapján!
b) Határozzuk meg és ábrázoljuk az Z valószínségi változó eloszlásfüggvényét!
c) Határozzuk meg a Z valószínségi változó srségfüggvényét és vessük össze grafikonját az a) feladatrészben kapott hisztogrammal!
d) Határozzuk meg az eloszlás és srségfüggvényt a Statistics csomag eljárásait használva is!
e) Számítsuk ki kétféleképpen a $P\left(\frac{1}{3} \le Z \le \frac{3}{4}\right)$ valószínséget!
f) Számítsuk ki Z várható értékét, varianciáját és szórását!
g) Mekkora valószínséggel lesz Z nagyobb, mint várható értéke?

Megoldás

```
> restart;
```

a) Ábrázoljuk Z eloszlását pontdiagramon, illetve hisztogramon egy 10000 elem minta alapján!

Definiáljuk a Z valószínségi változót a Statistics csomag eszközeivel!

```
> with(Statistics):
  randomize():
```

```
> Y := RandomVariable(Uniform(0, 1));
```

$$Y := \_R \qquad (2.2.6.1.1)$$

```
> Z := Y^2;
```

$$Z := \_R^2 \qquad (2.2.6.1.2)$$

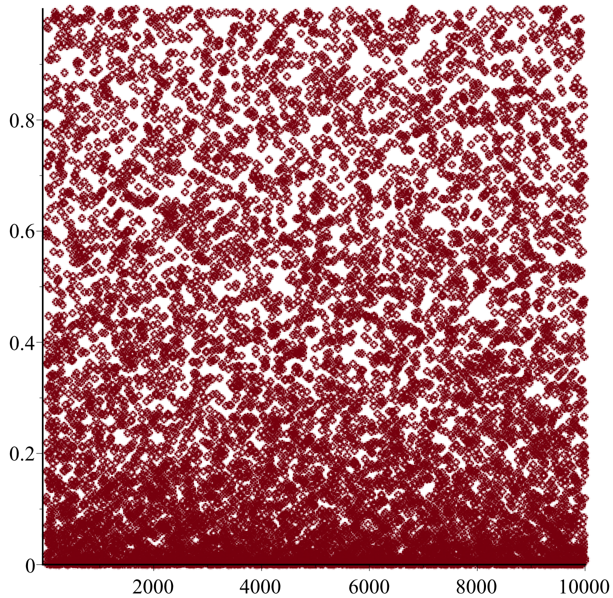
Vegyünk egy 10000 elem mintát Z-bl és ábrázoljuk pontdiagramon!

```
> n := 10000;
  minta := Sample(Z, n);
```

$$n := 10000$$

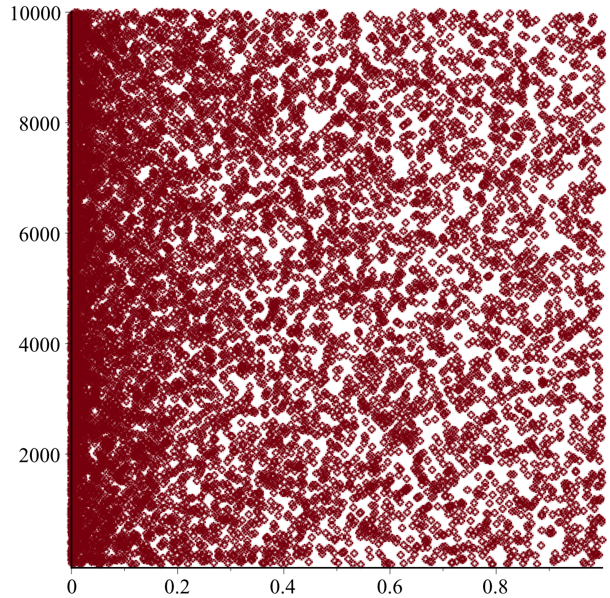
$$minta := \left[\begin{array}{c} \text{1 .. 10000 } Vector_{row} \\ \text{Data Type: } float_8 \\ \text{Storage: rectangular} \\ \text{Order: Fortran\_order} \end{array}\right] \qquad (2.2.6.1.3)$$

```
> pontok := [seq([i, minta[i]], i = 1..n)]:
  plot(pontok, style = point);
```



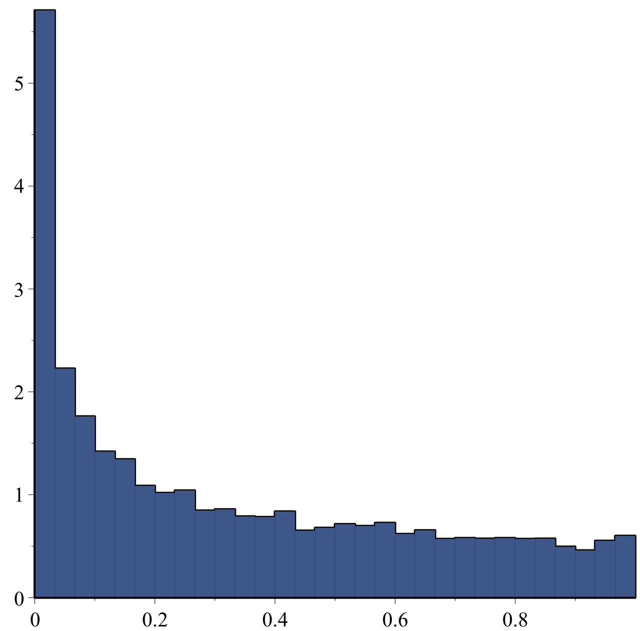
Fordított tengelykiosztással:

```
> pontok_forditott := [seq([minta[i], i], i = 1..n)]:
  plot(pontok_forditott, style = point);
```



Szembetn, hogy bár a kapott értékek most is a $[0, 1]$ intervallumon helyezkednek el a 4. Feladatban bemutatott egyenletes eloszláshoz hasonlóan (hiszen $[0, 1]$-bl választott értékek négyzete is ugyanott van), nem egyenletesen oszlanak el az intervallumon, hanem jól érezheten az alsó végén srsödnek.
Könnyen érthet például, hogy átlagosan az értékek fele most az alsó negyedbe esik, hiszen az egyenletes eloszlás szerint választott, 50 % eséllyel $\frac{1}{2}$ alatti értékek négyzete már $\frac{1}{4}$ alatt lesz.
Végül a hisztogram:

```
> H := Histogram(minta): H;
```



A hisztogramnál most nem adtuk meg explicite a az oszlopok számát (*bincount*), ráhagytuk a Maple-re annak ésszer beállítását.

b) Határozzuk meg és ábrázoljuk az Z valószínségi változó eloszlásfüggvényét!

A c) részben megkíséreljük megállapítani azt az egzakt srségfüggvényt, amit a hisztogram közelít.

Gyakran az eloszlásfüggvény meghatározása a könnyebb - itt is ezzel kezdjük, a srségfüggvény utána már ennek deriválásával adódik.

Feladatunk az $F(x) = P(Z < x)$ valószínség meghatározása tetszleges x valós számra. Emlékezzünk rá, hogy Z-t egy $[0, 1]$-en egyenletes eloszlású Y valószínségi változó négyzeteként kaptuk!

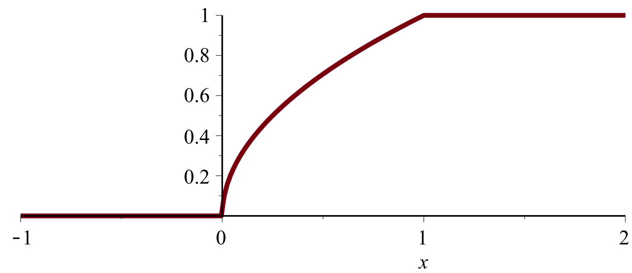
$F(x) = P(Z < x) = P\left(Y^2 < x\right) = P\left(Y < \sqrt{x}\right)$, ha $x \geq 0$. ($x < 0$ esetén könnyen látható, hogy $F(x) = 0$.)

Mivel maga Y egyenletes eloszlású, ezért $F(x) = P\left(Y < \sqrt{x}\right) = \sqrt{x}$ a valószínség geometriai kiszámítási módja alapján.

Ez természetesen a $[0, 1]$ intervallumra vonatkozik, tle balra konstans 0, jobbra konstans 1. A teljes eloszlásfüggvény:

```
> F := x -> piecewise(x <= 0, 0, x <= 1, sqrt(x), 1);
  plot(F(x), x = -1..2, scaling = constrained, thickness =
  3);
```

$$F := x \rightarrow piecewise\left(x \leq 0, 0, x \leq 1, \sqrt{x}, 1\right)$$



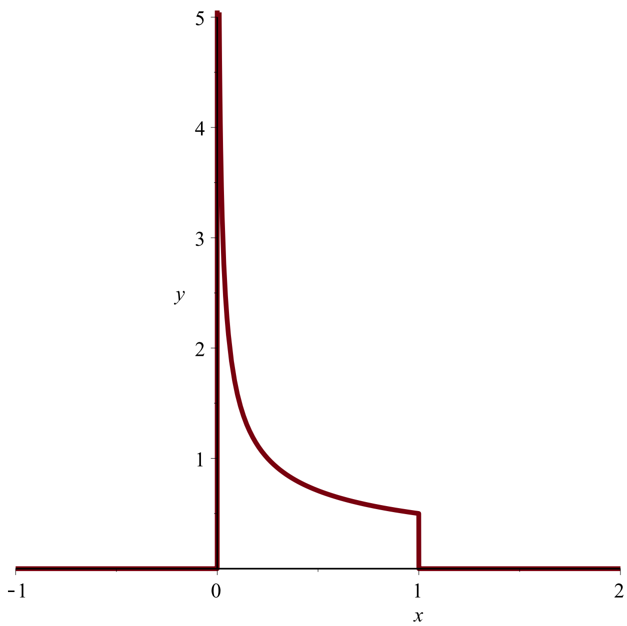
c) Határozzuk meg a *Z* valószínségi változó srségfüggvényét és vessük össze grafikonját az a) feladatrészben kapott hisztogrammal!
Az eloszlásfüggvény deriváltja lesz a srségfüggvény (láthatóan szakaszonként deriválható), melyhez a hisztogram közelít.

```
> f := x -> diff(F(x), x); `f(x)` = f(x);
```

$$f := x \rightarrow \frac{\mathrm{d}}{\mathrm{d}x} F(x)$$

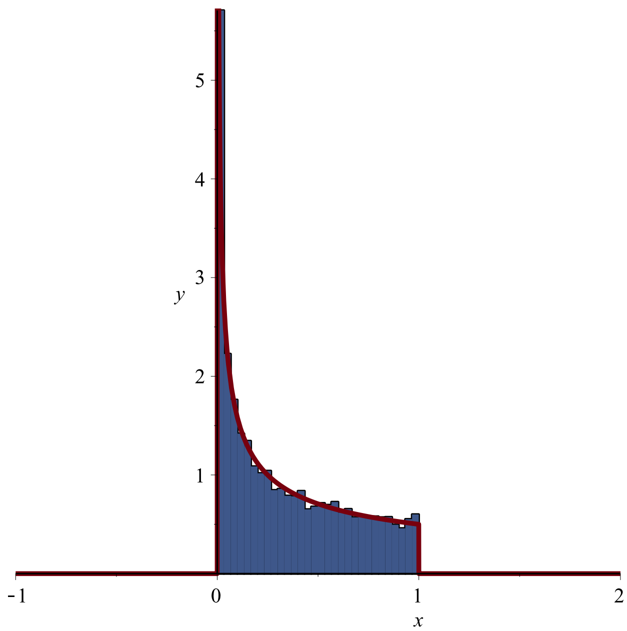
$$f(x) = \begin{cases} 0 & x < 0 \\ undefined & x = 0 \\ \frac{1}{2\sqrt{x}} & x < 1 \\ undefined & x = 1 \\ 0 & 1 < x \end{cases} \qquad \textbf{(2.2.6.1.4)}$$

```
> suruseg_plot := plot(f(x), x = -1..2, y= 0..5, thickness
  = 3): suruseg_plot;
```



Összemásolva a hisztorgammal meggyz az illeszkedés:

```
> plots[display](H, suruseg_plot);
```

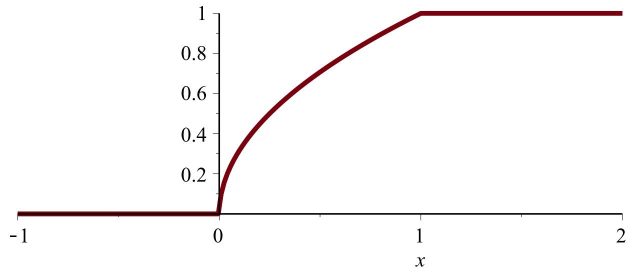


d) Határozzuk meg az eloszlás és srségfüggvényt a Statistics csomag eljárásait használva is!

Az eloszlás ill. srségfüggvényt úgy is megkaphatjuk, hogy a korábbi feladatokban már megismert *CDF* és *PDF* eljárásokat meghívjuk a *Z* valószínségi változóra:

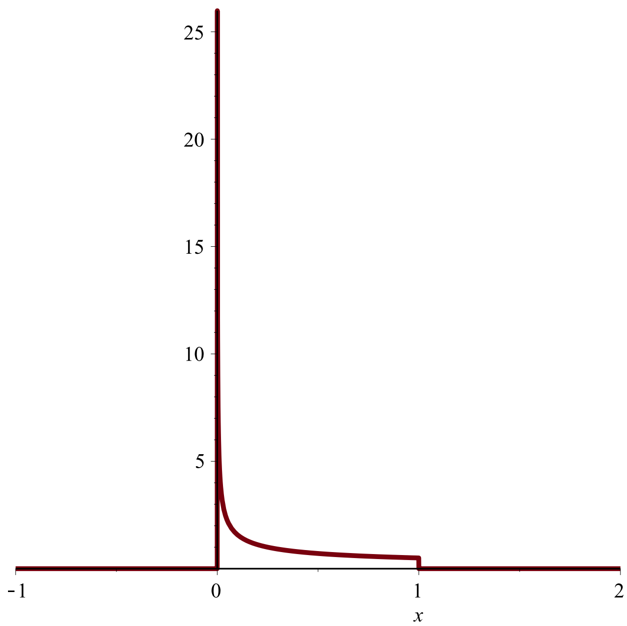
```
> F := x -> CDF(Z, x);
  plot(F(x), x = -1..2, scaling = constrained, thickness =
  3);
```

$$F := x \rightarrow Statistics\text{:-}CDF(Z, x)$$




```
> f := x -> PDF(Z, x);
  plot(f(x), x = -1..2, thickness = 3);
```

$$f := x \rightarrow Statistics\text{:-}PDF(Z, x)$$



e) Számítsuk ki kétféleképpen a $P\left(\frac{1}{3} \le Z \le \frac{3}{4}\right)$ valószínséget!

Jelölje p a kérdéses valószínséget! Az eloszlás- és a srségfüggvényt is használhatjuk p kiszámításához. Lássuk elször az eloszlásfüggvényt!

```
> p := F(3/4) - F(1/3); evalf(%);
```

$$p := \frac{1}{6}\sqrt{3}$$

$$0.2886751347 \qquad \textbf{(2.2.6.1.5)}$$

Tehát 0.289 a valószínsége, hogy Z az $\left[\frac{1}{3}, \frac{3}{4}\right]$ intervallumba esik. Mivel Z értékei a 0 körül srsödnek, ezért p kisebb, mint az intervallum hossza, ami $\frac{3}{4} - \frac{1}{3} = \frac{5}{12} \approx 0.417$.

A másik mód p kiszámítására a srségfüggvény integrálása az $\left[\frac{1}{3}, \frac{3}{4}\right]$ intervallumon.

```
> p := int(f(x), x = 1/3..3/4); evalf(%);
```

$$p := \frac{1}{6}\sqrt{3}$$

$$0.2886751347 \qquad \textbf{(2.2.6.1.6)}$$

Ugyanazt az értéket kaptuk, ahogy az várható is volt.

f) Számítsuk ki *Z* várható értékét, varianciáját és szórását!

A várható érték definíció szerinti integrállal:

```
> EZ := int(x*f(x), x = 0..1); evalf(%); # mivel f(x) a
  [0, 1] intervallumból vehet fel értékeket, ezért elég
  itt integrálni
```

$$EZ := \frac{1}{3}$$

$$0.3333333333 \quad (2.2.6.1.7)$$

A variancia (szórásnégyzet):

```
> VarZ := int((x-EZ)^2*f(x), x = 0..1); evalf(%); # a
  várható értékhez hasonlóan szintén elég a [0, 1]
  intervallumon integrálni
```

$$VarZ := \frac{4}{45}$$

$$0.08888888889 \quad (2.2.6.1.8)$$

Igazoljuk, hogy a variancia az alternatív $Var(Z) = E(Z^2) - E(Z)^2$ képlettel is meghatározható!

```
> EZ2 := int(x^2*f(x), x = 0..1); evalf(%);
```

$$EZ2 := \frac{1}{5}$$

$$0.2000000000 \quad (2.2.6.1.9)$$

```
> VarZ = EZ2-EZ^2;
```

$$\frac{4}{45} = \frac{4}{45} \quad (2.2.6.1.10)$$

A szórás a variancia négyzetgyöke:

```
> sigmaZ := sqrt(VarZ); evalf(%);
```

$$sigmaZ := \frac{2}{15}\sqrt{5}$$

$$0.2981423969 \quad (2.2.6.1.11)$$

Ezeket a mutatókat a *Statistics* csomag függvényeinek a segítségével közvetlenül is meghatározhatjuk:

```
> ExpectedValue(Z);
```

$$\frac{1}{3} \quad (2.2.6.1.12)$$

```
> Variance(Z);
```

$$\frac{4}{45} \quad (2.2.6.1.13)$$

```
> StandardDeviation(Z);
```

$$\frac{2}{15}\sqrt{5} \quad (2.2.6.1.14)$$

g) Mekkora valószínséggel lesz *Z* nagyobb, mint várható értéke?

A $P(Z > E(Z))$ valószínséget keressük! Ezt a srségfüggvény megfelel intervallumon való integrálásával (vagy az eloszlásfüggvény segítségével) határozhatjuk meg:

```
> `P(Z > EZ)` := int(f(x), x = EZ..1); evalf(%);
```

$$P(Z > EZ) := 1 - \frac{1}{3}\sqrt{3}$$

(2.2.6.1.15)

$$0.4226497307 \qquad (2.2.6.1.15)$$

A Statistics csomag *Probability* eljárásával ugyanerre az eredményre juthatunk:

```
> Probability(Z > EZ);
```

$$1 - \frac{1}{3}\sqrt{3} \qquad (2.2.6.1.16)$$

6. Feladat (Összeg eloszlása)

Válasszunk véletlenszeren a [0,1] intervallumból két számot, jelölje *Z* az összegüket.
a) Ábrázoljuk *Z* eloszlását pontdiagramon illetve hisztogramon egy 1000 elem minta alpján!
b) Határozzuk meg *Z* eloszlás- és srségfüggvényét! Hasonlítsuk össze a srségfüggvényt az elz pontban kapott hisztogrammal!
c) Határozzuk meg *Z* várható értékét és szórását definíció szerinti kiszámítással és a Maple beépített eljárásaival is!

Megoldás

```
> restart;
```

a) Ábrázoljuk *Z* eloszlását pontdiagramon illetve hisztogramon egy 1000 elem minta alpján!

```
> with(Statistics):
  randomize():
```

Készítünk két egyenletes eloszlású valószínségi változót a $[0, 1]$ intervallumon és vegyük az összegüket!

```
> X := RandomVariable(Uniform(0, 1));
  Y := RandomVariable(Uniform(0, 1));
  Z := X + Y;
```

$$X := \_R$$
$$Y := \_R0$$
$$Z := \_R + \_R0 \qquad (2.2.7.1.1)$$

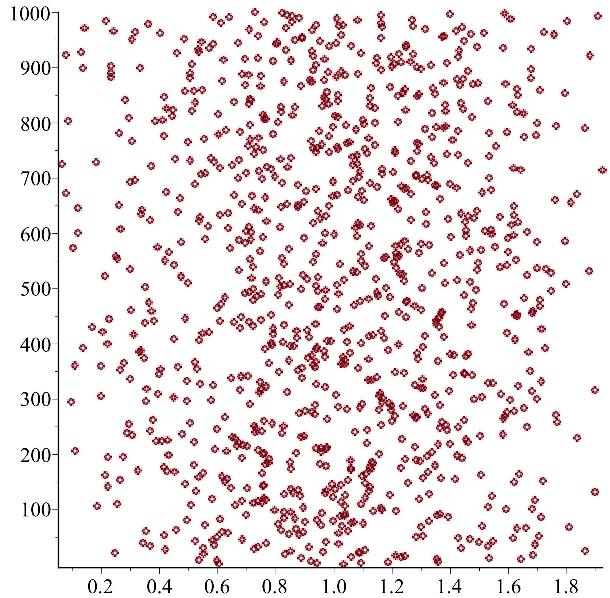
Generáljunk egy 1000 elem mintát *Z*-re, majd ábrázoljuk pontdiagramon!

```
> n := 1000; minta := Sample(Z, n);
```

$$n := 1000$$

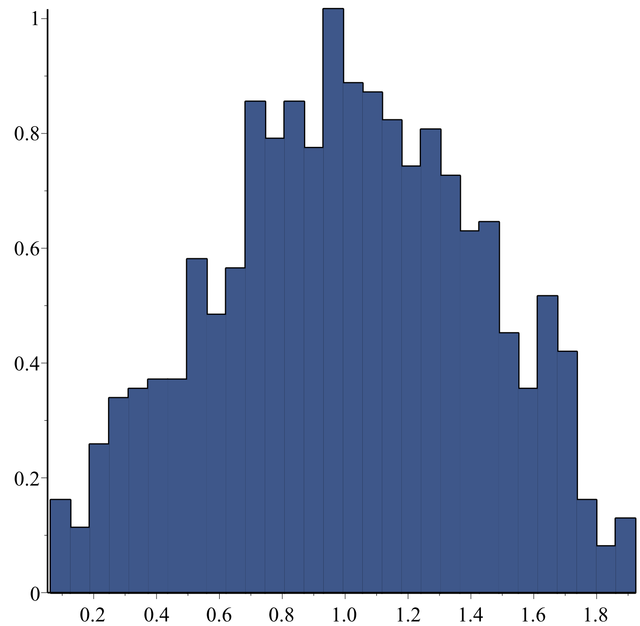
$$minta := \left[\begin{array}{c} \text{1 .. 1000 } Vector_{row} \\ \text{Data Type: } float_{8} \\ \text{Storage: rectangular} \\ \text{Order: Fortran\_order} \end{array}\right] \qquad (2.2.7.1.2)$$

```
> pontok := [seq([minta[i], i], i = 1..n)]:
  plot(pontok, style = point);
```



Az látszik, hogy a pontok most a $[0, 2]$ intervallumban oszlanak el és a közepe felé srsödnek. Nézzük meg ezt hisztogramon is!

```
> H := Histogram(minta): H; # a bincount opció beállítását
  a Maple-re bízzuk
```



A hisztogram megersíti az elbbi észrevételünket. A minta elemszámának növelésével a hisztogram egyre inkább "kisímul", de 100 %-ban pontos képet a háttérben meghúzódó eloszlásról csak a srségfüggvény meghatározásával kaphatunk.

b) Határozzuk meg Z eloszlás- és srségfüggvényét! Hasonlítsuk össze a srségfüggvényt az elz pontban kapott hisztogrammal!

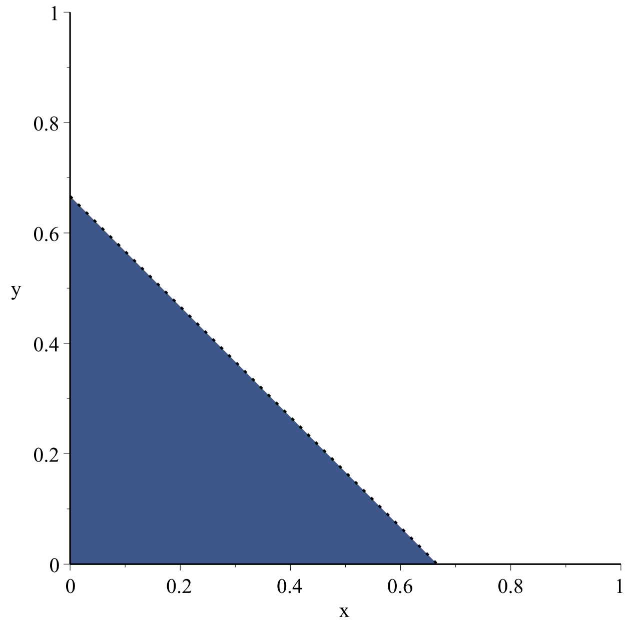
Kezdjük az eloszlásfüggvénnyel, amihez a $P(Z < z)$ valószínségeket kell meghatározni minden z valós számra! Mivel Z két független mennyiség (X és Y) összegeként áll el, érdemes áttérni a 2-dimenziós szemléletre és geometriai valószínséggel számolni. Az eseménytér: $\Omega = \{(x, y) | 0 \leq x \leq 1, 0 \leq y \leq 1\} = [0, 1] \times [0, 1]$, geometriai mértéke $m(\Omega) = 1$. A kedvez esetek halmaza $z = c$ választással:

$$P(Z < c) = P(X + Y < c) = P(Y < c - X),$$

ami az egységnégyzet "$y = c - x$" egyenlet egyenes alatti tartománya. Nézzünk néhény példát!

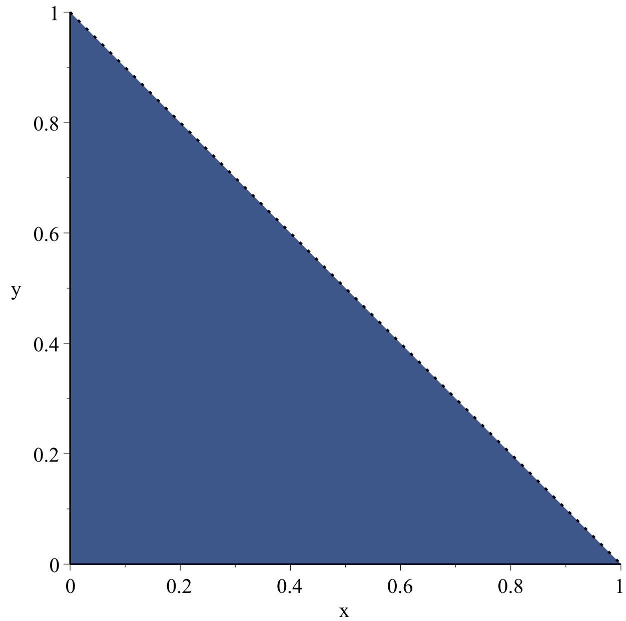
$c = \frac{2}{3}$:

```
> with(plots):
  inequal({y < 2/3 - x}, x = 0..1, y = 0..1);
```



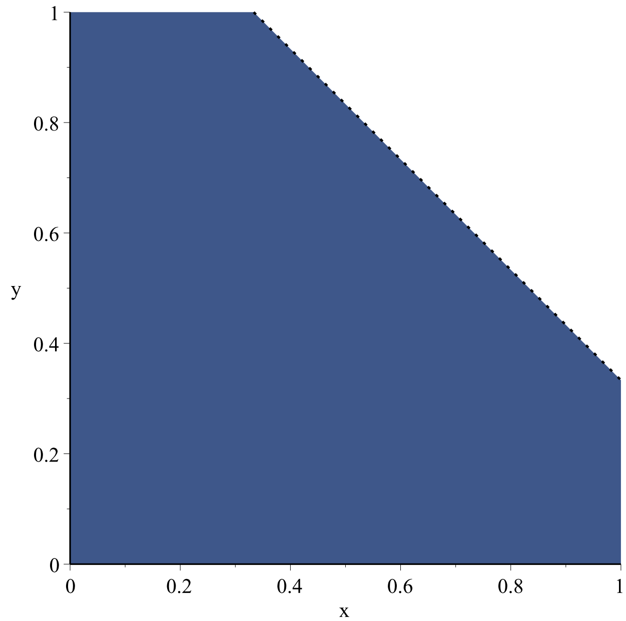
$c = 1$:

```
> inequal({y < 1 - x}, x = 0..1, y = 0..1);
```



$c = \frac{4}{3}$:

```
> inequal({y < 4/3 - x}, x = 0..1, y = 0..1);
```



Általánosságban elmondható, hogy

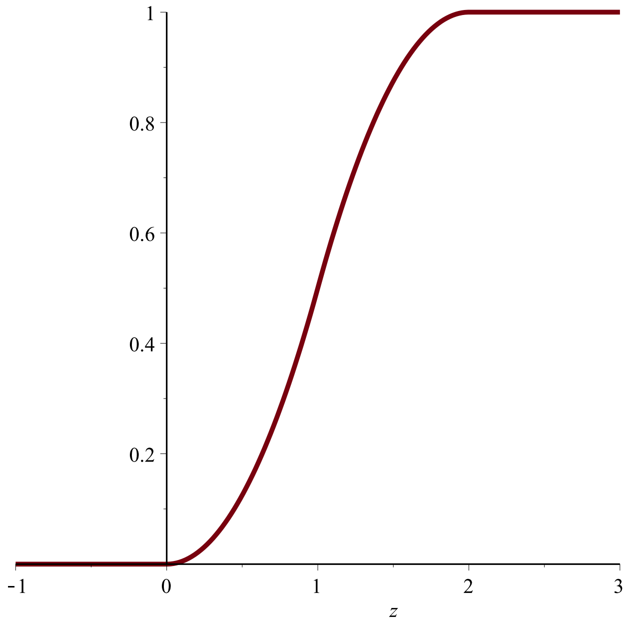
- ha $0 < c \leq 1$, akkor a kedvez tartomány területe $T_c = \frac{c^2}{2}$ (c befogójú derékszög háromszög) és az eloszlásfüggvény értéke $F(c) = \frac{T_c}{m(\Omega)} = \frac{T_c}{1} = \frac{c^2}{2}$;
- ha $1 < c \leq 2$, akkor $T_c = 1 - \frac{(2-\mathrm{c})^2}{2}$ (az egységnégyzet területébl kivonva egy $(2-\mathrm{c})$ befogójú derékszög háromszög területe) és az eloszlásfüggvény értéke $F(c)$ $= \frac{T_c}{m(\Omega)} = 1 - \frac{(2-\mathrm{c})^2}{2}$.

Könnyen látható továbbá, hogy $c \leq 0$ esetén $F(c) = 0$, $c > 2$ esetén pedig $F(c) = 1$.
Összefoglalva, az eloszlásfüggvény:

```
> F := z -> piecewise(z <= 0, 0, z <= 1, z^2/2, z <= 2, 1
  - (2-z)^2/2, 1);
```

$$F := z \rightarrow piecewise\left(z \leq 0, 0, z \leq 1, \frac{1}{2}\,z^2, z \leq 2, 1 - \frac{1}{2}\,(2-z)^2, 1\right) \qquad \textbf{(2.2.7.1.3)}$$

```
> plot(F(z), z = -1..3, thickness = 3);
```



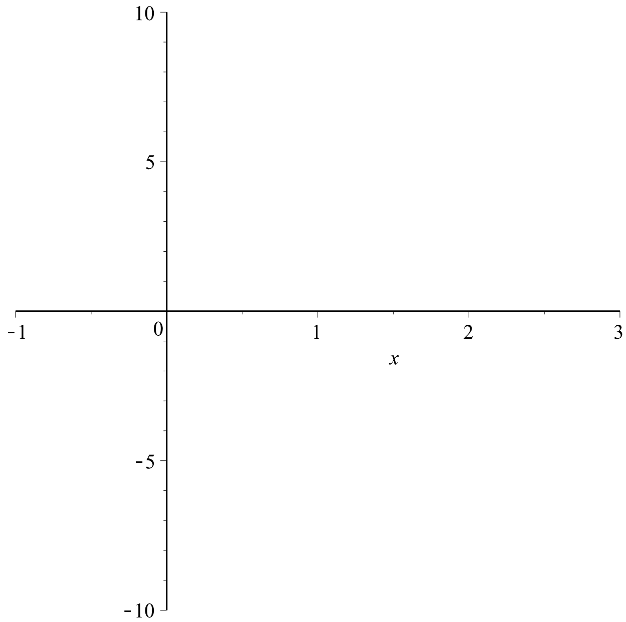
A srségfüggvény az eloszlásfüggvény deriváltja:

```
> f := z -> diff(F(z), z);
```

$$f := z \rightarrow \frac{d}{dz} F(z) \qquad \textbf{(2.2.7.1.4)}$$

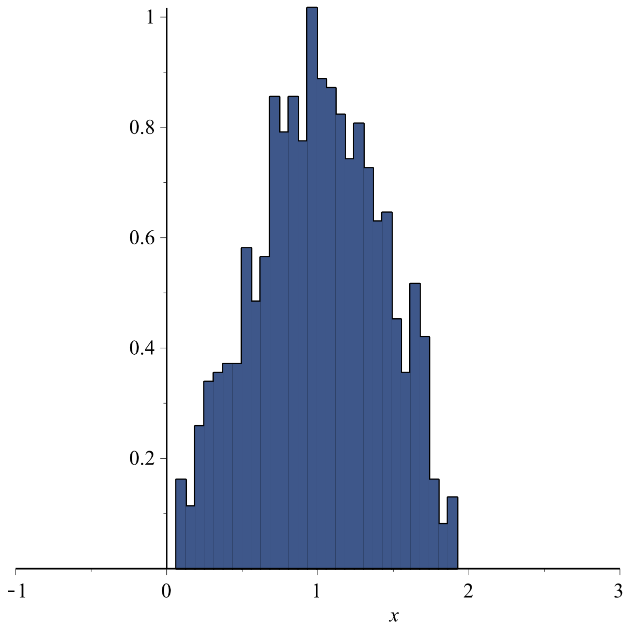
```
> suruseg_plot := plot(f(z), x = -1..3, thickness = 3):
  suruseg_plot;
```

Warning, expecting only range variable x in expression piecewise(z <= 0,0,z <= 1,z,z <= 2,2-z,2 < z,0) to be plotted but found name z



Utolsó lépésként ábrázoljuk közös diagramon az a) feladatrészben kapott hisztogramot és a srségfüggvényt

```
> plots[display](H, suruseg_plot);
```



Meggyzen látszik, hogy a hisztogram illeszkedik a srségfüggvényhez.

c) Határozzuk meg *Z* várható értékét és szórását definíció szerinti kiszámítással és a Maple beépített eljárásaival is!

Z várható értékét a srségfüggvény *z*-szeresének integrálásával kapjuk:

```
> EZ := int(z*f(z), z = 0..2);
```

$$EZ := 1 \quad \textbf{(2.2.7.1.5)}$$

Mivel a srségfüggvény szimmetrikus $z = 1$-re, ezért geometriai szemlélet alapján is könnyen látható, hogy $E(Z) = 1$.

A várható érték additivitását alkalmazva is ugyanerre az eredményre jutunk, és ha tudjuk fejbl az egyenletes eloszlás várható értékét, akkor számolnunk sem kell:

$E(Z) = E(X) + E(Y) = \frac{1}{2} + \frac{1}{2} = 1.$

Számoljuk ki most a varianciát és a szórást!

```
> VarZ := int((z-EZ)^2*f(z), z = 0..2);
```

$$VarZ := \frac{1}{6} \quad \textbf{(2.2.7.1.6)}$$

A számoláshoz kihasználhatjuk azt is, hogy a variancia független valószínségi változók összegére additív: $Var(Z) = Var(X) + Var(Y) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$.

Végül a szórás:

```
> sigmaZ := sqrt(VarZ);
```

$$sigmaZ := \frac{1}{6}\sqrt{6} \quad \textbf{(2.2.7.1.7)}$$

A Maple beépített eszközeivel (Statistics csomag) is megkaphatjuk a várható értéket, varianciát és szórást:

```
> ExpectedValue(Z);
```

$$1 \quad \textbf{(2.2.7.1.8)}$$

```
> Variance(Z);
```

$$\frac{1}{6} \quad \textbf{(2.2.7.1.9)}$$

```
> StandardDeviation(Z);
```

$$\frac{1}{6}\sqrt{6} \quad \textbf{(2.2.7.1.10)}$$

7. Feladat (Ötöslottó húzás)

Az ötöslottó-sorsoláson az 1 - 90 számok közül húznak ki 5-öt.
a) Várhatóan mennyi lesz a kihúzott számok összege?
b) Várhatóan hány 1 jegy szám lesz a kihúzottak között?

Megoldás

```
> restart;
```

a) Várhatóan mennyi lesz a kihúzott számok összege?

Jelölje X a kihúzott számok összegét. Ez egy diszkrét valószínségi változó, melynek eloszlását azonban igen komplikált lenne megadni, ezért a várható érték meghatározására a definíció alapján történ kiszámítástól különböz utat kell találnunk: bontsuk fel X-et egyszerbb változók összegére, és használjuk ki a várható érték additivitását!

Legyen X_i az i-edikre húzott szám, $i = 1, 2, 3, 4, 5$. Ekkor $X = \sum_{i=1}^{5} X_i$ a húzott számok összege.

Észrevehetjük, hogy az X_i valószínségi változók mind diszkrét egyenletes eloszlásúak az 1-t1 90-ig terjed egész számok halmazán: mivel nincs kitüntetett szám, ezért bármelyik húzásra bármelyik szám azonos valószínséggel adódhat (persze a húzások nem függetlenek egymástól –ha pl. elsőre kihúzzák a 10-et, akkor másodszorra már nem húzhatják ugyanezt–, így az előző húzásokra vonatkoztatott *feltételes* valószínségek már eltérnek az egyenletes $\frac{1}{90}$-tl). Határozzuk meg elbb az egyes húzások $E(X_i)$ várható értékét, majd alkalmazzuk a várható érték additivitását:

$$E(X) = \sum_{i=1}^{5} E(X_i) = 5 \cdot E(X_1)$$

(X_i-k azonos eloszlásúak, így várható értékük is megegyezik).

```
> xi := [seq(j, j = 1..90)]:
  pxi := [seq(1/90, j = 1..90)]:
> EXi := sum(xi[j]*pxi[j], j = 1..90); evalf(%);
```

$$EXi := \frac{91}{2}$$

$$45.50000000 \quad \textbf{(2.2.8.1.1)}$$

Azt kaptuk, hogy $E(X_i) = 45.5$, ami szimmetriai meggondolással is könnyen látható: az 1-90 számok átlaga az $[1, 90]$ intervallum felezpontja, ami $\frac{1+90}{2}$.

Az X várható értéke:

```
> EX1 := EXi:
  EX := 5*EX1; evalf(%);
```

$$EX := \frac{455}{2}$$

$$227.5000000 \qquad (2.2.8.1.2)$$

Ugyanez a számolás elvégezhet a Statistics csomag eszközeivel is:

```
> with(Statistics):
> XX := [seq(RandomVariable(DiscreteUniform(1,90)), i = 1..5)]; # 5 független valószínségi változó az 5 húzásra
```

$$XX := [\_R, \_R0, \_R1, \_R2, \_R3] \qquad (2.2.8.1.3)$$

```
> X := sum(XX[i], i = 1..5);
```

$$X := \_R + \_R0 + \_R1 + \_R2 + \_R3 \qquad (2.2.8.1.4)$$

```
> EX1 := ExpectedValue(XX[1]); # Mean(XX[1]) is használható
```

$$EX1 := \frac{91}{2} \qquad (2.2.8.1.5)$$

```
> EX := ExpectedValue(X);
```

$$EX := \frac{455}{2} \qquad (2.2.8.1.6)$$

b) Várhatóan hány 1 jegy szám lesz a kihúzottak között?

Jelölje Y a kihúzott egyjegy számok darabszámát! $E(Y)$ kiszámítására ugyanazt a trükköt alkalmazzuk, mint az a) feladatrészben.

Legyen Y_i annak az eseménynek az *indikátor változója*, hogy az i-edik alkalommal kihúzott szám egyjegy ($i = 1, 2, 3, 4, 5$), azaz

$$Y_i = \begin{cases} 1 & \text{ha az i-edik kihúzott szám egyjegyű} \\ 0 & \text{különben} \end{cases}.$$

Megint láthatjuk, hogy az Y_i-k (i-tl függetlenül) azonos eloszlásúak, hiszen minden húzásnál ugyanakkora valószínséggel adódhat egyjegy szám. Y_i ún. p-paraméter *Bernoulli* eloszlású valószínségi változó, ahol $p > 0$ az $\{Y_i = 1\}$ eseményhez tartozó valószínség, ami jelen esetben az egyjegy szám húzásának esélye: $p = \frac{9}{90} = \frac{1}{10}$ (klasszikus valószínségi mez, 9 egyjegy szám jó a 90-b1).

Nem nehéz belátni, hogy $Y = \sum_{i=1}^{5} Y_i$ (annyiszor adunk össze 1-et, ahányszor egyjegy szám került kihúzásra), amibl következik

$$E(Y) = \sum_{i=1}^{5} E(Y_i) = 5 \cdot E(Y_1).$$

Y_i eloszlása és várható értéke:

```
> yi := [0, 1];
  pyi := [9/10, 1/10]; # P(Yi = 0) = 1 - P(Yi = 1)
```

$$yi := [0, 1]$$

$$pyi := \left[\frac{9}{10}, \frac{1}{10}\right] \quad (2.2.8.1.7)$$

```
> EYi := sum(yi[j]*pyi[j], j = 1..2); evalf(%);
```

$$EYi := \frac{1}{10}$$

$$0.1000000000 \quad (2.2.8.1.8)$$

Azt kaptuk, hogy Y_i várható értéke $\frac{1}{10}$ (és általában is igaz, hogy p-paraméter Bernoulli eloszlás várható értéke egyenl p-vel). Ebbl megkapjuk Y várható értékét:

```
> EY1 := EYi:
  EY := 5*EY1; evalf(%);
```

$$EY := \frac{1}{2}$$

$$0.5000000000 \quad (2.2.8.1.9)$$

Ezt úgy is értelmezhetjük, hogy 2 lottósorsolás során összesen várhatóan 1 darab egyjegy szám fog felbukkanni.
Ugyanez a számolás a Statistics csomag eszközeivel:

```
> YY := [seq(RandomVariable(Bernoulli(1/10)), i = 1..5)];
```

$$YY := [\_R4, \_R5, \_R6, \_R7, \_R8] \quad (2.2.8.1.10)$$

```
> Y := sum(YY[i], i = 1..5);
```

$$Y := \_R4 + \_R5 + \_R6 + \_R7 + \_R8 \quad (2.2.8.1.11)$$

```
> EY1 := ExpectedValue(YY[1]);
```

$$EY1 := \frac{1}{10} \quad (2.2.8.1.12)$$

```
> EY := ExpectedValue(Y);
```

$$EY := \frac{1}{2} \quad (2.2.8.1.13)$$

8. Feladat (Elhajított pontszer test)

Egy elhajított pontszer test földfelszín feletti magasságát a dobás óta eltelt id függvényében a $h(t) = -5\,t^2 + t + 42$ függvény írja le (méterben mérve). Megfigyeljük helyzetét egy véletlenszer T idpontban, melynek várható értéke $E(T) = 1.5$, szórása pedig $\sigma(T) = \frac{\sqrt{3}}{2}$. Mennyi a megfigyelt $H = h(T)$ magasság várható értéke?

Megoldás

```
> restart;
```

Ez a feladat jól példázza, hogyan lehet adott bizonytalansággal ($\sigma(T)$) ismert fizikai mennyiségekkel számolni. Rögzítsük az adatokat!

```
> h := t -> -5*t^2 + t + 42;
```

$$h := t \rightarrow -5\,t^2 + t + 42 \quad (2.2.9.1.1)$$

```
> ET := 1.5;
  sigmaT := sqrt(3)/2;
```

$$ET := 1.5$$

(2.2.9.1.2)

$$sigmaT := \frac{1}{2}\sqrt{3} \qquad \mathbf{(2.2.9.1.2)}$$

H a T véletlen idpontban mért magasság:

```
> H := h(T);
```

$$H := -5\,T^2 + T + 42 \qquad \mathbf{(2.2.9.1.3)}$$

Az $E(H)$ várható értéket kell meghatároznunk. A várható érték tulajdonságai (additivitás és linearitás) miatt

$$E(H) = E(-5\,T^2) + E(T) + E(42) = -5E(T^2) + E(T) + 42.$$

Az $E(T^2)$ második momentumot a tanult $\sigma^2(T) = E(T^2) - E^2(T)$ azonosság alkalmazásával számolhatjuk ki.

```
> EH := -5*ET2+ET+42;
```

$$EH := -5\,ET2 + 43.5 \qquad \mathbf{(2.2.9.1.4)}$$

```
> isolate(sigmaT^2 = ET2 - ET^2, ET2); assign(%);
```

$$ET2 = 3.000000000 \qquad \mathbf{(2.2.9.1.5)}$$

```
> EH; # a megoldás
```

$$28.50000000 \qquad \mathbf{(2.2.9.1.6)}$$

Tehát a megfigyeléskor mért magasság várható értéke 28.5 *m*.

9. Feladat (Magyar kártya párok)*

Egy 32 lapos magyar kártyapakliból kiválasztjuk a piros és a zöld lapokat (azaz 8-8 darabot), majd közülük véletlenszeren kiveszünk 8-at (visszatevés nélkül). Várhatóan hány pár (2 azonos érték, de különböz szín lap) lesz a kihúzott lapok között?

Megoldás

```
> restart;
```

Rögzítsük az adatokat!

```
> n := 16; # zöld és piros lapok száma összesen
  k := 8; # kihúzott lapok száma
```

$$n := 16$$
$$k := 8 \qquad \mathbf{(2.2.10.1.1)}$$

Legyen X a kihúzott párok száma! Összesen 8 pár van a piros és zöld lapok között (VII-es, VIII-as, IX-es, X-es, Alsó, Fels, Király, Ász).

Jelölje X_i ($i = 1, 2, \ldots, 8$) annak az eseménynek az indikátor-változóját, hogy az i-edik pár a kihúzott 8 lap között van, azaz

$$X_i = \begin{cases} 1 & \text{ha az i-edik párt kihúztuk} \\ 0 & \text{különben} \end{cases}$$

X_i ún. p-paraméter Bernoulli-eloszlású valószínségi változó, ahol $p = P(X_i = 1)$.

Vegyük észre, hogy $X = \sum_{i=1}^{8} X_i$ (minden kihúzott párra 1-et adunk hozzá az összeghez). A várható érték additív tulajdonsága miatt ekkor $E(X) = \sum_{i=1}^{8} E(X_i) = 8 \cdot E(X_1)$ (az utolsó egyenlség abból adódik, hogy az X_i-k azonos eloszlásúak, bármely pár kihúzásának ugyanakkora esélye van). Határozzuk meg elbb $E(X_1)$-et, majd $E(X)$-et!

$E(X_1) = 1 \cdot P(\text{kihúztuk az i-edik párt}) + 0 \cdot P(\text{nem húztuk ki az i-edik párt})$

$$= P(\text{kihúztuk az i-edik párt}) = p = \frac{\textit{kedvező esetek}}{\textit{összes eset}}$$

, mivel az elemi események (8-lap húzások) azonos valószínségek.

Az összes esetek száma: $n = 16$ lapból húzunk $k = 8$-at $\rightarrow N = \binom{16}{8}$ (ismétlés nélküli kombináció).

```
> N := binomial(16, 8);
```

$$N := 12870 \qquad \textbf{(2.2.10.1.2)}$$

A kedvez esetek száma: $n = 16$ lapból kihúzzuk az els párt (2 lap) és még 6 lapot $\rightarrow$ $K = \binom{16-2}{6}$ (ismétlés nélküli kombináció).

```
> K := binomial(14, 6);
```

$$K := 3003 \qquad \textbf{(2.2.10.1.3)}$$

A kérdéses valószínség:

```
> p := K/N;
```

$$p := \frac{7}{30} \qquad \textbf{(2.2.10.1.4)}$$

X_1 várható értéke ezzel megegyezik:

```
> EX1 := p;
```

$$EX1 := \frac{7}{30} \qquad \textbf{(2.2.10.1.5)}$$

X várható értéke, azaz a kihúzott párok várható száma a fenti okoskodás alapján:

```
> EX := 8*EX1; evalf(%);
```

$$EX := \frac{28}{15}$$

$$1.866666667 \qquad \textbf{(2.2.10.1.6)}$$

Tehát a kihúzott párok várható száma 1.867, vagyis átlagosan megközelítleg 2 párra számíthatunk.

10. Feladat (Játék korrektsége)

Egy játék 3 szabályos pénzérme feldobásából áll. A játékos 3 fej dobásakor 6 forintot, 2 fej dobásakor 3 forintot, 1 fej dobásakor 2 forintot kap, egyébként pedig 3 írás dobásakor a játékos fizet a banknak 16 forintot. Állapítsuk meg, hogy a játékos szempontjából kedvez vagy kedveztlen-e a játék!

Megoldás

```
> restart;
```

Az adatok:

```
> n := 3; # pénzérmék száma
  p := 1/2; # egy érmével fej dobásának esélye
```

$$n := 3$$

$$p := \frac{1}{2} \qquad \textbf{(2.2.11.1.1)}$$

```
> F := [0, 1, 2, 3]; # fejek lehetséges száma
```

$$F := [0, 1, 2, 3] \qquad \textbf{(2.2.11.1.2)}$$

```
> x := [-16, 2, 3, 6]; # fejek számához tartozó nyeremények (a játékos szempontjából)
```
$$x := [-16, 2, 3, 6] \quad (2.2.11.1.3)$$

```
> linalg[augment](Matrix(2, 1, [[fejek], [nyeremény]]), Matrix(2, 4, [F, x]));
```
$$\begin{bmatrix} fejek & 0 & 1 & 2 & 3 \\ nyeremény & -16 & 2 & 3 & 6 \end{bmatrix} \quad (2.2.11.1.4)$$

Jelölje X a játékos nyereményét! Határozzuk meg X eloszlását, azaz a $P(X=k)$ valószínségeket, ahol $k=0, 1, 2, 3$. A kísérlethez az a matematikai modell illeszkedik, melyben különböznek tekintjük az érméket (pl. nem egyszerre, hanem egymás után kerülnek feldobásra) és az elemi események a 3-hosszú fej-írás sorozatok (ekkor kapunk klasszikus valószínségi mezt). Pontosan k fej dobásának valószínségét egy (n, p)-paraméter *binomiális eloszlás* írja le: $P(X=k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \binom{3}{k} \cdot \frac{1}{2^n}$ - a dobások egymástól függetlenek, így a valószínségek szorzástétele szerint annak valószínsége, hogy az els k dobás fej, a maradék $n-k$ dobás írás $p^k \cdot (1-p)^{n-k}$, és a k fej $\binom{n}{k}$-féleképpen helyezkedhet el a sorrendben. (A $\frac{\textit{kedvező eset}}{\textit{összes eset}}$ képlettel számolva is ugyanerre az eredményre juthatunk.)

```
> vsz := k -> binomial(n, k)*p^k*(1 - p)^(n - k);
```
$$vsz := k \rightarrow \mathrm{binomial}(n, k)\, p^k\, (1-p)^{n-k} \quad (2.2.11.1.5)$$

```
> px := map(vsz, F); # kimenetelek valószínségei
```
$$px := \left[\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}\right] \quad (2.2.11.1.6)$$

```
> linalg[augment](Matrix(2, 1, ['x', 'px']), Matrix(2, 4, [x, px]));
```
$$\begin{bmatrix} x & -16 & 2 & 3 & 6 \\ px & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{bmatrix} \quad (2.2.11.1.7)$$

A játék *korrekt*, ha a nyeremény várható értéke 0, *kedvez*, ha > 0, *kedveztlen*, ha < 0. Az $E(X)$ várható értéket X eloszlásának ismeretében könnyen kiszámíthatjuk.

```
> EX := sum(x[i + 1]*px[i + 1], i = 0..n);
```
$$EX := \frac{5}{8} \quad (2.2.11.1.8)$$

Mivel $E(X) = \frac{5}{8} > 0$, ezért a játék kedvez a játékos szempontjából.

11. Feladat (Együttes eloszlás, határeloszlás, kovariancia)*

Tegyük fel, hogy az X és Y valószínségi változók együttes eloszlása az alábbi táblázattal adott! Határozzuk meg X és Y eloszlását, várható értékét, szórását, kovarianciáját és korrelációját! Vizsgáljuk meg X és Y függetlenségét!

a)

$X \backslash Y$	-3	2	4		$\sum_{sor}$

1	**0.1**	**0.2**	**0.2**	**0.5**
3	**0.3**	**0.1**	**0.1**	**0.5**
	$\sum_{oszlop}$ **0.4**	**0.3**	**0.3**	**1**

b)

$X \setminus Y$	**-4**	**2**	**7**	$\sum$
1	**30/16**	**69/16**	**21/16**	**?**
5	**20/16**	**1/16**	**19/16**	**?**
$\sum$	**?**	**?**	**?**	**1**

Megoldás

```
> restart;
```

a)

$X \setminus Y$	**-3**	**2**	**4**	
1	**0.1**	**0.2**	**0.2**	**0.5**
3	**0.3**	**0.1**	**0.1**	**0.5**
	0.4	**0.3**	**0.3**	**1**

Az együttes eloszlástáblázat értelmezése:

$X \setminus Y$	y_1	y_2	y_3	
x_1	p_{11}	p_{12}	p_{13}	$p_{1\bullet}$
x_2	p_{21}	p_{22}	p_{23}	$p_{2\bullet}$
	$p_{\bullet 1}$	$p_{\bullet 2}$	$p_{\bullet 3}$	1

ahol $x_1, x_2, \ldots$ ill. $y_1, y_2, \ldots$ az X ill. Y valószínségi változók értékkészlete, $p_{ij} = P(X = x_i \text{ és } Y = y_j)$ az ún. együttes valószínségek, $p_{1\bullet}, p_{2\bullet}, \ldots$ a sorösszegek, melyek X eloszlását adják és $p_{\bullet 1}, p_{\bullet 2}, \ldots$ az oszlopösszegek, melyek Y eloszlását adják. Az utóbbi kettt hívják marginális-, perem, vagy határeloszlásoknak is, mivel a táblázat szélein helyezkednek el (és az együttes eloszlásfüggvény ∞-ben vett határértékeként állnak elő).

Egy ilyen táblázat tehát két diszkrét eloszlású valószínségi változó összes lehetséges értékkombinációinak valószínségeit adja meg, és az egyes változók külön-külön vett eloszlása is leolvasható belle.

X eloszlása:

```
> m := 2; # sorok száma
  x := [1, 3];
  px := [0.5, 0.5];
```

$$m := 2$$
$$x := [1, 3]$$
$$px := [0.5, 0.5] \quad \textbf{(2.2.12.1.1)}$$

```
> linalg[augment](Matrix(2, 1, ['x'[i], 'px'[i]]), Matrix
  (2, m, [x, px]));
```

$$\begin{bmatrix} x_i & 1 & 3 \\ px_i & 0.5 & 0.5 \end{bmatrix} \quad \textbf{(2.2.12.1.2)}$$

Y eloszlása:

```
> n := 3; # oszlopok száma
  y := [-3, 2, 4];
  py := [0.4, 0.3, 0.3];
```

$$n := 3$$
$$y := [-3, 2, 4]$$
$$py := [0.4, 0.3, 0.3] \quad \textbf{(2.2.12.1.3)}$$

```
> linalg[augment](Matrix(2, 1, ['y'[i], 'py'[i]]), Matrix
  (2, n, [y, py]));
```

$$\begin{bmatrix} y_i & -3 & 2 & 4 \\ py_i & 0.4 & 0.3 & 0.3 \end{bmatrix} \quad \textbf{(2.2.12.1.4)}$$

A várható érték és a szórás:

```
> EX := sum(x[i]*px[i], i = 1..m);
```

$$EX := 2.0 \quad \textbf{(2.2.12.1.5)}$$

```
> sigmaX := sqrt(sum((x[i] - EX)^2*px[i], i = 1..m));
```

$$sigmaX := 1.0 \quad \textbf{(2.2.12.1.6)}$$

```
> EY := sum(y[i]*py[i], i = 1..n);
```

$$EY := 0.6 \quad \textbf{(2.2.12.1.7)}$$

```
> sigmaY := sqrt(sum((y[i] - EY)^2*py[i], i = 1..n));
```

$$sigmaY := 3.039736831 \quad \textbf{(2.2.12.1.8)}$$

A kovariancia és a korreláció kiszámításához készítsünk egy mátrixot a fenti táblázat "közepébl"!

```
> P := Matrix(m, n, [[0.1, 0.2, 0.2], [0.3, 0.1, 0.1]]);
```

$$P := \begin{bmatrix} 0.1 & 0.2 & 0.2 \\ 0.3 & 0.1 & 0.1 \end{bmatrix} \quad \textbf{(2.2.12.1.9)}$$

A kovariancia a két valószínségi változó várható értékétl való eltérése szorzatának várható értéke: $Cov(X, Y) = E((X - E(X)) \cdot (Y - E(Y)))$

```
> CovXY := add(add((x[i] - EX)*(y[j] - EY)*P[i,j], i = 1..
  m), j = 1..n); # a sum itt nem mködik, a Maple
```

```
speciális kiértékelési szabályai miatt az add-ot kell
használnunk
```
$$CovXY := -1.200 \tag{2.2.12.1.10}$$

A kovariancia kiszámítható a tételként tanult $Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$ képlettel is, melyet gyakorlati szempontból jobban kezelhet.

```
> EXY := add(add(x[i]*y[j]*P[i,j], i = 1..m), j = 1..n);
```
$$EXY := 0. \tag{2.2.12.1.11}$$

```
> CovXY = EXY - EX*EY;
```
$$-1.200 = -1.20 \tag{2.2.12.1.12}$$

A korrelációs együttható a regresszióanalízisben játszik fontos szerepet és a két változó közötti lineáris összefüggés ersségét és irányát adja meg: ha $Corr(X, Y)$ közel van az 1-hez, akkor Y értéke X növekedésével arányosan n, ellenben ha $Corr(X, Y)$ közel van az -1-hez, akkor Y értéke X növekedésével arányosan csökken.
A korrelációt úgy kapjuk, hogy a kovarianciát leosztjuk a szórásokkal:
$Corr(X, Y) = \dfrac{Cov(X, Y)}{\sigma(X) \cdot \sigma(Y)}$.

```
> CorrXY := CovXY/(sigmaX*sigmaY);
```
$$CorrXY := -0.3947710169 \tag{2.2.12.1.13}$$

Ez azt jelenti, hogy a két változó között közepes ersség, fordított lineáris kapcsolat van.
A függetlenség vizsgálatához érdemes tudni, hogy független valószínségi változók kovarianciája (és korrelációja) 0. Mivel itt a kovariancia nem nulla, ezért X és Y nem lehet független! (De vigyázzunk: $Cov(X, Y) = 0$-ból még nem feltétlenül következik a függetlenség, mint azt a b) feladatrészben látni fogjuk!)

b)

X \ ***Y***	**-4**	**2**	**7**	**∑**
1	**30/1(**	**69/1(**	**21/1(**	**?**
5	**20/1(**	**1/1(**	**19/1(**	**?**
∑	**?**	**?**	**?**	**1**

A számolás az a) feladatrészhez hasonlóan történik, azzal a kivétellel, hogy itt nem adja meg a táblázat a határeloszlásokat, ezért ezeket is nekünk kell meghatározni.
Vegyük fel az együttes eloszlás mátrixát, mely a táblázat "közepét" tartalmazza!

```
> restart;
> m := 2; # sorok száma
  n := 3; # oszlopok száma
  P := Matrix(m, n, [[30/160, 69/160, 21/160], [20/160,
  1/160, 19/160]]);
```
$$m := 2$$
$$n := 3$$

(2.2.12.1.14)

$$P := \begin{bmatrix} \frac{3}{16} & \frac{69}{160} & \frac{21}{160} \\ \frac{1}{8} & \frac{1}{160} & \frac{19}{160} \end{bmatrix} \quad (2.2.12.1.14)$$

X értékei:

```
> x := [1, 5];
```

$$x := [1, 5] \quad (2.2.12.1.15)$$

A hozzájuk tartozó valószínségek (marginális eloszlás) a P mátrix sorösszegei:

```
> px := [seq(add(P[i,j], j = 1..n), i = 1..m)];
```

$$px := \left[\frac{3}{4}, \frac{1}{4}\right] \quad (2.2.12.1.16)$$

Táblázatban összefoglalva X eloszlása:

```
> linalg[augment](Matrix(2, 1, ['x'[i], 'px'[i]]), Matrix
  (2, m, [x, px]));
```

$$\begin{bmatrix} x_i & 1 & 5 \\ px_i & \frac{3}{4} & \frac{1}{4} \end{bmatrix} \quad (2.2.12.1.17)$$

Hasonlóan Y-ra:

```
> y := [-4, 2, 7];
```

$$y := [-4, 2, 7] \quad (2.2.12.1.18)$$

```
> py := [seq(add(P[i,j], i = 1..m), j = 1..n)];
```

$$py := \left[\frac{5}{16}, \frac{7}{16}, \frac{1}{4}\right] \quad (2.2.12.1.19)$$

```
> linalg[augment](Matrix(2, 1, ['y'[i], 'py'[i]]), Matrix
  (2, n, [y, py]));
```

$$\begin{bmatrix} y_i & -4 & 2 & 7 \\ py_i & \frac{5}{16} & \frac{7}{16} & \frac{1}{4} \end{bmatrix} \quad (2.2.12.1.20)$$

X és Y várható értéke és szórása:

```
> EX := sum(x[i]*px[i], i = 1..m); evalf(%);
```

$$EX := 2$$

$$2. \quad (2.2.12.1.21)$$

```
> sigmaX := sqrt(sum((x[i] - EX)^2*px[i], i = 1..m));
  evalf(%);
```

$$sigmaX := \sqrt{3}$$

$$1.732050808 \quad (2.2.12.1.22)$$

```
> EY := sum(y[i]*py[i], i = 1..n); evalf(%);
```

$$EY := \frac{11}{8}$$

$$1.375000000 \quad (2.2.12.1.23)$$

```
> sigmaY := sqrt(sum((y[i] - EY)^2*py[i], i = 1..n));
  evalf(%);
```

$$sigmaY := \frac{1}{8}\sqrt{1095}$$

$$4.136348028 \qquad \textbf{(2.2.12.1.24)}$$

A kovariancia és a korreláció:

```
> CovXY := add(add((x[i] - EX)*(y[j] - EY)*P[i,j], i = 1..
  m), j = 1..n); # a sum itt nem mködik, a Maple
  speciális kiértékelési szabályai miatt az add-ot kell
  használnunk
```

$$CovXY := 0 \qquad \textbf{(2.2.12.1.25)}$$

```
> CorrXY := CovXY/(sigmaX*sigmaY);
```

$$CorrXY := 0 \qquad \textbf{(2.2.12.1.26)}$$

Ez azt jelenti, hogy nincs lineáris összefüggés a két valószínségi változó között. Vizsgáljuk meg, függetlenek-e! A függetlenség feltétele diszkrét esetben az, hogy az együttes eloszlás előálljon a peremeloszlások (diadikus) szorzataként: $p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$, minden (i, j) index-párra. Teszteljük az $i = 1, j = 1$ helyettesítéssel!

```
> P[1,1] = px[1]*py[1]; evalb(%);
```

$$\frac{3}{16} = \frac{15}{64}$$

$$false \qquad \textbf{(2.2.12.1.27)}$$

Tehát X és Y nem független, annak ellenére, hogy $Cov(X, Y) = 0$!

12. Feladat (Pont kiválasztása körlapon)

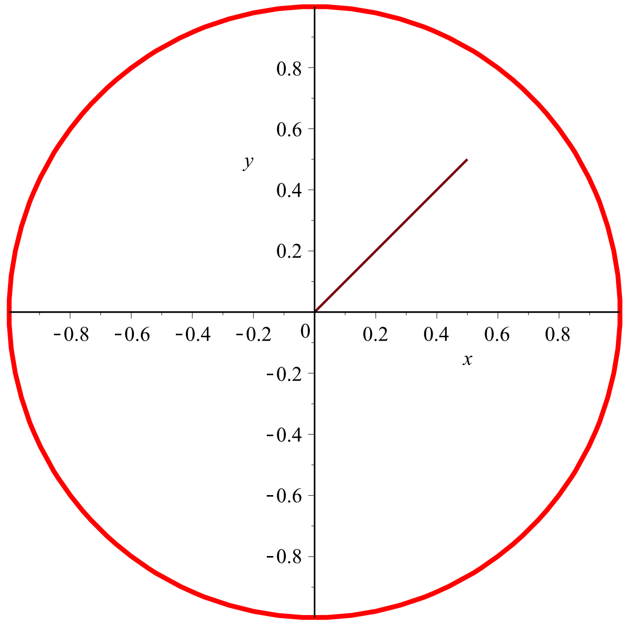
Válasszunk ki véletlenszeren egy *P* pontot egy egységnyi sugarú körlapon és jelöljük *Z*-vel a kör középpontjától mért távolságát! Határozzuk meg *Z* eloszlás és srségfüggvényét valamint számítsuk ki a várható értékét és varianciáját!

Megoldás

```
> restart;
```

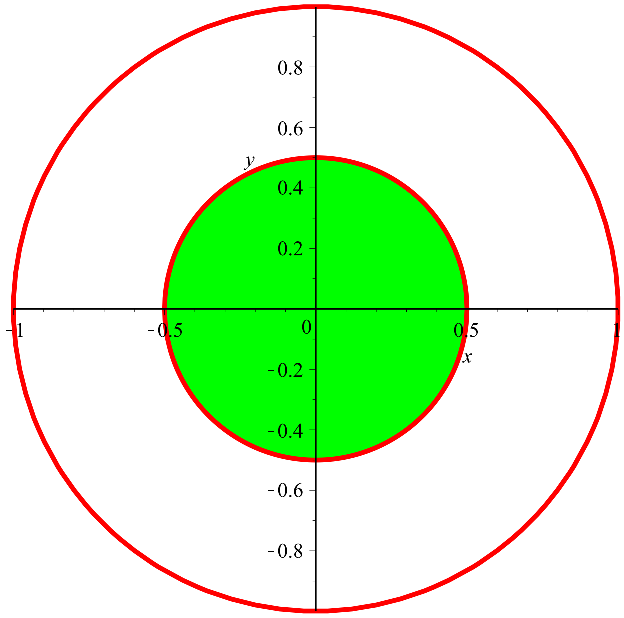
Az eseménytér: $\Omega = \{P(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \le 1\}$, a valószínségi mez geometriai. Az eseménytér geometriai mértéke: $m(\Omega) = 1^2 \cdot \pi = \pi$ Ezen definiáljuk a Z valószínségi változót, mely egy $P(x, y) \in \Omega$ ponthoz a $Z(P) = \sqrt{x^2 + y^2}$ értéket rendeli hozzá.

```
> P1 := plots[implicitplot](x^2 + y^2 = 1, x = -1..1, y =
  -1..1, scaling = constrained, color = red, thickness =
  3):
  P2 := plots[pointplot]([0.5, 0.5], color = black, symbol
  = cross, symbolsize = 30, thickness = 1):
  P3 := plot([[0, 0], [0.5, 0.5]]):
  plots[display](P1, P2, P3);
```



Az eloszlásfüggvény megadásához meg kell határozni az $F(z) = P(Z < z)$ valószínségeket, minden $z \in \mathbb{R}$ esetén (nyilván csak a $z \in]0;\, 1\,]$ tartomány az érdekes). Például az alábbi grafikonon látható a kedvez esetek halmazának geometriai reprezentációja $z = 0.5$ esetén, mely egy 0.5 sugarú kör:

```
> P4 := plot({sqrt(1/4 - x^2), -sqrt(1/4 - x^2)}, x =
  -0.5..0.5, color = red, filled = [color = green,
  transparency = 0.5], thickness = 3):
  P5 := plots[pointplot]([0.2, 0.2], color = black, symbol
  = cross, symbolsize = 30, thickness = 1):
  plots[display](P1, P4, P5, plot({}, x = -1..1));
```



Hasonlóan meggondolható, hogy a valószínség geometriai kiszámítási módja alapján $P(Z < z) = \dfrac{z^2 \cdot \pi}{m(\Omega)} = z^2$. Innen az eloszlásfüggvény:

$$F(z) = \begin{cases} 0 & z \le 0 \\ z^2 & 0 < z \le 1 \\ 1 & 1 < z \end{cases}$$

```
> F := z -> piecewise(z <= 0, 0, z <= 1, z^2, 1): 'F(z)' =
  F(z);
```

$$F(z) = \begin{cases} 0 & z \le 0 \\ z^2 & z \le 1 \\ 1 & \textit{otherwise} \end{cases} \quad \textbf{(2.2.13.1.1)}$$

A srségfüggvény az eloszlásfüggvény deriváltja:

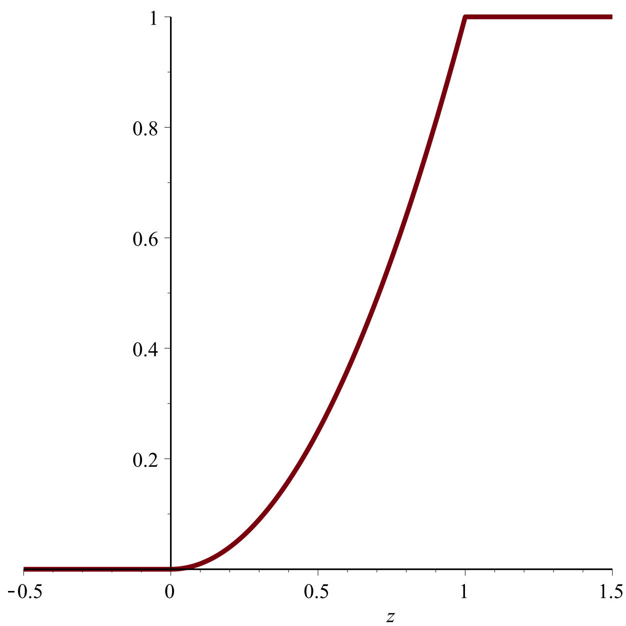
```
> f := z -> diff(F(z), z): 'f(z)' = f(z);
```

(2.2.13.1.2)

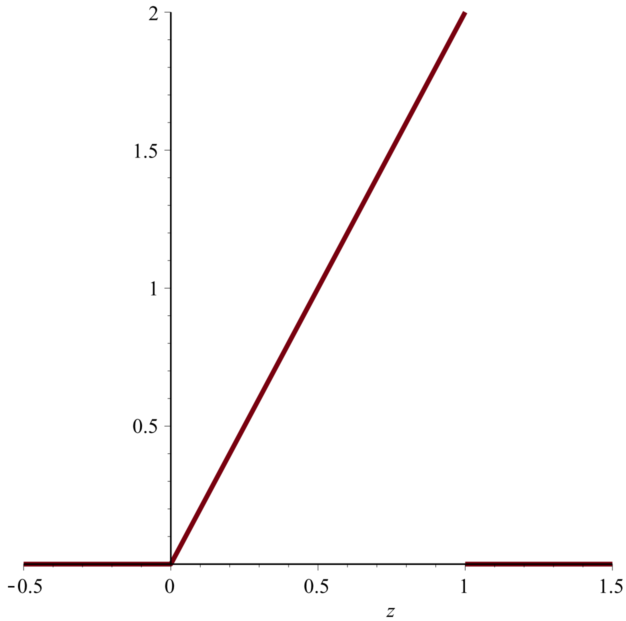
$$f(z) = \begin{cases} 0 & z \leq 0 \\ 2\,z & z < 1 \\ \mathit{undefined} & z = 1 \\ 0 & 1 < z \end{cases} \tag{2.2.13.1.2}$$

Láthatjuk, hogy a derivált véges sok pont kivételével létezik, ezért a X folytonos eloszlású.
Ábrázoljuk az eloszlás és srségfüggvényt!

```
> plot(F(z), z = -0.5..1.5, thickness = 3);
```




```
> plot(f(z), z = -0.5..1.5, thickness = 3, discont = true)
;
```



A várható értéket ill. a varianciát definíció szerint a srségfüggvény z-szeresének ill. $(z - E(Z))^2$-szeresének integrálja adja.

```
> EZ := int(z*f(z), z = 0..1); # Elég a [-infinity, infinity] intervallum helyett csak a [0, 1] intervallumon integrálni, mert azon kívül f(z) = 0.
```

$$EZ := \frac{2}{3} \quad \textbf{(2.2.13.1.3)}$$

Tehát az egységkör véletlen pontjának origótól való várható távolsága $\frac{2}{3}$ (és nem $\frac{1}{2}$!), ami jobban érthet, ha arra gondolunk, hogy a körvonalhoz lényegesen több pont van közel, mint a kör középpontjához. Nézzük most a varianciát:

```
> VarZ := int((z - EZ)^2*f(z), z = 0..1);
```

$$VarZ := \frac{1}{18} \quad \textbf{(2.2.13.1.4)}$$

Tehát a távolság varianciája $\frac{1}{18}$.

Gyakorló feladatok

Gy/1. Feladat (Két dobókocka minimuma / 2.)

Dobjunk fel két szabályos dobókockát és jelölje X a dobott számok közül a kisebbiket (egyenl dobás esetén bármelyiket)!
e) Számítsuk ki X várható értékét, a varianciáját és a szórását! Ábrázoljuk az eloszlást az X-p síkban pálcikadiagramon, a várható érték és a szórás fetüntetésével!
f) Számítsuk ki, hogy mekkora valószínséggel lesz Y kisebb a várható értékénél!

Megoldás

```
> restart;
```

Gy/2. Feladat (Összeg standardizáltja)

Legyenek X_1, X_2, ..., X_{100} független, azonos eloszlású valószínségi változók $\mu = 10$ várható értékkel és $\sigma = 2$ szórással. Adjuk meg az $S_{100} = \sum_{i=1}^{100} X_i$ és az $\overline{X} = \frac{S_{100}}{100}$ valószínségi változók várható értékét, szórását és standardizáltját!

Megoldás

```
> restart;
```

Gy/3. Feladat (Geometriai eloszlás)

Az X diszkrét valószínségi változó $p > 0$ paraméter geometriai eloszlású, ha $P(X=k) = p \cdot (1-p)^{k-1}$, ahol $k = 1, 2, 3, \ldots$. Határozzuk meg X várható értékét, szórását és második momentumát!

Megoldás

```
> restart;
```

Gy/4. Feladat (Normális eloszlás momentuma)

Határozzuk meg az (m, σ)-paraméter normális eloszlás várható értékét, szórását és 2. momentumát!

Megoldás

```
> restart;
```

Gy/5. Feladat (Összeg szórása, kovarianciája)

Legyenek X és Y független valószínségi változók, melyekrl tudjuk, hogy $E(X) = 3$, $E(Y) = -1$, $\sigma^2(X) = 1$ és $\sigma^2(Y) = 2$ (a szórásnégyzet megegyezik a varianciával, azaz $\sigma^2(X) = Var(X)$)! Határozzuk meg a $W = -5X + 2Y + 10$ valószínségi változó várható értékét és szórását! Mennyi $Cov(X, W)$ és $Corr(W, Y)$?

Megoldás

```
> restart;
```

Gy/6. Feladat (Korrelációs együttható lineáris függvényre)

Mutassuk meg, hogy ha X egy valószínségi változó és $Y = a \cdot X + b$, akkor X és Y korrelációs együtthatója 1, ha $a > 0$ és (-1), ha $a < 0$.

Megoldás

```
> restart;
```

Gy/7. Feladat (Négyzetes kifejezés várható értéke)

Egy X valószínségi változó várható értéke 12, szórása 5. Határozzuk meg X második momentumát és az $Y = 2\,X^2 - 4\,X + 7$ valószínségi változó várható értékét!

Megoldás

```
> restart;
```

Gy/8. Feladat (Várható érték és szórás eloszlástáblázatból)

Számítsuk ki a várható értéket, a varianciát és a szórást az alábbi táblázatokkal adott eloszlásokhoz!

a)

x_i	2	3	11
p_i	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

b)

x_i	-5	-4	1	2
p_i	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

Megoldás

```
> restart;
```

Gy/9. Feladat (Korreláció összeggel)

Legyenek X és Y független valószínségi változók, melyekre $Var(X) = \frac{3}{7}$ és $Var(Y) = \frac{4}{7}$.

Számítsuk ki X és $X + Y$ korrelációs együtthatóját!

Megoldás

```
> restart;
```

Gy/10. Feladat (Baráti társaság)

Egy szociológiai felmérésben három, egymással kapcsolatban nem lév baráti társaságot vizsgálnak. Az *A* társaságnak 5, a *B* társaságnak 7, a *C* társaságnak pedig 10 tagja van. Mekkora a baráti társaságok átlagos mérete? Várhatóan hány barátja van egy véletlenszeren kiválasztott személynek (egy társaságon belül mindenki mindenkinek barátja)?

Megoldás

```
> restart;
```