5. Některá důležitá rozdělení pravděpodobnosti

5.1. Diskrétní rozdělení pravděpodobnosti

Čas ke studiu kapitoly: 50 minut

Cíl Po prostudování tohoto odstavce budete umět

- charakterizovat Bernoulliho pokusy a z nich odvozené jednotlivé typy diskrétních rozdělení: binomické, geometrické, negativně binomické
- charakterizovat Poissonův proces a z něj vycházející Poissonovo rozdělení
- popsat vzájemnou souvislost mezi diskrétními rozděleními

VÝKLAD

Existuje mnoho typů diskrétních náhodných veličin. My si nyní shrneme základní poznatky o těch nejběžnějších.
Bernoulliho pokusy:

- posloupnost *nezávislých* pokusů majících pouze 2 možné výsledky (událost nastane-nenastane; úspěch-neúspěch; popřípadě 1-0)
- pravděpodobnost výskytu události (úspěchu) p je konstantní v každém pokuse

Binomická náhodná veličina:

$X \to B(n, p)$
počet výskytů události (úspěchů, jedniček) v posloupnosti „n“ nezávislých pokusů.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}; \; 0 \le k \le n$$

$$EX = \sum_{k=0}^{n} k \cdot P(X = k) = \sum_{k=0}^{n} k \cdot \frac{n!}{(n-k)!\,k!} \cdot p^k (1-p)^{n-k} = n \cdot p \sum_{k=0}^{n} \frac{(n-1)!}{(n-k)!\cdot(k-1)!} \cdot p^{k-1} \cdot (1-p)^{n-k} = n \cdot p$$

$DX = EX^2 - (EX)^2$

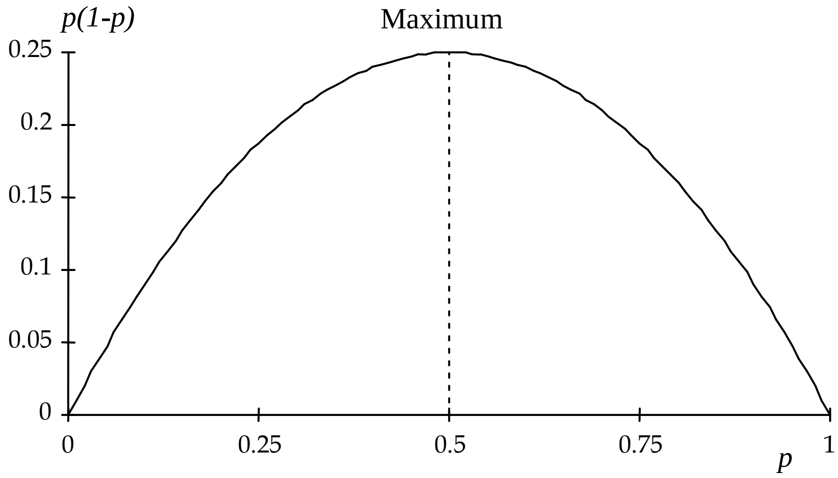
$$EX^2 = \sum_{k=1}^{n} k^2 \cdot P(X = k) = \sum_{k=1}^{n} k \cdot (k-1) \cdot \frac{n!}{(n-k)!\,k!} \cdot p^k \cdot (1-p)^{n-k} + EX =$$

$$= n \cdot (n-1) \cdot p^2 \cdot \sum_{k=2}^{n} \frac{(n-2)!}{(n-k)!\,(k-2)!} \cdot p^{k-2} \cdot (1-p)^{n-k} + EX =$$

$$= n \cdot (n-1) \cdot p^2 + n \cdot p = (np)^2 - np^2 + np$$

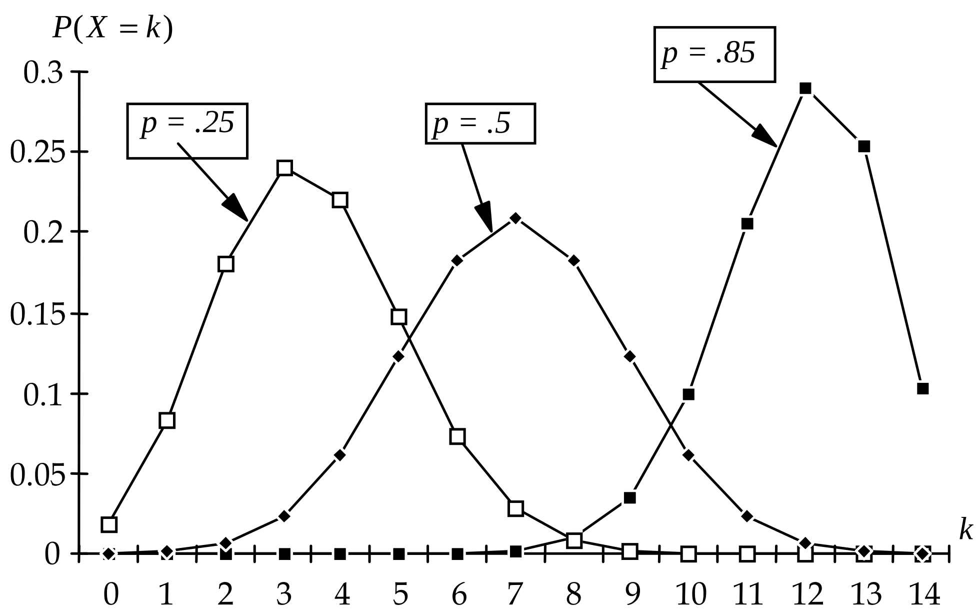
$DX = EX^2 - (EX)^2 = n.p.(1 - p)$

Z následujícího obrázku je patrné, že rozptyl je maximální pro $p=0.5$



Příklad:

Některé příklady demonstrace pravděpodobnostní funkce binomického rozdělení pro $n = 14$ pokusů jsou znázorněny na následujícím obrázku. Všimněme si, že pokud p roste, rozdělení se posouvá k vyšším hodnotám na x-ové ose. Dále, pokud $p = .5$, rozdělení je symetrické okolo hodnoty 7.5.



Geometrická náhodná veličina:

$X \rightarrow G(p)$

G(p) … počet Bernoulliho pokusů do prvního výskytu události (úspěchu) včetně něj

$$P(X = k) = p(1-p)^{k-1};\ 1 \le k < \infty$$

$$EX = \sum_{k=1}^{\infty} k \cdot P(X = k) = \sum_{k=1}^{\infty} k \cdot p \cdot (1-p)^{k-1} = p \cdot \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} = p \cdot \frac{\partial}{\partial(1-p)} \cdot \left(\sum_{k=1}^{\infty} (1-p)^k \right) = \frac{1}{p}$$

Poznámka: V předposledním výrazu nejdříve sečteme danou geometrickou řadu a posléze derivujeme.

Tedy střední hodnota počtu pokusů do prvního úspěchu je převrácená hodnota pravděpodobnosti úspěchu v každém pokuse, což znamená, že pokud 10% pokusů je úspěšných, pak potřebujeme průměrně 10 pokusů k tomu, abychom dosáhli úspěchu.

$$DX = EX^2 - (EX)^2 = \frac{1-p}{p^2}$$

Odvození:

$$E(X^2) = \sum_{k=1}^{\infty} k^2\, p(1-p)^{k-1}$$

$$= p(1-p) \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} + p \sum_{k=2}^{\infty} k(1-p)^{k-1}$$

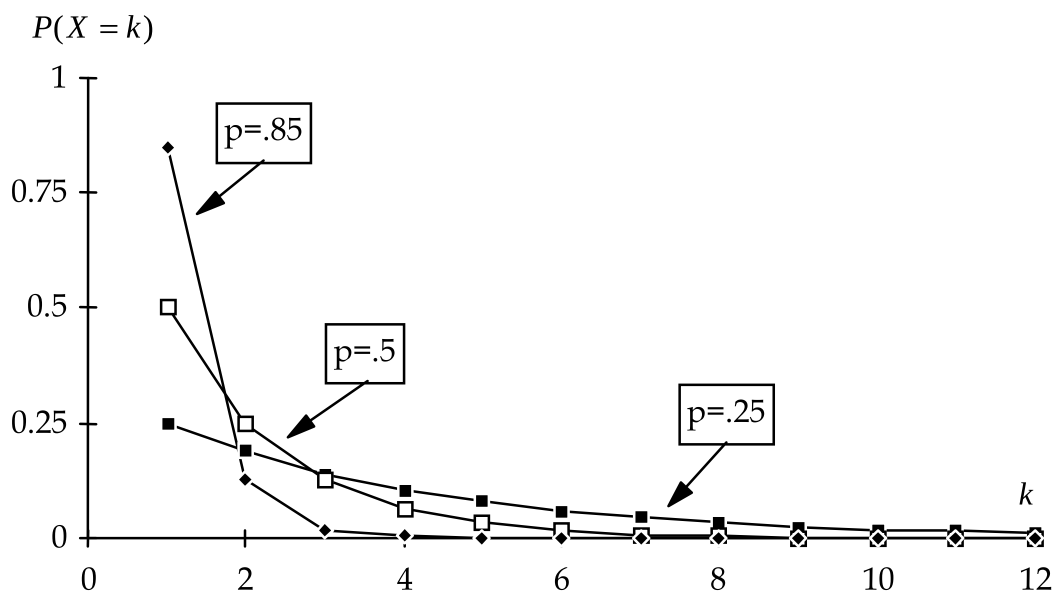
$$= p(1-p) \frac{\partial^2 \sum_{k=1}^{\infty} (1-p)^k}{\partial^2 (1-p)} + p \frac{\partial \sum_{k=1}^{\infty} (1-p)^k}{\partial (1-p)}$$

$$= p(1-p) \frac{\partial^2 \left(\frac{1-p}{p} \right)}{\partial^2 (1-p)} + p \frac{\partial \left(\frac{1-p}{p} \right)}{\partial (1-p)} = \frac{2(1-p)}{p^2} + \frac{1}{p}$$

Z čehož po dosazení dostáváme vztah pro rozptyl.

Příklad

Některé příklady geometrického rozdělení pro různé hodnoty *p* jsou ilustrovány níže. Z odvozeného vztahu vyplývá, že s klesající pravděpodobností výskytu úspěchu rozptyl vzrůstá.



Negativně binomická náhodná veličina

$X \rightarrow NB(k, p) \ldots$ počet pokusů do k-tého výskytu události (úspěchu) včetně tohoto k-tého výskytu.

$$P(X=n)=\binom{n-1}{k-1} p^k (1-p)^{n-k};\ \ k \le n < \infty$$

Negativní binomickou náhodnou veličinu si můžeme představit jako součet nezávislých k geometrických náhodných veličin:

$$W_i \rightarrow G(p); 1 \le i \le k$$

potom

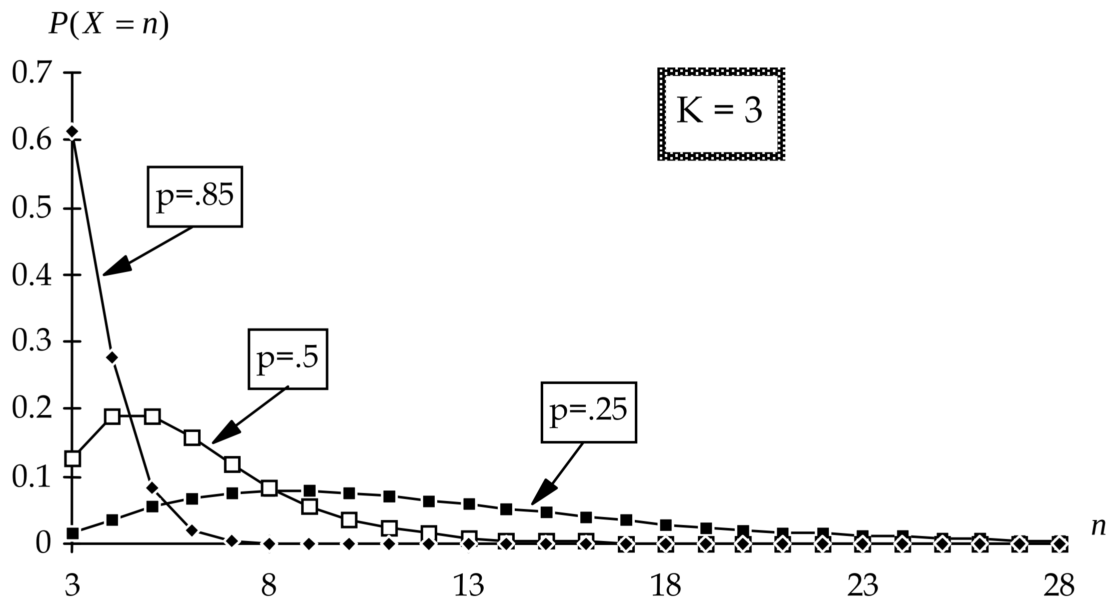
$$X = \sum_{i=1}^{k} W_i$$

$$E(X) = \sum_{i=1}^{k} E(W_i) = \frac{k}{p}$$

$$D(X) = \sum_{i=1}^{k} D(W_i) = \frac{k(1-p)}{p^2}$$

Příklad:

Následující obrázek ilustruje některé příklady NB rozdělení pro $k = 3$ a různé hodnoty p. Pokud p je v blízkosti hodnoty *0.5*, NB rozdělení má jednoduchý modus poblíž hodnoty 5. Tento modus se vzdaluje směrem od počátku a přitom se jeho pravděpodobnostní hodnota zmenšuje, pokud p klesá, což znamená růst rozptylu pro klesající p. NB rozdělení má podobný tvar jako geometrické rozdělení pro velké hodnoty p.



***Poznámka*:**
Porovnání binomického a NB rozdělení.

Ačkoliv se může na první pohled zdát, že obě rozdělení mají podobnou pravděpodobnostní funkci, existují významné rozdíly:

Binomické rozdělení

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}; 0 \le k \le n$$

V tomto vztahu je *k* náhodné a *n* deterministické.

Negativně binomické rozdělení

$$P(X = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}; k \le n < \infty$$

V tomto vztahu je *n* náhodné a *k* deterministické.

Poissonův proces

Poissonův proces je další z obecných modelů schémat sběru dat, který má široké využití v praxi. Lze ho chápat jako zobecnění Berhoulliho posloupnosti pokusů ve spojitém čase. Poissonův proces popisuje výskyt náhodných událostí na nějakém pevném časovém intervalu. Poissonův proces předpokládá, že **rychlost výskytu událostí je konstantní** v průběhu celého intervalu (nebo nějaké vymezené prostorové oblasti). Dalším předpokladem je **nezávislost** událostí.

λ … lambda $\rightarrow$

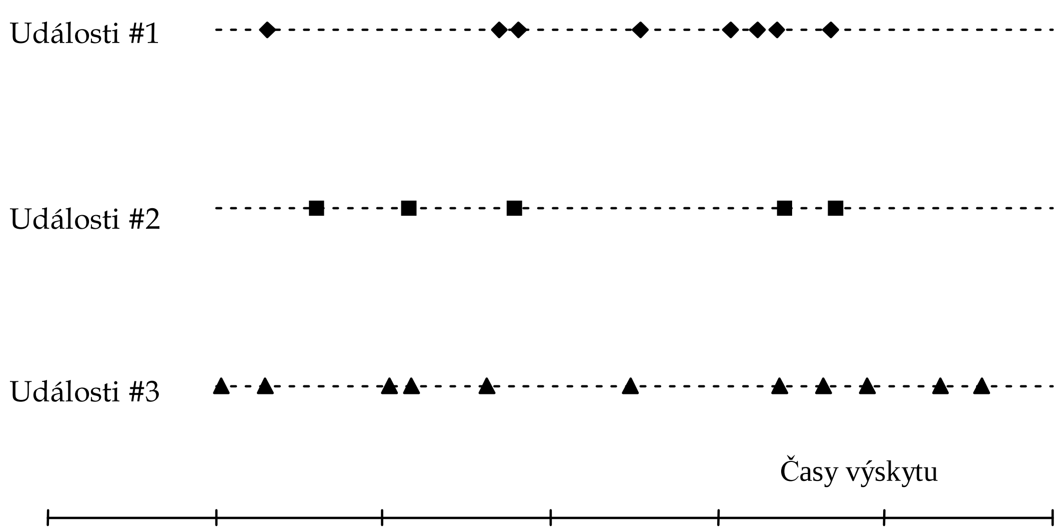
rychlost výskytu událostí (je úměrná pravděpodobnosti výskytu jedné události za jednotku času)

Poissonův proces popisuje události, které se vyskytnou náhodně v průběhu nějakého intervalu nebo v nějaké prostorové oblasti.

Obecným názvem pro takové procesy je ***bodový proces***. Poissonův proces je speciální případ bového procesu.

Příklady:

- počet studentů vstupujících do budovy VŠB TUO od 8:00 do 9:00 hod.
- počet pacientů ošetřených během dopoledních ordinačních hodin
- počet mikrodefektů v zadaném vzorku materiálu, atd.



Poissonovo rozdělení pravděpodobnosti

- **Jak vypadá pravděpodobnostní funkce a základní číselné charakteristiky Poissonova rozdělení ?**

$X \rightarrow P(\lambda t)$

Uvažujme Poissonův proces, který je pozorován v průběhu času *t*. Předpokládejme, že rychlost výskytu událostí je λ. Potom pravděpodobnost výskytu událostí během intervalu od 0 do t bude úměrná hodnotě λt. Nyní rozdělíme interval *t* na *n* subintervalů stejné délky *t/n*. Výskyt událostí v každém z těchto intervalů bude nezávislý a pravděpodobnost výskytu událostí během jednoho tohoto malého intervalu bude úměrná hodnotě $\lambda\ t/n$. Pokud *n* je dostatečně velké číslo, pak délka intervalu *t/n* bude dostatečně malá natolik, že pravděpodobnost výskytu více než jedné události v tomto intervalu je téměř nulová a pravděpodobnost výskytu jedné události je úměrná $\lambda t\ /n$. Potom pravděpodobnostní rozdělení počtu událostí vyskytlých během celého intervalu délky *t* bude možno aproximovat binomickým rozdělením s parametry *n* a $\lambda t\ /n$. Tedy,

$$P(X=k) = \binom{n}{k}\left(\frac{\lambda t}{n}\right)^{k}\left(1-\frac{\lambda t}{n}\right)^{n-k}$$

Po provedení limity dostáváme,

$$\lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1-\frac{\lambda t}{n}\right)^{n-k} = \frac{(\lambda t)^k}{k!}\lim_{n\to\infty}\frac{n!}{(n-k)!\,n^k}\left(1-\frac{\lambda t}{n}\right)^{n-k} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad ,$$

neboť celý první zlomek s faktoriály má po vykrácení limitu 1. Takže můžeme vyjádřit pravděpodobnostní funkci Poissonova rozdělení jako:

$$P(X=k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!};\ 0\le k<\infty$$

Snadno se můžeme přesvědčit o splnění normalizační podmínky,

$$\sum_{k=0}^{\infty}P(X=k)=\sum_{k=0}^{\infty}\frac{(\lambda t)^k e^{-\lambda t}}{k!}=e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}=1$$

Výpočet střední hodnoty:

$$E(X)=\sum_{k=0}^{\infty}kP(X=k)=\sum_{k=0}^{\infty}k\frac{(\lambda t)^k e^{-\lambda t}}{k!}=\lambda t\sum_{k=1}^{\infty}\frac{(\lambda t)^{k-1}e^{-\lambda t}}{(k-1)!}=\lambda t$$

Dále,

$$E(X^2)=\sum_{k=0}^{\infty}k^2P(X=k)=\sum_{k=2}^{\infty}k(k-1)\frac{(\lambda t)^k e^{-\lambda t}}{k!}+\mathrm{E}(\mathrm{X})=$$

$$=(\lambda t)^2\sum_{k=2}^{\infty}\frac{(\lambda t)^{k-2}e^{-\lambda t}}{(k-2)!}+\lambda t = (\lambda t)^2+\lambda t$$

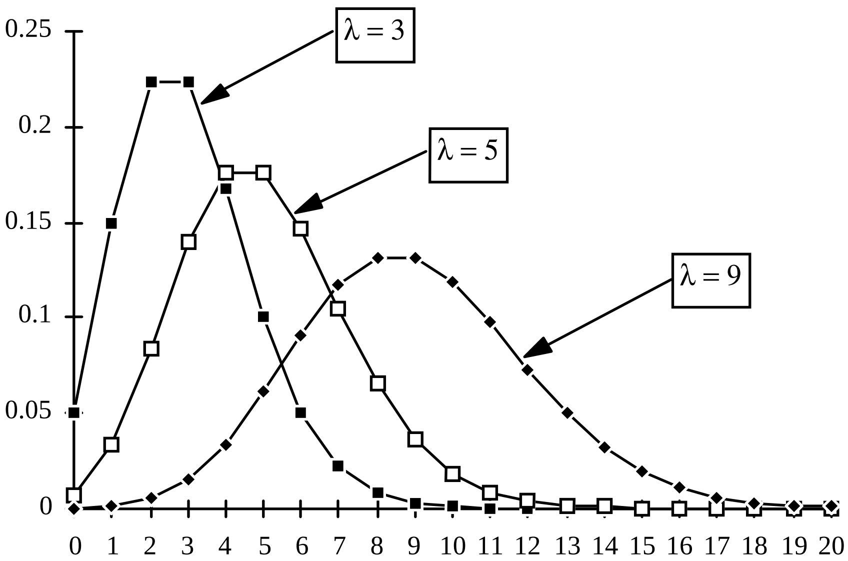
Pro rozptyl pak dostáváme,

$$D(X)=E[(X-E(X)]^2)=\mathrm{E}(\mathrm{X}^2)-(\mathrm{E}(\mathrm{X}))^2=\lambda t$$

Čili zajímavost tohoto rozdělení spočívá v tom, že střední hodnota je stejná jako rozptyl.

Příklady

Obrázek ilustruje příklady Poissonova rozdělení pro různé hodnoty λ, při *t=1*. Poznamenejme, že pro $\lambda = 9$ je rozdělení téměř symetrické.



Obrázek: Pravděpodobnostní funkce Poissonova rozdělení pravděpodobnosti

5.2. Spojitá rozdělení pravděpodobnosti

Čas ke studiu kapitoly: 50 minut

Cíl Po prostudování tohoto odstavce budete umět

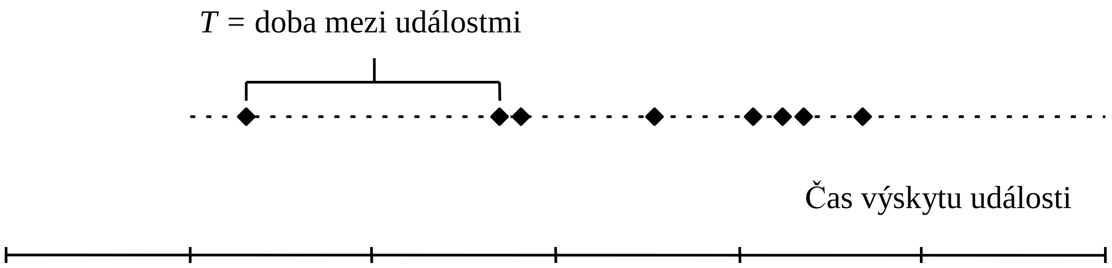
- charakterizovat jednotlivé typy spojitých rozdělení: exponenciální, Gamma, Weibullovo
- popsat jejich vzájemnou souvislost

VÝKLAD

- **Jak vypadá základní popis pro exponenciální, Gamma a Weibullovo rozdělení ?**

Exponenciální rozdělení

$T \rightarrow E(\lambda)$



T má exponenciální rozdělení

Popisuje náhodnou veličinu T ... dobu do výskytu první události nebo dobu mezi dvěma po sobě jdoucími událostmi v Poissonově procesu. T je tedy **spojitá** náhodná veličina.
N_t … počet událostí Poissonova procesu v intervalu $(0,t)$ $N_t \rightarrow P(\lambda t)$
Nejdříve zapíšeme ekvivalenci dvou jevů:
$N_t \geq 1 \Leftrightarrow T < t$
Pravděpodobnost jevu na pravé straně ekvivalence je definice distribuční funkce T.

$$F(t)=P(T < t) = P(N_t \geq 1) = 1 - P(N_t < 1)$$
$$= 1 - e^{-\lambda t}$$

Z čehož dále plyne vztah pro hustotu

$f(t) = \lambda e^{-\lambda t}; t \geq 0$

$E(T) = \int_{t=0}^{\infty} \lambda t\, e^{-\lambda t}\, dt = \frac{1}{\lambda}$

$DT = ET^2 - (ET)^2 = \ldots = \frac{1}{\lambda^2}$

Hazardní funkce - Intenzita poruch

- **Odvození intenzity (hazardní funkce) poruch exponenciálního rozdělení**

➢ Intenzitu poruch odvodíme z definičního vztahu:

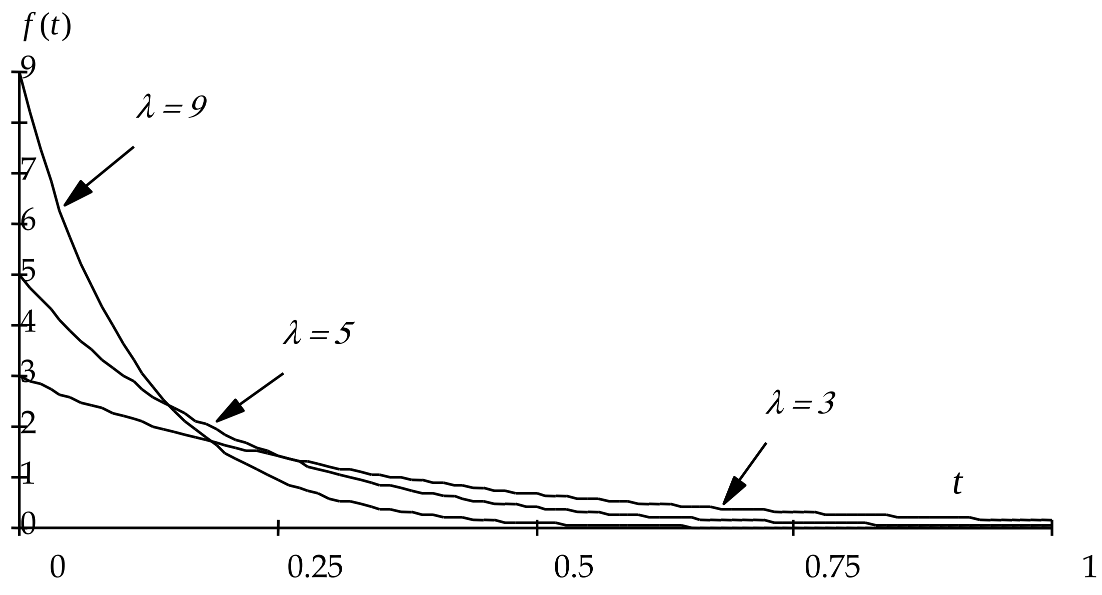
$$\lambda(t) = \frac{f(t)}{1 - F(t)}; \qquad F(t) < 1,\, tj.\, t > 0$$

$$\lambda(t) = \frac{\lambda \cdot e^{-\lambda t}}{1 - \left(1 - e^{-\lambda t}\right)} = \lambda = \frac{1}{EX}$$

λ = *konst.* $\Rightarrow$ charakteristická vlastnost exponenciálního rozdělení, která bývá interpretována jako skutečnost, že toto rozdělení „nemá paměť “.

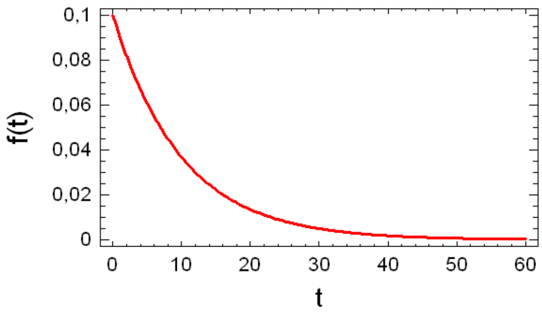
Příklad

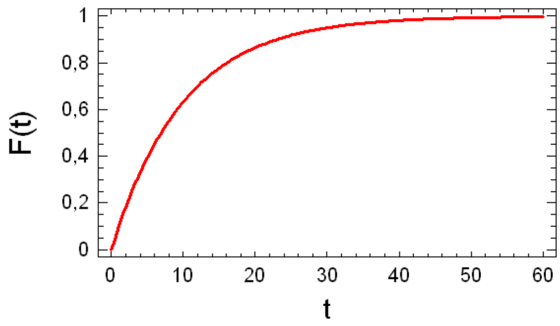
Následující graf ilustruje některé příklady hustoty pravděpodobnosti pro různé hodnoty parametru λ. Stojí za povšimnutí, že tvar hustoty je podobný jako tvar pravděpodobnostní funkce geometrického rozdělení. Exponenciální rozdělení je spojitým ekvivalentem diskrétního geometrického rozdělení pravděpodobnosti.



Hustota pravděpodobnosti exponenciálního rozdělení pravděpodobnosti

Grafické znázornění hustoty pravděpodobnosti a distribuční funkce exponenciálního rozdělení





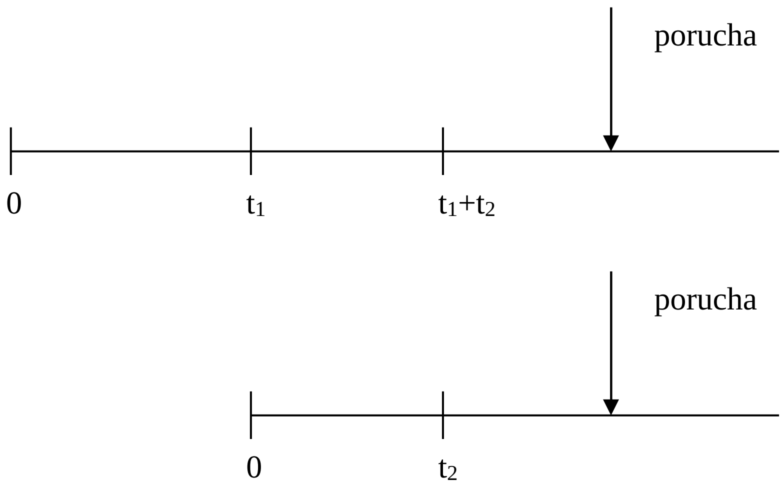
Exponenciální rozdělení = „rozdělení bez paměti“

Exponenciální rozdělení bývá někdy nazýváno "**rozdělení bez paměti**". Tento název znamená, že:

$$P\left(X > (t_1 + t_2) \middle| X > t_1\right) = \mathrm{P}(\mathrm{X} > \mathrm{t}_2); \qquad \mathrm{t}_1; \mathrm{t}_2 \geq 0$$

Co si představit pod tímto vztahem?

Považujme exponenciální náhodnou veličinu X za dobu do poruchy nějakého zařízení. Pak pravděpodobnost, že zařízení, které pracovalo bez poruchy po dobu t_1, bude pracovat bez poruchy ještě alespoň po dobu t_2, je rovna pravděpodobnosti, že zařízení, které dosud nebylo v provozu, bude pracovat alespoň po dobu t_2.

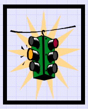


X … doba do poruchy

Zdá se jako by toto zařízení "zapomnělo" na dříve odpracovanou dobu. (Považovali-li bychom dobu do poruchy Vašeho monitoru za exponenciální náhodnou veličinu, pak pravděpodobnost, že se Váš monitor porouchá za více než 200hodin od této chvíle, by nijak nezávisela na jeho stáří (době jeho předcházejícího provozu)).

Tato vlastnost vysvětluje použití exponenciálního rozdělení v teorii spolehlivosti. **Exponenciální rozdělení popisuje dobře rozdělení doby života zařízení, u kterých dochází k poruše ze zcela náhodných příčin a nikoliv v důsledku opotřebení** (mechanické opotřebení, únava materiálu apod.). Zároveň tato vlastnost exponenciálního rozdělení vysvětluje proč je jeho intenzita poruch konstantní (není závislá na délce předcházejícího provozu zařízení).

Má-li doba do výskytu události exponenciální rozdělení, pak informace o tom, že událost nenastala po dobu t_1, nemění pravděpodobnost výskytu události v následujícím období délky t_2.



Řešený příklad:

Výrobce žárovky XX ví, že průměrná životnost žárovek XX je 10.000 h. V rámci své propagační kampaně chce garantovat dobu T, do níž se nespálí více než 3% žárovek. Určete tuto dobu.

Řešení:

X ... životnost žárovky (doba do poruchy) má exponenciální rozdělení

$$X \to E(\lambda)$$

- Určíme parametr λ:

$$\begin{aligned} EX &= \frac{1}{\lambda} \\ EX &= 10.000\,h \end{aligned} \quad \Rightarrow \quad \lambda = 10^{-4}\,h^{-1}$$

- Na základě zadané pravděpodobnosti nejdeme dobu T:

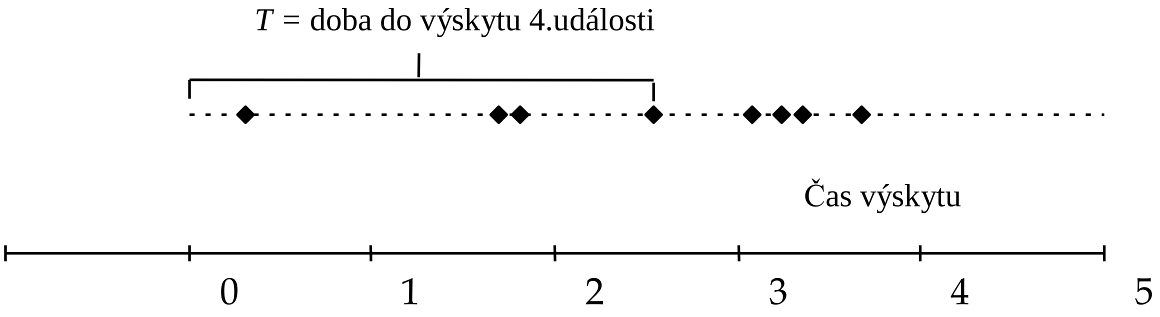
$$\begin{aligned} P(X < T) &\le 0{,}03 \\ F(T) &\le 0{,}03 \\ 1 - e^{-\lambda T} &\le 0{,}03 \\ 0{,}97 &\le e^{-\lambda T} \\ -\frac{\ln(0{,}97)}{\lambda} &\ge T \\ T \le -10^4 \cdot \ln(0{,}97) &\Rightarrow \underline{\underline{T \cong 304\,h}} \end{aligned}$$

Výrobce může tvrdit, že více než 97% žárovek má životnost delší než 304 hodin.

Gamma rozdělení

$T \rightarrow Ga\,(k, \lambda)$

Popisuje náhodnou veličinu T ... dobu do výskytu k-té události v Poissonově procesu.
Nechť k = 4:



T má Gamma rozdělení

Náhodnou veličinu s Gamma rozdělením si můžeme představit jako součet k nezávislých exponenciálních náhodných veličin. Jestliže exponenciální rozdělení je spojitým ekvivalentem diskrétního geometrického rozdělení, pak Gamma rozdělení lze chápat jako spojitý ekvivalent diskrétního NB rozdělení.

Odvození Gamma rozdělení:

$\{N_t \geq k\} \Leftrightarrow \{(T_k = \sum_{i=1}^{k} X_i) < t\}$ ekvivalence dvou jevů, $N_t \rightarrow P(\lambda t)$

$T_k = \sum_{i=1}^{k} X_i$ .... Gama náhodná veličina = součet k dob do výskytu první události (tedy součet k exponenciálních náhodných veličin).

Z výše uvedené ekvivalence lze odvodit distribuční funkci tohoto rozdělení:
$F(t) = P(T_k < t) =$

$$P\left(\sum_{i=1}^{k} X_i < t\right) = P(N_t \geq k) = 1 - P(N_t < k) = 1 - \sum_{j=0}^{k-1} e^{-\lambda t} \cdot \frac{(\lambda t)^j}{j!} \quad = 1 - e^{-\lambda t}\left[\sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!}\right]$$

Derivací $F(t)$ dostaneme funkci hustoty pravděpodobnosti:

$$f(t) = \lambda e^{-\lambda t}\left[\sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!}\right] - e^{-\lambda t}\left[\sum_{j=1}^{k-1} \frac{\lambda(\lambda t)^{j-1}}{(j-1)!}\right]$$

$$= \lambda^k e^{-\lambda t}\left[\frac{t^{k-1}}{(k-1)!}\right]; \quad t > 0$$

$T_k = X_1 + X_2 + X_3 + \ldots + X_k$
$X_i \rightarrow E(\lambda)$

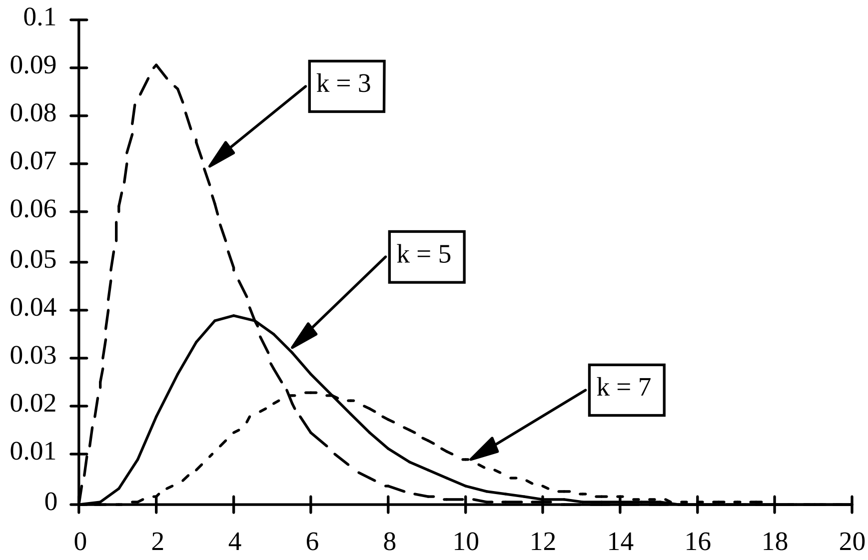
$$ET_k = EX_1 + EX_2 + \ldots + EX_k = \frac{1}{\lambda} + \ldots + \frac{1}{\lambda} = \frac{k}{\lambda}$$ ... střední (očekávaná) hodnota T_k

$$DT_k = \ldots . = \frac{k}{\lambda^2}$$ .... rozptyl T_k

Pokud *k* je celočíselné, toto rozdělení mívá označení jako **Erlangovo**.

Příklad

Na obrázku jsou příklady hustoty Gamma rozdělení pro $\lambda = 1$ a různé hodnoty *k*. Poznamenejme, že s rostoucím *k* roste jednak rozptyl tohoto rozdělení a navíc rozdělení je více symetrické.



Hustota pravděpodobnosti pro Gamma rozdělení při uvedených hodnotách *k*

Hazardní funkce – intenzita poruch

- **Odvození intenzity poruch**

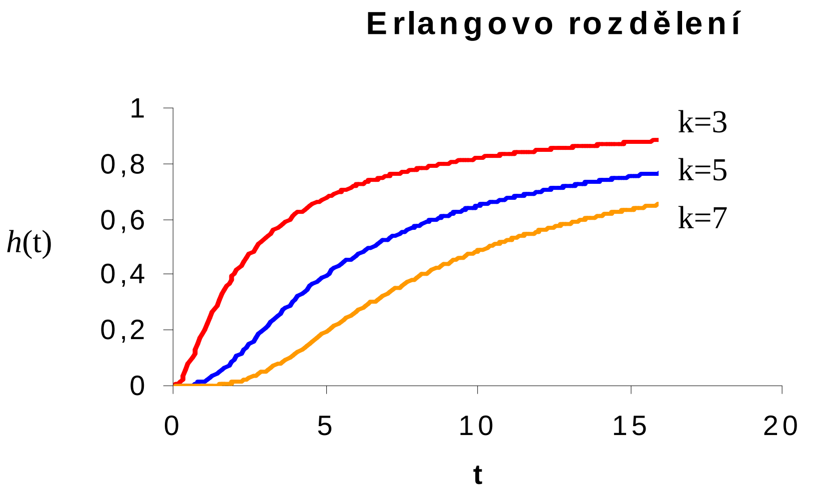
$$h(t) = \frac{f(t)}{1 - F(t)}$$

$$= \frac{\lambda \cdot e^{-\lambda t} \cdot \frac{(\lambda t)^{k-1}}{(k-1)!}}{e^{-\lambda t} \cdot \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!}} = \frac{\lambda}{(k-1)! \cdot \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{(\lambda t)^{k-1} \cdot j!}} = \frac{\lambda}{(k-1)! \cdot \sum_{j=0}^{k-1} \frac{(\lambda t)^{j-k+1}}{j!}} = \frac{\lambda}{(k-1)! \cdot \sum_{j=0}^{k-1} \frac{1}{(\lambda t)^{k-1-j} \cdot j!}} =$$

$$= \frac{\lambda}{(k-1)! \cdot \sum_{j=0}^{k-1} \frac{1}{(\lambda t)^j \cdot (k-1-j)!}}$$

$$h(t) = \frac{\lambda}{(k-1)!\sum_{j=0}^{k-1}\frac{1}{(k-1-j)!\,(\lambda t)^{j}}}$$

Graf intenzity poruch Erlangova rozdělení pro λ = 1; k = 3; 5; 7



Intenzita poruch *h*(t) je v případě Erlangova rozdělení rostoucí funkce a proto je toto **rozdělení vhodné pro modelování procesů stárnutí**.
h(t) je ostře rostoucí pro k > 1 ⇒ rozdělení je vhodné pro modelování procesů stárnutí a opotřebení

Weibullovo rozdělení

Weibullovo rozdělení je velmi flexibilní (díky parametru β) a proto se jím zejména v teorii spolehlivosti popisují spojité náhodné veličiny definované jako **doba do poruchy** (doba bezporuchovosti). Používá se zejména při popisu komponent, které jsou **v období ranných poruch nebo v období stárnutí** (tj. tam kde se projevuje mechanické opotřebení nebo únava materiálu).

Weibullovo rozdělení má dva parametry: **Θ** – parametr měřítka (scale, Θ > 0, závisí na materiálu, namáhání a podmínkách užívání) a **β** – parametr tvaru (shape, β > 0, na jeho hodnotě závisí tvar intenzity poruch a tím i vhodnost použití pro určité období doby života).

Má-li náhodná veličina X Weibullovo rozdělení, značíme to takto:

$$X \rightarrow W(\Theta, \beta)$$

Distribuční funkce:

$$F(t) = 1 - e^{-\left(\frac{t}{\Theta}\right)^{\beta}}; \qquad t > 0; \Theta > 0; \beta > 0$$

Hustota pravděpodobnosti:

$$f(t) = \frac{\beta}{\Theta}\left(\frac{t}{\Theta}\right)^{\beta-1} e^{-\left(\frac{t}{\Theta}\right)^{\beta}}; \qquad t > 0; \Theta > 0; \beta > 0$$

Intenzita poruch (Hazardní funkce):

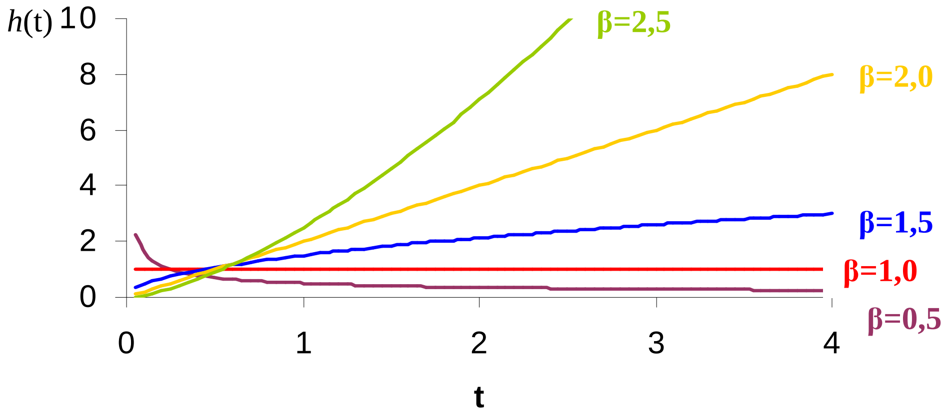
$$h(t) = \frac{\beta}{\Theta}\cdot\left(\frac{t}{\Theta}\right)^{\beta-1}; \qquad t > 0; \Theta > 0; \beta > 0$$

Ze vztahu pro intenzitu poruch Weibullova rozdělení je zřejmé, že:

$$h(t) = konst.\cdot t^{\beta-1}$$

a proto tvar intenzity poruch závisí na volbě parametru β.

Některé příklady intenzity poruch Weibullova rozdělení (Θ=1):



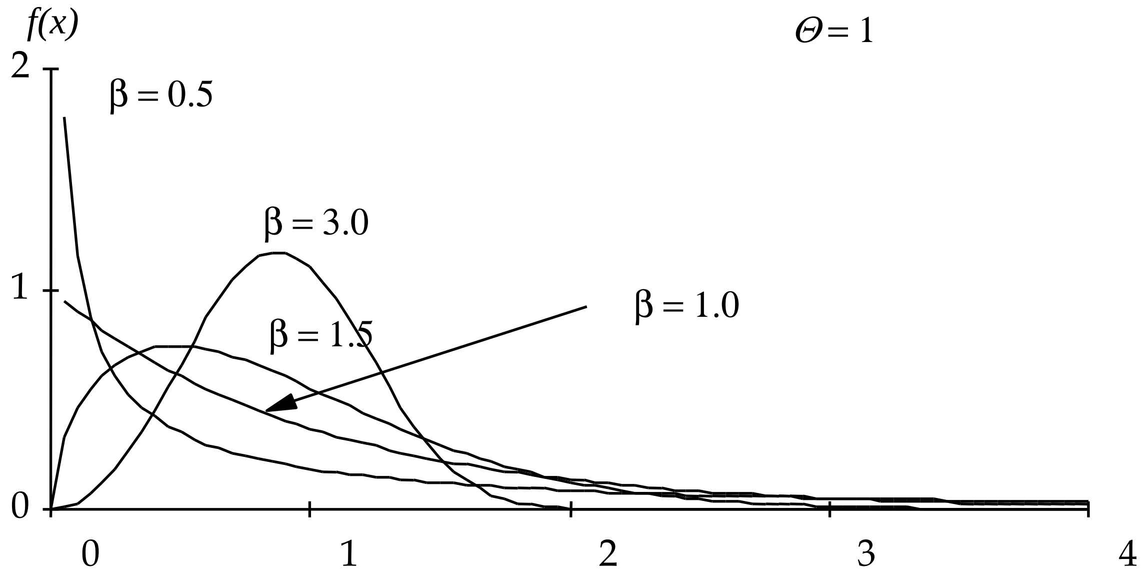
Všimněme si, že pro β=1, přejde Weibullovo rozdělení v rozdělení exponenciální (konstantní intenzita poruch) s parametrem $\lambda = \frac{1}{\Theta}$.

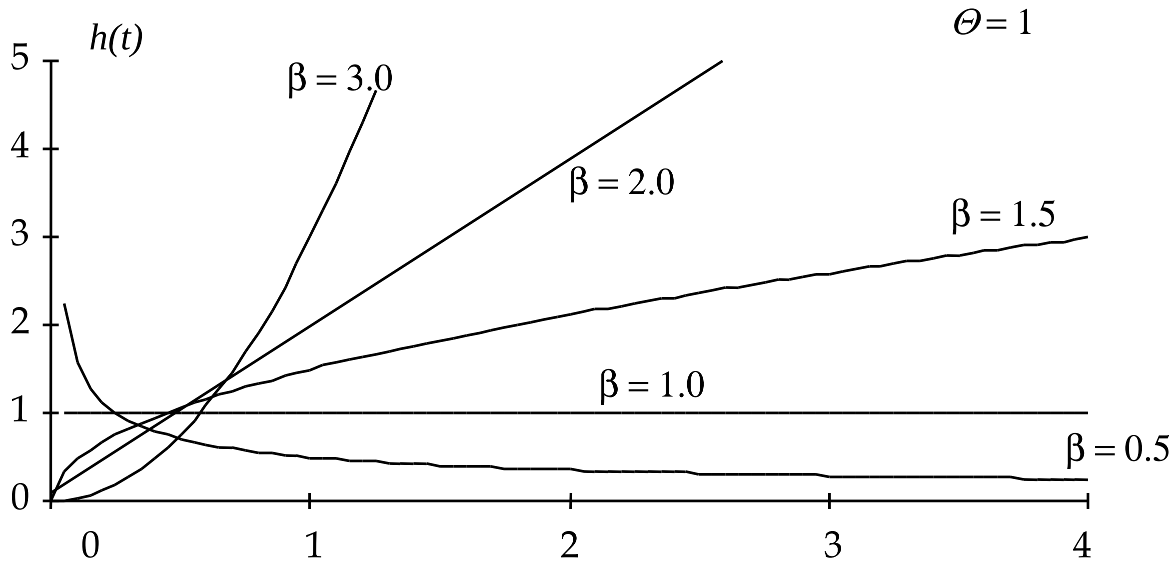
$$\beta = 1 \quad \Rightarrow \quad W(\Theta;1) \to E\left(\frac{1}{\Theta}\right)$$

Z výše uvedeného grafu je rovněž zřejmé použití Weibullova rozdělení v závislosti na parametru β:

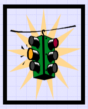
$0 < \beta < 1$	období dětských nemocí	*h*(t) ... klesající funkce
$\beta = 1$	období stabilního života	$h(t) = konst. = \frac{1}{\Theta} = \lambda$ (exp. rozdělení)
$1 < \beta < 2$	období stárnutí	*h*(t) ... konvexní, rostoucí funkce
$\beta = 2$	období stárnutí	*h*(t) ... lineárně rostoucí funkce
$\beta > 2$	období stárnutí	*h*(t) ... konkávní, rostoucí funkce

Další příklady hustoty pravděpodobnosti a intenzity poruch Weibullova rozdělení:





Weibullovo rozdělení je velmi flexibilní a používá se zejména v teorii spolehlivosti pro modelování náhodné veličiny „doby do poruchy.



Řešený příklad:

Předpokládejme, že doba do poruchy určitého systému je modelována Weibullovým rozdělením s lineárně rostoucí intenzitou poruch. (Θ =50)

a) Jaká je intenzita poruch systému po deseti hodinách funkce?
b) Jaká je pravděpodobnost, že systém bude pracovat bez poruchy během počátečních 100 hodin?

Řešení:

$$\text{X ... doba do poruchy, } (X \to W(50;\beta))$$

Hodnotu parametru β určíme na základě poznámky, že intenzita poruch je lineárně rostoucí. Obecný tvar intenzity poruch Weibullova rozdělení je:

$$\lambda(t) = \frac{\beta}{\Theta} \cdot \left(\frac{t}{\Theta}\right)^{\beta-1}; \qquad t > 0; \Theta > 0; \beta > 0$$

_z čehož vyplývá, že β = 2.

$$X \to W(50;2)$$

ada) Hledanou intenzitu poruch určíme dosazením do obecného vztahu:

$$\underline{\underline{\lambda(10)}} = \frac{2}{50} \cdot \left(\frac{10}{50}\right)^{2-1} = \underline{\underline{0{,}008}}$$

Intenzita poruch daného systému je po 10 hodinách provozu 0,008. Tj. pokud byl systém po 10 hodin bezporuchový, pak pravděpodobnost, že v následujícím velmi krátkém časovém intervalu Δt dojde k poruše, je 0,008 .Δt.

adb) Pravděpodobnost, že systém bude prvních 100 hodin bezporuchový určíme přes jev opačný, jehož pravděpodobnost udává distribuční funkce.

$$F(t) = 1 - e^{-\left(\frac{t}{\Theta}\right)^{\beta}}; \qquad t > 0; \Theta > 0; \beta > 0$$

$$\underline{\underline{P(X > 100)}} = 1 - F(100) = 1 - \left[1 - e^{-\left(\frac{100}{50}\right)^{2}}\right] = e^{-\left(\frac{100}{50}\right)^{2}} = e^{-4} = \underline{\underline{0{,}018}}$$

Pravděpodobnost, že daný systém bude prvních 100 hodin bezporuchový je 1,8%.

Σ Shrnutí pojmů

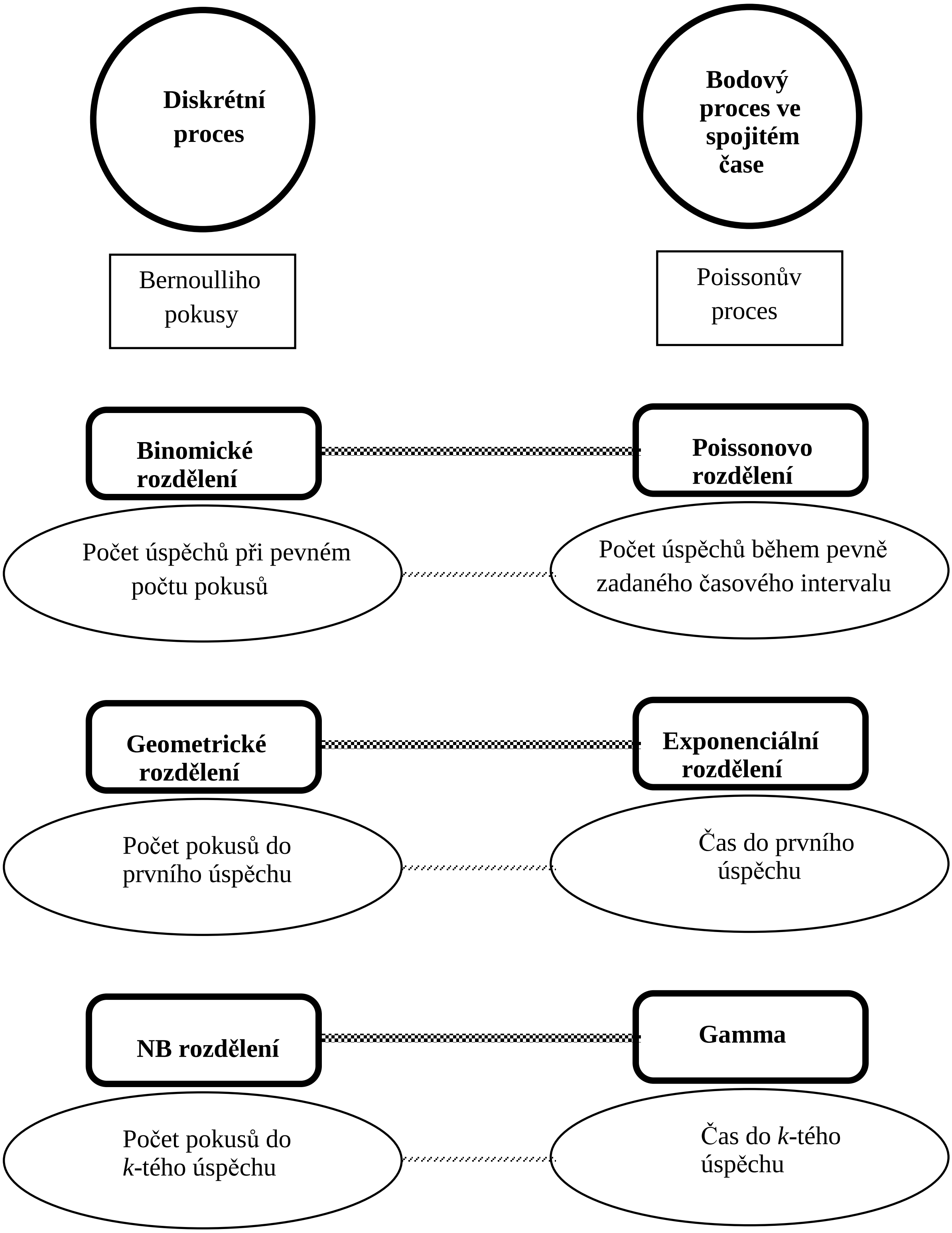
Posloupnost **Bernoulliho pokusů** je posloupnost nezávislých pokusů majících pouze 2 možné výsledky (událost nastane-nenastane; úspěch-neúspěch; popřípadě 1-0) a pravděpodobnost výskytu události (úspěchu) *p* je konstantní v každém pokuse. Na základě předpokladů těchto pokusů lze definovat následující náhodné veličiny: **binomickou**, **geometrickou** a **negativně binomickou**.

Počet výskytů událostí na nějakém deterministickém intervalu od 0 do t je za jistých předpokladů popsán **Poissonovým** rozdělením pravděpodobnosti. Náhodná veličina *T* ... doba do výskytu první události nebo doba mezi dvěma po sobě jdoucími událostmi v Poissonově procesu je **spojitá** náhodná veličina, která je zpravidla popsána **exponenciálním** rozdělením pravděpodobnosti. Doba do výskytu k-té události je popsána G**amma** rozdělením. **Weibullovo** rozdělení pravděpodobnosti je zobecněním exponenciálního rozdělení, které je velmi flexibilní.

? Otázky

1. Jaká diskrétní a spojitá rozdělení pravděpodobnosti znáte ?
2. Charakterizujte Bernoulliho pokusy a z nich odvozené jednotlivé typy diskrétních rozdělení: binomické, geometrické, negativně binomické. Vypočtěte střední hodnotu binomické náhodné veličiny.
3. K čemu slouží Gamma rozdělení a jak souvisí s exponenciálním rozdělením ?
4. Jakou hodnotu musí mít parametr tvaru Weibullova rozdělení, aby intenzita poruch byla lineárně rostoucí ?
5. Proveďte komentář k následujícímu schématu. Popište vztah mezi uvedenými rozděleními pravděpodobnosti.

Schéma: vzájemná souvislost mezi rozděleními



Úlohy k řešení

Př. 1: Pravděpodobnost úspěchu je 0.1. Určete pravděpodobnost, že do prvního úspěchu provedeme:

a) méně než 5 pokusů
b) více než 10 pokusů
c) mezi 6 a 8 pokusy
d) právě 7 pokusů.

{a) 0.344; b) 0.349; c) 0.160; d) 0.053}

Př. 2: Víme, že pravděpodobnost vady výrobku je 17%. Určete pravděpodobnost, že mezi 20 výrobky bude:

a) více než 5 vadných výrobků
b) méně než dva vadné výrobky
c) mezi 4 a 8 vadnými výrobky
d) právě 3 vadné výrobky

{a) 0.110; b) 0.123; c) 0.446; d) 0.236}

Př. 3: Předpokládejme, že pravděpodobnost výskytu všech krevních skupin je stejná. Jaká je pravděpodobnost, že proto abychom našli 3 dárce se skupinou 0 budeme muset vyšetřit:

a) více než 20 dárců krve
b) méně než 12 dárců krve
c) mezi 14 a 18 dárci krve
d) právě 15 dárců krve

{a) 0.091; b) 0.545; c) 0.197; d) 0.045}

Př. 4: Kolikrát (průměrně) musíme hodit mincí, aby nám 5x padl lev?

{10}

Př. 5: Továrna produkuje integrované obvody XX. Při jedné fázi výroby dochází často k závadě, proto je 25% výrobků vadných. Jaká je pravděpodobnost, že mezi 12 integrovanými obvody budou:

a) 4 vadné
b) méně než 4 vadné
c) Jaká je střední hodnota a rozptyl počtu vadných IO, budeme-li testovat 15 vzorků?
d) Nyní uvažme, že bylo vyrobeno pouze 48 IO a my vybereme 12 z nich. Jaká je nyní pravděpodobnost, že mezi vybranými IO budou právě 4 vadné?

{a) 0.19; b) 0.65; c) EX=3.75, DX=2.81; d) 0.22}

Př. 6: Distributor prodává knihu XY po telefonu. 12% hovorů je úspěšných (tj. objednají si knihu). Jaká je pravděpodobnost, že distributor předtím než bude úspěšný bude muset uskutečnit:

a) 5 hovorů
b) méně než 5 hovorů
c) více než 8 hovorů

Předpokládejme, že distributor musí splnit denní kvótu - prodat 10 knih.

d) Jaká je pravděpodobnost, že distributor bude pro splnění denní kvóty potřebovat méně než 30 telefonátů?
e) Určete střední hodnotu a rozptyl počtu telefonátů potřebných pro splnění denní kvóty.

Uvažme nyní, že ne každý z těch, kdo si telefonicky objednají danou knihu, ji skutečně odebere. Přesněji řečeno - 65% osob objednanou knihu skutečně zaplatí. Distributor je podle této skutečnosti ohodnocen. Dostává 30,- Kč za každou objednávku a dalších 50,- Kč ve chvíli, kdy je objednávka převzata.

f) Jaká je pravděpodobnost, že výdělek distributora ve chvíli, kdy splní svou denní kvótu, bude vyšší než 500,- Kč?

g) Jaký je jeho průměrný výdělek (a rozptyl jeho výdělku) při splnění denní kvóty?

{a) 0.06; b) 0.47; c) 0.32; d) 0.001; e) EX = 83.33, DX = 611.11; f) 0.91; g) (625 $\pm$ 75)Kč}

Př. 7: Celník na hranici se Slovenskem má za úkol kontrolovat projíždějící vozidla. Víme, že 25% vozidel veze kontraband a 40% z nich celník odhalí. Jaká je pravděpodobnost, že celník, předtím než objeví první vozidlo s kontrabandem, bude muset prohlédnout:

a) 5 aut
b) více než 10 aut
c) Určete střední hodnotu a rozptyl počtu aut, jenž musí celník prohlédnout předtím než objeví první automobil s kontrabandem.

Nadřízený tohoto celníka vydal příkaz, že celník může jít domů poté co nalezne 5 aut s kontrabandem. Předpokládejme, že prohlédnutí jednoho auta trvá celníkovi 10 minut.

d) Jaká je pravděpodobnost, že tento příkaz prodlouží celníkovi pracovní den (8 hodin) ?
e) Jaká je nyní průměrná pracovní doba (a její rozptyl) celníka?

{a) 0.059; b) 0.31; c) EX = 10, DX = 90; d) 0.47; e) EY = 8h 20min, σ_Y = 3h 32min}

Př. 8: Bankovní úředník provádějící kontrolu návrhů půjček zjistil, že se v nich nachází 0.5 chyby na návrh. Jaká je pravděpodobnost, že úředník najde v deseti návrzích:

a) 6 chyb
b) více než 6 chyb
c) ani jednu chybu

V 35% chyb je nutno chybu přičíst úmyslné chybné prezentaci dat.

d) Jaký je průměrný počet chyb způsobených chybnou prezentací v celkovém množství 100 návrhů ?
e) Pokud všechny chybné návrhy vyřadíme, jaká je pravděpodobnost, že více než 2 návrhy z deseti budou vyřazeny vlivem úmyslné chybné prezentace dat ?

{binom. NV - a) 0.21; b) 0.17; c) 0.001; d)17.5; e) 0.74}

Př. 9: Počet návštěvníků Fitness Centra VŠB je v průměru 10 na hodinu. Určete:

a) Pravděpodobnost, že v určitou hodinu je ve Fitcentru přesně 10 lidí
b) Pravděpodobnost, že v určitou hodinu je ve Fitcentru méně než 5 lidí
c) Pravděpodobnost, že v určitou hodinu je ve Fitcentru mezi 8 a 15 osobami

{a) 0.125; b) 0.029; c) 0.731}

Př. 10: Bankovní úředník provádějící kontrolu návrhů půjček zjistil, že se v nich nachází 0.5 chyby na návrh. Jaká je pravděpodobnost, že úředník najde v deseti návrzích:

a) 6 chyb
b) více než 6 chyb
c) ani jednu chybu

V 35% chyb je nutno chybu přičíst úmyslné chybné prezentaci dat.

d) Jaký je průměrný počet chyb způsobených chybnou prezentací v celkovém množství 100 návrhů ?
e) Pokud všechny chybné návrhy vyřadíme, jaká je pravděpodobnost, že více než 2 návrhy z deseti budou vyřazeny vlivem úmyslné chybné prezentace dat ?

{poisson. NV - a) 0.146; b) 0.238; c) 0.007;(srovnej s 6.9); d) 17.5; e) 0.738}

Př. 11: Každá tramvaj DPMO Ostrava je jednou měsíčně podrobena technické kontrole, přičemž doba prohlídky je samozřejmě silně závislá na typu závady. Předpokládejme, že doba prohlídky tramvaje má exponenciální rozdělení se střední hodnotou 3 hodiny. Určete:

a) Pravděpodobnost, že doba prohlídky tramvaje bude kratší než 2 hodiny
b) Pravděpodobnost, že prohlídka tramvaje bude trvat déle než 4 hodiny
c) Pravděpodobnost, že mechanici během osmihodinového pracovního dne prohlédnou právě 3 tramvaje
d) Pravděpodobnost, že mechanici během osmihodinového pracovního dne prohlédnou méně než 3 tramvaje

{a) 0.486; b) 0.264; c) 0.220; d) 0.502}

Př. 12: Předpokládejme, že doba do poruchy určitého systému je modelována Weibullovým rozdělením s klesající intenzitou poruch, parametry: $\lambda = 0.02$; $\beta = 0.5$.

a) Jaká je intenzita poruch systémů po deseti hodinách funkce?
b) Jaká je intenzita poruch systémů po 200 hodinách funkce?
c) Jaká je pravděpodobnost, že systém bude pracovat bez poruchy během prvních 10 hodin?
d) Jaká je pravděpodobnost, že systém bude pracovat bez poruchy během prvních 200 hodin?

{a) h(10) = 0.022; b) h(200) = 0.005; c) 0.905; c) 0.135}

Další spojitá rozdělení pravděpodobnosti

Rovnoměrné rozdělení

Již dříve, v některém z předchozích řešených příkladů, jsme se setkali s rovnoměrným (rektangulárním) rozdělením. Jde o rozdělení, jehož hustota pravděpodobnosti je konstantní na nějakém intervalu $\langle a;b\rangle$ a všude jinde je nulová.

X … náhodná veličina s rovnoměrným rozdělením na intervalu $\langle a;b\rangle$

$$X \to R(a;b)$$

Hustota pravděpodobnosti:

$$f(x) = \begin{matrix} \dfrac{1}{b-a} & x \in \langle a;b\rangle \\ 0 & jinde \end{matrix}$$

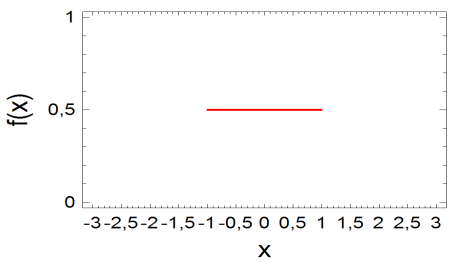
Distribuční funkce:

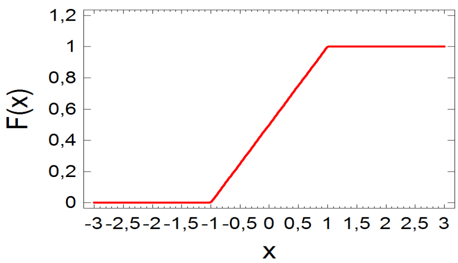
$$F(x) = \begin{matrix} 0 & x \in (-\infty;a) \\ \dfrac{x-a}{b-a} & x \in \langle a;b\rangle \\ 1 & x \in (b;\infty) \end{matrix}$$

Střední hodnota: $EX = \dfrac{a+b}{2}$

Rozptyl: $DX = \dfrac{(a-b)^2}{12}$

Grafické znázornění hustoty pravděpodobnosti a distribuční funkce rovnoměrného rozdělení na intervalu $\langle -1;1\rangle$





Průvodce studiem:

- **Jak jsme přišli na to, že hustota pravděpodobnosti rovnoměrného rozdělení je definována jako**:

$$f(x) = \begin{matrix} \dfrac{1}{b-a} & x \in \langle a;\mathrm{b} \rangle \\ 0 & jinde \end{matrix} \qquad ?$$

Odvození:

Uvedli jsme si, že rovnoměrné rozdělení na intervalu $\langle a;b \rangle$ je takové, jehož hustota je konstantní na daném intervalu a všude jinde je nulová.

Z toho vyplývá, že vztah pro hustotu pravděpodobnosti můžeme zapsat ve tvaru:

$$f(x) = \begin{matrix} c & x \in \langle a;\mathrm{b} \rangle \\ 0 & jinde \end{matrix}, \qquad \text{kde } c \in R$$

Zbývá nám nalézt konstantu c:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_a^b c dx = [cx]_a^b = c(b-a)$$

$$c(b-a) = 1$$

$$\underline{\underline{c = \frac{1}{b-a}}}$$

A proto:

$$f(x) = \begin{matrix} \dfrac{1}{b-a} & x \in \langle a;\mathrm{b} \rangle \\ 0 & jinde \end{matrix}$$

- **Odvození distribuční funkce rovnoměrného rozdělení**

$$x \in (-\infty; a) \qquad F(x) = \int_{-\infty}^{x} 0 dt = 0$$

$$x \in \langle a;b \rangle \qquad F(x) = \int_{-\infty}^{a} 0 dt + \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}$$

$$x \in (b;\infty) \qquad F(x) = \int_{-\infty}^{a} 0 dt + \int_a^b \frac{1}{b-a} dt + \int_b^{\infty} 0 dt = \frac{b-a}{b-a} = 1$$

Výklad:

- **Odvození střední hodnoty a rozptylu:**

$$EX = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

$$\underline{\underline{EX}} = \int_{a}^{b} x \cdot \frac{1}{b-a}dx = \frac{1}{b-a} \cdot \left[\frac{x^2}{2}\right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \underline{\underline{\frac{a+b}{2}}}$$

$$EX^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx$$

$$EX^2 = \int_{a}^{n} x^2 \cdot \frac{1}{b-a}dx = \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}$$

$$DX = EX^2 - (EX)^2$$

$$\underline{\underline{DX}} = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} = \frac{a^2 - 2ab + b^2}{12} =$$

$$= \underline{\underline{\frac{(a-b)^2}{12}}}$$

Normální rozdělení

Normální rozdělení je nejdůležitějším pravděpodobnostním rozdělením, popisujícím chování velkého množství náhodných jevů v technice, přírodních vědách i ekonomii.

Klasickým příkladem tohoto rozdělení je rozdělení náhodných chyb vzniklých při měření nějaké veličiny (tažnosti 0,5 palcových trubek). Při opakovaném měření téže veličiny za stejných podmínek způsobují náhodné (neovlivnitelné) vlivy odchylky od skutečné hodnoty měřené veličiny.

Lze říci, že **normální rozdělení je vhodným pravděpodobnostním modelem tehdy, působí-li na kolísání náhodné veličiny velký počet nepatrných a vzájemně nezávislých vlivů**.

Značný význam normálního rozdělení spočívá rovněž v tom, že **za určitých podmínek lze pomocí něj aproximovat řadu jiných spojitých i nespojitých rozdělení**.

Normální rozdělení má dva parametry: $\boldsymbol{\mu}$ – střední hodnotu, charakterizující polohu tohoto rozdělení a $\boldsymbol{\sigma^2}$ – rozptyl, charakterizující rozptýlení hodnot kolem střední hodnoty.

POZOR!
V anglosaské literatuře (a v některých statistických paketech) jsou jako parametry normálního rozdělení uváděny střední hodnota $\boldsymbol{\mu}$ a směrodatná odchylka $\boldsymbol{\sigma}$.

Normální rozdělení (hustota pravděpodobnosti) je jednomodální rozdělení, symetrické kolem střední hodnoty μ. Střední hodnota je rovna modu a mediánu. Náhodná veličina X, jež se rímto rozdělením řídí, může nabývat libovolné hodnoty z R. Křivka hustoty pravděpodobnosti (Gaussova křivka) má zvonovitý tvar s maximem ve střední hodnotě a „šířkou“ úměrnou směrodatné odchylce.
To, že se náhodná veličina X řídí normálním rozdělením se střední hodnotou μ a rozptylem σ^2 zapisujeme:

$$X \to N\left(\mu; \sigma^2\right)$$

Hustota pravděpodobnosti:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2}; \qquad -\infty < x < \infty$$

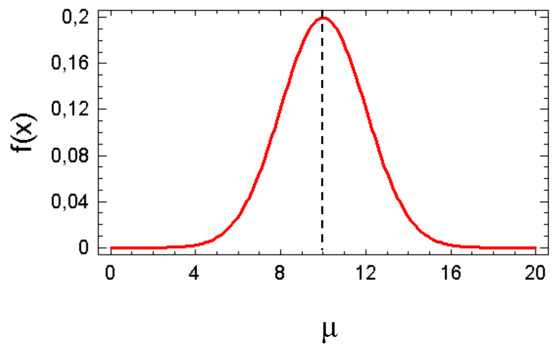
Distribuční funkce:

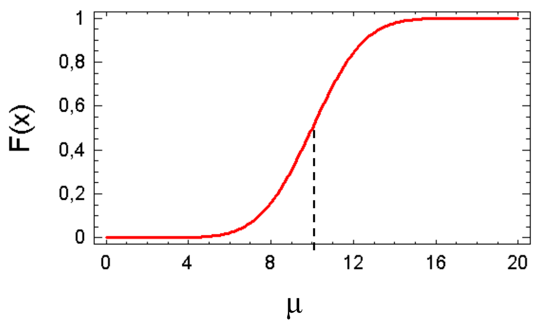
$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)^2} dt$$

Střední hodnota: $EX = \mu$

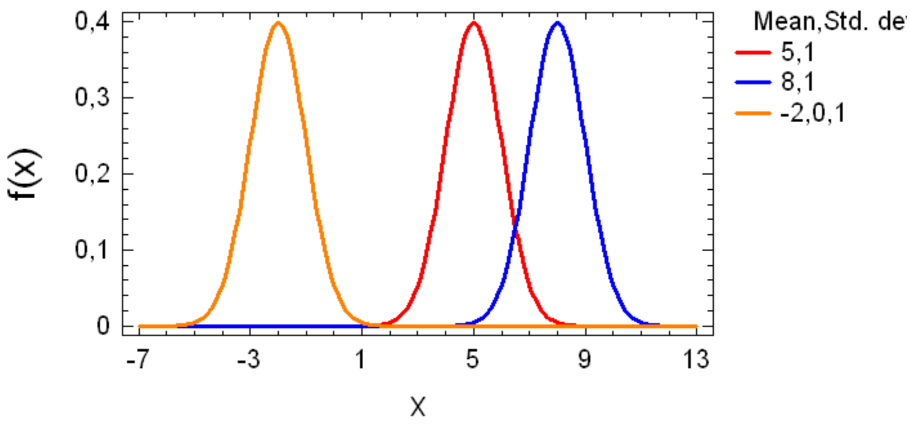
Rozptyl: $DX = \sigma^2$

Grafické znázornění hustoty pravděpodobnosti a distribuční funkce:

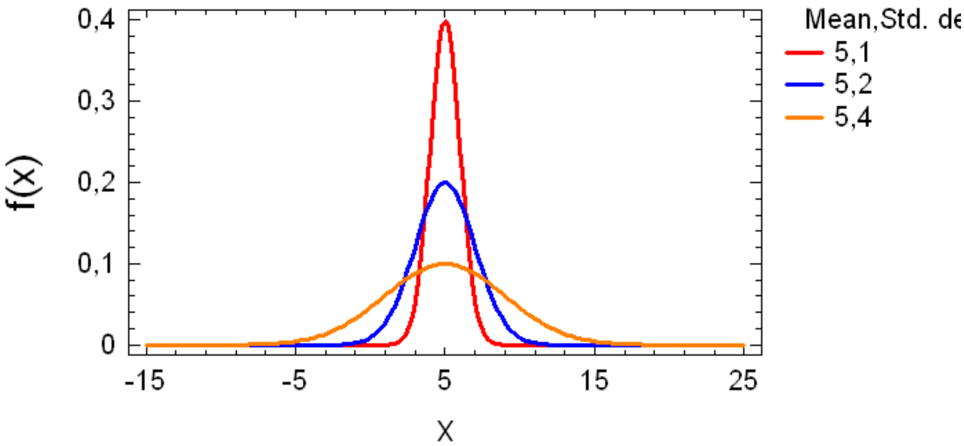




Vliv μ na křivku hustoty pravděpodobnosti



Vliv σ na křivku hustoty pravděpodobnosti



Výpočet distribuční funkce je analyticky nemožný a proto se využívá možnosti vyjádřit distribuční funkci normální náhodné veličiny pomocí distribuční funkce normované náhodné veličiny, tj. normální náhodné veličiny s parametry $\mu = 0$, $\sigma^2 = 1$. Distribuční funkce normované náhodné veličiny je přitom tabelována (případně spočteme pomocí vhodného software).

Normované (standardizované) normální rozdělení

Jak již jsme se zmínili, jde o speciální typ normálního rozdělení se střední hodnotou rovnou nule a jednotkovým rozptylem.

To, že má náhodná veličina Z (obvyklé značení pro tuto náhodnou veličinu) normované normální rozdělení, značíme:

$$Z \rightarrow N(0;1)$$

Důležitost tohoto rozdělení ukazuje i nestandardní značení pro distribuční funkci (Φ(x)) a hustotu pravděpodobnosti (φ(x)).

Hustota pravděpodobnosti:

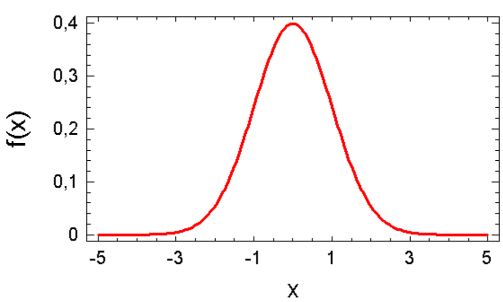
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}; \qquad -\infty < x < \infty$$

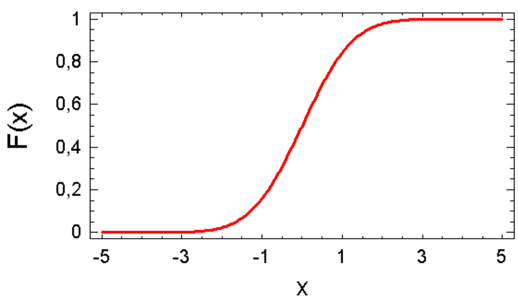
Distribuční funkce:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dt$$

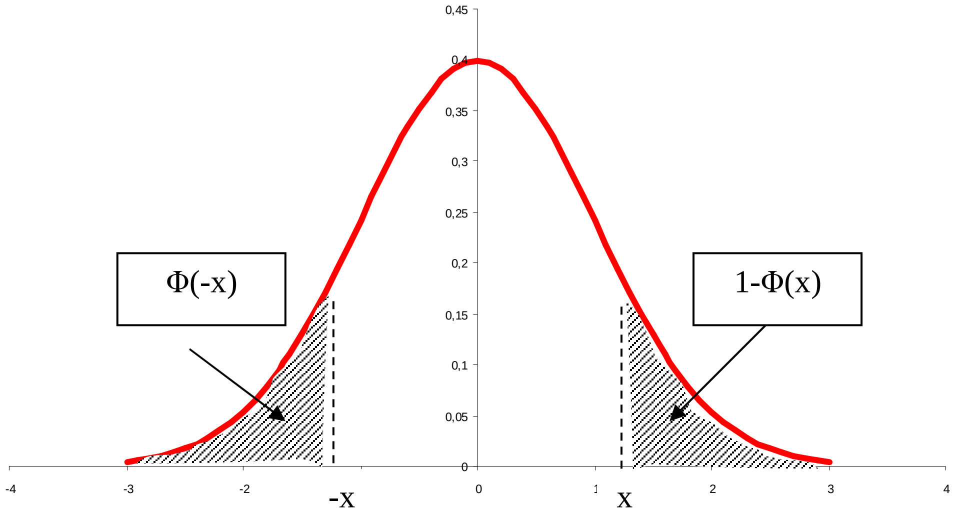
Střední hodnota: $EZ = 0$
Rozptyl: $DZ = 1$

Grafické znázornění hustoty pravděpodobnosti a distribuční funkce:





Určení distribuční funkce Φ(x):



Hustota pravděpodobnosti normovaného normálního rozdělení je symetrická kolem „0“ a platí pro ni tedy:

$$\varphi(z) = \varphi(-z); \qquad -\infty < z < \infty$$

A opět ze symetrie dostáváme pro distribuční funkci (viz. výše uvedený obrázek):

$$\Phi(z) = 1 - \Phi(-z); \qquad -\infty < z < \infty$$

Zároveň lze dokázat, že pro kvantily normovaného normálního rozdělení platí vztah:

$$z_p = -z_{1-p}$$

Důležitost tohoto rozdělení spočívá zejména v tom, že jeho **distribuční funkce je tabelována**. V tabulkách najdeme distribuční funkci normovaného normálního rozdělení pro $z \geq 0$, pro $z < 0$ určíme distribuční funkci na základě převodního vztahu mezi Φ(z) a Φ(-z).

Standardizace normálního rozdělení

Jak jsme již uvedli výše, distribuční funkci normální náhodné veličiny nedokážeme analyticky nalézt a proto pro její určení používáme distribuční funkce normované (standardní) normální náhodné veličiny.

Nechť:

$$X \to N\left(\mu; \sigma^2\right)$$

Pak definujme náhodnou veličinu Z:

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

Náhodná veličina Z má normované normální rozdělení, $Z \to N(0;1)$.

Mezi distribuční funkci normální a normované normální náhodné veličiny platí tento převodní vztah:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Důkaz:

$$F(x) = P(X < x) = P(Z \cdot \sigma + \mu < x) = P\left(Z < \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Pravidlo 6σ

Pravidlo 6σ je jedním ze základních principů na nichž stojí kontrola kvality a jakosti (SPC – Statisitics Process Control, ISO normy). Toto pravidlo říká, že **máme-li data pocházející z normálního rozdělení o parametrech μ, σ² (hodnoty normální náhodné veličiny X,** $X \to N(\mu,\sigma^2)$**), pak téměř všechna (99,8% z nich) leží v intervalu** $(\mu \pm 3\sigma)$**.** Protože délka tohoto intervalu je 6σ, hovoří se o pravidle šesti sigma.

Důkaz:

$$X \to N(\mu,\sigma^2)$$

Chceme dokázat, že: $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0{,}998$

Řešený příklad:

$L:\quad P(\mu - 3\sigma < X < \mu + 3\sigma) = F(\mu + 3\sigma) - F(\mu - 3\sigma) = \Phi\left(\frac{(\mu+3\sigma)-\mu}{\sigma}\right) - \Phi\left(\frac{(\mu-3\sigma)-\mu}{\sigma}\right) =$

$$= \Phi(3) - \Phi(-3) = \Phi(3) - [1 - \Phi(3)] = 2\cdot\Phi(3) - 1 = 2\cdot 0{,}999 - 1 =$$
$$= 0{,}998$$

$P:\quad 0{,}998$

$L = P$

Řešený příklad:

Stanovme pravděpodobnost, že náhodná veličina X mající rozdělení $N\left(\mu,\sigma^2\right)$ nabude hodnoty z intervalu $\left(\mu - k\cdot\sigma;\, \mu + k\cdot\sigma\right)$ pro dané kladné k.

Řešení:

Pro k>0:

$$\underline{\underline{P(\mu - k\sigma < X < \mu + k\sigma)}} = F(\mu + k\sigma) - F(\mu - k\sigma) = \Phi\left(\frac{(\mu + k\sigma) - \mu}{\sigma}\right) - \Phi\left(\frac{(\mu - k\sigma) - \mu}{\sigma}\right) =$$
$$= \Phi(k) - \Phi(-k) = \Phi(k) - [1 - \Phi(k)] = \underline{\underline{2\cdot\Phi(k) - 1}}$$

Následující tabulka uvádí hodnoty této pravděpodobnosti pro některé hodnoty k:

k	$P(\mu - k\sigma < \mathrm{X} < \mu + k\sigma)$
1	0,683
1,64	0,900
1,96	0,950
2,58	0,990
3	0,998

Logaritmicko-normální rozdělení

Jestliže má náhodná veličina Y, Y = ln X, normální rozdělení s parametry μ a σ^2, pak náhodná veličina X má logaritmicko-normální rozdělení se stejnými parametry, což zapisujeme:

$$X \to LN\left(\mu;\sigma^2\right)$$

Z definice je zřejmé, že náhodná veličina s logaritmicko-normálním rozdělením může nabývat pouze kladných hodnot (definiční obor ln x). Proto nachází uplatnění při **popisu náhodných veličin nabývajících pouze kladných hodnot a to zejména v případech, kdy hustota pravděpodobnosti je asymetrická** (šikmost není nulová) s jedním vrcholem. Značný význam tohoto rozdělení tedy nacházíme v teorii spolehlivosti (různé parametry součástek nabývají pouze kladných hodnot – životnost, rozměry, tažnost, …) a v ekonomii při popisu příjmů (příjmová rozdělení).

Hustota pravděpodobnosti:

$$f(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}}\cdot e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}; & \text{pro x} > 0 \\ 0 & \text{pro x} \le 0 \end{cases}$$

Distribuční funkce:

Distribuční funkci log.-normálního rozdělení nalezneme prostřednictvím distribuční funkce normovaného normálního rozdělení.

$$F(x) = \begin{cases} \Phi\left(\dfrac{\ln x - \mu}{\sigma}\right); & \text{pro } x > 0 \\ 0 & \text{pro } x \leq 0 \end{cases}$$

Střední hodnota: $EX = e^{\mu + \frac{\sigma^2}{2}}$

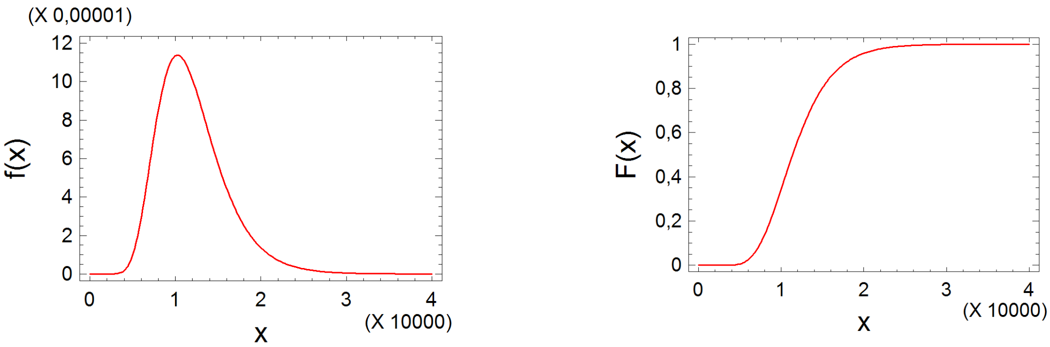
Rozptyl: $DX = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right)$

100p%-ní kvantil: $x_p = e^{\mu + \sigma \cdot z_p}$,

kde z_p je 100p%-ní kvantil normovaného normálního rozdělení

Grafické znázornění hustoty pravděpodobnosti a distribuční funkce:

$$X \to LN\left(12.000; 4.000^2\right)$$



Při praktickém používání tohoto rozdělení postupujeme tak, že náhodnou veličinu X nejdříve převedeme na Y = ln X a potom již postupujeme stejně jako u normálního rozdělení.

- **Odvození distribuční funkce logaritmicko-normálního rozdělení:**

Nechť:

$$Y = \ln X$$

$$X \to LN\left(\mu; \sigma^2\right) \quad \Leftrightarrow \quad Y \to N\left(\mu; \sigma^2\right)$$

$F_X(x)$ (resp. $F_Y(y)$) je distribuční funkce náhodné veličiny X (resp. Y)

$$\forall x > 0: \quad F_X(x) = P(X < x) = P\left(e^Y < x\right) = P(Y < \ln x) = F_Y(\ln x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

$$\forall x \leq 0: \quad F_X(x) = 0$$

- **Odvození hustoty pravděpodobnosti logaritmicko-normálního rozdělení:**

$f_X(x)$ … hustota pravděpodobnosti náhodné veličiny X

$$\forall x > 0: \quad f_X(x) = \frac{dF_X(x)}{dx} = \frac{d\Phi\left(\frac{\ln x - \mu}{\sigma}\right)}{dx} = \varphi\left(\frac{\ln x - \mu}{\sigma}\right) \cdot \frac{1}{x \cdot \sigma} = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{\ln x - \mu}{\sigma}\right)^2}{2}} =$$

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

$$\forall x \leq 0: \quad f_X(x) = 0$$

- **Odvození vztahu pro výpočet 100p%-ního kvantilu:**

$$\begin{aligned}
P(X < \mathrm{x_p}) &= p \\
F(x_p) &= p \\
\Phi\left(\frac{\ln x_p - \mu}{\sigma}\right) &= p \\
\frac{\ln x_p - \mu}{\sigma} &= z_p \qquad (\Phi(z_p) = p) \\
\ln x_p &= \sigma \cdot z_p + \mu \\
x_p &= e^{(\mu + \sigma \cdot z_p)}
\end{aligned}$$

Řešený příklad:

Nechť X je náhodná veličina s logaritmicko-normálním rozdělením s parametry: $\mu=2$; $\sigma^2=9$. Určete:

a) pravděpodobnost, že náhodná veličina X je z intervalu (0;30)
b) medián daného rozdělení
c) střední hodnotu a rozptyl náhodné veličiny X

Řešení:

$$X \to LN(2;9)$$

ada) Pravděpodobnost, že náhodná veličina X je z intervalu (0;30) můžeme určovat rovněž jako pravděpodobnost, že náhodná veličina X je menší než 30, neboť log.-normální náhodná veličina může nabývat pouze kladných hodnot.

Připomeňme si postup při určování distribuční funkce log.-normální náhodné veličiny:

$$F(x) = \begin{cases} \Phi\left(\frac{\ln x - \mu}{\sigma}\right); & \text{pro } x > 0 \\ 0 & \text{pro } x \leq 0 \end{cases}$$

A nyní již přejděme k určení hledané pravděpodobnosti:

$$\underline{\underline{P(0 < X < 30)}} = F(30) - F(0) = \Phi\left(\frac{\ln 30 - 2}{\sqrt{9}}\right) - 0 = \Phi(0{,}47) \underline{\underline{= 0{,}681}}$$

nebo

$$\underline{\underline{P(0 < X < 30)}} = P(X < 30) = F(30) = \Phi\left(\frac{\ln 30 - 2}{\sqrt{9}}\right) = \Phi(0{,}47) \underline{\underline{= 0{,}681}}$$

adb) Pro určení mediánu můžeme použít vztah pro 100p%-ní kvantil, který byl odvozen v Průvodci studiem:

$$x_p = e^{\mu + \sigma \cdot z_p}$$

$$z_{0,5} = 0 \quad \Rightarrow \quad \underline{\underline{x_{0,5}}} = e^{2 + \sqrt{9} \cdot 0} = \underline{\underline{e^2 \cong 7{,}4}}$$

adc) Střední hodnotu a rozptyl určíme na základě výše uvedených vztahů:

$$EX = e^{\mu + \frac{\sigma^2}{2}} \quad \Rightarrow \quad \underline{\underline{EX}} = e^{2 + \frac{9}{2}} = \underline{\underline{e^{\frac{13}{2}} \cong 665{,}1}}$$

$$DX = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right) \quad \Rightarrow \quad \underline{\underline{DX}} = e^{2 \cdot 2 + 9}\left(e^9 - 1\right) \underline{\underline{\cong 3{,}6 \cdot 10^9}}$$

Další úlohy k řešení

1. Doba vypracování testu má normální rozdělení se střední hodnotou 60minut a směrodatnou odchylkou 10minut.
   a) Kolik % studentů dokončí test do hodiny a čtvrt?
   b) Jaká doba by měla být stanovena, aby test dokončilo průměrně 95% studentů?

2. Výrobní zařízení má poruchu v průměru jednou za 2000 hodin. Jaká je pravděpodobnost, že přístroj bude pracovat déle než 550 hodin?

3. Životnost žárovky má exponenciální rozdělení se střední hodnotou 400h. S jakou pravděpodobností bude žárovka svítit dalších 100 hodin, jestliže již svítila 600 hodin?

4. Odhadujeme, že střední životnost určitého přístroje je 110 dnů. S jakou pravděpodobností bude životnost náhodně vybraného přístroje mezi 100 a 150 dny?

5. Při kontrole jakosti přebíráme součástku pouze tehdy, jestliže se její rozměr pohybuje v mezích 26-27mm. Rozměry součástek mají normální rozdělení se střední hodnotou 26,4mm a směrodatnou odchylkou 0,2mm. Jaká je pravděpodobnost, že rozměr součástky náhodně vybrané ke kontrole bude v požadovaných mezích?

6. Průměrná doba mezi příjezdy nákladních automobilů s betonovou směsí je 10 minut. Jaká je pravděpodobnost, že doba mezi příjezdy dvou vozidel bude kratší než 7 minut?

7. Firma získá z každého prodaného výrobku 100,-Kč. Za výměnu během záruční lhůty zaplatí 300,-Kč. Životnost výrobku v letech má normální rozdělení N(3;1). Jakou záruční dobu v měsících má firma stanovit, aby střední (průměrný) zisk byl alespoň 60,-Kč/výrobek?

8. Doba do vybití baterie se řídí exponenciálním rozdělením.
   a) Jaká je střední doba do vybití, víme-li, že 4000 hodin přežije 1% těchto baterií?
   b) Je-li střední doba do vybití 3.150 hodin, kolik procent těchto baterii přežije 4000 hodin?

9. Chybu při měření určité veličiny modelujeme normálním rozdělením s nulovou střední hodnotou a s rozptylem 1,5. Určete interval (souměrný podle počátku), ve kterém se bude nacházet chyba v 90% měření.

10. Obsah nečistot v odpadních vodách je popsán normálním rozdělením se střední hodnotou 0,18 a směrodatnou odchylkou 0,03. Vypočtěte:
    6. procento zkoušek, při kterých obsah nečistot překročí hodnotu 0,24.
    7. hodnotu obsahu nečistot, která bude překročena v 1% zkoušek.

Řešení:

1. a) $0{,}933 = 93{,}3\%$
   b) 1 hodina 17 minut

2. $e^{-\frac{550}{2000}} = 0{,}760 = 76\%$

3. $e^{-\frac{1}{4}} = 0{,}779 = 77{,}9\%$

4. $e^{-\frac{100}{110}} - e^{-\frac{150}{110}} \cong 0{,}147 = 14{,}7\%$

5. 0,976 = 97,6%

6. $1 - e^{-\frac{7}{10}} \cong 0{,}503 = 50{,}3\%$

7. $T \leq 2{,}89\ let \Rightarrow \mathrm{T} = 34\ m\text{ěsíce}$

8. 869 hodin

9. $P(-\,2{,}01 < X < 2{,}01) = 0{,}9$

10. a) $0{,}977 = 97{,}7\%$
    b) 0,25