Univerzita Pardubice

Dopravní Fakulta Jana Pernera

Generátory pseudonáhodných čísel rozdělení pravděpodobností

Tomáš Nadrchal

Bakalářská práce

2013

UPA046511

Univerzita Pardubice
Dopravní fakulta Jana Pernera
Akademický rok: **2012/2013**

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Tomáš Nadrchal**
Osobní číslo:	**D10014**
Studijní program:	**B3709 Dopravní technologie a spoje**
Studijní obor:	**Aplikovaná informatika v dopravě**
Název tématu:	**Generátory pseudonáhodných čísel rozdělení pravděpodobností**
Zadávající katedra:	**Katedra informatiky v dopravě**

Zásady pro vypracování:

Cílem práce je návrh programového nástroje na generování hodnot náhodné proměnné. Nástroj bude možné použít v oblasti statistického zpracování dat, v simulacích reálných systémů, případně ve výuce předmětů matematické statistiky a teorie pravděpodobností. Práce bude vycházet ze studia modelovacích technik, zejména techniky simulace. Diplomant provede kritickou analýzu dostupných nástrojů a možností jejich využití při modelování technologických procesů v dopravních a logistických systémech. Funkčnost navrženého simulátoru bude ověřena na konkrétních vybraných příkladech z praxe pokrývajících případy náhodné proměnné typických pro dopravní systémy.

Rozsah grafických prací:

Rozsah pracovní zprávy: **30 normostran**

Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam odborné literatury:

1. **KUBANOVÁ, Jana a Bohdan LINDA. Kritické hodnoty a kvantily vybraných rozdělení pravděpodobností. 2. vyd. Pardubice: Univerzita Pardubice, 2009, 53 s. ISBN 978-80-7395-222-8.**

2. **FORBES, Catherine, Merran EVANS, Nicholas HASTINGS a Brian PEACOCK. Statistical Distributions. 4th ed. Hoboken, New Jersey: Wiley, 2011, xviii, 212 p. ISBN 978-047-0390-634.**

3. **ČERNY, Jan,Pavol KLUVÁNEK. Zaklady matematickej teórie dopravy. Bratislava: Veda, vydavateľstvo Slovenskej akadémie vied, 1991, 279 p. ISBN 80-224-0099-8.**

Vedoucí bakalářské práce: **doc. Ing. Josef Volek, CSc.**
Katedra informatiky v dopravě

Datum zadání bakalářské práce: **6. prosince 2012**
Termín odevzdání bakalářské práce: **31. května 2013**

prof. Ing. Bohumil Culek, CSc.
děkan

L.S.

doc. Ing. Josef Volek, CSc.
vedoucí katedry

V Pardubicích dne 6. prosince 2012

Prohlašuji:

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 17. 3. 2013

Tomáš Nadrchal

PODĚKOVÁNÍ

Tímto bych chtěl poděkovat panu doc. Ing. Josefu Volkovi, CSc. za vedení práce, paní Věře Záhorové, Ph.D. a panu Ing. Filipu Víznerovi, Ph.D. za přínosné konzultace a panu Ing. Josefu Bulíčkovi, Ph.D. za poskytnutí souboru naměřených dob zpoždění.

ANOTACE

Práce se zabývá generováním pseudonáhodných čísel spojitých rozdělení pravděpodobnosti. Dále je práce doplněna výpočty hlavních charakteristik implementovaných rozdělení pravděpodobnosti a vykreslováním grafů hustot pravděpodobnosti a distribučních funkcí. Bakalářská práce může být použita uživateli z praxe při statistickém vyhodnocování dat, může též posloužit jako vhodná studijní opora pro posluchače předmětu teorie pravděpodobnosti a matematická statistika.

KLÍČOVÁ SLOVA

Generátory čísel, rozdělení pravděpodobností, náhodná veličina, náhodná čísla, pseudonáhodná čísla

TITLE

Pseudorandom number generators for statistical distributions

ANNOTATION

Bachelor´s thesis deals with the generation of pseudorandom numbers of continuous probability distributions. Further work is extended by the calculation of the main numerical characteristics of the implemented distributions and plotting of the graphs of probability density and distribution functions. Bachelor´s thesis can be used by users from practise for statistical evaluation of data, but it can also serve as suitable learning support for listeners of Probability Theory and Mathematical Statistics.

KEYWORDS

Number generators, probability distributions, random variable, random number, pseudorandom number

OBSAH

SEZNAM ILUSTRACÍ

SEZNAM TABULEK

0 ÚVOD

Tématem bakalářské práce je programový nástroj pro generování náhodných čísel dle spojitých rozdělení pravděpodobnosti. Ten bude možné připojit k projektům vyvíjeným v programovacím jazyku Object Pascal. Hodnoty náhodné proměnné lze generovat s tzv. násadou i bez ní. Ukázka funkčnosti je demonstrována na zkušební aplikaci, která umožňuje vygenerovat soubor pseudonáhodných čísel do textového souboru.

Zkušební aplikace je rozšířena o výpočty charakteristik implementovaných spojitých rozdělení pravděpodobnosti. Jednou z možností je výpočet střední hodnoty, rozptylu, směrodatné odchylky, distribuční funkce a hustoty pravděpodobnosti. Poslední dvě funkce jsou znázorněny i graficky, tedy spojnicovým grafem a jejich výpočet probíhá buď dynamicky, kdy jsou hodnoty počítány podle polohy kurzoru v oblasti grafu, nebo staticky, dle pevně zadané hodnoty. Druhou možností je porovnávání více rozdělení v jednom grafu. Jednotlivá rozdělení jsou barevně odlišena. Navíc je aplikace doplněna o kalkulátor složitějších funkcí, které jsou nezbytné pro výpočet některých charakteristik. Je to beta funkce, beta neúplná funkce, gama funkce, gama neúplná funkce, faktoriál a kumulativní distribuční funkce normálního normovaného rozdělení. Ve většině z nich se vyskytuje nevlastní integrál.

Tato aplikace může být tedy použita jak pro získání posloupnosti pseudonáhodných čísel řídících se požadovaným rozdělením pravděpodobnosti, tak i jako studijní opora v předmětech zabývajících se nejen statistikou a pravděpodobností.

Pro připomenutí problematiky je v dokumentaci věnována část se základním teoretickým aparátem, kde je vzpomenuta pravděpodobnost jako taková a její definice, problematika jednorozměrné náhodné veličiny a jejich charakteristik, dále jsou zmíněna implementovaná rozdělení pravděpodobnosti a je zakončena generováním náhodných čísel a jejich testováním.

Další dvě kapitoly se zabývají samotnou aplikací. První slouží jako návod pro její ovládání. Je doplněna o náhledy a poznámky, jež ještě více usnadní práci s přiloženou aplikací. Druhá má za úkol přiblížení vnitřní struktury programu, jednotlivých tříd a použitých datových struktur. To pomůže při vývoji aplikací za použití mého nástroje pro generování.

V poslední kapitole je ověřena kvalita generátoru. Výstupní hodnoty jsou porovnány s naměřenými dobami zpoždění, s generátorem čísel v tabulkovém procesoru Excel a generátorem aplikace Matlab.

1 TEORETICKÝ APARÁT

Cílem této kapitoly je připomenutí problematiky, kterou se práce zabývá. To pomůže uživateli k oprášení dříve nabytých znalostí, a tedy i ke snadnějšímu ovládání přiložené aplikace.

1.1 Pravděpodobnost

Matematické teorie studující náhodně se vyskytující jevy, případně náhodně kolísající veličiny, se nazývají **teorie pravděpodobnosti** a **matematická statistika**. Statistika vychází z teorie pravděpodobnosti a věnuje se sběru statistických dat, jejich zpracování a následným vyhodnocením. Teorie pravděpodobnosti se zabývá vytvářením tzv. pravděpodobnostních modelů. Odvozuje ze známého pravděpodobnostního modelu průběh odpovídajícího procesu. U statistiky se jedná o postup opačný, kdy je na základě sebraných dat odhadován alternativní model sledovaného procesu.

1.1.1 Axiomatická definice pravděpodobnosti

„Nechť Ψ je pole náhodných jevů. Libovolnou množinovou funkci P, definovanou na Ψ, nazveme **pravděpodobností** (pravděpodobnostní funkcí, pravděpodobnostní mírou, rozdělením pravděpodobnosti), jestliže splňuje tyto podmínky:

1. $A \in \Psi \Rightarrow P(A) \geq 0$
2. $P(S) = 1$
3. $\{A_i\}_i \subset \Psi, A_{i \cap A_j} = \emptyset, i \neq j \Rightarrow P(\bigcup_i A_i) = \sum_i P(A_i)$.“[1]

„Číslo $P(A)$ nazýváme **pravděpodobností náhodného jevu** A.“[1]

„Vztahům 1 – 3 říkáme **axiomy** pravděpodobnosti. Trojice (S, Ψ, P) se nazývá **pravděpodobnostní prostor**, někdy také **pravděpodobnostní pole**. Tento přístup k budování teorie pravděpodobnosti pochází od A. N. Kolmogorova, proto se uvedené trojici říká **Kolmogorovo pravděpodobnostní pole**.“[1]

Výše zmíněná definice je z matematického hlediska zcela korektní a poskytuje základ pro vybudování exaktní teorie pravděpodobnosti. Ovšem z praktického hlediska, tedy pro výpočet konkrétních hodnot pravděpodobností náhodných jevů, má pouze malé využití. Pro tyto výpočty je možné použít níže zmíněné modely.[1]

1.1.2 Klasická definice pravděpodobnosti

„U jistého typu náhodných pokusů lze při počítání pravděpodobností náhodných jevů použít model, který je v literatuře nazýván **klasická definice pravděpodobnosti**, vycházející z teoretických úvah o podstatě náhodného pokusu.“[1]

„Tento model předpokládá, že prostor všech elementárních jevů S má jenom konečný počet výsledků a každý elementární jev má stejnou možnost nastat po vykonání náhodného pokusu. Označíme-li $N[\cdot]$ počet elementárních jevů, kterými je tvořen náhodný jev uvedený v hranatých závorkách, tak potom pro pravděpodobnost $P(A)$ libovolného náhodného jevu platí

$$P(A) = \frac{N[A]}{N[S]}.\text{“[1]}$$

Číslo $N[A]$ nazýváme **počet příznivých případů** a číslo $N[S]$ jako **počet všech možných případů**.[1]

1.1.3 Geometrická definice pravděpodobnosti

„Klasickou definici pravděpodobnosti lze použít jenom pro pokusy mající konečný počet výsledků. Pro jistou třídu náhodných pokusů s nespočetně mnoha výsledky lze použít následující model, který je nazýván **geometrická definice pravděpodobnosti**. Uvažujeme náhodný pokus mající nespočetně mnoho výsledků, které lze interpretovat jako body n-rozměrného reálného prostoru R_n. Body reprezentující prostor všech elementárních jevů S potom vytvoří v R_n geometrický útvar nenulové míry $m(S)$. Náhodné jevy jsou podmnožiny S a budou reprezentovány geometrickými útvary ležícími v S. Dále se předpokládá, že jestliže $A \subset S$ je náhodný jev, tak možnost, že bod reprezentující výsledek náhodného pokusu bude ležet v množině A, je úměrná její míře $m(A)$ bez ohledu na její tvar a polohu v S. Potom pro pravděpodobnost náhodné jevu A platí

$$P(A) = \frac{m(A)}{m(S)}.\text{“[1]}$$

1.1.4 Statistická definice pravděpodobnosti

„Statistická definice pravděpodobnosti byla zavedena matematikem R. Misesem, který odmítal tzv. apriorní výpočet pravděpodobnosti, tj. výpočet, který se provádí na základě teoretického poznání struktury náhodného pokusu, aniž by byl pokus proveden. Pro výpočet pravděpodobnosti náhodného jevu použil tento model:“[1]

„Nechť n je počet opakování náhodného pokusu $\mathcal{P}$ a nechť m udává, kolikrát v dané sérii pokusů nastal náhodný jev A. Potom pravděpodobnost náhodného jevu A

$$P(A) = \lim_{n\to\infty} \frac{m}{n}.\text{“[1]}$$

„Tento model výpočtu (lépe odhadu, protože nelze provést nekonečně mnoho pokusů) pravděpodobnosti se v literatuře nazývá **statistická definice pravděpodobnosti**.“[1]

1.2 Jednorozměrná náhodná veličina

Předpoklad, že všechny náhodné elementární jevy náhodného pokusu nastanou se stejnou pravděpodobností, nemusí být vždy splněn. V těchto případech můžeme k popsání pravděpodobnosti použít právě **náhodné veličiny**. „Náhodná veličina je definována jako funkce, která každému elementárnímu jevu z množiny všech možných výsledků téhož pokusu přiřadí reálné číslo. Náhodné veličiny se obvykle značí velkými písmeny latinské abecedy, jejich konkrétní hodnoty písmeny malými.“[2]

„Nechť $(S,\ \Psi,\ P)$ je nějaký pravděpodobnostní prostor. **Jednorozměrnou náhodnou veličinou** definovanou na pravděpodobnostním prostoru $(S,\ \Psi,\ P)$ budeme nazývat zobrazení X základního prostoru S do množiny reálných čísel R $(X: S \to R)$ takové, že pro každé reálné x množina $X_{-1}((-\infty;\ x)) \epsilon \Psi$.“[1]

Navíc musíme rozlišovat mezi diskrétní a spojitou náhodnou veličinou. „Zjednodušeně lze říci, že **diskrétní náhodné veličiny** nabývají spočetně mnoha hodnot, kdežto všechny hodnoty **spojité náhodné veličiny** vyplňují na reálné přímce souvislý interval. V praxi se snažíme aproximovat každou náhodnou veličinu právě jedním z těchto typů.“[1]

1.3 Charakteristiky náhodné veličiny

„Znalost rozdělení pravděpodobnosti náhodné veličiny nám poskytuje všechny informace o této náhodné veličině. V případě, že rozdělení pravděpodobnosti má komplikovaný tvar, nemusí být snadné udělat si představu o jejím chování. Z tohoto důvodu se vybrané důležité vlastnosti náhodných veličin vyjadřují jedním číslem. Z velikosti těchto čísel si pak lze utvořit představu o náhodné veličině.“[1]

1.3.1 Střední hodnota

„Střední hodnota je jednou z charakteristik polohy (úrovně). Můžeme si ji představit jako souřadnici polohy těžiště obrazce, který odpovídá grafu rozdělení pravděpodobnosti uvažované náhodné veličiny.“[2]

U diskrétní náhodné veličiny je střední hodnota dána vztahem $E(X) = \sum_i x_i \cdot P(x_i)$.[2] Střední hodnota spojité náhodné veličiny se v podstatě neliší od té diskrétní. Musíme ale nahradit pravděpodobnost v bodě $P(x_i)$ pravděpodobností v okolí bodu, $P(x_i) \approx f(x)dx$. Střední hodnota spojité náhodné veličiny je potom dána vztahem $E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$.[3]

1.3.2 Rozptyl

„Rozptyl patří do skupiny charakteristik variability (rozptýlení). Vyjadřuje, jak dalece se navzájem liší jednotlivé hodnoty náhodné veličiny.“[2]

U diskrétní náhodné veličiny je rozptyl dán vztahem $D(X) = \sum_i \big(x_i - E(X)\big)^2 \cdot P(x_i)$. Častěji se však používá vztah $D(X) = E(X^2) - \big(E(X)\big)^2$, kde $E(X^2) = \sum_i x_i^2 \cdot P(x_i)$.[2] Vztah pro výpočet rozptylu spojité náhodné veličiny je upraven obdobně jako u střední hodnoty, tedy nahrazením bodu jeho okolím, pak $D(X) = \int_{-\infty}^{\infty} \big(x - E(X)\big)^2$ f(x) dx. I zde se častěji používá vztah $D(X) = E(X^2) - \big(E(X)\big)^2$, kde ale $E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot P(x)\, dx$.[3]

1.3.3 Směrodatná odchylka

„Rozptyl má jednotku odpovídající druhé mocnině jednotky, ve které měříme sledovanou náhodnou veličinu. To znesnadňuje jeho interpretaci. Proto zavádíme ještě **směrodatnou odchylku** $\sigma(X) = \sqrt{D(X)}$.“[2]

1.3.4 Pravděpodobnostní funkce, hustota pravděpodobnosti

„Pravděpodobnosti, se kterými náhodná veličina nabývá určitých hodnot, určuje v případě diskrétních veličin **pravděpodobnostní funkce** $P(x_i) = P(X = x_i)$, která popisuje pravděpodobnost, že náhodná veličina nabude určité hodnoty x_i. V případě spojitých náhodných veličin nemá smysl ptát se na pravděpodobnost v určitém bodě, neboť ta se limitně blíží k nule. Proto jsme nuceni definovat **hustotu pravděpodobnosti** $f(x)$, která sama o sobě nemá význam pravděpodobnosti, ale lze ji použít k vyjádření pravděpodobnosti, že náhodná veličina nabude hodnoty z určitého intervalu prostřednictvím vztahu $P\big(x \in (a,b)\big) = \int_a^b f(x)\, dx$.“[4]

1.3.5 Distribuční funkce

„Univerzálním nástrojem k popisu rozdělení pravděpodobnosti uvažované náhodné veličiny je **distribuční funkce**, která udává pravděpodobnost, že náhodná veličina bude mít hodnotu menší nebo rovnou hodnotě x. V případě diskrétních veličin platí $F(X) = P(X \leq x) = \sum_{x_i \leq x} P(x_i)$.“[4]

Jedná-li se o spojitou veličinu, platí zde vztah $F(X) = P(X \leq x) = \int_{-\infty}^{x} f(x)\, dx$.[4]

1.4 Vybraná rozdělení pravděpodobnosti

1.4.1 Rovnoměrné rozdělení

„Rovnoměrné rozdělení je charakterizováno dvěma parametry a, b, pro které platí $-\infty < a < b < \infty$.“[1] Právě v tomto intervalu nabývá náhodná veličina nenulové pravděpodobnosti.[4]

- **Střední hodnota**

$$E(X) = \frac{a+b}{2}$$

- **Rozptyl**

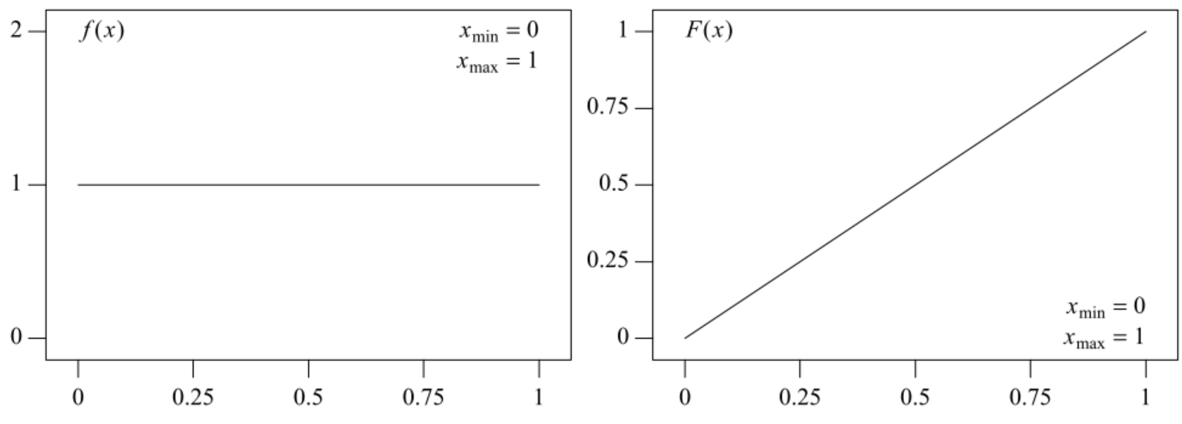
$$D(X) = \frac{(b-a)^2}{12}$$

- **Hustota pravděpodobnosti**

$$\begin{aligned} f(x) &= \frac{1}{b-a}, && x \in (a,b) \\ &= 0, && jinak \end{aligned}$$

- **Distribuční funkce**

$$\begin{aligned} F(x) &= 0, && x < a \\ &= \frac{x-a}{b-a}, && a \leq x \leq b \\ &= 1, && x \geq b \end{aligned}$$



Obr. 1: Hustota pravděpodobnosti a distribuční funkce rovnoměrného rozdělení [5]

„Používá se v případech, kdy předpokládáme, že se všechny možné hodnoty sledované náhodné veličiny vyskytují se stejnou pravděpodobností. Tímto rozdělením by se měla řídit například náhodná čísla.“[4]

1.4.2 Normální rozdělení

„Normální rozdělení pravděpodobnosti má dva parametry μ a σ, kde $-\infty < \mu < \infty$ a $0 < \sigma < \infty$. K jeho značení se používá symbol $N(\mu,\ \sigma)$.“[1] Je považováno za nejdůležitější rozdělení pravděpodobnosti spojité náhodné veličiny.

„Pro veličiny řídící se normálním rozdělením platí tzv. **pravidlo 3 σ**, které říká, že v intervalu $(\mu - 3\sigma,\ \mu + 3\sigma)$ leží skoro všechny hodnoty, kterých náhodná veličina nabývá.“[4]

- **Střední hodnota**

$$E(X) = \mu$$

- **Rozptyl**

$$D(X) = \sigma^2$$

- **Hustota pravděpodobnosti**

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

Je-li parametr $\mu = 0$ a parametr $\sigma = 1$, jedná se o **normální normované rozdělení pravděpodobnosti**. Po dosazení těchto hodnoty získáme následující zjednodušený vztah:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad x \in (-\infty, \infty)$$

- **Distribuční funkce**

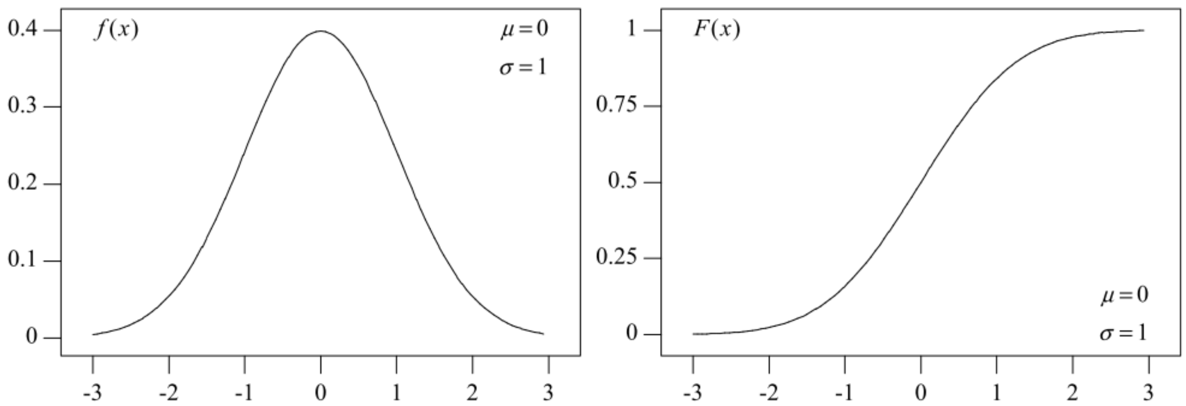
$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \qquad x \in (-\infty, \infty)$$

Vztah pro výpočet distribuční funkce normálního normovaného rozdělení pravděpodobnosti je následující:

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

Integrály na pravých stranách vzorců pro výpočet distribuční funkce normálního rozdělení neumíme vyjádřit pomocí elementárních funkcí. Proto jsou hodnoty distribuční funkce počítány numericky a jsou tabelované. Pro libovolné parametry μ a σ by však bylo nutné sestavit velké množství tabulek. Využívá se tedy vztahu mezi $N(\mu,\ \sigma)$ a $N(0,1)$.[1]

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



Obr. 2: Hustota pravděpodobnosti a distribuční funkce normálního rozdělení [5]

„Normálním rozdělením pravděpodobnosti se v praxi řídí náhodné veličiny, které nebývají své hodnoty v důsledku součtu velkého množství nezávislých vlivů, z nichž žádný nemá dominující význam. Jedná se především o náhodné chyby vznikající při různých měřeních a šetřeních.“[1]

1.4.3 Trojúhelníkové rozdělení

Trojúhelníkové rozdělení je charakterizováno třemi parametry a, b, c, pro které platí $-\infty < a \leq c \leq b < \infty$ a zároveň $a < b$, kde a, b jsou krajní meze intervalu, ve kterém náhodná veličina nabývá nenulové pravděpodobnosti a parametr c značí místo, ve kterém je pravděpodobnost nejvyšší. Graf hustoty pravděpodobnosti tedy svým tvarem připomíná ostroúhlý trojúhelník.

- **Střední hodnota**

$$E(X) = \frac{a + b + c}{3}$$

- **Rozptyl**

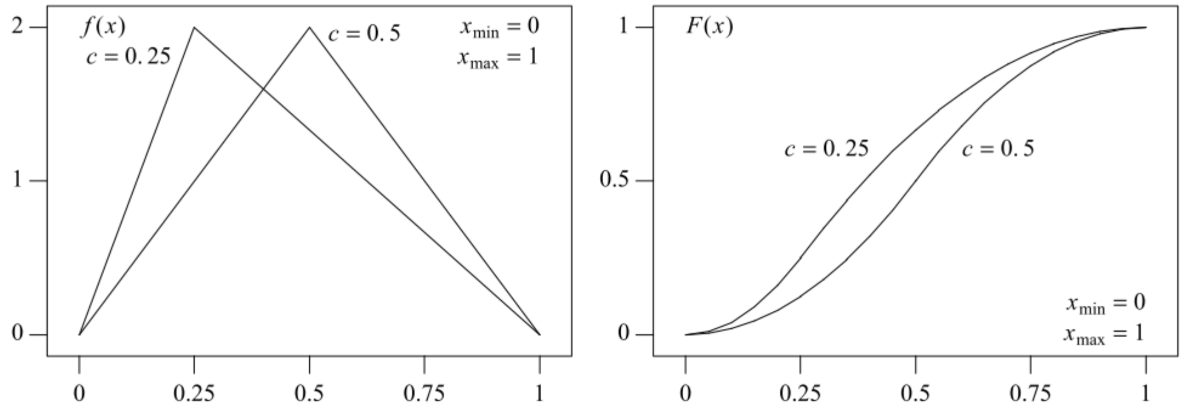
$$D(X) = \frac{3(b-a)^2 + (a + b - 2c)^2}{72}$$

- **Hustota pravděpodobnosti**

$$\begin{aligned} f(x) &= \frac{2}{b-a}\frac{x-a}{c-a}, && a \leq x < c \\ &= \frac{2}{b-a}\frac{b-x}{b-c}, && c \leq x \leq b \\ &= 0, && jinak \end{aligned}$$

- **Distribuční funkce**

$$
\begin{aligned}
F(x) &= 0, && x < a \\
&= \frac{(x-a)^2}{(b-a)(c-a)}, && a \leq x < c \\
&= 1 - \frac{(b-x)^2}{(b-a)(b-c)}, && c \leq x \leq b \\
&= 1, && x > b
\end{aligned}
$$



Obr. 3: Hustota pravděpodobnosti a distribuční funkce trojúhelníkového rozdělení [5]

Toto rozdělení se používá především v ekonomických aplikacích.

1.4.4 Exponenciální rozdělení

Jednoparametrické exponenciální rozdělení je dáno střední hodnotou μ, kde $\mu > 0$. Dvouparametrické exponenciální rozdělení pravděpodobnosti je navíc rozšířeno o parametr A, kde $-\infty < A < \infty$, jenž určuje posunutí.

- **Střední hodnota**

$$E(X) = \frac{1}{\mu} + A$$

- **Rozptyl**

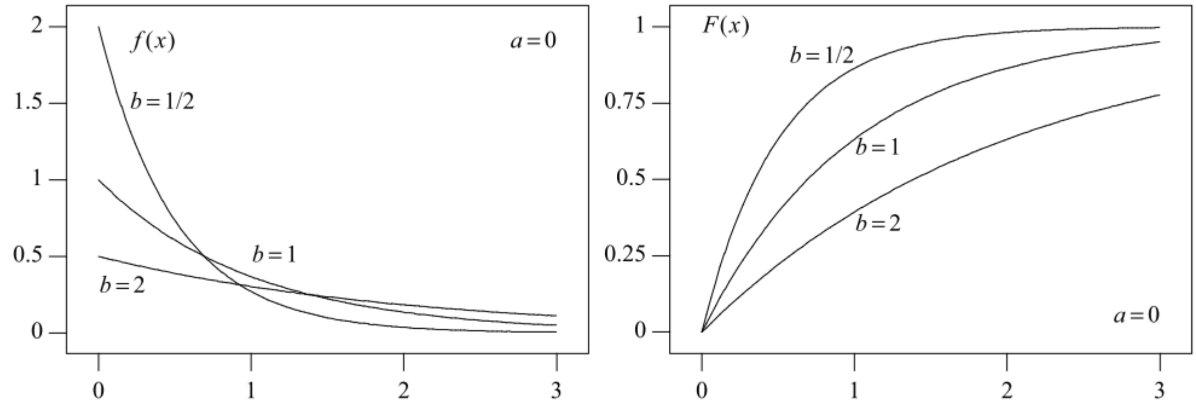
$$D(X) = \frac{1}{\mu^2}$$

- **Hustota pravděpodobnosti**

$$
\begin{aligned}
f(x) &= 0, && x < A \\
&= \mu e^{-\mu(x-A)}, && x \geq A
\end{aligned}
$$

- **Distribuční funkce**

$$F(x) = 0, \qquad x < A$$
$$= 1 - e^{-\mu(x-A)}, \qquad x \geq A$$



Obr. 4: Hustota pravděpodobnosti a distribuční funkce exponenciálního rozdělení [5]

„Exponenciálním rozdělením se často řídí životnost zařízení, která se neopravují, ale při selhání se vyměňují za nová, intervaly mezi poruchami nějakého zařízení nebo intervaly mezi příchody zákazníků do systému hromadné obsluhy. Je také používáno v neživotním pojištění při modelování nároků, když pravděpodobnost jejich extrémně velkých hodnot je zanedbatelná.“[1] „Pokud existuje doba A, po kterou sledovaná událost nemůže nastat, je vhodné použít dvouparametrické exponenciální rozdělení.“[4]

1.4.5 Logaritmicko-normální rozdělení

„Logaritmicko-normální rozdělení pravděpodobnosti má dva parametry μ a σ, kde $-\infty < \mu < \infty$ a $0 < \sigma < \infty$.“[1] Toto rozdělení mívají veličiny, které nabývají svých hodnot v důsledku velkého množství různých vlivů, mající interval možných hodnot omezen zdola.[4]

- **Střední hodnota**

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

- **Rozptyl**

$$D(X) = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

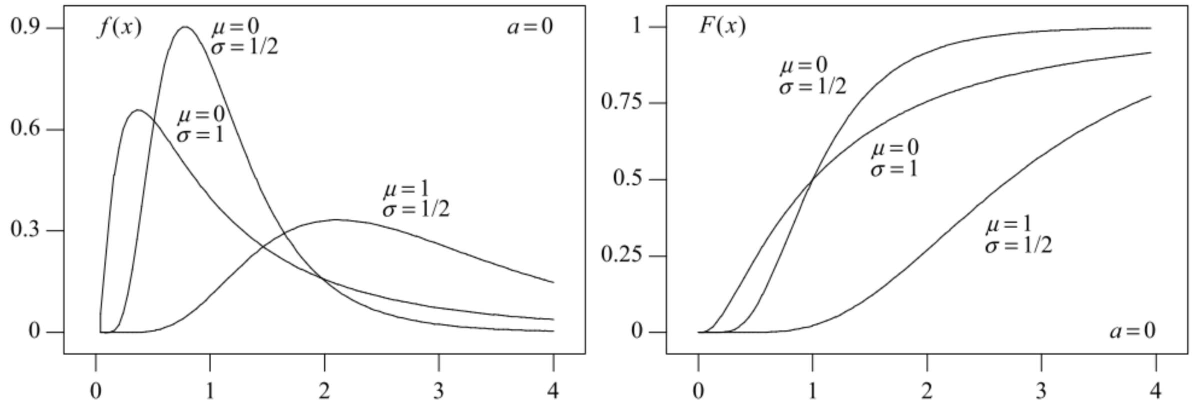
- **Hustota pravděpodobnosti**

$$f(x) = \frac{1}{\sqrt{2\pi} x\sigma} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \qquad x > A$$
$$= 0, \qquad x \leq A$$

- **Distribuční funkce**

$$F(x) = 0, \qquad x < A$$
$$= \int_0^x \frac{1}{\sqrt{2\pi} t\sigma} e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}} dt, \qquad x \geq A$$

„Chceme-li získat hodnoty distribuční funkce tohoto rozdělení, využívá se často skutečnosti, že má-li náhodná veličina X logaritmicko-normální rozdělení, pak náhodná veličiny $Y = ln X$ má normální rozdělení.“[4]



Obr. 5: Hustota pravděpodobnosti a distribuční funkce lognormálního rozdělení [5]

„Logaritmicko-normální rozdělení je vhodné k modelování výšky pojistných náhrad např. v havarijním pojištění, v požárním pojištění zděných budov a v pojištění proti vichřici. Dále je používáno v teorii spolehlivosti a při popisování velkých částic sypkých materiálů.“[1]

1.4.6 Erlangovo rozdělení

„Erlangovo rozdělení má dva parametry λ a k, kde $\lambda > 0$ a $k \epsilon \mathbb{N}$.“[1] Je-li $k = 1$, přechází toto rozdělení v exponenciální rozdělení pravděpodobnosti s parametrem λ.[1]

- **Střední hodnota**

$$E(X) = \frac{k}{\lambda}$$

- **Rozptyl**

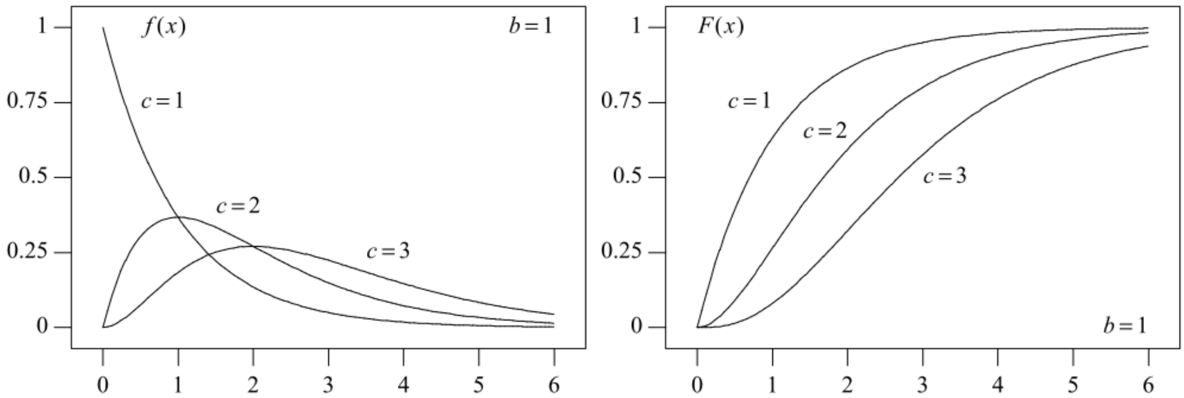
$$D(X) = \frac{k}{\lambda^2}$$

- **Hustota pravděpodobnosti**

$$f(x) = 0, \qquad x < 0$$
$$= \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{(k-1)!}, \qquad x \geq 0$$

- **Distribuční funkce**

$$F(x) = 0, \qquad\qquad x < A$$
$$= 1 - e^{-\lambda x}\sum_{i=0}^{k-1}\frac{(\lambda x)^i}{i!}, \qquad x \geq 0$$



Obr. 6: Hustota pravděpodobnosti a distribuční funkce Erlangova rozdělení [5]

„Erlangovo rozdělení pravděpodobnosti se uplatňuje hlavně v teorii hromadné obsluhy a v podobných situacích jako exponenciální rozdělení. Díky dvěma parametrům je však „ohebnější“, tudíž v některých případech vhodnější pro aproximaci reálných situací.“[1]

1.4.7 Beta rozdělení

„Standartní beta rozdělení (B-rozdělení) pravděpodobnosti je charakterizováno dvěma parametry a a b, pro které platí $a > 0$, $b > 0$.“[1]

- **Střední hodnota**

$$E(X) = \frac{a}{a+b}$$

- **Rozptyl**

$$D(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

- **Hustota pravděpodobnosti**

$$f(x) = \frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}, \qquad 0 < x < 1$$
$$= 0, \qquad\qquad jinak$$

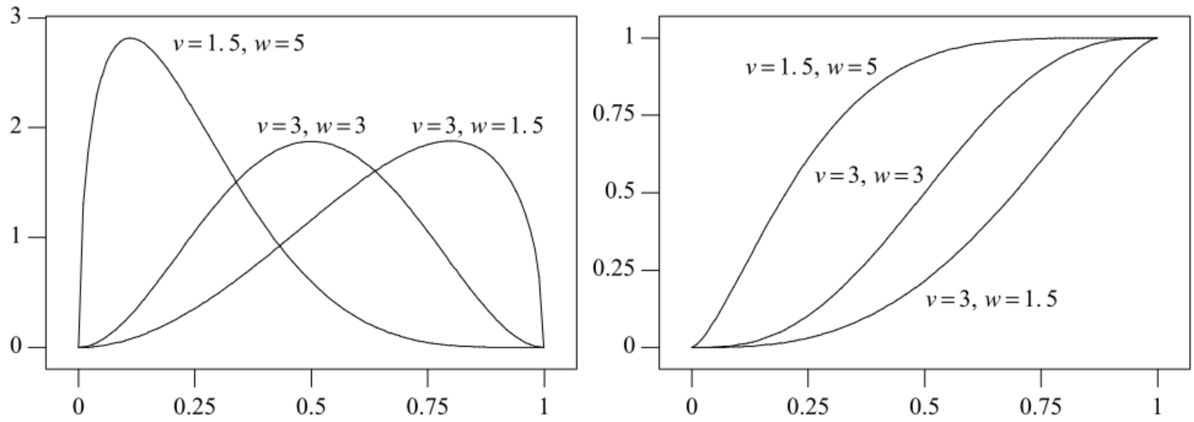
- **Distribuční funkce**

$$F(x) = \int_0^x \frac{1}{B(a,b)} t^{a-1}(1-t)^{b-1} dt, \qquad 0 < x < 1$$

$$= 0, \qquad jinak$$

B(a, b) je beta funkce. Jedná se o složitou funkci, o které si v případě zájmu musí čtenář sám dohledat další informace. Tato funkce je definována níže uvedeným vztahem.

$$\mathrm{B}(a,b) = \int_0^1 x^{a-1}\,(1-x)^{b-1}\,dx, \quad a > 0, b > 0$$



Obr. 7: Hustota pravděpodobnosti a distribuční funkce beta rozdělení [5]

„Tímto rozdělením se řídí mnoho náhodných veličin používaných v ekonomii, jejichž hodnoty jsou omezené shora i zdola a u nichž předpokládáme existenci jediného modu ležícího uvnitř intervalu možných hodnot. Také slouží jako teoretický model rozdělení pravděpodobnosti doby trvání činnosti v síťové analýze při aplikace metody PERT.“[1]

1.4.8 Gama rozdělení

Gama rozdělení (Γ-rozdělení) pravděpodobnosti je charakterizováno dvěma parametry a a b, pro které platí $a > 0$, $b > 0$.[1]

- **Střední hodnota**

$$E(X) = \frac{a}{b}$$

- **Rozptyl**

$$D(X) = \frac{a}{b^2}$$

- **Hustota pravděpodobnosti**

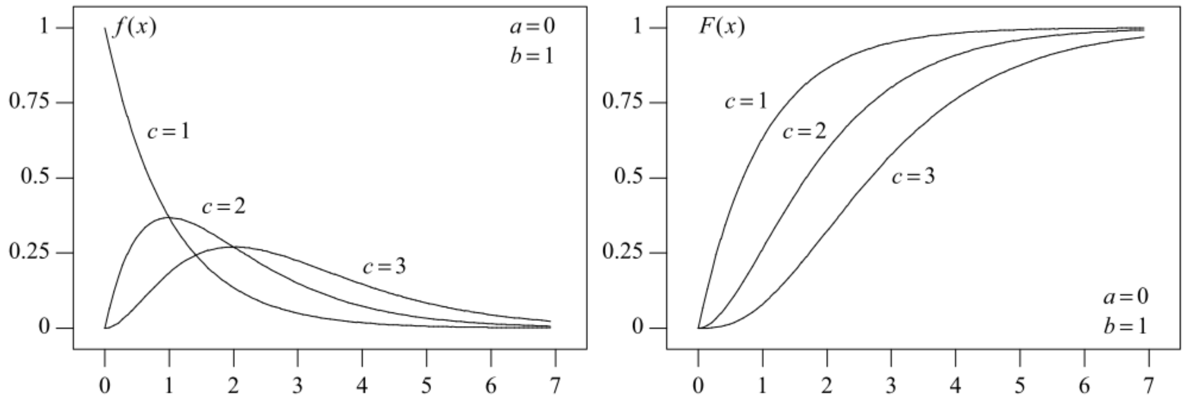
$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \qquad x > 0$$
$$= 0, \qquad x \leq 0$$

- **Distribuční funkce**

$$F(x) = \int_0^x \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt} dt, \qquad x > 0$$
$$= 0, \qquad x \leq 0$$

Γ(a) je gama funkce. Jedná se o složitou funkci, o které si v případě zájmu musí čtenář sám dohledat další informace. Tato funkce je definována níže uvedeným vztahem.

$$\Gamma(a) = \int_0^\infty x^{a-1}\, e^{-x}\, dx, \qquad a > 0$$



Obr. 8: Hustota pravděpodobnosti a distribuční funkce gama rozdělení [5]

„*Γ*-rozdělení pravděpodobnosti je vhodné pro simulaci jízdní doby dopravního prostředku. V oblasti pojišťovnictví se toto rozdělení používá při modelování výšky pojistných plnění motorových vozidel, kde variabilita výšky pojistných náhrad není příliš velká.“[1]

1.4.9 Chí-kvadrát rozdělení

„Chí-kvadrát (χ-kvadrát) rozdělení pravděpodobnosti se v literatuře často označuje symbolem χ^2_n. Má jeden parametr n, $n \in \mathbb{N}$. Nazýváme jej **počet stupňů volnosti**. Toto rozdělení pravděpodobnosti se nazývá i Pearsonovo rozdělení pravděpodobnosti s n stupni volnosti.“[1]

- **Střední hodnota**

$$E(X) = \mathrm{n}$$

- **Rozptyl**

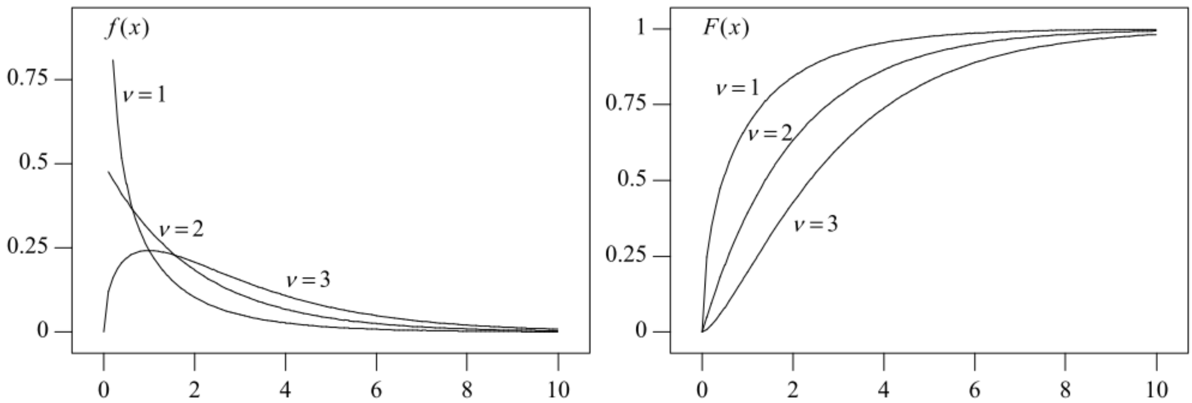
$$D(X) = 2n$$

- **Hustota pravděpodobnosti**

$$f(x) = 0, \qquad x \leq 0$$
$$= \frac{1}{2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)} x^{\frac{n-2}{2}} e^{-\frac{x}{2}}, \qquad x > 0$$

- **Distribuční funkce**

$$F(x) = 0, \qquad x \leq 0$$
$$= \int_0^x \frac{1}{2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)} x^{\frac{n-2}{2}} e^{-\frac{x}{2}} dx\,, \qquad x > 0$$



Obr. 9: Hustota pravděpodobnosti a distribuční funkce chí-kvadrát rozdělení [5]

Počet stupňů volnosti značí počet náhodných veličin v součtu druhých mocnin majících rozdělení $N(0, 1)$, které se často vyskytují v matematické statistice, z čehož plyne význam tohoto rozdělení.[1]

1.4.10 Weibullovo rozdělení

„Weibullovo rozdělení pravděpodobnosti má dva parametry $b > 0$ a $c > 0$.“[1]

- **Střední hodnota**

$$E(X) = \frac{\Gamma\left(\frac{c+1}{c}\right)}{b}$$

- **Rozptyl**

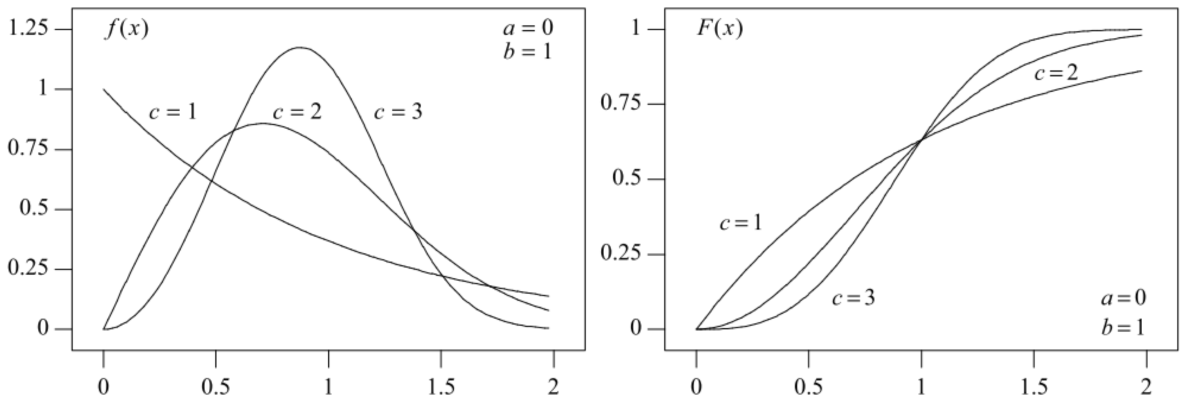
$$D(X) = \frac{\Gamma\left(\frac{c+2}{c}\right) - \Gamma^2\left(\frac{c+1}{c}\right)}{b^2}$$

- **Hustota pravděpodobnosti**

$$f(x) = b^c\, cx^{c-1}\, e^{-(bx)^c}, \qquad x > 0$$
$$= 0, \qquad x \leq 0$$

- **Distribuční funkce**

$$F(x) = 1 - e^{-(bx)^c}, \qquad x > 0$$
$$= 0, \qquad x \leq 0$$



Obr. 10: Hustota pravděpodobnosti a distribuční funkce Weibullova rozdělení [5]

„Weibullovo rozdělení nachází uplatnění především v technických aplikacích. Popisují se jím doby bezporuchové činnosti, životnosti, oprav a prostojů různých strojů a technických zařízení.“[1]

1.4.11 Rayleighovo rozdělení

„Rayleighovo rozdělení pravděpodobnosti má jeden parametr c, $c > 0$.“[1] Navíc je toto rozdělení speciálním případem Weibullova rozdělení s parametry $a = 2$ a $b = \frac{1}{2c^2}$.[1]

- **Střední hodnota**

$$E(X) = c\sqrt{\frac{\pi}{2}}$$

- **Rozptyl**

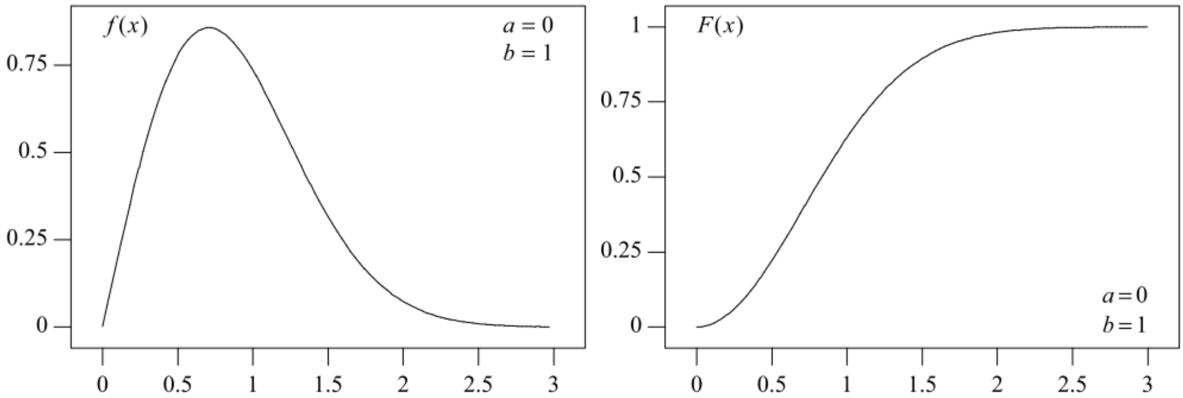
$$D(X) = \left(2 - \frac{\pi}{2}\right)c^2$$

- **Hustota pravděpodobnosti**

$$f(x) = \frac{x}{c^2}\, e^{-\frac{x^2}{2c^2}}, \qquad x \geq 0$$
$$= 0, \qquad x < 0$$

- **Distribuční funkce**

$$F(x) = 1 - e^{-\frac{x^2}{2c^2}}, \qquad x \geq 0$$
$$= 0, \qquad x < 0$$



Obr. 11: Hustota pravděpodobnosti a distribuční funkce Rayleighova rozdělení [5]

„Rayleighovo rozdělení pravděpodobnosti nachází uplatnění především v dynamice pružných soustav.“[1]

1.4.12 Cauchyovo rozdělení

„Cauchyovo rozdělení má dva parametry a a b, $-\infty < a < \infty$ a $0 < b < \infty$.“[1]

- **Střední hodnota**

$$Není\ definováno.$$

- **Rozptyl**

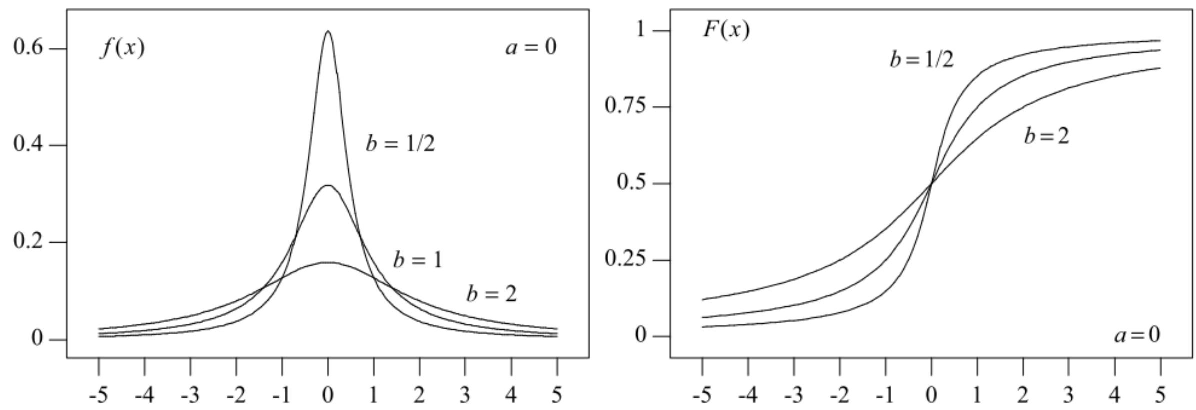
$$Není\ definováno.$$

- **Hustota pravděpodobnosti**

$$f(x) = \frac{1}{\pi} \frac{b}{b^2 + (x-a)^2}, \qquad x \in (-\infty, \infty)$$

- **Distribuční funkce**

$$F(x) = \frac{1}{2} + \frac{1}{\pi} arctg\left(\frac{x-a}{b}\right), \qquad x \in (-\infty, \infty)$$



Obr. 12: Hustota pravděpodobnosti a distribuční funkce Cauchyova rozdělení [5]

„Cauchyovo rozdělení má mnoho aplikací. Uplatňuje se při studiu chování kapalin. Je využíváno při zkoumání ohnisek zemětřesení. Velký význam má v ekonomii, kde mnoho ekonomických ukazatelů je podílem dvou nezávislých náhodných veličin majících $N(0, 1)$ rozdělení pravděpodobnosti.“[1] Pro velký počet extrémních hodnot je toto rozdělení doporučováno pro testování simulačních modelů.[6]

1.4.13 Fisher-Snedecorovo F-rozdělení

„Často je používán zkrácený název F-rozdělení pravděpodobnosti. Má dva parametry $n_1, n_2 \in N$ nazývané stupně volnosti.“[1]

- **Střední hodnota**

$$E(X) = \frac{n_2}{n_2 - 2}, \quad n_2 > 2$$

- **Rozptyl**

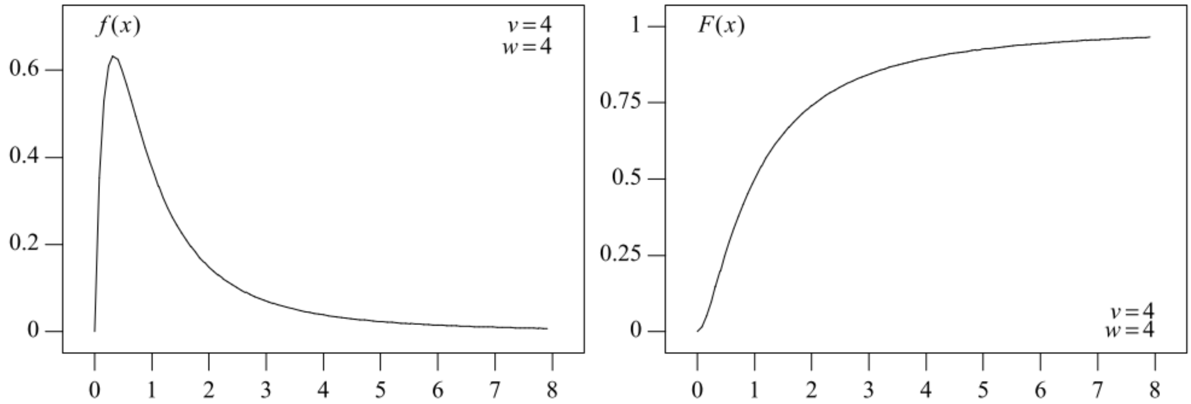
$$D(X) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}, \quad n_2 > 4$$

- **Hustota pravděpodobnosti**

$$\begin{aligned} f(x) &= \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1-2}{2}}\left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}, \qquad x > 0 \\ &= 0, \qquad x \leq 0 \end{aligned}$$

- **Distribuční funkce**

$$\begin{aligned} F(x) &= \int_0^x \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} t^{\frac{n_1-2}{2}}\left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} dt, \qquad x > 0 \\ &= 0, \qquad x \leq 0 \end{aligned}$$



Obr. 13: Hustota pravděpodobnosti a distribuční funkce F-rozdělení [5]

„F-rozdělení pravděpodobnosti má význam pro matematickou statistiku. Řídí se jím náhodná veličina $F = \frac{X_1}{X_2}\frac{\sqrt{n_2}}{\sqrt{n_1}}$, kde X_1 a X_2 jsou nezávislé náhodné veličiny mající $\chi^2_{n_1}$ a $\chi^2_{n_2}$ rozdělení pravděpodobnosti.“[1]

1.4.14 Studentovo t-rozdělení

„Studentovo rozdělení pravděpodobnosti (často krátce t-rozdělení) má jediný parametr n, $n \in \mathbb{N}$, který je nazýván počet stupňů volnosti.“[1]

- **Střední hodnota**

$$E(X) = 0$$

- **Rozptyl**

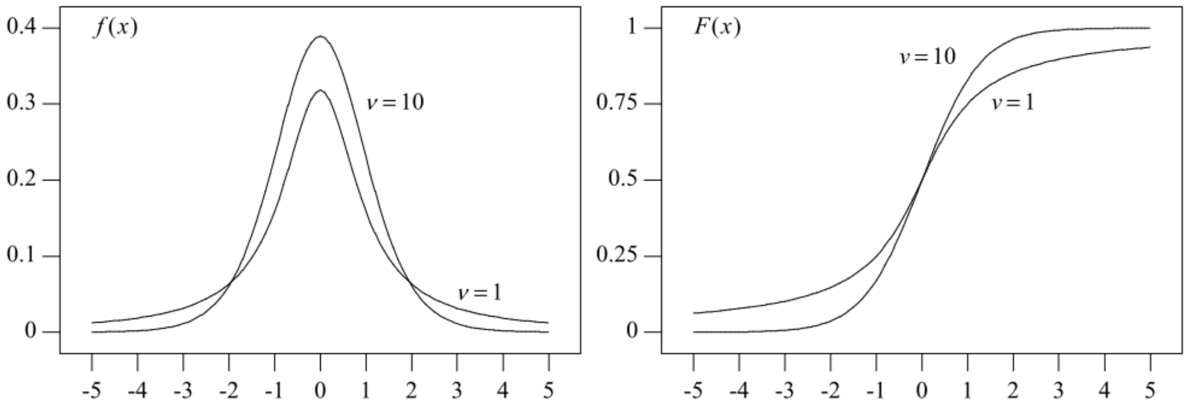
$$D(X) = \frac{n}{n-2}, \qquad n > 2$$

- **Hustota pravděpodobnosti**

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\sqrt{n}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \qquad x \in (-\infty, \infty)$$

- **Distribuční funkce**

$$F(x) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\sqrt{n}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt, \qquad x \in (-\infty, \infty)$$



Obr. 14: Hustota pravděpodobnosti a distribuční funkce t-rozdělení [5]

Studentovo rozdělení pravděpodobnosti má význam především v matematické statistice, ve které se často vyskytují náhodné veličiny typu $t = \frac{X}{\sqrt{Y}}\sqrt{n}$, kde X a Y jsou nezávislé náhodné veličiny. Náhodná veličina X má rozdělení $N(0,\ 1)$ a náhodná veličina Y se řídí χ_n^2 rozdělením.[1]

1.4.15 Logistické rozdělení

„Logistické rozdělení pravděpodobnosti je charakterizováno parametry a a b, $a \epsilon (-\infty, \infty)$, $b \epsilon (0, \infty)$.“[1]

- **Střední hodnota**

$$E(X) = -\frac{a}{b}$$

- **Rozptyl**

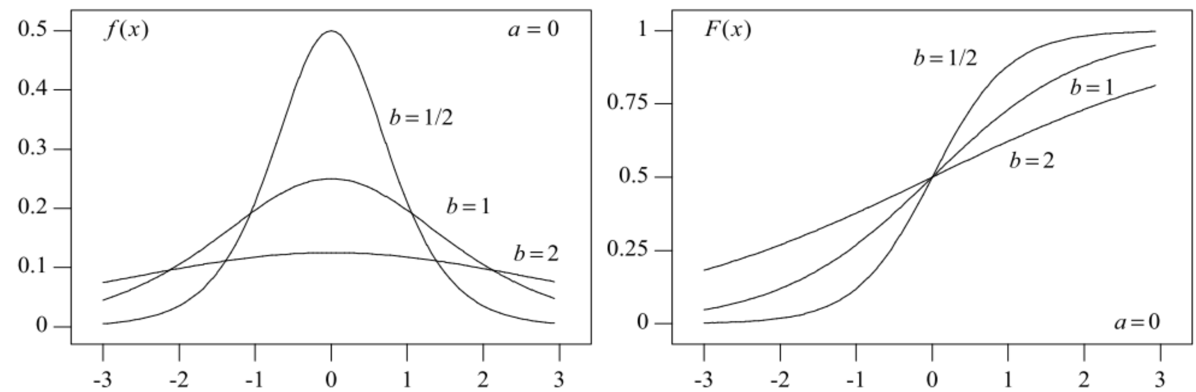
$$D(X) = \frac{\pi^2}{3b^2}$$

- **Hustota pravděpodobnosti**

$$f(x) = \frac{b\ e^{-(a+bx)}}{(1 + e^{-(a+bx)})^2}, \qquad x \in (-\infty, \infty)$$

- **Distribuční funkce**

$$F(x) = \frac{1}{1 + e^{-(a+bx)}}, \qquad x \in (-\infty, \infty)$$



Obr. 15: Hustota pravděpodobnosti a distribuční funkce logistického rozdělení [5]

„Logistické rozdělení je využíváno v ekonometrii (logitový model). Často se také vyskytuje při hodnocení dat při pokusech v biologii.“[1]

1.4.16 Laplaceovo rozdělení

„Toto rozdělení pravděpodobnosti má dva parametry a a b, pro které platí: $-\infty < a < \infty$, $b \in (0, \infty)$.“[1] Řídí se jím například Brownův pohyb.

- **Střední hodnota**

$$E(X) = \mathrm{a}$$

- **Rozptyl**

$$D(X) = 2b^2$$

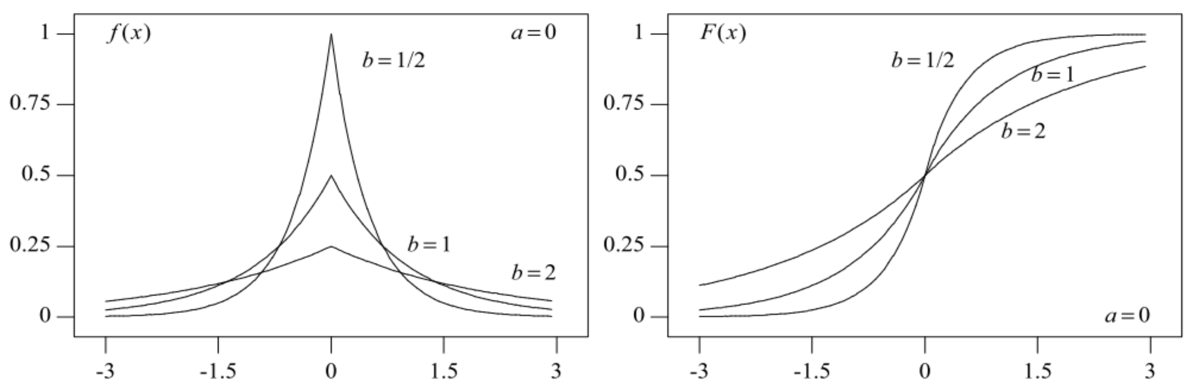
- **Hustota pravděpodobnosti**

$$f(x) = \frac{1}{2b} e^{-\frac{|x-a|}{b}}, \qquad x \in (-\infty, \infty)$$

- **Distribuční funkce**

$$F(x) = \frac{1}{2} e^{\frac{x-a}{b}}, \qquad x \leq \mathrm{a}$$

$$= 1 - \frac{1}{2} e^{-\frac{x-a}{b}}, \qquad x > \mathrm{a}$$



Obr. 16: Hustota pravděpodobnosti a distribuční funkce Laplaceova rozdělení [5]

1.5 Generování hodnot náhodných veličin

„Modely systémů, jejichž činnost je ovlivňována náhodnými faktory, obvykle využívají hodnoty náhodných veličin různých typů. Tyto hodnoty by v zásadě bylo možné v některých případech registrovat (na reálném systému) a pak použít jako součást vstupních dat simulačního programu. Pro simulační experimenty běžného rozsahu je však většinou třeba několik tisíc až set tisíc takových hodnot, takže je výhodné, jestliže na základě pozorovaných údajů stanovíme pravděpodobnostní zákonitosti (např. určíme typ rozdělení náhodné veličiny a odhadneme její parametry), které co možná nejlépe odpovídají disponibilním údajům, a na základě těchto pravděpodobnostních zákonitostí hodnoty generujeme v průběhu chodu simulačního programu samotným simulačním programem.“[6]

„Je zajímavou skutečností, že hodnoty různých typů náhodných veličin (v širším slova smyslu) lze získávat pomocí deterministických transformací hodnot jedné velmi jednoduché náhodné veličiny – rovnoměrného rozdělení na intervalu (0, 1).“[6]

1.5.1 Náhodná čísla

Termínem náhodná čísla budeme rozumět nezávislé hodnoty rovnoměrného rozdělení pravděpodobnosti na intervalu (0, 1). Ty můžeme získat hned několika způsoby – z tabulek náhodných čísel, za pomoci fyzikálních generátorů náhodných čísel či použitím aritmetických algoritmů.[6]

Pro výpočty malého rozsahu lze využít oficiálně vydávaných **tabulek náhodných čísel**. Jejich výhodou je zaručená náhodnost. Nevýhodou je počet náhodných čísel, který je i pro nerozsáhlé simulační modely příliš malý. Příkladem mohou být Tippetovy tabulky z roku 1927 (40 000 číslic) nebo tabulky RAND Corp. z roku 1955 (1 milion číslic).[6]

Další možností je použití **fyzikálních generátorů** náhodných čísel. Jejich podstata spočívá v registraci charakteristik určitých fyzikálních pochodů. To může být například měření délky intervalů mezi po sobě následujícími dopady částic na registrační plochu. Nevýhodou je ta skutečnost, že nelze posloupnost náhodných čísel reprodukovat, není umožněno identické opakování fyzikálního procesu.[6]

V simulačních modelech se dnes používá výhradně **aritmetických algoritmů**, jež vytvářejí náhodná čísla pomocí jednoduchých rekurentních výpočtů, v nichž následující číslo deterministicky závisí na jednom či více předchozích čísel. Z toho plyne, že tato čísla nejsou vůbec náhodná. Označují se tedy jako **pseudonáhodná čísla**. V dnešní době však existují algoritmy, které vykazují všechny podstatné vlastnosti náhodných čísel.[6]

1.5.2 Generování náhodných čísel

„Historicky první metodou získávání náhodných čísel pomocí aritmetických operací je von Neumannova metoda „prostředních řádů druhé mocniny“, navržená v roce 1946.“[6] Její princip zjednodušeně spočívá ve vhodném určení počátečního čísla, jeho umocnění a ponechání prostředních k míst. Tato metoda má celou řadu nedostatků – pro většinu výchozích hodnot je počet různých čísel předtím, než dojde k opakování, malý. Další nevýhodou je, že vygenerované posloupnosti čísel nevyhovují některým testům náhodnosti.[6]

Nejpoužívanější jsou dnes **lineárně kongruenční generátory** fungující na principu zavedeném Lehmerem v roce 1948.[6]

1.5.3 Testování náhodných čísel

„Před prováděním výpočtů, jež využívají posloupnosti náhodných čísel, je vhodné ověřit, zda použitý generátor poskytuje hodnoty, které lze považovat za nezávislé hodnoty rozdělení $R(0, 1)$. K tomuto účelu byla vytvořena celá řada testů (nejpoužívanějších je zhruba 20), které ověřují, zda poskytované posloupnosti čísel neodporují různým kritériím náhodnosti, jichž lze formulovat velmi mnoho. Přitom je zřejmé, že jestliže generovaná posloupnost čísel je ve shodě s testy $T_1, T_2, \ldots, T_n$, nemůžeme si být obecně vůbec jisti, jak se bude chovat vzhledem k dalšímu testu T_{n+1}. Každý test, kterým posloupnost úspěšně prošla, nám může pouze poněkud zvýšit míru důvěry v náhodnost čísel. V praxi se doporučuje použít asi 5 – 6 testů, a pokud jimi čísla poskytovaná generátorem projdou, lze považovat generátor za vyhovující.“[6]

K tomuto účelu slouží empirické a teoretické testy. **Empirické** vyhodnocují vygenerované posloupnosti čísel a porovnávají je s hodnotami statistik vypočtenými za předpokladu náhodnosti. Některé z těchto testů jsou zmíněny dále. **Teoretické** testy využívají výsledků teorie čísel a jejich závěry jsou platné pro celou periodu.[6]

K ověření rovnoměrnosti rozdělení generovaných čísel se používá **frekvenční test**. Pravděpodobnost, že náhodná veličina s rovnoměrným rozdělením pravděpodobnosti na intervalu (0, 1) nabude hodnoty z intervalu (α, β), kde $0 \leq \alpha \leq \beta \leq 1$, je rovna rozdílu $\beta - \alpha$. Rozdělíme jednotkový interval na k úseků a generujeme N náhodných čísel. Porovnáváme četnost výskytu čísel v jednotlivých úsecích s počty očekávanými. Zde můžeme použít například χ^2-test dobré shody.[6]

Dalším testem může být **test náhodnosti výskytu číslic**. Převedením posloupnosti čísel na řadu číslic a testováním výskytu jednotlivých číslic můžeme stanovit pravděpodobnost, že mezi dvěma stejnými číslicemi je k jiných číslic. I zde testujeme, zda se shodují předpokládané pravděpodobnosti s těmi zjištěnými. Zde je možné použít Kolmogorovův-Smirnovův test.[6]

Poker test je odvozen od známé karetní hry. Je v něm testována frekvence výskytu takových kombinací číslic, které jsou analogické s některými figurami v této hře. Pro pětimístná náhodná čísla (či pětice náhodných čísel) můžeme například testovat, zda se empiricky zjištěné četnosti výskytu určitých kombinací významně neodlišují od teoretických četností vypočtených za předpokladu náhodnosti. K tomuto testování můžeme použít χ^2-test.[6]

Test výskytu úplných sad číslic zkoumá délku posloupností, v které se vyskytuje každá z číslic dané soustavy alespoň jednou. U desítkové soustavy se tady zjišťuje, kolik prvků posloupnosti je třeba projít, aby se každá z číslic 0, 1, …, 9 vyskytla alespoň jednou. Ověřuje se, zda se zjištěné počty posloupností různých délek významně neliší od očekávaných četností.[6]

K odhalení případných vazeb mezi prvky posloupnosti náhodných čísel je určen **test autokorelace**. Zjišťuje se tedy, jaký je vztah mezi prvky r_i a r_{i+k}.[6]

V případě hlubšího zájmu je třeba upozornit na **spektrální test**, který spojuje některé vlastnosti teoretických a empirických testů. Je založen na poměrně náročné matematické teorii, ale zato je považován za dosud nejúčinnější test.[6]

1.5.4 Transformace náhodných čísel

Jak bylo vzpomenuto výše, existují i jiná rozdělení pravděpodobnosti než rovnoměrné. Abychom mohli generovat i hodnoty jiných rozdělení, je třeba právě rovnoměrné rozdělení transformovat. To lze uskutečnit několika metodami, jejichž princip je stručně naznačen dále.

Metoda inverzní transformace spočívá v nalezení inverzní funkce F^{-1} funkce distribuční. Existuje-li F^{-1} v explicitním tvaru, můžeme použít vztah $x = F^{-1}(r)$, kde $r \epsilon (0,\ 1)$ je náhodné číslo. Po dosazení r získáme x, což je hodnota náhodné veličiny X.[6]

Zamítací metoda předpokládá, že je hustota pravděpodobnosti ohraničena v intervalu $\langle a,\ b \rangle$, mimo něj je rovna nule. Princip této metody spočívá v generování bodů $[x,\ y]$ rovnoměrně rozdělených v obdélníku, v němž je určen graf funkce $f(x)$. V případě, že $y \leq f(x)$, považujeme x za vygenerovanou hodnotu, jinak generujeme další bod $[x, y]$.[6]

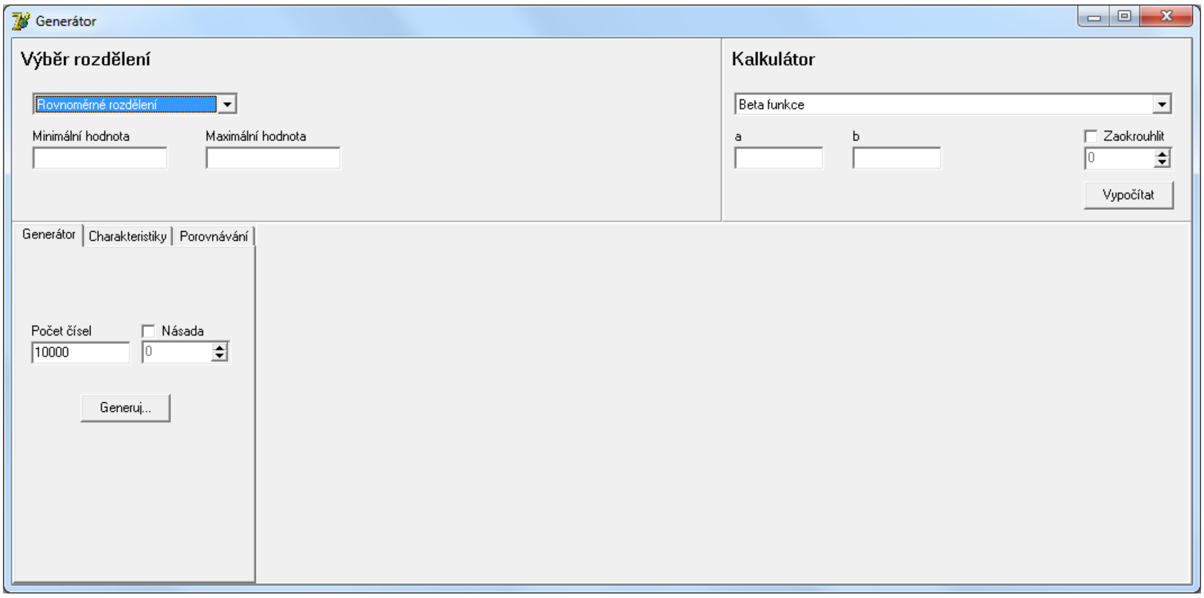
Podstatou **kompoziční metody** je vyjádření hustoty pravděpodobnosti náhodné veličiny X ve tvaru $f(x) = \sum_i p_i f_i(x)$, kde $p_i > 0$ a $\sum_i p_i = 1$.[6]

2 OVLÁDÁNÍ APLIKACE

V této kapitole uživatel nalezne podrobný návod k ovládání přiložené aplikace. Nejdříve je popsáno okno aplikace jako celek a následně jsou uvedeny instrukce pro ovládání jeho jednotlivých částí.

Jak je vidět na níže uvedeném náhledu, je okno aplikace složeno z několika přehledných částí, jejichž snahou je zajistit intuitivní a uživatelsky přívětivé prostředí. První část slouží k výběru rozdělení pravděpodobnosti a zadání jeho parametrů, druhá je určena k výpočtu funkcí a třetí zapouzdřuje ovládací panely generování čísel, charakteristik a porovnání rozdělení se segmenty pro vykreslení grafů hustoty pravděpodobnosti a distribuční funkce. Tyto části jsou blíže popsány dále.

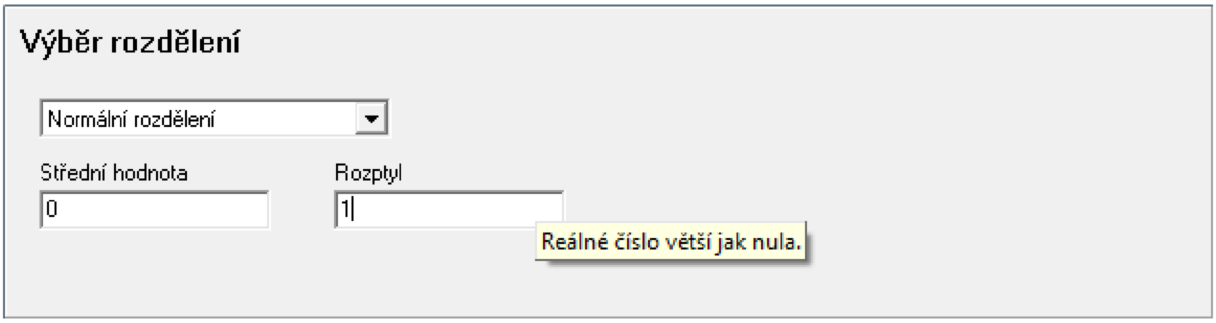
K ukončení programu slouží tlačítko *Křížek* v pravém horním rohu okna nebo stisknutí kombinace kláves *Alt + F4*.



Obr. 17: Okno aplikace

2.1 Seznam rozdělení pravděpodobností

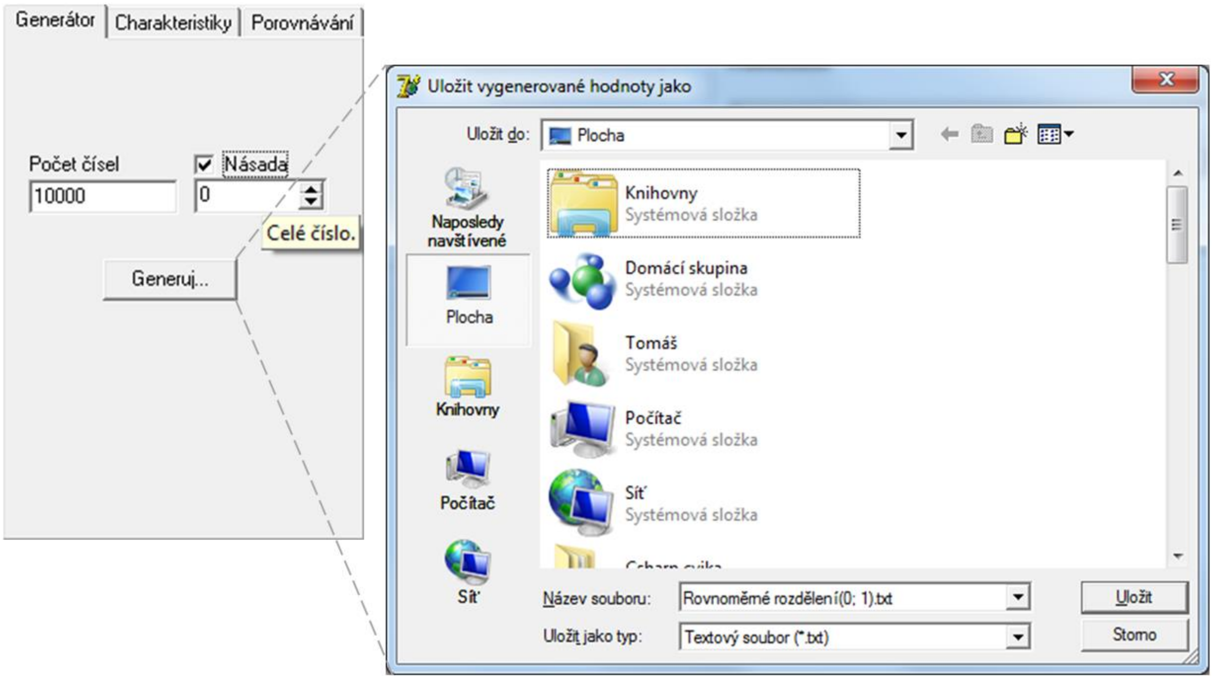
Aby bylo možné generovat čísla, počítat charakteristiky a porovnávat jednotlivá rozdělení pravděpodobnosti, je nutné nejdříve vybrat jedno z dostupných rozdělení a vyplnit jeho parametry. Při výběru požadovaného rozdělení se automaticky aktualizují pole pro zadání vstupních údajů. Každé toto pole obsahuje nápovědu, která se zobrazí po najetí myší, což usnadní uživateli práci. Obsahem této nápovědy je informace o intervalu, v němž se může daná hodnota pohybovat. Detail na tuto oblast je zobrazen na níže uvedeném náhledu.



Obr. 18: Výběr rozdělení pravděpodobnosti

2.2 Generování pseudonáhodných čísel

Po nastavení rozdělení pravděpodobnosti se přesuneme do části s parametry generování. Zadáme požadovaný počet generovaných čísel a zvolíme, zda chceme použít násadu, případně nastavíme její hodnotu. Při použití násady je vygenerována vždy stejná posloupnost hodnot závislá na zvolené hodnotě násady. I zde jsou editační pole doplněna o nápovědu.



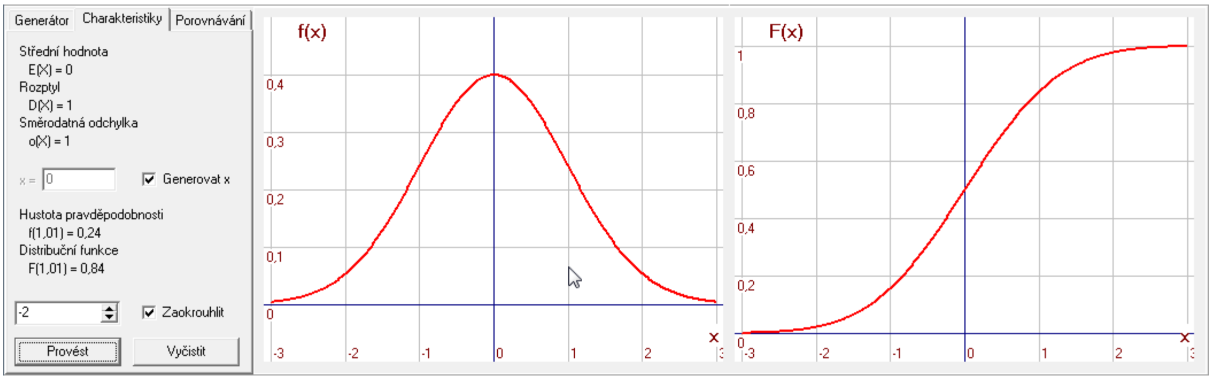
Obr. 19: Práce s generátorem

Odpovídají-li vyplněné hodnoty našim požadavkům, stiskneme tlačítko *Generuj*. To vyvolá dialogové okno, pomocí kterého nastavíme umístění a název souboru, do kterého se sekvence pseudonáhodných čísel uloží. Podle vybraného rozdělení pravděpodobnosti a jeho parametrů se automaticky vyplní název souboru. Ten je případně možné změnit. Není-li na konci názvu uvedena koncovka, je doplněna automaticky. Jedná se o textový soubor s koncovkou **.txt*. Data jsou v něm uspořádána v řádcích tak, že každé číslo je na řádku samo. Při stisknutí tlačítka *Storno* se celé generování zruší. Tlačítkem *Uložit* se celý proces dokončí a vytvoří se soubor s hodnotami. Při špatně zadaných parametrech rozdělení pravděpodobnosti se vypíše chybové hlášení a generování se zruší obdobně jako při stisknutí tlačítka *Storno*. Je-li vše správně, objeví se informační okno s informací, že generování proběhlo úspěšně.

2.3 Charakteristiky rozdělení pravděpodobnosti

Po nastavení rozdělení pravděpodobnosti se přesuneme do části charakteristik. Po stisku tlačítka *Provést*, jsou-li správně zadány vstupní parametry požadovaného rozdělení, se vypíší jednotlivé charakteristiky a zobrazí se graf hustoty pravděpodobnosti a distribuční funkce, viz následující obrázek. Pokud vstupní údaje nesouhlasí s těmi povolenými, je vypsáno chybové hlášení a ve vypisování výsledků se dále nepokračuje.

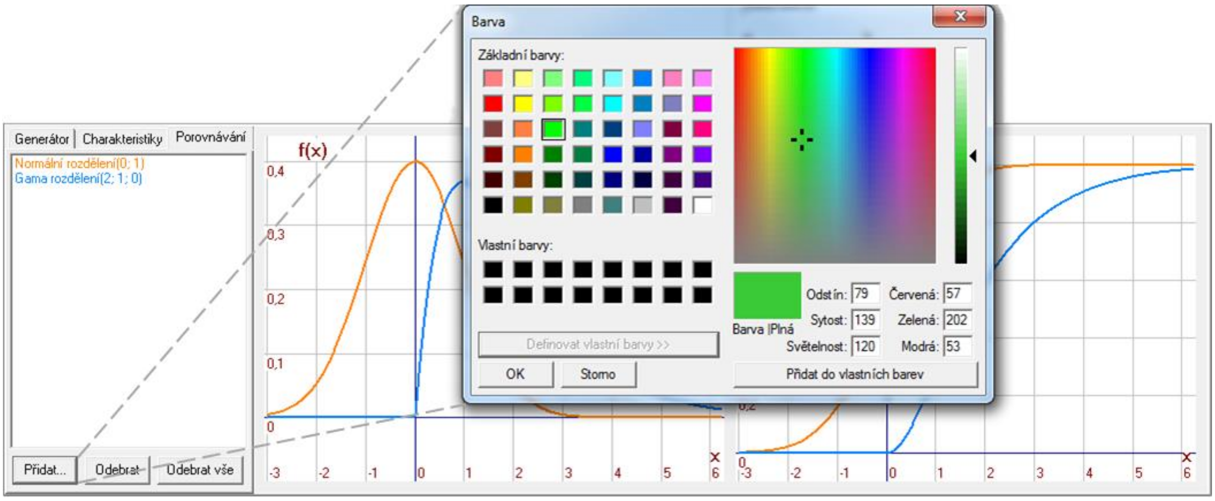
Navíc je možné nastavit, zda se budou hodnoty vykreslených funkcí vypisovat dynamicky podle polohy kurzoru na ploše grafu, nebo budou počítány z uživatelem zadané hodnoty. To můžeme nastavit pomocí ovládacího prvku *Generovat x*. Všechny zobrazené výsledky je možné před vypsáním na obrazovku zaokrouhlit. K tomu slouží ovládací prvek *Zaokrouhlit*. Zaokrouhlení se nastavuje pomocí celočíselné hodnoty, která určuje počet desetinných míst za desetinnou čárkou, je-li hodnota záporná. V případě opačném číslo určuje zaokrouhlení vlevo od desetinné čárky.



Obr. 20: Výpis charakteristik

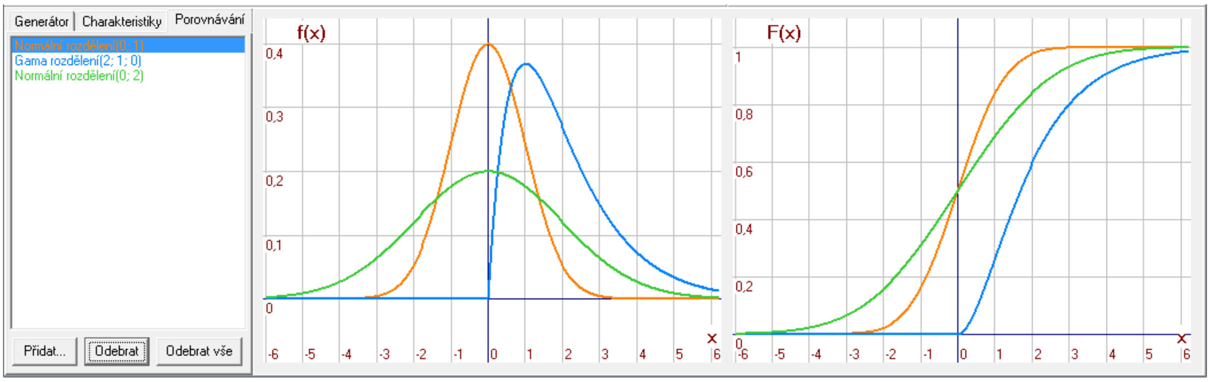
2.4 Porovnání rozdělení pravděpodobnosti

Po nastavení rozdělení pravděpodobnosti se přesuneme do části porovnávání. Po stisku tlačítka *Přidat* se otevře dialogové okno pro výběr barvy, kterou bude požadované rozdělení reprezentováno. Po výběru barvy, jsou-li správně zadány vstupní parametry, je rozdělení přidáno do seznamu a vykresleno.



Obr. 21: Přidání rozdělení

Chceme-li nějaké rozdělení odebrat, stačí na něj kliknout levým tlačítkem myši, tím se označí a pomocí tlačítka *Odebrat* ho odstraníme jak ze seznamu, tak i z oblasti grafů.



Obr. 22: Odebrání rozdělení

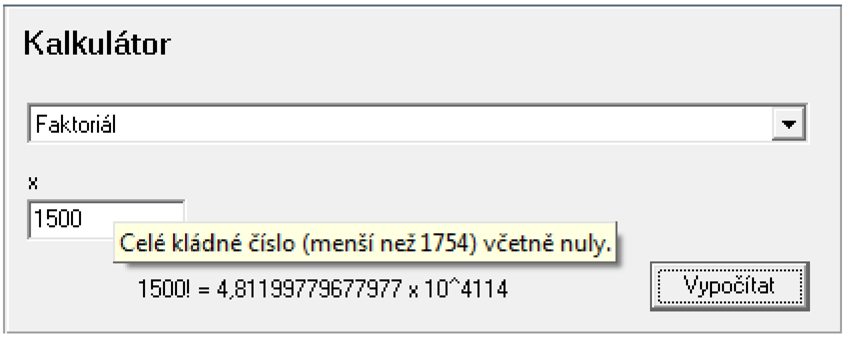
Chceme-li odebrat všechna rozdělení, stačí stisknout klávesu *Odebrat vše*. Tím se odstraní všechna rozdělení.

V případě, že máme připraveno několik rozdělení, a potřebujeme pracovat například s generátorem čísel, nemusíme se bát, že by se seznam vymazal. Můžeme se tedy vrátit k porovnávání později. Seznam rozdělení je smazán pouze stisknutím tlačítka *Odebrat vše* nebo vypnutím aplikace.

2.5 Kalkulátor funkcí

Kalkulátor jako jediný nepožaduje vyplnění informací o rozdělení pravděpodobnosti. Pracuje tedy nezávisle na zbytku programu. Po výběru jedné z šesti funkcí se automaticky aktualizují pole pro zadání vstupních hodnot. I zde jsou editační pole doplněna o nápovědu. Poté stiskneme tlačítko *Vypočítat* a jsou-li hodnoty vyplněny korektně, v textovém poli se objeví výsledek. Pokud ne, objeví se chybové hlášení.

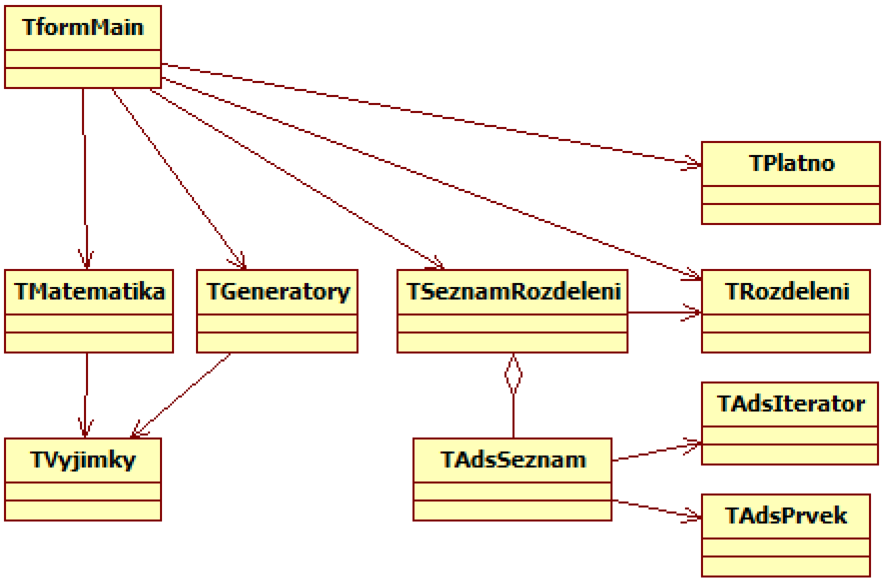
U všech funkcí kromě faktoriálu je možné zaokrouhlení výsledné hodnoty. To funguje obdobně jako dříve zmíněných charakteristik rozdělení pravděpodobnosti. Pokud je ovšem výsledek příliš velké nebo malé číslo, může nastat při zaokrouhlování chyba způsobená maximální velikostí datového typu. V tomto případě je vypsáno chybové hlášení, které na tuto skutečnost upozorní. Poté je třeba pro zadanou funkci s těmito parametry zaokrouhlování vypnout. Z tohoto důvodu není u faktoriálu dostupné zaokrouhlování vůbec.



Obr. 23: Kalkulátor funkcí

3 IMPLEMENTACE APLIKACE

Obsahem této kapitoly je přiblížení vnitřní struktury aplikace a použitých datových struktur. Nejdříve je popsána aplikace jako celek a poté jsou blíže rozebrány její stěžejní části. Vnitřní uspořádání je zobrazeno pomocí zjednodušeného UML diagramu tříd, který jsem vytvořil v aplikaci StarUML.



Obr. 24: Zjednodušený UML diagram tříd

Třída `TformMain` zapouzdřuje okno aplikace s ovládacími prvky a propojení s dalšími třídami, které jsou nezbytné pro smysluplnou práci programu. Zajišťuje tedy komunikaci se třídami pro generování čísel – `TGeneratory`, výpočty charakteristik, matematických i statistických funkcí – `TMatematika` a vykreslování grafů na kreslící plátna – `TPlatno`. Pro práci s jednotlivými rozděleními pravděpodobnosti je vytvořena třída `TRozdeleni` a pro zapouzdření více rozdělení, která jsou pak porovnávána, je určena třída `TSeznamRozdeleni`.

Třída `TSeznamRozdeleni` obsahuje tedy především seznam jednotlivých rozdělení a metody pro jejich vkládání, odebírání a prohlížení. Je založen na abstraktní datové struktuře (dále ADS) obousměrný acyklický seznam s hlavou – `TAdsSeznam`. Aby bylo možné s touto datovou strukturou pracovat, bylo třeba vytvořit ADS prvek – `TAdsPrvek` a ADS umožňující procházení seznamu – `TAdsIterator`.

Podstatou ADS je především jejich univerzálnost použití. Není potřeba dopředu vědět, s jakými daty se bude pracovat. `TAdsPrvek` tedy zapouzdřuje kromě atributů s odkazy na předchozí a následující prvek, také atribut s odkazem na data, která jsou typu `TObject`, což zajišťuje znovupoužitelnost této struktury. Jak bylo zmíněno výše, k procházení dat slouží třída `TAdsIterator`. Iterátor při práci uchovává odkaz na prvek, kterým začíná procházení seznamu, a na prvek aktuální, tedy na ten, pomocí kterého jsou data zpřístupňována. Základní metodou této struktury je zpřístupnění prvku tak, aby uživatel nemohl měnit strukturu samotného seznamu, ale pouze data prvku. Následně se prohlídka posune na další prvek.

Třída `TRozdeleni` zapouzdřuje informace o rozdělení pravděpodobnosti, a to především s důrazem na data důležitá pro vykreslení grafů hustoty pravděpodobnosti a distribuční funkce. Pro tento účel je třeba atributů pro uchování hranic intervalu, v kterém jsou funkce zajímavé, název rozdělení jako identifikátor, vstupní parametry rozdělení a další.

Další důležitou částí pro zajištění korektních výstupů aplikace je třída `TVyjimky`. Ta je používána třídami `TMatematika` a `TGeneratory` k tomu, aby nedocházelo k jakýmkoliv výpočtům s parametry, které neodpovídají povoleným vstupním hodnotám. Dostanou-li se tedy chybné vstupní údaje až k místům, kde dochází k samotným výpočtům, je vyvolána výjimka typu `EInvalidArgument` jednotky `Math` s odpovídajícím chybovým hlášením. Ty jsou v případě této aplikace ošetřovány hlavním formulářem `TformMain`. Nemělo by se tedy stát, že by se k uživateli dostaly nesmyslné informace.

3.1 Generování pseudonáhodných čísel

Stěžejní funkcí celé aplikace je generování pseudonáhodných čísel. O to se stará třída `TGenerátory`, která zapouzdřuje jak funkci pro generování hodnot rovnoměrného rozdělení $R(0, 1)$, tak i metody transformující tyto hodnoty na jiná spojitá rozdělení.

Pro vývojáře používajícího tuto třídu je připraveno 16 metod pro generování hodnot spojité náhodné veličiny. Ty jsou přetíženy, takže je možné využít i tzv. generování s násadou, kdy je výsledná posloupnost hodnot za použití stejné násady vždy totožná.

Metody jsou pojmenovány názvy rozdělení. Jejich parametry jsou pojmenovány tak, že uživatel pozná, co který parametr reprezentuje. První skupina parametrů představuje vstupní hodnoty daného rozdělení pravděpodobnosti. Ty jsou všechny typu `Real`, jsou kontrolovány pomocí metod třídy `TVyjimky` a jejich počet je dán rozdělením pravděpodobnosti. Poslední parametr je jednorozměrné dynamické pole typu `Real` a je volané odkazem. Do tohoto pole se zapisují výstupní hodnoty generátoru. Počet těchto hodnot je určen velikostí pole. U přetížených metod, jež jsou doplněny o násadu, je posledním parametrem právě hodnota této násady, která je typu `Integer`.

Tyto veřejné metody poté ověří správnost vstupních hodnot a pomocí soukromých metod naplní pole požadovanými hodnotami.

Soukromá část třídy funguje tak, že je vygenerováno pole náhodných čísel $R(0, 1)$, které je poté přetransformováno na požadované rozdělení pravděpodobnosti se zadanými parametry. Algoritmy pro generování hodnot jsem čerpal z literatury [5] a [6].

3.2 Matematické funkce

Aby mohla být aplikace použita i jako studijní opora, bylo nutné vytvořit třídu, která dokáže pomocí metod vracet výsledky matematických a statistických funkcí. V tomto případě šlo především o střední hodnotu, rozptyl, hustotu pravděpodobnosti a distribuční funkci u každého rozdělení. K tomuto účelu bylo navíc potřeba implementovat další složité matematické funkce. I ty jsou prostřednictvím této třídy k dispozici.

Funkce pro výpočty charakteristik rozdělení pravděpodobností jsou pojmenovány názvem rozdělení doplněným o název dané charakteristiky. Vstupní parametry jsou implementovány obdobně jako u třídy `TGeneratory`. U metod pro výpočet distribuční funkce a hustoty pravděpodobnosti je jako první parametr hodnota x a až poté vstupní hodnoty rozdělení. Návratová hodnota je typu `Real`. Vzorce pro tyto výpočty jsem čerpal z literatury [1],[5] a [6].

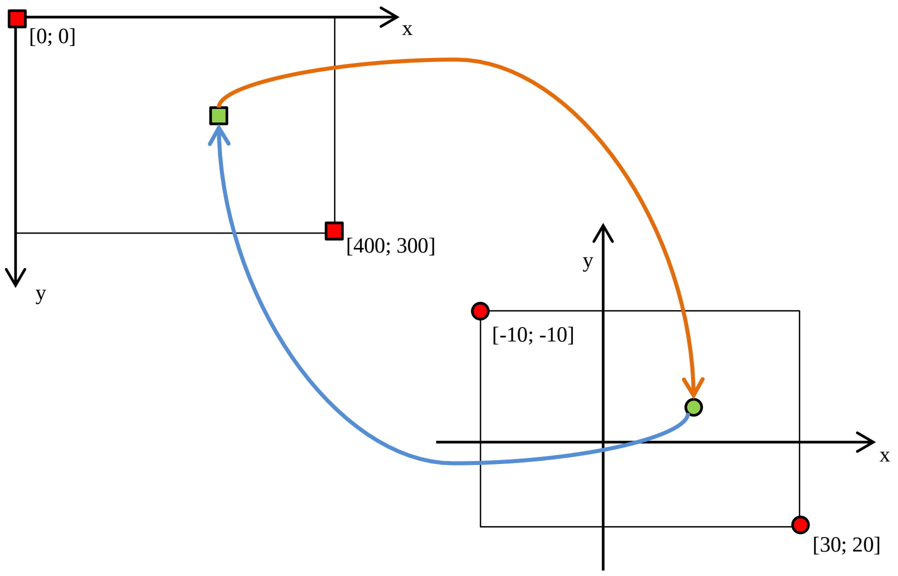
Pomocné funkce jsou pojmenovány názvem dané funkce. Jedná se o beta funkci, beta neúplnou funkci, gama funkci, gama neúplnou funkci, faktoriál a kumulativní distribuční funkci normovaného normální rozdělení – v Excelu známou pod názvem NORMSDIST.

Faktoriál jsem implementoval na základě vlastních znalostí, funkci NORMSDIST jsem nalezl na internetových stránkách [7] a ostatní čtyři funkce v literatuře [8].

3.3 Vykreslování grafů

Aby bylo možné vykreslit hodnoty hustoty pravděpodobnosti a distribuční funkce do grafu, vytvořil jsem třídu `TPlatno`. Při jejím vytvoření se do konstruktoru vloží jako parametr kreslící plátno třídy `TPaintBox`. Poté je možné na kreslící plochu pomocí implementovaných procedur kreslit.

V této třídě jsou pro aplikaci stěžejní čtyři metody. První nastaví plátno. To spočívá v určení hodnot pro přepočet z reálné souřadné soustavy na tu kreslenou. Kreslící plátno má implicitně bod [0; 0] v levém horním rohu a bod [X_{MAX}; Y_{MAX}] je dán velikostí plátna. Díky těmto přepočtům můžeme na plátno vykreslit graf dle vlastních požadavků, nezajímá nás velikost kreslící plochy a dostaneme se i na záporné hodnoty. Toto nastavení se provede na základě vstupních parametrů metody, které nastavují krajní hranice plátna. Viz obrázek.



Obr. 25: Přechod mezi reálným a kresleným souřadným systémem [9]

Druhá metoda je určena k překreslení plátna. Parametrem této metody je barva typu `TColor`, kterou chceme plátno překreslit. Po zavolání této metody se celé plátno přebarví na požadovanou barvu, dojde tedy k vymazání plátna.

Třetí metoda se stará a vykreslení hlavních a vedlejších os. Hlavní osy se vykreslí jako přímky procházející reálnou souřadnici [0; 0]. Vedlejší osy se vykreslí rovnoběžně s těmi hlavními, a to ve vzdálenostech zadaných jako parametry procedury. Chce-li uživatel k hlavním osám přidat i popisky, použije k tomu další dva parametry. Nechce-li je, vloží prázdný řetězec. Korektní zobrazení je zaručeno pro několikaznakové popisky.

Čtvrtou důležitou metodou je vykreslení grafu. Vstupní parametry jsou pole hodnot x a pole hodnot y. Tato pole musí mít stejnou velikost a jsou obě typu `Real`. Pomocí příkazu `MoveTo` se přemístí „kreslící hrot" na pozici se souřadnicemi prvního prvku polí. Poté se prochází celé pole a příkazem `LineTo` se postupně kreslí přímky přes všechny body načtené z polí. Další dva parametry umožňují změnit barvu a tloušťku čáry. Parametr barvy je typu `TColor` a tloušťka čáry je typu `Byte`.

4 OVĚŘENÍ NÁHODNÝCH ČÍSEL

Jelikož je pro generování náhodných čísel použita implicitní funkce `Random` jednotky `System` programovacího jazyka Object Pascal, lze předpokládat, že tyto vygenerované hodnoty rovnoměrného rozdělení $R(0, 1)$ odpovídají kritériím náhodnosti a není je tedy potřeba zkoumat pomocí výše uvedených testů. Ověřování náhodných čísel je tedy zaměřeno na porovnání vygenerovaných hodnot jak s hodnotami skutečnými, tak i s těmi vygenerovanými jinými sofistikovanými aplikacemi. Také je prováděno testování hypotéz, zda hodnoty pochází z daného rozdělení pravděpodobnosti. Ověřování je provedeno pro některá rozdělení důležitá v dopravě. Vzorce použité pro výpočet distribučních funkcí jsou uvedeny v kapitole Teoretický aparát.

4.1 Exponenciální rozdělení

Vygenerované hodnoty exponenciálního rozdělení jsou porovnány s naměřenými hodnotami zpoždění osobního vlaku. Nejdříve je odhadnuta střední hodnota souboru dob zpoždění a je ověřeno, zda se naměřené hodnoty řídí exponenciálním rozdělením. Poté jsou na základě odhadnuté střední hodnoty vygenerována čísla, která jsou také podrobena testu, zda se řídí exponenciálním rozdělením. Pokud oba soubory čísel projdou testem, bude uskutečněno jejich vzájemné porovnání.

Soubor hodnot je tvořen 250 naměřenými dobami zpoždění osobního vlaku. Nejdříve jsem tedy odhadnul střední hodnotu naměřených hodnot. Jelikož by se tento soubor měl řídit exponenciálním rozdělením, použil jsem vzorec $\mu = \frac{1}{\bar{x}} \doteq 0{,}6$, kde $\bar{x}$ je aritmetický průměr naměřených hodnot.[10]

Poté jsem naměřené hodnoty roztřídil do tříd, jejichž počet jsem odhadnul na základě vzorce $k = 1 + 3{,}3 \cdot \log n \doteq 8{,}9$, kde k je počet tříd a n celkový počet hodnot.[11] Po analýze naměřených hodnot jsem počet tříd upravil na 11. Aby měl test dobrou vypovídající hodnotu, je třeba splnit podmínku, že v každé třídě je alespoň 5 hodnot, a je nutné, aby žádná z tříd nebyla „prázdná“.[10] Nyní jsem mohl jednotlivé naměřené hodnoty n_i^e rozřadit do tříd.

V dalším kroku jsem vypočítal předpokládané doby zpoždění podle vztahu $n_i^t = np_i^t = n[F(x_i) - F(x_{i-1})]$, kde n_i^t je předpokládaná doba zpoždění, n celkový počet naměřených dob zpoždění, p_i^t předpokládaná pravděpodobnost nastání zpoždění z dané třídy zpoždění a $F(x)$ distribuční funkce, do které dosazujeme x_i, x_{i-i} jako meze dané třídy.[10] Tedy doby korespondující právě s exponenciálním rozdělením.

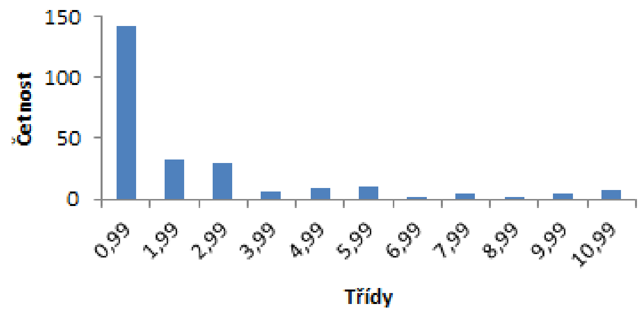
Nyní mohu porovnat naměřené doby zpoždění s předpokládanými. K tomu jsem použil X^2 test shody s $k - r - 1$ stupni volnosti, kde k je počet tříd a r počet odhadovaných parametrů, a to na hladině významnosti $\alpha = 0{,}05$. Testovací kritérium R má tvar $R = \sum_{i=1}^{k} \frac{(n_i^e - n_i^t)^2}{n_i^t}$.[10] Kritická hodnota je v tomto případě tedy $\mathrm{X}^2_{0{,}95;\,9} = 16{,}919$. Výsledky jednotlivých kroků jsou uvedeny v tabulce.

Navíc je přidán histogram naměřených hodnot. Hranice tříd jsem určil tak, aby mohly být použity i pro vygenerované hodnoty. Naměřené hodnoty byly uvedeny jako celá čísla, kdežto ty vygenerované jako čísla desetinná. Zaokrouhlením těchto hodnot by se změnila střední hodnota celého souboru.

Tab. 1: Vypočtené hodnoty – naměřené doby zpoždění

Hranice tříd		n_i^e	p_i^t	$n_i^t = np_i^t$	$\frac{(n_i^e - n_i^t)^2}{n_i^t}$
Předchozí	0,99	143	0,447886	111,9714	8,598393
0,99	1,99	32	0,249108	62,2769	14,71959
1,99	2,99	29	0,136713	34,17829	0,784552
2,99	3,99	6	0,07503	18,75744	8,67668
3,99	4,99	9	0,041177	10,2943	0,162733
4,99	5,99	11	0,022599	5,649633	5,066954
5,99	6,99	2	0,012402	3,100584	0,390664
6,99	7,99	4	0,006807	1,701637	3,104349
7,99	8,99	2	0,003736	0,933878	1,217093
8,99	9,99	5	0,00205	0,512523	39,29081
9,99	Další	7	0,002494	0,623417	65,22245
Součet		**250**	**1**	**250**	**147,2343**

Jak je vidět v tabulce, hodnota testovacího kritéria je výrazně vyšší než kritická hodnota! Můžeme tedy říct, že hodnoty zpoždění tohoto souboru se neřídí dle exponenciálního rozdělení. Tento fakt byl nejspíše způsoben znatelným množstvím extrémních hodnot a tím, že velké množství vlaků jelo dle jízdního řádu. Doby zpoždění jsou zobrazeny na histogramu.



Obr. 26: Histogram – naměřené doby zpoždění

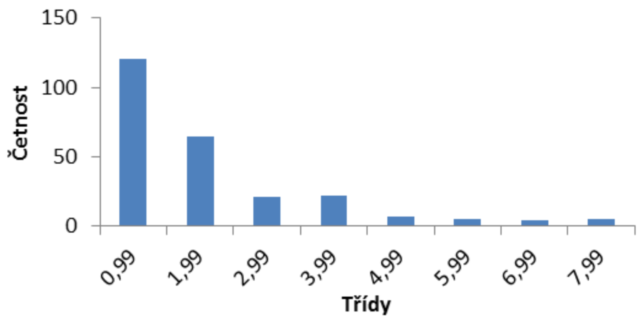
Jelikož se soubor naměřených hodnot neřídí předpokládaným rozdělením, není potřeba jej porovnávat s vygenerovanými hodnotami. Zkusíme tedy aspoň potvrdit hypotézu, že vygenerované hodnoty se řídí exponenciálním rozdělením pravděpodobnosti.

Nejdříve vygeneruji doby zpoždění. Jako střední hodnotu použiji tu výše odhadnutou, tedy $\mu = 0{,}6$. Počet dob bude také zachován, tedy 250. Nyní budu pokračovat obdobně jako v předcházejícím případě. Jelikož se v tomto vygenerovaném souboru nevyskytují extrémní hodnoty, jako tomu bylo ve skutečnosti, je potřeba změnit počet tříd a tím dodržet pravidla pro testování pomocí X^2 testu shody. Po analýze hodnot jsem určil jako vhodný počet 8 tříd. Změní se tedy i kritická hodnota, pak $\mathrm{X}^2_{0,95;\,6} = 12{,}592$.

Tab. 2: Vypočtené hodnoty – vygenerované doby zpoždění

Hranice tříd		n_i^e	p_i^t	$n_i^t = np_i^t$	$\frac{(n_i^e - n_i^t)^2}{n_i^t}$
Předchozí	0,99	121	0,447886	111,9714	0,728004
0,99	1,99	65	0,249108	62,2769	0,11907
1,99	2,99	21	0,136713	34,17829	5,081215
2,99	3,99	22	0,07503	18,75744	0,560534
3,99	4,99	7	0,041177	10,2943	1,054217
4,99	5,99	5	0,022599	5,649633	0,074699
5,99	6,99	4	0,012402	3,100584	0,260902
6,99	Další	5	0,015086	3,771455	0,400196
Součet		**250**	**1**	**250**	**8,278838**

Jelikož je hodnota testovacího kritéria menší než hodnota kritická, můžeme potvrdit naší hypotézu. To znamená, že hodnoty vygenerované generátorem se řídí exponenciálním rozdělením pravděpodobnosti. Tyto hodnoty jsou zobrazeny na histogramu.



Obr. 27: Histogram – vygenerované doby zpoždění

4.2 Normální rozdělení

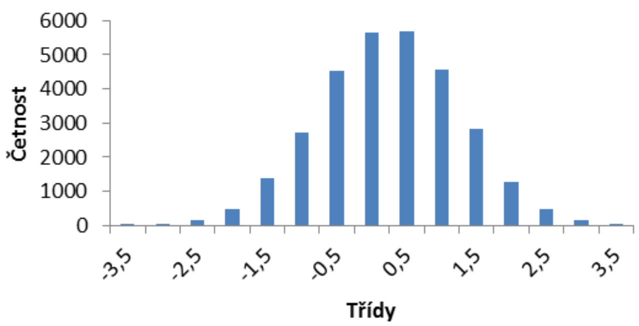
Vygenerované hodnoty normálního rozdělení jsou porovnány s hodnotami vygenerovanými tabulkovým procesorem Excel od firmy Microsoft. Nejdříve je testováno, zda se oba soubory hodnot řídí normálním rozdělením pravděpodobnosti. Pokud ano, jsou vzájemně porovnány.

Oběma generátory jsem vygeneroval 30 000 hodnot normálního normovaného rozdělení. Dále jsem postupoval obdobně jako při ověřování generátoru exponenciálního rozdělení. Vygenerované hodnoty jsem rozčlenil do 15 tříd. Kritická hodnota je v tomto případě $\mathrm{X}^2_{0,95;\,13} = 22{,}362$. Nejprve jsem testoval hodnoty vygenerované aplikací Excel. Vypočítané hodnoty jsou uvedeny v tabulce a náhodná čísla jsou znázorněna histogramem.

Tab. 3: Vypočtené hodnoty – Excelem vygenerované hodnoty $N(0, 1)$

Hranice tříd		$\boldsymbol{n_i^e}$	$\boldsymbol{p_i^t}$	$\boldsymbol{n_i^t = np_i^t}$	$\boldsymbol{\frac{(n_i^e - n_i^t)^2}{n_i^t}}$
Předchozí	-3,5	4	0,000229	6,876942	1,203558
-3,5	-3	28	0,001117	33,51807	0,908438
-3	-2,5	158	0,00486	145,793	1,022068
-2,5	-2	489	0,01654	496,214	0,104878
-2	-1,5	1380	0,044057	1321,712	2,570516
-1,5	-1	2723	0,091848	2755,442	0,381956
-1	-0,5	4520	0,149882	4496,469	0,123148
-0,5	0	5658	0,191462	5743,874	1,283858
0	0,5	5699	0,191462	5743,874	0,350575
0,5	1	4574	0,149882	4496,469	1,336855
1	1,5	2820	0,091848	2755,442	1,512567
1,5	2	1291	0,044057	1321,712	0,713644
2	2,5	463	0,01654	496,214	2,223173
2,5	3	153	0,00486	145,793	0,356262
3	Další	40	0,001117	33,51807	1,253516
Součet		**30 000**	**1**	**30 000**	**15,34501**

Jelikož je hodnota testovacího kritéria menší než hodnota kritická, můžeme potvrdit, že hodnoty vygenerované aplikací Excel se řídí normálním normovaným rozdělením pravděpodobnosti.



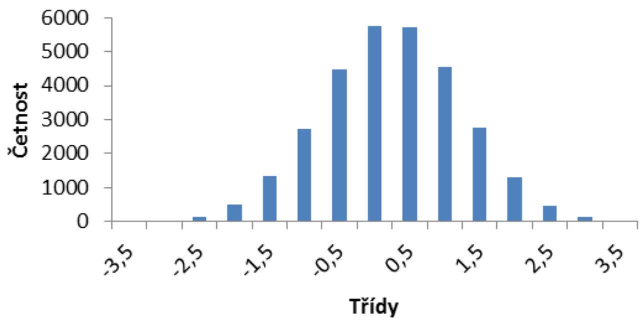
Obr. 28: Histogram – Excelem vygenerované hodnoty $N(0, 1)$

Poté jsem testoval hodnoty vygenerované mnou vytvořenou aplikací. I zde jsou vypočítané hodnoty uvedeny v tabulce a náhodná čísla jsou znázorněna histogramem.

Tab. 4: Vypočtené hodnoty – mou aplikací vygenerované hodnoty $N(0, 1)$

Hranice tříd		n_i^e	p_i^t	$n_i^t = np_i^t$	$\frac{(n_i^e - n_i^t)^2}{n_i^t}$
Předchozí	-3,5	9	0,000229	6,876942	0,655433
-3,5	-3	25	0,001117	33,51807	2,164728
-3	-2,5	136	0,00486	145,793	0,657804
-2,5	-2	500	0,01654	496,214	0,028886
-2	-1,5	1332	0,044057	1321,712	0,080079
-1,5	-1	2734	0,091848	2755,442	0,166849
-1	-0,5	4493	0,149882	4496,469	0,002676
-0,5	0	5751	0,191462	5743,874	0,008841
0	0,5	5710	0,191462	5743,874	0,199767
0,5	1	4567	0,149882	4496,469	1,106354
1	1,5	2765	0,091848	2755,442	0,033157
1,5	2	1324	0,044057	1321,712	0,00396
2	2,5	473	0,01654	496,214	1,086003
2,5	3	148	0,00486	145,793	0,033409
3	Další	33	0,001117	33,51807	0,008007
Součet		**30 000**	**1**	**30 000**	**6,235953**

Jelikož je hodnota testovacího kritéria menší než hodnota kritická, můžeme potvrdit, že hodnoty vygenerované mou aplikací se řídí normálním normovaným rozdělením pravděpodobnosti, a to s lepším výsledkem než Excel.



Obr. 29: Histogram – mou aplikací vygenerované hodnoty $N(0, 1)$

Protože se oba generátory řídí normálním rozdělením pravděpodobnosti, porovnal jsem vygenerované hodnoty navzájem. Použil jsem X^2 test shody, kde byl soubor hodnot vygenerovaný mým generátorem použit jako zjištěné hodnoty a soubor hodnot aplikace Excel jako hodnoty předpokládané.

Tab. 5: Vypočtené hodnoty – porovnání hodnot $N(0, 1)$

Hranice tříd		n_i^t	n_i^e	$\frac{(n_i^e - n_i^t)^2}{n_i^t}$
Předchozí	-3,5	4	9	6,25
-3,5	-3	28	25	0,321429
-3	-2,5	158	136	3,063291
-2,5	-2	489	500	0,247444
-2	-1,5	1380	1332	1,669565
-1,5	-1	2723	2734	0,044436
-1	-0,5	4520	4493	0,161283
-0,5	0	5658	5751	1,528632
0	0,5	5699	5710	0,021232
0,5	1	4574	4567	0,010713
1	1,5	2820	2765	1,072695
1,5	2	1291	1324	0,843532
2	2,5	463	473	0,215983
2,5	3	153	148	0,163399
3	Další	40	33	1,225
Součet		**30 000**	**30 000**	**16,83863**

Neboť i tentokrát vyšla hodnota testovacího kritéria nižší než ta kritická, můžeme potvrdit, že hodnoty mého generátoru mohou být použity místo těch, které byly získány tabulkovým procesorem Excel.

4.3 Beta rozdělení

Vygenerované hodnoty beta rozdělení jsou porovnány s hodnotami vygenerovanými aplikací Matlab. Nejdříve je testováno, zda se oba soubory hodnot řídí beta rozdělením pravděpodobnosti. Pokud ano, jsou vzájemně porovnány.

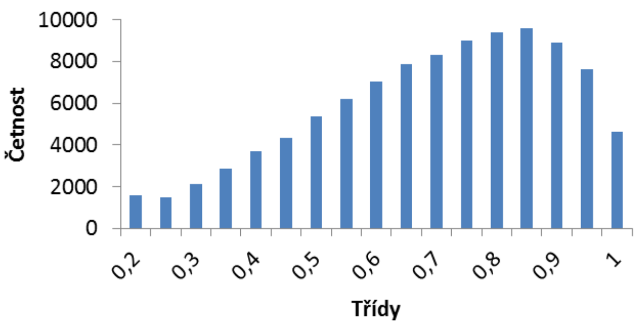
Oběma generátory jsem vygeneroval 100 000 hodnot beta rozdělení s parametry $\mathrm{B}(3; 1{,}5)$. Dále jsem postupoval obdobně jako při ověřování generátoru normálního rozdělení. Vygenerované hodnoty jsem roztřídil do 17 tříd. Kritická hodnota je v tomto případě $\mathrm{X}^2_{0{,}95;\,15} = 24{,}996$.

Nejprve jsem testoval hodnoty vygenerované aplikací Matlab. Vypočítané hodnoty jsou uvedeny v tabulce a pseudonáhodná čísla jsou znázorněna histogramem.

Tab. 6: Vypočtené hodnoty – Matlabem vygenerované hodnoty $B(3; 1{,}5)$

Hranice tříd		$\boldsymbol{n_i^e}$	$\boldsymbol{p_i^t}$	$\boldsymbol{n_i^t = np_i^t}$	$\boldsymbol{\frac{(n_i^e - n_i^t)^2}{n_i^t}}$
Předchozí	0,2	1582	0,01613	1613,009	0,596127
0	0,25	1488	0,014666	1466,57	0,313147
0,2	0,3	2114	0,021164	2116,382	0,002681
0,25	0,35	2879	0,028502	2850,21	0,290799
0,3	0,4	3703	0,036498	3649,809	0,775201
0,35	0,45	4325	0,044952	4495,183	6,442989
0,4	0,5	5359	0,053642	5364,178	0,004999
0,45	0,55	6195	0,062319	6231,891	0,218388
0,5	0,6	7039	0,070698	7069,847	0,134594
0,55	0,65	7853	0,078448	7844,763	0,008649
0,6	0,7	8326	0,085166	8516,617	4,266338
0,65	0,75	8984	0,090354	9035,447	0,292929
0,7	0,8	9410	0,093356	9335,569	0,593422
0,75	0,85	9606	0,093239	9323,946	8,532296
0,8	0,9	8905	0,088525	8852,517	0,311149
0,85	0,95	7621	0,076309	7630,906	0,01286
0,9	Další	4611	0,046032	4603,156	0,013368
Součet		**100 000**	**1**	**100 000**	**22,80994**

Jelikož je hodnota testovacího kritéria menší než hodnota kritická, můžeme potvrdit, že vygenerované hodnoty aplikací Matlab se řídí beta rozdělením pravděpodobnosti.



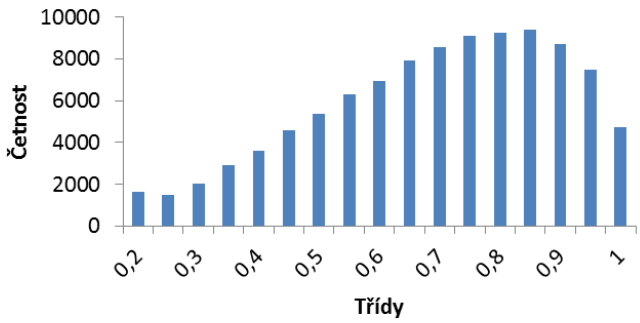
Obr. 30: Histogram – Matlabem vygenerované hodnoty $B(3; 1{,}5)$

Poté jsem testoval hodnoty vygenerované mnou vytvořenou aplikací. I zde jsou vypočítané hodnoty uvedeny v tabulce a náhodná čísla jsou znázorněna histogramem.

Tab. 7: Vypočtené hodnoty – mou aplikací vygenerované hodnoty $B(3; 1{,}5)$

Hranice tříd		n_i^e	p_i^t	$n_i^t = np_i^t$	$\frac{(n_i^e - n_i^t)^2}{n_i^t}$
Předchozí	0,2	1624	0,01613	1613,009	0,074893
0	0,25	1491	0,014666	1466,57	0,406958
0,2	0,3	2048	0,021164	2116,382	2,209471
0,25	0,35	2936	0,028502	2850,21	2,582211
0,3	0,4	3606	0,036498	3649,809	0,525832
0,35	0,45	4563	0,044952	4495,183	1,023113
0,4	0,5	5355	0,053642	5364,178	0,015704
0,45	0,55	6315	0,062319	6231,891	1,108339
0,5	0,6	6920	0,070698	7069,847	3,176057
0,55	0,65	7933	0,078448	7844,763	0,992486
0,6	0,7	8564	0,085166	8516,617	0,263622
0,65	0,75	9093	0,090354	9035,447	0,366601
0,7	0,8	9234	0,093356	9335,569	1,105055
0,75	0,85	9387	0,093239	9323,946	0,426413
0,8	0,9	8720	0,088525	8852,517	1,983707
0,85	0,95	7503	0,076309	7630,906	2,14391
0,9	Další	4708	0,046032	4603,156	2,388004
Součet		**100 000**	**1**	**100 000**	**20,79238**

Jelikož je hodnota testovacího kritéria menší než hodnota kritická, můžeme potvrdit, že hodnoty vygenerované mou aplikací se řídí beta rozdělením pravděpodobnosti, a to s lepším výsledkem než Matlab.



Obr. 31: Histogram – mou aplikací vygenerované hodnoty $B(3; 1{,}5)$

Protože se oba generátory řídí beta rozdělením pravděpodobnosti, porovnal jsem vygenerované hodnoty navzájem. Použil jsem X^2 test shody, kde byl soubor hodnot vygenerovaný mým generátorem použit jako zjištěné hodnoty a soubor hodnot aplikace Matlab jako hodnoty předpokládané.

Tab. 8: Vypočtené hodnoty – porovnání hodnot $B(3; 1{,}5)$

Hranice tříd		$\boldsymbol{n_i^t}$	$\boldsymbol{n_i^e}$	$\boldsymbol{\frac{(n_i^e - n_i^t)^2}{n_i^t}}$
Předchozí	0,2	1582	1624	1,115044
0,2	0,25	1488	1491	0,006048
0,25	0,3	2114	2048	2,060549
0,3	0,35	2879	2936	1,128517
0,35	0,4	3703	3606	2,540913
0,4	0,45	4325	4563	13,09688
0,45	0,5	5359	5355	0,002986
0,5	0,55	6195	6315	2,324455
0,55	0,6	7039	6920	2,011791
0,6	0,65	7853	7933	0,814975
0,65	0,7	8326	8564	6,803267
0,7	0,75	8984	9093	1,322462
0,75	0,8	9410	9234	3,291817
0,8	0,85	9606	9387	4,992817
0,85	0,9	8905	8720	3,843346
0,9	0,95	7621	7503	1,827057
0,95	Další	4611	4708	2,040555
Součet		**100 000**	**100 000**	**49,22348**

Neboť tentokrát nevyšla hodnota testovacího kritéria nižší než ta kritická, nemůžeme potvrdit, že hodnoty mého generátoru mohou být použity místo těch, které byly získány aplikací Matlab. To je s největší pravděpodobností způsobeno použitím jiného algoritmu pro generování, než jsem použil já.

5 ZÁVĚR

Programový nástroj pro generování hodnot náhodné proměnné pracuje správně a může být nasazen v praxi. Toto tvrzení je založenu na provedeném testování tří rozdělení pravděpodobnosti důležitých v dopravě. U všech těchto souborů byla prokázána závislost na daném rozdělení pravděpodobnosti. V příloze A jsou uvedeny histogramy jako ukázka vygenerovaných hodnot všech rozdělení.

Tento nástroj je demonstrován ve zkušební aplikaci, která umožňuje generování pseudonáhodných čísel do textového souboru. Na výběr je celkem z 16 spojitých rozdělení pravděpodobnosti. Tato aplikace byla doplněna o výpočet základních charakteristik a vykreslování grafů hustoty pravděpodobnosti i distribuční funkce implementovaných rozdělení. Navíc je umožněno grafické porovnávání výše uvedených funkcí různých rozdělení najednou. Dále byl doplněn kalkulátor složitějších funkcí, které ve většině případů obsahují nevlastní integrál. Jednou z předností je uživatelsky přívětivé prostředí s intuitivním ovládáním.

U některých rozdělení bylo potřeba z důvodu univerzálnosti upravit algoritmy pro generování hodnot a vzorce pro výpočet charakteristik. Jedním z těchto případů bylo například beta rozdělení pravděpodobnosti, které je definováno na uzavřeném intervalu <0; 1>. Z důvodu zmíněné univerzálnosti jsem musel zakomponovat funkce pro přepočet na libovolný interval <min; max>. U rozdělení, která nejsou dána minimální a maximální hodnotou, bylo potřeba implementovat možnost posunutí – logaritmicko-normální, Studentovo a některá další rozdělení. Tyto úpravy jsem úspěšně dokončil.

Zkušební aplikace lze tedy použít jak pro generování hodnot náhodné proměnné, tak i jako studijní opora pro předměty zabývající se rozděleními pravděpodobnosti. Mohu tedy říct, že jsem splnil zadání práce, a to v celém rozsahu.

Aplikace by mohla být dále doplněna o diskrétní a případně i vícerozměrná rozdělení pravděpodobnosti. Jednou z vhodných úprav by bylo přepracování struktury programu dle zásad třívrstvého návrhu aplikace. Dalším možným krokem by bylo převedení jednotlivých tříd do jiných, dnes více frekventovaných, programovacích jazyků.

6 POUŽITÁ LITERATURA

1. LINDA, Bohdan. *Pravděpodobnost*. Vyd. 1. Pardubice: Univerzita Pardubice, Fakulta ekonomicko-správní, 2010, 167 s. ISBN 978-80-7395-303-4.
2. ZÁHOROVÁ, Věra. *Diskrétní náhodná veličina a její charakteristiky* [intranet]. Pardubice, 2011 [cit. 2013-05-19]. Dostupné z: https://portal.upce.cz/portal/moje-studium/materialy.html
3. ZÁHOROVÁ, Věra. *Spojitá náhodná veličina a její charakteristiky* [intranet]. Pardubice, 2011 [cit. 2013-05-19]. Dostupné z: https://portal.upce.cz/portal/moje-studium/materialy.html
4. ZÁHOROVÁ, Věra. *Často používaná rozdělení pravděpodobnosti náhodných veličin* [intranet]. Pardubice, 2011 [cit. 2013-05-19]. Dostupné z: https://portal.upce.cz/portal/moje-studium/materialy.html
5. SAUCIER, Richard. *Computer Generation of Statistical Distributions* [online]. 2000 [cit. 2013-05-19]. ISBN [9781423540281]. Dostupné z: http://ftp.arl.mil/random/random.pdf
6. HUŠEK, Roman a Josef LAUBER. *Simulační modely*. Praha: SNTL - Nakladatelství technické literatury, 1987. DT 330.116.1(075.8).
7. KEEN, Sitt Chee. *Excel Function: NORMSDIST(z)*. CodeProject: For those who code [online]. 2012 [cit. 2013-05-19]. Dostupné z: http://www.codeproject.com/Articles/408214/Excel-Function-NORMSDIST-z
8. PRESS, William H., Saul A. TEUKOLSKY, William T. VETTERLING a Brian P. FLANNERY. *Numerical Recipes: The Art of Scientific Computing*. 3rd ed. Cambridge: Cambridge University Press, 2007, xxi, 1235 s. ISBN 978-0-521-88068-8.
9. VESELÝ, Petr. *Transformace 2D* [intranet]. Pardubice, 2011 [cit. 2013-05-19]. Dostupné z: https://portal.upce.cz/portal/moje-studium/materialy.html
10. ZÁHOROVÁ, Věra. *Testy shody* [intranet]. Pardubice, 2011 [cit. 2013-05-19]. Dostupné z: https://portal.upce.cz/portal/moje-studium/materialy.html
11. KAVIČKA, Antonín. *Modelování a simulace* [intranet]. Pardubice, 2012 [cit. 2013-05-19]. Dostupné z: http://fei-learn.upceucebny.cz

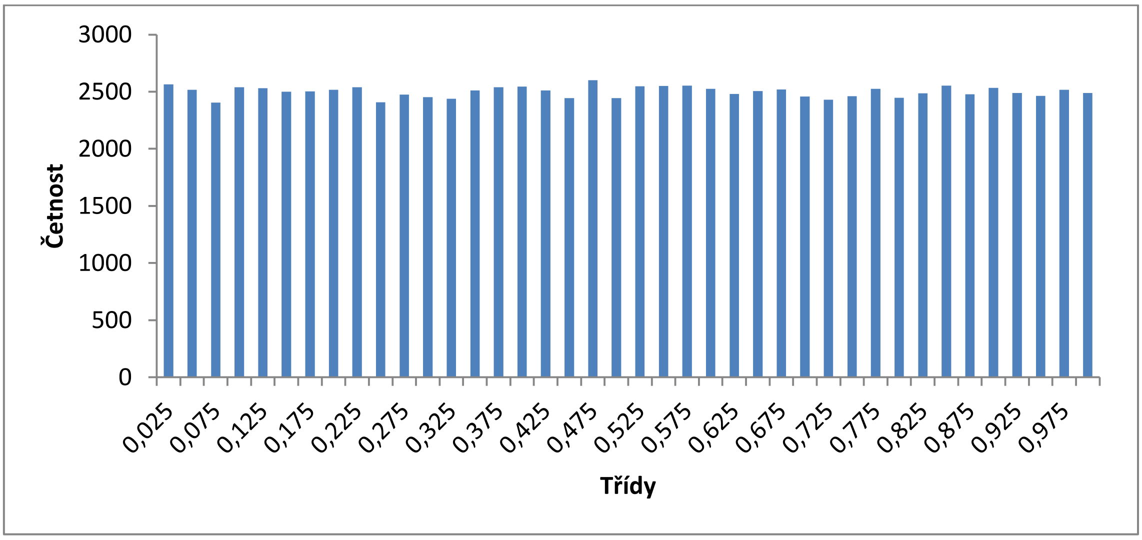
7 PŘÍLOHY

PŘÍLOHA A *Ukázka výstupu generátoru*

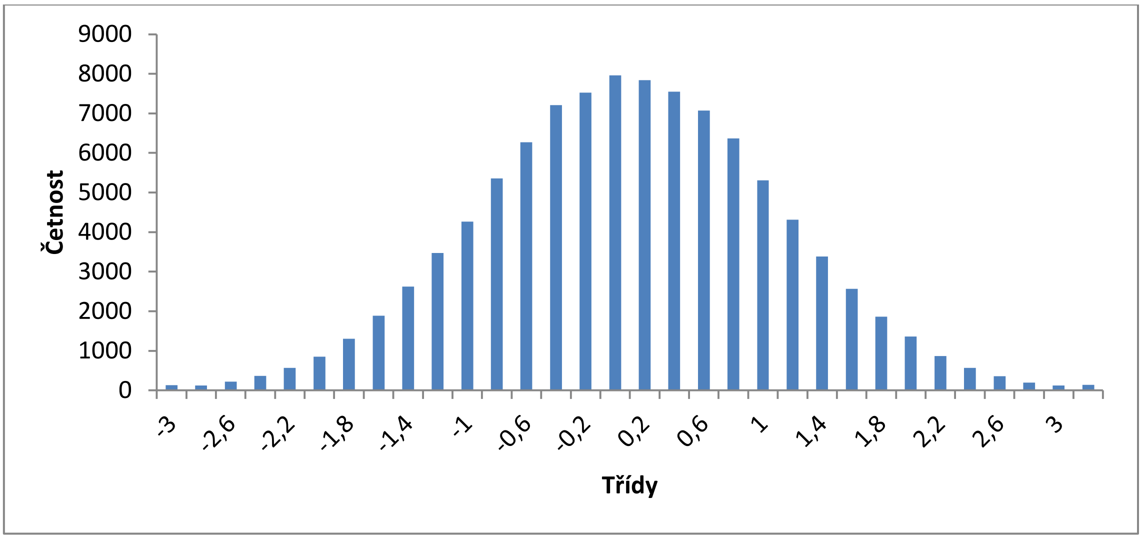
Tato příloha obsahuje histogramy vygenerovaných hodnot všech implementovaných rozdělení pravděpodobnosti. Histogramy jsou vytvořeny v tabulkovém procesoru Excel firmy Microsoft. Vždy je uveden název rozdělení s hodnotami jeho parametrů v závorce.

Na histogramu Cauchyova rozdělení pravděpodobnosti je vidět, že nebývá velkého množství extrémních hodnot. Právě proto je vhodné pro testování simulačních modelů.

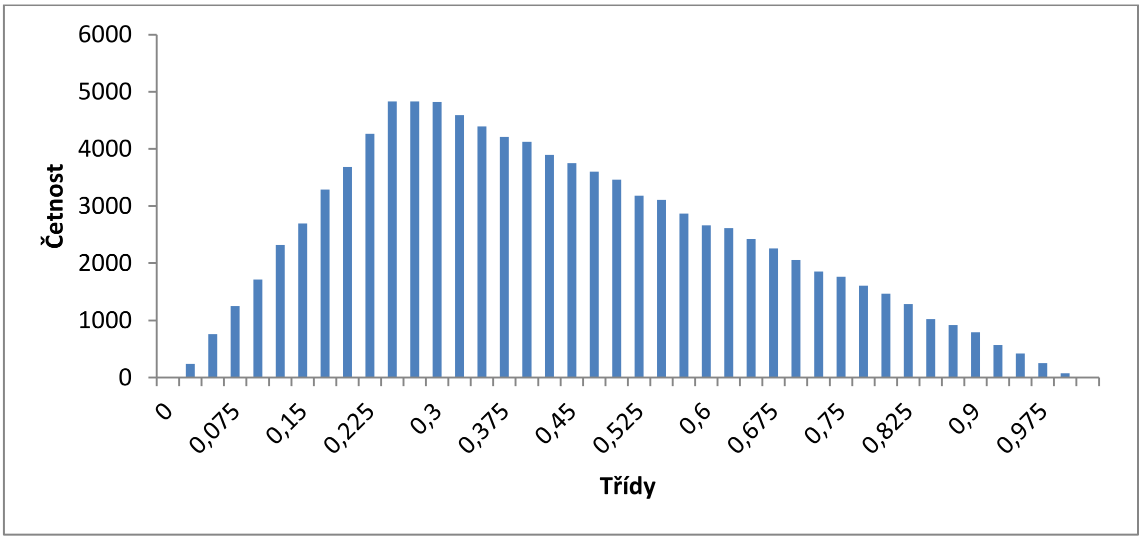
Rovnoměrné rozdělení (0; 1)



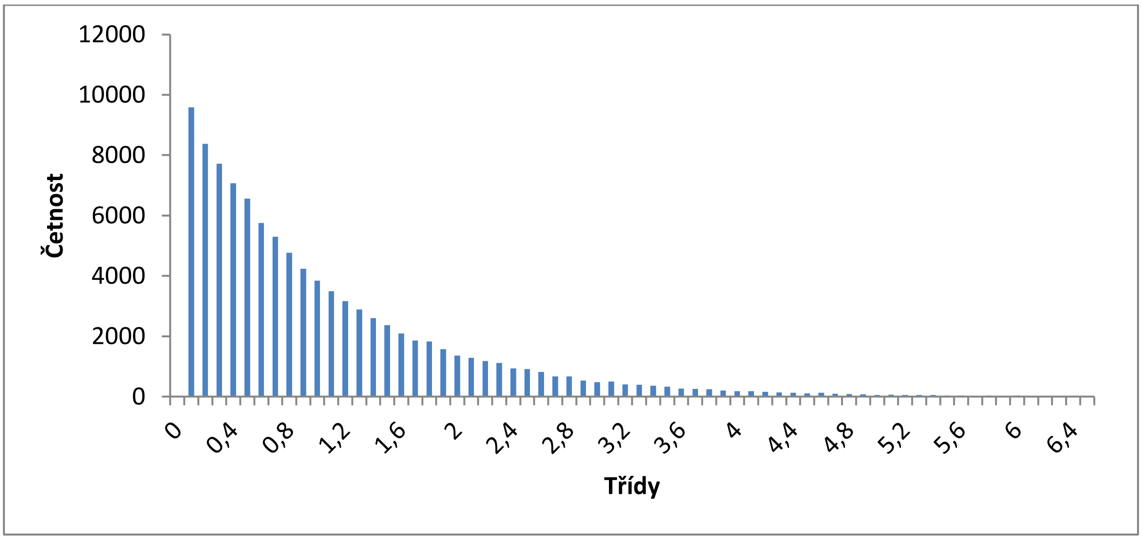
Normální rozdělení (0; 1)



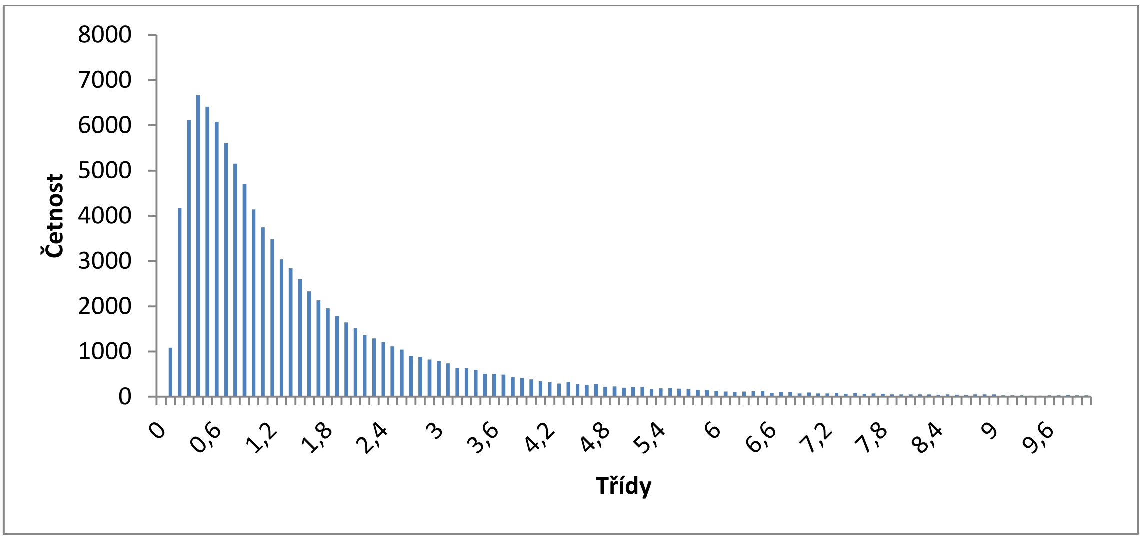
Trojúhelníkové rozdělení (0; 1; 0,25)



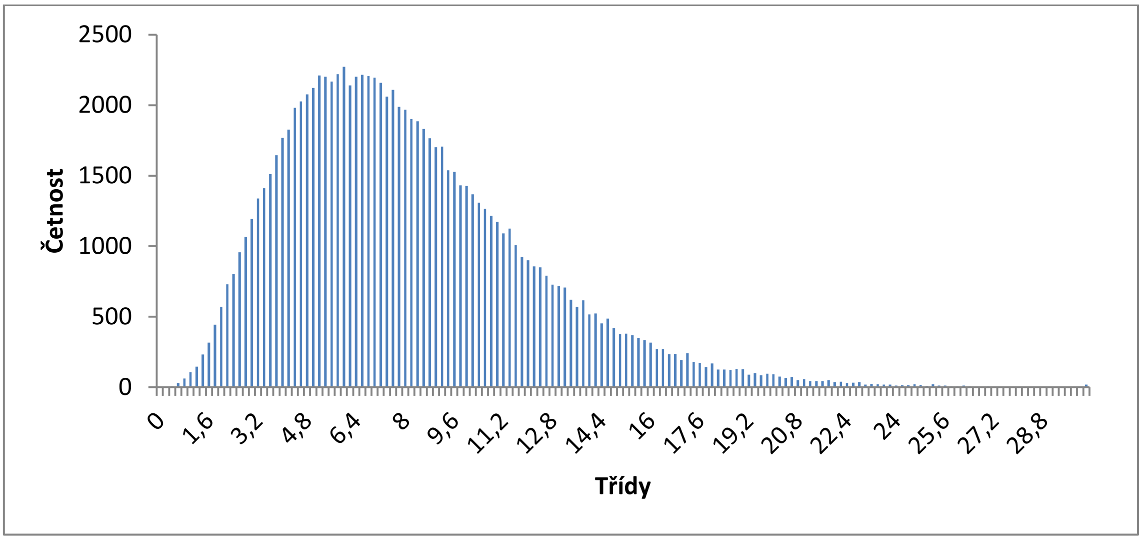
Exponenciální rozdělení (1; 0)



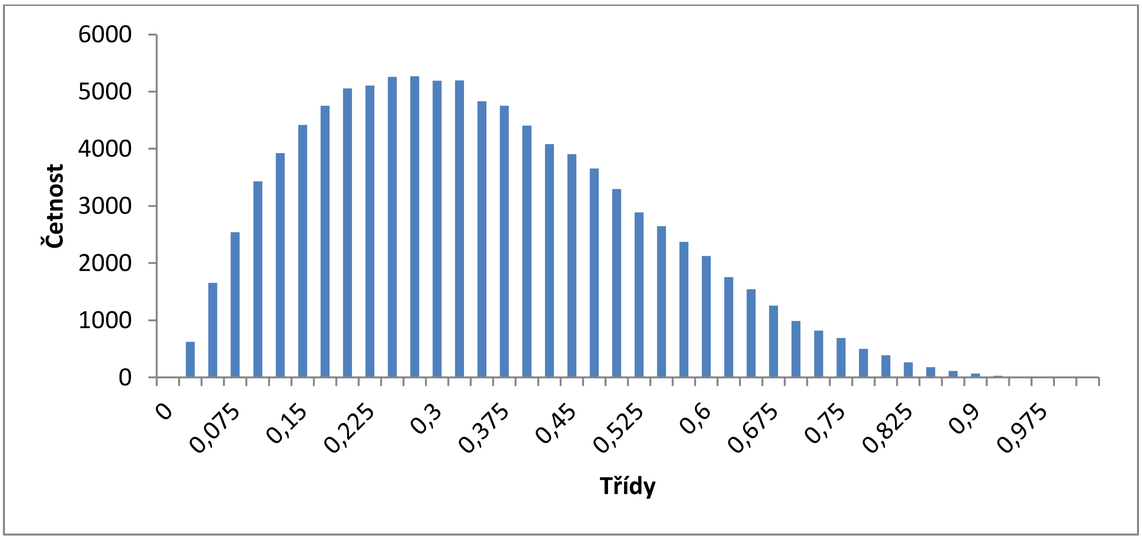
Logaritmicko-normální rozdělení (0; 1)



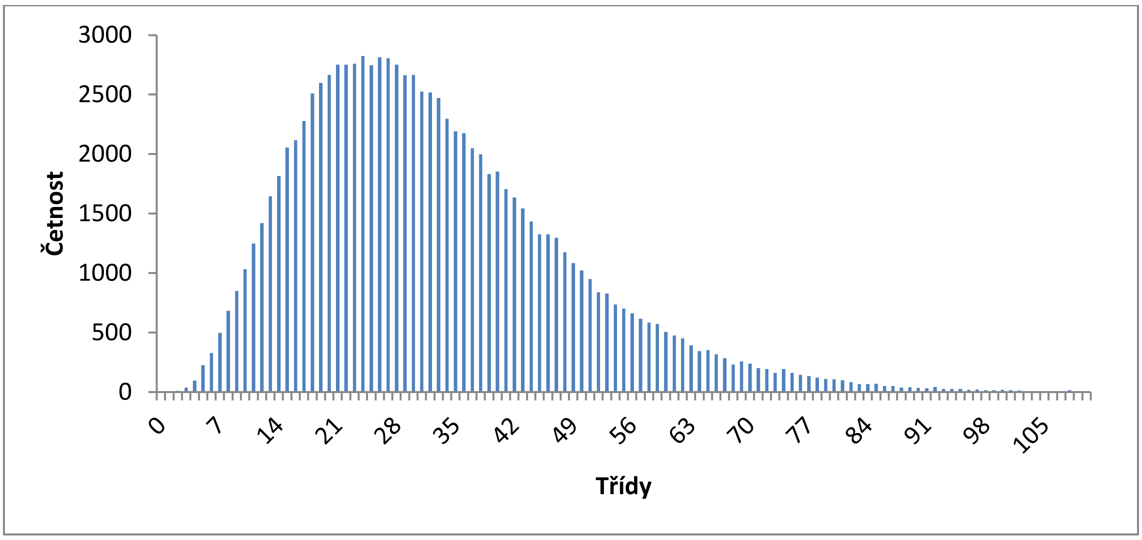
Erlangovo rozdělení (4; 2)



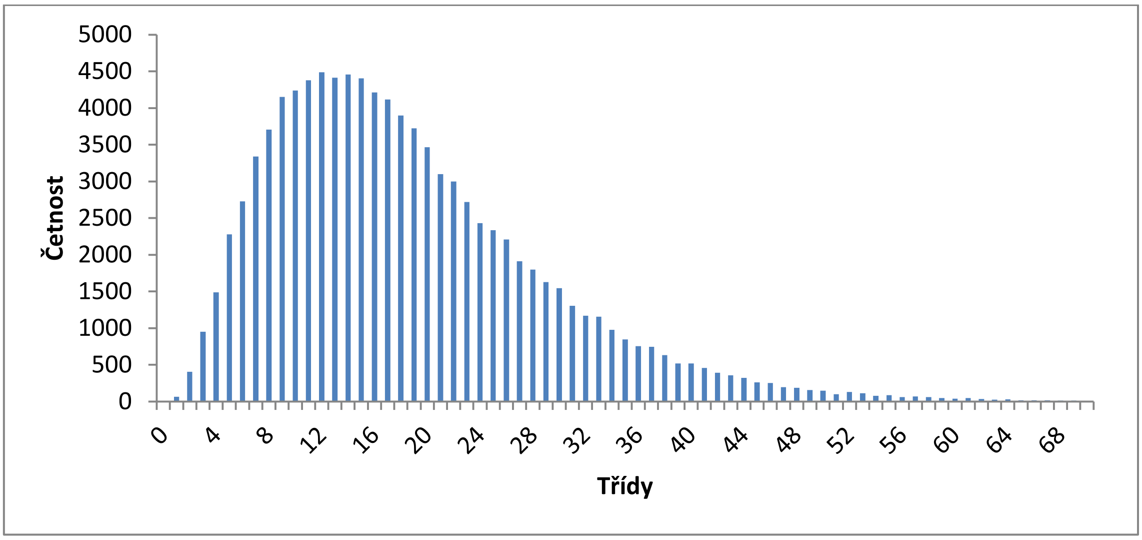
Beta rozdělení (2; 4)



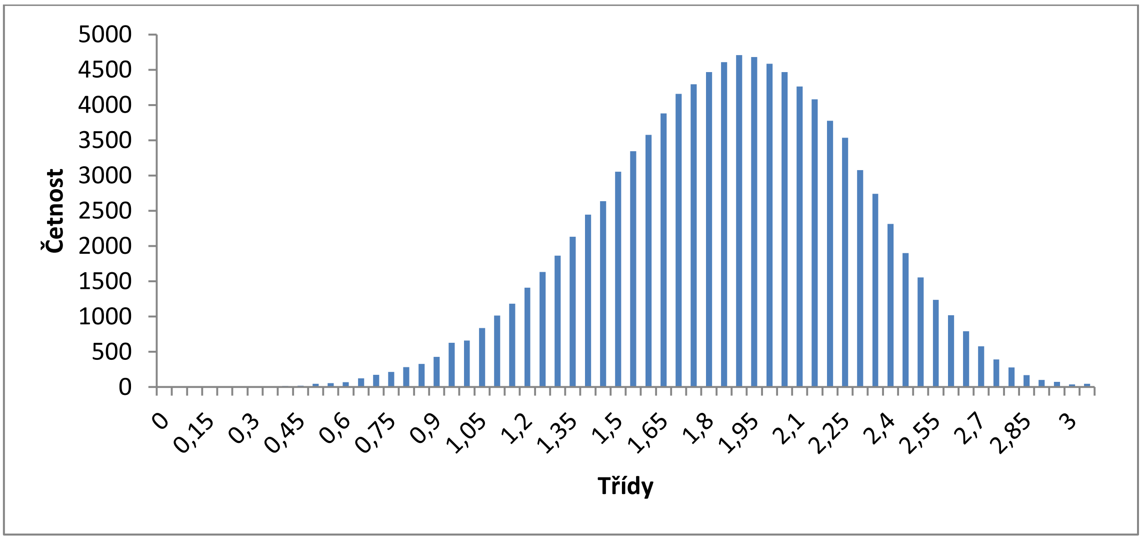
Gama rozdělení (4; 8)



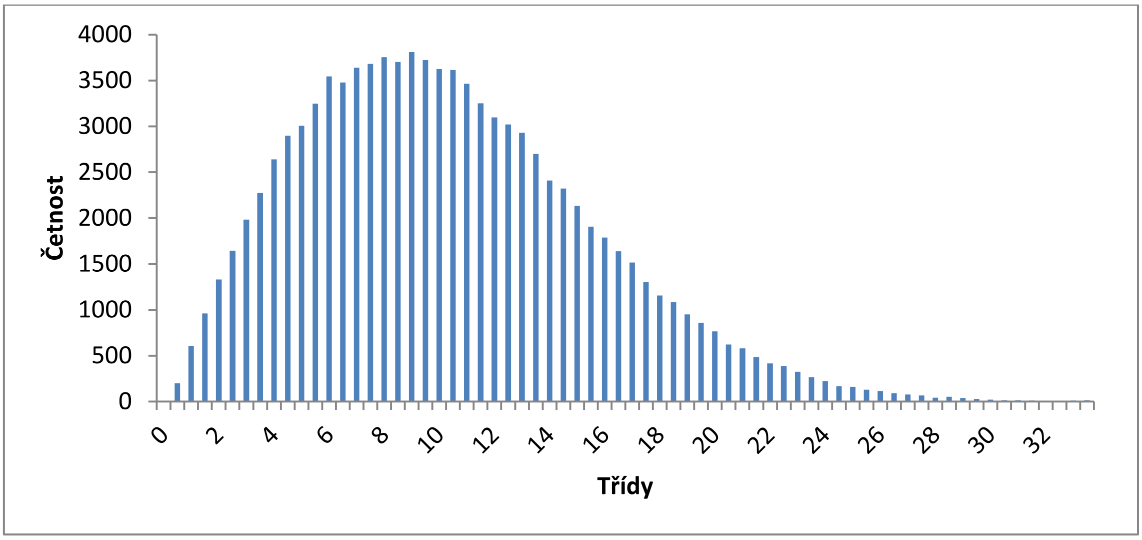
Chí-kvadrát rozdělení (6)



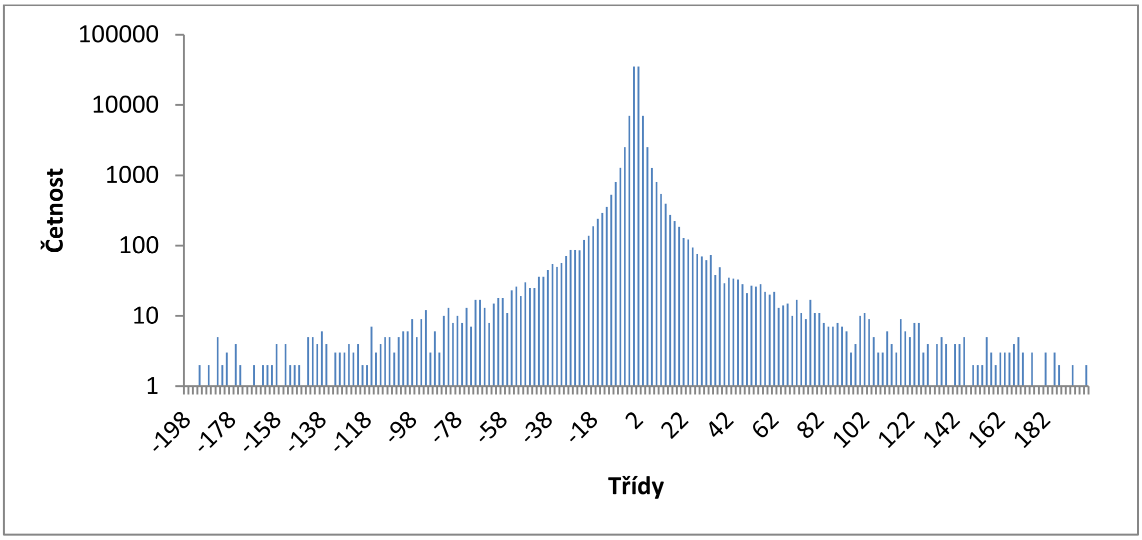
Weibullovo rozdělení (2; 5)



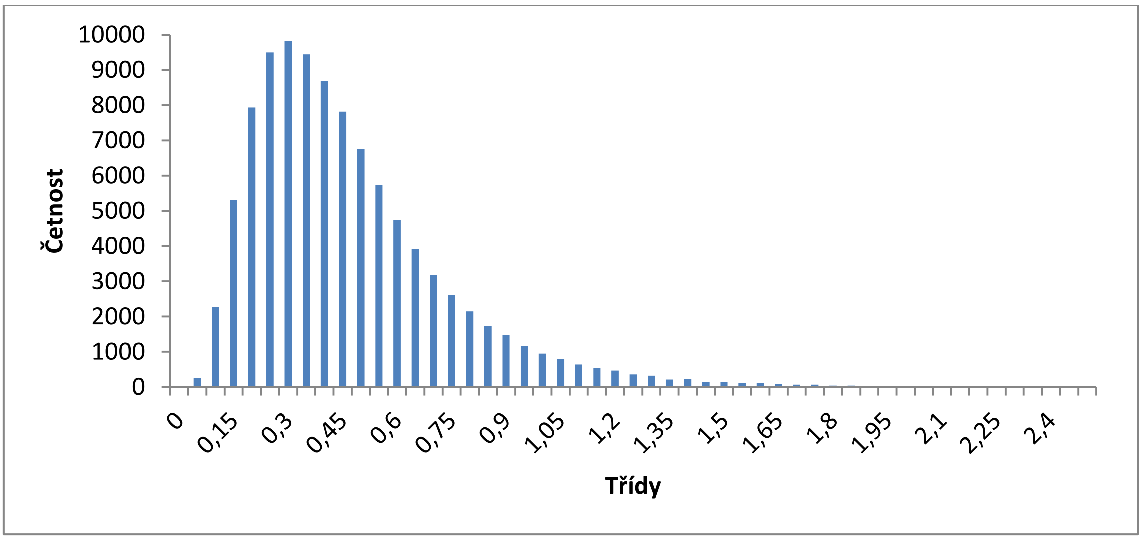
Rayleighovo rozdělení (8)



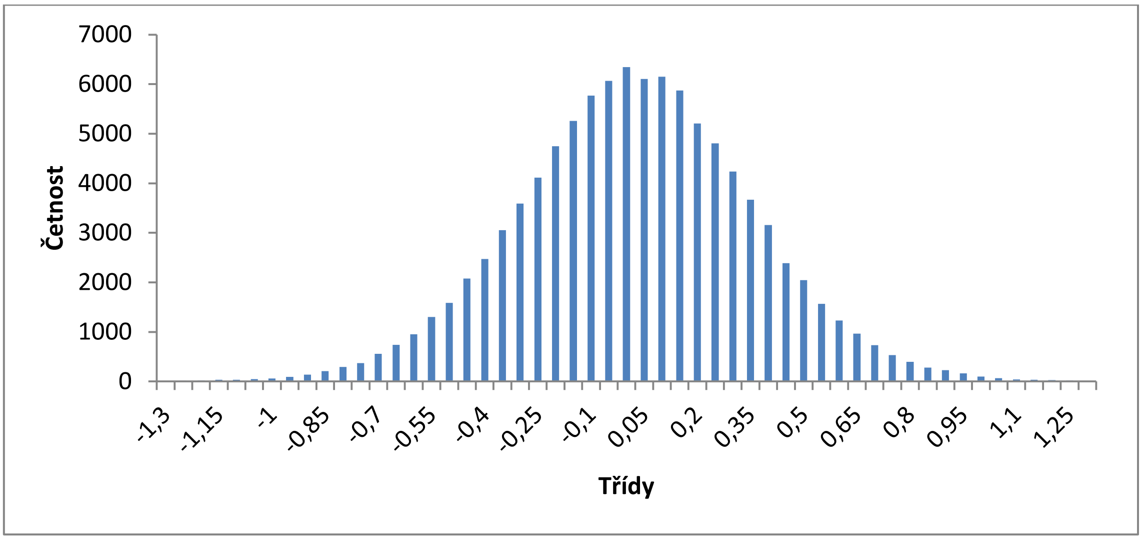
Cauchyovo rozdělení (0; 1)



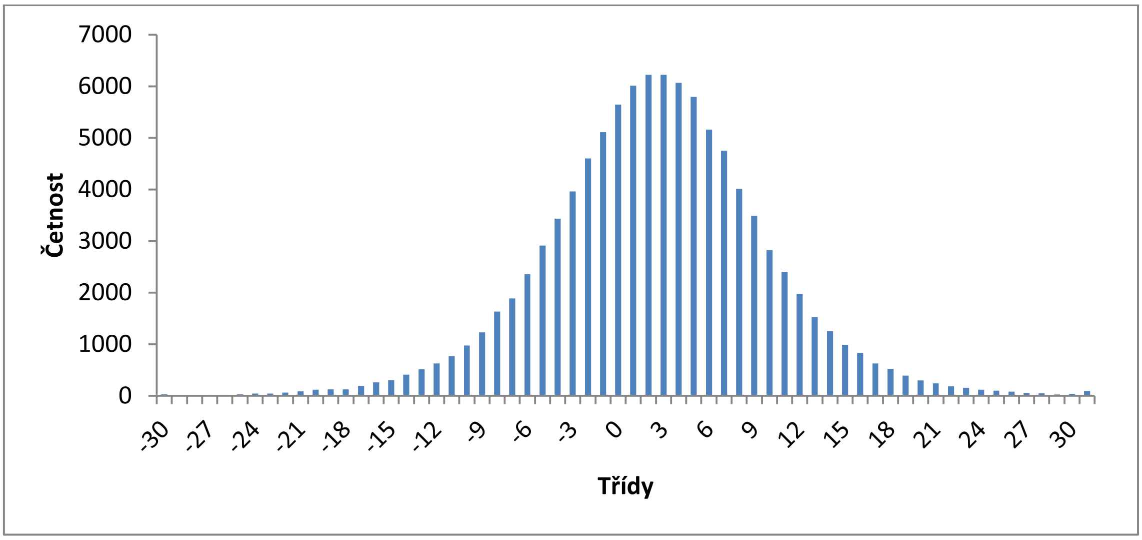
F rozdělení (8; 20)



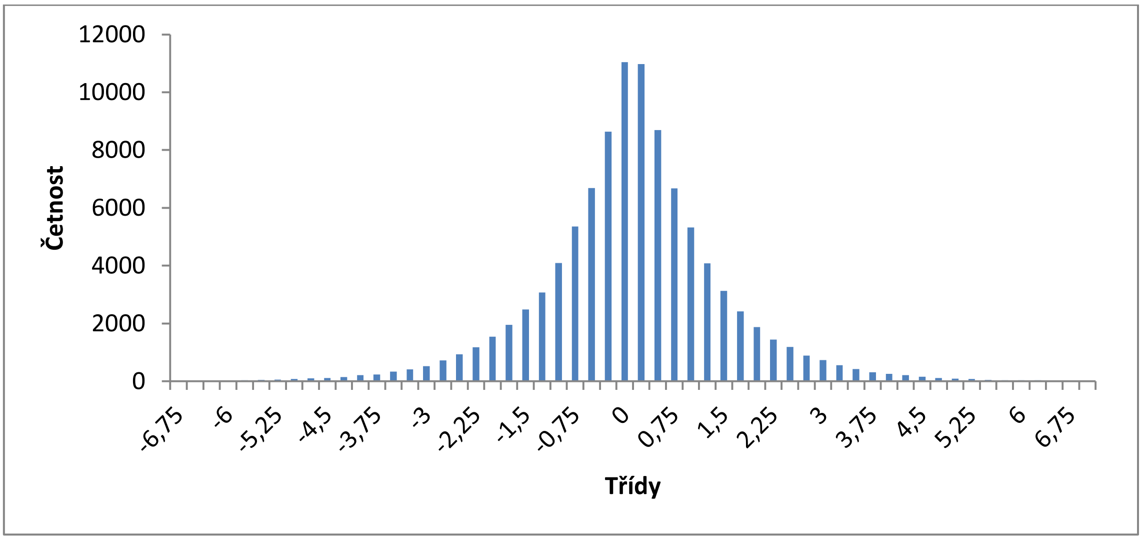
Studentovo rozdělení (20)



Logistické rozdělení (2; 4)



Laplaceovo rozdělení (0; 1)



PŘÍLOHA B *Přiložené datové médium*

K bakalářské práci je přiloženo datové médium se zkušební aplikací a kopií této práce ve formátu **.pdf*.