Kuinka tarkka on riittävän tarkka?

Anne-Maria Ernvall-Hytönen
Åbo Akademi

Erilaisiin arviointiongelmiin törmää päivittäin: kaupassa on ihan kiva tietää, mikä ostosten loppusumma tulee suurin piirtein olemaan, tai jos kassalla summa vaikuttaa merkilliseltä, niin on mukava tarkistaa se. Harvemmin ainakaan itse lasken ostoksia sentin tarkkuudella. Yleensä pyöristän lähimpään euroon, tai yritän löytää kuitista asioita, joiden summa olisi suurin piirtein tasoja kymmeniä euroja, esimerkiksi siis bataatit 4,28 euroa ja juusto 5,95 euroa summautuvat karkeasti ottaen kymppiin. Myös niin, että pyöristetään lähimpiin euroihin ja sitten lasketaan yhteen, saadaan näille ostoksille kymppi loppusummaksi: eli bataatit 4 euroa ja juusto 6 euroa on yhteensä 10 euroa. Tarkka summahan olisi $4{,}28 + 5{,}95 = 10{,}23$, mutta hyvin harvoin tämä tarkkuus on tarpeen.

Pyöristysongelmaan olen törmännyt myös tutkimusprojekteissa: Jos halutaan vaikka tietokoneella piirtää kuvaajia erilaisten funktioiden ja lukujonojen käytöksestä, niin mikä tarkkuus riittää? Miten tiheästi pitää funktion arvoja laskea, jotta tietokoneen kuvaaja antaisi hyvän käsityksen funktion käytöksestä? Mistä kohtaa voi irrationaaliluvun desimaaliesityksen katkaista menettämättä älytöntä määrää tarkkuudesta? Näihin kysymyksiin vastaukset tietenkin riippuvat ainakin funktiosta, irrationaaliluvun luonteesta ja suuruudesta sekä vaaditusta tarkkuudesta.

Olen itse käyttänyt taannoin aikaa arvioidakseni, miten monta desimaalia tarvitaan Riemannin zetafunktion nollakohtien esityksistä, jotta ns. Lin kertoimien kuvaaja pysyy hallinnassa. Toisessa vaakakupissa oli siis laskenta-aika, toisessa puolestaan kuvaajan luotettavuus. Jouduimme arvioimaan esimerkiksi termin

$$\left(\frac{\rho}{\rho-\tau}\right)^n$$

käytöstä, kun τ on pieni vakio, n saa olla varsin iso ja ρ on Riemannin zeta-funktion nollakohta. Hyvin nopeasti kävi ilmi, että jos ρ on itseisarvoltaan iso, ei tarvita kovinkaan monen desimaalin tarkkuutta. Jos taas ρ on pieni, on virhe ihan valtava, jos n on suuruudeltaan vaikka sata. Päädyimme käyttämään kaksijakoista lähestymistapaa: suurimmalle o

Tässä tekstissä on tarkoitus parin esimerkin kautta näyttää, miten esimerkiksi voi laskujen lopputuleman tarkkuutta arvioida.

Luvun $\frac{1}{e}$ tarkkuus

Oletetaan tietävämme luvun e desimaalit tarkkuudella ϵ, eli toisin sanoen, käytämme luvulle e approksimaatiota θ, jonka suuruudesta tiedämme, että $|\theta - e| \leq \epsilon$. Jos esimerkiksi $\theta = 2{,}7$, niin $|\theta - e| < 0{,}018282$, jolloin luku ϵ voi olla esimerkiksi 0,018282.

Nyt voimme laskea todellisen osamäärän ja arvioosamäärän erotuksen:

$$\left|\frac{1}{e} - \frac{1}{\theta}\right| = \frac{1}{\theta} - \frac{1}{e} = \frac{e-\theta}{e\theta} < \frac{\epsilon}{e\theta}.$$

Esimerkissämme olisi siis

$$\frac{e-\theta}{e\theta} = \frac{e-2{,}7}{e\theta} < \frac{0{,}018282}{2{,}7^2} < 0{,}00251,$$

eli nyt tarkkuus jopa parani!

Mikäli luvun e sijaan olisimme olleet kiinnostuneita Eulerin vakion $\gamma \approx 0{,}5772$ arvioinnista vaikkapa luvulla 0,58, olisi käynyt heikommin: Alkuperäinen arviotarkkuus olisi ollut suurin piirtein $0{,}58 - 0{,}5772 = 0{,}0028$, mutta osamäärien tarkkuus olisikin ollut

$$\frac{1}{\gamma} - \frac{1}{0{,}58} \approx 0{,}0083,$$

eli tarkkuus olisi heikentynyt. Selitys tälle on se, että kun lukua ρ arvioidaan luvulla θ ja arvion tarkkuus on ϵ, pätee

$$\left|\frac{1}{\rho} - \frac{1}{\theta}\right| = \frac{|\theta-\rho|}{\theta\rho} < \frac{\epsilon}{\theta\rho},$$

joka on pienempi kuin ϵ, jos $\theta\rho > 1$, ja suurempi kuin ϵ, jos $\theta\rho < 1$, eli alle ykkösen olevien lukujen osamäärien arviot ovat heikompia kuin lähtöarvio, yli ykkösen puolestaan parempia.

Tämä on luonnollista: jos tarkkuus heikkenisi, kun jakajat olisivat yli yhden, tarkoittaisi se, että lopputulos olisi suhteellisesti todella heikko: erotus olisi pieni ja virhe verrattain iso.

Suhteellinen virhe

Verrataan vielä virheen kokoa tuloksen kokoon: onhan ihan eri asia, jos virhe on vaikkapa kymmenen suuruusluokkaa, jos tuloksen koko on biljoona tai jos tulos onkin esimerkiksi sadasosa.

Virheen olemme äsken todenneet olevan korkeintaan

$$\frac{\epsilon}{\theta\rho}.$$

Tuloksen koko on $\frac{1}{\rho}$. Suhteellinen virhe on siis näiden osamäärä

$$\frac{\frac{\epsilon}{\theta\rho}}{\frac{1}{\rho}} = \frac{\epsilon\rho}{\theta\rho} = \frac{\epsilon}{\theta},$$

eli myös suhteellinen virhe käyttäytyy samoin kuin absoluuttinen virhe: yli ykkösen olevien lukujen osamäärän arvio on parempi, alle ykkösen olevien taas huonompi.

Luvun π desimaalit

Ajatellaanpa, että tehtävänä on laskea π^{10} ykkösen tarkkuudella, eli löytää sellainen kokonaisluku n, että $|n - \pi^{10}| < 1$. Selvästikään ei riitä arvioida, että π on kolme, sillä

$$3^{10} = 59049$$

ja

$$3{,}1^{10} \approx 81962{,}8$$

ja π^{10} on vielä tätäkin suurempi. Yksi keino olisi tietenkin testata: Todetaan, että

$$\pi \approx 3{,}14159265358979323846264338\ldots$$

Tästä voisi sitten kokeilla, että riittäisikö vaikkapa kuusi desimaalia:

$$3{,}141592^{10} \approx 93647{,}85$$

ja

$$3{,}141593^{10} \approx 93648{,}15.$$

Nyt kävikin hyvä tuuri, ja tarkkuus on riittävä. Tämä tehtävä olisi siis nopeasti ohi kokeilemalla, mutta entä jos haluttaisiinkin vaikka laskea π^{100} kokonaisluvun tarkkuudella? Haarukointi ja testaus voisi viedä aika paljon pidempään.

Tarkastellaan nyt kuitenkin yksinkertaisuuden vuoksi alkuperäistä ongelmaa. Ajatellaan, että käytetään luvulle π likiarvoa θ, jolloin $|\pi - \theta| = \epsilon$. Kuinka pieni ϵ on varmasti riittävän pieni?

Lasketaan erotus

$$\left|\pi^{10} - (\pi \pm \epsilon)^{10}\right|.$$

Aloitetaan tilanteesta $|\pi^{10} - (\pi+\epsilon)^{10}|$. Koska $\pi + \epsilon > \pi$, voidaan kirjoittaa $(\pi+\epsilon)^{10} - \pi^{10}$. Esitellään nyt erilaisia tapoja lähestyä ongelmaa. Tilanne, jossa on miinus plus-merkin tilalla voidaan käsitellä samoin, ja välillä mahdollisesti jopa hieman helpommin.

Sulut auki

Yksi vaihtoehto olisi kertoa sulut auki. Yleensä en ole tämän lähestymistavan fani, mutta nyt se kuitenkin toimii.

$$(\pi+\epsilon)^{10} - \pi^{10} = \sum_{k=1}^{10} \binom{10}{k} \epsilon^k \pi^{10-k}.$$

Termejä on nyt yhteensä $2^{10} - 1 = 1023$ kappaletta. Jotta näihin saa mitään otetta, pitää olettaa jotain luvun ϵ suuruudesta suhteessa luvun π suuruuteen. Helppo oletus olisi vaikkapa olettaa, että $\epsilon < \frac{1}{\pi}$, mutta nyt esimerkiksi termille

$$\binom{10}{5} \epsilon^5 \pi^5$$

ei saa parempaa arviota kuin $\binom{10}{5}$, mikä ei missään nimessä riitä. Oletus $\epsilon < \frac{1}{\pi^{10}}$ voisi toimia paremmin. Nyt lauseke muuttuisi muotoon

$$\sum_{k=1}^{10} \binom{10}{k} \epsilon^k \pi^{10-k} < \sum_{k=1}^{10} \binom{10}{k} \pi^{10-10k-k},$$

mutta tämäkään ei ihan riitä, sillä

$$\sum_{k=1}^{10} \binom{10}{k} \pi^{10-10k-k} > \binom{10}{1} \pi^{10-10-1} = \frac{10}{\pi}.$$

Näin voi jatkaa ja yrittää lisätä oletuksia luvun ϵ suuruudelle, kunnes arvio on lopulta riittävän pieni. Lähestymistapa on melkoisen hidas ja kaikkien termien hyvä arviointi ei välttämättä ole ihan triviaalia, vaikkakin binomikerrointen kohdalla usein onnistaa. Parempi on lähteä hieman toisesta suunnasta: kunhan

$$\binom{10}{k} \pi^{10-k} \epsilon^k > 2 \binom{10}{k+1} \pi^{10-k-1} \epsilon^{k+1},$$

eli

$$\pi > 2\epsilon \frac{10-k}{k+1},$$

joka on varmasti tosi (sillä tiedämme jo varmasti testailujen pohjalta, että on vaadittava $\epsilon < \pi^{-10}$), huomaamme, että

$$2\epsilon \frac{10-k}{k+1} < 20\epsilon < \frac{20}{\pi^{10}} < \pi,$$

jolloin koko summaa voi arvioida geometrisena summana:

$$\sum_{k=1}^{10} \binom{10}{k} \epsilon^k \pi^{10-k} < 2 \binom{10}{1} \pi^{10-1} \epsilon = 20\pi^9 \epsilon.$$

Nyt tiedämmekin, että

$$|(\pi+\epsilon)^{10} - \pi^{10}| < 20\pi^9 \epsilon < 1,$$

kun $\epsilon < \frac{1}{20\pi^9}$. Kokonaislukupotensseille tämä tapa ei ole hullumpi, mutta yleisille funktioille mahdollisesti jopa mahdoton (aina ei ole sulkuja, jotka voisi kertoa auki, ja vaikka olisi, ei tapa välttämättä toimi arvioinnissa).

Ottaen tekijän

Tätä ongelmaa voimme lähestyä myös tekijän ottaen. Käytetään kaavaa

$$a^n - b^n = (a-b) \sum_{k=0}^{n-1} a^k b^{n-1-k},$$

jolloin saadaan

$$(\pi+\epsilon)^{10} - \pi^{10} = (\pi+\epsilon-\pi) \sum_{k=0}^{9} \epsilon^k \pi^{9-k} = \epsilon \sum_{k=0}^{9} \epsilon^k \pi^{9-k}.$$

Tämä on paljon helpompi esitys kuin sulut aukikertoen saatu. Nyt nimittäin huomaamme välittömästi, että summa muodostaa geometrisen jonon, jonka suhdeluku on $\frac{\epsilon}{\pi}$. Summa ei ainakaan pienene, jos kasvatamme sen äärettömän pitkäksi:

$$\epsilon \sum_{k=0}^{9} \epsilon^k \pi^{9-k} < \epsilon \sum_{k=0}^{\infty} \epsilon^k \pi^{9-k} = \epsilon\pi^9 \sum_{k=0}^{\infty} \left(\frac{\epsilon}{\pi}\right)^k.$$

Muutos äärettömän pitkäksi sarjaksi tehtiin, jotta summa olisi helpompi laskea. Nyt

$$\epsilon\pi^9 \sum_{k=0}^{\infty} \left(\frac{\epsilon}{\pi}\right)^k = \frac{\epsilon\pi^9}{1-\frac{\epsilon}{\pi}} = \frac{\epsilon\pi^{10}}{\pi-\epsilon} < 1,$$

kun

$$\epsilon\pi^{10} < \pi - \epsilon,$$

eli

$$\epsilon < \frac{\pi}{\pi^{10}+1},$$

eli aavistuksen luvun π^9 käänteislukua pienempi luku riittää, varmuudella esimerkiksi $\frac{1}{\pi^{10}}$.

Derivoiden ja integroiden

Yritetään nyt kuitenkin hieman toisenlaista, ihan yhtä suoraviivaista tapaa lähestyä ongelmaa: kirjoitetaan erotus derivaatan integraalina:

$$(\pi+\epsilon)^{10} - \pi^{10} = \int_{\pi}^{\pi+\epsilon} 10x^9 dx.$$

Koska integraalin suuruus on korkeintaan integroitavan funktion maksimiarvo integrointivälillä kertaa integraalin pituus, saadaan

$$\int_{\pi}^{\pi+\epsilon} 10x^9 dx \le 10\epsilon \max_{\pi \le x \le \pi+\epsilon} x^9 = 10\epsilon(\pi+\epsilon)^9.$$

Seuraava ongelma onkin arvioida siedettävällä tarkkuudella tämä lauseke. Toisin sanoen, arvioida lukua π jotenkin mukavasti ylöspäin. Jos arviosta haluaa selvitä ilman laskinta, voi todeta vaikkapa, että $\pi + \epsilon < 3{,}15$, kun esimerkiksi $\epsilon < \frac{1}{1000}$ ja $3{,}15^2 < 10$, joten

$$(\pi+\epsilon)^{10} - \pi^{10} \le 10 \cdot \epsilon \cdot 10^4 \cdot 3{,}15 = 315000\epsilon.$$

Jos tämä on alle yhden, niin myös alkuperäinen lauseke on varmasti alle yhden, joten kirjoitetaan

$$315000\epsilon < 1,$$

eli

$$\epsilon < \frac{1}{315000} \approx 0{,}00000317.$$

Vastaavasti voidaan laskea myös erotus $\pi^{10} - (\pi-\epsilon)^{10}$. Tämä jätetään lukijalle harjoitustehtäväksi.

Kannattaa huomioida, että vaikka näin saadaan vaadittu tarkkuus laskuihin, ei tämä suoraan esimerkiksi kerro, että π^{10} ja $(\pi-\epsilon)^{10}$ pyöristyisivät samaan kokonaislukuun, ei vaikka vaadittaisiin paljon parempikin tarkkuus, sillä vaikka lukujen erotus olisi kuinka pieni tahansa, voisivat ne silti sijaita eri puolin jotain muotoa $k + \frac{1}{2}$ olevaa lukua, missä k on kokonaisluku. Jos siis haluttaisiin tarkastella mihin kokonaislukuun jokin luku pyöristyy, vaadittaisiin vielä lisää työtä.