Bayesísk tölfræði og raðkvarðar

Helgi Tómasson

8. október 2008

1 Inngangur

Markmið þessarar greinar er í fyrsta lagi að gera grein fyrir aðferðafræði við mælingar, sem eru raðtölur. Í öðru lagi að lýsa hvernig megi hagnýta bayesískar aðferðir við mat á líkönum fyrir raðtölur.

Í mörgum vísindagreinum, mikið innan félagsvísinda, eru notaðir raðkvarðar (ordinal scales). Í markaðskönnunum oft óskað eftir því að svarandi kvarði svar sitt, til dæmis með tölunum 1,2,3,4 og 5. Einnig gefa sumir sérfræðingar kvörðuð svör, t.d. fyrsta, annað, eða þriðja stigs bruni. Tölfræðileg líkön fyrir greiningu slíkra gagna byggja á raðaðri aðhvarfsgreiningu (ordinal-regression). Í slíkri aðhvarfsgreiningu er háða breytan heiltala og gert er ráð fyrir röðun, þ.e. 2 er stærri en 1. Vel þekktar útfærslur á slíkri aðhvarfsgreiningu eru notast svið cumulative-logits, proportional-odds, ordered-probit o.s.frv. Sjá nánar í Agresti (2002). Ef sami einstaklingur er látin gefa mörg röðuð svör opnast möguleiki á margvíðri greiningu og þar með að bera einstaklinginn saman við sjálfan sig. Margvítt líkan er eðli málsins samkvæmt flóknara en einvítt og því verða túlkanir og mat erfiðara.

Bayesísk tölfræði gengur út frá því að vísindamaður setji fram fordóma sína á líkindafræðilegu formi og álykti síðan með reglu Bayes. Hugmyndafræðin er önnur en í hefðbundinni tölfræði. Tölulegar útkomur eru oft svipaðar. Klassískar bækur um bayesíska tölfræði eru til dæmis Zellner (1971) og Box & Tiao (1973). Nútíma Bayesísk tölfræði er afar reiknifrek og byggir meðal annars á þróuðum hermunaraðferðum svo sem MCMC. Þær aðferðir bjóða upp á að meta líkön sem eru illmetanleg með hefðbundnum aðferðum. Þetta tæknilega atriði hefur gert það að verkum að margir hefðbundnir tölfræðingar notað slíkar aðferðir við flókin líkön. Baysískar aðferðir eru því í tísku og hafa ratað inn í kennslubækur svo sem Koop (2003) og Lancaster (2004). Einnig hafa komið fram hreinar aðferðafræðibækur eins og Gelman, Carlin, Stern & Rubin (2004) og Canova (2007). Gerð er lauslega grein fyrir því hvernig bayesísk tölfræði er byggð upp.

Í allri gagnavinnslu er gengið út frá líkani. Það er mjög mikilvægt að átta sig á því að nauðsynlegt er að skilgreina líkan þannig að allir mikilvægir þættir séu með. Annars er nánast öruggt að ályktanir verða villandi. Þetta eru alþekkt sannindi sem ekki virðast áréttuð nægilega oft. Það að einstaklingar nota kvarða á mismunandi hátt getur truflað ályktanir. Notkun margvíðra líkana opna möguleika á að leiðrétta fyrir því að einstaklingar eru mismunandi. Sýnd eru dæmi hvernig herma má gögn og meta með R-tölfræðiforritinu (R Development Core Team, 2005). Einnig er sýnt tölulegt dæmi sem byggir á svarendum í kennslukönnun Háskóla Íslands.

2 Grunnatriði bayesískrar tölfræði

Tilgangur tölfræði er að álykta/alhæfa út frá mælingum. Tölfræði byggir á líkindafræði sem er hrein stærðfræðigrein. Sem slík er líkindafræðin algerlega „exact" vísindi. Um aldamótin 1800 klofnar hagnýting líkindafræðinnar í tvennt sem annars vegar ákvörðunarfræði (decision theory) og hins vegar mælingafræði/metría, sem er oft kölluð tölfræði (statistics). Sumir nota einnig orðið tölfræði yfir talnasafn eða heiti á gagnabönkum sem innihalda staðreyndir skráðar í talnaformi.

Sem stærðfræði eru líkur mælikvarði á stærð mengja sem oft eru kölluð atburðir. Þegar búið er að setja fram viðeigandi skilgreiningar er líkindamálið fall sem úthlutar atburði tölu á bilinu [0,1]. Hvað sú tala þýðir er hins vegar túlkunaratriði. Í tölfræði hafa þróast tvær megin túlkanir á þessari tölu. Annars vegar „tíðnitúlkun (frequentistic)", sem gengur út á að túlka eigi líkur á atburði sem hlutfallslega tíðni atburðarins í endurteknum tilraunum. Hins vegar er túlkun sem gengur út á að einnig megi túlka líkur sem mælikvarða á vissu. Á 20. öld hefur tíðnitúlkunin verið hin ríkjandi túlkun. Það að túlka líkur sem mælikvarða á vissu er nefnt „bayesísk" tölfræði og hefur mestan hluta 20. aldar verið eins konar sértrúarsöfnuður. Bayesísk tölfræði er kennd við Thomas Bayes sem uppi var á 18. öld og setti fram reglu:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{1}$$

Jafna (1) er kölluð regla Bayes og táknar líkur á atburðinum A gefið atburðurinn B sem fall af líkunum á atburðinum B gefið atburðurinn A. Þessi jafna var grunnur að því sem Laplace kallaði „inverse-probability". Stiegler (1986) rekur hugsunargang Laplace.

Í allri tölfræði er gengið út frá líkani. Þ.e. að gert er ráð fyrir að til sé satt líkan sem hafi búið til mælingarnar. Líkanið sé að mestu þekkt en að mælingar séu notaðar til að giska á óþekkta hluta líkansins. Líkanið er til dæmis sett fram sem þéttifall (density function):

$$f(\boldsymbol{x}|\boldsymbol{\theta}),$$

þar sem $\boldsymbol{x}$ er hugsanlegt gildi mælingar og $\boldsymbol{\theta}$ er stiki (parameter) sem stýrir líkindadreifingunni. Ágiskunin um stikann $\boldsymbol{\theta}$ er nefnt mat (estimate). Tíðnitúlkun gengur út frá úrtaksfræði (sampling theory), þar sem reiknuð er út dreifing falls af hendingum, sem gætu haft fengnar mælingar sem útkomu.

Vinnubrögðin í bayesískri tölfræði eru, að sett er fram líkan, $f(\boldsymbol{x}|\boldsymbol{\theta})$, fyrir eiginleika hendingarinnar $\boldsymbol{X}$. Fyrirframvissan (óvissan) er sett fram með „a priori"dreifingunni:

$$\pi(\boldsymbol{\theta}).$$

Mælingum er safnað í vektorinn $\boldsymbol{x} = (x_1, \ldots, x_n)$ og sennileikafallið (likelihood function), $l(\boldsymbol{\theta}|\boldsymbol{x})$ er reiknað. Eftirávissan er fundin með reglu Bayes sett fram sem „a posteriori"dreifingin:

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{l(\boldsymbol{x}|\boldsymbol{\theta})}{m(\boldsymbol{x})} \tag{2}$$

$$m(\boldsymbol{x}) = \int l(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \tag{3}$$

Nefnari í jöfnu (2), jafna (3), tryggir að fallið $\pi(\theta|\boldsymbol{x})$ er vel skilgreind líkindadreifing. Það er einungis fyrir vissar samsetningar á líkani og fyrirframvissu sem eftirávissan verður þægileg formúla. Fyrir slíkt par af dreifingum er talað um „conjugate prior". Hugsanlegt Bayes-mat á $\boldsymbol{\theta}$ er til dæmis:

$$\hat{\boldsymbol{\theta}}_{Bayes} = E_{\pi(\boldsymbol{\theta}|\boldsymbol{x})}(\boldsymbol{\theta})) = \int \pi(\boldsymbol{\theta}|\boldsymbol{x})\,\mathrm{d}\boldsymbol{\theta} \quad \text{væntanlegt gildi eða} \tag{4}$$

$$\tilde{\boldsymbol{\theta}}_{Bayes} = \arg\max_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta}|\boldsymbol{x}) \tag{5}$$

Væntanlega gildið í (4) og hámarkið (5)er einungis til sem þægileg formúlu fyrir nokkur sértilfelli. Til dæmis ef líkanið er:

$$X|\mu,\sigma \sim N(\mu,\sigma^2) \quad \text{og a priori}$$

$$\pi(\mu) \sim N(a,\tau) \quad \sigma \text{ þekkt, og gögn úr slembiúrtaki}$$

$$x_1,\ldots,x_n \quad \text{þá er}$$

$$\pi(\mu|x_1,\ldots,x_n) \sim N(\bar{x}\frac{\tau}{\sigma^2/n+\tau} + a\frac{\sigma^2/n}{\sigma^2/n+\tau}, \frac{1}{n/\sigma^2+1/\tau})$$

Meðaltal eftirávissunnar er því vegið meðaltal af úrtaksmeðaltali og meðaltali fyrirframvissunnar. Ljóst er að ef n er stórt þá er útkoman mjög lík því sem fæst með hefðbundnum aðferðum, þ.e. $\bar{x}$ er eðlilegt ágiskun μ.

Einfaldar formúlur fyrir $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ eru aðeins þekktar fyrir nokkur einföld dæmi. Í öðrum tilfellum verður að treysta á tölulegar (númerískar) aðferðir. Hermanir (simulation) eru ráðandi í hagnýtingu á Bayes aðferðum. Forrit sem velja slembitölur (pseudo-random), $\boldsymbol{\theta}_l, \quad l = 1,\ldots,L$ og síðan er t.d væntanlegt gildi $\boldsymbol{\theta}$.

$$E_{\pi(\boldsymbol{\theta}|\boldsymbol{x})}(\boldsymbol{\theta}) = \int \boldsymbol{\theta}\pi(\boldsymbol{\theta}|\boldsymbol{x})\,\mathrm{d}\boldsymbol{\theta}, \qquad \text{metið með}$$

$$\frac{1}{L} = \sum_{l=1}^{L} \boldsymbol{\theta}_l$$

Þar sem $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ er yfirleitt flókin dreifing þá er ekki einfalt að gera forrit sem velja slembitölur úr henni. Lausnin er því oft á því formi að hönnuð er Markov-keðja sem hefur $\pi(\boldsymbol{\theta}|\boldsymbol{x})$, sem jafnvægisdreifningu. Tölulega vinnan er að herma Markov-keðjuna og er nefnd MCMC=Markov-Chain-Monte-Carlo. Mikið notaðar útfærslur af MCMC aðferðum eru Gibbs sampling og Metropolis-Hastings. Gangurinn er sá að keðjan er sett af stað, $\boldsymbol{\theta}_j$ valin og síðan er treyst á að eftir ákveðin tíma (burn-in) sé dreifing hermuðu gildanna orðin nægilega nálægt jafnvægisdreifingunni.

Til að skilja Gibbs sampling er hægt að hugsa sér normal líkan, $N(\mu, \sigma^2)$, og mælingar sem hafa fengist með slembiúrtaki, $\boldsymbol{x} = (x_1, \ldots, x_n)'$. Stikinn er hér tvívíður $\boldsymbol{\theta} = (\mu, \sigma)'$. Þá er sennileikafallið:

$$l(\boldsymbol{\theta}|\boldsymbol{x}) \propto \frac{\sqrt{n}}{\sigma} \exp(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}) \tag{6}$$

Ef litið er á hægri hliðina í (6) sem fall af μ þá er það í hlutfalli við þéttifall normaldreifingar (í μ). Bayesistar líta á $\boldsymbol{\theta}$ sem hendingu, sem hefur dreifingu sem lýsir vissunni um stikann $\boldsymbol{\theta}$. Ef $\bar{x}$ og σ eru þekkt þá má velja slembitölu úr normaldreifingunni sem hægri hliðin í (6) lýsir. Þá tölu má nota sem ágiskun á μ. Hægri hlið (6) sem fall af $1/\sigma^2$ er einnig í hlutfalli við þekkta líkindadreifingu, þ.e. gamma dreifingu. Ef $\bar{x}$ og μ eru þekkt má velja slembitölu úr þeirri gammadreifingu og nota útkomuna sem ágiskun á $1/\sigma^2$. Þetta er síðan endurtekið þangað til að jafnvægi er náð. Venjan er að margfalda (6) með fyrirframvissu, $\pi(\boldsymbol{\theta})$. Ef $\pi(\boldsymbol{\theta})$ er á heppilegu formi, þ.e. inniheldur þætti eins og $(1/\sigma^2)^\alpha \exp(-\beta/\sigma^2)$ og $\exp(-(\mu-a)^2/)/(2\tau^2)$, þá verða útreikningarnir að ofan einfaldir.

Aðferðin er mjög einföld:

1. Set $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0)$.
2. Bý til nýtt gildi á μ. Auðvelt. því $\mu|\sigma, \boldsymbol{x}$ er normal.
3. Nota það til að búa til nýtt gildi á σ. Auðvelt, því $1/\sigma^2|\mu, \boldsymbol{x}$ er gamma.
4. Bý til nýtt gildi á μ o.s.frv.

Í þessu tilfelli er notaður normal fyrirframvissa um μ og IG(inverse-gamma) fyrirframvissa um σ^2. Ef líkanið er svona einfalt eru til formúlur og MCMC aðferðir óþarfar. Í kennslubókum er stundum notað að $1/\sigma^2$ sé χ^2 dreift. Það er að sjálfsögðu það sama því að χ^2 dreifingin er sértilfelli af gammadreifingu. Oft er hægt að setja upp hluta líkans á þennan hátt og nota síðan hliðstæðar aðferðir við annan hluta.

3 Nauðsynlegt er að leiðrétta fyrir öllum þáttum

Það er vel þekkt að einstaklingar nota kvarða á mismunandi hátt. Nauðsynlegt er að leiðrétta fyrir því að einstaklingar eru mismunandi. Aðalverkefni líkanasmiðs er því að hanna líkan sem sigtar frá einstaklingsbundinn breytileika. Í allri tölfræðivinnu er nauðsynlegt að leiðrétta fyrir truflandi þáttum. Í kennslubókum er þetta stundum sýnt með því að láta nemendur sýna það að ef mikilvægri skýribreytu í aðhvarfslíkani er sleppt þá muni leiða til bjagaðs mata á þýðingu annar skýristærða. Þ.e. ef sanna líkanið er:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

og metið líkan er:

$$y = \alpha_0 + \alpha_1 x_1 + \epsilon.$$

Þá fær nemandinn það verkefni, að meta hversu villandi mynd metið α_1 gefur á sannri þýðingu breytunnar x_1, þ.e. β_1. Þessi gerð villu er útbreidd í daglegri tölfræðivinnu og hefur höfundur varað við henni bæði í kennslu í tölfræði og í riti. Sumir halda til dæmis að það halli á konur í launum á vinnumarkaði og draga fram niðurstöður aðhvarfsgreiningar máli sínu til stuðnings. Höfundur hefur í grein (Tómasson, 2005) lýst hvernig slíkt getur gefið villandi ályktanir. Tölfur 1 og 2 eru 2x2 töflur sem gefa villandi ályktanir. Þær báðar í raun útkoma út aðhvarfslíkani þar sem mikilvægri skýribreytu hefur verið sleppt. Þessar töflur eru hluti kennslubókardæmis þar sem sanna líkanið, sem inniheldur margar skýristærðir, er látið vera konum í hag en allar 2x2 töflur gefa til kynna að kerfið sé körlum hagstætt. Sama á við ef x-breyta er mæld með mæliskekkju því að mæliskekkjan er ómæld breyta sem vantar í sanna líkanið.

	starf=1	starf=2
karlar	154.000	241.429
konur	126.667	200.000

Tafla 1: Meðallaun eftir starfi og kyni

	aldur=1	aldur=2
karlar	145.000	235.000
konur	122.500	190.000

Tafla 2: Meðallaun eftir aldri og kyni

	Há laun	Lág laun
Karlar	18	12
Konur	7	3

Tafla 3: Launadreifing í fyrirtæki A. 70% kvenna með há laun, 60% karla með há laun.

	Há laun	Lág laun
Karlar	2	8
Konur	9	21

Tafla 4: Launadreifing í fyrirtæki B. 30% kvenna með há laun, 20% karla með há laun.

Það að leggja saman mismunandi hóp er önnur algeng villa. Í áðurnefndri grein er tekið einfalt dæmi það að varasamt sé að slá saman mismunandi hópum. Í töflum 3 til 5 er sýnt að hugsanlegt er að tvö fyrirtæki mismuni körlum í óhag en þegar þau eru lögð saman þá virðist summan mismuna konum í óhag. Almennt gildir um aðvharfslíkön að allar mikilvægar breytur verða að vera með, bæði áhugaverðar og óáhugaverðar. Gera verður áætlun um hvernig varast beri áhrif ómældra miklivægra breyta og þýðið sem rannsaka á verður að hafa ákveðna einsleitni (homegeneity).

	Há laun	Lág laun
Karlar	20	20
Konur	16	24

Tafla 5: Launadreifing í fyrirtækjum A+B. 40% kvenna með há laun, 50% karla með há laun.

4 Nokkur orð um líkön fyrir raðkvarða

Aðhvarfslíkan er grundvallarhugtak í hagnýtri tölfræðivinnu. Í hefðbundnu aðhvarfslíkani er háða breytan samfelld. Með aukinni tölvutækni hafa útfærslur það sem háða breytan tekur t.d. 0 eða 1 (logit/probit), eða talningarbreyta (0,1,2, ...), orðið viðráðanlegar. Ein útvíkkunin er sú að háða breytan taki gildi á röðuðum kvarða, 1,2,3. Með röðuðum kvarða (ordinal scale) er átt við að gildin hafa eðlilega stærðarröðun, góður, betri, bestur. Ýmsar útfærslur er til af líkanasmíði fyrir raðaðra breyta. Einföld hugmynd er að útvíkka logit/probit formið. Logit/probit líkan skýrir breytu, Y, sem einungis getur tekið gildið 0 eða 1. Þ.e. að skýra líkurnar:

$$\begin{aligned}
&P(Y=1|\boldsymbol{X}=\boldsymbol{x}) = 1 - P(Y=0|\boldsymbol{X}=\boldsymbol{x}) = F(\boldsymbol{X\beta}),\\
&F(z) = \frac{e^z}{1+e^z} \quad \text{logit, eða}\\
&F(z) = \Phi(z) \quad \text{probit, dreififall staðlaðrar normalhendingar.}
\end{aligned}$$

Cumulative-logit hugmyndin er að skrifa:

$$P(Y \le y|\boldsymbol{X}=\boldsymbol{x}) = p_1(\boldsymbol{x}) + \cdots + p_y(\boldsymbol{x}), \qquad \text{þannig að} \tag{7}$$

$$\log \frac{P(Y \le y|\boldsymbol{X}=\boldsymbol{x})}{1 - P(Y \le y|\boldsymbol{X}=\boldsymbol{x})} = \frac{p_1(\boldsymbol{x}) + \cdots + p_y(\boldsymbol{x})}{p_{y+1} + \cdots + p_k(\boldsymbol{x})} \tag{8}$$

Líkanið í jöfnu (7) býður upp á þann möguleika að áhrif $\boldsymbol{X}$-breytunnar á að hoppa á milli flokka séu mismunandi eftir flokkum, þ.e. að allur $\boldsymbol{\beta}$-vektorinn sé háður flokki. Það að einungis skurðpunkturinn sé háður flokki er nefnt proportional-odds líkan.

$$\log \frac{P(Y \le y|\boldsymbol{X}=\boldsymbol{x})}{1 - P(Y \le y|\boldsymbol{X}=\boldsymbol{x})} = \alpha_j + \boldsymbol{\beta x}$$

Vinsæl hugmynd er að mælda breytan, y, sé klippt útgáfa af samfelldri ómældri (latent) breytu $y*$. Ef breytna y tekur gildi frá 1 til k þá,

$$y = \begin{cases} 1 & y^* < c_1 \\ 2 & c_1 < y^* < c_2 \\ \vdots & \vdots \\ k & c_{k-1} < y^* \end{cases}.$$

Punktarnir $c_1, \ldots, c_{k-1}$ eru þröskuldar í breytunni y^* sem gefa röðuðu breytunni y gildi. Breytan y^* er ekki mæld en gert er ráð fyrir hún fylgi einhverri dreifingu, t.d.:

$$y^* \sim N(0,1)$$

Hægt er að bæta við skýribreytum:

$$y^* = \boldsymbol{X\beta} + \varepsilon \quad \varepsilon \sim N(0,1)$$

Ef fyrir hendi eru gögn á forminu:

$$\begin{array}{ccccc} y_1 & x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots & \vdots & \\ y_n & x_{n1} & x_{n2} & \cdots & x_{np} \end{array}$$

þá er vandinn að meta $\boldsymbol{\beta}$ og $(c_1, \cdots, c_{k-1})$. Líkanið gefur þá mat á líkunum,

$$P(y \leq j | \boldsymbol{X} = \boldsymbol{x}) = \Phi(c_j - \boldsymbol{x\beta}), \tag{9}$$

þar sem Φ er dreififall staðlaðrar normalhendingar. Hugmyndin er að mældi kvarðinn sé klippt útgáfa af ómældri breytu, y_i^*. Ljóst er að flokkunin helst óbreytt ef þröskuldarnir/skurðpunktarnir, $(c_1, \cdots, c_{k-1})$, og y_i^* er margfölduð með fasta. Til að stikar sé greinanlegir (identifiable) þarf því einhvers konar skorður. Dæmi um slíkar skorður er að setja staðalfrávik $\boldsymbol{y}^*$ jafnt 1. Jafna (9) sýnir að útreikningur sennileikafallsins (likelihood function) er einfaldur og því hægt að finna mat mesta sennileika (ML-mat) á $\boldsymbol{\beta}$ og $(c_1, \cdots, c_{k-1})$.

5 Margvíð líkön

Með hugtakinu margvítt líkan (multivariate model) er átt við að fyrir hvern einstakling eru mældar margar breytur, t.d. hæð og þyngd. Þegar verið er að mæla fólk er oft nauðsynlegt að leiðrétta fyrir til kyni, aldri og hugsanlega fleiri breytum. Algengt form er margvíð normal aðhvarfslíkan:

$$\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ip})' = \boldsymbol{x}_i \boldsymbol{B} + \boldsymbol{E}_i$$
$$\boldsymbol{B} = \begin{bmatrix} \beta_{11} & \cdots & \cdots & \beta_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{k1} & \cdots & \cdots & \beta_{kp} \end{bmatrix}$$
$$V(\boldsymbol{E}_i) = V((\varepsilon_{i1}, \ldots, \varepsilon_{ip})') = \Sigma =$$
$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \vdots & \ddots & & \vdots \\ \sigma_{1p} & \cdots & \cdots & \sigma_{pp} \end{bmatrix}$$

Hér er $\boldsymbol{y}_i$ vektor af breytum sem skýra á hjá einstaklingi i. Skýristærðirnar eru í vektornum $\boldsymbol{x}_i$. Samdreifnistikanir (covariance), σ_{ij} í Σ fylkinu lýsa tengslum hnita i og j í mælda vektornum þegar leiðrétt hefur verið fyrir $\boldsymbol{x}_i$. Ef engin hliðarskilyrði eru á stikafylkinu $\boldsymbol{B}$ þá má meta p einvíðar aðhvarfsjöfnur og síðan meta Σ með:

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_j' \tag{10}$$

Margvítt, línulegt normal líkan er algengt tól til greiningar á flóknum tengslum. Forsendur eru að gefið vektorinn $\boldsymbol{x}_i$ þá fylgi vektorinn $\boldsymbol{y}_i$ margvíðri normaldreifingu.

Ef hnit í mældum vektor $\boldsymbol{y}_i$ samanstanda af gildum af raðkvörðum þá er breytan ekki samfelld og má ætla að normaldreifingin sé ekki góð nálgun. Vel má hugsa sér svipaða nálgun og í einvíða tilfellinu, þ.e. að raðkvarðinn sé klippt útgáfa að ómældri undirliggjandi breytu $\boldsymbol{y}^*$. Í slíku líkani yrðu þá þröskuldar í hverri hnit. Eins og í einvíða tilfellinu þarf að setja skorður (hliðstætt því að setja staðalfrávik ómældu stærðarinnar jafnt 1). Margvítt líkan er flóknara en einvítt og koma margar útfærslur af slíkum skorðum til greina. Mörgum mismunandi útfærslum er lýst af t.d. Agresti (2002), Congdon (2005) og Rossi, Allenby & McCulloch (2005). Þar er gengið út frá því að einstaklingur i svari J spurningum. Svör einstaklings i við spurningu j eru táknuð með breytunni y_{ij}, sem tekur gildi frá 1 upp í k.

Hugmyndin er að hvert hnit í y_i sé skorin útgáfa af y_i^* sem fylgi margvíðri normaldreifingu:

$$y_i^* \sim N(\mu_i^*, \Sigma_i^*).$$

Líkanið leyfir því mismunandi kvarðanotkun einstaklinga með því að vera með einstaklingsbundin μ_i^* og Σ^*. Að sjálfsögðu þarf ýmis hliðarskilyrði til að hægt sé að meta þessa stika. Útfærslan hjá Rossi, Gilula & Allenby (2001) gerir t.d. ráð fyrir að Σ_i^* sé á forminu $\sigma_i\Sigma$, þ.e. að fylgni milli einstakra hnita sé sú sama hjá öllum einstaklingum og breytileiki samdreifnifylkisins milli einstaklinga sé eingöngu í formi skrefstærðar. Meðaltalið μ_i^* er valið sem fasti fyrir hvern einstakling og ekki háð skýristærðum. Ljóst er að tæknilega er hægt að bæta við skýristærðum. Rossi, Gilula & Allenby (2001) leggja til bayesíska nálgun á tölulegu mati á svona líkani. Jafnvel þó að hægt væri að skrifa niður sennileikafallið er ljóst að í mörgum tilfellum gæti verið erfitt að hámarka það. Með því að velja veika fyrirframvissu fæst mat þar sem eiginleikar gagna verða ráðandi.

	Númer spurningar					
	1	2	3	4	5	6
Einstaklingur 1	4	5	4	5	4	5
Einstaklingur 2	2	3	2	3	2	5
Einstaklingur 3	1	5	1	5	1	5

Tafla 6: Sammála einstaklingar.

Í töflu 6 eru sýnd svör þriggja einstaklinga við 6 ímynduðum spurningum. Þessir einstaklingar eru allir sammála um að gefa hærri tölu við spurningu númer 2 en númer 1, hærri tölu við spurningu númer 4 en númer 3, o.s.frv. Hins vegar er ljóst að einstaklingarnir nota kvarðann á ólíkan hátt. Einstaklingur 2 er hliðruð útgáfa af einstaklingi 1 en einstaklingur 3 tekur stærri skref á sínum kvarða. Margvíð líkön bjóða upp á þann möguleika að leiðrétta fyrir því að einstaklingar nota kvarðann á mismunandi hátt og greina einstaklingshreinsað mynstur í sambandi breytanna sem mynda hnitir í mælivektornum.

6 Einfalt einvítt dæmi

Til að sýna að hvernig nota má nýtísku Bayes-aðferðafræði við útskýringu á raðkvörðum er hér tekið einfalt sýnidæmi úr R-tölfræðipakkanum R Development Core Team (2005). Gert er ráð fyrir að tvær mældar skýristærðir séu, X_1 og X_2. Gert er ráð fyrir að sanna líkanið sé á forminu:

$$y_i^* = \beta_0 + \beta_1 x_{i1} + \beta_{i2} x_{i2} + \varepsilon_i \quad \varepsilon_i \sim N(0,1)$$
$$y_i = \begin{cases} 1 & -\infty < y_i^* < c_1 \\ 2 & c_1 < y_i^* < c_2 \\ \vdots & \vdots \\ 5 & c_4 < y_i^* < \infty \end{cases}$$

Skýristærðirnar X_1 og X_2 eru jafndreifðar $U(0,5)$. Vektorinn $\boldsymbol{\beta} = (0.5, 1, -0.5)$. Gert er ráð fyrir 300 mælingum og að y sé kvörðuð, 1,2,3,4,5. Gögn gætu til dæmis litið út eins og sýnt er í töflu 7. Hægt er að meta líkanið með hefðbundnum ML aðferð. Niðurstöðurnar eru sýndar í töflu 8. Með því að byrja með mjög veikar fyrirfram upplýsingar fæst bayesískt mat með Gibbs-MCMC aðferð. Í töflu 9 er sýnt úrtaksmeðaltal og staðalfrávik Markov-keðjunnar. Reiknuð voru 2000 skref og fyrstu 200 hent. Ljóst er að tölulega eru niðurstöðurnar í töflum 8 og 9 svipaðar. Slíks er að vænta þegar ML aðferð virkar og byrjað með veikar fyrirfram upplýsingar. Í allri vinnu með líkön þarf að framkvæma einhvers kon-

i	y	x_1	x_2
1	5	4.95	2.26
2	4	4.15	2.66
$\vdots$	$\vdots$	$\vdots$	$\vdots$
299	2	2.42	4.31
300	1	1.14	1.00

Tafla 7: Dæmi um gögn í einföldu dæmi.

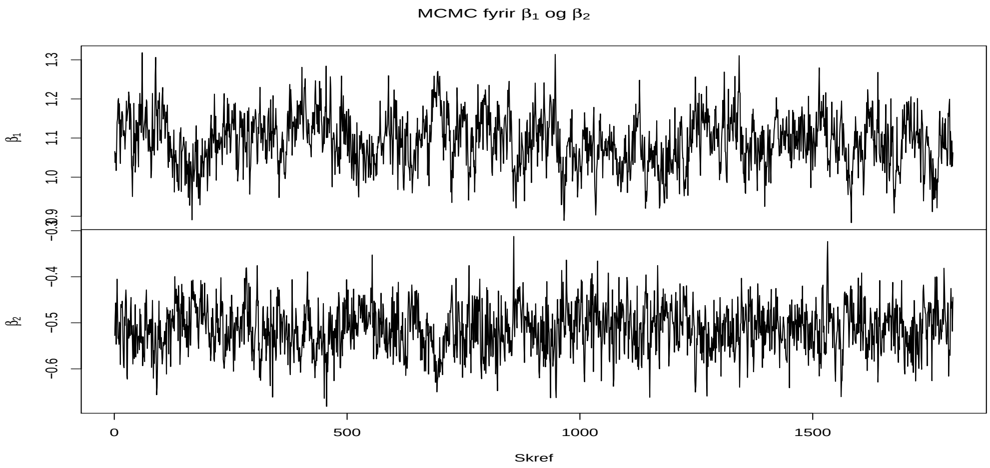
	$\hat{\beta}_{ML}$	s.e.-β	t-gildi
X_1	1.09	0.072	15.24
X_2	-0.51	0.054	-9.43

Tafla 8: Niðurstaða ML-mats í einföldu dæmi.

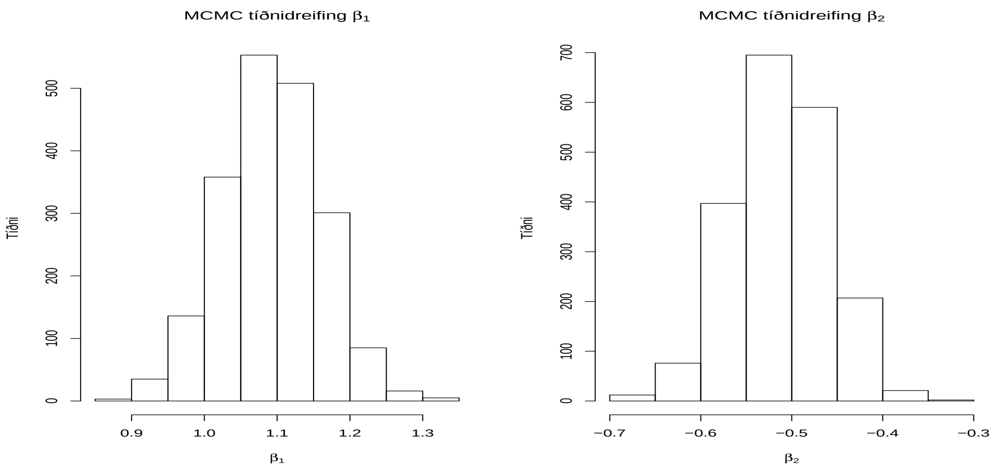
	$E(\beta)$	$\sqrt{V(\beta)}$
X_2	1.08	0.072
X_3	-0.51	0.054

Tafla 9: Meðaltal og staðalfrávik bayesísks mats.

ar gæðaeftirlit sjúkdómsgreiningu („diagnostics"). Gæðaeftirlitið er tvenns konar, annars vegar tæknilegt, til dæmis hvort að töluleg hámörkunaraðferð hafi í raun fundi hámark og hins vegar fræðilegt, þ.e. hvort forsendur metins líkans virðast bresta. Við ML-mat er tæknilega spurningin hvort um hámark sé að ræða og þá skoðað formerki annarar afleiðu sennileikafallsins. Í bayesískum MCMC aðferðum er tæknilega spurningin hvort að hermun Markov-keðjunnar hafi staðið nægjanlega lengi yfir til að jafnvægisdreifingu hafi verið náð. Fyrstu 2000 gildi í tvívíðri MCMC-keðju fyrir (β_1, β_2) eru sýnd í mynd 1. Ekki virðist um neina þróun að ræða og virðist keðjan sveiflast um fasta. Í mynd 2 eru sýnd súlurit af gildum keðjunnar fyrir (β_1, β_2). MCMC er ekki slembiúrtak og því nauðsynlegt að vita hversu tengdar hermanir eru milli skrefa í keðjunni. Því er hefð að skoða metin sjálffylgniföll. Í mynd 3 eru sýnd sjálffylgniföll fyrir keðjurnar sem herma β_1 og β_3. Samkvæmt myndinni virðist sjálffylgnin deyja út eftir fáeina tugi skrefa. Algengast er að keðjan taki gildi nálægt sönnu gildunum, $(\beta_1 = 1, \beta_2 = -0.5)$. Staðalfrávik eftirádreifingarinar í töflu 9 eru svipuð og ML-staðalfrávikin í töflu 8. Þröskuldarnir í ómældu breytunni (c_1, c_2, c_3, c_4) eru einnig metnir. Bayes-aðferðin meðaltal eftirávissu, (0,1.1,1.9.3.4) og ML-aðferðin gefur (0,1.0,1.88,3.4). Sönnu gildin voru (0,1.0,1.8,3.2). Staðalfrávikin voru í báðum tilfellum af stærðargráðunni 0.2.



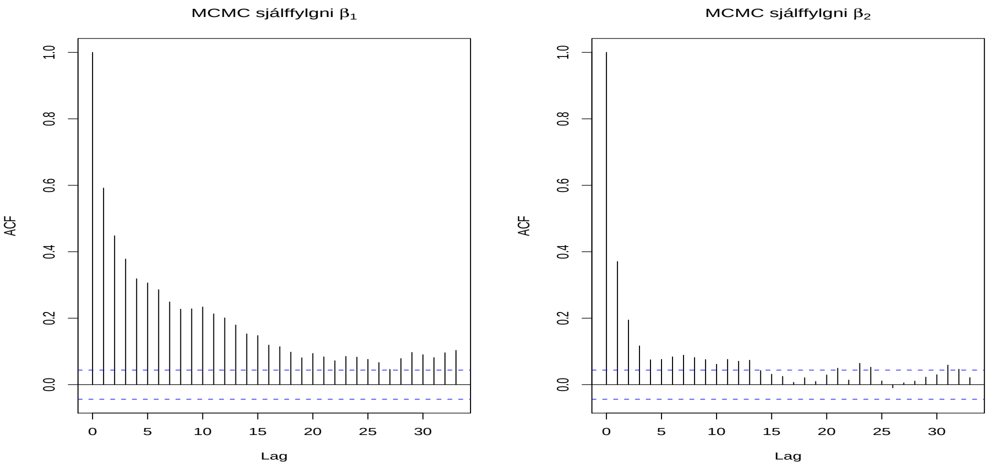
Mynd 1: MCMC keðjur fyrir β_1 og β_2.



Mynd 2: Tíðnidreifing MCMC fyrir β_1 og β_2.

7 Gögn úr kennslukönnun

Í markaðskönnun er það tíðkað að svarendur eru látnir kvarða ýmis atriði með spurningalista. Dæmi um slíkt eru kennslukannanir sem framkvæmdar eru í Háskóla Íslands og eiga sér hliðstæður í öðrum háskólum víða um heim. Að áliti höfundar er notkun viðeigandi líkana til ályktana út frá slíkum mælingum vanþróuð. Kannanirnar fara þannig fram að safnað er svörum við ákveðnum spurningarlista frá nemendum. Hverjum nemanda býðst að svara slíkum lista fyrir hvert námskeið sem viðkomandi er skráður í. Það er ekki skylda að svara og ekkert eftirlit er með því hvort nemandinn svarar einn og óháður eða hvernig



Mynd 3: Sjálffylgni MCMC-keðju fyrir β_1 og β_2.

hann ber sig að við svörin. Ljóst er því að gæði mælinganna er mjög misjöfn. Höfundur hefur undir höndum gögn um könnunina haustið 2005. Henni svöruðu 4053 nemendur frá einu svari per nemanda upp í 21 svör per nemanda. Fjöldi svara per nemenda er sýndur í mynd 4. Þar sést að í kringum 400-500, þ.e. samtals um 20%, svara 4 eða 5 námskeiðum. Svarendur voru sjálfum sér samkvæmir með kyn sitt, þ.e. innan hvers einstaklings þá var kynið óbreytt. Hugsanlega er þetta leiðrétt með vélrænum hætti. Karlkyns svarendur voru 1384, kvenkyns svarendur voru 2603 og kyn óuppgefið 66. Alls bárust 17723 svör.

Kyn er möguleg skýribreyta og sömuleiðis er hugsanlegt að nota spurninguna: Hvað einkunn telur þú að þú fáir í námskeiðinu?, sem skýrirbreytu. Þýðing þessara breyta var metin fyrir eina spurningu hjá í einu námskeiði. Alls 96 svör. Niðurstöður úr ML-mati eru sýndar í tölfu 10. Sömu gögn voru metin með bayesískri aðferð þar sem veikar fyrirfram upplýsingar voru notaðar. Matið sem sýnt er í töflu 11 er reiknað með 10000 umferðum í Gibbs-sampler þar sem fyrstu 1000 er sleppt (burn-in). Eins og búast má við þegar veik fyrirframvissa er notuð þá ber niðurstöðunum vel saman. Í báðum tilfellum skýrir uppgefin væntanleg einkunn nemanda mikið af þeirri raðtölu sem hann gefur sem svar við spurningunni. Einnig er eðlilegt að gera ráð fyrir að mikilvægar skýrirstærðir svo sem ástundun, undirbúningur hafi áhrif.

Eðlilegt er að gera rá fyrir að nemendur noti kvarðann á mismunandi hátt. Heitin

	$\hat{\beta}_{ML}$	s.e.-β	t-gildi
kyn	0.15	0.22	0.67
einkunn	0.31	0.11	2.89

Tafla 10: Niðurstaða ML-mats fyrir eina spurningu í einu námskeiði.

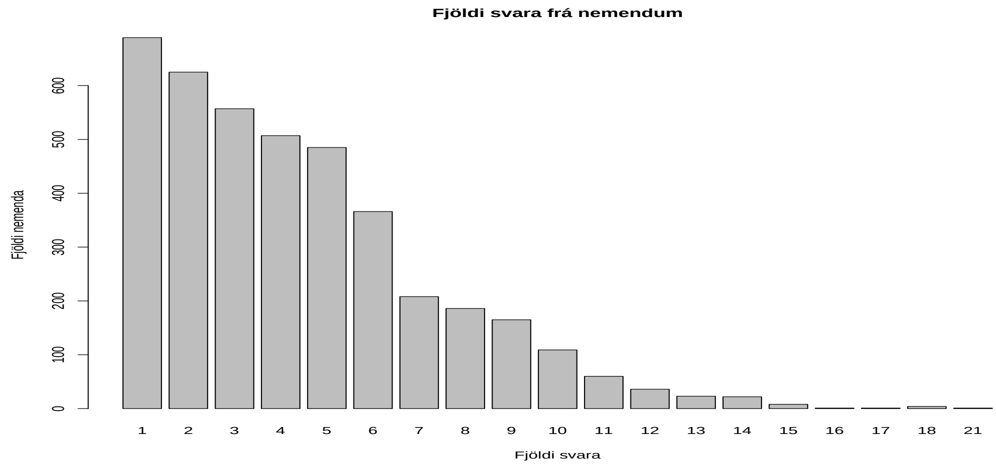
	$E(\beta)$	$\sqrt{V(\beta)}$
kyn	0.15	0.22
einkunn	0.31	0.11

Tafla 11: Meðaltal og staðalfrávik bayesísks mats fyrir eina spurningu í einu námskeiði.

„repeated-measures", „panel-data" og „longitudinal-data" eru samheiti á því fyrirbæri að margar mælingar eru á hverjum einstaklingi. Í þeim fræðum er mikilvægur þáttur að leiðrétta fyrir einstaklingsáhrifum. Hugtök „first-difference"(FD) og „fixed-effecs"(FE) eru tæknilega nálganir sem hafa það að markmiði að leiðrétta fyrir einstaklingsáhrifum. Um þær aðferðir má lesa í kennslubókum eins og t.d. Cameron & Trivedi (2005); Wooldridge (2002); Pinheiro & Bates (2000). Einfalt og algengt er að notast við einstaklingsstöðluð svör,

$$(x_{ij} - \bar{x}_i)/s_i,$$

þar sem x_{ij} er svar einstaklings i við spurningu j, $\bar{x}_i$ er meðaltal hjá einstakling númer i og s_i er staðalfrávik hjá einstakling númer i. Það að draga $\bar{x}_i$ frá leiðréttir fyrir því að einstaklingar noti skalann mismunandi hátt á þann hátt að þeir staðsetji sig á mismunandi hátt. Það að deila með s_i er viðleitni til að leiðrétta fyrir því að skrefstærðin í kvarðanum sé mismunandi eftir einstaklingum, t.d. svipað og að sumir mældu hitann á Fahrenheit og aðrir á Celius. Það að vinna með staðalfrávik í raðkvarða er hugsanlega hæpið eins og Rampichini, Grilli & Petrucci (2004) ræða. Rossi, Gilula & Allenby (2001) taka einnig á þeim vanda og segja að það þurfi að taka á því að breytan sé ósamfelld (*no standard transform can change the discreteness of the data*). Þeir stinga síðan upp á bayesískri aðferðafræði til að nálgast vandann. Til að sýna hvernig nota megi aðferðafræði Rossi,



Mynd 4: Dreifing svarfjölda eftir nemendum.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.00	0.06	0.39	0.01	0.10	0.24	0.04	−0.07	0.10	0.05	0.02	−0.14	−0.05	0.03	−0.03	−0.00
0.06	1.00	0.26	0.14	0.26	0.22	0.29	0.24	0.07	−0.09	−0.01	0.09	0.19	−0.18	−0.24	−0.16
0.39	0.26	1.00	0.05	0.12	0.35	0.12	0.02	0.06	−0.03	−0.10	−0.15	−0.11	0.01	−0.02	0.02
0.01	0.14	0.05	1.00	0.09	−0.02	0.13	0.21	0.18	0.23	0.20	0.22	0.26	−0.06	−0.07	−0.11
0.10	0.26	0.12	0.09	1.00	0.17	0.33	0.32	0.04	−0.11	0.12	0.10	0.12	−0.21	−0.23	−0.27
0.24	0.22	0.35	−0.02	0.17	1.00	0.40	0.11	−0.05	−0.09	0.02	−0.10	−0.00	−0.06	−0.19	−0.08
0.04	0.29	0.12	0.13	0.33	0.40	1.00	0.48	0.05	−0.13	0.01	0.02	0.26	−0.27	−0.33	−0.21
−0.07	0.24	0.02	0.21	0.32	0.11	0.48	1.00	0.03	0.01	0.06	0.12	0.25	−0.17	−0.19	−0.15
0.10	0.07	0.06	0.18	0.04	−0.05	0.05	0.03	1.00	0.11	−0.05	−0.04	−0.03	−0.17	−0.07	−0.11
0.05	−0.09	−0.03	0.23	−0.11	−0.09	−0.13	0.01	0.11	1.00	0.21	0.14	0.14	−0.00	0.01	0.03
0.02	−0.01	−0.10	0.20	0.12	0.02	0.01	0.06	−0.05	0.21	1.00	0.40	0.36	−0.05	−0.06	−0.10
−0.14	0.09	−0.15	0.22	0.10	−0.10	0.02	0.12	−0.04	0.14	0.40	1.00	0.53	−0.19	−0.07	−0.09
−0.05	0.19	−0.11	0.26	0.12	−0.00	0.26	0.25	−0.03	0.14	0.36	0.53	1.00	−0.23	−0.19	−0.17
0.03	−0.18	0.01	−0.06	−0.21	−0.06	−0.27	−0.17	−0.17	−0.00	−0.05	−0.19	−0.23	1.00	0.58	0.53
−0.03	−0.24	−0.02	−0.07	−0.23	−0.19	−0.33	−0.19	−0.07	0.01	−0.06	−0.07	−0.19	0.58	1.00	0.56
−0.00	−0.16	0.02	−0.11	−0.27	−0.08	−0.21	−0.15	−0.11	0.03	−0.10	−0.09	−0.17	0.53	0.56	1.00

Tafla 12: Einstaklingshreinsað mat á fylgni svara.

Gilula & Allenby (2001) var valið eitt námskeið með 96 gildum svörum við 16 spurningum.

$$
y_{ij} = \begin{cases} 1 & -\infty < y^*_{ij} < c_1 \\ \vdots & \vdots \\ 5 & c_4 \leq y^*_{ij} < \infty \end{cases}
$$

$$
y^*_i = (y_{i1}, \ldots, y_{iJ})', \quad J = 16.
$$

$$
y^*_i \sim (\mu^*_i, \Sigma^*_i).
$$

Til að gera stika greinanlega eru settar ýmsar skorður, t.d.:

$$
\mu^*_i = \mu + \tau_i + \sigma_i z_i, \quad z_i \sim N(0, \Sigma)
$$

þar sem z_i er 16 víð ómæld breyta með samdreifnifylki Σ. Gibbs sampler er settur upp í nokkrum þrepum. Þrepin eru valin þannig að skilyrtar dreifingar verði þægilegar. T.d. er gert ráð fyrir að skilyrt dreifing μ, gefið annað í líkani, sé normal, dreifing fylkisins Σ sé inverse-Wishart. Inverse-Wishart hendingin er einskonar margvíð χ^2 hending og eru gildi hennar fylki. Það að nota inverse-Wishart dreifingu sem fyrirframvissu um Σ í margvíðu normal líkani er hliðstætt því að nota inverse-gamma sem fyrirframvissu um σ^2 í einvíðu normal líkani.

8 Lokaorð

Þróun í tölvutækni hefur valdið því að margar aðferðir sem áður voru einungis heimspekilegs eðlis eru nú reiknanlegur raunveruleiki. Ef fyrirframvissan í bayesískri nálgun er valin þannig að jákvæðar líkur eru á öllu stikarúmin þá munu eiginleikar bayesmats í stórum úrtökum verða þeir sömu og eiginleikar ML. Ef veik fyrirframvissa er valin þá munu eiginleikar gagna fljótt verða ráðandi miðað við fyrirframvissuna. ML-aðferð getur haft leiðinlega eiginleika í litlum úrtökum. Ef til dæmis gögnin eru þannig að engin notar lægsta gildi kvarðans þá er ekki víst að sennileikafallið hafi hámark. Í flóknum líkönum, þ.e. stikinn $\boldsymbol{\theta}$ mjög margvíður, þá er aukin hætta á að ML-mat í einstökum hnitum misheppnist. Einnig er tæknilega erfitt að hámarka með tilliti til mjög margra stærða samtímis. Í slíkum tilfellum eru Bayes-aðferðir tæknilegur valkostur fyrir þá sem ekki aðhyllast Bayes-heimspekina.

Hér hefur verið rakið hvernig setja megi upp bayesískt líkan fyrir einvíða og margvíða raðkvarðahendingu. Augljóslega er hægt að bæta við skýristærðum, t.d. aldri, kyni o.s.frv. Í markaðskönnunum er gögnum stundum safnað með spurningalistum þar sem svarendur eru látnir kvarða svör sín. Ef svarendur hafa ekki samræmt sig fyrirfram er eðlilegt að gera ráð fyrir að þeir noti kvarðann mismunandi. Rossi, Gilula & Allenby (2001) segja: „*we focus on two major uses of customer-statisfaction-measurement ratings data: (1) measurement of the relationship between overall satisfaction with the specific product attributes and (2) identification of customers with extreme views. Scale usage heterogeneity can substantially bias analysis aimed at either use*". Því er oft nauðsynlegt að leiðrétta fyrir því að einstaklingar noti kvarðann á mismunandi hátt. Slík leiðrétting er þó alls ekki allra meina bót. Það má ekki slá misleitum hópum saman. Í kennslubók í markaðsfræðum segir Churchill (1995) „*it has been observed that there are large cultural or cross-country differences in scale usage, making it difficult to combine data across cultural or international boundaries*". Hæfni, vilji og smekkur svarenda ræður gæðum gagna í spurningakönnun. Morris (1978) lýsir líkindafræðilegu líkani sem meta á hæfni einstaklinga til að greina skemmt epli frá óskemmdu. Hann nefnir einnig vínsmökkun eins og Randall (1989). Ljóst er að sumir einstaklingar geta ekki greint skemmt frá óskemmdu og sumir sem geta greint á milli gætu metið skemmt epli betur. Ef vit á að vera í markaðskönnunum verður að nota líkan sem ræður við svona atriði. Það má ekki slá saman hópum sem hugsanlega snúa kvarðanum sitt í hvora áttina, samanber dæmið með fyrirtæki A og fyrirtæki B. Í því dæmi virtist sambandið í summunni vera öðruvísi en í hvoru fyrir sig.

Jafnvel þó að allar forsendur séu uppfylltar þarf að túlka útkomur með varúð. Höfundur þessarar greinar er algerlega sammála Parducci (1968) og Rossi, Gilula & Allenby (2001), „*we do not believe that even properly adjusted ratings data can provide **ratio scale information.** For example, if a respondent gives only ratings at the top end of the scale,* *we cannot infer that he or she is extremely satisfied*", þ.e. ekki má álykta að sá sem svarar öllum spurningum með hæsta gildi kvarðans afar ánægður. Einstaklingsleiðréttingin býður upp á einstaklingshreinsað mat á fylgnifylki (eða samdreifnifylki) margvíðu, ómældu hendingarinnar sem liggur að baki kvörðuðum svörum. Það mætti hugsanleg skoða slíkt fylgnifylki með einhvers konar þáttagreiningu (factor analysis). Að ofansögðu ætti að vera ljóst að ég tel ekki rétt að framkvæma þáttagreiningu á óleiðréttum gögnum. Það má aldrei sleppa mikilvægum skýristærðum. Rampichini, Grilli & Petrucci (2004) segja: ... *it should be recognized that the satisfaction of a student, as expressed by the ratings, depends not only on the course characteristics of interest, but also on the student's traits and expectations. Therefore a fair comparison among courses requires the calculation of net measures that adjust for individual characterisics. Such measures can be obtained, among others, by means of multilevel models* ... og síðar í sömu grein: *The descriptive indicators proposed in the previous section are mainly useful for the analysis of a single course. However, when it comes to comparison among courses, the are not appropriate since the evaluations expressed by the students are influenced by individual expectations and traits of the student themselves. Any comparison among courses that is based on the descriptive indicators, could be misleading. .. To make fair comparison among courses it is necessary to isolate the influence of these components on the expressed evaluations. In particular, it is important to net out the effects related to the individual characteristics of the students in order to obtain fair rankings of courses, schools, universities, etc.*

Hugleiðingar þessara höfunda um kennslukannanir eiga við um markaðskannanir og alla aðra hagnýta tölfræði. Ljóst er að í markaðskönnun þarf að huga vel að gæðum svara, þ.e. að svarendur geti og vilji svara. Einnig þeir snúi kvarðanum eins, þ.e. ekki að sumir mæli gæði með vaxandi gildi og aðrir með minnkandi gildi. Líkön til að fá einstaklingshreinsað mat á tengslum þátta á spurningalistum verða fljótt flókin. Bayesísk aðferðafræði opnar möguleika á vinnu með flókin líkön án þess að rannsakendur þurfi að skuldbindast bayesísku heimspekinni. Bayesískar aðferðir hafa svipaða eiginleika og ML-aðferðir í stórum úrtökum og oft heppilegri eiginleika í litlum úrtökum.

Heimildir

Agresti, A. (2002). *Categorical Data Analysis, 2nd edition.* John Wiley & Sons: New York.

Box, G. & Tiao, G. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley.

Cameron, A. & Trivedi, P. (2005). *Microeconometrics: Methods and Applications.* Cambridge University Press.

Canova, F. (2007). *Methods for Applied Macroeconomic Research.* Princeton University Press.

Churchill, G. (1995). *Marketing Research: Methodological Foundations* (6th ed.). Orlando: Dryden Press.

Congdon, P. (2005). *Bayesian Models for Categorical Data.* John Wiley & sons.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian Data Analysis* (2 ed.). Chapman & Hall.

Koop, G. (2003). *Bayesian Econometrics.* John Wiley & sons.

Lancaster, T. (2004). *An Introduction to Modern Baysian Econometrics.* Blackwell Publishing.

Morris, D. (1978). A probability model for forced binary choice. *The American Statistician*, *32*(1), 23–25.

Parducci, A. (1968). The relevatism in absolute judgement. *Scientific American*, *219*, 84–90.

Pinheiro, J. & Bates, D. (2000). *Mixed-effects models in S and S-Plus.* Springer.

R Development Core Team (2005). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rampichini, C., Grilli, L., & Petrucci, A. (2004). Analysis of university course evaluations: from descriptive measures to multilevel models. *Statistical Methods & Applications*, *13*, 357–373.

Randall, J. (1989). The analysis of sensory data by generalized linear models. *Biometrical Journal*, *31*, 783–791.

Rossi, P., Allenby, G., & McCulloch, R. (2005). *Bayesian Statistics and Marketing.* John Wiley & sons.

Rossi, P., Gilula, Z., & Allenby, G. (2001). Overcoming scale usage heterogeneity: A bayesian hierarchical approach. *Journal of the American Statistical Association*, *96*(453), 20–31.

Stiegler, S. (1986). Laplace's 1774 memoir on inverse probability. *Statistical Science*, *1*(3), 359–378.

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

Wooldridge, J. (2002). *Econometric Analysis of cross section and panel data.* MIT press.

Zellner, A. (1971). *An Introduction to Bayesian Inference.* John Wiley & sons.