ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Lönnrotinkatu 4 B, 00120 Helsinki 12, Finland, tel. 601322

Keskusteluaiheita
Discussion papers

Yrjö O. Vartia*:

TUNNETUISSA PISTEISSÄ EPÄSÄÄN-
NÖLLISEN FUNKTION NUMEERINEN
INTEGROINTI

No 179 21.10.1985

* Tilastotieteen vt. professori
(Helsingin Kauppakorkeakoulu),
tutkimusjohtaja (Logit Ky)

ISSN 0781-6847

Tunnetuissa pisteissä epäsäännöllisen funktion numeerinen integrointi

1. Johdanto

Elinkeinoelämän tutkimuslaitoksessa on valmistumassa tulojakauma-ohjelmisto, joka mahdollistaa luokitellussa muodossa olevan tulojakauman monipuolisen analysoinnin. Luokitellusta frekvenssijakaumasta lasketaan jakauman ns. splinefunktio-esitys[1], josta edelleen määrätään mm. jakaumaa kuvaava tiheysfunktio, tunnusluvut, ns. markkojen kertymäfunktio ja tiheysfunktio sekä Lorenz-käyrä eri tunnuslukuineen numeerista integrointia käyttäen.

Tulojakaumaohjelmiston kehittämisessä törmättiin integroituvien funktioiden epäsäännöllisyyden (erityisesti epäjatkuvuuspisteiden) aiheuttamiin vaikeuksiin. Tavanomaiset numeeriset integrointimenetelmät edellyttävät integroitavan funktion säännöllisyyden: mm. epäjatkuvuuspisteiden ympäristössä tavalliset menetelmät antavat virheellisiä tuloksia.

Teknisissä ja luonnontieteellisissä sovellutuksissa epäjatkuvia funktioita harvemmin joudutaan tarkastelemaan, mutta taloudellisissa sovellutuksissa niitä kyllä esiintyy. Tulonsaajien ja markkojen tiheysfunktioesitykset (esim. yksinkertaiset hintagrammat ja niiden eräät yleistykset) ovat usein joissakin pisteissä epäjatkuvia tai nii-

---

1) Ks. Y. Vartia (1980): Interpolation of cumulative frequency curves by cubic splines, Helsingin yliopiston tilastotieteen laitoksen "Research report" No. 22.

den derivaatioilla on epäjatkuvuuskohtia. Vastaavanlaisia epäsäännöllisyyspisteitä esiintyy myös esim. veroasteikkoihin liittyvissä tarkasteluissa (ks. 3 luku).

Laskentatyön ja analyysin tehostamiseksi on osoittautunut hyödylliseksi tavanomaisten varsin epätarkkojen likimääräisten mallien ja mihin perustuvien (taskulaskimella laskettavien) karkeiden arvioiden sijaan käyttää tarkempia mutta mutkikkaampia malleja (esim. valtion tuloverofunktioille). Näihin liittyviä laskelmia varten on laskenta-algoritmit syytä ohjelmoida yleispätevässä muodossa tietokoneelle, jolloin päästään aiempaa huomattavasti monipuolisempaan, tarkempiin ja realistisempiin tuloksiin.

Tässä raportissa esitetään, kuinka klassinen Simpsonin numeerinen integrointimenetelmä voidaan yleistää algoritmiksi, joka sallii integroitavalle funktiolle äärellisen määrän epäsäännöllisyyspisteitä. Yleistetty ns. Simpson-Vartia -algoritmi (SV-algoritmi) on lähes yhtä helposti ohjelmoitavissa kuin Simpsonin menetelmäkin.

Luvussa 2 esitellään ongelma ja tarvittavat merkinnät. Luvussa 3 kuvataan kaksi verotukseen liittyvää sovellutusmahdollisuutta. Varsinainen uusi numeerinen integrointimenetelmä esitetään 4. luvussa.

2. Johdanto

Tarkastellaan funktiota $f^*(x)$, jonka epäjatkuvuus- ja muut epäsäännöllisyyspisteet ovat tiedossa. Järjestetään ko. epäjatkuvuuspisteet, $f^*(x)$:n derivaatan epäjatkuvuuspisteet ja muut mahdolliset erikoispisteet suuruusjärjestykseen $x(1) < x(2) < \ldots < x(K)$ ja talletetaan ne vektoriin

(1) $e = (x(1), \ldots, x(K))$.

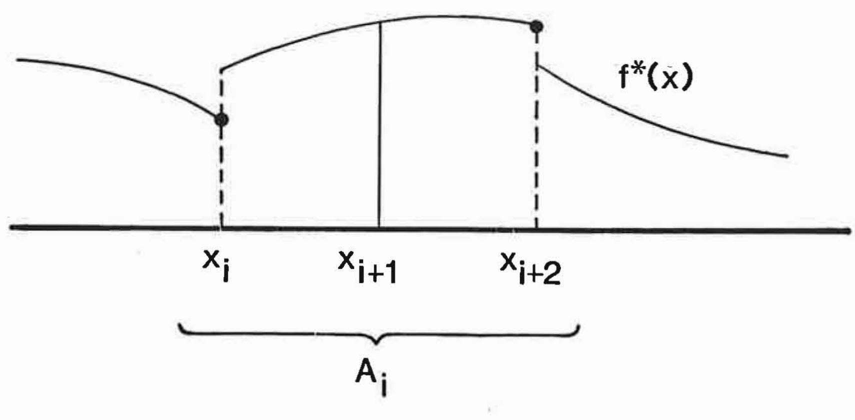
Funktion $f^*(x)$ kulku on tällöin säännöllistä jokaisella osavälillä $[x(i),\ x(i+1)]$. Tehtävänä on laskea numeerisesti $f^*(x)$:n integraali

(2) $$\int_{x(1)}^{x(K)} f^*(x)dx \ .$$

Jaetaan väli $[x(1), x(K)] = A$ osaväleihin $A_i = [x_i, x_{i+2}]$, $i=1,3,5,\ldots$ siten, ettei yksikään luvuista $x(k)$ $(k = 1,\ldots,K)$ ole minkään A_i:n sisäpiste eli luvut $x(k)$ voivat esiintyä vain välien A_i päätepisteinä. Välin A_i keskipiste olkoon x_{i+1}, jolloin siis lukujen x_i, x_{i+1} ja x_{i+2} peräkkäiset etäisyydet ovat samat: $x_{i+1} - x_i = x_{i+2} - x_{i+1} = \Delta$. Funktiolla $f^*(x)$ tai sen derivaatalla voi olla epäjatkuvuuspiste vain välien $A_i = [x_i, x_{i+2}]$ päätepisteissä.

Hankalimmassa tapauksessa $f^*(x)$ on molemmissa päätepisteissä epäjatkuva, ks. kuvio 1.

<u>Kuvio 1.</u> Välin A_i päätepisteissä epäjatkuva funktio $f^*(x)$.



Koska $A = A_1 \cup A_3 \cup \ldots A_I$ niin

(3) $$\int_{x(1)}^{x(K)} f^*(x)dx = \int_{A_1} f^*(x)dx + \ldots + \int_{A_I} f^*(x)dx$$

Kukin osaintegraali arvioidaan erikseen ja (2) niiden summana.

3. Sovellutusesimerkkejä

Esimerkiksi tulonsaajien keskimääräinen rajaveroaste m^0 voidaan arvioida integraalina

$$m^0 = \int_0^\infty m(y)f(y)dy ,$$

jossa $f(y)$ on tulonsaajien tiheys tulon y kohdalla eli $f(y) = F'(y)$, jossa $F(y)$ = tulonsaajien lukumäärän kertymä tuloon y saakka. Koska $m(y)$ on paloittain vakio ja $f(y)$:n arviotkin voivat olla epäjatkuvia tuloluokkien rajapisteissä on $m(y)f(y)$ epäjatkuva pisteissä, joissa $m(y)$ tai $f(y)$ on epäjatkuva. Lisäksi funktion $m(y)f(y)$ derivaatta voi olla epäjatkuva (ja määrittelemätön) vielä joissain muissakin pisteissä.

Vastaavasti keskimääräinen verojousto e^0 voidaan arvioida kaavalla

$$e^0 = \int_0^\infty e(y)w(y)dy ,$$

jossa $w(y) = T(y)f(y)/\int_0^{\infty} T(y)f(y)dy$ on maksettujen veromarkkojen tiheys tulon y kohdalla. Integroitava funktio $e(y)w(y)$ on varsin mutkikas funktio, joka on mm. vain paloittain jatkuva. Se hyppää aina tiettyjen tulorajojen kohdalla. Myös sen derivaatalla voi olla epäjatkuvuuspisteitä, jotka estävät yksinkertaisten, polynomeihin perustuvien numeeristen integrointikaavojen käytön ko. erikoispisteiden ympäristössä.

4. Numeerinen integrointi

Tarkastellaan kuvion 1 osaintegraalia

$$(4) \qquad I_i = \int_{A_i} f^*(x)dx = \int_{x_i}^{x_{i+2}} f^*(x)dx .$$

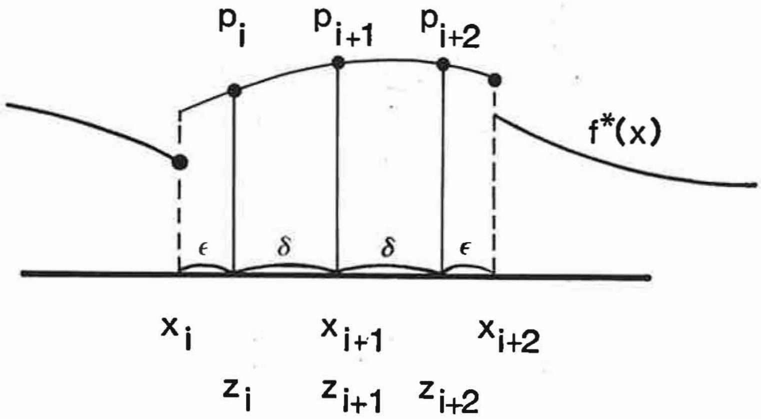
Funktiolla $f^*(x)$ tai sen derivaatalla voi olla epäjatkuvuuskohta pisteissä x_i taija x_{i+2}. Jos $f^*(x)$ on määritelty paloittain, saattavat sen eri 'paloihin' liittyvät määritelmät antaa ko. välien päätepisteissä (solmupisteissä) keskenään ristiriitaisia tuloksia. Esimerkiksi funktion arvoksi sen epäjatkuvuuspisteessä voidaan asettaa kumpi tahansa kahdesta arvosta. On vaikea etukäteen suunnitella, miten tällaisiin monikäsitteisyyksiin tietokoneohjelmassa tulisi reagoida. Selkeään ratkaisuun päästään välttämällä funktion arvojen laskemista ko. erikoispisteissä. Koska pisteet x_i ja x_{i+2} voivat olla erikoispisteitä, esitetään seuraavassa yksinkertainen menetelmä, jossa $f^*(x)$:n arvoja näissä välin A_i päätepisteissä ei tarvitse suoraan laskea.

Valitaan apupisteet z_i, z_{i+1} ja z_{i+2} siten, että

$$(5)\quad \begin{cases} z_i = x_i + \varepsilon \\ z_{i+1} = x_{i+1} \\ z_{i+2} = x_{i+2} - \varepsilon \end{cases}$$

ja ε on etäisyyteen Δ nähden pieni luku. Kutsumme ε:lla merkittyjen välien kohdalla olevia integroitavan funktion osia räystäiksi.

Kuvio 2.



Määrätään $f^*(x)$:n arvot pisteissä (5) eli räystäiden seinänpuoleisissa pisteissä:

$$(6)\quad \begin{cases} p_i = f^*(z_i) \\ p_{i+1} = f^*(z_{i+1}) \\ p_{i+2} = f^*(z_{i+2}) \end{cases}$$

Käytämme p-merkintää eo. pisteille, koska niiden kautta määrätään kulkemaan parabeli.

Näiden kolmen pisteen kautta kulkeva parabeli on

$$(7) \qquad p(x) = p_i + a_1(x-z_{i+1}) + a_2(x-z_{i+1})(x-z_{i+2}) ,$$

jonka kertoimet voidaan lausua seuraavasti käyttäen Newtonin ns. jaettua (divided) differenssia:

$$(8) \qquad a_1 = [z_i, z_{i+1}] = \frac{p_i - p_{i+1}}{z_i - z_{i+1}} = -(p_i-p_{i+1})/\delta$$

$$a_2 = [z_i, z_{i+1}, z_{i+2}]$$

$$= \frac{[z_i, z_{i+1}] - [z_{i+1}, z_{i+2}]}{z_i - z_{i+2}}$$

$$= \frac{1}{2\delta^2}(p_i - 2p_{i+1} + p_{i+2})$$

ja $z_i - z_{i+1} = z_{i+1} - z_{i+2} = -\delta$.

Parabeli $p(x)$ on siten funktion $f^*(x)$ käyttökelpoinen ja helposti määrättävä approksimaatio välin $[z_i, z_{i+2}]$ lisäksi myös räystäiden kohdalla eli koko välillä $A_i = [x_i, x_{i+2}]$. Voimme arvioida funktion $f^*(x)$ oikean puoleisen raja-arvon $f_i^* = \lim f^*(x)$ kun $x \to x_i + 0$ ja vasemmanpuoleisen raja-arvon $f_{i+2}^* = \lim f^*(x)$ kun $x \to x_{i+2} - 0$ parabelimme avulla seuraavasti:

$$(9) \qquad \begin{cases} f_i = p(x_i) & (\approx f_i^* = f^*(x_i)) \\ f_{i+2} = p(x_{i+2}) & (\approx f_{i+2}^* = f^*(x_{i+2})). \end{cases}$$

Arvioiden f_i ja f_{i+1} laskemisessa eivät $f^*(x)$:n mahdolliset epäjatkuvuudet räystäiden päissä pääse aiheuttamaan mitään vaikeuksia.

Lopullisena arviona pinta-alalle (4) saadaan Simpsonin kaavan mukaisesti

$$(10) \quad \int_{x_i}^{x_{i+2}} f^*(x)dx \approx \int_{x_i}^{x_{i+2}} p(x)dx$$

$$= \frac{h}{6}(f_i + 4f_{i+1} + f_{i+2})$$

jossa $h = x_{i+2} - x_i = 2\Delta$.

Tämän yksinkertaisen kaavan ohjelmoiminen ei tuota mitään hankaluuksia. Arvot f_i ja f_{i+2} ovat (9):n mukaiset arviot $f^*(x)$:lle räystäiden reunoilla. Ne lasketaan approksimoivan parabelin (7) avulla. Simpsonin menetelmään verrattuna on siten lisäksi määrättävä ja ohjelmoitava vain eo. parabeli $p(x)$, joka on alkuperäistä funktiota $f^*(x)$ yksinkertaisempi funktio. Räystäiden leveyttä ε lyhentämällä päästään esitetyllä Simpson-Vartia-menetelmällä tarkasteltujen epäsäännölisten funktioiden integroinnissa samaan tarkkuuteen kuin Simpsonin menetelmällä säännöllisiä funktioita integroitaessa.

Funktion $f^*(x)$ erikoispisteiden ja epäjatkuvuuksien aiheuttamat vaikeudet numeerisessa integroinnissa voidaan näin välttää tinkimättä numeerisen integroinnin tarkkuudesta. Laskentamenetelmä voidaan säilyttää lähes samana kuin integroitaessa jatkuvaa ja jatkuvasti derivoituvaa funktiota tavanomaisella Simpsonin menetelmällä. Lisäyksenä Simpsonin menetelmään on Simpson-Vartia menetelmässä vain pidettävä kirjaa erikoispisteistä (1) ja ohjelmoitava parabelin (7) lauseke.

ELINKEINOELÄMÄN TUTKIMUSLAITOS (ETLA)
The Research Institute of the Finnish Economy
Lönnrotinkatu 4 B, SF-00120 HELSINKI 12 Puh./Tel. (90) 601 322

KESKUSTELUAIHEITA - DISCUSSION PAPERS ISSN 0781-6847

No 167 JUHA KETTUNEN, Sairausvakuutuskorvauksien ja niiden rahoituksen kehityksestä. 12.11.1984. 36 s.

No 168 OSMO FORSSELL, Changes in the Structure of the Finnish Economy 1970-1980. 20.11.1984. 17 p.

No 169 JUHANI RAATIKAINEN - KARI TAKALA, Kausaalisuustestejä suomalaisilla rahamarkkinamuuttujilla. 11.02.1985. 53 s.

No 170 ROBERT A. HART, Wage Supplements through Collective Agreement or Statutory Requirement? 04.02.1985. 26 p.

No 171 AKINARI HORII, Financial Innovation and its Implications for the Japanese Financial System. 18.03.1985. 8 p.

No 172 PAUL D. McNELIS, Indexing and Macroeconomics: A Survey of Theoretical Developments and International Experience. 02.04.1985. 44 p.

No 173 EERO PYLKKÄNEN - PENTTI VARTIA, Some Comments on Fine-Tuning Macro-Model Forecasts. 29.05.1985. 22 p.

No 174 CHRISTIAN EDGREN, Automatic versus Discretionary Tax-Policy Effects Revisited. 05.06.1985. 44 p.

No 175 TAUNO KALLINEN, Työn tuottavuuden taso ja ansiotaso Suomen, Ruotsin, Norjan ja Tanskan tehdasteollisuudessa vuosina 1973 ja 1982. 11.06.1985. 44 s.

No 176 J. LESKELÄ - J. MOILANEN, A Comparison of Recent Medium-Term Forecasts. 17.06.1985. 34 p.

No 177 PAAVO SUNI, Efektiivisen nimellisen valuuttakurssin mittaaminen ja mittarien tulkinta. 20.06.1985. 42 s.

No 178 MARKKU LAMMI, Alkoholi- ja polttoaineverojen kertymästä. 27.09.1985. 49 s.

No 179 YRJÖ O. VARTIA, Tunnetuissa pisteissä epäsäännöllisen funktion numeerinen integrointi. 21.10.1985. 8 s.

Elinkeinoelämän Tutkimuslaitoksen julkaisemat "Keskusteluaiheet" ovat raportteja alustavista tutkimustuloksista ja väliraportteja tekeillä olevista tutkimuksista. Tässä sarjassa julkaistuja monisteita on rajoitetusti saatavissa ETLAn kirjastosta tai ao. tutkijalta.

Papers in this series are reports on preliminary research results and on studies in progress; they can be obtained, on request, by the author's permission.

0033A/21.10.1985