Poslední úprava dokumentu: 18. března 2025.

Pravděpodobnostní rozdělení, centrální limitní věta

Spusťte RStudio, otevřete si nový skript, uložte si ho a nastavte pracovní adresář.

Prostudujte si prosím text *Náhodná veličina a její rozdělení* (dále pod zkratkou *NVR*), který najdete na mých stránkách mezi materiály ke cvičení z 19.3.2025, popřípadě si příslušné pojmy připomeňte z přednášky či z učebnice.

1 Binomické rozdělení

Připomeňte si význam veličiny s binomickým rozdělením (např. *NVR* str. 7) a základní charakteristiky tohoto rozdělení.

Příklad - ženy vs. muži

Pravděpodobnost, že náhodně vybraný jedinec v populaci je žena, je 0.55. Jaká je pravděpodobnost, že mezi 3 náhodně vybranými jedinci bude právě 1 žena?

1) První možnost je vyřešit příklad pomocí pravidla o násobení pravděpodobností pro nezávislé jevy (viz vzorec (3) v pracovním listu 07):

- Pravděpodobnost, že mezi třemi po sobě vybranými jedinci je na prvním místě žena, je:

  $$P(\text{na 1. místě je žena \& na 2. je muž \& na 3. je muž}) = 0.55 \times 0.45 \times 0.45 \doteq 0.11$$

  (Znak & čtěte jako „a současně“ a značí průnik příslušných náhodných jevů. Tyto jevy jsou navíc nezávislé, neboť jedince považujeme za nezávislé (viz poznámka níže). Na přednášce jste si průnik značili jako $\cap$, ale & mi zde přijde názornější.)
- Stejný výsledek dostaneme i pro pravděpodobnost, že žena je na 2. nebo 3. místě. Celkem tedy máme:

  $$P(\text{mezi 3 vybranými jedinci je právě jedna žena}) = 3 \times 0.55 \times 0.45 \times 0.45 \doteq 0.33$$

- Při výpočtu tiše předpokládáme, že populace je dostatečně velká, tj. vybrání jednoho muže či ženy nám nijak neovlivní pravděpodobnost vybrání může či ženy v dalším kroku.
- Navíc předpokládáme, že jedinci jsou navzájem nezávislí, tj. to, že jsme na prvním místě vybrali jedince daného pohlaví, nijak neovlivní pohlaví jedinců na dalších místech.

2) Druhou možností je si uvědomit, že *počet žen mezi 3 jedinci* je veličina s binomickým rozdělením s parametry $n = 3$ a $p = 0.55$. Označíme-li si tuto veličinu písmenem X, lze to symbolicky zapsat jako $X \sim Bi(3, 0.55)$.

- Pravděpodobnost, že mezi 3 lidmi je právě jedna žena, je pak vlastně dotaz na pravděpodobnost konkrétní hodnoty binomického rozdělení. Z teorie víme, že

  $$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1}$$

✧ Pro nás tedy

$$P(X = 1) = \binom{3}{1}(0.55)^1(1 - 0.55)^{3-1} = 3 \times 0.55 \times (0.45)^2 \doteq 0.33.$$

✧ V R to můžeme jednoduše vypočíst pomocí příkazu

```
dbinom(1, 3, 0.55)
```

Příklad s krajtou

U krajt je úspěšnost líhnutí kolem 70 % až 90 %, pokud jsou vajíčka inkubována v ideálních podmínkách. Předpokládejme, že se z vajíčka krajty vylíhne živý jedinec s pravděpodobností 0.7. V následujících úkolech nás bude zajímat pravděpodobnost, s jakou se z 10 vajíček vylíhne určitý počet živých krajt. Veličinou, která nás zajímá je tedy počet úspěchů (= živých krajt) z 10 pokusů (= vajíček). To je přesně veličina s binomickým rozdělením a parametry $n = 10$ a $p = 0.7$. Označíme-li si tuto veličinu jako X, pak můžeme symbolicky psát $X \sim Bi(10, 0.7)$.

1) S jakou pravděpodobností se z 10 vajíček vylíhne **právě** 5 živých krajt?

✧ Zajímá nás $P(X = 5)$. Dle vzorečku (1) víme, že je to

$$P(X = 5) = \binom{10}{5}(0.7)^5(1 - 0.7)^{(10-5)} = \binom{10}{5}(0.7)^5(0.3)^5,$$

✧ což v R můžeme jednoduše spočítat pomocí

```
dbinom(5, 10, 0.7)
```

✧ Odkud se vzal název příkazu? Zkratka `binom` v příkazu samozřejmě odkazuje na binomické rozdělení, předpona `d` pak zkracuje *density* (hustota), neboť pravděpodobnosti $P(X = k)$ jsou vlastně analogií hustoty u spojitých rozdělení.

2) S jakou pravděpodobností se z 10 vajíček vylíhne **nejvýše** 5 živých krajt?

✧ Tady je dotaz na $P(X \leq 5)$, což je hodnota distribuční funkce binomického rozdělení v bodě 5. Z teorie víme, že

$$P(X \leq 5) = \sum_{k=0}^{5} P(X = k) = \sum_{k=0}^{5} \binom{10}{k}(0.7)^k(1 - 0.7)^{(10-k)},$$

✧ což v R vypočteme jako

```
pbinom(5, 10, 0.7)
```

(Předpona `p` v názvu příkazu odkazuje na *probability*, ale nedá se říct, že by toto označení bylo nějak zvlášť šťastné...)

3) S jakou pravděpodobností se z 10 vajíček vylíhne **alespoň** 5 živých krajt?

✧ Zde je dotaz na $P(X \geq 5)$, což můžeme jednoduše převést na předchozí případ pomocí

$$P(X \geq 5) = 1 - P(X \leq 4).$$

✧ V R to pak vypočteme jako

```
1 - pbinom(4, 10, 0.7)
```

4) Všechny pravděpodobnosti $P(X = k)$ pro $k = 0, 1, \ldots, 10$ lze vypsat příkazem:

```
dbinom(0:10, 10, 0.7)
```

Příklad - kostka

Házíme 20-krát symetrickou kostkou. Zajímá nás jaká je pravděpodobnost, že z těchto 20 hodů padně určitý počet šestek. Rozmyslete si, že veličina X představující počet šestek ve 20 hodech má binomické rozdělení $Bi(20, \frac{1}{6})$.

- S jakou pravděpodobností padnou právě 4 šestky?
- S jakou pravděpodobností padnou nejvýše 4 šestky?
- S jakou pravděpodobností padne alespoň 5 šestek?

2 Poissonovo rozdělení

V textu *NVR*, str. 8, si přečtěte sekci o Poissonovu rozdělení a o možnosti jeho využití k aproximaci binomického rozdělení.

Pixely

Z deseti miliónů pixelů obrazovky jsou v průměru dva vadné. Uvažujte obrazovku o rozměru 1280×1024 pixelů.

1) Porovnejte pravděpodobnosti hodnot 0,1,...,5, když uvažujete binomické rozdělení a jeho aproximaci rozdělením Poissonovým.

```
n <- 1280*1024
p <- 2/10000000
cbind("k"=0:5,"binom."=dbinom(0:5,n,p),"Pois"=dpois(0:5,n*p))
```

2) Jaký je očekávaný počet vadných pixelů na obrazovce?

3) Nechte si vykreslit pravděpodobnosti jednotlivých hodnot. Která hodnota je nejpravděpodobnější?

```
plot(0:5,dbinom(0:5,n,p),col="red",pch=16,xlab="k",
ylim=c(0,max(dbinom(0:5,n,p))),ylab="dbinom(k,n,p)")
points(0:5,dpois(0:5,n*p),pch=3,col="blue")
legend("topright",legend=c("binomické","Poissonovo")
```

Geigerův-Müllerův čítač

Geigerův-Müllerův čítač je detektor ionizujícího záření, který měří počet částic daného záření zachycených za daný časový interval. Při pokusu se $^{90}\mathrm{Sr}$ byly během desetisekundových intervalů naměřeny tyto četnosti, které naleznete v souboru `GM_counter_data.txt` ve složce `V:/turcicova/MatStat_cv/data`.

```
GMdata <- read.table("GM_counter_data.txt", header=FALSE)
GMdata <- GMdata[ ,1]
```

1) Jakým rozdělením lze modelovat počet částic detekovaných během desetisekundového intervalu?
Jelikož se jedná o počet událostí za krátký časový interval, bude nejvhodnějším modelem Poissonovo rozdělení $Po(\lambda)$ pro nějaký parametr λ.

2) Odhadněte parametr tohoto rozdělení.
Všimneme si, že střední hodnota Poissonova rozdělení je rovna přímo jeho parametru, tj. že $EX = \lambda$. Střední hodnotu EX přitom můžeme odhadnout pomocí výběrového průměru z našich dat. Tento výběrový průměr bude tedy zároveň odhadem parametru λ.

```
odhad.lambda <- mean(GMdata)
```

3) Za pomoci tohoto odhadu spočtěte pravděpodobnost, že bychom během desetisekundového intervalu detekovali více než 190 částic.

```
1-ppois(190, odhad.lambda)
```

3 Exponenciální rozdělení

Doba čekání na autobus

Doba čekání na autobus (v min) se řídí exponenciálním rozdělením s parametrem 1/5.

1) Nakreslete si graf distribuční funkce. Připomeňte si základní vlastnosti distribučních funkcí.

```
curve(pexp(x,1/5))
```

2) Vykreslete si graf hustoty daného rozdělení. Jaké jsou vlastnosti hustoty?

```
curve(dexp(x,1/5))
```

3) Jaká je pravděpodobnost, že budeme čekat méně než dvě minuty?

```
pexp(2,1/5)
```

4) Jaká je pravděpodobnost, že budeme čekat déle než 5 minut?

```
1-pexp(5, 1/5)
```

5) Jaká je pravděpodobnost, že čekáním strávíme dobu z intervalu 1 až 5 minut?

```
pexp(5,1/5) - pexp(1,1/5)
```

6) Zjistěte medián doby strávené čekáním na autobus.

```
qexp(0.5, 1/5)
```

7) Jaká je očekávaná doba strávená čekáním?
$EX = 1/\lambda = 5$

4 Normální rozdělení

V textu *NVR*, str. 10, si přečtěte sekci o normálním rozdělení a jeho charakteristikách.

Na normálním rozdělení je mimo jiné zajímavé i to, že má zvlášť parametr pro střední hodnotu a zvlášť parametr pro rozptyl. Jeho střední hodnota a rozptyl spolu tedy nejsou žádným způsobem provázané. To u jiných rozdělení nenajdeme.

Vykreslete si „gaussovku“

Hustota normálního rozdělení (slangově nazývaná „gaussovka“) je funkce daná předpisem

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

kde μ a σ^2 jsou střední hodnota a rozptyl, kteréžto jsou parametry normálního rozdělení.

Vykreslete si „gaussovku“ pro různé hodnoty střední hodnoty a rozptylu (resp. směrodatné odchylky) a pozorujte vliv těchto parametrů na její tvar.

```
strhod = 1        # stredni hodnota, zkousejte ruzne hodnoty
smerodch = 3       # smerodatna odchylka, zkousejte ruzne hodnoty
curve(dnorm(x,strhod,smerodch),xlim=c(-30,30))
```

✧ **Pro zájemce:** Příkaz `dnorm` poskytuje hodnotu hustoty normálního rozdělení v daném bodě s danými parametry. Předpona `d` je zkrácením slova *density* (hustota). Prvním argumentem funkce `dnorm` je bod, ve kterém nás hustota zajímá (zde x, neboť nás to zajímá ve všech bodech a funkce `curve` přijímá pouze funkce proměnné x). Druhým argumentem je střední hodnota a třetím argumentem je směrodatná odchylka.

✧ Příkaz `curve` dokáže vykreslit libovolnou křivku proměnné x. Jeho argument `xlim` pak udává limity na ose x, v nichž má být křivka vykreslena.

✧ Jak jste zpozorovali, střední hodnota normálního rozdělení udává střed hustoty (tj. střed „kopečku“). Rozptyl pak udává, jak bude křivka „roztažená“ do stran. Možná jste si také všimli, že při větší hodnotě rozptylu se výška kopečku sníží (hodnoty na ose y klesnou). Je to proto, že celková plocha pod hustotou musí být vždy rovna jedné, a tudíž je nutno tuto plochu i při větším rozptylu (a tedy „roztaženější“ křivce) zachovat.

Příklad - IQ

Hodnota IQ koeficientu je tabelována tak, aby bylo jeho rozdělení normální se střední hodnotou 100 a směrodatnou odchylkou 15 (tímto tiše předpokládáme, že každý člověk je charakterizován hodnotou IQ určenou s neomezenou přesností).

Označme si veličinu představující IQ náhodně vybraného člověka jako X. Pak $X \sim N(100, 15^2)$. (Druhým parametrem normálního rozdělení je tradičně rozptyl, nikoli směrodatná odchylka, proto musíme 15 umocnit na druhou).

1) Určeme pravděpodobnost, že IQ náhodně vybraného člověka je nižší než 120.

✧ Zajímá nás tedy $P(X \leq 120)$, což je hodnota distribuční funkce normálního rozdělení $N(100, 15^2)$ v bodě 120. Žádný explicitní vzorec pro distribuční funkci normálního rozdělení bohužel neexistuje, lze ji zapsat pouze pomocí integrálu z hustoty, tj.

$$P(X \leq z) = \int_{-\infty}^{z} f(x)dx = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx,$$

kde $f(x)$ je hustota normálního rozdělení $N(\mu, \sigma^2)$, kterou jsme již viděli ve vzorci (2). Pro nás tedy

$$P(X \leq 120) = \int_{-\infty}^{120} \frac{1}{\sqrt{2\pi 15^2}} e^{-\frac{(x-100)^2}{2\cdot 15^2}} dx.$$

✧ Naštěstí tento integrál nemusíme počítat ručně, ale spočítá ho za nás R:

```
pnorm(120, 100, 15)
```

2) Určeme pravděpodobnost, že IQ náhodně vybraného člověka je vyšší než 120.

- Zde je dotaz na $P(X > 120)$, což můžeme opět spočítat jako:

$$P(X > 120) = 1 - P(X \leq 120),$$

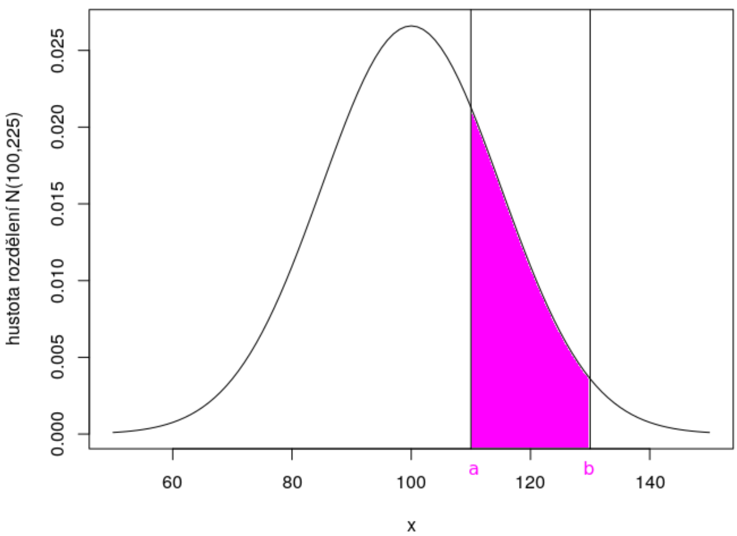
- což v R spočteme jako

```
1 - pnorm(120, 100, 15)
```

3) Určete pravděpodobnost, že IQ náhodně vybraného člověka leží v intervalu (110, 130).

- Pravděpodobnost, že veličina X nabyde hodnoty z nějakého intervalu (a, b) je rovna ploše po křivkou hustoty v tomto intervalu. A tato plocha se vypočte jakožto určitý integrál z hustoty $f(x)$ přes interval (a, b):

$$P(X \in (a, b)) = P(a < X < b) = \int_a^b f(x)dx.$$



- Místo počítání integrálu si ale můžeme rozmyslet, že

$$P(a < X < b) = P(X < b) - P(X < a),$$

tedy jde o rozdíl distribuční funkce v bodě b a v bodě a.

- V našem případě je $a = 110$ a $b = 130$. Rozdíl distribučních funkcí pak v R spočítáme jako

```
pnorm(130, 100, 15) - pnorm(110, 100, 15)
```

- Možná vás zarazilo, že distribuční funkcí zde nazýváme pravděpodobnost $P(X < z)$, místo tradičního $P(X \leq z)$. Pravda je, že u spojitých rozdělení je to jedno. Je to díky tomu, že u spojitých rozdělení je pravděpodobnost jednoho konkrétního bodu rovna nule, tj. $P(X = z) = 0$. A tudíž $P(X \leq z) = P(X = z) + P(X < z) = P(X < z)$. Podrobněji se o tom můžete dočíst níže.

4) Určete pravděpodobnost, že IQ náhodně vybraného člověka je rovno 120.

✧ Správná odpověď je

$$P(X = 120) = 0.$$

✧ Někdo bude možná zkoušet: `dnorm(120, 100, 15)`
ALE POZOR! Pro spojitá rozdělení je $\mathrm{P}(X = x) = 0$! Hustota je pro spojitá rozdělení sice jakousi „náhražkou“ za výraz $\mathrm{P}(X = x)$ (který se vyskytuje u rozdělení diskrétních), avšak má význam pouze při určení pravděpodobnosti, že náhodná veličina X leží v nějakém intervalu (tato pravděpodobnost je pak rovna ploše pod křivkou hustoty v daném intervalu). Hustotu tedy nelze použít pro výpočet pravděpodobnosti, že se X rovná určité hodnotě, tj. $f(x) \neq P(X = x)$. Příkaz dnorm(120,100,15) nám sice dá nějaké číslo (neboť hustota $f(x)$ je funkcí proměnné x, takže $f(120)$ je „nějaké číslo“), avšak toto číslo nemá význam pravděpodobnosti jevu „IQ=120“, tj. $f(120) \neq P(\mathrm{IQ} = 120)$.

5 Maxwellovo rozdělení

Hustota Maxwellova rozdělení s parametrem $a > 0$ má tvar

$$\begin{aligned} f(x) &= \frac{2}{a^3\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2a^2}}, \quad \text{pro } x \geq 0, \\ &= 0, \quad \text{pro } x < 0. \end{aligned} \tag{3}$$

Žádný explicitní vzorec pro distribuční funkci Maxwellova rozdělení bohužel neexistuje, lze ji zapsat pouze pomocí integrálu z hustoty, tj.

$$P(X \leq z) = \int_{-\infty}^{z} f(x)dx = \int_{-\infty}^{z} \frac{2}{a^3\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2a^2}} dx,$$

a tudíž medián i kvantily lze spočíst jen numericky. Střední hodnota je rovna

$$\mathrm{E}\, X = \sqrt{\frac{8}{\pi}} a.$$

Někdy se toto rozdělení nazývá též Maxwell-Boltzmannovo.

1) **Příklad - ideální plyn:** Při snaze o fyzikální popis souboru mnoha částic se ukázalo, že rozdělení jejich rychlostí v ideálním plynu o teplotě T lze velmi dobře popsat funkcí:

$$g(v) = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-\frac{mv^2}{2kT}} v^2, \tag{4}$$

což je hustota Maxwellova rozdělení výše s parametrem $a = \sqrt{kT/m}$, kde T značí teplotu v Kelvinech, m hmotnost částice v kilogramech a k je Boltzmannova konstanta.

Předpokládejme, že máme nádobu s 10 moly CO_2 o teplotě 300 K. V případě CO_2 je hodnota parametru a rovna

$$a = \sqrt{kT/m} \doteq \sqrt{189T}.$$

✧ Vykreslete si hustotu rozdělní (4) pro různé hodnoty teploty T z intervalu 200 až 500 K.

```
install.packages("shotGroups")
library(shotGroups)
T1 <- 200;   a <- sqrt(189*T1)
curve(dMaxwell(x,a),from = 0, to = 1000, xlab="v [m/s]", ylab= "f(v)",
                          ylim = c(0,0.003), main = "Maxwellovo rozdělení")
T2 <- 500;   a <- sqrt(189*T2)
curve(dMaxwell(x,a),col = 3, add= TRUE)
legend("topright", legend=c(paste(T1,"K"),paste(T2,"K")),col=c(1,3),lty=c(1,1))
```

Nejspíš jste zpozorovali, že při vyšší teplotě se výška kopečku sníží (hodnoty na ose y klesnou). Je to proto, že celková plocha pod hustotou musí být vždy rovna jedné, a tudíž je nutno tuto plochu i při větším „roztažení“ křivky zachovat. To tedy znamená, že při vyšší teplotě je menší pravděpodobnost, že částice bude cestovat menší rychlostí a naopak jsou upřednostňovány vyšší rychlosti. Při poklesu teploty je naopak velmi pravděpodobné, že částice bude cestovat nízkou rychlostí.

✧ Spočtěte nejpravděpodobnější rychlost v_p a zakreslete si ji do grafu hustoty.

Nejpravděpodobnější rychlost je ta, kde funkce $g(x)$ nabývá maxima, to najdeme jednoduše z podmínky $g'(v) = 0$. Derivace funkce g se rovná

$$g'(v) = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-\frac{mv^2}{2kT}} \left[2v - \frac{mv^3}{kT}\right].$$

Položíme-li tento výraz roven 0, zjistíme, že

$$v_p = \sqrt{\frac{2kT}{m}} = \sqrt{2}a.$$

V případě CO_2 o teplotě 300 K je v_p rovna 336.73 m/s.

```
T <- 300
a <- sqrt(189*T)
curve(dMaxwell(x,a),from = 0, to = 800, xlab="v [m/s]", ylab= "f(v)",
main = paste(T,"K"), xaxt="n")
axis(side=1, at=c(0,200,600,800), labels = TRUE)
vp <- sqrt(2)*a
abline(v=vp, lty=2)
axis(side=1, at=vp, labels = expression("v"[p]))
```

✧ Spočtěte střední rychlost $\mathrm{E}\,v$ a přidejte ji do grafu.

Střední rychlost bychom spočítali jako

$$\mathrm{E}\,v = \int_0^\infty vg(v)dv = ... = \sqrt{\frac{8}{\pi}}a = \sqrt{\frac{8kT}{\pi m}} = \frac{2}{\sqrt{\pi}}v_p.$$

V případě CO_2 o teplotě 300 K je $\mathrm{E}\,v$ rovna 379.96 m/s.

Jelikož $\mathrm{E}\,v > v_p$, tak vidíme, že ačkoli se většina částic pohybuje rychlostí v_p, je tam mnoho částic které se pohybují výrazně rychleji.

```
Ev <- 2/sqrt(pi)*vp
abline(v=Ev, lty=3)
axis(side=1, at=Ev, labels = "E(v)")
```

✧ Spočtěte střední kvadratickou rychlost $\mathrm{E}\,v^2$ a přidejte její odmocninu do grafu.

Tato rychlost vlastně představuje 2. moment Maxwellova rozdělení a spočteme ji (při integrování by se využily polární souřadnice) jako

$$\mathrm{E}\,v^2 = \int_0^\infty vg(v)dv = ... = \frac{3kT}{m}.$$

Tato rychlost je v praxi velmi důležitá, neboť pomocí ní lze určit střední kinetickou energii částice (molekuly): $\mathrm{E}(E_k) = \frac{1}{2}m\,\mathrm{E}(v^2)$. Aby byla tato rychlost porovnatelná se střední a nejpravděpodobnější rychlostí, zavádí se její odmocnina (*root-mean-square velocity*)

$$v_{rms} = \sqrt{\frac{3kT}{m}} = \sqrt{\frac{3}{2}}v_p.$$

V případě CO_2 o teplotě 300 K je v_{rms} rovna 412.41 m/s.

```
v_rms <- sqrt(3/2)*vp
abline(v=v_rms, lty=4)
axis(side=1, at=v_rms, labels = expression("v"[rms]))
```

Poznámka: Všimněte si, že poměry rychlostí jsou

$$\frac{v_{rms}}{v_p} = \sqrt{\frac{3}{2}}, \quad \frac{v_{rms}}{\mathrm{E}\, v} = \sqrt{\frac{3\pi}{8}},$$

a tedy jsou konstantní pro každý plyn bez ohledu na teplotu!

✧ Jaká je pravděpodobnost nalezení molekuly CO_2, která se pohybuje rychlostí mezi 500 a 520 m/s? Jaký počet částic v naší nádobě se touto rychlostí pohybuje?

Pravděpodobnost, že rychlost částice nabyde hodnoty z intervalu $(500, 520)$ m/s je rovna ploše pod křivkou hustoty v tomto intervalu, tj.

$$P(v \in (500, 520)) = \int_{500}^{520} g(v)dv = P(v < 520) - P(v < 500),$$

což lze v R snadno spočítat pomocí příkazu

```
a <- sqrt(189*300)
pMaxwell(520,a) - pMaxwell(500,a)
```

Výsledek jsou asi 3 %. Pokud bychom spočítali pravděpodobnost, že částice cestuje rychlostí $(v_p - 10, v_p + 10)$, dostali bychom hodnotu skoro 5 %. To, že je tato pravděpodobnost větší, je patrné už z tvaru hustoty rozdělení.

Abychom zjistili počet molekul v naší nádobě, které cestují rychlostí 500–520 m/s, musíme nejprve zjistit celkový počet částic v nádobě. Ten je roven

$$N = n \cdot N_A = 10 \cdot 6.023 \cdot 10^{23} = 6.023 \cdot 10^{24},$$

a tudíž počet molekul pohybujících se rychlostí v rozmezí 500–520 m/s je $0.03 \cdot N \doteq 1.8 \cdot 10^{23}$.

6 Centrální limitní věta

Teorii k centrální limitní větě si můžete prostudovat v textu *CLV.pdf*, který najdete na mých stránkách.

Dále si stáhněte skript „ukazkaCLT.R“ a uložte si ho do svého pracovního adresáře. Skript si načtěte do R pomocí příkazu:

```
source("ukazkaCLT.R")
```

Nejprve si prohlédněte, jak vypadají hustoty některých spojitých rozdělení, o kterých jste slyšeli na přednášce.

```
ukazkaCLT(n=0)
```

Jako ilustraci centrální limitní věty si prohlédněte, jak se chová rozdělení výběrových průměrů z výše uvedených rozdělení, měníme-li rozsah výběru n.

```
ukazkaCLT(n=1)
```

✧ Volte postupně n: 1, 2, 5, 20, 100

✧ Povšimněte si, že měřítka na y-ové ose pro jednotlivá rozdělení jsou různá.

✧ Důvodem je fakt, že limitní (normální) rozdělení mají různé parametry pro jednotlivá rozdělení.

Galtonova deska

https://www.youtube.com/watch?v=EvHiee7gs9Y

Příklad se zaokrouhlováním čísel

Bylo sečteno 300 čísel zaokrouhlených na jedno desetinné místo. Vyšetříme chybu tohoto součtu vzniklou zaokrouhlováním sčítanců. Předpokládejme, že zaokrouhlovací chyby (označme si je $X_1, X_2, \ldots, X_{300}$) jednotlivých sčítanců jsou nezávislé náhodné veličiny s rovnoměrným rozdělením na intervalu (-0.05, 0.05).

1) Označme si jako X_k chybu k-tého sčítance. Víme, že $X_k \sim R(-0.05, 0.05)$. Pak nutně platí, že $EX_k = 0$ a $var(X_k) = 1/1200$ (připomeňte si vzoreček pro střední hdonotu a rozptyl rovnoměrného rozdělení).

2) Chyba součtu je náhodná veličina $Y = \sum_{k=1}^{300} X_k$.

3) Normovaná chyba je náhodná veličina

$$Z = \frac{Y}{\sqrt{300/1200}} = 2Y$$

a dle Centrální limitní věty má tato veličina rozdělení $N(0, 1)$.

4) Pravděpodobnost, že chyba součtu nepřekročí v absolutní hodnotě dané číslo $\varepsilon > 0$ je

$$P(|Y| \leq \varepsilon) = P(|Z| \leq 2\varepsilon) = P(-2\varepsilon \leq Z \leq 2\varepsilon) = \Phi(2\varepsilon) - (1 - \Phi(-2\varepsilon)) = 2\Phi(2\varepsilon) - 1,$$

kde Φ značí distribuční funkci rozdělení $N(0, 1)$. Např. pro $\varepsilon = 1$ tato pravděpodobnost rovna 0.9545.