

# Mat á lengdardreifingum

Birgir Hrafnkelsson<sup>1</sup> og Gunnar Stefánsson<sup>2</sup>

<sup>1,2</sup>Háskóla Íslands og <sup>2</sup>Hafrannsóknastofnun

Vefútgáfa: 18. október 2003

**Ágrip** – Hér er kynnt nýtt tölfræðilíkan sem byggir á fjölkostalíkaninu. Samdreifnifylki líkansins er sveigjanlegur en samdreifnifylki fjölkostalíkansins. Fjölkostalíkanið hefur verið notað fram að þessu til að lýsa gögnum af fjölda fiska í lengdarflokki úr rannsóknaleiðangrum eða úr lönduðum afla. Fjölkostalíkanið nær engan veginn að lýsa samdreifni gagna af þessu tagi meðan nýja líkanið nær að lýsa henni vel. Kostnaðarsamir rannsóknaleiðangrar eru farnir árlega frá hafrannsóknastofnunum víða um heim til að safna gögnum um fiskistofna. Oftast er framkvæmt eitt tog á hverri stöð, þar sem stöðvarnar eru dreifðar á ákveðinn hátt um svæðið sem á skoða. Á hverri stöð og fyrir hverja fiskitegund, er tekið handahófsúrtak og fiskarnir í úrtakinu eru lengdarmældir. Við þetta fæst tíðnin í hverjum lengdarflokki á hverri stöð. Eitt af markmiðunum er að meta lengdardreifingu ákveðinnar tegundar á ákveðnum svæðum. Þrjú líkön fyrir gögn um tíðni í lengdarflokki eru kynnt og borin saman. Notið eru gögn um rækju við samanburðinn. Líkönin eru fjölkostalíkanið og tvö stigskipt líkön (e. hierarchical models). Í stigskiptum líkönunum er gert ráð fyrir að á hverri stöð fylgi fjöldinn í lengdarflokki fjölkostalíkani með einhvern fjölda mældra fiska og með einhvern tíðnivígi. Tíðnivígararnir eru breytilegir á milli stöðva en koma frá sameiginlegri dreifingu með ákveðið væntigildi og samdreifnifylki. Dreifingar tíðnivígranna í stigskiptu líkönunum eru Dirichlet-dreifingin og margvið Gauss-dreifing sem er varpað yfir í hymu. Gögnin hafa meiri dreifni en fjölkostalíkanið og fylgnin í gögnunum er sterkari en fylgnin í líkaninu sem byggir á fjölkostalíkaninu og Dirichlet-dreifingunni. Líkanið sem byggir á fjölkostalíkaninu og margviðu Gauss-dreifingunni virðist ná að lýsa dreifninni og fylgninni í gögnunum viðunandi vel. Stikarnir í líkönunum eru metnir með Bayes-matsaðferðum.

## 1. Inngangur

Eitt af aðalviðfangsefnum hafrannsókna eru líkön fyrir fiskistofna. Rannsóknirnar krefjast mikillar gagnasöfnunar um fiskistofnana. Tvær grundvallartegundir gagna um fiskistofna eru lengdarmælingar stakra fiska og magnvísitölur (e. abundance indices). Þessum gögnum má safna í rannsóknaleiðangrum eða úr lönduðum afla. Önnur gögn geta einnig verið mikilvæg í ákveðnum tilfellum en önnur tegundin af þessum grundvallargögnum er alltaf hluti af greiningu á fiskistofnum.

Tölfræðilegir eiginleikar magnvísitalna hafa verið rannsakaðir umtalsvert, sjá t.d. Pennington (1983), Gunnar Stefánsson (1996) og Jacobson (1996). Lengdarmælingar á stökum fiskum eru þannig að fiskarnir eru flokkaðir í ákveðna lengdarflokka, t.d. 1 cm

eða 1 mm flokka, (flokunargögn). Þrátt fyrir að gögn um lengdarflokka séu afar mikilvæg fyrir stofnstærðarmat, þá hafa tölfræðilegir eiginleikar gagnanna ekki verið rannsakaðir til hlítar og einföld tölfræðileg líkön hafa verið notuð. Í þessari grein er sýnt fram á hversu langt frá forsendum einföldu líkananna gögnin eru og kynnt tölfræðilíkan sem lýsir gögnunum mun betur.

Fjölkostalíkanið hefur mikið verið notað til að lýsa fjölda fiska í lengdarflokki, sjá t.d. MacDonald og Pitcher (1979). Líkanið gerir ráð fyrir að einstaklingur úr ákveðnum flokki sé valinn af handahófi með ákveðnum líkum. Í MacDonald og Pitcher (1979) er gert ráð fyrir að líkurnar breytist ekki á milli staða (stöðva). Í flestum tilfellum hegða fiskar sér ekki sem fullkomlega sjálfstæðir einstaklingar heldur er um einhvers konar hóphegðun að ræða. Þetta er meðal annars komið til vegna fæðuvals og hryggningar-

aldurs. Af þessum sökum kemur fram meiri dreifni en fjölkostalíkanið segir til um. Fjallað er um flokk-unarbreytulíkön sem gera ráð fyrir meiri dreifni en fjölkostalíkanið í köflum 4-5 í McCullagh og Nelder (1989).

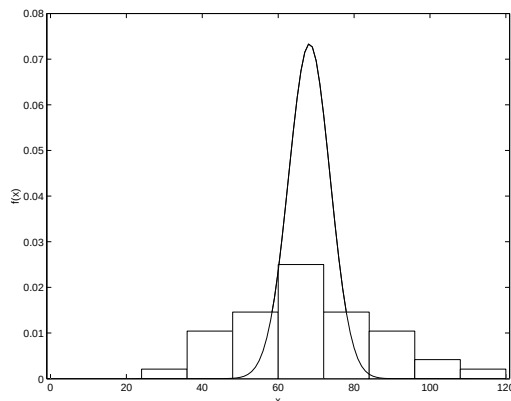
Önnur forsenda í fjölkostalíkaninu, sem ekki á við um lengdarflokknargögn um fiska, er forsendan um fylgni milli fjölda í lengdarflokki. Í gögnum má sjá að samliggjandi flokkar hafa sterka jákvæða fylgni á meðan fjölkostalíkanið getur aðeins gefið neikvæða fylgni. Það er viðbúið að fiskar af svipaðri lengd hegði sér á svipaðan hátt og haldi því hópinn, en þetta kallar á jákvæða fylgni milli samliggjandi flokka. Líffræðilegar ástæður fyrir þessu geta verið fæðuval, þar sem t.d. stærð rándýrsins setur ákveðnar skorður á ákjósanlega stærð bráðarinnar, sem leiðir af sér að fiskar af svipaðri stærð finnast á sama svæði.

Klassískar aðferðir við greiningu gagna um fiskistofna byggja á að nota samantektarstærðir, þar sem stærðum frá mismunandi stöðum er slegið saman. Það er ljóst að tölfraðileg mæliskekkja í samantektarstærðunum er ekki sú sem að einfalt líkan eins og fjölkostalíkanið mundi spá fyrir um. Þetta fer að skipta máli þegar nota á nokkur gagnasöfn í einu og val á vogtölum fyrir mismunandi þætti sennileikafallsins getur haft mikil áhrif á punktmatið, sjá Gunnar Stefánsson (1998). Nýjustu líkönin fyrir fiskistofna eru afar flókin og krefjast að gagnasöfn séu vigtuð í sennileikafallinu, sjá Gunnar Stefánsson og Ólafur K. Pálsson (1998). Því er ástæða til að finna leiðir til að skilgreina sennileikabátt hvers gagnsafns vel og þannig að ekki þurfi að nota vogtölur.

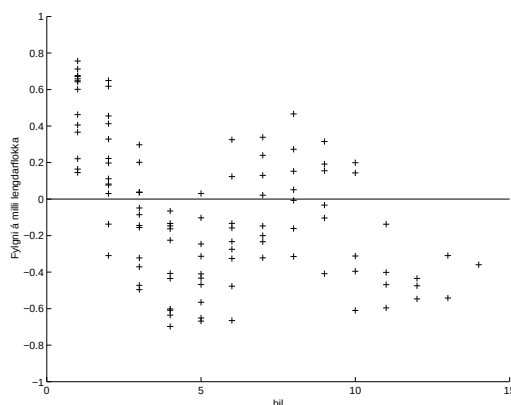
Greinin er þannig uppbyggð að í kafla 2 er gögnunum lýst og athugað hversu langt frá forsendum fjölkostalíkansins gögnin eru. Í kafla 3 eru kynnt þrjú líkön sem er beitt á gögnin. Kafli 4 fjallar um greiningu á gögnunum og í kafla 5 er samantekt.

## 2. Gögnin

Gögnin sem eru notuð í þessari grein eru lengdarmælingar á karlkynsrækju og eru dæmigerð fyrir lengdarflokknargögn úr rannsóknaleiðangrum. Á hverri stöð er framkvæmt eitt tog og úr hverju togi er tekið handahófsúrtak af rækju sem er lengdarmælt. Aðeins eru notuð gögn frá 43 stöðvum í Ísafjarðardjúpi úr leiðangri sem var farinn í október 2000. Fjöldi lengdarmældrar rækju á hverri stöð er á bilinu 60 til 394 með miðgildi 264. Í upphaflegu gögnunum er rækj-



**Mynd 1.** Tíðnirit af fjölda rækju sem eru minni en 13.5 mm að lengd (af 120 á hverri stöð). Byggt á gögnum frá 40 stöðvum þar sem 120 eða fleiri rækjur voru lengdarmældar. Nákvæmlega 120 rækjur voru valdar af handahófi frá þessum 40 stöðvum. Línan sýnir tvíkostadreifingu með stika  $n = 120$  og  $p = 0.568$ .



**Mynd 2.** Fylgni á milli fjölda í lengdarflokki. Byggt á gögnum frá 40 stöðvum þar sem 120 eða fleiri rækjur voru lengdarmældar. Nákvæmlega 120 rækjur voru valdar af handahófi frá þessum 40 stöðvum til að reikna fylgni-stuðlana.

unni skipað í 0.5 mm flokka. Flokkunum frá 6 mm til 11 mm og flokkunum frá 18 mm til 20.5 mm er slegið saman, en flokkarnir á milli 11 mm og 18 mm eru óbreyttir. Samtals eru 15 lengdarflokkar.

Til sjá hvort dreifni sé meiri en fjölkostalíkanið segir til um skiptum við gögnunum í tvo flokka, rækja minni en 13.5 mm og rækja stærri en 13.5 mm. Því næst eru valdar 120 mælingar af handahófi á öllum þeim 40 stöðvum þar sem 120 eða fleiri rækjur voru

mældar. Þegar nákvæmlega sami fjöldi mælinga er á hverri stöð er auðveldara að skoða eiginleika gagnanna. Ef fjölkostalíkanið á við gögnin þá fylgir fjöldi rækju sem er 13.5 mm að lengd eða minni tvíkostadreifingu með stika  $n = 120$  og eitthvert  $p$ . Mynd 1 sýnir tíðnirit fjöldans í flokknum með rækju 13.5 mm og undir, ásamt tvíkostadreifingu sem kemst næst gögnunum. Það er ljóst að það er mun meiri dreifni í gögnunum en tvíkostalíkanið segir til um, því nokkrar mælingar falla á bilið núll til 40 og á bilið 90 til 120 en tvíkostalíkanið spáir mjög litlum líkum á báðum bilum. Einfalt kí-kvaðrat-próf styður tilgátuna um að tvíkostalíkanið eigi ekki við en  $p$ -gildi prófsins er núll. Niðurstaðan er því sú að líkan sem getur gert ráð fyrir meiri dreifni er nauðsynlegt til að lýsa gögnunum.

Ef notað er sama undirsafn af gögnunum og hér að ofan, þar sem fjöldinn er sá sami á hverri stöð, má fá vigra með fjölda rækju í hverjum af flokkunum 15. Með þessum vigrum má meta fylgni á milli fjöldans í tveimur flokkum án þess að heildarfjöldi á hverri stöð bjagi matið á fylgninni. Þessi fylgni er sýnd á mynd 2 sem fall af fjarlægð á milli flokka mæld í bilum (ekki í mm). Hver punktur sýnir fylgni á milli fjöldans í tveimur flokkum. Það er ljóst af mynd 2 að fylgnin er ekki sú sama og fjölkostalíkanið segir fyrir um, en fylgnin í fjölkostalíkaninu er veik og neikvæð á meðan mynd 2 sýnir aðeins jákvæða fylgni fyrir samliggjandi flokka. Það er afar ólíklegt að fylgni-stuðlarnir fyrir samliggjandi flokka sem reiknaðir eru út frá gögnunum séu allir jákvæðir þegar undirliggjandi líkan segir fyrir um neikvæða fylgni. Eftir að hafa skoðað dreifina og fylgnina í gögnunum er ljóst að líkan sem hefur sveigjanlegra samdreifnifylki en fjölkostalíkanið er nauðsynlegt til að lýsa gögnunum.

### 3. Þrjú líkön fyrir flokkunargögn

Í þessum kafla eru kynnt þrjú líkön fyrir flokkunargögn og fjallað um hvernig stikarnir eru metnir með Bayes-matsaðferðum sem byggja á Markov-keðju Monte Carlo (MCMC), sjá t.d. Gilks, Richardson og Spiegelhalter (1996) og Carlin og Louis (2000). Gert er ráð fyrir að hver mæling sé vigur þar sem hvert stak gefur fjöldann í tilsvarendi flokki. Summan yfir hvern vigur er ekki sú sama, en greiningin á gögnunum er skilyrt á summuna yfir hvern vigur.

Fyrsta líkanið er hið vel þekkta fjölkostalíkan þar sem tíðnivigurinn er sá sami fyrir allar mæl-

ingar, sjá kafla 5 í McCullagh og Nelder (1989). Hér verður þetta líkan kallað staðlaða fjölkostalíkanið. Annað líkanið er Dirichlet-fjölkostalíkanið, sjá t.d. Mosimann (1962). Líkanið er stigskipt þar sem hver mæling fylgir fjölkostadreifingu með einhvern tíðnivigur en tíðnivigrarnir fylgja staðlaðri Dirichlet-dreifingu, sjá t.d. Johnson og Kotz (1972), bls. 231-235. Þriðja líkanið er eins og annað líkanið nema að tíðnivigrarnir hafa sömu dreifingu og margvíd Gauss-slembibreyta sem er varpað á hyrnu. Aitchison (1985) fjallar um slembistærðir á hyrnu. Hér verður þetta líkan kallað Gauss-fjölkostalíkanið en það hefur ekki verið kynnt í ritum hingað til. Líkanið hefur sveigjanlegra samdreifnifylki heldur en Dirichlet-fjölkostalíkanið. Sem dæmi um greinar sem fjalla um stigskipt Bayes-líkön má nefna, Diggle, Tawn, og Moyeed (1998) og Waller, Carlin, Xia, og Gelfand (1997).

Mestur áhugi er á meðaltíðnivigrinum, táknaður með  $\theta$ , þar sem í öllum þremur líkönum gildir að

$$E(X_q|m_q) = m_q\theta, \quad q = 1, \dots, Q,$$

þar sem  $X_q$  táknar  $q$ -ta fjölkostavigurinn,  $m_q$  er fjöldi hluta eða einstaklinga sem eru flokkaðir og  $Q$  er fjöldi vigra. Í stigskiptu líkönum má líta á tíðnivigurinn fyrir  $q$ -ta fjölkostavigurinn, táknaður með  $p_q$ , sem slembibreytu þar sem

$$E(p_q) = \theta, \quad q = 1, \dots, Q.$$

Í staðlaða fjölkostalíkaninu má líta á  $p_q$  sem fasta þar sem  $p_q = \theta$  fyrir öll  $q$ .

Hér táknar  $X_{qj}$  fjöldann í  $j$ -ta flokknum fyrir  $q$ -tu mælinguna og  $p_{qj}$  er tilsvarendi tíðni og  $X_q = (X_{q1}, \dots, X_{qJ})^T$  og  $p_q = (p_{q1}, \dots, p_{qJ})^T$ . Summan yfir  $X_q$  er  $m_q$ , það er,  $m_q = \sum_{j=1}^J X_{qj}$ . Ennfremur er  $X = (X_1^T, \dots, X_Q^T)^T$ ,  $m = (m_1, \dots, m_Q)^T$  og  $p = (p_1^T, \dots, p_Q^T)^T$ . Fengin gildi á  $X_q$  og  $X$  eru táknuð með  $x_q$  og  $x$ .

#### 3.1. Staðlaða fjölkostalíkanið

Strjáli slembivigurinn  $X_q$  fylgir stöðluðu fjölkostalíkani með stika  $m_q$  og  $\theta$  ef samdreifing  $X_q$  er á forminu

$$P(X_q = x_q) = f(x_q|m_q, \theta) = \frac{m_q!}{x_{q1}! \dots x_{qJ}!} \theta_1^{x_{q1}} \dots \theta_J^{x_{qJ}}, \quad q = 1, \dots, Q,$$

þar sem  $\theta_j > 0$ ,  $j = 1, \dots, J$  og  $\sum_{j=1}^J \theta_j = 1$ . Við ritum

$$(X_q|m_q, \theta) \sim \text{Mult}_J(m_q, \theta), \quad q = 1, \dots, Q,$$

sem táknar að  $X_q$  fylgir ofangreindri dreifingu. Væntigildi og samdreifni  $X_q$  gefið  $m_q$ , eru

$$E(X_q|m_q) = m_q\theta, \quad \text{Var}(X_q|m_q) = m_q\Psi, \\ q = 1, \dots, Q,$$

þar sem

$$\Psi = \Psi(\theta) = \text{diag}(\theta) - \theta\theta^T,$$

og  $\text{diag}(z)$  táknar hornalínufylki með vigurinn  $z$  á hornalínunni. Fylgnin á milli fjöldans í flokki  $j$  og fjöldans í flokki  $i$  er

$$\text{Corr}(X_{qj}, X_{qi}) = -\sqrt{\frac{\theta_j\theta_i}{(1-\theta_j)(1-\theta_i)}}, \\ i, j = 1, \dots, J, \quad i \neq j, \quad q = 1, \dots, Q. \quad (1)$$

Af (1) má sjá að fylgnin er neikvæð og veik ef bæði  $\theta_j$  og  $\theta_i$  eru lítil.

Fyrirframdreifing (e. prior distribution) stika er sú dreifing sem að endurspeglar þá vitneskju sem til er um stikann áður en gögnin eru kynnt. Látum fyrirframdreifingu  $\theta$  vera Dirichlet-dreifingu á  $(J-1)$ -víðri hyrnu með stika  $\beta = (\beta_1, \dots, \beta_J)^T$ . Dreifing  $\theta$  er táknueð með  $\pi(\theta)$  og er gefin með

$$\pi(\theta) = \frac{\Gamma(\beta_0)}{\Gamma(\beta_1) \cdots \Gamma(\beta_J)} \theta_1^{\beta_1-1} \cdots \theta_J^{\beta_J-1}, \\ \beta_0 = \sum_{j=1}^J \beta_j.$$

Við ritum  $\theta \sim \text{Dir}_J(\beta)$  sem táknar að  $\theta$  fylgir ofangreindri fyrirframdreifingu. Eftirárdreifing (e. posterior distribution) stika er sú dreifing sem að lýsir vitneskju um stikann eftir að gögnin hafa verið notuð til að uppfæra fyrirframdreifingu stikans. Eftirárdreifing  $\theta$  gefið  $m$  og  $X = x$ , táknueð með  $\pi(\theta|m, x)$ , er í hlutfalli við dreifingu  $X$  margfaldaða við fyrirframdreifingu  $\theta$

$$\pi(\theta|m, x) \propto \left\{ \prod_{q=1}^Q f(x_q|m_q, \theta) \right\} \pi(\theta).$$

Þá má sýna að eftirárdreifingin hér að ofan er Dirichlet-dreifing, það er

$$(\theta|m, x) \sim \text{Dir}_J(\beta + \sum_{q=1}^Q x_q).$$

Það er auðvelt að fá úrtak frá dreifingunni hér að ofan og ekki þarf að beita MCMC aðferðum.

### 3.2. Dirichlet-fjölkostalíkanið

Dirichlet-fjölkostalíkan með stikavigur  $\alpha = (\alpha_1, \dots, \alpha_J)^T$  má rita sem stigskipt líkan á forminu

$$(X_q|m_q, p_q) \sim \text{Mult}_J(m_q, p_q), \\ (p_q|\alpha) \sim \text{Dir}_J(\alpha), \quad q = 1, \dots, Q,$$

sem felur í sér að samdreifing  $X_q$  og  $p_q$  er

$$f(x_q, p_q|m_q, \alpha) = f(x_q|m_q, p_q)\pi(p_q|\alpha) \\ = \frac{m_q!}{x_{q1}! \cdots x_{qJ}!} p_{q1}^{x_{q1}} \cdots p_{qJ}^{x_{qJ}} \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_J)} \\ \times p_{q1}^{\alpha_1-1} \cdots p_{qJ}^{\alpha_J-1}, \quad q = 1, \dots, Q,$$

þar sem  $\pi(p_q|\alpha)$  er dreifing  $p_q$  (Dirichlet-dreifing með stikavigur  $\alpha$ ) og  $\alpha_0 = \sum_{k=1}^J \alpha_k$ . Jaðardreifing  $X_q$  fæst með því að heilda  $f(x_q, p_q|m_q, \alpha)$  yfir  $p_q$  vigurinn en hún er til á lokuðu formi, sjá Mosimann (1962). Tíðnivigrarnir,  $p_q$ , fylgja Dirichlet-dreifingu og því er væntigildi og samdreifni tíðnivigranna

$$E(p_q) = \theta, \quad \text{Var}(p_q) = \tau\Psi, \quad q = 1, \dots, Q, \quad (2)$$

þar sem

$$\theta = \alpha/\alpha_0, \quad \text{og} \quad \tau = (\alpha_0 + 1)^{-1}, \quad 0 < \tau < 1, \quad (3)$$

sjá t.d. Johnson og Kotz (1972), bls. 231-235. Fylgnin á milli  $p_{qj}$  og  $p_{qi}$  er sú sama og fylgnin á milli  $X_{qj}$  og  $X_{qi}$  í (1). Væntigildi og samdreifni  $X_q$  gefið  $m_q$ , eru

$$E(X_q|m_q) = m_q\theta, \\ \text{Var}(X_q|m_q) = m_q\{1 + \tau(m_q - 1)\}\Psi, \\ q = 1, \dots, Q.$$

Í jöfnunni hér að ofan má sjá hvernig stikinn  $\tau$  stýrir dreifninni í  $X_q$ . Fylgnin á milli stakanna í  $X_q$  er sú sama og fylgnin í staðlaða fjölkostalíkaninu, sjá (1). Jaðarvæntigildi og jaðarsamdreifni í tveggja laga

röðuðu líkani ákvarðast af  $E(X) = E\{E(X|Y)\}$  og  $\text{Var}(X) = E\{\text{Var}(X|Y)\} + \text{Var}\{E(X|Y)\}$ , sjá t.d. Casella og Berger (1990), bls. 153-160.

Fyrirframdreifing  $\alpha$  er táknuð með  $\pi(\alpha)$ . Gert er ráð fyrir að  $\pi(\alpha)$  sé þannig að stök  $\alpha$  séu óháð. Fyrirframdreifing  $\alpha_j$  er táknuð með  $\pi(\alpha_j)$ ,  $j = 1, \dots, J$ , og því er  $\pi(\alpha) = \prod_{j=1}^J \pi(\alpha_j)$ . Hér er veldisvísisdreifing með væntigildi  $\gamma_j$  valin sem fyrirframdreifing  $\alpha_j$ . Eftirádreifing  $\alpha$  og  $p$  gefið  $m$  og  $X = x$ , er

$$\pi(\alpha, p|m, x) \propto \left\{ \prod_{q=1}^Q f(x_q|m_q, p_q) \pi(p_q|\alpha) \right\} \pi(\alpha),$$

Til að draga handahófsúrtök frá eftirádreifingunni hér að ofan, eru dregin handahófsúrtök frá eftirfarandi skilyrtu dreifingum til skiptis

$$\pi(\alpha_j|m, x, \alpha_{-j}, p) \propto \left\{ \prod_{q=1}^Q f(x_q|m_q, p_q) \pi(p_q|\alpha) \right\} \pi(\alpha_j),$$

fyrir  $j = 1, \dots, J$  og

$$\pi(p_q|m, x, \alpha, p_{-q}) \propto f(x_q|m_q, p_q) \pi(p_q|\alpha),$$

fyrir  $q = 1, \dots, Q$ , þar sem  $p_{-q}$  táknar vigur með öllum stökum  $p$  að undanskildum stökum  $p_q$  og  $\alpha_{-j}$  táknar vigur með öllum stökum  $\alpha$  að undanskildu  $j$ -ta staki  $\alpha$ . Dreifingarnar hér að ofan mynda svokallaðan Gibbs safnara, sjá Carlin and Louis (2000) bls. 141-152. Gerð er grein fyrir skilyrtu dreifingunum hér að ofan í viðauka. Þegar búið er að safna úrtaki frá eftirádreifingu  $\alpha$  þá má varpa því úrtaki yfir í úrtak af  $\theta$ ,  $\tau$  samkvæmt (3) og yfir í úrtak af fylgni samkvæmt (1).

### 3.3. Gauss-fjölkostalíkanið

Gauss-fjölkostalíkanið má rita sem stigskipt líkan á forminu

$$(X_q|m_q, p_q) \sim \text{Mult}_J(m_q, p_q),$$

$$p_{qj}(\eta_q) = \frac{\exp(\eta_{qj})}{1 + \sum_{k=1}^{J-1} \exp(\eta_{qk})}, \quad j = 1, \dots, J-1,$$

$$p_{qJ}(\eta_q) = \frac{1}{1 + \sum_{k=1}^{J-1} \exp(\eta_{qk})},$$

$$(\eta_q|\mu, W) \sim \text{Gau}(\mu, W^{-1}), \quad q = 1, \dots, Q, \quad (4)$$

þar sem  $\eta_q = (\eta_{q1}, \dots, \eta_{q, J-1})^T$ . Látum einnig  $\eta = (\eta_1^T, \dots, \eta_Q^T)^T$ . Líkanið hér að ofan felur í sér að samdreifing  $X_q$  og  $\eta_q$  er

$$f(x_q, \eta_q|m_q, \mu, W) = f(x_q|m_q, p_q(\eta_q))\pi(\eta_q|\mu, W)$$

$$= \frac{m_q}{x_{q1}! \dots x_{qJ}!} p_{q1}^{x_{q1}}(\eta_q) \dots p_{qJ}^{x_{qJ}}(\eta_q) (2\pi)^{-(J-1)/2}$$

$$\times |W|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2}(\eta_q - \mu)^T W (\eta_q - \mu) \right\},$$

þar sem  $\pi(\eta_q|\mu, W)$  er Gauss-dreifing með væntigildi  $\mu$  og samdreifnifylki  $W^{-1}$ . Eins og í Dirichlet-fjölkostalíkaninu, gildir að  $E(p_q) = \theta$ ,  $q = 1, \dots, Q$ , þar sem  $j$ -ta stak  $\theta$  er gefið með

$$\theta_j = E_{\mu, W} \left( \frac{\exp(\psi_j)}{1 + \sum_{k=1}^{J-1} \exp(\psi_k)} \right), \quad j = 1, \dots, J, \quad (5)$$

þar sem væntigildið í (5) er yfir  $\psi$  þar sem  $\psi$  fylgir Gauss-dreifingu með meðalgildi  $\mu$  og samdreifni  $W^{-1}$ . Í líkaninu er samdreifnifylki sérhvers  $p_q$  táknað með  $\Omega$ , það er,  $\text{Var}(p_q) = \Omega$ ,  $q = 1, \dots, Q$ , þar sem  $(i, j)$ -ta stak  $\Omega$  er gefið með

$$\Omega_{ij} = E_{\mu, W} \left( \frac{\exp(\psi_i + \psi_j)}{\{1 + \sum_{k=1}^{J-1} \exp(\psi_k)\}^2} \right) - \theta_i \theta_j,$$

$$i, j = 1, \dots, J. \quad (6)$$

Ritum  $j$ -ta hornalínustak  $\Omega$  sem  $\Omega_{jj} = \tau_j \theta_j (1 - \theta_j)$ ,  $j = 1, \dots, J$ . Stikinn  $\tau_j$  stýrir því hversu mikil dreifni er í  $j$ -ta flokknum umfram  $m_q \theta_j (1 - \theta_j)$ . Væntigildi og samdreifni  $X_q$  gefið  $m_q$  sem fall af  $\theta$ ,  $\Psi$  og  $\Omega$ , eru

$$E(X_q|m_q) = m_q \theta,$$

$$\text{Var}(X_q|m_q) = m_q \Psi + m_q(m_q - 1) \Omega, \quad q = 1, \dots, Q. \quad (7)$$

Jöfnurnar í (7) eru fundnar á sama hátt og tilsvareandi jöfnur í Dirichlet-fjölkostalíkaninu í kafla 3.2.

Látum  $\pi(\mu, W)$  vera fyrirframdreifingu  $\mu$  og  $W$  þar sem fyrirframdreifingin er þannig að  $\mu$  og  $W$  eru óháðar. Því má rita  $\pi(\mu, W) = \pi(\mu)\pi(W)$  þar sem

$\pi(\mu)$  er fyrirframdreifing  $\mu$  og  $\pi(W)$  er fyrirframdreifing  $W$ . Eftirádreifing  $\mu$ ,  $W$  og  $\eta$  gefið  $m$  og  $X = x$ , er

$$\pi(\mu, W, \eta | m, x) \propto \left\{ \prod_{q=1}^Q f(x_q | m_q, p_q(\eta_q)) \pi(\eta_q | \mu, W) \right\} \pi(\mu) \pi(W).$$

Til að draga handahófsúrtök frá eftirádreifingunni hér að ofan, eru dregin handahófsúrtök frá eftirfarandi skilyrtu dreifingum til skiptis

$$\begin{aligned} \pi(\mu | m, x, W, \eta) &\propto \left\{ \prod_{q=1}^Q \pi(\eta_q | \mu, W) \right\} \pi(\mu), \\ \pi(W | m, x, \mu, \eta) &\propto \left\{ \prod_{q=1}^Q \pi(\eta_q | \mu, W) \right\} \pi(W), \\ \pi(\eta_q | m, x, \mu, W, \eta_{-q}) &\propto f(x_q | m_q, p_q(\eta_q)) \\ &\quad \times \pi(\eta_q | \mu, W), \quad q = 1, \dots, Q, \end{aligned}$$

þar sem  $\eta_{-q}$  táknar vigur sem inniheldur öll stök  $\eta$  að undanskildum stökum vigursins  $\eta_q$ . Gerð er grein fyrir skilyrtu dreifingunum hér að ofan í viðauka.

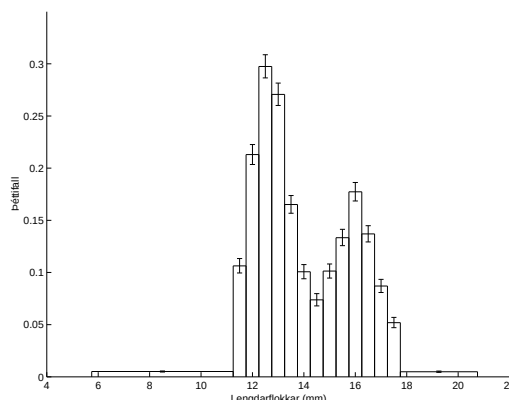
Til að draga handahófsúrtök af  $\theta$  og  $\Omega$  frá eftirádreifingunni er notuð slembin heildun fyrir hvert stak af  $\mu$  og  $W$ . Það er, fyrir  $i$ -ta stakið af  $\mu$  og  $W$  frá eftirádreifingunni, sem er hér táknað með  $(\mu^{(i)}, W^{(i)})$ , drögum við 100,000 handahófsúrtök frá Gauss-dreifingu með væntigildi  $\mu^{(i)}$  og samdreifnifylki  $\{W^{(i)}\}^{-1}$ . Þessum 100,000 vigurum er varpað með sömu vörpun og er notuð til að varpa  $\eta_q$  yfir í  $p_q$  í Gauss-fjölkostalíkaninu í (4) en þá fást tilsvareandi vigrar í hyrnunni. Því næst er meðaltal og samdreifni vigranna í hyrnunni reiknað. Meðaltalið er nálgun á  $\theta^{(i)} = \theta(\mu^{(i)}, W^{(i)})$  þar sem  $\theta$  er skilgreint eins og í (5) og samdreifnin er nálgun á  $\Omega^{(i)} = \Omega(\mu^{(i)}, W^{(i)})$  þar sem  $\Omega$  er skilgreint eins og í (6).

#### 4. Tölfræðileg greining

Hér er farið yfir tölfræðilega greiningu á gögnunum um lengdardreifingu rækju sem lýst var í kafla 2. Líkönunum þremur úr kafla 3 er beitt á gögnin. Safnað var 25,000 handahófsúrtökum frá eftirádreifingu staðlaða fjölkostalíkansins. Fyrir Dirichlet-fjölkostalíkanið voru notaðar 4 MCMC keðjur að lengd 2,000 eftir að leitarferli (e. burn-in) lauk (ákveðið sjónrænt). Fyrir Gauss-fjölkostalíkanið voru

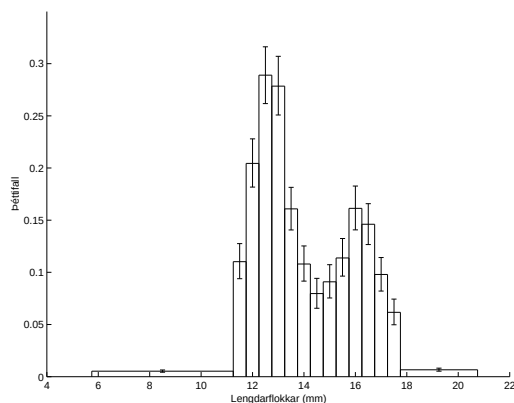
notaðar 4 MCMC keðjur að lengd 250,000 eftir að leitarferli (e. burn-in) lauk og hver 25-ta mæling var geymd. Til að meta hvort keðjurnar í Dirichlet- og Gauss-fjölkostalíkonunum séu nógu langar til að gefa rétta mynd af eftirádreifingum stikanna, var reiknuð Gelman og Rubin (1992) samleitnistærðin fyrir stikana. Byggt á samleitnistærðinni fyrir hvern stika og sjónrænu mati voru keðjurnar taldar gefa rétta mynd af eftirádreifingunum.

Við greininguna voru notaðar fyrirframdreifingar sem hafa mikla dreifni. Fyrirframdreifing  $\theta$  í staðlaða fjölkostalíkaninu er Dirichlet-dreifing með stikavigurinn  $\beta$  þar sem öll stök  $\beta$  eru 0.5. Fyrirframdreifing  $\alpha_j$  í Dirichlet-fjölkostalíkaninu er veldisvísisdreifing með stika  $\gamma_j = 100$ ,  $j = 1, \dots, J$ . Fyrirframdreifing  $\mu$  í Gauss-fjölkostalíkaninu er Gauss-dreifing með meðalgildi  $\gamma_\mu = 0$  og samdreifni  $W_\mu^{-1}$  þar sem  $W_\mu = \delta I_{J-1}$ ,  $\delta = 10^{-5}$  og  $I_{J-1}$  er  $(J-1)$ -vítt einingafylki. Fyrirframdreifing  $W$  í Gauss-fjölkostalíkaninu er óheildanleg (e. improper) Wishart-dreifing, sjá Mardia, Kent og Bibby (1979), bls. 85, en dreifingunni er lýst nánar í viðauka.



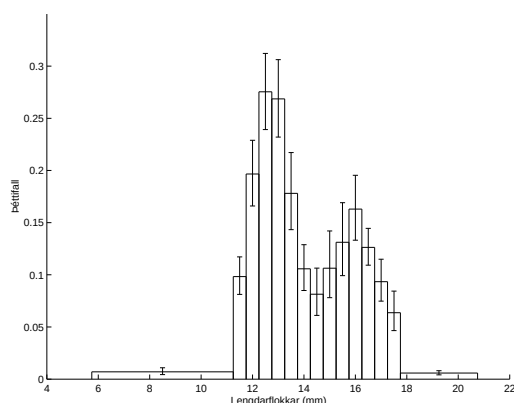
**Mynd 3.** Mat á lengdardreifingu rækju ásamt 90% Bayes-jaðarlíkindasvæðum þegar gert er ráð fyrir að staðlaða fjölkostalíkanið eigi við.

Myndir 3, 4 og 5 sýna lengdardreifinguna þegar gert er ráð fyrir að gögnin komi frá staðlaða fjölkostalíkaninu, Dirichlet-fjölkostalíkaninu og Gauss-fjölkostalíkaninu. Á þessum þremur myndum er hæð súlanna punktmata sem er byggt á væntigildi eftirádreifingar  $\theta$  (e. posterior mean), og skekkjumörkin eru 90% Bayes-jaðarlíkindasvæði (e. Bayesian



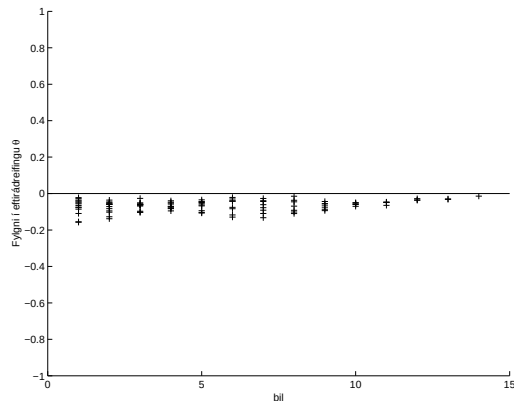
**Mynd 4.** Mat á lengdardreifingu rækju ásamt 90% Bayesjaðarlíkindasvæðum þegar gert er ráð fyrir að Dirichlet-fjölkostalíkanið eigi við.

marginal credible sets). Það ljóst af þessum myndum að punktmatið er ekki það sama þegar gert er ráð fyrir þessum þremur líkönum og að staðlaða fjölkostalíkanið gefur þrengstu jaðarlíkindasvæðin en Gauss-fjölkostalíkanið þau víðustu. Myndir 6, 7 og 8

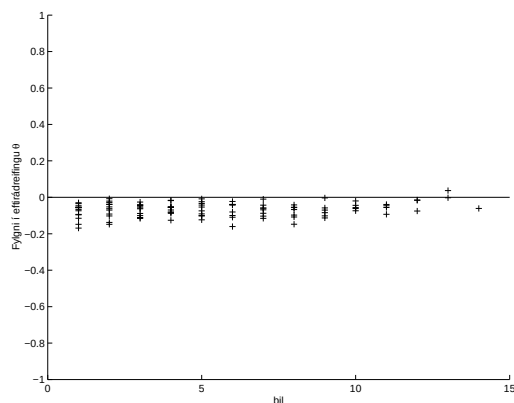


**Mynd 5.** Mat á lengdardreifingu rækju ásamt 90% Bayesjaðarlíkindasvæðum þegar gert er ráð fyrir að Gauss-fjölkostalíkanið eigi við.

sýna fylgni í eftiræðringu  $\theta$  sem fall af fjölda bila á milli flokka fyrir staðlaða fjölkostalíkanið, Dirichlet-fjölkostalíkanið og Gauss-fjölkostalíkanið. Mynd 8 sýnir bæði sterka jákvæða og neikvæða fylgni á meðan myndir 6 og 7 sýna nær eingöngu veika neikvæða fylgni. Því er ljóst að eftiræðring  $\theta$  er mjög ólík undir líkönum þremur þegar horft er á fylgnina.

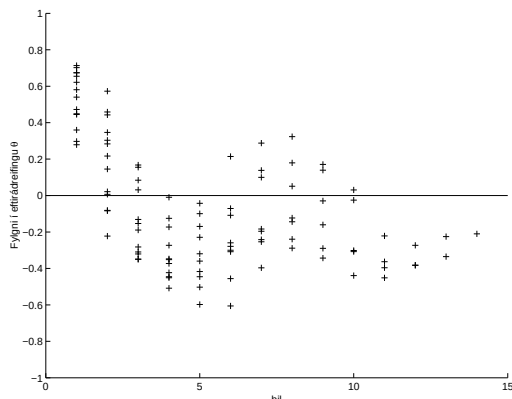


**Mynd 6.** Fylgnin í eftiræðringu  $\theta$  sem fall af bilinu á milli lengdarflokka þegar gert er ráð fyrir að staðlaða fjölkostalíkanið eigi við.

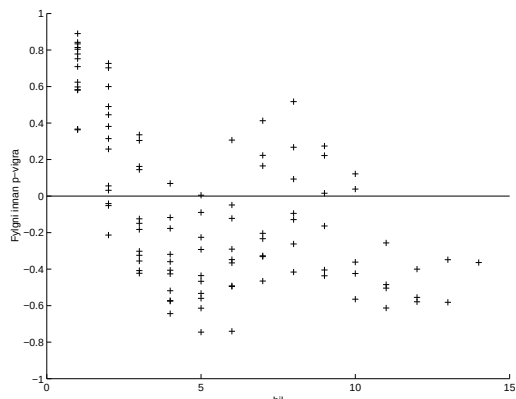


**Mynd 7.** Fylgnin í eftiræðringu  $\theta$  sem fall af bilinu á milli lengdarflokka þegar gert er ráð fyrir að Dirichlet-fjölkostalíkanið eigi við.

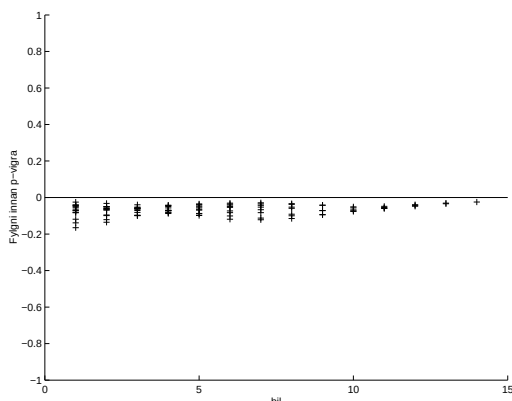
Myndir 9 og 10 sýna mat á fylgninni á milli tíðnivigranna (byggt á miðgildum í eftiræðringunni) þegar gert er ráð fyrir að Dirichlet-fjölkostalíkanið eigi við annars vegar og þegar gert er ráð fyrir að Gauss-fjölkostalíkanið eigi við hins vegar. Þegar gert er ráð fyrir að Dirichlet-fjölkostalíkanið eigi við, þá er fylgnin reiknuð samkvæmt (1) en fylgnin er reiknuð út frá  $\Omega$  í (6) þegar gert er ráð fyrir að Gauss-fjölkostalíkanið eigi við. Þegar myndir 9 og 10 eru bornar saman við mynd 2 er ljóst að það er sterk jákvæð og neikvæð fylgni til staðar í gögnunum sem Dirichlet-fjölkostalíkanið nær ekki að lýsa á meðan Gauss-fjölkostalíkanið virðist vera nægjanlega sveigjanlegt til að lýsa fylgninni í gögnunum viðundandi



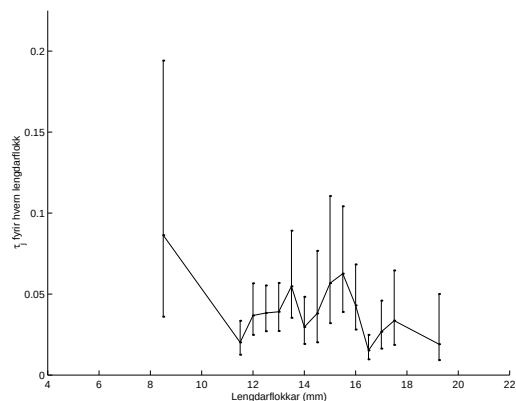
**Mynd 8.** Fylgnin í eftiradreifingu  $\theta$  sem fall af bilinu á milli lengdarflokka þegar gert er ráð fyrir að Gauss-fjölkostalíkanið eigi við.



**Mynd 10.** Mat á fylgninni innan tíðnivigurs sem fall af bilinu á milli lengdarflokka þegar gert er ráð fyrir að Gauss-fjölkostalíkanið eigi við.



**Mynd 9.** Mat á fylgninni innan tíðnivigurs sem fall af bilinu á milli lengdarflokka þegar gert er ráð fyrir að Dirichlet-fjölkostalíkanið eigi við.



**Mynd 11.** Dreifnistikarnir  $\tau_j$ ,  $j = 1, \dots, J$ , fyrir lengdarflokkanum í Gauss-fjölkostalíkaninu, ásamt 90% Bayes-jaðarlíkindasvæðum.

vel. Þess ber að geta að fylgni á mynd 10 er reiknuð út frá mati á  $\Omega$  eingöngu, og er byggð á öllum gögnunum meðan mynd 2 sýnir fylgnina innan hvers  $X_q$  og er byggð á punktmati á  $\text{Var}(X_q) = m_q \Psi + m_q(m_q - 1)\Omega$  og undirsafni úr gögnunum ( $m_q = 120$  fyrir öll  $q$ , alls 40 stöðvar). Mynd 11 sýnir mat á stikunum  $\tau_j$ ,  $j = 1, \dots, J$ , í Gauss-fjölkostalíkaninu, byggt á miðgildi eftiradreifingarinnar og skekkjumörkum byggðum á 90% Bayes-jaðarlíkindasvæðum. Á mynd 11 má sjá að það er ósennilegt að öll  $\tau_j$ ,  $j = 1, \dots, J$ , séu þau sömu þar sem nokkur skekkjumörk skarast ekki. Sambærilegur stíki í Dirichlet-fjölkostalíkaninu er stíkin  $\tau$ . Miðgildi eftiradreifingarinnar  $\tau$  er 0.0209 og 90% Bayes-jaðarlíkindasvæði er (0.0185, 0.0234).

Til að prófa hvort gögnin komi frá Dirichlet-fjölkostalíkani er eftirfarandi tölfræðipróf notað. Gildi úr eftiradreifingu  $\alpha$  eru notuð til að búa til gögn samkvæmt Dirichlet-fjölkostalíkaninu sem hafa sömu stærð og undirsafnið í kafla 2 ( $m_q = 120$  fyrir öll  $q$ , alls 40 stöðvar). Fyrir hvert gagnasafn eru fylgnistuðlar á milli flokka reiknaðir. Fylgnistuðlarnir eru settir í annað veldi og meðaltal þeirra svo tekið. Þá fæst dreifing fyrir meðaltal fylgnistuðlanna í öðru veldi miðað við Dirichlet-fjölkostalíkanið. Meðalgildið á fylgnistuðlunum í öðru veldi sem fæst úr undirsafninu úr kafla 2 er borið saman við dreifinguna og gefur sá samanburður  $p$ -gildi upp á núll. Ef við tökum þessa staðreynd og staðreyndirnar hér að ofan saman



bendir allt til þess að sveigjanlegt líkan eins og Gauss-fjölkostalíkanið sé nauðsynlegt til að lýsa gögnunum nægjanlega vel. Staðlaða fjölkostalíkanið á alls ekki við gögnin og Dirichlet-fjölkostalíkanið er ekki nægjanlega sveigjanlegt til að lýsa gögnunum, sérstaklega þegar kemur að fylgninni.

## 5. Samantekt

Í þessari grein hefur verið kynnt nýtt líkan sem við köllum Gauss-fjölkostalíkanið. Líkanið lýsir vel gögnunum sem var fjallað um í kafla 2, það er flokkunargögn um lengd rækju sem hafa meiri dreifni og sterkari fylgni en staðlaða fjölkostalíkanið segir fyrir um. Dirichlet-fjölkostalíkanið er oft notað til að lýsa gögnum sem hafa meiri dreifni en staðlaða fjölkostalíkanið segir fyrir um. Hér kemur í ljós að Dirichlet-fjölkostalíkanið er ekki nægjanlega sveigjanlegt til að lýsa fylgni gagnanna. Gögn um lengd þorska gáfu sömu niðurstöður, það er, Dirichlet-fjölkostalíkanið átti ekki við gögnin en Gauss-fjölkostalíkanið var nægjanlega sveigjanlegt. Þetta gefur til kynna að Gauss-fjölkostalíkanið sé gagnlegt líkan í fiskifræði.

Þegar eftirádreifing  $\theta$ , hefur verið fundin með MCMC þar sem gert er ráð fyrir að Gauss-fjölkostalíkanið eigi við, má nota úrtakið til að finna nálgun á logaritmanum af sennileikafalli  $\theta$  sem er hluti af stærra líkani. Vigurinn  $\theta$  er hugsanlega stíkaður í stóra líkaninu sem  $\theta = f(\psi)$ , þar sem  $f$  er eitt-hvert fall og  $\psi$  er vigur með stikum. Nálgun á logaritmanum af sennileikafalli  $\psi$  gæti verið á forminu

$$l(\psi) = [h\{\theta\} - E\{h(f(\psi))|m, x\}]^T \\ \times [\text{Var}\{h(\theta)|m, x\}]^{-1} [h\{\theta\} - E\{f(\psi)|m, x\}],$$

þar sem  $h(x)$  er þannig að  $h_j(x) = \log(x_j) - \log(1 - \sum_{k=1}^{J-1} x_k)$ ,  $j = 1, \dots, J-1$ , og  $E\{h(\theta)|m, x\}$  og  $\text{Var}\{h(\theta)|m, x\}$  eru væntigildi og samdreifni eftirádreifingar  $h(\theta)$ . Þó að þessi nálgun byggji á bæði Bayes-tölfræði og sennileikaaðferðum, þá getur nálgun sem þessi verið nauðsynleg þegar um er að ræða stór líkön sem byggja á nokkrum mismunandi gagnasöfnum.

Ef sennileikafall  $\psi$  væri nálgað með jaðareftirádreifingu  $h(\theta)$  byggða á Dirichlet-fjölkostalíkaninu fengist nálgun sem hefði annað hágildi en nálgun byggða á Gauss-fjölkostalíkaninu. Þetta sést þegar myndir 4 og 5 eru bornar saman, meðalgildi  $\theta$  er ekki það sama í þessum tveimur eftirádreifingum og sama gildi um meðalgildi  $h(\theta)$ . Nálgunin tæki heldur ekki

tillit til fylgni í eftirádreifingu  $h(\theta)$ . Fylgnin í eftirádreifingu  $h(\theta)$  er mjög svipuð og fylgnin í eftirádreifingu  $\theta$ . Myndir 7 og 8 sýna eftirádreifingu  $\theta$  þegar gert er ráð fyrir Dirichlet-fjölkostalíkaninu annars vegar og þegar gert er ráð fyrir Gauss-fjölkostalíkaninu hins vegar. Á myndunum má sjá að mikill munur er á fylgninni í eftirádreifingu  $\theta$ . Því er ljóst að ef sennileikafall  $\psi$  væri nálgað með jaðareftirádreifingu  $\theta$  byggða á Dirichlet-fjölkostalíkaninu fengist bjagað mat á  $\psi$ .

## Þakkir

Rannsóknirnar sem liggja hér að baki eru studdar með Evrópubandalagsrannsóknastyrk QLK-CT1999-01609 og voru unnar á meðan Birgir starfaði við Hafrannsóknastofnun.

## Viðauki: Gibbs safnarar

### Dirichlet-fjölkostalíkanið

Skref 1: Skilyrta dreifing  $\alpha_j$  gefið  $m$ ,  $X$ ,  $p$  og  $\alpha_{-j}$ , er

$$\pi(\alpha_j|m, x, p, \alpha_{-j}) \propto \\ \left\{ \prod_{q=1}^Q f(x_q|m_q, p_q) \pi(p_q|\alpha) \right\} \pi(\alpha_j), \quad j = 1, \dots, J.$$

þar sem  $(X_q|m_q, p_q) \sim \text{Mult}_J(m_q, p_q)$ ,  $(p_q|\alpha) \sim \text{Dir}_J(\alpha)$ ,  $q = 1, \dots, Q$ , og fyrirframdreifing  $\alpha_j$  er veldisvísisdreifing  $\alpha_j \sim \text{Exp}(\gamma_j)$ ,  $j = 1, \dots, J$ . Skilyrta dreifing  $\alpha_j$  hefur ekki þekkt form og því er notað Metropolis–Hastings-skref til að draga handahófsúrtak frá dreifingunni, sjá Carlin og Louis (2000) bls. 152-158.

Skref 2: Skilyrta dreifing  $p_q$  gefið  $m$ ,  $X$ ,  $p_{-q}$  og  $\alpha$ , er

$$\pi(p_q|m, x, p_{-q}, \alpha) \propto f(x_q|m_q, p_q) \pi(p_q|\alpha), \\ q = 1, \dots, Q.$$

Margfeldi dreifinganna hér að ofan er Dirichlet-dreifing með stika  $(\alpha + x_q)$ , þannig að

$$(p_q|m, x, p_{-q}, \alpha) \sim \text{Dir}_J(\alpha + x_q), \quad q = 1, \dots, Q.$$

### Gauss-fjölkostalíkanið

Skref 1: Skilyrta dreifing  $\mu$  gefið  $m$ ,  $X$ ,  $W$  og  $\eta$ , er

$$\pi(\mu|m, x, W, \eta) \propto \left\{ \prod_{q=1}^Q \pi(\eta_q|\mu, W) \right\} \pi(\mu)$$

þar sem  $(\eta_q|\mu, W) \sim \text{Gau}(\mu, W^{-1})$ ,  $q = 1, \dots, Q$ , og fyrirframdreifing  $\mu$ ,  $\pi(\mu)$ , er Gauss-dreifing  $\mu \sim \text{Gau}(\gamma_\mu, W_\mu^{-1})$ . Þá verður skilyrta dreifing  $\mu$  Gauss-dreifing

$$(\mu|m, x, W, \eta) \sim \text{Gau}(\{QW + W_\mu\}^{-1} \times \{W \sum_{q=1}^Q \eta_q + W_\mu \gamma_\mu\}, \{QW + W_\mu\}^{-1}).$$

**Skref 2:** Skilyrta dreifing  $W$  gefið  $m, X, \mu$  og  $\eta$ , er

$$\pi(W|m, x, \mu, \eta) \propto \left\{ \prod_{q=1}^Q \pi(\eta_q|\mu, W) \right\} \pi(W)$$

þar sem fyrirframdreifing  $W$  er hér óheildanleg (e. improper) Wishart dreifing, sjá Mardia, Kent og Bibby (1979) bls. 85

$$\pi(W) \propto |W|^{-J/2} \exp\{-\text{tr}(P_0 W)/2\}$$

þar sem  $P_0 = \epsilon I_{J-1}$ ,  $I_{J-1}$  er einingafylki af vídd  $(J-1)$ , og  $\epsilon = 0.01$ . Skilyrta dreifing  $W$  er Wishart dreifing,

$$(W|m, x, \eta, \mu) \sim \text{Wis}_{J-1}(\{\sum_{q=1}^Q (\eta_q - \mu)(\eta_q - \mu)^T + P_0\}^{-1}, Q).$$

Dreifingin hér að ofan er heildanleg (e. proper) Wishart dreifing að því gefnu að  $Q \geq (J-1)$ .

**Skref 3:** Skilyrta dreifing  $\eta_q$  gefið  $m, X, \mu, W$  og  $\eta_{-q}$ , er

$$\pi(\eta_q|m, x, \mu, W, \eta_{-q}) \propto f(x_q|m_q, p_q(\eta_q)) \times \pi(\eta_q|\mu, W), \quad q = 1, \dots, Q.$$

Skilyrta dreifing  $\eta_q$  hefur ekki þekkt form, og því er notað Metropolis–Hastings skref til að draga handa-hófsúrtak frá dreifingunni.

**Summary:** An extension of the multinomial model of counts is presented to account for over-dispersion and different correlation structure. Such models are needed in biological applications such as the analysis of length measurements from surveys of heterogeneous populations, used for assessments of marine resources. One of the goals of such a survey is to estimate the length distribution of pre-specified species within a particular area. Using data on shrimp in

Icelandic waters, it is demonstrated that the assumptions used in practice for categorical length data are seriously violated. The length data on shrimp exhibit variances that are larger than those of the standard multinomial model and correlation coefficients that are greater than those of the Dirichlet-multinomial model. To alleviate these problems a hierarchical model based on the multinomial distribution and the logistically transformed multivariate Gaussian distribution is proposed. It is illustrated that this model captures the complex covariance structure of the data. The parameters in the models are estimated using a Bayesian estimation procedure based on Markov chain Monte Carlo.

## Heimildir

- [1] J. Aitchison (1985), A general class of distributions on the simplex, *Journal of the Royal Statistical Society Ser. B*, **47**, 136-146.
- [2] B. P. Carlin og T. A. Louis (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, 2. útg., Chapman & Hall/CRC.
- [3] G. Casella og R. L. Berger (1990), *Statistical Inference*, Duxbury.
- [4] P. J. Diggle, J. A. Tawn og R. A. Moyeed (1998), Model-based geostatistics, (with discussion), *Applied Statistics*, **47**, 299-350.
- [5] A. Gelman og D. B. Rubin (1992), Inference From Iterative Simulation Using Multiple Sequences, *Statistical Science*, **7**, 457-472.
- [6] W. R. Gilks, S. Richardson og D. J. Spiegelhalter (1996), *Markov Chain Monte Carlo in Practice*, Chapman & Hall.
- [7] L. D. Jacobson, S. Ralston og A. D. MacCall (1996), "Historical larval abundance indices for bocaccio rockfish (*Sebastes paucispinis*) from CalCOFI data," U.S. NOAA, NMFS, SWFSC Skýrsla LJ-96-09, 30 bls.
- [8] N. L. Johnson og S. Kotz (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley.
- [9] P. D. M. MacDonald og T. J. Pitcher (1979), "Age-groups from size-frequency data: A versatile and efficient method of analyzing distribution mixtures," *Journal of the Fisheries Research Board of Canada* **36**, 987-1001.
- [10] K. V. Mardia, J. T. Kent og J. M. Bibby (1979), *Multivariate Analysis*, Academic Press.
- [11] P. McCullagh og J. A. Nelder (1989), *Generalized Linear Models*, 2. útg., Chapman & Hall.
- [12] J. E. Mosimann (1962), On the compound multinomial distribution, the multivariate  $\beta$ -distribution, and correlations among proportions, *Biometrika*, **49**, 65-82.

- [13] M. Pennington (1983), Efficient Estimators of Abundance for Fish and Plankton Surveys, *Biometrics*, **39**, 281-286.
- [14] Gunnar Stefánsson (1996), Analysis of groundfish survey data: Combining the GLM and delta approaches, *ICES Journal of Marine Science*, **53**, 577-588.
- [15] Gunnar Stefánsson (1998), Comparing different information sources in a multispecies context, in *Fishery Stock Assessment Models. Alaska Sea Grant College Program*, AK-SG-98-01, 741-758.
- [16] Gunnar Stefánsson og Ólafur K. Pálsson (1998), A framework for multispecies modelling of Boreal systems, *Reviews in Fish Biology and Fisheries*, **8**, 101-104.
- [17] L. A. Waller, B. P. Carlin, H. Xia og A. E. Gelfand, (1997), Hierarchical Spatiotemporal Mapping of Disease Rates, *Journal of the American Statistical Association*, **92**, 607-617.

**Um höfundana:** Birgir Hrafnkelsson lauk C.S. prófi í vélaverkfræði árið 1993 og M.S. prófi í vélaverkfræði árið 1995 frá verkfræðideild Háskóla Íslands. Birgir hóf nám við tölfræðideild Texas A & M University árið 1995 og lauk þaðan doktorsprófi árið 1999. Verkefnið hans var á sviði tímaráðagreiningar. Frá 1999 til 2000 vann Birgir við rannsóknir við tölfræðideild The Ohio State University. Rannsóknir hans beindust að rúmtölfræði og Bayesískum stigskiptum líkönum. Birgir starfaði við Hafrannsóknastofnun frá 2000 til 2001. Frá 2001 til 2003 vann hann hjá Íslenskri Erfðagreiningu. Birgir starfar nú sem stundakennari við vélaverkfræðiskor verkfræðideildar Háskóla Íslands.

Gunnar Stefánsson lauk B.S. prófi í stærðfræði frá Háskóla Íslands árið 1978 og doktorsprófi í tölfræði frá Ohio State University 1983. Hann var forstöðumaður reiknideildar Hafrannsóknastofnunar 1983-1998, en er nú prófessor í stærðfræði við Háskóla Íslands með rannsóknaaðstöðu á Hafrannsóknastofnun. Hann stundar rannsóknir á sviði tölfræðilegrar líkanagerðar með áherslu á fjölstofnarannsóknir.

---

Verkfræðideild Háskóla Íslands  
Hjarðarhaga 6  
IS-107 Reykjavík  
birgirhr@hi.is

Raunvísindastofnun Háskólans  
Dunhaga 3  
IS-107 Reykjavík  
gunnar@hi.is

Móttékin: 14. maí 2002