Nokkur atriði um launadreifingar

Helgi Tómasson
helgito@hi.is
Háskóli Íslands

11. mars 2009

1 Inngangur

Hugtakið dreifing er notað til að lýsa því að ekki eru allir einstaklingar í tilteknu þýði (population) eins. Í þessari grein er markmiðið að velta upp nokkrum atriðum er varða launa- og/eða tekjudreifingar og hvað þarf að hafa í huga við umfjöllun þeirra. Hugtakið, launa- og/eða tekjudreifing er pólítískt rafmagnað og stundum gefið í skyn að form dreifingarinnar sé einhvers konar mælistika á ranglæti. Hér verður ekkert fjallað um hvað sé réttlát dreifing, heldur einungis hlutlægar skilgreiningar og hvers vænta má vegna tölfræðilegra lögmála. Til að geta tjáð sig með hlutlægum hætti er nauðsynlegt hafa nákvæmar skilgreiningar. Gera þarf grein fyrir hvað sé átt við með hugtökum eins og launum og tekjum. Með launum er venjulega átt við tekjuflæði sem ákvarðast af framlagi einstaklings á vinnumarkaði. Með tekjum er átt við heildarflæðið þegar eigna/fjármagnstekjum hefur verið bætti við launin. Laun eru í eðli sínu jákvæð stærð, en eignatekjur geta falið í sér söluhagnað eða sölutap. Það að eignatekjur geti verið neikvæðar gerir meðhöndlun þeirra flóknari.

Í þessari grein eru ekki meðhöndluð raunveruleg gögn, í stað þess er áherslan á þá krafta sem kunna að liggja baki raunverulegum gögnum. Ekki er fjallað um áhrif af hugsanlegum neikvæðum eignatekjum. Þegar rætt eru um dreifingar þarf að liggja fyrir hvað dreifing er.Dreifing er til dæmis skipting endanlegra gæða milli einstaklinga. Allar slíkar skiptingar eru í eðli sínu skölուð líkindindadreifing, þ.e. að magn heildargæða er 1. Hugtakið líkindadreifing er stærðfræðihugtak. Það má líta á allar skiptingar á magni sem líkindadreifingar. Notkun á hugtökum úr líkindafræði býður upp á stærðfræðilegar rökvissar aðferðir við að lýsa dreifingum almennt. Vísindamenn nota líkön (model) til að skerpa tjáningu sína um tiltekin fyrirbæri. Líkan á að geta endurspeglað ákveðna eiginleika tengda vísindalegu vandamáli. Ýmis grundvallaratriði verður að hafa í huga þegar líkan er hannað. Þegar talað er um tölfræðileg líkan er stundum átt við einhvers konar líkindadreifingu sem notuð er til að gera grein fyrir breytileika í umhverfinu. Hagnýt tölfræðivinna gengur út á að tengja mælingar og fræðilega líkindadreifingu. Ekki er raunhæft að ætla að fyrir sérhvert fyrirbæri sé hægt að finna einfalda stærðfræðiformúlu sem gefi tæmandi lýsingu á fyrirbærinu. Hins vegar er í mörgum tilfellum hægt að finna viðunandi nálgun fyrir afmörkuð vandamál. Einnig stuðlar líkanasmíðin að öguðum kerfisbundnum vinnubrögðum. Einn af frægustu tölfræðingum 20. aldar lét hafa eftir sér: „All models are wrong, some are useful" (Box, 1979).

Mikilvægt er að átta sig á hver tilgangurinn er með líkanasmíðinni. Draper & Smith (1966) hugsa sér þrjár meginstefnur sem vísindamaður eigi að hafa í huga við gerð líkans:

1. The functional model
2. The control model
3. The predictive model

Með flokki 1 er átt við líkan sem lýsir algerlega einhverju kerfi. Með flokki 2 er átt við líkan sem nota má við stýringu, þ.e. að einangraður er þáttur sem hefur áhrif á tiltekna útkomu.

Með flokki 3 er átt við að hægt sé að hanna líkan sem í einhverjum skilningi segir til um framtíðina. Við rannsóknir á tekjudreifingum og þróun þeirra hefur nálgunin oftast byggst á einu af þremur eftirfarandi atriðum: 1) Stærðfræðilegur lipurleiki í fyrirrúmi, þ.e. að dreifing sé skilgreind með fáum stikum, fallform dreifingarinnar sé meðfærilegt o.s.frv. 2) Að dreifingin sem valin er sé vel studd fræðilega, með hagfræðilegum og stærðfræðilegum rökum. 3) Að dreifingin falli vel að mælingum.

Í þessari grein er farin eins konar kennslubókarleið, þ.e. að nota einfaldað viðmiðunarlíkan til að lýsa þáttum sem geta verið mikilvægir fyrir myndun og þróun tekjudreifinga. Tilgangurinn er að reyna að varpa ljósi á hvað atriði skipta máli þegar rætt eru um tekjudreifingar.

Hvað er jöfnuður/ójöfnuður? Hvernig má lýsa breytileika með einni tölu? Síðar eru rakin nokkur atriði sem vert er að hafa í huga varðandi það þann vanda að lýsa jöfnuði í dreifingu með einni tölu. Vinsæll kvarði, Gini-stuðullinn, er útskýrður með viðmiðunardæmi. Gini-stuðullinn er hlutfallslegur eins og ýmsir aðrir kvarðar. Þegar talað er um hlutfallsleg fátæktarmörk þá er ekki tekið tillit til þess hvernig þeir einstaklingar sem eru undir fátæktarmörkum lifa. Vel mætti hugsa sér kvarða sem ekki eru hlutfallslegir, eins og til dæmis það hlutfall einstaklinga sem ekki hefur efni á vatnsglasi og brauðsneið á dag.

Hvað er hreyfing? Þróun í tíma er samsett úr fyrirsjáanlegum atriðum annars vegar og ófyrirsjáanlegum atriðum hins vegar. Fyrirsjáanleg atriði eru sett fram með falli af tíma án allrar óvissu. Ófyrirsjáanleg atriði er breytur með óvissri útkomu. Runur slíka breyta eru stundum nefndar slembiferli (stochastic process). Einfaldasta gerð slembiferlis er runa af óháðum happdrættum. Óháður hér þýðir að útkomur úr fyrri happdrættum hafa ekki áhrif á vinningslíkur í næstu happdrættum. Ef liðnar útkomur hafa áhrif á vinningslíkur framtíðar er ferlið ekki lengur runa af óháðum happdrættum. Ef vísindamenn telja að slík tengsl séu raunhæf þarf að byggja slíkt inn í líkan.

Vinnumarkaður er samsettur úr misstórum kynslóðum. Hæfni og vilji einstaklinga til að afla tekna er háður aldri. Því er nauðsynlegt að hugleiða lýðfræðilega þróun þegar tekjudreifingar eru skoðaðar.

Einfalt líkan hefur þann kost að það er auðskilið, en jafnframt þann ókost að vera ekki mjög raunhæft. Raunhæfara líkan, byggt á einfaldri stærðfræði, verður fljótt flókið sem gerir yfirsýn mun erfiðari. Með flóknari stærðfræði má setja flóknari veruleika fram með skýrum hætti. Slík framsetning gerir hins vegar meiri kröfur til lesandans. Síðar er lauslega reifuð hugsanleg braut í líkanasmíði þegar gert er ráð fyrir að laun taki gildi á bili og hreyfist í samfelldum tíma.

Gögn eru ekki upplýsingar. Upplýsingar eru mælikvarði á hversu nákvæmlega er hægt að álykta um stika í tölfræðilegu líkani. Óáhugaverðar breytur geta truflað ályktun. Nauðsynlegt er að leiðrétta fyrir öllum þáttum sem geta haft áhrif. Leiðrétting fyrir truflandi þáttum er framkvæmd þannig að stikar eru metnir eða leystir út úr líkani með vörpunum (transformations). Sé ekki leiðrétt fyrir truflandi þáttum bjagast mat á áhrifum áhugaverðra breyta.

2 Hlutfallslegir kvarðar

Sömu atriði og gilda almennt um líkindadreifingar gilda um launadreifingar. Eins og um aðrar dreifingar eru hugtök eins og væntanlegt gildi, μ, og staðalfrávik, σ, mikilvæg:

$$\mu = E(X) = \int x f(x) dx,$$

$$\sigma = \sqrt{V(X)} = \sqrt{E(X-\mu)^2} = \sqrt{\int (x-\mu)^2 f(x) dx}.$$

Hér er fallið f, þéttifall (density function) hendingarinnar (random variable) X. Mest notaði einstaki mælikvarðinn á breytileika (ójöfnuð) er án efa staðalfrávik, $\sigma = \sqrt{E(X-\mu)^2}$. Ýmis konar vægi koma til greina, svo sem meðalfrávik frá meðaltali $E(|X-\mu|)$. Einnig hafa hlutfallskvarðar verið notaðar, t.d. hlutfall staðalfráviks af meðaltali, σ/μ (fráviks-hlutfalls, e: coefficient of variation). Slíkir kvarðar hafa þann galla að vægin (moments), meðaltöl, staðalfrávik o.s.frv. verða að vera til.

Kvantíll dreifingar F, x_q er skilgreint með $x_q = F^{-1}(q)$, þar sem F^{-1} er andhverfa dreififalls dreifingarinnar. Þetta þýðir 100q% eru fyrir neðan x_q. Kvantílabil (x_q, x_{1-q}) eða kvantílahlutföll x_{1-q}/x_q eru hugsanlegir kvarðar á dreifingu. Kvantílar hafa þann eiginleika að þeir breytast ekki þó teygt sé á dreifingunni t.d. fyrir ofan x_{1-q}, þ.e. þeir eru ónæmir fyrir hreyfingum innan þeirra flokka sem þeir skilgreina. Þeir lýsa því mjög takmörkuðum eiginleikum dreifingarinnar. Því hafa verið þróaðir kvarðar sem vega saman fleiri eiginleika dreifingarinnar.

Langþekktasti kvarðinn sem notaður er við lýsingu á tekjudreifingum er án efa Gini-stuðullinn: Ef meðaltal jákvæðrar hendingar X er μ þá mælir fallið:

$$G(x) = \frac{\int_0^x s f(s) ds}{\mu},$$

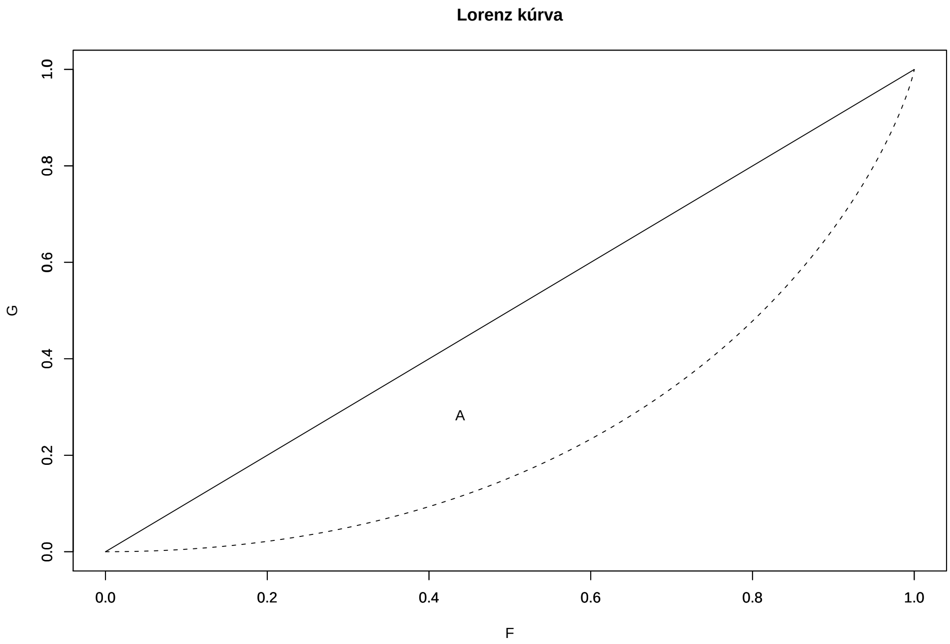
meðaltal þeirra sem eru með lægra gildi en x sem hlutfall af meðaltali allra. Grafið $(F(x), G(x))$ þar sem

$$F(x) = \int_0^x f(s) ds,$$

er dreififall X, er kallað Lorenz-kúrfa, sjá mynd 1. Gini stuðullinn, g, er tvöfalt flatarmál svæðis á milli Lorenz-kúrvunnar og beinnar línu með hallatölu 1 (Lorenz-kúrvu jafnrar dreifingar). Með hlutheildun má fá að:

$$g = 1 - \frac{\int_0^\infty (1-F(s))^2 ds}{\int_0^\infty (1-F(s)) ds}.$$

Auðvelt er því að sjá að Gini-stuðull fyrir allar veldisdreifingar er 1/2, óháð gildi á stika dreifingarinnar. Mynd 1 sýnir Lorenz-kúrvur veldisdreifingar og jafnrar dreifingar. Gini-stuðul má túlka sem samanburð á tiltekinni dreifingu og jafnri dreifingu. Gini-stuðullinn



Mynd 1: Óbrotna línan er Lorenz-kúrva jafnrar dreifingar, brotna línan er Lorenz-kúrva veldisdreifingar. A táknar flatarmálið á milli kúrvanna. Gini=2A sem er 1/2 fyrir allar exponential dreifingar.

er tvöfalt flatarmál milli Lorenz-kúrfu dreifinganna. Fyrir Pareto dreifingu gildir að:

$$g = 1 - \frac{2(\alpha - 1)}{2\alpha - 1}.$$

Til að væntanlegt gildi Pareto dreifingar sé til er $\alpha > 1$ nauðsynlegt og til að staðalfrávik sé til er $\alpha > 2$ nauðsynlegt, þ.e. $g < 1/3$. Gini stuðull hefur einungis merkingu fyrir dreifingar þar sem meðaltal er skilgreint og því ljóst að fyrir Pareto-dreifingu með skilgreint meðaltal gildir að $g < 1/2$. Notkun á Pareto dreifingu setur því ákveðnar skorður á hvaða gildi Gini-stuðullinn getur tekið. Fyrir Weibull dreifingu er Gini-stuðullinn

$$g = 1 - (1/2)^{1/\gamma}.$$

Veldisdreifing er sértilfelli af Weibull dreifingu, $\gamma = 1$, og því er Weibull dreifingin jafnari en veldisdreifing (þ.e. $g < 1/2$), ef $\gamma > 1$. T.d. ef $\gamma = 2$ þá er $g \simeq 0,29$.

Ginistuðullinn er sköluð meðalfjarlægð milli tveggja einstaklinga,

$$g = \frac{\frac{1}{2}E(|X - Y|)}{\mu}.$$

Hér ef fjarlægð mæld með tölugildisfráviki. Tölugildisfrávik er oft kallað L_1-norm. Í hagnýtri tölfræði er oft notast við kvaðratfrávik, sá fjarlægðarkvarði, er til samræmis nefndur L_2-norm. Stuðull sem er hliðstæður Gini-stuðlinum en byggir á L_2-normi er:

$$CV = \frac{\sqrt{\frac{1}{2}E(X-Y)^2}}{\mu},$$

sem er vel þekkt undir heitinu „coefficient-of-variation". Augljóslega má alhæfa þessa formúlu og reikna

$$\frac{[E\frac{1}{2}|X-Y|^p]^{1/p}}{\mu},$$

eða að byggja á L_∞ mælikvarða og reikna

$$\frac{\frac{1}{2}(x_{max}-x_{min})}{\mu},$$

þar sem x_{max} er stærsta mögulega gildi og x_{min} er minnsta mögulega gildi. Gini-stuðullinn hefur þann kost að ef meðaltal er til þá er hægt að reikna hann. CV-stuðullin krefst þess að $E(X^2) < \infty$ sé til. Nauðsynleg forsenda þess að L_p mælikvarði sé nothæfur þá þarf $E(X^p)$ að vera til. Augljóst er að sá kvarði sem byggir L_∞ kvarða er ekki nothæfur nema að til séu hæsta gildi og lægsta gildi. Samanburðarhæfni kvarða sem byggja á L_p fjarlægðarmáli er háð því að fjarlægðir séu endanlegar. Ef dreifing hefur „þungan hala", þ.e. $E(|X|^p) = \infty$, er ekki hægt að tala um meðalmun milli tveggja einstaklinga mældan með L_p kvarða.

Aðrir kvarðar sem stungið hefur verið upp á fyrir tekjudreifingar eru t.d. Atkinsson-stuðullinn:

$$A = 1 - \frac{1}{\mu}\left[\int_0^\infty x^{1-\varepsilon} f(x)dx\right]^{1/(1-\varepsilon)},$$

Kolm-stuðullinn:

$$K = \frac{1}{\beta}log\left[\int_0^\infty exp(\beta(\mu - x))f(x)dx\right],$$

og Theil-stuðullinn:

$$T = \frac{\beta}{1+\beta}\int_0^\infty (\frac{x}{\mu} - 1)^{\beta+1} f(x)dx.$$

Stuðlarnir ε og β setja ákveðnar vogir á tiltekna hluta dreifingarinnar.

Í bókum um tekjudreifingar eru tilgreindir ýmsir æskilegir eiginleikar sem mælikvarðar á jöfnuð (ójöfnuð) ættu að hafa:

- Að vera ekki fall af meðaltali, þ.e. ef öll laun eru tvöfölduð ætti það ekki að hafa áhrif á kvarðann.

- Stærð þýðis á ekki að skipta máli.

- Ef tveir einstaklingar skiptast á launum á það ekki að hafa áhrif á kvarðann.
- Það að færa frá „ríkum" til „fátæks" á að minnka mældan ójöfnuð
- Það ætti að vera hægt að sundurliða breytileikann, t.d. eitthvað í líkingu við ANOVA.

Gini-stuðullinn uppfyllir fyrstu fjögur skilyrðin en ekki það fimmta. Það eru til mælikvarðar sem uppfylla öll skilyrðin.Sumir þeirra, eins og t.d. Kolm og Atkinson, sem byggja á upplýsingafræði (enska: information theory) eru e.t.v. ekki eins auðskildir og Gini-stuðullinn, en geta verið hentugri, t.d. ef að sundurliða á þýði í hópa eða tekjur í tegundir tekna. Cowell & Victoria-Feser (1996) sýna nánari útfærslu og setja fram nokkra möguleika sem eru sundurliðanlegir.

Einfalt dæmi um Gini-stuðul

Gert er ráð fyrir að einungis séu tveir launaflokkar, l og h, $0 < l < h$. Ef líkur á að einstaklingur X sé í launaflokki h eru $P(X = h) = p_h = p$ og í launaflokki l $P(X = l) = 1 - p_h = 1 - p$, þá er Gini-stuðull dreifingarinnar

$$g = \frac{(h-l)p(1-p)}{ph + (1-p)l}.$$

Ef launadreifingin er lík útkomu í happdrætti þar sem vinningurinn h er í hlufalli við $1/p_h \simeq 0$ og $l \simeq 0$, þá er ljóst að:

$$g = \frac{h \quad 1/h}{1/h \quad h+0} \simeq 1,$$

þ.e. að vinningshafinn í happdrættinu fær nánast allan pottinn.

3 Hreyfimynstur

Laun einstaklinga þróast yfir ævina. Tilfærsla einstaklings milli launaflokka byggir á blöndu af heppni og uppsafnaðri reynslu. Það að líta á launaþróun einstaklings sem runu af óháðum happdrættum er ekki raunhæft þar sem laun einstaklings á hverjum tímapunkti eru háð launum hans á fyrri tímapunktum. Sá einstaklingur sem er mjög hæfur í dag, er það líklega einnig í nánustu framtíð. Til að geta beitt stærðfræðilegum röksemdum er hreyfimynstri oft lýst með líkönum sem byggja á einhvers konar diffurjöfnum eða mismunajöfnum. Það að nota slembin (stochastic) líkön er stærðfræðileg leið til að nálga eðli óvissrar þróunar.

Í þessari grein er gengið út frá einföldu dæmi þar sem aðeins er gert ráð fyrir tveim launaflokkum, l og h, þar sem $l < h$. Einstaklingur getur farið á milli flokkanna en gert er ráð fyrir ákveðinni tregðu, þ.e. að hálaunaður einstaklingur hefur tilhneigingu til að vera áfram hálaunaður. Möguleg einföldun á slíku hreyfimynstri er að gera ráð fyrir að minnið sé takmarkað. Með takmörkuðu minni er átt við að einungis nýliðin fortíð skipti máli. Í líkindafræði er þetta kallað Markov-eiginleiki. Hefðbundin framsetningarmáti er:

$$\underbrace{P(X_{t+1} = x_{t+1}|X_t = x_t, X_{t-1} = x_{t-1} \cdots)}_{\text{A}} = \underbrace{P(X_{t+1} = x_{t+1}|X_t = x_t)}_{\text{B}}$$

Hér táknar vinstri hlið jöfnunnar, A, líkur á að tekjurnar X_{t+1}, á tíma $t+1$ taki gildið x_{t+1}, gefin öll fortíðin. Hægri hliðin, B, táknar líkur á að þær taki gildið, x_{t+1}, gefið að síðasta gildi sé x_t. Markov-eiginleikinn segir að $A = B$, þ.e. að allar upplýsingar um áhrif sögunnar séu í nýjasta gildinu.

Hreyfimynstur í einföldu dæmi

Gert er ráð fyrir að laun séu annað hvort há, h, eða lág l. Einnig er gert ráð fyrir að líkur á að vera í háum flokki á tíma $t+1$ séu fall af þeim launaflokki sem viðkomandi var í á tíma t. Líkur á að hálaunaður einstaklingur á tíma t sé einnig hálaunaður á tíma $t+1$ eru táknaðar q_{hh} og líkur á að láglaunaður einstaklingur á tíma t sé einnig láglaunaður á tíma $t+1$ eru q_{ll}. Dreifingin á tíma 0 er táknuð með $p_{h,0}$ og $p_{l,0} = 1 - p_{h,0}$. Einföld líkindafræði gefur að líkur á að einstaklingur valinn af handahófi sé hálaunaður á tíma $t = 1$ séu

$$p_{h,1} = q_{hh}p_{h,0} + (1 - q_{ll})p_{l,0},$$

og láglaunaður á tíma 1,

$$p_{l,1} = (1 - q_{hh})p_{h,0} + q_{ll}p_{l,0}.$$

Þetta er þægilegt að setja fram á á fylkjaformi:

$$\underbrace{\begin{bmatrix} p_{h,1} \\ p_{l,1} \end{bmatrix}}_{\boldsymbol{p}_1} = \underbrace{\begin{bmatrix} q_{hh} & 1 - q_{ll} \\ 1 - q_{hh} & q_{ll} \end{bmatrix}}_{\boldsymbol{T}} \underbrace{\begin{bmatrix} p_{h,0} \\ p_{l,0} \end{bmatrix}}_{\boldsymbol{p}_0}.$$

Við það að fara úr tíma $t = 0$ og í tíma $t = 1$ hefur tilfærslufylkið, $\boldsymbol{T}$, breytt dreifingunni úr $\boldsymbol{p}_0$ í $\boldsymbol{p}_1$. Hafi upphafsdreifingin til dæmis verið $\boldsymbol{p}_0 = (0, 1)$, þ.e. að allir hafi verið með lág laun þá er dreifingin á tíma 1 $\boldsymbol{p}_1 = (1 - q_{ll}, q_{ll})$. Breytileiki dreifingarinnar sem var enginn í $\boldsymbol{p}_0$ er því jákvæður í $\boldsymbol{p}_1$. Auðvelt er að reikna Gini-stuðul fyrir $\boldsymbol{p}_0$ og $\boldsymbol{p}_1$. Í þessu Markov líkani er til jafnvægisdreifing, þ.e. dreifing sem tilfærslufylkið, $\boldsymbol{T}$ heldur fastri. Þá dreifingu má finna með því að leysa jöfnuna

$$\boldsymbol{T}\boldsymbol{p} = \boldsymbol{p}.$$

Eiginvektor fylkisins $\boldsymbol{T}$ með eigingildi 1 verður því jafnvægisdreifingin. Auðvelt er að sýna að sú dreifing, $\boldsymbol{p}_\infty$ er:

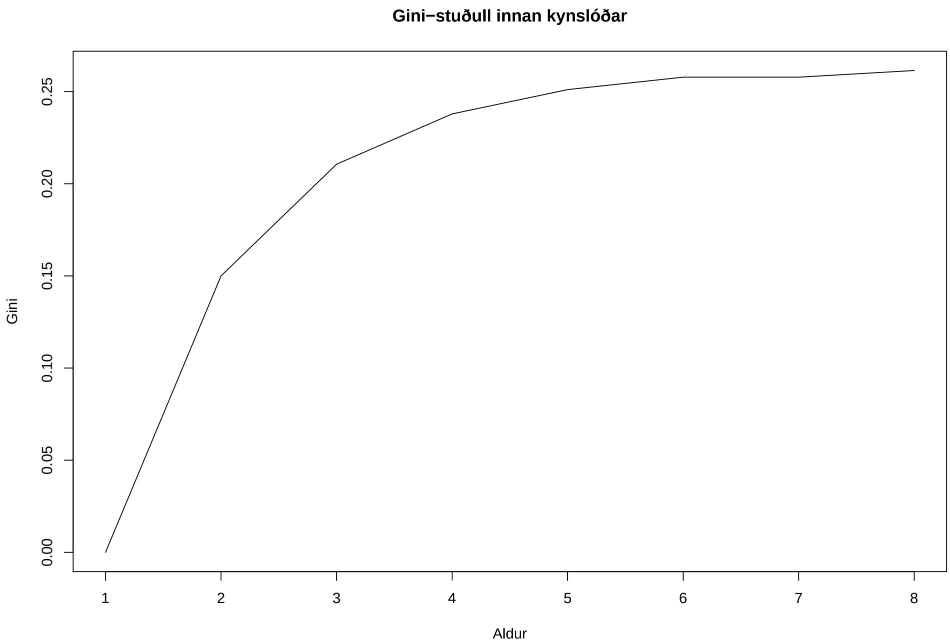
$$\boldsymbol{p}_\infty = \begin{bmatrix} \dfrac{1 - q_{ll}}{2 - q_{ll} - q_{hh}} \\ \dfrac{1 - q_{hh}}{2 - q_{ll} - q_{hh}} \end{bmatrix}.$$

Til dæmis má hugsa sér kerfi þar sem líkur á að hálaunaður einstaklingur haldist hálaunaður milli tímabila séu $q_{hh} = 0,8$ og að láglaunaður einstaklingur haldist láglaunaður milli tímabila séu $q_{ll} = 0,9$. Þróun dreifingar þessa kerfis fyrstu 8 tímabilanna er lýst í töflu 1.

	p_0	p_1	p_2	p_3	p_4	p_5	p_6	p_7
h	0,0	0,10	0,17	0,22	0,25	0,28	0,28	0,29
l	1,0	0,90	0,83	0,78	0,75	0,72	0,72	0,71

Tafla 1: Þróun launadreifingar í tíma.

Vel má hugsa sér að dálkarnir 8 í töflu 1 séu 5 ára tímabil, þ.e. að taflan lýsi launadreifingu í 40 ára sögu kynslóðar. Jafnvægisdreifingin er að 0,33 séu hálaunaðir. Mynd 2 sýnir þróun Gini-stuðulsins innan kynslóðar miðað við þetta hreyfimynstur og upphafsdreifinguna „allir láglaunaðir".



Mynd 2: Þróun Gini-stuðuls innan kynslóðar.

Champernowne (1953) rekur flóknara dæmi í frægri grein. Hann gengur út frá ákveðnu hreyfimynstri í logaritma af tekjum einstaklinga í þýði og sýnir að jafnvægisdreifingin verður Pareto-dreifing. Með hugtakinu jafnvægisdreifing er átt við þá dreifingu sem hreyfimynstrið heldur fastri. Í sumum tilfellum er hægt að leiða þá jafnvægisdreifingu út (Bhattacharya & Waymire, 1990). Í D-ritgerð (eins konar mastersritgerð) fékkst höfundur við Markov líkan fyrir tekjuþróun sænskra karlkyns launþega (Tómasson, 1980). Þar var reiknuð út jafnvægisdreifing og virtist hún svipuð og mæld dreifing.

Eðlisfræðingarnir Dragulescu & Yakovenko (2003) flytja mál sitt á þann hátt að jafnvægisdreifing orku fylgi lögmáli Boltzmann-Gibbs, sem er veldisdreifing (exponential distribution). Það byggir á lögmálinu um varðveislu orkunnar, þ.e. að orka fyrir og eftir breytingu sé sú sama. Þ.e. $m_1 + m_2 = m_1' + m_2'$ og slíkt hið sama gildi um fjármuni. Þeir hugsa sér síðan að Bandaríkin séu þróað land með tveim fyrirvinnum í heimili. Tekjudreifing hvorrar fyrirvinnu ætti að vera veldisdreifð samkvæmt lögmáli Boltzmann-Gibbs og summa tveggja óháðra veldisdreifðra hendinga er ákveðin gammadreifing. Sú gammadreifing hefur ginistuðul 3/8. Þeir athuguðu síðan gögn frá bandarískum skattayfirvöldum og fengu að úrtaks gini-stuðullinn fyrir fjölskyldutekjur er nálægt 3/8. Þeir hafa einnig unnið svipað verk með breskum gögnum (Dragulescu & Yakovenko, 2001)

Á hverjum tíma samanstendur þýði af safni misgamalla einstaklinga. Á hverju ári sem líður eldist einstaklingur um eitt ár. Ef $n(a,t)$ táknar fjölda einstaklinga á aldri a á tíma t, má lýsa þróun í mannfjölda með

$$n(a+1, t+1) = n(a,t)(1 - d(a,t)),$$

þar sem $d(a,t)$ er hlutfall einstaklinga á aldri a sem deyja á tíma t (þeirra sem eru lifa tíma t en deyja fyrir tíma $t+1$).

Áhrif misstórra kynslóða í einföldu dæmi

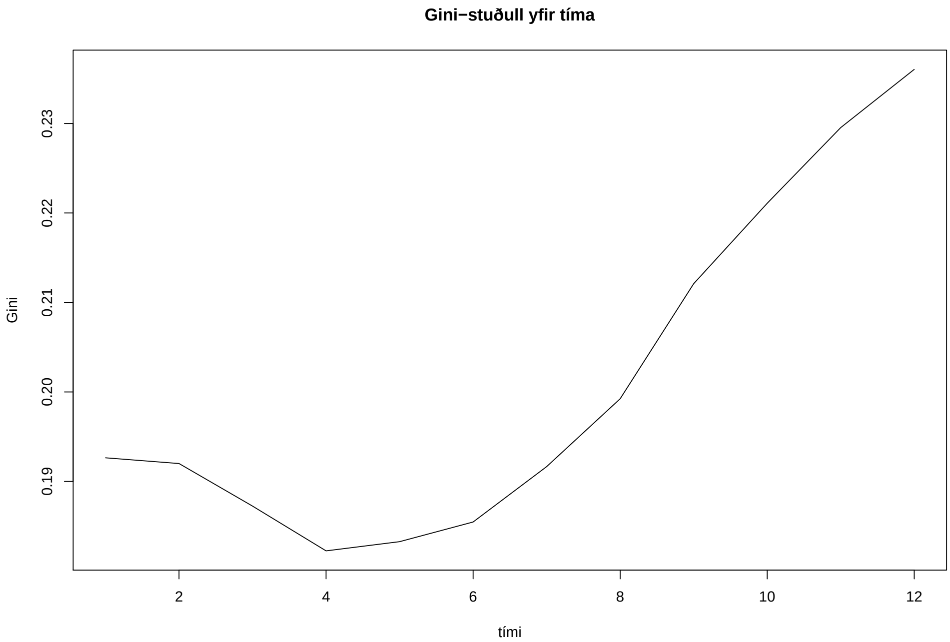
Stærð kynslóða er breytileg og því orsakar þróun í kynslóðarstærð þróun í jöfnuði. Gert er ráð fyrir að hreyfimynstur einstaklinga innan kynslóðar sé það sem liggur að baki töflu 1. Gert er ráð fyrir að 1% fækkun í hverri kynslóð á hverju tímabili. Þ.e. ef fjöldi einstaklinga á aldri a á tíma t er $n(a,t)$ þá er

$$n(a+1, t+1) = 0,99n(a,t).$$

Í byrjun, á tíma $t = 1$, er gert ráð fyrir að aldursdreifing sé með þeim hætti að yngsti hópurinn sé langstærstur. Fyrri hluti tímabilsins lýsir sér í miklum vexti í nýliðun á vinnumarkaði. Seinni hluti tímabilsins einkennist af því að nýliðar á vinnumarkaði eru færri en tímabilið á undan. Af þessum sökum breytist aldursdreifingin á tímabilinu. Aldursdreifing í lok tímabilsins er því nánast jöfn. Ef gert er ráð fyrir að tíminn sé mældur í 5 ára tímabilum þá er mannfjöldaþróunin hliðstæð því sem var á Íslandi á seinni hluta 20. aldar. Mynd 3 lýsir þróun Gini-stuðuls fyrir þetta kerfi. Í fyrstu eykst jöfnuður vegna þess að nýliðar á lágmarkslaunum hrúgast inn í kerfið, þegar líður á er eru hóparnir á miðjum aldri og eldri hlutfallslega stærri hluti þýðis og þá einkennist þróunin af því að þýðið er að eldast og ójöfnuður eykst með aldri.

4 Flóknari líkön og skilgreiningar á breytum

Dæmið hér að framan er að sjálfsögðu fráleit einföldun. Það voru aðeins tvö ástönd, h og l. Auðvelt væri að hugsa sér að h og l væru föll af skýristærðum, t.d. aldri, starfi og menntun. Það er augljóslega ekki raunhæft að segja að há laun fyrir 25 ára einstakling



Mynd 3: Þróun Gini-stuðuls í tíma.

séu jöfn háum launum fyrir alla aldurshópa. Hægt er að gera líkanið almennara með því að láta tilfærslufylkið vera háð aldri, menntun, starfi, o.s.frv. Það væri einnig hægt að fjölga launaflokkunum. Ef þeim er fjölgað þarf að hugleiða hvernig það er gert. Er það gert þannig að sami krónufjöldi sé á milli flokkanna? Eða er sama hlutfall á milli hverra tveggja? Eða er einhver allt önnur flokkun valin? Í dæminu að framan var tíminn látinn ganga í skrefum. Tilgangurinn með því að hafa dæmið einfalt var að skýra áhrif af Markov-hreyfimynstri og misstórum kynslóðum. Eftir því sem líkanið verður raunhæfara þeim mun erfiðara verður að hafa yfirsýn.

Launaflokkarnir í dæminu sem hér hefur verið lýst er dæmi um þrepaskipt ástandsrúm (state space). Hægt er að skilgreina Markov líkön fyrir samfellt ástandsrúm og samfelldan tíma. Í því tilfelli eru engin þrep í launum, heldur unnið með laun sem samfellda breytu. Einnig líður tíminn í samfelldum straumi í stað þess að ganga í skrefum. Form líkana verður einfaldara en stærðfræðin að baki þeim flóknari. Kostur þess að nota flóknari stærðfræði er að öflugri verkfæri verða handbær, galli er hins vegar að það getur verið erfiðara að skilja líkönin og því í sumum tilfellum erfiðara að sannfæra aðra um boðskapinn.

Frá 1997, þegar Merton og Scholes fengu Nóbelsverðlaun í hagfræði fyrir fjármálastærðfræði, hefur þekking á slembnum diffurjöfnum (stochastic differential equations) vaxið. Í fjármálastærðfræði er hreyfimynstri lýst með slembnum diffurjöfnum:

$$dX(t) = \underbrace{a(X(t),t)dt}_{A} + \underbrace{b(X(t),t)dW(t)}_{B}$$

Í þessari jöfnu er A fyrirsjáanlegur hluti hreyfimynstursins, og B óviss hluti. Í framhaldsskólum er nemendum kennt að leysa diffurjöfnur þar sem B-hlutanum hefur verið sleppt. Í óvissa hlutanum er dW, tákn fyrir óspáanlega stærð, hvítt suð (W er skammstöfun fyrir Wiener ferli). Ef tiltekin skilyrði eru uppfyllt þá er til jafnvægisdreifing. Vel þekkt líkan í fjármálastærðfræði er CEV (constant elasticity variance) líkanið.:

$$dX(t) = \kappa(\alpha - X(t))dt + \sigma X(t)dW(t) \tag{1}$$

Í þessu tilfelli gildir að ef upphafsgildið $X(0)$ er jákvætt þá mun ferlið alltaf verða jákvætt og til er jafnvægisdreifing sem er inverse-gamma (IG), Hún hefur það sameiginlegt með Pareto-dreifingunni að vera með þungan hala, þ.e. mjög stór gildi eru möguleg. (Ef hending X fylgir IG-dreifungu þá er $1/X$ gamma-dreifð). Túlkun stikanna í CEV líkaninu er að α er meðaltal jafnvægisdreifingarinnar, κ lýsir aðlögunarhraða að jafnvægi og σ lýsir skrefstærð í breytingu. Það að óvissi hlutinn hafi vægið $\sigma X(t)$, segir að þegar $X(t)$ er stórt, þá er skrefstærð breytingar stærri, þ.e. hálaunaðir hreyfast í stærri stökkum.

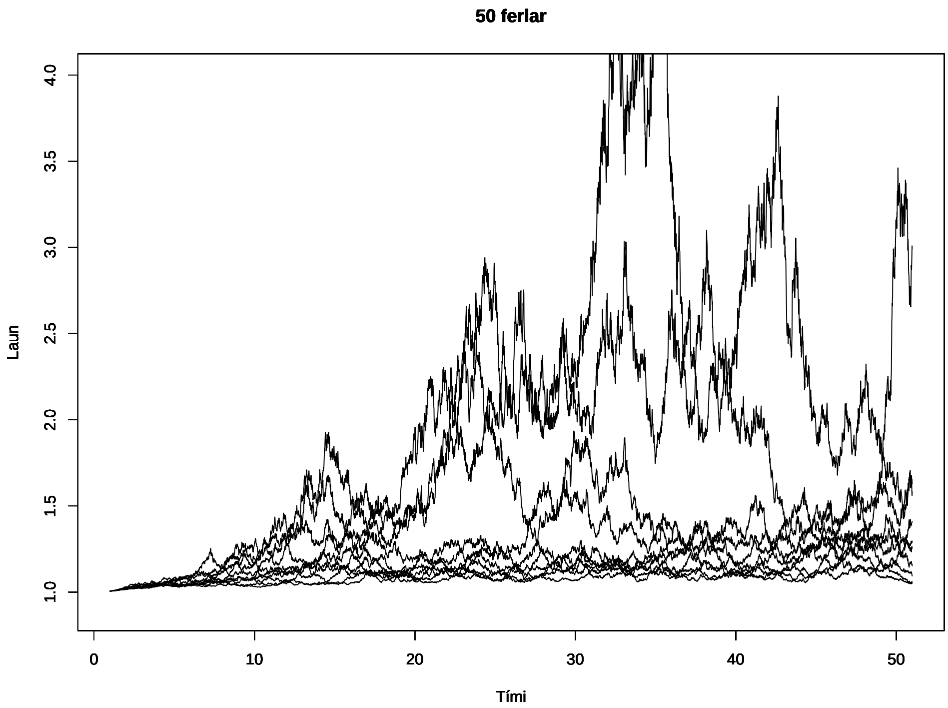
Þegar launaflokkarnir voru sundurslitnir þurfti að ákveða bilið á milli þeirra, t.d. hvort það væri fast í krónum, sama hlutfall væri á milli flokka, eða álíka. Þegar unnið er með laun sem samfellda stærð þarf að ákveða kvörðun, t.d. hvort X sé krónur, eða logaritmi af krónum.

Hér hefur verið rætt um laun. Laun myndast í einhvers konar mannauðsferli. Hægt er að hugsa sér þá einföldun að fyrir einsleitt þýði stýrist mannauðshreyfimynstrið af einhverri útgáfu af slembnu diffurjöfnunni (1). Til dæmis að verðskulduð laun, Laun(t), hreyfist samkvæmt eftirfarandi jöfnu:

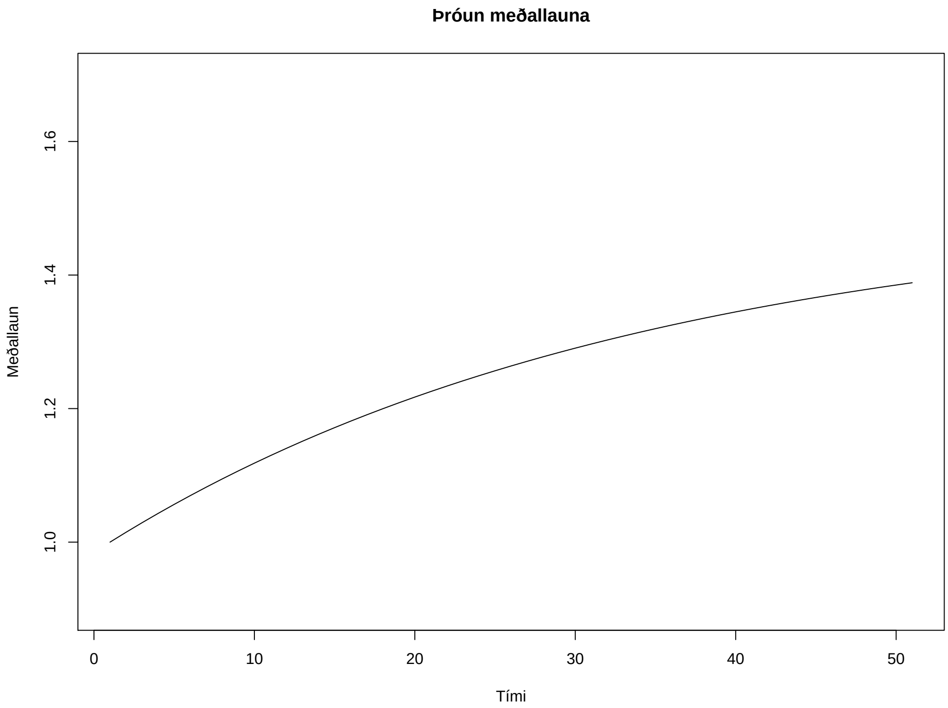
$$\mathrm{Laun}(t) = (1 + X(t))\mathrm{Laun}(0). \tag{2}$$

Þar sem $X(t)$ fylgir CEV hreyfimynstri. CEV-ferlið lýsir í þessu tilfelli hlutfallslegri viðbót á byrjunarlaun. Á mynd 4 eru sýndir mögulegir ferlar fyrir 50 einstaklinga þar sem $X(t)$ fylgir CEV líkaninu úr jöfnu (1) og laun eru fengin með jöfnu (2). Á henni var er $\kappa = 0,03, \alpha = 0,5$, og $\sigma = 0,3$. Byrjunarlaun eru þau sömu ($X(0)$=0, Laun(0)=1,0) hjá öllum og fræðilegt meðaltal jafnvægisdreifingar er 1,5. Væntanleg aðlögun, þ.e. þróun meðallauna samkvæmt þessu er sýnd á mynd 5. Túlkun á $\alpha = 0,5$ er sú að í jafnvægi séu meðallaun 50% ofan á byrjunarlaunin L(0). Ef aðlögunin κ er lág tala þýðir það að jafnvægisdreifingin hefur þyngri hala en ef κ er há og einhverjir eru með mjög hátt álag á byrjunarlaun. Kraftarnir í hreyfimynstrinu eru þannig að sérhver einstaklingur dregst að meðaltalinu α. Ef laun einstaklings hafa tiltekið frávik frá meðaltalinu α er væntanlegur helmingunartími fráviksins $\log(2)/\kappa$. Ef $X(t)$ lýsir virði mannauðs þá hefur þetta líkan þann raunhæfa eiginleika að á hverjum tíma þá eru einhverjir einstaklingar verðmætari en aðrir. Hins vegar hefur þetta líkan þann eiginleika að aðlögunarhraði að jafnvægi er sá sami fyrir há laun og lág laun, og að mjög há laun, sem eru langt frá jafnvægi, lækki hratt.

Til að leggja áherslu á sveigjanleika Markov líkana í samfelldum tíma var hér stuðst við lágan aðlögunarhraða κ, og háan breytileikastuðul, σ. Ginistuðull jafnvægisdreifingarinnar er um það bil 0,2. Jafnvægisdreifingin hefur þungan hala (heavy-tail) og hefur ekki staðalfrávik. Þetta má túlka þannig að gífurlega há gildi eru möguleg en sjaldgæf. Þetta líkan leyfir því „ofurlaun". Mynd 4 gefur til kynna að meðal þessara 50 einstaklinga var einn sem fjórfaldaði byrjunarlaunin á um það bil 30 tímabilum (árum eða mánuðum, o.s.frv.).



Mynd 4: Dæmi um 50 mögulega ferla úr CEV-líkani.



Mynd 5: Þróun meðallauna í CEV-líkani.

Ef til vill er raunhæfara að líta á launin á tilteknum tíma t sem fall af hámarksmannauði upp að tíma t og gera ráð fyrir að miklu erfiðara sé að lækka laun en að hækka þau:

$$\mathrm{Laun}(t) = (1 + \max_{s\in[0,t]} X(s))\mathrm{Laun}(0).$$

Einnig væri hugsanlegt að breyta CEV líkaninu þannig að aðlögunarhraði að jafnvægi, κ, væri minnkandi fall af álaginu $X(t)$ á byrjunarlaun Báðar hugmyndirnar kalla á mun flóknari stærðfræði.

Vinnulaun í markaðshagkerfi myndast í jafnvægi eftirspurnar eftir eiginleikum launþegans og vilja (þarfa) hans til vinnu. Tekjur hvers einstaklings eru samsettar úr launa- og fjármagnstekjum. Ef gert er ráð fyrir að hreyfimynstur launatekna stjórnist af hreyfimynstri mannauðs yfir starfsævina þá er ljóst að hreyfimynstur fjármagnstekna stjórnast af allt öðrum þáttum. Stærðir eins og vextir, söluhagnaður og erfðir eru grundvallarþættir fjármagnstekna. Í frægri grein eftir Wold & Whittle (1957) og kennslbók eftir Whittle (1992) er lýst einföldu dæmi þar sem gert er ráð fyrir að allar tekjur séu fjármagnstekjur og að eignir fari einungis á milli manna með erfðum. Gert er ráð fyrir að allir eigi jafnmarga erfingja og að eina uppspretta ójöfnuðar sé að menn lifi mislengi. Gert er ráð fyrir afar einföldu líftímalíkani, þ.e. að líftími sé veldisdreifður. Líftímaferillinn er því minnislaus. Þetta kerfi er Markov-kerfi og dæmið gengur út á að finna jafnvægisdreifingu tekna. Í þessu tilfelli verður jafnvægisdreifingin Pareto-dreifing, Í kennslubókinni eftir Whittle (1992) er dæmið útvíkkað með því að leyfa breytilegan fjölda erfinga sem er enn önnur uppspretta ójöfnuðar. Í því tilfelli fæst einnig jafnvægisdreifing með þungan hala. Það er því rök fyrir því að vegna tölfræðilegra lögmála þá muni alltaf verða til einstaklingar sem séu „mjög" ríkir. Í kennslubókinni eftir Whittle (1992) er nefnt að halar í fjármagstekjudreifingum séu yfirleitt þyngri en halar í launadreifingum. Líklegt má telja að þeir einstaklingar sem njóta hárra fjármagnstekna hafi aðrar þarfir og annan vilja til launaðrar vinnu.

Lokaorð

Hér hafa verið rakin nokkur atriði sem tengjast greiningu á launa- og tekjudreifingum. Aðeins mjög einföld líkön hafa verið reifuð. Sú athugun sýnir að viðfangsefnið er flókið og mikilvægt að taka tillit til ýmissa truflandi þátta, svo sem stærðar kynslóða, hreyfimynsturs einstaklinga o.s.frv. Mikilvægt er að val á líkani henti rannsóknarspurningu í hverju tilfelli fyrir sig. Þegar unnið er með gögn um fólk, t.d. launagögn, þarf að huga að því hvernig laun verða til. Einstaklingur kemur út á vinnumarkað á tilteknum aldri og eldist þar eftir því sem tíminn líður. Ljóst er því að taka verður tillit til lýðfræðilegra þátta þegar fylgja á þróun í tíma. Gera þarf glöggan mun á fræðilegum dreifingum og úrtaksdreifingu. Gögn fást með úrtaki og ef reikna á úrtaksstærðir, t.d. úrtaks Gini-stuðul þá þarf að huga að úrtaksdreifingu hans. Jafnvel þó að fræðilega dreifingin væri þekkt er hætta á að gögn séu blanda úr raunverulegri dreifingu og einhverri annarri truflandi dreifingu, t.d. mæli eða flokkunarskekkju. Eitt af vandamálum við notkun Gini-stuðulsins er að hann er viðkvæmur fyrir „mengun" (contamination) í gögnum. Með mengun er átt við að annarri dreifingu er blandað við dreifinguna sem verið er að rannsaka. Þetta getur verið mæliskekkja eða einhvers konar kerfisbundin atriði, t.d. að launin séu kvörðuð í bil og t.d. hæsti flokkurinn hafður opinn. Victoria-Feser (2000) varar við notkun Gini-stuðuls í fjölþjóðlegum samanburði og stingur upp á aðferðum sem eru þolnari (robust).

Ef finna á mat á þýðingu einstakra skýristærða er nauðsynlegt að tekið sé tillit til allra mikilvægra skýristærða samtímis. Ef slíkt er ekki gert er hætt við að villandi ályktanir verði dregnar. Alþýðleg dæmi um þetta eru til dæmis rakin í grein Tómasson (2005). Þegar þróun tekjudreifinga er skoðuð er nauðsynlegt að tekið sé tillit til hreyfimynsturs einstaklinga, hvernig þýði er samsett og hvaða tekjuhugtök sé verið að vinna með. Æskilegt er að nota aðferðir sem bjóða upp á sundurliðanleika, t.d. að hægt sé að skoða mismunandi hópa og mismunandi tegundir tekna. Í nýlegri grein reyna Daly & Valletta (2006) að meta hluta lýðfræðilegra þátta í breytingu á tekjudreifingu í Bandaríkjunum á tímabilinu 1969-1998. Þeir sýna myndir af metnum dreifingum og álykta að ef lýðfræðileg samsetning Bandaríkjanna 1998 væri eins og hún var árið 1969 væri tekjudreifingin 1998 mjög svipuð og hún var á árinu 1969. Þ.e. að meginskýring á ójafnari tekjudreifingu á árinu 1998 væri sú að meira af miðaldra fólki væri á vinnumarkaði, fjölskyldur væru barnfærri, o.s.frv. Auk lýðfræðilegra þátta, þarf að taka tillit til hegðunarþátta eins og til dæmis hlutverki hjónabands. Ef aðilar eru í hjónabandi þá er hugsanlegt að þeir myndi með sér einhvers konar efnahagsbandalag þannig að annar aðilinn er tilbúinn að minnka sínar tekjur í þeim tilgangi að heildartekjur parsins aukist. Hlutir eins og einstaklingshæfni og uppsöfnuð reynsla eru atriði sem eru metin á vinnumarkaði en erfitt er að mæla með beinum hætti.

Nauðsynlegt er að huga að launa- og tekjuhugtakinu. Er verið að tala um árstekjur eða ævitekjur? Ef einstaklingur fer í langskólanám styttist vinnuævin. Þó að engar tekjur sé bókfærðar á námstíma eru væntanlegar tekjur að loknu námi hærri en ella. Því má spyrja hvort ekki þyrfti að jafna ævitekjunum niður á námsárin. Slíkt myndi augljóslega jafna tekjudreifinguna. Á sama hátt er ekki sjálfgefið á hvaða tímapunkti ætti að bókfæra arf eða söluhagnað. Ef fyrirtæki hefur verið byggt upp í áratugi og síðan selt, hefur það einhvern aðdraganda og aðilar er farnir að nota söluvirðið áður en það er bókfært. Á síðustu tveim áratugum hafa verðbréfamarkaðir eflst í Evrópu og víðar. Gamalgróin fyrirtæki hafa verið seld, upprunalegir stofnendur og/eða erfingjar þeirra, hafa fundið leið út úr þeim og aðrir hafa getað keypt sig inn. Á þessu tímabili hafa fjármagnstekjur því verið áberandi í bókhaldi. Í öllum greiningum á tekjudreifingu þarf að átta sig á eðli tekjuþátta. Eðlilegt er að gera ráð fyrir að hreyfimynstur fjármagnstekna sé annað en launatekna og því nauðsynlegt að nota líkön sem aðgreina fjármagnstekjur og launatekjur.

Heimildir

Bhattacharya, R. & Waymire, E. (1990). *Stochastic Processes with Applications.* John Wiley & Sons.

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & E. G N. Wilkinson (Eds.), *Robustness in Statistics.* Akademic Press: New York.

Champernowne, D. (1953). A model for income distribution. *Economic Journal*, *53*, 318–351.

Cowell, F. & Victoria-Feser, M.-P. (1996). Robustness properties of inequality measures. *Econometrica*, *64*, 77–101.

Daly, M. C. & Valletta, R. G. (2006). Inequality and poverty in United States: The effects of rising dispersion of men's earnings and changing family behaviour. *Economica*, *73*, 75–98.

Dragulescu, A. & Yakovenko, V. (2001). Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A: Statistical Mechanics and its Applications*, *299*(1-2), 213–221.

Dragulescu, A. & Yakovenko, V. (2003). *Statistical Mechanics of Money, Income and Wealth: A Short Survey.* Seventh Granada Lectures.

Draper, N. & Smith, H. (1966). *Applied Regression Analysis.* John Wiley & Sons.

Tómasson, H. (1980). Inkomstförändringar för anställda män 1963-1973: Markov analys. Seminar paper, Göteborgs Universitet.

Tómasson, H. (2005). Tölfræðigildrur og launamunur kynja. Þjóðmál, 1. árg, 2. hefti.

Victoria-Feser, M.-P. (2000). Robust estimation for the analysis of income distribution, inequality and poverty. *International Statistical Review*, *68*(3), 277–293.

Whittle, P. (1992). *Probability via Expectation* (Third ed.). Springer-Verlag.

Wold, H. & Whittle, P. (1957). A model explaining the Pareto distribution of wealth. *Econometrica*, *25*, 591–595.