[Náhodná veličina 3](#_Toc119939461)

[Distribuční funkce – F(x), F(X≤x) 3](#_Toc119939462)

[Příklady 4](#_Toc119939463)

[Pravděpodobnostní funkce – p(x) 4](#_Toc119939464)

[Hustota pravděpodobnosti – f(x) 4](#_Toc119939465)

[Příklady 5](#_Toc119939466)

[Bodové ukazatele náhodné veličiny (spojité funkce) 5](#_Toc119939467)

[Příklady 5](#_Toc119939468)

[Střední hodnota 𝐸(X) 5](#_Toc119939469)

[Příklady 6](#_Toc119939470)

[Rozptyl D(X) 6](#_Toc119939471)

[Příklady 6](#_Toc119939472)

[Směrodatná odchylka 𝜎 6](#_Toc119939473)

[Medián 𝜎(X) 6](#_Toc119939474)

[Příklady 6](#_Toc119939475)

[Šikmost 6](#_Toc119939476)

[Příklady 7](#_Toc119939477)

[Kvantily 7](#_Toc119939478)

[Příklady 7](#_Toc119939479)

[Charakteristiky numerických proměnných (data) 7](#_Toc119939480)

[Příklady 7](#_Toc119939481)

[Aritmetický průměr 8](#_Toc119939482)

[Výběrový rozptyl 8](#_Toc119939483)

[Výběrová směrodatná odchylka 8](#_Toc119939484)

[Kvantily, kvartily a medián 9](#_Toc119939485)

[Příklady 9](#_Toc119939486)

[MAD 9](#_Toc119939487)

[Výběrová šikmost 9](#_Toc119939488)

[Modus 9](#_Toc119939489)

[Identifikace odlehlých měření 10](#_Toc119939490)

[Příklady 10](#_Toc119939491)

[Metoda z-souřadnice 10](#_Toc119939492)

[Metoda vnitřních hradeb 10](#_Toc119939493)

[𝑥0.5 souřadnice 10](#_Toc119939494)

[Diskrétní rozdělení pravděpodobnosti 11](#_Toc119939495)

[Hypergeometrické rozdělení 11](#_Toc119939496)

[Příklady 12](#_Toc119939497)

[Binomické rozdělení 13](#_Toc119939498)

[Příklady 14](#_Toc119939499)

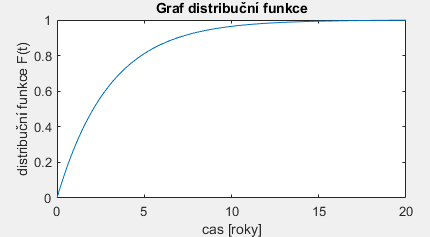
[Alternativní rozdělení 14](#_Toc119939500)

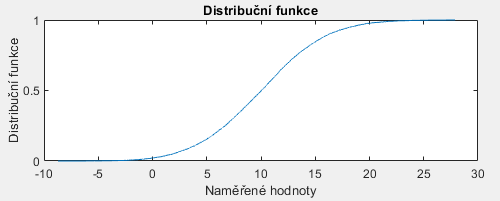
[Geometrické rozdělení 14](#_Toc119939501)

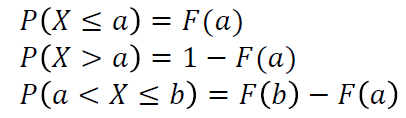
# Náhodná veličina

## Distribuční funkce – F(x), F(X≤x)

* Graf popisující pravděpodobnost, že náhodná veličina X nabude menší nebo rovna hodnotě x.
  + **Pravděpodobnost, že životnost výrobku je menší než 𝑡 (3 roky).**
  + Pravděpodobnost, že bude člověk vážit méně než 50 kg.



(pravděpodobnost 80%, že životnost výrobku je menší než 5 let) 

* **Vlastnosti funkce:**
  + 0 ≤ 𝐹(𝑥) ≤ 1 - Funkce je omezena
  + Funkce je neklesající
  + Funkce začíná v nule a končí v 1 (X může být i záporný)
  + 

### Příklady

Z naměřených dat výrobte distribuční funkci:

x = importdata('data.mat');

x = sort(x);

S = size(x,1);

V = 1/S;

y = 0+V/2:V:1-V/2;

plot(x,y);

title('Distribuční funkce');

xlabel('Naměřené hodnoty');

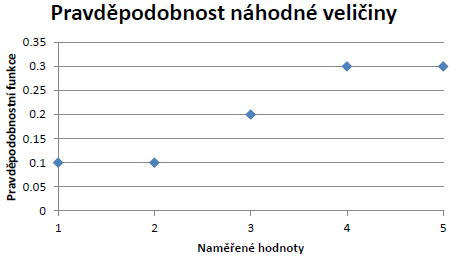
ylabel('Distribuční funkce');

figure;

cdfplot(x); %nebo použití Matlab funkce

## Pravděpodobnostní funkce – p(x)

* Pro diskrétní náhodnou veličinu.
* Pravděpodobnost, že náhodná veličina bude nabývat přímo hodnoty x.
  + Pravděpodobnost, že při hodu šestistěnnou kostkou mi padne 3.



## Hustota pravděpodobnosti – f(x)

* Pro spojitou náhodnou veličinu.
* Graf **derivace distribuční funkce**.
* Pravděpodobnost, že náhodná veličina bude nabývat přímo hodnoty x.
  + Pravděpodobnost, že životnost výrobku bude 3 roky.
* **Vlastnosti funkce:**
  + reálná nezáporná funkce o celkové ploše = 1
  + 
  + Pravděpodobnost, že se výrobek porouchá v intervalu <𝑎,𝑏>.
    - 
  + Pravděpodobnost, že se výrobek porouchá do času 𝑡.
    - 
  + Pravděpodobnost, že životnost výrobku je vyšší než 𝑎
    - 

### Příklady

Určete, že životnost výrobku je delší než 3 časové jednotky:

syms t;

f = 0.5\*exp(-0.5\*t); %hustota pravděpodobnosti

F = int(f,3,inf);

F = double(F);

Zjistěte distribuční funkci z hustoty pravděpodobnosti:

syms t C;

f = 0.5\*exp(-0.5\*t); %hustota pravděpodobnosti

F = int(f) + C; %integrace

F(t) = F; %časově závislé

r1 = F(0) == 0; %rovnice pro určení konstanty

r2 = F(inf) == 1; %rovnice pro určení konstanty

C = solve(r1,r2,C); %určení integrační konstanty

F(t) = subs(F(t)); %dosazení

Zjistěte hustoty pravděpodobnosti z distribuční funkce:

syms t;

F = 1-exp(-0.5\*t); %distribuční funkce

f = diff(F,t); %hustota pravděpodobnosti

# Bodové ukazatele náhodné veličiny (spojité funkce)

### Příklady

Náhodná veličina má distribuční funkci 𝐹(𝑥), pro 𝑥∈〈0,2〉. Určete střední hodnotu, rozptyl, směrodatnou odchylku, medián náhodné veličiny. Jaká je pravděpodobnost, že náhodná veličina bude mít výsledek v intervalech 〈0.5,1.5〉:

syms x;

F(x) = (x^3)/8; %distribuční funkce

f = diff(F) %hustota pravděpodobnosti

EX = int(x.\*f,0,2) %střední hodnota

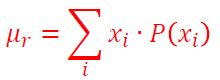
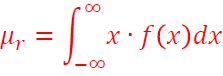
DX = int((x-EX).^2.\*f,0,2) %rozptyl

median = solve(F(x)==0.5,'Real',true); %median

smerod = sqrt(DX); %směrodatná odchylka

P = F(1.5) - F(0.5); %pravděpodobnost

## Střední hodnota 𝐸(X)

* Průměr všech realizací náhodné veličiny.
* 
* 

### Příklady

Vypočítejte střední hodnotu z hustoty pravděpodobnosti:

syms x;

fx = 0.25\*exp(-x/4);

EX = int(x\*fx,0,inf);

## Rozptyl D(X)

* Parametr vyjadřující variabilitu (rozptýlenost) realizací od střední hodnoty.
  + Jednotkou rozptylu je kvadrát střední hodnoty [𝑚2,𝑘𝑔2,°𝐶2] apod.
* 
* 
* 

### Příklady

Vypočítejte rozptyl z hustoty pravděpodobnosti:

syms x;

fx = 0.25\*exp(-x/4);

EX = int(x\*fx,0,inf);

DX = int((x-EX)^2\*fx,0,inf);

## Směrodatná odchylka 𝜎

* Zavádí se z důvodu nevhodných jednotek rozptylu.
* 

## Medián 𝜎(X)

### Příklady

Určete medián z distribuční funkce

syms x;

Fx=x\*x/4;

median = solve(Fx==0.5);

## Šikmost

* Šikmost je mírou symetrie rozdělení pravděpodobnosti.
* Definována je jako podíl 3. centrálního momentu a třetí mocniny směrodatné odchylky.
* 
* 𝛼3=0 symetrické rozdělení
* 𝛼3<0 negativně zešikmené rozdělení
* 𝛼3>0 pozitivně zešikmené rozdělení

### Příklady

Vypočítejte šikmost z hustoty pravděpodobnosti:

syms x;

fx = 0.25\*exp(-x/4);

EX = int(x\*fx,0,inf);

DX = int((x-EX).^2.\*fx,0,inf);

smerod = sqrt(DX);

treti\_moment = int((x-EX).^3.\*fx,0,inf);

sikmost = treti\_moment./smerod.^3;

## Kvantily

* Představuje takovou hodnotu, že pravděpodobnost, že náhodná veličina nabude hodnoty menší než 𝑥𝑝 je 100\*p procent.
* 50% kvantil je nazýván medián.
  + Rozděluje datový soubor na dvě skupiny takové, že v první je 50 % dat menších než medián a obdobně v druhé je 50 % dat větších než medián.
* 25% a 75% kvantil je nazýván dolní / horní kvartil.
* 1% kvantil je nazýván percentil.

### Příklady

Určete dolní kvartil z distribuční funkce:

syms t;

Fx = 1-exp(-0.1\*t); %distribuční funkce

kvantil = solve(Fx==0.25); %dolní kvartil

# Charakteristiky numerických proměnných (data)

### Příklady

Máte data a určete z nich střední hodnotu, rozptyl, směrodatnou odchylku, medián a šikmost:

x = importdata('data.xlsx');

x = x.data.List1;

stredni\_hod = nanmean(x); %střední hodnota

rozptyl = nanvar(x); %rozptyl

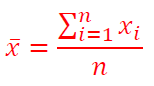
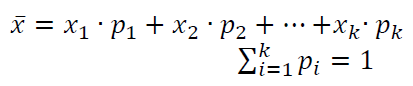
smerod = nanstd(x); %směrodatná odchylka

median = quantile(x,0.50); %medián

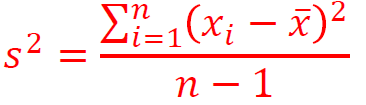
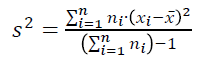
median = nanmedian(x); %nebo medián

sikmost = skewness(x); %šikmost

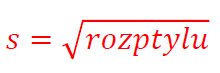
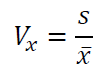
## Aritmetický průměr

* Vstupem jsou data:
  + 𝑥𝑖 - vstupní data, 𝑛 – počet dat
  + 
  + mean(x)
  + nanmean(x) – počítá průměr pouze z číselných hodnot
* Vstupem je četnostní tabulka:
  + 𝑥𝑖 – data, 𝑛𝑖 - představuje buď váhu, četnost, nebo procento výskytu
  + 
* Vstupem jsou pravděpodobnosti jevů:
  + 

## Výběrový rozptyl

* Popisuje variabilitu dat kolem střední hodnoty
  + Jak jsou data blízké či vzdálené.
* Vstupem jsou data:
  + 
* Vstupem je četnostní tabulka:
  + 
* Vstupem jsou pravděpodobnosti jevů:
  + 
* var(x)
* Určení libovolného centrálního momentu: moment(x, řád momentu)

## Výběrová směrodatná odchylka

* 
* std(x)
* **Variační koeficient**
  + Vhodný pro porovnání míry variabilit mezi sebou.
  + 

## Kvantily, kvartily a medián

* 100∙𝑝 % kvantil proměnné 𝑥 je hodnota, která rozděluje soubor na 100∙𝑝 % menších hodnot od zbytku.
* Kvantily tvoří inverzní funkci k distribuční funkci.
* 50% kvantil je nazýván medián.
  + median(x), nanmedian(x)
  + Rozděluje datový soubor na dvě skupiny takové, že v první je 50 % dat menších než medián a obdobně v druhé je 50 % dat větších než medián.
* 25% a 75% kvantil je nazýván dolní / horní kvartil.
* 1% kvantil je nazýván percentil.
* **Interkvartilové rozpětí**
  + Rozdíl mezi horním a dolním kvartilem.
  + Pomocí interkvartilového rozpětí lze porovnávat variabilitu souboru.
    - Jak jsou hodnoty veličiny u jednotlivých prvků souboru vzájemně blízké či vzdálené.
  + iqr(x)

### Příklady

Určete 5% a 95% kvantil a dolní a horní kvartil a medián a vykreslete distribuční funkci F(X):

x = importdata('data.mat');

Kq5procent = quantile(x,0.05) %5% kvantil

Kq50procent = quantile(x,0.50) %median

Kq95procent = quantile(x,0.95) %95% kvantil

Kq25procent = quantile(x,0.25) %dolní kvantil

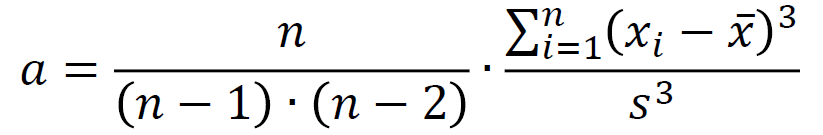
Kq75procent = quantile(x,0.75) %horní kvantil

F = cdfplot(x) %distribuční funkce

## MAD

* a) Střední hodnota absolutních odchylek od střední hodnoty.
  + mad(x,0)
* b) Medián absolutních odchylek od mediánu.
  + mad(x,1)

## Výběrová šikmost

* Vyjadřuje symetrii rozložení hodnot kolem průměru.
* 
* 𝑎 = 0 data jsou kolem průměru rozloženy symetricky.
* skewness(x)

## Modus

* Jedná se o nejčetnější hodnotu.
* mode(x)

# Identifikace odlehlých měření

### Příklady

Nalezněte odlehlé data a ty odstraňte:

x = [5 6 4 6 5 25 6 -30 4 6 5 6 4 4 5 6 5 5]; %data

xdolni = quantile(x,0.25)-1.5\*iqr(x) %vnitřní hradby

xhorni = quantile(x,0.75)+1.5\*iqr(x) %vnitřní hradby

x(x > xhorni) = []; %odstranění odlehlých dat

x(x < xdolni) = []; %odstranění odlehlých dat

## Metoda z-souřadnice

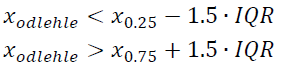
* Data z normálního rozdělení, symetrická.
* 

%z souradnice

xdolni = mean(x)-3\*std(x)

xhorni = mean(x)+3\*std(x)

## Metoda vnitřních hradeb

* Data nesymetrická.
* 

%vnitřní hradby

xdolni = quantile(x,0.25)-1.5\*iqr(x)

xhorni = quantile(x,0.75)+1.5\*iqr(x)

## 𝑥0.5 souřadnice

* Data nesymetrická.
* 

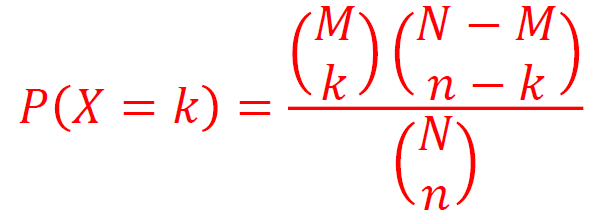
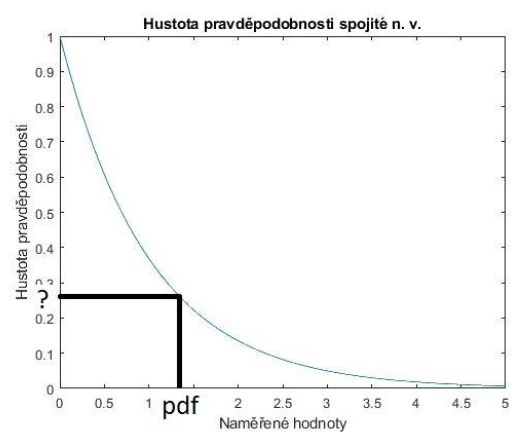
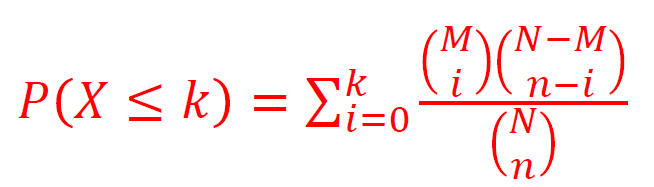
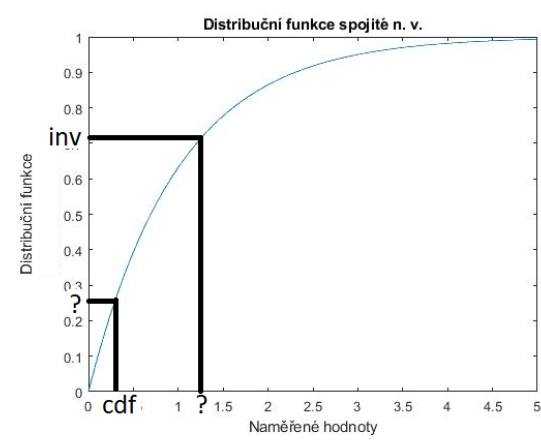
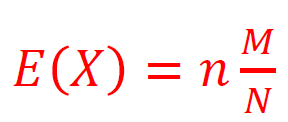
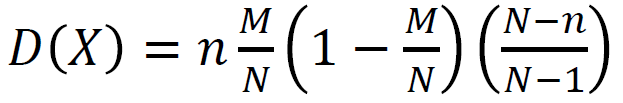
%0.5 souřadnice

xdolni = median(x)-3\*1.483\*mad(x)

xhorni = median(x)+3\*1.483\*mad(x)

# Diskrétní rozdělení pravděpodobnosti

## Hypergeometrické rozdělení

* Pravděpodobnost určitého počtu úspěchů v n **závislých** pokusech.
* **Závislé** = pokusy **bez vracení**, pravděpodobnost se při každém pokusu mění
  + N prvků obsahuje M prvků s určitou vlastností a zbylých N-M prvků tuto vlastnost nemá.
  + Náhodně se ze souboru vybere n prvků, z nichž se žádný nevrací zpět.
  + Pravděpodobnost, že z vylosovaných n prvků má právě k prvků danou vlastnost:
  + 
  + N – celkový počet prvků
  + M – celkový počet prvků s danou vlastností
  + n – počet vybraných prvků
  + k – počet vybraných prvků s danou vlastností
  + P = hygepdf(k,N,M,n)
  + 
* Distribuční funkce:
  + 
  + 
  + F = hygecdf(k,N,M,n)
  + Inverzní distribuční funkce
    - x = hygeinv(pravd,N,M,n)
* Střední hodnota:
  + 
  + [MN,var] = hygestat(N,M,n)
* Rozptyl:
  + 
  + [MN,var] = hygestat(N,M,n)
* Generování náhnodných čísel:
  + hygernd(N,M,n)

### Příklady

V osudí je 10 černých míčků a 15 bílých. Z osudí vylosujeme 4 míčky, které nevracíme. Určete pravděpodobnost, že **právě** 3 míčky budou bílé a jeden černý:

N = 25; %(celkem je v osudí 25 míčků)

M = 15; %(15 bílých míčků, míčky s určitou vlastností)

n = 4; %(losujeme 4 míčky)

k = 3; %(3 míčky budou bílé)

P = hygepdf(k,N,M,n);

Určete pravděpodobnost, že vylosujete **2 a méně** bílých míčků:

F = hygecdf(2.5,25,15,4); %2.5 pro lepší pocit

Vypočtěte pravděpodobnost, že z 32 karetního balíčku budou při vylosování 3 karet právě 2 esa.

N = 32;

M = 4;

n = 3;

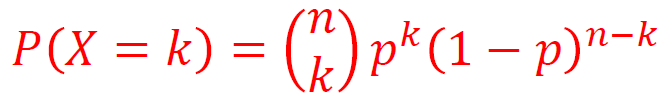
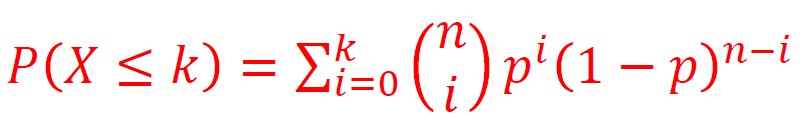
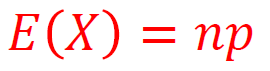
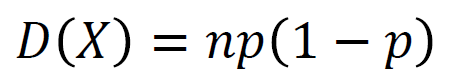
k = 2;

P = hygepdf(k,N,M,n);

Vypočtěte pravděpodobnost, že při výběru 10 karet z 32 karetního balíčku bude právě 8 vyšších karet (spodek, filek, král nebo eso). Balíček obsahuje 4 spodky, 4 filky, 4 krále a 4 esa.

P = hygepdf(8,32,16,10);

## Binomické rozdělení

* Pravděpodobnost určitého počtu úspěchů v n **nezávislých** pokusech.
* **Nezávislé** = pokusy **s vracením**, pravděpodobnost se nemění
  + 
  + n – počet náhodných pokusů
  + p – pravděpodobnost úspěšného pokusu
  + k – počet úspěšných pokusů
  + P = binopdf(k,n,p)
* Distribuční funkce:
  + 
  + F = binocdf (k,n,p)
  + Inverzní distribuční funkce
    - x = binoinv(pravd,n,p)
* Střední hodnota:
  + 
  + [MN,var] = binostat(n,p)
* Rozptyl:
  + 
  + [MN,var] = binostat(n,p)
* Generování náhnodných čísel:
  + binornd(n,p)
* Odhad parametrů rozdělení
  + p = binofit(k,n)

### Příklady

V osudí je 10 černých míčků a 15 bílých. Z osudí vylosujeme 4 míčky, které vracíme. Určete pravděpodobnost, že 3 míčky budou bílé a jeden černý.

n = 4; %počet náhodných pokusů

p = 15/25; %pravděpodobnost úspěšného pokusu

k = 3; %počet úspěšných pokusů

P = binopdf(k,n,p);

Pětkrát hodíme mincí. Určete pravděpodobnost, že orel padne **právě** dvakrát:

P = binopdf(2,5,0.5);

Určete pravděpodobnost, že padne **alespoň** 4krát (padne 4krát a více):

P = 1-binocdf(3.5,5,0.5); %jedna mínus pravděpodobnost že padne 3krát a méně

Zásilka obsahuje 80 % kvalitních a 20 % nekvalitních výrobků. Náhodně s vracením vybereme 5 výrobků. Určete pravděpodobnost, že:

a) **právě 3** budou kvalitní

P = binopdf(3,5,0.8);

b) **2 a méně** budou kvalitní

P = binocdf(2,5,0.8);

c) **alespoň 3** a více budou kvalitní

P = 1-binocdf(2,5,0.8);

d) **více než** **3** budou kvalitní

P = 1-binocdf(3,5,0.8);

Mějme mariášové karty (32 karet, které obsahují 4 esa, 4 krále, 4 filky,…, 4 sedmičky). Losujete karty z balíčku a vracíte je zpět. Určete pravděpodobnost, že z prvních 7 vylosovaných karet dostanete právě 4 esa nebo krále. A poslední osmá vylosovaná karta bude 7.

%prvnich sedm tahu

P = binopdf(4,7,8/32);

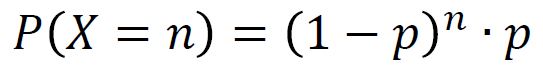
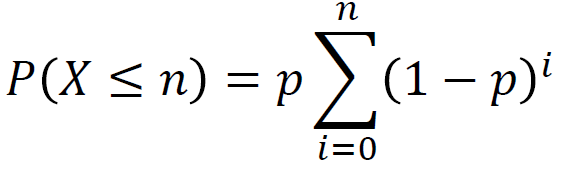
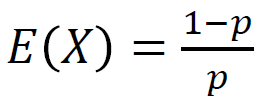
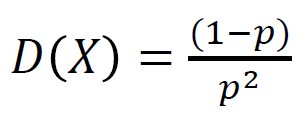
%osmy tah 7

vysledek = P\*4/32;

## Alternativní rozdělení

* Popisuje pravděpodobnost jednoho náhodného pokusu.
* Binomické rozdělení, kde parametr n=1.

## Geometrické rozdělení

* Popisuje počet neúspěšných pokusů před prvním úspěchem.
* Prvních n pokusů je bez úspěchu, právě (n+1) pokus je úspěšný.
* X – počet neúspěšných pokusů
* p – pravděpodobnost.
* Pravděpodobnostní funkce
  + 
  + P = geopdf(X,p)
* Distribuční funkce
  + 
  + F = geocdf(X,p)
  + Inverzní distribuční funkce
    - x = geoinv (pravd,p)
* Střední hodnota
  + 
  + [M,var] = geostat(p)
* Rozptyl
  + 
  + [M,var] = geostat(p)
* Generování náhnodných čísel:
  + geornd(p)

### Příklady

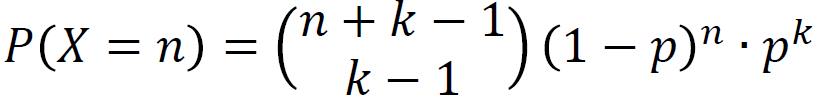
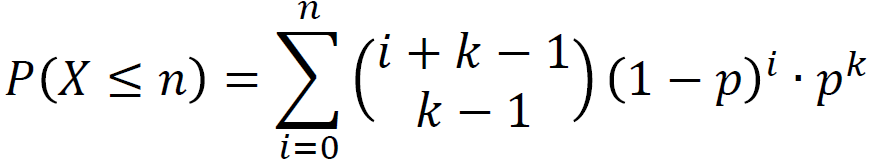
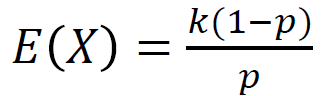
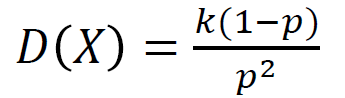
Určete pravděpodobnost, že **právě** u pátého hodu šestistěnou kostkou Vám padne poprvé 6:

P = geopdf(4,1/6); %pravděpodobnost úspěchu je 1/6

Určete pravděpodobnost, že **do** pátého hodu šestistěnou kostkou padne 6:

P = geocdf(4,1/6); %pravděpodobnost úspěchu je 1/6

## Negativně binomické rozdělení

* Prvních (n+k-1) pokusů má (k-1) úspěchů, právě (n+k) pokus je úspěšný.
* n – počet neúspěšných pokusů
* k – počet úspěšných pokusů
* p – pravděpodobnost úspěšného pokusu
* Pravděpodobnostní funkce
  + 
  + P = nbinpdf(n,k,p)
* Distribuční funkce
  + 
  + F = nbincdf (n,k,p)
  + Inverzní distribuční funkce
    - x = nbininv(pravd,k,p)
* Střední hodnota
  + 
  + [M,var] = nbinstat(k,p)
* Rozptyl
  + 
  + [M,var] = nbinstat(k,p)
* Generování náhnodných čísel:
  + nbinrnd(k,p)

### Příklady

Určete pravděpodobnost, že **do** 10 hodu šestistěnnou kostkou Vám padne právě 3x šestka:

n = 7; %počet neúspěšných pokusů

k = 3; %počet úspěšných pokusů

p = 1/6; %pravděpodobnost úspěšného pokusu

P = nbincdf(n,k,p);

Určete minimální počet hodů, abyste na šestistěnné kostce s pravděpodobností 0.95 obdrželi třikrát 6.

P = nbininv(0.95,3,1/6); %=33 neúspěšných hodů = 36 hodů

K = nbincdf(33,3,1/6); %jaká je pravděpodobnost, aby jsme **do** 36 hodech kostkou obdrželi právě 3x šestku = 0.95

Pravděpodobnost výskytu krevní skupiny A+ je 0.35. V nemocnici potřebují najít 3 dárce s touto krevní skupinou. Dárcové však neznají svojí krevní skupinu. Jaká je pravděpodobnost, že pro nalezení právě 3. dárce s krevní skupinou A+ budou muset vyšetřit:

a) **právě 10** dárců

P = nbinpdf(7,3,0.35);

b) **více jak 9** dárců

P = 1-nbincdf(6,3,0.35);

c) **aspoň 6** (včetně) a **nejvýše 10** dárců (včetně).

P = nbincdf(7,3,0.35) - nbincdf(2,3,0.35);

d) chceme vědět, že mezi 10 dárci budou **právě** 3 s krevní skupinou A+.

P = binopdf(3,10,0.35)