Náhodná veličina

Výsledek náhodného pokusu, daný reálným číslem je hodnotou náhodné veličiny.

Náhodná veličina je libovolná reálná funkce *X* definovaná na množině elementárních E pravděpodobnostního prostoru S.

- **Diskrétní náhodná veličina**
  může nabývat pouze spočetně mnoha hodnot
  (počet aut v náhodně vybraná domácnost, výsledek hodu kostkou)
- **Spojitá náhodná veličina**
  může nabývat všech hodnot z nějakého intervalu
  (doba bezporuchového chodu zařízení, výška náhodně vybraného člověka)

Náhodná veličina

- Proměnná, jejíž hodnota je určena výsledkem náhodného pokusu.
- Každému el. jevu *E* z prostoru všech jevů S přiřadíme reálné číslo X(E), takové, že pro každé reálné číslo *a* je jevem i množina

$$A = \{E; X(E) \leq a\}$$

- **Diskrétní náhodná veličina** (množina hodnot je konečná, nebo spočetná) je popsána pravděpodobnostní funkcí $P(a)=P(X(E)=a)$, nebo diskrétní distribuční funkcí
- **Spojitá náhodná veličina** (množina hodnot je interval $I \subset R$)
  – distribuční funkce $F(x)$

$$F(x) = P(X \leq x)$$

  – hustota pravděpodobnosti $f(x)$

$$F(x) = \int_{-\infty}^{x} f(u)du \qquad F(b) - F(a) = \int_{a}^{b} f(u)du$$

Střední hodnota a rozptyl

- Střední hodnota
  - diskrétní náhodné veličiny $$E[x] = \sum_i x_i P(X = x_i)$$
  - spojité náhodné veličiny $$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

vlastnosti střední hodnoty:
E[c.X] = c.E[X]
E[X+Y] = E[X] + E[Y]
E[X .Y] = E[X] . E[Y] pro X, Y nezávislé

- Rozptyl
  - diskrétní náhodné veličiny $$V[x] = \sum_i \left(x_i - E[X]\right)^2 P(X = x_i)$$
  - spojité náhodné veličiny $$V[X] = \int_{-\infty}^{\infty} \left(x - E(X)\right)^2 \cdot f(x) dx$$

Charakteristiky polohy a variability

Cíl je jedním číslem charakterizovat velikost všech číselných hodnot ve statistickém souboru.

Charakteristiky polohy nám umožňují srovnávat úroveň zkoumaného jevu u dvou nebo více souborů.

- **aritmetický průměr**
  `mean(X)`, `trimmean(X,25)*`
- **medián**
  `median(X)`
- **modus**
  `mode(X)`
- **Směrodatná odchylka**
  `std(X)`
- **Rozptyl**
  `var(X)`
- **Kvartilové rozpětí**
  `iqr(X)*`

`*Statistical toolbox`

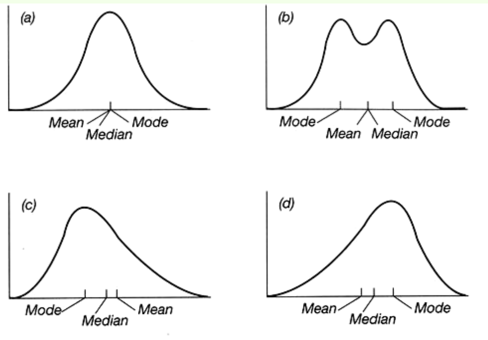


Figure 3.2 Frequency distributions showing measures of central tendency. Values of the variable are along the abscissa (horizontal axis), and the frequencies are along the ordinate (vertical axis). Distributions (a) and (b) are symmetrical, (c) is positively skewed, and (d) is negatively skewed. Distributions (a), (c), and (d) are unimodal, and distribution (b) is bimodal. In a unimodal asymmetric distribution, the median lies about one-third the distance between the mean and the mode.*

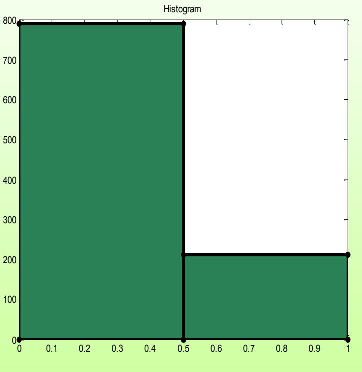
Statistical toolbox

```
pdf ('jméno', data, param)
cdf ('jméno', data, param)
random('jméno', data, param)
```

- Diskrétní rozdělení pravděpodobnosti
  - Alternativní
  - Binomické 'Binomial' n,p
  - Geometrické 'Geometric' p
  - Poissonovo 'Poisson' λ
- Spojité rozdělení pravděpodobnosti
  - Rovnoměrné 'Uniform' a, b
  - Normální 'Normal' μ, σ
  - Exponenciální 'Exponential' λ

Alternativní rozdělení

- http://home.zcu.cz/~friesl/hpsb/alt.html
- náhodná proměnná,
  s pstí *p* nabývá hodnoty 1
  a s pstí *1-p* nabývá hodnoty 0
  - `Alt=rand(n,1) < p;`
  - `hist(Alt,2)`
  - `p_est=sum(Alt)/n`



Geometrické rozdělení diskrétní náhodné veličiny

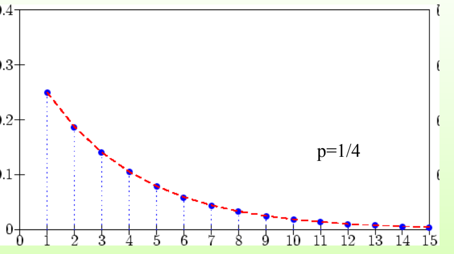
Počet pokusů do prvního úspěšného výsledku.

Experimenty jsou nezávislé, pravděpodobnost úspěchu je dána *p*.

$$p_i = P(X = i) = p\ (1-p)^{i-1}$$

$$E[X] = \frac{1}{p}$$

$$V[X] = \frac{1-p}{p^2}$$



Příklad: Zařízení kontrolujeme pravidelně jednou za hodinu. Pravděpodobnost, že se za hodinu zařízení nepokazí je 0,9. Určete pravděpodobnost, že k chybě zařízení dojde při šesté kontrole. Určete průměrnou dobu bezvadného chodu.

$$p = 0{,}1$$

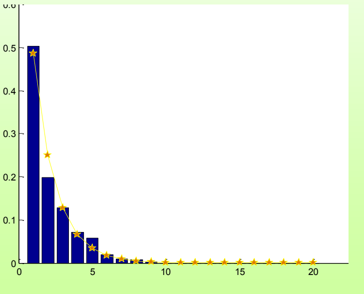
$$P(X = 6) = 0{,}1 \cdot (0{,}9)^5 = 0{,}059$$

$$E[X] = \frac{1}{0.1} = 10\ (hod)$$

Geometrické rozdělení

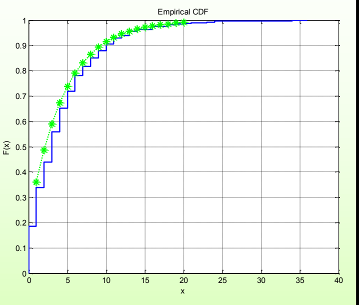
```
- y=zeros(n,1);
- for i = 1:n
-     while rand > p %failure
-     y(i)=y(i)+1;
-     end
-     y(i)=y(i)+1; %success
- end
```

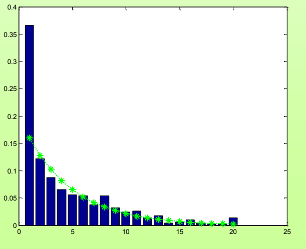
```
nbins=1:20;
freq=hist(y,nbins);
bar(nbins,freq/n)
```



Geometrické rozdělení – Statistical Toolbox

```
p=0.2;
y=random('Geometric',p,500,1);
cdfplot(y)
hold on
plot(1:20,cdf('Geometric',1:20,p),'g*')
hold off

%%%%%%%%%%%%%%%%%%%
nbins=1:20;
freq=hist(y,nbins);
bar(nbins,freq/500)
hold on
plot(nbins,pdf('Geometric',nbins,p),'g*')
hold off
```



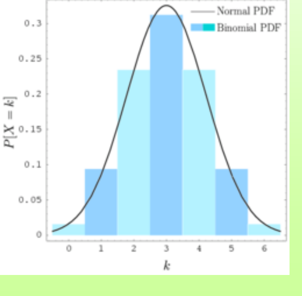


Binomické rozdělení diskrétní náhodné veličiny

- Nechť i je počet úspěšných výsledků z *n* provedených nezávislých experimentů, kde pravděpodobnost úspěšného výsledku je dána *p*, potom

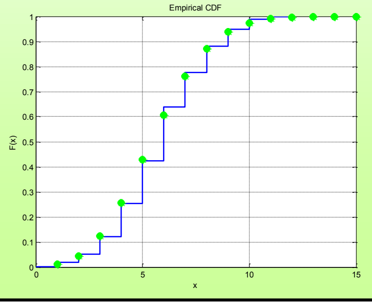
$$p_i = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad E[X] = n \cdot p \qquad V[X] = np(1-p)$$

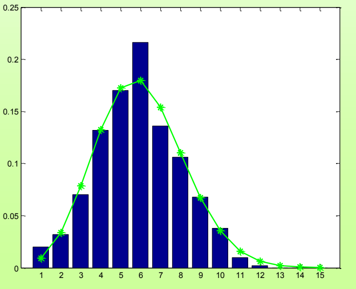
```
p=1/5;n=30;
ns=500;
for i = 1:ns
    y(i)= sum(rand(n,1) < p);
end

fprintf('Mean = %2.2f,\v',mean(y))
```



Binomické rozdělení – statistical toolbox

```
p=0.2; n=30;
y=random('Binomial',n,p,500,1);
cdfplot(y)
hold on
plot(1:15,cdf('Binomial',1:15,n,p),'g*')
hold off
```





Binomické rozdělení diskrétní náhodné veličiny

- Nechť i je počet úspěšných výsledků z *n* provedených nezávislých experimentů, kde pravděpodobnost úspěšného výsledku je dána *p*, potom

$$p_i = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad E[X] = n \cdot p$$
$$V[X] = np(1-p)$$

Příklad: Test obsahuje 10 otázek s výběrem z 5 možných odpovědí.

1. Určete pravděpodobnost, že student nezaškrtne ani jednu odpověď správně

$$P(X=0) = \left(1-\frac{1}{5}\right)^{10} \doteq$$

2. Určete pravděpodobnost, že všechny odpovědi budou správné

$$P(X=0) = \left(\frac{1}{5}\right)^{10} \doteq \quad 01$$

3. Určete pravděpodobnost, že student zvolí alespoň 7 odpovědí správně

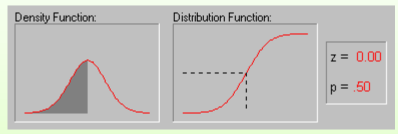
$$P(X \geq 7) = P(X=7) + P(X=8) + P(X=9) + P(X=10) \doteq \quad 4$$

4. Určete průměrný počet správných odpovědí

$$E(X) = 10 \cdot \frac{1}{5} = 2$$

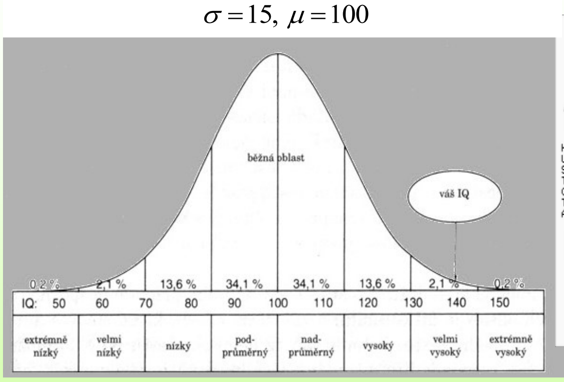
Normální rozdělení spojité náhodné veličiny $X \sim N(\mu, \sigma^2)$

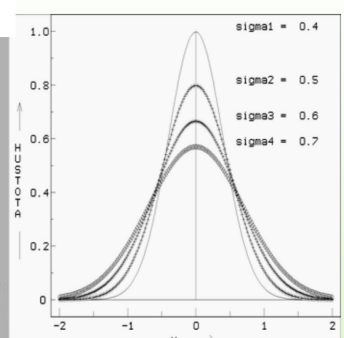
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Počet bodů z testu inteligence

$\sigma = 15, \ \mu = 100$






```
ynorm=randn(500,1);
y=ynorm.*s+nu; %Linearni transformace
fprintf('Mean = %2.2f\n', mean(y));
fprintf('Standard deviation = %2.3f\n',std(y));
hist(y,40)
```

Studentovo rozdělení

Pokud má proměnná *X* normální rozdělení, pak proměnná

$$Z = \frac{X - \mu}{\sigma}$$

má **normované normální rozdělení**.

Průměr náhodného výběru má **normální rozdělení** se směrodatnou odchylkou

$$\bar{\sigma} = \frac{\sigma}{\sqrt{n}}$$

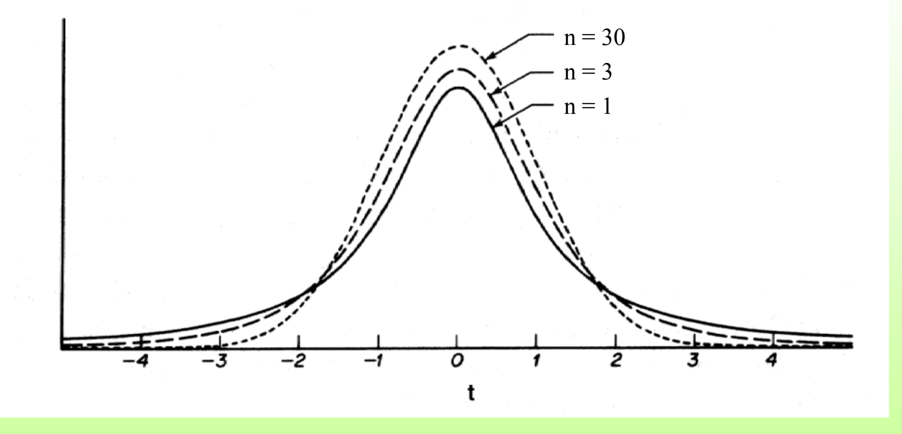
Odchylku σ většinou neznáme, ale můžeme ji odhadnout pomocí výběrové směrodatné odchylky s. Proměnná t

$$t = \frac{\overline{X} - \mu}{\bar{s}}$$

má **Studentovo t rozdělení**.

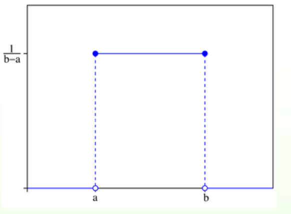
Studentovo rozdělení

$$X = \frac{U}{\sqrt{\frac{V}{n}}} \qquad U \approx N(0,1) \qquad V \approx \chi^2(n)$$



Obr. *t*-distribuce pro různé stupně volnosti n. Pro n=∞, *t* distribuce je identická s normální distribucí.

Rovnoměrné rozdělení spojité náhodné veličiny



$$f(x)=\frac{1}{b-a};\quad a\le x\le b$$

$$E[X]=\frac{a+b}{2};\quad V[X]=\frac{(a+b)^2}{12}$$

```
y=rand(1,n)*(b-a)+a;
prb_upto1=sum(y<1)/n % P(Xi<1)
```

Příklad: Tramvaje jezdí pravidelně každých 5 minut. Na zastávku přijdeme náhodně.

1. Určete pravděpodobnost, že budeme čekat nejvýš 1 minutu

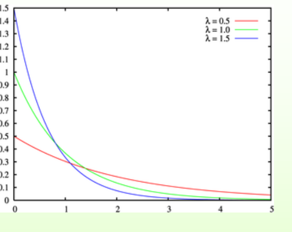
$$f(x)=\frac{1}{5};\quad 0\le x\le 5$$

$$P(x<1)=\int_0^1\frac{1}{5}dx=\left[\frac{1}{5}x\right]_0^1=\frac{1}{5}$$

2. Určete průměrnou dobu čekání

$$E[X]=\frac{5}{2};$$

Exponenciální rozdělení spojité náhodné veličiny



$$f(x)=\lambda e^{-\lambda x};\quad 0\le x\quad F(x)=1-e^{-\lambda x}$$

$$E[X]=\frac{1}{\lambda};\quad V[X]=\frac{1}{\lambda^2}$$

```
y=-log(1-rand(1,n))/lambda;
```

Příklad:

1. Doba dvou po sobě následujících jevů je exponenciálně rozdělená náhodná veličina s parametrem λ. Určete průměrný počet jevů za časovou jednotku.

$$E[T]=\frac{1}{\lambda}$$

2. Doba bezvadného chodu nového automobilu je náhodná veličina λ=1/10 [rok].
a) Určete pravděpodobnost, že se do 5 let neobjeví žádná závada.

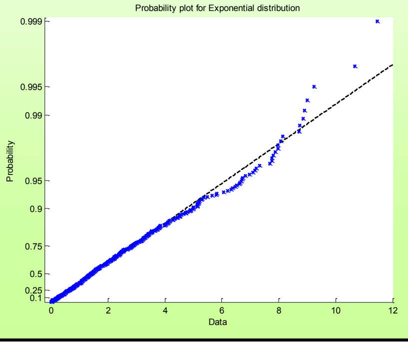
$$P(X\ge 5)=1-F(5)=1-\left(1-e^{-\frac{1}{10}x}\right)=e^{-\frac{1}{2}}=0{,}606$$

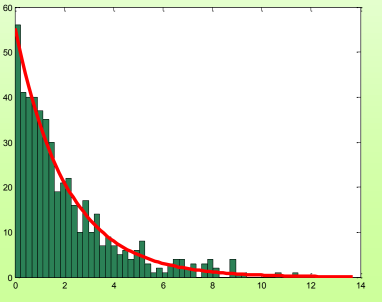
b) Určete průměrnou dobu bezvadného chodu auta.

$$E[T]=\frac{1}{0.1}=10\ (let)$$

Exponenciální rozdělení – statistický toolbox

```
lambda=2;
y=random('Exponential',lambda,500,1);
histfit(y,50,'Exponential')
figure
probplot('Exponential',y)
```





Erlangovo rozdělení spojité náhodné veličiny $X \sim Erlang(\lambda,k)$

$$f(x)=\lambda e^{-\lambda x}\frac{(\lambda x)^{k-1}}{(k-1)!};\quad 0\le x\quad E[X]=\frac{k}{\lambda};\quad V[X]=\frac{k}{\lambda^2}$$

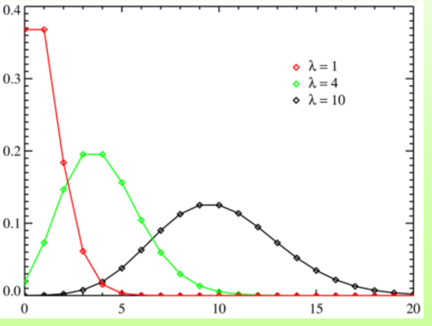
```
function[y]=erlang(n,lambda,k)
%help
for i=1:n
   x=-log(1-rand(1,k))/lambda
   y(i)=sum(x);
end
```

Součet k nezávislých náhodných veličin, jež mají všechny exponenciální rozdělení $X_i \sim \exp(\lambda)$ je Erlangovo rozdělení $X \sim Erlang(\lambda,k)$

Poissonovo rozdělení diskrétní náhodné veličiny

- počet výskytů sledovaného jevu v určitém časovém intervalu t, jestliže posloupnost časových okamžiků sledovaného jevu tvoří ordinální homogenní proces s nezávislými přírůstky (Elementární tok).

$$P(N(t)=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t} \qquad E[X]=\lambda\cdot t$$



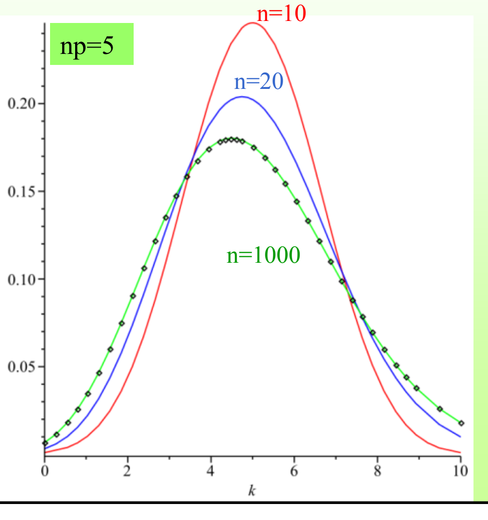
Poissonovo rozdělení diskrétní náhodné veličiny

- Pokud n→∞, pak náhodná veličina s binomickým rozdělením konverguje k Poissonovu rozdělení, *np*=λ

$$Binom(n,p)\rightarrow Poisson(n\cdot p)$$

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$$

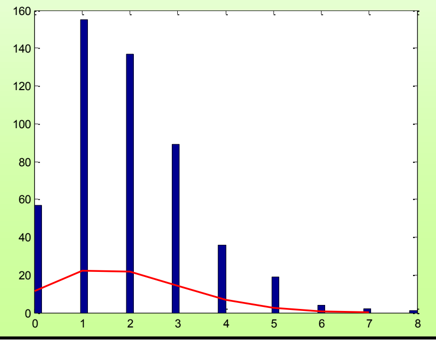
$$E[X]=\lambda$$

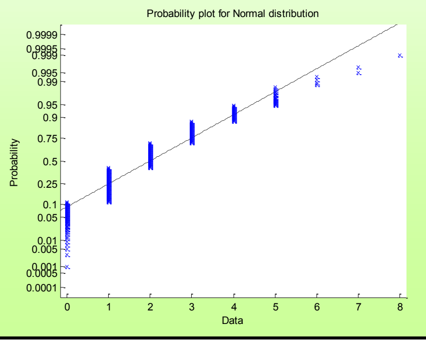


Poissonovo rozdělení diskrétní náhodné veličiny

- Pokud n→∞, pak náhodná veličina s binomickým rozdělením konverguje k Poissonovu rozdělení, $np=\lambda$

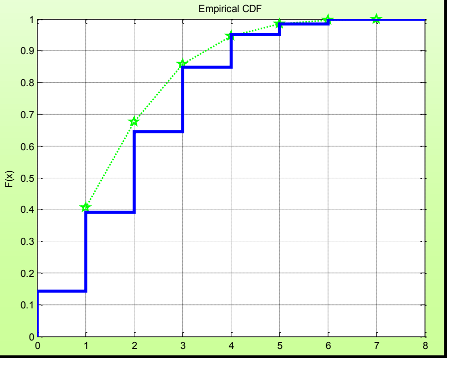
```
lambda=2;
y=random('Poisson',lambda,500,1); % data vector
histfit(y,50,'poisson')
figure
probplot('normal',y) % normality test
```





Poissonovo rozdělení diskrétní náhodné veličiny

```
lambda=2;
y=random('Poisson',lambda,500,1);
cdfplot(y)
hold on
plot(1:7,cdf('Poisson',1:7,lambda),'g+')
hold off
```



Centrální limitní věta

Lévyho-Lindebergova věta.

Pokud je náhodná veličina X součtem n vzájemně nezávislých náhodných veličin X_1, X_2, ...X_n se shodným rozdělením libovolného typu, s konečnou střední hodnotou μ a s konečným rozptylem σ^2, pak pro normovanou náhodnou veličinu

$$U = \frac{X - n\mu}{\sqrt{n\sigma^2}}$$

platí vztah

$$\lim_{n\to\infty} P(U < u) = \Phi(u)$$

kde Φ(u) je distribuční funkce normovaného normálního rozdělení N(0,1).

- Př: Doba životnosti auta má exponenciální rozdělení s parametrem (1/15). Potom normovaný tvar průměru dob životnosti nezávisle vyráběných aut

  $$U = \frac{\overline{X} - 15}{15\sqrt{n}}$$

  je možné aproximovat normálním rozdělením N(0,1)

Centrální limitní věta

Centrální limitní věta označuje tvrzení, podle něhož (za určitých podmínek) se rozdělení výběrového průměru blíží k normálnímu rozdělení. O náhodné veličině s uvedeným chováním říkáme, že má asymptoticky normální rozdělení.

WikipediE

- Údaje, které jsou ovlivňovány velkým počtem malých a na sobě nezávislých efektů budou rozděleny přibližně normálně
- Čím větší je rozsah výběru, tím více se rozdělení průměrů blíží normálnímu rozdělení

  $$N(\mu, \frac{\sigma^2}{n})$$

Vygenerujte 100 hodnot náhodné veličiny s rovnoměrným rozdělením na intervalu (a,b) a vypočtěte průměr vzorku. Výběr 50 x opakujte. Nakreslete histogram výběrových průměrů, vypočtěte průměr a směrodatnou odchylku.