Rozdělení pravděpodobnosti pro aplikace

Úvodem. V kapitole věnované náhodným veličinám jsme se obecně zabývali funkčními a číselnými charakteristikami náhodných veličin a jejich vlastnostmi. Nyní se budeme zajímat o typické náhodné veličiny, přesněji o **typická rozdělení pravděpodobnosti náhodných veličin.** Místo „rozdělení pravděpodobnosti náhodných veličin" budeme dále často krátce uvádět „rozdělení".

Připomeňme si, že v 1. semestru studia na FSI VUT jsou v předmětu Matematika 1 probírány reálné funkce jedné reálné proměnné a jejich vlastnosti jsou rozebírány nejen obecně, ale i pro jednotlivé skupiny elementárních funkcí, které se pak používají v řadě inženýrských aplikací.

Podobně za kapitolou věnovanou náhodným veličinám a rozdělením pravděpodobnosti z obecného pohledu nyní následuje kapitola, ve které nás budou zajímat typická rozdělení pravděpodobnosti, jež jsou vhodnými matematickými modely častých náhodných pokusů.

Rozdělení pravděpodobnosti můžeme zjednodušeně klasifikovat jako diskrétní, spojitá a smíšená (viz kapitola zavádějící náhodné veličiny). V této kapitole se **omezíme na několik základních diskrétních a spojitých rozdělení pravděpodobnosti v rozsahu podle sylabu předmětu Matematika 4.** Rozšiřující informace o dalších rozděleních jsou **odlišeny značením.** Čtenář zajímající se o podrobnější informace a další rozdělení pak najde více informací v odborné literatuře [17] a na internetu [20].

A. Typická diskrétní rozdělení pravděpodobnosti

1. Alternativní (Bernoulliovo) rozdělení

Značení. Skutečnost, že náhodná veličina X má alternativní rozdělení pravděpodobnosti zapisujeme $X \sim A(p)$, kde $p \in \langle 0; 1\rangle$ a zápis čteme **„náhodná veličina X má alternativní rozdělení pravděpodobnosti s parametrem p"**. Písmeno A tedy značí název rozdělení, písmeno p označuje jeho parametr. V konkrétních příkladech je označení parametru nahrazeno jeho číselnou hodnotou. Někteří autoři vyhrazují písmeno p pouze pro označení pravděpodobnostní funkce a pro odlišení používají značení $A(\pi)$. My budeme předpokládat, že ze souvislosti bude zřejmé, v jakém významu byl použit symbol p.

Obecně o značení. Poznamenejme, že obdobné značení je obvyklé i pro ostatní rozdělení pravděpodobnosti. Písmeno označující náhodnou veličinu (zde X) je následováno symbolem $\sim$, který čteme „má rozdělení", a dále následuje označení konkrétního rozdělení písmenem (nebo více písmeny) a symbolický zápis uzavírá seznam parametrů rozdělení pravděpodobnosti v kulatých závorkách, oddělený čárkami nebo středníky, pokud by mohlo dojít k záměně s desetinnou čárkou.

Jiný název. V cizojazyčné literatuře najdeme alternativní rozdělení pod názvem **Bernoulliovo rozdělení** a značí se $X \sim \mathrm{Be}(p)$.

Použití. Rozdělení je používáno k modelování jednoho náhodného pokusu, který má dva možné číselné výsledky: **výsledek** 1 (úspěch) **s pravděpodobností** p a **výsledek** 0 (neúspěch) **s pravděpodobností** $1 - p$. Vidíme, že parametr rozdělení p má význam pravděpodobnosti úspěchu, tj. $p = p(1) = P(X = 1)$.

Funkční charakteristiky. Jak již víme, každé rozdělení je jednoznačně určeno funkční charakteristikou. Pro diskrétní rozdělení pravděpodobnosti budeme uvádět zejména pravděpodobnostní funkce $p(x)$. Připomeňme, že $p(x)$ a $F(x)$ byly zavedeny tak, že jejím definičním oborem je množina všech reálných čísel $\mathbb{R}$. Dále pro pravděpodobnostní prostor $(\Omega, \mathcal{S}, \mathcal{P})$ a $X : \Omega \longrightarrow \mathbb{R}$ takové, že $\{\omega \mid X(\omega) = x\} \in \mathcal{S}$ a $\{\omega \mid X(\omega) < x\} \in \mathcal{S}$ platí, že $\forall x \in \mathbb{R} : p(x) = P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\})$ a $F(x) = P(X < x) = P(\{\omega \in \Omega \mid X(\omega) < x\})$.

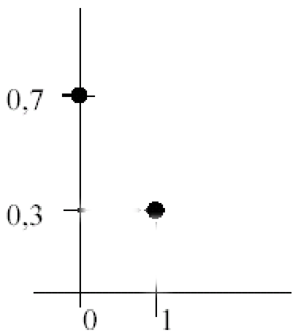
Pravděpodobnostní funkce:

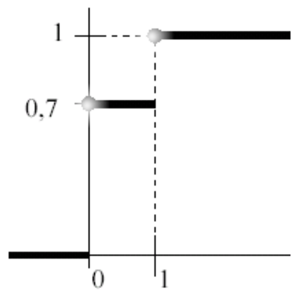
$$p(x) = P(X = x) = \begin{cases} 1-p & \text{pro } \ x = 0 \\ p & \text{pro } \ x = 1 \\ 0 & \text{jinde} \end{cases} \tag{1}$$

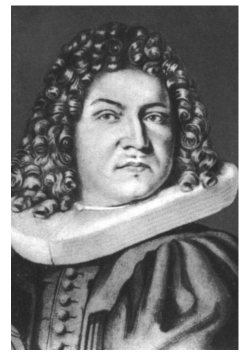
Bývá používán i kompaktnější zápis pro $x \in \{0;1\}$: $p(x) = p^x(1-p)^{1-x}$.

Distribuční funkce:

$$F(x) = P(X < x) = \begin{cases} 0 & \text{pro } \ x \leq 0 \\ 1-p & \text{pro } \ x \in (0,1\rangle \\ 1 & \text{pro } \ x > 1 \end{cases} \tag{2}$$







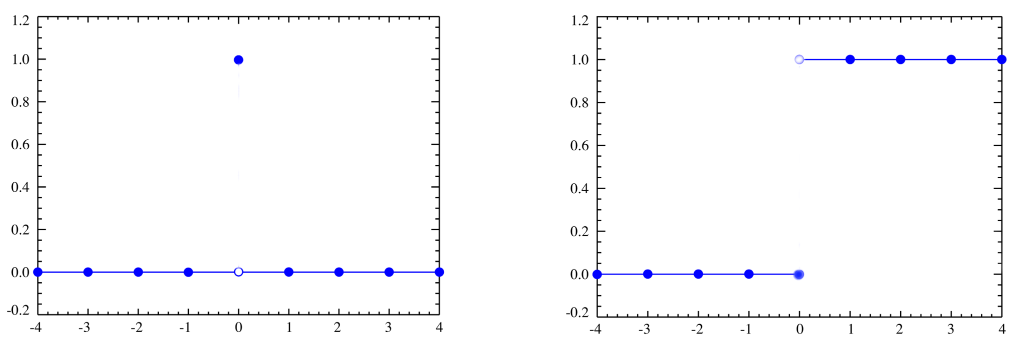
Obrázek 1: $p(x)$ pro $A(0,3)$ $F(x)$ pro $A(0,3)$ Jakob Bernoulli

Číselné charakteristiky. Je výhodou, že pro jednotlivá rozdělení pravděpodobnosti lze často odvodit vztahy pro výpočet číselných charakteristik pomocí parametrů rozdělení a nemusíme pak při konkrétních výpočtech opakovaně používat obecné vzorce.

$$\textbf{střední hodnota:} \quad E(X) = p \tag{3}$$

$$\textbf{rozptyl:} \quad D(X) = p(1-p) \tag{4}$$

Degenerované rozdělení. V případě $X \sim A(0)$ hovoříme o degenerovaném rozdělení a na obrázku 2 vidíme náčrt grafů pravděpodobnostní a distribuční funkce.



Obrázek 2: Degenerované rozdělení, nástin grafu $p(x)$ a $F(x)$.

Zajímavosti. Rozdělení dostalo své druhé a exoticky znějící zahraniční jméno po švýcarském matematikovi Jakobu Bernoullim (1654–1705) [21], který se zabýval posloupnostmi nezávislých náhodných pokusů modelovatelných tímto rozdělením. Jakob Bernoulli patřil mezi osm známých matematiků rodiny Bernoulliových. Studoval teologii, matematiku a astronomii. Během svých cest po Evropě (1676–1682) se seznámil s výsledky předních vědců té doby. Osvojil si zejména zejména výsledky Gottfrieda Leibnize v oblasti kalkulu (diferenciálního a integrálního počtu). Jako vzpomínku na předmět Matematika III uveďme, že v roce 1690, Jakob Bernoulli vyvinul postup řešení obyčejných diferenciálních rovnic se separovanými proměnnými. Jeho nejznámější prací je Ars Conjectandi, publikovaná posmrtně v roce 1713, shrnující výsledky v oblasti teorie pravděpodobnosti a uvádějící původní důkazy.

Příklady:

(1) Vlastním výpočtem odvoďte vztah (3) pro $E(X)$.

(2) Vlastním výpočtem odvoďte vztah (4) pro $D(X)$ a směrodatnou odchylku $\sqrt{D(X)}$.

(3) Modelujte náhodný hod symetrickou mincí pomocí alternativního rozdělení. Určete hodnotu p. Svůj názor podpořte 100 hody vhodně zvolenou vlastní mincí (Autor textu neodpovídá za ztrátu mince při konání pokusu.).

(4) Pomocí náhodné veličiny Y a lineární transformace náhodné veličiny X s alternativním rozdělením popište náhodný pokus u kterého výsledek 2 nastane s pravděpodobností $0,3$ a výsledek 5 nastane s pravděpodobností $0,7$.

Závěrem. Jedná se o značně triviální rozdělení, které je ale základem pro složitější rozdělení. Otázka položená u zkoušky na toto rozdělení obvykle svědčí o naprostém vyčerpání zkoušejícího a o zoufalé snaze pomoci studentovi uspět. Je pak velmi smutné, pokud student zatvrzelým mlčením tuto pomocnou ruku odmítne.

2. Klasické (diskrétní rovnoměrné) rozdělení

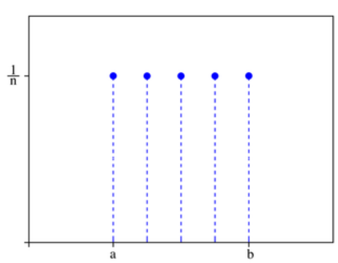
Značení. Skutečnost, že náhodná veličina X má klasické rozdělení, značíme $X \sim C(n)$, kde $n \in \mathbb{N}$.

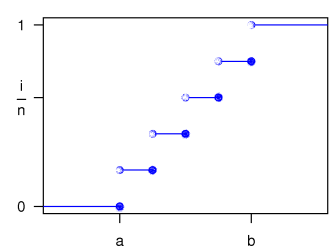
Použití. Klasické rozdělení pravděpodobnosti získalo své jméno, protože nám umožňuje modelovat úlohy klasické pravděpodobnosti. Připomeňme, že se jedná o úlohy, ve kterých se k výpočtu pravděpodobnosti jevu A používá vztah

$$P(A) = \frac{|A|}{|\Omega|} = \frac{m}{n} = \frac{\text{počet příznivých výsledků}}{\text{počet všech možných výsledků}}$$

a jedná se o modelování náhodných pokusů, které mají konečný počet n stejně možných výsledků.

Funkční charakteristiky. Z výše uvedeného vyplývá pravděpodobnost $\frac{1}{n}$ pro elementární (zde jednoprvkové) jevy, a tedy i následující vztah (1) pro pravděpodobnostní funkci. Distribuční funkci zde neuvádíme, pouze ukázku jejího grafu (viz obr. 3), tak jako pro $p(x)$.





Obrázek 3: $p(x)$ $F(x)$

Pravděpodobnostní funkce:

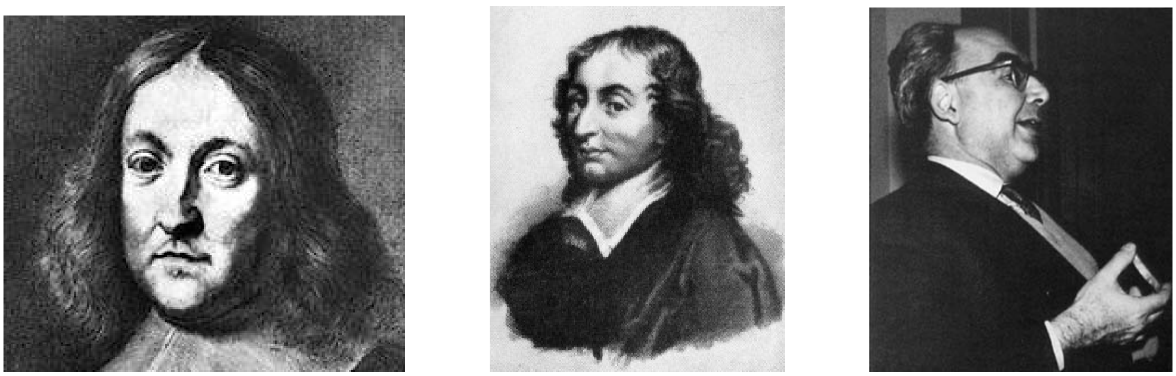
$$p(x) = \begin{cases} \frac{1}{n} & \text{pro } x = 1, 2, \ldots, n \\ 0 & \text{jinde} \end{cases} \tag{5}$$

Číselné charakteristiky.

$$\textbf{střední hodnota:}\quad E(X) = \frac{n+1}{2} \tag{6}$$

$$\textbf{rozptyl:}\quad D(X) = \frac{n^2-1}{12} \tag{7}$$

Zajímavosti. Začátky klasické pravděpodobnosti bývají spojovány s komunikací matematiků Fermata a Pascala v 17.století. Jejich korespondenci se pokusil dovedně „rekonstruovat" vynikající maďarský odborník v oblasti teorie pravděpodobnosti Rényi [5], autor „Dialogů o matematice". Zájemcům o hlubší poznání myšlenek teorie pravděpodobnosti je vřele doporučujeme (Rényi, A.: Dialogy o matematice. Praha, Mladá fronta, 1980).



Obrázek 4: Pierre Fermat (1601-1665), Blaise Pascal (1623-1662), Alfréd Rényi (1921-1970)

Posloupnost realizací hodnot diskrétního rovnoměrného rozdělení čtenář obdrží, pokud ve většině programovacích jazyků využije základní náhodný generátor (přesněji ve většině případů kongruenční generátor posloupnosti pseudonáhodných čísel).

Zájemcům o problematiku náhodných generátorů doporučujeme zprvu prohledat aktuální internetové zdroje (viz např. `http://mathworld.wolfram.com/topics/RandomNumbers.html`).

Příklady:

(1) Ukažte, že pro vhodnou volbu p lze zápis $X \sim A(p)$ interpretovat jako zápis $X + 1 \sim C(2)$.

(2) Vlastním výpočtem odvoďte vztah (6) pro $E(X)$.

(3) Vlastním výpočtem odvoďte vztah (7) pro $D(X)$ a směrodatnou odchylku $\sqrt{D(X)}$.

(4) Modelujte náhodný hod pravidelnou kostkou pomocí klasického rozdělení. Určete hodnoty $p(x)$. Svůj názor podpořte 60 hody hrací kostkou (Kostku neztraťte, budete ji ještě potřebovat).

(5) Řešte některé úlohy klasické pravděpodobnosti tak, že určíte náhodnou veličinu $X \sim C(n)$ popisující náhodný pokus a žádanou pravděpodobnost $P(X \in B)$ vypočtete pomocí $\sum_{x \in B} p(x)$.

(6) Vyhledejte ve svém oblíbeném programovacím jazyce (existuje-li takový) generátor posloupnosti „náhodných" čísel. Jeho opakovaným použitím získejte větší statistický soubor (více než 30 hodnot) a pomocí vhodného softwaru (MS Excel, Statistica, Minitab) získejte tříděním tabulku relativních četností a histogram. Posuďte, zda jeho tvar splňuje Vaše očekávání (byla-li nějaká).

3. Hypergeometrické rozdělení

Značení. Skutečnost, že náhodná veličina X má hypergeometrické rozdělení, značíme $X \sim H(N, M, n)$, kde $N, M, n \in \mathbb{N}$ a $1 \leq n < N, 1 \leq M < N$ (ostré nerovnosti zajišťují, že se nebudeme zabývat triviálními případy).

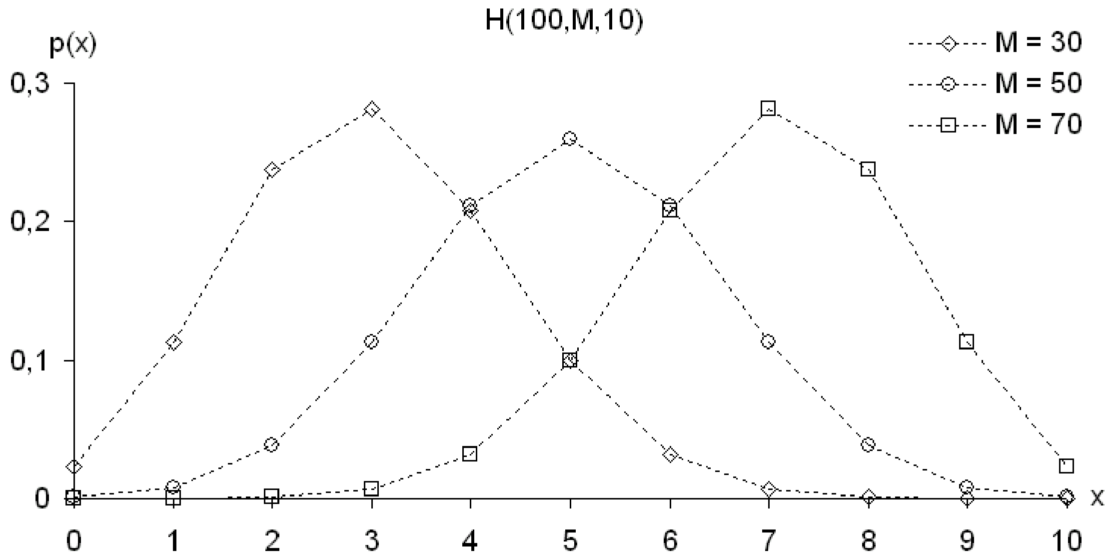
Použití. Hypergeometrické rozdělení pravděpodobnosti se používá k modelování náhodných pokusů, při kterých **náhodně vybíráme** (najednou nebo postupně) **a nevracíme** n jednotek (výrobků, součástek, kuliček, aj.) z konečného souboru N jednotek, ze kterých je M s určitou vlastností (zmetky, bílé kuličky, aj.) a zbývajících $N - M$ tuto vlastnost nemá (dobré výrobky, černé kuličky, aj.). Přitom se ptáme, jaká je pravděpodobnost, že mezi n vybranými je určitý počet jednotek (obvykle x, pokud nás zajímá hodnota pravděpodobnostní funkce $p(x)$) se zmíněnou vlastností (zmetků, bílých kuliček, aj.). **Náhodná veličina X pak označuje náhodný počet jednotek se zmíněnou vlastností mezi vybranými při výběru bez vracení.**

Funkční charakteristiky. Z výše uvedeného a na základě zkušeností s výpočtem příkladů na náhodný výběr bez vracení v klasické pravděpodobnosti vyplývá následující vztah (8) pro pravděpodobnostní funkci. Distribuční funkci $F(x)$ zde neuvádíme. Pokud ale není u diskrétních rozdělení $F(x)$ uvedena, stačí dosadit do obecného vztahu $F(x) = \sum_{t<x} p(t)$, který byl uveden v kapitole věnované zavedení náhodných veličin a jejich charakteristik.

Pravděpodobnostní funkce:

$$p(x) = \begin{cases} \dfrac{\dbinom{M}{x}\dbinom{N-M}{n-x}}{\dbinom{N}{n}} & \text{pro } \; x = \max\{0, M-N+n\}, \ldots, \min\{M, n\} \\ 0 & \text{jinde} \end{cases} \tag{8}$$

Hodnoty pravděpodobnostní funkce H rozdělení lze počítat přímo podle (8), lze je rovněž najít ve statistických tabulkách ([13], [14]) a zejména řada statistických softwarů nabízí jejich rychlý výpočet. Pro pravděpodobnostní funkci (8) uvádíme ukázky jejich grafů (viz obr. 5), kde volíme $N = 100$, $n = 10$ a různé hodnoty M. Poznamenejme, že přerušovaná čára je použita pouze pro odlišení jednotlivých pravděpodobnostních funkcí.



Obrázek 5: $p(x)$

Číselné charakteristiky.

střední hodnota:
$$E(X) = n\frac{M}{N} \tag{9}$$

rozptyl:
$$D(X) = n\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1} \tag{10}$$

modus:
$$\hat{x} \in \left\langle \frac{(M+1)(n+1)}{N+2} - 1, \frac{(M+1)(n+1)}{N+2} \right\rangle \tag{11}$$

Zajímavosti. Hypergeometrické rozdělení obvykle „kráčí ruku v ruce" s rozdělením binomickým, protože jedno se používá při modelování náhodného výběru bez vracení (hypergeometrické) a druhé lze použít pro modelování náhodného výběru s vracením (binomické). Doporučujeme čtenáři, aby začal již nyní promýšlet, jaké chyby se dopustí, pokud místo výběru bez vracení omylem uvažuje výběr s vracením. Logickou otázkou je, zda za určitých okolností není tato chyba zanedbatelná. Budeme se studentů ptát, zda a kdy případně ano a kdy ne?

Základní motivací pro hledání odpovědi může být skutečnost, že výpočet pomocí hypergeometrického rozdělení je pracnější (3 kombinační čísla ve vztahu (8) pro výpočet $p(x)$), zatímco pro binomické rozdělení se výpočet jeví jako jednodušší (1 kombinační číslo ve výpočtu $p(x)$ podle (12) – viz dále).

Doplňme praktickou motivaci: Při statistické kontrole kvality, kontrolor náhodně vybere z mnoha výrobků ke kontrole určitou skupinu, aby zjistil procento zmetků. Má tyto výrobky vybírat jednotlivě a vždy vrátit? Nebo má vybrat celou skupinu naráz a vybrané výrobky nevracet? Jaké vzorce má použít? Jsou některé někdy zaměnitelné?

Konkrétnější příklad: V jednom závodě představují výrobky 5% celkové výroby. Sestavte tabulku pravděpodobnosti počtů zmetků při náhodné kontrole tří výrobků. **Otázka:** Vidíte nějaký problém v zadání? Je úplné?

Příklady:

(1) Formulujte a řešte pomocí hypergeometrického rozdělení příklad na náhodný výběr bez vracení, který jste řešili v kapitole věnované klasické pravděpodobnosti.

(2) Najděte v literatuře nebo na internetu odvození vztahu (9) pro $E(X)$.

(3) Najděte v literatuře nebo na internetu odvození vztahu (10) pro $D(X)$ a směrodatnou odchylku $\sqrt{D(X)}$.

(4) Vysvětlete, proč vzorec (8) pro pravděpodobnostní funkci $p(x)$ hypergeometrického rozdělení není definován pro $x = 0, 1, \dots, n$, ale je nutné vzít do úvahy i M a N?

(5) Vysvětlete, jak se zjednoduší výpočet pomocí hypergeometrického rozdělení, pokud $n = N$ nebo $M = N$?

(6) Vyhledejte ve svém oblíbeném statistickém programu (existuje-li takový) generátor posloupnosti náhodných čísel získaných výběrem z hypergeometrického rozdělení pravděpodobnosti. Jeho opakovaným použitím získejte větší statistický soubor (více než 30 hodnot) a znázorněte histogram rozdělení četností výsledků. Porovnejte získané relativní četnosti s vlastním výpočtem získanými hodnotami pravděpodobnostní funkce. Vysvětlete rozdíly.

Závěrem. Z pohledu zkoušky se jedná o rozdělení důležité (viz barva nadpisu této části). Lze rovněž očekávat zadání zaměřené na náhodný výběr bez vracení již při zkoušení poznatků klasické pravděpodobnosti.

4. Binomické rozdělení

Značení. Skutečnost, že náhodná veličina X má binomické rozdělení, značíme $X \sim Bi(n,p)$, kde $n \in \mathbb{N}$ a $p \in (0;1)$ (neuvažujeme triviální případy $p \in \{0;1\}$).

Použití. Pro srovnání s hypergeometrickým rozdělením nejprve uveďme, jak se binomické rozdělení používá k modelování **nezávislých** náhodných pokusů, při kterých **postupně náhodně vybíráme** n jednotek (výrobků, součástek, kuliček, aj.). **Zdůrazněme, že vybrané jednotky vracíme a mohou být znovu vybrány.** Velikost souboru, ze kterého vybíráme, nemusí být známa, ale musí být známo, jaký je podíl p jednotek s určitou vlastností (zmetky, bílé kuličky, aj.). Ostatní jednotky tuto vlastnost nemají a jejich podíl je tedy $q = 1-p$. Přitom se ptáme, jaká je pravděpodobnost, že mezi n vybranými je určitý počet jednotek (obvykle x, pokud nás zajímá hodnota pravděpodobnostní funkce $p(x)$) se zmíněnou vlastností (zmetků, bílých kuliček, aj.). **Náhodná veličina X pak označuje náhodný počet jednotek se zmíněnou vlastností mezi vybranými při výběru s vracením.**

Funkční charakteristiky. Z výše uvedeného a na základě zkušeností s výpočtem příkladů na náhodný výběr s vracením v klasické pravděpodobnosti vyplývají následující vztahy (12) a (13) pro pravděpodobnostní a distribuční funkci ($\mathbb{N}_0$ značí množinu přirozených čísel včetně nuly).

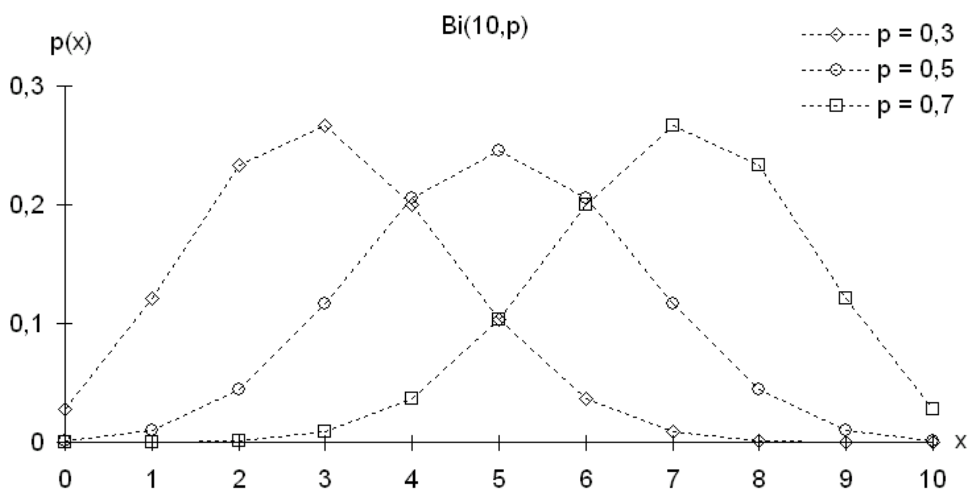
Pravděpodobnostní funkce:

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{pro} \quad x = 0, \dots, n \\ 0 & \text{jinde} \end{cases} \tag{12}$$

Distribuční funkce:

$$F(x) = \begin{cases} 0 & \text{pro} \quad x \le 0 \\ \sum\limits_{t<x, t\in\mathbb{N}_0} \binom{n}{t} p^t (1-p)^{n-t} & \text{pro} \quad x \in (0,n\rangle \\ 1 & \text{pro} \quad x > n \end{cases} \tag{13}$$

Hodnoty pravděpodobnostní funkce Bi rozdělení lze počítat přímo podle (12), lze je rovněž najít ve statistických tabulkách a téměř všechny statistické softwary nabízí jejich rychlý výpočet. Pro pravděpodobnostní funkci (12) uvádíme ukázky jejich grafů (viz obr. 6), kde volíme $n = 10$ a různé hodnoty p. Zdůrazněme, že přerušovaná čára je použita pouze pro odlišení jednotlivých pravděpodobnostních funkcí. Pro $p = 0,5$ je rozdělení symetrické a pro $p > 0,5$ ($p < 0,5$) je záporně (kladně) asymetrické.



Obrázek 6: $p(x)$

Číselné charakteristiky.

$$\begin{aligned} &\textbf{střední hodnota:} & E(X) &= np & (14)\\ &\textbf{rozptyl:} & D(X) &= np(1-p) = npq & (15)\\ &\textbf{modus:} & \hat{x} &\in \langle (n+1)p-1, (n+1)p \rangle & (16)\\ &\textbf{koeficient šikmosti (asymetrie):} & A(X) &= \frac{1-2p}{\sqrt{np(1-p)}} & (17) \end{aligned}$$

Název rozdělení. Název rozdělení pochází ze skutečnosti, že pravděpodobnosti $p(x)$ jsou členy binomického rozvoje $1 = 1^n = (p + (1-p))^n = \sum p(x)$.

Bernoulliovská posloupnost nezávislých pokusů. Tvar pravděpodobnostní funkce, který pro nového čtenáře může vypadat nezvykle, se často vysvětluje pomocí posloupnosti nezávislých pokusů:
1) Máme n nezávislých náhodných pokusů, např. opakovaný náhodný výběr výrobku ze skupiny dobrých výrobků a zmetků s vracením vybraného výrobku.
2) Podíl zmetků ve skupině označíme p a je to tedy i pravděpodobnost vytažení zmetku v jednom tahu.
3) Pravděpodobnost vytažení nejprve právě x zmetků, a potom právě $n-x$ dobrých výrobků je díky nezávislosti pokusů $p^x(1-p)^{n-x}$.
4) Jenže výběr s právě x zmetky lze získat více způsoby než tak, že nejprve budou vytaženy zmetky. Přesněji, počet těchto možností je dán počtem rozmístění x zmetků na n pořadí, ve kterých mohou být taženy. Těchto možností je $\binom{n}{x}$.
5) Vynásobením pravděpodobností jedné možnosti z 3) počtem možností v 4) tedy získáme pravděpodobnost $p(x)$ z (12).

Řečnická otázka: Jaké z nám známých rozdělení obdržíme, pokud zvolíme $n = 1$ u $X \sim Bi(n, p)$? Odpověď: Ano, platí rovněž, že $X \sim A(p)$.

Vlastnosti. Předcházející zjištění můžeme využít k pochopení následujících vlastností binomického rozdělení:

1) Náhodná veličina $X = X_1 + \ldots + X_k$, kde náhodné veličiny X_j, $j = 1, \ldots, k$, jsou stochasticky nezávislé (viz náhodný vektor) a mají binomická rozdělení $Bi(n_j, p)$ se stejným parametrem p, **má opět binomické rozdělení $Bi(n, p)$**, kde $n = n_1 + \ldots + n_k$.

2) Speciálně, **součet n stochasticky nezávislých náhodných veličin s alternativním rozdělením $A(p)$ má binomické rozdělení $Bi(n, p)$.**

Obecnější použití. Nyní již můžeme shrnout, že binomické rozdělení používáme v případě **posloupnosti n nezávislých náhodných pokusů** (viz $A(p)$), kdy **rozlišujeme mezi úspěchem v jednotlivém pokusu**, který nastane **s pravděpodobností p (stejnou pro každý jednotlivý pokus)** a neúspěchem, který nastane s pravděpodobností $1-p$. Náhodná veličina $X \sim Bi(n, p)$ pak **určuje počet úspěchů v n pokusech. Obecněji se často hovoří ne o úspěchu, ale o výskytu určitého jevu.**

Příklady:

(1) Formulujte a řešte pomocí binomického rozdělení příklad na náhodný výběr s vracením, který jste řešili v kapitole věnované klasické pravděpodobnosti.

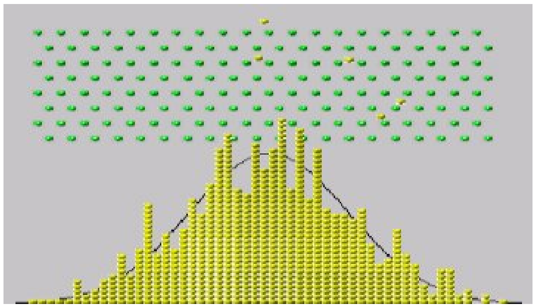
(2) Najděte v literatuře nebo na internetu přímé odvození vztahu (14) pro $E(X)$.

(3) Najděte v literatuře nebo na internetu přímé odvození vztahu (15) pro $D(X)$ a směrodatnou odchylku $\sqrt{D(X)}$.

(4) Zamyslete se nad zdůvodněním, které vyplývá z vlastností střední hodnoty součtu náhodných veličin a rozptylu stochasticky nezávislých náhodných veličin (nápověda: viz náhodný vektor).

Zajímavosti. Mnoho poutavých informací najde laický čtenář v dosud čtivostí nepřekonané knize H. Swobody [10]. Příkladem autorovy představivosti může být nalezení souvislostí mezi binomickým rozdělením a římskou fontánou.

Francis Galton (1822-1911) byl anglický vědec a vzdálený příbuzný Charlese Darwina. Zabýval se statistickým výzkumem v oblasti dědičnosti. Z jeho prací pochází termín regrese jako procesu návratu k průměru. Realizoval rovněž myšlenku „Galtonova stroje" (srovnejte s některými hracími automaty a jejich napodobeninami pro děti), který ilustruje, jak náhodné odrazy žlutých kuliček v bludišti zelených kolíčků vedou k jejich rozmístění „přibližně" podle obrysů násobku pravděpodobnostní funkce $Bi(n; 0,5)$. Vysvětlení „proč to funguje" necháme laskavému čtenáři k přemýšlení a diskusi se cvičícími.



Obrázek 7: Galtonův stroj

Příklady:

(5) Vysvětlete, jak se zjednoduší Bi, když $n = 1$?

(6) Vyhledejte ve svém oblíbeném statistickém programu generátor posloupnosti náhodných čísel získaných výběrem z binomického rozdělení. Jeho opakovaným použitím získejte větší statistický soubor (více než 30 hodnot) a znázorněte histogram rozdělení četností výsledků. Porovnejte získané relativní četnosti s vlastním výpočtem získanými hodnotami pravděpodobnostní funkce. Vysvětlete rozdíly.

(7) Pokud víte, že $X \sim Bi(n, p)$, uveďte funkční a číselné charakteristiky pro rozdělení transformované náhodné veličiny $Y = X/n$. Zvolte vhodné p a malé n, výsledky konkretizujte a načrtněte grafy $p(x)$ a $F(x)$.

(8) Může mít binomické rozdělení právě dvě různé hodnoty $\hat{x}$?

(9) **Typické zadání:** V dodávce 50 výrobků je 5 zmetků. Z dodávky jsou náhodně vybrány 3 výrobky. Počet zmetků mezi vybranými výrobky je náhodná veličina X. Určete typ jejího rozdělení pravděpodobnosti, její pravděpodobnostní funkci $p(x)$, střední hodnotu $E(X)$, rozptyl $D(X)$, směrodatnou odchylku $\sqrt{D(X)}$, koeficient šikmosti $A(X)$, medián $x_{0,5}$, modus $\hat{x}$ a $P(1 < X \leq 3)$. Předpokládejte, že každý vybraný výrobek se vrátí nazpět do dodávky, takže jde o náhodný výběr s vracením.

(10) Tři rovnocenní hráči A, B, C hrají společenskou hru. Který z případů, že hráč A vyhraje 3 ze 4 partií anebo 5 z 8 partií, je pravděpodobnější? Zkuste nejprve uhádnout s pomocí „zdravého rozumu", a potom ověřte svoji úvahu výpočtem.

(11) Rozhodující test u přijímacích zkoušek na VŠ obsahuje právě 20 otázek. Na každou z nich jsou možné 4 odpovědi, z nichž jedna je správná. Student odpovídá náhodně. Jaká je pravděpodobnost, že student bude přijat, když je požadováno 13 správných odpovědí?

(12) Promyslete, zda u testu v příkladě (11) je výhodnější strategií v případě neznalosti mlčet nebo odpovídat náhodně.

(13) Domníváte se, že u testu v příkladě (11) v případě neznalosti mají náhodně odpovídající student a student, který neodpovídá vůbec, rovné podmínky? Navrhněte řešení! Nápověda: zvažte penalizaci špatných odpovědí.

Závěrem. Z pohledu zkoušky se jedná o rozdělení pravděpodobnosti, které je ale opravdu **velmi důležité**.

5. Poissonovo rozdělení

Značení. Skutečnost, že náhodná veličina X má Poissonovo rozdělení, značíme $X \sim Po(\lambda)$, kde $\lambda > 0$.

Použití. Poissonovo rozdělení se obvykle užívá pro vyjádření **pravděpodobnosti počtu nastoupení sledovaného jevu v určitém časovém intervalu** (počet poruch, nehod, katastrof, zmetků, apod.) s malou pravděpodobností výskytu.

Rada. Při řešení úloh na typická rozdělení je pro mnohé studenty problémem správně zvolit typ rozdělení. Jestliže při volbě H nebo Bi bývá obvykle rozhodující, zda jde o výběr bez vracení nebo s vracením, u Poissonova rozdělení mnozí studenti tápou. Ve snaze pomoci jim nabízíme následující zapamatovatelnou pomůcku.

Všimněme si typického zadání příkladu na Po **rozdělení**: „Určitá radioaktivní látka vyzařuje průměrně 30 částic α za minutu. Vypočítejte pravděpodobnost, že v průběhu jedné sekundy vyzáří látka právě 2 částice."

A srovnejme nyní se zadáním na Bi **rozdělení**: „Náhodně vybíráme 10 výrobků z krabice, ve které je $100p\%$ zmetků, vybrané výrobky vracíme, vypočítejte pravděpodobnost jevu, že mezi vybranými budou právě 2 zmetky."

Pomocná otázka: U Bi rozdělení zní smysluplně otázka: „Jestliže 2 zmetky byly vybrány, kolik jich nebylo vybráno?"

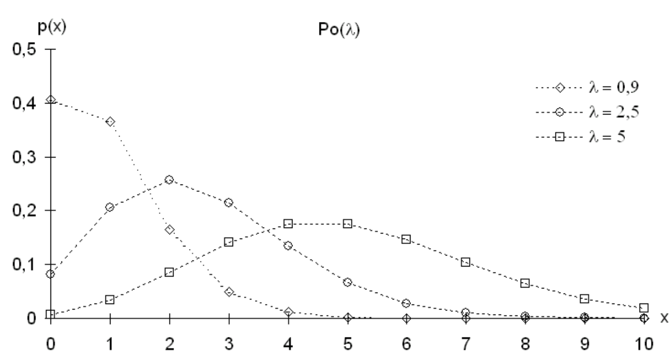
U Po rozdělení naproti tomu otázka: **„Jestliže přiletěly 2 částice, kolik jich nepřiletělo?"** zaujme svojí nesmyslností, a to nás upozorňuje, abychom použili Poissonovo rozdělení, u kterého se úspěch vztahuje k zanedbatelnému časovému okamžiku uvnitř časového intervalu, zatímco u Bi rozdělení úspěch se vztahuje k jednomu z konečně mnoha pokusů.

Funkční charakteristiky. Než uvedeme $p(x)$ pro $Po(\lambda)$, zdůrazněme, že Poissonovo rozdělení je první námi uvedené diskrétní rozdělení pravděpodobnosti s nekonečným nosičem, tj. množina x, pro která $p(x) > 0$ je nekonečná, přesněji spočetná, zapisujeme $|\{x \mid p(x) > 0\}| = \aleph_0$.

Pravděpodobnostní funkce:

$$p(x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!} & \text{pro } \ x = 0, 1, \ldots \\ 0 & \text{jinde} \end{cases} \tag{18}$$

Hodnoty pravděpodobnostní funkce Po rozdělení lze počítat přímo podle (18) a statistické softwary většinou nabízí jejich rychlý výpočet. Hodnoty distribuční funkce bývají tabelovány. Pro $p(x)$ (18) uvádíme ukázky jejich grafů (viz obr. 8), kde volíme různé hodnoty parametru rozdělení λ. Zdůrazněme, že přerušovaná čára je použita pouze pro odlišení jednotlivých pravděpodobnostních funkcí.



Obrázek 8: $p(x)$

Siméon-Denis Poisson

Číselné charakteristiky.

střední hodnota:	$E(X)$	$=$	λ	(19)
rozptyl:	$D(X)$	$=$	λ	(20)
modus:	$\hat{x}$	$\in$	$\langle \lambda - 1, \lambda \rangle$	(21)
koeficient šikmosti:	$A(X)$	$=$	$\sqrt{\lambda}$	(22)

Příklady:

(1) Ověřte, že funkce (18) je pravděpodobnostní funkcí. Nápověda: Využijte svých znalostí nekonečných řad z předmětu Matematika 3 a ukažte, že $\sum_{x \in \mathbb{R}} p(x) = 1$.

(2) Odvoďte vztah (19) pro $E(X)$.

(3) Odvoďte vztah (20) pro $D(X)$ a směrodatnou odchylku $\sqrt{D(X)}$.

(4) Rozlište případy, kdy λ je a není přirozené číslo. Nápověda: Všimněte si počtu modů $\hat{x}$.

(5) Určete modus $\hat{x}$ pro $\lambda < 1$.

Vlastnosti. Náhodná veličina $X = X_1 + \ldots + X_k, k \geq 2$, kde náhodné veličiny X_j, $j = 1, \ldots, k$, jsou stochasticky nezávislé (viz náhodný vektor) a mají Poissonovo rozdělení $Po(\lambda_j)$, **má opět Poissonovo rozdělení** $Po(\lambda)$, kde $\lambda = \lambda_1 + \ldots + \lambda_k$.

Zajímavosti. Zajímavý příklad aplikace Poissonova rozdělení najde čtenář v již zmiňované knize H. Swobody [10]. Autor uvádí, že Poissonovo rozdělení měl počet úmrtí na základě kopnutí koněm v pruské armádě v 19.století.

Rozdělení bylo zavedeno Siméonem-Denisem Poissonem (1781–1840) a publikováno společně s jeho dalšími výsledky v teorii pravděpodobnosti v roce 1838 v jeho práci s poutavým názvem „Recherches sur la probabilité des jugements en matieres criminelles et matiere civile” nebo anglicky „Research on the Probability of Judgments in Criminal and Civil Matters” (překlad alespoň názvu ponecháváme laskavému čtenáři).

Příklady:

(6) Vyhledejte ve svém oblíbeném statistickém programu generátor posloupnosti náhodných čísel získaných výběrem z Poissonova rozdělení. Jeho opakovaným použitím získejte větší statistický soubor (více než 30 hodnot) a znázorněte histogram rozdělení četností výsledků. Porovnejte získané relativní četnosti s vlastním výpočtem získanými hodnotami pravděpodobnostní funkce. Vysvětlete rozdíly.

(7) Zvolte vhodné λ a načrtněte grafy $p(x)$ a $F(x)$.

(8) Uveďte vztah určující distribuční funkci $F(x)$ pro Po rozdělení.

(9) **Typické zadání:** Statistickým průzkumem bylo zjištěno, že během jedné minuty navštíví prodejnu průměrně 3 zákazníci. Najděte vhodný typ rozdělení pravděpodobnosti náhodné veličiny X vyjadřující počet zákazníků, kteří navštíví prodejnu během jedné minuty. Určete její pravděpodobnostní funkci $p(x)$, střední počet zákazníků $E(X)$, rozptyl $D(X)$ a směrodatnou odchylku $\sqrt{D(X)}$ počtu zákazníků, koeficient šikmosti $A(X)$ a nejpravděpodobnější počet zákazníků za jednu minutu. Určete dále pravděpodobnost, že během jedné minuty přijde a) právě 1 zákazník, b) alespoň 1 zákazník, c) medián $x_{0,5}$ počtu zákazníků.

Závěrem. Poissonovo rozdělení patří z pohledu zkoušky mezi základní rozdělení.

6. Geometrické rozdělení a další

Značení. Skutečnost, že náhodná veličina X má geometrické rozdělení, značíme $X \sim Ge(p)$, kde parametr $p \in (0; 1)$.

Použití. Příkladem náhodné veličiny s geometrickým rozdělením je počet neúspěšných nezávislých náhodných pokusů (s pravděpodobností úspěchu p), které předcházejí prvnímu úspěchu.

Rada. Pomůckou pro rozeznání aplikovatelnosti Ge rozdělení je dostatečně drastická formulace zadání pomocného příkladu: „ Představme si například testování nového typu padáku opakovanými nezávislými seskoky. Zajímá nás počet úspěchů (opak formulace výše), zde s pravděpodobností $1 - p$, které předchází první neúspěch."

Funkční charakteristiky. Uvedeme $p(x)$ a upozorňujeme, že geometrické rozdělení je další diskrétní rozdělení pravděpodobnosti s nosičem splňujícím $|\{x \mid p(x) > 0\}| = \aleph_0$.

Pravděpodobnostní funkce:

$$p(x) = \begin{cases} p(1-p)^x & \text{pro } x = 0, 1, \ldots \\ 0 & \text{jinde} \end{cases} \tag{23}$$

Vidíme, že rozdělení se nazývá právem geometrické, protože pravděpodobnosti $p(x)$ tvoří geometrickou posloupnost s kvocientem $q = 1 - p$ a prvním členem p. Pravděpodobnostní funkce tedy určuje pravděpodobnost toho, že první úspěch v sérii nezávislých pokusů nastane právě po x neúspěších.

Číselné charakteristiky.

střední hodnota: $$E(X) = \frac{1-p}{p} \tag{24}$$

rozptyl: $$D(X) = \frac{1-p}{p^2} \tag{25}$$

modus: $$\hat{x} \in 0 \tag{26}$$

Příklady:

(1) Ověřte, že funkce (23) je pravděpodobnostní funkcí. Nápověda: Využijte svých znalostí nekonečných řad z předmětu Matematika 3 a ukažte, že $\sum_{x \in \mathbb{R}} p(x) = 1$.

(2) Odvoďte vztah (24) pro $E(X)$.

(3) Odvoďte vztah (25) pro $D(X)$ a směrodatnou odchylku $\sqrt{D(X)}$.

(4) Vysvětlete, proč je v (26) vždy $\hat{x} = 0$.

Závěrem. Geometrické rozdělení poutá pozornost zkoušejícího proto, že řada výpočtů je založena na využívání studentovy znalosti (neznalosti) geometrické posloupnosti a řady.

Poznámka. Dosud uvedený přehled diskrétních rozdělení pravděpodobnosti je značně neúplný, ale dle našeho názoru dává studentovi dostatečné základy, aby se rychle dokázal seznámit s dalšími rozděleními a jejich aplikacemi. Proto na závěr této části uveďme jen několik názvů dalších rozdělení s krátkými komentáři: **Negativně binomické rozdělení** jednak zahrnuje jako speciální případy další rozdělení (Pascalovo, Pólyovo - modelování náhodných katastrof v klimatologii). Dále je lze použít jako robustní alternativu k Poissonovou, protože lze prostřednictvím jeho parametru kontrolovat jejich odchylku. Rovněž zobecňuje Ge rozdělení, protože je rozdělením počtu neúspěchů předcházejících $r$tý úspěch (u Ge $r = 1$) v Bernoulliovské posloupnosti nezávislých pokusů s pravděpodobností úspěch v jednotlivém pokusu rovnou p. **Zipfovo-Mandelbrotovo** rozdělení patří mezi diskutabilní empirická rozdělení. Používá se v jazykovědě a fanoušky fraktálů by mohl jméno v názvu vést k hlubšímu samostudiu. Naopak seriózně na fyziku zaměření čtenáři jistě prozkoumají **rozdělení Boltzmannovo** a související rozdělení Gibbsovo, Maxwellovo-Boltzmannovo, Boseho-Einsteinovo a Fermiho-Diracovo. Pro sportovní fanoušky a specialisty na rozpoznávání obrazu je připraveno **Skellamovo rozdělení**.

B. Typická spojitá rozdělení pravděpodobnosti

1. Rovnoměrné rozdělení

Značení. Skutečnost, že náhodná veličina X má rovnoměrné rozdělení, značíme $X \sim R(a, b)$, kde $a, b \in \mathbb{R}$ a $a < b$. V zahraniční literatuře najdeme značení $X \sim U(a, b)$, písmeno U je použito jako zkratka slova z anglického názvu rozdělení „uniform distribution".

Použití. Rovnoměrné rozdělení je vhodným modelem pro ty úlohy klasické pravděpodobnosti, kde jsme použili výpočet pomocí **geometrické pravděpodobnosti** (nezaměnit s geometrickým rozdělením!). Tj. použili jsme vzorec:

$$P(A) = \frac{\mu(A)}{\mu(\Omega)},$$

kde $\mu(\cdot)$ určuje délkovou, obsahovou nebo objemovou míru velikosti množiny (hlubší teoretické úvahy vynecháváme a odkazujeme na [5]).

Funkční charakteristiky. Z výše uvedeného vyplývají následující vztahy (42) a (28) pro hustotu rozdělení pravděpodobnosti (dále zkracujeme na „hustotu") a distribuční funkci.

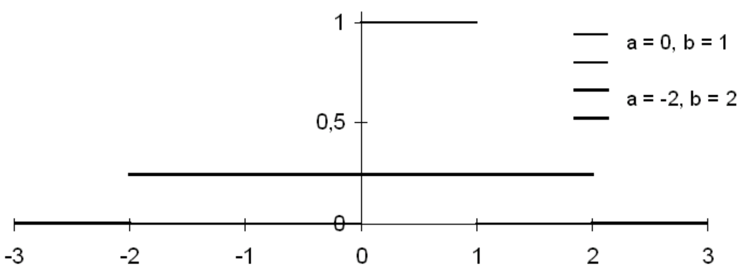
Hustota:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pro } \ x \in \langle a, b\rangle \\ 0 & \text{jinde} \end{cases} \tag{27}$$

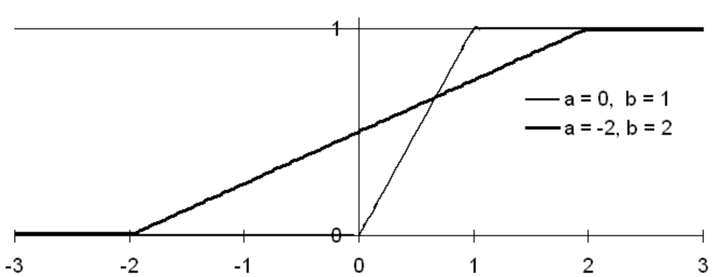
Distribuční funkce:

$$F(x) = \begin{cases} 0 & \text{pro } \ x \leq a \\ \frac{x-a}{b-a} & \text{pro } \ x \in (a, b\rangle \\ 1 & \text{pro } \ x > b \end{cases} \tag{28}$$

Výhodou je, že hodnoty funkcí $p(x)$ a $F(x)$ pro R rozdělení lze počítat přímo podle (42) a (28). Pro hustotu a distribuční funkci uvádíme vždy ukázky dvou jejich grafů (viz obr. 9 a 10) a volíme různé meze $\langle a, b\rangle$.



Obrázek 9: $p(x)$



Obrázek 10: $F(x)$

Číselné charakteristiky.

$$\textbf{střední hodnota:} \quad E(X) = \frac{a+b}{2} \tag{29}$$

$$\textbf{rozptyl:} \quad D(X) = \frac{(b-a)^2}{12} \tag{30}$$

$$\textbf{modus:} \quad \hat{x} \in \langle a, b\rangle \tag{31}$$

$$\textbf{koeficient šikmosti:} \quad A(X) = 0 \tag{32}$$

Název rozdělení. Název rovnoměrné rozdělení se užívá proto, že pravděpodobnost toho, že náhodná veličina X, která má $R(a,b)$ rozdělení, nabude hodnoty z intervalu určité délky (např. d), je **úměrná pouze délce tohoto intervalu** a nezávisí na jeho umístění uvnitř intervalu $\langle a, b\rangle$. Pojem **rovnoměrné rozdělení se obvykle zobecňuje i pro náhodný vektor**, viz např. [1].

Příklady:

(1) K přerušení optického kabelu o délce 500 m může dojít v libovolné vzdálenosti od jeho počátku, přičemž pravděpodobnost náhodného jevu, že dojde k přerušení v nějakém úseku je přímo úměrná délce úseku a nezávisí na jeho poloze. Určete rozdělení pravděpodobnosti náhodné veličiny X vyjadřující vzdálenost místa přerušení od počátku, její hustotu pravděpodobnosti a základní číselné charakteristiky a pravděpodobnost, že k přerušení kabelu dojde v úseku od 300 m do 400 m.

(2) V rovině jsou narýsovány rovnoběžky vzdálené od sebe střídavě $1,5$ a 8 cm. Na rovinu je náhodně vržen kruh o poloměru $2,5$ cm. Určete pravděpodobnost toho, že nebude protťat žádnou rovnoběžkou.

(3) Některé slovní úlohy, podle toho jak čtenář chápe slovo „náhodně", mohou vést k různým výsledkům i paradoxům. V případě zájmu doporučujeme seznámit se s diskusí v [5].

(4) **Buffonova úloha o jehle (18.století) viz obr. 11:** Na rovinu rozdělenou na pásy soustavou rovnoběžných přímek o vzdálenosti $2a$ se hází náhodným způsobem jehla délky $2b < 2a$. Určete pravděpodobnost, že jehla protne některou přímku soustavy, nepředpokládáme, že se jehla zapíchne.

Řešení: Poloha jehly je popsána pomocí úhlu, který jehla svírá s přímkou (náhodná veličina X), a vzdálenosti středu jehly od nejbližší rovnoběžky (náhodná veličina Y). Tyto veličiny lze považovat za stochasticky nezávislé a nabývají s pravděpodobností 1 svých hodnot z množiny $\Omega = \langle 0,\pi\rangle \times \langle 0, a\rangle$. Házíme-li jehlu náhodně, znamená to, že ji házíme tak, že všechny dvojice výsledků můžeme považovat za stejně možné, a tedy náhodný vektor $(X,Y)^\top$ má **dvojrozměrné rovnoměrné rozdělení** na obdélníku Ω. Protože jeho obsah je roven $a\pi$, je hustota vektoru $f(x,y) = \frac{1}{a\pi}$ na obdélníku a nulová jinde.

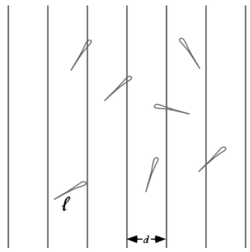
Nyní nás zajímá, kdy nastane jev A, kdy jehla protne přímku. S pomocí obrázku 11 vidíme, že to je tehdy, když je splněna podmínka $Y \le b\sin X$. Můžeme tedy určit, že $A = \{(x,y) \in \Omega \mid y \le b \sin x\}$. Nyní již zbývá jen dosadit do vztahů a vypočítat pravděpodobnost:

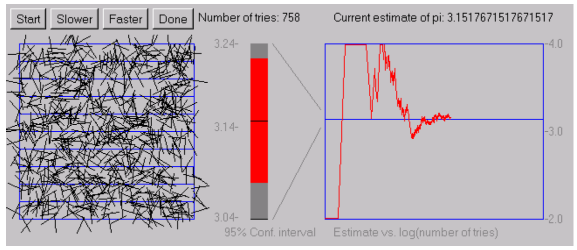
$$P((X,Y) \in A) = \int_A f(x,y)dxdy = \int_0^\pi \left(\int_0^{b\sin x} \frac{1}{a\pi} \right) dx = \frac{1}{a\pi} \int_0^\pi b \sin x = \frac{2b}{a\pi}.$$

Zajímavosti - metody Monte Carlo. Předešlý příklad (Buffonova jehla) a jeho výsledek vypadají poněkud vyumělkovaně a nepoužitelně. Zdání ovšem klame, protože historicky stojí u základů důležité skupiny výpočetních metod nazvaných „Monte Carlo", protože jsou založeny na generování posloupností náhodných čísel. Můžeme uvážit, že vypočtená pravděpodobnost

$$P((X,Y) \in A) = \frac{2b}{a\pi}. \quad \text{Logicky lze psát} \quad \pi = \frac{2b}{aP((X,Y) \in A)}.$$

A nyní přichází pointa. V klasické pravděpodobnosti se hovoří o tom, že pravděpodobnost P často odhadujeme statistickým experimentem (např. odhad pravděpodobnosti pádu rubu při házení mincí pomocí





Obrázek 11: hody jehly, počítačová simulace, Buffon na známce

relativní četnosti). Podobný experiment může nyní vést k pokusu o **experimentální odhad** čísla π. Označíme-li si n počet všech hodů jehly a m kolikrát protťala přímku (nastal jev A), potom po dosazení $\frac{m}{n}$ za P dostáváme:

$$\pi \text{ „je přibližně rovno" } \frac{2bn}{am}.$$

Poznamenejme, že v kapitolách věnovaných matematické statistice se čtenář dozví, v jakém smyslu je zde míněno „je přibližně rovno" a co je to „statistický odhad". Již nyní poznamenejme, že tento **odhad je tím přesnější, čím je n větší** a pro $n \longrightarrow \infty$ tento „odhad konverguje k π s pravděpodobností 1" (Zájemce, toužící po hlubším pochopení odlišností v chápání jemu známých numerických chyb a nově zmíněných statistických chyb, se bude muset začíst do podstatně náročnějšího textu [8], orientační pohled nabízí grafické znázornění počítačové simulace, viz obr.11).

Uvedená úloha je jednou z prvních úloh, které sloužily jako základ pro rozvoj výpočetních metod Monte Carlo. Principem těchto metod je, že řešený problém není aproximován numericky, ale je nahrazen „ekvivalentním" problémem stochastickým, a tedy lze hovořit i o tom, že část výpočtu je nahrazena náhodnými pokusy. **Význam těchto metod byl doceněn teprve se zavedením počítačů**, jistě únavné náhodné experimenty byly konány i předtím (nejen hazardní hráči různých středověkých her, ale např. je citováno opakované házení mincí prováděné Pearsonem).

U zrodu myšlenky využití výpočetní techniky pro náhodné generování (místo např. lidské ruky) stáli matematici Ulam a von Neumann (viz www zdroje). Základní otázka, která se od té doby řeší je, jak získat dostatečně dlouhou a reprezentativní posloupnost realizací z daného rozdělení.

Připomeňme, že zmíněné algebraické kongruenční generátory generují výběr z $C(n)$. Pokud ovšem zvolíme n velké a dělíme získané hodnoty n, získáme rozumný výběr z $R(0;1)$.

Pro řadu dalších rozdělení se pak používá obrat, který využívá distribuční funkce rovnoměrného rozdělení a vlastností transformací náhodných veličin. Umožňuje nám pak pomocí tzv. **kvantilové funkce** (získané z inverzní relace k distribuční funkci) transformovat rovnoměrně generované hodnoty na hodnoty reprezentující jiné rozdělení (viz např. odvození v [1], str. 95).

Zdůrazněme, že **metody Monte Carlo se používají v případech, kdy z důvodu složitosti nebo velké dimenze řešených úloh nelze použít klasické analytické nebo numerické postupy** (např. výpočet 10 rozměrného integrálu - viz aplikace ve spolehlivosti). Nelze ovšem očekávat, že poskytnou stejnou nebo dokonce lepší kvalitu tam, kde přesnější postupy selhaly. Proto je také nevhodné Monte Carlo metody používat tam, kde se osvědčily klasické postupy. Mezi typické oblasti jejich aplikací patří výpočty integrálů a hledání extrémů.

Příklady:

(5) Navrhněte postup výpočtu dvojného integrálu (např. obsah kruhu) pomocí metody Monte Carlo.

(6) Seznamte se s problematikou transformace náhodných veličin a kvantilových funkcí (značíme F^{-1}, protože pro rostoucí $F(x)$ je funkcí inverzní) a ukažte, že $X \sim R(0,1)$ a pro rostoucí distribuční funkci F náhodné veličiny Y platí, že transformací $Y = F^{-1}(X)$ hodnot generovaných z rovnoměrného rozdělení získáme rozdělení s distribuční funkcí $F(x)$.

Závěrem. Z pohledu zkoušky se jedná o rozdělení pravděpodobnosti, které umožňuje vyzkoušet principy s pomocí jednoduchých výpočtů. Proto je **velmi důležité**. Lze jen doporučit **provést výpočty i pro po částech rovnoměrné rozdělení** (jeho hustota je po částech konstantní - graf připomíná histogram), které zde záměrně nebylo zmíněno.

2. Normální rozdělení

Značení. Skutečnost, že náhodná veličina X má normální rozdělení, značíme $X \sim N(\mu, \sigma^2)$, kde $\mu, \sigma^2 \in \mathbb{R}$ a $\sigma^2 > 0$. V zahraniční literatuře najdeme názvy rozdělení: Gaussovo (Německo), Laplaceovo (Francie - to ovšem správně vyhrazujeme pro dvojitě exponenciální rozdělení). My se přidržíme „česko-anglického” názvu. Závěrem poznamenejme, že pro náhodný vektor se zavádí vícerozměrné normální rozdělení.

Použití. Normální rozdělení je používaným modelem v aplikacích, které interpretují **náhodné výsledky jako aditivní výsledek mnoha nezávislých vlivů** (např. chyba měření, odchylka rozměru výrobku od požadované hodnoty, apod.).

Funkční charakteristiky. Jestliže jsme se dosud odkazovali na předchozí znalosti, případně „zdravý rozum” čtenáře, teď pouze konstatujeme, že hustota normálního rozdělení pravděpodobnosti je definována níže. Pokud na laskavého a neinformovaného čtenáře působí vzorec jako blesk z čistého nebe a vede k reakci „Kde se to vzalo?”, je to reakce správná a odpovídá popularitě a jisté tajemnosti obklopující N rozdělení. My však budeme poměrně necitelní a část tajemství kolem tohoto rozdělení čtenářům odhalíme.

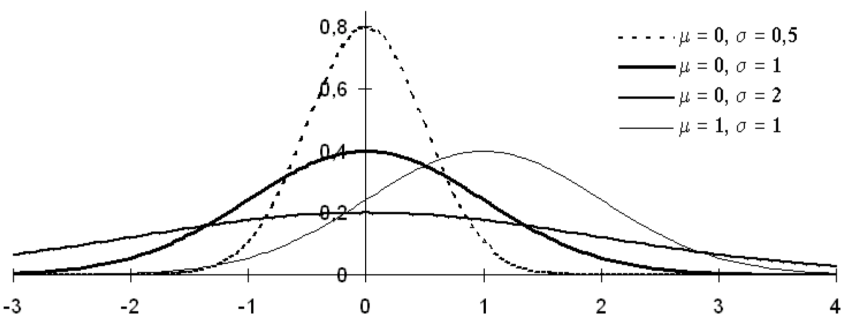
Hustota:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R} \tag{33}$$

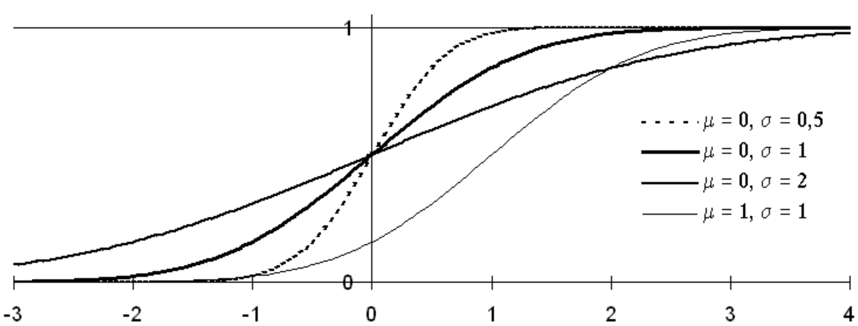
Distribuční funkce:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad x \in \mathbb{R} \tag{34}$$

Na obr. 12 jsou grafy hustot pravděpodobnosti a na obr. 13 grafy odpovídajících distribučních funkcí normálního rozdělení pro různé hodnoty parametrů μ a σ^2.



Obrázek 12: $p(x)$



Obrázek 13: $F(x)$

Číselné charakteristiky.

$$\begin{aligned} &\textbf{střední hodnota:} & E(X) &= \mu & (35)\\ &\textbf{rozptyl:} & D(X) &= \sigma^2 & (36)\\ &\textbf{modus:} & \hat{x} &= \mu & (37)\\ &\textbf{medián:} & x_{0,5} &= \mu & (38)\\ &\textbf{koeficient šikmosti:} & A(X) &= 0 & (39) \end{aligned}$$

Vlastnosti i příklady. Vlastnosti hustoty normálního rozdělení lze zkoumat i z pohledu matematické analýzy (viz Matematika I – průběh funkce). Doporučujeme čtenáři ověřit vše vlastním výpočtem:

1) $f(x)$ je **spojitá funkce** nabývající vesměs kladných hodnot (kdo chce hlubší rozbor a komplikace, nechť „do toho šlápne" a otevře např. [7]).

2) $f(x)$ **je symetrická podle osy** $x = \mu$ (viz $x_{0,5}$).

3) $f(x)$ je diferencovatelná na $\mathbb{R}$ (má navíc derivace všech řádů), rostoucí pro $x < \mu$, klesající pro $x > \mu$. V bodě $x = \mu$ nabývá svého **lokálního maxima**, které je zároveň globálním (viz $\hat{x}$).

4) Asymptotické chování $f(x)$ je popsáno vztahy $\lim\limits_{n \longrightarrow -\infty} f(x) = 0$ a $\lim\limits_{n \longrightarrow \infty} f(x) = 0$.

5) $f(x)$ má inflexní body $\mu - \sigma$ a $\mu + \sigma$. Je konkávní na intervalu $\langle \mu - \sigma, \mu + \sigma \rangle$ a konvexní vně tohoto intervalu. Připomínáme samozřejmost, že $\sigma = \sqrt{\sigma^2}$, která později u jiných řeckých písmen a jiných rozdělení nemusí být tak zřejmá.

6) Jednou z posledních otázek, nikoliv však významem, je: „proč v zápise $F(x)$ vystupuje integrál? Proč není vypočítaný?" Odpověď je prostá: **K integrálu, kterým je dána distribuční funkce $F(x)$, neexistuje primitivní funkce v konečném tvaru**. K výpočtu hodnoty $F(x)$ lze ale použít např. nekonečné řady.

7) V aplikacích (při řízení jakosti výroby, apod.) se často užívá tzv. **pravidlo tří sigma** (plus/minus 3σ), založené na tom, že:

$$P(X \in \langle \mu - 3\sigma, \mu + 3\sigma \rangle) = F(\mu + 3\sigma) - F(\mu - 3\sigma) = \int\limits_{\mu - 3\sigma}^{\mu + 3\sigma} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}\, dx = 0,997$$

Toto pravidlo znamená, že při velkém počtu pozorování náhodné veličiny X s normálním rozdělením $N(\mu, \sigma^2)$ můžeme očekávat, že cca 99,7% pozorovaných hodnot x bude ležet v intervalu $\langle \mu - 3\sigma, \mu + 3\sigma \rangle$.

Normované (základní, standardní) normální rozdělení. Velmi důležitým případem je situace, kdy $\mu = 0$ a $\sigma^2 = 1$, tedy $U \sim N(0; 1)$. Říkáme, že **náhodná veličina U má normované normální rozdělení**. Značení odlišující obecné a normované rozdělení je vhodné při uvádění dalších poznatků.

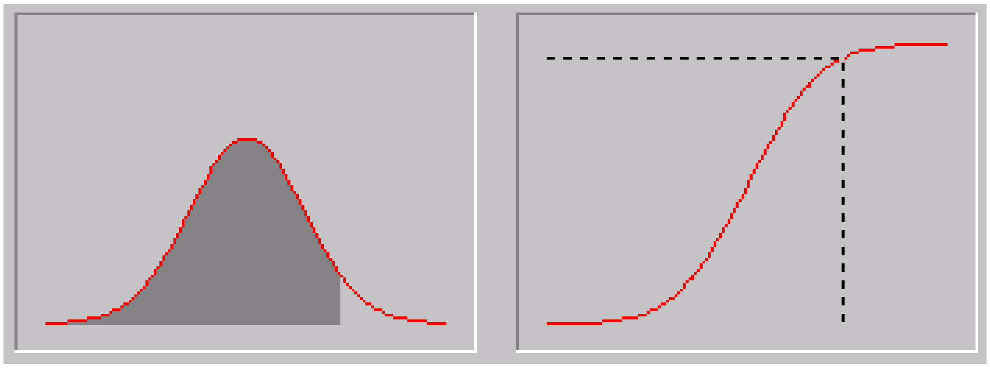
Funkční charakteristiky. Hustotu $\varphi(u)$ (40) i distribuční funkci $\Phi(u)$ (41) normovaného normálního rozdělení N(0;1) získáme dosazením $\mu = 0$ a $\sigma^2 = 1$ do (33) a (34):

Hustota:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{u^2}{2}}, \qquad u \in \mathbb{R} \tag{40}$$

Distribuční funkce:

$$\Phi(u) = \int\limits_{-\infty}^{u} \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{t^2}{2}}\, dt, \qquad u \in \mathbb{R} \tag{41}$$



Obrázek 14: software Statistica: $U \sim N(0;1)$, $\varphi(u)$, $\Phi(u)$, u_P.

Vlastnosti. Vlastnosti hustoty normálního rozdělení $N(\mu, \varsigma^2)$ uvedené výše lze přímo aplikovat na normované normální rozdělení $N(0;1)$: $E(X) = \hat{x} = x_{0,5} = 0, D(X) = \sqrt{D(X)} = 1, A(X) = 0$.

1) Základní otázkou je, **zda pro výpočty $\Phi(u)$ a u_P potřebujeme počítat výše uvedené integrály**. Překvapivá odpověď zní, že nikoliv. K výpočtu hodnot $\Phi(u)$ a kvantilů u_P **lze použít statistické tabulky** (viz `www.mat.fme.vutbr.cz`, kde hodnoty distribuční funkce $\Phi(u)$ normované náhodné veličiny U jsou tabelovány v tabulce T1; další zdroje jsou [2], [14], [13]). Je možné rovněž použít vhodný software (např. Statistica, Statgraphics, Minitab, Excel aj.). Obvykle lze s výhodou redukovat rozsah tabulek a využít symetrie zmíněné dříve, díky které platí $\Phi(-u) = 1 - \Phi(u)$ a pro kvantily U platí $u_{1-P} = -u_P$, $0 < P < 1$. Otázkou tedy je, jak určit obecné hodnoty $F(x)$ a x_P. Všimněme si dalších poznatků:

2) Jestliže náhodná veličina X má normální rozdělení $N(\mu, \sigma^2)$, pak náhodná veličina $Y = aX + b$, kde $a, b \in \mathbb{R}$, $a \neq 0$, má normální rozdělení $N(a\mu + b, a^2\sigma^2)$.

3) Aplikací 2) zjistíme, že transformací náhodné veličiny X s normálním rozdělením $N(\mu, \sigma^2)$ na náhodnou veličinu $U = \frac{X-\mu}{\sigma}$ dostaneme normované normální rozdělení $N(0;1)$ s distribuční funkcí $\Phi(u)$.

4) **Pro výpočty je důležité**, že jestliže náhodná veličina X má normální rozdělení $N(\mu, \sigma^2)$, potom její distribuční funkce $F(x) = \Phi(\frac{x-\mu}{\sigma})$ a její kvantily jsou $x_P = \mu + \sigma u_P, 0 < P < 1$.

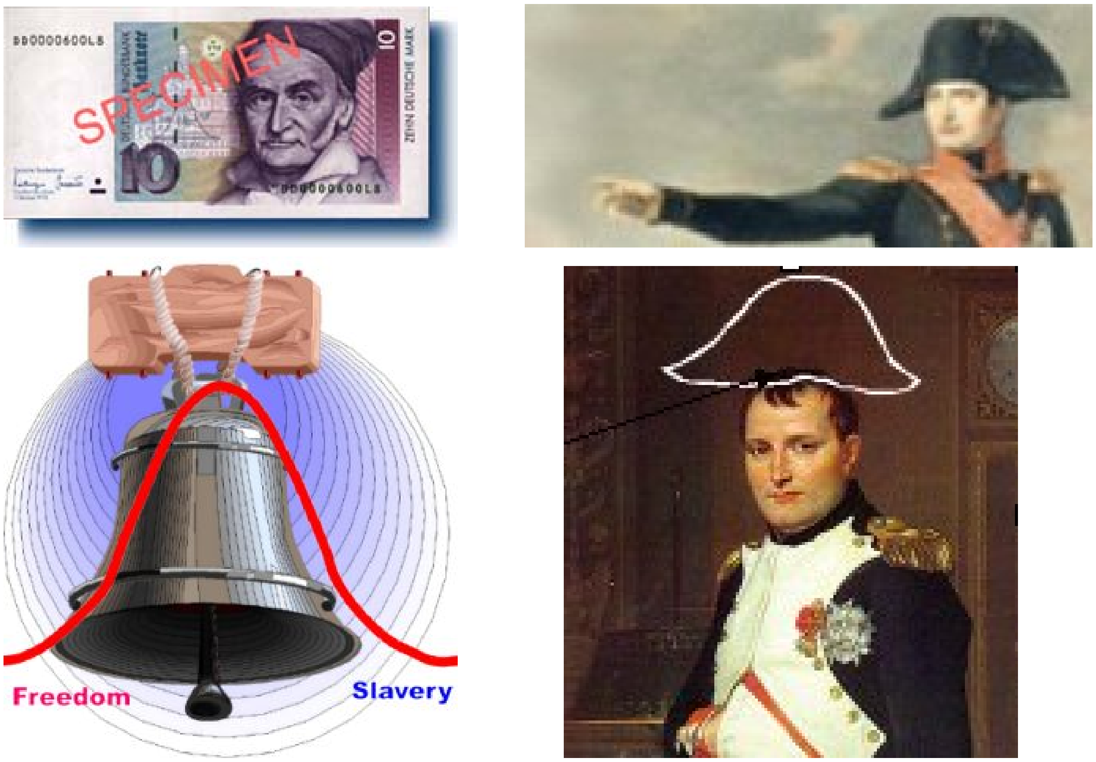
5) Stačí tedy **převést problém výpočtu pravděpodobnosti $P(X \in B)$ pro $X \sim N(\mu, \sigma^2)$ na hledání hodnot distribuční funkce $F(x)$, po transformaci $\Phi(u)$ a nakonec hledat v tabulkách**.

Příklady:

(1) Pro náhodnou veličinu X s rozdělením $N(0;1)$, určete pravděpodobnost, že nabude hodnoty: a) větší než $2,68$, b) menší nebo rovné $1,73$, c) větší nebo rovné $-0,66$, d) menší než $-1,88$, e) v mezích od $1,05$ do $1,65$, f) v mezích od $-0,05$ do $1,05$. Řešte pomocí hledání ve statistických tabulkách (viz T1).

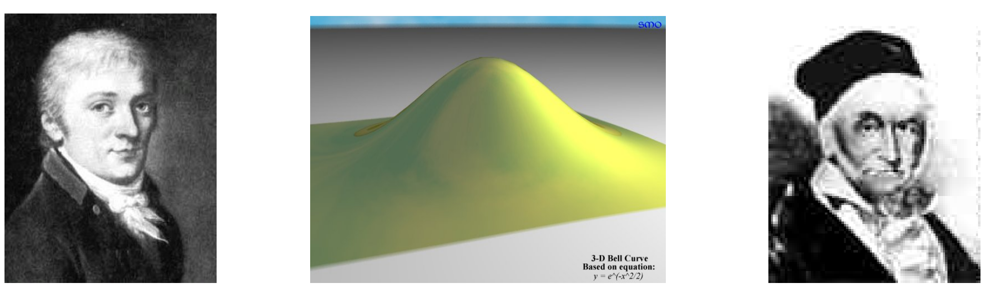
(2) Životnost baterie je náhodnou veličinou s normálním rozdělením s parametry $\mu = 300$ hod. a $\sigma = 35$ hod. Určete a) pravděpodobnost, že baterie bude mít životnost větší než 320 hodin, b) jakou hodnotu překročí životnost baterie s pravděpodobností větší než $0,75$.

Zajímavosti. Vraťme se nyní k diskusi o názvu rozdělení. Nejprve připomeňme některé poutavé názvy: Neutrální název *graf hustoty normálního rozdělení* bývá v souvislosti s IQ testy anglicky píšícími autory nahrazován názvem „**the bell curve**" (zvonovitá křivka), viz obr. 15. Francouzi používají název **křivka policejního (Napoleonského) klobouku** viz obr. 15. Vynechejme dále neortodoxní názvy a spíše se zeptejme na to, který název byl použit jako první, zda normální nebo Gaussovo rozdělení. Jestliže čekáme, že různé názvy prosazovaly různé skupiny jejich příznivců, nebo dokonce Gauss sám, hluboce se mýlíme. Způsobil to jeden člověk, byť statistický velikán – Karl Pearson, ve svých pracích z období 1893-97 dospěl od pojmenování normální křivky k normálnímu rozdělení, aby v roce 1905 uvedl pojem Gaussova křivka. A dvojznačnost byla na světě. Navíc ještě nedávno byla podložena „tvrdou měnou". (Poznamenejme, že přetisk sice bankovku hyzdí, ale dovoluje zveřejnění obrázku 15. Věřím, že zajímající se čtenář najde na internetu i původní vyobrazení.) Další logickou otázkou může být, kdo si všiml normálního rozdělení, byť tehdy bez názvu, jako první. Pierre-Simon Laplace (1749-1827) francouzský matematik zabývající se teorií pravděpodobnosti více než 50 let je zmínil v „Théorie Analytique des Probabilités" v roce 1812.



Obrázek 15: zvonová křivka, dvakrát napoleonský klobouk, „Gaussova” bankovka

Carl Friedrich Gauss (1777-1855) jeden z největších matematiků všech dob sice zveřejnil své poznatky později, ale doložitelně se rozdělením zabýval kolem roku 1809, kdy postavil původně Legendreovu **metodu nejmenších čtverců** na pevné základy analýzy normálního rozdělení chyb. Protože své výsledky ale poskytl ve značně mlhavé podobě, Legendre jej obvinil z plagiátorství. Proto se Laplace v pracích období 1810-1812 snažil Gaussovy postupy upřesnil. Nicméně pro Gausse v dnešní době plné potřeby ohlasu v médiích hovoří fenomenální úspěch při předpovědi, kdy na základě mála historických dat a svých znalostí nebeské mechaniky **určil kde očekávat tehdy objevenou planetku Ceres**. Dramatický příběh začíná 1.1. 1801, kdy italský astronom Piazzi spatřil nový světlý bod. Proběhla pozorování, formulace předpovědí kde bude ke spatření příště, ale v létě 1801 nikdo nic nenašel. Až v prosinci se objevil text mladého Gausse, kterým předčil další velká jména (Eulera, Laplace, Lagrange, atd.) a navrhl hledat planetku úplně jinde (6 úhlových stupňů se tak jeví v astronomii). 24letý mladík uspěl a „trefil se”. Podle vlastního sdělení použil při výpočtu i metodu nejmenších čtverců, kterou objevil již v 18ti letech. Sláva byla zaručena a doživotní pozice profesora na universitě v Göttingenu také. Takže jak dále? Budeme se konzervativně držet neutrálního názvu **normální**, protože po Gaussovi je pojmenována řada dalších pojmů.



Obrázek 16: Různě staří Gaussové a hustota normálního rozdělení pro (X, Y).

Závěrem. Asi se shodneme se čtenářem, že můžeme shrnout, že **normální rozdělení je velmi důležité, ale jeho používání je vlastně docela snadné** (potřebujete statistické tabulky a počítat lineární transformace). N rozdělení především představuje příprava na matematickou statistiku. Samozřejmě v případě snahy pochopit všechny hluboké souvislosti je nutné kvalitní matematické zázemí.

3. Exponenciální rozdělení

Značení. Skutečnost, že náhodná veličina X má exponenciální rozdělení, značíme $X \sim Exp(A, \delta)$, kde $A \in \mathbb{R}$ a $\delta > 0$. V literatuře a programech se vyskytuje i jednoparametrická varianta rozdělení, kdy $A = 0$.

Použití. Exp rozdělení je vhodným modelem, kdy nás zajímá rozdělení doby X do poruchy nějakého zařízení, a přitom pravděpodobnost poruchy v následujících x hodinách není ovlivněna předcházející historií zařízení. Rozdělení se také nazývá rozdělením „bez paměti". **Exponenciální rozdělení popisuje dobře rozdělení života zařízení, u nichž dochází k poruše z náhodných příčin**, nikoliv v důsledku opotřebení nebo únavy materiálu.

Funkční charakteristiky. Uvádíme pouze hustotu $f(x)$, distribuční funkci $F(x)$ doporučujeme čtenáři spočítat samostatně. Rovněž doporučujeme, aby si čtenář sám načrtl grafy $f(x)$ a $F(x)$ pro vhodnou volbu parametrů rozdělení.

Hustota:

$$f(x) = \begin{cases} \frac{1}{\delta} e^{-\frac{x-A}{\delta}} & \text{pro } x > A \\ 0 & \text{jinde} \end{cases} \tag{42}$$

Číselné charakteristiky.

střední hodnota:	$E(X)$	$=$	$A + \delta$	(43)
rozptyl:	$D(X)$	$=$	δ^2	(44)
koeficient šikmosti (asymetrie):	$A(X)$	$=$	2	(45)

Příklady:

(1) Kromě vlastního odvození $f(x), F(x), E(X), D(X), A(X)$ doporučujeme odvodit $x_{0,5}$.

(2) Určete rozdělení transformované náhodné veličiny $Y = \frac{X-A}{\delta}$, když víte, že $X \sim Exp(A, \delta)$.

(3) Projděte si dosud probraná rozdělení pravděpodobnosti a zamyslete se nad tím, jak souvisí parametry rozdělení (obvykle zadané) s číselnými charakteristikami (obvykle vypočtenými na základě vzorců).

(4) Seznamte se s **Laplaceovým rozdělením** (dvojitě exponenciálním, rozdělením „s těžkými konci") a po probrání regresní analýzy a metody nejmenších čtverců hledejte k jakému kritériu povede situace, kdy chybové členy budou mít Laplaceovo rozdělení.

Závěrem. Exponenciální rozdělení je vyučujícími velmi ceněno, protože je vhodné pro zadání ke zkoušce.

4. Weibullovo rozdělení

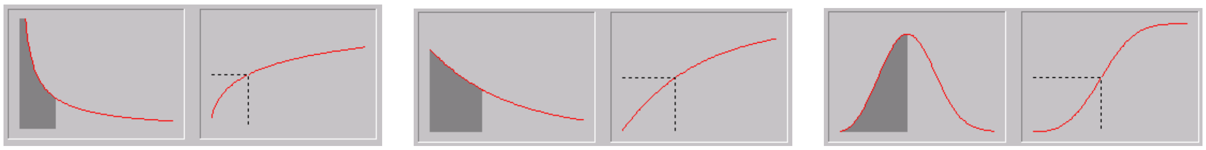
Značení. Skutečnost, že náhodná veličina X má Weibullovo rozdělení, značíme $X \sim W(\delta, b, k)$, kde $b \in \mathbb{R}$ a $\delta, k > 0$. V literatuře a programech se vyskytuje i dvouparametrická varianta rozdělení, kdy $b = 0$.

Použití. Weibullovo rozdělení je vhodným modelem pro úlohy, kdy zkoumáme životnost nějakého zařízení. Jestliže X značí životnost zařízení, **pro $k > 1$ je vhodným modelem pro životnost zařízení podléhajícího opotřebení nebo únavě, pro $k < 1$ je vhodným modelem pro životnost zařízení, u něhož dochází k poruchám v důsledku vad**, pro $k = 1$ dostaneme exponenciální rozdělení (viz jeho použití).

Funkční charakteristiky. Uvádíme hustotu $f(x)$, pro čtenáře může být zajímavým úkolem najít zdůvodnění jejího tvaru. Dále uvádíme grafy $f(x)$ a $F(x)$ pro $b = 0$, $\delta = 1$ a $k = 0,5; 1; 3$. Číselné charakteristiky neuvádíme, řada z nich je vyjádřena pomocí Γ funkce, a proto jejich vyhledání ponecháváme na zajímajícím se čtenáři.

Hustota:

$$f(x) = \begin{cases} \frac{k}{\delta}(x-b)^{k-1} & \text{pro } \; x \geq b \\ 0 & \text{jinde} \end{cases} \tag{46}$$



Obrázek 17: software Statistica: Weibullovo rozdělení $b = 0, \delta = 1$ a bráno zleva $k = 0,5; 1; 3$.

Zajímavosti. Švédský statistik Wallodi Weibull (1887-1979) působil mnoho let jako námořní důstojník a inženýr zabývající se zkoumáním mořského dna. První článek o W rozdělení publikoval v roce 1939. Přes mnohá ocenění, která obdržel, jeden z jeho žáků v roce 2000 konstatoval, že pouze 3 univerzity v USA zahrnuly do výuky statistiky Weibullovo rozdělení (na FSI VUT je vyučováno a používáno od osmdesátých let minulého století ...).

Závěrem. Weibullovo rozdělení je uvedené jako důležité více pro inženýrskou budoucnost čtenáře než ve vztahu ke zkoušce.

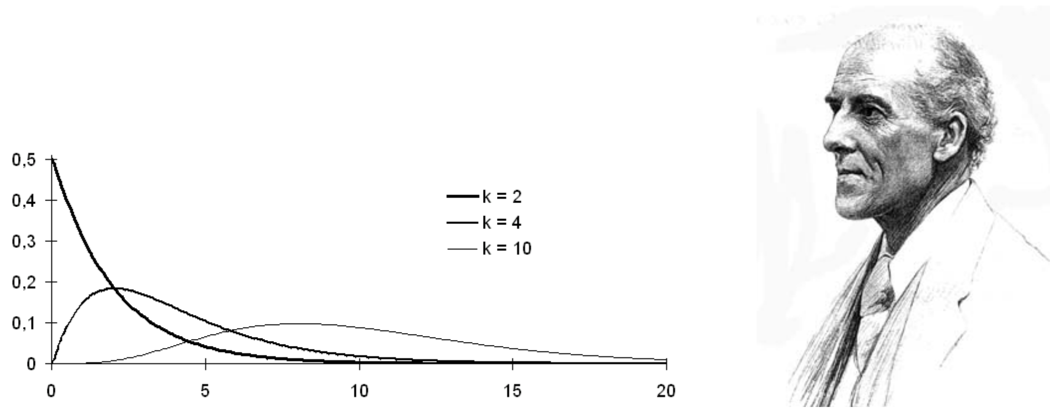
5. Některá další spojitá rozdělení

Poznámka. Podobně jako u diskrétních rozdělení i v tomto odstavci uvedeme názvy a aplikační oblasti pro některá spojitá rozdělení pravděpodobnosti. **Cauchovo** (Lorentzovo) rozdělení se používá ve fyzice, pro nás má tu zajímavou vlastnost, že $E(X)$ není definovaná (ověřte sami, viz [20]). Další aplikace v přírodních a technických vědách mají Maxwellovo a Rayleighovo rozdělení. **Beta** rozdělení a **trojúhelníkové** rozdělení se často používají při řízení projektů metodami stochastické optimalizace. V jiné oblasti operačního výzkumu (teorie front) se používá **Erlangovo** rozdělení, které je speciálním případem **Gamma** rozdělení, které je zobecněním několika dalších rozdělení. Zajímavou konstrukcí jsou **useknutá rozdělení** (např. useknuté normální), která se snaží vyloučit odlehlé nerealistické hodnoty. **Paretovo** (Bradfordovo) rozdělení bylo určeno pro ekonomické aplikace a původně popisovalo distribuci bohatství ve společnosti.

6. Pearsonovo χ^2 rozdělení

Značení. Skutečnost, že náhodná veličina X má Pearsonovo χ^2 rozdělení, značíme $X \sim \chi^2(k)$, kde $k \in \mathbb{N}$. Parametr k se nazývá počet stupňů volnosti a ve statistice je obvykle volen jako počet nezávislých informačních jednotek (např. počet měření minus počet odhadnutých parametrů). Zdůrazněme, že χ^2 je jeden symbol!

Použití. Toto a další dvě rozdělení (t a F) jsou **příklady rozdělení používaných později v matematické statistice** (viz intervalové odhady a testování hypotéz). Používají se zejména jejich kvantily, ale jejich **potřebné hodnoty jsou buď tabelovány (viz T3 v [2]) nebo je poskytne statistický software**. Můžeme tedy říci, že pro studenty Matematiky 4 by bylo uvedení hustot rozdělení pravděpodobnosti pouze informativní a proto uvedeme jen obrázky hustot a jakou transformací příslušné rozdělení vzniká, zájemce odkazujeme na [2].



Obrázek 18: Grafy hustot χ^2 rozdělení a jeho autor.

Vlastnosti. Jestliže $U_1, ..., U_k$ jsou nezávislé náhodné veličiny s normovaným normálním rozdělením $N(0;1)$, pak náhodná veličina $\sum_{i=1}^{k} U_i^2$ má Pearsonovo rozdělení $\chi^2(k)$. Jeho základní číselné charakteristiky jsou $E(X) = k, D(X) = 2k$.

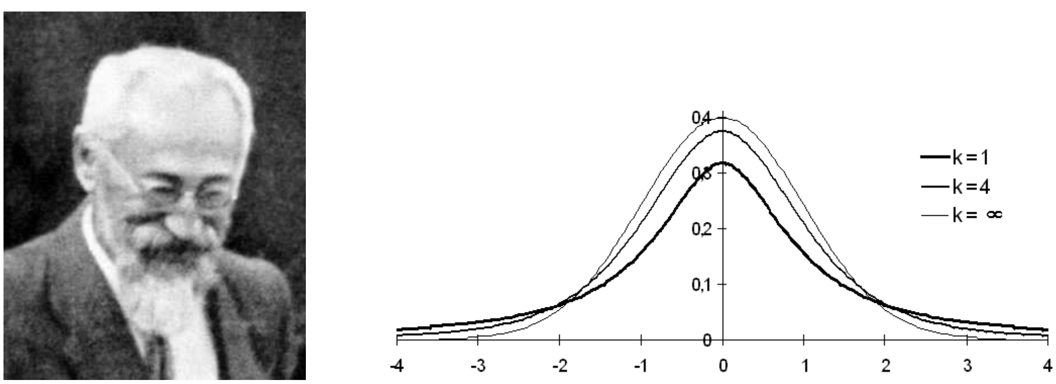
Zajímavosti. Karl Pearson (1857-1936) byl anglický statistik, žák sira Francise Galtona. Důležité jsou jeho výsledky v oblasti korelační a regresní analýzy, v klasifikací rozdělení pravděpodobnosti vcetně zavedení χ^2 rozdělení pro χ^2 test.

Závěrem. Pearsonovo χ^2 rozdělení je velmi důležité pro matematickou statistiku, a proto je nutné si osvojit hledání hodnot kvantilů v tabulkách.

7. Studentovo t rozdělení

Značení. Skutečnost, že náhodná veličina X má Studentovo t rozdělení, značíme $X \sim S(k)$, kde $k \in \mathbb{N}$.

Použití. Studentovo t rozdělení je příkladem rozdělení používaného později v matematické statistice (viz intervalové odhady a testování hypotéz). Používají se zejména jeho kvantily, ale **potřebné hodnoty jsou buď tabelovány (viz T2 v [2]) nebo je poskytne statistický software**. Opět vynecháme zavedení hustoty, zájemce odkazujeme na [2].



Obrázek 19: „Student" a grafy hustot t rozdělení.

Vlastnosti. Graf hustoty t rozdělení je symetrický vzhledem k x = 0. Jeho základní číselné charakteristiky jsou: $E(X) = 0$ pro $k > 1$ (pro $k = 1$ dostaneme Cauchyho rozdělení), $D(X) = \frac{k}{k-2}$ pro $k > 2$, $A(X) = 0$ pro $k > 3$, $x_{0,5} = 0$. Jestliže U a V jsou nezávislé náhodné veličiny, přičemž U má normované normální rozdělení $N(0;1)$ a V má Pearsonovo rozdělení $\chi^2(k)$, pak náhodná veličina $\frac{U}{\sqrt{V}}\sqrt{k}$ má Studentovo rozdělení $S(k)$.

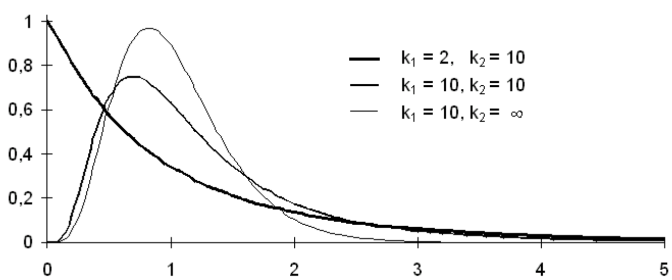
Zajímavosti. Čtenáři jistě vrtá hlavou, kdo je ten vousatý pán na obrázku a zda se jmenoval Student. Odpověď je nikoliv, byl to pan William Gosset (1876-1937), který publikoval pod pseudonymem Student. Naskýtá se otázka proč, když jména význačných statistiků nacházíme u mnoha pojmů. Proč ta skromnost, když jeho rozdělení umožňuje efektivně pracovat v rámci matematické statistiky s malými statistickými soubory (viz t-test)? Odpovědí může být dnes již legendární bonmot anonymního kolegy: „Výzkum se dělí na publikovatelný a použitelný a obvykle jedno vylučuje druhé." Pan Gosset totiž **pracoval jako sládek v pivovaru Guiness a po svých studijních cestách si nepřál, aby se vědělo, že jeho pivovar používá statistické metod**. A tak pro některé v Anglii a Evropě bylo skutečné jméno Studenta takovým tajemstvím, jako pro jiné skutečná totožnost Jacka Rozparovače. Naštěstí dnes víme více alespoň o panu Gossetovi a můžeme smeknout před jeho aplikačními schopnostmi nad sklenici Guinessu.

Závěrem. Studentovo t rozdělení je velmi důležité pro matematickou statistiku, a proto je nutné si osvojit hledání hodnot kvantilů v tabulkách.

8. Fisherovo-Snedecovorovo F rozdělení

Značení. Skutečnost, že náhodná veličina X má Fisher-Snedecorovo F rozdělení, značíme $X \sim F(k_1, k_2)$, kde $k_1, k_2 \in \mathbb{N}$.

Použití. F rozdělení je příkladem rozdělení používaného později v matematické statistice (viz intervalové odhady a testování hypotéz). Používají se zejména jeho kvantily, ale **potřebné hodnoty jsou buď tabelovány (viz T4 v [2]) nebo je poskytne statistický software**. Opět vynecháme zavedení hustoty, zájemce odkazujeme na [2].



Obrázek 20: Jeden z autorů R.A. Fisher a grafy hustot F rozdělení

Vlastnosti. Jestliže V_1 a V_2 jsou nezávislé náhodné veličiny, přičemž V_1 má Pearsonovo rozdělení $\chi^2(k_1)$ a V_2 má Pearsonovo rozdělení $\chi^2(k_2)$, pak náhodná veličina $\frac{V_1/k_1}{V_2/k_2}$ má Fisherovo-Snedecorovo rozdělení $F(k_1, k_2)$.

Zajímavosti. Angličan Ronald Aylmer Fisher (1890-1962) patřil k nejvýznamnějším statistikům 20. století. Ze spolupracovníka se stal Pearsonovým rivalem. Zabýval se genetikou, uspořádal výsledky v oblasti testování hypotéz (proto F rozdělení), stál u zrodu metody maximální věrohodnosti, analýzy rozptylu, plánování experimentu a neparametrických testů.

George Waďell Snedecor (1882-1974) byl zakladatel prvního statistického ústavu v USA. Spolupracoval s Fisherem zejména během a v návaznosti na jeho návštěvu v USA, F test patří k jejich společným výsledkům.

Závěrem. F rozdělení je velmi důležité pro matematickou statistiku, a proto je nutné si osvojit hledání hodnot kvantilů v tabulkách.

C. Aproximace rozdělení

Úvodem. V této části se budeme krátce zabývat možnostmi aproximovat jedno rozdělení pravděpodobnosti druhým. Konečně uvidíme, proč hraje normální rozdělení tak významnou roli.

K výpočtu hodnot pravděpodobnostních funkcí a distribučních funkcí binomického rozdělení $Bi(n,p)$, hypergeometrického rozdělení $H(N,M,n)$ a Poissonova rozdělení $Po(\lambda)$ na PC používáme vhodný software (např. Statgraphics, Statistica, Minitab, Excel aj.). Tato diskrétní **rozdělení však můžeme za jistých podmínek také vzájemně aproximovat** nebo aproximovat pomocí normálního rozdělení. Vycházíme přitom z vlastností uvažovaných rozdělení a tzv. limitních vět z teorie pravděpodobnosti (viz [1], [8]).

1. Aproximace hypergeometrického rozdělení

Postup. Pro malé hodnoty $\frac{n}{N}$ (přibližně $\frac{n}{N} < 0,1$) lze hypergeometrické rozdělení $H(N,M,n)$ aproximovat rozdělením binomickým $Bi(n,p)$ s parametry n a $p = \frac{M}{N}$.

Pro $\frac{n}{N} < 0,1$, $\frac{M}{N} < 0,1$ a $n > 30$ aproximujeme hypergeometrické rozdělení $H(N,M,n)$ Poissonovým rozdělením $Po(\lambda)$, kde položíme $\lambda = \frac{nM}{N}$.

Příklad:

(1) V dodávce 50 výrobků je 5 zmetků. Z dodávky jsou náhodně vybrány 3 výrobky. Počet zmetků mezi vybranými výrobky je náhodná veličina X. Předpokládejte nejprve, že se vybraný výrobek nevrací nazpět do dodávky, takže jde o náhodný výběr bez vracení a určete $p(x)$. Potom výpočet opakujte pro případ, že vybrané výrobky jsou vraceny. Obě pravděpodobnostní funkce porovnejte.

2. Aproximace binomického rozdělení Poissonovým

Postup. Binomické rozdělení $Bi(n,p)$, kde $p < 0,1$ a $n > 30$, aproximujeme Poissonovým rozdělením $Po(\lambda)$, kde položíme $\lambda = np$.

Možná se již čtenář osmělil a je připraven po tomto ryze praktickém návodu na trochu teorie. Nabídněme ji:

Věta (Poissonova): Mějme posloupnost $X_1, \ldots, X_n \ldots$ náhodných veličin, které mají rozdělení $Bi(j, p_j), j = 1, \ldots, n, \ldots$ Nechť $p_n \longrightarrow 0$ pro $n \longrightarrow \infty$ a přitom $\lim\limits_{n \longrightarrow \infty} np_n = \lambda$, kde $\lambda > 0$. Potom platí, že

$$\lim_{n \longrightarrow \infty} \binom{n}{x} p_n^x (1-p_n)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, \ldots$$

Můžeme pak říci, že za uvedených podmínek Poissonovo rozdělení je limitním rozdělením binomického rozdělení.

Prakticky to znamená, že pro dostatečně velké n a dostatečně malé p můžeme binomické rozdělení aproximovat rozdělením Poissonovým. Pro $n > 30$ a $p < 0,1$ se dopouštíme chyby menší než 10^{-2}.

Příklad:

(1) Opakujeme n-krát nezávisle tentýž pokus, jehož kladný výsledek má pravděpodobnost p.

a) Uvažujeme tři varianty hodnot n a p:

(1) n = 20, p = 0,2 ; (2) n = 200, p = 0,02; (3) n = 2000, p = 0,002.

Pro všechny tři varianty vypočtěte hodnotu pravděpodobnosti vždy pro tu hodnotu x, která je při daných n pokusech nejpravděpodobnější.

b) K témuž výpočtu použijte aproximaci pomocí Poissonova rozdělení pravděpodobnosti a výsledky porovnejte.

3. Aproximace binomického rozdělení normálním

Postup. Jestliže $np(1-p) > 9$, můžeme binomické rozdělení $Bi(n,p)$ náhodné veličiny X , aproximovat normálním rozdělením $N(\mu, \sigma^2)$, kde klademe $m = np$ a $\sigma^2 = np(1-p)$. Pro celá nezáporná čísla a, b potom je (s ohledem na skutečnost, že jde o diskrétní rozdělení)

$$P(a \le X \le b) \approx \Phi\left(\frac{b + 0,5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 0,5 - np}{\sqrt{np(1-p)}}\right).$$

Nabídněme nyní slavnou větu:

Věta (Moivre-Laplace): Nechť $X_1, \ldots, X_n \ldots$ je posloupnost vzájemně nezávislých náhodných veličin takových, že $\forall i \in \mathbb{N}: \; X_i \sim A(p)$. Pak náhodná veličina

$$Y_n = \frac{1}{\sqrt{np(1-p)}} \left(\sum_{i=1}^{n} X_i - np \right)$$

má asymptoticky normální rozdělení $N(0;1)$.

Protože $Z_n = \sum_{i=1}^{n} X_i$ je náhodná veličina s rozdělením $Bi(n,p)$, plyne z Moivre-Laplaceovy věty, že lze rozdělení Z_n aproximovat rozdělením $N(np, np(1-p))$. Vhodnost aproximace je tím lepší, čím je p blíží $0,5$. Zlepšuje se s rostoucím rozptylem a rovněž se zavádí oprava na spojitost (viz výše).

Příklad:

(1) Pravděpodobnost, že výrobek nebyl prověřen technickou kontrolou je $0,2$. Určete pravděpodobnost, že mezi 400 náhodně vybranými výrobky bude 70 až 100 neprověřených výrobků. Nápověda X má binomické rozdělení $Bi(400; 0,2)$, které aproximujeme normálním rozdělením $N(80; 64)$.

Otázka: Kam to dále vede? Jak asi čtenář tuší, stojí těsně před otevřením dveří do komnaty s nádhernou a náročnou matematikou. Může zůstat klidný, dveře neotevřeme, pouze kapitolu uzavřeme poznámkami, zájemce si cestu ke zdrojům informací jistě najde sám.

Formulace výše uvedené Moivre-Laplaceovy věty zní zajímavě, ale pozorný čtenář si všimne, že jsme neřekli co to je „asymptoticky” normální rozdělení. Učinili jsme tak záměrně.

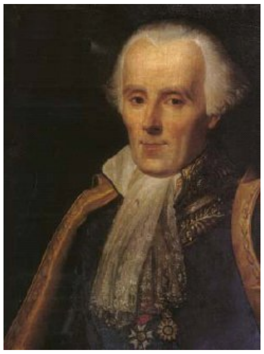
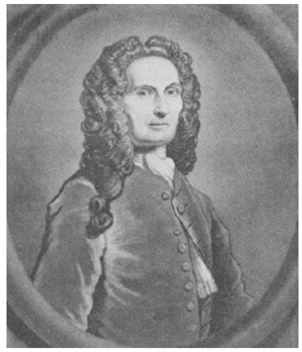
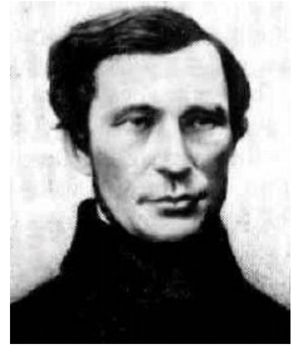
Zde naznačená oblast **souvisí s pojmem konvergence**, ale tentokrát náhodných veličin. Lze opět zavést řadu konvergencí: v distribuci a podle pravděpodobnosti [1].

Exaktni přístup pak umožnil formulovat **řadu zobecnění Moivre-Laplaceovy** věty. Poznamenejme, že existující obecnější verze se nazývají centrální limitní věty a dokonce se nepožaduje, aby sečítané náhodné veličinyX_i měly stejné rozdělení pravděpoodobnosti, a přesto za poměrně slabých předpokladů lze náhodnou veličinu $\sum_{i=1}^{n} X_i$ **aproximovat opět vhodným normálním rozdělením**. A to je příčina té časté frekvence, se kterou se N rozdělení objevuje v aplikacích. Můžeme parafrázovat Galtona: „Vše spěje k normálu.”

Koneckonců i řada dříve uvedených rozdělení při určitém chování parametrů opět konverguje k normálnímu rozdělení. Jako příklad uveďme, že Studentovo rozdělení S(k) konverguje k normovanému normálnímu rozdělení N(0; 1) pokud $k \longrightarrow \infty$.

Zajímavosti. Abraham de Moivre (1667-1754) uprchl do Anglie ve 20 letech. De Moivre studoval práci Huygense a od roku 1711 zveřejňoval své výsledky, nejznámější je kniha „The Doctrine of Chances: or, a Method of Calculating the Probability of Events in Play”. **De Moivre získal normální aproximaci k binomickému rozdělení** a byl blízko k nalezení Poissonova rozdělení.

Pierre-Simon Laplace (1749-1827) psal o pravděpodobnosti mnoho let. První knihou byla v roce 1774 „Mémoire sur la probabilité des causes par les évenemens”. Laplace byl velice aktivní (viz výše uvedená Moivre-Laplaceova věta). Jak jeho nástupci inovovali a zpřesňovali pojmy, jeho odkaz začal být zapomínán.

Obrázek 21: Pierre-S. Laplace (1749-1827), Abraham de Moivre (1667-1754), P. L. Čebyšev (1821-94)

P. L. Čebyšev (1821-94) byl jedním z nejdůležitějších matematiků 19. století. Položil základy ruské tradice v pravděpodobnosti, na kterou pak navazovali, Markov, Lyapunov, Kolmogorov, Chinčin, Širjajev a další. V roce 1867 Čebyšev zveřejnil práci obsahující tzv. **Čebyševovu nerovnost** a využil jí k důkazu tzv. Zákona velkých čísel. Zobecnil rovněž Moivre-Laplaceovu větu díky jinému přístupu v důkazu.

D. Cvičení

Základní okruhy otázek. Nápověda: K prohloubení znalostí lze využít i skripta [2] str. 85-95:
a) Funkční a číselné charakteristiky pro vybraná diskrétní rozdělení pravděpodobnosti: A, Bi, H, Po.
b) Typická slovní zadání pro uvedená diskrétní rozdělení pravděpodobnosti.
c) Funkční a číselné charakteristiky pro vybraná spojitá rozdělení pravděpodobnosti: R, N, E.
d) Typická slovní zadání pro uvedená spojitá rozdělení pravděpodobnosti.
e) Hledání ve statistických tabulkách a transformace výsledků pro N rozdělení.
f) Aproximace rozdělení pravděpodobnosti.

Kontrolní otázky.

1. Uveďte 3 příklady na Bi rozdělení a popište význam jejich parametrů i číselných charakteristik.
2. Uveďte 3 příklady na H rozdělení a popište význam jejich parametrů i číselných charakteristik.
3. Uveďte 3 příklady na Po rozdělení a popište význam jejich parametrů i číselných charakteristik.
4. Uveďte 3 příklady na N rozdělení a popište význam jejich parametrů i číselných charakteristik.
5. Proč se používají aproximace rozdělení pravděpodobnosti?

Typická zadání. Na základě slovního zadání určete typ rozdělení pravděpodobnosti a vypočítejte:
a) hodnoty pravděpodobnostní funkce (případně hustoty rozdělení pravděpodobnosti);
b) hodnoty distribuční funkce a načrtněte její graf;
c) vypočtěte vybrané číselné charakteristiky (např. střední hodnotu, modus, medián, rozptyl, atd.);
d) vypočtěte pravděpodobnost zadaného náhodného jevu.

Poděkování. Autor kapitoly děkuje doc.Zdeňku Karpíškovi za mnoho let cenných diskusí a za nezištné poskytnutí vlastních textů i rozsáhlého osobního archívu. Bez jeho pomoci by tato kapitola byla rozhodně fádnější a kratší.

Autor děkuje nejmenovaným norským kolegům, kteří mu pomohli obejít se bez rekonstrukce celého textu a zachránili většinu dat po havárii operačního systému počítače na zahraniční cestě a slibuje jim, že si opravdu bude průběžně zálohovat své soubory na další média.

Autor děkuje svým kolegům, autorům dalších kapitol této studijní opory věnované Matematice 4. Bez jejich včasného a inspirujícího zvládnutí zejména rozvržení a barevnosti jejich textů by **tato kapitola byla rozhodně „méně strakatá"**.

Reference

[1] Šikulová, M. - Zdeněk Karpíšek, Z.: Matematika 4. Pravděpodobnost a matematická statistika, 6. vyd., VUT Brno, 1997.
[2] Karpíšek, Z.: Matematika 4. Statistika a pravděpodobnost. CERM, Brno, 2003.
[3] Karpíšek, Z. - Popela, P. - Bednář, J.: Statistika a pravděpodobnost. Učební pomůcka - studijní opora pro kombinované studium, FSI VUT, Brno 2002.
[4] Karpíšek, Z. - Popela, P. - Bednář, J.: Statistika a pravděpodobnost - přehled vzorců a poznatků, Brno 2006.
[5] Rényi, A.: Teorie pravděpodobnosti. ACADEMIA, Praha, 1972.
[6] Likeš J. - Machek, J.: Počet pravděpodobnosti. SNTL Praha, 1981.
[7] Anděl J.: Matematická statistika, SNTL/Alfa Praha 1978.
[8] Anděl, J.: Matematika náhody, matfyzpress, Praha 2003.
[9] Zvára, K. - Štěpán, J.: Pravděpodobnost a matematická statistika. MAT.FYZ.press, MFF-UK, Praha, 1997.
[10] Swoboda, H.: Moderní statistika, Svoboda, Praha, 1977.
[11] Riečan B. - Lamoš F. - Lenárt C.: Pravdepodobnosť a matematická štatistika, Alfa/SNTL 1984.
[12] Štěpán ANO
[13] Janko, J.: Statistické tabulky. Praha, Nakladatelství ČSAV 1958
[14] Likeš, J. – Laga, J.: Základní statistické tabulky. Praha, SNTL/Alfa 1978.
[15] Montgomery, D. C. et al.: Probability and Statistics in Engineering, 4th Edition, Wiley, 2002.
[16] Koutková, H. - Moll,I.: Úvod do pravděpodobnosti a matematické statistiky. FAST VUT Brno, 2001.
[17] Čermák, V.: Diskrétní a spojitá rozdělení - vzorce, grafy, tabulky. Skripta. Praha, VŠE 1993.
[18] Rogalewicz, V.: Pravděpodobnost a statistika pro inľenýry, skripta, Vydavatelství ČVUT, 1998.
[19] Briš R. - Litschmannová M.: Statistika I. pro kombinované a distanční studium, Ostrava 2004.
[20] `http://www.wikipedia.org`
[21] `http://www-history.mcs.st-andrews.ac.uk/Biographies`

Anonym (Itálie, 16.století): „Čtěte! Vše již bylo napsáno Vašimi předchůdci."