

Matematika 3

Garant předmětu:

RNDr. Břetislav Fajmon, PhD

Autoři textu:

Mgr. Irena Růžičková

RNDr. Břetislav Fajmon, PhD

Obsah

I	NUMERICKÉ METODY	9
1	Chyby při numerických výpočtech	9
1.1	Zdroje a typy chyb	9
1.2	Definice chyb	9
1.3	Zaokrouhlování. Šíření chyb při výpočtu	10
1.4	Podmíněnost numerických úloh a numerická stabilita algoritmů	12
2	Exkurze do funkcionální analýzy	13
2.1	Metrický prostor	13
2.2	Úplný metrický prostor	14
2.3	Pevný bod zobrazení, iterační proces	16
2.4	Normovaný vektorový prostor	18
3	Numerické řešení soustavy lineárních rovnic	21
3.1	Přímé metody	21
3.1.1	Cramerovo pravidlo	21
3.1.2	Gaussova eliminační metoda	22
3.1.3	Eliminace s výběrem hlavního prvku	24
3.2	Iterační metody	25
3.2.1	Jacobiho metoda	25
3.2.2	Gauss-Seidelova metoda	30
3.3	Srovnání přímých a iteračních metod	33
4	Numerické metody řešení nelineárních rovnic	34
4.1	Numerické metody řešení jedné nelineární rovnice	34
4.1.1	Metoda půlení intervalu	35
4.1.2	Metoda regula falsi	37
4.1.3	Metoda sečen	39
4.1.4	Newtonova metoda (metoda tečen)	40
4.1.5	Metoda prosté iterace	43
4.2	Numerické metody řešení soustav nelineárních rovnic	47
4.2.1	Metoda prosté iterace	47
4.2.2	Newtonova metoda	49
5	Aproximace funkcí	53
5.1	Interpolace algebraickými polynomy	53
5.1.1	Existence a jednoznačnost interpolačního polynomu	54
5.1.2	Konstrukce interpolačního polynomu, Lagrangeův interpolační polynom	54
5.1.3	Newtonův interpolační polynom	55
5.1.4	Odhad chyby	58

5.2	Interpolace pomocí splajnů	59
5.3	Metoda nejmenších čtverců	63
6	Numerické derivování a integrování	67
6.1	Numerické derivování	67
6.1.1	Některé často používané vzorce pro numerické derivování	68
6.2	Numerické integrování	70
6.2.1	Newton-Cotesovy vzorce	70
6.2.2	Složené kvadrurní vzorce	74
7	Numerické řešení diferenciálních rovnic	77
7.1	Počáteční úlohy	77
7.2	Okrajové úlohy	77
II	PRAVDĚPODOBNOST	78
8	Pravděpodobnostní modely	78
8.1	Klasická pravděpodobnost	79
8.2	Geometrická pravděpodobnost	85
8.3	Diskrétní pravděpodobnost	87
8.4	Spojité pravděpodobnost	92
8.5	Shrnutí pojmů	95
9	Střední hodnota a rozptyl	97
9.1	Empirické a teoretické rozdělení pravděpodobnosti	97
9.2	Empirické charakteristiky popisu dat	104
9.3	Teoretické charakteristiky popisu dat	108
9.4	Shrnutí pojmů	112
10	Binomické rozdělení pravděpodobnosti	114
10.1	Vlastnosti binomického rozdělení	114
10.2	Základní principy statistického testu	121
10.3	Znaménkový test	124
10.4	Shrnutí pojmů	127
11	Poissonovo a exponenciální rozdělení pravděpodobnosti	129
11.1	Odvození	129
11.2	Příklady užití	133
11.3	Teorie front	137
11.3.1	Fronty typu $(M M 1) : (GD \infty \infty)$	140
11.3.2	Fronty typu $(M M 1) : (GD N \infty)$	141
11.3.3	Fronty typu $(M M c) : (GD \infty \infty)$	143
11.3.4	Fronty typu $(M M c) : (GD N \infty)$	144
11.4	Náhodné generování hodnot Po a Exp na počítači	144

11.5 Shrnutí pojmů	145
12 Rovnoměrné a normální rozdělení pravděpodobnosti	147
12.1 Rovnoměrné rozdělení pravděpodobnosti	147
12.2 Normální rozdělení pravděpodobnosti	148
12.3 U -rozdělení	151
12.4 U -test	157
12.4.1 Kvalitativní test	158
12.4.2 Jednostranný test	160
12.4.3 Oboustranný test	161
12.5 Shrnutí pojmů	163
13 Statistický test střední hodnoty normálního rozdělení při známém roz-	
ptylu	165
13.1 Teoretické rozdělení parametru empirického rozdělení	165
13.2 Teoretické rozdělení průměru \bar{X}	165
13.3 Testy o střední hodnotě průměru při známém rozptylu	168
13.3.1 Test „ $\mu = \text{konst}$ ”	168
13.3.2 Test „ $\mu_1 = \mu_2$ ”	170
13.4 Shrnutí pojmů	171

Seznam obrázků

2.1	„Vzdálenost“ bodů A, B podle metriky 2.2	14
2.2	„Vzdálenost“ dvou spojitých funkcí v metrice 2.4	14
2.3	Pevné body reálné funkce	16
2.4	Funkce, která je kontraktivní	17
2.5	Funkce, která není kontraktivní	18
4.1	Ilustrace k větě 4.1	35
4.2	K příkladu 4.1 - separace kořenů rovnice	35
4.3	Metoda půlení intervalu	36
4.4	Metoda regula falsi	38
4.5	Metoda sečen	39
4.6	Metoda sečen může divergovat.	40
4.7	Newtonova metoda	41
4.8	Newtonova metoda může divergovat	42
4.9	Metoda prosté iterace	44
4.10	Metoda prosté iterace může divergovat	44
4.11	Grafický význam řešení dvou nelineárních rovnic	48
4.12	K příkladu 4.7 - odhad polohy kořenů.	50
5.1	Funkce a interpolační polynom	53
5.2	K příkladu 5.1: Zadané body a výsledný interpolační polynom	55
5.3	K příkladu 5.2: Srovnání funkce a interpolačního polynomu	57
5.4	Nevhodná aproximace interpolačním polynomem	59
5.5	Nahrazení funkce lineárním splajnem	60
5.6	Nahrazení funkce $f(x) = \frac{1}{1+x^2}$ přirozeným kubickým splajnem.	61
5.7	K příkladu 5.5: zadané body a nalezená přímka	67
6.1	Numerické derivování - ilustrace ke vzorci 6.2	69
6.2	Numerické derivování - ilustrace ke vzorci 6.4	70
6.3	Připomenutí významu určitého integrálu	71
6.4	Obdélníková metoda	71
6.5	Lichoběžníková metoda	72
6.6	Simpsonova metoda	73
6.7	Složené lichoběžníkové pravidlo	74
8.1	K př. 8.11: Množina všech možných výsledků.	85
8.2	K př. 8.11: Množina všech příznivých výsledků.	86
8.3	K př. 8.12: Hodnoty pravděpodobnostní funkce $p(x)$	90
8.4	K př. 8.12: Graf distribuční funkce $F(x)$ diskrétního rozdělení.	91
8.5	K př. 8.13: Graf hustoty $f(x)$ spojitého rozdělení.	93
8.6	K př.8.13: Pravděpodobnost u spojitě veličiny je rovna obsahu šrafované plochy.	94
8.7	K př.8.13: Graf distribuční funkce $F(x)$ daného normálního rozdělení.	95
9.1	K příkladu 9.1: Histogram četností veličiny X	98
9.2	K př. 9.1: Histogram pravděpodobností veličiny X	99

9.3	K př. 9.1: I v diskrétním případě lze pravděpodobnost vyjádřit jako obsah jisté plochy.	99
9.4	K př. 9.1: Graf pravděpodobnostní funkce $p(x)$	100
9.5	K př. 9.1: Graf distribuční funkce $F(x)$ rozdělení veličiny X	100
9.6	K př. 9.2: Histogram pravděpodobnosti teoretického rozdělení veličiny X	102
9.7	K příkladům 9.1, 9.2: Empirický histogram pravděpodobností veličiny X , který se hodně liší od teoretického.	103
10.1	Histogram pravděpodobností binomického rozdělení pro $N = 3, p = 0.5$	116
10.2	Histogram pravděpodobností binomického rozdělení pro $N = 6, p = 0.5$	117
10.3	Histogram pravděpodobností binomického rozdělení pro $N = 10, p = 0.5$	117
10.4	Histogram pravděpodobností binomického rozdělení pro $N = 4, p = 0.1$	117
10.5	Histogram pravděpodobností binomického rozdělení pro $N = 4, p = 0.9$	118
10.6	Histogram pravděpodobností binomického rozdělení pro $N = 10, p = 0.1$	118
10.7	Histogram pravděpodobností binomického rozdělení pro $N = 40, p = 0.1$	118
10.8	Histogram pravděpodobností binomického rozdělení pro $N = 16, p = 0.5$ s hodnotami relativních četností.	121
11.1	Graf hustoty $f(t)$ rozdělení $Exp(12)$	134
11.2	Graf distribuční funkce $F(t)$ rozdělení $Exp(12)$	135
11.3	Graf pravděpodobnostní funkce p rozdělení $Po(12)$	135
11.4	Graf distribuční funkce $F(t)$ rozdělení $Po(12)$: funkce s nekonečně mnoha schody, která vyjadřuje kumulativní pravděpodobnosti $F(t) = P(Y < t)$	136
11.5	Pro kladná t je distribuční funkce $F(t)$ prostá, a proto pro $p \in (0; 1)$ existuje jediná hodnota $t_p \in (0, \infty)$ tak, že $F(t_p) = p$	145
12.1	Hustota rovnoměrného rozdělení pravděpodobnosti $Ro(0; 6)$	148
12.2	Distribuční funkce rovnoměrného rozdělení pravděpodobnosti $Ro(0; 6)$	148
12.3	Hustota normálního rozdělení pro různé střední hodnoty μ	149
12.4	Hustota normálního rozdělení pro různé rozptyly σ^2	149
12.5	Obsah šrafované plochy je roven pravděpodobnosti, že X nabude hodnot z intervalu $< 75; 77 >$	151
12.6	Obsah šrafované plochy je roven pravděpodobnosti, že U nabude hodnot z intervalu $< 0; 0.4 >$. Tento obsah je stejný jako obsah šrafované plochy z obr. 12.5.	152
12.7	Obsah šrafované plochy je roven funkční hodnotě distribuční funkce $\Phi(u)$ rozdělení U	153
12.8	Obsahy ploch A a B jsou stejné.	153
12.9	K př. 12.5a) - výpočet pravděpodobnosti u normálního rozdělení je roven obsahu šrafované plochy.	154
12.10	K př. 12.7 - aproximovaná plocha.	156
12.11	K př. 12.7 - nevhodná aproximace Bi pomocí No	157
12.12	K př. 12.7 - vhodná aproximace Bi pomocí No užitím korekce.	157
12.13	Obsah šrafované plochy je roven $P(U \geq 1.64) = \alpha = 0.05$	159
12.14	Obsah šrafované plochy je roven pravděpodobnosti β chyby druhého druhu.	160
12.15	Význam kritických hodnot oboustranného testu - obsah každého z obou šrafovaných konců je roven $\frac{\alpha}{2}$	162

13.1 Graf hustoty rozdělení průměru je pro $N = 25$ užší než pro $N = 1$	167
13.2 Výpočet pravděpodobností v příkladu 13.3.	168
13.3 Ad př. 13.4 - hustota rozdělení veličiny \bar{X} za předpokladu, že platí H_0 . . .	170

Úvod

Tato skripta jsou napsána jako doplňující text do předmětu MATEMATIKA 3 pro 2.ročník bakalářského studia FEKT. Daný předmět se skládá ze dvou odlišných oblastí matematiky - numerických metod, jejichž cílem je představit základy numerického řešení úloh praxe, a pravděpodobnosti, jejichž úkolem je seznámit studenty s pravděpodobnostními modely popisujícími konkrétní situace. Autorkou první části je Mgr. Irena Růžičková, autorem druhé RNDr. Břetislav Fajmon, PhD.

Studenti by po absolvování kursu měli být schopni numericky řešit rovnice a systémy rovnic, aproximovat hodnoty pomocí metody nejmenších čtverců a interpolačních polynomů, používat vzorce numerického derivování a numerické integrace a řešit numericky některé diferenciální rovnice. Dále v oblasti pravděpodobnostních modelů se studenti seznámí s tím, jaké situace daný model popisuje, a použijí jej v konkrétních úlohách. V rámci úvodu do statistiky se setkají s některými základními statistickými testy.

V další fázi plánujeme za účelem materiálu pro dálkové studium každou kapitolu textu rozšířit o kontrolní otázky sloužící k samostatné kontrole znalostí a o příklady určené k samostatnému řešení.

Ad numerické metody

V praxi má velký význam matematické modelování a simulace nejrůznějších procesů. Při tom je potřeba řešit různé matematické úlohy, mnoho dějů je např. popsáno diferenciálními rovnicemi. Nalezení přesného řešení takovýchto problémů bývá často náročné, někdy i úplně nemožné. Často je lepší nehledat řešení v uzavřeném tvaru, ale pomocí konečného počtu kroků určitého postupu najít řešení přibližné. K tomu právě slouží numerické metody.

I hledání přibližného řešení bývá ovšem dosti pracné a jen málo úloh lze s uspokojivou přesností vyřešit „ručně“. Proto jsou numerické metody těsně spjaty s programováním a rozkvět některých oblastí numerických metod přišel teprve s rozvojem výpočetní techniky.

V první části těchto skript se studenti mohou seznámit se základními a nejjednoduššími numerickými metodami pro řešení lineárních a nelineárních rovnic, aproximaci funkcí, numerické derivování a integrování a pro řešení diferenciálních rovnic.

Některé kapitoly by si zasloužily mnohem obsáhlejší teoretický úvod. Na ten však zde bohužel není prostor. Pokusila jsem se všude alespoň naznačit, proč uvedené metody fungují.

Snažila jsem se o srozumitelnost a současně o zachování matematické přesnosti. Zkušenější čtenáři mi snad prominou jistou nepořádnost v uvádění předpokladů.

Poznámka k řešeným příkladům

Všechny mezivýsledky v příkladech řešených v těchto skriptech jsou zapisovány po zaokrouhlení. Při dalším výpočtu však byly použity původní, přesnější hodnoty. Proto se může stát, že bude-li někdo tyto příklady přepočítávat a použije k tomu mezivýsledky zde uvedené, může dojít k výsledkům poněkud odlišným.

Irena Růžičková, Brno 2002

Ad pravděpodobnost

Co je pravda? Tuto otázku se zeptal Pilát chvíli předtím, než vydal příkaz k ukřižování Ježíše Krista (viz Bible). Nevěděl, že Ježíš o sobě prohlásil: Já jsem ta cesta, pravda i život. Pilát neměl dost trpělivosti hledat odpověď na svou otázku, a tak vydal příkaz k popravě nevinného, protože byl pro něj pohodlnější vlastní klid než spravedlnost. Možná že i dnes si nedáváme dost času k hledání odpovědi, a tak je možné, že ve svém životě křížujeme to dobré ve prospěch určitého dočasného klidu.

Jiná odpověď na naši otázku: Pravda je soubor mýtů, které se lidem ještě nepodařilo vyvrátit. Toto humorné prohlášení trochu představuje historii vědy, protože například při fyzikálním popisu skutečnosti se setkáváme s tím, že model sloužící k popisu jisté situace se v jiné situaci ukázal nevhodným, což vedlo ke hledání nových souvislostí.

Zkrátka a dobře, ve svém poznání světa máme jisté rezervy, a tak nám místo pravdy zůstává spíše pravděpodobnost - jakási míra víry, že určitá věc je skutečností. Všichni jsme odkázáni k tomu, že musíme věřit.

Druhá část textu - pravděpodobnost - je uspořádána do šesti kapitol, které mohou docela dobře sloužit jako doplnění šesti přednášek v daném tematickém rámci (první kapitola je nejobsáhlejší, doporučuji partii o podmíněné pravděpodobnosti a Bayesově vzorci z přednášky vypustit a probrat pouze na cvičení, aby se dostalo na poměrně důležitější poslední dvě kapitoly textu, kde je vyloženo normální rozdělení a statistický U -test).

Text je založen na uvádění příkladů - v průběhu příkladů jsou nově uváděny matematické pojmy. A proto příklady nelze při studiu přeskakovat - respektive kdo bude přeskakovat příklady, tomu toho ke studiu moc nezbyde.

Břetislav Fajmon, Brno 2002

Část I

NUMERICKÉ METODY

1 Chyby při numerických výpočtech

Protože základem numerických metod je získávání přibližných výsledků, je nutné mít vždy představu, jaký rozdíl může být mezi přesným řešením dané úlohy a řešením získaným použitou numerickou metodou.

1.1 Zdroje a typy chyb

Pomineme-li jako zdroj chyb člověka dopouštějícího se omylů, můžeme chyby rozdělit na několik základních druhů:

- **chyby matematického modelu** – vznikají nahrazením reálné fyzikální situace matematickým modelem. Může se jednat například o popis nějakého fyzikálního děje pomocí diferenciální rovnice.
- **chyby vstupních dat** – jsou způsobeny nepřesnostmi při měření fyzikálních veličin.
- **chyby numerické metody** – vznikají při náhradě původní matematické úlohy jednodušší úlohou numerickou. Často se jedná o náhradu nekonečného procesu procesem konečným, např. při výpočtu hodnoty některé elementární funkce pomocí součtu několika prvních členů její nekonečné Taylorovy řady nebo při aproximaci určitého integrálu součtem konečného počtu funkčních hodnot. Odhad této chyby je důležitou součástí řešení každé numerické úlohy.
- **chyby zaokrouhlovací** – vznikají tím, že při výpočtech pracujeme s čísly zaokrouhlenými na určitý, relativně nevelký počet míst. Tyto chyby se při výpočtu mohou kumulovat, nebo naopak navzájem rušit. Při velkém počtu operací je posouzení jejich vlivu velmi náročné.

1.2 Definice chyb

Je-li \hat{x} přesná hodnota nějakého čísla a x její aproximace, jejich rozdíl

$$E(x) = \hat{x} - x$$

nazýváme **absolutní chyba** aproximace. Obvykle se budeme zabývat odhadem této chyby, ale je-li přesná hodnota veličiny velmi malá nebo velmi velká, má větší význam užívat **relativní chybu**

$$RE(x) = \frac{\hat{x} - x}{x},$$

kteřá se též často vyjadřuje v procentech.

Například absolutní chyba 10^6 se může na první pohled zdát velmi velká. Je-li ovšem přesná hodnota veličiny 10^{15} , už se chyba tak závažná nejeví. Tento fakt lze nejlépe vyjádřit pomocí relativní chyby, v tomto případě je $RE = 10^{-9} = 10^{-7} \%$.

Přesnou hodnotu chyb zpravidla neznáme. Proto jsou důležité odhady chyb.

Každé nezáporné číslo $ME(x)$, pro které platí

$$|\hat{x} - x| \leq ME(x), \text{ tj. } \hat{x} \in [x - ME(x), x + ME(x)]$$

nazýváme **odhad absolutní chyby** aproximace x nebo **mezní absolutní chyba**.

Každé nezáporné číslo $MR(x)$, pro které platí

$$\frac{|\hat{x} - x|}{|x|} \leq MR(x), \quad x \neq 0$$

nazýváme **odhad relativní chyby** nebo **mezní relativní chyba**.

Často užíváme symbolických zápisů

$$\hat{x} = x \pm ME(x), \text{ resp. } \hat{x} = x(1 \pm MR(x)).$$

1.3 Zaokrouhlování. Šíření chyb při výpočtu

Je-li x reálné číslo, které má obecně nekonečné dekadické vyjádření, pak číslo $x^{(d)}$, které má d desetinných míst, je **správně zaokrouhlenou hodnotou** čísla x , platí-li

$$|x - x^{(d)}| \leq \frac{1}{2} 10^{-d} \quad (1.1)$$

(Tedy např. má-li být $x^{(1)}$ správně zaokrouhlená hodnota čísla x na jedno desetinné místo, nesmí se od x lišit o více než o $\frac{1}{2} 10^{-1} = 0,05$.)

Jestliže číslo x , které chceme zaokrouhlit na d desetinných míst, má právě $d+1$ desetinných míst, z nichž poslední je pětka, často se používá pravidlo (čtenáři snad známé ze základní školy), že pětka po liché číslici se zaokrouhluje nahoru, po sudé dolů. Lze ale také (a některé počítačové programy tak činí) volit vždy zaokrouhlení nahoru nebo vždy zaokrouhlení dolů.

Při numerických výpočtech pracujeme se zaokrouhlenými čísly. Výsledky početních operací s těmito čísly jsou opět zaokrouhlovány a dále se s nimi pracuje. Tím se zaokrouhlovací chyby šíří. Budeme se nyní zabývat tím, co se děje při základních aritmetických operacích. Nechť x a y jsou aproximace čísel \hat{x} a \hat{y} .

Pro **chybu součtu a rozdílu** platí

$$\begin{aligned} |E(x \pm y)| &= |(\hat{x} \pm \hat{y}) - (x \pm y)| = |(\hat{x} - x) \pm (\hat{y} - y)| = \\ &= |E(x) \pm E(y)| \leq |E(x)| + |E(y)| \leq ME(x) + ME(y) \end{aligned} \quad (1.2)$$

Odhad chyby součinu a rozdílu je o něco pracnější. Pro **chybu součinu** platí

$$\begin{aligned} |E(x \cdot y)| &= |\hat{x}\hat{y} - xy| = |E(x) \cdot y + E(y) \cdot x + E(x) \cdot E(y)| \leq \\ &\leq |y| \cdot ME(x) + |x| \cdot ME(y) + ME(x) \cdot ME(y) \end{aligned} \quad (1.3)$$

Protože součin $ME(x) \cdot ME(y)$ bývá vzhledem k ostatním sčítancům zanedbatelný, dostáváme pro **relativní chybu součinu**

$$|RE(xy)| \approx \left| \frac{E(x) \cdot y + E(y) \cdot x}{xy} \right| \leq MR(x) + MR(y) \quad (1.4)$$

Podobně pro **chybu podílu** platí

$$\left| E\left(\frac{x}{y}\right) \right| = \left| \frac{x + E(x)}{y + E(y)} - \frac{x}{y} \right| = \left| \frac{E(x) \cdot y - x \cdot E(y)}{y(y + E(y))} \right| \leq \frac{|y|ME(x) + |x|ME(y)}{|y|(|y| - ME(y))} \quad (1.5)$$

a je-li $ME(y)$ zanedbatelné vzhledem k y , pak pro **relativní chybu podílu** dostaneme

$$\left| R\left(\frac{x}{y}\right) \right| \leq MR(x) + MR(y)$$

Nyní se ještě zmíníme obecně o **chybě při výpočtu funkční hodnoty**. Máme stanovit, jaké chyby se dopustíme při výpočtu hodnoty funkce $f(x_1, x_2, \dots, x_n)$ v bodě $[\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n]$, jestliže přesné hodnoty \hat{x}_i nahradíme přibližnými hodnotami x_i . Chybu i -té proměnné označíme E_i . Platí

$$f(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) = f(x_1, x_2, \dots, x_n) + \sum_{i=1}^n E_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \left(\sum_{i=1}^n E_i \frac{\partial}{\partial x_i} \right)^2 f + \dots$$

kde parciální derivace se berou v bodě $[x_1, x_2, \dots, x_n]$. Protože obvykle budeme moci předpokládat, že členy obsahující součiny chyb jsou malé ve srovnání s ostatními členy na pravé straně, můžeme psát

$$f(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) - f(x_1, x_2, \dots, x_n) \approx \sum_{i=1}^n E_i \frac{\partial f}{\partial x_i} \quad (1.6)$$

Všimněme si, že 1.2, 1.3 a 1.5 jsou speciálními případy tohoto vzorce.

Zde je na místě zmínit se o tom, že při odečítání dvou sobě blízkých čísel se může velmi zvětšit relativní chyba. Pokud pak takto získaný výsledek použijeme dále jako dělitele, může dojít i k podstatnému zvětšení absolutní chyby. Tento jev ukážeme na příkladech.

Příklad 1.1 *Nechť $x = 2,78493$ a $y = 2,78469$ jsou aproximace čísel \hat{x} a \hat{y} získané zaokrouhlením těchto čísel na pět desetinných míst. Určete odhady absolutní a relativní chyby rozdílu $x - y$.*

Řešení: Mezní absolutní chyby x a y jsou podle 1.1 $ME(x) = ME(y) = \frac{1}{2} 10^{-5}$. Tedy podle 1.2 $|E(x - y)| \leq 10^{-5} = ME(x - y)$.

Mezní relativní chyba x je $MR(x) = \frac{\frac{1}{2} 10^{-5}}{2,78493} \doteq 1,8 \cdot 10^{-6}$ ($MR(y)$ vyjde skoro stejně), zatímco pro rozdíl může být relativní chyba řádově vyšší, její odhad je roven $\frac{ME(x-y)}{x-y} = \frac{10^{-5}}{0,00024} \doteq 4,2 \cdot 10^{-2}$.

Příklad 1.2 *Nechť $z = 1,23456$ je aproximace čísla \hat{z} získaná zaokrouhlením tohoto čísla na pět desetinných míst. Určete odhad chyby podílu $\frac{z}{x-y}$, kde x a y jsou čísla z příkladu 1.1*

Řešení: Z příkladu 1.1 známe odhad chyby jmenovatele. Dále víme, že $ME(z) = \frac{1}{2} 10^{-5}$. Pro odhad chyby podílu stačí dosadit do 1.5:

$$\begin{aligned} \left| E \left(\frac{z}{x-y} \right) \right| &\leq \frac{|x-y| \cdot ME(z) + |z| \cdot ME(x-y)}{|x-y|(|x-y| - ME(x-y))} = \\ &= \frac{0,00024 \cdot \frac{1}{2} \cdot 10^{-5} + 1,23456 \cdot 10^{-5}}{0,00024 \cdot (0,00024 - 10^{-5})} \doteq 2,2 \cdot 10^2 \end{aligned}$$

Tedy zatímco vstupní hodnoty x, y a z měly chybu řádově v stotisícinách, výsledek může mít chybu řádově ve stovkách!

Proto, je-li to možné, je žádoucí se odečítání blízkých čísel vyvarovat.

1.4 Podmíněnost numerických úloh a numerická stabilita algoritmů

Při numerickém řešení různých úloh musíme zkoumat, jaký vliv na výsledek mají malé změny ve vstupních hodnotách a zaokrouhlování během výpočtu.

Řešení numerických úloh můžeme považovat za postup, kterým přiřazujeme vstupním údajům výstupní data. Je-li toto přiřazení spojitě zobrazení, pak říkáme, že numerická úloha je **korektní úloha**, v opačném případě se jedná o úlohu nekorektní.

Pro tyto úlohy má zásadní význam relativní citlivost výsledku na malé změny ve vstupních parametrech úlohy. **Korektní úloha je dobře podmíněná**, jestliže malým relativním změnám vstupních údajů odpovídají malé relativní změny výstupních údajů. Číslo

$$C_p = \frac{\text{relativní chyba výstupních údajů}}{\text{relativní chyba vstupních údajů}}$$

nazýváme **číslo podmíněnosti úlohy**. Pro dobře podmíněné úlohy je číslo C_p blízké číslu 1.

Pokud malé relativní změny na vstupu způsobí velké relativní změny na výstupu, pak mluvíme o **špatně podmíněné úloze**. Řešení špatně podmíněných úloh je nejlépe se vyhnout, protože výsledky jakéhokoli algoritmu jsou velmi nespolehlivé.

Podobně řekneme, že je **algoritmus dobře podmíněný**, je-li málo citlivý na poruchy ve vstupních datech. Kromě nepřesností ve vstupních údajích ovlivňuje výsledek použitého algoritmu i zaokrouhlování čísel během výpočtu. Je-li vliv zaokrouhlovacích chyb na výsledek malý, mluvíme o **numericky stabilním algoritmu**. Algoritmus dobře podmíněný a numericky stabilní se nazývá **stabilní**.

2 Exkurze do funkcionální analýzy

Tato kapitola tvoří teoretický základ pro metody probírané v dalších dvou kapitolách. Protože prostor, který lze této problematice věnovat, je velmi omezený, pokusíme se zde vysvětlit jen nejnnutnější pojmy. Pokud by někoho odrazovala přílišná teoretičnost a „vědeckost“ této kapitoly a spokojil by se s tím, že metody popsané v kapitolách 3 a 4 fungují, aniž by se zajímal o to, proč fungují, mohl by snad následující text přeskočit.

2.1 Metrický prostor

Studenti určitě umí vypočítat vzdálenost dvou reálných čísel na číselné ose nebo vzdálenost dvou bodů v rovině či v prostoru. Podobně se dá určovat „vzdálenost“ různých jiných objektů. Této zobecněné vzdálenosti se říká metrika.

Definice. Buď X množina (prvků jakéhokoli typu). Řekneme, že na této množině je definována **metrika** d , jestliže každým dvěma prvkům $x, y \in X$ je přiřazeno reálné číslo $d(x, y)$ tak, že

- 1) $d(x, y) \geq 0 \quad \forall x, y \in X \quad , \quad d(x, y) = 0 \Leftrightarrow x = y$
- 2) $d(x, y) = d(y, x) \quad \forall x, y \in X$
- 3) $d(x, z) \leq d(x, y) + d(y, z) \quad \forall x, y, z \in X$ (trojúhelníková nerovnost)

Množinu X s metrikou d pak nazýváme **metrický prostor**.

Příklady metrických prostorů

Asi nejjednodušším příkladem metrického prostoru je množina všech reálných čísel \mathbb{R} s metrikou d definovanou jako $d(x, y) = |x - y|$.

Jako množinu X však nemusíme brát celé \mathbb{R} , může to být i jakákoli jeho podmnožina, např. interval nebo množina všech racionálních čísel \mathbb{Q} .

Jiným příkladem je množina všech uspořádaných n -tic reálných čísel. Je-li $\mathbf{x} = (x_1, x_2, \dots, x_n)$ a $\mathbf{y} = (y_1, y_2, \dots, y_n)$, metriku d můžeme definovat různě. Jako nejprirozenější se jeví obvyklá vzdálenost dvou bodů:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}, \quad (2.1)$$

existují však i jiné možnosti, např.

$$d(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n| \quad (2.2)$$

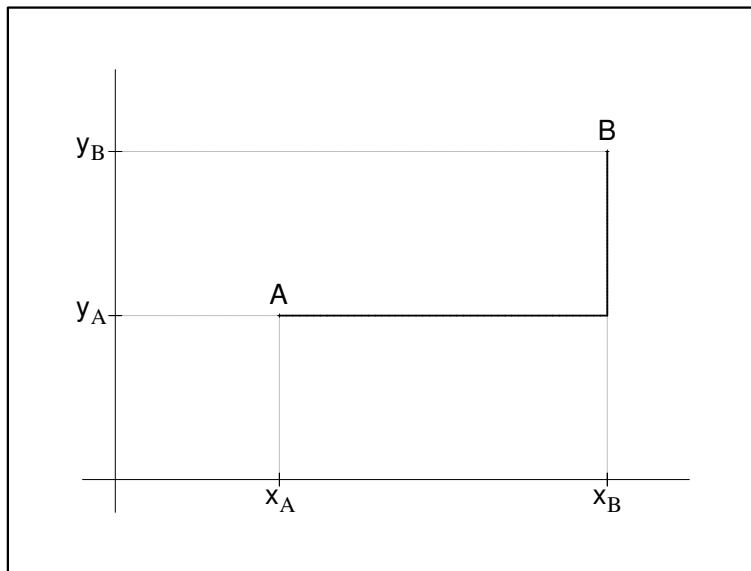
nebo

$$d(\mathbf{x}, \mathbf{y}) = \max(|x_1 - y_1|, |x_2 - y_2|, \dots, |x_n - y_n|). \quad (2.3)$$

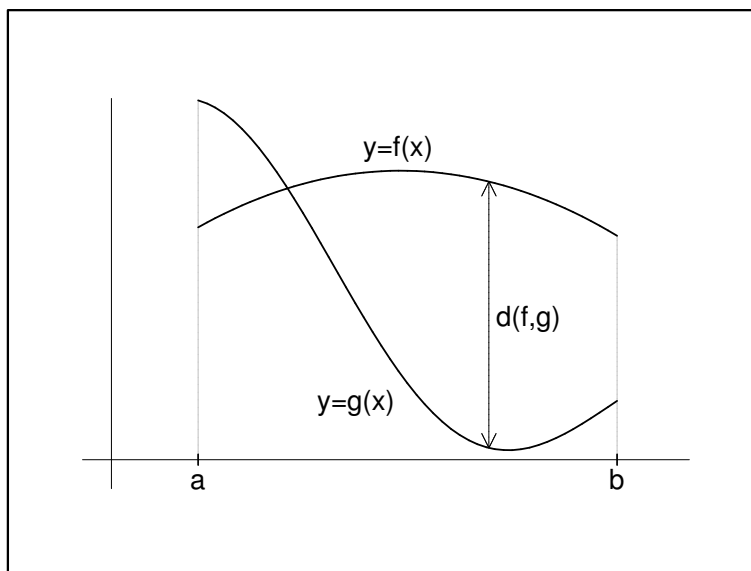
Jako poslední příklad uvedeme množinu všech funkcí definovaných a spojitých na intervalu $\langle a, b \rangle$, která se označuje jako $C(\langle a, b \rangle)$. Jsou-li $f, g \in C(\langle a, b \rangle)$, definujeme

$$d(f, g) = \max_{x \in \langle a, b \rangle} |f(x) - g(x)|. \quad (2.4)$$

Obrázky 2.1 a 2.2 poslouží k objasnění některých uvedených metrik.



Obrázek 2.1: „Vzdálenost” bodů A, B podle metriky 2.2 je délka silné černé čáry.



Obrázek 2.2: „Vzdálenost” dvou spojitých funkcí v metrice 2.4

2.2 Úplný metrický prostor

Již na střední škole se studenti seznámili s posloupnostmi reálných čísel a (snad) i s jejich limitami.

Připomeňme, že limita posloupnosti reálných čísel $(a_n)_{n=1}^{\infty}$ je, populárně řečeno, takové číslo a , ke kterému se členy posloupnosti pro n jdoucí do nekonečna přibližují. Přesněji:

Reálné číslo a se nazývá limitou posloupnosti $(a_n)_{n=1}^{\infty}$, jestliže ke každému $\varepsilon > 0$ existuje přirozené číslo N tak, že pro všechna $n > N$ platí $|a_n - a| < \varepsilon$.

Neboli: ať zvolíme ε libovolně malé, od jistého indexu N se členy posloupnosti budou od a lišit méně než o ε .

Posloupnosti však můžeme sestavovat i z jiných objektů než z reálných čísel. Stejně tak můžeme u takových posloupností říci, zda mají, nebo nemají limitu. Pro posloupnosti sestavené z prvků obecného metrického prostoru se limita definuje velmi podobně, jen je třeba zobecnit ono „lišení se o méně než ε “. To se provede pomocí metriky.

Definice. Buď X metrický prostor s metrikou d a $(x_n)_{n=1}^{\infty}$ posloupnost prvků z X . Řekneme, že $x \in X$ je **limitou** této posloupnosti, píšeme $\lim_{n \rightarrow \infty} x_n = x$, jestliže ke každému $\varepsilon > 0$ existuje přirozené číslo N tak, že pro všechna $n > N$ platí $d(x_n, x) < \varepsilon$.

Posloupnost, která má limitu, se nazývá **konvergentní**.

Nyní definujeme další vlastnost posloupností.

Definice. Buď X metrický prostor s metrikou d a $(x_n)_{n=1}^{\infty}$ posloupnost prvků z X . Řekneme, že tato posloupnost je **cauchyovská**, jestliže ke každému $\varepsilon > 0$ existuje přirozené číslo N tak, že pro všechna $n > N$ a každé přirozené číslo k platí $d(x_n, x_{n+k}) < \varepsilon$.

Dá se říci, že cauchyovská posloupnost je taková, jejíž členy se výše popsaným způsobem zahušťují.

Dá se dokázat, že každá konvergentní posloupnost je cauchyovská. Intuitivně by se mohlo zdát, že to musí být i naopak. Existují ale prostory, v nichž najdeme cauchyovské posloupnosti, které v daném prostoru limitu nemají. Ukážeme to na následujícím příkladu: Mějme například množinu všech reálných čísel a v něm posloupnost $a_1 = 3.1, a_2 = 3.14, a_3 = 3.141, a_4 = 3.1415, \dots$. Tato posloupnost má limitu π a tedy je cauchyovská.

Nyní vezměme tutéž posloupnost, ale v množině všech racionálních čísel \mathbb{Q} . Je to posloupnost cauchyovská, ale limitu v \mathbb{Q} nemá (protože $\pi \notin \mathbb{Q}$).

Existují tedy prostory, v nichž „něco schází“, neobsahují limity některých posloupností, které se jinak chovají tak, jako by limitu mít měly. Tím se dostáváme k definici úplného prostoru.

Definice. Metrický prostor se nazývá **úplný**, jestliže každá cauchyovská posloupnost v něm má limitu.

Příklady úplných a neúplných prostorů

Množina \mathbb{R} s metrikou $d(x, y) = |x - y|$ je úplný metrický prostor.

Jakýkoli uzavřený interval $\langle a, b \rangle$ s toutéž metrikou je také úplný prostor.

Otevřený interval s toutéž metrikou není úplný. To můžeme ukázat na příkladu intervalu $(0, 1)$ a posloupnosti $x_n = \frac{1}{n}$. Tato posloupnost je cauchyovská a přitom v intervalu $(0, 1)$ nemá limitu ($0 \notin (0, 1)$).

Dá se dokázat, že prostor všech uspořádaných n -tic s kteroukoli z metrik 2.1, 2.2, 2.3 je úplný.

2.3 Pevný bod zobrazení, iterační proces

Definice. Řekneme, že F je zobrazení množiny X do množiny Y , píšeme $F : X \rightarrow Y$, jestliže každému prvku $x \in X$ je pomocí F přiřazen právě jeden prvek $y \in Y$, $y = F(x)$.

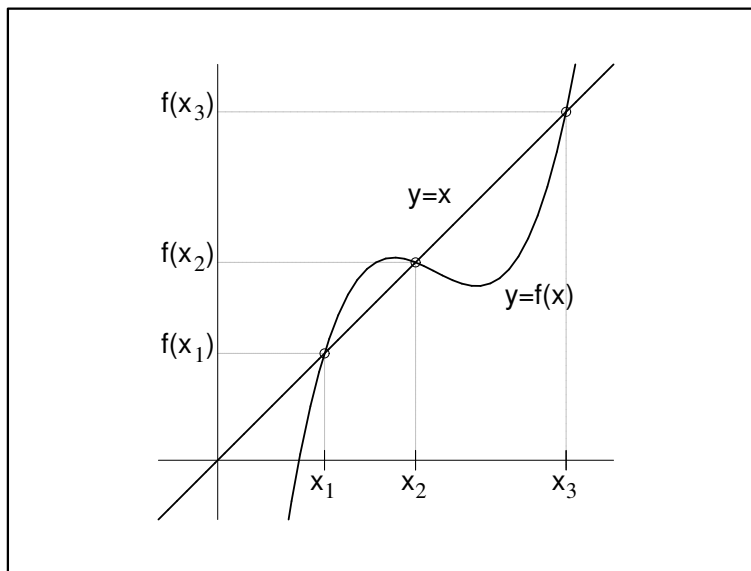
Budeme se zabývat hlavně zobrazeními množiny do sebe sama, tj. zobrazení $F : X \rightarrow X$. Takové zobrazení přiřazuje každému prvku $x \in X$ opět (obecně jiný) prvek z X . Nás bude zajímat, jestli existuje takový prvek x , který se zobrazí sám na sebe, případně jak takový prvek najít.

Definice. Prvek $x \in X$ se nazývá **pevný bod** zobrazení $F : X \rightarrow X$, jestliže platí

$$F(x) = x.$$

Jestliže za množinu X vezmeme \mathbb{R} , pak zobrazení $F : \mathbb{R} \rightarrow \mathbb{R}$ je obyčejná funkce jedné proměnné. Na obrázku 2.3 jsou vyznačeny pevné body jisté funkce f . Jsou to body, v nichž se protne graf funkce f s přímkou $y = x$.

Příklad. Funkce $f(x) = x^2$ má právě dva pevné body, a to $x = 0$ a $x = 1$, protože $0^2 = 0$ a $1^2 = 1$.



Obrázek 2.3: Pevné body reálné funkce

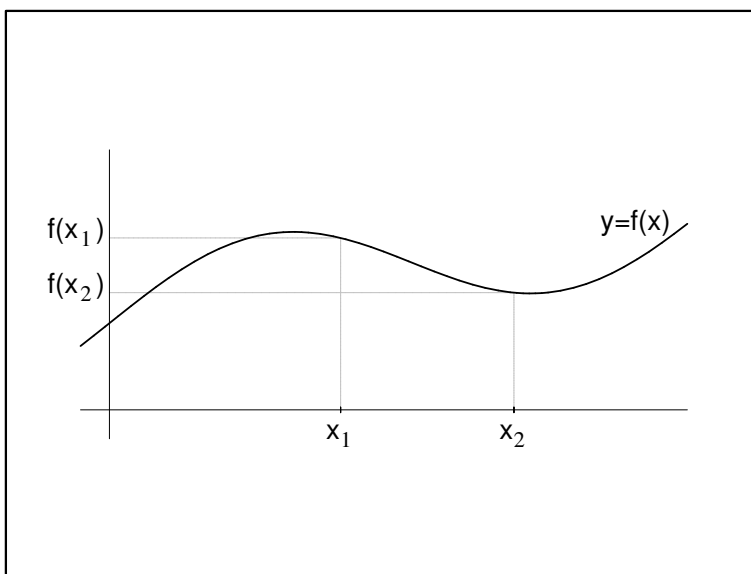
Hledání pevného bodu zobrazení má v numerické matematice velký význam. Některé úlohy, jejichž zadání zpočátku vypadá úplně jinak, lze převést právě na problém nalezení pevného bodu. Proto se nyní budeme zabývat otázkou, jak ověřit, že nějaké zobrazení pevný bod má a jak jej najít. Dá se dokázat, že jistý druh zobrazení má pevný bod vždy a existuje postup, který nás k němu dovede.

Definice. Bud' X metrický prostor. Řekneme, že zobrazení $F : X \rightarrow X$ je **kontraktivní (kontrakce)**, jestliže existuje $\alpha \in (0, 1)$ tak, že pro každé dva prvky $x, y \in X$ platí

$$d(F(x), F(y)) \leq \alpha d(x, y) \quad (2.5)$$

Číslo α nazýváme koeficient kontrakce.

„Kontrakce“ česky znamená „stažení“. Dá se tedy, byť poněkud nepřesně říct, že kontraktivní zobrazení je takové, u něž jsou si obrazy (funkční hodnoty) bližší než byly vzory. Na obrázcích 2.4 a 2.5 jsou příklady kontraktivní a nektraktivní funkce.



Obrázek 2.4: Funkce, která je kontraktivní

Věta 2.1 Bud' X úplný metrický prostor a $F : X \rightarrow X$ kontraktivní zobrazení. Pak existuje právě jeden pevný bod tohoto zobrazení \hat{x} , pro něžž platí

$$\hat{x} = \lim_{n \rightarrow \infty} x_n, \quad (2.6)$$

kde $(x_n)_{n=1}^{\infty}$ je tzv. posloupnost postupných aproximací, která je definována takto: x_0 je libovolný prvek z X a další členy posloupnosti jsou definovány předpisem

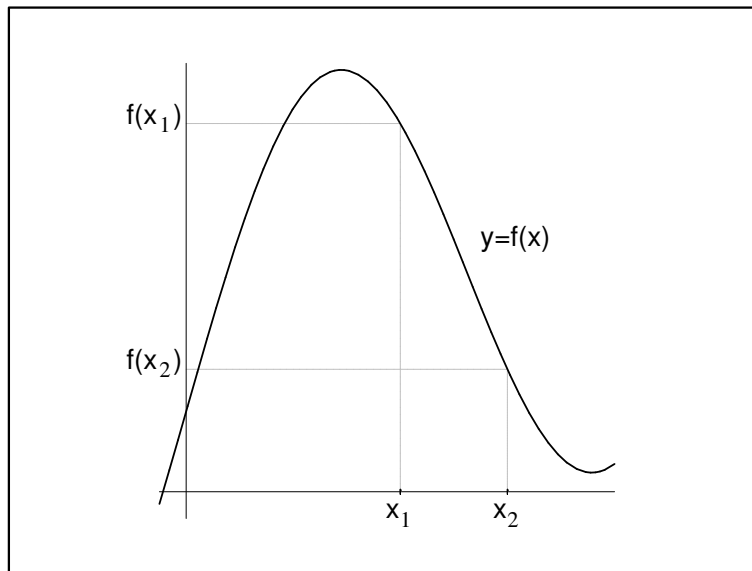
$$x_{k+1} = F(x_k), \quad k = 0, 1, \dots \quad (2.7)$$

Dále pro všechna přirozená čísla n platí:

$$d(\hat{x}, x_n) \leq \frac{\alpha}{1 - \alpha} d(x_n, x_{n-1}) \quad (2.8)$$

$$d(\hat{x}, x_n) \leq \frac{\alpha^n}{1 - \alpha} d(x_0, x_1), \quad (2.9)$$

kde α je koeficient kontrakce.



Obrázek 2.5: Funkce, která není kontraktivní

Tato věta nám dává návod, **jak pevný bod** zadaného zobrazení alespoň přibližně **najít**. Zvolíme $x_0 \in X$. Tomuto bodu se říká **počáteční aproximace**.

Pak počítáme další členy posloupnosti podle předpisu 2.7. Tomuto výpočtu se říká **iterační proces**, k-tý člen posloupnosti, x_k , se nazývá **k-tá aproximace**.

Protože podle 2.6 je pevný bod limitou posloupnosti $(x_n)_{n=1}^{\infty}$, postupné aproximace se k němu budou přibližovat. Kdybychom v iteračním procesu mohli pokračovat donekonečna, dostali bychom se nakonec k pevnému bodu. To ale není možné, a proto se v určitý moment zastavíme a řekneme, že pevný bod \hat{x} je přibližně roven poslednímu vypočtenému členu posloupnosti.

Kdy iterační proces zastavit, rozhodneme podle toho, s jakou přesností chceme mít pevný bod vypočtený. Můžeme k tomu použít např. odhad 2.8, který říká, jak je n-tá aproximace nanejvýš vzdálena od pevného bodu. K tomu ovšem musíme znát hodnotu koeficientu kontrakce α , která může být u některých úloh velmi obtížně zjistitelná. Proto se častěji používají empirická kritéria, jež pro konkrétní úlohy později popíšeme.

2.4 Normovaný vektorový prostor

V prvním semestru se studenti seznámili s vektorovými prostory.

Prvky vektorových prostorů mohou být objekty nejrůznějšího typu. Nemusí to být pouze „vektory“ v tom smyslu, jaký si člověk obvykle pod tímto pojmem představí (tj. uspořádané n-tice reálných čísel).

Nejjednodušším příkladem vektorového prostoru je množina všech reálných čísel \mathbb{R} s obvyklými operacemi $+$ a \cdot .

Vektorovým prostorem je i množina všech matic typu (m, n) s operacemi $+$ (sčítání matic)

$a \cdot$ (násobení matice konstantou).

Vektorový prostor může být tvořen též funkcemi jedné nebo více proměnných s určitou vlastností. V některých oblastech matematiky se často setkáváme např. s prostorem všech funkcí spojitých na daném intervalu $\langle a, b \rangle$, či s prostorem všech funkcí na intervalu $\langle a, b \rangle$ integrovatelných.

Studenti jistě vědí, co je absolutní hodnota čísla nebo délka vektoru. Tyto veličiny udávají velikost daného čísla, resp. vektoru bez ohledu na jeho znaménko, resp. směr. „Velikost“ lze různým způsobem určovat i u jiných objektů. Jakési zobecnění velikosti, které zachovává její přirozené vlastnosti, se nazývá norma.

Definice. Bud' V vektorový prostor. Řekneme, že na tomto prostoru je definována **norma**, jestliže každému prvku $v \in V$ je přiřazeno reálné číslo $\|v\|$ (norma v) tak, že

- 1) $\|v\| \geq 0 \quad \forall v \in V \quad , \quad \|v\| = 0 \Leftrightarrow v = 0$
- 2) $\|k \cdot v\| = |k| \cdot \|v\| \quad \forall v \in V, \forall k \in \mathbb{R}$
- 3) $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\| \quad \forall v_1, v_2 \in V$ (trojúhelníková nerovnost)

Prostor V pak nazýváme **normovaný vektorový prostor**.

Je známo, že absolutní hodnota rozdílu dvou reálných čísel udává vzdálenost těchto čísel na číselné ose. Podobně si lze normu rozdílu dvou prvků vektorového prostoru $\|u - v\|$ představit jako vzdálenost těchto dvou prvků. To znamená, že na vektorovém prostoru můžeme definovat metriku předpisem

$$d(v_1, v_2) = \|v_1 - v_2\|. \quad (2.10)$$

Příklady normovaných vektorových prostorů:

Na množině všech reálných čísel \mathbb{R} lze zavést normu jako $\|x\| = |x|$, $\forall x \in \mathbb{R}$.

Na vektorovém prostoru V_n můžeme zavést normu různým způsobem.

Je-li $\mathbf{v} = (v_1, v_2, \dots, v_n) \in V_n$, pak jeho norma může být např. definována jako délka tohoto vektoru

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}. \quad (2.11)$$

Tato norma se často značí jako $\|\mathbf{v}\|_2$ a nazývá se eukleidovská norma. Existují však i jiné možnosti. V dalším textu se setkáme s normami

$$\|\mathbf{v}\|_1 = |v_1| + |v_2| + \dots + |v_n| \quad (2.12)$$

$$\|\mathbf{v}\|_\infty = \max(|v_1|, |v_2|, \dots, |v_n|) \quad (2.13)$$

U matic lze normu počítat podobně jako u vektorů. V kapitole 3 budeme pracovat s následujícími normami (\mathbf{A} je matice typu (m, n) s prvky a_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$):

$$\|\mathbf{A}\|_\infty = \max_{i=1, \dots, m} \sum_{j=1}^n |a_{ij}| \quad \text{řádková norma} \quad (2.14)$$

$$\|\mathbf{A}\|_1 = \max_{j=1, \dots, n} \sum_{i=1}^m |a_{ij}| \quad \text{sloupcová norma} \quad (2.15)$$

Příklad 2.1 *Vypočtěte řádkovou a sloupcovou normu matice*

$$\mathbf{A} = \begin{pmatrix} -3 & 2 & 5 \\ 1 & -4 & -2 \\ 3 & -1 & 4 \end{pmatrix}$$

Řešení: Řádková norma matice je maximum ze součtů absolutních hodnot prvků v jednotlivých řádcích.

Součet absolutních hodnot prvků v prvním řádku matice je $|-3|+|2|+|5| = 10$, ve druhém řádku je součet roven 7 a ve třetím 8. Největší z těchto čísel je 10 a proto $\|\mathbf{A}\|_{\infty} = 10$. Sloupcová norma je maximum ze součtů absolutních hodnot prvků v jednotlivých sloupcích. Tedy $\|\mathbf{A}\|_1 = \max(7, 7, 11) = 11$.

Čtenář si možná povšiml značné podobnosti norem 2.11, 2.12 a 2.13 s metrikami uvedenými v kapitole 2.1. Skutečně, všechny tyto metriky můžeme dostat z výše uvedených norem pomocí 2.10.

Nabízí se otázka, proč jsme označili řádkovou normu matice 2.14 stejně jako normu vektoru 2.13 a sloupcovou normu matice 2.15 stejně jako normu vektoru 2.12.

Tyto normy skutečně mají mnoho společného. Představíme-li si vektor \mathbf{v} dimenze n jako **sloupec**, můžeme jej považovat za matici o n řádcích a jediném sloupci. Vypočteme-li nyní řádkovou normu této matice, dostaneme právě normu vektoru 2.13, vypočteme-li sloupcovou normu matice, dostaneme normu vektoru 2.12.

Dále platí, a to je pro další úvahy podstatnější, že

$$\begin{aligned} \|\mathbf{A}\mathbf{v}\|_{\infty} &\leq \|\mathbf{A}\|_{\infty} \cdot \|\mathbf{v}\|_{\infty} \\ \|\mathbf{A}\mathbf{v}\|_1 &\leq \|\mathbf{A}\|_1 \cdot \|\mathbf{v}\|_1 \end{aligned}$$

Můžeme říct, že řádková norma matice je **přidružená** vektorové normě 2.13 a sloupcová norma matice je přidružená vektorové normě 2.12.

(Obecně se maticová norma přidružená vektorové normě definuje docela složitě, o tom zde mluvit nebudeme. Např. maticová norma přidružená eukleidovské normě vektoru se počítá zcela odlišně.)

3 Numerické řešení soustavy lineárních rovnic

Řešení soustav lineárních rovnic patří mezi nejdůležitější části numerické matematiky. Mnoho praktických úloh nakonec vede k řešení takovýchto soustav, často velmi rozsáhlých. K obrovským soustavám rovnic dospějeme např. při hledání rozložení nějaké fyzikální veličiny v určitém tělese. Problém se, velmi zhruba řečeno, může řešit tak, že hledáme hodnoty této veličiny pouze v konečném počtu bodů (a čím více těchto bodů bude, tím lépe), a to právě jako řešení soustavy lineárních rovnic.

Budeme se zabývat řešením soustavy n lineárních rovnic

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \tag{3.1}$$

s neznámými x_1, x_2, \dots, x_n .

Připomeňme, že matice $\mathbf{A} = (a_{ij}), i, j = 1, \dots, n$, se nazývá **matice soustavy** a sloupcový vektor $\mathbf{b} = (b_1, \dots, b_n)^T$ **vektor pravých stran**.

Soustavu můžeme zapsat maticově ve tvaru

$$\mathbf{A} \mathbf{x} = \mathbf{b} \tag{3.2}$$

Všude v dalším textu budeme předpokládat, že matice soustavy je regulární, tj. že řešená soustava má právě jedno řešení. (V technických úlohách, kde se problém řešení soustavy lineárních rovnic může vyskytnout, to tak zpravidla bývá.)

V prvním semestru se studenti seznámili s Gaussovou eliminační metodou a s Cramerovým pravidlem. Obě tyto metody patří mezi tzv. **metody přímé**. Druhou skupinou metod řešení soustav lineárních rovnic jsou **metody iterační**.

3.1 Přímé metody

Přímé metody vedou k řešení soustavy po konečném počtu kroků. Takto nalezené řešení by bylo přesné, kdybychom se v průběhu výpočtu nedopouštěli zaokrouhlovacích chyb.

Připomeneme metody, které by studenti měli znát z prvního semestru a uvedeme některé další.

3.1.1 Cramerovo pravidlo

Je-li matice soustavy 3.2 regulární, tj. její determinant je nenulový, pak řešení soustavy lze vypočítat jako

$$x_1 = \frac{D_1}{D}, \quad x_2 = \frac{D_2}{D}, \quad \dots, \quad x_n = \frac{D_n}{D}$$

kde D je determinant matice soustavy \mathbf{A} a $D_k, k = 1, \dots, n$ jsou determinanty matic, které vzniknou z matice \mathbf{A} nahrazením k -tého sloupce této matice vektorem pravých stran \mathbf{b} .

Cramerovo pravidlo je vhodné pouze pro velmi malé soustavy rovnic, např. pro soustavu dvou rovnic s „ošklivými“ koeficienty. Pro větší soustavy by bylo nutné počítat mnoho determinantů vysokého řádu, což je velmi pracné. Proto se pro řešení velkých soustav rovnic tato metoda nepoužívá.

3.1.2 Gaussova eliminační metoda

Základem této metody je úprava soustavy na trojúhelníkový tvar pomocí elementárních úprav.

Přidáme-li v soustavě 3.2 vektor pravých stran \mathbf{b} jako $(n+1)$ -ní sloupec k matici \mathbf{A} , můžeme soustavu přepsat ve tvaru

$$\begin{array}{cccccc} a_{11} x_1 & + & a_{12} x_2 & + & \cdots & + & a_{1n} x_n & = & a_{1n+1} \\ a_{21} x_1 & + & a_{22} x_2 & + & \cdots & + & a_{2n} x_n & = & a_{2n+1} \\ \vdots & & & & & & & & \vdots \\ a_{n1} x_1 & + & a_{n2} x_2 & + & \cdots & + & a_{nn} x_n & = & a_{nn+1} \end{array}$$

Nyní se pomocí přičítání vhodných násobků první rovnice budeme snažit z ostatních rovnic eliminovat x_1 . (Je-li $a_{11} = 0$, vyměníme první rovnici s první takovou rovnicí, která na prvním místě nulu nemá.)

Odečteme-li postupně první rovnici, vynásobenou číslem $\frac{a_{i1}}{a_{11}}$, od i -té rovnice, pro $i = 2, 3, \dots, n$, dostaneme

$$\begin{array}{cccccc} a_{11} x_1 & + & a_{12} x_2 & + & \cdots & + & a_{1n} x_n & = & a_{1n+1} \\ & & a_{22}^{(1)} x_2 & + & \cdots & + & a_{2n}^{(1)} x_n & = & a_{2n+1}^{(1)} \\ & & \vdots & & & & \vdots & & \\ & & a_{n2}^{(1)} x_2 & + & \cdots & + & a_{nn}^{(1)} x_n & = & a_{nn+1}^{(1)} \end{array}$$

Nové koeficienty jsou vypočteny jako $a_{ij}^{(1)} = a_{ij} - \frac{a_{i1}}{a_{11}} a_{1j}$, $i = 2, 3, \dots, n, j = 2, 3, \dots, n+1$. Nyní budeme pomocí vhodných násobků druhé rovnice eliminovat x_2 v třetí, čtvrté, ... n -té rovnici. (Opět, je-li $a_{22}^{(1)} = 0$, vyměníme druhou rovnici s první z dalších rovnic, ve které u x_2 nula není.)

Tím dostaneme

$$\begin{array}{cccccc} a_{11} x_1 & + & a_{12} x_2 & + & a_{13} x_3 & + & \cdots & + & a_{1n} x_n & = & a_{1n+1} \\ & & a_{22}^{(1)} x_2 & + & a_{23}^{(1)} x_3 & + & \cdots & + & a_{2n}^{(1)} x_n & = & a_{2n+1}^{(1)} \\ & & & & a_{33}^{(2)} x_3 & + & \cdots & + & a_{3n}^{(2)} x_n & = & a_{3n+1}^{(2)} \\ & & & & \vdots & & & & \vdots & & \\ & & & & a_{n3}^{(2)} x_3 & + & \cdots & + & a_{nn}^{(2)} x_n & = & a_{nn+1}^{(2)} \end{array}$$

kde $a_{ij}^{(2)} = a_{ij}^{(1)} - \frac{a_{i2}^{(1)}}{a_{22}^{(1)}} a_{2j}^{(1)}$, $i = 3, 4, \dots, n, j = 3, 4, \dots, n+1$.

Pokračujeme-li dále stejným způsobem, dostaneme po $n-1$ krocích soustavu v trojúhelníkovém tvaru

$$\begin{aligned}
 a_{11} x_1 + a_{12} x_2 + a_{13} x_3 + \cdots + a_{1n} x_n &= a_{1n+1} \\
 a_{22}^{(1)} x_2 + a_{23}^{(1)} x_3 + \cdots + a_{2n}^{(1)} x_n &= a_{2n+1}^{(1)} \\
 a_{33}^{(2)} x_3 + \cdots + a_{3n}^{(2)} x_n &= a_{3n+1}^{(2)} \\
 &\vdots \\
 a_{nn}^{(n-1)} x_n &= a_{nn+1}^{(n-1)}
 \end{aligned} \tag{3.3}$$

Z této soustavy snadno určíme hledané řešení:

$$\begin{aligned}
 x_n &= \frac{a_{nn+1}^{(n-1)}}{a_{nn}^{(n-1)}} \\
 x_{n-1} &= \frac{1}{a_{n-1n-1}^{(n-2)}} \left(a_{n-1n+1}^{(n-2)} - a_{n-1n}^{(n-2)} x_n \right) \\
 &\vdots \\
 x_1 &= \frac{1}{a_{11}} \left(a_{1n+1} - a_{12} x_2 - a_{13} x_3 - \cdots - a_{1n} x_n \right)
 \end{aligned} \tag{3.4}$$

Postup vedoucí k soustavě 3.3 se nazývá **Gaussova eliminace**, výpočet neznámých dle 3.4 **zpětná substituce** nebo též **zpětný chod**. Číslo $a_{kk}^{(k-1)}$ nazýváme hlavní prvek.

Příklad 3.1 Pomocí Gaussovy eliminace vyřešte soustavu rovnic

$$\begin{aligned}
 1,67 x_1 - 0,15 x_2 + 2,51 x_3 &= -0,84 \\
 2,15 x_1 + 3,02 x_2 - 0,17 x_3 &= 2,32 \\
 1,71 x_1 - 2,83 x_2 + 1,45 x_3 &= 1,26
 \end{aligned}$$

Řešení: Koefficienty soustavy opíšeme do matice:

$$\begin{pmatrix} 1,67 & -0,15 & 2,51 & -0,84 \\ 2,15 & 3,02 & -0,17 & 2,32 \\ 1,71 & -2,83 & 1,45 & 1,26 \end{pmatrix}$$

Od druhého řádku odečteme první řádek, vynásobený $\frac{2,15}{1,67}$ a od třetího vynásobený $\frac{1,71}{1,67}$ (všechny mezivýsledky jsou zaokrouhlovány na pět desetinných míst):

$$\begin{pmatrix} 1,67 & -0,15 & 2,51 & -0,84 \\ 0 & 3,21311 & -3,40144 & 3,40144 \\ 0 & -2,67641 & -1,12012 & 2,12012 \end{pmatrix}$$

Nyní od třetího řádku odečteme druhý, vynásobený $\frac{-2,67641}{3,21311}$. Tím dostaneme

$$\begin{pmatrix} 1,67 & -0,15 & 2,51 & -0,84 \\ 0 & 3,21311 & -3,40144 & 3,40144 \\ 0 & 0 & -3,95339 & 4,95339 \end{pmatrix},$$

což už odpovídá soustavě v trojúhelníkovém tvaru

$$\begin{array}{rclcl} 1,67x_1 & - & 0,15x_2 & + & 2,51x_3 & = & -0,84 \\ & & 3,21311x_2 & - & 3,40144x_3 & = & 3,40144 \\ & & & & -3,95339x_3 & = & 4,95339 \end{array}$$

Řešení této soustavy je

$$\begin{aligned} x_3 &= \frac{4,95339}{-3,95339} \doteq -1,25295 \\ x_2 &= \frac{1}{3,21311} \left(3,40144 + 3,40144 \cdot (-1,25295) \right) \doteq -0,26777 \\ x_1 &= \frac{1}{1,67} \left(-0,84 + 0,15 \cdot (-0,26777) - 2,51 \cdot (-1,25295) \right) \doteq 1,35613 \end{aligned}$$

3.1.3 Eliminace s výběrem hlavního prvku

Eliminace s výběrem hlavního prvku je modifikace Gaussovy eliminační metody, která slouží ke zmenšení zaokrouhlovacích chyb.

Je-li absolutní hodnota některého z dělitelů $a_{ii}^{(i-1)}$ malá ve srovnání s absolutní hodnotou prvků $a_{ki}^{(i-1)}$, $k > i$, může hrozit nebezpečí velkých zaokrouhlovacích chyb. Zaokrouhlovací chyba v absolutní hodnotě malého čísla způsobí totiž velkou chybu v jeho převrácené hodnotě a tedy i v číslech, jimiž násobíme řádky při eliminaci.

Abychom se vyhnuli dělení čísly, která jsou malá vzhledem k ostatním veličinám, použijeme postup zvaný **výběr hlavního prvku**:

V prvním kroku eliminace najdeme rovnici, která má u x_1 v absolutní hodnotě největší koeficient. Vyměníme ji s první rovnicí a pak pomocí jejích násobků eliminujeme x_1 z ostatních rovnic. Ve druhém kroku najdeme mezi všemi rovnicemi kromě první tu rovnici, která má v absolutní hodnotě největší koeficient u x_2 . Vyměníme ji s druhou rovnicí a pomocí jejích násobků eliminujeme x_2 z dalších rovnic. Obecně v k -tém kroku eliminace najdeme mezi posledními $n - k + 1$ rovnicemi tu, která má největší koeficient u x_k , vyměníme ji s k -tou rovnicí a pak pomocí ní eliminujeme.

Příklad 3.2 *Soustavu z příkladu 3.1 řešte eliminací s výběrem hlavního prvku.*

Řešení: Postupujeme podobně jako v předchozím příkladu. Vybraný hlavní prvek je vždy v rámečku.

$$\begin{aligned} \left(\begin{array}{cccc} 1,67 & -0,15 & 2,51 & -0,84 \\ \boxed{2,15} & 3,02 & -0,17 & 2,32 \\ 1,71 & -2,83 & 1,45 & 1,26 \end{array} \right) &\sim \left(\begin{array}{cccc} 2,15 & 3,02 & -0,17 & 2,32 \\ 0 & -2,49577 & 2,64205 & -2,64205 \\ 0 & \boxed{-5,23195} & 1,58521 & -0,58521 \end{array} \right) \sim \\ \left(\begin{array}{cccc} 2,15 & 3,02 & -0,17 & 2,32 \\ 0 & -5,23195 & 1,58521 & -0,58521 \\ 0 & 0 & 1,88586 & -2,36289 \end{array} \right) \end{aligned}$$

Následovala by zpětná substituce.

Právě popsanou metodu bychom mohli nazvat výstižněji eliminační metodou s **částečným výběrem hlavního prvku**.

Úplný výběr hlavního prvku spočívá v tom, že v k -tém kroku volíme za hlavní prvek ten, který je největší v absolutní hodnotě v submatici vytvořené vynecháním prvních $k - 1$ řádků a sloupců v upravované matici. Nutnost hledat největší prvek v celé submatici a vyměňovat řádky i sloupce způsobuje větší časovou (a programátorskou) náročnost této metody. Gaussova eliminační metoda s částečným výběrem je proto obvykle efektivnější než metoda s úplným výběrem hlavního prvku.

3.2 Iterační metody

Iterační metody, na rozdíl od přímých metod, nevedou k přesnému řešení po konečném, předem daném počtu kroků. U iteračních metod zvolíme počáteční aproximaci řešení a určitým postupem ji v každém kroku metody zlepšíme. K řešení se přibližujeme postupně a obecně ho dosáhneme až v limitě. Protože výpočet nelze provádět do nekonečna, po jisté době jej ukončíme. Výsledkem bude přibližné řešení soustavy.

3.2.1 Jacobiho metoda

Nejprve popíšeme, jak se Jacobiho metodou soustavy rovnic řeší a kdy se touto metodou řešit mohou. Na konci kapitoly teoreticky zdůvodníme, proč Jacobiho metoda funguje. (Aby čtenář děsící se jakékoli teorie mohl konec kapitoly přeskóčit a nebyl hned zpočátku zastrašen.)

Budeme opět pracovat se soustavou lineárních rovnic

$$\begin{aligned} a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n &= b_1 \\ a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n &= b_2 \\ \vdots & \\ a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{nn} x_n &= b_n \end{aligned}$$

Z první rovnice vyjádříme x_1 , ze druhé rovnice x_2 atd. Dostaneme

$$\begin{aligned} x_1 &= \frac{1}{a_{11}} (b_1 - a_{12} x_2 - a_{13} x_3 - \cdots - a_{1n} x_n) \\ x_2 &= \frac{1}{a_{22}} (b_2 - a_{21} x_1 - a_{23} x_3 - \cdots - a_{2n} x_n) \\ &\vdots \\ x_n &= \frac{1}{a_{nn}} (b_n - a_{n1} x_1 - a_{n2} x_2 - \cdots - a_{nn-1} x_{n-1}) \end{aligned} \tag{3.5}$$

Řešení soustavy budeme hledat následujícím způsobem:

Libovolně zvolíme počáteční aproximaci řešení $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^T$.

Tato čísla dosadíme do pravé strany 3.5. Tím dostaneme novou aproximaci řešení $\mathbf{x}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)})^T$. Tu opět dosadíme do pravé strany 3.5 atd.

Obecně každou další aproximaci řešení získáme podle předpisu

$$\begin{aligned} x_1^{(r+1)} &= \frac{1}{a_{11}} \left(b_1 - a_{12} x_2^{(r)} - a_{13} x_3^{(r)} - \cdots - a_{1n} x_n^{(r)} \right) \\ x_2^{(r+1)} &= \frac{1}{a_{22}} \left(b_2 - a_{21} x_1^{(r)} - a_{23} x_3^{(r)} - \cdots - a_{2n} x_n^{(r)} \right) \\ &\vdots \\ x_n^{(r+1)} &= \frac{1}{a_{nn}} \left(b_n - a_{n1} x_1^{(r)} - a_{n2} x_2^{(r)} - \cdots - a_{nn-1} x_{n-1}^{(r)} \right), \end{aligned} \quad (3.6)$$

Za jistých (dále popsaných podmínek) se tímto postupem budeme přibližovat k přesnému řešení soustavy.

Ve výpočtu pokračujeme, dokud se nedosáhne určité předem dané přesnosti, např. dokud se aproximace řešení neustálí na požadovaném počtu desetinných míst, nebo dokud není překročen předem daný maximální počet kroků.

Jacobiho metodou nemusíme řešení soustavy najít vždy. V některých případech posloupnost postupných aproximací k řešení soustavy nekonverguje. Uvedeme nyní podmínky, které zaručí, že metoda konverguje (tj. najdeme pomocí ní přibližné řešení).

Definice. Matice A se nazývá **řádkově ostře diagonálně dominantní** právě tehdy, když

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}| \quad \text{pro } i = 1, \dots, n \quad (3.7)$$

(neboli když je v každém řádku matice absolutní hodnota prvku na diagonále větší než součet absolutních hodnot všech ostatních prvků v onom řádku)

a **sloupcově ostře diagonálně dominantní** právě tehdy, když

$$|a_{jj}| > \sum_{i=1, i \neq j}^n |a_{ij}| \quad \text{pro } j = 1, \dots, n \quad (3.8)$$

(neboli když je v každém sloupci matice absolutní hodnota prvku na diagonále větší než součet absolutních hodnot všech ostatních prvků v onom sloupci).

Na konci této kapitoly dokážeme, že:

Je-li matice soustavy 3.2 ostře řádkově nebo sloupcově diagonálně dominantní, Jacobiho metoda konverguje.

Jestliže matice soustavy 3.2 není diagonálně dominantní, Jacobiho metoda konvergovat může a nemusí.

Existuje podmínka pro konvergenci Jacobiho metody nutná a dostatečná (tj. pokud je splněna, metoda konverguje a pokud není splněna, metoda diverguje), jenže je pro velké matice prakticky neověřitelná.

Proto, nejsme-li si jisti konvergencí metody, je vhodné stanovit maximální počet kroků a je-li překročen, výpočet ukončit s tím, že metoda diverguje. Pak je potřeba zvolit jinou metodu nebo soustavu nějak upravit.

Příklad 3.3 *Jacobiho metodou řešte soustavu*

$$\begin{aligned} 15x_1 - x_2 + 2x_3 &= 30 \\ 2x_1 - 10x_2 + x_3 &= 23 \\ x_1 + 3x_2 + 18x_3 &= -22 \end{aligned}$$

Řešení: Matice soustavy je diagonálně dominantní, protože platí

$$|15| > |-1| + |2|, \quad |-10| > |2| + |1|, \quad |18| > |1| + |3|.$$

Proto je konvergence metody zaručena. Vypíšeme iterační vztahy:

$$\begin{aligned} x_1^{(r+1)} &= \frac{1}{15} \left(30 + x_2^{(r)} - 2x_3^{(r)} \right) \\ x_2^{(r+1)} &= -\frac{1}{10} \left(23 - 2x_1^{(r)} - x_3^{(r)} \right) \\ x_3^{(r+1)} &= \frac{1}{18} \left(-22 - x_1^{(r)} - 3x_2^{(r)} \right) \end{aligned}$$

Jako počáteční aproximaci zvolíme $\mathbf{x} = (0, 0, 0)^T$. Postupně získávané aproximace řešení budeme zapisovat do tabulky:

r	$x_1^{(r)}$	$x_2^{(r)}$	$x_3^{(r)}$
0	0	0	0
1	2	-2,3	-1,2222
2	2,0096	-2,0222	-0,9500
3	1,9918	-1,9930	-0,9968
4	2,0000	-2,0013	-1,0007

Je vidět, že posloupnost postupných aproximací konverguje k řešení soustavy (2,-2,-1). Kdybychom chtěli získat řešení s přesností $\varepsilon = 0,01$, mohli bychom nyní výpočet zastavit, protože

$$\begin{aligned} |x_1^{(4)} - x_1^{(3)}| &= |2,0000 - 1,9918| < 0,01 \\ |x_2^{(4)} - x_2^{(3)}| &= |-2,0013 - (-1,9930)| < 0,01 \\ |x_3^{(4)} - x_3^{(3)}| &= |-1,0007 - (-0,9968)| < 0,01, \end{aligned}$$

zatímco kdybychom požadovali přesnost $\varepsilon = 0,001$, museli bychom ve výpočtu pokračovat, protože např. $|x_1^{(4)} - x_1^{(3)}| > 0,001$.

Ukázka divergence Jacobiho metody

Kdybychom rovnice z předcházejícího příkladu přepsali v jiném pořadí, např.

$$\begin{aligned} x_1 + 3x_2 + 18x_3 &= -22 \\ 15x_1 - x_2 + 2x_3 &= 30 \\ 2x_1 - 10x_2 + x_3 &= 23, \end{aligned}$$

příslušné iterační vztahy by vypadaly takto:

$$\begin{aligned}x_1^{(r+1)} &= -22 - 3x_2^{(r)} - 18x_3^{(r)} \\x_2^{(r+1)} &= -30 + 15x_1^{(r)} + 2x_3^{(r)} \\x_3^{(r+1)} &= 23 - 2x_1^{(r)} + 10x_2^{(r)}.\end{aligned}$$

Podmínka konvergence metody není splněna. Podívejme se, jak se budou chovat postupné aproximace řešení:

r	$x_1^{(r)}$	$x_2^{(r)}$	$x_3^{(r)}$
0	0	0	0
1	-22	-30	23
2	-346	-314	-233
3	5114	-5686	-2425

Na první pohled je zřejmé, že k řešení soustavy $(2, -2, -1)$ touto cestou nedojdeme, metoda diverguje.

Jacobiho metoda z teoretického hlediska

Nyní ukážeme, proč Jacobiho metoda funguje a proč konverguje zrovna za výše uvedených podmínek.

Rovnice 3.5 se dají zapsat maticově jako

$$\mathbf{x} = \mathbf{C}_J \mathbf{x} + \mathbf{d},$$

kde \mathbf{C}_J je tzv. **iterační matice** Jacobiho metody. Prvky matice \mathbf{C}_J a vektoru \mathbf{d} jsou

$$\begin{aligned}c_{ij} &= -\frac{a_{ij}}{a_{ii}} \quad \text{pro } i \neq j, \quad c_{ii} = 0 \\d_i &= \frac{b_i}{a_{ii}}.\end{aligned}$$

Tím, že jsme původní soustavu rovnic $\mathbf{A} \mathbf{x} = \mathbf{b}$ upravili na tento tvar, se úkol najít řešení soustavy rovnic převedl na hledání pevného bodu zobrazení

$$F(\mathbf{x}) = \mathbf{C}_J \mathbf{x} + \mathbf{d}, \tag{3.9}$$

protože řešením původní soustavy rovnic je právě takový vektor \mathbf{x} , pro nějž platí $F(\mathbf{x}) = \mathbf{x}$.

V kapitole 2 jsme předvedli obecný postup, který vede k nalezení pevného bodu. Je to tzv. metoda postupných aproximací, iterační proces.

Proto řešení hledáme výše popsáním způsobem, tj. zvolíme libovolně počáteční aproximaci $\mathbf{x}^{(0)}$ a další aproximace počítáme jako

$$\mathbf{x}^{(r+1)} = F(\mathbf{x}^{(r)}) = \mathbf{C}_J \mathbf{x}^{(r)} + \mathbf{d}. \tag{3.10}$$

Dále jsme v kapitole 2 uvedli, za jakých podmínek je jisté, že pevný bod zobrazení existuje a že metodou postupných aproximací k němu dojdeme. Prozkoumáme nyní, jak vypadají tyto obecné podmínky pro naši konkrétní situaci.

Máme zobrazení $F : V_n \rightarrow V_n$, kde V_n je prostor všech uspořádaných n -tic reálných čísel. Na tomto prostoru můžeme zavést metriku předpisem

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|,$$

kde $\|\cdot\|$ je některá z norem 2.12, 2.13. Prostor V_n s touto metrikou je úplný. Zjistíme, kdy bude zobrazení F kontraktivní.

Platí

$$\begin{aligned} d(F(\mathbf{x}), F(\mathbf{y})) &= \|F(\mathbf{x}) - F(\mathbf{y})\| = \|\mathbf{C}_J \mathbf{x} + \mathbf{d} - (\mathbf{C}_J \mathbf{y} + \mathbf{d})\| = \|\mathbf{C}_J (\mathbf{x} - \mathbf{y})\| \leq \\ &\leq \|\mathbf{C}_J\| \cdot \|\mathbf{x} - \mathbf{y}\| = \|\mathbf{C}_J\| \cdot d(\mathbf{x}, \mathbf{y}), \end{aligned}$$

kde $\|\mathbf{C}_J\|$ je norma matice přidružená použité normě vektoru.

Je-li tedy $\|\mathbf{C}_J\| < 1$, je zobrazení F kontraktivní s koeficientem kontrakce $\alpha = \|\mathbf{C}_J\|$ a je zaručeno, že posloupnost postupných aproximací získaná podle předpisu 3.10 konverguje k pevnému bodu zobrazení 3.9.

(Je-li $\|\mathbf{C}_J\| > 1$, o konvergenci či divergenci iteračního procesu nevíme nic.)

Nyní se podíváme na to, jak podmínka $\|\mathbf{C}_J\| < 1$ souvisí s diagonální dominantností matice soustavy \mathbf{A} .

Předpokládejme, že matice \mathbf{A} je ostře řádkově diagonálně dominantní.

Počítáme-li řádkovou normu matice \mathbf{C}_J , bereme součty absolutních hodnot prvků v jednotlivých řádcích a z nich pak vybíráme maximum. Součet absolutních hodnot prvků prvního řádku je

$$\left| -\frac{a_{12}}{a_{11}} \right| + \left| -\frac{a_{13}}{a_{11}} \right| + \cdots + \left| -\frac{a_{1n}}{a_{11}} \right| = \frac{|a_{12}| + |a_{13}| + \cdots + |a_{1n}|}{|a_{11}|}.$$

Protože je \mathbf{A} řádkově diagonálně dominantní, musí být

$$|a_{11}| > |a_{12}| + |a_{13}| + \cdots + |a_{1n}|$$

a tedy součet absolutních hodnot prvků prvního řádku matice \mathbf{C}_J musí být menší než 1. Úplně stejně se ukáže, že i součty v ostatních řádcích jsou menší než jedna.

Řádková norma matice \mathbf{C}_J , coby největší z čísel menších než jedna, bude určitě také menší než jedna. Proto, je-li \mathbf{A} řádkově diagonálně dominantní, je zaručeno, že Jacobiho metoda konverguje.

Zcela analogicky se dá ukázat, že je-li \mathbf{A} ostře sloupcově diagonálně dominantní, je sloupcová norma matice \mathbf{C}_J menší než 1.

V případě, že je $\|\mathbf{C}_J\| < 1$, platí odhady 2.8 a 2.9 z věty 2.1. Zde jsou přepsány speciálně pro naši úlohu:

$$\|\mathbf{x}^{(r)} - \mathbf{x}\| \leq \frac{\|\mathbf{C}_J\|}{1 - \|\mathbf{C}_J\|} \|\mathbf{x}^{(r)} - \mathbf{x}^{(r-1)}\| \quad (3.11)$$

$$\|\mathbf{x}^{(r)} - \mathbf{x}\| \leq \frac{\|\mathbf{C}_J\|^r}{1 - \|\mathbf{C}_J\|} \|\mathbf{x}^{(0)} - \mathbf{x}^{(1)}\| \quad (3.12)$$

Pomocí odhadu 3.11 můžeme rozhodnout, kdy zastavit iterační proces, chceme-li mít jistotu, že se přibližné řešení od přesného v použité normě neliší víc než o předem dané ε . Odhad 3.12 může posloužit k určení počtu kroků metody, který bude stačit pro dosažení přesnosti ε .

Protože však pro velké soustavy rovnic je vypočítat normu matice \mathbf{C}_J pracné, pro zastavení výpočtu se spíše používá kritérium

$$\|\mathbf{x}^{(r)} - \mathbf{x}^{(r-1)}\| < \varepsilon,$$

i když jeho splněním není zaručeno, že bude i $\|\mathbf{x}^{(r)} - \mathbf{x}\| < \varepsilon$.

(Toto kritérium se objevilo již v příkladu 3.3, použita byla norma $\|\cdot\|_\infty$.)

Příklad 3.4 Odhadněte, o kolik se nanejvýš liší přibližné řešení získané v příkladu 3.3 od přesného řešení v normě $\|\cdot\|_\infty$.

Řešení: K odhadu chyby použijeme vzorec 3.11. K tomu musíme vypočítat řádkovou normu iterační matice \mathbf{C}_J . Nejprve vypíšeme samotnou iterační matici:

$$\mathbf{C}_J = \begin{pmatrix} 0 & \frac{1}{15} & -\frac{2}{15} \\ \frac{2}{10} & 0 & \frac{1}{10} \\ -\frac{1}{18} & -\frac{3}{18} & 0 \end{pmatrix}$$

$$\|\mathbf{C}_J\|_\infty = \max\left(\frac{3}{15}, \frac{3}{10}, \frac{4}{18}\right) = \frac{3}{10} = 0,3.$$

Dále vypočteme normu rozdílu posledních dvou získaných aproximací

$$\mathbf{x}^{(3)} = (1,9918; -1,9930; -0,9968) \text{ a } \mathbf{x}^{(4)} = (2,0000; -2,0013; -1,0007) :$$

$$\|\mathbf{x}^{(4)} - \mathbf{x}^{(3)}\|_\infty = \max(|0,0082|; |-0,0095|; |-0,0039|) = 0,0095$$

Nyní dosadíme do 3.11

$$\|\mathbf{x}^{(4)} - \mathbf{x}\|_\infty \leq \frac{0,3}{1-0,3} \cdot 0,0095 \doteq 0,0041$$

To znamená, že každá ze složek přibližného řešení $\mathbf{x}^{(4)}$ se od odpovídající složky přesného řešení může lišit nanejvýš o 0,0041.

3.2.2 Gauss-Seidelova metoda

Gauss-Seidelova metoda je velmi podobná metodě Jacobiho. Liší se od ní pouze v tom, že při výpočtu další aproximace řešení použijeme vždy nejnovější přibližné hodnoty x_1, x_2, \dots, x_n , které máme k dispozici.

Podrobněji: $x_1^{(r+1)}$ vypočteme stejně jako u Jacobiho metody a při výpočtu $x_2^{(r+1)}$ je ihned použijeme (zatímco u Jacobiho metody jsme použili staré $x_1^{(r)}$). Při výpočtu $x_3^{(r+1)}$ použijeme nové $x_1^{(r+1)}$ a $x_2^{(r+1)}$ atd.

Obecně iterační vztahy vypadají takto:

$$\begin{aligned}
 x_1^{(r+1)} &= \frac{1}{a_{11}} \left(b_1 - a_{12} x_2^{(r)} - a_{13} x_3^{(r)} - \dots - a_{1n} x_n^{(r)} \right) \\
 x_2^{(r+1)} &= \frac{1}{a_{22}} \left(b_2 - a_{21} x_1^{(r+1)} - a_{23} x_3^{(r)} - \dots - a_{2n} x_n^{(r)} \right) \\
 x_3^{(r+1)} &= \frac{1}{a_{33}} \left(b_3 - a_{31} x_1^{(r+1)} - a_{32} x_2^{(r+1)} - \dots - a_{3n} x_n^{(r)} \right) \\
 &\vdots \\
 x_n^{(r+1)} &= \frac{1}{a_{nn}} \left(b_n - a_{n1} x_1^{(r+1)} - a_{n2} x_2^{(r+1)} - \dots - a_{nn-1} x_{n-1}^{(r+1)} \right),
 \end{aligned} \tag{3.13}$$

Dá se dokázat, že **je-li matice soustavy 3.2 ostře řádkově diagonálně dominantní, Gauss-Seidelova metoda konverguje.**

V jiném kritériu konvergence se objevuje pojem pozitivně definitní matice. Protože není jisté, zda se s ním studenti již setkali, řekneme, co to je.

Definice. Symetrická matice \mathbf{A} řádu n se nazývá **pozitivně definitní**, jestliže pro každý nenulový sloupcový vektor $\mathbf{x} = (x_1, \dots, x_n)^T$ platí

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$$

Příklad. Pozitivně definitní je např. matice

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$$

protože pro každý vektor $\mathbf{x} = (x_1, x_2)^T \neq (0, 0)^T$ platí

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = x_1^2 + 4x_1x_2 + 5x_2^2 = (x_1 + 2x_2)^2 + x_2^2 > 0,$$

zatímco matice

$$\mathbf{B} = \begin{pmatrix} -1 & 2 \\ 2 & 5 \end{pmatrix}$$

není pozitivně definitní, protože např. pro $\mathbf{x} = (1, 0)^T$ platí

$$(1, 0) \mathbf{B} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (1, 0) \begin{pmatrix} -1 \\ 2 \end{pmatrix} = -1 < 0.$$

Platí: **Je-li matice soustavy 3.2 pozitivně definitní, Gauss-Seidelova metoda konverguje.**

Ověření toho, že je daná matice pozitivně definitní, je náročné a pro velké matice prakticky neproveditelné. Naštěstí je u některých úloh z povahy řešeného problému předem jasné, že matice soustavy pozitivně definitní bude.

Poznámka. Vynásobíme-li libovolnou regulární čtvercovou matici \mathbf{A} zleva maticí k ní trasponovanou, vzniklá matice $\mathbf{A}^T \mathbf{A}$ bude symetrická a pozitivně definitní.

Proto, vynásobíme-li soustavu rovnic $\mathbf{Ax} = \mathbf{b}$ s regulární maticí \mathbf{A} zleva maticí \mathbf{A}^T , dostaneme novou soustavu

$$\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b},$$

jejíž matice je pozitivně definitní a je tedy zaručeno, že Gauss-Seidelova metoda bude pro tuto novou soustavu konvergovat.

V případě takto získaných soustav však Gauss-Seidelova metoda může konvergovat velmi pomalu.

Příklad 3.5 Gauss-Seidelovou metodou řešte tutéž soustavu jako v příkladu 3.3, t.j.

$$\begin{aligned} 15x_1 - x_2 + 2x_3 &= 30 \\ 2x_1 - 10x_2 + x_3 &= 23 \\ x_1 + 3x_2 + 18x_3 &= -22 \end{aligned}$$

Řešení: Již jsme ověřili, že podmínka konvergence je splněna. Vypíšeme iterační vztahy:

$$\begin{aligned} x_1^{(r+1)} &= \frac{1}{15} \left(30 + x_2^{(r)} - 2x_3^{(r)} \right) \\ x_2^{(r+1)} &= -\frac{1}{10} \left(23 - 2x_1^{(r+1)} - x_3^{(r)} \right) \\ x_3^{(r+1)} &= \frac{1}{18} \left(-22 - x_1^{(r+1)} - 3x_2^{(r+1)} \right) \end{aligned}$$

Jako počáteční aproximaci zvolíme opět $\mathbf{x} = (0, 0, 0)^T$.

r	$x_1^{(r)}$	$x_2^{(r)}$	$x_3^{(r)}$
0	0	0	0
1	2	-1,9	-1,0167
2	2,0089	-1,9999	-1,0005
3	2,0001	-2,0000	-1,0000
4	2,0000	-2,0000	-1,0000

Vidíme, že se k řešení soustavy přibližujeme rychleji než pomocí Jacobiho metody.

I obecně se dá říci, že Gauss-Seidelova metoda obvykle konverguje rychleji než metoda Jacobiho. Proto se používá častěji. Další její výhodou oproti Jacobiho metodě je, že pro uložení přibližného řešení v paměti počítače nám stačí jediné pole, jehož složky postupně přepisujeme, zatímco u Jacobiho metody si musíme pamatovat pole dvě: starou a novou aproximaci řešení.

Existuje a v praxi se používá i mnoho dalších iteračních metod. Seznámil-li se čtenář s Jacobiho a Gauss-Seidelovou metodou, bude pro něj jednodušší pochopit i jiné iterační metody, pokud se s nimi někdy setká.

3.3 Srovnání přímých a iteračních metod

Eliminační metoda je velmi náročná z časového i paměťového hlediska. Máme-li řešit n rovnic, je potřeba vykonat přibližně $n^3/3$ aritmetických operací.

Proto se hodí nejlépe pro nepřilíš rozsáhlé soustavy s plnou maticí.

Dnes existují profesionální programy i pro řešení velkých soustav rovnic s řídkou maticí koeficientů (řídkou maticí se rozumí taková matice, která má v každém řádku jen malý počet nenulových prvků).

Iterační metody jsou vhodné pro řešení velkých soustav s řídkou maticí koeficientů. Pro řešení malého počtu rovnic vhodné nejsou, tam lépe poslouží eliminace.

4 Numerické metody řešení nelineárních rovnic

4.1 Numerické metody řešení jedné nelineární rovnice

Budeme se zabývat řešením nelineární rovnice

$$f(x) = 0, \quad (4.1)$$

tj. hledáním takových bodů $\xi \in \mathbb{R}$, že $f(\xi) = 0$. Takovéto body budeme nazývat **kořeny rovnice 4.1**.

Při hledání kořenů rovnice 4.1 nejprve zjistíme, kolik kořenů rovnice má a najdeme intervaly obsahující právě jeden kořen rovnice. Tato část řešení se nazývá **separace kořenů rovnice**.

Pak budeme pomocí některé z dále popsaných metod hledat přibližnou hodnotu vybraného kořene rovnice.

Při hledání kořenů je užitečná následující věta, jejíž význam je patrný z obrázku 4.1

Věta 4.1 *Je-li funkce f spojitá na intervalu $\langle a, b \rangle$ a platí-li*

$$f(a) \cdot f(b) < 0, \quad (4.2)$$

pak v intervalu $\langle a, b \rangle$ leží alespoň jeden kořen rovnice $f(x) = 0$.

Poznámka. Podmínka 4.2 znamená, že znaménka funkčních hodnot v krajních bodech intervalu $\langle a, b \rangle$ jsou opačná.

Kořenů rovnice může být v uvedeném intervalu i více, o jejich počtu věta nic neříká. Na druhou stranu, není-li podmínka 4.2 splněna, neznamená to, že v intervalu $\langle a, b \rangle$ žádný kořen rovnice neleží.

Pro nalezení počtu a polohy kořenů je vhodné prozkoumat vlastnosti funkce f a načrtnout (nebo si pomocí vhodného prostředku nechat načrtnout) její graf.

U některých úloh je možné upravit rovnici 4.1 na tvar

$$f_1(x) = f_2(x),$$

kde f_1 a f_2 jsou funkce, jejichž grafy umíme nakreslit. V bodech, kde se grafy funkcí f_1 a f_2 protnou, se nacházejí kořeny původní rovnice.

Příklad 4.1 *Najděte počet kořenů rovnice*

$$e^x + x^2 - 3 = 0$$

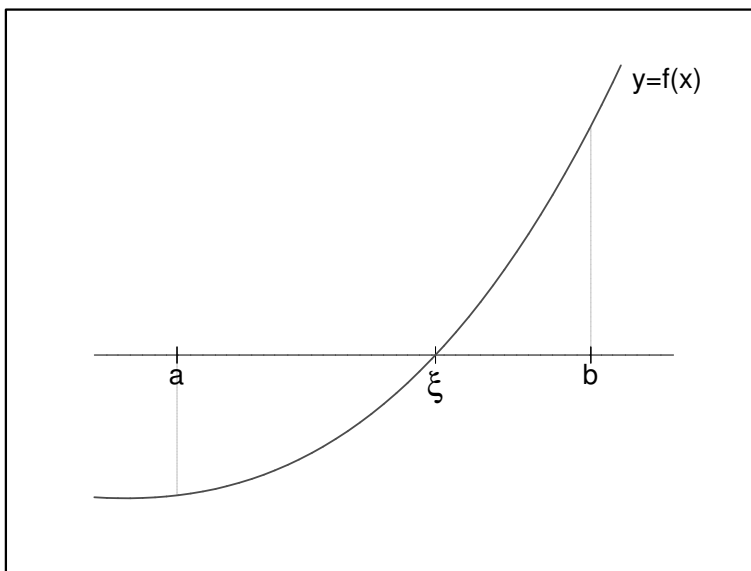
a intervaly, v nichž tyto kořeny leží.

Řešení: Zadanou rovnici můžeme upravit na tvar

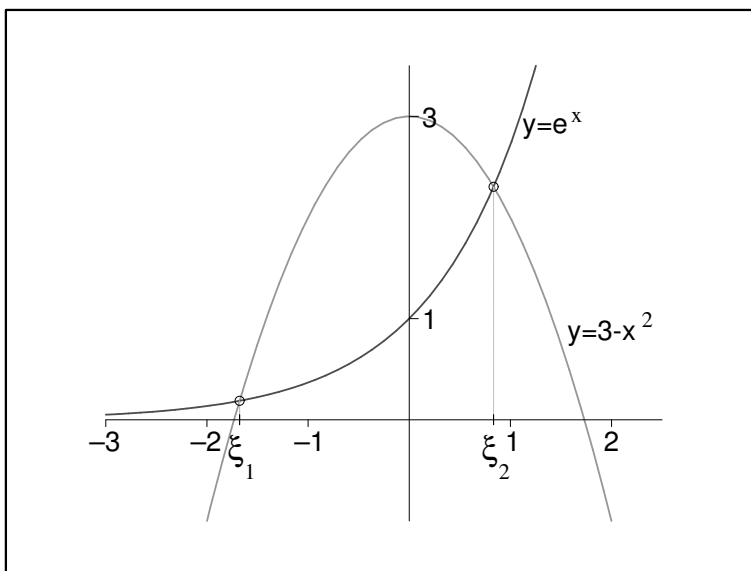
$$e^x = 3 - x^2.$$

Grafy funkcí $f_1(x) = e^x$ a $f_2(x) = 3 - x^2$ umíme načrtnout - viz obrázek 4.2. Z obrázku vidíme, že rovnice má právě dva kořeny ξ_1 a ξ_2 , $\xi_1 \in \langle -2, -1 \rangle$, $\xi_2 \in \langle 0, 1 \rangle$.

Nyní postupně probereme metody, které lze použít pro nalezení kořenů rovnice 4.1. **Všude dál v této kapitole budeme předpokládat, že funkce f je na zkoumaném intervalu spojitá.**



Obrázek 4.1: Ilustrace k větě 4.1



Obrázek 4.2: K příkladu 4.1 - separace kořenů rovnice

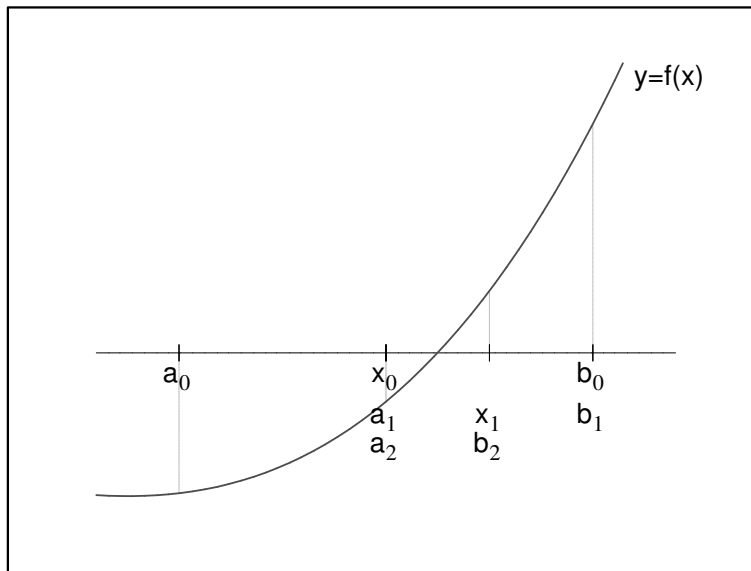
4.1.1 Metoda půlení intervalu

Metoda půlení intervalu je nejjednodušší z metod řešení nelineárních rovnic.

Mějme interval $\langle a, b \rangle$ takový, že $f(a) \cdot f(b) < 0$, tj. leží v něm alespoň jeden kořen rovnice $f(x) = 0$. Tento výchozí interval označíme jako $\langle a_0, b_0 \rangle$. Interval rozpůlíme. Jeho střed je $x_0 = \frac{a_0 + b_0}{2}$. Z intervalů $\langle a_0, x_0 \rangle$, $\langle x_0, b_0 \rangle$ vybereme ten, ve kterém je zaručena existence kořene. Který z nich to je, rozeznáme podle znamének funkčních hodnot v

krajních bodech. Je-li $f(a_0) \cdot f(x_0) < 0$, budeme pokračovat s intervalem $\langle a_0, x_0 \rangle$, v opačném případě s intervalem $\langle x_0, b_0 \rangle$. (Platí-li $f(x_0) = 0$, našli jsme kořen rovnice a výpočet ukončíme.) Nový interval poloviční délky označíme $\langle a_1, b_1 \rangle$, opět jej rozpůlíme a stejným způsobem pokračujeme.

Takto postupně sestrojíme posloupnost intervalů $\langle a_0, b_0 \rangle$, $\langle a_1, b_1 \rangle$, $\langle a_2, b_2 \rangle$, ... Každý další interval získáme tak, že z předchozího (na základě znamének funkčních hodnot v krajních bodech a uprostřed) vybereme tu jeho polovinu, která obsahuje kořen rovnice - viz obrázek 4.3.



Obrázek 4.3: Metoda půlení intervalu

V půlení pokračujeme tak dlouho, dokud nenarazíme na kořen rovnice, nebo dokud se interval nezúží na předem danou délku 2ε , neboli dokud pro nějaké k neplatí

$$b_k - a_k < 2\varepsilon$$

Za přibližnou hodnotu kořene pak vezmeme střed posledního nalezeného intervalu

$$x_k = \frac{a_k + b_k}{2}$$

Protože kořen se určitě nachází uvnitř posledního intervalu, může se x_k od přesné hodnoty kořene lišit nanejvýš o polovinu jeho délky, tj. o ε ,

$$|x_k - \xi| < \varepsilon.$$

Touto metodou kořen rovnice 4.1 nalezneme vždy. Obsahuje-li výchozí interval $\langle a, b \rangle$ více kořenů, najdeme jeden z nich. Nevýhodou metody půlení intervalu je, že konverguje (přibližuje se ke kořeni) dosti pomalu. Proto je vhodné použít ji na zúžení původního intervalu a pak pokračovat jinou, rychlejší metodou.

Příklad 4.2 Metodou půlení intervalu najděte kladný kořen rovnice z příkladu 4.1

$$e^x + x^2 - 3 = 0$$

s přesností $\varepsilon = 0,01$.

Řešení: Kladný kořen zadané rovnice leží v intervalu $\langle 0, 1 \rangle$. Postupně vypočítávané hodnoty a_k, b_k, x_k budeme zapisovat do tabulky. Je vhodné si také zapisovat znaménka funkčních hodnot funkce $f(x) = e^x + x^2 - 3$ v těchto bodech.

k	a_k	b_k	x_k	$f(a_k)$	$f(b_k)$	$f(x_k)$
0	0	1	0,5	-	+	-
1	0,5	1	0,75	-	+	-
2	0,75	1	0,875	-	+	+
3	0,75	0,875	0,8125	-	+	-
4	0,8125	0,875	0,84375	-	+	+
5	0,8125	0,84375	0,828125	-	+	-
6	0,828125	0,84375	0,8359375			

Nyní můžeme výpočet ukončit, protože $b_6 - a_6 < 2 \cdot 0,01$. Řešení rovnice $e^x + x^2 - 3 = 0$ s přesností 0,01 je $x_6 \doteq 0,836$.

Příklad 4.3 Kolik dalších kroků metody půlení intervalu by bylo potřeba provést v předchozím příkladu, kdybychom chtěli najít řešení s přesností 0,001 ?

Řešení: V každém kroku se interval zkrátí na polovinu. Vyjdeme-li z intervalu délky l , po k krocích se zúží na $\frac{l}{2^k}$. V našem případě vycházíme z intervalu $\langle a_6, b_6 \rangle$ délky 0,015625.

Hledáme tedy k tak, aby platilo $\frac{0,015625}{2^k} < 2 \cdot 0,001$. Odtud $k > \frac{\ln \frac{0,015625}{0,002}}{\ln 2} \doteq 2,97$. Musíme tedy udělat ještě tři kroky.

Je vidět, že počet kroků metody půlení intervalu nutný k nalezení kořene se zadanou přesností vůbec nezávisí na řešené rovnici. Dá se ukázat (podobně jako v řešení příkladu 4.3), že k zpřesnění výsledku o jedno desetinné místo je vždy potřeba udělat 3-4 kroky této metody.

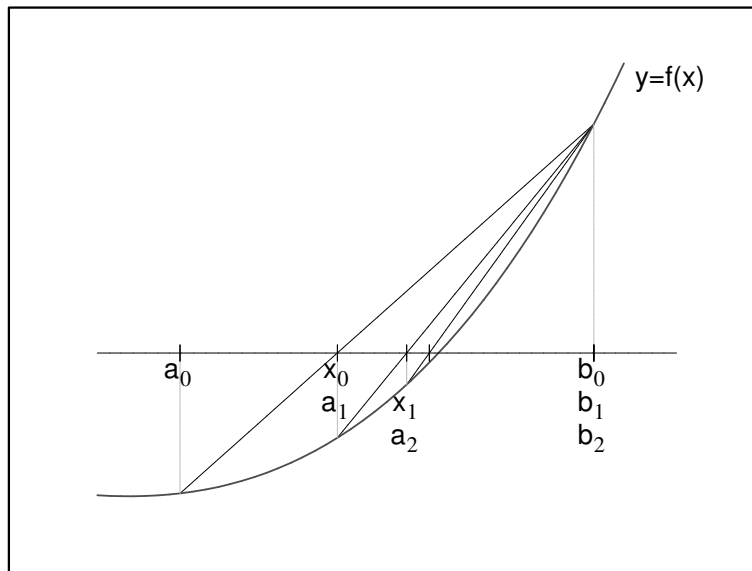
4.1.2 Metoda regula falsi

Princip metody regula falsi je velmi podobný jako u metody půlení intervalu. Opět postupně zužujeme interval obsahující kořen rovnice 4.1. Tentokrát ale dělicím bodem není polovina intervalu, nýbrž průsečík sečny vedené body $[a_k, f(a_k)]$ a $[b_k, f(b_k)]$ s osou x - viz obrázek 4.4.

Tento průsečík vypočteme podle vzorce

$$x_k = b_k - \frac{b_k - a_k}{f(b_k) - f(a_k)} f(b_k) \quad (4.3)$$

Z intervalů $\langle a_k, x_k \rangle$, $\langle x_k, b_k \rangle$ pak vybereme ten, v jehož krajních bodech mají funkční hodnoty funkce f opačná znaménka.



Obrázek 4.4: Metoda regula falsi

Platí-li $f(a_k) \cdot f(x_k) < 0$, položíme $a_{k+1} = a_k, b_{k+1} = x_k$, platí-li $f(b_k) \cdot f(x_k) < 0$, položíme $a_{k+1} = x_k, b_{k+1} = b_k$. V případě, že $f(x_k) = 0$, našli jsme kořen rovnice a výpočet ukončíme.

Ve výpočtu pokračujeme tak dlouho, dokud nenarazíme na kořen, nebo dokud neplatí

$$|x_k - x_{k-1}| < \varepsilon,$$

kde $\varepsilon > 0$ je předem dané číslo. Splněním tohoto kriteria ale bohužel není zaručeno, že přesná hodnota kořene ξ se od jeho aproximace x_k liší o méně než ε . Chceme-li se přesvědčit, že $|x_k - \xi| < \varepsilon$, můžeme vypočítat $f(x_k + \varepsilon)$ a $f(x_k - \varepsilon)$. Platí-li $f(x_k) \cdot f(x_k + \varepsilon) < 0$, resp. $f(x_k) \cdot f(x_k - \varepsilon) < 0$, je jisté, že kořen ξ leží v intervalu $\langle x_k, x_k + \varepsilon \rangle$, resp. $\langle x_k - \varepsilon, x_k \rangle$ a tedy se od x_k nemůže lišit o více než ε .

Metoda regula falsi je vždy konvergentní (vždy najde kořen). Bývá rychlejší než půlení intervalu, ale existují případy, kdy je pomalejší.

Příklad 4.4 Metodou regula falsi najděte kladný kořen rovnice z příkladu 4.1

$$e^x + x^2 - 3 = 0$$

s přesností $\varepsilon = 0,01$.

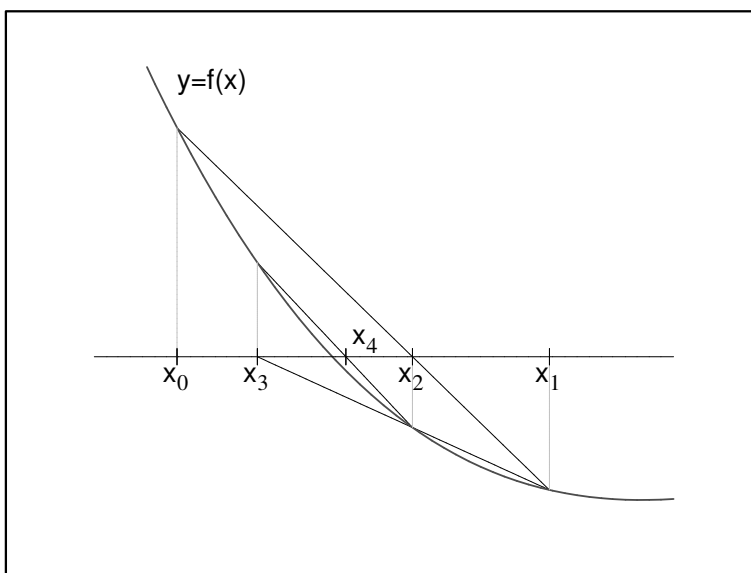
Řešení: Mohli bychom vyjít z intervalu nalezeného metodou půlení v příkladu 4.2, ale pro srovnání obou metod začneme opět s intervalem $\langle 0, 1 \rangle$. U metody regula falsi budeme potřebovat i funkční hodnoty v bodech a_k, b_k a x_k , nejen jejich znaménka.

k	a_k	b_k	x_k	$f(a_k)$	$f(b_k)$	$f(x_k)$
0	0	1	0,73576	-2	0,71828	-0,37159
1	0,73576	1	0,82585	-0,37159	0,71828	-0,03414
2	0,82585	1	0,83375	-0,03414	0,71828	-0,00291

Platí $|x_2 - x_1| < 0,01$, proto výpočet ukončíme. Přibližné řešení rovnice je $x_2 \doteq 0,834$.

4.1.3 Metoda sečen

Metoda sečen je velmi podobná jako metoda regula falsi. Vyjdeme z intervalu $\langle a, b \rangle$, obsahujícího kořen rovnice. Označíme $x_0 = a$ a $x_1 = b$. Vedeme sečnu body $[x_0, f(x_0)]$ a $[x_1, f(x_1)]$ a najdeme její průsečík s osou x . Ten označíme x_2 . Na rozdíl od metody regula falsi však nyní nevybíráme interval obsahující kořen, ale vedeme sečnu body $[x_1, f(x_1)]$, $[x_2, f(x_2)]$, její průsečík označíme x_3 , pak vedeme sečnu body $[x_2, f(x_2)]$ a $[x_3, f(x_3)]$ atd. - viz obrázek 4.5.



Obrázek 4.5: Metoda sečen

V k -tém kroku metody počítáme aproximaci kořene podle vzorce

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} f(x_k), \quad (4.4)$$

kde $x_0 = a, x_1 = b$.

Výpočet ukončíme, když je splněna podmínka

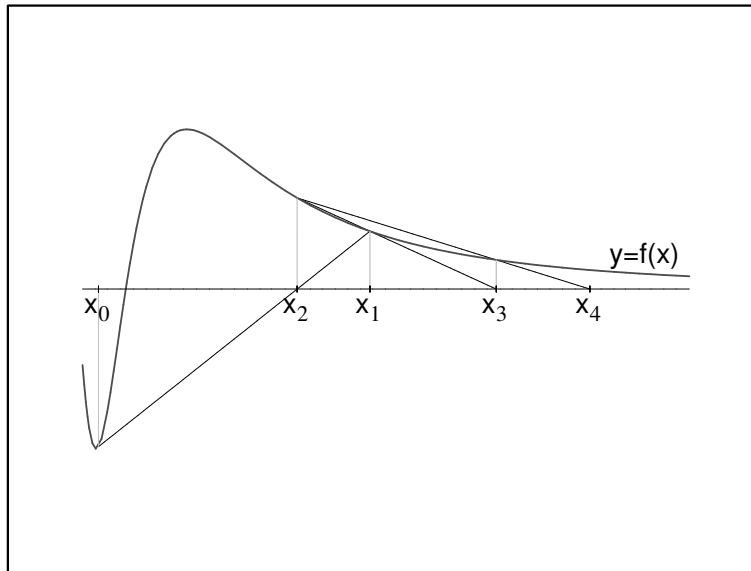
$$|x_k - x_{k-1}| < \varepsilon,$$

nebo když narazíme přímo na kořen rovnice.

Připomeňme, že daná podmínka nezaručuje, že platí $|x_k - \xi| < \varepsilon$.

Metoda sečen je rychlejší než metoda regula falsi, nemusí ale vždy konvergovat - viz obrázek 4.6.

Protože je obtížné předem zjistit, zda metoda pro danou rovnici konverguje nebo diverguje, je vhodné zadat při výpočtu maximální počet kroků. Je-li tento počet překročen a kořen rovnice jsme nenašli, výpočet ukončíme s tím, že metoda diverguje. Pak je nutno změnit počáteční aproximace nebo zvolit jinou metodu.



Obrázek 4.6: Metoda sečen může divergovat.

4.1.4 Newtonova metoda (metoda tečen)

Už sám název metody říká, že budeme pracovat s tečnami ke grafu funkce f . Proto všude v této kapitole budeme předpokládat, že funkce f má derivaci.

Newtonovu metodu můžeme popsat graficky takto:

Zvolíme počáteční aproximaci kořene x_0 . Bodem $[x_0, f(x_0)]$ vedeme tečnu ke grafu funkce f . Její průsečík s osou x označíme x_1 . Pak vedeme tečnu bodem $[x_1, f(x_1)]$, její průsečík s osou x označíme x_2 atd. - viz obrázek 4.7.

Průsečík tečny s osou x vypočteme jako

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (4.5)$$

Výpočet provádíme tak dlouho, dokud není splněna podmínka

$$|x_k - x_{k-1}| < \varepsilon$$

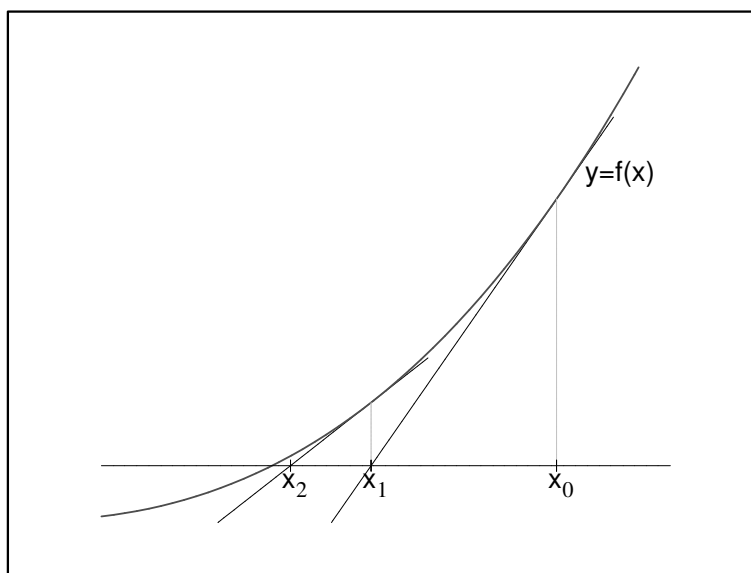
Při splnění této podmínky však nemusí platit $|x_k - \xi| < \varepsilon$.

Newtonovu metodu lze odvodit i pomocí Taylorova vzorce. Ukážeme nyní jak, protože stejný postup později zobecníme i pro soustavu rovnic.

Předpokládejme, že známe k -tou aproximaci řešení x_k . Pak můžeme psát

$$f(\xi) = f(x_k) + f'(x_k)(\xi - x_k) + R,$$

kde R je zbytek v Taylorově vzorci.



Obrázek 4.7: Newtonova metoda

Zanedbáme-li tento zbytek a uvědomíme-li si že $f(\xi) = 0$ (protože ξ je kořenem rovnice $f(x) = 0$), můžeme z předchozí rovnice přibližně vyjádřit kořen ξ jako

$$\xi \doteq x_k - \frac{f(x_k)}{f'(x_k)},$$

což je právě x_{k+1} nalezené dříve popsáním způsobem.

Z Taylorova vzorce lze také odvodit odhady chyby k -té aproximace kořene získané Newtonovou metodou. Má-li funkce na intervalu I obsahujícím x_k i kořen ξ druhou derivaci, platí

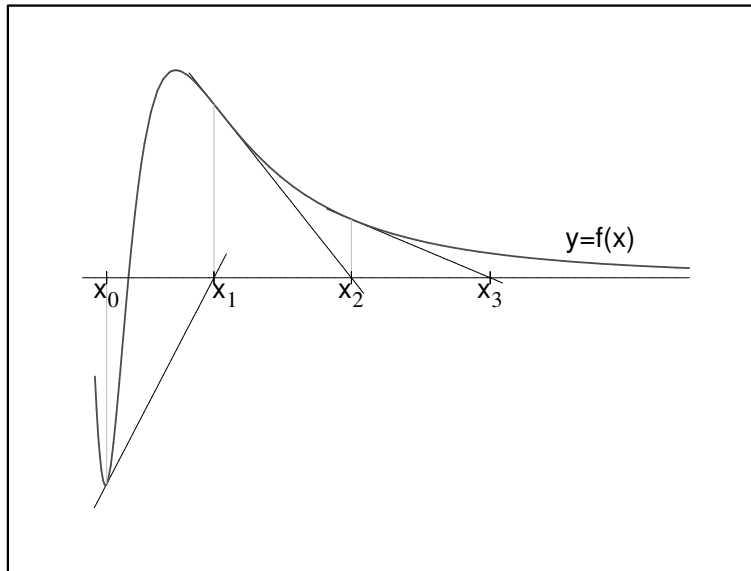
$$|\xi - x_k| \leq \frac{M_2}{2m_1} (x_k - x_{k-1})^2$$

$$|\xi - x_k| \leq \frac{M_2}{2m_1} (\xi - x_{k-1})^2,$$

kde $M_2 = \max |f''(x)|$ a $m_1 = \min |f'(x)|$ pro $x \in I$.

Newtonova metoda je z metod pro řešení nelineárních rovnice nejefektivnější, nemusí však konvergovat - viz obrázek 4.8.

Jestli Newtonova metoda konvergovat bude, nebo nebude, závisí do značné míry také na tom, jak zvolíme počáteční aproximaci x_0 . Při pohledu na obrázek 4.7 je zřejmé, že zde byla počáteční aproximace zvolena vhodně. Kdybychom jako x_0 zvolili např. levý krajní bod zobrazeného intervalu, konvergence už by zaručena (ovšem ani vyloučena) nebyla. Tím se dostáváme k podmínkám, při jejichž splnění bude jisté, že Newtonova metoda konverguje.



Obrázek 4.8: Newtonova metoda může divergovat

Věta 4.2 (*Fourierova podmínka*)

Nechť v intervalu $\langle a, b \rangle$ leží jediný kořen rovnice $f(x) = 0$ a necht' $f'(x)$ a $f''(x)$ jsou spojité a nemění znaménko na intervalu $\langle a, b \rangle$. Zvolíme-li za počáteční aproximaci $x_0 \in \langle a, b \rangle$ tak, aby byla splněna podmínka

$$f(x_0) \cdot f''(x_0) > 0, \quad (4.6)$$

Newtonova metoda bude konvergovat.

Připomeňme v souvislosti s předpoklady věty 4.2 některé poznatky z prvního semestru. To, že $f'(x)$ nemění znaménko na intervalu $\langle a, b \rangle$, znamená, že funkce f buď na celém intervalu $\langle a, b \rangle$ roste, nebo na celém intervalu klesá.

To, že znaménko nemění $f''(x)$, znamená, že funkce f je buď na celém intervalu $\langle a, b \rangle$ konvexní (nad tečnou), nebo je na celém intervalu konkávní (pod tečnou).

Podmínka 4.6 znamená, že ze x_0 vybereme bod, v němž má funkční hodnota stejné znaménko jako druhá derivace.

Funkce, jejíž graf je na obrázku 4.7, je na celém zobrazeném intervalu rostoucí a konvexní. To znamená, že její druhá derivace je na tomto intervalu kladná. Proto se jako počáteční aproximace zvolil bod, v němž byla i funkční hodnota kladná.

Čtenář si může zkusit představit další možné situace, např. funkci na celém intervalu rostoucí a konkávní - zde by se jako x_0 zvolil levý krajní bod - a podobně.

Příklad 4.5 Newtonovou metodou najděte záporný kořen rovnice z příkladu 4.1

$$e^x + x^2 - 3 = 0$$

s přesností $\varepsilon = 0,01$.

Řešení: Víme, že kořen leží v intervalu $\langle -2, -1 \rangle$. Ověříme, že na tomto intervalu jsou splněny předpoklady věty 4.2.

Vypočteme první a druhou derivaci funkce $f(x) = e^x + x^2 - 3$:

$$f'(x) = e^x + 2x \quad , \quad f''(x) = e^x + 2$$

Na celém intervalu $\langle -2, -1 \rangle$ je $f'(x) < 0$ a $f''(x) > 0$ (tzn. ani první, ani druhá derivace zde nemění znaménko).

Nyní vybereme počáteční aproximaci x_0 tak, aby byla splněna podmínka 4.6. Protože $f(-2) = e^{-2} + 1 > 0$ a $f(-1) = e^{-1} - 2 < 0$, zvolíme $x_0 = -2$.

Další aproximace řešení budeme počítat pomocí iteračního vztahu

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{e^{x_k} + x_k^2 - 3}{e^{x_k} + 2x_k}$$

Dostaneme

$$x_0 = -2$$

$$x_1 \doteq -1,70623$$

$$x_2 \doteq -1,67752$$

$$x_3 \doteq -1,67723$$

Nyní můžeme výpočet zastavit, protože $|x_3 - x_2| < 0,01$. Všimněme si, že tři kroky by nám stačily i pro dosažení přesnosti 0,001. Newtonova metoda je obvykle velice rychlá.

Přibližné řešení rovnice je $x_3 \doteq -1,677$.

Nejsme-li schopni ověřit podmínky z věty 4.2, můžeme Newtonovu metodu přesto použít. Pokud tyto podmínky neplatí, Newtonova metoda konvergovat může a nemusí. Proto je při výpočtu vhodné stanovit maximální počet kroků metody a je-li překročen, výpočet ukončit a zvolit jinou počáteční aproximaci, resp. jinou metodu řešení.

4.1.5 Metoda prosté iterace

Metoda prosté iterace pro řešení jedné nelineární rovnice je další aplikací obecné metody postupných aproximací, popsané v kapitole 2.

Rovnici $f(x) = 0$ upravíme na tvar

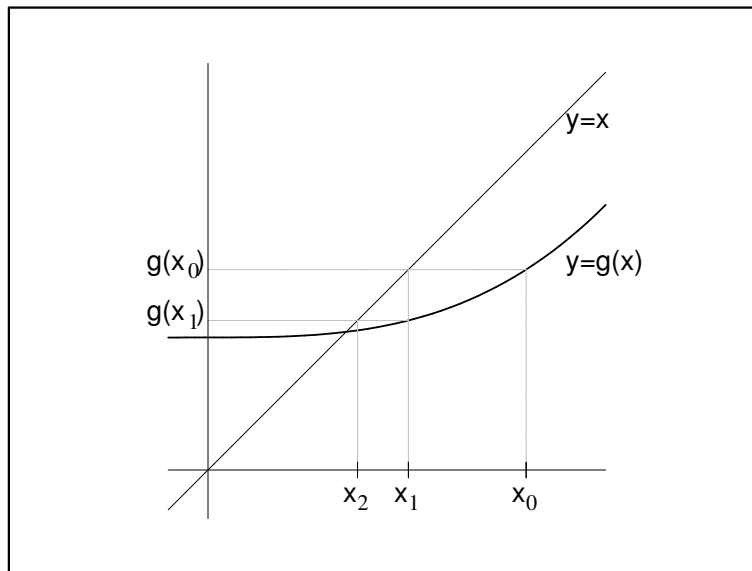
$$x = g(x).$$

Funkce g se nazývá **iterační funkce**. Nyní budeme místo kořene původní rovnice hledat pevný bod funkce $g(x)$. Uděláme to postupem uvedeným v kapitole 2.

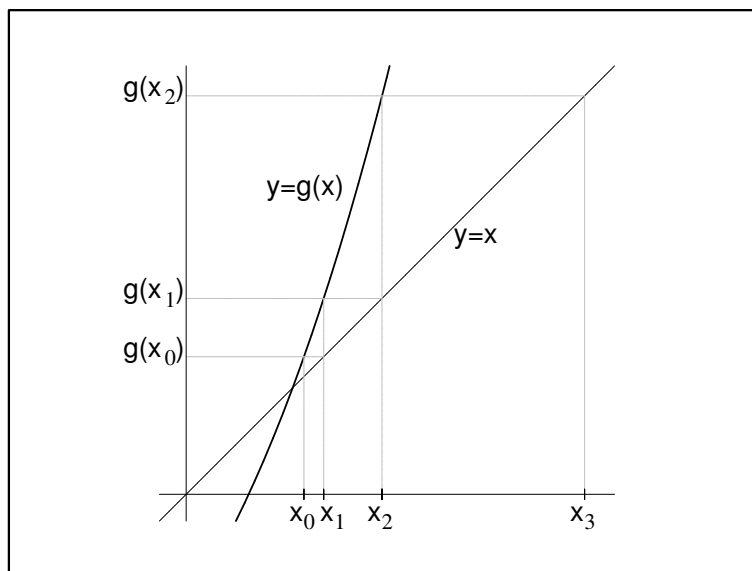
Zvolíme počáteční aproximaci x_0 a další aproximace pevného bodu (neboli řešení původní rovnice) budeme počítat jako

$$x_{k+1} = g(x_k) \tag{4.7}$$

Tímto způsobem můžeme a nemusíme dojít k pevnému bodu funkce g - viz obrázky 4.9 (kde se pevný bod najde) a 4.10 (kde metoda diverguje, i když počáteční aproximace byla pevnému bodu velmi blízko)



Obrázek 4.9: Metoda prosté iterace



Obrázek 4.10: Metoda prosté iterace může divergovat

Nyní řekneme, kdy je zaručeno, že metoda prosté iterace konverguje.

V kapitole 2 jsme se dozvěděli, že metoda postupných aproximací konverguje, je-li zobrazení, jehož pevný bod hledáme, kontraktivní. U funkce jedné proměnné kontraktivita úzce souvisí s rychlostí růstu této funkce - viz obrázky 2.4 a 2.5. Proto platí

Věta 4.3 *Nechť funkce g zobrazuje interval $\langle a, b \rangle$ do sebe a má na tomto intervalu*

derivaci. Jestliže existuje číslo $\alpha \in \langle 0, 1 \rangle$ tak, že

$$|g'(x)| \leq \alpha \quad \forall x \in \langle a, b \rangle, \quad (4.8)$$

pak v intervalu $\langle a, b \rangle$ existuje pevný bod ξ funkce g a posloupnost postupných aproximací získaná předpisem 4.7 k němu konverguje pro libovolnou počáteční aproximaci $x_0 \in \langle a, b \rangle$.

Dále platí

$$|x_k - \xi| \leq \frac{\alpha}{1 - \alpha} |x_k - x_{k-1}| \quad (4.9)$$

$$|x_k - \xi| \leq \frac{\alpha^k}{1 - \alpha} |x_1 - x_0| \quad (4.10)$$

Odhad 4.9 lze použít při rozhodování o zastavení iteračního procesu. Protože však ověření podmínky 4.8 a nalezení α může být obtížné, jako kritérium pro zastavení výpočtu se opět spíše používá podmínka

$$|x_k - x_{k-1}| < \varepsilon$$

(která opět nezaručuje, že $|x_k - \xi| < \varepsilon$).

Také je vhodné stanovit maximální počet kroků a je-li překročen, výpočet ukončit. Pak je potřeba zvolit jinou iterační funkci nebo jinou metodu.

Příklad 4.6 Metodou prosté iterace najděte záporný kořen rovnice z příkladu 4.1

$$e^x + x^2 - 3 = 0$$

s přesností $\varepsilon = 0,01$.

Řešení: Víme, že kořen leží v intervalu $\langle -2, -1 \rangle$. Budeme hledat vhodnou iterační funkci g .

Jedna možnost, jak ze zadané rovnice vyjádřit x , je

$$x^2 = 3 - e^x \quad \Rightarrow \quad x = \pm \sqrt{3 - e^x}.$$

Protože hledáme záporný kořen, je

$$g(x) = -\sqrt{3 - e^x}$$

Ověříme, je-li splněna podmínka 4.8. K tomu je potřeba funkci g zderivovat. Dostaneme

$$g'(x) = \frac{e^x}{2\sqrt{3 - e^x}}.$$

Nyní budeme hledat maximum $|g'(x)|$ na intervalu $\langle -2, -1 \rangle$. Na tomto intervalu je $|g'(x)| = \frac{e^x}{2\sqrt{3 - e^x}}$. Derivace této funkce je $\frac{e^x(6 - e^x)}{4(3 - e^x)^{\frac{3}{2}}}$. To je funkce na intervalu $\langle -2, -1 \rangle$ kladná, tedy $|g'(x)|$ je na tomto intervalu rostoucí a svého maxima nabývá v pravém krajním bodě tohoto intervalu. Hodnota maxima je $|g'(-1)| = \frac{e^{-1}}{2\sqrt{3 - e^{-1}}} \leq 0,12 < 1$. To znamená, že podmínka 4.8 je splněna.

Ještě bychom měli ověřit, že funkce g zobrazuje interval $\langle -2, -1 \rangle$ do sebe. Protože je na tomto intervalu $g'(x) > 0$, je funkce g rostoucí a stačí ověřit, že hodnoty g v krajních bodech intervalu do tohoto intervalu patří. (Kdyby g nebyla monotonní, museli bychom hledat její maximum a minimum na zkoumaném intervalu, nestačilo by dosadit krajní body.)

Protože $g(-2) \doteq -1,69 \in \langle -2, -1 \rangle$ a $g(-1) \doteq -1,62 \in \langle -2, -1 \rangle$, funkce g zobrazuje zkoumaný interval do sebe.

Konvergence iteračního procesu je tedy zaručena.

Můžeme zvolit např. $x_0 = -2$.

Další aproximace pak budeme počítat podle předpisu

$$x_{k+1} = g(x_k) = -\sqrt{3 - e^{x_k}}$$

Dostaneme

$$x_0 = -2$$

$$x_1 \doteq -1,69253$$

$$x_2 \doteq -1,67808$$

$$x_3 \doteq -1,67728$$

Nyní můžeme výpočet zastavit, protože $|x_3 - x_2| < 0,01$. Iterační metoda v tomto případě konverguje docela rychle, protože hodnota $\alpha = 0,12$ je malá. Obecně platí, že čím je derivace funkce g v absolutní hodnotě v okolí pevného bodu menší, tím rychleji metoda prosté iterace konverguje.

Přibližné řešení rovnice je $x_3 \doteq -1,677$

Jiná možnost, jak z rovnice vyjádřit x , je

$$x = \ln(3 - x^2) \quad , \text{ tj. } \quad g(x) = \ln(3 - x^2).$$

V tomto případě by na intervalu $\langle -2, -1 \rangle$ podmínky konvergence splněny nebyly. Podívejme se, jak se budou chovat postupné aproximace, zvolíme-li $x_0 = -1$:

$$x_0 = -1$$

$$x_1 \doteq 0,69315$$

$$x_2 \doteq 0,92408$$

$$x_3 \doteq 0,76364$$

$$x_4 \doteq 0,88247$$

⋮

Nakonec bychom našli kladný kořen rovnice, který již jsme hledali metodou půlení a metodou regula falsi.

Poznámka. Způsobů, jak z rovnice $f(x) = 0$ vyjádřit x , je nekonečně mnoho. Jedna z možností je vydělit rovnicí $f(x) = 0$ derivací funkce f , pak rovnicí vynásobit -1 a nakonec na obě strany přičíst x . Dostaneme

$$x = x - \frac{f(x)}{f'(x)},$$

vztah, který by nám měl být povědomý.

Newtonova metoda je tedy speciálním (a obvykle nejvhodnějším) případem metody prosté iterace.

4.2 Numerické metody řešení soustav nelineárních rovnic

Budeme se zabývat řešením soustavy n nelineárních rovnic o n neznámých

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0 \\ f_2(x_1, x_2, \dots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0 \end{aligned} \tag{4.11}$$

kterou můžeme přepsat vektorově jako

$$\mathbf{F}(\mathbf{x}) = \mathbf{o}, \tag{4.12}$$

kde $\mathbf{F} = (f_1, \dots, f_n)^T$, $\mathbf{x} = (x_1, \dots, x_n)^T$ a \mathbf{o} je nulový vektor.

Přesné řešení této soustavy opět budeme značit $\xi = (\xi_1, \dots, \xi_n)^T$.

Ukážeme zde metodu prosté iterace a Newtonovu metodu. Obě tyto metody vypadají velice podobně jako pro jedinou nelineární rovnici. Ve skutečnosti je ale vícedimenzionální případ mnohem složitější, protože na rozdíl od jediné rovnice je velmi nesnadné získat dobré informace o poloze kořene. Podmínky konvergence obou uvedených metod se také ověřují mnohem obtížněji než u jediné rovnice.

V případě, že řešíme dvě rovnice, hledáme vlastně průsečíky dvou křivek v rovině daných implicitně rovnicemi $f_1(x, y) = 0$ a $f_2(x, y) = 0$ - viz obrázek 4.11

4.2.1 Metoda prosté iterace

Soustavu 4.11 upravíme na tvar

$$\begin{aligned} x_1 &= g_1(x_1, x_2, \dots, x_n) \\ x_2 &= g_2(x_1, x_2, \dots, x_n) \\ &\vdots \\ x_n &= g_n(x_1, x_2, \dots, x_n) \end{aligned} \tag{4.13}$$

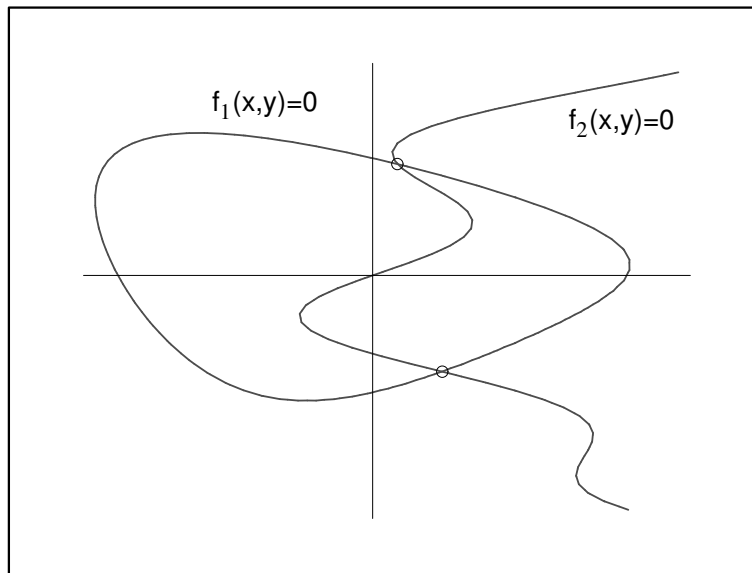
což můžeme zapsat vektorově jako

$$\mathbf{x} = \mathbf{G}(\mathbf{x}), \tag{4.14}$$

kde $\mathbf{G} = (g_1, \dots, g_n)^T$

Podobně jako u jedné rovnice zvolíme počáteční aproximaci $\mathbf{x}^{(0)}$ a počítáme posloupnost postupných aproximací z iteračního vztahu

$$\mathbf{x}^{(k+1)} = \mathbf{G}(\mathbf{x}^{(k)}) \tag{4.15}$$



Obrázek 4.11: Grafický význam řešení dvou nelineárních rovnic

Jsou-li funkce g_1, \dots, g_n diferencovatelné, lze vyslovit podmínky konvergence pro metodu prosté iterace, podobné těm z věty 4.3.

Protože pracujeme s n funkcemi n proměnných, v roli derivace zde bude vystupovat matice

$$\mathbf{G}' = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \dots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \dots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \dots & \frac{\partial g_n}{\partial x_n} \end{pmatrix}$$

Věta 4.4 *Nechť \mathbf{G} zobrazuje uzavřenou oblast D do sebe a je v této oblasti diferencovatelná. Jestliže existuje číslo $\alpha \in (0, 1)$ tak, že*

$$\|\mathbf{G}'\| \leq \alpha \quad \forall \mathbf{x} \in D, \quad (4.16)$$

kde $\|\mathbf{G}'\|$ je řádková nebo sloupcová norma matice \mathbf{G}' , pak v oblasti D existuje pevný bod ξ zobrazení G a posloupnost postupných aproximací získaná předpisem 4.15 k němu konverguje pro libovolnou počáteční aproximaci $\mathbf{x}^{(0)} \in D$.

Pro odhad chyby platí podobné vztahy jako 4.9, 4.10 u jedné rovnice.

Pro zastavení výpočtu se používá kritérium

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| < \varepsilon,$$

kde $\|\cdot\|$ je některá z norem 2.13, 2.12.

Protože ověření podmínek konvergence může být dost problematické, je vhodné předem stanovit maximální počet kroků metody a je-li překročen, výpočet ukončit s tím, že metoda

diverguje. Pak je potřeba zvolit jinou počáteční aproximaci, jiné iterační funkce, nebo jinou metodu.

Poznamenejme, že **najít vhodné iterační funkce může být velmi obtížné.**

4.2.2 Newtonova metoda

Předpokládejme, že již máme aproximaci řešení $\mathbf{x}^{(k)}$.

Podobně jako u diferencovatelné funkce jedné proměnné platilo pro x_k blízké ke kořeni ξ

$$f(\xi) \doteq f(x_k) + f'(x_k)(\xi - x_k),$$

platí pro n -tici diferencovatelných funkcí n proměnných $\mathbf{F} = (f_1, \dots, f_n)^T$

$$\mathbf{F}(\xi) \doteq \mathbf{F}(\mathbf{x}^{(k)}) + \mathbf{F}'(\mathbf{x}^{(k)}) \cdot (\xi - \mathbf{x}^{(k)}),$$

kde

$$\mathbf{F}' = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

a \cdot značí násobení matic.

Uvědomíme-li si, že $\mathbf{F}(\xi) = \mathbf{o}$, můžeme odtud ξ přibližně vyjádřit, čímž získáme jeho další aproximaci $\mathbf{x}^{(k+1)}$. Dostaneme

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{F}'(\mathbf{x}^{(k)}))^{-1} \cdot \mathbf{F}(\mathbf{x}^{(k)}) \quad (4.17)$$

Při výpočtu další aproximace řešení vzorec 4.17 nepoužíváme. Museli bychom počítat inverzní matici, což je velmi pracné, zvláště pro matice velkých rozměrů. Místo toho postupujeme následovně:

Vzorec 4.17 přepíšeme na tvar

$$\mathbf{F}'(\mathbf{x}^{(k)}) \cdot (\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) = -\mathbf{F}(\mathbf{x}^{(k)}).$$

Označíme

$$\delta^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} = (\delta_1^{(k)}, \dots, \delta_n^{(k)})^T \quad (4.18)$$

a vyřešíme soustavu rovnic

$$\mathbf{F}'(\mathbf{x}^{(k)}) \cdot \delta^{(k)} = -\mathbf{F}(\mathbf{x}^{(k)}) \quad (4.19)$$

s neznámými $\delta_1^{(k)}, \dots, \delta_n^{(k)}$.

Řešíme-li dvě rovnice, hodí se pro řešení soustavy 4.19 Cramerovo pravidlo.

Máme-li velký počet rovnic, použijeme některou z dalších metod popsanych v kapitole 3.

Novou aproximaci řešení pak vypočteme z 4.18 jako

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta^{(k)}.$$

Ve výpočtu pokračujeme tak dlouho, dokud není splněna podmínka

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| < \varepsilon, \text{ neboli } \|\delta^{(k-1)}\| < \varepsilon,$$

nebo dokud není překročen předem stanovený maximální počet kroků (v takovém případě je nutno zvolit jinou počáteční aproximaci).

V každém kroku Newtonovy metody musíme vyřešit soustavu lineárních rovnic.

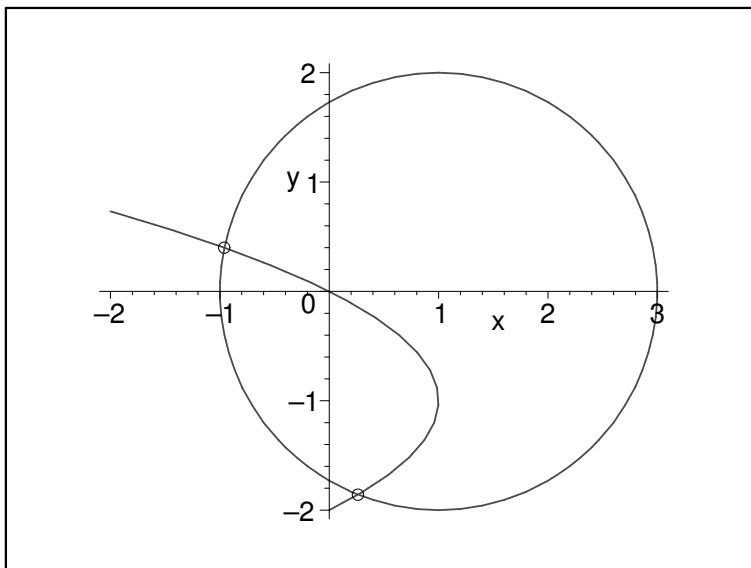
Z toho je vidět, že Newtonova metoda je pracná a časově náročná. Na druhou stranu, začneme-li blízko kořene, konverguje obvykle velmi rychle.

Příklad 4.7 *Newtonovou metodou najděte řešení soustavy rovnic*

$$\begin{aligned} (x-1)^2 + y^2 - 4 &= 0 \\ x + (y+1)^2 - 1 &= 0 \end{aligned}$$

s přesností $\varepsilon = 0,01$.

Řešení: Počet a polohu kořenů můžeme v tomto případě odhadnout graficky. První rovnice je rovnicí kružnice a druhá rovnice je rovnicí paraboly - viz obrázek 4.12.



Obrázek 4.12: K příkladu 4.7 - odhad polohy kořenů.

Vidíme, že soustava má dvě řešení. Budeme hledat např. kořen ležící ve čtvrtém kvadrantu. Jako počáteční aproximaci můžeme zvolit $\mathbf{x}^{(0)} = (0, -2)$.

Dále musíme vypočítat matici parciálních derivací funkcí

$$f_1(x, y) = (x - 1)^2 + y^2 - 4 \quad , \quad f_2(x, y) = x + (y + 1)^2 - 1$$

Dostaneme

$$\mathbf{F}' = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} = \begin{pmatrix} 2(x - 1) & 2y \\ 1 & 2(y + 1) \end{pmatrix}$$

1. krok

Dosadíme bod $\mathbf{x}^{(0)} = (0, -2)$ do matice derivací a do funkcí f_1 a f_2 :

$$\mathbf{F}'(0, -2) = \begin{pmatrix} -2 & -4 \\ 1 & -2 \end{pmatrix} \quad , \quad \mathbf{F}(0, -2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Soustava rovnic pro neznámé δ_1 a δ_2 (horní index, označující krok, pro přehlednost vynecháme, je ale nutno mít na paměti, že v každém kroku budeme počítat jiné δ_1 a δ_2) bude

$$\begin{aligned} -2\delta_1 - 4\delta_2 &= -1 \\ \delta_1 - 2\delta_2 &= 0 \end{aligned}$$

Snadno zjistíme, že řešením této soustavy je $\delta_1 = \frac{1}{4} = 0,25$, $\delta_2 = \frac{1}{8} = 0,125$. Odtud $\mathbf{x}^{(1)} = (0 + 0,25; -2 + 0,125) = (0,25; -1,875)$.

2. krok

$$\mathbf{F}'(0,25; -1,875) = \begin{pmatrix} -1,5 & -3,75 \\ 1 & -1,75 \end{pmatrix} \quad , \quad \mathbf{F}(0,25; -1,875) \doteq \begin{pmatrix} 0,07812 \\ 0,01562 \end{pmatrix}$$

Budeme řešit soustavu

$$\begin{aligned} -1,5\delta_1 - 3,75\delta_2 &= -0,07812 \\ \delta_1 - 1,75\delta_2 &= -0,01562 \end{aligned}$$

Řešení této soustavy můžeme najít pomocí Cramerova pravidla:

$$\delta_1 = \frac{\begin{vmatrix} -0,07812 & -3,75 \\ -0,01562 & -1,75 \end{vmatrix}}{\begin{vmatrix} -1,5 & -3,75 \\ 1 & -1,75 \end{vmatrix}} \doteq 0,01225 \quad , \quad \delta_2 = \frac{\begin{vmatrix} -1,5 & -0,07812 \\ 1 & -0,01562 \end{vmatrix}}{\begin{vmatrix} -1,5 & -3,75 \\ 1 & -1,75 \end{vmatrix}} \doteq 0,01593$$

Odtud $\mathbf{x}^{(2)} = (0,25 + 0,01225; -1,875 + 0,01593) = (0,26225; -1,85907)$.

3. krok

$$\begin{aligned} \mathbf{F}'(0,26225; -1,85907) &\doteq \begin{pmatrix} -1,47549 & -3,71814 \\ 1 & -1,71814 \end{pmatrix} \\ \mathbf{F}(0,26225; -1,85907) &\doteq \begin{pmatrix} 0,00040 \\ 0,00025 \end{pmatrix} \end{aligned}$$

Budeme řešit soustavu

$$\begin{aligned} -1,47549 \delta_1 - 3,71814 \delta_2 &= -0,00040 \\ \delta_1 - 1,71814 \delta_2 &= -0,00025 \end{aligned}$$

Řešení této soustavy pomocí Cramerova pravidla:

$$\delta_1 = \frac{\begin{vmatrix} -0,00040 & -3,71814 \\ -0,00025 & -1,71814 \end{vmatrix}}{\begin{vmatrix} -1,47549 & -3,71814 \\ 1 & -1,71814 \end{vmatrix}} \doteq -0,00004, \quad \delta_2 = \frac{\begin{vmatrix} -1,47549 & -0,00040 \\ 1 & -0,00025 \end{vmatrix}}{\begin{vmatrix} -1,47549 & -3,71814 \\ 1 & -1,71814 \end{vmatrix}} \doteq 0,00012$$

Odtud $\mathbf{x}^{(3)} = (0,26221; -1,85894)$.

Protože $|\delta_1| < 0,01$ i $|\delta_2| < 0,01$ (tj. $\|\delta\|_\infty < 0,01$), můžeme výpočet ukončit. Přibližné řešení je $(0,262; -1,859)$.

5 Aproximace funkcí

Čtenář se již určitě mnohokrát setkal s různými funkcemi a s výpočtem jejich hodnot. U některých funkcí se funkční hodnota vypočítá snadno, u jiných by to člověk „ručně“ nezvládl a musí použít kalkulačku. Některé funkce jsou zadány tak složitým předpisem (viz část o statistice), že jejich hodnoty je jednodušší nalézt v tabulce, než je počítat. Někdy též máme funkci, která není zadána vůbec žádným předpisem, ale známe pouze její hodnoty v určitých bodech, např. získané nějakým měřením. Naskytá se otázka, jak zjistit hodnotu takové funkce v netabulkovém bodě, jak vypočítat hodnotu její derivace v určitém bodě nebo jak ji zintegrovat.

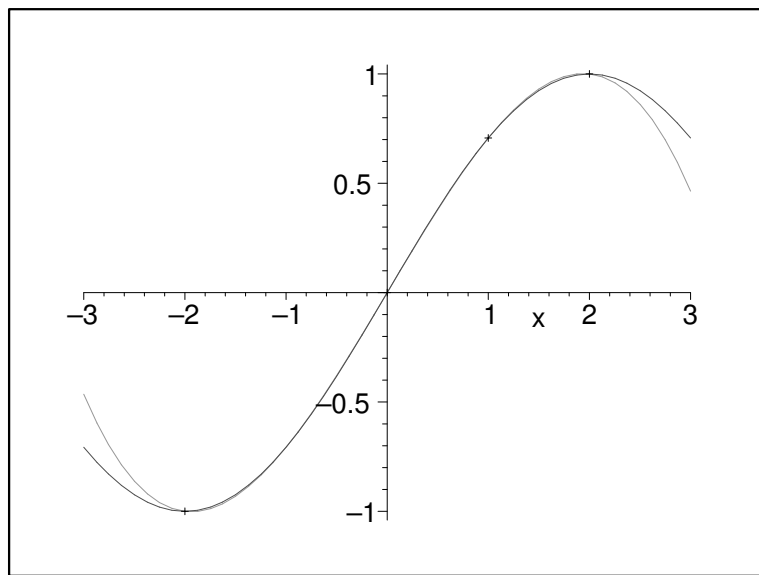
Řešením je nahradit zkoumanou funkci funkcí jinou, která se jí jakýmsi způsobem podobá a se kterou se lépe pracuje.

Nejčastěji touto „náhradní“ funkcí bývá algebraický polynom, protože v tomto případě jsou všechny výše uvedené výpočty skutečně velmi jednoduché.

Požadavky, podle nichž vybíráme onu náhradní funkci, mohou být různé. Zde si blíže všimneme **interpolace**, kde se požaduje, aby aproximující funkce měla s funkcí původní v určitých bodech stejné hodnoty a **metody nejmenších čtverců**, kde má aproximující funkce procházet zadaným bodům v jistém smyslu nejbliže, ale přímo jimi procházet nemusí.

5.1 Interpolace algebraickými polynomy

Při interpolaci zní základní úloha takto: Máme $n+1$ navzájem různých bodů x_0, x_1, \dots, x_n , kterým říkáme uzlové body nebo uzly interpolace a dále funkční hodnoty v těchto bodech $f_0 = f(x_0), f_1 = f(x_1), \dots, f_n = f(x_n)$. Hledáme polynom $P_n(x)$ stupně nejvýše n takový, že v uzlových bodech nabývá týchž hodnot jako funkce f , tj. $P(x_i) = f_i, \quad i = 0, \dots, n$.



Obrázek 5.1: Funkce a interpolační polynom

Poznámka. Někdy se též hledá polynom, který má se zadanou funkcí nejen stejné funkční hodnoty v uzlových bodech, ale i stejné hodnoty derivací až do určitého řádu.

5.1.1 Existence a jednoznačnost interpolačního polynomu

Věta 5.1 *Nechť jsou dány body $[x_i, f_i], i = 0, \dots, n$. Pak existuje právě jeden polynom P_n stupně nanejvýš n takový, že $P_n(x_i) = f_i, i = 0, \dots, n$.*

Důkaz. Existenci interpolačního polynomu dokážeme tím způsobem, že předvedeme postup, kterým jej lze pro libovolné navzájem různé uzlové body zkonstruovat. Tomu bude věnován další odstavec této kapitoly.

To, že interpolační polynom procházející danými body existuje právě jeden, dokážeme sporem. Předpokládejme, že existují dva polynomy stupně nanejvýš n , označme je $P_n(x)$ a $R_n(x)$ takové, že $P_n(x_i) = f_i, i = 0, \dots, n$ i $R_n(x_i) = f_i, i = 0, \dots, n$. Ukážeme, že tyto dva polynomy jsou shodné. Za tím účelem označme $Q_n(x) = P_n(x) - R_n(x)$. Je vidět, že $Q_n(x)$ je opět polynom stupně nejvýše n a navíc $Q_n(x_i) = 0, i = 0, \dots, n$. Máme tedy polynom stupně nejvýše n , který má $n + 1$ kořenů. To je možné jedině tak, že $Q_n(x)$ je identicky roven nule, $Q_n(x) \equiv 0$ a tedy $P_n(x) \equiv R_n(x) \forall x \in \mathbb{R}$

5.1.2 Konstrukce interpolačního polynomu, Lagrangeův interpolační polynom

Interpolační polynom daný body $[x_i, f_i], i = 0, \dots, n$ sestavíme pomocí polynomů $l_i(x)$ takových, že

$$l_i(x_j) = \begin{cases} 1 & \text{pro } i = j \\ 0 & \text{pro } i \neq j \end{cases}$$

Čtenář snadno ověří, že polynom

$$l_0(x) = \frac{(x - x_1)(x - x_2) \dots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \dots (x_0 - x_n)}$$

má v x_0 hodnotu 1 a v ostatních uzlových bodech hodnotu 0.

Podobně dostaneme i ostatní polynomy $l_i, i = 0, \dots, n$:

$$l_i(x) = \frac{(x - x_0) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_0 - x_1)(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}$$

Interpolační polynom $L_n(x)$ nyní dostaneme snadno jako kombinaci $l_i(x)$:

$$\begin{aligned} L_n(x) &= f_0 l_0(x) + f_1 l_1(x) + \dots + f_n l_n(x) = & (5.1) \\ &= f_0 \frac{(x - x_1)(x - x_2) \dots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \dots (x_0 - x_n)} + f_1 \frac{(x - x_0)(x - x_2) \dots (x - x_n)}{(x_1 - x_0)(x_1 - x_2) \dots (x_1 - x_n)} + \dots \\ &\quad \dots + f_n \frac{(x - x_0)(x - x_1) \dots (x - x_{n-1})}{(x_n - x_0)(x_n - x_1) \dots (x_n - x_{n-1})} \end{aligned}$$

Interpolační polynom ve tvaru 5.1 se nazývá **Lagrangeův interpolační polynom**.

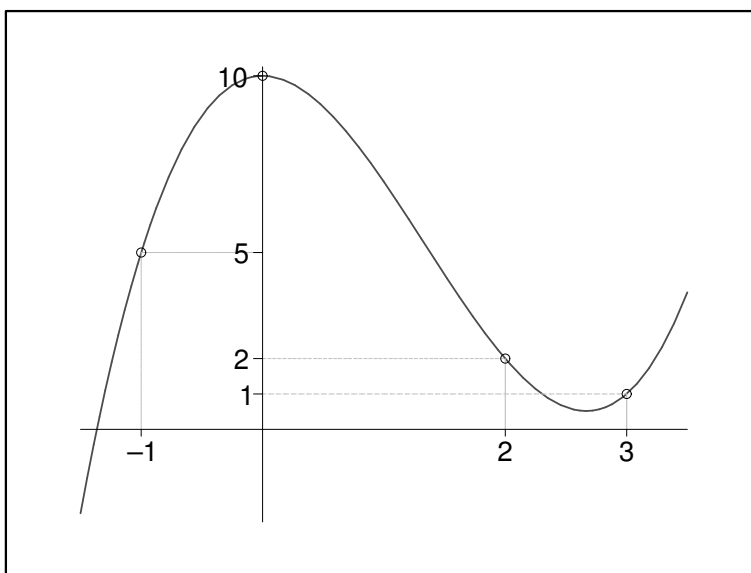
Příklad 5.1 Najděte Lagrangeův interpolační polynom daný body

x_i	-1	0	2	3
f_i	5	10	2	1

Řešení: Máme zadány 4 body, interpolační polynom bude tedy stupně nejvýše třetího. Pro jeho konstrukci použijeme vzorec 5.1:

$$L_3(x) = 5 \frac{(x-0)(x-2)(x-3)}{(-1-0)(-1-2)(-1-3)} + 10 \frac{(x-(-1))(x-2)(x-3)}{(0-(-1))(0-2)(0-3)} + 2 \frac{(x-(-1))(x-0)(x-3)}{(2-(-1))(2-0)(2-3)} + 1 \frac{(x-(-1))(x-0)(x-2)}{(3-(-1))(3-0)(3-2)} = x^3 - 4x^2 + 10$$

Výsledný interpolační polynom je spolu se zadanými body znázorněn na obrázku 5.2.



Obrázek 5.2: K příkladu 5.1: Zadané body a výsledný interpolační polynom

5.1.3 Newtonův interpolační polynom

Interpolační polynom v Lagrangeově tvaru má tu nevýhodu, že chceme-li přidat další uzlový bod, musíme celý polynom přepočítat znovu. Také výpočet hodnoty tohoto polynomu v určitém bodě je dosti pracný. Proto je někdy výhodnější hledat interpolační polynom v jiném tvaru než 5.1. Jako vhodný se ukazuje tvar

$$N_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1) \dots (x - x_{n-1}) \quad (5.2)$$

Koeficienty a_0, a_1, \dots, a_n lze získat řešením soustavy rovnic vzniklé rozepsáním podmínek $N_n(x_i) = f(x_i)$, $i = 0, 1, \dots, n$, ale přehlednější a méně pracné je vypočítat je pomocí takzvaných **poměrných diferencí**.

Pro danou funkci f a uzlové body x_i , $i = 0, \dots, n$ nazveme podíly

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}, \quad i = 0, 1, \dots, n-1$$

poměrnými diferencemi prvního řádu

Pomocí poměrných diferencí prvního řádu definujeme poměrné difference druhého řádu jako

$$f[x_i, x_{i+1}, x_{i+2}] = \frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i}, \quad i = 0, 1, \dots, n-2$$

a obecně **poměrné difference k-tého řádu** pro $k \leq n$ definujeme takto:

$$f[x_i, x_{i+1}, \dots, x_{i+k}] = \frac{f[x_{i+1}, x_{i+2}, \dots, x_{i+k}] - f[x_i, x_{i+1}, \dots, x_{i+k-1}]}{x_{i+k} - x_i}, \quad i = 0, \dots, n-k.$$

Dá se dokázat, že pro koeficienty a_i , $i = 0, 1, \dots, n$ v 5.2 platí

$$\begin{aligned} a_0 &= f(x_0) \\ a_1 &= f[x_0, x_1] \\ a_2 &= f[x_0, x_1, x_2] \\ &\vdots \\ a_n &= f[x_0, x_1, \dots, x_n] \end{aligned}$$

Dosazením těchto hodnot do 5.2 dostaneme **Newtonův interpolační polynom**

$$\begin{aligned} N_n(x) &= f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots \\ &\quad \dots + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1) \dots (x - x_{n-1}) \end{aligned} \quad (5.3)$$

Poznámka. Jestliže vzdálenosti mezi sousedními uzlovými body jsou konstantní, tj. platí-li $x_{i+1} - x_i = h$ pro všechna $i = 1, \dots, n$, kde $h \in \mathbb{R}$ je konstanta, lze odvodit jiný, jednodušší tvar Newtonova i Lagrangeova interpolačního polynomu, viz např. [2]. Uzlovým bodům s touto vlastností říkáme **ekvidistantní**.

Příklad 5.2 *Aproximujte funkci $f(x) = \frac{1}{x}$ Newtonovým interpolačním polynomem v uzlech*

x_i	1	2	2,5	3,2	4
-------	---	---	-----	-----	---

a pak pomocí něj vypočtete přibližnou hodnotu funkce f v bodech $x = 3$ a $x = 10$.

Řešení: Abychom mohli sestavit Newtonův interpolační polynom, musíme vypočítat poměrné difference funkce f až do řádu 4. Budeme je postupně, po sloupcích, zapisovat do tabulky. Podtržené hodnoty pak použijeme pro interpolační polynom.

i	x_i	$f(x_i) = \frac{1}{x_i}$	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_i, \dots, x_{i+3}]$	$f[x_0, \dots, x_4]$
0	1	<u>1</u>	<u>-0,5</u>	<u>0,2</u>	<u>-0,0625</u>	<u>0,015625</u>
1	2	0,5	-0,2	0,0625	-0,015625	
2	2,5	0,4	-0,125	0,03125		
3	3,2	0,3125	-0,078125			
4	4	0,25				

Nyní dosadíme do vzorce 5.3

$$N_4(x) = 1 - 0,5(x-1) + 0,2(x-1)(x-2) - 0,0625(x-1)(x-2)(x-2,5) + \\ + 0,015625(x-1)(x-2)(x-2,5)(x-3,2)$$

Přibližnou hodnotu funkce f v bodě $x = 3$ vypočteme dosazením do interpolačního polynomu $N_4(x)$. Pro výpočet funkčních hodnot interpolačního polynomu v Newtonově tvaru je vhodné si tento polynom poněkud upravit. Můžeme vytknout $(x-1)$, pak ve zbytku $(x-2)$ a tak dále, až nakonec dostaneme

$$N_4(x) = 1 - (x-1) \left(-0,5 + (x-2) \left(0,2 + (x-2,5) \left(-0,0625 + (x-3,2) 0,015625 \right) \right) \right)$$

Dosazovat se hodí „zevnitř“.

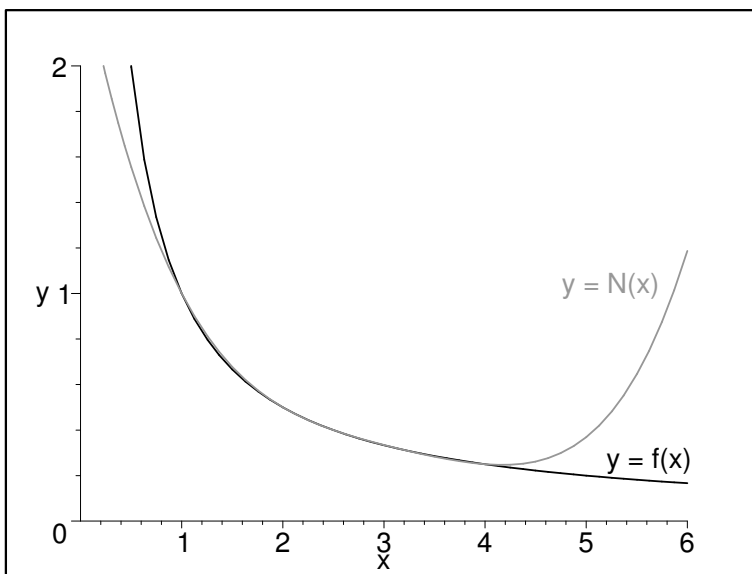
Při použití tohoto tvaru se značně sníží počet výpočetních operací nutných pro získání výsledku. Je-li čtenář obeznámen s Hornerovým schématem, možná najde jistou podobnost s tímto postupem.

V našem případě dostaneme $N_4(3) \doteq 0,334$, zatímco přesná hodnota je $\frac{1}{3} \doteq 0,333$.

Pro $x = 10$ vyjde $P_4(10) = 34,525$, zatímco přesná hodnota je $\frac{1}{10} = 0,1$.

Vidíme, že v bodě, který byl zhruba uprostřed uzlových bodů, je aproximace dobrá, hodnoty interpolačního polynomu a zadané funkce jsou blízké. Naopak v bodě, který leží daleko vně intervalu $\langle 1, 4 \rangle$, je aproximace velmi špatná.

Situace je dobře patrná z obrázku 5.3, kde je vykreslen graf funkce f spolu s vypočteným interpolačním polynomem. Můžeme si všimnout, že na intervalu $\langle 1, 4 \rangle$ interp. polynom dobře vystihuje chování funkce f , ale mimo tento interval se od sebe hodnoty funkce f a interpolačního polynomu značně liší.



Obrázek 5.3: K příkladu 5.2: Srovnání funkce a interpolačního polynomu

Poznámka. Bod $x = 10$ ležel vně intervalu ohraničeného nejmenším a největším uzlovým bodem. V takovém případě mluvíme o **extrapolaci**. Obecně je extrapolaci vhodné používat pouze v bodech blízkých nejmenšímu nebo největšímu uzlovému bodu. O tom, čím je způsobena velká odchylka funkce a interpolačního polynomu v bodech vzdálených od uzlových bodů a jakou přesnost lze při interpolaci očekávat, pojednává další odstavec.

5.1.4 Odhad chyby

Věta 5.2 *Nechť interval I obsahuje body x_0, x_1, \dots, x_n a nechť f je $(n+1)$ -krát diferencovatelná funkce na I . Nechť $P_n(x)$ je interpolační polynom n -tého stupně určený hodnotami funkce f v bodech x_0, \dots, x_n . Potom pro libovolné $x \in I$ existuje $\xi \in I$ takové, že pro chybu interpolace $E(x)$ platí*

$$E(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-x_0)(x-x_1)\dots(x-x_n). \quad (5.4)$$

Důkaz není úplně jednoduchý a lze jej nalézt např. v [2].

Bod ξ je pro každé $x \in I$ jiný a jeho nalezení je prakticky nemožné. Chybu interpolace však můžeme alespoň shora odhadnout:

Označíme-li $M_{n+1} = \max_{t \in I} |f^{(n+1)}(t)|$, platí

$$|E(x)| = |f(x) - P_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |(x-x_0)(x-x_1)\dots(x-x_n)| \quad (5.5)$$

Najít veličinu M_{n+1} však také nemusí být zrovna jednoduché.

Poznámka. Odhad 5.5 lze použít např. v případě, kdy chceme sestavit tabulku hodnot nějaké funkce $f(x)$ s konstantním krokem mezi hodnotami x a ptáme se, jak tento krok zvolit, aby chyba při interpolaci polynomem daného stupně n nepřevýšila dané ε .

Příklad 5.3 *Odhadněte chybu interpolace z příkladu 5.2 v bodě $x = 3$.*

Poznámka. Tento příklad slouží spíše k ozřejvení jednotlivých veličin ve vzorci 5.5 a jako ukázka, že vzorec „funguje“, protože v tomto případě můžeme určit i přesnou hodnotu chyby a nemusíme nic odhadovat.

Řešení: Pro odhad chyby potřebujeme vypočítat pátou derivaci interpolované funkce $f(x) = \frac{1}{x}$ (protože n je v tomto případě 4) a najít maximum její absolutní hodnoty na intervalu $I = \langle 1, 4 \rangle$ (I je nejmenší interval obsahující všechny uzlové body a bod, v němž chceme odhadovat chybu).

$$\text{Vyjde } f^{(5)}(x) = -\frac{120}{x^6}$$

Je vidět, že $|f^{(5)}(x)| = \frac{120}{x^6}$, což je funkce na I klesající. Svého maxima na tomto intervalu proto dosahuje v bodě $x = 1$ a jeho hodnota je $M_5 = \frac{120}{1^6} = 120$.

Nyní dosadíme do 5.4:

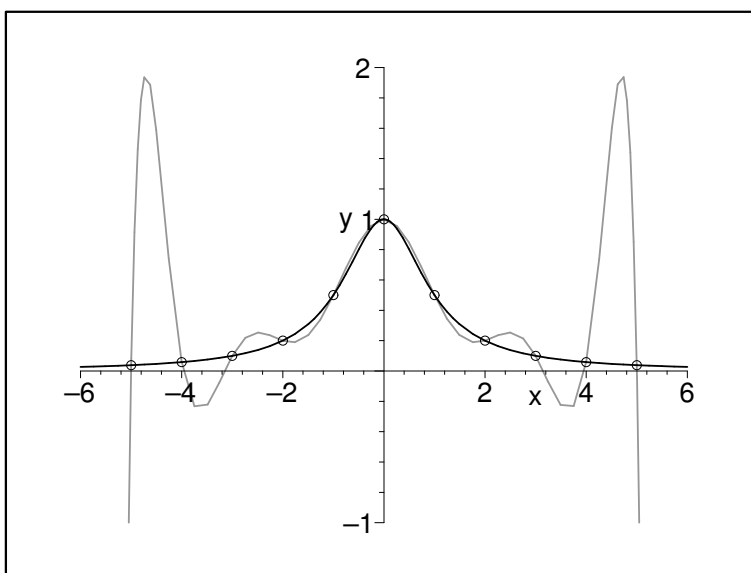
$$|E(3)| \leq \frac{120}{5!} |(3-1)(3-2)(3-2,5)(3-3,2)(3-4)| = |2 \cdot 1 \cdot 0,5 \cdot (-0,2) \cdot (-1)| = 0,2$$

Odhad chyby je v tomto případě dosti nadsazený, chyba v bodě $x = 3$ je ve skutečnosti mnohem menší než 0,2, viz řešení příkladu 5.2

To, že teoretický odhad chyby je příliš pesimistický, je poměrně časté i u jiných metod.

V bodech vzdálených uzlovým bodům nabývá výraz $(x - x_0)(x - x_1) \dots (x - x_n)$, který se vyskytuje v odhadu chyby, velkých hodnot. Proto se interpolační polynom pro výpočet přibližných hodnot funkce v takovýchto bodech nehodí.

Aproximace ale v některých případech nemusí být dobrá ani v bodech relativně blízkých uzlovým bodům. To ilustruje obrázek 5.4, na němž je graf funkce $f(x) = \frac{1}{1+x^2}$ a interpolační polynom daný vyznačenými uzlovými body.



Obrázek 5.4: Nevhodná aproximace interpolačním polynomem

Situace by se příliš nezlepšila, ani kdybychom přidali více uzlových bodů.

Zde je velká odchylka funkce a polynomu taktéž způsobena velkými hodnotami součinu $(x - x_0)(x - x_1) \dots (x - x_n)$, především poblíž konců interpolačního intervalu.

Proto je někdy vhodné nenahrazovat funkci, zvláště chceme-li ji aproximovat na delším intervalu, jedním interpolačním polynomem, ale interval rozdělit na malé části a na každé z nich funkci nahradit polynomem nízkého stupně. To bude námětem následující kapitoly.

5.2 Interpolace pomocí splajnů

Základní myšlenka interpolace pomocí splajnů je obdobná jako u Lagrangeovy interpolace. Máme zadány uzlové body $a = x_0 < x_1 < \dots < x_n = b$ a funkční hodnoty v nich, které označíme f_0, f_1, \dots, f_n . Stejně jako předtím hledáme funkci $S(x)$ takovou, že platí $S(x_i) = f_i$, $i = 0, 1, \dots, n$, ale tentokrát je funkce $S(x)$ po částech polynom (obecně na každém intervalu $\langle x_i, x_{i+1} \rangle$, $i = 0, 1, \dots, n-1$ jiný) a splňuje určité požadavky hladkosti (tj. spojitosti derivací).

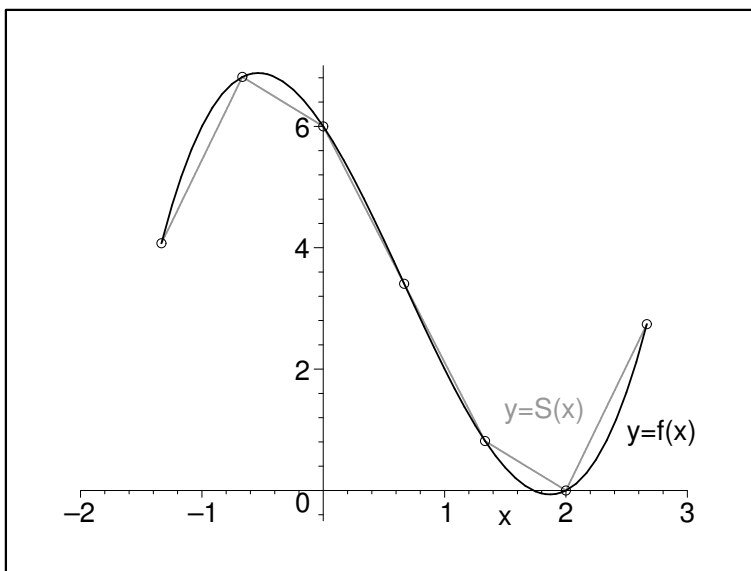
Konkrétně **splajnem řádu k** pro uzly $a = x_0 < x_1 < \dots < x_n = b$ rozumíme funkci, která je v každém intervalu $\langle x_i, x_{i+1} \rangle$, $i = 0, \dots, n-1$, polynom stupně k a která má v celém intervalu $\langle a, b \rangle$ spojité derivace až do řádu $k-1$ včetně.

Poznámka. Slovo „splajn“ pochází z anglického „spline“, což znamená pružné konstruktérské pravítko. V české literatuře se někdy píše splajn a někdy spline.

Nejjednodušším příkladem je splajn řádu 1, **lineární splajn**. Funkce je na každém subintervalu $\langle x_i, x_{i+1} \rangle$, $i = 0, \dots, n-1$, nahrazena úsečkou, jejíž rovnice je

$$S_i(x) = f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} (x - x_i), \quad x \in \langle x_i, x_{i+1} \rangle$$

U splajnu 1. řádu požadujeme spojitost derivací do řádu 0 včetně, tj. spojitost samotné funkce $S(x)$. Snadno se přesvědčíme, že hodnoty jednotlivých funkcí $S_i(x)$ v krajních bodech příslušného intervalu $\langle x_i, x_{i+1} \rangle$ jsou rovny $f(x_i)$, resp. $f(x_{i+1})$, čímž je zaručeno, že na sebe tyto funkce v uzlových bodech spojitě navazují (viz obrázek 5.5). Zlepšení aproximace dosáhneme zjemněním intervalů mezi uzlovými body.



Obrázek 5.5: Nahrazení funkce lineárním splajnem

Nejčastěji užívané jsou tzv. **kubické splajny**, kdy $k=3$.

Definice a konstrukce kubického splajnu

Kubický splajn pro funkci f s uzlovými body x_0, x_1, \dots, x_n je funkce $S(x)$, která je kubický polynom označený $S_i(x)$ na každém subintervalu $\langle x_i, x_{i+1} \rangle$, $i = 0, 1, \dots, n-1$, vyhovuje podmínkám

$$S_i(x_i) = f(x_i), \quad i = 0, \dots, n \quad (5.6)$$

$$S_i(x_{i+1}) = S_{i+1}(x_{i+1}), \quad i = 0, \dots, n-2 \quad (5.7)$$

$$S'_i(x_{i+1}) = S'_{i+1}(x_{i+1}), \quad i = 0, \dots, n-2 \quad (5.8)$$

$$S''_i(x_{i+1}) = S''_{i+1}(x_{i+1}), \quad i = 0, \dots, n-2 \quad (5.9)$$

a okrajovým podmínkám a), b) nebo c)

$$\text{a) } S''(x_0) = S''(x_n) = 0$$

$$\text{b) } S''(x_0) = f''_0, \quad S''(x_n) = f''_n$$

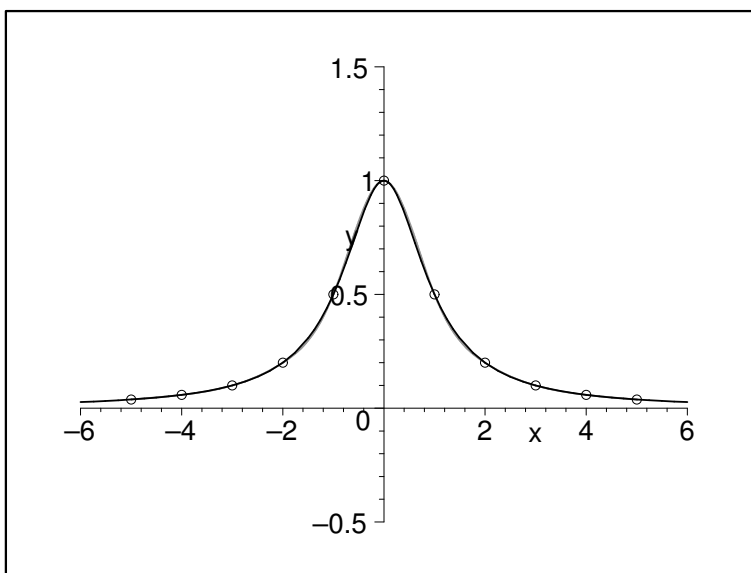
$$\text{c) } S'(x_0) = f'_0, \quad S'(x_n) = f'_n$$

(f''_0, f''_n, f'_0 a f'_n jsou předem zadané konstanty).

Podmínky 5.7 znamenají spojitost funkce S v uzlových bodech, podmínky 5.8 a 5.9 spojitost prvních, resp. druhých derivací.

Kubický splajn splňující okrajové podmínky a) se nazývá **přirozený kubický splajn**.

Na obrázku 5.6 je znázorněna aproximace funkce $f(x) = \frac{1}{1+x^2}$ pomocí přirozeného kubického splajnu. Můžeme porovnat s obrázkem 5.4, kde byla tatáž funkce nahrazena interpolačním polynomem daným stejnými uzlovými body.



Obrázek 5.6: Nahrazení funkce $f(x) = \frac{1}{1+x^2}$ přirozeným kubickým splajnem.

Nyní se budeme zabývat problémem, jak k zadaným uzlovým bodům a hodnotám funkce v nich sestavit přirozený kubický splajn. (Splajn vyhovující jiným okrajovým podmínkám by se našel podobně.)

Na jednotlivých intervalech $\langle x_i, x_{i+1} \rangle$, $i = 0, 1, \dots, n-1$, budeme splajn hledat ve tvaru

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$

Z podmínek 5.6 dostaneme $a_i = f(x_i), i = 0, 1, \dots, n-1$. Odtud, z podmínek 5.7, 5.8, 5.9 a z okrajových podmínek $S_0''(x_0) = S_{n-1}''(x_n) = 0$ lze po jistém úsilí odvodit soustavu rovnic s neznámými $c_i, i = 0, \dots, n$

$$h_{i-1}c_{i-1} + 2(h_{i-1} + h_i)c_i + h_i c_{i+1} = 3 \left(\frac{\Delta f_i}{h_i} - \frac{\Delta f_{i-1}}{h_{i-1}} \right), \quad i = 1, \dots, n-1 \quad (5.10)$$

$$c_0 = c_n = 0$$

kde $h_i = x_{i+1} - x_i$ a $\Delta f_i = f(x_{i+1}) - f(x_i), i = 0, \dots, n-1$.

Po rozepsání a dosazení za c_0 a c_n soustava vypadá takto:

$$\begin{array}{rcl} 2(h_0 + h_1)c_1 & + & h_1c_2 & = & 3\left(\frac{\Delta f_1}{h_1} - \frac{\Delta f_0}{h_0}\right) \\ h_1c_1 & + & 2(h_1 + h_2)c_2 & + & h_2c_3 & = & 3\left(\frac{\Delta f_2}{h_2} - \frac{\Delta f_1}{h_1}\right) \\ & & & & \ddots & & \vdots \\ & & h_{n-2}c_{n-2} & + & 2(h_{n-2} + h_{n-1})c_{n-1} & = & 3\left(\frac{\Delta f_{n-1}}{h_{n-1}} - \frac{\Delta f_{n-2}}{h_{n-2}}\right) \end{array} \quad (5.11)$$

Jedná se o třídiagonální soustavu rovnic a lze ji vyřešit např. pomocí Gaussovy eliminační metody přizpůsobené pro třídiagonální soustavu.

Koeficienty b_i a d_i pak dopočítáme pomocí c_i ze vztahů (také odvozených z podmínek 5.6 – 5.9)

$$b_i = \frac{f(x_{i+1}) - f(x_i)}{h_i} - \frac{c_{i+1} + 2c_i}{3} h_i \quad i = 0, \dots, n-1 \quad (5.12)$$

$$d_i = \frac{c_{i+1} - c_i}{3h_i} \quad i = 0, \dots, n-1 \quad (5.13)$$

Příklad 5.4 Funkci $f(x) = \sqrt{x}$ aproximujte přirozeným kubickým splajnem s uzlovými body

x_i	1	1,69	2,25	2,89	4
-------	---	------	------	------	---

 a pak pomocí tohoto splajnu vypočtěte přibližně hodnotu $f(2)$.

Řešení: Dopočítáme funkční hodnoty v uzlových bodech a pak vypočteme $h_i, i = 0, 1, 2, 3$, tj. délky jednotlivých intervalů a $\Delta f_i, i = 0, 1, 2, 3$. Vypočtené hodnoty jsou zapsány v následující tabulce

i	0	1	2	3	4
x_i	1	1,69	2,25	2,89	4
$f(x_i) = \sqrt{x_i}$	1	1,3	1,5	1,7	2
h_i	0,69	0,56	0,64	1,11	
Δf_i	0,3	0,2	0,2	0,3	

Víme, že $c_0 = 0$. Pro neznámé c_1, c_2, c_3 dostaneme podle 5.11 soustavu rovnic

$$\begin{array}{rcl} 2,5c_1 & + & 0,56c_2 & = & -0,232919 \\ 0,56c_1 & + & 2,4c_2 & + & 0,64c_3 & = & -0,133929 \\ & & 0,64c_2 & + & 3,5c_3 & = & -0,126689 \end{array}$$

Řešením této soustavy je $c_1 = -0,087085, c_2 = -0,027155, c_3 = -0,031231$.

Koeficienty b_i a $d_i, i = 0, 1, 2, 3$, dopočítáme podle vzorců 5.12 a 5.13.

Tedy např. $b_0 = \frac{1,3 - 1}{0,69} - \frac{-0,087085 + 2 \cdot 0}{3} \cdot 0,69 \doteq 0,454812$

Ostatní koeficienty by se vypočítaly podobně. Vyjde:

i	0	1	2	3
a_i	1	1,3	1,5	1,7
b_i	0,454812	0,394724	0,330749	0,293381
c_i	0	-0,087085	-0,027155	-0,031231
d_i	-0,042070	0,035672	-0,002123	0,009379

Výsledný přirozený kubický splajn je tedy

$$S(x) = \begin{cases} S_0(x) = 1 + 0,454812(x-1) - 0,042070(x-1)^3 & x \in \langle 1; 1,69 \rangle \\ S_1(x) = 1,3 + 0,394724(x-1,69) - 0,087085(x-1,69)^2 + 0,035672(x-1,69)^3 & x \in \langle 1,69; 2,25 \rangle \\ S_2(x) = 1,5 + 0,330749(x-2,25) - 0,027155(x-2,25)^2 - 0,002123(x-2,25)^3 & x \in \langle 2,25; 2,89 \rangle \\ S_3(x) = 1,7 + 0,293381(x-2,89) - 0,031231(x-2,89)^2 + 0,009379(x-2,89)^3 & x \in \langle 2,89; 4 \rangle \end{cases}$$

Přibližnou hodnotu funkce f v bodě $x = 2$ nyní vypočteme jako $S_1(2) \doteq 1,415058$ (protože $2 \in \langle 1,69; 2,25 \rangle$). Pro srovnání, přesná hodnota je $\sqrt{2} \doteq 1,414214$.

5.3 Metoda nejmenších čtverců

V předchozích částech této kapitoly jsme požadovali, aby interpolační polynom, resp. splajn nabýval v uzlových bodech stejných hodnot jako funkce, již se snažíme aproximovat. V případě, že jsou funkční hodnoty získány experimentálně, např. jako výsledky nějakého měření, je interpolace nevhodná. Výsledky jsou totiž zatíženy chybami a interpolační funkce by tyto chyby kopírovala, což je přesně to, čeho se chceme vyvarovat. Kromě toho povaha experimentů nevyklučuje možnost několika měření při nezměněné hodnotě x , tj. nemusí být všechny uzlové body navzájem různé. Vzhledem k těmto okolnostem není dobré požadovat, aby aproximační funkce nabývala v uzlových bodech předem daných hodnot. V mnoha případech máme určitou představu o povaze funkce, jejíž hodnoty jsme naměřili, např. může se jednat o lineární nebo kvadratickou závislost. Pak hledáme mezi všemi funkcemi tohoto známého typu takovou, která prochází k zadaným bodům v jistém smyslu nejlépe.

Formulace problému

Jsou dány body x_i , $i = 0, \dots, n$ a funkční hodnoty v nich y_i . Dále jsou dány funkce φ_i , $i = 0, \dots, m$, $m < n$. Mezi všemi funkcemi tvaru

$$P_m(x) = c_0\varphi_0(x) + c_1\varphi_1(x) + \dots + c_m\varphi_m(x),$$

c_0, \dots, c_m jsou reálná čísla, hledáme takovou, pro niž veličina

$$\rho^2(c_0, \dots, c_m) = \sum_{i=0}^n (y_i - P_m(x_i))^2$$

nabývá minimální hodnoty.

Takovou funkci pak nazýváme **nejlepší aproximací** experimentálních dat y_0, \dots, y_n v dané třídě funkcí **ve smyslu metody nejmenších čtverců**.

Veličina ρ^2 se nazývá **kvadratická odchylka**.

Nalezení nejlepší aproximace

Kvadratická odchylka

$$\rho^2 = \sum_{i=0}^n (y_i - c_0\varphi_0(x_i) - c_1\varphi_1(x_i) - \dots - c_m\varphi_m(x_i))^2$$

je funkcí koeficientů c_0, \dots, c_m . Z diferenciálního počtu funkcí více proměnných je známo, že nutnou podmínkou pro to, aby ρ^2 nabývala minima, je splnění rovnic

$$\frac{\partial}{\partial c_j}(\rho^2) = \frac{\partial}{\partial c_j} \left[\sum_{i=0}^n (y_i - c_0\varphi_0(x_i) - c_1\varphi_1(x_i) - \dots - c_m\varphi_m(x_i))^2 \right] = 0, \quad j = 0, \dots, m$$

Zderivováním dostaneme

$$\sum_{i=0}^n 2(y_i - c_0\varphi_0(x_i) - c_1\varphi_1(x_i) - \dots - c_m\varphi_m(x_i))(-\varphi_j(x_i)) = 0, \quad j = 0, \dots, m.$$

Rovnice vydělíme -2 a rozdělíme na jednotlivé sumy:

$$\sum_{i=0}^n y_i\varphi_j(x_i) - \sum_{i=0}^n c_0\varphi_0(x_i)\varphi_j(x_i) - \dots - \sum_{i=0}^n c_m\varphi_m(x_i)\varphi_j(x_i) \quad j = 0, \dots, m.$$

Z každé sumy můžeme vytknout odpovídající koeficient c_k . Snadnou úpravou pak dostaneme tzv. **normální rovnice** pro neznámé c_0, \dots, c_m :

$$c_0 \sum_{i=0}^n \varphi_0(x_i)\varphi_j(x_i) + \dots + c_m \sum_{i=0}^n \varphi_m(x_i)\varphi_j(x_i) = \sum_{i=0}^n y_i\varphi_j(x_i) \quad j = 0, \dots, m.$$

Tato soustava rovnic po rozepsání vypadá takto:

$$\begin{aligned} c_0 \sum_{i=0}^n \varphi_0^2(x_i) &+ c_1 \sum_{i=0}^n \varphi_1(x_i)\varphi_0(x_i) + \dots + c_m \sum_{i=0}^n \varphi_m(x_i)\varphi_0(x_i) = \sum_{i=0}^n y_i\varphi_0(x_i) \\ c_0 \sum_{i=0}^n \varphi_0(x_i)\varphi_1(x_i) &+ c_1 \sum_{i=0}^n \varphi_1^2(x_i) + \dots + c_m \sum_{i=0}^n \varphi_m(x_i)\varphi_1(x_i) = \sum_{i=0}^n y_i\varphi_1(x_i) \\ &\vdots \\ c_0 \sum_{i=0}^n \varphi_0(x_i)\varphi_m(x_i) &+ c_1 \sum_{i=0}^n \varphi_1(x_i)\varphi_m(x_i) + \dots + c_m \sum_{i=0}^n \varphi_m^2(x_i) = \sum_{i=0}^n y_i\varphi_m(x_i) \end{aligned} \quad (5.14)$$

Získaná soustava rovnic vypadá možná poněkud hrozivě a nepřehledně, ale uvidíme, že s konkrétními funkcemi φ_i se situace vyjasní.

Aproximace metodou nejmenších čtverců algebraickými polynomy

Velmi častá volba funkcí φ_i je $\varphi_i(x) = x^i$, $i = 0, 1, \dots, m$, tj.

$$\varphi_0(x) = 1, \varphi_1(x) = x, \dots, \varphi_m(x) = x^m$$

Aproximující funkce P_m je pak tvaru

$$P_m(x) = c_0 + c_1x + \dots + c_mx^m$$

a jednotlivé sumy v soustavě normálních rovnic vyjdou

$$\sum_{i=0}^n \varphi_0^2(x_i) = \sum_{i=0}^n 1 = \underbrace{1 + 1 + \dots + 1}_{n+1} = n + 1, \quad \sum_{i=0}^n \varphi_1(x_i)\varphi_0(x_i) = \sum_{i=0}^n x_i \cdot 1 = \sum_{i=0}^n x_i, \dots$$

obecně

$$\sum_{i=0}^n \varphi_k(x_i)\varphi_l(x_i) = \sum_{i=0}^n x_i^k \cdot x_i^l = \sum_{i=0}^n x_i^{k+l}, \quad k, l = 0, \dots, m$$

Soustava normálních rovnic pak vypadá následovně

$$\begin{aligned} c_0(n+1) + c_1 \sum_{i=0}^n x_i + \dots + c_m \sum_{i=0}^n x_i^m &= \sum_{i=0}^n y_i \\ c_0 \sum_{i=0}^n x_i + c_1 \sum_{i=0}^n x_i^2 + \dots + c_m \sum_{i=0}^n x_i^{m+1} &= \sum_{i=0}^n x_i y_i \\ &\vdots \\ c_0 \sum_{i=0}^n x_i^m + c_1 \sum_{i=0}^n x_i^{m+1} + \dots + c_m \sum_{i=0}^n x_i^{2m} &= \sum_{i=0}^n x_i^m y_i \end{aligned} \quad (5.15)$$

Speciálně pro aproximaci přímkou $P_1(x) = c_0 + c_1x$ dostaneme soustavu

$$\begin{aligned} c_0(n+1) + c_1 \sum_{i=0}^n x_i &= \sum_{i=0}^n y_i \\ c_0 \sum_{i=0}^n x_i + c_1 \sum_{i=0}^n x_i^2 &= \sum_{i=0}^n x_i y_i \end{aligned} \quad (5.16)$$

a pro aproximaci parabolou $P_2(x) = c_0 + c_1x + c_2x^2$ soustavu

$$\begin{aligned} c_0(n+1) + c_1 \sum_{i=0}^n x_i + c_2 \sum_{i=0}^n x_i^2 &= \sum_{i=0}^n y_i \\ c_0 \sum_{i=0}^n x_i + c_1 \sum_{i=0}^n x_i^2 + c_2 \sum_{i=0}^n x_i^3 &= \sum_{i=0}^n x_i y_i \\ c_0 \sum_{i=0}^n x_i^2 + c_1 \sum_{i=0}^n x_i^3 + c_2 \sum_{i=0}^n x_i^4 &= \sum_{i=0}^n x_i^2 y_i \end{aligned}$$

Příklad 5.5 Funkci zadanou následující tabulkou bodů aproximujte metodou nejmenších čtverců pomocí přímky.

x_i	0,2	0,5	0,9	1,6	2,0	2,9	3,5
y_i	16,58	19,30	18,12	20,94	20,90	24,66	24,50

Řešení: Bylo zadáno 7 bodů, proto $n = 6$. Koefficienty přímky získáme jako řešení soustavy rovnic 5.16. Pro přehlednost si všechny potřebné hodnoty zapíšeme do tabulky:

i	x_i	y_i	x_i^2	$x_i y_i$
0	0,2	16,58	0,04	3,316
1	0,5	19,30	0,25	9,650
2	0,9	18,12	0,81	16,308
3	1,6	20,94	2,56	33,504
4	2,0	20,90	4,00	41,800
5	2,9	24,66	8,41	71,514
6	3,5	24,50	12,25	85,750
Σ	11,6	145,00	28,32	261,842

Nyní můžeme sestavit normální rovnice:

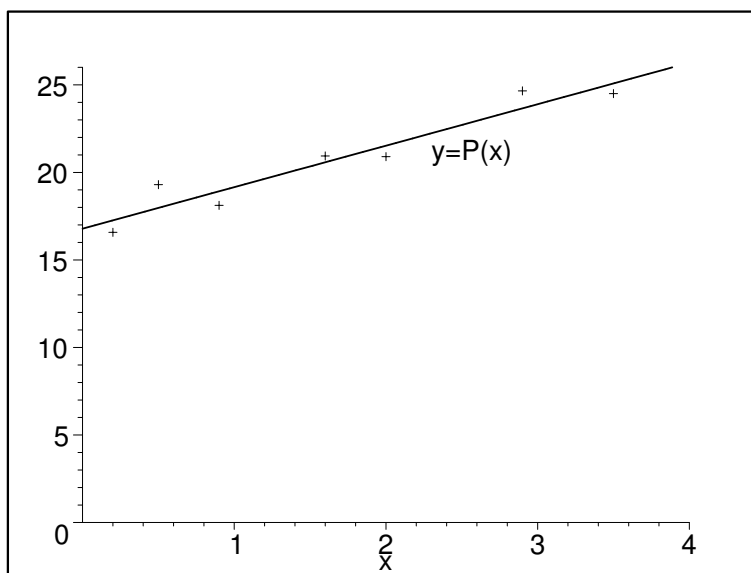
$$\begin{aligned} 7c_0 + 11,6c_1 &= 145 \\ 11,6c_0 + 28,32c_1 &= 261,842 \end{aligned}$$

Jejich řešením je $c_0 \doteq 16,788$, $c_1 \doteq 2,370$.

Hledaná přímka je tedy $P_1(x) = 16,788 + 2,370x$. Zadané body jsou spolu s touto přímkou zobrazeny na obrázku 5.7.

Poznámka. Pokud naměřené hodnoty vykazují periodické chování, je vhodnější je pomocí metody nejmenších čtverců aproximovat trigonometrickými polynomy. Za funkce φ_i můžeme volit

$$\varphi_0(x) = 1, \quad \varphi_1(x) = \cos x, \quad \varphi_2(x) = \sin x, \quad \varphi_3(x) = \cos 2x, \quad \varphi_4(x) = \sin 2x, \quad \dots$$



Obrázek 5.7: K příkladu 5.5: zadané body a nalezená přímka

6 Numerické derivování a integrování

V této kapitole se budeme zabývat otázkou, jak vypočítat derivaci a integrál z funkce, která je zadána pouze tabulkou bodů nebo pro kterou by byl analytický výpočet příliš složitý.

Základní myšlenkou je nahradit funkci interpolačním polynomem, popřípadě jinou aproximací, a derivovat či integrovat aproximující funkci.

6.1 Numerické derivování

Jak již bylo řečeno v úvodu, budeme řešit problém, jak vypočítat hodnotu derivace dané funkce v určitém bodě nikoli analyticky, ale pouze přibližně, a to pomocí známých funkčních hodnot v určitých bodech.

Můžeme k tomu použít interpolační polynom. Hodnotu derivace funkce nahradíme hodnotou derivace interpolačního polynomu. Tedy, je-li $P_n(x)$ interpolační polynom daný funkcí $f(x)$ a uzlovými body x_0, x_1, \dots, x_n , položíme

$$f'(x) \doteq P'_n(x).$$

Podobně pro derivace vyšších řádů (ovšem pouze do řádu n , pro vyšší už ne) můžeme položit

$$f^{(s)}(x) \doteq P^{(s)}(x).$$

Poznamenejme, že v uzlových bodech se hodnoty derivací funkce a interpolačního polynomu nemusejí shodovat. Pro ilustraci může posloužit opět 5.4, na kterém je dobře vidět, že zatímco funkční hodnoty v uzlových bodech jsou u funkce a interpolačního polynomu

stejně, směrnice tečen k těmto dvěma grafům (tj. hodnoty derivací) jsou v uzlových bodech velmi odlišné.

Odhad chyby pro první derivaci lze vyjádřit poměrně složitým vzorcem, proto jej uvedeme pouze speciálně pro uzlové body a pro případ, že uzly jsou **ekvidistantní**.

Připomeňme, že to jsou uzly takové, že vzdálenost mezi každými dvěma sousedními uzly je stejná. Tuto vzdálenost značíme h . Platí, že $x_i = x_0 + ih$, $i = 0, \dots, n$.

Pro takovéto uzly pak platí

$$f'(x_k) - P'_n(x_k) = \frac{(-1)^{n-k} h^n}{(n+1) \binom{n}{k}} f^{(n+1)}(\xi_k), \quad k = 0, \dots, n,$$

kde ξ_k jsou body ležící uvnitř intervalu $\langle x_0, x_n \rangle$.

6.1.1 Některé často používané vzorce pro numerické derivování

Uvedeme zde některé jednodušší, často užívané vzorce pro první a druhou derivaci v uzlových bodech. V tomto textu se s nimi ještě setkáme v kapitolách věnovaných numerickému řešení diferenciálních rovnic.

Jako poslední je v každém vzorci uveden chybový člen, který při samotném výpočtu zanedbáváme.

Čím vyšší mocnina kroku h se v něm vyskytuje, tím je chyba menší (a tedy vzorec lepší), neboť h bývá zpravidla malé číslo, $h \ll 1$, a pro taková čísla platí $h > h^2 > h^3 > \dots$.

Nejjednodušší vzorec pro derivaci prvního řádu dostaneme zderivováním interpolačního polynomu prvního stupně daného uzly x_0 a $x_1 = x_0 + h$.

Má-li funkce f druhou derivaci na intervalu $\langle x_0, x_1 \rangle$, pak existují body $\xi_0, \xi_1 \in \langle x_0, x_1 \rangle$ tak, že platí

$$f'(x_0) = \frac{f(x_1) - f(x_0)}{h} - \frac{h}{2} f''(\xi_0) \quad (6.1)$$

$$f'(x_1) = \frac{f(x_1) - f(x_0)}{h} - \frac{h}{2} f''(\xi_1). \quad (6.2)$$

Tyto vzorce lze též odvodit pomocí Taylorova rozvoje funkce f .

Derivováním interpolačního polynomu druhého stupně daného uzly $x_0 = x_1 - h, x_1$ a $x_2 = x_1 + h$ dostaneme přesnější vzorce pro první derivaci v těchto uzlových bodech.

Má-li funkce f čtvrtou derivaci na intervalu $\langle x_0, x_2 \rangle$, pak existují body $\xi_0, \xi_1, \xi_2 \in \langle x_0, x_2 \rangle$ takové, že

$$f'(x_0) = \frac{-3f(x_0) + 4f(x_1) - f(x_2)}{2h} + \frac{h^2}{3} f'''(\xi_0) \quad (6.3)$$

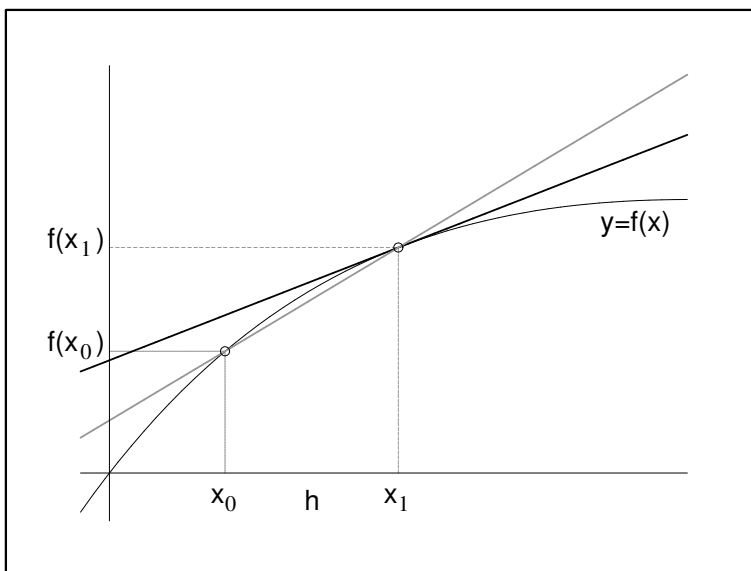
$$f'(x_1) = \frac{f(x_2) - f(x_0)}{2h} - \frac{h^2}{6} f'''(\xi_1) \quad (6.4)$$

$$f'(x_2) = \frac{f(x_0) - 4f(x_1) + 3f(x_2)}{2h} + \frac{h^2}{3} f'''(\xi_2) \quad (6.5)$$

Pomocí druhé derivace téhož interpolačního polynomu dostaneme vzorec pro druhou derivaci funkce f v bodě x_1 .

Má-li funkce f pátou derivaci na intervalu $\langle x_0, x_2 \rangle$, pak existuje bod $\xi \in \langle x_0, x_2 \rangle$ takový, že

$$f''(x_1) = \frac{f(x_0) - 2f(x_1) + f(x_2)}{h^2} - \frac{h^2}{12} f^{(4)}(\xi) \quad (6.6)$$



Obrázek 6.1: Numerické derivování - ilustrace ke vzorci 6.2

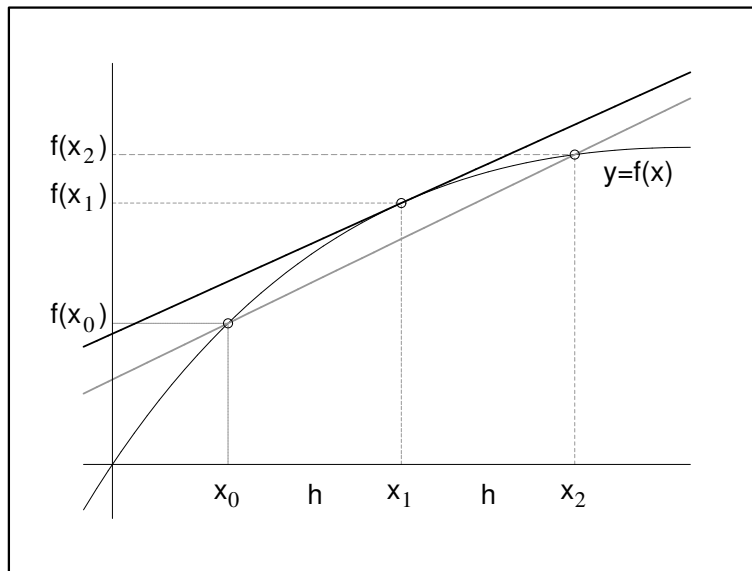
Na obrázcích 6.1 a 6.2 je zachycen geometrický význam vzorců 6.2 a 6.4. Hodnota derivace funkce f v bodě x_1 , tj. směrnice tečny ke grafu funkce v tomto bodě (tečna je na obrázcích nakreslena černě), je přibližně rovna směrnici sečny dané body x_0 a x_1 , resp. x_0 a x_2 (tyto sečny jsou na obrázcích nakresleny šedě).

Poznámka o zaokrouhlovací chybě při numerickém derivování

Mohlo by se zdát, že zmenšováním kroku h lze dosáhnout při numerickém derivování libovolné přesnosti. Bohužel se však ukazuje, že při příliš malém h může velmi narůst vliv zaokrouhlovací chyby.

To je vidět už z nejjednoduššího vzorce 6.2. Pro malé h může být $f(x_0) \doteq f(x_1)$ a tedy v čitateli zlomku odčítáme dvě sobě velmi blízká čísla, výsledek pak navíc opět dělíme malým číslem. To jsou operace vzhledem k zaokrouhlovací chybě velmi riskantní, viz kapitolu o chybách.

Naopak, při velkém kroku h nelze očekávat velkou přesnost vzhledem k chybě metody. Proto je potřeba volit kompromis, více o tom v [5].



Obrázek 6.2: Numerické derivování - ilustrace ke vzorci 6.4

V případě funkcí, jejichž hodnoty byly získány např. experimentálně a jsou zatíženy nezanedbatelnými chybami, se doporučuje nejprve tyto hodnoty metodou nejmenších čtverců „vyrovnat“ a potom teprve funkci derivovat.

6.2 Numerické integrování

Určení primitivní funkce k dané funkci $f(x)$ může být nesnadné, jak si čtenář jistě vzpomene z prvního semestru matematiky, někdy je to zcela nemožné. V případě, že jsou hodnoty funkce f dány tabulkou, pojem primitivní funkce úplně ztrácí smysl. Přesto můžeme chtít z takové funkce integrál vypočítat.

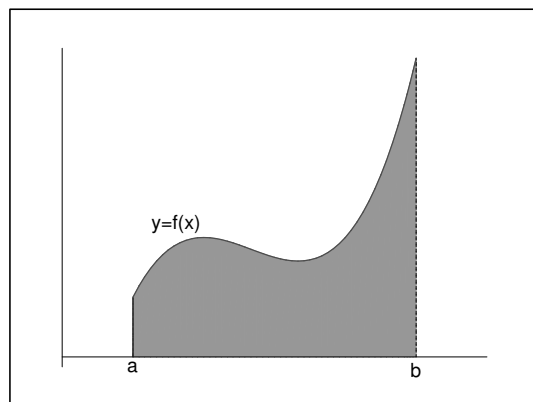
Zde se budeme zabývat výpočtem určitého integrálu $\int_a^b f(x)dx$. Jak si jistě všichni vzpomenu, pomocí tohoto integrálu se vypočítá obsah plochy pod grafem funkce $f(x)$ na intervalu $\langle a, b \rangle$, viz obrázek 6.3.

Numerický výpočet tohoto integrálu se nazývá numerická kvadratura. Jedna z možných cest je nahrazení funkce f na intervalu $\langle a, b \rangle$ interpolačním polynomem. Ten již se pak zintegruje snadno.

6.2.1 Newton-Cotesovy vzorce

Newton-Cotesovy kvadrurní vzorce (kvadrurní formule) obdržíme integrováním interpolačních polynomů s ekvidistantními uzly. Můžeme je rozdělit do dvou skupin:

- uzavřené vzorce, kde krajní body intervalu bereme za uzly kvadratury
- otevřené vzorce, kde krajní body nebereme za uzly kvadratury a uzly jsou položeny symetricky podle středu intervalu.

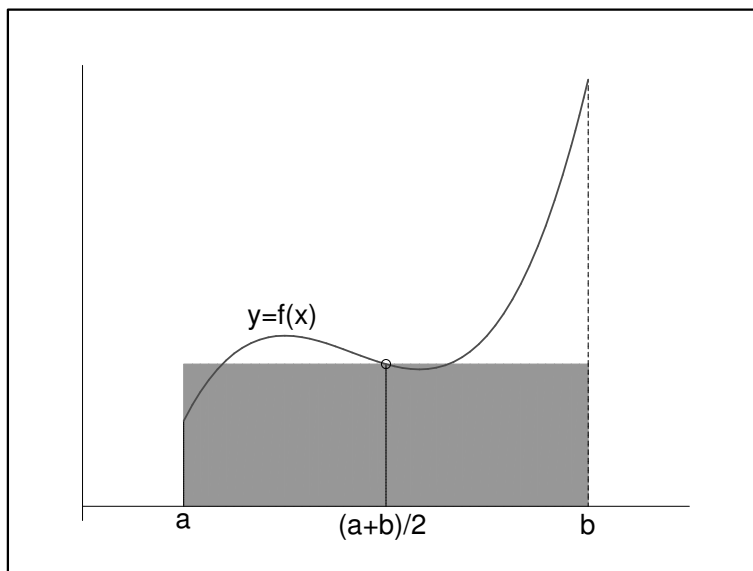


Obrázek 6.3: Připomenutí významu určitého integrálu

Bližší se zde budeme zabývat uzavřenými formulemi, z otevřených se můžeme zmínit o nejjednodušší z nich, a tou je tzv. **obdélníková metoda**.

Za jediný uzel interpolace bereme střed intervalu $\langle a, b \rangle$, vlastně funkci na tomto intervalu nahradíme konstantou $f(\frac{a+b}{2})$ a integrál je pak přibližně roven obsahu obdélníka, viz obrázek 6.4.

$$\int_a^b f(x) dx \doteq (b-a)f\left(\frac{a+b}{2}\right). \quad (6.7)$$



Obrázek 6.4: Obdélníková metoda

Z uzavřených vzorců je nejjednodušší **lichoběžníková metoda** (nebo též lichoběžníkové pravidlo).

Funkci $f(x)$ nahradíme na intervalu $\langle a, b \rangle$ lineárním interpolačním polynomem daným uzly a, b (zde zapsaným v Lagrangeově tvaru):

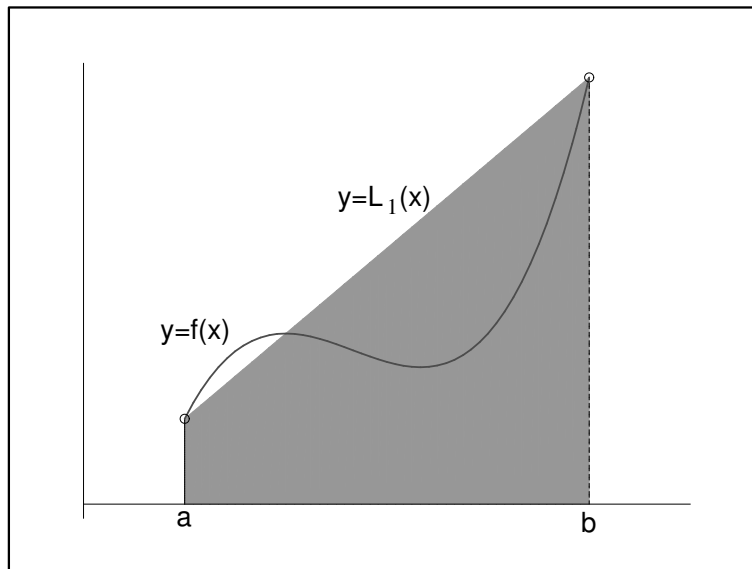
$$L_1(x) = f(a) \frac{x-b}{a-b} + f(b) \frac{x-a}{b-a}.$$

Integrací tohoto polynomu po použití jednoduchých úprav dostaneme

$$\int_a^b f(x) dx \doteq \int_a^b L_1(x) dx = \frac{b-a}{2} (f(a) + f(b)). \quad (6.8)$$

V tomto případě nahrazujeme obsah podgrafu funkce f obsahem příslušného lichoběžníka, viz obrázek 6.5, odtud název metody.

Poznámka. Vzorec 6.8 můžeme dostat i použitím známého vztahu pro obsah lichoběžníka $S = \frac{1}{2}(A + C)v$, kde A a C jsou délky podstav lichoběžníka a v je jeho výška. Musíme si ovšem uvědomit, že v tomto případě je lichoběžník obrácen, jeho podstavy jsou svisle.



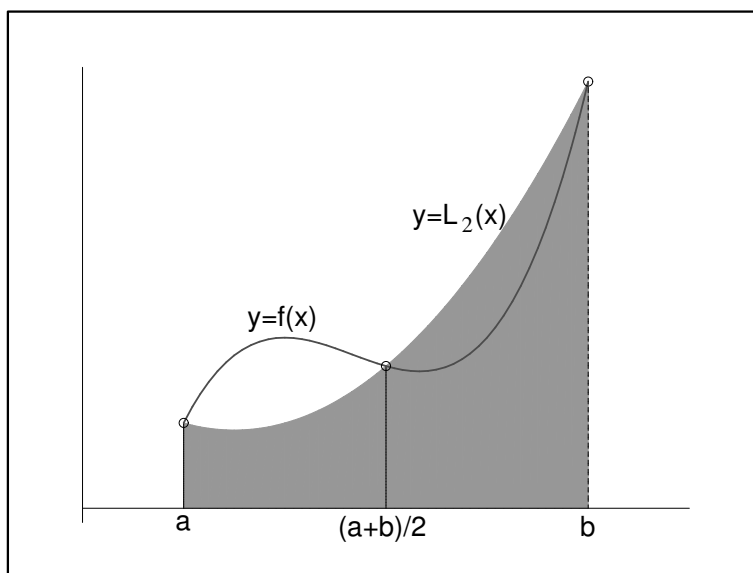
Obrázek 6.5: Lichoběžníková metoda

Na integraci interpolačního polynomu druhého stupně, za jehož uzly bereme a, b a střed integračního intervalu, tj. $\frac{a+b}{2}$, je založena tzv. **Simpsonova metoda** (viz obrázek 6.6):

$$\int_a^b f(x) dx \doteq \frac{b-a}{6} (f(a) + 4f(\frac{a+b}{2}) + f(b)). \quad (6.9)$$

Podobně bychom mohli integrovat interpolační polynomy vyšších stupňů. Přibližná hodnota integrálu vždy vyjde jako součet určitých násobků funkčních hodnot v uzlech. Obecně je **uzavřený Newton-Cotesův vzorec** tvaru

$$\int_a^b f(x) dx \doteq (b-a) \sum_{i=0}^n H_i f(x_i), \quad (6.10)$$



Obrázek 6.6: Simpsonova metoda

kde n je stupeň použitého interpolačního polynomu, H_i jsou tzv. Cotesovy koeficienty a x_i jsou uzly, pro něž platí $x_i = a + ih, i = 0, \dots, n$, ($h = \frac{b-a}{n}$ je krok mezi uzly). Přehled Cotesových koeficientů až do $n = 8$ lze nalézt např. v [2].

Chyba E Newton-Cotesových vzorců se vypočte integrací chyby interpolace 5.4,

$$E = \frac{1}{(n+1)!} \int_a^b f^{(n+1)}(\xi)(x-x_0) \cdots (x-x_n) dx$$

Zjednodušení tohoto výrazu je dosti obtížné, je ho potřeba provést zvlášť pro n sudé a pro n liché. Podrobnosti lze nalézt v [5].

Pro n sudé platí

$$E = \frac{f^{(n+2)}(\eta)}{(n+2)!} \int_a^b x(x-x_0) \cdots (x-x_n) dx, \quad (6.11)$$

a pro n liché

$$E = \frac{f^{(n+1)}(\eta)}{(n+1)!} \int_a^b (x-x_0) \cdots (x-x_n) dx, \quad (6.12)$$

kde $\eta \in [a, b]$. Integrály v těchto vzorcích lze pro konkrétní n vypočítat (byť je to poněkud pracné).

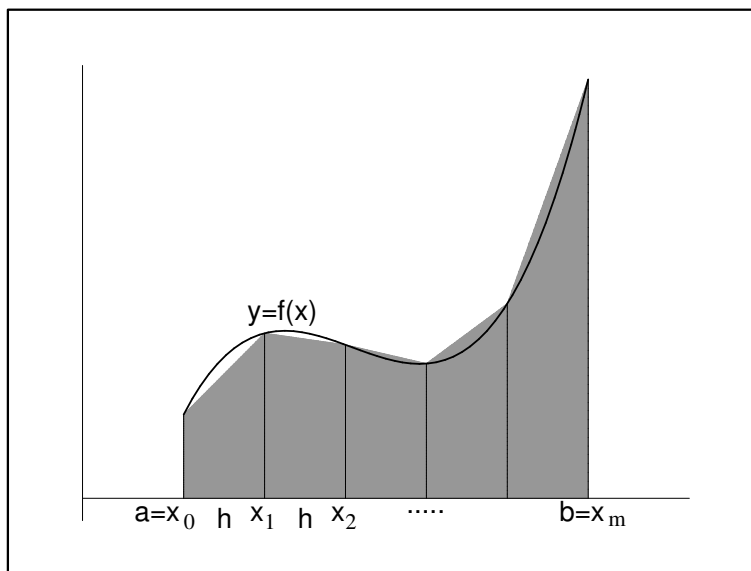
Např. chyba lichoběžníkové metody pomocí vzorce 6.12 vyjde

$$E = -\frac{1}{12} (b-a)^3 f''(\eta). \quad (6.13)$$

V kapitole o interpolaci jsme ukázali, že interpolační polynomy vyšších stupňů mohou oscilovat a nemusejí dobře vystihnout chování interpolované funkce. Také výpočet Cotesových koeficientů je pro velká n složitý. Proto se Newton-Cotesových vzorců vysokých řádů užívá zřídka.

6.2.2 Složené kvadrurní vzorce

Již z obrázku je vidět, že chyba integrace pomocí uvedených Newton-Cotesových vzorců nízkých řádů může být značná. Proto je lepší interval $\langle a, b \rangle$ rozdělit na větší počet stejných dílků a na každém z nich použít vybraný jednoduchý kvadrurní vzorec.



Obrázek 6.7: Složené lichoběžníkové pravidlo

Rozebereme si nyní podrobněji **složené lichoběžníkové pravidlo**.

Interval $\langle a, b \rangle$ rozdělíme na m subintervalů délky $h = \frac{b-a}{m}$ - viz obrázek 6.7. Na každém subintervalu použijeme jednoduché lichoběžníkové pravidlo. Platí

$$\begin{aligned} \int_a^b f(x) dx &= \int_{x_0}^{x_1} f(x) dx + \int_{x_1}^{x_2} f(x) dx + \cdots + \int_{x_{m-1}}^{x_m} f(x) dx \doteq \\ &\doteq \frac{h}{2} (f(x_0) + f(x_1)) + \frac{h}{2} (f(x_1) + f(x_2)) + \cdots + \frac{h}{2} (f(x_{m-1}) + f(x_m)) \end{aligned}$$

Celkem tedy

$$\int_a^b f(x) dx \doteq h \left(\frac{1}{2} f(x_0) + f(x_1) + \cdots + f(x_{m-1}) + \frac{1}{2} f(x_m) \right) = L_m \quad (6.14)$$

Je zřejmé, že čím jemněji interval $\langle a, b \rangle$ nadělíme, tím přesnější bude výsledek.

Chyba integrace na každém dílčím intervalu $\langle x_{i-1}, x_i \rangle$ je podle 6.13 $E_i = -\frac{1}{12} h^3 f''(\eta_i)$.

Celková chyba je tedy

$$E = -\frac{h^3}{12} \left(f''(\eta_1) + f''(\eta_2) + \dots + f''(\eta_m) \right).$$

Je-li funkce f'' na intervalu $[a, b]$ spojitá, existuje bod $\eta \in \langle a, b \rangle$ tak, že platí

$$f''(\eta_1) + f''(\eta_2) + \dots + f''(\eta_m) = m f''(\eta)$$

Dohromady dostaneme pro **chybu složeného lichoběžníkového pravidla**

$$E = -\frac{h^3}{12} m f''(\eta) = -\frac{(b-a)^3}{12 m^3} m f''(\eta) = -\frac{(b-a)^3}{12 m^2} f''(\eta). \quad (6.15)$$

Podobně jako u chyby interpolace je prakticky nemožné určit bod η . Lze-li nalézt $M_2 = \max_{t \in \langle a, b \rangle} |f''(t)|$, můžeme chybu alespoň shora odhadnout. Platí totiž

$$|E| \leq \frac{(b-a)^3}{12 m^2} M_2 \quad (6.16)$$

Tento odhad lze použít též pro určení vhodného počtu dělení m , chceme-li, aby chyba integrace nepřesáhla nějaké zadané ε .

Spíše než odhad chyby se ovšem pro dosažení žádané přesnosti ε používá jiný postup. Můžeme konstruovat posloupnost L_1, L_2, L_4, \dots . Její výpočet je velmi úsporný, protože všechny funkční hodnoty použité v nějakém L_m se použijí i při výpočtu L_{2m} . Platí

$$L_{2m} = \frac{1}{2} L_m + \frac{b-a}{2m} \left(f(x_1) + f(x_3) + \dots + f(x_{2m-1}) \right),$$

kde v závorce je pouze součet funkčních hodnot v nových dělicích bodech, které původní dělení zjemňují.

Výpočet zastavíme, jakmile je splněna podmínka $|L_{2m} - L_m| < \varepsilon$.

(Splněním této podmínky ale není zaručeno, že se L_{2m} od přesné hodnoty integrálu liší o méně než ε .)

Zcela analogicky jako složené lichoběžníkové pravidlo můžeme odvodit **složené Simpsonovo pravidlo**.

Interval $\langle a, b \rangle$ rozdělíme na **sudý** počet m dílků délky $h = \frac{b-a}{m}$ a postupně na dvojicích sousedních dílků použijeme jednoduché Simpsonovo pravidlo.

Po úpravě dostaneme

$$\begin{aligned} \int_a^b f(x) dx &\doteq & (6.17) \\ &\doteq \frac{h}{3} \left(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \dots + 2f(x_{m-2}) + 4f(x_{m-1}) + f(x_m) \right) = S_m \end{aligned}$$

Pro **odhad chyby** E se použije vzorec 6.11 a podobné úvahy jako při odvozování chyby složeného lichoběžníkového pravidla. Vyjde

$$E = -\frac{(b-a)^5}{180 m^4} f^{(4)}(\eta), \quad (6.18)$$

$\eta \in [a, b]$.

Příklad 6.1 Vypočítejte přibližnou hodnotu integrálu $\int_0^2 e^{-x^2} dx$ pomocí složeného lichoběžníkového pravidla pro $m = 4$. Odhadněte, jaké nanejvýš chyby se při tomto výpočtu dopustíme.

Řešení: Dosadíme do vzorce 6.14. Délka kroku h je v tomto případě $\frac{2-0}{4} = 0,5$. Přibližná hodnota integrálu je tedy

$$\begin{aligned} L_4 &= 0,5 \cdot \left(\frac{1}{2}f(0) + f(0,5) + f(1) + f(1,5) + \frac{1}{2}f(2) \right) = \\ &= 0,5 \cdot \left(\frac{1}{2}e^0 + e^{-0,25} + e^{-1} + e^{-2,25} + \frac{1}{2}e^{-4} \right) \doteq 0,8806 \end{aligned}$$

Odhad chyby dostaneme pomocí vzorce 6.16.

Musíme vypočítat druhou derivaci funkce $f(x) = e^{-x^2}$. Ta vyjde $f''(x) = e^{-x^2}(4x^2 - 2)$. Nyní najdeme maximum její absolutní hodnoty na intervalu $\langle 0, 2 \rangle$. Využitím poznatků z prvního semestru matematiky zjistíme, že funkce $f''(x)$ nabývá lokálního minima v bodě $x = 0$ a lokálního maxima v bodech $x = \pm \frac{\sqrt{6}}{2}$. Nás však zajímá maximum absolutní hodnoty na intervalu $\langle 0, 2 \rangle$. Vypočteme hodnoty f'' ve všech „podezřelých“ bodech:

$$f''(0) = -2 \quad f''\left(\frac{\sqrt{6}}{2}\right) \doteq 0,89 \quad f''(2) \doteq 0,26$$

V absolutní hodnotě je z těchto čísel největší -2 , tedy $M_2 = |-2| = 2$.

Celkem je tedy absolutní hodnota chyby nanejvýš rovna $\frac{(2-0)^3}{12 \cdot 4^2} \cdot 2 = \frac{1}{12} = 0,08\overline{33}$

Příklad 6.2 Zjistěte, jakou délku kroku je třeba zvolit při výpočtu integrálu $\int_0^2 e^{-x^2} dx$ (téhož jako v příkladu 6.1) pomocí složeného lichoběžníkového pravidla, chceme-li, aby chyba integrace nebyla větší než $0,001$.

Řešení: Přehlednější je najít nejprve vhodný počet dělení m , z něj již délku kroku určíme snadno.

Víme, že pro chybu E platí $|E| \leq \frac{(b-a)^3}{12m^2} M_2$. V příkladu 6.1 jsme zjistili, že $M_2 = 2$. Najdeme-li m tak, aby výraz na pravé straně předchozí nerovnosti byl menší než $0,001$, bude zaručeno, že i chyba E bude dostatečně malá. Má tedy platit

$$\frac{(2-0)^3}{12m^2} \cdot 2 \leq 0,001$$

Odtud snadno dostaneme, že

$$\begin{aligned} m^2 &\geq \frac{8 \cdot 2}{12 \cdot 0,001} \\ m &\geq 36,51 \end{aligned}$$

Zvolíme-li tedy $m = 37$ (nebo jakékoli větší), je zaručeno, že chyba bude menší než $0,001$. Hledaná délka kroku může být tedy $\frac{2}{37}$.

Poznamenejme, že takto získaný počet dělení m může být zbytečně velký. V tomto příkladu by ve skutečnosti pro dosažení zadané přesnosti stačilo už $m = 5$ - to ale bez znalosti přesné hodnoty integrálu nejsme schopni rozeznat. S počtem dělení získaným právě

předvedeným postupem máme sice možná více práce, ale zato jistotu, že výsledek bude dost přesný.

Poznámka. Kromě Newton-Cotesových kvadrurních vzorců existuje i mnoho dalších. Důležité jsou např. Gaussovy kvadrurní formule. V nich se přibližná hodnota integrálu opět počítá jako lineární kombinace funkčních hodnot,

$$\int_a^b f(x) dx \doteq \sum_{i=0}^n H_i f(x_i).$$

Koeficienty $H_i \in R$ a uzly $x_i \in \langle a, b \rangle$ jsou určeny tak, aby vzorec byl přesný pro integrování polynomů do stupně $2n + 1$ včetně.

Poznámka. Numerický výpočet neurčitého integrálu $\int f(x) dx$ spočívá v nalezení funkce $y(x) = \int_{x_0}^x f(t) dt$.

Tato úloha je ekvivalentní s nalezením řešení Cauchyovy počáteční úlohy

$$y' = f(x), \quad y(x_0) = 0.$$

Metodám numerického řešení takovýchto úloh bude věnována kapitola 7.

7 Numerické řešení diferenciálních rovnic

Bude doplněno.

7.1 Počáteční úlohy

7.2 Okrajové úlohy

Část II

PRAVDĚPODOBNOST

8 Pravděpodobnostní modely

Nyní se ve studiu přeneseme někam trochu jinam - opustíme numerické metody a vrhneme se do studia pravděpodobnosti. Cílem této kapitoly je představit čtenáři čtyři základní pojetí pravděpodobnosti, která jsou užívána v technické praxi. Uvidíme, že pojetí 7.1 je speciálním případem pojetí 8.3 a pojetí 8.2 speciálním případem pojetí 8.4. Důležitý je pojem náhodné veličiny (náhodné proměnné) X . Popsat, jak se veličina X chová, je úkolem teorie pravděpodobnosti.

Co je to pravděpodobnost? Souběžně v tomto textu se mluví i o statistice, tedy druhá otázka, která s tou první souvisí, je: Co je to statistika?. Statistika a pravděpodobnost jsou jako dvě strany jedné mince. Teorie pravděpodobnosti se ptá: Pokud vycházíme z konkrétního stavu světa, jaké důsledky budou pravděpodobně následovat? A teorie statistiky se ptá: Pokud vycházíme z jisté skupiny důsledků (např. měření), jaký stav světa asi tyto důsledky způsobí?

Příklad 8.1 *Házíme hrací kostkou. Pokud je kostka z homogenního materiálu (vycházíme z určitého stavu světa), tj. je regulérní hrací kostka a není falešná, pravděpodobnost, že padne šestka (=určitý důsledek), je rovna $\frac{1}{6}$. Tj. usuzujeme, že šestka padne asi v $\frac{1}{6} \cdot 100 = 16,66\%$ případů hodu kostkou. To je pravděpodobnost.*

Kdyby naopak nám ze 150 hodů kostkou šestka padla ve 47 případech (= měření), usoudili bychom, že buď se jedná o náhodu, že šestka padala v $\frac{47}{150} \cdot 100 = 31,33\%$ hodů, nebo kostka není homogenní a obsahuje nějaké olůvko, které ji nutí k tomu, aby šestka padala častěji (usuzujeme na určitý stav světa). To je statistika.

Tématem této ilustrace byla náhodná veličina X , která udává, jaký počet ok padne při hodu kostkou.

Dříve než představíme jednotlivá pojetí pravděpodobnosti, musíme zavést určité označení. Písmeno Ω bude značit množinu všech hodnot, kterých náhodná veličina X může nabývat. Bude to zpravidla množina všech možných výsledků experimentu nebo hry. Velkými písmeny (např. A, B, \dots) budeme označovat nějaké podmnožiny množiny Ω a budeme jim říkat *náhodné jevy*. Když řekneme, že *nastal jev A* , budeme tím rozumět, že náhodná veličina X nabývá hodnoty z množiny A . Symbol $P(A)$ bude označovat *pravděpodobnost, že nastane jev A* . Pravděpodobnost splňuje následující vlastnosti:

- (i) $0 \leq P(A) \leq 1$.
- (ii) Ω označuje *jev jistý*, jehož pravděpodobnost je $P(\Omega) = 1$, prázdná množina \emptyset znamená *jev nemožný*, pro který $P(\emptyset) = 0$.

- (iii) Pokud náhodné jevy A_1, A_2, \dots, A_n jsou po dvou disjunktní, tj. $A_i \cap A_j = \emptyset$ pro $i \neq j$, pak pravděpodobnost jejich sjednocení je rovna součtu jednotlivých pravděpodobností, tj.

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

Dále $\bar{A} = \Omega - A$ znamená *opačný jev k jevu A*. Jev \bar{A} tedy nastane, pokud nenastane jev A . Sjednocení jevů $A \cup B$ znamená, že nastane aspoň jeden z jevů A, B . Průnik jevů $A \cap B$ říká, že jevy A, B nastanou současně. Z vlastností (i) až (iii) lze odvodit další důležité vztahy, které platí:

(iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(v) Pokud $A \subseteq B$, tak $P(A) \leq P(B)$.

(vi) $P(\bar{A}) = 1 - P(A)$.

To jsme tedy charakterizovali pravděpodobnost a můžeme se pustit do studia jednotlivých typů pravděpodobnostních modelů.

8.1 Klasická pravděpodobnost

Vážení přátelé, ano. Důvodem vzniku pravděpodobnosti je rozvoj hazardních her. To je tzv. klasické pojetí. **Klasická pravděpodobnost jevu A se definuje jako podíl počtu příznivých výsledků (=hodnot ležících v množině A= počtu prvků množiny A) ku počtu všech možných výsledků (= počtu prvků množiny Ω).** Označíme-li počet prvků množiny svislými čarami, platí

$$P(A) = \frac{|A|}{|\Omega|} \text{ (svislé čáry označují počet prvků množiny).}$$

Příklad 8.2 *Uvažujme jednoduchou hazardní hru, která spočívá v hodů mincí dvakrát za sebou. Přitom náhodná veličina (X, Y) neudává vzdálenost, do které jsme minci hodili, nýbrž všímá si, kolikrát a v jakém pořadí padl na minci rub nebo líc. Jedná se vlastně o dvourozměrnou veličinu - její první souřadnice X charakterizuje první hod, druhá souřadnice Y druhý hod mince. Množina všech možných výsledků je zde $\Omega = \{(L, R), (R, L), (L, L), (R, R)\}$. Když jev A například znamená, že v naší hře padl líc při prvním hodů, tento výsledek nastane ve dvou případech: $A = \{(L, R), (L, L)\}$. Tedy*

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2}{4} = 0,5.$$

To znamená, že když naši jednoduchou hru budeme několikrát opakovat, tak pokud mince není falešná a je dobře vyvážená, jev A nastane přibližně v 50% případech.

Důležitý je následující rámeček, kde je charakterizováno, kdy lze klasickou pravděpodobnost použít:

Klasickou pravděpodobnost můžeme užít jen tehdy, když Ω (= množina všech možných výsledků pokusu) je konečná a všechny výsledky hry nebo pokusu nastávají se stejnou pravděpodobností (= jsou stejně pravděpodobné).

Příklad 8.3 *Uvažujme jednoduchý experiment třech hodů mincí. Jaká je pravděpodobnost jevu $A =$ dvakrát padne líc a jednou rub (přitom nezáleží na pořadí, ve kterém padnou)?*

Řešení: Množina všech možných výsledků experimentu je

$$\Omega = \{LLL, LLR, LRL, RLL, LRR, RLR, RRL, RRR\}.$$

Množinu A lze psát $A = \{LLR, LRL, RLL\}$. Podle definice klasické pravděpodobnosti tedy $P(A) = \frac{3}{8} = 0.375$.

Aby bylo vidět, že klasickou pravděpodobnost lze užít i k něčemu praktičtějšímu, zavedeme pojem podmíněné pravděpodobnosti. O co se jedná, vysvětlím na příkladu.

Příklad 8.4 *Ze 120 studentů v přednáškové skupině jich 90 spočetlo příklady zadané za dobrovolnou domácí samostatnou práci. Pak 75 studentů složilo zkoušku v řádném termínu, z toho 70 bylo těch, co spočítali zadané příklady. Student XY se přišel zeptat na výsledek zkoušky. Zkoušející jej nezná, ale XY prozradí, že si spočítal zadané příklady. Zkoušející nešťastnou náhodou zapomeněl zkouškovou zprávu doma, ale na základě předchozích souhrnných údajů (které zná z paměti) studentovi je schopen říct pravděpodobnost, s jakou složil zkoušku. Určete ji i vy.*

*Řešení: Označme $S =$ náhodně vybraný student spočítal zadané příklady; $Z =$ náhodně vybraný student složil zkoušku. Naším úkolem je určit **podmíněnou pravděpodobnost** $P(Z|S)$ (čti: podmíněná pravděpodobnost jevu Z , pokud už víme, že nastala podmínka S ; podmínku S při tomto typu zápisu píšeme vždy za svislou čarou). Abychom tuto podmíněnou pravděpodobnost mohli určit, zcela zapomeneme ty studenty, kteří nesplňují podmínku, o které víme, že už nastala - zúžíme tedy své další uvažování pouze na množinu těch 90 studentů, kteří spočítali zadané příklady. Pak v duchu klasické pravděpodobnosti použité na těchto 90 studentů dostáváme*

$$P(Z|S) = \frac{70}{90} = 0.778 = \frac{|Z \cap S|}{|S|}.$$

v rámci hesla „vyjádří jednoduchou skutečnost tak, aby jí nikdo nerozuměl“ nyní ještě upravíme poslední zlomek v příkladu tak, že čitatele i jmenovatele vydělíme hodnotou $|\Omega|$ (koneckonců je to povolená úprava, takže si to můžeme dovolit):

$$P(Z|S) = \frac{|Z \cap S|}{|S|} = \frac{\frac{|Z \cap S|}{|\Omega|}}{\frac{|S|}{|\Omega|}} = \frac{P(Z \cap S)}{P(S)}.$$

Poslední vztah v předchozím odvození se uvádí jako základní vzorec pro výpočet podmíněné pravděpodobnosti:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (8.1)$$

Porovnáním pravděpodobností $P(Z) = \frac{75}{120} = 0.625$ a $P(Z|S) = 0.778$ vidíme, že spočítání domácí úlohy naznačuje, že student dopadl u zkoušky lépe. Podmíněná pravděpodobnost udává, jak se změní $P(Z)$ dodáním podmínky S , tj. jakým způsobem ovlivní podmínka S pravděpodobnost jevu Z .

Někdy dodání další podmínky pravděpodobnost jevu neovlivní - pak říkáme, že jev A je na podmínce B nezávislý, nebo že **jevy A, B jsou nezávislé**:

$$P(A|B) = P(A)$$

Příklad 8.5 *V situaci z příkladu 8.3 sestává 120 studentů, o kterých je řeč, z 24 dívek (z nichž 15 složilo zkoušku) a 96 kluků (z nichž 60 složilo zkoušku). Závísí úspěch u zkoušky na tom, zda je student dívka nebo kluk?*

Řešení. Označme D = náhodně vybraný student je dívka; K = náhodně vybraný student je kluk. Pak

$$P(Z|D) = \frac{P(Z \cap D)}{P(D)} = \frac{\frac{15}{120}}{\frac{24}{120}} = 0.625 = P(Z);$$

$$P(Z|K) = \frac{P(Z \cap K)}{P(K)} = \frac{\frac{60}{120}}{\frac{96}{120}} = 0.625 = P(Z).$$

Vidíme, že jev Z nezávisí na jevu D , ani na jevu K . Tj. úspěch u zkoušky nezávisí na tom, zda je student dívka nebo kluk.

Zatím se zdálo, že dosazovat do vzorce 8.1 je poněkud vykonstruované, protože dosazujeme dva stejné jmenovatele, které pak zkrátíme, ale tento vztah má skutečně užítí - například lze z něj zase něco odvodit, a sice vztah pro výpočet průniku dvou jevů:

$$P(A \cap B) = P(A) \cdot P(B|A); \quad (8.2)$$

a protože při operaci průniku nezáleží na pořadí množin, platí též

$$P(A \cap B) = P(B \cap A) = P(B) \cdot P(A|B).$$

Zkrátka a dobře, při výpočtu pravděpodobnosti průniku jevů lze použít libovolný ze dvou právě uvedených vzorců podle toho, do kterého umíme jednodušeji dosadit. Pokud jevy A, B jsou nezávislé, na základě toho, co už bylo řečeno, víme, že platí

$$P(A \cap B) = P(A) \cdot P(B). \quad (8.3)$$

Příklad 8.6 Semináře se účastní šest lidí, z toho čtyři muži a dvě ženy. Během prvních šesti týdnů semináře má každý účastník jednou vystoupit s referátem. Pořadí referátů je sestaveno náhodně, tj. každý týden je náhodně vybrán jeden z těch, co ještě nereferovali. Jaká je pravděpodobnost, že první dva týdny budou mít referát ženy?

Řešení: Označme F_1 = první týden má referát žena, F_2 = druhý týden má referát žena. Pak podle vzorce 8.2 $P(F_1 \cap F_2) = P(F_1) \cdot P(F_2|F_1)$. Podle klasické pravděpodobnosti $P(F_1) = \frac{2}{6}$. Při výpočtu $P(F_2|F_1)$ musíme brát v úvahu platnost podmínky, že první týden byla vybrána žena. Proto tedy druhý týden můžeme vybrat už jen z pěti kandidátů pouze zbývající ženu, co ještě nereferovala, tj. $P(F_2|F_1) = \frac{1}{5}$. Celkem $P(F_1 \cap F_2) = \frac{2}{6} \cdot \frac{1}{5} = 0.066$.

Příklad 8.7 Současně házeme kostkou i mincí. Jaká je pravděpodobnost, že na kostce padne pětka a na minci současně padne líc?

Řešení: Protože hod mincí je nezávislý na hodu kostkou, využijeme vztah 8.3:

$$P(5 \cap L) = P(5) \cdot P(L) = \frac{1}{6} \cdot \frac{1}{2} = 0.083.$$

Zatím jsme se stále nevzdálili od celkem nepraktického házení kostkou nebo mincí. Ale jak už to bývá, abychom se přiblížili popisu složitějších skutečností, budeme muset ještě projít trochu složitější teorie. Uvažujte se mnou následující situaci: Ω stále značí množinu všech možných výsledků experimentu (z nichž každý nastává se stejnou pravděpodobností). Vezměme libovolné **disjunktní pokrytí množiny** Ω - tím rozumíme takový systém podmnožin H_1, H_2, \dots, H_k množiny Ω , kde

$$H_i \cap H_j = \emptyset \text{ pro } i \neq j, \text{ a dále } \bigcup_{i=1}^k H_i = \Omega.$$

Čili množinu Ω jsme rozdělili na disjunktní systém podmnožin. Když nyní vezmeme libovolnou podmnožinu A množiny Ω , platí následující bizarní vztah:

$$A = (H_1 \cap A) \cup (H_2 \cap A) \cup \dots \cup (H_k \cap A), \quad (8.4)$$

slovně vyjádřeno - množina A má s každou z množin H_1, \dots, H_k nějaký průnik (třeba i prázdný), a když se všechny ty průniky sjednotí, dostaneme zase množinu A . Kdo tomu nevěří, ať si nakreslí obrázek třeba pro $k = 4$ (nakreslete nejprve množinu Ω , pak ji rozdělte na disjunktní systém množin H_1, H_2, H_3, H_4 , a nakonec množinu A).

Využijme nyní pro výpočet $P(A)$ bizarního vztahu 8.4:

$$\begin{aligned} P(A) &= P(H_1 \cap A) + P(H_2 \cap A) + \dots + P(H_k \cap A) = \\ &= P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) + \dots + P(H_k) \cdot P(A|H_k) \end{aligned}$$

(první rovnost platí na základě vlastnosti (iii) pravděpodobnosti disjunktního sjednocení z úvodu kapitoly, druhá rovnost je pouze přepis s využitím vzorce 8.2). Uvedený vztah se nazývá **věta o úplné pravděpodobnosti** - přepišme jej ještě jednou:

$$P(A) = P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) + \dots + P(H_k) \cdot P(A|H_k). \quad (8.5)$$

Úplná pravděpodobnost zde je právě $P(A)$, kterou dostaneme součtem jistých dílčích pravděpodobností - odtud název věty. Na otázku, k čemu je tento vztah dobrý, odpovídám, že paradoxně je někdy jednodušší vypočítat $P(A)$ pomocí tohoto vzorce, protože pravděpodobnosti $P(H_i)$ a $P(A|H_i)$ jsou celkem snadno zjistitelné.

Příklad 8.8 Ze zkušenosti se ví, že Tomáš zasáhne basketbalový koš s pravděpodobností 0.8, Jana s pravděpodobností 0.5 a Honza s pravděpodobností 0.4. Jaká je pravděpodobnost, že náhodně vybraný hráč trefí koš?

Řešení. Klíčem úspěchu těchto a podobných příkladů je všechny jevy si dobře označit. To někdy studenti podcení, rychle něco spočtou, a pak neví, co vlastně spočetli - tak se snadno vyrobí chyba. Nejprve musíme označit disjunktivní pokrytí množiny možných výsledků: H_1 = vybraný hráč je Tomáš, H_2 = vybraný hráč je Jana, H_3 = vybraný hráč je Honza. To je disjunktivní pokrytí - jednotlivé situace se navzájem vylučují (nemohou nastat současně) a žádná další situace nastat nemůže. Dále A = náhodně vybraný hráč trefí koš. Protože $P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}$, známe vše potřebné pro dosazení do vzorce:

$$\begin{aligned} P(A) &= P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) + P(H_3) \cdot P(A|H_3) = \\ &= \frac{1}{3} \cdot 0.8 + \frac{1}{3} \cdot 0.5 + \frac{1}{3} \cdot 0.4 = 0.566. \end{aligned}$$

Ten, kdo na předchozí příklad přišel i bez vzorce 8.5, necht' prosím promine, že se snažím zamtlžit jednoduché skutečnosti složitými vzorci. Ono se opravdu jedná o prosté úvahy vyplývající z vlastností pravděpodobnosti.

A ještě poslední odvození na téme klasické pravděpodobnosti: kombinací vzorce pro podmíněnou pravděpodobnost, pravděpodobnost průniku a věty o úplné pravděpodobnosti dostaneme:

$$P(H_i|A) = \frac{P(H_i \cap A)}{P(A)} = \frac{P(H_i) \cdot P(A|H_i)}{P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) + \dots + P(H_k) \cdot P(A|H_k)}$$

Tento vzorec se nazývá **Bayesův vzorec** - přepišme jej ještě jednou:

$$P(H_i|A) = \frac{P(H_i) \cdot P(A|H_i)}{P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) + \dots + P(H_k) \cdot P(A|H_k)} \quad (8.6)$$

Příklad 8.9 Víme, že pravděpodobnost, že Honza na pálce při baseballu dobře odpálí míč, je 0.1. Pravděpodobnost, že kdokoli jiný z jeho týmu dobře odpálí, je rovna 0.3. Z rádia se dovídáme, že Honzův tým je na pálce, a slyšíme: Je to zásah! Jaká je pravděpodobnost, že rozhlasový reportér mluví o Honzovi (v jednom týmu je celkem devět hráčů)?

Řešení: Jádrem správného použití Bayesova vzorce je nalezení disjunktivního pokrytí a označení jevu A - zbytek už jen spočívá v dosazení. Tak tedy: H_1 = Honza je na pálce, H_2 = někdo jiný z Honzova týmu je na pálce. Tyto dva jevy tvoří disjunktivní pokrytí, protože vyčerpávají všechny situace, které nás zajímají, a přitom nemohou nastat současně. Dále A = Honzův tým zasáhl míč. Naším úkolem je zjistit $P(H_1|A)$:

$$P(H_1|A) = \frac{P(H_1) \cdot P(A|H_1)}{P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2)} = \frac{\frac{1}{9} \cdot 0.1}{\frac{1}{9} \cdot 0.1 + \frac{8}{9} \cdot 0.3} = 0.04.$$

Složitější využití Bayesova vzorce probírá následující příklad.

Příklad 8.10 Výrobce dodává svůj produkt v sadách o pevném počtu kusů. Díky poruchám ve výrobním procesu je v některých sadách nepřijatelné množství zmetků. Pravděpodobnost výskytu této špatné sady (se zmetkovitostí 15%) je $P(B) = 0.05$, kdežto dobré sady (se zmetkovitostí 4%) $P(G) = 0.95$. Výrobce ví, že prodej špatné sady může být pokutován. Samozřejmě si může myslet, že pravděpodobnost výroby špatné sady je tak malá, že může k dodávce zvolit jakoukoliv sadu. Ale pokud provede kontrolu např. pěti výrobků z dané sady, tato dodatečná informace může ovlivnit jeho rozhodnutí (jedná se o tzv. **aposteriorní Bayesovské rozhodování**, *aposteriori* = po (provedení kontroly, experimentu, apod.) - na rozdíl od *apriorního* rozhodnutí, *apriori* = před). Označíme-li

Y_0 = z pěti kontrolovaných výrobků dané sady jsou všechny v pořádku;

Y_1 = z pěti kontrolovaných výrobků dané sady je jeden zmetek;

Y_2 = z pěti kontrolovaných výrobků dané sady jsou dva zmetky;

Y_3 = z pěti kontrolovaných výrobků dané sady jsou tři zmetky;

Y_4 = z pěti kontrolovaných výrobků dané sady jsou čtyři zmetky;

Y_5 = z pěti kontrolovaných výrobků dané sady je všech pět vadných,

vypočtete $P(G|Y_i)$ pro $i = 0, 1, \dots, 5$.

Řešení. Příslušná apriorní pravděpodobnost je $P(G) = 0.95$. Vypočteme nyní aposteriorní pravděpodobnosti, které v sobě obsahují už výsledek kontroly pěti výrobků z dané sady. Příslušné disjunktní pokrytí je právě B = daná sada je špatná (bad), G = daná sada je dobrá (good). Využijeme tedy Bayesova vzorce

$$P(G|Y_i) = \frac{P(G) \cdot P(Y_i|G)}{P(G) \cdot P(Y_i|G) + P(B) \cdot P(Y_i|B)}$$

(pro zmatení nepřítelů index i v celém vzorci zůstává stejný). Pak počet zmetků v dobré sadě z pěti vybraných má rozdělení $Bi(N = 5, p = 0.04)$, počet zmetků ve špatné sadě z pěti vybraných rozdělení $Bi(N = 5, p = 0.15)$. Pomocí těchto modelů určíme pravděpodobnosti $P(Y_i|B)$, $P(Y_i|G)$. Dosazením máme

$$\begin{aligned} P(G|Y_0) &= \frac{0.95 \cdot 0.96^5}{0.95 \cdot 0.96^5 + 0.05 \cdot 0.85^5} = 0.972; \\ P(G|Y_1) &= \frac{0.95 \cdot \binom{5}{1} \cdot 0.04 \cdot 0.96^4}{0.95 \cdot \binom{5}{1} \cdot 0.04 \cdot 0.96^4 + 0.05 \cdot \binom{5}{1} \cdot 0.15 \cdot 0.85^4} = 0.892; \\ P(G|Y_2) &= 0.661; \\ P(G|Y_3) &= 0.315; \\ P(G|Y_4) &= 0.098; \\ P(G|Y_5) &= 0.025. \end{aligned}$$

Vidíme tedy, že rostoucí počet zmetků ve výběru podstatně mění původní apriorní pravděpodobnost $P(G) = 0.95$. Kdyby například při kontrole pěti výrobků byly už čtyři vadné, jedná se o dobrou sadu s pravděpodobností menší než jedna desetina a výrobce by měl raději k dodávce zvolit sadu jinou.

Předchozí příklad rozebírá teoretické zázemí za jistým typem podnikové či firemní kontroly - přinejmenším dobrá ukázka toho, že i pomocí klasické pravděpodobnosti lze popsat určité situace praxe.

8.2 Geometrická pravděpodobnost

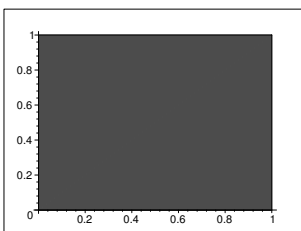
Příklad 8.11 Honza a Marek se domluvili, že se setkají na jistém místě mezi osmou a devátou hodinou, kam každý z nich v tu dobu přijde. Ale řekli si, že ten, kdo přijde první, bude na toho druhého čekat jen 15 minut, a pak odejde. Jaká je pravděpodobnost, že se setkají?

Řešení: Označme

x ... čas příchodu Honzy (v hodinách);

y ... čas příchodu Marka.

Víme, že oba přijdou určitě do devíti hodin, tedy $0 \leq x \leq 1$, $0 \leq y \leq 1$. Každý výsledek jejich příchodu lze vyjádřit jako uspořádanou dvojici (x, y) , což lze znázornit - a uvidíme, že to bude pomocí - jako bod v rovině, jehož obě souřadnice leží v intervalu $(0, 1)$. Všechny tyto body modelující možný výsledek příchodů vytvářejí tedy čtverec v rovině. Tento čtverec $\Omega = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$ je množinou všech možných výsledků dané situace (viz obrázek 8.1).



Obrázek 8.1: K př. 8.11: Množina všech možných výsledků.

Počet všech možných případů je sice nekonečný, ale jsme schopni spočítat obsah čtverce: $S(\Omega) = 1 \cdot 1 = 1$.

Označme dále

A ... Honza a Marek se setkají

Příznivým případům jevu A odpovídají ty příchody (x, y) obou studentů, ve kterých se

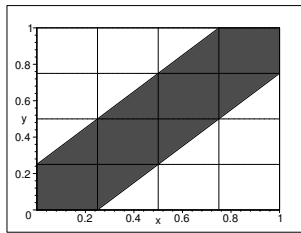
x od y liší nanejvýš o 15 minut, což je asi $\frac{1}{4}$ hodiny. Pro tyto „příznivé“ body čtverce Ω tedy musí platit nerovnost

$$|y - x| \leq \frac{1}{4}.$$

Vyřešme tuto nerovnost. Při odstraňování absolutní hodnoty musíme rozlišit dvě situace:

- Pro $y - x \geq 0$ se znaménka nemění, tj $y - x \leq \frac{1}{4}$, odtud $y \leq x + \frac{1}{4}$.
- Pro $y - x < 0$ musíme při odstraňování absolutní hodnoty na levé straně nerovnosti změnit znaménka: $-y + x \leq \frac{1}{4}$, odtud $y \geq x - \frac{1}{4}$.

Body splňující některou z uvedených dvou situací lze znázornit vyšrafovanou částí na obrázku 8.2:



Obrázek 8.2: K př. 8.11: Množina všech příznivých výsledků.

Jev A lze tedy vyjádřit jako množinu bodů v rovině:

$$A = \left\{ (x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1, y \leq x + \frac{1}{4}, y \geq x - \frac{1}{4} \right\}.$$

Příznivých případů je také nekonečně mnoho, ale jsme schopni vypočítat míru této nekonečnosti, konkrétně řečeno obsah množiny A : nejjednodušeji $S(A)$ vypočteme z grafického znázornění na obrázku 8.2, když budeme brát v úvahu rozdělení čtverce Ω na šestnáct menších čtverečků o straně délky $\frac{1}{4}$. Je vidět, že množina A zabírá plochu sedmi z těchto čtverečků, a protože $S(\Omega) = 1$, máme $S(A) = \frac{7}{16} \cdot S(\Omega) = \frac{7}{16}$.

Pravděpodobnost jevu A teď určíme jako podíl míry množiny příznivých případů a míry množiny všech možných případů:

$$P(A) = \frac{S(A)}{S(\Omega)} = \frac{\frac{7}{16}}{1} = \frac{7}{16}.$$

V tomto příkladu jsme se zabývali opět dvourozměrnou veličinou (X, Y) , aby byl krásně zřetelný geometrický rozměr tohoto pravděpodobnostního modelu. Ovšem definování některých dále uváděných pojmů pro vícerozměrné veličiny by zabralo čas. Zkrátka a dobře, tímto příkladem v našem krátkém šestipřednáškovém kursu pravděpodobnosti dvourozměrné veličiny opustíme. Spokojíme se s faktem, že existují, a necháme je na pokoji.

V právě uvedeném příkladu jsme už nepoužili klasickou pravděpodobnost, ale jakési její přirozené rozšíření - geometrickou pravděpodobnost. Protože počty prvků množin A a Ω jsou nekonečné, nelze je dosazovat do zlomku. Ale pokud místo počtu prvků dosazujeme míry množin, podíl

$$P(A) = \frac{m(A)}{m(\Omega)}$$

má vlastnosti pravděpodobnosti. **Geometrickou pravděpodobnost jevu A definujeme jako podíl míry množiny příznivých výsledků (=míry množiny A) a míry množiny všech možných výsledků (= míry množiny Ω).** Vzhledem k tomu, že míra množiny je velmi složitý pojem, jehož přesné zavedení by zabralo i několik přednášek, spokojme se s tvrzením, že mírou intervalu rozumíme jeho délku, mírou části roviny rozumíme její obsah a mírou části prostoru její objem.

Ovšem nesmíme zde zapomenout zdůraznit (pěkně do rámečku), ve kterých případech lze geometrickou pravděpodobnost použít:

Geometrickou pravděpodobnost můžeme užít jen tehdy, když Ω (= množina všech možných výsledků pokusu) je nekonečná a všechny výsledky hry nebo pokusu nastávají se stejnou pravděpodobností (= jsou stejně pravděpodobné).

Přiznávám se, že v právě uvedeném rámečku jsem se dopustil nepřesnosti ve slově „nekonečná“. Množina Ω musí být nespočetně nekonečná oblast kladné míry (pojem *oblasti* viz 1.ročník - diferenciální a integrální počet funkcí více proměnných),

V obou dosud uvažovaných modelech se vyskytovala důležitá podmínka, že každé dva různé výsledky jisté situace musí být stejně pravděpodobné. To ovšem někdy není skutečností, a díky tomu vznikly další dva modely pro popis pravděpodobnosti.

8.3 Diskrétní pravděpodobnost

Už jsme v teorii pravděpodobnosti tak zběhlí, že můžeme začít i třeba něčím tak důležitým, jako je rámeček:

Diskrétní pravděpodobnost můžeme užít tehdy, když Ω (= počet všech možných výsledků pokusu) je konečná ($\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$) nebo spočetná ($\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \omega_{n+1}, \dots\}$); přitom výsledky w_i nemusí nastat se stejnou pravděpodobností. Musí ovšem vždy platit, že $\sum_{\omega_i \in \Omega} P(\omega_i) = 1$.

Jednotlivé elementární výsledky experimentu v případě diskrétní pravděpodobnosti mohou, ale nemusí být stejně pravděpodobné. **Diskrétní pravděpodobnost jevu A definujeme jako součet pravděpodobností těch elementárních jevů ω_i , které jsou**

prvky množiny A :

$$P(A) = \sum_{\omega_i \in A} P(\omega_i).$$

Takto zavedená funkce P na podmnožinách množiny Ω splňuje vlastnosti (i), (ii), (iii) z úvodu této kapitoly, a je to tedy pravděpodobnost. Například třeba platí $P(\Omega) = \sum_{\omega_i \in \Omega} P(\omega_i) = 1$, což je součást vlastnosti (ii).

Příklad 8.12 *Pravděpodobnost, že zařízení pracuje celý den bez poruchy, je rovna $\frac{1}{5}$. Tato pravděpodobnost je stejná každý den a nezávisí na tom, zda ve dnech předchozích došlo k poruše nebo ne. Pravděpodobnost, že v některý den dojde k poruše, vyčerpává všechny ostatní situace, které mohou ten den nastat kromě bezporuchového provozu, a je tudíž rovna $1 - \frac{1}{5}$, což je $\frac{4}{5}$. Náhodná veličina X udává počet dnů nutný k tomu, aby nastala první porucha (sleduje tedy spolehlivost zařízení - hodnoty veličiny X snížené o jedničku nám říkají, kolik dnů zařízení pracovalo bez poruchy).*

- a) *Určete rozdělení veličiny X (tj. určete elementární jevy ω_i a jejich pravděpodobnosti $P(\omega_i)$).*
- b) *Vypočítejte pravděpodobnost, že k poruše zařízení nedojde prvních pět dní jeho provozu.*

Řešení:

- ad a)** *Nejnižší možná hodnota veličiny X , kterou můžeme naměřit, je hodnota 1, a to tehdy, když k poruše zařízení dojde už první den provozu. To může nastat s pravděpodobností $\frac{4}{5}$. Tento fakt budeme zkráceně zapisovat*

$$P(X = 1) = \frac{4}{5} = 0.8 \text{ (čti: pravděpodobnost, že } X \text{ nabude hodnoty 1, je rovna } \frac{4}{5}\text{)}.$$

Dále může veličina X nabýt hodnoty 2 - a to tehdy, když první den nedojde k poruše (to nastane s pravděpodobností $\frac{1}{5}$), ale druhý den ano (a sice s pravděpodobností $\frac{4}{5}$). Tedy výsledná pravděpodobnost této situace je rovna součinu pravděpodobností v jednotlivých dnech:

$$P(X = 2) = \frac{1}{5} \cdot \frac{4}{5} = 0.16.$$

Samozřejmě se také může stát, že naměříme hodnotu $X = 3$, a sice s pravděpodobností $\frac{1}{5}$ (že první den nedojde k poruše) krát $\frac{1}{5}$ (že druhý den nedojde k poruše) krát $\frac{4}{5}$ (že k poruše dojde třetí den). A tak

$$P(X = 3) = \frac{1}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} = 0.032.$$

Teoreticky je prostě možné, že veličina X nabude jakékoli přirozené hodnoty k , a sice s pravděpodobností

$$P(X = k) = \underbrace{\frac{1}{5} \cdot \frac{1}{5} \cdot \dots \cdot \frac{1}{5}}_{(k-1) \text{ krát}} \cdot \frac{4}{5} = \left(\frac{1}{5}\right)^{k-1} \cdot \frac{4}{5}.$$

Například pravděpodobnost, že veličina X nabude hodnoty 100 (tj. k první poruše dojde až po 100 dnech provozu) je sice hodně malá ($P(X = 100) = 6.3 \cdot 10^{-70}$), ale stále ještě různá od nuly.

Právě jsme popsali rozdělení veličiny, kde jednotlivé elementární hodnoty $1, 2, 3, 4, \dots$ nastávají s různou pravděpodobností. Těchto hodnot je nekonečně mnoho a víme, že musí splňovat vztah

$$\sum_{k=1}^{\infty} P(X = k) = 1,$$

protože pravděpodobnost všech možných případů, které mohou při měření veličiny X nastat, je vždy rovna jedné - to je jedna ze základních vlastností pravděpodobnosti.

*Veličina X se nazývá **diskrétní náhodná veličina** - nikoliv proto, že je nenápadná, ale že nabývá tzv. diskrétních hodnot, což jsou například takové hodnoty, které se liší o násobek určité konstanty (v našem případě konstanty 1). Funkce, jejíž hodnoty jsme právě určili, se nazývá **pravděpodobnostní funkce** a označuje se většinou $p(x)$, což je ještě více zkrácený zápis:*

$$p(x) = P(X = x)$$

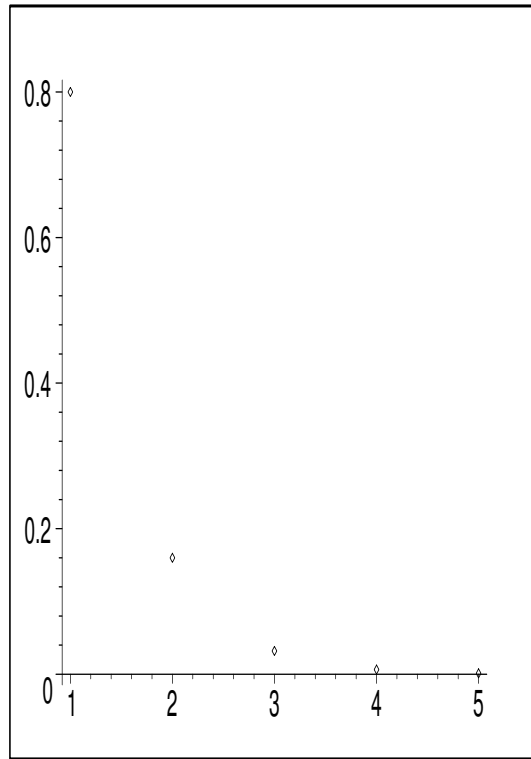
(čti: pravděpodobnost, že „velké X “ nabývá hodnoty „malé x “). Od nynějška tedy záleží na tom, zda je napsáno velké X (kterým budeme mít na mysli veličinu X) nebo malé x (označující jednu konkrétní hodnotu veličiny „velké X “). V našem případě

$$p(x) = \begin{cases} \left(\frac{1}{5}\right)^{x-1} \cdot \frac{4}{5} & \text{pro } x \in \{1, 2, 3, \dots\} \\ 0 & \text{jinak.} \end{cases}$$

Na obrázku 8.3 je vidět, že hodnoty jednotlivých pravděpodobností se pro rostoucí x blíží rychle k nule. Pokud zaokrouhlujeme výsledky na tři desetinná místa (což je přesnost postačující pro pravděpodobnostní výpočty), už pro $x \geq 6$ je $p(x)$ prakticky rovno nule.

*Pro popis rozdělení náhodných veličin se definuje tzv. **distribuční funkce** $F(x)$ předpisem*

$$F(x) = P(X < x).$$



Obrázek 8.3: K př. 8.12: Hodnoty pravděpodobnostní funkce $p(x)$.

Aby nedošlo k nedorozumění, tento vztah čteme: hodnota funkce F v bodě „malé x “ je rovna pravděpodobnosti, že náhodná veličina „velké X “ nabude hodnoty menší než „malé x “, tj. hodnoty z intervalu $(-\infty, x)$.

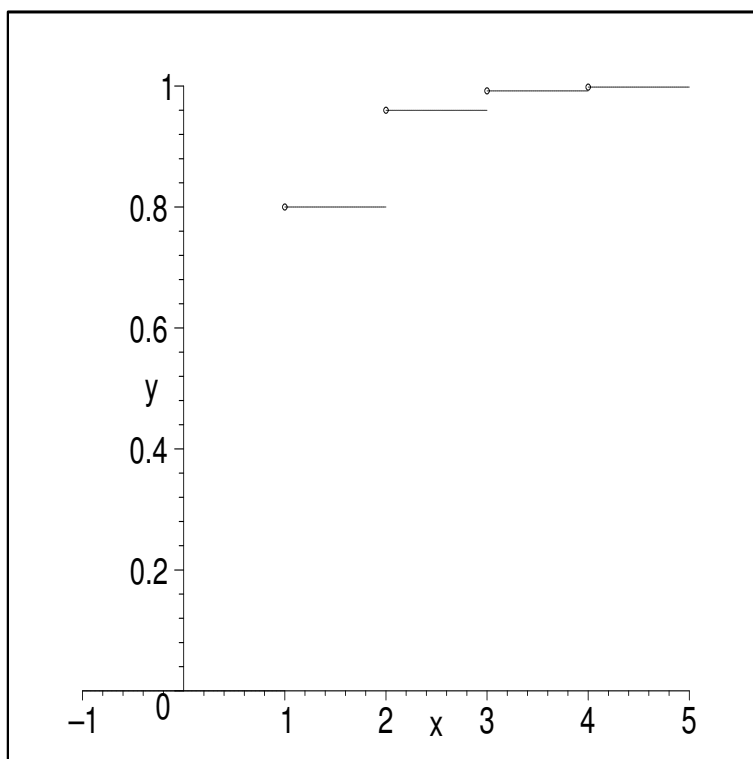
Pro diskrétní veličinu lze dosadit do pravé strany tohoto definičního vztahu:

$$F(x) = P(X < x) = \sum_{k < x} p(k).$$

Distribuční funkce v našem příkladu je zachycena na obrázku 8.4.

U diskrétní veličiny je distribuční funkce schodového tvaru - jedná se o funkci, která je po částech konstantní, pouze v bodech $1, 2, 3, \dots$ dochází ke změně (ke schodu), kde velikost změny (= výška schodu) v bodě k je rovna právě hodnotě $p(k)$. Body vyznačené na levém konci každého ze schodů prázdným kolečkem (na obrázku to nelze rozeznat - jsou tam jen jakési černé tečky) naznačují, že funkční hodnota distribuční funkce v bodě schodu je definována ne v bodě prázdného kolečka, ale dole u paty nižšího schodu (ještě nezvýšená). Například $F(2) = 0.8$. Distribuční funkce je tedy zleva spojitá funkce, tj.

$$\lim_{x \rightarrow k} F(x) = F(k),$$



Obrázek 8.4: K př. 8.12: Graf distribuční funkce $F(x)$ diskrétního rozdělení.

kde k je bod, ve kterém dochází ke změně výšky schodu. V našem příkladu se jedná o nekonečně dlouhé schodiště, ale většina z nekonečně mnoha schodů (to už obrázek nezachycuje, ale jsou tam) za pátým schodem má velmi malou výšku.

Rozdělení pravděpodobnosti v tomto příkladu má i svůj název - je to tzv. geometrické rozdělení s parametrem p (pozor, je to něco jiného než geometrická pravděpodobnost - geometrická pravděpodobnost je obecný název pro celou třídu pravděpodobnostních modelů u nespočetně mnoha možných výsledků, z nichž každý nastane se stejnou pravděpodobností, kdežto geometrická pravděpodobnost je konkrétní model diskrétního typu pro nejvýše spočetně mnoho různých výsledků nastávajících s různou pravděpodobností - toto názvosloví má svůj původ v historii a ustálilo se tak, a proto by si tyto pojmy nikdo neměl zaměnit).

ad b) *Máme určit pravděpodobnost, že k poruše dojde nejdříve šestý den od zahájení provozu. To znamená, že k první poruše může dojít šestý den, sedmý den, osmý den nebo kdykoliv později. Hledaná pravděpodobnost se tedy rovná*

$$p = p(6) + p(7) + p(8) + \dots,$$

zkrátka a dobře se jedná o součet nekonečné řady. Nekonečnou řadu někdy není snadné sečíst - to potvrdí každý, kdo se o to někdy pokoušel. Ale v našem případě

využijeme faktu, že součet všech nenulových hodnot pravděpodobnostní funkce je roven jedné, a místo sečítání nekonečné řady odečteme od hodnoty 1 pravděpodobnosti těch elementárních jevů, které v této řadě nejsou obsaženy:

$$p = \sum_{k=6}^{\infty} p(k) = 1 - \sum_{k=1}^5 p(k) = 1 - (0.8 + 0.16 + 0.032 + 0.0064 + 0.00128) = 0.00032.$$

Vidíme tedy, že pravděpodobnost, že k první poruše dojde nejdříve šestý den, je skutečně malá. Nicméně fintu s odečtením zbývajících pravděpodobností od jedničky si můžeme pamatovat - hodí se vždy, když tím ušetříme počet dosazení do pravděpodobnostní funkce (a využíváme ji i v případech, kdy diskrétní veličina nenabývá spočetného, ale jen konečného počtu hodnot).

8.4 Spojitá pravděpodobnost

Některé veličiny nenabývají diskrétní hodnoty, ale hodnoty z určitého intervalu reálných čísel. Například při měření veličiny udávající teplotu vzduchu můžeme naměřit libovolnou reálnou hodnotu z intervalu 0 až 25°C (jsme omezeni pouze přesností svého teploměru). Veličiny nabývající hodnoty z jistého intervalu se nazývají **spojité náhodné veličiny**. A jejich pravděpodobnostní zákonitosti popisuje spojitě rozdělení pravděpodobnosti.

Spojitě rozdělení k popisu veličiny X můžeme užít tehdy, když X nabývá hodnot z množiny Ω , která je nespočetně nekonečná (zpravidla $\Omega = \mathbf{R}$); přitom jednotlivých hodnot nemusí nabývat se stejnou pravděpodobností; různost, s jakou nabývá jednotlivých hodnot, je určena funkcí $f(x)$, které říkáme **hustota**. Musí přitom vždy platit, že

$$\int_{\Omega} f(x) dx = 1.$$

Spojitou pravděpodobnost jevu, že veličina X nabude hodnoty z intervalu $\langle a, b \rangle$, kde $a \leq b$, definujeme jako integrál z hustoty:

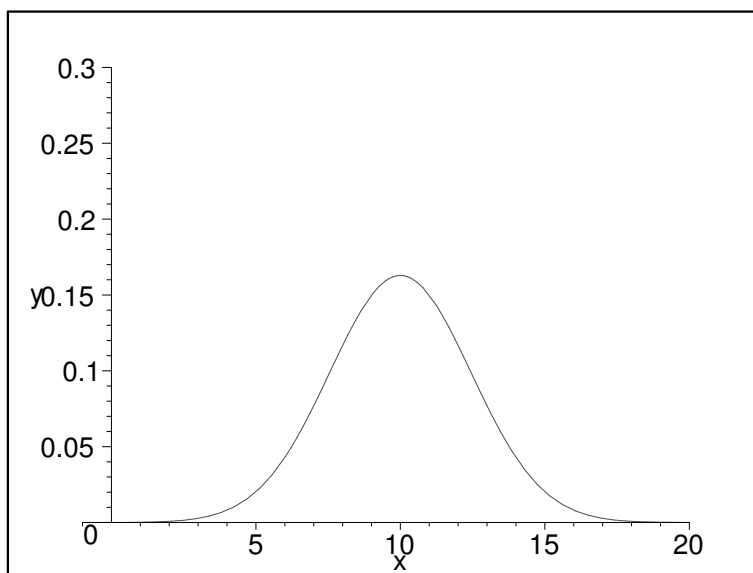
$$P(X \in \langle a, b \rangle) = \int_a^b f(x) dx.$$

Příklad 8.13 Životnost jistého druhu velmi speciálních žárovek je spojitá náhodná veličina s hustotou

$$f(x) = \frac{1}{\sqrt{12\pi}} \cdot e^{-\frac{(x-10)^2}{12}}.$$

Vypočtete pravděpodobnost, že koupená žárovka vydrží

- a) 9 až 12 hodin provozu.
- b) přesně 10 hodin provozu.



Obrázek 8.5: K př. 8.13: Graf hustoty $f(x)$ spojitého rozdělení.

Řešení: Uvedené rozdělení má svůj název - je to tzv. normální rozdělení pravděpodobnosti a jeho hustota je uvedena na obrázku 8.5.

Křivce hustoty se někdy říká Gaussova křivka, protože za jejího objevitele je považován německý matematik, fyzik, geofyzik a astronom Carl Friedrich Gauss (1777 - 1855). Říká se, že tento člověk předběhl svou dobu. A skutečně, je obdivuhodné, jak mohl najít funkci tak podivného vzorce a krásného vzezření, která nabývá nenulové hodnoty pro každé reálné číslo, a přesto je integrál z ní roven jedné. Až dopíšu tato skripta, zkusím v knihovně najít, jak k tomu došlo.

Řešení:

ad a) Označme X veličinu udávající životnost žárovky. Pak

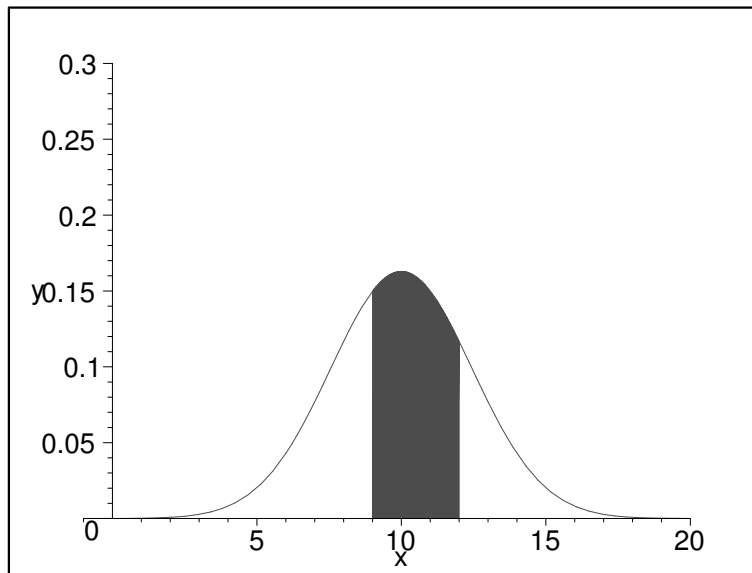
$$P(X \in \langle 9; 12 \rangle) = \int_9^{12} f(x) dx = 0.453.$$

Z matematické analýzy všichni vědí, že určitý (Riemannův) integrál z nezáporné funkce je roven obsahu plochy pod grafem funkce na daném intervalu. Platí to i v tomto případě - vypočtená pravděpodobnost je rovna obsahu šrafované plochy na obrázku 8.6.

Vlastnímu výpočtu integrálu se budeme věnovat až v kapitole 12, která se zabývá normálním rozdělením hlouběji. Zde se spokojíme pouze s výsledkem.

Podobně jako u diskrétní pravděpodobnosti i zde se definuje distribuční funkce, a sice stejným způsobem:

$$F(x) = P(X < x).$$



Obrázek 8.6: K př.8.13: Pravděpodobnost u spojitě veličiny je rovna obsahu šrafované plochy.

Nyní ovšem se ke konkrétnímu výpočtu funkční hodnoty užívá hustoty $f(x)$:

$$F(x) = P(X < x) = P(X \in (-\infty, x)) = \int_{-\infty}^x f(t)dt.$$

Mezi hustotou a distribuční funkcí u spojitěho rozdělení pravděpodobnosti platí zajímavý vztah, a sice hustota je derivací distribuční funkce:

$$F'(x) = f(x)$$

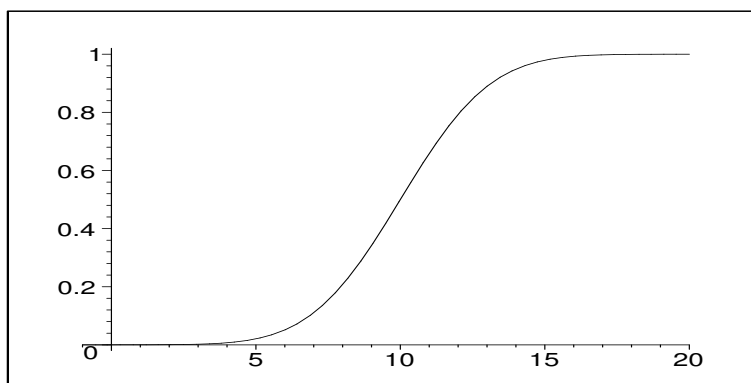
v těch bodech x , kde existuje derivace funkce $F(x)$. Graf distribuční funkce v našem příkladu je uveden na obrázku 8.7.

Podobně jako u diskrétního rozdělení, i u spojitěho rozdělení leží funkční hodnoty distribuční funkce v intervalu $< 0; 1 >$ (protože se jedná o hodnoty jisté pravděpodobnosti); dále platí

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

ad b) Podle části a) můžeme určit pravděpodobnost, že životnost žárovky bude přesně 10 hodin:

$$P(X = 10) = \int_{10}^{10} f(x)dx = 0.$$



Obrázek 8.7: K př.8.13: Graf distribuční funkce $F(x)$ daného normálního rozdělení.

Tímto se liší spojitá veličina od diskrétní veličiny: u diskrétní veličiny existuje nenulová pravděpodobnost, že X nabude konkrétní hodnoty. Kdežto u spojitě veličiny pravděpodobnost, že X nabude jisté konkrétní hodnoty, je vždy rovna nule. Do volte mi pokusit se vysvětlit tento jev. Jeho podstata tkví v integrálním počtu, ale vysvětlíme jej úvahou. Dejme tomu, že bychom chtěli mezi vyráběnými žárovkami najít některou, jejíž životnost je rovna přesně 10 hodin. Tuto dobu životnosti bychom měřili pomocí mechanických hodin se vteřinovou ručičkou (přesnost je na sekundy), stopkami (přesnost na setinu sekundy) a ještě jedním měřidlem přesnějším než stopky, které měří sekundy s přesností na 4 desetinná místa. Pokud bychom našli žárovku, jejíž životnost by byla 10 hodin měřená hodinkami s ručičkou, je dost malá pravděpodobnost, že by na stopkách nebyla žádná odchylka od 10 hodin v setinách sekundy. Ale i kdyby to nastalo, tak je dost málo pravděpodobné, že by na třetím měřidle nebyla odchylka při měření s přesností na 4 desetinná místa. Pokud bychom použili ještě přesnější měřidlo, pravděpodobnost, že při zvyšujícím se počtu desetinných míst přesnosti měření je životnost rovna přesně 10 hodin, je stále menší. Celkem můžeme uzavřít, že pravděpodobnost, že bychom našli žárovku s životností 10 hodin a přesností na nekonečně mnoho desetinných míst, je rovna nule.

Komu se toto vysvětlení stále ještě nezdá, musí se spokojit s konstatováním, že pravděpodobnost naměření životnosti přesně 10 hodin je hodně, hodně malá.

8.5 Shrnutí pojmů

Pokud výsledky jistého pokusu, hry nebo experimentu mohou nastat se stejnou pravděpodobností, používáme k jeho popisu klasickou (8.1) nebo geometrickou (8.2) pravděpodobnost. Ovšem pokud některé z elementárních výsledků nastávají častěji než jiné, situaci znázorníme pomocí diskrétní (8.3) nebo spojitě (8.4) pravděpodobnosti.

Naše exkurze po základních pravděpodobnostních modelech je u konce. Studovali jsme přitom vždy rozdělení jisté náhodné veličiny. Intuitivně je jasné, o co se jedná. Matematicky se náhodná veličina definuje jako jisté zobrazení:

Pokud S je množina jevů nad prostorem Ω , nazveme zobrazení $X : \mathbf{S} \rightarrow \mathbf{R}$ *náhodnou veličinou*, když pro libovolné $x_0 \in \mathbf{R}$ je množina $X^{-1}((-\infty, x_0))$ prvkem množiny \mathbf{S} (množinou $X^{-1}((-\infty, x_0))$ rozumíme sjednocení všech množin z \mathbf{S} , které zobrazení X zobrazí na hodnotu menší než x_0).

Nechci nyní trávit čas objasňováním této definice. Spokojím se s tím, že upozorním čtenáře na to, co od něj budu vyžadovat především. Když studujeme jistou veličinu, jako první věc bychom si měli uvědomit, zda se jedná o veličinu diskrétní (ta nabývá hodnot z konečné (např. $\{1, 2, 3, 4, 5, 6\}$) nebo spočetné (např. \mathbf{N} , \mathbf{Z}) množiny Ω) nebo spojitou (ta nabývá hodnot z reálného intervalu $\Omega = \langle a, b \rangle$ nebo z celé množiny reálných čísel). Popis těchto dvou typů veličin se totiž v některých věcech liší. A používané vzorce nebo způsob popisu se neustále odvíjí od jednoho z těchto dvou typů. V následujících kapitolách (a i v úlohách praxe) se potřebuje občas určit pravděpodobnost, že náhodná veličina nabývá hodnot z jistého intervalu $(a, b \rangle$. S ohledem na typ veličiny budeme užívat vzorec

$$P(X \in (a, b \rangle)) = P(a < X \leq b) = \begin{cases} \sum_{a < x \leq b, x \in \Omega} p(x) & \text{pro diskrétní veličinu } X, \\ \int_a^b f(x) dx & \text{pro spojitou veličinu } X. \end{cases} \quad (8.7)$$

Jak už bylo řečeno, v diskrétním případě funkce $p(x)$ se nazývá pravděpodobnostní funkce, ve spojitém případě funkce $f(x)$ hustota. U obou typů veličin se definuje tzv. distribuční funkce $F(x)$. Pokud známe její hodnoty, můžeme místo vzorce 8.7 použít u obou typů veličin

$$P(X \in (a, b \rangle) = F(b) - F(a). \quad (8.8)$$

9 Střední hodnota a rozptyl

Zatímco v kapitole 8 jsme se zabývali různými matematickými přístupy k pravděpodobnosti, nyní se podíváme zejména na zpracování konkrétních dat. Pokud získáme měřením soubor hodnot určité veličiny, existují různé metody, kterými naměřená data zpracováváme a popisujeme. K základnímu popisu patří průměr naměřených dat. Uvidíme, že průměr souvisí s pojmem střední hodnoty v teorii pravděpodobnosti. Dále se seznámíme s některými dalšími charakteristikami naměřených dat, mezi nimiž je nejdůležitější tzv. rozptyl. Důležitou součástí této kapitoly jsou také další kroky v objasňování rozdílu mezi teorií a praxí - vztahu pravděpodobnosti a statistiky.

9.1 Empirické a teoretické rozdělení pravděpodobnosti

Dříve než přistoupíme ke konkrétnímu popisu souboru naměřených dat, je důležité si uvědomit rozdíl a souvislost mezi empirickým a teoretickým rozdělením pravděpodobnosti. Vysvětlíme ji na následujících dvou příkladech. **Empirické rozdělení pravděpodobnosti** je to rozdělení, které získáme z naměřených dat (z anglického empire = říše, impérium; tedy empirické rozdělení popisuje konkrétní měření - jak nám v tom našem českém impériu hážou kostky, mince, porouchávají se zařízení, apod.)

Příklad 9.1 *Byla získána data tím způsobem, že každá z dvaceti osob hodila čtyřikrát korunou. V tabulce 9.1 jsou zaznamenány počty líců ve čtyřech hodech u každé z osob. Určete empirické rozdělení pravděpodobnosti veličiny X .*

Tabulka 9.1: K př. 9.1: Naměřené hodnoty veličiny X .

osoba	1	2	3	4	5	6	7	8	9	10
X -hodnota	3	1	1	3	1	2	0	2	4	4
osoba	11	12	13	14	15	16	17	18	19	20
X -hodnota	1	2	2	1	2	1	2	3	3	3

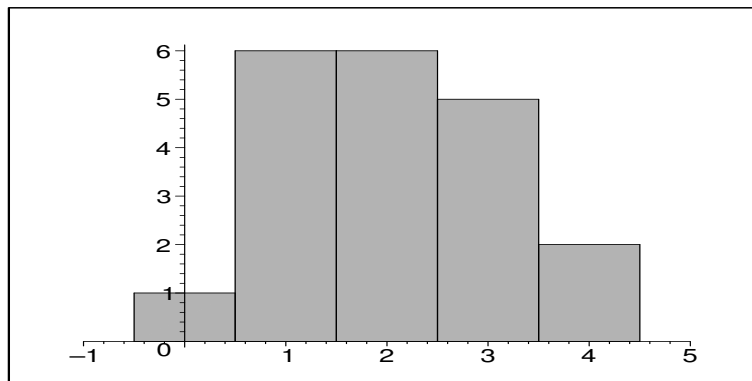
Řešení: Nejprve si všimněme, že naše veličina X je diskrétní, protože nabývá pouze pěti hodnot - 0, 1, 2, 3 nebo 4. Zpracování této úlohy je založeno na pojmu **četnost**, který udává počet výskytů dané hodnoty v našem souboru. Například ze všech dvaceti měření je jen jedna hodnota 0, tj. veličina X nabývá hodnoty 0 s četností 1 (budeme značit $c(0) = 1$). Hodnota 1 se vyskytuje s četností 6, atd. Všechny četnosti jsou zaznamenány v tabulce 9.2:

Musí platit jednoduchá kontrola, že součet všech četností ve druhém řádku tabulky je roven počtu hodnot (v našem případě 20), protože každou hodnotu jsme počítali právě jednou.

Tabulka 9.2: K př. 9.1: Tabulka empirických četností hodnot veličiny X .

X -hodnota	0	1	2	3	4
četnost	1	6	6	5	2

Uvedené četnosti lze také znázornit v tzv. **histogramu četností** - viz obr. 9.1, kde výšky jednotlivých obdélníčků jsou rovny konkrétním četnostem a délka základny každého z obdélníčků je rovna 1.

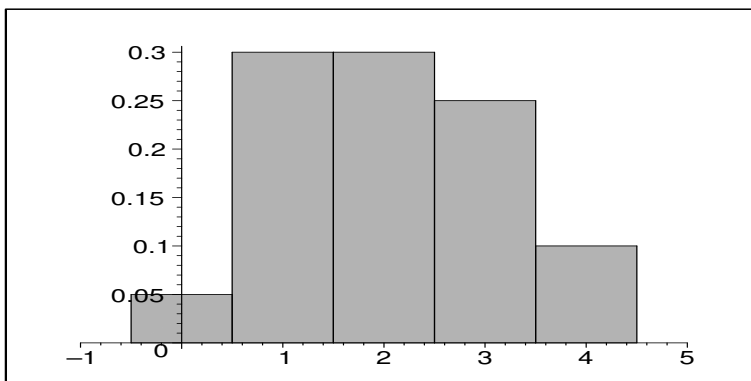
**Obrázek 9.1:** K příkladu 9.1: Histogram četností veličiny X .

K určení empirického rozdělení pravděpodobnosti nám zbývá poslední krok - vydělit četnosti délkou souboru (= počtem hodnot), v našem případě číslem 20. Tak dostaneme tabulku 9.3 relativních četností vzhledem k počtu měření.

Tabulka 9.3: K př. 9.1: Funkce $p(x)$ empirického rozdělení pravděpodobnosti veličiny X .

X -hodnota	0	1	2	3	4
$p(x)$	0.05	0.3	0.3	0.25	0.1

Součet těchto relativních četností je roven jedné, jsou tedy splněny všechny podmínky diskrétní pravděpodobnosti - našli jsme pravděpodobnostní funkci $p(x)$ tohoto rozdělení. Při zpracování dat se někdy místo pravděpodobnostní funkce užívá grafického znázornění v podobě **histogramu pravděpodobnosti** (pravděpodobnostního histogramu) viz obr. 9.2. Jediný rozdíl mezi obrázkem 9.1 a 9.2 je v tom, že v prvním případě se na osu y nanáší hodnoty četností a ve druhém případě pravděpodobnosti. Na pravděpodobnostním histogramu



Obrázek 9.2: K př. 9.1: Histogram pravděpodobností veličiny X .

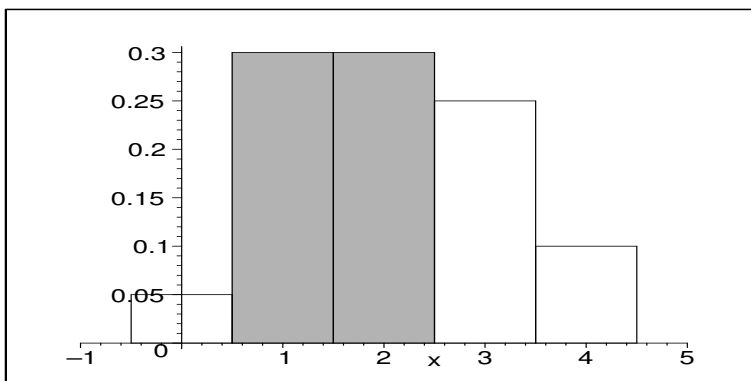
je zajímavé to, že součet obsahů všech obdélníků na obrázku je roven jedné, čili jedná se o jakýsi geometrický model analogický situaci spojité pravděpodobnosti, kde víme, že platí

$$P(X \in \langle a, b \rangle) = \int_a^b f(x)dx = \text{obsah plochy pod křivkou } f(x) \text{ na intervalu } \langle a, b \rangle.$$

Pokud chceme s využitím histogramu pravděpodobnosti v našem diskrétním případě vyčíslit třeba pravděpodobnost, že při 4 hodech mincí padl líc jednou nebo dvakrát, dostáváme

$$P(X \in \langle 1, 2 \rangle) = P(X = 1) + P(X = 2) = 0.3 + 0.3 = 0.6,$$

což je rovno součtu obsahů obdélníků histogramu nad hodnotami 1 a 2 (viz obr. 9.3).

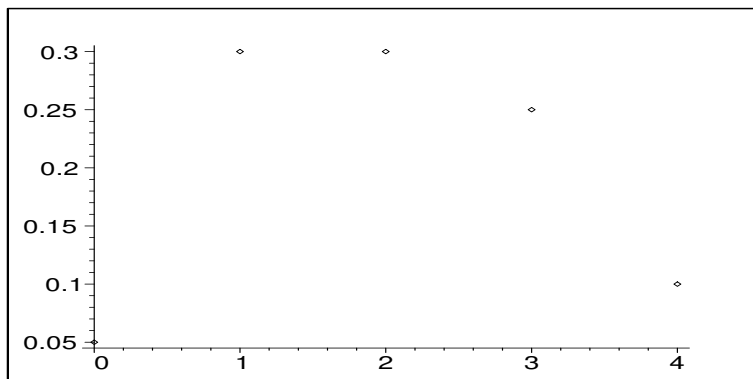


Obrázek 9.3: K př. 9.1: I v diskrétním případě lze pravděpodobnost vyjádřit jako obsah jisté plochy.

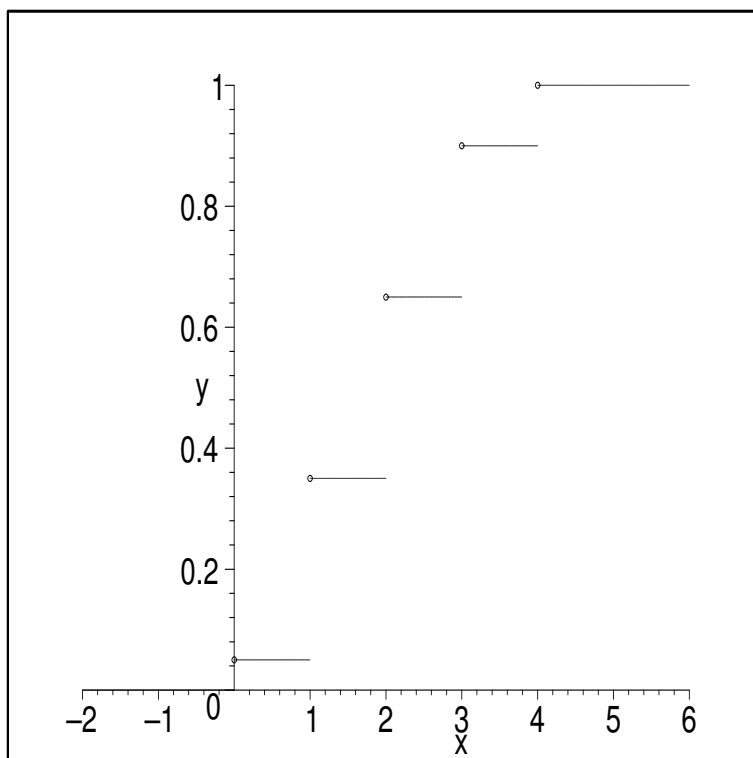
Pokud tedy uvažujeme u spojité veličiny hustotu a u diskrétní veličiny histogram pravděpodobností, lze v obou případech vyjádřit pravděpodobnost, že veličina X nabude hodnot z jistého intervalu, jako obsah určité plochy (v případě histogramu musí platit důležitý předpoklad, který zde ještě jednou připomenou: základna každého z elementárních obdélníků histogramu

musí mít délku 1).

Můžeme také pro formu nakreslit graf pravděpodobnostní funkce $p(x)$ (obrázek 9.4), popřípadě graf distribuční funkce $F(x)$ (9.5). V tomto případě se distribuční funkce skládá z pěti schodů, z nichž ten poslední má výšku 1 a nekonečnou délku.



Obrázek 9.4: K př. 9.1: Graf pravděpodobnostní funkce $p(x)$.



Obrázek 9.5: K př. 9.1: Graf distribuční funkce $F(x)$ rozdělení veličiny X .

Veličina X v tomto příkladu je možná ještě vhodnějším reprezentantem diskrétního rozdělení než veličina z příkladu 8.12, protože nabývá konečně mnoha hodnot s různou pravděpodobností (aby si někdo po absolvování příkladu 8.12 nemyslel, že diskrétní veličina může nabývat jen nekonečně mnoha hodnot).

Pojem **teoretické rozdělení pravděpodobnosti** je asi každému jasný - určíme rozdělení teoreticky, nikoliv na základě měření. Ale zajímavé bude nalézt teoretické rozdělení ve stejné situaci, kterou jsme právě uvažovali.

Příklad 9.2 Nalezněte teoretické rozdělení veličiny X , která udává počet líců při čtyřech hodech mincí.

Řešení: Podrobíme naši situaci teoretickým úvahám za předpokladu, že mince je vyvážená a vyrobená ze stejnorodého materiálu. V tabulce 9.4 jsou uvedeny všechny možné výsledky čtyř hodů mincí (druhý sloupec udává vždy počet líců v dané variantě):

Tabulka 9.4: K př. 9.2: přehled všech možných výsledků při čtyřech hodech mincí.

výsledek	počet líců	výsledek	počet líců
LLLL	4	LRRL	2
LLLR	3	RLRL	2
LLRL	3	RRLR	2
LRLR	3	LRRR	1
RLLL	3	RLRR	1
LLRR	2	RRLR	1
LRLR	2	RRRL	1
RLLR	2	RRRR	0

Bystrému pozorovateli asi neušlo, že všech možných výsledků je 16. A protože líc padá s pravděpodobností $\frac{1}{2}$, každý z těchto 16 výsledků je stejně pravděpodobný. A proto můžeme z tabulky určit četnosti počtu líců (viz tabulka 9.5)

Tabulka 9.5: K př. 9.2: Tabulka teoretických četností hodnot veličiny X .

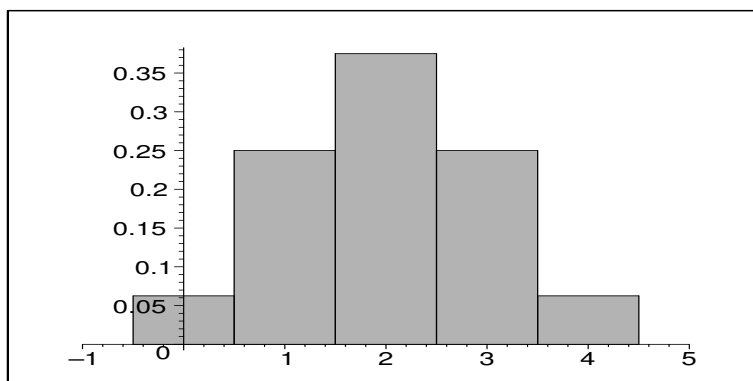
X -hodnota	0	1	2	3	4
četnost	1	4	6	4	1

Tabulka 9.6: K př. 9.2: Funkce $p(x)$ teoretického rozdělení pravděpodobnosti veličiny X .

X -hodnota	0	1	2	3	4
$p(x)$	0.0625	0.25	0.375	0.25	0.0625

a vydělením hodnotou 16 pak i relativní četnosti, které už jsou hodnotami hledané teoretické pravděpodobnostní funkce $p(x)$ (viz tabulka 9.6).

Příslušný histogram pravděpodobnosti je znázorněn na obrázku 9.6.

**Obrázek 9.6:** K př. 9.2: Histogram pravděpodobnosti teoretického rozdělení veličiny X .

K teoretickému rozdělení pravděpodobnosti v příkladu 9.2 lze jednoduše sestrojít teoretické rozdělení četnosti, a dokonce si můžeme vybrat, kolikrát se má experiment „prakticky“ provádět. Například pro 128 opakování experimentu čtyř hodů mincí má teoretické rozdělení četnosti stejný tvar jako pravděpodobnostní histogram 9.6, jen na osu y vynášíme hodnoty reprezentující četnost $c(i)$ (obrázek zde už není uveden, od 9.6 se liší jen měřítkem svislé osy):

$$c(0) = p(0) \cdot 128 = 0.0625 \cdot 128 = 8$$

$$c(1) = p(1) \cdot 128 = 0.25 \cdot 128 = 32$$

$$c(2) = p(2) \cdot 128 = 0.375 \cdot 128 = 48$$

$$c(3) = p(3) \cdot 128 = 0.25 \cdot 128 = 32$$

$$c(4) = p(4) \cdot 128 = 0.0625 \cdot 128 = 8$$

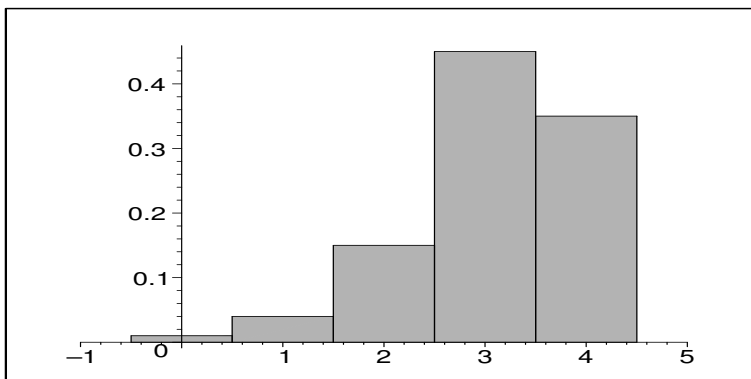
Čili kdybychom učinili 128 pokusů, z nichž jeden sestává ze čtyř hodů mincí, náš nejlepší teoretický odhad je ten, že v 8 pokusech by nepadl žádný líc, ve 32 pokusech jeden líc, atd.

Teoretické rozdělení pravděpodobnosti je jakési očekávané rozdělení, které nastane za jistých předpokladů. Například při pokusu 4 hodů mincí těmito předpoklady jsou:

- Mince je vyrobena tak, že rub a líc padá se stejnou pravděpodobností.
- Mincí je házeno „normálně“, ne nějakým divným stylem, který by zvýhodňoval buď rub, nebo líc.
- Každý účastník pokusu pravdivě nahlásí své výsledky.

Rozdělení získané empiricky v příkladu 9.1 „zhruba“ odpovídá teoretickému rozdělení z příkladu 9.2. Zdá se tedy rozumné uzavřít, že se světem je všechno v pořádku: mince je pravděpodobně dobře vyvážená, lidé jí házou dobrým způsobem a nahlašují výsledky poctivě.

Pokud by data z příkladu 9.1 vedla na empirické rozdělení pravděpodobnosti uvedené na obrázku 9.7,



Obrázek 9.7: K příkladům 9.1, 9.2: Empirický histogram pravděpodobností veličiny X , který se hodně liší od teoretického.

bylo by patrné, že tři nebo čtyři líce padaly ve čtyřech hodech mnohem častěji, než jsme očekávali, na úkor výsledků 0 líců, 1 líc, 2 líce. To by zpochybnilo některý z našich předpokladů. Uzavřeli bychom, že buď je mince nějak divně vyvážená, nebo lidé jí házejí divným stylem.

V tom tedy tkví podstata statistického usuzování: Před experimentem se určí, jaký tvar má za jistých předpokladů teoretické rozdělení pravděpodobnosti. Pak se provede experiment a z naměřených dat získáme empirické rozdělení pravděpodobnosti. Jestliže se teoretické a empirické rozdělení shodují, uzavíráme, že předpoklady, které jsme učinili, jsou pravděpodobně správné. Na druhé straně, když se teoretické rozdělení od empirického

významně liší, uzavíráme, že jeden nebo více předpokladů je pravděpodobně nesprávných. Podrobněji o tom bude řeč při konkrétních statistických testech v následujících kapitolách. Zde byly uvedeny jen příklady vysvětlující, k čemu pravděpodobnostní rozdělení slouží.

9.2 Empirické charakteristiky popisu dat

Pustíme se nyní už do konkrétního zpracování naměřených dat. Pokud máme jistý počet měření veličiny, lze z těchto měření určit následující jednoduché charakteristiky:

- **Průměr** z naměřených hodnot x_1, x_2, \dots, x_n : $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$. Označení \bar{X} je celkem standardní a používá se ve fyzice i dalších vědách k vyjádření průměrné hodnoty.
- **Medián** z hodnot x_1, x_2, \dots, x_n je prostřední z těchto hodnot vzhledem k jejich uspořádání podle velikosti.
- **Modus** z hodnot x_1, x_2, \dots, x_n je ta z hodnot, která se vyskytuje s nejvyšší četností.

Příklad 9.3 Jsou získány výsledky kvizového skóre u 11 osob:

8, 5, 7, 9, 8, 1, 3, 4, 7, 7, 7.

Průměr těchto hodnot je $\bar{X} = \frac{1}{11} \cdot \sum x_i = \frac{66}{11} = 6$. Modus tohoto souboru je hodnota, která se vyskytuje nejčastěji, čili číslo 7. A abychom mohli určit medián, musíme hodnoty seřadit podle velikosti (například vzestupně):

1, 3, 4, 5, 7, 7, 7, 7, 8, 8, 9.

Prostřední z těchto hodnot je na šesté pozici, čili mediánem je číslo 7.

Příklad 9.4 Mějme jiný soubor hodnot, už uspořádaný podle velikosti, například sestupně:

7, 6, 5, 5, 4, 2, 1, 1.

Protože počet měření je sudý (budeme též říkat, že soubor měření má sudou délku), medián určíme jako průměr dvou prostředních hodnot: $\frac{1}{2}(5 + 4) = 4.5$.

Příklad 9.5 Soubor měření může mít více modů (= druhý pád od slova modus). Například soubor

8, 6, 6, 5, 4, 3, 3

je tzv. bimodální soubor, protože nejčastěji (= dvakrát) se v něm objevují hodnoty 6 (=modus1) a 3 (=modus2). Při třech modech mluvíme o trimodálním, při čtyřech o kvatromodálním souboru, atd. Některé učebnice ignorují možnost více modů a za modus označují největší nejčastěji nabývanou hodnotu, což by v našem případě bylo 6.

Z uvedených tří charakteristik je většinou nejužitečnější průměr - až na následující příklad, kde se vyskytuje v souboru měření tzv. odkloněná hodnota, což je hodnota, která se hodně liší od všech ostatních.

Tabulka 9.7: K příkladu 9.6: Soubor měření získaných v experimentu.

otázka	odpověď	doba reakce (v sekundách)
ovoce na „h”	hruška	0.6
stát na „F”	Francie	0.4
mužské jméno na „H”	Horymír	0.6
roční období na „p”	podzim	0.7
část těla na „z”	záda	10.0

Příklad 9.6 Uvažujme experiment, ve kterém měříme čas reakce náhodně vybraného studenta na otázku, respektive čas, který uběhne mezi naší otázkou a jeho odpovědí. Průběh experimentu je zaznamenán v tabulce 9.7.

Z naměřených dat $\bar{X} = 2.46$, $\text{modus} = \text{medián} = 0.6$. Velký rozdíl mezi mediánem a průměrem je způsoben odkloněnou hodnotou 10.0. V tomto případě je k popisu souboru měření užitečnější užít medián (anebo odkloněnou hodnotu vypustit, a pak teprve spočítat průměr).

Kromě průměru nás často zajímá, jakým způsobem se data od průměru liší, tj. jak velká je odchylka hodnot od průměru. Lze určovat různé typy odchylek - podívejme se na ně pro konkrétní data.

Příklad 9.7 Uvažujme soubor měření z příkladu 9.3. Pro tato data se definují různé typy odchylek uvedené v tabulce 9.8.

Pro každou hodnotu měření x_i lze určit její **odchylku** od průměru $x_i - \bar{X}$, **absolutní hodnotu této odchylky** $|x_i - \bar{X}|$ (tzv. **absolutní odchylku**) a **kvadratickou odchylku** $(x_i - \bar{X})^2$. Nám by se ovšem kromě průměru \bar{X} získaného ze všech hodnot v souboru hodila další míra odchýlení od průměru vypočtená ze všech hodnot souboru najednou.

Touto mírou odchýlení od průměru nemůže být **průměrná odchylka** $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})$, protože ta je vždy rovna nule, čili žádnou informaci o rozptylu hodnot z něj nezískáme. Kdo tomu nevěří, ať upravuje spolu se mnou:

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X}) = \frac{1}{n} (\sum x_i) - \frac{1}{n} (\sum \bar{X}) = \bar{X} - \frac{1}{n} \cdot n \cdot \bar{X} = 0.$$

Dalším kandidátem na rozptyl je **průměrná absolutní odchylka** $\frac{1}{n} \sum |x_i - \bar{X}|$. V našem příkladu je rovna 2 a už sděluje jakousi informaci o rozptylu: náhodně vybraná hodnota měření je od průměru \bar{X} odchýlená asi o 2 jednotky. S touto měrou rozptylu se v některých matematických popisech už setkáváme. Ale vzhledem k tomu, že součet absolutních hodnot je obtížně matematicky zpracovatelný (např. obtížně se derivuje, apod.), nejčastějšího používání se těší **průměrná kvadratická odchylka** $\frac{1}{n} \sum (x_i - \bar{X})^2$. S ní se už čtenář

Tabulka 9.8: K příkladu 9.7: Různé typy odchylek od průměru.

x_i	$x_i - \bar{X}$	$ x_i - \bar{X} $	$(x_i - \bar{X})^2$
8	2	2	4
5	-1	1	1
7	1	1	1
9	3	3	9
8	2	2	4
1	-5	5	25
3	-3	3	9
4	-2	2	4
7	1	1	1
7	1	1	1
7	1	1	1

setkal v první části tohoto skriptu - u metody nejmenších čtverců. I v pravděpodobnosti a statistice se používá spíše tato míra odchýlení. Budeme ji označovat S^2 a nazývat **empirický rozptyl** (pokud bude ze souvislosti jasné, že se jedná o soubor empiricky získaných hodnot, slovo „empirický“ někdy vynecháme).

V našem příkladu $S^2 = 5.455$. Jedná se o veličinu, jejíž rozměr je vzhledem k získaným datům umocněný na druhou. Protože někdy budeme potřebovat charakteristiku stejného fyzikálního rozměru, označme

$S := \sqrt{S^2}$; veličina S se nazývá **empirická směrodatná odchylka**.

Pro naše data $S = \sqrt{5.455} = 2.336$. Dovídáme se tedy, že náhodně vybraná hodnota ze souboru je od průměru odchýlená asi o 2.336. Tato míra rozptýlení je tedy mírně vyšší než průměrná absolutní odchylka - u průměrné kvadratické odchylky můžeme tedy mluvit o větší velikosti.

Další mírou rozptylu hodnot může být intervalový rozsah hodnot $\langle x_{\min}, x_{\max} \rangle$. V našem příkladu vidíme, že $X \in \langle 1; 9 \rangle$.

Protošim budeme užívat zejména empirický rozptyl S^2 , na chvíli se u něj zastavme. Pokud budete spolu se mnou upravovat definiční vztah, dospějeme ke vzorci, který budeme pro výpočet S^2 používat:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2 = \frac{1}{n} \sum (x_i^2 - 2\bar{X} \cdot x_i + \bar{X}^2) =$$

$$\begin{aligned}
&= \frac{1}{n}(\sum x_i^2) - \frac{2\bar{X}}{n}(\sum x_i) + \frac{n}{n} \cdot \bar{X}^2 = \frac{1}{n}(\sum x_i^2) - 2\bar{X}^2 + \bar{X}^2 = \\
&= \frac{1}{n}(\sum x_i^2) - \bar{X}^2.
\end{aligned}$$

Poslední řádek odvození se někdy čte jako „průměr čtverců minus čtverec průměru“, což je i pomůckou k zapamatování vzorce.

Následující příklad je klíčovým příkladem této kapitoly - jsou zde uvedeny vzorce, které jsou podkladem teoretických charakteristik oddílu 9.3.

Příklad 9.8 *Náhodná veličina X udává počet líců při čtyřech hodech mincí. Měřením se získalo těchto dvacet hodnot veličiny:*

3, 1, 1, 3, 1, 2, 0, 2, 4, 4, 1, 2, 2, 1, 2, 1, 2, 3, 3, 3.

Určete průměr a empirický rozptyl souboru měření.

a) Klasické řešení: *Jedná se o stejná data jako v příkladu 9.1. Vypočteme průměr, empirický rozptyl i empirickou směrodatnou odchylku:*

$$\begin{aligned}
\bar{X} &= \frac{1}{20} \sum_1^{20} x_i = 2.05; \\
S^2 &= \frac{1}{20} \left(\sum_1^{20} x_i^2 \right) - 2.05^2 = 1.1475; \\
S &= \sqrt{1.1475} = 1.0712.
\end{aligned}$$

Vidíme tedy, že při čtyřech hodech mincí padalo průměrně 2.05 líců (hodnota průměru se ve statistice zpravidla nezaokrouhluje), přitom náhodně vybraná hodnota se od tohoto průměru odchyluje asi o 1.07 líců (tato hodnota se rovněž nezaokrouhluje).

b) Řešení pomocí rozdělení četnosti: *Máme li data zpracována v podobě četností - viz tabulka 9.9, kde ν_i jsou hodnoty, kterých veličina X nabývá (ν je písmeno řecké abecedy a čte se „ný“) -*

můžeme k výpočtu průměru a rozptylu dat využít vzorce, které obsahují četnosti:

$$\bar{X} = \frac{1}{n} \sum_{\nu_i} \nu_i \cdot c(\nu_i); \quad S^2 = \frac{1}{n} \left(\sum_{\nu_i} \nu_i^2 \cdot c(\nu_i) \right) - \bar{X}^2.$$

Dosažením se přesvědčíme, že dostaneme stejný výsledek jako v případě klasických vzorců (a ono je i vidět, že vzorce pro četnosti dostaneme z klasických vzorců jednoduchou úvahou - četnost $c(\nu_i)$ vyjadřuje, kolikrát se hodnota ν_i v souboru vyskytuje, a proto se jedná jen o přepsání jednoho a téhož vzorce).

Tabulka 9.9: K příkladu 9.8: Tabulka četností souboru měření veličiny X .

ν_i	ν_i^2	četnost $c(\nu_i)$
0	0	1
1	1	6
2	4	6
3	9	5
4	16	2

c) **Řešení pomocí rozdělení pravděpodobnosti:** *Sledujte se mnou následující úvahu: Vklouzneme-li se zlomkem $\frac{1}{n}$ ve vzorcích užívajících četnost z řešení b) za sumu, uvnitř dostaneme zlomky $\frac{c(\nu_i)}{n}$. Tyto zlomky vlastně vyjadřují relativní četnosti hodnot ν_i , tedy jejich empirické pravděpodobnosti:*

$$\frac{c(\nu_i)}{n} = p(\nu_i).$$

Odtud můžeme psát vzorce pro výpočet průměru a rozptylu ve tvaru

$$\begin{aligned}\bar{X} &= \sum_{\nu_i} \nu_i \cdot \frac{c(\nu_i)}{n} = \sum_{\nu_i} \nu_i \cdot p(\nu_i); \\ S^2 &= \left(\sum_{\nu_i} \nu_i^2 \cdot \frac{c(\nu_i)}{n} \right) - \bar{X}^2 = \left(\sum_{\nu_i} \nu_i^2 \cdot p(\nu_i) \right) - \bar{X}^2.\end{aligned}$$

S využitím tabulky 9.10 empirických pravděpodobností pak dosazením do těchto vzorců dostaneme tentýž výsledek jako v případě a) a b).

I v tomto případě se stále jedná o pouhé přepsání stejných vzorců a) nebo b) s využitím označení pomocí pravděpodobnosti.

9.3 Teoretické charakteristiky popisu dat

Někomu sem možná zdály vzorce z odstavce c) příkladu 9.8 příliš vykonstruované, ale tyto úvahy jsou základem pro definici charakteristik teoretického rozdělení pravděpodobnosti. Právě u teoretického rozdělení nemáme totiž k dispozici ani četnosti, ani měření, ale pouze teoretické pravděpodobnosti. Právě ty dosadíme do vzorců místo pravděpodobností empirických.

Tabulka 9.10: K příkladu 9.8: Tabulka empirických pravděpodobností.

ν_i	ν_i^2	$p(\nu_i)$
0	0	0.05
1	1	0.3
2	4	0.3
3	9	0.25
4	16	0.1

Uvažujme nejprve diskrétní náhodnou veličinu X . **Střední hodnotu** EX veličiny X definujeme vztahem

$$EX = \sum_{\nu_i} \nu_i \cdot p(\nu_i).$$

Označení pomocí písmene E pochází z anglického expected value (= očekávaná hodnota). Střední hodnota podle odstavce c) příkladu 9.8 tedy není nic jiného než průměr hodnot, které bychom získali při platnosti daných teoretických předpokladů. Je to tedy jakýsi teoretický průměr, který bychom získali při praktickém měření, kdyby měřená veličina odpovídala danému teoretickému popisu.

Rozptyl DX veličiny X definujeme jako střední hodnotu čtverce odchylky veličiny X od své střední hodnoty EX :

$$DX = E(X - EX)^2.$$

Označení pomocí písmene D pochází z anglického dispersion (=rozptyl). Jiné anglické slovo pro rozptyl je variance, odtud v některých učebnicích se rozptyl označuje jako $var X$. Ale my se v dalším budeme držet označení DX . Podívejme se definici rozptylu na zoubek: umocněním závorečky a dosazením za střední hodnotu veličin X^2 a X dostaneme

$$\begin{aligned} DX &= E(X - EX)^2 = E(X^2 - 2X \cdot EX + (EX)^2) = \\ &= EX^2 - 2EX \cdot EX + (EX)^2 = EX^2 - (EX)^2 = \\ &= \sum_{\nu_i} \nu_i^2 p(\nu_i) - \left(\sum_{\nu_i} \nu_i \cdot p(\nu_i) \right)^2. \end{aligned}$$

Druhý řádek odvození je právě vzorec pro výpočet S^2 z odstavce c) příkladu 9.8. Tj. rozptyl je definován naprosto přirozeně jako hodnota, kterou bychom vypočetli jako S^2 pro soubor měření veličiny, která by odpovídala teoretickému rozdělení.

Také analogicky definujeme **směrodatnou odchylku** veličiny X jako \sqrt{DX} .

Příklad 9.9 Vypočítejte střední hodnotu a rozptyl počtu líců ze čtyř hodů v příkladu 9.2.

Řešení. Dosazením do vzorce pro EX máme

$$EX = \sum \nu_i \cdot p(\nu_i) = 0 \cdot 0.0625 + 1 \cdot 0.25 + 2 \cdot 0.375 + 3 \cdot 0.25 + 4 \cdot 0.0625 = 2.$$

Nyní vypočteme ještě EX^2 , protože to budeme potřebovat pro výpočet rozptylu:

$$EX^2 = \sum \nu_i^2 \cdot p(\nu_i) = 0 \cdot 0.0625 + 1 \cdot 0.25 + 4 \cdot 0.375 + 9 \cdot 0.25 + 16 \cdot 0.0625 = 5.$$

A nyní

$$DX = EX^2 - (EX)^2 = 5 - 2^2 = 1.$$

Tedy nejvíce očekávaná hodnota počtu líců je 2, a odchylka od této hodnoty je většinou \sqrt{DX} , což je 1 líc.

Příklad 9.10 Vypočítejte střední hodnotu veličiny X z příkladu 8.12.

Řešení: EX v tomto případě udává očekávaný počet dnů, po kterém dojde k první poruše zařízení. Dosazením do vzorce dostáváme

$$EX = \sum \nu_i \cdot p(\nu_i) = \sum_{k=1}^{\infty} k \cdot p(k) = \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{5}\right)^{k-1} \cdot \frac{4}{5}.$$

A jsme v pěkné bryndě, protože máme sečíst nekonečnou řadu. Zde nepomůže jen se usmívat a pohodlně dosadit vzorec pro součet geometrické řady. Tak jednoduché to nebude. Trochu musíme zapracovat a vylovit v paměti něco o integrování nekonečné řady člen po členu. Ale začněme tou geometrickou řadou. Platí následující vztah, který by si měl odnést do života každý absolvent VUT (nikdy nevíte, kdy se vám bude hodit - ale teď vážně, v životě jsou přece nejdůležitější ty věci, které si myslíme, že vůbec nepotřebujeme, například nějaký kamarád, který nám nepřestane důvěřovat, když uděláme v životě nějakou chybu, nebo vzorec pro součet geometrické řady):

$$1 + x + x^2 + x^3 + x^4 + x^5 + \dots = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad \text{pro } |x| < 1.$$

Uvedená řada se nazývá geometrická, protože každý další člen řady je x -násobkem předchozího členu. Říkáme, že x je kvocient. Tato řada má konečný součet jen pro $|x| < 1$. Bylo by fajn, kdybychom tento krásný vzorec mohli použít i v našem případě. Po nějakých úpravách zjistíme, že to jde. Začněme označením:

$$EX = \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{5}\right)^{k-1} \cdot \frac{4}{5} = \frac{4}{5} \cdot \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{5}\right)^{k-1} = \frac{4}{5} \cdot v\left(\frac{1}{5}\right),$$

kde $v(x) = \sum_{k=1}^{\infty} k \cdot x^{k-1}$. Nyní si můžeme dovolit číslovat sumu $v(x)$ od nuly, protože přičtením nuly se hodnota výrazu $v(x)$ nezmění:

$$v(x) = \sum_{k=1}^{\infty} k \cdot x^{k-1} = \sum_{k=0}^{\infty} k \cdot x^{k-1}.$$

Nyní se zintegrováním této rovnosti zbavíme konstanty k , která vystupuje v každém členu řady:

$$\int v(x)dx = \sum_{k=0}^{\infty} k \cdot \int x^{k-1}dx = \sum_{k=0}^{\infty} k \cdot \frac{x^k}{k} = \sum_{k=0}^{\infty} x^k.$$

Ted' jsme už schopni sumu sečíst podle vzorce pro součet geometrické řady:

$$\int v(x)dx = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

No a $v(x)$ ted' získáme zase derivací poslední rovnosti:

$$v(x) = \frac{d}{dx} \left(\frac{1}{1-x} \right) = \frac{1}{(1-x)^2}.$$

A jsme téměř u cíle. Nesmíme zapomenout, že celý postup funguje jen pro $|x| < 1$. Ale my potřebujeme znát $v(x)$ pro $x = \frac{1}{5}$, což splňuje tuto podmínku konvergence. Tak tedy:

$$EX = \frac{4}{5} \cdot v\left(\frac{1}{5}\right) = \frac{4}{5} \cdot \frac{1}{\left(1 - \frac{1}{5}\right)^2} = \frac{5}{4} = 1.25.$$

Ve spojitém případě se střední hodnota a rozptyl definují vlastně obdobně, s jediným rozdílem - sčítáme nespočetně mnoho nekonečně malých hodnot, takže místo sumy použijeme integrál. Pro spojitou veličinu X tedy

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

$$DX := E(X - EX)^2 = \int_{-\infty}^{\infty} (x - EX)^2 \cdot f(x)dx.$$

Úpravou definičního vztahu pro DX a využitím vzorce $\int_{-\infty}^{\infty} f(x)dx = 1$ bychom dospěli k témuž způsobu výpočtu jako v diskrétním případě:

$$DX = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \left[\int_{-\infty}^{\infty} x \cdot f(x)dx \right]^2 = EX^2 - (EX)^2.$$

Příklad 9.11 Honza získal na zkoušku 80 bodů, zatímco průměr je 75. Je jeho výsledek vynikající, nebo průměrný? Na tuto otázku právě dává odpověď rozptyl. V případě malého empirického rozptylu (např. většina ohodnocení se pohybuje mezi 73 a 77 body) je výsledek 80 bodů vynikající, až pozoruhodný. V případě velkého rozptylu (např. jsou zcela běžné hodnoty z intervalu 55 až 95) je jeho výsledek naprosto průměrný. O kvalitě výsledku nerozhoduje (nevypovídá) pouze jeho porovnání s průměrem, ale také uvážení rozptylu.

Pro určení kvality určitého výsledku je tedy důležitý jak průměr, tak i rozptyl. Zavedeme nyní jakousi transformaci hodnot veličiny X na hodnoty veličiny U , ve kterých je skryta informace jak o průměru, tak i o rozptylu. Pokud x_i jsou hodnoty veličiny X pro $i = 1, 2, \dots, n$, tak

$$u_i := \frac{x_i - \bar{X}}{S}, \quad i = 1, 2, \dots, n$$

jsou hodnoty **normované veličiny** U . Z definice normované veličiny například plyne, že

1. Pokud $u_i > 0$, znamená to, že $x_i > \bar{X}$.
2. Pokud $|u_i| > 1$, znamená to, že x_i se od průměru \bar{X} liší o více než S .

Veličina U tedy představuje převod jakékoli veličiny X na jakousi normovanou stupnici hodnot, ve které je skryta informace o průměru i o rozptylu současně (v některé literatuře, zejména anglické, se normovaná veličina označuje písmenem Z a mluvíme o z -hodnotě; ale české názvosloví celkem jednotně označuje písmenem U).

Příklad 9.12 *Kdybych vám řekl, že moje marťanská kamarádka je 100 cm vysoká, nemohli byste tuto výšku porovnat s výškou ostatních marťanů. Ale kdybych uvedl, že normovaná hodnota její výšky je -1 , věděli byste, že je na marťany dost malá. Z údaje, že normovaná hodnota jejího IQ je 2, byste usoudili, že je to vysoce inteligentní marťanka. Z normované hodnoty hmotnosti 0 se vidí, že její hmotnost je průměrná.*

Příklad 9.13 *Moje mladší sestra se rozhoduje, zda se stane pilotkou letadla nebo kuchařkou. pilotní zkoušky zvládla na 62%, kuchařské na 90%. Na co se víc hodí?*

Bylo by nemoudré, aby se rozhodovala na základě porovnání této procentuální úspěšnosti. Důležitější je porovnání hodnot normovaných:

$$\frac{62 - 50}{6} = 2 \quad 62 \text{ je o } 2 \cdot S \text{ více než průměr } 50.$$

$$\frac{90 - 85}{5} = 1 \quad 90 \text{ je o } S \text{ více než průměr } 85.$$

Vidíme, že se sestra více hodí na pilotku než na kuchařku, respektive má větší šance získat zaměstnání pilotky.

Normovaná hodnota je tedy v jednotkách směrodatné odchylky od průměru.

9.4 Shrnutí pojmů

V této kapitole jsme definovali dvě důležité charakteristiky pro popis dat jak naměřených, tak teoretických. Jsou to střední hodnota a rozptyl. Znovu je při výpočtu těchto charakteristik důležité si uvědomit, zda je veličina X diskrétní, nebo spojitá. Podle typu veličiny pak dosazujeme do vzorce:

$$EX = \begin{cases} \sum_{x_i \in \Omega} x_i \cdot p(x_i) & \text{pro diskrétní veličinu } X; \\ \int_{-\infty}^{\infty} x \cdot f(x) dx & \text{pro spojitou veličinu } X. \end{cases} \quad (9.1)$$

Pojem rozptylu definujeme už s využitím pojmu střední hodnoty. Pro diskrétní i spojitou veličinu lze pro výpočet rozptylu užít vzorec

$$DX = E(X^2) - (EX)^2. \quad (9.2)$$

Při odvozování tohoto vzorce v diskrétním případě jsme užili jistých pravidel pro počítání se střední hodnotou: pokud a, b jsou reálná čísla a X, Y náhodné veličiny, platí vztah

$$E(aX - bY) = a \cdot EX - b \cdot EY.$$

Čili konstantu lze vytknout před střední hodnotu. Dále platí $E(aX + b) = aEX + b$ (tj. střední hodnota konstanty je rovna konstantě samotné). Pro výpočet rozptylu složených výrazů platí jiná pravidla, a sice

$$D(aX - bY) = a^2 \cdot DX + b^2 \cdot DY.$$

Čili pokud vytýkáme konstantu před rozptyl, musíme ji umocnit na druhou. To například znamená, že rozptyl rozdílu veličin je roven součtu, nikoliv rozdílu rozptylů: $D(X - Y) = DX + DY$. Dále platí $D(aX + b) = a^2 \cdot DX$, tj. rozptyl konstanty b je roven nule.

10 Binomické rozdělení pravděpodobnosti

V této a následujících dvou kapitolách projdeme podrobněji některá rozdělení, jež mají největší využití v technické praxi. Nejprve se seznámíme s binomickým rozdělením. Toto rozdělení pravděpodobnosti je základní a je východím pro odvození všech ostatních. A proto úvodní kurs pravděpodobnostních modelů musí obsahovat kapitolu o něm. Seznámíme se se základními vlastnostmi tohoto rozdělení, a pak uvidíme jeho využití ve statistice na znaménkovém testu. Tato kapitola tedy obsahuje také principy, které jsou společné všem statistickým testům.

10.1 Vlastnosti binomického rozdělení

Začneme hned definicí binomického rozdělení, kterou pak osvětlíme na několika příkladech. Uvažujme experiment takové povahy, že mohou nastat jen dva různé výsledky, které se navzájem vylučují (nemůže k nim dojít současně): „úspěch“ a „neúspěch“ („úspěch“ nemusí znamenat nic světoborného; označuje se tímto termínem proto, že se jedná o ten ze dvou možných výsledků, na který se ve svých úvahách chceme zaměřit).

Pravděpodobnost úspěchu je p , pravděpodobnost neúspěchu $1 - p$. Náhodná veličina X , která udává počet výskytů úspěchu při N nezávislých opakováních experimentu, má tzv. binomické rozdělení pravděpodobnosti (s parametry N, p) a nabývá hodnot z množiny $\{0, 1, 2, \dots, N\}$ s pravděpodobnostmi

$$P(X = r) = \binom{N}{r} \cdot p^r \cdot (1 - p)^{N-r}.$$

Mluví se zde o nezávislých opakováních experimentu. Slovo „nezávislých“ znamená, že výskyt úspěchu při prvním opakování experimentu nemá vliv na to, zda při druhém a dalších opakováních nastane úspěch nebo ne. Skutečnost, že veličina X má binomické rozdělení s parametry N, p , budeme označovat

$$X \sim Bi(N, p).$$

Podívejme se nyní na konkrétní příklady.

Příklad 10.1 *Házeme čtyřikrát kostkou. Veličina X udává, kolikrát přitom padne šestka. Jaké je rozdělení pravděpodobnosti veličiny X ?*

Řešení: Pravděpodobnost, že při jednom hodu padne šestka, je rovna $p = \frac{1}{6}$. Hody jsou navzájem nezávislé, tj. pokud v prvním hodu padla šestka, nemá to vliv na to, zda ve druhém hodu padne nebo ne. Tedy veličina X , která měří počet šestek při čtyřech hodech, má binomické rozdělení pravděpodobnosti s parametry $N = 4$, $p = \frac{1}{6}$. Podívejme se konkrétně na pravděpodobnosti, s jakými veličina X nabývá konkrétní hodnoty. Bude odtud zřejmé i odvození vzorce pro jejich výpočet.

$$P(X = 0) = P(\text{ne } 6) \cdot P(\text{ne } 6) \cdot P(\text{ne } 6) \cdot P(\text{ne } 6) = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = 0.482;$$

$$\begin{aligned}
P(X = 1) &= P(\text{jednou padne 6, jinak něco jiného než 6}) = \\
&= P(6 \text{ padne jako první, jinak ne}) + P(6 \text{ padne jako druhá, jinak ne}) + \\
&\quad + P(6 \text{ padne jako třetí, jinak ne}) + P(6 \text{ padne jako čtvrtá, jinak ne}) = \\
&= \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} + \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} = \\
&= (\text{všechna možná pořadí výskytu jednoho úspěchu}) \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \\
&= \binom{4}{1} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = 0.386; \\
P(X = 2) &= P(\text{dvakrát padne šestka, jinak ne}) = \\
&= (\text{všechny možnosti výběru 2 pořadí ze 4}) \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \\
&= \binom{4}{2} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = 0.116; \\
P(X = 3) &= \binom{4}{3} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} = 0.015; \\
P(X = 4) &= \binom{4}{4} \cdot \left(\frac{1}{6}\right)^4 = 0.001.
\end{aligned}$$

Všimněte si, že součet těchto pěti pravděpodobností je roven jedné. Při výpočtu jsme zaokrouhlovali na tři desetinná místa.

Příklad 10.2 Senátor Swenson před volbami tvrdí, že pro něj bude hlasovat 70% voličů. Agentura STEN chce provést průzkum u 20 lidí. Náhodná veličina X udává počet Swensonových voličů z dvaceti dotázaných. Určete

- teoretické rozdělení veličiny X (před provedením průzkumu);
- pravděpodobnost, že Swensona bude volit přesně 14 lidí z 20 dotázaných;
- pravděpodobnost, že Swensona bude volit maximálně 14 lidí z 20 dotázaných.

Řešení:

ad a) Dané teoretické rozdělení je binomické s parametry $N = 20$ a $p = 0.7$. Veličina X nabývá hodnot z množiny $\{0, 1, 2, \dots, 20\}$ s pravděpodobnostmi

$$P(X = r) = \binom{20}{r} \cdot 0.7^r \cdot 0.3^{20-r}.$$

ad b) Dosazením do vzorce a) máme

$$P(X = 14) = 0.192,$$

pokud zaokrouhlujeme na tři desetinná místa.

ad c) Zde využijeme finty použité poprvé v příkladu 8.12: abychom ušetřili několik sčítanců, vypočteme pravděpodobnost opačného jevu a odečteme ji od jedničky:

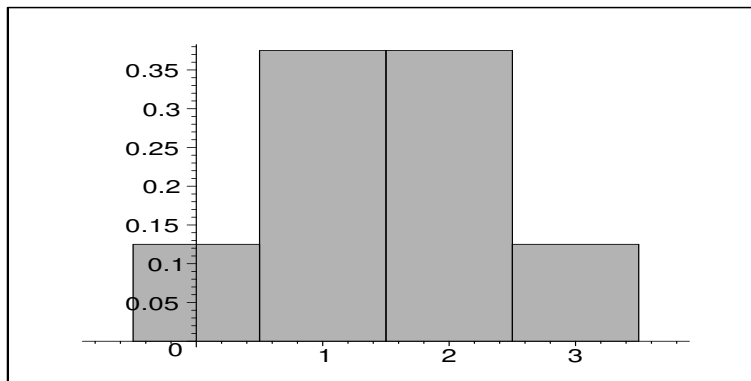
$$\begin{aligned} P(X \leq 14) &= 1 - P(X > 14) = 1 - (p(15) + p(16) + p(17) + p(18) + p(19) + p(20)) = \\ &= 1 - (0.179 + 0.13 + 0.072 + 0.028 + 0.007 + 0.001) = 0.583. \end{aligned}$$

Pokud by agentura STEN v předchozím příkladu zjistila, že „pro“ bylo jen 8 lidí z 20, pak některý z teoretických předpokladů nebyl v pořádku:

- vzorek dotázaných lidí nebyl náhodný (byl z antiswensonovské oblasti státu);
- odpovědi nebyly nezávislé (odpovídající mezi sebou navzájem diskutovali o Swensonovi);
- STEN pracovala dobře, ale Swenson byl příliš optimistický se svým odhadem (to je nejpravděpodobnější problém).

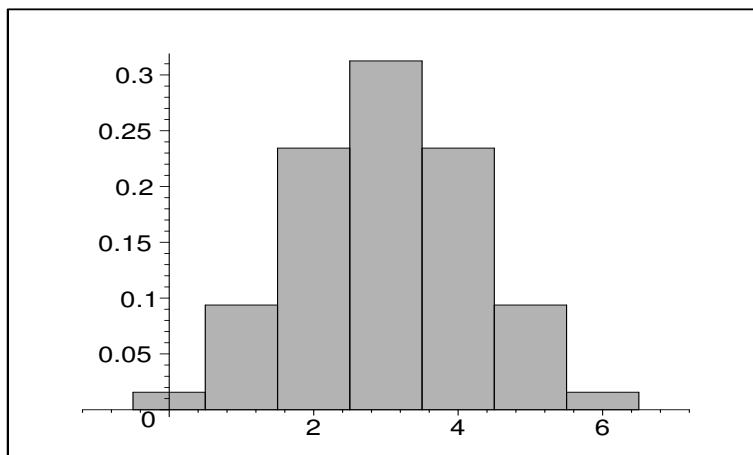
Ukažme si ještě graficky tvar binomického rozdělení, například pomocí pravděpodobnostního histogramu.

a) Pokud $p = 0.5$, rozdělení je vždy symetrické (viz obr. 10.1, 10.2, 10.3).

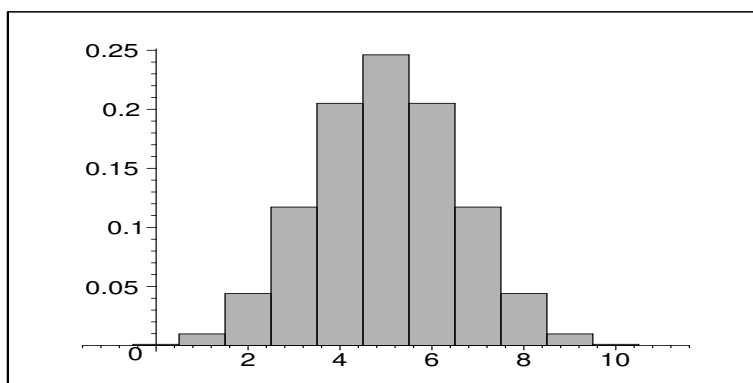


Obrázek 10.1: Histogram pravděpodobností binomického rozdělení pro $N = 3$, $p = 0.5$.

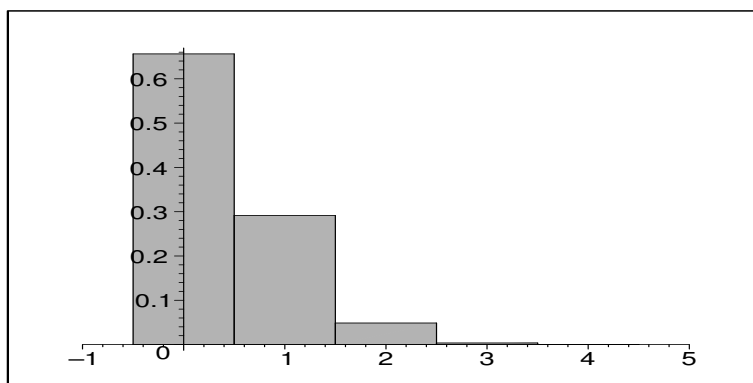
b) Pro $p \neq 0.5$ a malé N je rozdělení asymetrické, ale pro rostoucí N se stává více a více symetrickým (viz obr. 10.4, 10.5, 10.6, 10.7 - na obrázku 10.7 jsou pravděpodobnosti nenulové pro hodnoty 0 až 40, ale při zaokrouhlování na tři desetinná místa jsou hodnoty v bodech 12 a více už rovny nule; Je vidět, že histogram je už poměrně symetrický na rozdíl od obrázku 10.6.).



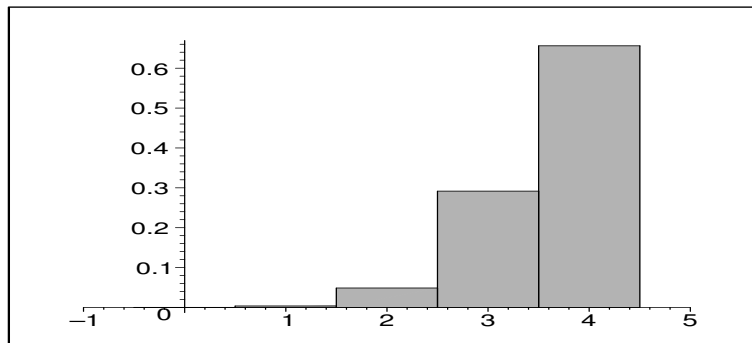
Obrázek 10.2: Histogram pravděpodobností binomického rozdělení pro $N = 6$, $p = 0.5$.



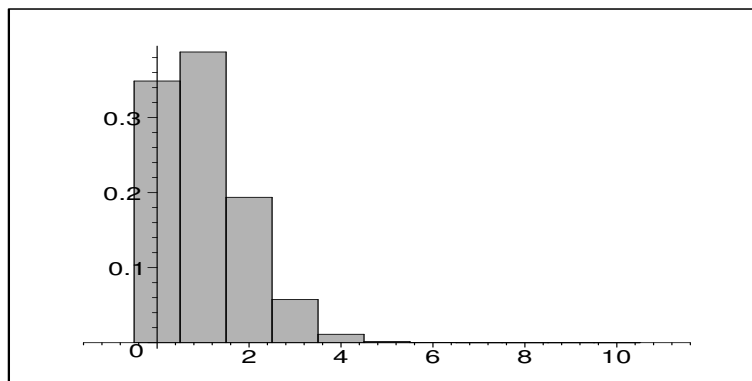
Obrázek 10.3: Histogram pravděpodobností binomického rozdělení pro $N = 10$, $p = 0.5$.



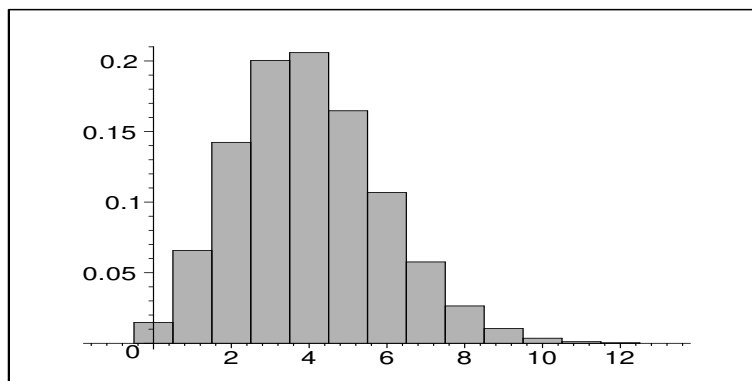
Obrázek 10.4: Histogram pravděpodobností binomického rozdělení pro $N = 4$, $p = 0.1$.



Obrázek 10.5: Histogram pravděpodobností binomického rozdělení pro $N = 4$, $p = 0.9$.



Obrázek 10.6: Histogram pravděpodobností binomického rozdělení pro $N = 10$, $p = 0.1$.



Obrázek 10.7: Histogram pravděpodobností binomického rozdělení pro $N = 40$, $p = 0.1$.

Vypočteme nyní střední hodnotu a rozptyl veličiny X s binomickým rozdělením $Bi(N, p)$.

$$\begin{aligned}
EX &= \sum_{i=0}^N i \cdot p(i) = \sum_{i=1}^N i \cdot \binom{N}{i} \cdot p^i \cdot (1-p)^{N-i} = \\
&= \sum_{i=1}^N i \cdot \frac{N!}{(N-i)! \cdot i!} \cdot p^i \cdot (1-p)^{N-i} = \\
&= N \cdot p \cdot \sum_{i=1}^N \frac{(N-1)!}{(N-i)! \cdot (i-1)!}
\end{aligned}$$

Nejprve jsme dosadili do vzorce pro střední hodnotu diskrétní veličiny, vyjádřili kombinační číslo podle definice s využitím faktoriálů, zkrátili i a vyhodili N a jedno p před sumu. Nyní ještě označíme

$$M := N - 1, \quad j := i - 1.$$

Pak totiž

$$EX = N \cdot p \cdot \sum_{j=0}^M \frac{M!}{(M-j)! \cdot j!} \cdot p^j \cdot (1-p)^{M-j} = Np \sum_{j=0}^M \binom{M}{j} \cdot p^j \cdot (1-p)^{M-j}$$

a na pravé straně posledního vztahu dostáváme sumu, která vyjadřuje součet hodnot pravděpodobnostní funkce binomického rozdělení s parametry M a p , tj. podle jedné ze základních vlastností pravděpodobnosti je rovna jedné. A tak nám zůstává pouze

$$EX = N \cdot p.$$

Při odvození hodnoty rozptylu použijeme stejnou taktiku: pokusíme se před sumu něco vytknout, aby zbylé sumování bylo rovno jedné:

$$\begin{aligned}
DX &= \left(\sum_{i=0}^N i^2 \cdot p(i) \right) - (EX)^2 = \left(\sum_{i=1}^N i^2 \cdot \frac{N!}{(N-i)! \cdot i!} \cdot p^i \cdot (1-p)^{N-i} \right) - N^2 \cdot p^2 = \\
&= N \cdot p \cdot \left(\sum_{i=1}^N i \cdot \frac{(N-1)!}{(N-i)! \cdot (i-1)!} \cdot p^{i-1} \cdot (1-p)^{N-i} \right) - N^2 \cdot p^2;
\end{aligned}$$

Označíme-li nyní $M := N - 1$, $j := i - 1$, dostaneme

$$DX = Np \left(\sum_{j=0}^M (j+1) \cdot \frac{M!}{(M-j)! \cdot j!} \right) - N^2 p^2.$$

Závorku $(j+1)$ v posledním výrazu rozdělíme do součtu dvou sum - v té první sumě bude j , ve druhé bude 1:

$$DX = Np \sum_{j=0}^M j \cdot \binom{M}{j} p^j (1-p)^{M-j} + Np \sum_{j=0}^M \binom{M}{j} p^j (1-p)^{M-j} - N^2 p^2.$$

Nyní už je snadné sečíst obě poslední sumy v právě dosaženém výrazu, protože ta první je podle definice rovna střední hodnotě binomického rozdělení s parametry M a p (což je Mp), ta druhá je rovna součtu pravděpodobností binomického rozdělení s parametry M a p (čili jedné). Celkem dostáváme

$$\begin{aligned} DX &= Np \cdot Mp + Np \cdot 1 - N^2 p^2 = \\ &= Np \cdot (N-1)p + Np - N^2 p^2 = Np - Np^2 = \\ &= Np(1-p). \end{aligned}$$

Binomické rozdělení je příkladem toho, že výpočet rozptylu dá vždy víc práce než výpočet střední hodnoty (respektive střední hodnota je jedním z členů při výpočtu rozptylu).

Někdy se hodnoty veličiny s binomickým rozdělením uvádějí nikoliv v četnostech i (např. 12 úspěchů ze 20 pokusů), ale v podílech úspěšnosti $\frac{i}{N}$ (např. $\frac{12}{20}$). Toto **binomické rozdělení podílů úspěšnosti** má stejné parametry N , p , ale díky jiným hodnotám, kterých nabývá, je zde jiná střední hodnota a rozptyl:

$$\begin{aligned} EX &= \sum_0^N \frac{i}{N} \cdot p(i) = \frac{1}{N} \cdot (\text{střední hodnota veličiny četností}) = \frac{1}{N} \cdot Np = p. \\ DX &= \left(\sum_0^N \frac{i^2}{N^2} \cdot p(i) \right) - (EX)^2 = \frac{1}{N^2} \left(\sum_0^N i^2 p(i) \right) - p^2 = \\ &= \frac{1}{N^2} \cdot (\text{první člen při výpočtu rozptylu veličiny četností}) - p^2 = \\ &= \frac{1}{N^2} (Np \cdot (N-1)p + Np) - p^2 = p^2 - \frac{p^2}{N} + \frac{p}{N} - p^2 = \frac{p(1-p)}{N}. \end{aligned}$$

Příklad 10.3 Na obrázku 10.8 je histogram pravděpodobnostní funkce binomické veličiny pro $p = 0.5$, která nabývá hodnot $0, \frac{1}{16}, \frac{2}{16}, \dots, \frac{16}{16}$. Od binomického rozdělení s hodnotami $0, 1, 2, \dots, 16$ se liší jen jiným značením hodnot na vodorovné ose; jinak jsou příslušné histogramy stejné.

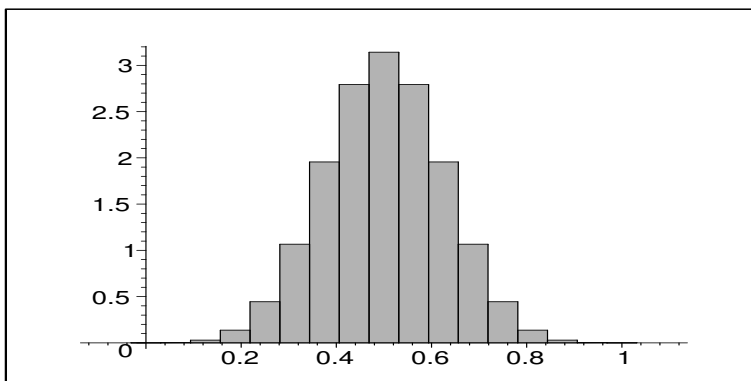
Příklad 10.4 Hodíme 400-krát mincí. Náhodná veličina udávající počet líců v těchto pokusech má binomické rozdělení s parametry $N = 400$, $p = 0.5$. Příslušné teoretické rozdělení má tyto charakteristiky:

a) Hodnoty X jsou v četnostech:

$$EX = Np = 200; \quad DX = Np(1-p) = 100; \quad \sqrt{DX} = 10.$$

b) Hodnoty X jsou v podílech (= relativních četnostech):

$$EX = p = 0.5; \quad DX = \frac{p(1-p)}{N} = 0.000625; \quad \sqrt{DX} = 0.025.$$



Obrázek 10.8: Histogram pravděpodobností binomického rozdělení pro $N = 16$, $p = 0.5$ s hodnotami relativních četností.

Protože charakter histogramu pravděpodobností je stejný (rozdíl je pouze v označení hodnot na ose x), sobě odpovídající normované hodnoty se rovnají: Například pokud ze 400 hodů padne 210 líců, příslušná normovaná hodnota je

$$\frac{210 - 200}{10} = 1;$$

210 lícům odpovídá relativní četnost $\frac{210}{400} = 0.525$, příslušná normovaná hodnota je

$$\frac{0.525 - 0.5}{0.025} = 1.$$

Jediné, na co si musíme dávat pozor, je tedy jiná střední hodnota a rozptyl v každém z přístupů a), b).

10.2 Základní principy statistického testu

Jedno z využití binomického rozdělení je ve statistickém znaménkovém testu. Dříve než k němu přistoupíme, na příkladu vysvětlíme jednotlivé kroky statistického testu obecně.

Příklad 10.5 Soudní proces jako příklad rozhodovacího procesu. Uvažujme jednoduchý soudní proces, ve kterém existuje pouze jediný možný trest a soud rozhodne, zda se tomuto trestu obžalovaný podrobí nebo ne. A navíc proti rozhodnutí soudu neexistuje žádné odvolání. Jedná se o jakýsi rozhodovací proces, u kterého mohou nastat čtyři možné výsledky:

1. Obžalovaný je vinen a soud jej odsoudí.
2. Obžalovaný je nevinný a soud jej osvobodí.
3. Obžalovaný je nevinný a soud jej odsoudí. Jedná se o chybné rozhodnutí - tuto chybu budeme označovat jako chybu prvního druhu.

4. *Obžalovaný je vinen a soud jej osvobodí. Toto rozhodnutí je rovněž chybné - budeme tuto chybu označovat chybou druhého druhu.*

V každém soudním procesu se musí hledat jistá rovnováha mezi tvrdostí a mírností. Jedním extrémem je liberální soudce, který k usvědčení obžalovaného vyžaduje velké množství důkazů. Takový soudce jen zřídka odsoudí nevinného (zřídka se dopustí chyby prvního druhu), ale dosti často osvobodí viníka (chyba druhého druhu). Druhým extrémem je konzervativní soudce, kterému k usvědčení stačí jen několik důkazů. Takový soudce posílá do vězení i jen při stínu podezření, čili častěji odsoudí nevinného (chyba prvního druhu), ale zřídka osvobodí darebáka (= zřídka se dopustí chyby druhého druhu). Slova „konzervativní“ a „liberální“ jsou termíny z politiky. V dnešní době už nikdo neví, co znamenají. Tato jejich „statistická“ definice navrhuje jejich význam, ale také upozorňuje na nebezpečí každého z těchto postojů.

Je otázkou, která z chyb je závažnější - zda chyba prvního druhu, nebo chyba druhého druhu. Všeobecně se má za to, že závažnější je uvěznit nevinného, než odsoudit darebáka. A proto se chybě odsouzení nevinného přisuzuje druh číslo 1 a věnuje se jí větší pozornost. Ale někde musí být stanovena jistá hranice, po jejímž překročení už soud přistoupí k rozhodnutí „vinen“ a bez skrupulí člověka potrestá.

Všimněme si jedné věci, která platí jako obecný princip. Pokud se soudce snaží být benevolentní a odsoudí člověka až po nahromadění velkého množství důkazů (snižuje tím možnost výskytu chyby prvního druhu), současně narůstá nebezpečí, že i když je obžalovaný vinen, potřebné množství důkazů se nenajde a soud jej osvobodí (roste možnost výskytu chyby druhého druhu). Není to nic světoborného, ale už jsme dlouho neměli žádný rámeček, a proto jej aspoň uvnitř příkladu můžeme použít:

Snižováním možnosti výskytu chyby prvního druhu roste možnost výskytu chyby druhého druhu - a naopak: pokud zvyšujeme možnost výskytu chyby prvního druhu, snižuje se možnost výskytu chyby druhého druhu.

Z uvedeného rámečku je vidět, že žádnou z chyb není možné naprosto vyrušit: pokud totiž snižujeme možnost výskytu chyby prvního druhu až téměř na nulu, roste tím možnost výskytu chyby druhého druhu do obhludných rozměrů a rozhodnutí učiněná tímto stylem jsou nerozumná, až nemoudrá. Strategii v rozhodovacích procesech tohoto typu je tedy zvolit pravděpodobnost výskytu chyby prvního druhu malou, ale ne příliš malou.

Shrňme předchozí úvahy do pěti kroků, které popisují celý soudní proces:

1. *Stojí proti sobě dvě možná rozhodnutí soudu:*

$$\begin{aligned} H_0 & \dots \text{obžalovaný je nevinný} \\ H_1 & \dots \text{obžalovaný je vinen} \end{aligned}$$

Soud musí rozhodnout právě jednu z těchto variant a toto rozhodnutí je nezvratné, neexistuje proti němu odvolání.

2. *Vystoupí žalobce, který předloží nashromážděné důkazy pro platnost H_1 .*
3. *Vystoupí obhájce a vysvětlí všechny souvislosti za předpokladu, že platí H_0 . Snaží se vidět a vysvětlit všechny argumenty obžaloby ve světle toho, že obžalovaný je nevinen.*
4. *Porota soudu se odebere k rokování. Bere v úvahu jak množství důkazů a jejich závažnost, tak i argumenty obhajoby a možnost, že tyto důkazy neznamenaají nutně vinu obžalovaného, ale v jeho neprospěch hrají jen náhodou.*
5. *Porota se vrací a vyslovuje svůj verdikt: pokud byla překročena míra závažnosti důkazů pro platnost H_1 , obžalovaný je vinen. pokud ne, obžalovaný je osvobozen. Toto rozhodnutí soudu je nezvratné.*

Právě uvedených pět kroků v příkladu 10.5 se vyskytuje v mnoha rozhodovacích procesech, které nazýváme **statistické testy**. Tyto principy platí obecně, vyslovme je tedy obecně, už oproštěni od příkladu soudce a obžalovaného (ovšem analogie se soudním procesem zde existuje velice přímá):

- (K1) Statistický test obvykle rozhoduje o tom, zda platí hypotéza H_0 (tzv. **nulová hypotéza**) nebo H_1 (tzv. **alternativní hypotéza**). Tyto dvě hypotézy přitom stojí ve vzájemném rozporu. Ve většině testů H_0 tvrdí, že jistá veličina *nezávisí* na hodnotách určité další veličiny, kdežto H_1 tvrdí, že naopak *závisí* (pro ty, kdo by si chtěli udržet souvislost mezi statistickým testem a soudním procesem, což doporučuji, pomůcka k zapamatování: H_0 testu říká **nezávisí**, a H_0 soudního procesu **nevinen**).
- (K2) Stanovíme **kritérium** (zpravidla určitou funkci), které ukazuje na míru platnosti alternativní hypotézy H_1 (určuje „závažnost důkazů“ pro H_1). Pak provedeme experiment, ve kterém změříme data potřebná pro dosažení hodnot do našeho kritéria.
- (K3) Kritériem bývá jistá funkce, která při různých měřeních nabývá různých hodnot, je to tedy náhodná veličina. Určíme **teoretické rozdělení kritéria** za předpokladu, že platí hypotéza H_0 . Jinými slovy, popíšeme vlastnosti kritérijní veličiny ve světle toho, že platí H_0 .
- (K4) Na základě teoretického rozdělení kritérijní veličiny stanovíme určitý interval hodnot, kam když padne empirická hodnota kritéria, tak nezviklá naše přesvědčení o platnosti H_0 , ale eventuelní dopad hodnoty kritéria mimo tento interval nás povede k názoru, že byla překročena jistá **kritická míra**, takže usoudíme, že H_0 neplatí. Kritickou míru zpravidla určujeme tak, aby pravděpodobnost výskytu chyby prvního druhu (tj. že rozhodneme, že H_0 neplatí, když ve skutečnosti H_0 platí) byla dostatečně malá, např. rovna 0.05 (to se chyby prvního druhu dopustíme nejvýše v pěti procentech případů), ale ne příliš malá, aby nerostla možnost výskytu chyby druhého druhu (tj. že rozhodneme, že H_0 platí, když ve skutečnosti H_0 neplatí) do nerozumných rozměrů.

(K5) Porovnáme empirickou hodnotu kritéria s kritickou mírou. Pokud je kritická míra překročena (hodnota kritéria leží mimo interval nalezený v bodě 4), zamítáme hypotézu H_0 ve prospěch alternativní hypotézy H_1 . Pokud není kritická míra překročena, hypotézu H_0 nezamítáme.

Pro ty, co nedávali pozor nebo jsou unavení, ještě jednou definice chyby prvního a druhého druhu - s využitím tabulky 10.1:

Tabulka 10.1: Čtyři možné výsledky statistického testu.

	skutečnost: H_0 platí	skutečnost: H_1 platí
rozhodnutí: H_0 nezamítáme	O.K.	chyba 2.druhu
rozhodnutí: H_0 zamítáme	chyba 1.druhu	O.K.

Další standardní označení se používá pro pravděpodobnost výskytu chyby 1.druhu (značí se α) a pravděpodobnost výskytu chyby 2.druhu (značíme β).

10.3 Znaménkový test

Teď už známe potřebnou terminologii, a proto se pustíme do jednoduchého statistického testu, kterým je **znaménkový test** (anglicky - the sign test). Vše bude vysvětleno v následujícím příkladu.

Příklad 10.6 *Chceme ověřit hypotézu, že zvýšení motivace má vliv na lidskou paměť. Abychom získali určitá data, nebudeme zkoumat všechny lidi na zeměkouli, ale náhodně vybereme 10 lidí, provedeme s nimi test a jeho výsledek vztáhneme na celé lidstvo (tento test vzorku a vztahení jeho výsledku na celek je pro statistiku charakteristický). U vybraných lidí provedeme následující experiment:*

1. Každému z vybraných lidí se pomalu přečte 20 slov, a po pěti minutách má zopakovat všechna, která se mu vybaví. Za každé správně zopakované slovo dostává 10 Kč.
2. Přečte se jiných 20 slov a dotazovaný člověk si jich po pěti minutách má opět co nejvíce vybavit - nyní ale za každé správně zapamatované slovo dostává 200 Kč.
3. Znaménkovým testem zjistíme, zda se při zvýšení finanční motivace významně zvýšila vybavovací schopnost daného vzorku 10 lidí.

Řešení: Získala se data v tabulce 10.2.

Budeme nyní přesně procházet pět kroků testu představených v předchozím oddílu:

Tabulka 10.2: K příkladu 10.6: Data získaná testovým měřením.

člověk	počet zapamatovaných slov za 10 Kč	počet zapamatovaných slov za 200 Kč	zlepšení?
1	7	8	+
2	5	7	+
3	6	5	-
4	5	9	+
5	6	7	+
6	5	9	+
7	3	5	+
8	4	5	+
9	8	11	+
10	2	4	+

(K1) Stanovme hypotézy H_0 a H_1 :

H_0 : Vybavovací schopnost člověka **nezávisí** na velikosti motivace v tom smyslu, že zvýšení motivace nevede ke zvýšení schopnosti zapamatování

H_1 : Vybavovací schopnost člověka **závisí** na velikosti motivace v tom smyslu, že se zvýšením motivace roste i zapamatovací schopnost

(K2) Kritériem našeho testového rozhodování bude počet lidí, u kterých nastalo zlepšení při zvýšení finanční motivace, tj. počet kladných znamének v posledním sloupci tabulky 10.2. Označme tento počet kladných znaménkových změn jako T . Veličina T tedy udává počet kladných znamének v deseti nezávislých měřeních.

(K3) Určeme rozložení pravděpodobnosti náhodné veličiny T za předpokladu, že platí H_0 - tj. vysvětleme chování veličiny ve světle toho, že zapamatování nezávisí na motivaci. V takovém případě výskyt kladného znaménka je naprosto náhodný a stejně dobře se místo kladného znaménka může u konkrétního člověka objevit záporné znaménko. Čili pokud je výskyt kladného znaménka u konkrétního člověka náhodný, může k němu dojít s takovou pravděpodobností, s jakou při hodu korunou padne líc - čili s pravděpodobností $\frac{1}{2}$. Tedy veličina T při platnosti hypotézy H_0 udává počet kladných znamének z deseti situací, přičemž v každé situaci k tomu dojde s pravděpodobností $\frac{1}{2}$ - ale to nám něco připomíná. To přece znamená, že veličina T má za předpokladu platnosti H_0 binomické rozdělení s parametry $N = 10$, $p = 0.5$.

(K4) Musíme určit kritickou mez T_k počtu kladných znamének, při jejichž dosažení už přestaneme věřit, že platí H_0 , a usoudíme, že počet kladných znamének je statisticky významný a ukazuje na platnost hypotézy H_1 . Z toho důvodu musíme blíže prozkoumat pravděpodobnostní funkci $p(x)$ naší diskrétní náhodné veličiny T - viz tabulka 10.3

Tabulka 10.3: K příkladu 10.6: hodnoty funkce $p(r)$ a kumulativní pravděpodobnostní funkce $P(T \geq r)$ zaokrouhleny na tři des. místa.

r	$p(r) = P(T = r)$	$P(T \geq r)$
10	0.001	0.001
9	0.010	0.011
8	0.044	0.055
7	0.117	0.172
6	0.205	0.377
5	0.246	0.623
4	0.205	0.828
3	0.117	0.945
2	0.044	0.989
1	0.010	0.999
0	0.001	1.000

Pro určení kritické hodnoty je rozhodující právě kumulativní pravděpodobnostní funkce v posledním sloupci tabulky 10.3. Nyní pravděpodobnost, že k výskytu deseti kladných znamének dojde naprostou náhodou, nikoliv na základě závislosti pamatování na motivaci, je rovna

$$P(T \geq 10) = 0.001;$$

Čili kdybychom hypotézu H_0 zamítli v případě výskytu 10 kladných znamének, měli bychom šanci se dopustit chyby prvního druhu s pravděpodobností 0.001. To je dost nízká pravděpodobnost, což znamená, že riziko výskytu chyby druhého druhu (= H_0 neplatí, ale my ji nezamítneme) je naopak veliké. Proto jdeme v tabulce kumulativních pravděpodobností dále: pravděpodobnost, že k výskytu devíti a více kladných znamének dojde naprostou náhodou, nikoliv na základě závislosti pamatování na motivaci, je rovna

$$P(T \geq 9) = 0.011;$$

Čili kdybychom H_0 zamítli pro kritickou hodnotu $T_k = 9$, dopustili bychom se chyby prvního druhu s pravděpodobností 0.011. A tak dále, zkrátka snažíme se najít kritickou hodnotu pro takové riziko α výskytu chyby prvního druhu, které je dost malé (např $\alpha \leq 0.05$), ale ne zas příliš malé. Proto se zarazíme u takové kumulativní četnosti, která je menší než 0.05, ale přitom je to největší možná kumulativní pravděpodobnost s touto vlastností. Protože

$$P(T \geq 8) = 0.055 > 0.05,$$

vrátíme se zpět k nejbližší nižší hodnotě, tj. $T_k = 9$ a pravděpodobnost výskytu chyby prvního druhu je rovna $\alpha = 0.011$ (tj. pokud při $T \geq T_k = 9$ zamítneme H_0 , máme šanci dopustit se chyby prvního druhu na 1.1%).

(K5) Naměřený počet kladných znamének $T = 9$ je roven kritické hodnotě $T_k = 9$, a tedy zamítáme H_0 o nezávislosti ve prospěch alternativní hypotézy H_1 . Říkáme, že závislost pamatování na motivaci je statisticky významná. V případě, kdy by počet kladných znamének byl menší než $T_k = 9$, bychom H_0 nezamítli.

Je otázkou, jaké znaménko přiřadit v předchozím příkladu člověku, který má stejnou hodnotu zapamatovaných slov v obou motivačních situacích (to v našich datech nenastalo, ale je to možné). Existují dvě alternativy řešení: buď můžeme stejnou hodnotu u obou finančních podmínek označit znaménkem „minus“ (koneckonců o zlepšení se nejedná, čili daného člověka můžeme započítat jako případ potvrzující náhodnost, tj. nezávislost obou veličin), nebo měření u tohoto člověka z testu úplně vypustit (to je asi nejférovější řešení - stejná hodnota u obou podmínek nehovoří pro, ani proti kladné změně). Tato jemnost je příkladem úvah, které musíme někdy provést před konkrétním výpočtovým provedením testu.

Vzhledem k tomu, jak byly formulovány hypotézy H_0 a H_1 , se jednalo o tzv. jednostranný test, kdy jsme si všimli pouze významně vyššího počtu kladných znamének. Oboustranný test v případě daného experimentu by bral v potaz i možnost, že zvýšení finanční motivace vede u člověka k degradaci paměti, což se projeví na extrémně malém počtu kladných změn. V případě oboustranného testu jsou kritické hodnoty dvě (levá a pravá mez jistého intervalu). ovšem v naší situaci je rozumné předpokládat, že zvýšení finanční motivace člověka nedeprimuje, ale naopak povzbudí k lepšímu pamatování, tj. bylo vhodné použít jednostranný test. K oboustrannému testu se vrátíme v kapitole 12.

10.4 Shrnutí pojmů

V této kapitole jsme se seznámili s prvním typem rozdělení pravděpodobnosti, které má široké využití v praxi. Veličina X s rozdělením $Bi(N, p)$ nabývá hodnot z množiny $\Omega = \{0, 1, 2, \dots, N\}$ s pravděpodobnostmi

$$p(k) = P(X = k) = \binom{N}{k} \cdot p^k \cdot (1 - p)^{N-k}. \quad (10.1)$$

Teoreticky je pravděpodobnostní funkce $p(x)$ tohoto diskrétního rozdělení pravděpodobnosti definována pro každé reálné x , ale hodnot jiných než z množiny Ω nabývá veličina X s nulovou pravděpodobností (tj. $p(x) = 0$, pokud $x \notin \Omega$).

Seznámili jsme se s pěti kroky statistického testu, které jsou stavebními kameny i ostatních statistických testů, nejen testu znaménkového. V kapitolách 12 a 13 budeme dále studovat test využívající normálního rozdělení pravděpodobnosti (což je nejčastější případ spojitého rozdělení v úlohách praxe).

V dalším budeme označovat řeckým písmenem α pravděpodobnost výskytu chyby prvního druhu v daném statistickém testu, β pravděpodobnost výskytu chyby druhého druhu. Číslo α slouží k určení kritických hodnot testu, a má proto svůj název - říká se mu **hladina významnosti testu**. Kromě hladiny významnosti se někdy definuje další pojem charakterizující statistický test daného typu, a sice **síla testu**:

$$\text{Síla daného testu} = 1 - \beta,$$

což je pravděpodobnost, že správně zamítneme H_0 v situaci, kdy skutečně hypotéza H_0 neplatí. Jedná se o pozitivní pojem - čím je síla testu větší, tím je tento test vhodnější k nalezení závislosti mezi danými proměnnými. Ovšem sílu testu většinou neznáme, protože pravděpodobnost β často nedokážeme určit. O síle testu by se dalo mluvit déle, ale v tomto základním kursu na to není čas, ani prostor. Takže u síly testu na shledanou až v inženýrském studiu.

Se silou testu souvisí i následující věc: pokud naměřená hodnota kritéria nepřekročí teoretické kritické hodnoty, říkáme, že „hypotézu H_0 nezamítáme”, nikoliv „hypotézu H_0 přijímáme”. Pokud totiž náš použitý statistický test měl malou sílu, mohlo se stát, že ačkoliv závislost mezi veličinami nenalezl, ona ve skutečnosti existuje a H_0 neplatí (co si budeme nalhávat, do jisté míry závisí všechno na všem). Z tohoto důvodu se používá tato „opatrná” terminologie.

Další obrat jsme v příkladu už také použili: pokud zamítáme H_0 , někdy se říká, že výsledek testu je statisticky významný (resp. závislost mezi studovanými veličinami je statisticky významná, nebo vliv jedné veličiny na druhou je významný).

Celkem neformálně budeme užívat slova **vzorek**, ovšem v jiném významu než v elektrotechnických předmětech. Ve statistice je vzorkem označována vybraná skupina lidí (nebo jiných jednotek) z celé populace, a potažmo to bude znamenat zejména soubor měření provedený u této vybrané skupiny. Tj. *délka vzorku* bude označovat počet měření provedený v dané situaci.

11 Poissonovo a exponenciální rozdělení pravděpodobnosti

V této kapitole se seznámíme s dalšími dvěma typy rozdělení pravděpodobnosti, které jsou využívány v úlohách technické praxe. I když Poissonovo rozdělení je diskrétní a exponenciální rozdělení spojité, existuje mezi nimi blízký vztah - každé z nich sice používáme k popisu jiné veličiny, ale hodnoty těchto veličin měříme v jedné a téže situaci. Podíváme se také na teorii front, kde se využívá několik pravděpodobnostních modelů, zejména právě Poissonovo a exponenciální rozdělení. Zejména v této kapitole je vidět bohatost užití pravděpodobnostních modelů při popisu reálných situací.

11.1 Odvození

Jeden student mi kdysi řekl, že ti matematici si ten vzorec vždy nějak vycucají z prstu. To není pravda. Matematici si vzorec nevymyslí, nýbrž jej objeví. V tomto oddílu společně „objevíme“ dvě důležitá pravděpodobnostní rozdělení, a užijeme si tak opravdové matematiky.

Uvažujme situaci, ve které dochází k výskytu jistého typu náhodné události - touto událostí může být například příchod zákazníka do fronty, příjezd automobilu na parkoviště, přijetí zprávy SMS, narození dítěte v jisté porodnici, apod. V této situaci opakovaného výskytu náhodné události budeme měřit hodnoty dvou veličin - veličinu X , která udává dobu mezi dvěma po sobě jdoucími výskytů události, a veličinu Y , která měří počet výskytů události za časovou jednotku. Určíme nyní rozdělení pravděpodobnosti obou těchto veličin. Označme $p_n(t)$ pravděpodobnost, že v časovém intervalu délky t nastane právě n událostí popsánoho typu. Celé odvození vychází z následujících tří předpokladů:

1. Pravděpodobnost výskytu události v intervalu $(t, t + h)$ závisí pouze na h , nikoli na počtu událostí, které nastaly před okamžikem t , ani na t samotném. Říkáme, že veličina X má nezávislé stacionární přírůstky. Tento předpoklad lze vyjádřit rovnicí

$$p_0(t + h) = p_0(t) \cdot p_0(h) \quad (11.1)$$

2. Platí: $0 < p_0(h) < 1$. Jinými slovy, pravděpodobnost, že v časovém intervalu délky h k výskytu žádné události nedojde, je kladná, ale menší než 1.
3. Pro malá h nastane v intervalu délky h nejvýše jedna událost, tj. platí

$$p_0(h) + p_1(h) = 1. \quad (11.2)$$

Podíváme-li se blíže na funkcionální rovnici 11.1, vidíme, že se jedná o vlastnost exponentu: základ umocněný na součet je roven součinu základů umocněných na jednotlivé členy. Tedy jejím řešením je funkce, která má argument v exponentu - exponenciální funkce. Napišme ji ve tvaru

$$p_0(t) = e^{-\lambda t}$$

pro $t \geq 0$ a kladnou konstantu λ . Pokud tuto exponenciální funkci rozvineme podle známého vzorce (viz 1.ročník - Taylorova řada) v nekonečnou řadu

$$p_0(t) = e^{-\lambda t} = 1 - \lambda t + \frac{(\lambda t)^2}{2!} - \frac{(\lambda t)^3}{3!} + \dots$$

a pro malá t zanedbáme členy obsahující t^2 , t^3 , atd., máme

$$p_0(t) = 1 - \lambda t.$$

Díky vztahu 11.2 tedy pro malá t platí

$$p_1(t) = \lambda t.$$

Z charakteru veličiny X je vidět, že její rozdělení je spojitě (doba mezi dvěma výskyty může být rovna libovolnému kladnému reálnému číslu). Abychom popsali její rozdělení, stačí najít hustotu $f(t)$ veličiny X . Jak naznačuje název tohoto rozdělení - **exponenciální rozdělení pravděpodobnosti** - hustotou bude exponenciální funkce. Protože X může nabývat jen kladných hodnot, musí platit $f(t) = 0$ pro $t < 0$. Zbývá najít $f(t)$ pro $t \geq 0$. V tomto případě bude jednodušší najít nejdříve distribuční funkci $F(t)$ veličiny X , a pak využít toho, že hustota je derivací distribuční funkce (viz kapitola 9).

$$F(t) = P(X < t) = \int_{-\infty}^t f(x) dx = \begin{cases} 0 & \text{pro } t < 0; \\ \int_0^t f(x) dx & \text{pro } t \geq 0. \end{cases}$$

Vypočteme nyní pravděpodobnost, že v čase t k bezprostředně následujícímu výskytu události ještě nedošlo. Použijeme přitom známou fintu z kapitoly 9, že pravděpodobnost určitého jevu lze určit jako 1 minus pravděpodobnost jevu opačného:

$$P(X \geq t) = 1 - P(X < t) = 1 - F(t) = p_0(t) = e^{-\lambda t},$$

tj.

$$1 - F(t) = e^{-\lambda t}.$$

Derivací tohoto vztahu dostaneme

$$0 - f(t) = -\lambda \cdot e^{-\lambda t},$$

tj.

$$f(t) = \lambda \cdot e^{-\lambda t} \text{ pro } t \geq 0.$$

Celkem tedy pro hustotu $f(t)$ exponenciálního rozdělení platí

$$f(t) = \begin{cases} 0 & \text{pro } t < 0; \\ \lambda \cdot e^{-\lambda t} & \text{pro } t \geq 0. \end{cases}$$

Známe tedy už rozdělení veličiny X . Použitím integrace "per partes" lze spočítat podle vzorců z kapitoly 9, že

$$EX = \frac{1}{\lambda}, \quad DX = \frac{1}{\lambda^2}.$$

Naše vymodelované exponenciální rozdělení pravděpodobnosti tedy říká, že k výskytu náhodné události dochází průměrně jednou za $\frac{1}{\lambda}$ časových jednotek, tj. λ -krát za časovou jednotku (takový je význam konstanty λ).

Pokračujme nalezením rozdělení náhodné veličiny Y . K určení nekonečně mnoha hodnot pravděpodobnostní funkce diskrétní veličiny Y (s názvem **Poissonovo rozdělení pravděpodobnosti**) budeme potřebovat určit už dříve označené $p_n(t)$ pro $n \geq 2$.

Z výchozího předpokladu číslo 1 platí pro malá h také

$$\begin{aligned} p_n(t+h) &= P(\text{v intervalu } (0;t) \text{ } n \text{ výskytů, pak mezi } t \text{ a } t+h \text{ žádný výskyt}) \\ &\quad + P(\text{v int. } (0;t) \text{ } (n-1) \text{ výskytů, pak mezi } t \text{ a } t+h \text{ jeden výskyt}) \\ &= p_n(t) \cdot p_0(h) + p_{n-1}(t) \cdot p_1(h). \end{aligned} \quad (11.3)$$

Dosadíme-li za p_0, p_1 , můžeme přepsat rovnice 11.1, 11.3 ve tvaru

$$\begin{aligned} p_0(t+h) &= p_0(t) \cdot (1-h), \\ p_n(t+h) &= p_n(t) \cdot (1-h) + p_{n-1}(t) \cdot \lambda h \end{aligned}$$

pro dostatečně malá h . Úpravou

$$\begin{aligned} \frac{p_0(t+h) - p_0(t)}{h} &= -\lambda \cdot p_0(t), \\ \frac{p_n(t+h) - p_n(t)}{h} &= -\lambda \cdot p_n(t) + \lambda \cdot p_{n-1}(t), \quad n = 1, 2, \dots \end{aligned}$$

a limitním přechodem pro h jdoucí k nule dostáváme systém diferenciálních rovnic

$$\begin{aligned} p_0'(t) &= -\lambda \cdot p_0(t) \\ p_n'(t) &= -\lambda \cdot p_n(t) + \lambda \cdot p_{n-1}(t), \quad n = 1, 2, 3, \dots \end{aligned} \quad (11.4)$$

Systém 11.4 vyřešíme pomocí jedné elegantní metody, na kterou si možná vzpomenete z prvního ročníku: pomocí Z -transformace. V našem případě Z -obrazem posloupnosti $(p_n'(t))_{n=0}^{\infty}$ je funkce komplexní proměnné

$$F(z) = \sum_{n=0}^{\infty} \frac{p_n'(t)}{z^n}.$$

Dosazením máme

$$\begin{aligned} F(z) &= \frac{-\lambda \cdot p_0(t)}{z^0} + \sum_{n=1}^{\infty} \frac{(-\lambda \cdot p_n(t) + \lambda \cdot p_{n-1}(t))}{z^n} = \\ &= -\lambda \cdot \left(\sum_{n=0}^{\infty} \frac{p_n(t)}{z^n} \right) + \lambda \cdot \sum_{n=1}^{\infty} \frac{p_{n-1}(t)}{z^n} = \\ &= -\lambda \cdot \left(\sum_{n=0}^{\infty} \frac{p_n(t)}{z^n} \right) + \frac{\lambda}{z} \cdot \sum_{n=1}^{\infty} \frac{p_{n-1}(t)}{z^{n-1}}. \end{aligned}$$

Dostali jsme tedy rovnici

$$\sum_{n=0}^{\infty} = F(z) = -\lambda \cdot \left(\sum_{n=0}^{\infty} \frac{p_n(t)}{z^n} \right) + \frac{\lambda}{z} \cdot \sum_{n=1}^{\infty} \frac{p_{n-1}(t)}{z^{n-1}}. \quad (11.5)$$

Označme dále

$$P(z, t) := \sum_{n=0}^{\infty} \frac{p_n(t)}{z^n}$$

(čili $P(z, t)$ je Z -obrazem posloupnosti $(p_n(t))_{n=0}^{\infty}$). Pak derivací podle proměnné t dostáváme

$$\frac{\partial P(z, t)}{\partial t} = \sum_{n=0}^{\infty} \frac{p'_n(t)}{z^n}.$$

Nyní dosazením $P(z, t)$ a $\frac{\partial P(z, t)}{\partial t}$ se rovnice 11.5 zjednoduší na

$$\begin{aligned} \frac{\partial P(z, t)}{\partial t} &= -\lambda \cdot P(z, t) + \frac{\lambda}{z} \cdot P(z, t); \\ \frac{\partial P(z, t)}{P(z, t)} &= \lambda \left(\frac{1}{z} - 1 \right) dt. \end{aligned}$$

Integrací obou stran podle t dostaneme

$$\begin{aligned} \ln |P(z, t)| &= \lambda \cdot t \cdot \left(\frac{1}{z} - 1 \right) \\ |P(z, t)| &= e^{\lambda t \left(\frac{1}{z} - 1 \right)} \\ P(z, t) &= K \cdot e^{\lambda t \left(\frac{1}{z} - 1 \right)}, \quad \text{kde } K \in \{-1; 1\}. \end{aligned}$$

Protože platí $P(z, 0) = p_0(0) = 1$, vidíme, že $K = 1$, tj.

$$P(z, t) = e^{\lambda t \left(\frac{1}{z} - 1 \right)}.$$

Nyní při výpočtu zpětné transformace Z^{-1} máme

$$\begin{aligned} Z^{-1}\left(e^{\lambda t\left(\frac{1}{z}-1\right)}\right) &= e^{-\lambda t} \cdot Z^{-1}\left(e^{\frac{\lambda t}{z}}\right) = \\ &= e^{-\lambda t} \cdot Z^{-1}\left(1 + \frac{\lambda t}{z} + \frac{\lambda^2 t^2}{z^2 \cdot 2!} + \frac{\lambda^3 t^3}{z^3 \cdot 3!} + \dots\right), \end{aligned}$$

a tedy

$$p_n(t) = \frac{(\lambda t)^n}{n!} \cdot e^{-\lambda t}, \quad n = 0, 1, 2, 3, \dots$$

A jsme hotovi. Nalezli jsme (respektive objevili) hledané pravděpodobnosti. Většinou se objevené vzorce uvádějí pro $t = 1$, kde pak $p_n(1)$ je pravděpodobnost, že za časovou jednotku $t = 1$ dojde k n výskytům události. Pokud veličina Y udává počet výskytů události za časovou jednotku $t = 1$, její rozdělení se nazývá **Poissonovo rozdělení pravděpodobnosti**: veličina Y nabývá hodnot $0, 1, 2, 3, \dots$ s pravděpodobnostmi

$$p_k = P(Y = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad \text{pro } k = 0, 1, 2, 3, \dots$$

Podobnou strategií jako v kapitole 10 (vytýkáním před sumu a sečítáním nekonečné řady) lze ověřit, že pro střední hodnotu a rozptyl veličiny s Poissonovým rozdělením pravděpodobnosti platí

$$EY = DY = \lambda.$$

Tohle je celkem výjimečný fakt - Poissonovo rozdělení je na rozdíl od většiny jiných takové, že jeho střední hodnota je stejná jako jeho rozptyl. Konstanta λ má přitom též význam jako u veličiny X - označuje průměrný počet výskytů události za časovou jednotku $t = 1$.

11.2 Příklady užití

Příklad 11.1 *Zdravotnický úřad shromažďuje údaje o nově narozených dětech. Průměrně každé dvě hodiny se narodí další dítě. Určete*

- a) *Průměrný počet narozených dětí za rok.*
- b) *Pravděpodobnost, že v daném dnu se nenarodí žádné dítě.*
- c) *Pravděpodobnost, že v jednom dnu se narodí 20 dětí.*
- d) *Pravděpodobnost, že za 4 hodiny se narodí aspoň 5 dětí.*

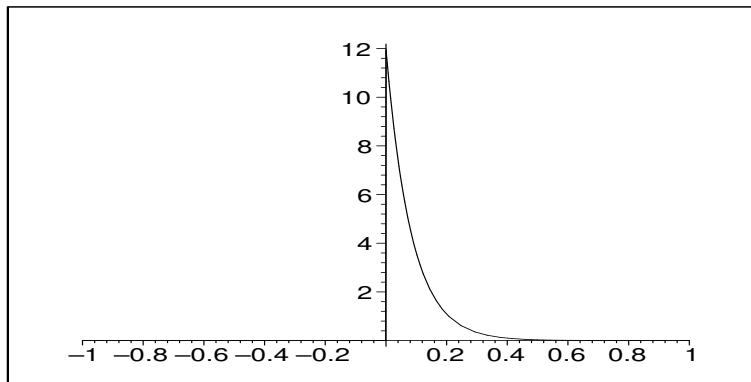
Řešení:

ad a) *Z tohoto úkolu nebudeme dělat vědu. Průměrně jedno dítě za hodinu dává dvanáct dětí za den a $365 \cdot 12 = 4380$ dětí za rok.*

ad b) Základem dobrého využití exponenciálního nebo Poissonova popisu je zvolit si vhodnou časovou jednotku. Pokud hledáme určitý údaj za den, zvolme časovou jednotku jeden den. Druhým krokem po volbě časové jednotky je vypočtení parametru λ . V našem případě $\lambda = 12$ dětí za den (jedná se o průměrný údaj za časovou jednotku). V některých případech, například zde, máme možnost použít buď exponenciální, nebo i Poissonovo rozdělení. pro ilustraci uijeme obou cest. Nejprve tedy označme X dobu mezi dvěma po sobě jdoucími výskyty narození dítěte. Podle podrobného odvození v předchozím oddílu má veličina X exponenciální rozdělení s parametrem $\lambda = 12$. Pak pravděpodobnost, že daný den se nenarodí nikdo, je rovna

$$P(X \geq 1) = 1 - P(X < 1) = 1 - F(1) = 1 - (1 - e^{-12 \cdot 1}) = e^{-12} = 0.00000614$$

(využili jsme raději distribuční funkce $F(t)$ než hustoty $f(t)$, abychom se vyhli integraci; pro $t \geq 0$ platí $F(t) = 1 - e^{-\lambda t}$; pro ilustraci - graf hustoty $f(t)$ rozdělení $Exp(12)$ je uveden na obrázku 11.1 (pro záporná t je rovna nule, pro $t = 0$ je rovna hodnotě parametru λ , pak klesá a asymptoticky se blíží k ose t . Platí $\int_0^{\infty} f(t)dt = 1$.), graf příslušné distribuční funkce $F(t)$ na obrázku 11.2 (pro záporná t je rovna nule, pak začíná konkávně růst a asymptoticky se blíží k hodnotě $y = 1$)).



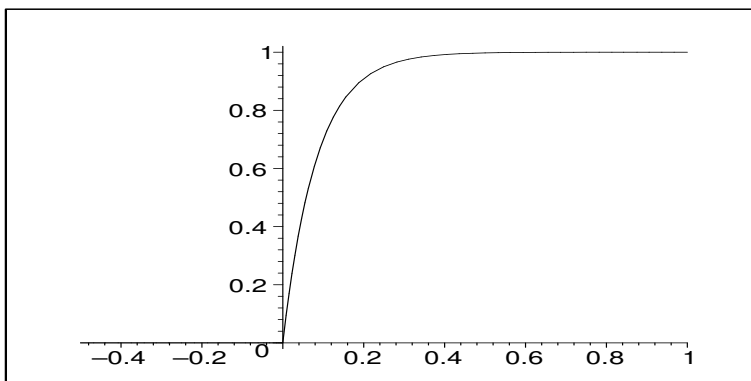
Obrázek 11.1: Graf hustoty $f(t)$ rozdělení $Exp(12)$.

Druhá možná cesta je užít veličiny Y , která udává počet narození za jeden den. Y má Poissonovo rozdělení s parametrem $\lambda = 12$, čili hledaná pravděpodobnost je rovna

$$P(Y = 0) = \frac{12^0}{0!} \cdot e^{-12} = 0.00000614.$$

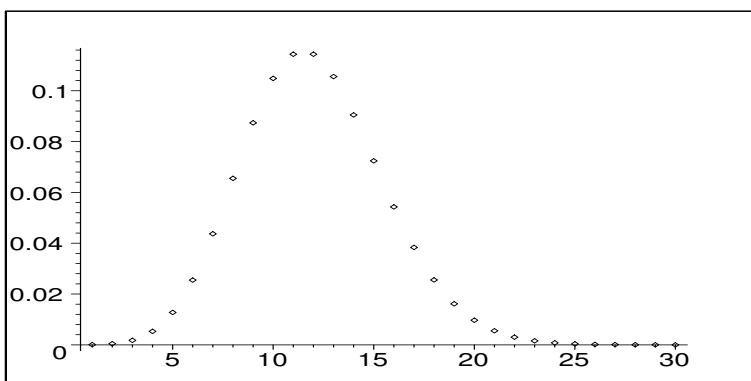
ad c) Využijeme veličiny Y zavedené v b) a dosadíme:

$$P(Y = 20) = \frac{12^{20}}{20!} \cdot e^{-12} = 0.00968$$



Obrázek 11.2: Graf distribuční funkce $F(t)$ rozdělení $Exp(12)$.

Pro ilustraci - graf pravděpodobnostní funkce p Poissonova rozdělení je uveden na obrázku 11.3, graf příslušné distribuční funkce F na obr. 11.4.

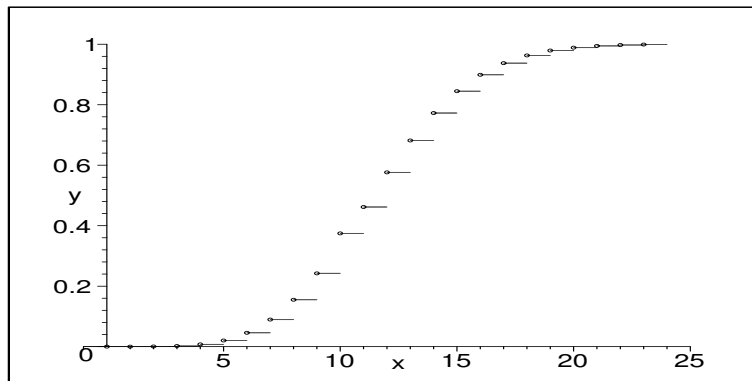


Obrázek 11.3: Graf pravděpodobnostní funkce p rozdělení $Po(12)$.

ad d) *Poslední úkol tohoto příkladu je analogický, ovšem otázka je položena tak, že nás zajímá údaj dosažený za 4 hodiny. Musíme tedy změnit časovou jednotku na 4 hodiny. Tím pádem se mění průměrný počet narození za časovou jednotku na $\lambda = 2$. Označíme-li nyní $Y =$ počet dětí narozených za 4 hodiny, platí $Y \sim Po(\lambda = 2)$. A tedy*

$$\begin{aligned} P(Y \geq 5) &= p(5) + p(6) + p(7) + \dots = 1 - (p(0) + p(1) + p(2) + p(3) + p(4)) = \\ &= 1 - e^{-2} \cdot \left(\frac{12^0}{0!} + \frac{12^1}{1!} + \frac{12^2}{2!} + \frac{12^3}{3!} + \frac{12^4}{4!} \right) = 0.05265 \end{aligned}$$

(místo sečítání nekonečné řady jsme opět odečetli pravděpodobnost opačného jevu od jedničky). Jak je uvedeno na posledním řádku výpočtu, pokud sečítáme několik pravděpodobností Poissonova rozdělení (zejména při písemce na kalkulačce), je vhodné člen $e^{-\lambda}$ vytknout - ušetříme si práci.



Obrázek 11.4: Graf distribuční funkce $F(t)$ rozdělení $Po(12)$: funkce s nekonečně mnoha schody, která vyjadřuje kumulativní pravděpodobnosti $F(t) = P(Y < t)$.

Příklad 11.2 *Na poště mají být instalovány automaty na prodej známek, které po vhození mince vydají přesně za deset sekund žádanou známku. Předpokládáme, že průměrně bude chtít použít automat šest osob za minutu. Kolik automatů bychom měli instalovat, aby s pravděpodobností 0.95 byl i v době největší frekvence obsloužen každý zájemce bez čekání?*

Řešení: V dnešní hektické době jsou i ekonomické požadavky neúprosné: čekat deset sekund je nepřijatelné, na 95% musí být automat k dispozici okamžitě. Klíčem k tomuto příkladu je zjistit, s jakou pravděpodobností přijde jistý počet lidí za deset sekund - to je totiž doba, kdy automat eventuelně někoho obsluhuje a každý další příchozí musí čekat. Zvolme tedy v první řadě časovou jednotku rovnou deseti sekundám. Ve druhé řadě pro tuto časovou jednotku určíme průměrný počet příchozích zákazníků: jestliže průměrně přijde šest za minutu, za deset sekund přijde jeden, čili $\lambda = 1$. Označme $Y =$ počet příchozích zákazníků během deseti sekund. Bystrý čtenář již tuší, že na následujícím řádku prohlásím, že podle přechodného podrobného odvození má veličina Y rozdělení Poissonovo s parametrem $\lambda = 1$.

Položme si nyní následující otázku: Jaká je pravděpodobnost, že během deseti sekund nepřijde více než jeden zákazník (a tedy k okamžitému obsloužení stačí jeden automat)?

$$p = P(Y \leq 1) = P(Y = 0) + P(Y = 1) = e^{-1} \cdot \left(\frac{1^0}{0!} + \frac{1^1}{1!} \right) = e^{-1} \cdot (1 + 1) = 0.73.$$

Tedy jediný automat je dostatečný v 73% času. Ovšem v ostatních 27% příchozí zákazník musí čekat, a to je nepřijatelné. Podívejme se, co říká teorie pro dva nainstalované automaty: Pravděpodobnost, že během deseti sekund přijdou maximálně dva zákazníci, je rovna

$$P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) = 0.73 + P(Y = 2) = 0.92.$$

Tedy v 92% času nový příchozí nemusí čekat. To je ovšem podle našeho zadání stále málo. Spočtěme dále pravděpodobnost, že během deseti sekund přijdou maximálně tři:

$$P(Y \leq 3) = 0.92 + P(Y = 3) > 0.95,$$

a tedy k uspokojení požadavku ze zadání stačí tři automaty.

Příklad 11.3 Výrobní zařízení má poruchu v průměru jednou za 2000 hodin. Veličina X představující dobu čekání na poruchu má exponenciální rozdělení. Určete dobu T tak, aby pravděpodobnost, že přístroj bude pracovat delší dobu než T , byla 0.99.

Řešení. Pravděpodobnost 0.99 je dost vysoká - proto doba T bezporuchového provozu s touto pravděpodobností bude mnohem nižší než 2000 hodin. Určeme nyní T přesně.

V první řadě stanovíme časovou jednotku. Nabízí se jednotka 2000 hodin, tj. budeme teď počítat s čísly, kdy $1 = 2000\text{hod}$. Za druhé stanovíme λ , tj. průměrný počet poruch za časovou jednotku: v našem případě $\lambda = 1$. A tak $X \sim \text{Exp}(\lambda = 1)$. Hledejme teď takovou dobu T , aby $P(X \geq T) = 0.99$. Využijeme opět distribuční funkce $F(t)$, protože její hodnoty jsou přímo rovny jistým kumulativním pravděpodobnostem - a jednu z nich můžeme do posledního vztahu dosadit:

$$\begin{aligned} P(X \geq T) &= 0.99 \\ 1 - P(X < T) &= 0.99 \\ 1 - F(T) &= 0.99 \\ F(T) &= 0.01 \\ 1 - e^{-\lambda T} = 1 - e^{-T} &= 0.01 \\ T &= 0.01005034 \end{aligned}$$

(mezi posledními dvěma řádky je několik kroků vynecháno, ale absolvent prvního ročníku by si s nimi měl poradit). Našli jsme tedy dobu T , po kterou zařízení bude pracovat bez poruchy na 99%. Ovšem musíme tento údaj prezentovat v rozumnějších jednotkách: Pokud $1 = 2000$ hodin, tak

$$T = 0.01005034 = 2000 \cdot 0.01005034 \text{ hodin} = 20.1 \text{ hodin.}$$

11.3 Teorie front

Přirozením rozšířením předchozích úvah je teorie front. Zde bude náhodnou událostí příchod zákazníka do fronty („přirozeným“ ne v tom smyslu, že by člověka hned napadlo se tím zabývat, ale že mnohé modely teorie front z Poissonova a exponenciálního rozdělení vycházejí). Pod frontou nebudeme chápat okluzní frontu nebo válečnou frontu, ale frontu na maso, na mobil, na přístup k tiskárně, frontu u holiče nebo kadeřníka, apod. S tímto druhem front se každý den setkáváme. Při popisu fronty je potřeba modelovat situaci, kdy do fronty lidé přicházejí a současně z ní odcházejí - ne pryč, ale do jednotky obsluhy (tj. ten, kdo je obsluhován, už není ve frontě). Počet příchodů do fronty za časovou jednotku lze dobře popsat Poissonovým rozdělením. Tempo obsluhy (tj. odchody z fronty) lze dobře popsat exponenciálním rozdělením (doba mezi dvěma po sobě jdoucími odchody z fronty je rovna době obsluhy jednoho zákazníka). Ovšem ve skutečných frontách se vyskytují ještě další parametry, nejen příchody a odchody. V následujícím si budeme všimnout různých situací vzhledem k šesti různým parametrům takzvaného Kendallova-Leeova rozšířeného označení $(a|b|c) : (d|e|f)$. Vysvětleme nyní jejich význam:

- a.. Typ rozdělení veličiny X popisující počet příchodů do fronty za jednotku času (pokud hodnota tohoto parametru je M , označuje to tzv. Markovského typ příchodů, což znamená, že X má Poissonovo rozdělení; my se zde budeme bavit pouze o tomto typu příchodů do fronty - průměrný počet zákazníků příšlých do fronty za jednotku času budeme značit λ).
- b.. Typ rozdělení veličiny Y popisující dobu obsluhy jednoho zákazníka (hodnota M označuje tzv. Markovského typ obsluhy, kdy Y má exponenciální rozdělení - pak průměrný počet zákazníků obslužených za jednotku času má pro změnu opět Poissonovo rozdělení - parametr tempa obsluhy budeme označovat μ).
- c.. Počet paralelních serverů = obsluhovacích pultů.
- d.. Typ fronty (např. FIFO = first in first out = první člověk ve frontě je ten, kdo bude nejdřív obsluhován; LIFO = last in first out = ten, kdo přišel poslední, bude obslužen nejdřív (tzv. zásobníková fronta); apod.)
- e.. Maximální dovolený počet zákazníků v systému (systém = fronta + obsluha).
- f.. Velikost zdroje, ze kterého zákazníci přicházejí do fronty.

Samozřejmě, že matematické modely mají své slabiny - například v tomto oddílu nebudeme uvažovat, že zákazník, který se do fronty jednou zařadí, pak za chvíli změní názor a odejde ještě dříve, než je obslužen. Kdybychom i tohle chtěli brát v potaz, situace by byla ještě složitější než ty, kterými se budeme zabývat. To je jasný princip - čím přesnější chceme, aby model popisující realitu byl, tím je složitější (a většinou vždy obsahuje jistou míru pravděpodobnosti).

Všechny následující modely uvažují tento systém: Zákazníci přicházejí do jediné fronty a řadí se za sebe. Jakmile se uvolní místo v obsluze, která sestává z jednoho nebo více paralelních serverů, ten, kdo je ve frontě první, odchází z fronty do obsluhy a začne být obsluhován. Obslužený zákazník odchází pryč.

Důležitou otázkou těchto modelů je, zda existuje tzv. ustálený stav (ten neexistuje, pokud tempo λ příchodů do fronty za časovou jednotku je větší než tempo $\mu \cdot c$ ($= \mu$ krát počet serverů) obslužených zákazníků za jednotku času - v tom případě je systém zahlcen). Pokud existuje ustálený stav systému, budeme se zabývat jeho následujícími charakteristikami:

- p_n .. pravděpodobnost, že v ustáleném stavu je v systému (= frontě + obsluze) právě n zákazníků;
- L_s .. očekávaný (střední, průměrný) počet zákazníků v systému;
- L_q .. očekávaný počet zákazníků ve frontě (q je z anglického fronta = queue);
- W_s .. očekávaná doba strávená zákazníkem v systému;

W_q .. očekávaná doba strávená zákazníkem ve frontě.

Nejdůležitější z uvedených **charakteristik ustáleného stavu** jsou bezesporu pravděpodobnosti p_n , protože pomocí nich určíme všechny ostatní uvedené parametry jako střední hodnoty jistých veličin:

a) Z definice střední hodnoty diskrétní náhodné veličiny plyne

$$L_s = \sum_{n=0}^{\infty} n \cdot p_n;$$

$$L_q = \sum_{n=c}^{\infty} (n - c) \cdot p_n.$$

b) Vztah mezi L a W : $L_s = \lambda \cdot W_s$, $l_q = \lambda \cdot W_q$.

c) Průměrná doba strávená v systému se rovná součtu průměrné doby čekání ve frontě a průměrné doby obsluhy, tj.

$$W_s = W_q + \frac{1}{\mu}.$$

Odtud vynásobením λ a užitím b) máme

$$L_s = L_q + \frac{\lambda}{\mu}.$$

Z posledního vztahu mimo jiné plyne $\lambda = \mu \cdot (L_s - L_q)$.

d) Někdy se díky omezení délky fronty (např. počtem parkovacích míst, počtem telefonátů „v pořadí“ ve frontě, apod.) další zákazníci nemohou do fronty připojit. Pak zavádíme tzv. ovlivněné (cizím slovem efektivní) tempo příchoďů λ_{eff} vyjadřující, že ze všech příchozích se do fronty přidá jen jisté procento, tj.

$$\lambda_{eff} = \lambda \cdot \beta, \text{ kde } \beta \in (0; 1).$$

V takovém případě platí

$$L_s = \lambda_{eff} \cdot W_s;$$

$$L_q = \lambda_{eff} \cdot W_q;$$

$$L_s = L_q + \frac{\lambda_{eff}}{\mu}.$$

Podívejme se nyní na některé konkrétní modely front.

11.3.1 Fronty typu $(M|M|1) : (GD|\infty|\infty)$

První dvě M v identifikaci označují Markovského typ fronty (Markovského typ příchodů do fronty ... tempem λ za jednotku času; Markovského typ obsluhy ... tempem μ za jednotku času) popsaný Poissonovým (popřípadě exponenciálním) rozdělením pravděpodobnosti.

Třetí parametr 1 říká, že obsluha sestává z jedné jednotky (serveru, pokladny, apod.). Čtvrtý parametr GD označuje nespécifikovaný typ fronty, kde určení pravděpodobnosti p_n závisí pouze na střední hodnotě doby čekání ve frontě (pokud bychom chtěli studovat vlastnosti fronty, které závisí nejen na střední hodnotě, ale i na konkrétním rozdělení doby čekání, museli bychom typ fronty specifikovat).

Dále $e = \infty$... délka fronty není nijak omezena; $f = \infty$... velikost zdroje, ze kterého přicházejí do fronty zákazníci, není nijak omezena (ve skutečnosti je tato hodnota vždy omezena např. počtem obyvatel v republice, apod., ale toto číslo je tak velké, že je můžeme označovat třeba jako ∞).

Celkem složitým způsobem (Saaty 1961) se odvodí diferenční rovnice pro p_n za ustáleného stavu $\rho = \frac{\lambda}{\mu} < 1$, které se celkem lehce vyřeší např. pomocí Z -transformace. Výsledek pro ustálený stav:

$$p_n = (1 - \rho) \cdot \rho^n, \quad n = 0, 1, 2, 3, \dots$$

(tzv. geometrické rozdělení pravděpodobnosti - s ním jsme se už jednou setkali v příkladu 8.12 a zde vidíme další jeho využití). Pak

$$\begin{aligned} L_s &= \sum_0^{\infty} n \cdot p_n = \sum_0^{\infty} n(1 - \rho)\rho^n = \\ &= (1 - \rho) \cdot \rho \cdot \frac{d}{d\rho} \left(\sum_0^{\infty} \rho^n \right) = (1 - \rho) \cdot \rho \cdot \left(\frac{1}{1 - \rho} \right)' = \frac{\rho}{1 - \rho}; \\ L_q &= L_s - \frac{\lambda}{\mu} = \frac{\rho^2}{1 - \rho}; \\ W_s &= \frac{L_s}{\lambda} = \frac{1}{\mu(1 - \rho)}; \\ W_q &= \frac{\rho}{\mu(1 - \rho)}. \end{aligned}$$

Příklad 11.4 Do myčky aut přijíždějí auta na základě Poissonova rozložení se střední hodnotou 5 za hodinu. Doba mytí jednoho auta se řídí exponenciálním rozdělením se střední hodnotou 10 minut. Nemůže probíhat mytí více aut najednou. Určete

- Průměrný počet aut ve frontě.
- Kolik parkovacích míst je třeba, aby se aspoň 80% času všechna přijíždějící auta vešla do fronty.

- c) Průměrnou dobu, po kterou je linka nevyužitá.
 d) Průměrnou dobu strávenou zákazníkem v systému.

Řešení: $\lambda = 5$, $\mu = 6$ (za hodinu je průměrně umyto 6 aut). $\rho = \frac{5}{6}$, existuje tedy ustálený stav a má smysl určovat jeho charakteristiky.

ad a) $L_q = \frac{\rho^2}{1-\rho} = 4,17 \doteq 4$ auta. Průměrně budou ve frontě 4 auta.

ad b) Máme najít s tak, aby

$$p_0 + p_1 + p_2 + \dots + p_s \geq 0.8;$$

Dosažením a využitím vzorce pro součet prvních s členů geometrické posloupnosti máme

$$\begin{aligned} 1 - \rho + \rho(1 - \rho) + \dots + \rho^s(1 - \rho) &\geq 0.8 \\ (1 - \rho) \cdot \frac{1 - \rho^{s+1}}{1 - \rho} &\geq 0.8 \\ 0.2 &\geq \rho^{s+1} \\ s + 1 &\geq \frac{\ln 0.2}{\ln \rho} \doteq 6.8 \\ s &\geq 7.8 \end{aligned}$$

Tedy 8 parkovacích míst zajistí, že 80% času se všechna přijíždějící auta vejdou.

ad c) $p_0 = 1 - \rho \doteq 0.17$, tj. linka je nevyužitá 17% času.

ad d) $W_s = \frac{1}{\mu(1-\rho)} = 1$ hodina, tj. manažer by měl přemýšlet, jak zrychlit provoz myčky.

11.3.2 Fronty typu $(M|M|1) : (GD|N|\infty)$

Oproti modelu 11.3.1 je zde jediný rozdíl, a sice ten, že je-li v systému N zákazníků, dalším není dovoleno zařadit se do fronty, čili tempo λ_{eff} zařazení do fronty je menší než tempo λ přijíždějících zákazníků. Systém se někdy nazývá systémem se ztrátami, protože někteří zákazníci jsou ztraceni (= díky plné frontě jedou jinam).

V případě tohoto modelu nemusí platit $\rho < 1$, ustálený stav existuje vždy. Lze odvodit, že

$$p_n = \begin{cases} \frac{1}{N+1} & \dots \rho = 1 \\ \left(\frac{1-\rho}{1-\rho^{N+1}}\right) \cdot \rho^n & \dots \rho \neq 1 \end{cases}$$

pro $n = 0, 1, 2, \dots, N$. Odtud

$$L_s = \begin{cases} \frac{N}{2} & \dots \rho = 1; \\ \sum_0^N n p_n = \frac{1-\rho}{1-\rho^{N+1}} \cdot \sum_0^N n \rho^n = \frac{1-\rho}{1-\rho^{N+1}} \cdot \rho \cdot \frac{d}{d\rho} \left(\frac{1-\rho^{N+1}}{1-\rho}\right) = \frac{\rho(1-(N+1)\rho^N + N\rho^{N+1})}{(1-\rho)(1-\rho^{N+1})} & \dots \rho \neq 1. \end{cases}$$

Dále pravděpodobnost, že zákazník se už do fronty nepřipojí (jede jinam), se rovná p_N , a tedy pravděpodobnost, že příjíždějící zákazník se do fronty připojí, je rovna $1 - p_N$. Odtud

$$\lambda_{eff} = \lambda \cdot (1 - p_N).$$

Ostatní charakteristiky určíme ze vztahů

$$\begin{aligned} W_q &= \frac{L_q}{\lambda_{eff}} = \frac{L_q}{\lambda(1 - p_N)}; \\ L_s &= L_q + \frac{\lambda_{eff}}{\mu} = L_q + \frac{\lambda(1 - p_N)}{\mu}; \\ W_s &= W_q + \frac{1}{\mu} = \frac{L_s}{\lambda(1 - p_N)}. \end{aligned}$$

Také lze ukázat, že platí

$$\lambda_{eff} = \mu \cdot (L_s - L_q).$$

Model sice zachycuje skutečnost, že někteří zákazníci jednou jinam díky plné frontě, ale nepočítá se ztrátou dobré vůle zákazníků, tj. s tím, že zákazníci, kteří museli několikrát odjet díky plné frontě, už třeba příště nepřijedou vůbec.

Příklad 11.5 *Vraťme se k příkladu 11.4 myčky a uvažujme $N = 5 + 1$ (tj. pět parkovacích míst a jedno místo obsluhy v myčce). Určete*

- Kolik aut jede jinam díky plné frontě v průběhu osmihodinové pracovní doby.*
- Průměrnou dobu W_s strávenou zákazníkem v systému.*

Řešení:

$$\text{ad a)} \quad p_N = p_6 = \left(\frac{1 - \frac{5}{6}}{1 - \left(\frac{5}{6}\right)^7} \right) \cdot \left(\frac{5}{6}\right)^6 = 0.0774,$$

tj. počet odrazených zákazníků za hodinu je

$$\lambda - \lambda_{eff} = \lambda \cdot p_N = 5 \cdot 0.0774 = 0.387;$$

Tedy za osm hodin jedou asi $8 \cdot 0.387 \doteq 3$ auta jinam.

ad b)

$$\begin{aligned} L_s &= \sum_0^N np_n = \sum_0^6 np_n = \dots = 2.29 \text{ auta}; \\ W_s &= \frac{L_s}{\lambda_{eff}} = \frac{2.29}{5(1 - 0.0774)} = 0.496 \text{ hodin.} \end{aligned}$$

Tedy oproti neomezené délce fronty (příklad 11.4) byla doba strávená v systému zkrácena z jedné hodiny asi na polovinu za cenu tří ztracených aut denně.

11.3.3 Fronty typu $(M|M|c) : (GD|\infty|\infty)$

Tento typ fronty je analogický typu 11.3.1 s tím rozdílem, že může současně být obsluhováno c zákazníků (říkáme, že obsluha má c jednotek). Podmínka existence ustáleného stavu je

$$\frac{\rho}{c} < 1, \quad \text{tj.} \quad \frac{\lambda}{\mu c} < 1 \quad (\text{pro } \rho = \frac{\lambda}{\mu}).$$

Charakteristiky ustáleného stavu:

$$p_0 = \left[\left(\sum_0^{c-1} \frac{\rho^n}{n!} \right) + \frac{\rho^c}{c!(1-\frac{\rho}{c})} \right]^{-1};$$

$$p_n = \begin{cases} \left(\frac{\rho^n}{n!} \right) \cdot p_0 & \dots \quad 0 < n \leq c; \\ \frac{\rho^n}{c^{n-c} \cdot c!} \cdot p_0 & \dots \quad n > c. \end{cases}$$

Odtud dostaneme

$$L_q = \frac{\rho^{c+1}}{(c-1)! \cdot (c-\rho)^2} \cdot p_0 = \frac{c \cdot \rho}{(c-\rho)^2} \cdot p_c;$$

$$L_s = L_q + \rho;$$

$$W_q = \frac{L_q}{\lambda};$$

$$W_s = W_q + \frac{1}{\mu}.$$

Celkem komplikované výpočty lze apoximovat

$$\text{pro } \rho \ll 1 \quad : \quad p_0 \doteq 1 - \rho, \quad L_q \doteq \frac{\rho^{c+1}}{c^2};$$

$$\text{pro } \rho \doteq 1 \quad : \quad p_0 \doteq \frac{(c-\rho)(c-1)!}{c^c}, \quad L_q \doteq \frac{\rho}{c-\rho}.$$

Příklad 11.6 V malém městě provozovaly taxislužbu dvě firmy, z nichž každá vlastnila dvě auta. Byly koupeny jedním majitelem, který si položil otázku: Jsou oba dispečinky využity, nestačil by jeden? Na každém z dispečinků jsou objednávky stejně časté, asi $\lambda = 10$ za hodinu. Průměrná doba jedné jízdy je 11,5 minut.

Řešení: Otázka zní: co je lepší - dva systémy $(M|M|2)$, kde $\lambda = 10$ a $\mu = 5.217$ v každém z nich, nebo jeden systém $(M|M|4)$, kde $\lambda = 20$ a $\mu = 5.217$?

Poměr využitosti linek $\frac{\rho}{c}$ je stejný v obou situacích, ale jiné parametry jsou odlišné:

Model $(M|M|2)$... $p_0 = 0.0212$, $W_q = 2.16$ hod.

Model $(M|M|4)$... $p_0 = 0.0042$, $W_q = 1.05$ hod., čili pro zákazníky jasně vhodnější model.

11.3.4 Fronty typu $(M|M|c) : (GD|N|\infty)$

Na rozdíl od typu 11.3.3 je nyní N maximální počet zákazníků v systému, tj. $(N - c)$ je maximální délka fronty. Ustálený stav tedy existuje vždy a zde jsou jeho charakteristiky:

$$\begin{aligned}
 p_0 &= \begin{cases} \left[\left(\sum_0^{c-1} \frac{\rho^n}{n!} \right) + \frac{\rho^c}{c!} \cdot (N - c + 1) \right]^{-1} & \dots \frac{\rho}{c} = 1; \\ \left[\left(\sum_0^{c-1} \frac{\rho^n}{n!} \right) + \frac{\rho^c}{c!} \cdot \frac{1 - (\frac{\rho}{c})^{N-c+1}}{1 - \frac{\rho}{c}} \right]^{-1} & \dots \frac{\rho}{c} \neq 1; \end{cases} \\
 p_n &= \begin{cases} \frac{\rho^n}{n!} \cdot p_0 & \dots 0 \leq n \leq c; \\ \frac{\rho^n}{c! \cdot c^{n-c}} \cdot p_0 & \dots c \leq n \leq N; \end{cases} \\
 L_q &= \begin{cases} p_0 \cdot \frac{\rho^c (N-c)(N-c+1)}{2 \cdot c!} & \dots \frac{\rho}{c} = 1; \\ p_0 \cdot \frac{\rho^{c+1} c!}{(c-1)!(c-\rho)^2} \left[1 - \left(\frac{\rho}{c} \right)^{N-c} - (N-c) \left(\frac{\rho}{c} \right)^{N-c} \left(1 - \frac{\rho}{c} \right) \right] & \dots \frac{\rho}{c} \neq 1; \end{cases} \\
 L_s &= L_q + (c - \bar{c}) = L_q + \frac{\lambda_{eff}}{\mu},
 \end{aligned}$$

kde \bar{c} je očekávaný počet nečinných serverů ($\bar{c} = \sum_0^c (c - n)p_n$). $(c - \bar{c})$ je očekávaný počet využitých serverů a pro efektivní tempo příchodů λ_{eff} platí

$$\lambda_{eff} = \lambda(1 - p_N) = \mu(c - \bar{c}).$$

Příklad 11.7 Vraťme se k příkladu 11.6 a modelu typu $(M|M|4)$. Pokud se dispečink omluví, když je ve frontě už 16 žadatelů o odvoz (tj. $N = 16 + 4 = 20$), pak

$$\begin{aligned}
 p_0 &= 0.00753; \\
 L_q &= 5.85; \\
 p_{20} &= 0.03433 \implies \lambda_{eff} = \lambda(1 - p_{20}) = 19.31; \\
 W_q &= \frac{L_q}{\lambda_{eff}} = 0.303 \text{ hod} \doteq 18 \text{ minut}.
 \end{aligned}$$

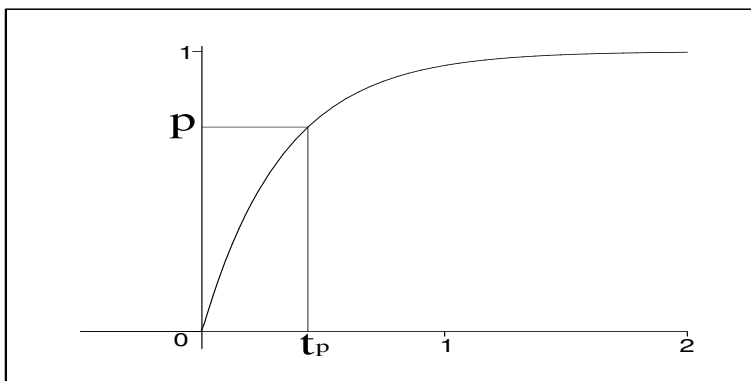
Doba čekání ve frontě je tedy dále zkrácena na úkor ztráty $p_{20} \cdot 100 = 3.4\%$ zákazníků. Samozřejmě tento model neříká nic o ztrátě dobré vůle některých zákazníků po dlouhodobém provozu (ve skutečnosti bude ztráta klientely větší než 3,4%).

11.4 Náhodné generování hodnot Po a Exp na počítači

V některých oborech (pokud chceme sestavit model fronty na počítači, nebo v jiných oblastech) se někdy využívá tzv. simulace, tj. hodnoty veličin získáváme náhodně. Při náhodném generování veličiny X , která má rozdělení $Exp(\lambda)$, využijeme její distribuční funkce

$$F(t) = \begin{cases} 0 & \dots t < 0; \\ 1 - e^{-\lambda t} & \dots t \geq 0. \end{cases}$$

Využijeme toho, že distribuční funkce představuje přechod mezi hodnotami pravděpodobnosti z intervalu $(0; 1)$ a hodnotami, kterých nabývá veličina X . Abychom získali hodnotu



Obrázek 11.5: Pro kladná t je distribuční funkce $F(t)$ prostá, a proto pro $p \in (0; 1)$ existuje jediná hodnota $t_p \in (0, \infty)$ tak, že $F(t_p) = p$.

veličiny X , náhodně vygenerujeme (to většinou počítač umí - funkce RANDOM) hodnotu p z intervalu $(0; 1)$. Tuto hodnotu p nabývá distribuční funkce v jediném bodě t_p , který se nazývá **p -kvantil** - viz obr. 11.5.

Z rovnice $p = F(t_p)$ tedy vypočteme t_p :

$$p = 1 - e^{-\lambda \cdot t_p}$$

$$t_p = -\frac{1}{\lambda} \cdot \ln(1 - p)$$

Při náhodném generování hodnot veličiny Y s Poissonovým rozdělením $Po(\lambda)$ využijeme vztahu mezi exponenciálním a Poissonovým rozdělením - opakovaně generujeme hodnoty veličiny X a sčítáme je, dokud nepřesáhnou danou časovou jednotku; pak náhodná hodnota veličiny Y je rovna počtu těchto opakování zmenšenému o jedničku.

Například pro $\lambda = 3$ generujeme postupně náhodné hodnoty veličiny X , získáváme

0.1626 0.0176 0.2447 0.1318 0.9436

(nyní součet těchto pěti hodnot přesáhl časovou jednotku 1, a proto náhodně získaná hodnota Y je rovna $5 - 1 = 4$).

11.5 Shrnutí pojmů

Exponenciální i Poissonovo rozdělení pravděpodobnosti jsou dva pravděpodobnostní modely popisující tutéž situaci. Příslušné veličiny však vyjadřují různé věci:

X ... doba mezi dvěma po sobě jdoucími výskyty události. Toto rozdělení se nazývá exponenciální (značíme: $X \sim Exp(\lambda)$). X je spojitá veličina, která nabývá kladných hodnot.

Y ... počet výskytů události za časovou jednotku. Toto rozdělení se nazývá Poissonovo (označujeme: $Y \sim Po(\lambda)$). Y je diskrétní veličina, která nabývá hodnot z množiny $\{0, 1, 2, 3, \dots\}$.

Parametr λ se dosazuje do obou rozdělení tentýž a udává průměrný počet výskytů události za jednotku času.

Kromě jednoduchých případů využití slouží obě rozdělení jako odrazový můstek matematického popisu teorie front. Exkurze zde zdaleka nebyla vyčerpávající. Lze odvodit popis dalších systémů, jako je samoobslužný model, fronty s prioritou, sériově řazené fronty, apod.

12 Rovnoměrné a normální rozdělení pravděpodobnosti

V minulé kapitole jsme se blíže seznámili už s jedním typem spojitého rozdělení pravděpodobnosti, v této kapitole budeme se spojitými modely pokračovat - čekají nás další dva. Budeme se zabývat zejména normálním rozdělením, protože to tvoří základ nejčastěji používaných statistických testů.

12.1 Rovnoměrné rozdělení pravděpodobnosti

Rovnoměrné rozdělení pravděpodobnosti je velmi jednoduchým typem spojitého rozdělení. Díky tomu je model při popisu konkrétních situací celkem nepřesný. Používáme jej jen zřídkka. Ovšem i jednoduché věci se mohou někdy hodit (třeba u zkoušky - když člověk nezná ani to jednoduché, jak potom může znát to složitější?).

Řekneme, že veličina X má **rovnoměrné rozdělení pravděpodobnosti**, pokud nabývá hodnot z intervalu $\langle a, b \rangle$ konečné délky a libovolná hodnota z tohoto intervalu je stejně pravděpodobná jako ty ostatní. Hustota této veličiny je dána vztahem

$$f(t) = \begin{cases} \frac{1}{b-a} & \dots t \in \langle a; b \rangle; \\ 0 & \dots \text{jinak,} \end{cases}$$

pro distribuční funkci $F(t)$ platí (mohli bychom to též odvodit, protože platí $F(t) = \int_{-\infty}^t f(x)dx$)

$$F(t) = \begin{cases} 0 & \dots t \leq a; \\ \frac{t-a}{b-a} & \dots t \in (a; b); \\ 1 & \dots t \geq b. \end{cases}$$

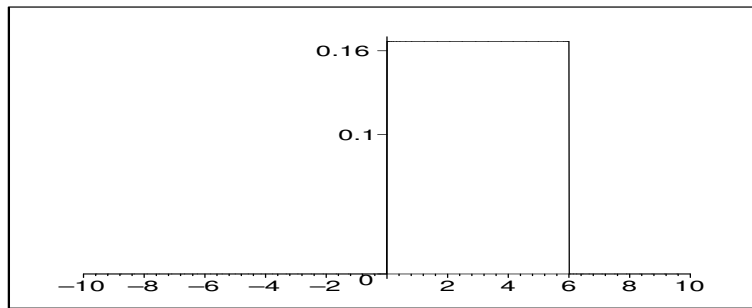
Označení rovnoměrného rozdělení je $Ro(a, b)$.

Příklad 12.1 *Nejmenovaný student bydlící v Bystřici jezdí výhradně tramvají číslo 1. Ovšem nikdy se doma nedívá do jízdního řádu, kdy tramvaj jede - to je pod jeho úroveň. Tramvaj jezdí v šestiminutových intervalech. Student přijde vždy na zastávku naprosto náhodně a čeká na svou oblíbenou „number one“. Dobu X jeho čekání na tramvaj lze popsat rovnoměrným rozdělením na intervalu $\langle 0; 6 \rangle$. Je pedagogické nakreslit grafy hustoty i distribuční funkce této veličiny (viz obr. 12.1 a 12.2).*

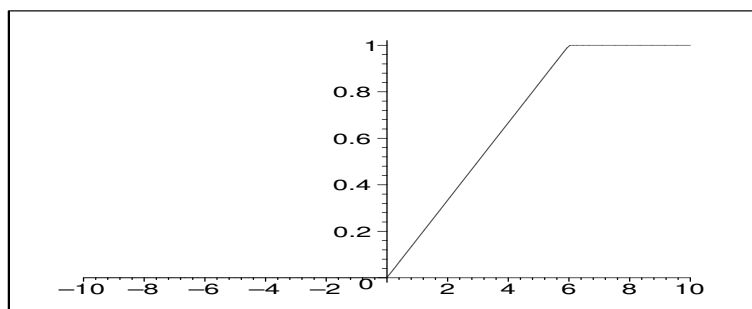
Když bychom nyní chtěli určit pravděpodobnost, že student bude na tramvaj čekat 4 až 6 minut, podle vzorců z kapitoly 9 máme

$$P(X \in (4; 6)) = \int_4^6 f(t)dt = \int_4^6 \frac{1}{6}dt = \frac{1}{3}.$$

Také platí, že $P(X \in (4; 6;)) = \frac{6-4}{6-0}$, tj. hledaná pravděpodobnost je rovna poměru délek dvou úseček. Jinými slovy, rovnoměrné rozdělení zachycuje právě situace geometrické pravděpodobnosti na intervalu (geometrická pravděpodobnost na intervalu je rovna poměru délek úseček - jak bylo řečeno v úvodní kapitole o pravděpodobnostních modelech, vidíme, že tento speciální případ je ve spojitých modelech zahrnut).



Obrázek 12.1: Hustota rovnoměrného rozdělení pravděpodobnosti $Ro(0; 6)$.



Obrázek 12.2: Distribuční funkce rovnoměrného rozdělení pravděpodobnosti $Ro(0; 6)$.

12.2 Normální rozdělení pravděpodobnosti

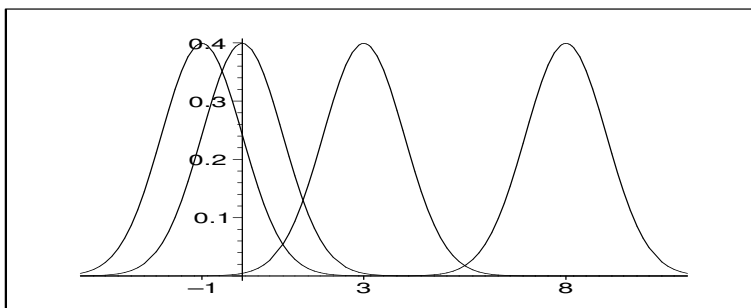
Normální rozdělení pravděpodobnosti je rozdělení pro veličiny spojitého typu a má hustotu

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

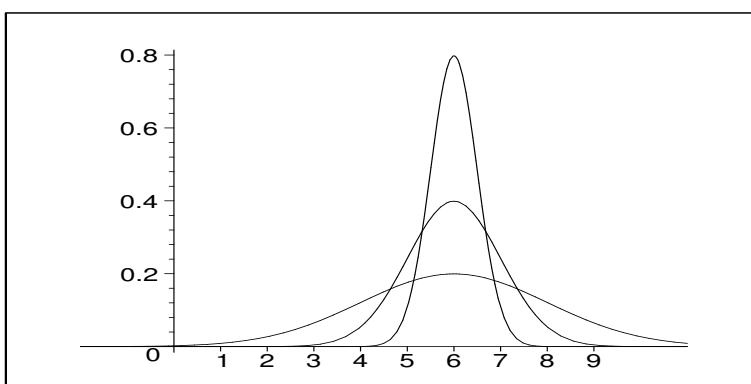
Vzorec této funkce na první pohled nemá příjemný tvar a asi by jej nikdo nechtěl potkat v noci na liduprázdné ulici. Dalo by se spočítat, že střední hodnota veličiny X s rozdělením zadaným touto hustotou je rovna parametru μ , rozptyl je roven parametru σ^2 . Proto budeme značit $No(\mu, \sigma^2)$. Na obr. 12.3 jsou uvedeny grafy hustoty pro σ^2 stále rovno jedné a různé střední hodnoty μ , na obr. 12.4 je $\mu = 6$ a mění se hodnoty rozptylu σ^2 (Při malém rozptylu je rameno grafu hustoty vysoké a úzké, pro větší rozptyl hustota nabývá nižších funkčních hodnot, ale interval s hodnotami významně odlišnými od nuly je širší). U všech těchto grafů hustot platí $\int_{-\infty}^{\infty} f(t)dt = 1$.

Normální rozdělení se stalo slavným díky tomu, co říká tzv. **centrální limitní věta**:

Jestliže X_1, X_2, \dots, X_N jsou navzájem nezávislé veličiny, které mají všechny stejné rozdělení (nemusí být normální, ale libovolné, jeho střední hodnota je $EX_i = \mu$ a rozptyl $DX = \sigma^2$), pak součtem těchto veličin je náhodná veličina Y (platí $Y = \sum_1^N X_i$) se střední hodnotou



Obrázek 12.3: Hustota normálního rozdělení pro různé střední hodnoty μ .



Obrázek 12.4: Hustota normálního rozdělení pro různé rozptyly σ^2 .

$EY = N \cdot \mu$ a rozptylem $DY = N \cdot \sigma^2$, která má pro dostatečně velké N ($N > 30$) normální rozdělení, tj. platí

$$P(Y \in (a; b)) = \int_a^b \frac{1}{\sqrt{2\pi} \text{sqrt} N\sigma} \cdot e^{-\frac{(t-N\mu)^2}{2N\sigma^2}} dt.$$

To, že hodně proměnných lze s velkou přesností popsat pomocí normálního rozdělení, je právě důsledkem centrální limitní věty. Následující dvě situace to dokreslují.

Příklad 12.2 Y_1 udává výšku borovic v daném lese (v metrech). Průměrná výška ($= \mu$) je 50 metrů. Vezměme nyní jeden konkrétní strom, jehož výška je 54 metrů. Co způsobilo, že vyrostl o 4 metry nad průměr? Hodně různých vlivů:

- Stromek byl zasazen v obzvlášť příznivém období roku, což způsobilo, že vyrostl o 1m nad průměr.
- Místo, kde strom roste, získává zdroje hnojiva navíc, což vede k růstu o 2.3m nad průměr.
- Nešťastnou náhodou byl stromek při sazení nalomen, což znamená, že narostl o 1.4m nižší, než mohl.

- d) *Strom má dobré místo na slunci, což mu pomohlo vyrůst o 2m nad průměr.*
- e) *Skupina příslušníků antagonistického hmyzu si vybrala strom za svůj domov, což mu vzalo šance vyrůst o 0.6m výš než ostatní stromy.*

atd.

Zkrátka a dobře, vychýlení 4m nad průměr je dáno součtem všech těchto možných kladných i záporných vlivů. Protože těchto vlivů je většinou poměrně dost, výslednou výšku stromu danou součtem všech těchto vlivů lze s velkou přesností popsat normálním rozdělením.

Příklad 12.3 Y_2 udává výsledek zkoušky z matematiky. Vezmeme nyní výsledek zkoušky jednoho konkrétního studenta. Co naň mělo vliv?

- a) *Honza měl den před zkouškou chřipku. To snížilo jeho výkon o 5 bodů.*
- b) *Honza si něco tipl a náhodou to trefil - přidalo mu to 2 body.*
- c) *Honza chyběl na klíčové přednášce a neměl u zkoušky její kopii - přišel o 5 bodů.*
- d) *Profesor byl v dobré náladě a při opravování Honzovi 3 body přidal zadarmo.*

atd.

Opět vidíme, že výsledek Honzovy zkoušky je dán součtem většího počtu navzájem nezávislých náhodných vlivů, a tedy jej lze s velkou přesností popsat normálním rozdělením.

Příklad 12.4 *Speciálně i binomické rozdělení lze pro dostatečně velké N dobře popsat (aproximovat, nahradit) normálním rozdělením:*

Uvažujme například veličinu X , která udává počet líců při 100 hodech korunou. Tato veličina má binomické rozdělení s parametry

$$N = 100, \quad p = \frac{1}{2}; \quad EX = Np = 50; \quad DX = Np(1 - p) = 25.$$

Tuto veličinu lze vyjádřit jako součet veličin X_1, X_2, \dots, X_{100} , kde X_i má binomické rozdělení s parametry $N = 1, p = \frac{1}{2}$, tj. udává počet líců v jediném hodu mincí (pro $N = 1$ se binomické rozdělení někdy nazývá alternativní rozdělení, protože veličina může zde nabývat pouze dvou alternativ: 0 nebo 1).

Jako součet stejně rozdělených nezávislých veličin lze tedy X s velkou přesností popsat normálním rozdělením s parametry (pro $N = 100$)

$$\mu = EX = N \cdot EX_i = Np = 50, \quad \sigma^2 = DX = N \cdot DX_i = Np(1 - p) = 25.$$

Čili pro dostatečně velké N lze binomické rozdělení s velkou přesností aproximovat normálním rozdělením se stejnou střední hodnotou a rozptylem.

12.3 U -rozdělení

Uvažujme náhodnou veličinu X udávající výsledky zkoušky z matematiky, kterou lze s velkou přesností popsat normálním rozdělením (viz příklad 12.3) s hustotou $f(t)$ a parametry

$$\mu_x = 75, \quad \sigma_x^2 = 25.$$

Její normované hodnoty (viz kap. 9) budeme chápat jako hodnoty veličiny U , kde

$$U = \frac{X - \mu_x}{\sigma_x} = \frac{X - 75}{5}$$

a platí

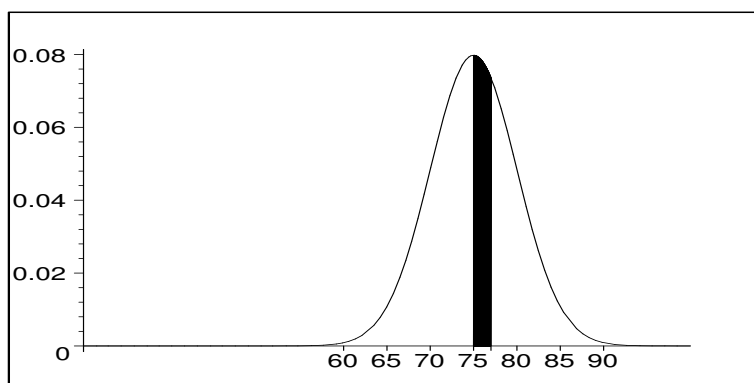
$$\begin{aligned} EU &= \int_{-\infty}^{\infty} \frac{t - \mu_x}{\sigma_x} \cdot f(t) dt = \frac{1}{\sigma_x} \left(\int_{-\infty}^{\infty} t \cdot f(t) dt - \mu_x \cdot \int_{-\infty}^{\infty} f(t) dt \right) = \\ &= \frac{1}{\sigma_x} (\mu_x - \mu_x \cdot 1) = 0; \end{aligned}$$

$$\begin{aligned} DU &= E(U^2) - E^2U = EU^2 - 0 = \int_{-\infty}^{\infty} \left(\frac{t - \mu_x}{\sigma_x} \right)^2 \cdot f(t) dt = \\ &= \frac{1}{\sigma_x^2} \int_{-\infty}^{\infty} (t - \mu_x)^2 \cdot f(t) dt = \frac{1}{\sigma_x^2} \cdot \sigma_x^2 = 1. \end{aligned}$$

Zajímá-li nás pravděpodobnost, s jakou student dosáhne výsledku mezi 75 a 77 body, musíme spočítat

$$P(75 \leq X \leq 77) = \int_{75}^{77} \frac{1}{\sqrt{2\pi} \cdot 5} \cdot e^{-\frac{(t-75)^2}{50}} dt,$$

což je obsah vyšrafované plochy na obrázku 12.5.

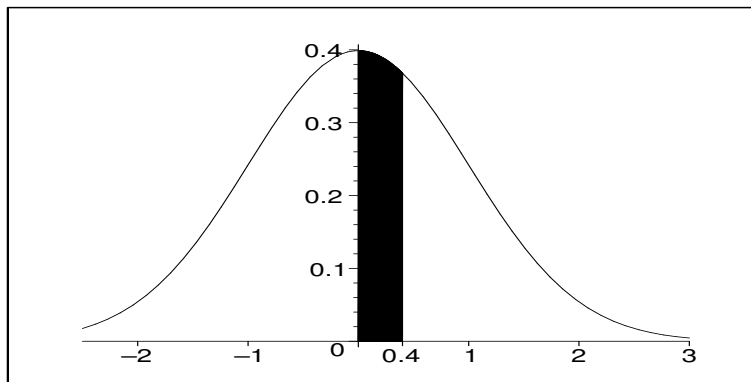


Obrázek 12.5: Obsah šrafované plochy je roven pravděpodobnosti, že X nabude hodnot z intervalu $\langle 75; 77 \rangle$.

Tato pravděpodobnost je stejná jako pravděpodobnost, že veličina U nabude hodnot z intervalu určeného příslušnými normovanými hodnotami

$$\frac{75 - 75}{5} \text{ a } \frac{77 - 75}{5}, \text{ tj. } P(75 \leq X \leq 77) = P(0 \leq U \leq 0.4) = \int_0^{0.4} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} du,$$

což je obsah šrafované plochy na obrázku 12.6.



Obrázek 12.6: Obsah šrafované plochy je roven pravděpodobnosti, že U nabude hodnot z intervalu $\langle 0; 0.4 \rangle$. Tento obsah je stejný jako obsah šrafované plochy z obr. 12.5.

Platí tedy

$$\int_{75}^{77} \frac{1}{\sqrt{2\pi} \cdot 5} \cdot e^{-\frac{(t-75)^2}{2 \cdot 25}} dt = \int_{\frac{75-75}{5}}^{\frac{77-75}{5}} f(u) du,$$

kde $f(u)$ je hustota U -rozdělení, tj. libovolný integrál z hustoty normálního rozdělení lze převést na integrál z hustoty rozdělení U .

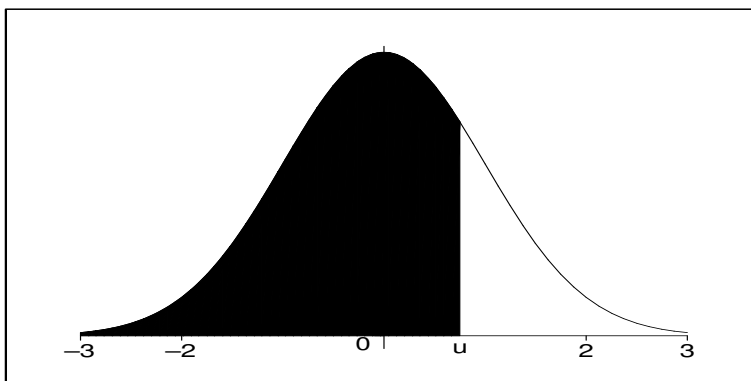
Velichina U má tedy normální rozdělení $No(\mu = 0; \sigma^2 = 1)$ a nazýváme ji **standardizované normální rozdělení** (v anglické literatuře Z -distribution a Z -value).

Výpočty uvedených integrálů jsou dosti pracné, a proto se s výhodou používá následujícího postupu: pravděpodobnostní výpočty obecného normálního rozdělení se převedou právě popsáním postupem na výpočet integrálu U -rozdělení, pro které byla vypočtena a sestavena tabulka integrálů

$$\Phi(u) = P(U < u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt$$

($\Phi(u)$ je označení distribuční funkce rozdělení U - jako pravděpodobnost má svůj geometrický význam, což znázorňuje obrázek 12.7).

Protože graf funkce $f(u)$ je symetrický vzhledem ke svislé ose (přímce $u = 0$), v tabulce nemusí být uvedeny hodnoty $\Phi(u)$ pro záporná u . Platí totiž pro $u > 0$:

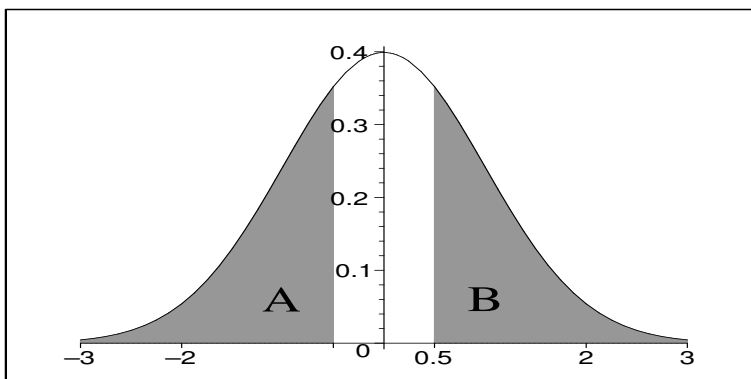


Obrázek 12.7: Obsah šrafované plochy je roven funkční hodnotě distribuční funkce $\Phi(u)$ rozdělení U .

$$\Phi(-u) = 1 - \Phi(u)$$

Pravdivost tohoto tvrzení je patrná z toho, že na obou stranách rovnosti v rámečku je obsah téže plochy. Např. $\Phi(-0.5) = 1 - \Phi(0.5)$, protože (viz obr. 12.8) funkce $f(u)$ je symetrická a celkový obsah plochy pod křivkou je roven jedné:

$$\Phi(-0.5) = S(A) = S(B) = 1 - \Phi(0.5)$$



Obrázek 12.8: Obsahy ploch A a B jsou stejné.

Hodnoty funkce $\Phi(u)$ jsou uvedeny v tabulce 12.1 a 12.2.

Příklad 12.5 Veličinu X udávající výsledek zkoušky lze popsat rozdělením $No(\mu = 75; \sigma^2 = 25)$. S jakou pravděpodobností je výsledek zkoušky

- v intervalu $< 69; 72 >$?
- menší než 65?

c) větší než 80?

d) v intervalu $< \mu_x - 3\sigma_x; \mu_x + 3\sigma_x > ?$

Řešení:

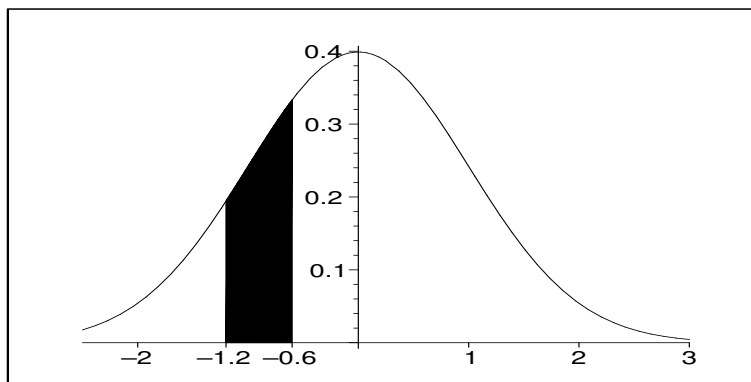
ad a)

$$\begin{aligned}
 P(69 \leq X \leq 72) &= P\left(\frac{69 - \mu_x}{\sigma_x} \leq \frac{X - \mu_x}{\sigma_x} \leq \frac{72 - \mu_x}{\sigma_x}\right) = \\
 &= P\left(\frac{69 - 75}{5} \leq U \leq \frac{72 - 75}{5}\right) = \\
 &= P(-1.2 \leq U \leq -0.6) = \Phi(-0.6) - \Phi(-1.2) = \\
 &= 1 - \Phi(0.6) - (1 - \Phi(1.2)) = \\
 &= \Phi(1.2) - \Phi(0.6) = 0.8849303 - 0.7257469 = 0.1591834,
 \end{aligned}$$

což je obsah plochy na obrázku 12.9. Pokud si zvědavý čtenář položil otázku, proč místo některých neostrých nerovností nejsou v tomto odvozování ostré a naopak, pak bych mu rád připomněl, že u spojitých veličin platí

$$P(X = t_0) = 0$$

pro libovolné t_0 . Díky tomu nezáleží na tom, zda u normálního rozdělení definujeme distribuční funkci předpisem $F(t) = P(X \leq t)$ nebo $F(t) = P(X < t)$ (tyto dva druhy definice se totiž objevují v matematické literatuře oba, ale žádný velký vliv to nemá - u spojitých veličin to nemá žádný vliv, u diskrétních veličin je schodová distribuční funkce v prvním případě zprava spojitá, ve druhém zleva spojitá, tj. v bodě skoku je v prvním případě funkční hodnota definována na horním schodu, ve druhém případě na dolním).



Obrázek 12.9: K př. 12.5a) - výpočet pravděpodobnosti u normálního rozdělení je roven obsahu šrafované plochy.

ad b)

$$\begin{aligned} P(X \leq 65) &= P\left(\frac{X - \mu_x}{\sigma_x} \leq \frac{65 - \mu_x}{\sigma_x}\right) = P\left(U \leq \frac{65 - 75}{5}\right) = \\ &= P(U \leq -2) = \Phi(-2) = 1 - \Phi(2) = 1 - 0.9772499 = 0.0227501. \end{aligned}$$

ad c)

$$\begin{aligned} P(X \geq 80) &= P\left(U \geq \frac{80 - 75}{5}\right) = P(U \geq 1) = 1 - P(U < 1) \\ &= 1 - \Phi(1) = 1 - 0.8413447 = 0.1586553. \end{aligned}$$

ad d)

$$\begin{aligned} P(\mu_x - 3\sigma_x \leq X \leq \mu_x + 3\sigma_x) &= P\left(\frac{\mu_x - 3\sigma_x - \mu_x}{\sigma_x} \leq U \leq \frac{\mu_x + 3\sigma_x - \mu_x}{\sigma_x}\right) = \\ &= P(-3 \leq U \leq 3) = \Phi(3) - \Phi(-3) = \Phi(3) - (1 - \Phi(3)) = \\ &= 2\Phi(3) - 1 = 0.9973002 \end{aligned}$$

Většina hodnot veličiny X leží tedy v intervalu $\langle \mu_x - 3\sigma_x, \mu_x + 3\sigma_x \rangle$. Veličina X nabude hodnoty z tohoto intervalu s pravděpodobností 99.7%.

Příklad 12.6 Firma vyrábí balíčky ořechů po 200ks, přičemž $\frac{3}{4}$ oříšků jsou burské a $\frac{1}{4}$ lískové, dokonale se promíchají, a pak se teprve sypou do balíčků. Jestliže koupíme jeden balíček ořechů, jaká je pravděpodobnost, že počet lískových ořechů je v intervalu $\langle 47; 56 \rangle$?

Řešení. Náhodná veličina X udávající počet lískových ořechů v jednom balíčku má rozdělení $Bi(N = 200, p = 0.25)$, čili $\mu_x = 50$, $\sigma_x^2 = 37.5$. Přímý výpočet

$$\begin{aligned} P(47 \leq X \leq 56) &= P(X = 47) + P(X = 48) + \dots + P(X = 56) = \\ &= \binom{200}{47} 0.25^{47} 0.75^{153} + \binom{200}{48} 0.25^{48} 0.75^{152} + \dots + \binom{200}{56} 0.25^{56} 0.75^{144} \end{aligned}$$

je příliš pracný (musíme vyčíslout velká kombinační čísla), a proto se využívá aproximace normálním rozdělením se stejnou střední hodnotou a rozptylem ($\sigma_x^2 = 37.5 \implies \sigma_x \doteq 6.12$):

$$\begin{aligned} P(47 \leq X \leq 56) &= P\left(\frac{47 - 50}{6.12} \leq U \leq \frac{56 - 50}{6.12}\right) = \Phi(0.98) - \Phi(-0.49) = \\ &= \Phi(0.98) - (1 - \Phi(0.49)) \doteq 0.524. \end{aligned}$$

Předchozí příklad dává obecnou představu o tom, jak funguje náhrada binomického rozdělení normálním rozdělením. Ovšem přece jen se dopouštíme jisté nepřesnosti, pramenící z faktu, že diskrétní rozdělení nahrazujeme spojitým. Tato nepřesnost je velká zejména pro malé N . Ale provedením tzv. **korekce** lze tuto nepřesnost snížit, jak ukazuje následující příklad.

Příklad 12.7 Vraťme se k příkladu 9.2, kde náhodná veličina X udává počet líců při čtyřech hodech mincí. Vypočteme například pravděpodobnost, že počet líců ve čtyřech hodech bude jeden nebo dva

- a) pomocí $Bi(N = 4, p = 0.5)$;
- b) pomocí normálního rozdělení;
- c) pomocí normálního rozdělení s korekcí.

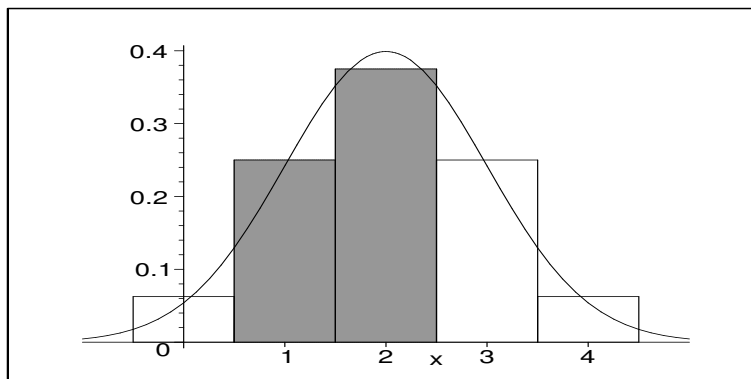
Řešení:

ad a) $P(1 \leq X \leq 2) = p_1 + p_2 = 0.25 + 0.375 = 0.625$.

ad b) Aproximujme binomické rozdělení normálním rozdělením $No(\mu_x = Np = 2, \sigma_x^2 = Np(1-p) = 1)$:

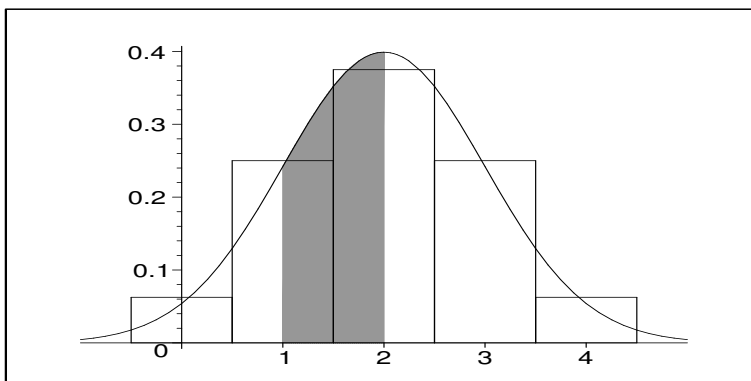
$$\begin{aligned} P(1 \leq X \leq 2) &= P\left(\frac{1 - \mu_x}{\sigma_x} \leq U \leq \frac{2 - \mu_x}{\sigma_x}\right) = P\left(\frac{1 - 2}{1} \leq U \leq \frac{2 - 2}{1}\right) = \\ &= \Phi(0) - \Phi(-1) = 0.341. \end{aligned}$$

ad c) Hodnota z b) se od hodnoty z a) významně liší!! Kde se udála tak velká chyba? V tom, že obsah plochy dvou obdélníků histogramu na obr.12.10

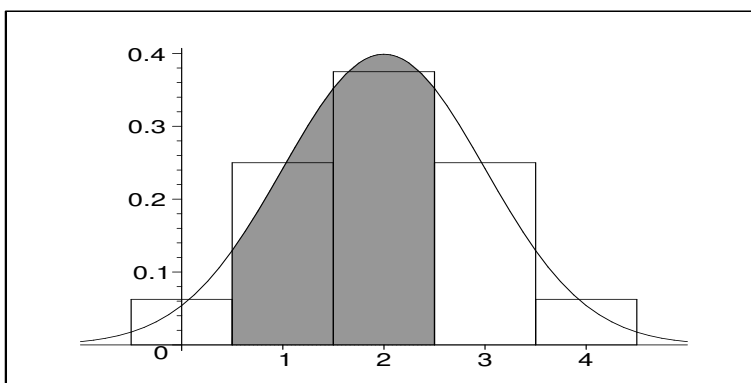


Obrázek 12.10: K př. 12.7 - aproximovaná plocha.

*jsme aproximovali pomocí obsahu plochy na obr. 12.11,
nikoliv pomocí šrafované plochy na obr. 12.12.*



Obrázek 12.11: K př. 12.7 - nevhodná aproximace Bi pomocí No .



Obrázek 12.12: K př. 12.7 - vhodná aproximace Bi pomocí No užitím korekce.

Aproximační chyba se zmenší, pokud výpočet pravděpodobnosti $P(t_1 \leq X \leq t_2)$ pomocí Bi nahradíme obsahem podgrafu hustoty No na intervalu stejné délky, tj. pravděpodobností $P(t_1 - 0.5 \leq X \leq t_2 + 0.5)$. Toto rozšíření intervalu o 0.5 na obou stranách nazýváme korekcí. V našem příkladu pak dostaneme užitím korekce:

$$\begin{aligned} P(1 - 0.5 \leq X \leq 2 + 0.5) &= P\left(\frac{1 - 0.5 - 2}{1} \leq U \leq \frac{2 + 0.5 - 2}{1}\right) = \\ &= \Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{3}{2}\right) = 0.624, \end{aligned}$$

což je docela dobrá aproximace přesné hodnoty 0.625. Je vidět, že pomocí korekce lze popsat binomické rozdělení normálním i pro malá N .

12.4 U -test

V anglické literatuře ... Z -test. Problematiku statistického testování v případě normálního rozdělení vysvětlíme na následujícím příkladu:

Příklad 12.8 Dlouhodobá praxe ukazuje, že žárovky ACME mají životnost, kterou lze popsat normálním rozdělením s parametry $\mu = 100$ hodin, $\sigma^2 = 25$. Vývojové oddělení firmy ACME se pokouší prakticky realizovat teoretický fakt, že jistý chemický povlakový proces zvyšuje životnost žárovky. Provádí se jednoduchý experiment: povlakový proces se realizuje na jednu žárovku a měří se její životnost. Pokud bude životnost větší než 100 hodin - řekněme 115 nebo 120, bude to potvrzovat, že povlakový proces zvyšuje životnost. Pokud žárovka vydrží jen asi 100 hodin, povede to k závěru, že povlakový proces nepřináší zlepšení životnosti.

V této situaci je možné statistické rozhodování provést na základě tří různých postupů užívajících U -rozdělení, které se liší v alternativní hypotéze H_1 . Na našem příkladu nyní provedeme všechny tři typy testu a porovnáme vhodnost užití toho kterého z nich. V každém z testů se provádí stejné obecné kroky jako v případě znaménkového testu v kapitole 10.

12.4.1 Kvalitativní test

Tento test bychom použili, kdyby teorie dokazovala, že povlakový proces zvýší životnost žárovky o jistou konkrétní hodnotu, např. o 8 hodin. Tj. žárovka, která by bez povlakového procesu vydržela 101 hodin, bude po nanesení povlaku pracovat 109 hodin, žárovka s životností 97 hodin zvýší povlakem svou životnost na 105 hodin, apod. Projdeme nyní pět kroků testu, respektive v případě kvalitativního testu je kroků šest.

(K1) Vyslovíme nulovou a alternativní hypotézu:

H_0 : Povlakový proces nemá vliv na životnost, tj. $\mu_p = 100$ (střední hodnota životnosti žárovky ošetřené povlakovým procesem je stále rovna původním 100 hodinám).

H_1 : Povlakový proces zvýší životnost o 8 hodin, tj. $\mu_p = 108$.

Alternativní hypotéza H_1 se nazývá konkrétní (nebo též hodnotová, kvalitativní), protože μ_p v ní se rovná konkrétní hodnotě.

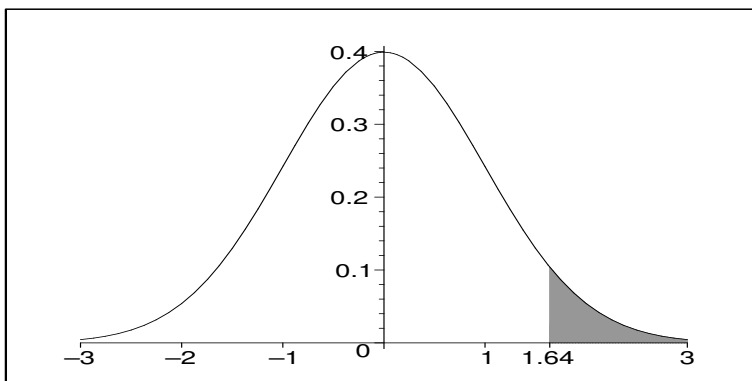
(K2) Kritériem testu bude naměřená doba T životnosti žárovky podrobené povlakovému procesu.

(K3) Pokud nulová hypotéza H_0 je pravdivá, veličina T má normální rozdělení s parametry $\mu_p = 100$, $\sigma_p^2 = 25$ (tj. rozdělení doby životnosti se povlakovým procesem nezmění).

(K4) Najdeme T_k tak, aby platilo $P(T \geq T_k) = \alpha = 0.05$:

(i) Převědeme rozdělení veličiny T na U -rozdělení: $U = \frac{T-100}{5}$.

(ii) V tabulce funkce Φ najdeme pro $\alpha = 0.05$ hodnotu $U_{1-\alpha}$ takovou, že (viz obr. 12.13)



Obrázek 12.13: Obsah šrafované plochy je roven $P(U \geq 1.64) = \alpha = 0.05$.

$$\begin{aligned} P(U \geq u_{1-\alpha}) &= \alpha \\ 1 - \Phi(u_{1-\alpha}) &= \alpha \\ 0.95 &= \Phi(u_{0.95}) \\ u_{0.95} &= 1.64 \end{aligned}$$

Podle tabulky sice $\Phi(1.64)$ není přesně rovno hodnotě 0.95, ale budeme s jistou přesností říkat, že to rovno je. Číslo 1.64 je vlastně 0.95-kvantil U -rozdělení (viz oddíl 11.4), protože pro ten právě platí, že $\Phi(1.64) = 0.95$. Obrázek 12.13 představuje geometrický význam kvantilu ve vztahu k hustotě: obsah plochy mezi hustotou a osou t na intervalu $< -\infty; U_{0.95} >$ je roven právě hodnotě 0.95. A tedy obsah zbytku podgrafu (= šrafovaná část) je roven 0.05.

(iii) Převedeme tuto U -hodnotu zpět na T -hodnotu:

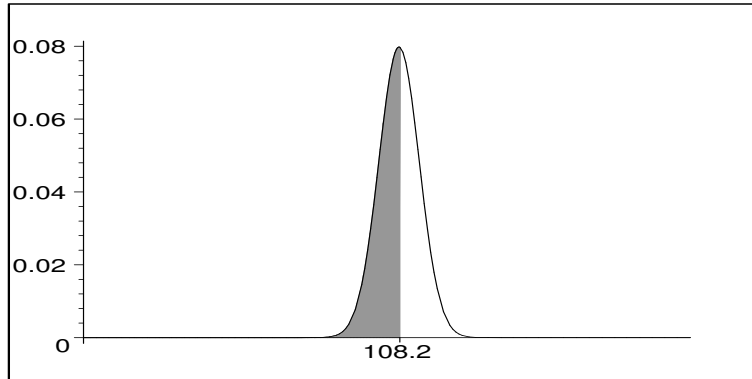
$$\begin{aligned} 1.64 &= \frac{T_k - 100}{5} \\ T_k &= 5 \cdot 1.64 + 100 = 108.2 \end{aligned}$$

Kritickou hodnotu vždy hledáme pro předem zvolenou hladinu významnosti testu α . To nemusí být vždy 0.05, ale třeba 0.01 nebo jiná hodnota - už o tom byla řeč v kapitole 10.

(K5) Rozhodnutí testu: pokud naměřená hodnota životnosti žárovky podrobené povlakovému procesu přesáhne kritickou hodnotu $T_k = 108.2$ (tj. odpovídající U -hodnota přesáhne hodnotu 1.64), zamítáme H_0 a uzavíráme, že povlakový proces zvyšuje životnost.

(K6) Alternativní hypotézu H_1 lze nyní využít k výpočtu pravděpodobnosti β výskytu chyby druhého druhu, tj. situace, kdy platí H_1 (životnost T žárovky s povlakem má normální rozdělení s parametry $\mu_p = 108$, $\sigma_p^2 = 25$), ale nezamítneme H_0 (naměřené $T < T_k = 108.2$). Pravděpodobnost, že nezamítneme H_0 , ačkoliv platí H_1 , je rovna

obsahu vyšrafované plochy na obrázku 12.14 - hustota na obrázku je hustotou platnosti H_1 , tj. svého extrému nabývá v bodě střední hodnoty $t = 108$.



Obrázek 12.14: Obsah šrafované plochy je roven pravděpodobnosti β chyby druhého druhu.

$$\begin{aligned}\beta &= P(T \leq 108.2) = P\left(U \leq \frac{108.2 - 108}{5}\right) = \\ &= P(U \leq 0.04) = \Phi(0.04) = 0.516\end{aligned}$$

Číslo $1 - \beta$ (= pravděpodobnost toho, že H_0 zamítneme správně (oprávněně), když platí H_1) se nazývá **síla testu**. Obecně je to pojem pozitivní, protože vyjadřuje jakousi úspěšnost testu.

12.4.2 Jednostranný test

Tento test bychom použili, kdyby teorie dokazovala, že povlakový proces zvýší životnost žárovky, ale neříkala, o kolik. Projdeme kroky testu:

(K1) H_0 : $\mu_p = 100$ (střední hodnota životnosti se povlakovým procesem nezmění).

H_1 : $\mu_p > 100$.

Alternativní hypotéza se zde nazývá jednostranná (nebo směrovaná).

(K2) až (K5) Tyto kroky jsou stejné jako v případě kvalitativního testu, protože konkrétní hodnota 108 kvalitativního testu splňuje alternativní hypotézu jednostranného testu $\mu_p > 100$. Při nerovnosti „>“ mluvíme o **pravostranném testu**. V případě alternativní hypotézy $\mu_p < 0$ bychom museli lehce obměnit výpočet kritické hodnoty, která by byla menší než 100 a testové rozhodnutí by zamítlo H_0 tehdy, když by platilo $T \leq T_k$. **Levostranný test** pro konkrétní hodnotu necháme na cvičení.

(K6) Hodnotu β chyby druhého druhu nemůžeme určit, protože neznáme, čemu se rovná μ_p . Víme jen, že $\mu_p > 100$. Jednostranné testy užíváme mnohem častěji než ty kvalitativní. Vskutečnosti téměř nikdy neznáme přesnou hodnotu μ_p - ale kvalitativní test má svou pedagogickou hodnotu, a proto byl uveden. V této souvislosti by se dalo více mluvit o pojmu síly testu, ale až někdy jindy - dnes už ne.

12.4.3 Oboustranný test

Tento test bychom použili, kdyby naše informace o chemickém povlakovém procesu byly tak nejasné, že bychom nevěděli, zda se povlakem životnost sníží nebo zvýší.

(K1) H_1 : $\mu_p = 100$ (střední doba životnosti se povlakem nezmění).

H_2 : $\mu_p \neq 100$ (střední doba se povlakem změní, ale nevíme, kterým směrem).

Alternativní hypotéza H_1 se nazývá oboustranná (nebo nesměrovaná).

(K2) Kritériem je doba T životnosti žárovky s povlakem.

(K3) Za předpokladu platnosti H_0 má veličina T rozdělení $No(\mu_p = 100, \sigma_p^2 = 25)$.

(K4) H_0 zamítneme tehdy, když naměřená hodnota veličiny T bude příliš malá nebo příliš velká, přičemž obojí má stejnou váhu. Tedy hledáme kritické hodnoty T_m, T_v tak, aby platilo

$$P(T \leq T_m) + P(T \geq T_v) = \alpha = 0.05$$

a oba členy na levé straně měly stejnou váhu, tj.

$$\begin{aligned} P(T \leq T_m) &= \frac{\alpha}{2} = 0.025; \\ P(T \geq T_v) &= \frac{\alpha}{2} = 0.025. \end{aligned}$$

(i) Převedeme T -rozdělení na U -rozdělení: $U = \frac{T-100}{5}$.

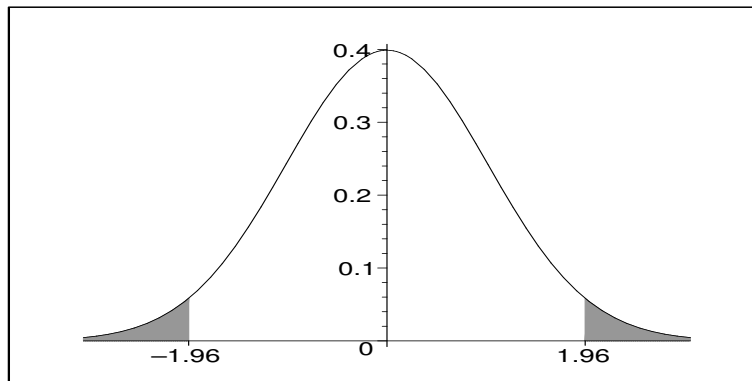
(ii) V tabulce hodnot funkce Φ najdeme $u_{\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}}$ tak, aby

$$P(U \leq u_{\frac{\alpha}{2}}) = \frac{\alpha}{2}; \quad P(U \geq u_{1-\frac{\alpha}{2}}) = \frac{\alpha}{2}.$$

Pro $\alpha = 0.05$ dostáváme (viz obr. 12.15)

$$\begin{aligned} 0.975 &= \Phi(u_{0.975}) \\ u_{0.975} &= 1.96 \end{aligned}$$

Ze symetrie hustoty U -rozdělení dostaneme $u_{0.025} = -1.96$.



Obrázek 12.15: Význam kritických hodnot oboustranného testu - obsah každého z obou šrafovaných konců je roven $\frac{\alpha}{2}$.

(iii) U -hodnoty převedeme na T -hodnoty:

$$-1.96 = \frac{T_m - 100}{5} \implies T_m = 100 - 5 \cdot 1.96 = 90.2$$

$$1.96 = \frac{T_v - 100}{5} \implies T_v = 100 + 5 \cdot 1.96 = 109.8$$

(K5) Rozhodnutí testu: Pokud je naměřená (= empirická) hodnota doby životnosti žárovky podrobené povlakovému procesu mimo interval $(90.2; 109.8)$, tj. odpovídající U -hodnota je mimo interval $(-1.96; 1.96)$, zamítáme H_0 a uzavíráme, že povlakový proces má vliv na životnost.

Pokud by životnost pozorované žárovky byla 109 hodin, jednostranný test (s mezní hodnotou $T_k = 108.2$) by zamítl H_0 , oboustranný test (s mezní hodnotou $T_v = 109.8$) by H_0 nezamítl. Jak je možné, že stejná data vedou při různých alternativních hypotézách k různým rozhodnutím?

Když nemáme žádný teoretický podklad toho, že povlakový proces zvyšuje životnost, musí se použít oboustranný test pro pravdivostní obor $(90.2; 109.8)$. a $\alpha = 0.05$. Kdybychom měli teoretický podklad o tom, že povlakový proces zvyšuje životnost, pro $\alpha = 0.05$ by platila kritická hodnota $T_k = 108.2$ pravostranného testu, tj. pravdivostní obor je interval $(0; 108.2)$.

Kdybychom měli teoretický důkaz o tom, že povlakový proces snižuje životnost, odvodilo by se pomocí jednostranného testu směřovaného na opačnou stranu (= levostranného testu), že pravdivostní obor pro nezamítnutí H_0 ($T_k = 100 - 5 \cdot 1.64 = 91.8$) je $(91.8; \infty)$ pro $\alpha = 0.05$.

Oba jednostranné testy tedy dávají jakýsi „přísnější“ pravdivostní obor $(91.8; 108.2)$ pro nezamítnutí H_0 , ale pak je celková chyba prvního druhu rovna součtu chyb obou jedno-

stranných testů, tj.

$$\alpha = 0.05 + 0.05 = 0.1$$

(chyby 0.05 se můžeme dopustit na obě strany). Odpověď v nastoleném dilematu tedy je: pokud nevíme nic o teorii a naměřená životnost žárovky je 109 hodin, pak

- a) nezamítáme H_0 na hladině významnosti 0.05;
- b) zamítáme H_0 na hladině významnosti 0.1.

Statistiku bychom nevhodně použili tehdy, když nevíme nic o teorii, životnost upravené žárovky naměříme 109 hodin, řekneme si: „aha, povlakový proces zvyšuje životnost“ a vymyslíme teorii, která naše tvrzení podporuje. Výsledky pak publikujeme v odborném časopise pro $\alpha = 0.05$. Ovšem ve skutečnosti je to jinak - bez teorie musíme publikovat výsledky pouze na hladině významnosti $\alpha = 0.1$. Jednostranný test pro $\alpha = 0.05$ lze použít jen při jasném teoretickém základu - např. tehdy, když je naprosto jasné, že povlakový proces nemůže vést k nižší životnosti.

Špatné použití statistiky tkví v tom, že na základě jednostranného testu se vyrábí teorie, a pak se publikuje na hladině významnosti 0.05 místo 0.1. Tímto špatným použitím statistiky lze „dokázat“ platnost čehokoliv - na určité hladině významnosti lze tvrdit jakýkoliv nesmysl: čím větší nesmysl, tím větší α se musí použít.

12.5 Shrnutí pojmů

Normální rozdělení je mocným nástrojem k popisu dějů a procesů praxe. A dokonce to má i teoretický podklad - díky centrální limitní větě můžeme říct, že mnohé veličiny závisí na velkém množství přibližně stejných vlivů, a tudíž se chovají „normálně“, tj. lze je popsat normálním rozdělením. I když každý člověk v naší republice je jiný, přece jen je v tom několikamiliónovém shromáždění jistý řád. A i v procesech náhodných, jako je rychlost větru nebo množství srážek, je řád. Je to zvláštní, že uprostřed náhody je „zakódován“ řád. Jako by to ani nebyla „náhoda“, ale jakýsi tanec podle jistých pravidel. V chaosu je řád. Neukazuje to na někoho většího, kdo stanovil pravidla našemu srdci i přírodním procesům? Škoda, že v hodinách matematiky se zamýšlíme jen nad otázkou „jak“, a ne „proč“.

Matematicky vzato, pracné pravděpodobnostní výpočty pomocí normálního rozdělení se převedou na výpočet pomocí standardizovaného normálního rozdělení $U = No(0; 1)$ a využíváme integrace jednou provždy zaznamenané do tabulek. Platí

$$P(t_1 \leq X \leq t_2) = P\left(\frac{t_1 - \mu_x}{\sigma_x} \leq U \leq \frac{t_2 - \mu_x}{\sigma_x}\right) = \Phi\left(\frac{t_2 - \mu_x}{\sigma_x}\right) - \Phi\left(\frac{t_1 - \mu_x}{\sigma_x}\right).$$

Statistický test popsáný v této kapitole je celkem chudý v tom, že zpracovává jen jediné měření. Abychom získali větší jistotu, že např. povlakový proces zvyšuje životnost jistého

druhu žárovek, neprovedeme měření s jednou žárovkou, ale s několika, a pak spočteme například průměr měřených parametrů. Toto přirozené a požadované rozšíření statistického testu na soubor hodnot je obsahem následující poslední kapitoly.

13 Statistický test střední hodnoty normálního rozdělení při známém rozptylu

Je slušností, aby poslední kapitola byla nejkratší. Budu se toho držet.

Až dosud byla řeč o dvou typech rozdělení, a sice teoretickém a empirickém. Nyní do svých úvah přibereme třetí typ rozdělení, který charakterizuje vztah mezi prvními dvěma typy: teoretické rozdělení parametru empirického rozdělení. Toto rozdělení hraje roli ve statistickém testu této kapitoly.

13.1 Teoretické rozdělení parametru empirického rozdělení

Příklad 13.1 *Uvažujme všechny studenty posledního ročníku čtyřletých středních škol v České republice. Všichni píší měsíc před maturitou souhrnný test z matematiky. Je známo, že střední hodnota ohodnocení testu je $\mu = 500$ bodů, směrodatná odchylka $\sigma = 100$ bodů (jedná se o teoretické rozdělení celé populace maturitních studentů - teoreticky předpokládáme, že rozdělení je stejné jako např. v minulém roce). Náhodně vybereme 9 studentů a z jejich ohodnocení vypočteme průměr $\bar{X}_1 = 513$ a empirickou směrodatnou odchylku $S_1 = 87$. Potom opět náhodně vybereme jiných 9 studentů a z jejich ohodnocení testu vypočteme průměr $\bar{X}_2 = 485$, empirická směrodatná odchylka $S_2 = 165$. Tímto způsobem jsme získali dvě empirická rozdělení počtu bodů vybraného vzorku devíti studentů, první má parametry $\bar{X}_1 = 513$, $S_1 = 87$, druhé má parametry $\bar{X}_2 = 485$, $S_2 = 165$.*

Soustředme se na některý z parametrů těchto empirických rozdělení, například na empirickou směrodatnou odchylku S . S má pro různé náhodně vybrané vzorky devíti studentů různé hodnoty: $S_1 = 87$, $S_2 = 165$, atd. Zkratka - je to náhodná veličina, a jako náhodná veličina má jisté rozdělení pravděpodobnosti. pokud bychom rozdělení veličiny S znali, mohli bychom spočítat například pravděpodobnost, že u náhodně vybraného vzorku devíti studentů bude směrodatná odchylka S ohodnocení testu větší než 110, apod.

Toto rozdělení veličiny S má jednu úžasnou vlastnost: nezávisí na datech konkrétního vzorku, ale platí pro celou populaci - říká něco o směrodatné odchylce vzorku, ale nezávisí na žádném konkrétním vybraném vzorku; naopak, obsahuje informace o parametrech všech možných vybratelných vzorků dané velikosti. Proto je toto rozdělení teoretické, i když popisuje směrodatnou odchylku rozdělení empirického!! Říká se mu teoretické rozdělení empirické směrodatné odchylky.

13.2 Teoretické rozdělení průměru \bar{X}

Zaměříme se na veličinu \bar{X} průměru vzorku délky N , protože ta bude základem statistického testu této kapitoly. Je to náhodná veličina, její rozdělení budeme nazývat teoretické rozdělení průměru. Pokud měříme hodnoty veličiny s normálním rozdělením $No(\mu, \sigma^2)$, průměr těchto hodnot má také normální rozdělení s parametry, které budeme označovat

$\mu_{\bar{X}}, \sigma_{\bar{X}}^2$. Vypočtěme tuto střední hodnotu a rozptyl:

$$\mu_{\bar{X}} = E\bar{X} = E\frac{1}{N}\sum_1^N X_i = \frac{1}{N}\sum_1^N EX_i = \frac{1}{N} \cdot N \cdot \mu = \mu,$$

kde μ je střední hodnota teoretického normálního rozdělení celé populace (protože $\mu_{\bar{X}} = \mu$, index \bar{X} budeme většinou vynechávat).

Nyní se musím zmínit o jisté nepřesnosti nebo kolizi značení - pokud si vzpomínáte, v kapitole 9 jsme dosazovali při výpočtu průměru hodnoty x_i , kdežto nyní jsem při výpočtu $E\bar{X}$ použil (velké) X_i - proč? (Malé) x_i značí konkrétní naměřenou hodnotu veličiny X , kterou celou dobu uvažujeme, kdežto velké X_i značí náhodnou veličinu, jejíž hodnotu malé x_i měříme (velké X_i je tedy jakási „teoretická naměřená hodnota“). V podstatě $X_i = X$, tedy \bar{X} je průměr N stejně rozdělených veličin X . Ale z jistého důvodu „srozumitelnosti“ dodáváme zde index i , aby bylo zřejmé, že počítáme průměr něčeho, co lze téměř nazvat hodnotami veličiny X . Tento posun zde vznikl právě tím, že místo konkrétních hodnot x_i jsme začali přemýšlet o vlastnostech průměru \bar{X} teoreticky. A při teoretickém zkoumání vlastností empirického rozdělení musíme uvažovat „teoretické“ hodnoty. Je to důležité, protože kdybychom tuto úvahu neprovedli a jen slepě dosazovali do vzorce, platí $Ex_i = x_i$ (protože x_i je konstanta), zatímco $EX_i = EX = \mu$ (protože X_i je náhodná veličina).

Tento rozdíl je odlišen i v terminologii: vektor (x_1, x_2, \dots, x_N) nazýváme **souborem hodnot**, vektor (X_1, X_2, \dots, X_N) nazýváme **náhodným výběrem**.

Při odvozování $\sigma_{\bar{X}}^2$ využijeme následující fakta:

- $DX_i = \sigma^2 = EX_i^2 - \mu^2 \implies EX_i^2 = \sigma^2 + \mu^2$.
- Pokud X_i, X_j jsou nezávislé veličiny, $EX_i X_j = EX_i \cdot EX_j = \mu \cdot \mu = \mu^2$.
- Suma $\sum_1^N X_i \cdot X_j$ má N členů pro $i = j$ a $N^2 - N$ členů, kde $i \neq j$.

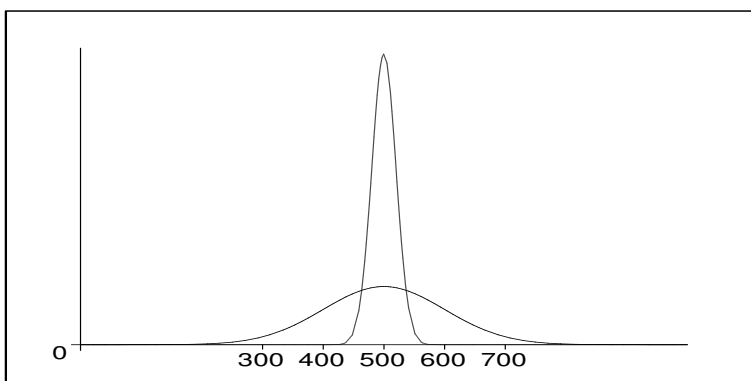
A nyní už k vlastnímu odvození:

$$\begin{aligned} \sigma_{\bar{X}}^2 &= E\bar{X}^2 - (E\bar{X})^2 = E\left(\frac{1}{N}\sum_1^N X_i\right)^2 - \mu^2 = \frac{1}{N^2} \cdot E(X_1 + \dots + X_N)^2 - \mu^2 = \\ &= \frac{1}{N^2} \cdot \left[E\sum_1^N X_i^2 + E\sum_{i \neq j} X_i X_j \right] - \mu^2 = \\ &= \frac{1}{N^2} \cdot [N(\sigma^2 + \mu^2) + (N^2 - N)\mu^2] - \mu^2 = \frac{\sigma^2}{N}. \end{aligned}$$

Vidíme, že rozptyl průměru \bar{X} je jiný než rozptyl průměru celé populace. Ilustrujme tento fakt na příkladu.

Příklad 13.2 Vrátime-li se k situaci v příkladu 13.1, studujme tvar rozdělení průměru \bar{X} souboru N hodnot pro různá N :

- a) Vybíráme-li vzorky studentů velikosti jedna ($N = 1$), získáváme například $\bar{X}_1 = 700$ (výjimečně inteligentní student), $\bar{X}_2 = 456$, $\bar{X}_3 = 498, \dots$. Průměr je vždy přímo roven jediné hodnotě vzorku. Teoretické rozdělení průměru je stejné jako původní teoretické rozdělení celé populace studentů před maturitou, tj. $\mu_{\bar{X}} = 500$, $\sigma_{\bar{X}}^2 = \frac{10000}{1} = 10000$.
- b) Pro $N = 25$ budou průměry vzorků pětadvaceti studentů stále přibližně na téže hodnotě $\mu_{\bar{X}} = 500$, zatímco rozptyl bude menší ($\sigma_{\bar{X}}^2 = \frac{10000}{25} = 400$). Na obrázku 13.1 jsou porovnány hustoty v případě a) a b) - je vidět, že hustota v případě b) nabývá hodnot podstatně odlišných od nuly na mnohem užším intervalu, tj. rozptyl je menší (což se projeví „užším“ grafem nabývajícím vyšších hodnot, aby byla zachována vlastnost $\int_{-\infty}^{\infty} f(t)dt = 1$).



Obrázek 13.1: Graf hustoty rozdělení průměru je pro $N = 25$ užší než pro $N = 1$.

- c) Pro $N = 200000$ (celá populace středoškolských studentů) $\mu_{\bar{X}} = 500$, $\sigma_{\bar{X}}^2 = \frac{10000}{200000} = 0.05$. U vzorků velikosti srovnatelné s velikostí celé populace je rozptyl téměř zanedbatelný - průměr vzorku se od střední hodnoty liší jen nepatrně.

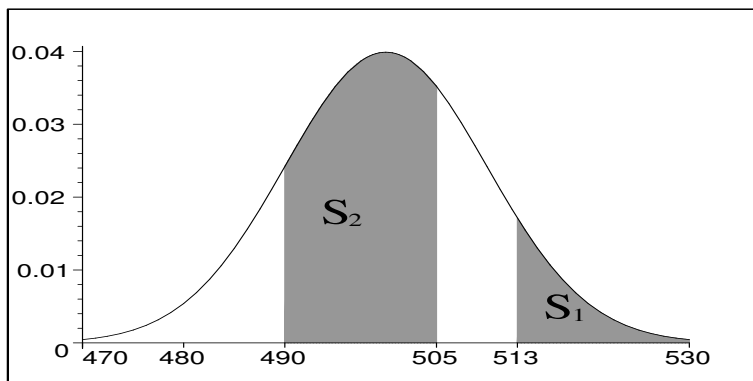
Rozptyl teoretického rozdělení průměru vzorku tedy pro rostoucí délku vzorku klesá od σ^2 k nule.

Příklad 13.3 Uvažujme situaci příkladů 13.1, 13.2 pro délku vzorku $N = 100$, tj. $\sigma_{\bar{X}}^2 = \frac{10000}{100} = 100$, čili $\sigma_{\bar{X}} = 10$.

- a) Jaká je pravděpodobnost, že průměr \bar{X} ohodnocení vybraného vzorku 100 studentů bude ≥ 513 ?
- b) Jaká je pravděpodobnost, že \bar{X} bude ležet v intervalu $< 490; 505 >$?

Řešení:

- ad a) Při řešení použijeme stejného postupu jako v předchozí kapitole: převedení na U -hodnoty a tabulku 12.1, 12.2. Hledaná pravděpodobnost je rovna obsahu plochy S_1 na obr. 13.2:



Obrázek 13.2: Výpočet pravděpodobností v příkladu 13.3.

$$\begin{aligned} P(\bar{X} \geq 513) &= P\left(\frac{\bar{X} - 500}{10} \geq \frac{513 - 500}{10}\right) = P(U \geq 1.3) = \\ &= 1 - \Phi(1.3) = 0.097. \end{aligned}$$

ad b) Pravděpodobnost je rovna obsahu plochy S_2 :

$$\begin{aligned} P(490 \leq \bar{X} \leq 505) &= P\left(\frac{490 - 500}{10} \leq U \leq \frac{505 - 500}{10}\right) = \\ &= \Phi(0.5) - \Phi(-1) = 0.532. \end{aligned}$$

13.3 Testy o střední hodnotě průměru při známém rozptylu

13.3.1 Test „ $\mu = \text{konst}$ ”

Kroky testu vysvětlíme na konkrétním příkladu. Základní filozofie je stejná jako u testů v předchozích dvou kapitolách.

Příklad 13.4 V situaci z příkladu 13.1 založili studenti FEKT firmu KAPPA a vyvinuli program INTEL, jehož cílem je zlepšit znalosti matematiky u středoškolských studentů, zejména pak zlepšit výsledky souhrnného testu.

Chtějí svůj program INTEL otestovat, a proto náhodně vybrali 25 studentů z ČR a program zaslali každému z nich. Po provedení testu z matematiky se ukázalo, že průměr ohodnocení daných 25 studentů je $\bar{X} = 540$. Otázka zní: lze nyní říct, že program INTEL zlepšuje výkon v testu, nebo se jen náhodou vybralo 25 studentů s vyšším výkonnostním průměrem v matematice? Jedná se o „skutečný” výsledek (= lze jej zobecnit pro celou populaci?), nebo bylo vyššího průměru dosaženo jen díky náhodným faktorům? Tyto otázky nás přivádějí ke statistickému testu, který rozhodne.

(K1) H_0 : $\mu = 500$ (program intel nemá vliv na zlepšení matematických schopností, tj. střední hodnota bodového ohodnocení testu celé populace studentů i po rozšíření programu všem (celé populaci) zůstane stejná).

H_1 : $\mu > 500$ (jednostranný test - můžeme předpokládat, že program znalosti matematiky nezhoršuje).

(K2) Kritériem volíme právě naměřenou hodnotu \bar{X} .

(K3) Za předpokladu platnosti H_0 má veličina \bar{X} parametry

$$\mu_{\bar{X}} = 500, \quad \sigma_{\bar{X}}^2 = \frac{\sigma}{N} = 400 \implies \sigma_{\bar{X}} = 20.$$

(K4) Stanovená kritická U -hodnota je pro $\alpha = 0.05$ stejná jako u jednostranného testu v předchozí kapitole: $U_{0.95} = 1.64$. Odtud kritická hodnota v rozměru veličiny \bar{X} je

$$T_k = \mu_{\bar{X}} + \sigma_{\bar{X}} \cdot 1.64 = 532.8;$$

(K5) Rozhodnutí testu: pokud příslušná U -hodnota průměru je ≥ 1.64 , zamítáme H_0 na hladině významnosti α . V našem případě $\bar{X} = 540$, tedy $U = \frac{540-500}{20} = 2 > 1.64$. Proto zamítáme H_0 a uzavíráme, že program „skutečně“ zlepšuje matematické schopnosti studentů.

V průběhu testu jsme vlastně počítali podmíněnou pravděpodobnost $P(\bar{X} \geq 540 | H_0 \text{ platí})$ (čti: pravděpodobnost, že \bar{X} nabude hodnoty větší nebo rovny 540, pokud H_0 platí; tomu, co v uvedeném zápisu následuje za svislou čarou, se říká **podmínka**;

podmíněná pravděpodobnost je pak pravděpodobnost události zaznamenané před svislou čarou vypočtená za předpokladu, že platí podmínka). Protože $\alpha = 0.05 = P(\bar{X} \geq 532.8 | H_0 \text{ platí})$, je očividné, že

$$P(\bar{X} \geq 540 | H_0 \text{ platí}) < \alpha;$$

přesněji (viz obr. 13.3)

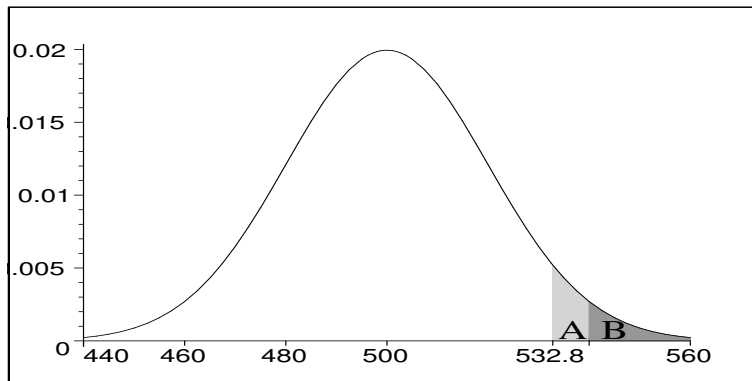
$$\alpha = 0.05 = P(532.8 \leq \bar{X} \leq 540 | H_0 \text{ platí}) + P(\bar{X} \geq 540 | H_0 \text{ platí}) = S(A) + S(B).$$

Protože podmíněná pravděpodobnost $P(\bar{X} \geq 540 | H_0 \text{ platí}) = S(B)$ je menší než naše $\alpha = 0.05 = S(A) + S(B)$, uzavíráme, že něco z našich výchozích předpokladů nebylo správné - to „něco“ je hypotéza H_0 . Samozřejmě, že kromě H_0 jsme měli i další výchozí předpoklady, např. naše data mohla být ovlivněna tím, že

a) Náš vzorek 25 studentů nebyl náhodný (byl z vysoce výběrové školy).

b) Kolega při opisování dat omylem zapsal některá ohodnocení vyšší než ve skutečnosti.

Ale vlivy typu a), b) mohou být vyloučeny správným naplánováním a provedením měření, takže se v podobných případech většinou uzavírá, že nízká pravděpodobnost $P(\bar{X} \geq 540 | H_0 \text{ platí})$ je důsledkem toho, že nesprávný byl předpoklad platnosti H_0 .



Obrázek 13.3: Ad př. 13.4 - hustota rozdělení veličiny \bar{X} za předpokladu, že platí H_0 .

13.3.2 Test „ $\mu_1 = \mu_2$ ”

Příklad 13.5 Vraťme se k situaci z příkladů 13.1 a 13.4. Ředitel firmy KAPPA zjistil, že konkurenční softwarová firma DELTA rovněž vyvinula program pro výuku matematiky (s názvem KILL). Zavolal si proto svého firemního psychologa a požádal ho, aby zjistil, který z obou konkurenčních programů INTEL a KILL je lepší, tj. který více zvyšuje úroveň matematických znalostí.

Psycholog získal kopie obou programů. První z nich předal 32 náhodně vybraným studentům, druhou jiným 32 náhodně vybraným studentům. Po provedení testu z matematiky získal od těchto 64 studentů výsledky jejich ohodnocení a spočetl průměry příslušných hodnot. U programu INTEL $\bar{X}_1 = 600$, u programu KILL $\bar{X}_2 = 533$ (v obou případech velikost vzorku $N = 32$).

Aby zjistil, do jaké míry je jeho měření reprezentativní a zda rozdíl průměrů není pouze náhodný (tj. způsobený např. tím, že program INTEL byl rozdán mezi studenty, kteří byli náhodou chytřejší, ale ne tím, že by INTEL byl lepší než KILL), sáhne ke statistickému testu:

(K1) $H_0: \mu_1 = \mu_2$ (kdyby se oba programy distribuovaly celé populaci, výsledná střední hodnota ohodnocení by byla u obou stejná).

$H_1: \mu_1 \neq \mu_2$ (musíme použít oboustranný test, protože nevíme, který z programů je lepší).

(K2) Hodnotou testového kritéria $\bar{X}_1 - \bar{X}_2$ bude rozdíl $600 - 533 = 67$.

(K3) Za předpokladu platnosti H_0 je rozdělení kritéria $\bar{X}_1 - \bar{X}_2$ normální, vypočteme jeho střední hodnotu a rozptyl:

$$E(\bar{X}_1 - \bar{X}_2) = E\bar{X}_1 - E\bar{X}_2 = \mu_1 - \mu_2 = 0,$$

poslední rovnost platí proto, že předpokládáme platnost H_0 , tj. $\mu_1 = \mu_2$. Při výpočtu rozptylu využíváme předpoklad nezávislosti veličin \bar{X}_1 , \bar{X}_2 , tj. platnosti vztahu

$$E(\bar{X}_1 \cdot \bar{X}_2) = E\bar{X}_1 \cdot E\bar{X}_2$$

$$\begin{aligned} D(\bar{X}_1 - \bar{X}_2) &= E(\bar{X}_1 - \bar{X}_2)^2 - E^2(\bar{X}_1 - \bar{X}_2) = \\ &= E(\bar{X}_1^2 - 2\bar{X}_1 \cdot \bar{X}_2 + \bar{X}_2^2) - (\mu_1 - \mu_2)^2 = \\ &= E\bar{X}_1^2 - 2\mu_1\mu_2 + E\bar{X}_2^2 - \mu_1^2 + 2\mu_1 - 1\mu_2 - \mu_2^2 = \\ &= (E\bar{X}_1^2 - \mu_1^2) + (E\bar{X}_2^2 - \mu_2^2) = D\bar{X}_1 + D\bar{X}_2 = \\ &= \sigma_1^2 + \sigma_2^2 = \frac{10000}{32} + \frac{10000}{32} = 625. \end{aligned}$$

Pokud $\sigma_{\bar{X}_1 - \bar{X}_2}^2 = 625$, tak $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{625} = 25$. Pro náš příklad nutné, aby obě vyšetřované skupiny měly stejný počet studentů - jiný počet studentů v každé skupině by se projevil pouze na tom, že v posledním řádku odvození by v obou jmenovatelích nebylo číslo 32, ale číslo vyjadřující velikost dané skupiny.

(K4) Pro $\alpha = 0.05$ jsou kritické U -hodnoty oboustranného testu stejné jako u oboustranného testu v kapitole 12: $U_m = -1.96$, $U_v = 1.96$.

(K5) Rozhodnutí testu: Pokud příslušná U -hodnota

$$\frac{\bar{X}_1 - \bar{X}_2 - 0}{25}$$

neleží v intervalu $(-1.96; 1.96)$, zamítáme H_0 na hladině významnosti α . V našem případě

$$\frac{\bar{X}_1 - \bar{X}_2 - 0}{25} = \frac{67}{25} = 2.68 \implies \text{zamítáme } H_0$$

o nezávislosti, program INTEL je lepší než program KILL.

Test v příkladu se liší od předchozího testu pouze krokem (K3), kde jsme museli určit rozdělení rozdílu dvou veličin.

13.4 Shrnutí pojmů

Testy uvedené v této kapitole jsou příkladem prvních „praktických“ statistických testů, které jsou užívány. Naměříme hodnotu jedné veličiny u jedné skupiny pozorování, popřípadě u dvou, vypočteme průměr měření v každé ze skupin a tento průměr podrobíme jednostrannému nebo oboustrannému statistickému testu.

Ovšem přitom v těchto testech tiše předpokládáme, že rozptyl σ^2 celé populace je známý.

To ale většinou není pravda a my jej musíme odhadnout (přibližně určit) z naměřených hodnot. Díky větší míře nejasnosti pak kritérium analogického statistického testu, který nepoužívá přímo σ^2 , ale jeho odhad S^2 (viz kapitola 9), nelze popsat normálním rozdělením, ale tzv. ***t-rozdělením*** - příslušný statistický test je v literatuře nazýván ***t-test***. To už je ale obsahem navazujícího kursu v magisterském studiu FEKT.

Závěr

Je poměrně náročným úkolem představit v jednosemestrovém předmětu dva celkem rozsáhlé obory matematiky, z nichž každý by mohl zabrat i třeba celý rok studia. Přesto jsme se o to museli pokusit. Text nemá encyklopedický charakter - mnohé metody a přístupy musely být vypuštěny za cenu toho, aby bylo možné ty základní vybrané vyložit podrobněji a v takovém stylu, že jsou snad pochopitelné i bez další literatury. A i vybrané partie musely být vyloženy v rychlém tempu, bez mnohých důkazů a odvození, text by jinak narostl do neúnosných rozměrů. Jsme přesvědčeni o tom, že věnovat každé z obou částí předmětu menší prostor není možné.

Co se týká matematických předmětů navazujícího magisterského studia FEKT, numerické metody, zejména řešení diferenciálních rovnic, bude prohloubeno v předmětu MODERNÍ NUMERICKÉ METODY. Některé další statistické testy budou probrány v navazujícím předmětu magisterského studia OPERAČNÍ VÝZKUM A EKONOMETRIE.

Ad numerické metody

První kapitola, o chybách, je v podstatě převzata ze skript [2]. Všechny další části jsou zkompileovány z různých zdrojů.

Studenti, kteří by se chtěli seznámit s numerickými metodami podrobněji, včetně některých důkazů, si mohou prohlédnout např. některou z knih [5] nebo [7]. Zvlášť první z nich je však psána spíše pro zkušenější čtenáře.

Ad pravděpodobnost

Kromě zápisků ze svého studia pravděpodobnosti na vysoké škole jsem vycházel zejména z učebnice [4], která je sice určena posluchačům netechnických škol, ale obsahuje srozumitelnou prezentaci pravděpodobnosti a statistiky, díky níž jsem pochopil mnohé. Skriptum [8] má širší záběr a lze v něm najít mnohé vztahy, rozdělení pravděpodobnosti a definice, o kterých v tomto textu není zmínka. Z učebnice [6] operačního výzkumu a optimalizačních metod pochází partie o teorii front v kapitole 11.

Tabulka 12.1: Hodnoty distribuční funkce $\Phi(u)$ - 1.část.

u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$
0.00	0.5000000	0.30	0.6179114	0.60	0.7257469	0.90	0.8159399	1.20	0.8849303
0.01	0.5039894	0.31	0.6217195	0.61	0.7290691	0.91	0.8185887	1.21	0.8868606
0.02	0.5079783	0.32	0.6255158	0.62	0.7323711	0.92	0.8212136	1.22	0.8887676
0.03	0.5119665	0.33	0.6293000	0.63	0.7356527	0.93	0.8238145	1.23	0.8906514
0.04	0.5159534	0.34	0.6330717	0.64	0.7389137	0.94	0.8263912	1.24	0.8925123
0.05	0.5199388	0.35	0.6368307	0.65	0.7421539	0.95	0.8289439	1.25	0.8943502
0.06	0.5239222	0.36	0.6405764	0.66	0.7453731	0.96	0.8314724	1.26	0.8961653
0.07	0.5279032	0.37	0.6443088	0.67	0.7485711	0.97	0.8339768	1.27	0.8979577
0.08	0.5318814	0.38	0.6480273	0.68	0.7517478	0.98	0.8364569	1.28	0.8997274
0.09	0.5358564	0.39	0.6517317	0.69	0.7549029	0.99	0.8389129	1.29	0.9014747
0.10	0.5398278	0.40	0.6554217	0.70	0.7580363	1.00	0.8413447	1.30	0.9031995
0.11	0.5437953	0.41	0.6590970	0.71	0.7611479	1.01	0.8437524	1.31	0.9049021
0.12	0.5477584	0.42	0.6627573	0.72	0.7642375	1.02	0.8461358	1.32	0.9065825
0.13	0.5517168	0.43	0.6664022	0.73	0.7673049	1.03	0.8484950	1.33	0.9082409
0.14	0.5556700	0.44	0.6700314	0.74	0.7703500	1.04	0.8508300	1.34	0.9098773
0.15	0.5596177	0.45	0.6736448	0.75	0.7733726	1.05	0.8531409	1.35	0.9114920
0.16	0.5635595	0.46	0.6772419	0.76	0.7763727	1.06	0.8554277	1.36	0.9130850
0.17	0.5674949	0.47	0.6808225	0.77	0.7793501	1.07	0.8576903	1.37	0.9146565
0.18	0.5714237	0.48	0.6843863	0.78	0.7823046	1.08	0.8599289	1.38	0.9162067
0.19	0.5753454	0.49	0.6879331	0.79	0.7852361	1.09	0.8621434	1.39	0.9177356
0.20	0.5792597	0.50	0.6914625	0.80	0.7881446	1.10	0.8643339	1.40	0.9192433
0.21	0.5831662	0.51	0.6949743	0.81	0.7910299	1.11	0.8665005	1.41	0.9207302
0.22	0.5870604	0.52	0.6984682	0.82	0.7938919	1.12	0.8686431	1.42	0.9221962
0.23	0.5909541	0.53	0.7019440	0.83	0.7967306	1.13	0.8707619	1.43	0.9236415
0.24	0.5948349	0.54	0.7054015	0.84	0.7995458	1.14	0.8728568	1.44	0.9250663
0.25	0.5987063	0.55	0.7088403	0.85	0.8023375	1.15	0.8749281	1.45	0.9264707
0.26	0.6025681	0.56	0.7122603	0.86	0.8051055	1.16	0.8769756	1.46	0.9278550
0.27	0.6064199	0.57	0.7156612	0.87	0.8078498	1.17	0.8789995	1.47	0.9292191
0.28	0.6102612	0.58	0.7190427	0.88	0.8105703	1.18	0.8809999	1.48	0.9305634
0.29	0.6140919	0.59	0.7224047	0.89	0.8132671	1.19	0.8829768	1.49	0.9318879

Tabulka 12.2: Hodnoty distribuční funkce $\Phi(u)$ - 2.část.

u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$
1.50	0.9331928	1.80	0.9640697	2.10	0.9821356	2.40	0.9918025	4.50	0.9999966
1.51	0.9344783	1.81	0.9648521	2.11	0.9825708	2.41	0.9920237	5.00	0.9999997
1.52	0.9357445	1.82	0.9656205	2.12	0.9829970	2.42	0.9922397	5.50	0.9999999
1.53	0.9369916	1.83	0.9663750	2.13	0.9834142	2.43	0.9924506		
1.54	0.9382198	1.84	0.9671159	2.14	0.9838226	2.44	0.9926564		
1.55	0.9394392	1.85	0.9678432	2.15	0.9842224	2.45	0.9928572		
1.56	0.9406201	1.86	0.9685572	2.16	0.9846137	2.46	0.9930531		
1.57	0.9417924	1.87	0.9692581	2.17	0.9849966	2.47	0.9932443		
1.58	0.9429466	1.88	0.9699460	2.18	0.9853713	2.48	0.9934309		
1.59	0.9440826	1.89	0.9706210	2.19	0.9857379	2.49	0.9936128		
1.60	0.9452007	1.90	0.9712834	2.20	0.9860966	2.50	0.9937903		
1.61	0.9463011	1.91	0.9719334	2.21	0.9864474	2.51	0.9939634		
1.62	0.9473839	1.92	0.9725711	2.22	0.9867906	2.52	0.9941323		
1.63	0.9484493	1.93	0.9731966	2.23	0.9871263	2.53	0.9942969		
1.64	0.9494974	1.94	0.9738102	2.24	0.9874545	2.54	0.9944574		
1.65	0.9505285	1.95	0.9744119	2.25	0.9877755	2.55	0.9946139		
1.66	0.9515428	1.96	0.9750021	2.26	0.9880894	2.56	0.9947664		
1.67	0.9525403	1.97	0.9755808	2.27	0.9883962	2.57	0.9949151		
1.68	0.9535213	1.98	0.9761482	2.28	0.9886962	2.58	0.9950600		
1.69	0.9544860	1.99	0.9767045	2.29	0.9889893	2.59	0.9952012		
1.70	0.9554345	2.00	0.9772499	2.30	0.9892759	2.60	0.9953388		
1.71	0.9563671	2.01	0.9777844	2.31	0.9895559	2.70	0.9965330		
1.72	0.9572838	2.02	0.9783083	2.32	0.9898296	2.80	0.9974449		
1.73	0.9581849	2.03	0.9788217	2.33	0.9900969	2.90	0.9981342		
1.74	0.9590705	2.04	0.9793248	2.34	0.9903581	3.00	0.9986501		
1.75	0.9599408	2.05	0.9798178	2.35	0.9906133	3.20	0.9993129		
1.76	0.9607961	2.06	0.9803007	2.36	0.9908625	3.40	0.9996631		
1.77	0.9616364	2.07	0.9807738	2.37	0.9911060	3.60	0.9998409		
1.78	0.9624620	2.08	0.9812372	2.38	0.9913437	3.80	0.9999277		
1.79	0.9632730	2.09	0.9816911	2.39	0.9915758	4.00	0.9999683		

Reference

- [1] Diblík, J., Baštinec, J. : Matematika IV. Skriptum FEI VUT Brno, 1991
- [2] Haluzíková, A. : Numerické metody. Skriptum FEI VUT Brno, 1989
- [3] Horová, I. : Numerické metody. Skriptum PřF MU Brno, 1999
- [4] Loftus, J., Loftus, E.: Essence of Statistics. Second Edition, Alfred A. Knopf, New York 1988.
- [5] Ralston, A. : Základy numerické matematiky. Praha, Academia 1978
- [6] Taha, H.A.: Operations research. An Introduction. Fourth Edition, Macmillan Publishing Company, New York 1989.
- [7] Vitásek, E. : Numerické metody. Praha, SNTL 1987
- [8] Zapletal, J.: Základy počtu pravděpodobnosti a matematické statistiky. Skriptum FEI VUT Brno, PC-DIR 1995.