8. Gyakorlat

Valószínségi változók

Elméleti összefoglaló

A valószínségi változó egy kísérlet(sorozat) eredményét számszersíti: minden kimenethez egy valós számot rendel. A kísérlet kimenete gyakran önmagában is számszer érték (pl. kockadobás értéke). Ha a kimenet közvetlenül nem számszer, általában akkor is számokká kódolható (pl. pénzfeldobásnál a *fej*/*írás* kimeneteket $0/1$ értékekre válthatjuk).
Úgy képzelhetjük, hogy –bármi legyen is a kísérleti konstrukció– minden egyes kísérlet elvégzésekor felvillan az eredménynek megfelel valós szám a számegyenesen. Ezzel lehetvé válik, hogy a konkrét kísérleti konstrukciótól függetlenül vizsgáljuk a folyamat néhány fontos paraméterét, és olyan általánosított elméleti modelleket hozzunk létre, melyek valós kísérleti konstrukciókra ráhúzva egyszersítik azok kezelését.

Ha a valószínségi változó diszkrét (elkülönül) értékeket vehet fel, akkor diszkrét valószínségi változóról beszélünk. Az ilyen változó értékkészlete gyakran véges (kockadobás, pénzfeldobás), de lehet megszámlálhatóan végtelen is, pl. egy pénzérmével hányadik dobásra kapunk elször *fej*et. Az utóbbi érték nem korlátos, mert elméletileg nem kizárható, hogy 10, 100, 1000, vagy akárhány alkalommal *írás*t dobjunk, és csak utána következzen egy *fej*.
A diszkrét valószínségi változó minden értéke mögött valamely elemi esemény bekövetkezése áll. A változó értékeihez a háttérben meghúzódó események valószínségei társíthatók, melyek összege ezek szerint 1. Ezen valószínségek együttese a *diszkrét* valószínségi változó *eloszlása*.
A valószínűségi változó megadásába –az értékkészlet mellett– többnyire beleértjük a valószínségi eloszlás megadását is.

A folytonos valószínségi változó esetében is úgy képzeljük, hogy minden elvégzett kísérlet után egy –az adott kimenethez tartozó– számértéket regisztrálunk a számegyenesen. A diszkrét valószínségi változóhoz képest azonban most az értékkészlet folytonos. Ilyen folytonos értékkészlettel rendelkezik a műszaki-tudományos folyamatok sok paramétere – pl. ilyenek lehetnek egy megfigyelésben a tér- és idkoordináták. Az is elképzelhet, hogy ha egy diszkrét valószínségi változó értékkészlete "elég sr", akkor modellezhet folytonos változóval - ill. épp a "srítéssel" adódó határesetként értelmezhet a folytonos modell.
Az értékkészlet folytonossága eleve azt jelenti, hogy nem véges, st, nem is diszkrét pontokból áll, hanem *kontinuum* végtelen számosságú. Az elemi események "pontszerek", azaz *nullmértékek*: egy-egy konkrét kimenetel valószínsége zérus (pl. geometriai valószínségi mez). Nincs értelme egy-egy konkrét kimenet valószínségét kérdezni, csakis egy adott (véges vagy végtelen) intervallumba esés valószínűségét. El kell vetnünk azt a –diszkrét esetben működő– lehetőséget, hogy fel tudjuk sorolni a kimeneteket a hozzájuk tartozó valószínűségekkel.

A legegyszerbb elméleti eset, ha pl. egy Y folytonos valószínségi változóval a $[0, 1]$ intervallumból egyenletes eloszlással véletlen számértékeket választunk. Mérték-szemléletünk alapján jól érezzük, hogy egy adott intervallumra esés valószínsége arányos az intervallum hosszával, annak a teljes $[0, 1]$ intervallum hosszához viszonyított aránya. Pl. a $P(0.3 < Y < 0.5)$ valószínség (a 0.3 és 0.5 értékek közül való választás valószínsége) 0.2,

mivel ennyi a megfigyelt intervallum hosszának a teljeshez viszonyított aránya.
De mennyi az $Y=0.5$ érték választásának valószínsége, azaz $P(Y=0.5)$? Ennek valószínségi (geometriai) mértéke nulla: az esemény *pontszer*, a kiválasztási valószínség zérus. A példa érdekessége, hogy ugyanakkor ez az esemény nem lehetetlen: nem kizárható, hogy akár ez is kijöhet eredményként (gondoljunk arra, hogy végülis minden ilyen sorsoláskor kijön egy érték, amelyre ugyanez elmondható). A nulla valószínség nem azt jelenti, hogy kizárt az esemény bekövetkezése, hanem csupán azt, hogy nagyon sok kísérlet mellett sem mernénk arra fogadni, hogy az esetek egy bármilyen kis hányadában ($\frac{1}{100}$, $\frac{1}{1000}$ stb.) átlagosan be fog következni.

A diszkrét eloszlás esetén a lehetséges kimeneteket (és a hozzájuk tartozó valószínségeket) fel tudjuk sorolni (akár végtelen értékkészlet esetén is egy végtelen sorozatként). Az eloszlás szemléltethet pálcikadiagrammal (=vonaldiagrammal) vagy (lépcss) eloszlásfüggvénnyel, ami eleinte csak egy újabb szemléltetési alternatívának tnik. Folytonos esetben azonban kiderül, hogy az eloszlásfüggvény jóformán az egyetlen lehetség a konstrukció modellezéséhez, ill. hogy a pálcikadiagramnak is létezik általánosítása –a sűrűségfüggvény– közöttük pedig kapcsolat van. A folytonos modell kezeléséhez analízisbl ismert eszköztárra (függvények, differenciál- és integrálszámítás) van szükség.

Kidolgozott példafeladatok

A *diszkret_eloszlasfuggveny* eljárás

Az alábbi eljárással adott diszkrét eloszláshoz tartozó eloszlásfüggvényeket ábrázolhatunk (a szakaszok nyílt/zárt végei üres/tömött karikákal vannak jelölve).

```
> 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:
```

1. Feladat (Kockadobás / 1.)

Dobjunk egy (6 oldalú) dobókockával, jelölje X a dobott számot.
a) Írjuk fel az X valószínségi változó eloszlását és szemléltessük pálcikadiagramon!
b) Végezzünk szimulációt 100 kockadobással és ábrázoljuk hisztogramon az eredményt! Hasonlítsuk össze az a) részben kapott eloszlással!
c) Rajzoljuk fel X eloszlásfüggvényét!

Megoldás

```
> restart;
```

a) Írjuk fel az X valószínségi változó eloszlását és szemléltessük pálcikadiagramon!

A kockadobás kísérlet közvetlenül egy 1-6 értékek közti számszer eredményt ad (ettl még nem kötelez, hogy a valószínségi változó értékei ezek legyenek, de itt most pont megfelel). Az X valószínségi változó lehetséges értékeit egy listában rögzítjük:

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

Az egyes értékek (mögött álló események) bekövetkezési valószínségeit szintén listában rögzíthetjük. Jelen esetben ez igen egyszer:

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

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
```

A valószínségek összegének minden eloszlás esetén 1-et kell adnia. Ellenrizzük!

```
> sum(px[i], i = 1..6);
```

A diszkrét valószínségi változó eloszlása jól szemléltethet pálcikadiagramon a *DynamicSystems* csomag *DiscretePlot* parancsával:

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

b) Végezzünk szimulációt 100 kockadobással és ábrázoljuk hisztogramon az eredményt! Hasonlítsuk össze az a) részben kapott eloszlással!

Definiáljunk egy valószínségi változót, melynek eloszlása megegyezik a feladatban megadottal:

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

Vegyünk 100 elem mintát X-bl és ábrázoljuk hisztogramon (szintén a Statistics csomag kell hozzá).

```
> minta := Sample(X, 100);
```

Diszkrét valószínségi változó esetén a hisztogram egy olyan oszlopdiagram, mely az értékek gyakoriságának eloszlását szemlélteti egy adott mintában. Az egyes oszlopok magassága gyakoriság-hisztogram esetén magával a darabszámmal egyenl, relatív gyakoriság hisztogramnál pedig a relatív gyakorisággal. Alapértelmezés szerint a Maple relatív gyakoriság hisztogramot rajzol, vagyis az oszlopok össz-magassága 1; az egységnyi magasság "kendik szét" az eloszlásra jellemz alakban.

```
> 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.

c) Rajzoljuk fel X eloszlásfüggvényét!

A véges diszkrét eloszlás összes adata tárolható két vektorban (x = értékek, px = valószínségek). Nem ez azonban az egyetlen mód az adatok tárolására, ráadásul folytonos változók eloszlásának leírására egyáltalán nem is alkalmas. Olyan matematikai objektumot kell konstruálni, mely képes leírni tetszleges valószínségi eloszlást. Erre alkalmas egy megfelelen definiált egyváltozós függvény, az ún. *eloszlásfüggvény*.

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.

Ennek megrajzolásához használjuk a fent definiált *diszkret_eloszlasfuggveny* eljárást (próbáljuk meg értelmezni a függvény mködését, majd értékeljük ki a definícióját ENTER-rel, hogy bekerüljön a memóriába).

```
> diszkret_eloszlasfuggveny(x, px);
```

Azt látjuk, hogy a kockadobás valószínségi változójának eloszlásfüggvénye lépcszetesen emelkedik 0-tól 1-ig (ez más egyéb diszkrét eloszlásoknál is így van). Azért ilyen, mert két egész érték (pl. 3 és 4) között nem vehet fel értéket a változó, így annak valószínsége, hogy az érték kisebb pl. 3.1-nél vagy 3.99-nél, egyaránt 0.5 (3 és 4 között végig ezt mutatja a függvény). A következ lépcs a 4 átlépésekor következik, hiszen pl. 4.1 esetén már $\frac{4}{6} = \frac{2}{3} \approx 0.67$ a valószínsége, hogy a dobott érték 4.1-nél kisebb. Minden lépcsfok magassága annyi, amennyi az adott lépésnél lév érték bekövetkezésének valószínsége. Esetünkben minden lépcsfok magassága $\frac{1}{6} \approx 0.17$.

Ráadás

A c) feladatrészt egyszerbben megoldhatjuk a Statistics csomag segítségével (bár a függvény láthatóan elnagyoltabb lesz, pl. nem lesz balról folytonos).

Állítsuk el az X valószínségi változó eloszlásfüggvényét, majd ábrázoljuk is azt!

```
> unassign('x'); # változóként szeretnénk használni x-et,
  ezért töröljük az értékét
  cdf_X := CDF(X, x); # az eloszlásfüggvény (CDF - Common
  Distribution Function)
```

```
> plot(cdf_X, x=-2..9, thickness = 3, discont = true); #
  eloszlásfüggvény ábrázolásánál mindig érdemes
  bekapcsolni a 'discont' opciót, mely a nem-folytonos
  függvények helyes ábrázolásáért felel
```

Az eloszlás is megkapható beépített függvénnyel, de a plot parancs sajnos nem tudja pálcikadiagramszeren ábrázolni.

```
> pdf_X := PDF(X, x); # a srségfüggvény (PDF -
  Probability Density Function)
```

```
> plot(pdf_X, x = -2..9, thickness = 3);
```

2. Feladat (Dobás két kockával / 1.)

Dobjunk két dobókockával egyszerre! Az *X* valószínségi változóban jegyezzük fel a két dobott érték maximumát!
a) Adjuk meg a valószínségi változó értékkészletét és az értékekhez tartozó valószínségeket!
b) Ábrázoljuk az eloszlást az *X* - *p* síkban pálcikadiagrammal!
c) Végezzünk szimulációt 1000 kísérlettel! Készítsük el a kapott minta hisztogramját és vessük össze a b) részben kapott pálcikadiagrammal!
d) Rajzoljuk meg az eloszlásfüggvényt!

Megoldás

```
> restart;
```

a) Adjuk meg a valószínségi változó értékkészletét és az értékekhez tartozó valószínségeket!

A valószínségi változó értékkészlete 1-t1 6-ig terjed (a feljegyzett maximum 1 is lehet, ha mindkét dobás 1-es, de persze ez viszonylag ritkán fordul el).

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

Mivel két kockával dobunk, a lehetséges kimenetelek szemléletesen összefoglalhatók egy mátrixban (2 dimenziós táblázatban):

```
> dobasok := matrix(6, 6, (i, j) -> [i, j]);
```

A modell megkülönböztethetnek feltételezi a két kockát. Minden elemi kimenetelbl (számpárból) a maximum jegyeztetik fel. Könnyen megjeleníthetjük ezt is egy mátrixban:

```
> maximumok := matrix(6, 6, (i, j) -> max(i, j));
```

Ebben a táblázatban jól látszik –és könnyen összeszámlálható–, hogy az egyenlő valószínség kimenetelek közül hány vezet 1-es, hány 2-es stb. maximumra. Ezek nyomán a valószínségi változó értékeihez tartozó valószínségek vektora:

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

b) Ábrázoljuk az eloszlást az *X* - *p* síkban pálcikadiagrammal!

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

Szépen láthatók az egyenletesen növekv valószínségek.

c) Végezzünk szimulációt 1000 kísérlettel! Készítsük el a kapott minta hisztogramját és vessük össze a b) részben kapott pálcikadiagrammal!

```
> with(Statistics):
  randomize():
> Y := RandomVariable(DiscreteUniform(1, 6));
  Z := RandomVariable(DiscreteUniform(1, 6));
> X := max(Y, Z);
```

Vegyünk 1000 elem mintát *X*-bl és ábrázoljuk hisztogramon (Statistics csomag kell hozzá).

```
> minta := Sample(X, 1000);
> 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.

d) Rajzoljuk meg az eloszlásfüggvényt!

Használjuk újra a munkalap elején definiált eljárást (újra értékeljük ki ENTER-rel a definícióját)!

```
> diszkret_eloszlasfuggveny(x, px);
```

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). Jól látszik, hogy az egyre növekv valószínségeknek megfelelen egyre magasabbak a lépcsfokok.

Az eloszlásfüggvényt ebben a példában nem tudjuk alternatív módon, a Statistics csomag segítségével meghatározni, mert a két valószínségi változó maximumának eloszlását a Maple nem tudja ilyen módon kiszámolni:

```
> cdf_X := CDF(X, t);
```

Elkészíthetjük viszont a kumulált valószínségek listáját, majd a *DynamicSystems DiscretePlot* eljárásával ábrázolhatjuk az eloszlásfüggvényt:

```
> N := 6;
  Fx := [seq(sum(px[i], i = 1..n), n = 1..N)]; # kumulatív
  (összegzett) valószínségek
> xx := [x[1]-3, op(x), x[N]+3];
  Fxx := [0, op(Fx), 1]; # az op() paranccsal sorozattá
  alakíthatunk egy listát, ami ezután könnyen
  beilleszthet egy vesszkkel elválasztott felsorolásba
> DiscretePlot(xx, Fxx, style = stair, color = blue,
  thickness = 3);
```

3. Feladat (Egyenletes eloszlás / 1.)

Válasszunk ki véletlenszeren (geometriai valószínség szerint) egy Y számot a $[0, 1]$ intervallumból!

a) Vegyünk egy $n=100$ elem mintát és ábrázoljuk annak eloszlását pontdiagramon, számegyenesen illetve hisztogramon!

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

c) Határozzuk meg és ábrázoljuk az Y valószínségi változó srségfüggvényét!

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

e) Vegyünk egy $n=100000$ elem mintát és ábrázoljuk közös grafikonon a hisztogramot és az (elméleti) srségfüggvényt!

Megoldás

```
> restart;
```

a) Vegyünk egy $n=100$ elem mintát és ábrázoljuk annak eloszlását pontdiagramon, számegyenesen illetve hisztogramon!

Amikor a $[0, 1]$ intervallumból véletlenszeren választunk egy számot, akkor egy ún. egyenletes eloszlású valószínségi változó modellje segítségével sorsolunk értékeket.

Létrehozzuk a követelménynek megfelel véletlen változót, hogy aztán kísérletezhessünk vele:

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

Generáljunk egy 100 elem mintát:

```
> n := 100; minta := Sample(Y, n);
```

Próbaként kiíratjuk az els elemet:

```
> minta[1];
```

(Megpróbálhatjuk, hogy a mintát generáló sort újra lefuttatva a mintákat tartalmazó lista megváltozik: a kiíratást ismét végrehajtva (ENTER) más érték van az 1. pozícióban.)
Láthatóvá tehetjük a kisorsolt értékeket, ha a sorszám függvényében ábrázoljuk a kisorsolt értékeket (pontdiagram). Ehhez számpárokból álló listát hozunk létre: minden i sorszám mellé a lista hozzá tartozó értékét (az i. lépésben kisorsolt véletlen számot) rendeljük.

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

Ha a sort kettspont helyett pontosvesszvel zárjuk le, kimenetként megjelenik a teljes generált lista, de éppen ennek elkerülésére alkalmas a kettspont - a parancs lefut, de nem terheli a képernyt nagy mennyiség adattal. Ha viszont valamit ellenrizni szeretnénk az adatok között, bármikor pontosvesszre cserélhetjük.
Ezt a koordinátalistát már fogadni tudja a rajzolóutasítás:

```
> plot(pontok, style = point);
```

A vízszintes tengelyen a sorszámozás halad, a függleges tengely a kisorsolt értékeket jelzi. Mivel a sorsoláshoz használt valószínségi változó egyenletes eloszlású a $[0, 1]$ intervallumon, a pontok egyenletesen oszlanak el - nem érezzük, hogy bárhol szignifikánsan srsödnének.
Megkísérelhetjük változtatni a minta elemszámát, és azzal idáig újra lefuttatni a lépéseket (kísérletezés után állítsuk vissza az eredeti ($n = 100$) értéket).
Fordított tengely-kiosztással is meg lehet csinálni ugyanezt:

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

Az áttekinthetségnek használ, ha az egyes értékeket a sorszám függvényében ábrázoljuk, de a kiválasztott értékek valójában a $[0, 1]$ szakaszon srsödnek:

```
> pontok_szamegyenesen := [seq([minta[i], 0], i = 1..n)]:
  pontok_plot := plot(pontok_szamegyenesen, style = point,
  symbolsize = 15, color = red): pontok_plot;
```

A generált mintából könnyen készíthetünk hisztogramot is:

```
> H := Histogram(minta, bincount = 10): H;
```

A hisztogram egy olyan oszlopdiagram, mely egy adott intervallumról választott értékek gyakoriságának eloszlását szemlélteti (folytonos eloszlás esetén). A szóban forgó $[0, 1]$ intervallumot *vízszintesen* vesszük fel, részekre osztjuk, majd megszámoljuk, hogy hány érték esik az egyes részintervallumokba. Ezek után a részintervallumokra mint alapokra olyan téglalapokat emelünk, melyek *területe* arányos az oda eső darabszámmal – gyakoriság-hisztogram esetén magával a darabszámmal egyenl, relatív gyakoriság hisztogramnál pedig a relatív gyakorisággal. Egy-egy oszlop magasságát tehát úgy kapjuk, hogy a megjelenítend értéket elosztjuk az oszlop alapjául szolgáló részintervallum hosszával (azonos hosszúságú részintervallumok esetén a területek aránya persze egyszeren a magasságok arányában mutatkozik meg). Alapértelmezés szerint a Maple relatív gyakoriság hisztogramot rajzol, vagyis az oszlopok össz-területe 1; az egységnyi terület "kendik szét" az eloszlásra jellemz alakban.

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

A kisorsolt pontokat rávetítve látjuk, hogy az intervallumon kisorsolt értékek összeszámlálásából hogy keletkeznek a hisztogram oszlopai.
A hisztogram imént vázolt konstrukciójából fakad, hogy az oszlopok össz-területe gyakoriság-hisztogramnál az összes adat darabszámával egyenl, relatív gyakoriság hisztogramnál viszont az össz-terület 1. Azonos paraméterekkel végrehajtott különböz kísérletsorozatok részben eltér hisztogramokat eredményeznek, de ezek össz-területe mindig megegyezik. Olyan, mintha egy adott állandó mennyiséget "kennénk szét" egy-egy kísérletsorozatban. Hasonló kísérletsorozatok hisztogramjai alakilag hasonlítanak egymásra, a hisztogramok alakja pedig jellemz a kísérletsorozat típusára.

Megjegyzés: A hisztogram felbontásának finomságát ésszeren, az adatmennyiséghez viszonyítva kell meghatározni. Esetünkben a 100 elem mintához választott 10 részes felbontás jól érzékelteti, hogy az eloszlás nagyjából egyenletes, de vannak ingadozások.
Durvább felbontásnál természetesen csökken az információtartalom:

```
> Histogram(minta, bincount = 2);
```

Az adatmennyiséghez képest túlzottan nagy felbontásnál viszont a hisztogram túlzottan "ideges" lesz, a lokális egyenetlenségek annyira felnagyítódnak, hogy már akadályozzák az alak lényegének, alapvet jellegének felismerését:

```
> Histogram(minta, bincount = 50);
```

b) Határozzuk meg és ábrázoljuk az *Y* valószínségi változó eloszlásfüggvényét!
Jelen esetben minden negatív értékre az eloszlásfüggvény értéke zérus, hiszen mivel Y értékei 0 és 1 közöttiek (zérus valószínséggel tud az érték kisebb lenni bármely negatív számnál). Az is könnyen adódik, hogy 1-nél nagyobb x értékekre az eloszlásfüggvény értéke már konstans 1, hiszen mivel Y maximális értéke 1, minden 1-nél nagyobb x értékre $100\ \%$ valószínséggel kisebb lesz az adott x-nél.
Egy $[0, 1]$-beli x-re $F(x) = P(Y < x)$ az a valószínség, mellyel a kisorsolt érték x-nél kisebb, márpedig ez a valószínség geometriai kiszámítási módja alapján a $[0, x]$ intervallum hosszának a teljes $[0, 1]$ hosszához viszonyított aránya, ami $\frac{(x-0)}{(1-0)} = \frac{x}{1} = x$. Vagyis $[0, 1]$-en $F(x) = x$. Mindezek alapján a $[0, 1]$-en egyenletes eloszlású Y valószínségi változó eloszlásfüggvénye:

```
> F := x -> piecewise(x <= 0, 0, x <= 1, x, 1); # a
  piecewise paranccsal adhatunk meg szakaszonként
  értelmezett függvényt
  plot(F(x), x = -1..2, y = 0..1, scaling = constrained,
  thickness = 3);
```

Az eloszlásfüggvény jellegzetessége, hogy a grafikonján balról jobbra haladva 0-ról monoton módon felér 1-ig (ezt a tulajdonságot mutatja a diszkrét eloszlások lépcss függvénye is, csak az nem folytonos, míg ez igen).
Megjegyzés: Az eloszlásfüggvény a $P(Y < x)$ típusú kérdésre közvetlenül választ tud adni: $P(Y < x) = F(x)$.
Közvetve viszont más típusúakra is:
- $P(Y \geq x) = 1 - P(Y < x) = 1 - F(x)$;
- $P(x_1 \leq Y < x_2) = P(Y < x_2) - P(Y < x_1) = F(x_2) - F(x_1)$.

c) Határozzuk meg és ábrázoljuk az *Y* valószínségi változó srségfüggvényét!
A hisztogram finomításával (a kísérletek számának és így a felbontásnak a folyamtos növelésével) az ábra "kisimul", az oszlopok teteje egy görbéhez (jelen esetben egyeneshez) közelít. Ez a grafikonja az Y valószínségi változó ún. srség-függvényének. A

srségfüggvény tehát a hisztogram határértékeként értelmezhet. Ha egy adott x_0 hely közelébe (az egyszerség kedvéért mondjuk jobbról) felveszünk egy x értéket a tengelyen, akkor egy hisztogramnak az $[x_0, x]$ intervallumra állított oszlopának *területe* az Y változó ezen intervallumba esésének valószínsége, ami az eloszlásfüggvény segítségével számolható: $P(x_0 \le Y < x) = P(Y < x) - P(Y < x_0) = F(x) - F(x_0)$

Az oszlop magassága a területének az alapjával vett hányadosa: $m = \dfrac{F(x) - F(x_0)}{x - x_0}$

A hisztogram folyamatos finomításával (az oszlopok keskenyedésével) az oszlopok magassága egyre jobban közelíti a határesetként adódó srségfüggvény értékéhez:

$$f(x_0) = \lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0} = F'(x_0)$$

A srségfüggvénynek ez a definíciója láthatóan megegyezik az eloszlásfüggvény x_0-nál vett differenciálhányadosával! Vagyis az derült ki, hogy a srségfüggvény az eloszlásfüggvény deriváltja (természetesen az eloszlásfüggvény differenciálható helyeinél). Jelen esetben az eloszlásfüggvény 0-nál és 1-nél láthatóan nem differenciálható (mert "szögletes"), a srségfüggvénynek ott szakadása van (de az eloszlásfüggvény legalább folytonos). (Ennél súlyosabb eset a diszkrét eloszlásoké, ahol a legfontosabb helyeknél az eloszlásfüggvény nemhogy nem differenciálható, de még csak nem is folytonos.) A srségfüggvényt ilyenkor külön objektummal (pálcikadiagrammal) tudjuk csak ábrázolni.)

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

A függvény szakaszonként folytonos, de a három folytonos szakasza között két (elsfajú) szakadása van.

A srségfüggvény jellegzetessége, hogy végig pozitív érték, "szélein" 0-hoz közelít, az alatta lév összes terület pedig 1. Úgy képzeljük, hogy a teljes 100 % valószínség "szét van kenve" a függvény alatt, és azok a gyakoribb elforduló értékek a számegyenesen (x tengely), ahol nagyobb a függvényérték.

Megjegyzés: Szakaszonként folytonos függvény a Maple-ben a következ, szakaszonkénti definícióval adható meg:

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

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

A Maple beépített eszközöket is tartalmaz valószínségi változó eloszlás- srségfüggvényének meghatározására.

Az eloszlásfüggvény (CDF - Common Distribution Function):

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

A srségfüggvény (PDF - Probability Density Function):

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

e) Vegyünk egy $n = 100000$ elem mintát és ábrázoljuk közös grafikonon a hisztogramot és az (elméleti) srségfüggvényt!

A kísérletek számának növelésével már finomabb felbontású hisztogram adódik, mely jobb képet mutat a háttérben meghúzódó eloszlásról. Egyúttal egyre inkább közeledünk a valószínségi eloszlás háttérben megbúvó srség-függvényéhez is:

```
> n := 100000;
  minta := Sample(Y, n);
> H := Histogram(minta, bincount = 50):
  suruseg_plot := plot(f(x), x = 0..1, thickness = 3):
  f(x);
  plots[display](H, suruseg_plot);
```

4. Feladat (Nem egyenletes eloszlás / 1.)

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!

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

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

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

> 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!

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

Mivel maga Y egyenletes eloszlású, ezért $F(x) = P(Y < \sqrt{x}) = \sqrt{x}$ 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);
```

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);
```

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

e) Számítsuk ki kétféleképpen a $P\left(\frac{1}{3} \leq Z \leq \frac{3}{4}\right)$ 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(%);
```

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(%);
```

Ugyanazt az értéket kaptuk, ahogy az várható is volt.

5. Feladat (Poisson-eloszlás)

Legyen a $\lambda > 0$ pozitív szám, és tekintsük a $p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ számokat, ahol $k = 0, 1, 2, \ldots$

a) Mutassuk meg, hogy a p_k $(k = 0, 1, 2, \ldots)$ valószínségek egy diszkrét eloszlást határoznak meg! Ezt az eloszlást nevezzük λ paraméter *Poisson-eloszlás*nak.

b) Rajzoljuk fel az eloszlást és az eloszlásfüggvényt a $[0, 12]$ intervallumon $\lambda = 0.5, 1, 2$ és 5 értékekre!

c) Mekkora a valószínsége, hogy $X \leq 5$, ha $\lambda = 2$?

Megoldás

```
> restart;
```

a) Mutassuk meg, hogy a p_k $(k = 0, 1, 2, \ldots)$ valószínségek egy diszkrét eloszlást határoznak meg! Ezt az eloszlást nevezzük λ paraméter *Poisson-eloszlás*nak.

Felírjuk függvényként az eloszlást és ellenrizzük, hogy a valószínségek összege 1-et ad.

```
> assume(lambda > 0); # ez a feltétel kell, különben lenne
  negatív valószínség is; az ilyen jelleg deklarált
  feltételek a Maple-nek is segítenek a bevitt képletek
  értelmezésében
> p := k -> lambda^k/k!*exp(-lambda);
> sum(p(k), k = 0..infinity);
```

b) Rajzoljuk fel az eloszlást és az eloszlásfüggvényt a $[0, 12]$ intervallumon $\lambda = 0.5, 1, 2$ és 5 értékekre!

Elször készítsük el a lehetséges értékek listáját és a hozzájuk tartozó valószínségek listáit $\lambda = 0.5, 1, 2$ és 5 értékekre!

```
> x := [seq(k, k = 0..12)];
> px := [seq(p(k), k = 0..12)]; # valószínségek általános
  lambdára; k=12-nél megállunk (elméletileg azonban
  végtelen a sorozat)
> `px_0.5` := eval(px, lambda=0.5); # behelyettesítjük
  lambda aktuális értékét az eval paranccsal
> px_1 := eval(px, lambda=1.0); # fontos, hogy lambda
  float típusú legyen, különben a DiscretePlot nem fog
  mködni
> px_2 := eval(px, lambda=2.0);
```

```
> px_5 := eval(px, lambda=5.0);
```

Ábrázoljuk az egyes eloszlásokat a *DiscretePlot* eljárással és az eloszlásfüggvényeket a korábban definiált *diszkret_eloszlasfuggveny* eljárással (elbb futtassuk le újra a definícióját).

$\lambda = 0.5$:

```
> with(DynamicSystems):
  DiscretePlot(x, `px_0.5`, style = stem, color = blue,
  thickness = 3);
> diszkret_eloszlasfuggveny(x, `px_0.5`);
```

$\lambda = 1$:

```
> DiscretePlot(x, px_1, style = stem, color = blue,
  thickness = 3);
> diszkret_eloszlasfuggveny(x, px_1);
```

$\lambda = 2$:

```
> DiscretePlot(x, px_2, style = stem, color = blue,
  thickness = 3);
> diszkret_eloszlasfuggveny(x, px_2);
```

$\lambda = 5$:

```
> DiscretePlot(x, px_5, style = stem, color = blue,
  thickness = 3);
> diszkret_eloszlasfuggveny(x, px_5);
```

c) Mekkora a valószínsége, hogy $X \leq 5$, ha $\lambda = 2$?

Mivel $\lambda = 2$, ezért a px_2 eloszlással kell számolni.

$X \leq 5$ azt jelenti, hogy X a $\{0, 1, 2, 3, 4, 5\}$ halmazból vehet fel értékeket (X értékkészlete a természetes számok halmaza).

$$P(x \leq 5) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

```
> `P(X <= 5)` := sum(px_2[k + 1], k = 0..5); # az index k
  + 1, mert az indexelés 1-tl indul, nem 0-tól
```

Tehát $\lambda = 2$ esetén $P(X \leq 5) = 0.983$.

6. Feladat (Eloszlások, eloszlás- és srségfüggvények)

a) Döntsük el, hogy az alábbi táblázatok közül melyik mutathatja egy diszkrét valószínségi változó eloszlását!

(A)

x_i	**0**	**5**	**6**	**10**
p_i	**0.2**	**0.4**	**0.4**	**0.1**

(B)

x_i	**0.3**	**0.5**	**0.7**	**0.9**
p_i	**0.1**	**0.4**	**0.3**	**0.2**

(C)

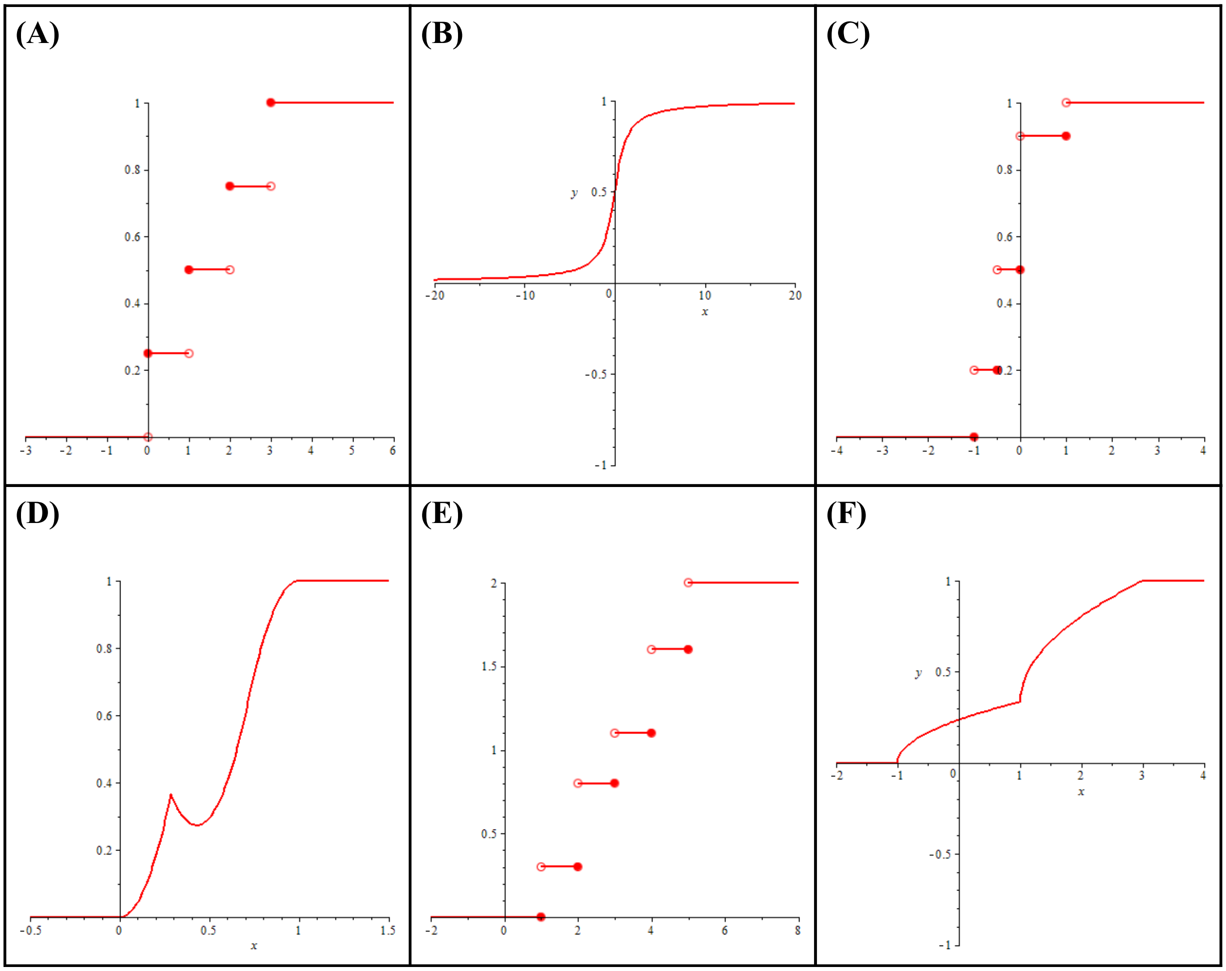
x_i	**−1**	**−0.5**	**0.5**	**1**

p_i	0.3	1.3	−0.8	0.2

(D)

x_i	−5
p_i	1.0

b) Válasszuk ki az alábbi grafikonok közül azokat, melyeken egy valószínségi változó eloszlásfüggvénye látható! A kiválasztottak közül melyek tartoznak diszkrét- és melyek folytonos eloszláshoz?



c) Válasszuk ki az alábbi függvények közül a srségfüggvényeket!

(A) $$f_1(x) = \begin{cases} 0 & x \le 0 \\ \frac{3}{2}\,\mathrm{e}^{-x} & 0 < x \end{cases}$$

(B) $$f_2(x) = \frac{1}{2}\tanh\left(\frac{x^3+1}{x^2+3}\right) + \frac{1}{2}$$

(C)

$$f_3(x) = \begin{cases} 0 & x \le 0 \\ \frac{24}{5}x & 0 < x \le \frac{1}{2} \\ \frac{16}{5} - \frac{16}{5}x & \frac{1}{2} < x \le 1 \\ 0 & 1 < x \end{cases}$$

Megoldás

```
> restart;
```

a) Döntsük el, hogy az alábbi táblázatok közül melyik mutathatja egy diszkrét valószínségi változó eloszlását!

(A) Rögzítsük a táblázat adatait két vektorban: x tárolja az értékeket, p pedig a hozzájuk tartozó valószínségeket

```
> n := 4;
  x := [0, 5, 6, 10];
  p := [0.2, 0.4, 0.4, 0.1];
```

Ahhoz, hogy eloszlást kapjunk két tulajdonságnak kell teljesülnie:

$$\textbf{(i)}\ \ 0 \le p_i \ (i = 1, 2, ..., n)$$

$$\textbf{(ii)}\ \sum_{i=1}^{n} p_i = 1$$

Könnyen látható, hogy p minden eleme nemnegatív. Az összegük viszont nagyobb, mint 1:

```
> sum(p[i], i=1..n);
```

Tehát a táblázat nem ad eloszlást.

(B) Hasonlóan ellenrizzük mint az (A) esetet.

```
> n := 4;
  x := [0.3, 0.5, 0.7, 0.9];
  p := [0.1, 0.4, 0.3, 0.2];
```

Minden valószínség nemnegatív. Az összegük:

```
> sum(p[i], i=1..n);
```

Így ez egy valószínségi eloszlás.

(C)

```
> n := 4;
  x := [-1, -0.5, 0.5, 1];
  p := [0.3, 1.3, -0.8, 0.2];
> sum(p[i], i=1..n);
```

Habár az összeg 1, ez nem lesz eloszlás, mert $p_3 < 0$ és valószínség nem lehet negatív!

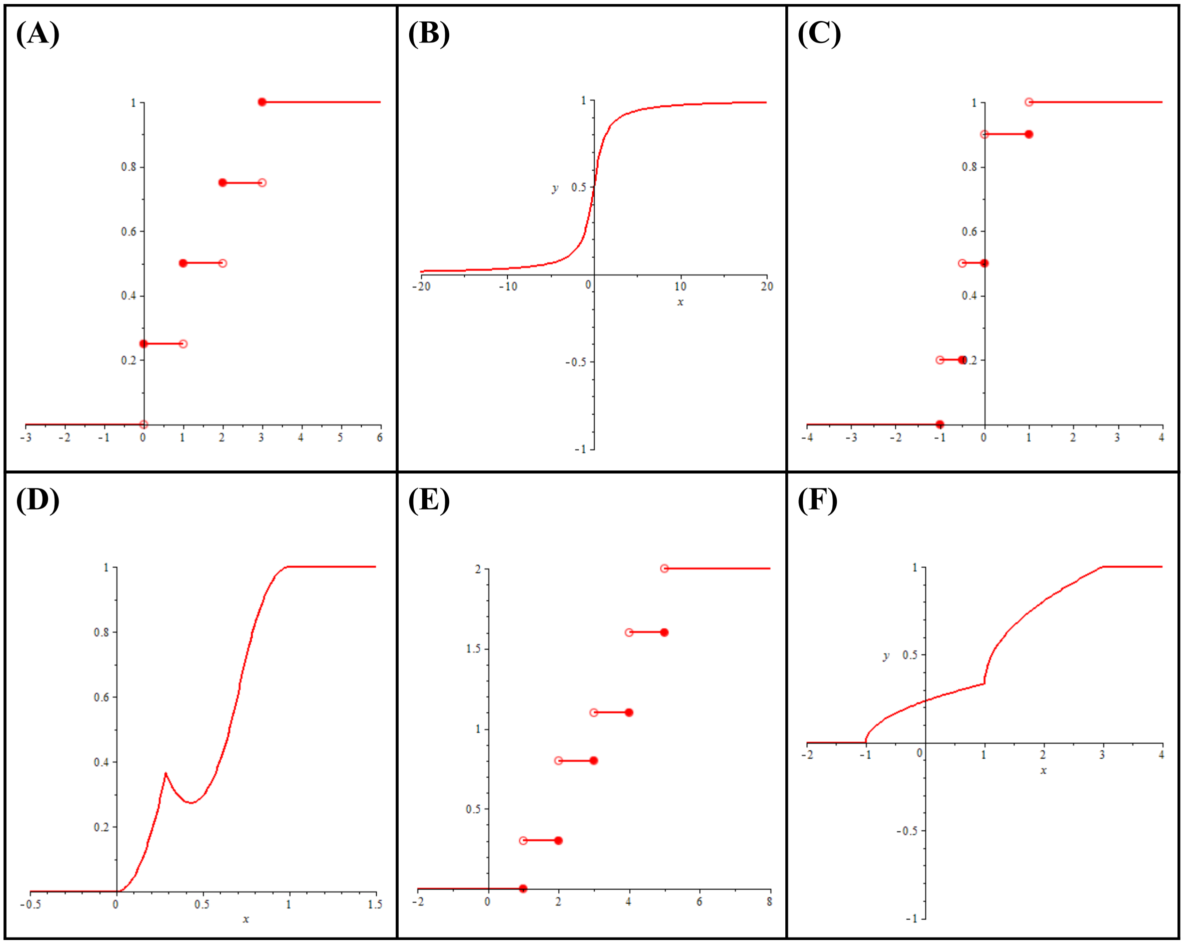
(D)

```
> n := 1;
  x := [-5];
  p := [1];
> sum(p[i], i=1..n); # ez persze nyilvánvalóan 1
```

Ez tehát eloszlás lesz! (Az nem probléma, hogy csak egy értéket vehet fel, a -5-öt!)

b) Válasszuk ki az alábbi grafikonok közül azokat, melyeken egy valószínségi változó eloszlásfüggvénye látható! A kiválasztottak közül melyek tartoznak diszkrét-

és melyek folytonos eloszláshoz?



Az eloszlásfüggvényekre vonatkozó szükséges és elégséges feltételek teljesülését kell ellenrizni. Ezek a következk:

(0) Értelmezési tartomány: $\mathbb{R}$ (valós számok halmaza)

(i) Értékkészlet: $0 \le F(x) \le 1$, minden $x \in \mathbb{R}$ esetén

(ii) Monotonitás: $F(x)$ monoton n (monoton nem csökken)

(iii) Végtelenben vett határértékek: $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$

(iv) Balról folytonosság: $\lim_{x \to a^-} F(x) = F(a)$, minden $a \in \mathbb{R}$ esetén

Megjegyzés: Az (i) feltétel következik (ii)-bl és (iii)-bl, ezért akár el is hagyható.

(A) Nem eloszlásfüggvény, mert nem balról folytonos (hanem jobbról). Ezt abból látjuk, hogy a folytonos szakaszok bal oldalán van tömött karika, nem a jobb oldalon.

(B) Erre a függvényre minden feltétel teljesül, ezért eloszlásfüggvény. Mivel nem szakad vagy törik meg, ezért minden pontban deriválható, így folytonos eloszláshoz tartozik (még az is beleférne, ha véges sok pontban megtörne, de szakadások nem megengedettek).

(C) Erre a függvényre is teljesülnek a fenti feltételek (különösebb meggondolást csak a balról-folytonosság igényel), ezért eloszlásfüggvény. Mivel a grafikon vízszintes "lépcskbl" áll, ezért diszkrét eloszláshoz tartozik.

(D) Nem eloszlásfüggvény, mert nem monoton növ (kb. $x = 0.3$ és $x = 0.45$ között szigorúan monoton csökken!).

(E) Habár ez hasonló a (C) grafikonhoz, nem lesz eloszlásfüggvény, mert a végtelenben 2-höz tart, nem 1-hez!

(F) Ez eloszlásfüggvény, mivel minden feltétel teljesül rá (a monotonitás is). A függvény nem szakad meg és csak véges sok töréspontja van, ezért véges sok pont kivételével differenciálható. Ez azt jelenti, hogy van srségfüggvénye és így folytonos eloszlást ír le. (A srségfüggvény létezésére vonatkozó precízebb matematikai feltételeket itt most nem tárgyaljuk, elég a lényeget érteni.)

c) Válasszuk ki az alábbi függvények közül a srségfüggvényeket!

(A) $f_1(x) = \begin{cases} 0 & x \leq 0 \\ \frac{3}{2}\,e^{-x} & 0 < x \end{cases}$

(B) $f_2(x) = \frac{1}{2}\tanh\left(\frac{x^3+1}{x^2+3}\right) + \frac{1}{2}$

(C) $f_3(x) = \begin{cases} 0 & x \leq 0 \\ \frac{24}{5}x & 0 < x \leq \frac{1}{2} \\ \frac{16}{5} - \frac{16}{5}x & \frac{1}{2} < x \leq 1 \\ 0 & 1 < x \end{cases}$

Ahhoz, hogy egy függvény srségfüggvény legyen, a következ feltételeket kell teljesítenie:

(0) Értelmezési tartomány: $\mathbb{R}$ **(véges sok pont kivételével)**
(i) Pozitivitás: $0 \leq f(x)$
(ii) Integrál: $\int_{-\infty}^{\infty} f(x)\,dx = 1$ **(azaz a függvény alatti terület 1)**

Megjegyzés: a (ii) tulajdonság következménye, hogy $\lim_{x\to-\infty} f(x) = \lim_{x\to\infty} f(x) = 0$.

```
> restart;
```

(A)

```
> f1 := x -> piecewise(x <= 0, 0, 3/2*exp(-x));
> plot(f1(x), x = -1..5);
```

$\frac{3}{2}e^{-x} \geq 0$, ezért $f_1(x) \geq 0$ is teljesül (ez a grafikonon is látszik). Az integrál viszont nem 1, ezért $f_1(x)$ nem srségfüggvény:

```
> int(f1(x), x = -infinity..infinity);
```

(B)

Megjegyzés: $\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ (ejtsd: *tangens-hiperbólikusz*)

```
> f2 := x -> 1/2*tanh((x^3+1)/(x^2+3)) + 1/2;
> plot(f2(x), x = -5..10);
```

Ez nem lesz sűrűségfüggvény, mert ∞-ben nem 0-hoz (hanem 1-hez) tart, így az integrálja (a függvény alatti terület) ∞.

(C)

```
> f3 := x -> piecewise(x <= 0, 0, x <= 1/2, 24/5*x, x <=
  1, 16/5-16/5*x, 0);
> plot(f3(x), x = -0.5..1.5, discont = true);
> int(f3(x), x = -infinity..infinity);
```

Tehát teljesül az összes feltétel, $f_3(x)$ srségfüggvény (az nem baj, hogy nem folytonos, 0.5-ben szakadása van).

7. Feladat (Valószínség eloszlás-/srségfüggvénybl)

Számítsuk ki a $P(X>1)$ és a $P(-2 \le x < 2)$ valószínségeket, ha tudjuk, hogy

a) X eloszlásfüggvénye $F(x) = \begin{cases} 0 & x \le -1 \\ \frac{1}{3} & -1 < x \le 1 \\ \frac{2}{3} & 1 < x \le 3 \\ 1 & 3 < x \end{cases}$

b) X srségfüggvénye $f(x) = \begin{cases} 0 & x \le -1 \\ \frac{1}{8} & -1 < x \le 1 \\ \frac{3}{8} & 1 < x \le 3 \\ 0 & 3 < x \end{cases}$

c) X eloszlásfüggvénye $F(X) = \begin{cases} 0 & x \le 0 \\ \sin x & 0 < x \le \frac{\pi}{2} \\ 1 & \frac{\pi}{2} < x \end{cases}$

d) X srségfüggvénye $f(x) = \begin{cases} 0 & x \le 0 \\ \frac{1}{4} x\, e^{-\frac{x}{2}} & x > 0 \end{cases}$

Megoldás

```
> restart;
```

Egy valószínségi változó adott intervallumba esésének valószínsgét sűrűségfüggvényből az intervallumon történő integrálással, eloszlásfüggvényből pedig –a Newton-Leibniz tételhez hasonlóan– a végpontokban felvett függvényértékek különbségének képzésével tudjuk számolni. Diszkrét eloszlásoknál figyelembe kell venni az intervallum zártságát/nyíltságát, folytonosaknál nem. Például: ($F(x_0^+)$ a jobboldali határértéket jelöli az x_0 pontban)

$$P(a \le X < b) = F(b) - F(a) \quad \left(= \int_a^b f(x)\,dx \text{ ha az eloszlás folytonos}\right)$$

$$P(a < X \le b) = F(b^+) - F(a^+) \quad ($$

$$= \int_a^b f(x)\,\mathrm{d}x \text{ ha az eloszlás folytonos)}$$

$$P(x < b) = F(b) \quad \left(= \int_{-\infty}^{b} f(x)\,\mathrm{d}x\text{, ha az eloszlás folytonos}\right)$$

$$P(a \le x) = 1 - F(a) \quad \left(= \int_{a}^{\infty} f(x)\,\mathrm{d}x\text{, ha az eloszlás folytonos}\right)$$

$$P(x \le b) = F(b^{+}) \quad \left(= \int_{-\infty}^{b} f(x)\,\mathrm{d}x\text{, ha az eloszlás folytonos}\right)$$

stb...

a) X **eloszlásfüggvénye** $F(x) = \begin{cases} 0 & x \le -1 \\ \frac{1}{3} & -1 < x \le 1 \\ \frac{2}{3} & 1 < x \le 3 \\ 1 & 3 < x \end{cases}$

A megadott eloszlásfüggvénybl számoljuk a valószínségeket. $F(x)$ diszkrét eloszlást határoz meg (mert lépcss függvény), ezért ügyelni kell az intervallum nyíltságára/zártságára!

```
> F := x -> piecewise(x <= -1, 0, x <= 1, 1/3, x <= 3,
  2/3, 1);
  'F(x)' = F(x);
```

```
> P0 := diszkret_eloszlasfuggveny([-1, 1, 3], [1/3, 1/3,
  1/3]): P0; # futtassuk újra a feladatsor elején az
  eljárást definiáló parancsot
```

$P(X > 1) = 1 - F(1^{+}) = 1 - \frac{2}{3} = \frac{1}{3}$ ($x = 1$-ben a jobboldali határérték a következ lépcs magassága (üres karika)). Lássuk ezt a számítást Maple-lel:

```
> `P(X > 1)` := 1 - limit(F(x), x = 1, right);
```

Az alábbi ábrán a függleges kék vonal hossza a kiszámolt valószínség.

```
> P1 := plot({[[0, 2/3],[1, 2/3]], [[0, 1],[3, 1]]},
  linestyle = dot, color = blue):
  P2 := plot([[0.2, 2/3],[0.2, 1]], color = blue,
  thickness = 3):
  plots[display](P0, P1, P2);
```

$P(-2 \le X < 2) = F(2) - F(-2) = \frac{2}{3} - 0 = \frac{2}{3}$

```
> `P(-2 <= X < 2)` := F(2) - F(-2);
```

```
> P3 := plot({[[-1, 0.01],[0.2, 0.01]], [[0, 2/3],[1, 2/3]
  ]}, linestyle = dot, color = blue):
  P4 := plot([[0.2, 0],[0.2, 2/3]], color = blue,
  thickness = 3):
  plots[display](P0, P3, P4);
```

b) X **srségfüggvénye** $f(x) = \begin{cases} 0 & x \le -1 \\ \frac{1}{8} & -1 < x \le 1 \\ \frac{3}{8} & 1 < x \le 3 \\ 0 & 3 < x \end{cases}$

A srségfüggvénybl integrálással számítjuk a valószínségeket.

```
> f := x -> piecewise(x <= -1, 0, x <= 1, 1/8, x <= 3,
  3/8, 0); 'f(x)' = f(x);
> P0 := plot(f(x), x = -2..4, discont = true, thickness =
  3): P0;
```

$P(X > 1) = \int_1^{\infty} f(x)\, dx$

```
> `P(X > 1)` := int(f(x), x = 1..infinity);
```

A következ ábra mutatja a srségfüggvény alatti területet.

```
> P1 := plot(f(x), x = 1..4, discont = true, thickness =
  3, filled = [color = blue, transparency = 0.5]):
  plots[display](P0, P1);
```

$P(-2 \le X < 2) = \int_{-2}^{2} f(x)\, dx$

```
> `P(-2 <= X < 2)` := int(f(x), x = -2..2);
```

A következ ábra mutatja a srségfüggvény alatti területet.

```
> P2 := plot(f(x), x = -2..2, discont = true, thickness =
  3, filled = [color = blue, transparency = 0.5]):
  plots[display](P0, P2);
```

Tehát $P(X > 1) = \frac{3}{4}$ és $P(-2 \le X < 2) = \frac{5}{8}$.

A valószínségeket a *Statistics* csomag segítségével egyszerbben is kiszámíthatjuk. Elsör definiáljunk egy olyan valószínségi változót, melynek $f(x)$ a srségfüggvénye:

```
> X := RandomVariable(Distribution(PDF = f));
> Probability(X > 1);
  Probability({X >= -2, X < 2});
```

c) X **eloszlásfüggvénye** $F(X) = \begin{cases} 0 & x \le 0 \\ \sin x & 0 < x \le \frac{\pi}{2} \\ 1 & \frac{\pi}{2} < x \end{cases}$

Az eloszlásfüggvény:

```
> F := x -> piecewise(x <= 0, 0, x <= Pi/2, sin(x), 1); 'F
  (x)' = F(x);
> P0 := plot(F(x), x = -1..(Pi/2 + 1), thickness = 3,
  discont = true): P0;
```

Ez a függvény folytonos eloszláshoz tartozik, mert folytonos és 1 pont kivételével differenciálható. A valószínséget az eloszlásfüggvénybl számoljuk (a jobb oldali határértékkel a folytonosság miatt most nem kell foglalkozni):

$P(X > 1) = 1 - F(1)$

```
> `P(X > 1)` := 1 - F(1); evalf(%);
```

```
> P1 := plot({[[0, F(1)],[1, F(1)]], [[0, 1],[Pi/2, 1]]},
  linestyle = dot, color = blue):
  P2 := plot([[0.1, F(1)],[0.1, 1]], color = blue,
  thickness = 3):
  plots[display](P0, P1, P2);
```

$P(-2 \le X < 2) = \mathrm{F}(2) - F(-2) = 1 - 0 = 1$

```
> `P(-2 <= X < 2)` := F(2) - F(-2); evalf(%);
```

```
> P3 := plot({[[0, 1],[Pi/2, 1]]}, linestyle = dot, color
  = blue):
  P4 := plot([[0.1, 0],[0.1, 1]], color = blue, thickness
  = 3):
  plots[display](P0, P3, P4);
```

A valószínségeket a *Statistics* csomag segítségével egyszerbben is kiszámíthatjuk. Elször definiáljunk egy olyan valószínségi változót, melynek $F(x)$ az eloszlásfüggvénye:

```
> X := RandomVariable(Distribution(CDF = F));
```

```
> Probability(X > 1);
  Probability({X >= -2, X < 2});
```

Megjegyzés: ez a módszer diszkrét eloszlásokra (pl. az a) feladatrészre) nem mindig mködik, mert nem veszi figyelembe az intervallum határait.

d) $\boldsymbol{X}$ **srségfüggvénye** $\boldsymbol{f(x) = \begin{cases} 0 & x \le 0 \\ \frac{1}{4} x \, \mathrm{e}^{-\frac{x}{2}} & x > 0 \end{cases}}$

A srségfüggvénybl integrálással számítjuk a valószínségeket.

```
> f := x -> piecewise(x <= 0, 0, 1/4*x*exp(-x/2)); 'f(x)'
  = f(x);
```

```
> P0 := plot(f(x), x = -3..15, thickness = 3): P0;
```

$$P(X > 1) = \int_1^{\infty} f(x) \, \mathrm{d}x$$

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

Az alábbi ábra mutatja a srségfüggvény alatti területet:

```
> P1 := plot(f(x), x = 1..15, thickness = 3, filled =
  [color = blue, transparency = 0.5]):
  plots[display](P0, P1);
```

$$P(-2 \le \mathrm{X} < 2) = \int_{-2}^{2} f(x) \, \mathrm{d}x$$

```
> `P(-2 <= X < 2)` := int(f(x), x = -2..2); evalf(%);
```

Az alábbi ábra mutatja a srségfüggvény alatti területet:

```
> P2 := plot(f(x), x = -2..2, thickness = 3, filled =
  [color = blue, transparency = 0.5]):
  plots[display](P0, P2);
```

Tehát $P(X > 1) = 0.910$ és $P(-2 \le X < 2) = 0.264$.

A valószínségeket a *Statistics* csomag segítségével egyszerbben is kiszámíthatjuk. Elször definiáljunk egy olyan valószínségi változót, melynek $f(x)$ az srségfüggvénye:

```
> X := RandomVariable(Distribution(PDF = f));
```

```
> Probability(X > 1);
  Probability({X >= -2, X < 2});
```

8. Feladat (Konstans meghatározása eloszlás-/srségfüggvényben / 1.)

a) Az alábbiakban egy X valószínségi változó eloszlásfüggvénye van megadva. Határozzuk meg a 'c' konstans értékét és ábrázoljuk az eloszlásfüggvényt! Diszkrét esetben adjuk meg és ábrázoljuk X valószínségi eloszlását, folytonos esetben pedig határozzuk meg és ábrázoljuk srségfüggvényét!

(A) $$F(x) = \begin{cases} 0 & x \le 5 \\ \frac{1}{3} & 5 < x \le 7 \\ c & 7 < x \end{cases}$$

(B) $$F(x) = \begin{cases} 0 & x \le 0 \\ c \arcsin\left(\sqrt{x}\right) & 0 < x < 1 \\ 1 & 1 \le x \end{cases}$$

b) Az alábbiakban egy X folytonos eloszlású valószínségi változó srségfüggvénye van megadva. Határozzuk meg a 'c' konstans értékét, valamint ábrázoljuk a srségfüggvényt és az eloszlásfüggvényt!

(C) $$f(x) = \frac{c}{\sqrt{2\pi}} e^{-\frac{x^2}{18}}$$

(D) $$f(x) = \begin{cases} 0 & x \le 1 \\ c - \frac{1}{4}x & 1 < x \le 2 \\ 0 & 3 < x \end{cases}$$

Megoldás

```
> restart;
```

a) Az alábbiakban egy X valószínségi változó eloszlásfüggvénye van megadva. Határozzuk meg a 'c' konstans értékét és ábrázoljuk az eloszlásfüggvényt! Diszkrét esetben adjuk meg és ábrázoljuk X valószínségi eloszlását, folytonos esetben pedig határozzuk meg és ábrázoljuk srségfüggvényét!

Úgy kell megválasztani c értékét, hogy $F(x)$ teljesítse az eloszlásfüggvényre vonatkozó szükséges és elégséges feltételeket (lásd 5. Feladat, b pont).

(A)

$$F(x) = \begin{cases} 0 & x \le 5 \\ \frac{1}{3} & 5 < x \le 7 \\ c & 7 < x \end{cases}$$

Mivel $x > 7$ esetén $F(x) = c$, ezért $\lim_{x \to \infty} F(x) = c$. De eloszlásfüggvény ∞ -ben vett határértékének 1-nek kell lennie, vagyis $c = 1$. Ábrázoljuk $F(x)$-et!

```
> c := 1;
  F := x -> piecewise(x <= 5, 0, x <= 7, 1/3, c);
> plot(F(x), x = 3..9, discont = true, symbolsize = 18,
  symbol = solidcircle, color = red);
```

Láthatjuk, hogy ez tényleg eloszlásfüggvény: monoton növ, balról folytonos és $-\infty$-ben vett határértéke 0. Ráadásul mivel lépcss függvény, ezért diszkrét eloszlást határoz meg. Az eloszlás lehetséges értékeit az eloszlásfüggvény ugrási helyei adják, a hozzájuk tartozó valószínségek pedig a lépcsfokok magasságai lesznek: $x_1 = 5$, $p_1 = \frac{1}{3}$ (els lépcsfok magassága), $x_2 = 7$, $p_2 = 1 - \frac{1}{3} = \frac{2}{3}$ (a 2. lépcsfoknál $\frac{1}{3}$-ról 1-re ugrik a függvény).

Táblázat beszúrásához válasszuk az *Insert* menübl a *Table* lehetséget:

x_i	5	7
p_i	$\frac{1}{3}$	$\frac{2}{3}$

Ellenrizhetjük, hogy a valószínségek összege tényleg 1:

```
> x := [5, 7];
  px := [1/3, 2/3];
> 'px[1] + px[2]' = px[1] + px[2];
```

Az eloszlás pálcikadiagramon ábrázolva:

```
> with(DynamicSystems):
  DiscretePlot(x, px, style = stem, color = blue,
  thickness = 3);
```

A munkalap elején szerepl *diszkret_eloszlasfuggveny* eljárással picit szebb ábrát is készíthetünk az eloszlásfüggvényrl:

```
> diszkret_eloszlasfuggveny(x, px);
```

(B)

$$F(x) = \begin{cases} 0 & x \leq 0 \\ c \arcsin(\sqrt{x}) & 0 < x < 1 \\ 1 & 1 \leq x \end{cases}$$

```
> restart;
> F := x -> piecewise(x <= 0, 0, x < 1, c*arcsin(sqrt(x)),
  1);
```

Próbáljunk ki néhány helyettesítési értéket c-re:

```
> plot(eval(F(x), c = 1), x = -1..2, discont = true,
  thickness = 3);
> plot(eval(F(x), c = 1/2), x = -1..2, discont = true,
  thickness = 3);
> plot(eval(F(x), c = -1/2), x = -1..2, discont = true,
  thickness = 3);
```

Mivel $\arcsin(\sqrt{x})$ pozitív a $]0, 1[$ intervallumon, ezért csak $c > 0$ lehet jó (máskülönben negatív értékeket venne fel az eloszlásfüggvény)!

Másrészt $c \cdot \arcsin(\sqrt{x}) \leq 1$-nek kell teljesülnie a $]0, 1[$ intervallumon ahhoz, hogy

$F(x)$ monoton növ legyen. Mivel $c \cdot \arcsin(\sqrt{x})$ folytonos és monoton növ függvény a $]0, 1]$ intervallumon, ezért $\sup_{x \in]0, 1[} c \cdot \arcsin(\sqrt{x}) = c \cdot \arcsin(\sqrt{1}) = c \cdot \frac{\pi}{2}$ ($x = 1$-ben érné el a maximális értéket). Tehát a fenti $c \cdot \arcsin(\sqrt{x}) \leq 1$ feltétel akkor és csak akkor teljesül, ha

$$c \cdot \frac{\pi}{2} \leq 1, \text{ azaz } c \leq \frac{2}{\pi}\,!$$

Ebbl úgy tnhet, hogy c-re több jó megoldás is van (pl. $c = \frac{2}{\pi}$, $c = \frac{1}{2}$, $c = 0$, stb.), de még figyelembe kell venni azt, hogy $F(x)$-nek balról folytonosnak kell lennie $x = 1$-ben (ebben a feladatrészben ez egy lényegi feltétel, ugyanis $F(x)$ úgy volt megadva, hogy $x = 1$-ben jobbról –de balról nem feltétlenül– folytonos).

$$F(1) = \lim_{x \to 1-} F(x) = \lim_{x \to 1-} c \cdot \arcsin(\sqrt{x}) = c \cdot \arcsin(\sqrt{1}) = c \cdot \frac{\pi}{2}$$

amibl $F(1) = 1$ miatt $c = \frac{2}{\pi}$ következik.

```
> unassign('c');
  isolate(F(1) = c*arcsin(sqrt(1)), c); assign(%);
  evalf(c);
> plot(F(x), x = -1..2, discont = true, thickness = 3);
```

A grafikon mutatja, hogy az eloszlásfüggvény folytonos és két pont kivételével differenciálható, ezért folytonos eloszlást határoz meg. A srségfüggvény az eloszlásfüggvény deriváltja: $f(x) = F'(x)$.

```
> f := x -> diff(F(x), x); 'f(x)' = f(x);
> plot(f(x), x = -1..2, discont = true, thickness = 3);
```

b) Az alábbiakban egy X folytonos eloszlású valószínségi változó srségfüggvénye van megadva. Határozzuk meg a 'c' konstans értékét, valamint ábrázoljuk a srségfüggvényt és az eloszlásfüggvényt!

(C)

$$f(x) = \frac{c}{\sqrt{2\pi}} e^{-\frac{x^2}{18}}$$

Úgy kell megválasztani c értékét, hogy $f(x)$ teljesítse az eloszlásfüggvényre vonatkozó szükséges és elégséges feltételeket (lásd 5. Feladat, c pont).

Az nyilvánvaló, hogy $c \geq 0$-nak teljesülnie kell, mert $f(x)$ nemnegatív. A másik feltétel szerint $f(x)$ integrálja a valós számegyenesen 1-et kell, hogy adjon.

```
> restart;
> f := x -> c/sqrt(2*Pi)*exp(-x^2/18);
> int(f(x), x = -infinity..infinity);
```

Ebbl $c = \frac{1}{3}$ adódik. Oldjuk meg Maple-lel is az egyenletet:

```
> unassign('c'); isolate(int(f(x), x = -infinity..
  infinity) = 1, c); assign(%);
```

Ábrázoljuk $f(x)$-et:

```
> plot(f(x), x = -12..12, thickness = 3);
```

Megjegyzés: ez a függvény egy normál eloszlás srségfüggvénye, *haranggörbé*nek is nevezik. A késbbiekben gyakran fogunk vele találkozni!

Az eloszlásfüggvény a srségfüggvény integrálfüggvénye: $F(x) = \int_{-\infty}^{x} f(t)\, \mathrm{d}t$

```
> F := x -> int(f(t), t = -infinity..x);
  F(x);
> plot(F(x), x = -12..12, thickness = 3);
```

(D)

$$f(x) = \begin{cases} 0 & x \leq 1 \\ c - \frac{1}{4}x & 1 < x \leq 2 \\ 0 & 3 < x \end{cases}$$

$f(x)$ egyedül az $[1, 2]$ intervallumon nem nulla, és ott monoton csökken, így a nemnegativitáshoz szükséges és elégséges feltétel $f(2) \geq 0$.

```
> restart;
> f := x -> piecewise(x <= 1, 0, x <= 2, c - 1/4*x, 0);
> isolate(f(2) >= 0, c);
```

Ebbl a feltételbl $c \geq \frac{1}{2}$ adódik. Lássuk most az integrált:

```
> int(f(x), x = -infinity..infinity);
```

Oldjuk meg az $\int_{-\infty}^{\infty} f(x)\, \mathrm{d}x = 1$ egyenletet!

```
> unassign('c');
  isolate(int(f(x), x = -infinity..infinity) = 1, c);
  assign(%);
```

$\frac{11}{8} \geq \frac{1}{2}$, így az els (pozitivitási) feltétel is teljesül, $c = \frac{11}{8}$ a helyes megoldás!

Ábrázoljuk grafikonon a srségfüggvényt!

```
> plot(f(x), x = 0..3, thickness = 3, discont = true);
```

Láthatjuk, hogy $f(x)$ nem folytonos és 1-nél nagyobb értéket is felvesz, de ez megengedett a srségfüggvényeknél.

Az eloszlásfüggvény ennek integrálfüggvénye:

```
> F := x -> int(f(t), t = -infinity..x); F(x);
> plot(F(x), x = 0..3, thickness = 3, discont = true);
```

Megjegyzés: a (C) és (D) pontban kapott eloszlások mind folytonosak, mert van srségfüggvényük (az eleve meg volt adva).

Gyakorló feladatok

Gy/1. Feladat (Két dobókocka minimuma)

Megoldás

Gy/2. Feladat (Pont a körön)

Megoldás

Gy/3. Feladat (Normál eloszlás)

Megoldás

Gy/4. Feladat (Környezetszennyezés)

Megoldás

Gy/5. Feladat (Eloszlás- és srségfüggvények)

Megoldás

Gy/6. Feladat (Konstans meghatározása eloszlás-/srségfüggvényben / 2.)

Megoldás