خطای نمایش اعداد

اگر aگرد شدهٔ عدد A تا n رقم اعشار باشد، باتوجه به نحوهٔ گرد کردن داریم:

$$|A - a| \leq ۵ \times ۱۰^{-(n+۱)}$$

نامساوی بالا نشان می‌دهد که هرچه n بزرگتر باشد a به A نزدیکتر خواهد بود.

خطای مطلق یک تقریب، دقت تقریب را مشخص نمی کند!

$A=1250, \quad a=1250.5 \longrightarrow e(a)=0.5$

$B=1, \quad b=1.5 \longrightarrow e(b)=0.5$

خطاهای مطلق b و a یکسان است. آیا دقت b و a نیز یکسان است؟

آنچه خطا را نشان می دهد، سنجش خطا در واحد کمیت است.

ارقام بامعنی درست یک تقریب

بدیهی است که تعداد ارقام مشترک یک تقریب مؤید دقت آن تقریب نیست.

فرض کنید $a'= ۸/۰۸ \quad , \quad a= ۷/۹۹۷ \quad , \quad A= ۸/۰۰۰$

مشاهده می‌شود که a' درست دو رقم مساوی با ارقام A دارد. اما، هیچیک از ارقام a مساوی ارقام A نیست.

آیا می‌توان گفت که ارقام درست a' بیشتر از ارقام درست a است؟ خواهیم دید که نه.

مفهوم ارقام با معنای درست هر تقریب رابطهٔ تنگاتنگ با دقت آن تقریب دارد. در اینجا

$e(a')= ۰/۰۸ \quad$ و $\quad e(a)= ۰/۰۰۳$

و در واقع a باید تعداد ارقام درست بیشتری داشته باشد! اما، تعداد ارقام با معنای درست چگونه به‌دست می‌آید؟

اگر a را تا سه رقم با معنا گرد کنید عدد A حاصل می‌شود. یعنی، a سه رقم با معنای درست دارد.

$$a= ۷/۹۹۷ \longrightarrow ۸/۰۰۰$$

اگر a' را تا یک رقم اعشار گرد کنید به ۸/۱ منجر می‌شود که مساوی A نیست یعنی، a' تنها یک رقم با معنای درست دارد (هرچند که دو رقم آن دقیقاً در بسط A ملاحظه می‌شود).

$$a'= ۸/۰۸ \longrightarrow ۸/۱$$

قضیه

اگر $a>0$ تقریبی از **عدد** A باشد به‌طوری که:

$$\delta(a) \le \frac{1}{2} \times 10^{-n}$$

آنگاه a حداقل n رقم با معنای درست دارد.

با توجه به این‌که خطای نسبی یک تقریب دقت آن تقریب را نشان می‌دهد، این قضیه به‌خوبی ارتباط بین دقت یک تقریب را با تعداد ارقام با معنای درست آن نشان می‌دهد.

خطای اعمال محاسباتی

در حالت کلی اگر A و B دو عدد و a و b به ترتیب، تقریبهایی از آنها باشند و $\otimes$ نماد یک عمل باشد، در ماشین محاسباتی این عمل با نماد $\otimes^*$ تقریب زده می‌شود و در واقع آنچه ماشین به ما می‌دهد $a\otimes^*b$ است و داریم:

$$|A\otimes B - a\otimes^* b| = |(A\otimes B - a\otimes b) + (a\otimes b - a\otimes^* b)|$$

$$\le \underbrace{|A\otimes B - a\otimes b|}_{\text{خطای منتشر شده}} + \underbrace{|a\otimes b - a\otimes^* b|}_{\text{خطای تولید شده}}$$

بنابراین، خطای کل از مجموع خطای منتشر شده و خطای تولید شده بیشتر نیست.

اکنون می خواهیم حداکثر خطای منتشر شده را برای چهارعمل اصلی به‌جای $\otimes$ محاسبه کنیم.
معمولاً در عمل به خطای تولید شده توجه زیادی نمی‌شود هرچند گاهی اوقات باعث به دست آمدن جوابهای غیرقابل قبول می‌شود.

در ادامه جمع، تفریق، ضرب و تقسیم اعداد تقریبی را بررسی می‌کنیم.

جمع اعداد تقریبی

$$\sqrt{2}+\sqrt{3} \simeq 1/41 + 1/73 = 3/14$$

مشاهده می‌شود که تقریبی‌از $\sqrt{2}$ با تقریبی از $\sqrt{3}$ جمع شده است. اکنون می‌خواهیم معین کنیم که خطای ۳/۱۴ حداکثر چقدر است و چه ارتباطی باخطاهای ۱/۴۱ و ۱/۷۳ دارد.

قضیه

اگر a و b تقریبهایی از A و B و این اعداد جملگی مثبت باشند آنگاه

$$e(a+b) < e(a) + e(b)$$

$$\delta(a+b) \le \max\{\delta(a), \delta(b)\}$$

برهان

بنابر تعریف خطای مطلق یک تقریب، چون $a + b$ به عنوان تقریبی از $A + B$ پذیرفته می‌شود داریم:

$$e(a+b) = |(A+B) - (a+b)| \le |A-a| + |B-b| = e(a) + e(b)$$

برای اثبات قسمت دوم حکم، قرار می‌دهیم

$$D = \max \{\delta(a) , \delta(b)\}$$

بنابر قسمت اول قضیه و تعریف خطای نسبی:

$$\delta(a+b) \simeq \frac{e(a+b)}{a+b} \le \frac{e(a)+e(b)}{a+b} = \frac{e(a)}{a+b} + \frac{e(b)}{a+b}$$

$$= \frac{e(a)}{a} \times \frac{a}{a+b} + \frac{e(b)}{b} \times \frac{b}{a+b} \le D\left[\frac{a}{a+b} + \frac{b}{a+b}\right] = D$$

نتیجه حداکثر خطای $a + b$ مجموع خطاهای a و b است.

تفریق اعداد تقریبی

در مورد تفریق اعداد تقریبی به راحتی می‌توان نشان داد که $e(a-b) \le e(a) + e(b)$

اما، بنا بر تعریف $\delta(a-b) \simeq \frac{e(a-b)}{|a-b|}$ و اگر $|a-b|$ کوچک باشد خطای نسبی $a - b$ می‌تواند بزرگ باشد، که در نتیجه $a - b$ نادقیق خواهد بود.

مثال اگر A و B نزدیک به هم باشند و هدف محاسبهٔ $\frac{1}{A-B}$ باشد، خطامی‌تواند فاحش باشد.

مثلاً، با حساب سه رقم اعشار صحیح داریم:

$$C = \frac{۱}{(\sqrt{۲}+\sqrt{۳}) - \pi} \simeq \frac{۱}{(۱/۴۱۴ + ۱/۷۳۲) - ۳/۱۴۲} = \frac{۱}{۰/۰۰۴} = ۲۵۰$$

در صورتی که، اگر به‌جای اعدادموجود تقریبهایی تا ۹ رقم اعشار قرار دهیم و جواب‌را گرد کنیم خواهیم داشت:

$$C = \frac{۱}{۰/۰۰۴۶۷۱۷۱۶} = ۲۱۴/۱$$

در حالت کلی باید، حتی‌المقدور، از تفریق اعداد تقریبی نزدیک به‌هم اجتناب کرد.

اصولاً، با توجه به ارتباط بین تعداد ارقام با معنای درست و دقت یک تقریب، علت اصلی نادقیق

بودن b − a کم شدن تعداد ارقام بامعناست که باید از وقوع آن جلوگیری کرد. مـثلاً، بـه‌جای محاسبهٔ $\sqrt{۲} - ۱$ بهتر است، $\frac{۱}{\sqrt{۲}+۱}$ که از نظر ریاضی با آن برابر است حساب شود، هرچـند محاسبات بیشتر می شود.

اعداد ۱/۴۱ و ۱ با این مفهوم دو عدد نزدیک به هم هستند زیرا: $\sqrt{۲} - ۱ = ۱/۴۱ - ۱ = ۰/۴۱$ و عدد ۰/۴۱ دو رقم با معنی و ۱/۴۱ سه رقم با معنی دارد. لذا برای پیدا کردن تقریبی از $\sqrt{۲} - ۱$ باید چنین عمل کنیم:

$$\sqrt{۲} - ۱ = \frac{۱}{\sqrt{۲}+۱} = \frac{۱}{۱/۴۱+۱} = \frac{۱}{۲/۴۱} = ۰/۴۱۵$$

بدیهی است که ۰/۴۱۵ به مقدار $\sqrt{۲} - ۱$ نزدیکتر است تا ۰/۴۱

همچنین اگر به جای $\sqrt{۲} - ۱$ قرار دهیم ۰/۴۱ و بخواهیم روی این عدد تعداد عملیات محاسباتی بیشتری انجام دهیم، داریم:

$$(\sqrt{۲} - ۱)^۳ = (۰/۴۱)^۳ = ۰/۰۶۸۹۲۱$$

$$(\sqrt{۲} - ۱)^۳ = \frac{۱}{(\sqrt{۲} + ۱)^۳} = \frac{۱}{(۲/۴۱)^۳} = ۰/۰۷۱۴۴۱$$

ضرب اعداد تقریبی

قضیه

اگر a و b به ترتیب، تقریبهایی از A و B و این اعداد جملگی مثبت باشند :

$$e(ab) \le a\,e(b) + b\,e(a)$$

$$\delta(ab) \le \delta(a) + \delta(b)$$

برهان

با توجه به تعریف خطای مطلق داریم $e(ab) = |AB - ab| = |AB - aB + aB - ab|$

$$\le B\,|A - a| + a\,|B - b| = B\,e(a) + a\,e(b)$$

اما، اگر فرض کنیم $B = b + \varepsilon_b$، که در آن $|\varepsilon_b| = e(b)$ ، در نتیجه $Be(a) = be(a) + \varepsilon_b e(a)$

$\varepsilon_b\, e(a)$ در مقایسه با $be(a)$ قابل اغماض است و می‌توان نوشت: $Be(a) \simeq be(a)$

که در نتیجه حکم اول قضیه حاصل می‌شود.

برای اثبات قسمت دوم حکم، با توجه به تعریف خطای نسبی و قسمت اول قضیه، داریم :

$$\delta(ab) \simeq \frac{e(ab)}{ab} \le \frac{a\,e(b) + b\,e(a)}{ab} = \frac{e(b)}{b} + \frac{e(a)}{a} \simeq \delta(a) + \delta(b)$$

نتیجه قسمت اول قضیه نشان می‌دهد که اگر a یا b بزرگ باشد خطای ab می‌تواند قابل ملاحظه باشد. از این‌رو، باید حتی‌المقدور از ضرب اعداد تقریبی بزرگ اجتناب کرد و در صورت اجبار باید دقت این اعداد را بالا برد، مثلاً با دقت مضاعف کار کرد.

تقسیم اعداد تقریبی

قضیه

اگر a و b به ترتیب، تقریبهایی از A و B باشند، به قسمی که: $A = a + \varepsilon_a$ و $B = b + \varepsilon_b$.

آنگاه:

$$e\left(\frac{a}{b}\right) \leq \frac{a\,e(b) + b\,e(a)}{b^{2}}$$

$$\delta\left(\frac{a}{b}\right) \simeq \left|\frac{\varepsilon_a}{a} - \frac{\varepsilon_b}{b}\right| \leq \delta(a) + \delta(b)$$

نتیجه قسمت اول قضیه نشان می‌دهد که اگر a یا b بزرگ باشد خطای ab می‌تواند قابل ملاحظه باشد. از این‌رو، باید حتی‌المقدور از ضرب اعداد تقریبی بزرگ اجتناب کرد و در صورت اجبار باید دقت این اعداد را بالا برد، مثلاً با دقت مضاعف کارکرد.

تقسیم اعداد تقریبی

قضیه

اگر a و b به ترتیب، تقریبهایی از A و B باشند، به قسمی که: $A= a + \varepsilon_a$ و $B= b + \varepsilon_b$. آنگاه:

$$e\left(\frac{a}{b}\right) \leq \frac{a\,e(b) + b\,e(a)}{b^۲}$$

$$\delta\left(\frac{a}{b}\right) \simeq \left|\frac{\varepsilon_a}{a} - \frac{\varepsilon_b}{b}\right| \leq \delta(a) + \delta(b)$$

مشاهده می شود که خطای مطلق چندان ساده بر حسب عوامل تقسیم بیان نمی شود و چنانچه خود را به داشتن خطای نسبی خارج قسمت قانع کنیم، می توان چنین نتیجه گرفت که خطای نسبی خارج قسمت دو مقدار تقریبی تقریباً برابر است با تفاضل خطاهای تقریب.

در انجام تقسیم معمولاً به گونه ای عمل می کنیم که منجر به عمل ضرب گردد.

$$\frac{\pi}{۲\sqrt{۷}} = \frac{\pi}{۲\sqrt{۷}}\,\frac{\sqrt{۷}}{\sqrt{۷}} = \frac{۱}{۱۴}\pi\sqrt{۷}$$

مثال، معادله درجه دوم زیر را در نظر بگیرید:

$$x^2 + 111.11x + 1.2121 = 0$$

فرض کنید محاسبه ریشه بزرگتر معادله مورد نظر است که از فرمول شناخته شده زیر حاصل می شود:

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

اگر از حساب ممیز شناور تا ۵ رقم اعشار استفاده کنیم داریم: $b^2 = 0.12345 \times 10^5$

$$b^2 - 4ac = 0.12340 \times 10^5 \longrightarrow \sqrt{b^2 - 4ac} = 0.11109 \times 10^3$$

لذا در محاسبه $\sqrt{b^2-4ac}+-b$ ارقام با معنی از دست می روند و خطا افزایش می یابد.

$$x=\frac{-111.11+111.09}{2}=-0.01000$$

برای رسیدن به دقت مطلوب چنین عمل می کنیم:

$$x=\frac{-b+\sqrt{b^2-4ac}}{2a}=\frac{(-b+\sqrt{b^2-4ac})(-b-\sqrt{b^2-4ac})}{2a(-b-\sqrt{b^2-4ac})}=\frac{-2c}{b+\sqrt{b^2-4ac}}$$

لذا مقدار ریشه بزرگتر عبارتست از: $x=-0.01091$

که تمامی ارقام آن با معنی و صحیح هستند.

عدم برقراری خاصیت جابجایی و توزیعی جمع و تفریق در ماشین محاسب!

$$225.1-(224.8+0.1572)=225.1-(225.0)=0.1000$$

$$(225.1-224.8)-0.1572=0.3000-0.1572=0.1428$$

با تغییر ساده در ترتیب محاسبات، جواب ها تغییر قابل ملاحظه ای می کند.

مثال، خطای محاسبه توابع (با حساب ۶ رقم اعشار با معنی)

$$f(x)=x\left[\sqrt{x+1}-\sqrt{x}\right]$$

x	$f(x)$ محاسبه شده با ماشین	$f(x)$ مقدار واقعی
1	0.414210	0.414214
10	1.54340	1.54347
100	4.99000	4.98756
1000	15.8000	15.8074
10000	50.000	49.9988
100000	100.000	158.113

برای $x = 100$ در ماشین محاسب داریم:

$$\sqrt{100} = 10.0000, \qquad \sqrt{101} = 10.0499$$

$$\sqrt{x+1} - \sqrt{x} = \sqrt{101} - \sqrt{100} = 0.0499000,$$

در حالی که مقدار واقعی عبارتست از: 4.98756

برای رفع این مورد چنین عمل می کنیم:

$$f(x) = x\left(\frac{\sqrt{x+1} - \sqrt{x}}{1}\right)\left(\frac{\sqrt{x+1} + \sqrt{x}}{\sqrt{x+1} + \sqrt{x}}\right) = \frac{x}{\sqrt{x+1} + \sqrt{x}}$$

$$f(100) = 4.98756$$