Renata Mikulaschková

Zaokrouhlovací chyba při numerickém počítání s hlediska statistického

Pokroky matematiky, fyziky a astronomie, Vol. 2 (1957), No. 6, 697--707

Persistent URL: http://dml.cz/dmlcz/137306

mětem jistého bodu $C^+ \in \lambda$ do roviny ϱ. Bodem C proložíme rovinu $\varrho^+ \parallel \varrho$. Tečné roviny k λ podél a, b protínají ϱ^+ v přímkách a^+, b^+. Zřejmě konstrukcí K je užitím bodů A, B, C^+ a přímek a^+, b^+ vytvořen útvar k^+ shodný s útvarem k (vytvořeným konstrukcí K užitím A, B, C, a, b). Avšak jest též $k^+ = \lambda \cap \varrho^+$ (Viz *obr. 8.*)

Věta 2 je tím dokázána.

Definice. *Společný průnik rotační kuželové plochy* $\varkappa$ [7]) *s nevrcholovou rovinou* ϱ, *nikoliv kolmou k ose plochy* $\varkappa$, *nazývá se hyperbolou, resp. parabolou, resp. elipsou, obsahuje-li právě dva nevlastní body, resp. jediný nevlastní bod, resp. neobsahuje-li žádný nevlastní bod.*

Hyperboly, paraboly, elipsy a kružnice označují se společným názvem jako (vlastní a jednoduché) kuželosečky.

Z vět 1, 2 plyne důležitý důsledek, který vyslovíme jako hlavní větu.

Hlavní věta. a) *Centrálním průmětem kuželosečky* (která neleží v rovině promítací) *do vlastní roviny je opět kuželosečka.*

b) *V lineárním zobrazení vlastní roviny* ϱ *na vlastní rovinu* ϱ' *odpovídá kuželosečce opět kuželosečka.*

Literatura

[1] J. Hadamard, *Leçons de géométrie élémentaire II*, Paris 1901 (§ 754—758 na str. 468—473).
[2] K. Havlíček, *Úvod do projektivní geometrie kuželoseček*, Praha 1956 (str. 159—166).
[3] I. M. Jaglom, *Geometričeskije preobrazovanija II*, Moskva 1956 (§ 3 na str. 70—76).
[4] J. Kounovský - Fr. Vyčichlo, *Deskriptivní geometrie*, Praha 1953 (§ 10, 12 na str. 313 až 315).
[5] E. Otto, *Geometria wykreślna*, Warszawa 1954 (§ 26—29 na str. 122—154).
[6] H. Prüfer, *Projektive Geometrie*, Leipzig 1953 (§ 10, str. 185, resp. § 12, str. 188—192).
[7] H. Rademacher - O. Toeplitz, *Von Zahlen und Figuren*, Berlin 1933 (kap. 21 na str. 150 až 160).
[8] K. Rohn - E. Papperitz, *Lehrbuch der darstellenden Geometrie I*, Leipzig 1913 (§ 237—248 na str. 181—191; pozn. 9. na str. 498; § 256—274 na str. 198—215).
[9] Ed. Stiefel, *Lehrbuch der darstellenden Geometrie*, Basel 1947 (str. 67, resp. str. 87—89, resp. str. 95).

ZAOKROUHLOVACÍ CHYBA PŘI NUMERICKÉM POČÍTÁNÍ S HLEDISKA STATISTICKÉHO

RENATA MIKULASCHKOVÁ

Velmi důležitým problémem při numerickém počítání je odhad chyby, které se dopustíme při numerickém řešení problému. Zatím co různé otázky jiné, na př. otázky konvergenční, jsou podrobně studovány, odhad chyby je otázkou poměrně málo propracovanou. To proto, že odhad, má-li být užitečný, musí být snadno proveditelný a musí dávat výsledky, které se příliš neliší od skutečnosti. Zatím se však ještě velmi často stává, je-li odhad vůbec proveditelný, že odhadnutá chyba je mnohem větší než chyba skutečná.

[7]) Poněvadž vyšetřujeme rozšířený euklidovský prostor, je rotační válcová plocha pouze speciálním případem rotační plochy kuželové.

Druhou stránkou, která prakticky znehodnocuje theoretický odhad chyby, je ta okolnost, že často počítáme jiným způsobem, než pro který byl proveden odhad. Tak na příklad při numerickém počítání byl odhad proveden za předpokladu, že počítáme přesně (to jest na nekonečný počet míst), zatím co prakticky počítáme na konečný počet míst. Jiným případem je, že předpokládáme zaokrouhlování, zatím co při nevhodném počítání na stroji potlačujeme poslední místa směrem dolů (to jest 9 zaokrouhlujeme na 0 a pod.).

Při každé numerické methodě se vyskytují prakticky dva druhy chyb. Jeden je chyba methody, kdy na příklad místo nekonečně mnoha postupných aprosimací bereme jen konečný počet, a druhý je chyba zaokrouhlovací, to jest chyba vzniklá tím, že počítáme na konečný počet desetinných míst. V tomto článku se budeme zabývat základní problematikou zaokrouhlovací chyby s hlediska statistického a v jednodušších případech formulovat konkretní závěry.

Každá základní učebnice numerických method se zabývá chybou zaokrouhlovací. Ve všech těchto monografiích se vychází z maximalistického pojetí, to jest, že odhadujeme největší chybu, které se theoreticky můžeme dopustit. Při tom předpokládáme vždy nejnepříznivější případ, to jest, že chyba zaokrouhlovací je stále 0,5 posledního místa a je téhož znaménka. Ve skutečnosti ve většině případů jsou chyby různého znaménka, vzájemně se ruší a jejich velikost není vždy maximální. Proto je celkem pochopitelné, že skutečná chyba je podstatně menší než chyba theoretická (odhadnutá). Je tedy zřejmé, že maximalistický odhad zvláště při větším počtu operací nemůže dát uspokojivé výsledky, to jest takové výsledky, které by byly ve shodě se skutečností.

Aby byla zřejmější základní myšlenka dalšího postupu, vyjdeme z experimentu a na základě dosažených výsledků budeme formulovat přesně předpoklady a závěry.

Počítejme z tabulek zaokrouhlovací chyby, jichž se dopustíme při zaokrouhlování čísel $e^{k_j + i\, 0{,}001}$ pro $i = 0, 1, \ldots, 24$, $j = 1, 2, 3$, $k_1 = 1,00$, $k_2 = 1,20$ a $k_3 = 1{,}60$ na 5 desetinných míst.

V prvých třech sloupcích *tab. 1* jsou tyto chyby násobené 10^6, ve sloupci čtvrtém je součet zaokrouhlovacích chyb.

Učiníme nyní několik poznámek k zaokrouhlování. Čísla 0, 1, 2, 3, 4 zaokrouhlujeme na 0 a čísla 5, 6, 7, 8, 9 na 10. Ve většině učebnic se doporučuje zaokrouhlovat číslo 5 na 0 když předchází číslo sudé, a na 10, když předchází číslo liché (na př. 2,5 zaokrouhluje se na 2,0 a 3,5 na 4,0). Ve většině případů však je třeba zaokrouhlovat číslo 5 nahoru, to jest na 10. To proto, že ve smyslu statistické theorie, které dále budeme užívat, je přibližně číslo rovné přesnému + nahodilé číslo z intervalu $\mp$ 0,5 posledního místa. Číslo 5 a za ním nekonečně mnoho nul nastává při zaokrouhlování s nulovou pravděpodobností a můžeme jeho vliv proto zanedbat. Většinou dostaneme číslo 5 a některé nenulové číslo za tím a toto číslo zaokrouhlujeme zřejmě ve shodě s theorií nahoru. Výjimkou, kdy zaokrouhlujeme číslo 5 podle druhého způsobu (uvedeného v učebnicích), je případ, kdy zaokrouhlované číslo je přesně poslední číslo, na př. při dělení 2 a počítání na celá čísla.

Vraťme se nyní k naší *tab. 1.* Zaokrouhlovací chybu v 1. sloupci lze vyjádřit zřejmě ve formě funkce argumentu i. 0,001, $i = 0, 1, 2, \ldots 24$. Vzhledem k tomu, že zřejmě posloupnost 0, 1, 2, ... 24 není posloupnost náhodných čísel, nejsou také ani zaokrouhlovací chyby nějakými náhodnými čísly. Podíváme-li se na sloupec 1, aniž bychom si tuto okolnost uvědomili, připadají nám čísla úplně náhodná. Vzniká zde nyní otázka, zda bychom nemohli dojít ke kvalitativně i kvantitativně správným závěrům, kdybychom tato čísla považovali za výběrové hodnoty náhodné veličiny. Je totiž zcela možné, že vlastnosti uvedené posloupnosti v sloupci 1 jsou stejné jako vlastnosti náhodného výběru, které jsou rozhodující pro některé závěry. Tento postup se často užívá i jinde, kde jisté pojmy nebo jevy vyjadřujeme analogicky jinými pojmy nebo jevy (t. zv. analogiemi). Tyto srovnávané jevy spolu nesouvisí, avšak jejich vnitřní stavba je stejná vzhledem k závěrům, které hledáme. Na př. v pružnosti lze uvažovat hydrodynamické analogie, v hydrodynamice elektrické analogie atd.

Tabulka 1.

ε_1	ε_2	ε_3	$\varepsilon_1 + \varepsilon_2 + \varepsilon_3$
+1,829	—3,077	+2,424	+1,176
+1,470	—1,300	—2,066	—1,896
+3,832	+3,801	—1,598	+6,035
—1,081	+2,229	+3,833	+4,981
—3,269	+3,987	+4,231	+4,949
—2,727	—0,922	—0,397	—4,046
+0,547	—2,494	—0,048	—1,995
—3,445	—0,726	—4,716	—8,887
—4,699	+4,386	—4,396	—4,709
—3,214	+2,844	+0,917	+0,547
+1,015	+4,653	+1,228	+6,896
—2,111	—0,185	—3,459	—5,755
—2,288	—1,667	—3,137	—7,092
+0,186	+0,212	+2,197	+2,595
—4,587	—4,545	+2,549	—6,583
+3,397	+4,064	—2,076	+5,385
+4,141	—3,956	—1,674	—1,489
—2,353	+1,398	+3,762	+2,807
+3,917	+0,128	+4,235	+8,280
+2,955	+2,239	—0,248	+4,946
+4,764	—2,266	+0,317	+2,815
+4,764	—2,266	+0,317	+2,815
—0,654	—3,384	+4,066	+0,028
—3,269	—1,112	—3,390	—7,771
—3,160	+4,554	+2,349	+3,743
—0,242	+3,618	+3,157	+6,533

	ε_1	ε_2	ε	$\varepsilon_1 + \varepsilon_2 + \varepsilon_3$
Střed. hod. chyby	—0,362	+0,499	+0,322	+0,460
Theor. střed. hod.	0	0	0	0
Interval	—5, +5	—5, +5	—5, +5	—15, +15
Rozptyl. skut.	8,589	8,445	8,111	25,815
Theor. rozptyl	8,33	8,33	8,33	25,00

Přesvědčíme-li se, že množina čísel v 1. sloupci má svoji vnitřní zákonitost, která je rozhodující pro naše závěry při početních operacích se zaokrouhlenými čísly, a která je stejná jako zákonitost statistického souboru, můžeme na zaokrouhlovací chyby s úspěchem aplikovat theorii pravděpodobnosti k charakterisaci (kvantitativní i kvalitativní) zaokrouhlovacích chyb i operací s nimi.

Je zřejmé, že ne každá zaokrouhlovací chyba musí mít charakter umožňující aplikaci statistických method při jejich studiu. Tak na př. v *tab. 2* jsou uvedeny zaokrouhlovací chyby těchže čísel jako v *tab. 1*, avšak pro případ, že zaokrouhlujeme na jedno desetinné místo. *Tab. 2*:

Tabulka 2.

Zaokrouhlovací chyby násobené 10^2 při zaokrouhlování na 1 desetinné místo

+1,82	+4,28	—3,52
+2,10	+4,56	—3,24
+2,37	+4,83	—2,96
+2,64	—4,90	—2,69
+2,91	—4,62	—2,41
+3,19	—4,34	—2,13
+3,46	—4,07	—1,85
+3,73	—3,79	—1,57
+4,01		

Vidíme zde zřejmě na první pohled, že tuto množinu čísel nemůžeme považovat za náhodný výběr. Předtím tedy, než je možno užít v otázce zaokrouhlovacích chyb statistické theorie, nutno se přesvědčit o základních vlastnostech těchto chyb, které jsou rozhodující pro případnou možnou aplikaci.

Tato otázka má dosti značný význam, což je vidět již z toho, že se můžeme přesvědčit, že 1. sloupec v *tab. 1* má žádané vlastnosti nutné pro aplikaci statistické theorie, zatím co v *tab. 2* tyto vlastnosti zřejmě splněny nejsou. Je proto pochopitelné, že na příklad v našem případě můžeme s úspěchem aplikovat statistickou theorii až při zaokrouhlování na $p \geqq p_0$ míst, kde p_0 je zřejmě větší než 1 a, jak jsme již řekli, menší nebo rovno 5.

S praktického hlediska ovšem může být případné testování tak pracné, že by prakticky znehodnotilo odhad (ve smyslu jak bylo výše řečeno). Ve většině případů však již při menší zkušenosti lze s dostatečnou přesností rozhodnout o možné aplikabilitě.

Zavedeme nyní některé pojmy a definice, které budeme v dalším užívat:

Definice 1.

Budeme značit $A, B, C, \ldots$ čísla a $a, b, c, \ldots$ jejich aproximace. Absolutní chybou ε_a, ε_b, ε_c budeme nazývat rozdíl mezi přesnou hodnotou A, B, C a jejich aproximací a, b, c. Tedy

$$\varepsilon_a = A - a\,, \quad \varepsilon_b = B - b\,, \quad \varepsilon_c = C - c.$$

Definice 2.

Relativní chybou δ_a, δ_b, $\delta_c \ldots$, budeme značit hodnoty

$$\delta_a = \frac{\varepsilon_a}{a}\,, \quad \delta_b = \frac{\varepsilon_b}{b}\,, \quad \delta_c = \frac{\varepsilon_c}{c}\,, \ldots$$

Definice 3.

Budeme říkat, že a je zaokrouhlení čísla A na p desetinných míst, když a má nejvýše p desetinných míst a $|\varepsilon_a| \leqq 0{,}5 \cdot 10^{-p}$, ε_a je zaokrouhlovací chyba.

Přicházíme nyní k základnímu principu, na němž je založeno statistické pojetí zaokrouhlovacích chyb.

Předpoklad 1.

Zaokrouhlovací chyba při zaokrouhlování na p desetinných míst je náhodná veličina s rovnoměrným rozdělením, to jest s hustotou

$$f(x) = \frac{1}{10^{-p}} \quad \text{pro } |x| \leqq 0{,}5 \cdot 10^{-p},$$

$$f(x) = 0 \quad \text{pro } |x| > 0{,}5 \cdot 10^{-p},$$

a zaokrouhlovací chyby jsou nezávislé náhodné proměnné.

To znamená, provádím-li zaokrouhlení n čísel na příklad jak bylo uvedeno v *tab. 1*, potom přibližně v $n \int_a^b f(x)\,dx$ případech bude chyba ε v mezích $\langle a, b \rangle$. Tento předpoklad umožňuje zavedení další definice:

Definice 4.

Budeme říkat, že počítáme na p desetinných míst, nebo že jde o zaokrouhlovací chybu $0{,}5 \cdot 10^{-p}$, s pravděpodobností q, jestliže platí

$$P\{|\varepsilon| < 0{,}5 \cdot 10^{-p}\} \geq q,$$

to zn. pravděpodobnost, že absolutní hodnota chyby ε bude menší než $0{,}5 \cdot 10^{-p}$, musí být větší nebo rovna q.

Jelikož jde o náhodné proměnné, to jest o pojmy z theorie pravděpodobnosti, uveďme zde některé věty z této theorie.

Věta 1.

Střední hodnota součtu náhodných proměnných je rovna součtu jednotlivých středních hodnot:

$$E[\xi_1 + \xi_2] = E[\xi_1] + E[\xi_2].$$

Věta 2.

Jsou-li náhodné proměnné nezávislé, rovná se rozptyl jejich součtu součtu jednotlivých rozptylů:

$$D[\xi_1 + \xi_2] = D[\xi_1] + D[\xi_2].$$

Věta 3.

Náhodná proměnná $\xi_1 = a\,\zeta_1$, kde a je reálné číslo, má tyto vlastnosti:

$$E[\xi_1] = aE[\zeta_1], D[\xi_1] = a^2 D[\zeta_1].$$

Věta 4.

Součet nezávislých náhodných proměnných s normálním rozdělením je opět náhodná proměnná s normálním rozdělením.

Věta 5.

Jestliže nezávislé náhodné proměnné $\xi_1, \xi_2, \ldots, \xi_n$ mají stejnou distribuční funkci a konečný od nuly různý rozptyl (dispersi), potom rozdělení součtu nezávislých náhodných proměnných konverguje pro $n \to \infty$ k normálnímu rozdělení:

$$P\left\{\frac{1}{\sqrt{B_n}} \sum_{k=1}^{n} (\xi_k - E[\xi_k]) < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz.$$

(B_n je rozptyl součtu, to jest $B_n = n\,B$, kde B je rozptyl veličiny ξ_i).

V dalším se budeme zabývat sčítáním nezávislých náhodných proměnných. V případě, že tyto veličiny budou mít normální rozdělení, bude mít jejich součet podle věty 4 také normální rozdělení. Proto s jistou chybou, kterou budeme dále analysovat, nahradíme předpoklad 1 předpokladem 1'.

Předpoklad 1'.

Zaokrouhlovací chyba při zaokrouhlování na p desetinných míst je náhodná proměnná s normálním rozdělením hustoty pravděpodobnosti, se střední hodnotou O a dispersí $D =$

$= \frac{1}{3} a^2$, *kde* $a = 0{,}5 \cdot 10$. *Zaokrouhlovací chyby jsou nezávislé náhodné proměnné.*

Pochopitelně záměnou předpokladu 1 předpokladem 1′ dopustíme se jisté chyby. Tato chyba, jak uvidíme dále, je však již při součtu tří veličin zanedbatelná.

Abychom mohli zkoumat tuto chybu, studujeme theoretické rozdělení součtu veličin s rovnoměrným rozdělením. Je známo, že hustota součtu nezávislých náhodných proměnných ξ_1, ξ_2 s hustotou $f_1(x), f_2(x)$ je

$$f(x) = \int_{-\infty}^{\infty} f_1(x - t) f_2(x)\,\mathrm{d}t.$$

Užijeme-li tohoto vzorce pro náš případ, dostaneme:

$$f_1(y) = \frac{1}{2a} \text{ v intervalu } <-a, a>,$$

$$f_1(y) = 0 \quad \text{vně intervalu } <-a, a>.$$

$$f_2(y) = \begin{cases} \dfrac{1}{2a}\left[\dfrac{y+2a}{2a}\right] & \text{v intervalu } <-2a,\ 0> \\[2ex] \dfrac{1}{2a}\left[\dfrac{y+2a}{2a} - 2\left(\dfrac{y+2a}{2a} - 1\right)\right] & \text{v intervalu } <0,\ 2a> \end{cases}$$

$$f_2(y) = 0 \qquad \text{vně intervalu } <-2a, 2a>,$$

$$f_3(y) = \begin{cases} \dfrac{1}{2a}\left[\dfrac{1}{2}\left(\dfrac{y+3a}{2a}\right)^2\right] & \text{v intervalu } <-3a, -a> \\[2ex] \dfrac{1}{2a}\left[\dfrac{1}{2}\left\{\left(\dfrac{y+3a}{2a}\right)^2 - 3\left(\dfrac{y+3a}{2a} - 1\right)^2\right\}\right] & \text{v intervalu } <-a, a>, \\[2ex] \dfrac{1}{2a}\left[\dfrac{1}{2}\left\{\left(\dfrac{y+3a}{2a}\right)^2 - 3\left(\dfrac{y+3a}{2a} - 1\right)^2 + 3\left(\dfrac{y+3a}{2a} - 2\right)^2\right\}\right] & \text{v int.} <a, 3a> \end{cases}$$

$$f_3(y) = 0 \qquad \text{vně intervalu } <-3a, 3a>,$$

a obecně lze napsat: vzorec pro součet n nezávislých náhodných proměnných:

$$fn(y) = \frac{1}{2a}\left\{\frac{1}{(n-1)!}\left[\left(\frac{y+na}{2a}\right)^{n-1} - \binom{n}{1}\left(\frac{y+(n-2)a}{2a}\right)^{n-1} + \binom{n}{2}\left(\frac{y+(n-4)a}{2a}\right)^{n-1} \dots\right]\right\}$$

kde y je v intervalu $<-na, na>$ a součet se provádí tak dlouho, pokud argumenty $\left(\frac{y+na}{2a}\right), \left(\frac{y+(n-2)a}{2a}\right), \dots$ jsou kladné.

Ukažme ještě na konkretním případě význam záměny předpokladu 1 předpokladem 1′. Při zaokrouhlování na 3 desetinná místa se dospustíme s pravděpodobností 0,95 při součtu $n = 1$ až 4 čísel chyby v mezích:

Tabulka 3.

n	Rovnoměrné r.	Normální r.	Maximální chyby
1	$0{,}48 \cdot 10^{-3}$	$0{,}54 \cdot 10^{-3}$	$0{,}5 \cdot 10^{-3}$
2	$0{,}77 \cdot 10^{-3}$	$0{,}80 \cdot 10^{-3}$	$1{,}0 \cdot 10^{-3}$
3	$0{,}97 \cdot 10^{-3}$	$0{,}98 \cdot 10^{-3}$	$1{,}5 \cdot 10^{-3}$
4	$1{,}119 \cdot 10^{-3}$	$1{,}127 \cdot 10^{-3}$	$2{,}0 \cdot 10^{-3}$

Z uvedené tabulky je patrno, že pro součet již dvou veličin je chyba asi 4 % vzhledem k praxi numerického počítání úplně zanedbatelná. Nahrazení předpokladu 1 předpokladem 1′ můžeme považovat prakticky za plně oprávněné od součtu 2 nezávislých náhodných veličin.

Užitím vět 1, 2, 3, 4 a předpokladu 1′ můžeme již snadno analysovat chybu. Budeme nyní aplikovat získané výsledky z theorie počtu pravděpodobnosti na základní aritmetické úkony. Zabývejme se nejprve sčítáním.

Zaokrouhlovací chyby při sčítání

Sčítáme-li n zaokrouhlených čísel, potom chyba součtu bude zřejmě rovna součtu zaokrouhlovacích chyb. Poněvadž zaokrouhlovací chyby jsou podle předpokladu 1′ nezávislé náhodné proměnné, bude chyba součtu náhodná proměnná, vzniklá jako součet nezávislých náhodných proměnných s normálním rozdělením, mít také normální rozdělení.

Uvedeme příklad.

1. Na kolik desetinných míst je třeba počítat hodnoty b_i (pro všechna b_i na stejný počet desetinných míst), jestliže s pravděpodobností q potřebujeme hodnotu A na p desetinných míst

$$A = \sum_{i=1}^{N} b_i\, 2^i.$$

Číslo A potřebujeme s pravděpodobností q na p_A míst přesně. Chyba A je náhodná proměnná s rozptylem D_A a střední hodnotou 0. Abychom měli zaručenu žádanou přesnost, to jest aby s pravděpodobností větší nebo rovnou q byla chyba menší než $0{,}5 10^{-p}$, musí být

$$D_A = D_0,$$

(D_0 značí daný rozptyl, který je dán žádanou přesností). Tuto podmínku snadno určíme z tabulek normálního rozdělení. Na př. pro $q = 0{,}95$ je $D_0 = \dfrac{(0{,}5 \cdot 10^{-p})^2}{1{,}96}$.

Z rovnice (1) a vět 1, 2, 3, 4 plyne

$$D_A = D_{b_i} \sum_{j=1}^{N} 2^j = D_{b_i} \frac{2^{2N}-1}{(4-1)} \cdot 2^2 = D_{b_i}\, \alpha,$$

kde

$$D_{b_i} = \frac{1}{3} a_{b_i}^2 \quad \text{a} \quad \alpha = \frac{2^{2N}-1}{(4-1)},$$

a proto

$$D_{b_i} \leq \frac{D_0}{\alpha},$$

a tedy

$$a_{b_i} \leq \frac{\sqrt{3}\,\sqrt{D_0}}{\sqrt{\alpha}}.$$

Potřebujeme tedy počítat na $p_{b_i} = \left\{-\log \frac{2\sqrt{3}\,\sqrt{D_0}}{\sqrt{\alpha}}\right\}$ míst, kde závorka značí nejmenší přirozené číslo větší nebo rovné p_{b_i}.

Jak uvidíme dále, prakticky nejvýhodnější by byl postup takový, že bychom počítali hodnoty b_i na nestejný počet desetinných míst tak, aby $b_i 2^i$ měly stejný počet míst pro všechna i. Uvedený příklad slouží jako ilustrace postupu (viz str. 000).

Přejděme nyní k druhé aritmetické operaci, k násobení.

Zaokrouhlovací chyby při násobení

Při násobení nelze theoreticky úplně přesně postupovat. Musíme zde provést další zjednodušení, a to zanedbat čtverec zaokrouhlovací chyby jako veličiny vyššího řádu. Je možno nyní postupovat dvěma způsoby: pomocí absolutních chyb a pomocí relativních chyb.

1. *Postup pomocí absolutních chyb*:

Zde vycházíme ze známého vzorce

$$\prod_{i=1}^{N} [b_i + \varepsilon_i]^{p(i)} = \prod_{i=1}^{N} b_i p^{(i)} + \sum_{i=1}^{N} \varepsilon_i b_1^{p(1)} b_2^{p(2)} \ldots b_{i-1}^{p(i-1)} b_{i+1}^{p(i+1)} \ldots b_N b_i^{p(1)-1} \cdot p(i),$$

který je platný po zanedbání veličin vyšších řádů. Tím je převeden problém násobení na problém sčítání, který již dovedeme řešit. Vzorec platí pochopitelně pouze při možnosti zanedbat veličiny vyšších řádů. Tento předpoklad je třeba alespoň přibližně verifikovat. Při praktickém počítání je verifikace snadná, neboť relativní chyby jednotlivých činitelů jsou malé. Z toho pak plyne možnost zanedbání členů vyššího řádu.

2. *Postup užitím relativních chyb*:

Zde vycházíme ze vzorce:

$$\lg A\,(1 + \delta) \doteq \lg A + \delta,$$

a tedy

$$\lg \prod_{i=1}^{N} [b_i\,(1 + \delta_i)]^{p(i)} = \sum_{i=1}^{N} \lg b_i^{p(i)} + \sum_{i=1}^{N} p\,(i)\,\delta_i.$$

Je tedy relativní chyba součinu $\prod_{i=1}^{N} [b_i\,(1 + \delta_i)]^{p(i)}$ rovna $\sum_{i=1}^{N} p\,(i)\,\delta_i$. Tím je opět převeden problém součinu na problém součtu, který již dovedeme řešit. V souvislosti se sčítáním dovedeme proto řešit i příklady kombinované.

Uveďme opět příklad:

Na kolik desetinných míst je třeba počítat jednotlivé činitele (všechny mají stejný počet míst), jestliže s pravděpodobností q potřebujeme určit hodnotu A na p desetinných míst.

$$A = \sum_{i=1}^{N} b_i \sqrt{i}.$$

Podle toho, co bylo řečeno výše je:

$$\varepsilon_A = \sum_{i=1}^{N} (\varepsilon_{b_i} + \varepsilon_{\sqrt{i}}\, b_i).$$

Mají-li absolutní chyby stejný rozptyl, dojdeme k závěru

$$D_{\varepsilon_A} = D_\varepsilon \left(\sum_{i=1}^{N} i + \sum_{i=1}^{N} b_i^2\right).$$

Z této rovnice již snadno určíme ε a počet desetinných míst. Poněvadž jsme užili vzorec pro násobení s absolutními chybami, musíme verifikovat předpoklad o zanedbání veličin vyšších řádů, na příklad tím, že požadujeme také, aby relativní chyby jednotlivých b_i a $\sqrt{i}$ nebyly příliš veliké.

Podobným způsobem můžeme užít i relativních chyb, takže počítáme pomocí relativní chyby nejprve absolutní chyby každého sčítance a potom odhadujeme chybu součtu.

Jak bylo výše ukázáno, je hlavním principem převod funkcionálních závislostí na závislosti lineární. Toho můžeme užít i při příkladech složitějších, na př. při vypočítávání vzorců s tabelovanými funkcemi sh$x$ a p.

Zabývejme se nyní ještě některými otázkami praktického postupu a naznačme myšlenky, na jejichž základě můžeme tyto otázky studovat.

Položme si tento problém:

Na kolik desetinných míst je třeba počítat čísla a i b, aby $(a + b)$ bylo s pravděpodobností q vypočteno na p desetinných míst přesně. Z předcházejícího příkladu vychází tato podmínka pro určení přesnosti čísel a i b:

$$D_a + D_b = D \quad (D \text{ je daný rozptyl}).$$

Tuto podmínku můžeme splnit řadou řešení. Je třeba určit poměr $\dfrac{D_a}{D_b}$. Tento poměr se může pohybovat v rozmezí od 0 do ∞. S hlediska praktického je rozhodující pracnost. Definujeme ji v nejjednodušším případě tak, že pracnost je charakterisována počtem užitých cifer. Potom z podmínky minima pracnosti vychází extrém s podmínkovou rovnicí:

$$10^{-2p_a} + 10^{-2p_b} = C,$$

$$(\alpha_a + p_a) + (\alpha_b + p_b) = \text{minimum},$$

(α je počet cifer před desetinnou čárkou a $(\alpha_a + p_a)$ je počet cifer celého čísla). Snadným výpočtem dostaneme, že $p_a = p_b$. Docházíme zde k známému závěru, který se používá ve všech učebnicích, a to, že při sčítání je třeba brát všechny sčítance na stejný počet míst. Tento závěr platí za předpokladu, že pracnost je úměrná počtu cifer.

Podobným způsobem, ať již použijeme absolutních nebo relativních chyb, docházíme k závěru, že při násobení je třeba brát všechny činitele se stejnou relativní chybou, ovšem opět za předpokladu, že pracnost je úměrná počtu cifer.

Tato okolnost nám pomůže při praktickém výpočtu a zřejmě preferuje výpočet chyb součinů podle relativních chyb.

S tohoto hlediska pochopitelně je třeba v příkladě 1 počítat výraz

$$A = \sum_{i=1}^{N} b_i\, 2^i,$$

jiným způsobem, než jsme uvedli. Plyne totiž z našich podmínek pracnosti, aby bylo $D_{b_i} 2^{2i} = D_b$, kde D_b již nezávisí na i; je tedy

$$D_{b_i} = D_b \frac{1}{2^{2i}}$$

a dále

$$D_{b_i} \leq \frac{D_0}{N\, 2^{2i}};$$

z toho snadno vypočteme

$$a_{b_i} = \frac{\sqrt{3 D_0}}{\sqrt{N}\, 2^i}.$$

Přejděme nyní k znázornění účinnosti statistického pojetí zvláště ve srovnání s maximalistickou theorií.

Ukažme na příkladě, který již byl počítán, shodu našich závěrů se skutečností. Vraťme te zpět a podívejme se opět na *tab. 1.* Jsou zde vyčísleny střední hodnoty a rozptyl skusečný a rovněž i theoretické hodnoty. Zároveň jsou zde uvedeny intervaly zaokrouhlovacích chyb. Pomocí těchto hodnot byla počítána k dané maximální chybě theoretická (na základě theoretického rozptylu a normálního rozložení) pravděpodobnost a relativní četnost. Hodnoty byly sestaveny do *tab. 4.*

Tabulka 4.

Pravděpodobnost	Theor. max. chyba	Relativní četnost
0,95	9,79	1,00
0,90	8,24	0,92
0,80	6,41	0,72

Vidíme, srovnáme-li theoretickou pravděpodobnost a relativní četnost, že je shoda velmi dobrá, uvážíme-li ještě, že jde o výběr rozsahu 25 a že jeden člen znamená četnost 0.04.

Methoda statistického chápání chyb byla užita i na problematiku složitější. Tak na př. Rademacher [1] aplikoval tuto theorii v případě integrování obyčejné diferenciální rovnice, A. A. Abramov [2] a L. A. Ljusternik [3] studovali zaokrouhlovací chybu při methodě sítí. Huskey [4] naopak ukázal na nutnou opatrnost při aplikaci statistické methody, neboť předpoklad o rovnoměrné hustotě může také vést k podhodnocení chyby. Statistická theorie byla také aplikována v souvislosti s velkými samočinnými počítači, kde zaokrouhlování se realisovalo skutečně jako náhodné veličiny (Forsythe) [5].

V článku omezili jsme se na nejjednodušší případy aplikace statistických method při numerickém počítání. Problematika s tím souvisící je podstatně širší. Cílem tohoto článku však bylo ukázat postup a hlavní myšlenky na nejjednodušších případech.

Literatura

[1] H. A. Rademacher, *On the accumulation of errors in processes of integration*, Ann. of Comp. Lab. of Harward Univ. Nat. Bur. Stand. Los Angeles.

[2] A. A. Abramov, *O vlijanii ošibok okruglenija pri rešenii uravněnija Laplacea*, Vyčislitělnaja matem. i techn., 1 (1953), 37—40.

[3] L. A. Ljusternik, *O schodimosti pri slučajnych načalnych,dannych i nakoplenii ošibok iteracionnogo processa rešenija sistěmi algebraičeskich uravněnij*, tamtéž, 41—45.

[4] H. D. Huskey, *On the procesion of a certain procedure of numerical integration, with an appendix by Douglas R., Hartree J.*, Res. Nat. Bur. Stand 42, 57—62 (1949).

[5] G. E. Forsythe, *Note on rounding-off errors*, Nat. Bur. Stand at Los Angeles. Prepublication copy (1950).

[6] B. V. Gněděnko, *Kurs těorii verojatnostěj*.

[7] H. Cramér, *Mathematical Methods of Statistics*.

O VYJADŘOVÁNÍ INTENSITY OSVĚTLENÍ

DANA MATOUŠOVÁ

Pro lepší prostorovou představu o zobrazovaných tělesech používáme často v technických aplikacích osvětlení, a to obyčejně osvětlení rovnoběžného. Ještě lepší představu o zobrazovaném tělese docílíme, když na zobrazované těleso narýsujeme čáry, spojující body o stejné intensitě osvětlení, buď skutečné, isofoty, nebo zdánlivé, isofengy. Jak se tyto křivky sestrojují je uvedené v některých učebnicích deskriptivní geometrie. Plošky, které vzniknou po sestrojení intensitních čar mezi sousedními isofotami či isofengami, pokládají se neutrální šedí. Pokládá se vždy více a více poloh od části nejvíce osvětlené až k mezi vlastního stínu. Za mezí vlastního stínu předpokládá se osvětlení paprsky odraženými od okolních předmětů, takže množství poloh od meze vlastního stínu zase ubývá. Množství poloh se řídí podle intensitní stupnice. Existuje několik takových stupnic, na př. Tilšerova, Wienerova, lze však odvoditi tuto stupnici i exaktně.

Tento způsob vyjádření intensity polohováním má však značné nevýhody. Výsledek je závislý na volbě šedé barvy. Volíme-li základní barvu na polohování příliš světlou, má to za následek, že plošky ležící mezi dvěma isofotami poblíže meze vlastního stínu nutno polohovat 20—30 ×, což má za potom následek rozmývání předcházejících poloh.

Bylo by tedy výhodnější použít pro vyjádření intensity osvětlení způsobu, který by byl sice exaktní, ale neměl při tom nevýhody nahoře uvedené. Tyto požadavky splňuje způsob vyjádření intensity osvětlení šrafováním. Ze zkušenosti víme, že ploška vyšrafovaná čarami tloušťky b ve vzájemné čisté vzdálenosti a činí z určité vzdálenosti dojem šedé. Chceme-li tímto způsobem vyjádřit intensitu osvětlení, musíme tento dojem přesně definovat.

Na vyšrafovanou plošku dopadají světelné paprsky. Paprsky dopadlé na pruh šířky a se odrážejí, paprsky dopadlé na pruh černé šrafy jsou pohlceny. Množství paprsků,