11. Gyakorlat

Nevezetes folytonos eloszlások

Elméleti összefoglaló

Exponenciális eloszlás

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

Tekintsünk egy olyan konstrukciót, ahol egy (végtelen hosszúságú) *egyenes mentén pontok szóródnak valamilyen átlagos egyenletes sûrûséggel.* A gyakorlatban sok, ezzel modellezhetô helyzet van; akár térben, akár idôben (idôtengelyen) szóródhatnak a megfigyelt, pontszerûnek tekinthetô események (ill. azok valószínûségi változóval számszerûsített értékei).
Ilyen modell szerint rendezôdnek el pl. egy adott pillanatban egy útszakaszon az autók - ha légifelvételt készítünk az útszakaszról, akkor térben, de ha az amúgy mozgó jármûveket egy megfigyelési ponton számláljuk, akkor idôben is. Ezen analógia szerint érthetôen ugyanilyen modell szerint jönnek velünk szemben a járókelôk az utcán, megfigyelt meteorok a légkörbe, vagy mûszerrel érzékelt részecskék egy radioaktív forrásból.
A modell tipikusan arra az esetre használható, ha egy (térbeli vagy idôbeli) szakaszon hosszabb távon *az átlagos pontsûrûség ismert* (elméleti úton meghatározható, vagy kísérletileg megmérhetô). A pontok száma persze (adott hosszúságú szakaszokon újra és újra megszámolva) ezen átlag körül ingadozásokat mutat (a lehetséges darabszámokhoz tartozó megfelelô valószínûségekkel). Ezen eloszlás viszont nemhogy maga is pontosan meghatározott a konstrukció által, de ráadásul (érdekes és egyáltalán nem triviális módon) olyan szoros és egyértelmû összefüggésben van a saját átlagával, hogy a teljes szerkezete maradéktalanul jellemezhetô vele, mint egyetlen paraméterrel (λ) – a megfigyelt szakaszon mért átlagos darabszámon kívül semmilyen más információra nincs szükség az elôfordulási darabszámok valószínûségeinek leírásához.
A konstrukcióban a tipikusan feltehetô kérdések:
1. Egy adott szakaszon hány pontra milyen valószínûséggel lehet számítani? (Adott hosszúságú szakaszon mérhetô darabszám valószínûségi eloszlása.)
2. Az egyes pontok között mekkora távolságra milyen valószínûséggel lehet számítani? (Pontok közti távolság valószínûségi eloszlása.)
Az elsô kérdésre a *Poisson-eloszlás*, a másodikra a most sorra kerülô *exponenciális eloszlás* ad választ. A kettô jelentôsen eltér már abban is, hogy elôbbi diszkrét, utóbbi folytonos, mégis összefüggnek a közös modell révén.
Vizsgáljuk meg ezúttal a 2. kérdést! (A elôzô heti forgalomszámlálós feladatot tekintve ez az a kérdés lenne, hogy mennyi idô telik el a –közvetlenül egymás után haladó– jármûvek érkezése között.)
A szakaszra esô darabszámot leíró valószínûségi változó értékkészlete egész számokból áll (diszkrét), viszont az idô folytonos, ezért a szomszédos pontok közti távolságot leíró valószínûségi változó is folytonos.
A Poisson-eloszlás felhasználásával könnyen fel lehet írni a szomszédos pontok távolságának eloszlásfüggvényét. Ennek definíció szerinti felírásához azt kell kiszámolnunk, hogy egy adott ponttól mérve milyen valószínûséggel van a következô (ténylegesen T távolságra levô) pont egy adott, a függvény változójaként szereplô x távolságon belül. Egyszerûbb viszont az ellentét-eseményt vizsgálni, vagyis hogy milyen valószínûséggel *nincs* újabb pont x-en belül. Erre pedig a választ a Poisson-eloszlás szolgáltatja, az x hosszúságú szakaszra számolt darabszám várható értéke, mint paraméter, és $k = 0$ elôfordulási darabszám helyettesítésével.

Ha feltételezzük, hogy az elôfordulási darabszám várható értéke egységnyi hosszon éppen λ, akkor x hosszon éppen $\lambda \cdot x$. Ezek szerint a λ paraméterû exponenciális eloszlású T valószínûségi változó eloszlásfüggvénye:

$$F_{\mathrm{T}}(x) = P(T < x) = 1 - F_{\mathrm{Poisson}(\lambda \cdot x)}(0) = 1 - \mathrm{e}^{-\lambda \cdot x}$$

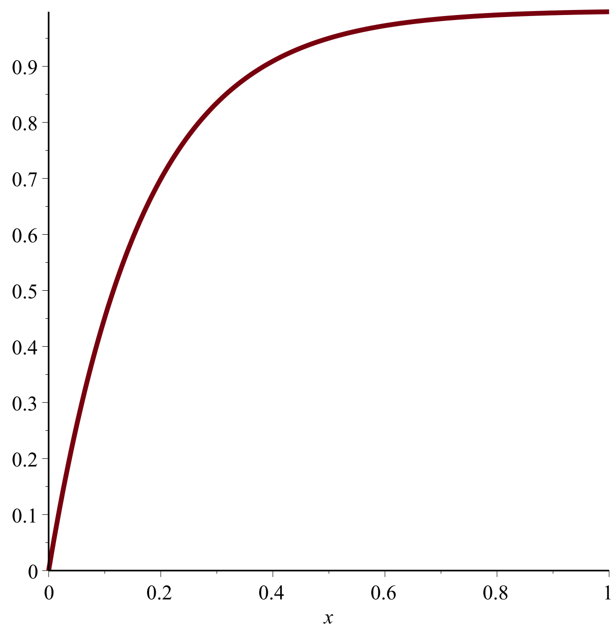
```
> F := (lambda, x) -> 1 - exp(-lambda*x);
```

$$F := (\lambda, x) \rightarrow 1 - \mathrm{e}^{-\lambda x} \qquad \textbf{(1.1.1.1)}$$

Az exponenciális eloszlás éppen errôl az eloszlásfüggvényében (és sûrûségfüggvényében) megjelenô exponenciális függvényrôl kapta a nevét. A forgalomszámlálós példánkban szereplô adattal:

```
> lambdaperc := 6;
  plot(F(lambdaperc, x), x = 0..1, thickness = 3);
```

$$lambdaperc := 6$$



A grafikon vízszintes tengelye percek szerint van skálázva. Látszik, hogy 1 percnél többet szinte soha nem kell várni a következô autóra.
Egész pontosan az 1 percnél hosszabb várakozás valószínûsége az 1 percnél rövidebb várakozás komplementereként számolva:

$$P(T > 1) = 1 - P(T < 1) = 1 - F_T(1)$$

```
> `P(T > 1)` := 1 - F(lambdaperc,1); evalf(%);
```

$$P(T > 1) := \mathrm{e}^{-6}$$

$$0.002478752177 \quad (1.1.1.2)$$

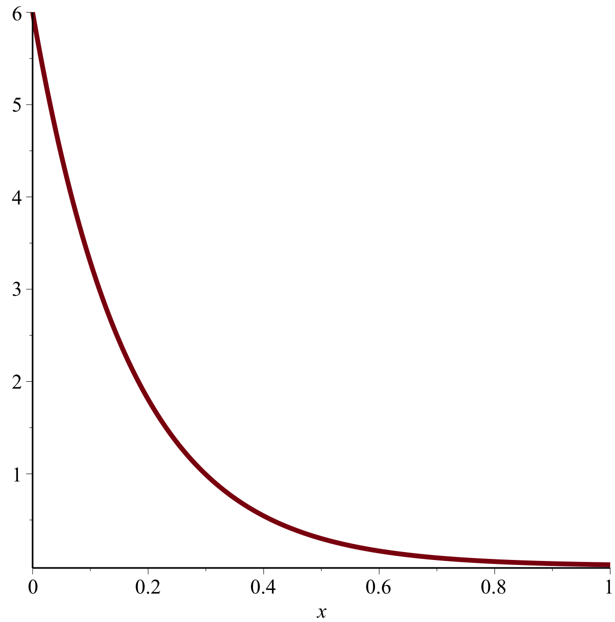
Vagyis csak kb. 2.5 ezrelék valószínûséggel lesz 1 percnél hosszabb hézag két áthaladó autó között.
Az exponenciális eloszlás sûrûségfüggvénye:

$$f_T(x) = \frac{d}{dx} F_T(x) = \lambda \cdot e^{-\lambda \cdot x}$$

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

$$f := (\lambda, x) \to \frac{\partial}{\partial x} F(\lambda, x) \quad (1.1.1.3)$$

```
> plot(f(lambdaperc, x), x = 0..1, thickness = 3);
```



Ha ez valóban egy valószínûségi sûrûségfüggvény, akkor egységnyi terület van "szétkenve" a görbe alatt. Errôl integrálással (improprius!) gyôzôdhetünk meg:

```
> assume(lambda > 0);
  int(f(lambda, x), x = 0..infinity);
```

$$1 \quad (1.1.1.4)$$

Vegyük észre, hogy a sûrûségfüggvény 0-ban jobbról vett határértéke éppen λ, vagyis az idôegység alatt várható események darabszáma. Jelen esetben ez 6, megfelelôen annak, hogy az idôegységként választott *perc* alatt átlagosan 6 jármû érkezik. Más idôegység választása esetén nyilván más λ-val kell számolnunk.
Az exponenciális eloszlás várható értéke és szórása egyaránt a λ paraméter reciproka:

$E(T) = \sigma(T) = \frac{1}{\lambda}$

A pontok közti távolság várható értékére kapott $\frac{1}{\lambda}$ érték összhangban van az egységszakaszon mért darabszám λ várható értékével: ha pl. ha az egységnyi hosszúságú szakaszra átlagosan 3 pont esik, akkor nem meglepô, hogy a köztük lévô távolság átlagos értéke $\frac{1}{3}$.

Normál eloszlás

A sok tényezô összegzett hatása által kialakított valószínûségi változók eloszlása nagyon sok esetben harangszerû alakot vesz fel, mely mögött a valószínûségszámítás és statisztika méltán legnevezetesebb eloszlása, az ún. *normál eloszlás* áll. A természetben a statisztikai vizsgálatokkal mérhetô paraméterek legtöbbje ilyen eloszlást közelít: legyen szó pl. a július hónap sokévi középhômérsékletérôl, az esôcseppek méretérôl, a boltban vásárolható kilós kenyerek pontos tömegérôl vagy pl. a felnôtt férfiak testmagasságáról.
A normál eloszlás sûrûségfüggvényének alakja (harang-görbe) jól meghatározott. Különbözô esetekben figyelve ugyan a csúcsa (várható értéke) kerülhet máshova, és a lapultsága (szórása) is változhat, de lineáris transzformációkkal (nyújtásokkal, zsugorításokkal, eltolásokkal) bármelyik bármely másikba átvihetô. A normál eloszlások bármelyike jellemezhetô mindössze két paraméterrel: a várható értékkel és a szórással.
Szokás kijelölni a sok közül egy szabványosat: az ún. *standard normál eloszlás*t, mely 0 várható értékkel és 1 szórással rendelkezik:

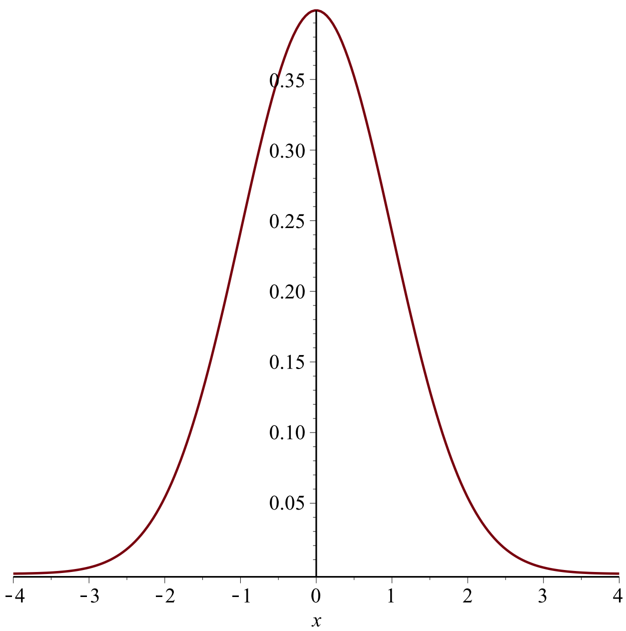
```
> N := RandomVariable(Normal(0, 1));
```

$$N := \_R \qquad \mathbf{(1.1.2.1)}$$

Ennek sûrûségfüggvénye:

```
> f := x -> PDF(N, x): 'f'(x) = f(x);
  plot(f(x), x = -4..4);
```

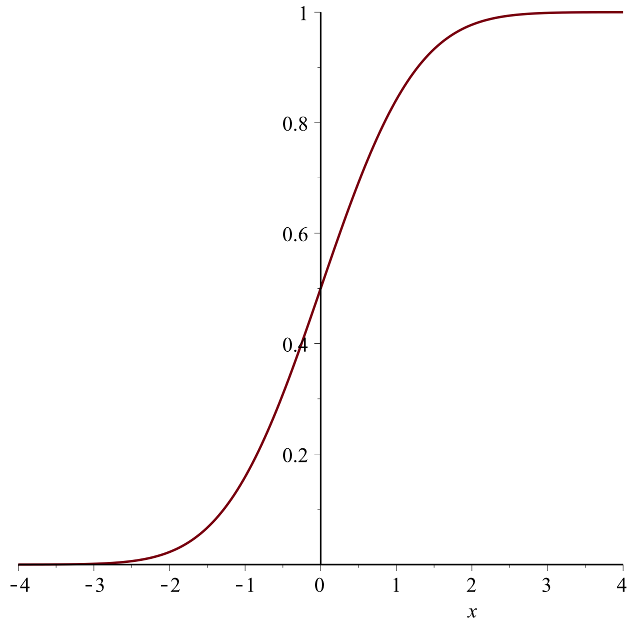
$$f(x) = \frac{1}{2}\,\frac{\sqrt{2}\,e^{-\frac{1}{2}x^2}}{\sqrt{\pi}}$$



A görbe alatti teljes terület bizonyíthatóan 1.
Az eloszlásfüggvény:

```
> F := x -> CDF(N, x): 'F'(x) = F(x);
  plot(F(x), x = -4 .. 4);
```

$$F(x) = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{1}{2}\,x\sqrt{2}\right)$$



Az eloszlásfüggvény a sûrûségfüggvény integrálja. Különös viszont, hogy neki, mint a matematika egyik legnevezetesebb függvényének nincs zárt alakja. Márpedig a normál eloszlásokkal kapcsolatos számításokhoz folyamatosan ezt a függvényt kell használni, de az értékek csak közelítéssel számíthatók (amúgy tetszôleges pontossággal, tehát a gyakorlatban nem jelent különösebb problémát, de elméleti szempontból fontos, megjegyzendô tény, hogy egzaktul számolható zárt formula nincs).

Bármilyen más normál eloszlás létrehozható a két paraméter, μ és σ megadásával:

```
> N1 := RandomVariable(Normal(2, 3));
  f1 := x -> PDF(N1, x);
```

$$N1 := \_R0$$

$$f1 := x \rightarrow Statistics\text{:-}PDF(N1, x) \quad \textbf{(1.1.2.2)}$$

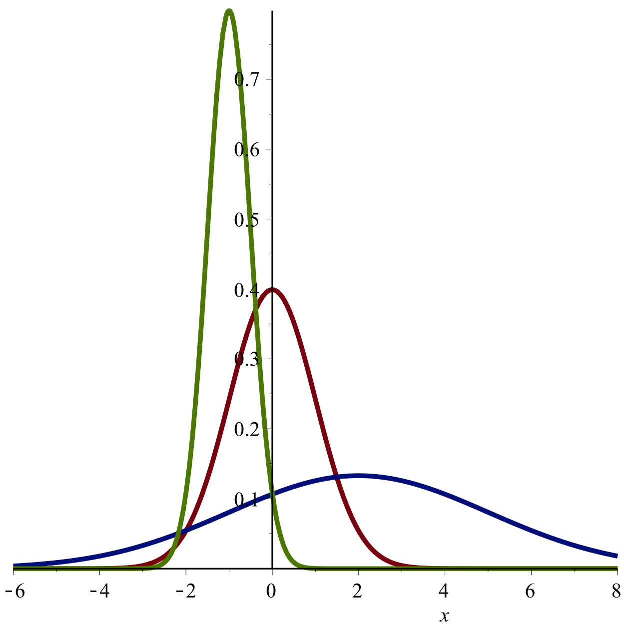
```
> N2 := RandomVariable(Normal(-1, .5));
  f2 := x -> PDF(N2, x);
```

$$N2 := \_R1$$

$$f2 := x \rightarrow Statistics\text{:-}PDF(N2, x) \quad \textbf{(1.1.2.3)}$$

A különbözô várható értékû és szórású normál eloszlásokat ábrázolhatjuk közös koordináta-rendszerben:

```
> plot([f(x), f1(x), f2(x)], x = -6..8, thickness = 3);
```



Mindegyik görbe eloszlásfüggvény, tehát mindegyik alatt a teljes 1 valószínûség oszlik el. Jól érthetô módon kisebb szóráshoz keskenyebb görbe és magasabb csúcsérték tartozik.

A sûrûségfüggvény pontos formulája: $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, ahol μ várható értéket, σ a szórást jelenti.
Ha egy normál eloszlással közelíthetô eloszlásnak tudjuk a várható értékét és szórását (vagy akár csak tapasztalati úton egy statisztikai mintának), akkor e két paraméter behelyettesítésével azonnal illeszthetünk is rá egy megfelelô normál eloszlást.

Kidolgozott feladatok

1. Feladat (Út forgalma / 2.)*

Egy útszakasz egy pontján megfigyelések szerint óránként átlagosan 360 jármû halad át.
a) Adjuk meg a két jármû érkezése között eltelt *T* idô eloszlását, várható értékét, varianciáját és szórását! Ábrázoljuk *T* eloszlás és sûrûségfüggvényét!
b) Mennyi a valószínûsége, hogy két jármû érkezése között 10 másodpercnél kevesebb idô telik el?
c) Mennyi a valószínûsége, hogy két jármû érkezése között 5 másodpercnél kevesebb idô telik el?

d) Mennyi a valószínûsége, hogy két jármû érkezése között 15 másodpercnél több idô telik el?
e) Mennyi a valószínûsége, hogy egy jármû után a következô az 5. és 10. másodperc között érkezik?

Megoldás

```
> restart;
```

a) Adjuk meg a két jármû érkezése között eltelt *T* idô eloszlását, várható értékét, varianciáját és szórását! Ábrázoljuk *T* eloszlás és sûrûségfüggvényét!

Miután példánk szerint percenként átlagosan $\frac{360}{60} = 6$ jármû érkezik, az átlagos várakozási idô 10 másodperc. Ha az idô mértékegységét *másodperc*nek választjuk, akkor eszerint a várakozási idôket leíró, $E(T) = \frac{1}{\lambda} = 10$ várható értékû exponenciális eloszlás paramétere $\lambda = \frac{1}{10}$ (másodpercenként átlagosan $\frac{1}{10}$ jármû érkezik):

```
> with(Statistics):
> lambdamasodperc := 1/10;
```

$$lambdamasodperc := \frac{1}{10} \quad \textbf{(1.2.1.1.1)}$$

```
> T := RandomVariable(Exponential(1/lambdamasodperc));
```

$$T := \_R \quad \textbf{(1.2.1.1.2)}$$

Vigyázzunk: a Maple-ben az exponenciális eloszlás paramétere –az általunk tanulttól eltérôen– λ helyett $\frac{1}{\lambda}$!

Az eloszlásból visszakapjuk a várható értéket:

```
> ET := ExpectedValue(T); # = 1/lambda
```

$$ET := 10 \quad \textbf{(1.2.1.1.3)}$$

A variancia és a szórás:

```
> VarT := Variance(T); # = 1/lambda^2
```

$$VarT := 100 \quad \textbf{(1.2.1.1.4)}$$

```
> sigmaT := StandardDeviation(T); # = 1/lambda
```

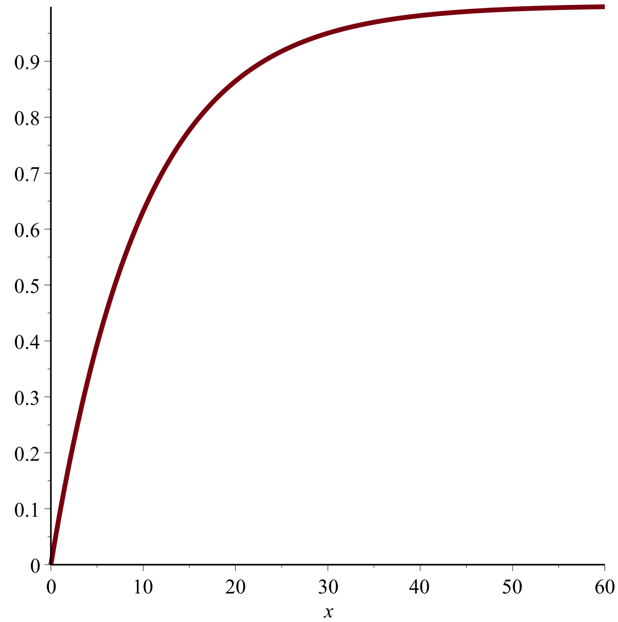
$$sigmaT := 10 \quad \textbf{(1.2.1.1.5)}$$

Rajzoljuk most meg *T* eloszlásfüggvényét!

```
> F := x -> CDF(T, x): 'F'(x) = F(x); # x -> piecewise(x
  <= 0, 0, 1 - e^(-lambdamasodperc*x))
```

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\frac{1}{10}x} & otherwise \end{cases} \quad \textbf{(1.2.1.1.6)}$$

```
> plot(F(x), x = 0..60, thickness = 3);
```

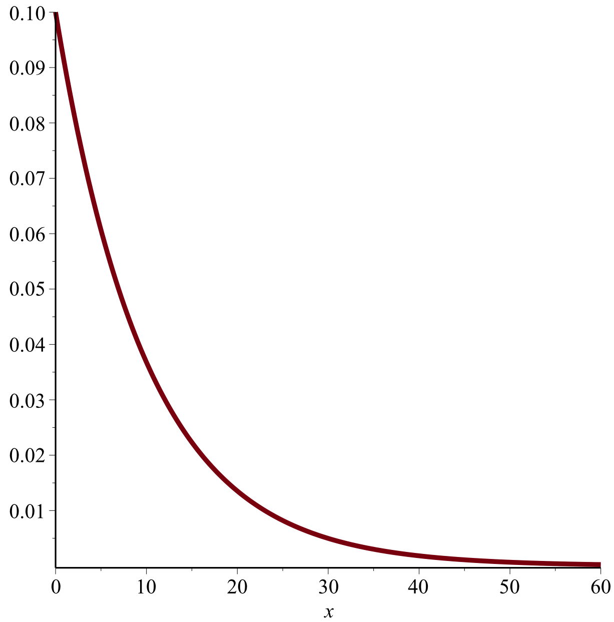


A sûrûségfüggvény ennek deriváltja, mely közvetlenül a Statistics csomag *PDF* eljárásával is számolható:

```
> f := x -> PDF(T, x): 'f'(x) = f(x); # x -> diff(F(x), x)
  = piecewise(x <= 0, 0, lambdamasodperc*e^(-
  lambdamasodperc*x))
```

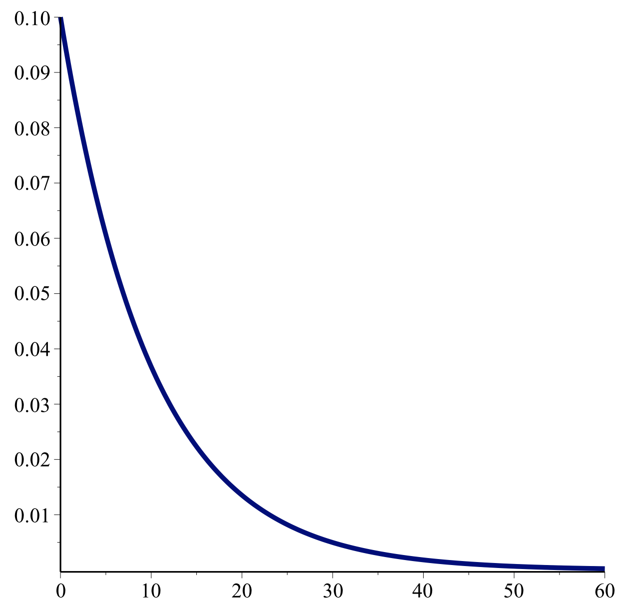
$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{10}\,\mathrm{e}^{-\frac{1}{10}x} & \mathit{otherwise} \end{cases} \qquad \textbf{(1.2.1.1.7)}$$

```
> surusegplot := plot(f(x), x = 0..60, thickness = 3):
  surusegplot;
```



A sûrûségfüggvény közvetlen ábrázolására is van külön eljárás a Statistics csomagban:

```
> DensityPlot(T, range = 0..60, thickness = 3);
```



Láthatjuk, hogy a sûrûségfüggvény 0-ban jobbról vett határértéke éppen λ, vagyis az idôegység alatt várható események darabszáma. Jelen esetben ez 0.1, megfelelôen annak, hogy az idôegységként választott *másodperc* alatt átlagosan 0.1 jármû érkezik (merthogy 10 másodpercenként érkezik átlagosan 1).

Megjegyzés: a grafikon adott idôtartamra vonatkoztatott alakja független az idô mértékegységétôl. (Eltekintve természetesen attól, hogy az *x* tengely skálázásának változása maga után vonja az *y* tengely skálázásának megváltozását, de egyszerû nyújtási transzformációkkal elérhetô, hogy a grafikonok egybevágóak legyenek.)

Mivel példánkban az autók érkezése közötti jellemzô idôtartam néhány másodperc nagyságrendû, az idô mértékéül válasszunk másodpercet (ahogy a fenti grafikonok megrajzolásánál is), és a kérdésekre eszerint adjunk választ.

b) Mennyi a valószínûsége, hogy két jármû érkezése között 10 másodpercnél kevesebb idô telik el?

Annak valószínûségét, hogy két egymás után érkezô jármû között adott idônél kevesebb telik el, az eloszlásfüggvény adja:

```
> `P(T < 10)` := F(10); evalf(%);
```

$$P(T < 10) := 1 - e^{-1}$$

$$0.6321205588 \qquad \mathbf{(1.2.1.1.8)}$$

Ezt meghatározhatjuk a *Probability* eljárással is:

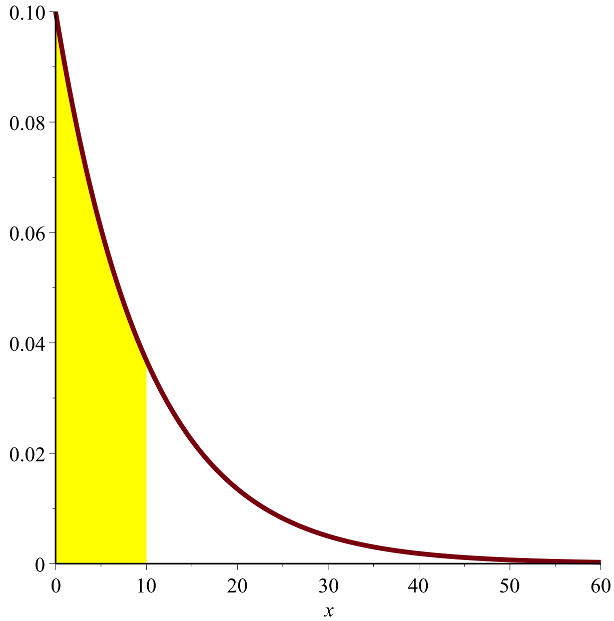
```
> Probability(T < 10); evalf(%);
```

$$1 - e^{-1}$$

$$0.6321205588 \quad \textbf{(1.2.1.1.9)}$$

Tehát kb. 63 % valószínûséggel telik el két jármû érkezése között (az amúgy átlagos) 10 másodpercnél kevesebb idô. Érdekes észrevenni, hogy bár az átlagos várakozási idô $\lambda = 10$ másodperc két jármû között, az esetek jóval több, mint felében a valódi várakozási idô kevesebb ennél. Az átlag úgy jöhet ki mégis 10 másodpercre, hogy bár összességében ritkábban fordulnak elô 10 másodperc feletti idôhézagok, mint alattiak, mivel az értékre elvileg nincs felsô korlát, olykor ezek elég naggyá válhatnak ahhoz, hogy kisebb elôfordulási valószínûség mellett is felhúzzák az átlagot.
Grafikusan szemléltetve a valószínûség a sûrûségfüggvény alatti (egységnyi) területnek az a része, ami a 10-es abszcisszától negatív irányba (balra) esik (elvileg mínusz végtelentôl integrálva, de mivel a negatív tartományon a függvény zérus, valójában $x = 0$-tól elég figyelni):

```
> szinezesplot := plot(f(x), x = 0..10, filled = [color =
  yellow, transparency = 0.5]):
  plots[display](surusegplot, szinezesplot);
```



Vegyük észre, hogy a színezett terület befoglaló téglalapja $10 \times 0.1 = 1$ területû - ehhez mérten reálisnak tûnik a rá számítással kapott ~63 % érték.
c) Mennyi a valószínûsége, hogy két jármû érkezése között 5 másodpercnél kevesebb idô telik el?

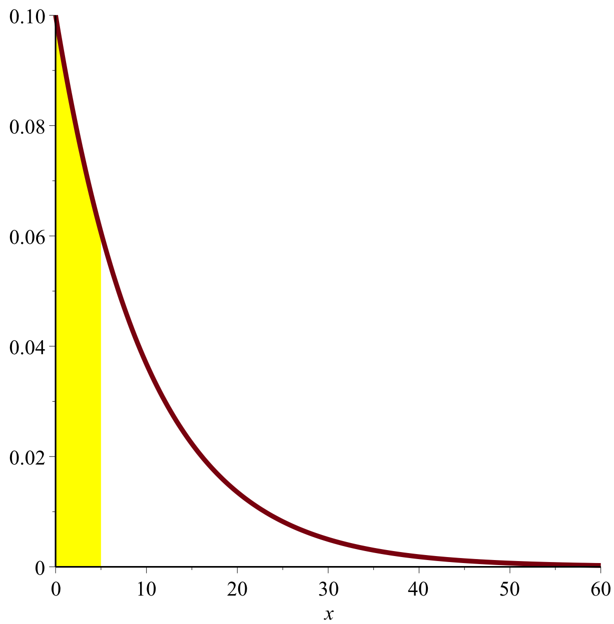
```
> `P(T < 5)` := F(5); evalf(%); # = Probability(T < 5)
```

$$P(T < 5) := 1 - e^{-\frac{1}{2}}$$

$$0.3934693403 \qquad \textbf{(1.2.1.1.10)}$$

Természetesen az 5 másodpercnél kisebb hézag valószínûsége kisebb, mint a 10 másodpercnél kisebb hézag valószínûsége, hiszen elôbbi esemény része utóbbinak. Ábrával szemléltetve:

```
> szinezesplot := plot(f(x), x = 0..5, filled = [color = yellow, transparency = 0.5]):
  plots[display](surusegplot, szinezesplot);
```



d) Mennyi a valószínûsége, hogy két jármû érkezése között 15 másodpercnél több idô telik el?

Az eloszlásfüggvény közvetlenül egy adott idônél *kevesebb* várakozási idô valószínûségét képes megmondani. Ha egy adott idônél *több* várakozási idô valószínûségére vagyunk kíváncsiak, azt komplementer eseményként kell számolni:

```
> `P(T < 15)` := F(15); evalf(%);
```

$$P(T < 15) := 1 - e^{-\frac{3}{2}}$$

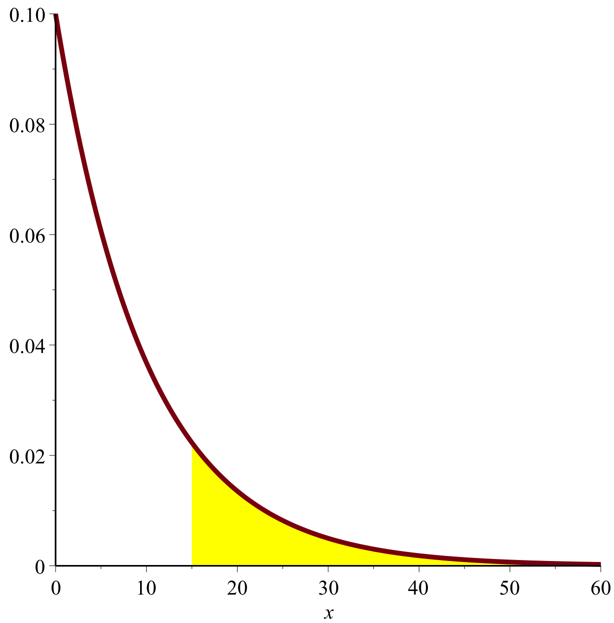
$$0.7768698399 \qquad \textbf{(1.2.1.1.11)}$$

```
> `P(T > 15)` := 1 - `P(T < 15)`; evalf(%);
```

$$P(T > 15) := e^{-\frac{3}{2}}$$

$$0.2231301601 \quad (1.2.1.1.12)$$

```
> szinezesplot := plot(f(x), x = 15..60, filled = [color = yellow, transparency = 0.5]):
plots[display](surusegplot, szinezesplot);
```



e) Mennyi a valószínûsége, hogy egy jármû után a következô az 5. és 10. másodperc között érkezik?

Egy adott intervallumba esés azt jelenti, hogy kisebb az érték a felsô végpontnál, de nem kisebb az alsónál. Vagyis az ide esés valószínûségét a felsô végpont alá és az alsó végpont alá esés valószínûségeinek különbsége adja:

```
> `P(5 <= T < 10)` := F(10) - F(5); evalf(%);
```

$$P(5 <= T < 10) := -e^{-1} + e^{-\frac{1}{2}}$$

$$0.2386512185 \quad (1.2.1.1.13)$$

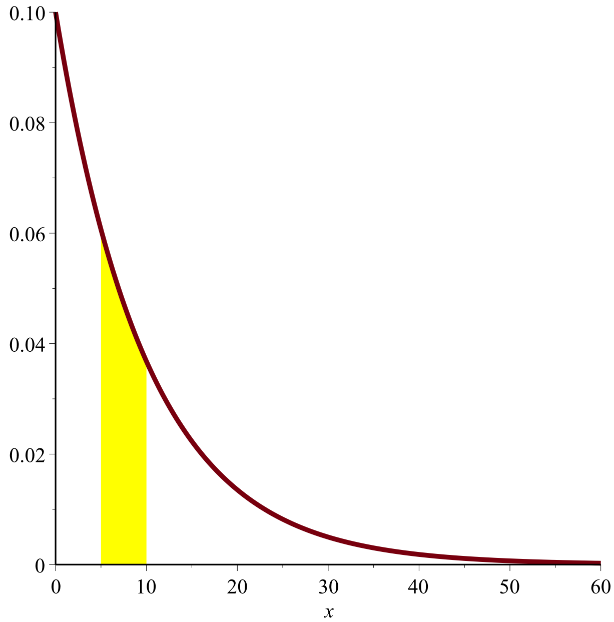
Vagy másként:

```
> Probability({5 <= T, T < 10}); evalf(%);
```

$$-e^{-1} + e^{-\frac{1}{2}}$$

$$0.2386512185 \quad (1.2.1.1.14)$$

```
> szinezesplot := plot(f(x), x = 5..10, filled = [color =
  yellow, transparency = 0.5]):
  plots[display](surusegplot, szinezesplot);
```



Érdekes észrevenni, hogy az idôhézag fentebb kiszámolt, elsô ötmásodperces tartományba esésének valószínûsége nagyobb, mint az iménti, második ötmásodperces tartományba esésé - annak ellenére, hogy az átlagos várakozási idôhöz (10 másodperc) az utóbbi tartomány van közelebb! Itt is ugyanaz a magyarázat, mint amit az a) pontnál említettünk.

2. Feladat (Dobások összege)*

Dobjunk egyszerre k dobókockával és jelölje X a dobott értékek összegét!
a) Készítsünk szimulációt $k = 1, 2, 3$ és 5 kocka esetén $n = 10000$ dobással és ábrázoljuk hisztogramon a tapasztalati eloszlást!
b) Illesszünk a $k = 5$ esetben kapott szimulált eloszlásra normál eloszlást és hasonlítsuk össze annak sûrûségfüggvényét a hisztogrammal!

Megoldás

```
> restart;
```

a) Készítsünk szimulációt $k = 1, 2, 3$ és 5 kocka esetén $n = 10000$ dobással és ábrázoljuk hisztogramon a tapasztalati eloszlást!

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

Felvesszük az adatokat: hány kockával dobunk, és hányszor tesszük ezt. A könnyû érthetôség kedvéért elôször mindössze két kockával és 10 dobással építsük fel a módszert:

```
> kockadb := 2; dobasdb := 10;
```

$$kockadb := 2$$

$$dobasdb := 10 \tag{1.2.2.1.1}$$

Megfelelô valószínûségi változóval szimuláljuk a kockadobást. Mivel több kockával játszunk, mindegyiknek külön létrehozunk egy valószínûségi változót – egy valószínûségi változókból álló tömböt készítünk:

```
> Y := [seq(RandomVariable(DiscreteUniform(1, 6)), i = 1..
  kockadb)];
```

$$Y := [\_R, \_R0] \tag{1.2.2.1.2}$$

A listában összegyûjtött valószínûségi változók mindegyikével generálunk egy mintát, és a megfelelô értékpárokat össze is adjuk.

```
> osszegek := add(Sample(Y[i], dobasdb), i = 1..kockadb);
  # vektorokat elemenként össze lehet adni az add
  paranccsal
```

$$osszegek := \left[\begin{array}{cccccccccc} 11. & 8. & 8. & 2. & 3. & 9. & 6. & 4. & 2. & 5. \end{array}\right] \tag{1.2.2.1.3}$$

Adjuk meg az összegek lehetséges értékeit és a hozzájuk tartozó valószínûségeket is! (Ezek már nem kellenek a hisztogram elkészítéséhez.)

```
> osszeg_ertekek := [seq(i, i = kockadb..6*kockadb)];
```

$$osszeg\_ertekek := [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12] \tag{1.2.2.1.4}$$

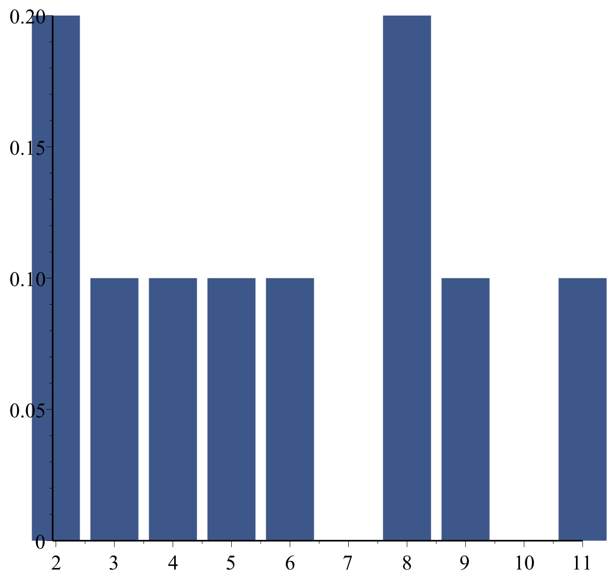
```
> osszeg_elofordulasok := [seq(0, i = kockadb..6*kockadb)]
  :
  for i to numelems(osszegek) do # numelems() kell, a nops
  () vektorokra nem az elemek számát adja
    ertekindex := round(osszegek[i] - kockadb + 1): #
  összeghez tartozó index
    osszeg_elofordulasok[ertekindex] :=
  osszeg_elofordulasok[ertekindex] + 1:
  end do:
  osszeg_elofordulasok;
  osszeg_valoszinusegek := osszeg_elofordulasok/dobasdb;
```

$$[2, 1, 1, 1, 1, 0, 2, 1, 0, 1, 0]$$

$$osszeg\_valoszinusegek := \left[\frac{1}{5}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, 0, \frac{1}{5}, \frac{1}{10}, 0, \frac{1}{10}, 0\right] \tag{1.2.2.1.5}$$

Ábrázoljuk az eloszlást hisztogramon!

```
> Histogram(osszegek, discrete = true, thickness = 30);
```



Érdemes összevetni a dobott összegeket tartalmazó vektor adatait a diagrammal.
A mechanizmus megértése kedvéért a fenti példában kevés dobást végeztünk, de az egész folyamat könnyen megismételhetô több dobással is. Legjobb, ha egy paraméterezhetô procedúrában foglaljuk össze a fentebb lépésenként kifejtett mûveletsort, miközben a köztes mûveletek végén kettôspontra cseréljük a pontosvesszôket, csak a hisztogram rajzoló utasítás mögött hagyjuk meg.

```
> kockadobalas := proc(kocka_db, dobas_db)
    local Y, szamok, osszegek:
    Y := [seq(RandomVariable(DiscreteUniform(1, 6)), i =
  1..kocka_db)]:
    osszegek := add(Sample(Y[i], dobas_db), i = 1..
  kocka_db):
    return Histogram(osszegek, discrete = true, thickness
  = round(330/(5*kocka_db + 1) - 3));
  end proc:
```

Ezzel már könnyedén kísérletezhetünk különbözô kocka- és dobás-számokkal.

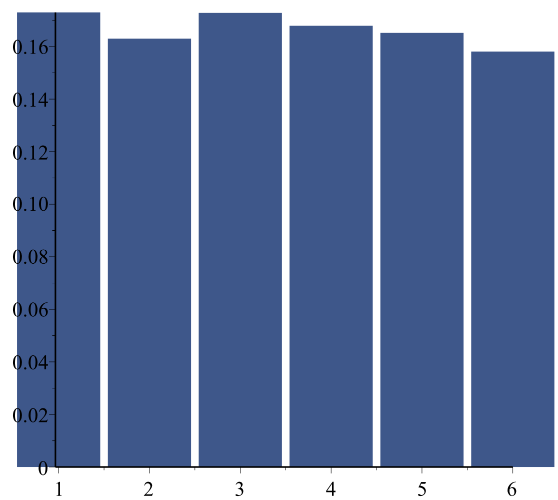
```
> n := 10000;
```

$$n := 10000 \quad \textbf{(1.2.2.1.6)}$$

$k = 1$ eset (egy kockával dobunk).

```
> k := 1;
  kockadobalas(k, n);
```

$$k := 1$$

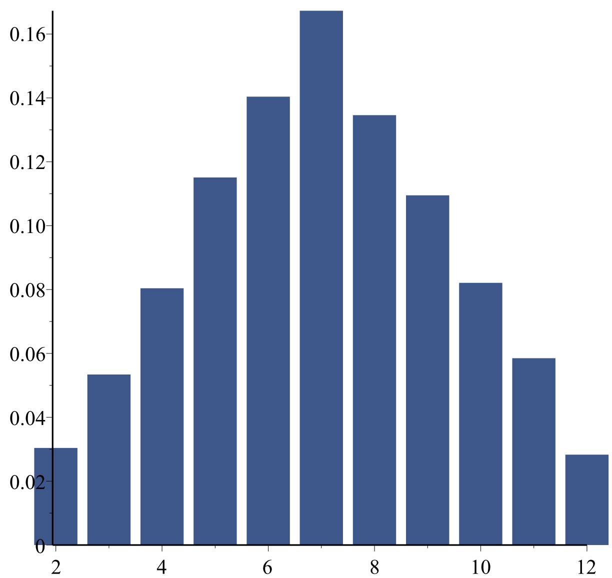


Látjuk, hogy nagyobb dobás-szám mellett a hisztogram kisimul, és közel egyenletes eloszlást mutat.

$k = 2$:

```
> k := 2;
  kockadobalas(k, n);
```

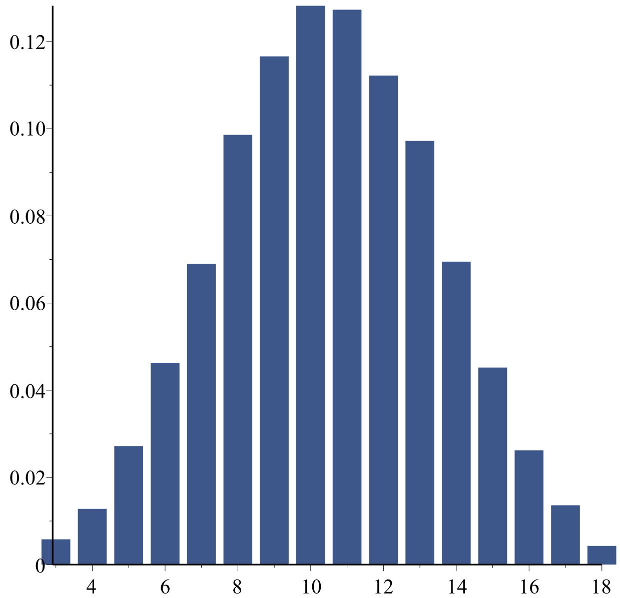
$$k := 2$$



A kellôen sok dobás miatt a hisztogram véletlenszerû ingadozásai csekélyek: szabályos alakot, de semmiképp sem egyenletes eloszlást mutat.
Látjuk, hogy a két kockával való dobáskor a középsô összegek gyakrabban jelennek meg, mint a szélsôk. Ez érthetô, hiszen a 2-es értékhez csakis két 1-es, a 12-eshez csakis két 6-os dobással lehet eljutni, míg pl. a 7-es érték többféleképp is összejöhet a két kockából: $1+6$, $2+5$, $3+4$, $4+3$, $5+2$, $6+1$. A hisztogramon is épp ez látszik: a középsô 7-es érték kb. 6-szor olyan gyakran fordul elô, mint a szélsôk.
Könnyen kipróbálható a 3 vagy még több kockás eset is.
$k = 3$:

```
> k := 3;
  kockadobalas(k, n);
```

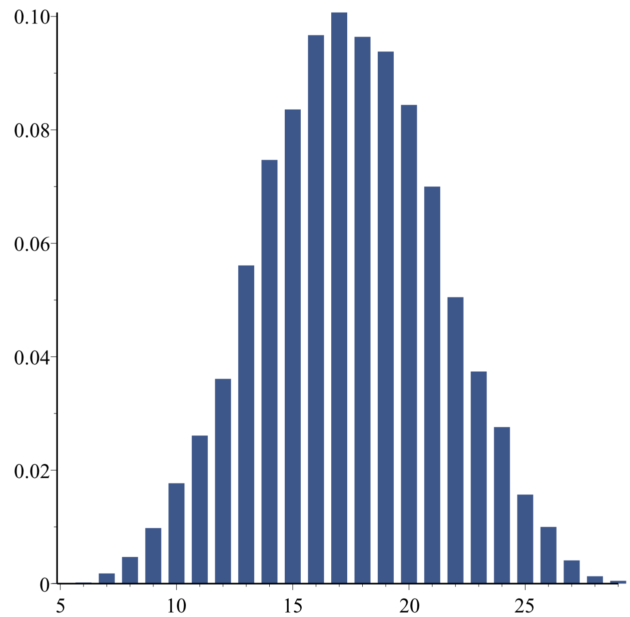
$$k := 3$$



$k = 5$:

```
> k := 5;
  H := kockadobalas(k, n): H;
```

$$k := 5$$



A fentebb kifejtett alapelv továbbra is igaz: a középsô értékek nagyobb valószínûségûek, mert többféle kombinációval jöhetnek létre, mint a szélsôk. A két kockás eset háromszög alakú grafikonjához képest viszont szépen kivehetô, hogy az ábra alakja haranghoz válik hasonlatossá: a hisztogram közepe konvex, de a szélek egy bizonyos ponttól kezdve már konkávok.

b) Illesszünk a $k = 5$ esetben kapott szimulált eloszlásra normál eloszlást és hasonlítsuk össze annak sûrûségfüggvényét a hisztogrammal!

A fenti kísérletben a k növelésével kirajzolódó haranggörbe létrejötte mögött a *centrális határeloszlás tétel* áll, melynek következménye, hogy véges szórás esetén sok független, azonos eloszlású valószínûségi változó összege jól közelíthetô az összeggel azonos várható értékû és szórású *normál eloszlású* változóval.

Jelölje Y_i, $(i = 1, 2, ..., 5)$ az i-edik kockával dobott számot és legyen Y ezek összege! Ekkor Y eloszlása jól közelíthetô egy $N(\mu, \sigma)$ normál eloszlással, ahol $\mu = E(Y)$ és $\sigma = \sigma(Y)$.

, így $E(Y_i) = \frac{6+1}{2} = 3.5$ és a várható érték additivitása miatt $E(Y) = 5 \cdot E(Y_1) = 17.5$.

Továbbá a kockák függetlenek, így a szórásnégyzet szintén additív:

$$\sigma^2(Y) = \sigma^2\left(\sum_{i=1}^{5} Y_i\right) = 5 \cdot \sigma^2(Y_1).$$

```
> EY1 := 3.5; EY := k*EY1;
```

$$EY1 := 3.5$$

$$EY := 17.5 \quad (1.2.2.1.7)$$

```
> VarY1 := add((i - EY1)^2*1/6, i = 1..6);
```

$$VarY1 := 2.916666668 \quad (1.2.2.1.8)$$

```
> VarY := k*VarY1;
  sigmaY := sqrt(VarY);
```

$$VarY := 14.58333334$$

$$sigmaY := 3.818813080 \quad (1.2.2.1.9)$$

Ugyanez a számolás a Maple beépített eszközeivel:

```
> unassign('i');
  Y := add(RandomVariable(DiscreteUniform(1,6)), i = 1..k)
  ; # sum() nem jó csak az add()
```

$$Y := \_R12 + \_R13 + \_R14 + \_R15 + \_R16 \quad (1.2.2.1.10)$$

```
> ExpectedValue(Y); evalf(%);
```

$$\frac{35}{2}$$

$$17.50000000 \quad (1.2.2.1.11)$$

```
> StandardDeviation(Y); evalf(%);
```

$$\frac{5}{6}\sqrt{21}$$

$$3.818813079 \quad (1.2.2.1.12)$$

Lássuk az illesztett normál eloszlás sûrûségfüggvényét:

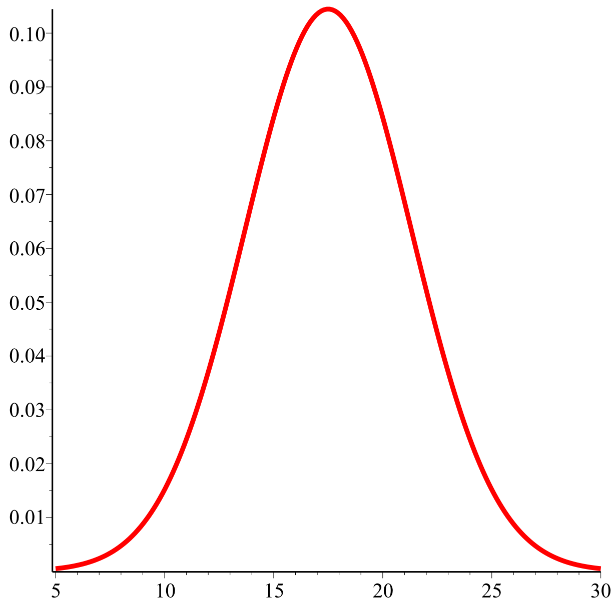
```
> X := RandomVariable(Normal(EY, sigmaY));
```

$$X := \_R17 \quad (1.2.2.1.13)$$

```
> 'f'(x) = PDF(X, x);
```

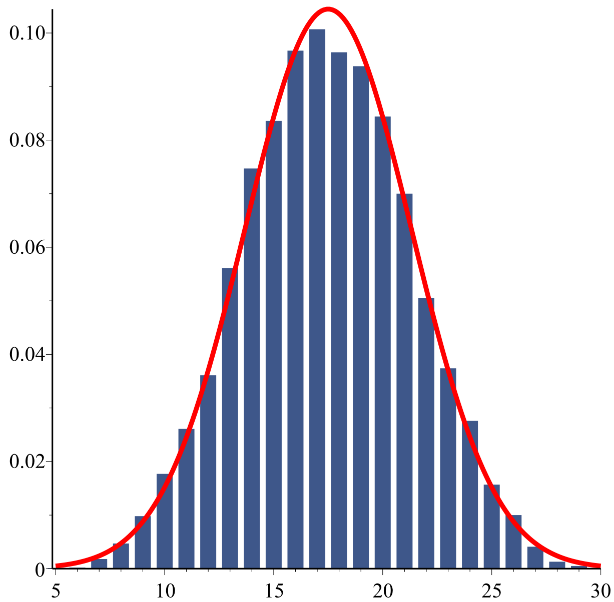
$$f(x) = \frac{0.1309307341\sqrt{2}\,\mathrm{e}^{-0.03428571427\,(x-17.5)^2}}{\sqrt{\pi}} \quad (1.2.2.1.14)$$

```
> P1 := DensityPlot(X, range = k..6*k, color = red,
  thickness = 3): P1;
```



Ezt rárajzolva a fenti hisztogramra:

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



Jól látszik a szoros illeszkedés.

3. Feladat (Egyenletes eloszlások összege)

Sorsoljunk egyenletes eloszlással számokat a [0, 1] intervallumból (egyszerre többet), és figyeljük ezek Y összegének eloszlását.
a) Készítsünk szimulációt $k=1, 2, 3$ ill. 5 szám összegére $n=10000$ kísérlettel! Ábrázoljuk hisztogramon a szimulált eloszlást és rajzoljuk rá az elméleti sûrûségfüggvényt!
b) Illesszünk a $k=5$ esetben kapott szimulált eloszlásra normál eloszlást és hasonlítsuk össze annak sûrûségfüggvényét a hisztogrammal!

Megoldás

```
> restart;
```

a) Készítsünk szimulációt $k=1, 2, 3$ ill. 5 szám összegére $n=10000$ kísérlettel! Ábrázoljuk hisztogramon a szimulált eloszlást és ralzoljuk rá az elméleti sûrûségfüggvényt!

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

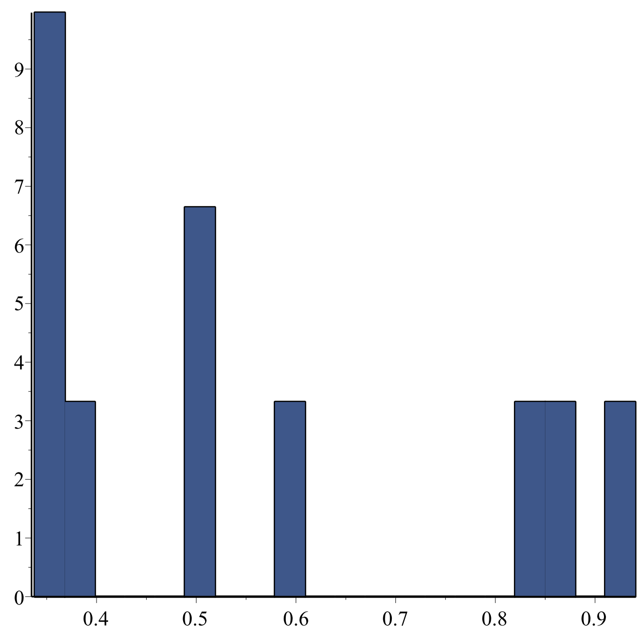
A fenti, több kockával való dobásról szóló (diszkrét) példa könnyen átvihetô folytonos esetre is.

Lényegében az ott látottakat ismételjük, talán még egyszerûbben is.

```
> egyenletesosszegzes := proc(valtozodb, sorsolasdb)
    local Y, osszegek:
    Y := [seq(RandomVariable(Uniform(0, 1)), i = 1..
  valtozodb)]:
    osszegek := add(Sample(Y[i], sorsolasdb), i = 1..
  valtozodb):
    return Histogram(osszegek, bincount = 20);
  end proc:
```

Tegyünk egy próbát kevés kísérlettel, egy változóra:

```
> egyenletesosszegzes(1, 10);
```

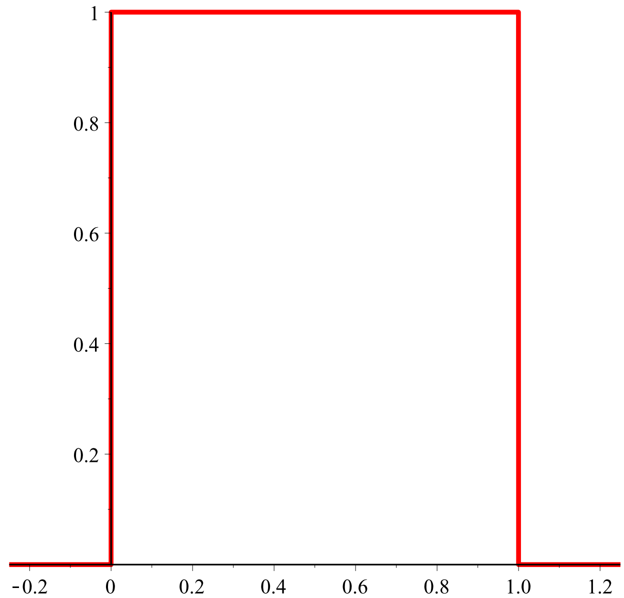


10000 kísérlettel egy változóra megkapjuk a várt egyenletes eloszlást:

```
> n := 10000;
```

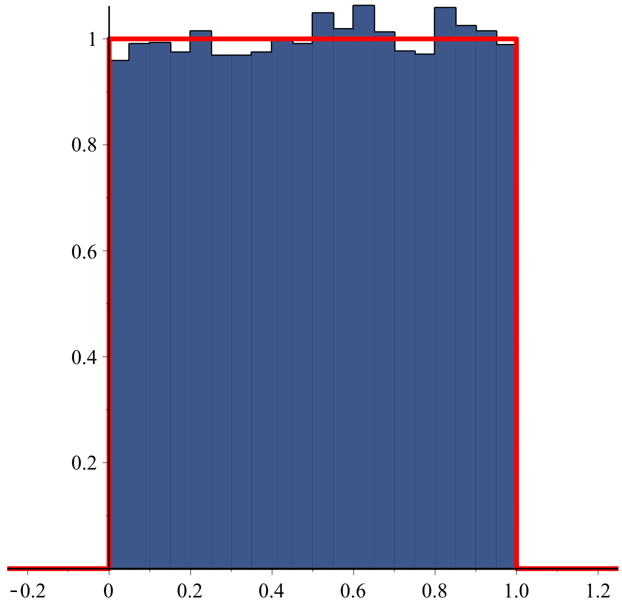
$$n := 10000 \tag{1.2.3.1.1}$$

```
> Y1 := RandomVariable(Uniform(0, 1)):
  P1 := DensityPlot(Y1, range = -0.25..1.25, color = red,
  thickness = 3): P1;
```



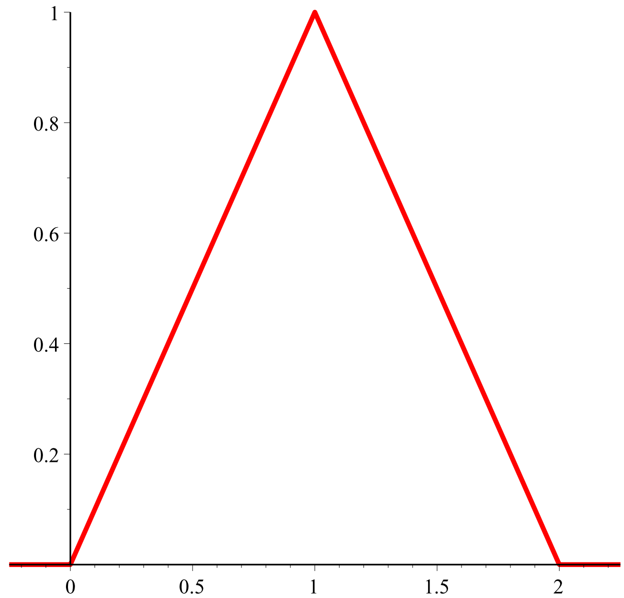

```
> k := 1;
  H1 := egyenletesosszegzes(k, n):
  plots[display](P1, H1);
```

$$k := 1$$



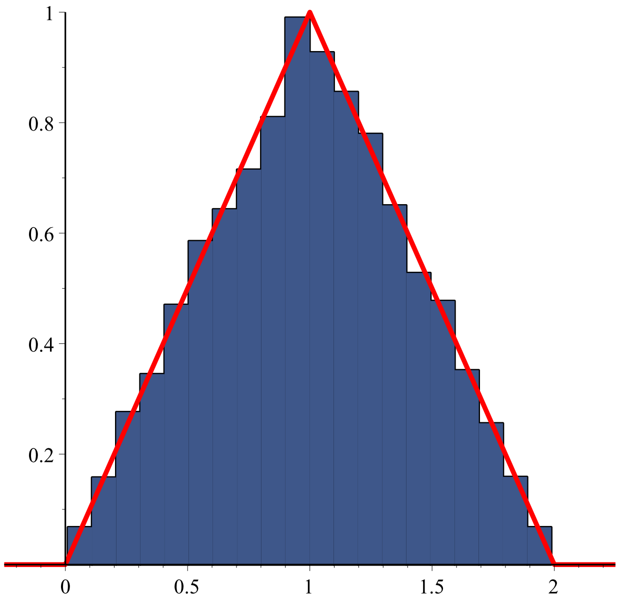
Nézzük hogy mutat ugyanez két változóval! A Maple képes közvetlenül megrajzolni a sûrûségfüggvényt, akár változók összegeként adódó újabb változókra is:

```
> Y2 := Y1 + RandomVariable(Uniform(0, 1)): # az elôzô
  változóhoz hozzáadunk még egyet
  P2 := DensityPlot(Y2, range = -0.25..2.25, color = red,
  thickness = 3): P2;
```



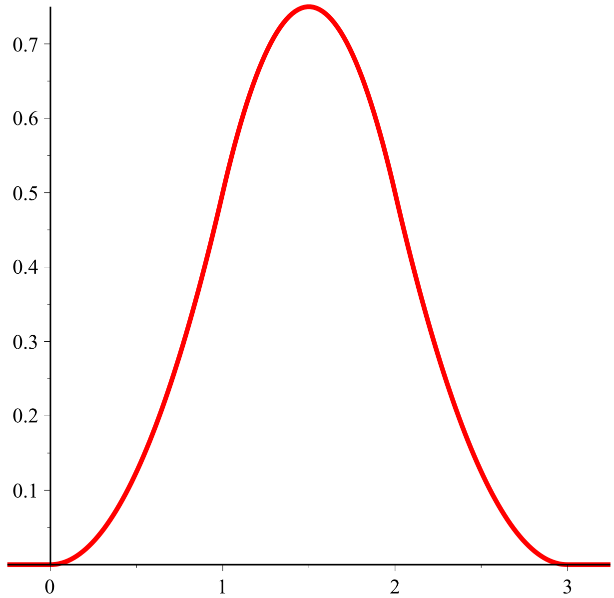

```
> k := 2;
  H2 := egyenletesosszegzes(k, n):
  plots[display](P2, H2);
```

$$k := 2$$



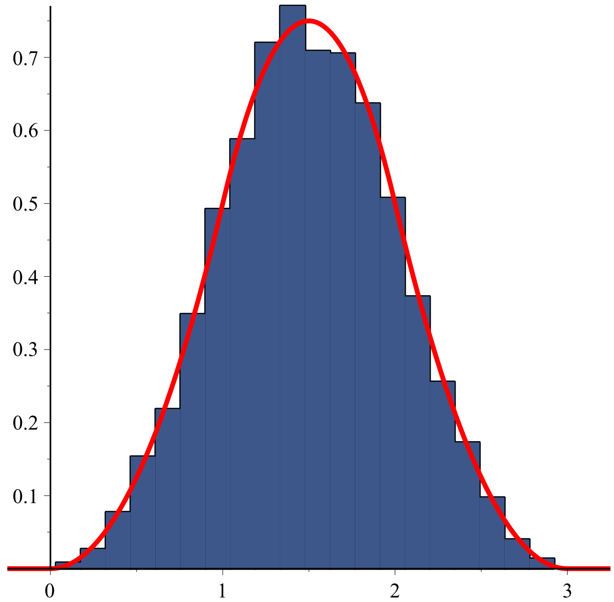
(Az értékkészlet így természetesen 0-tól 2-ig terjed.)
Három változóval:

```
> Y3 := Y2 + RandomVariable(Uniform(0, 1)): # az elôzô
  változóhoz hozzáadunk még egyet
  P3 := DensityPlot(Y3, range = -0.25..3.25, color = red,
  thickness = 3): P3;
```



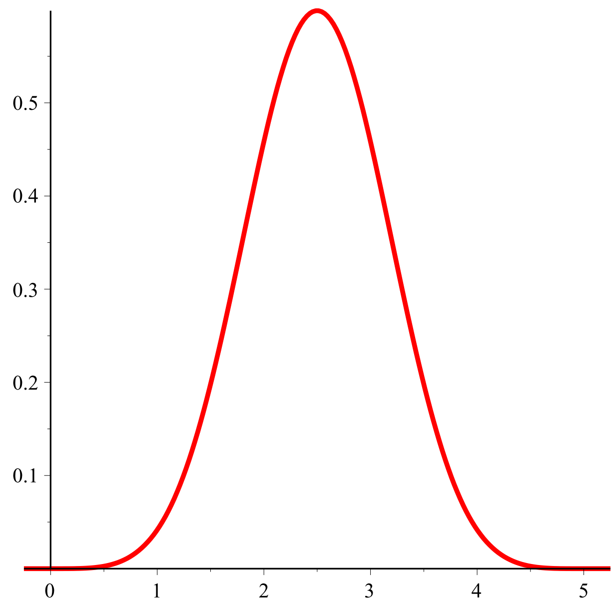

```
> k := 3;
  H3 := egyenletesosszegzes(k, n):
  plots[display](P3, H3);
```

$$k := 3$$



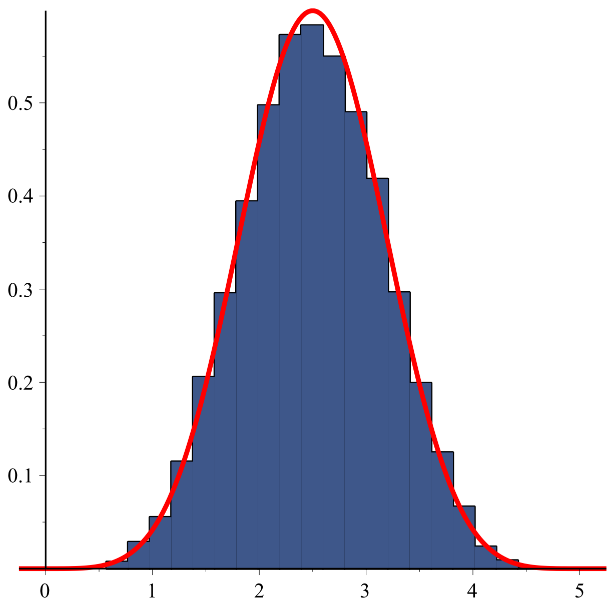
A $[0, 3]$ értékkészlet mellett az eloszlás már elkezdi felvenni a jellegzetes harang-formát.
Végül $k = 5$ esetén a hisztogram:

```
> Y4 := Y3 + RandomVariable(Uniform(0, 1)):
  Y5 := Y4 + RandomVariable(Uniform(0, 1)): # Vigyázzunk:
  ha egy sorban adjuk hozzá a két új egyenletes eloszlású
  valószínûségi változót, akkor azokat a Maple azonosnak
  fogja tekinteni, nem pedig függetlennek!
  P5 := DensityPlot(Y5, range = -0.25..5.25, color = red,
  thickness = 3): P5;
```




```
> k := 5;
  H5 := egyenletesosszegzes(k, n):
  plots[display](P5, H5);
```

$$k := 5$$



Látjuk, hogy ezek a görbék jól illeszkednek a nekik megfelelô hisztogramokra.
b) Illesszünk a $k=5$ esetben kapott szimulált eloszlásra normál eloszlást és hasonlítsuk össze annak sûrûségfüggvényét a hisztogrammal!
A 2. Feladat b) részéhez hasonlóan a k növelésével kirajzolódó haranggörbe mögött a *centrális határeloszlás tétel* húzódik meg, mely szerint Y nagy k esetén jól közelíthetô egy azonos várható értékû és szórású *normál eloszlású* változóval. Tehát csupán 2 paramétert kell meghatároznunk.
Jelölje Y_i $(i=1, 2, ..., 5)$ az i-edik véletlen számot. Ekkor $Y=Y_1+Y_2+...+Y_5$,
$E(Y_i)=\frac{0+1}{2}=0.5$, és a várható érték additivitása miatt

$$E(Y)=5\cdot E(Y_1)=2.5.$$

Továbbá a számokat függetlenül választottuk, így a szórásnégyzet szintén additív:

$$\sigma^2(Y)=\sigma^2\left(\sum_{i=1}^{5} Y_i\right)=5\cdot\sigma^2(Y_1)=5\cdot\frac{(1-0)^2}{12}=\frac{5}{12}.$$

(Itt felhasználtuk az $[a, b]$ intervallumon egyenletes eloszlású valószínûségi változó szórásnégyzetére tanult $\sigma^2=\frac{(b-a)^2}{12}$ képletet.)

```
> EY1 := 0.5; EY := k*EY1;
```

$$EY1 := 0.5$$

(1.2.3.1.2)

$$EY := 2.5 \quad (1.2.3.1.2)$$

```
> VarY1 := 1/12;
```

$$VarY1 := \frac{1}{12} \quad (1.2.3.1.3)$$

```
> VarY := k*VarY1;
  sigmaY := sqrt(VarY); evalf(%);
```

$$VarY := \frac{5}{12}$$

$$sigmaY := \frac{1}{6}\sqrt{15}$$

$$0.6454972245 \quad (1.2.3.1.4)$$

Ugyanez a számolás a Maple beépített eszközeivel:

```
> unassign('i');
  Y := add(RandomVariable(Uniform(0,1)), i = 1..k); # sum
  () nem jó csak az add()
```

$$Y := \_R16 + \_R17 + \_R18 + \_R19 + \_R20 \quad (1.2.3.1.5)$$

```
> ExpectedValue(Y); evalf(%);
```

$$\frac{5}{2}$$

$$2.500000000 \quad (1.2.3.1.6)$$

```
> StandardDeviation(Y); evalf(%);
```

$$\frac{1}{6}\sqrt{15}$$

$$0.6454972245 \quad (1.2.3.1.7)$$

Lássuk az illesztett normál eloszlás sûrûségfüggvényét:

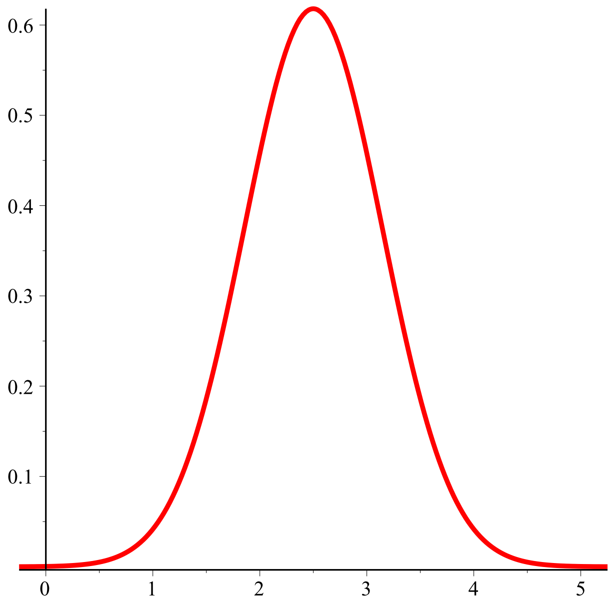
```
> X := RandomVariable(Normal(EY, sigmaY));
```

$$X := \_R21 \quad (1.2.3.1.8)$$

```
> 'f'(x) = PDF(X, x);
```

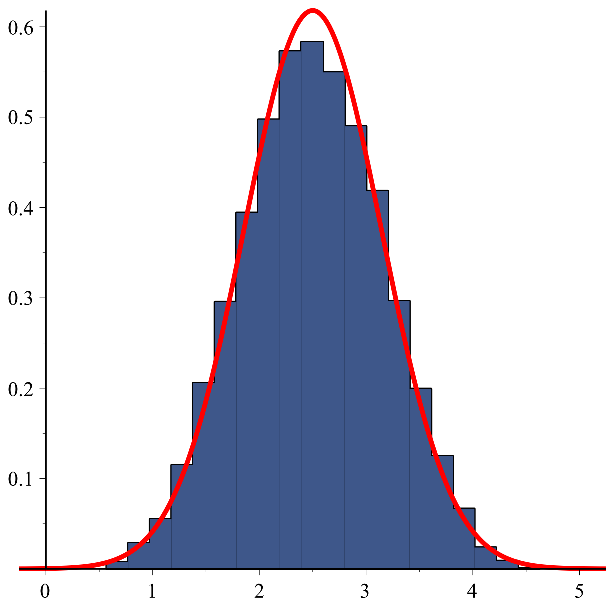
$$f(x) = \frac{1}{5} \frac{\sqrt{2}\sqrt{15}\,\mathrm{e}^{-\frac{6}{5}(x-2.5)^2}}{\sqrt{\pi}} \quad (1.2.3.1.9)$$

```
> PN := DensityPlot(X, range = -0.25..5.25, color = red,
  thickness = 3): PN;
```



Ezt rárajzolva a fenti hisztogramra:

```
> plots[display](H5, PN);
```



A normál sûrûségfüggvény görbéje jól láthatóan illeszkedik a hisztogramra, bár tudjuk, hogy nem teljesen egyezik az elméleti eloszlással, csak közelíti azt.

4. Feladat (Duzzasztógát)*

Egy duzzasztógát által felduzzasztott átlagos vízmélység 21 méter, és szórása 5 méter. A mindenkori *X* vízmélység méterben mérve normál eloszlású véletlen változó.
A gát tetején található egy biztonsági túlfolyó, amely megakadályozza, hogy a víz a 30 méteres szint fölé növekedjen. Másrészt van egy csô a túlfolyó teteje alatt 15 méterrel, amely bevezeti a vizet a hidroelektromos generátorhoz. (Hagyjuk figyelmen kívül a csô véges átmérôjét!)
a) Rajzoljuk fel az *Y* vízmélység sûrûségfüggvényét**!**
b) A teljes idô hányadrészében fordul elô, hogy a túlfolyón folyik el feleslegesen a felduzzasztott víz!
c) Szemléltessük a b) pontban számolt százalékot a sûrûség függvény görbéje alatti megfelelô terület **befestésével!**
d) Tudjuk, hogy a túlfolyón nem folyik a víz. Mekkora a feltételes valószínûsége ekkor annak, hogy a duzzasztó táplálja a generátort vízzel?
e) Az esetek hányadrészében nem táplálja a felduzzasztott víz a generátort?
f) Milyen *h* méter mélyre kellene betenni a generátort tápláló csövet a túlfolyó alatt ahhoz, hogy az esetek legfeljebb 5 *%*-ában ne kapjon vizet a generátor?

g) Szemléltessük az f) pontban számolt *h* mélységet, az *Y* vízmélység eloszlás függvény**ének és a 0.05 valószínûségi szintnek a felrajzolásával!**

Megoldás

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

Tegyük fel, hogy adott egy mérce a vízszint leolvasásához. Ennek 0 szintjéhez viszonyítunk minden adatot a példa megoldásakor.

a) Rajzoljuk fel az *Y* vízmélység sûrûség függvényét!

A felduzzasztott vízszintet az *Y* változó mutatja, mely a feltételezés szerint *normál* eloszlású, adott várható értékkel és szórással. Hosszuk létre a megfelelô eloszlású változót:

```
> mu := 21; sigma := 5;
```

$$\mu := 21$$

$$\sigma := 5 \tag{1.2.4.1.1}$$

```
> Y := RandomVariable(Normal(mu, sigma));
```

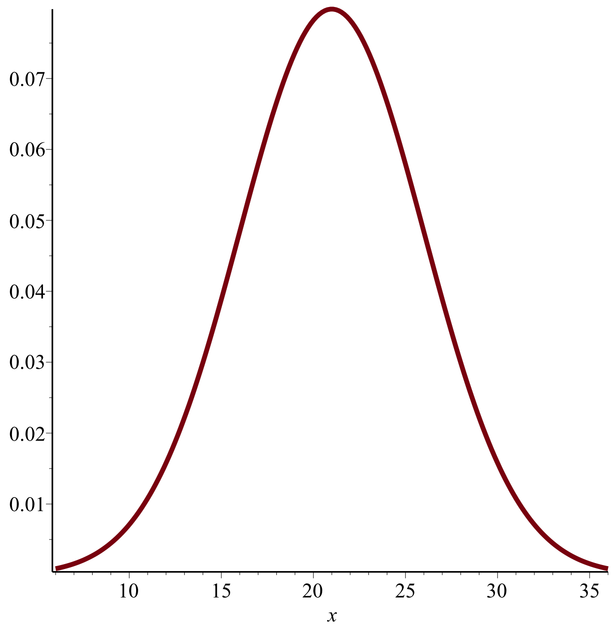
$$Y := \_R \tag{1.2.4.1.2}$$

Ebbôl már megrajzolható a sûrûségfüggvény:

```
> f := x -> PDF(Y, x);
```

$$f := x \rightarrow Statistics:\text{-}PDF(Y, x) \tag{1.2.4.1.3}$$

```
> surusegplot := plot(f(x), x = mu - 3*sigma..mu + 3*
  sigma, thickness = 3): surusegplot; # 3-szigma szabály
```



b) A teljes idô hányadrészében fordul elô, hogy a túlfolyón folyik el feleslegesen a felduzzasztott víz!
Itt azt kérdezzük, mekkora valószínûséggel van a vízszint 30 méter fölött: $P(Y > 30) = ?$
Az ilyen (adott intervallumba esés valószínûsége) típusú kérdésekre az eloszlásfüggvény és a sûrûségfüggvény használatával is lehet válaszolni. Vegyük most a valószínûségi változónk eloszlásfüggvényét:

```
> F := x -> CDF(Y, x);
```

$$F := x \rightarrow Statistics{:-}CDF(Y, x) \tag{1.2.4.1.4}$$

Ez a függvény - definíciója szerint - közvetlenül az adott érték *alá* esés valószínûségét képes mutatni, de közvetve adott érték fölé, vagy adott értékek közé esés valószínûsége is jól számolható vele. Például a kérdésünkben keresett valószínûséghez az is jó, ha a 30 alá esés valószínûségét tudjuk, mert annak komplementere lesz a megoldás:

```
> `P(Y > 30)` := evalf(1 - F(30));
```

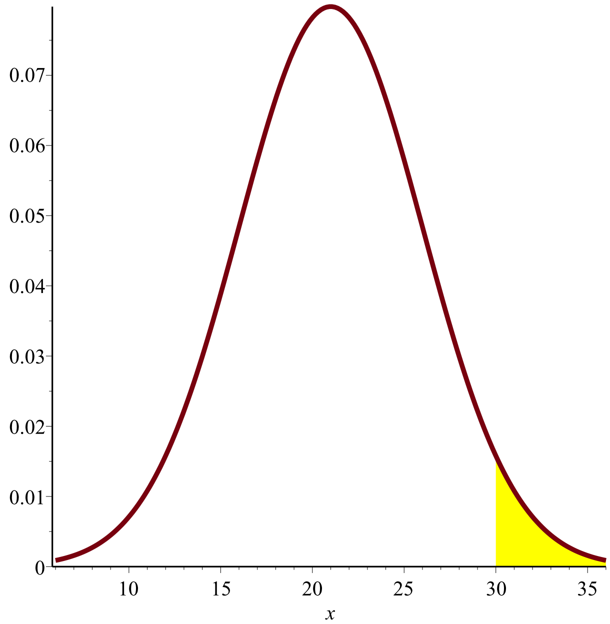
$$P(Y > 30) := 0.0359303192 \tag{1.2.4.1.5}$$

Eszerint ~3.6 % valószínûséggel (a mérési idôtartam átlagosan ennyi részében) lesz túlfolyás.

c) Szemléltessük a b) pontban számolt százalékot a sûrûség függvény görbéje alatti megfelelô terület befestésével!
Fenti eredmény szemléletesen azt jelenti, hogy a sûrûségfüggvény görbéje alatt "szétkent" egységnyi valószínûségbôl ennyi esik a 30-as abszcissza fölé. Ehhez készítünk egy szemléltetô ábrát.

Elkészítjük a befestést. majd összemásoljuk a görbe rajzával:

```
> szinezesplot := plot(f(x), x = 30..mu + 3*sigma, filled
  = [color = yellow, transparency = 0.5]):
  plots[display](szinezesplot, surusegplot);
```



d) Tudjuk, hogy a túlfolyón nem folyik a víz. Mekkora a feltételes valószínûsége ekkor annak, hogy a duzzasztó táplálja a generátort vízzel?

A generátort tápláló kifolyó a túlfolyó alatt 15 méterrel van, ez a mércén 15 méternek felel meg.

A szövegesen megfogalmazott kérdés a feltételes valószínûség tanult definícióját alkalmazva a következôképp formalizálható:

$$P\left(Y > 15 \middle| Y < 30\right) = \frac{P(Y > 15 \wedge Y < 30)}{P(Y < 30)} = \frac{P(15 < Y < 30)}{P(Y < 30)}$$

A feltételes valószínûség két, külön kiszámítandó valószínûség hányadosa:

```
> `P(15 < Y < 30)` := evalf(F(30)-F(15));
```

$P(15 < Y < 30) := 0.8490000105$ **(1.2.4.1.6)**

```
> `P(Y < 30)` := evalf(F(30));
```

$P(Y < 30) := 0.9640696808$ **(1.2.4.1.7)**

```
> `P(15 < Y < 30)`/`P(Y < 30)`;
```

0.8806417497 **(1.2.4.1.8)**

Tehát annak valószínûsége, hogy a duzzasztó táplálja a generátort vízzel, feltéve, hogy a túlfolyón nem folyik a víz, kb. 88 %.

e) Az esetek hányadrészében nem táplálja a felduzzasztott víz a generátort?

Mekkora valószínûséggel van a vízszint 15 méter alatt?

```
> evalf(F(15));
```

$$0.1150696703 \quad (1.2.4.1.9)$$

11.5 % az esélye annak, hogy nem táplálja a felduzzasztott víz a generátort.

f) Milyen *h* méter mélyre kellene betenni a generátort tápláló csövet a túlfolyó alatt ahhoz, hogy az esetek legfeljebb 5 %-ában ne kapjon vizet a generátor?

Az elôzô pontban láttuk, hogy az átlagosan 21 méterestôl egyre nagyobb eltéréseket egyre ritkábban produkáló vízszint 11.5 % valószínûséggel 15 méter alatt van. A generátor ilyenkor nem üzemel. Ha szeretnénk megnövelni az üzemelési idôt úgy, hogy 11.5 % helyett csak 5% legyen a kiesés, a kifolyási szintet mélyebbre kell tenni (ahol nagyobb valószínûséggel van víz). Kérdés, hogy hol legyen ez, milyen h szint esetén lenne az az alá esés valószínûsége 5 %? A formalizált feladat: $P(Y < h) = 0.05$. Vagyis meg kell oldani az $F(h) = 0.05$ egyenletet:

```
> h := fsolve(F(h) = 0.05, h); # solve helyett fsolve
  kell, ha csak a valós megoldásokat keressük
```

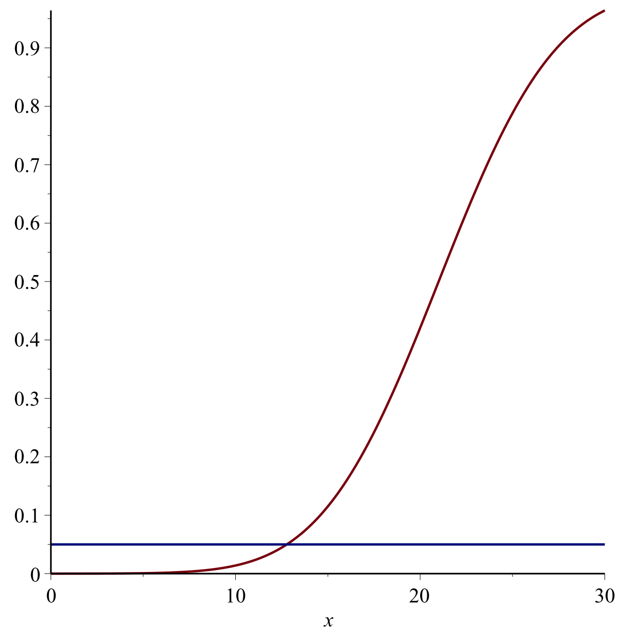
$$h := 12.77573187 \quad (1.2.4.1.10)$$

Ha tehát 12.77 méteres szint alá tesszük a kifolyót, akkor lehet garantálni átlagosan 5 %-nál kevesebb leállási idôt.

g) Szemléltessük az f) pontban számolt *h* mélységet, az *Y* vízmélység eloszlás függvényének és a 0.05 valószínûségi szint felrajzolásával!

A fenti eredményt szemléltethetjük grafikonon is:

```
> plot([F(x), 0.05], x = 0..30);
```



5. Feladat (Ipari fúrógép)*

Egy ipari fúrógép fúrófejének élettartama jól közelíthetô egy 10 óra várható értékû és 2 óra szórású valószínûségi változóval. Legalább hány fúrófej kell ahhoz, hogy a gépet legalább 95 %-os valószínûséggel 400 óráig mûködtetni tudjuk? (Feltehetô, hogy az egyes fúrófejek élettartama egymástól független.)

Megoldás

```
> restart;
```

Jelölje $X_1, X_2, X_3, \ldots$ a fúrófejek élettartamát. A feladat szövege szerint X_i-k független, azonos eloszlású valószínûségi változók, $\mu_X = \mathrm{E}(X_i) = 10$ h várható értékkel és $\sigma_X = \sigma(X_i) = 2$ h szórással. Legyen $S_n = X_1 + X_2 + \ldots + X_n$ az elsô n db fúrófej élettartamának összege. A feladat szövege szerint úgy kell n-et megválasztanunk, hogy

$$P(S_n \geq 400) \geq 0.95$$

teljesüljön.

```
> muX := 10;
  sigmaX := 2;
  t := 400;
  p := 0.95;
```

$$muX := 10$$
$$sigmaX := 2$$
$$t := 400$$
$$p := 0.95 \tag{1.2.5.1.1}$$

Fontos, hogy bár az X_i változók pontos eloszlása *nem ismert*, az S_n összeg eloszlását mégis jól tudjuk közelíteni egy megfelelô normál eloszlással, ha n elég nagy ($n \geq 30$ általában jó). Ez a matematikai statisztikában rendkívül fontos állítás a *Centrális Határeloszlás Tétel* (CHT). A tétel minden eloszlásra (diszkrétre és folytonosra is) igaz, melynek létezik szórása; az összeadandók azonos eloszlása és függetlensége szükséges feltétel!
A CHT-t általában akkor akalmazzuk, ha a kérdés az S_n összeg eloszlására vonatkozik, de az összeadandók eloszlása nem ismert, így S_n eloszlását sem tudjuk precízen meghatározni. Ekkor az egyetlen lehetôségünk, ha elég nagy n-et feltételezve normál határeloszlással közelítjük az összeg eloszlását. Ennek paraméterei (a konkrét példában):

$$\mu = E(S_n) = n \cdot E(X_1) = n \cdot \mu_X$$

és

$$\sigma = \sqrt{\sigma^2(S_n)} = \sqrt{n \cdot \sigma^2(X_1)} = \sqrt{n}\,\sigma(X_1) = \sqrt{n} \cdot \sigma_X$$

Azaz $S_n \sim N(10\,n, 2 \cdot \sqrt{n})$.
A gyakorlatban elterjedt, hogy ehelyett S_n standardizáltjával számolunk, és azt közelítjük standard normál eloszlással: $S_n^* = \dfrac{S_n - n \cdot \mu_X}{\sqrt{n} \cdot \sigma_X} \sim N(0, 1)$. Ez azért jó, mert így lényegében csak egy konkrét eloszlás (standard normál) ismeretére van szükség bármely kísérlet leírásához. A szakkönyvekben is csak a standard normál eloszlásra vonatkozó négyjegyû függvénytáblázattal találkozhatunk, abból ugyanis bármely normál eloszlásfüggvény értékei számolhatók a standardizálási trükkel. A számítógépes statisztikai programcsomagok elterjedésével ez manapság már nem jelent megkötést, de a normál eloszlás jobb megértéséhez mégis érdemes így végigszámolni a feladatot.

$$0.95 \leq P(S_n \geq 400) = P\left(\frac{S_n - n \cdot \mu_X}{\sqrt{n} \cdot \sigma_X} \geq \frac{400 - n \cdot \mu_X}{\sqrt{n} \cdot \sigma_X}\right) \quad \text{(standardizálás)}$$
$$= P\left(S_n^* \geq \frac{400 - 10\,n}{2 \cdot \sqrt{n}}\right) = P\left(S_n^* \geq \frac{200 - 5\,n}{\sqrt{n}}\right)$$
$$= 1 - P\left(S_n^* < \frac{200 - 10\,n}{2 \cdot \sqrt{n}}\right) \quad / \, S_n^* \sim N(0, 1)$$
$$= 1 - \Phi\left(\frac{200 - 10\,n}{2 \cdot \sqrt{n}}\right),$$

ahol $\Phi(x)$ a standard normál eloszlás eloszlásfüggvényét jelöli. Ez ekvivalens azzal a feltétellel, hogy $\Phi\left(\dfrac{200 - 10\,n}{2 \cdot \sqrt{n}}\right) \leq 0.05$, melybôl Φ inverzét véve

$$\frac{200 - 10\,n}{2 \cdot \sqrt{n}} \leq \Phi^{-1}(0.05)$$

Ebbôl kell meghatározni az ismeretlen n-et! Eloszlásfüggvény inverzét a *Quantile* eljárással kaphatjuk meg:

```
> with(Statistics):
  Phi_inverse := x -> Quantile(Normal(0, 1), x);
```

$$Phi\_inverse := x \rightarrow Statistics:\text{-}Quantile(Normal(0, 1), x) \quad \textbf{(1.2.5.1.2)}$$

```
> RealDomain[solve]((t - n*muX)/(sqrt(n)*sigmaX) <=
  Phi_inverse(1 - p), n); evalf(%); # a RealDomain[solve]
  eljárás csak a valós megoldásokat keresi
```

$$RealRange(42.13540794, \infty)$$

$$RealRange(42.13540794, \infty) \quad \textbf{(1.2.5.1.3)}$$

Tehát $n \geq 42.135$, vagyis $n = 43$ fúrófej kell legalább ahhoz, hogy a gépet legalább 95 %-os valószínûséggel 400 óráig mûködtetni tudjuk!
Megjegyzés: mivel $n \geq 30$, ezért a megoldáshoz használt normál közelítésnek elhanyagolható a hibája.

Megoldás standardizálás nélkül:

```
> unassign('n');
  Y := RandomVariable(Normal(n*muX, sqrt(n)*sigmaX)); #
  S_n normál közelítése CHT-vel
```

$$Y := \_R0 \quad \textbf{(1.2.5.1.4)}$$

```
> egyenlet := Probability(Y >= 400) = p; # a feltételünk
  (egyenlettel kell megadni), nincs standardizálás
```

$$egyenlet := \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{1}{4}\,\frac{(-10\,n + 400)\sqrt{2}}{\sqrt{n}}\right) = 0.95 \quad \textbf{(1.2.5.1.5)}$$

```
> n1, n2 := RealDomain[solve](egyenlet, n); # két
  megoldást ad
```

$$n1, n2 := 37.97281379, 42.13540794 \quad \textbf{(1.2.5.1.6)}$$

Így is megtaláltuk a megoldást, de sajnos bejött egy hamis gyök is, amit csak ellenôrzéssel tudunk kiszûrni:

```
> n := n1;
  Probability(Y >= 400); evalf(%);
```

$$n := 37.97281379$$

$$\frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{1}{4}\,\frac{(-10\,n + 400)\sqrt{2}}{\sqrt{n}}\right)$$

$$0.0499999996 \quad \textbf{(1.2.5.1.7)}$$

Ez rossz megoldás, mert 0.95-öt kell kapnunk!

```
> n := n2;
  Probability(Y >= 400); evalf(%);
```

$$n := 42.13540794$$

$$\frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{1}{4}\,\frac{(-10\,n + 400)\sqrt{2}}{\sqrt{n}}\right)$$

$$0.9499999996 \quad \textbf{(1.2.5.1.8)}$$

Ez a helyes megoldás! A felfelé kerekített egész darabszám lesz az, amivel már biztosan garantálni tudjuk a feltétel teljesülését, vagyis $n = 43$.
Ábrázoljuk a p valószínûséget a normál sûrûségfüggvény alatti megfelelô terület besatírozásával!

```
> n := 43;
```

(1.2.5.1.9)

$$n := 43 \tag{1.2.5.1.9}$$

Határozzuk meg a várható értéket és a szórást, hogy tudjuk, melyik tartományon kell ábrázolni!

```
> mu := ExpectedValue(Y); # = n*muX
```

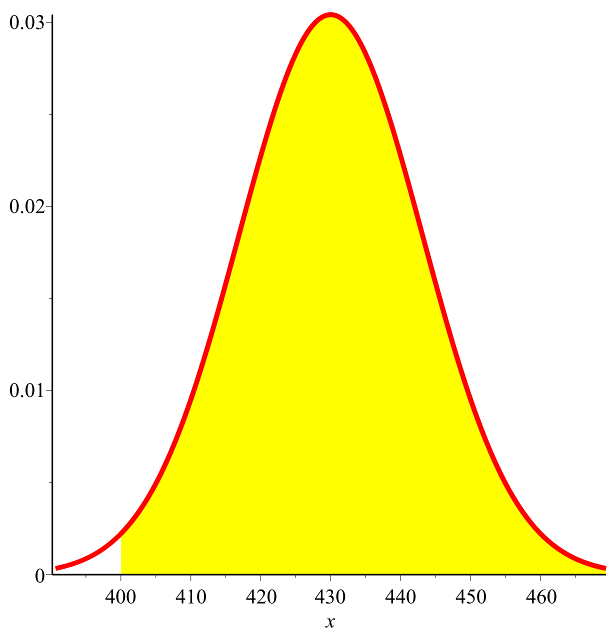
$$\mu := 430 \tag{1.2.5.1.10}$$

```
> sigma := StandardDeviation(Y); # = sqrt(n)*sigmaX
```

$$\sigma := 2\sqrt{43} \tag{1.2.5.1.11}$$

A tartomány kiválasztásánál a 3 σ-szabályt alkalmazzuk, azaz a $[\mu - 3\,\sigma, \mu + 3\,\sigma]$ intervallumon ábrázolunk.

```
> P1 := DensityPlot(Y, range = mu - 3*sigma..mu + 3*sigma, color = red, thickness = 3):
  P2 := plot(PDF(Y, x), x = 400..mu + 3*sigma, color = red, thickness = 3, filled = [color = yellow, transparency =0.5]): # DensityPlot-nál nem használható a 'filled' opció
  plots[display](P1, P2);
```



A satírozott terület n választása miatt legalább 95 %.

6. Feladat (Helymeghatározás)

Egy digitális helymeghatározó rendszer a $P_0(x_0, y_0, z_0)$ pontban elhelyezkedô jármûvet a $P(X, Y, Z)$ koordinátákkal jellemez, ahol, X, Y ill. Z független normál eloszlású valószínûségi változók x_0, y_0 ill. z_0 várható értékkel, és azonos $\sigma = 50$ cm szórással. Határozzuk meg annak a valószínûségét, hogy a helymeghatározó rendszer 1 m-nél többet téved

Megoldás

```
> restart;
```

Jelölje $\mathcal{E}$ a helymeghatározó rendszer méterben mért tévedését. Feladatunk a $P(\mathcal{E} > 1)$ valószínûség meghatározása.
$\mathcal{E}$ nem más, mint a P és P_0 pontok távolsága, amit a koordinátageometriából tanult képlettel adhatunk meg:

$$\mathcal{E} = |P - P_0| = \sqrt{(X - x_0)^2 + (Y - y_0)^2 + (Z - z_0)^2}$$

Ekkor

$$P(\mathcal{E} > 1) = P\left(\sqrt{(X - x_0)^2 + (Y - y_0)^2 + (Z - z_0)^2} > 1\right)$$
$$= P(\mathcal{E}^2 > 1^2) = P\left(\frac{\mathcal{E}^2}{\sigma^2} > \frac{1}{\sigma^2}\right) = P\left(Q > \frac{1}{\sigma^2}\right)$$

ahol

$$Q = \frac{\mathcal{E}^2}{\sigma^2} = \left(\frac{X - x_0}{\sigma}\right)^2 + \left(\frac{Y - y_0}{\sigma}\right)^2 + \left(\frac{Z - z_0}{\sigma}\right)^2 = (X^*)^2 + (Y^*)^2 + (Z^*)^2 \sim \chi^2(3),$$

azaz Q egy $n = 3$ szabadsági fokú χ^2-eloszlású valószínûségi változó, mert 3 standard normál eloszlású valószínûségi változó négyzetösszege (X^*, Y^*, Z^* éppen az X, Y, Z standardizáltja).
Vegyük fel az ennek megfelelô valószínûségi változót!

```
> sigma := 0.5; # méterben megadva
```

$$\sigma := 0.5 \tag{1.2.6.1.1}$$

```
> with(Statistics):
  Q := RandomVariable(ChiSquare(3));
```

$$Q := \_R \tag{1.2.6.1.2}$$

```
> Probability(Q > 1/sigma^2);
```

$$0.261464447426406 \tag{1.2.6.1.3}$$

Tehát az esetek kb. $\frac{1}{4}$ részében téved a helymeghatározó rendszer több, mint 1 métert.
Ábrázoljuk is a kapott valószínûséghez tartozó sûrûségfüggvény alatti területet! Ehhez elôbb az $\mathcal{E}$ változó eloszlás- és sûrûségfüggvényét kell meghatároznunk.

```
> F_Q := x -> CDF(Q, x);
```

$$F\_Q := x \rightarrow Statistics:\text{-}CDF(Q, x) \tag{1.2.6.1.4}$$

$F_{\mathcal{E}}(x) = P(\mathcal{E} < x) = P(\sigma \cdot \sqrt{Q} < x) = P\left(Q < \frac{x^2}{\sigma^2}\right) = F_Q\left(\frac{x^2}{\sigma^2}\right)$, ha $x > 0$, és $F_{\mathcal{E}}(x) = 0$ különben.

```
> F_E := x -> piecewise(x <= 0, 0, F_Q(x^2/sigma^2));
```

(1.2.6.1.5)

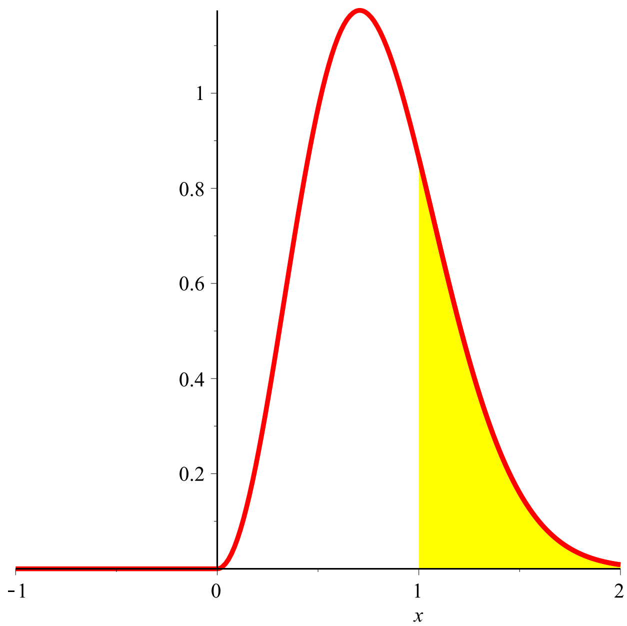
$$F\_E := x \rightarrow piecewise\left(x \le 0, 0, F\_Q\left(\frac{x^2}{\sigma^2}\right)\right) \quad \textbf{(1.2.6.1.5)}$$

A sûrûségfüggvény ennek deriváltja:

```
> f_E := x -> diff(F_E(x), x);
```

$$f\_E := x \rightarrow \frac{d}{dx} F\_E(x) \quad \textbf{(1.2.6.1.6)}$$

```
> P1 := plot(f_E(x), x = -1..2, color = red, thickness = 3):
  P2 := plot(f_E(x), x = 1..2, color = red, thickness = 3, filled = [color = yellow, transparency = 0.5]):
  plots[display](P1, P2);
```



Gyakorló feladatok

Gy/1. Feladat

Egy gyümölcsárusnál 5 almát vásárolunk. Egy alma tömege átlagosan 176 gramm, 50 gramm szórással és feltételezhetô hogy normál eloszlást követ.
a) Mekkora az esélye, hogy a vásárolt almák össztömege kevesebb, mint 650 gramm?

b) Mekkora az esélye, hogy az almák össztömege 800 és 900 gramm között van?
c) Egy almás süteményhez 1 kg alma kell. Hány almát kellene vennünk, hogy 95 % eséllyel el tudjuk készíteni a süteményt (ragaszkodunk a recept szerinti 1 kg-hoz)?

Megoldás

```
> restart;
```

Gy/2. Feladat

Egy elektromos készülék élettartamát 10 év várható értékû exponenciális eloszlású valószínûségi változó írja le.
a) Ábrázoljuk az élettartam eloszlás és sûrûségfüggvényét, és adjuk meg a szórását!
b) Melyiknek nagyobb az esélye: annak, hogy a készülék 5 éven belül elromlik, vagy annak, hogy 15 évig mûködôképes marad?
c) Hány év garanciát vállaljon a gyártó, ha azt szeretné, hogy termékeinek legfeljebb 10 %-át kelljen garanciaidôn belül javítani?

Megoldás

```
> restart;
```

Gy/3. Feladat

Egy adott népességben a felnôttek várható testmagassága $\mu = 170$ cm, $\sigma = 8$ cm szórással. Jelölje X 50 véletlenszerûen kiválasztott felnôtt átlagos magasságát! Számítsuk ki az alábbi valószínûségeket és ábrázoljuk azokat a sûrûségfüggvény alatti megfelelô tartomány színezésével!
a) $P(X > 169)$;
b) $P(X \leq 172)$;
c) $P(|X - \mu| > 1.5)$.

Megoldás

```
> restart;
```

Gy/4. Feladat

Egy síközpontban párhuzamosan több sífelvonó is üzemel. Mindig akkor indítják a következô kabint, ha pont 10 síelô összegyûlt.
a) Feltételezve, hogy a síelôk Poisson-modell szerint érkeznek, percenként átlagosan 4-en, írjuk fel a két kabin indulása között eltelt T idô eloszlását valamint ábrázoljuk eloszlás- és sûrûségfüggvényét!
b) Határozzuk meg T várható értékét és szórását!
c) Mekkora a valószínûsége, hogy 3 percnél többet kell várni a következô kabin indulására?
d) Közelítsük T-t egy normál eloszlású valószínûségi változóval és hasonlítsuk össze a két sûrûségfüggvényt! Mi lenne a válasz a c) részre, ha a normál közelítéssel számolnánk?

Megoldás

```
> restart;
```