

Uskottavuusfunktion approksimointi gaussisella
variaatiomenetelmällä Poisson-sekamallin ja -latentin
muuttujamallin tapauksessa

Viivi Haapaniemi

Tilastotieteen pro gradu -tutkielma

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
26. huhtikuuta 2017

JYVÄSKYLÄN YLIOPISTO

Matematiikan ja tilastotieteen laitos

Haapaniemi, Viivi: Uskottavuusfunktion approksimointi gaussisella variaatiomenetelmällä Poisson-sekamallin ja -latentin muuttujamallin tapauksessa

Tilastotieteen pro gradu -tutkielma, 19 sivua + liitteet 24 sivua, 26. huhtikuuta 2017

Tiivistelmä:

Tilastollisen mallin parametrit estimoidaan usein suurimman uskottavuuden menetelmällä. Havaitulla aineistolla ja voimassa olevan mallin vallitessa suurimman uskottavuuden estimointi valitsee malliparametreille arvot, jotka maksimoivat uskottavuusfunktion. Uskottavuusfunktion ollessa integraali yli latenttien eli havaitsemattomien muuttujien, ei uskottavuusfunktiota voida kuitenkaan kirjoittaa suljetussa muodossa. Tällöin uskottavuusfunktio voidaan approksimoida variaatiomenetelmällä.

Variaatiomenetelmässä uskottavuusfunktiolle lasketaan alaraja, jota maksimoidalla saadaan approksimaatiot malliparametrien suurimman uskottavuuden estimaattoreille. Variaatiomenetelmä tuottaa malliparametrien lisäksi ennusteet ja niiden keskivirheet latenteille muuttujille.

Tutkimuksessa sovelletaan variaatiomenetelmää Poisson-sekamalliin ja -latenttiin muuttujamalliin ordinaatiomenetelmälle. Sekamallilla pyritään selittämään simuloinneissa eri lajien määrää ja aineistoesimerkissä yhden lajin yksilömäärää eri mittauspaikoilla. Ordinaatiomenetelmässä pyritään havainnollistamaan aineisto kaksiulotteisena kuvana, jossa lajien suhteen samankaltaiset mittauspaikat ovat lähellä toisiaan. Molemmista tapauksista implementoidaan menetelmä R-ohjelmistolle. Variaatiomenetelmän soveltuvuutta laskenta-algoritmiksi tutkitaan simulointikokein. Lisäksi sekamalli ja latentti muuttujamalli sovitetaan biologian lukumääräaineistoihin, joista lasketaan estimaatit malliparametreille variaatiomenetelmällä.

Työssä osoitetaan että Poisson-jakauman tapauksessa variaatiomenetelmän mukainen uskottavuusfunktion alaraja voidaan kirjoittaa suljetussa muodossa, jolloin parametrien estimointi on yksinkertaista. Poisson-sekamallin tapauksessa lajien määrän lisääminen antaa tarkemman variaatioapproksimaation. Latentissa muuttujamallissa variaatiomenetelmä antaa hyvillä alkuarvoilla paremman tuloksen kuin perinteisesti käytetty moniulotteiseen skaalaukseen perustuva NMDS-menetelmä. Kuitenkin variaatiomenetelmä näyttäisi latentin muuttujamallin kohdalla reagoivan herkästi huonoihin alkuarvoihin. Kun Poisson-sekamalli sovitetaan lukumääräaineistoon, saadaan selville, että estimaatit ovat samansuuntaiset kuin perinteisesti käytetyllä Laplace-menetelmällä.

Avainsanoja: uskottavuuspäätely, variaatiomenetelmä, sekamalli, latentti muuttuja, ordinaatiomenetelmä, Poisson-jakauma.

Sisältö

1 Johdanto	1
2 Variaatiomenetelmä	2
2.1 Uskottavuusfunktion alaraja	2
2.2 Estimaattien keskivirheiden approksimointi	4
2.3 Satunnaisvaikutusten parhaimman ennusteen approksimaatio	4
3 Variaatiomenetelmän sovelluksia	5
3.1 Yksinkertainen Poisson-sekamalli	5
3.2 Latentti muuttujamalli Poisson-vasteille	7
4 Simulointi	9
4.1 Sekamalli	9
4.2 Latentti muuttujamalli	10
5 Esimerkki	13
5.1 Perhosaineisto	13
5.2 Hämähäkkiaineisto	15
6 Pohdinta	17
Viitteet	19
Liitteet	20
A Kaavan 2 todistus	20
B Log-uskottavuuden alarajan derivaatat Poisson-sekamallille	22
C Kaavan 3 todistus	23
D Alarajan derivaatat latentissa muuttujamallissa	27
E R-koodi	29

1 Johdanto

Uskottavuuspäätely on tilastotieteessä laajalti käytetty menetelmä parametrien estimoimiseksi tilastollisessa mallissa. Havaitulla aineistolla ja voimassa olevan mallin vallitessa suurimman uskottavuuden estimointi valitsee malliparametreille arvot, jotka maksimoivat uskottavuusfunktion. On kuitenkin tilanteita, joissa uskottavuusfunktiota ei voida kirjoittaa suljetussa muodossa. Näin voi käydä esimerkiksi kun uskottavuus sisältää integraalin latentin, havaitsemattoman, muuttujan yli. Tällöin sitä täytyy approksimoida esimerkiksi Laplace-approksimaatiolla, MCMC-menetelmällä tai (stokastisen) EM-algoritmin keinoin. Tässä esitelmässä uskottavuusfunktiota approksimoidaan variaatiomenetelmällä.

Variaatiomenetelmä on työkalu uskottavuuden approksimoimiseksi tilastollisissa malleissa. Variaatiomenetelmää on sovellettu muun muassa puheentunnistuksessa (Jordan, 2004) ja neuroverkoissa (Titterton, 2004). Tilastotieteessä variaatiomenetelmää voidaan soveltaa sekä uskottavuuspäätelyyn että Bayes-mallinnukseen. Jälkimmäisessä, jossa vaikeasti käsiteltäviä laskentaongelmia on runsaasti, menetelmää on käytetty enemmän. Bayes-mallinnuksessa variaatiomenetelmällä approksimoidaan posteriorijakaumaa toisilla laskennallisesti helpoilla jakaumilla. Frekventistisessä lähestymistavassa logaritmiselle uskottavuusfunktiolle lasketaan alaraja, jota maksimoidaan uskottavuuden sijaan. (Ormerod ja Wand, 2010.)

Tässä tutkielmassa sovelletaan variaatiomenetelmää yleistettyyn lineaariseen sekamalliin ja latenttiin muuttujamalliin ordinaatiomenetelmälle. Kummassakin tapauksessa uskottavuusfunktion tulee integraali latentin, todellisuudessa havaitsemattoman, muuttujan suhteen eikä uskottavuusfunktiota voida ratkaista suljetussa muodossa. Variaatiomenetelmän yksityiskohdat frekventistiselle lähestymistavalle on esitetty luvussa 2. Variaatiomenetelmän sopivuutta laskenta-algoritmiksi tutkitaan simulointikokein luvussa 4. Sekamalli ja latentti muuttujamalli sovitetaan luvussa 5 lukumääräaineistoihin, joista lasketaan variaatiomenetelmän mukaiset approksimaatiot malliparametrien suurimman uskottavuuden estimaateille. Mallidiagnostiikka suoritetaan jäännöstarkasteluilla.

2 Variaatiomenetelmä

Variaatiomenetelmää voidaan soveltaa parametrien approksimoimiseen tilastollisissa malleissa. Ormerod ja Wand (2010) soveltavat variaatiomenetelmää uskottavuuspäätelyyn. He toteavat variaatiomenetelmästä olevan hyötyä niiden mallien kohdalla, joissa uskottavuus sisältää latentin muuttujan.

2.1 Uskottavuusfunktion alaraja

Tarkastellaan perusajatusta, mihin variaatiomenetelmä perustuu. Ormerod ja Wand (2010) muotoilevat asian seuraavasti. Oletetaan, että \mathbf{y} on havaintovektori ja \mathbf{u} vektori, joka sisältää latentit muuttujat. Käytetään jatkossa tiheydelle geneeristä merkintää $p(\cdot)$, missä p on tilanteen mukaan joko satunnaismuuttujan tiheysfunktio tai pistetodennäköisyysfunktio. Nyt logaritminen uskottavuusfunktio (kiinteälle) malliparametrivektorille $\boldsymbol{\theta}$ on muotoa

$$l(\boldsymbol{\theta}) = \log p(\mathbf{y}; \boldsymbol{\theta}) = \log \int p(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta}) d\mathbf{u}$$

ja parametrivektorin $\boldsymbol{\theta}$ suurimman uskottavuuden estimaattori on

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} l(\boldsymbol{\theta}).$$

Uskottavuusfunktion sisältämän integraalin vuoksi sitä ei välttämättä pystytä kirjoittamaan suljetussa muodossa. Olkoon q satunnaisvektorin \mathbf{u} tiheysfunktio. Voidaan kirjoittaa

$$l(\boldsymbol{\theta}) = \log p(\mathbf{y}; \boldsymbol{\theta}) \cdot \int q(\mathbf{u}) d\mathbf{u} = \int q(\mathbf{u}) \cdot \log p(\mathbf{y}; \boldsymbol{\theta}) d\mathbf{u}.$$

Kirjoitetaan ehdollisen todennäköisyysjakauman $\mathbf{u} | \mathbf{y}; \boldsymbol{\theta}$ tiheysfunktio muotoon

$$p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta}) = \frac{p(\mathbf{u}, \mathbf{y}; \boldsymbol{\theta})}{p(\mathbf{y}; \boldsymbol{\theta})} \Leftrightarrow p(\mathbf{y}; \boldsymbol{\theta}) = \frac{p(\mathbf{u}, \mathbf{y}; \boldsymbol{\theta})}{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})}.$$

Nyt logaritminen uskottavuus voidaan kirjoittaa

$$\begin{aligned} l(\boldsymbol{\theta}) &= \int q(\mathbf{u}) \cdot \log \left\{ \frac{p(\mathbf{u}, \mathbf{y}; \boldsymbol{\theta})}{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})} \right\} d\mathbf{u} = \int q(\mathbf{u}) \cdot \log \left\{ \frac{p(\mathbf{u}, \mathbf{y}; \boldsymbol{\theta})/q(\mathbf{u})}{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})/q(\mathbf{u})} \right\} d\mathbf{u} \\ &= \int q(\mathbf{u}) \cdot \left(\log \left\{ \frac{p(\mathbf{u}, \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{u})} \right\} + \log \left\{ \frac{q(\mathbf{u})}{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})} \right\} \right) d\mathbf{u} \\ &= \int q(\mathbf{u}) \cdot \log \left\{ \frac{p(\mathbf{u}, \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{u})} \right\} d\mathbf{u} + \int q(\mathbf{u}) \cdot \log \left\{ \frac{q(\mathbf{u})}{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})} \right\} d\mathbf{u} \\ &\geq \int q(\mathbf{u}) \cdot \log \left\{ \frac{p(\mathbf{u}, \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{u})} \right\} d\mathbf{u}. \end{aligned}$$

Viimeisin epäyhtälö pitää paikkansa, sillä kaikille tiheysfunktioille q pätee

$$\int q(\mathbf{u}) \cdot \log \left\{ \frac{q(\mathbf{u})}{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})} \right\} d\mathbf{u} = \mathbb{E}_{q(\mathbf{u})} \left[\log \left\{ \frac{q(\mathbf{u})}{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})} \right\} \right] = \mathbb{E}_{q(\mathbf{u})} \left[-\log \left\{ \frac{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{u})} \right\} \right].$$

Koska funktio $-\log$ on konvekksi, seuraa Jensenin epäyhtälöstä

$$\begin{aligned} \mathbb{E}_{q(\mathbf{u})} \left[-\log \left\{ \frac{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{u})} \right\} \right] &\geq -\log \left\{ \mathbb{E}_{q(\mathbf{u})} \left[\frac{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{u})} \right] \right\} \\ &= -\log \int q(\mathbf{u}) \cdot \frac{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{u})} d\mathbf{u} = -\log \int p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta}) d\mathbf{u} \\ &= -\log 1 = 0. \end{aligned}$$

Edellisen epäyhtälön yhtäsuuruus pätee jos ja vain jos $q(\mathbf{u}) = p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})$ melkein kaikkialla (Kullback ja Leibler, 1951). Integraali

$$\int q(\mathbf{u}) \cdot \log \left\{ \frac{q(\mathbf{u})}{p(\mathbf{u} | \mathbf{y}; \boldsymbol{\theta})} \right\} d\mathbf{u}$$

tunnetaan myös nimellä Kullback-Leibler -etäisyys tiheysfunktioiden q ja $p(\cdot | \mathbf{y})$ välillä.

Uskottavuusfunktion logaritmillemme saadaan siis jakaumasta q riippuva alaraja

$$\underline{l}(\boldsymbol{\theta}; q) := \int q(\mathbf{u}) \cdot \log \left\{ \frac{p(\mathbf{u}, \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{u})} \right\} d\mathbf{u}.$$

Seuraavaksi valitaan q siten, että alaraja $\underline{l}(\boldsymbol{\theta}; q)$ saadaan helpommin käsiteltävään muotoon ja samalla pyritään minimoimaan Kullback-Leibler -etäisyys jakaumien q ja $p(\cdot | \mathbf{y})$ välillä. Korvataan yksi q jakaumaperheellä $q(\cdot; \boldsymbol{\xi})$. Kutsutaan tiheyden q valitsemisesta syntyneitä parametreja $\boldsymbol{\xi}$ variaatioparametreiksi. Alaraja saa nyt muodon

$$\underline{l}(\boldsymbol{\theta}, \boldsymbol{\xi}; q) := \int q(\mathbf{u}; \boldsymbol{\xi}) \cdot \log \left\{ \frac{p(\mathbf{u}, \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{u}; \boldsymbol{\xi})} \right\} d\mathbf{u}.$$

Nyt maksimointi täytyy suorittaa ensinnäkin variaatioparametrien $\boldsymbol{\xi}$ suhteen, jotta saadaan minimoitua Kullback-Leibler -etäisyys jakaumien q ja $p(\cdot | \mathbf{y})$ välillä. Toiseksi, maksimointi tulee suorittaa malliparametrien $\boldsymbol{\theta}$ suhteen, jotta saadaan maksimoitua approksimaatio logaritmiselle uskottavuudelle. Saadaan

$$\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\xi}} \right) = \underset{\boldsymbol{\theta}, \boldsymbol{\xi}}{\operatorname{argmax}} \underline{l}(\boldsymbol{\theta}, \boldsymbol{\xi}; q).$$

Alarajan $\underline{l}(\boldsymbol{\theta}, \boldsymbol{\xi}; q)$ maksimoiva $\hat{\boldsymbol{\theta}}$ on variaatioapproksimaatio suurimman uskottavuuden estimaattorille $\hat{\boldsymbol{\theta}}$.

Variaatioapproksimaation tarkkuudesta on kirjoitettu vain vähän. Hall et al. (2011) sekä Ormerod ja Wand (2012) mainitsevat muutamia tuloksia tapaukselle, jossa q on valittu multinormaalijakautuneeksi. Tarkastellaan seuraavaksi, kuinka voidaan laskea approksimaatiot estimaattien keskivirheille ja ennusteet satunnaisvaikutuksille, kun q on multinormaalijakauman $N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ tiheysfunktio. Tällöin siis $\boldsymbol{\xi} = (\boldsymbol{\mu}, \boldsymbol{\Lambda})$.

2.2 Estimaattien keskivirheiden approksimointi

Ormerod ja Wand (2012) tarjoavat approksimaation estimaattien keskivirheille. Olkoon $\underline{I}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \mathbb{E}[-\mathbf{H} \underline{l}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda})]$, missä \underline{I} on variaatioapproksimaatio Fisherin informaatiomatriisille ja \mathbf{H} on Hessen matriisioperaattori vektorin $(\boldsymbol{\theta}, \boldsymbol{\mu}, \text{vech } \boldsymbol{\Lambda})$ suhteen. Tässä $\text{vech } \boldsymbol{\Lambda}$ merkitsee matriisiin $\boldsymbol{\Lambda}$ vektorointia, joka suoritetaan matriisiin $\boldsymbol{\Lambda}$ alakolmio-osuudelle. Jos alarajaa $\underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ käsitellään kuten logaritmista uskottavuutta, ja variaatioparametreja $\boldsymbol{\mu}$ ja $\boldsymbol{\Lambda}$, kuten kiusaparametreja, niin uskottavuusteorian mukaan suurimman uskottavuuden estimaattorin variaatioapproksimaation $\hat{\boldsymbol{\theta}}$ asymptoottinen kovarianssimatriisi on Fisherin informaatiomatriisin käänteismatriisin parametria $\boldsymbol{\theta}$ vastaava osamatriisi. Keskivirheiden approksimaatiot saadaan asymptoottisen kovarianssimatriisin diagonaalialkioiden neliöjuurista, kun matriisiin sijoitetaan parametrien variaatioapproksimaatiot. Tässä tutkielmassa keskivirheet on laskettu numeerisesti käyttäen R-paketin `numDeriv` (Gilbert ja Varadhan, 2016) funktiota `hessian`.

2.3 Satunnaisvaikutusten parhaimman ennusteen approksimaatio

Satunnaisvaikutusten \mathbf{u} ennusteet ovat usein kiinnostuksen kohteena, esimerkiksi tehtäessä jäännöspohjaista mallidiagnostiikkaa. Ormerod ja Wand (2012) mukaan satunnaisvaikutusten \mathbf{u} paras ennuste (*best predictor*) on

$$\text{BP}(\mathbf{u}) = \mathbb{E}[\mathbf{u}|\mathbf{y}] = \int \mathbf{u} p(\mathbf{u}|\mathbf{y}) d\mathbf{u}.$$

Integraali on yleistetyin lineaarisen sekamallin tapauksessa vaikeasti käsiteltävä. Kuitenkin, maksimointi variaatioparametrien yli tarkoittaa Kullback-Leibler -etäisyyden minimoimista jakaumien $q(\mathbf{u})$ ja $p(\mathbf{y})$ välillä. Gaussiselle variaatiomenetelmälle $q(\mathbf{u}) = q(\cdot; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ on multinormaalijakauman $N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ tiheysfunktio, joten

$$\underline{\text{BP}}(\mathbf{u}) = \int \mathbf{u} q(\mathbf{u}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Lambda}}) = \hat{\boldsymbol{\mu}} \quad (1)$$

on sopiva approksimaatio parhaalle ennusteelle $\underline{\text{BP}}(\mathbf{u})$, kun $\hat{\boldsymbol{\mu}}$ ja $\hat{\boldsymbol{\Lambda}}$ ovat logaritmisien uskottavuusfunktion maksimoivat estimaatit variaatioparametreille.

Seuraavaksi tarkastellaan satunnaisvaikutusten parhaimman ennusteen approksimaation vaihtelua. McCullouch et al. (2008) tarjoavat tuloksen $\text{Cov}(\text{BP}(\mathbf{u}) - \mathbf{u}) = \mathbb{E}_{\mathbf{y}}[\text{Cov}(\mathbf{u}|\mathbf{y})]$. Kun korvataan $p(\mathbf{u}|\mathbf{y})$ tiheysfunktiolla $q(\mathbf{u}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Lambda}})$, saadaan estimoitu asymptoottinen kovarianssimatriisi

$$\widehat{\text{Cov}}(\underline{\text{BP}}(\mathbf{u}) - \mathbf{u}) = \hat{\boldsymbol{\Lambda}}.$$

Siispä, maksimoituja variaatioparametreja $\hat{\boldsymbol{\mu}}$ ja $\hat{\boldsymbol{\Lambda}}$ voidaan käyttää satunnaisvaikutusten ennustamiseen ja niiden vaihtelun mittaamiseen.

3 Variaatiomenetelmän sovelluksia

Yleistetty lineaarinen sekamalli ja latentti muuttujamalli ovat tapauksia, joissa uskottavuusfunktio sisältää integraalin latentin eli havaitsemattoman muuttujan yli, eikä uskottavuusfunktiota tällöin voida kirjoittaa suljetussa muodossa. Käydään seuraavassa läpi kaksi esimerkkiä, joissa uskottavuusfunktion approksimoimiseksi käytetään variaatiomenetelmää. Toimitaan kuten Hall et al. (2011) sekä Ormerod ja Wand (2012), eli valitaan molemmissa esimerkeissä tiheydeksi q tulo normaalijakaumien tiheysfunktioista ja käytetään menetelmästä tällöin nimeä gaussinen variaatiomenetelmä.

3.1 Yksinkertainen Poisson-sekamalli

Tarkastellaan yleistettyä lineaarista sekamallia, jossa vastemuuttuja oletetaan Poisson-jakautuneeksi ja satunnaistekijöistä mukana on satunnainen taso. Olkoon y_{ij} yksilön i , $i = 1, 2, \dots, m$, vasteen j , $j = 1, 2, \dots, n_i$, arvo. Kuhunkin vasteeseen liittyy k selittäjän arvo x_{kij} , missä $k = 1, 2, \dots, p$. Myöhempien lukujen simuloinnissa yksilönä on laji ja perhosaineistoesimerkissä mittauspaikka. Oletetaan, että

$$y_{ij}|u_i \sim Pois(\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + u_i))$$

ja

$$u_i \sim N(0, \sigma^2),$$

missä $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_p)$ on selittäviin muuttujiin liittyvä parametrivektori ja $\mathbf{x}'_{ij} = (1, x_{1ij}, x_{2ij}, \dots, x_{pij})$ ovat selittäjien arvot.

Latentit muuttujat, eli satunnaiset vaikutukset, u_i oletetaan riippumattomiksi. Vastemuuttujan arvot y_{ij} ehdolla latentit muuttujat u_i oletetaan myös riippumattomiksi. Nyt

$$p(y_{ij}|u_i; \boldsymbol{\beta}, \sigma^2) = \frac{(\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + u_i))^{y_{ij}}}{y_{ij}!} \cdot \exp(-\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + u_i))$$

ja

$$p(u_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{u_i^2}{2\sigma^2}\right).$$

Tällöin logaritmi uskottavuusfunktiosta voidaan kirjoittaa muodossa (McCullouch ja Searle, 2001)

$$l(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^m \log \int \prod_{j=1}^{n_i} \frac{(e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + u_i})^{y_{ij}}}{y_{ij}!} \exp(-e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + u_i}) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u_i^2}{2\sigma^2}\right) du_i.$$

Logaritminen uskottavuusfunktio sisältää hankalan integraalin, joten lasketaan

sille alaraja. Alaraja on muotoa (Hall et al. (2011); Ormerod ja Wand (2010))

$$\begin{aligned} \underline{l}(\boldsymbol{\beta}, \sigma^2; q) &= \int q(\mathbf{u}) \cdot \log \left\{ \frac{p(\mathbf{u}, \mathbf{y}, \boldsymbol{\beta}, \sigma^2)}{q(\mathbf{u})} \right\} d\mathbf{u} \\ &= \int_{\mathbb{R}^m} \left(\sum_{i=1}^m \left[\sum_{j=1}^{n_i} \left\{ y_{ij} (\mathbf{x}'_{ij} \boldsymbol{\beta} + u_i) - \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + u_i) - \log(y_{ij}!) \right\} - \frac{u_i^2}{2\sigma^2} \right] \right. \\ &\quad \left. - \frac{m}{2} \log(2\pi\sigma^2) - \log(q(u_1, \dots, u_m)) \right) \times q(u_1, \dots, u_m) du_1 \dots du_m. \end{aligned}$$

Valitaan $q(u_1, \dots, u_m) = q_1(u_1) \times \dots \times q_m(u_m)$ siten, että $u_i \sim N(\mu_i, \lambda_i)$, $\lambda_i > 0$ kaikilla $i = 1, 2, \dots, m$. Siis

$$q(u_1, \dots, u_m) = q_1(u_1) \times \dots \times q_m(u_m) = \prod_{i=1}^m (2\pi\lambda_i)^{-\frac{1}{2}} \exp\left(-\frac{(u_i - \mu_i)^2}{2\lambda_i}\right).$$

Merkitään $\boldsymbol{\mu}' = (\mu_1, \dots, \mu_m)$ ja $\boldsymbol{\lambda}' = (\lambda_1, \dots, \lambda_m)$.

Logaritmissen uskottavuusfunktion alaraja voidaan kirjoittaa muodossa

$$\begin{aligned} \underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}; q) &= \mathbb{E}_{q(\mathbf{u})} \left[\sum_{i=1}^m \left[\sum_{j=1}^{n_i} \left\{ y_{ij} (\mathbf{x}'_{ij} \boldsymbol{\beta} + u_i) - \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + u_i) - \log(y_{ij}!) \right\} - \frac{u_i^2}{2\sigma^2} \right] \right. \\ &\quad \left. - \frac{m}{2} \log(2\pi\sigma^2) - \log(q(u_1, \dots, u_m)) \right]. \end{aligned}$$

Tämä sievenee muotoon

$$\begin{aligned} \underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}; q) &= \sum_{i=1}^m \sum_{j=1}^{n_i} \left[y_{ij} (\mathbf{x}'_{ij} \boldsymbol{\beta} + \mu_i) - \exp\left(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mu_i + \frac{1}{2}\lambda_i\right) - \log(y_{ij}!) \right] \\ &\quad + \frac{m}{2} (1 - \log(\sigma^2)) + \frac{1}{2} \sum_{i=1}^m \left[\log(\lambda_i) - \frac{\lambda_i + \mu_i^2}{\sigma^2} \right]. \end{aligned} \quad (2)$$

Yksityiskohtaisempi todistus kaavalle 2 löytyy liitteestä A. (Hall et al., 2011; Ormerod ja Wand, 2010.)

Kuten aliluvussa 2.1 on esitetty, log-uskottavuusfunktion maksimointi parametrien $\boldsymbol{\beta}$ ja σ^2 suhteen antaa suurimman uskottavuuden estimaateille $\hat{\boldsymbol{\beta}}$ ja $\hat{\sigma}^2$ variaatioaprossimaatiot. Estimaattien keskihajonnat voidaan laskea kuten aliluvussa 2.2. Parhaat ennusteet satunnaisvaikutuksille sekä näiden keskihajonnat saadaan maksimoimalla log-uskottavuusfunktio parametrien $\boldsymbol{\mu}$ ja $\boldsymbol{\lambda}$ suhteen. Tässä työssä log-uskottavuuden maksimointiin käytetään kvasi-Newton -menetelmää (Broyden, 1967). Menetelmä käyttää apunaan optimoitavan funktion derivaattoja. Kaavan 2 mukaisen alarajan derivaatat löytyvät liitteestä B.

3.2 Latentti muuttujamalli Poisson-vasteille

Tarkastellaan yleistettyä lineaarista latenttia muuttujamallia, jossa vastemuuttuja on Poisson-jakautunut. Seuraava perustuu Huber et al. (2004) esitykseen. Oletetaan, että aineistossa on p vastemuuttujaa, esimerkiksi lajia, ja lajien lukumäärä on mitattu n eri paikasta. Olkoon y_{ij} vastemuuttujan arvo lajille j paikalla i , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$. Latenteja muuttujia $\mathbf{u}'_i = (u_{i1}, \dots, u_{im})$ paikalla $i = 1, 2, \dots, n$, ei havaita. Ne oletetaan kuitenkin riippumattomiksi ja standardinormaalijakautuneiksi. Oletetaan nyt

$$y_{ij} | \mathbf{u}_i \sim \text{Pois}(\eta_{ij}),$$

missä $\eta_{ij} = \exp(\alpha_i + \beta_j + \mathbf{u}'_i \boldsymbol{\gamma}_j)$. Parametrin α_i avulla vertaillaan eri mittauspaikkojen keskiarvoja ja parametrin β_j avulla voidaan vertailla eri lajien keskiarvoja. Lajiin liittyvä kerroin $\boldsymbol{\gamma}'_j = (\gamma_{j1}, \dots, \gamma_{jm})$ ilmaisee, kuinka kukin laji on yhteydessä latentteihin muuttujiin.

Latenttia muuttujamallia voidaan käyttää mallipohjaisena ordinaatiomenetelmänä. Ordinaatiomenetelmissä pyritään havainnollistamaan aineisto kaksiulotteisena kuvana, jossa lajien suhteen samankaltaiset mittauspaikat ovat lähellä toisiaan. Nyt kuva voidaan muodostaa kaksiulotteisten ($m = 2$) latenttien muuttujien ennusteiden hajontakuvana (Hui et al., 2015). Merkitään nyt $\boldsymbol{\alpha}' = (\alpha_1, \dots, \alpha_n)$, $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$, $\boldsymbol{\Gamma}' = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)$ ja $\boldsymbol{\theta} := (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma})$.

Vastemuuttujan arvot y_{ij} ehdolla latentit muuttujat \mathbf{u}_i oletetaan riippumattomiksi. Näin niiden yhteisjakauma on muotoa

$$p_i(\mathbf{y}_i | \mathbf{u}_i; \boldsymbol{\theta}) = \prod_{j=1}^p p_{ij}(y_{ij} | \mathbf{u}_i; \boldsymbol{\theta}),$$

missä

$$p_{ij}(y_{ij} | \mathbf{u}_i; \boldsymbol{\theta}) = \frac{(\eta_{ij})^{y_{ij}}}{y_{ij}!} \cdot \exp(-\eta_{ij}).$$

Vasteen \mathbf{y}_i ja latenttien muuttujien \mathbf{u}_i yhteisjakauma on tällöin

$$p_i(\mathbf{y}_i, \mathbf{u}_i; \boldsymbol{\theta}) = \left(\prod_{j=1}^p p_{ij}(y_{ij} | \mathbf{u}_i; \boldsymbol{\theta}) \right) \cdot p(\mathbf{u}_i)$$

missä $p(\mathbf{u}_i)$ on latentin muuttujan tiheysfunktio. Koska vektorit \mathbf{u}_i oletettiin m -ulotteisiksi standardinormaalijakautuneiksi ja keskenään riippumattomiksi, niin

$$p(\mathbf{u}_i) = (2\pi)^{-\frac{m}{2}} \exp\left(-\frac{1}{2} \mathbf{u}'_i \mathbf{u}_i\right).$$

Koska muuttujien \mathbf{u}_i arvoja ei havaita, $p_i(\mathbf{y}_i, \mathbf{u}_i; \boldsymbol{\theta})$ integroidaan muuttujien \mathbf{u}_i jakauman yli, toisin sanoen,

$$p_i(\mathbf{y}_i; \boldsymbol{\theta}) = \int \left(\prod_{j=1}^p p_{ij}(y_{ij} | \mathbf{u}_i; \boldsymbol{\theta}) \right) \cdot p(\mathbf{u}_i) d\mathbf{u}_i.$$

Muuttujien $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ ollessa riippumattomia logaritminen uskottavuusfunktio voidaan kirjoittaa muodossa

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{i=1}^n \log(p_i(\mathbf{y}_i; \boldsymbol{\theta})) \\ &= \sum_{i=1}^n \log \int \left(\prod_{j=1}^p \frac{(\eta_{ij})^{y_{ij}}}{y_{ij}!} \cdot \exp(-\eta_{ij}) \right) (2\pi)^{-\frac{m}{2}} \exp\left(-\frac{1}{2} \mathbf{u}_i' \mathbf{u}_i\right) d\mathbf{u}_i. \end{aligned}$$

Koska tämän ratkaiseminen suljetussa muodossa ei onnistu, lasketaan sille alaraja variaatiomenetelmällä. Valitaan

$$q(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = q_1(\mathbf{u}_1) \times q_2(\mathbf{u}_2) \times \dots \times q_n(\mathbf{u}_n)$$

siten, että jokainen $\mathbf{u}_i \sim N_m(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)$, missä keskiarvo on $\boldsymbol{\mu}_i' = (\mu_{i1}, \dots, \mu_{im})$ ja kovarianssimatriisi on diagonaalimatriisi $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{im})$ kaikilla $i = 1, 2, \dots, n$. Tällöin variaatiomenetelmän mukainen alaraja logaritmiselle uskottavuudelle on

$$\begin{aligned} \underline{l}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}; q) &= \sum_{i=1}^n \sum_{j=1}^p \left[y_{ij} (\alpha_i + \beta_j + \boldsymbol{\mu}_i' \boldsymbol{\gamma}_j) - \log(y_{ij}!) - \exp\left(\alpha_i + \beta_j + \boldsymbol{\mu}_i' \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}_j' \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right) \right] \\ &+ \frac{m \cdot n}{2} + \frac{1}{2} \sum_{i=1}^n [\log(\det(\boldsymbol{\Lambda}_i)) - \text{trace}(\boldsymbol{\Lambda}_i) - \boldsymbol{\mu}_i' \boldsymbol{\mu}_i]. \end{aligned} \quad (3)$$

Todistus kaavalle 3 on liitessä C. Yleisemmän muodon alarajalle tarjoaa Hui et al. (2017).

Kuten sekamallien tapauksessa, uskottavuusfunktion alaraja voidaan maksimoida Kvasi-Newton -menetelmällä. Kun seuraavien lukujen simuloinneissa ja esimerkeissä maksimoidaan alarajaa, kiinnitetään $\boldsymbol{\gamma}'_1 = (\gamma_{11}, 0)$. Tällöin $\boldsymbol{\Gamma}$ on täysin määritelty lukuunottamatta jokaisen sarakkeen etumerkkiä (Huber et al., 2004). Alarajan derivaatat löytyvät liitteestä D. Latentit muuttujat voidaan laskea käyttämällä variaatioparametrien estimaatteja kaavan (1) mukaisesti. Paikkojen välinen ordinaatiokuva saadaan aikaan, kun kuvataan latentit muuttujat. Lajien välisten ordinaation ollessa kiinnostuksen kohteena kuvataan $\hat{\boldsymbol{\Gamma}}$.

4 Simulointi

Tässä luvussa tarkastellaan aliluvuissa 3.1 ja 3.2 esitettyjen menetelmien teoreettisia ominaisuuksia simulointikokeiden avulla. Poisson-sekamallin kohdalla variaatiomenetelmän antamia estimaatteja vertaillaan Laplace-approksimaation antamiin estimaatteihin. Latenttiin muuttujamalliin perustuvia ordinaatiokuvia vertaillaan klassisen NMDS-menetelmän antamiin ordinaatiokuviin.

4.1 Sekamalli

Tarkastellaan ensin, kuinka mitattavien lajien määrän kasvattaminen vaikuttaa estimaattien tarkuuteen. Sitä varten simuloidaan havaintoja Poisson-sekamallista

$$y_{ij}|u_i \sim \text{Pois}(\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + u_i)),$$

missä $u_i \sim N(0, \sigma^2)$, $i = 1, 2, \dots, m$ ja $j = 1, 2, \dots, n_i$. Oletetaan asetelmat, missä $m = 50, 100, 200$, $n_i = 5$ kaikilla i , malliparametrit $\sigma^2 = 1$ ja $\boldsymbol{\beta}' = (-1, 1)$ sekä selittävänä muuttujana $\mathbf{x}'_{ij} = (1, x_{1ij})$, missä $x_{1ij} = -1$, kun i on parillinen ja $x_{1ij} = 1$, kun i on pariton. Tällöin kiinteällä i on x_{1ij} sama kaikilla j , kuten usein lukumääräaineistoissa. Simulointikierrosten N määrä on jokaisella asetelmalla 500. Lasketaan simuloiduista aineistoista estimaatit $\hat{\sigma}^2$ ja $\hat{\boldsymbol{\beta}}$ malliparametreille Laplace-menetelmällä ja variaatiomenetelmällä kolmella eri alkuarvolla. Variaatiomenetelmissä käytettävät alkuarvot ovat kiinteät alkuarvot (GVA1) $\boldsymbol{\beta}^0 = (0, 0)'$, $(\sigma^2)^0 = 1$, $\boldsymbol{\mu}^0 = (0, \dots, 0)'$ ja $\boldsymbol{\lambda}^0 = (1, \dots, 1)'$, Laplace-estimaatit alkuarvoina (GVA2) $\boldsymbol{\beta}^0 = \hat{\boldsymbol{\beta}}_L$, $(\sigma^2)^0 = \hat{\sigma}_L^2$, $\boldsymbol{\mu}^0 = \hat{\boldsymbol{u}}_L$ ja $\boldsymbol{\lambda}^0 = (1, \dots, 1)'$, sekä satunnaiset alkuarvot (GVA3) $\beta_k^0 \sim N(0, 1)$ kun $k = 0, 1$, $(\sigma^2)^0 \sim N(1, 1)$ siten, että $(\sigma^2)^0 > 0$, $\mu_i^0 \sim N(0, 1)$ ja $\lambda_i^0 \sim N(0, 1)$, $\lambda_i^0 > 0$ kun $i = 1, 2, \dots, m$.

Laplace-estimaatit lasketaan R-paketin `lme4` (Bates et al., 2015) funktion `glmer` avulla. Variaatiomenetelmän implementointi löytyy liitteestä E. Implementoinnissa log-uskottavuuden alaraja maksimoidaan ensin malliparametrien suhteen pitäen samalla variaatioparametrit vakioina. Alarajan maksimoivat arvot kiinnitetään malliparametreiksi ja maksimoidaan alarajaa variaatioparametrien suhteen. Saadut variaatioparametrien arvot kiinnitetään ja toistetaan maksimointia malli- ja variaatioparametrien suhteen kuten edellä, kunnes optimoitavan funktion arvo ei enää oleellisesti muutu.

Simuloiduista aineistoista lasketuille estimaateille lasketaan keskiarvot ja keskineliöpoikkeaman neliöjuuri (*root mean squared error*, RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{l=1}^N (\hat{\theta}_l - \theta_0)^2},$$

missä $\hat{\theta}_l$ on estimaatti parametrille θ_l laskettuna Laplace- ja variaatiomenetelmillä, θ_0 todellinen parametrin arvo ja N on simulointien määrä. Simulointitulokset ovat taulukossa 1. Taulukosta nähdään, että kun m , lukumääräaineistoissa esimerkiksi lajien

määrä kasvaa, niin keskineliöpoikkeaman neliöjuuri pienenee. Gaussisen variaatiomenetelmän ero perinteisesti käytettävään Laplace-menetelmään ei ole suuri. Variaatiomenetelmässä eri alkuarvoilla ei ole suurta vaikutusta. Parametrien β_0 ja σ^2 kohdalla kaikkein pienin keskineliöpoikkeama saadaan variaatiomenetelmällä, jossa alkuarvot ovat vakioita.

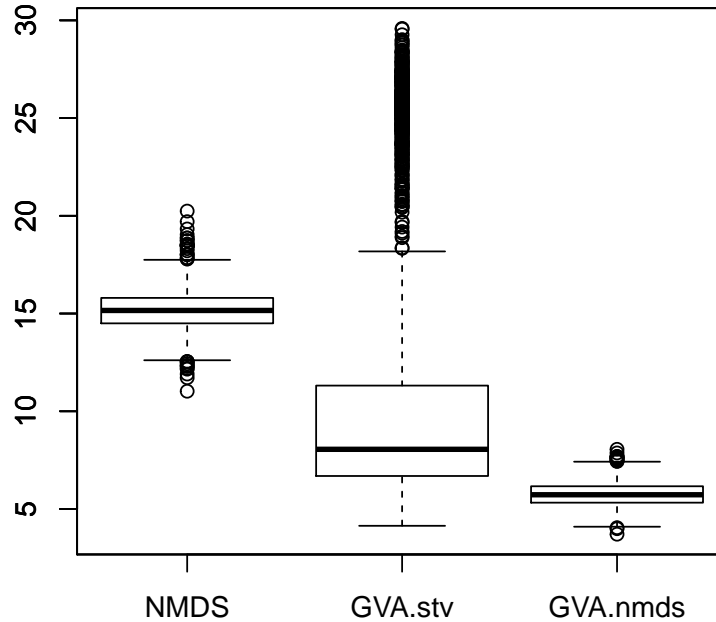
Taulukko 1: Keskiarvot ja keskineliöpoikkeaman neliöjuuri jokaiselle malliparametrille sekamallia simuloitaessa.

		β_0		β_1		σ^2	
	Menetelmä	keskiarvo	RMSE	keskiarvo	RMSE	keskiarvo	RMSE
m=50	Laplace	-1.008	0.206	1.000	0.189	0.932	0.316
	GVA1	-1.006	0.195	0.995	0.195	0.912	0.286
	GVA2	-1.058	0.208	0.963	0.186	0.953	0.305
	GVA3	-1.048	0.213	0.971	0.188	0.947	0.316
m=100	Laplace	-1.000	0.148	1.007	0.137	0.946	0.225
	GVA1	-1.000	0.140	1.001	0.139	0.927	0.202
	GVA2	-1.053	0.155	0.967	0.136	0.968	0.219
	GVA3	-1.039	0.154	0.975	0.136	0.957	0.225
m=200	Laplace	-1.007	0.104	1.003	0.090	0.968	0.165
	GVA1	-1.005	0.098	0.998	0.092	0.948	0.146
	GVA2	-1.062	0.119	0.961	0.094	0.992	0.162
	GVA3	-1.043	0.114	0.972	0.092	0.975	0.165

4.2 Latentti muuttujamalli

Tarkastellaan seuraavaksi latenttiin muuttujamalliin perustuvan ordinaation hyvyttä. Vertaillaan menetelmää klassiseen NMDS-menetelmään (Kruskal, 1964a, 1964b). NMDS-menetelmä, eli moniulotteinen skaalaus ilman metriikkaa (*non-metric multidimensional scaling*), on algoritmipohjainen tekniikka, joka iteratiivisesti asemoi ordinaatiopisteet kunnes ordinaation suhteellisilla etäisyyksillä on vahvin mahdollinen monotoninen sovitus paikkojen välisille parittaisille etäisyyksille, jotka on laskettu käyttäen jotakin erilaisuusmittaa (mahdollisesti muunnetusta) aineistosta.

Simuloidaan aineistoon perustuvasta asetelmasta. Simulointijakaumana käytetään hämähäkkiaineistosta aliluvussa 5.1 laskettuja estimaattien approksimaatioita. Simulointikierrosten määrä on 2500. Aliluvussa 3.2 määritellyille variaatioparametreille oletetaan rajoite $\Lambda_i = \lambda_i \mathbf{I}$. Simuloiduista aineistoista approksimoidaan latentit muuttajat gaussisella variaatiomenetelmällä kahdenlaisilla alkuarvoilla. Yhteisiä alkuarvoja molemmille approksimoineille ovat $\boldsymbol{\alpha}^0 = (0, \dots, 0)'$, $\boldsymbol{\beta}^0 = \boldsymbol{\beta}_{stv}$, $\boldsymbol{\Gamma}^0 = \boldsymbol{\Gamma}_{stv}$ ja $\boldsymbol{\lambda}^0 = (1, \dots, 1)'$. Alkuarvot eroavat variaatioparametrin $\boldsymbol{\mu}$ kohdalla. Approksimaatio (GVA.nmDS) käyttää NMDS-approksimaatiota alkuarvoina $\boldsymbol{\mu}^0 = \mathbf{u}_{NMDS}$. Ap-



Kuva 1: *Procrustes*-virhe estimoitujen ordinaatioiden ja todellisen ordinaation välillä.

proksimaatio (GVA.stv) taas käyttää `start.values` -funktion antamia alkuarvoja $\boldsymbol{\mu}^0 = \mathbf{u}_{stv}$.

Liitteessä E esitetty `start.values` -funktio arpoo ensin indeksejä kaksiulotteisesta normaalijakaumasta ja käyttää niitä selittäjinä sovittaessaan yleistetyn lineaarisen mallin jokaiselle lajille. Sovitus tehdään 10 kertaa ja niistä valitaan AIC-kriteerin perusteella paras. NMDS-approksimaatiot lasketaan paketin `vegan` (Oksanen et al., 2016) funktiolla `metaMDS`. Implementointi latentin muuttujamallin sovittamiselle variaatiomenetelmällä löytyy liitteestä E.

Jokaisen lähestymistavan kohdalla estimoitua ordinaatiota verrataan todelliseen niin sanotun *Procrustes*-virheen avulla. Siinä sovelletaan *Procrustes*-muunnosta minimoimaan neliöity ero sovitetun mallin ordinaation ja todellisen ordinaation välillä. *Procrustes*-virhe on muotoa

$$\text{Procrustes Error} = \sum_{i=1}^n \sum_{r=1}^m (u_{ir,fitted} - u_{ir,true})^2,$$

missä $u_{ir,fitted}$ on *Procrustes*-rotatoitu ordinaatiokoordinaatti paikalle i ja latentille muuttujalle r sovitetusta mallista ja $u_{ir,true}$ on vastaava todellinen ordinaatiokoordinaatti. Kuvasta 1 nähdään, että variaatiomenetelmällä latentille muuttuja mallille saadaan keskimäärin tarkempia tuloksia kuin NMDS-menetelmällä. Ero variaatiome-

netelmän eri alkuarvoilla tuotetun tulosten välillä kertonee variaatiomenetelmän sensitiivisyydestä huonoille alkuarvoille. *Procrustes*-virhe lasketaan paketin `vegan` funktiolla `procustes`.

5 Esimerkki

Tässä luvussa sovelletaan aikaisemmin esitettyjä malleja kahteen ekologian aineistoon.

5.1 Perhosaineisto

Aineistossa on mitattu perhosten lukumääriä Boulder County Open Space -alueelta Coloradosta Yhdysvalloista (Oliver et al., 2006). Tutkittavia perhoslajeja on 58 viidestä eri suvusta: *Hesperiidae* (17 lajia), *Papilionidae* (4 lajia), *Pieridae* (7 lajia), *Lycaenidae* (11 lajia), ja *Nymphalidae* (19 lajia). Mittauspaikkoja on 66, joissa kussakin perhosten lukumäärä on mitattu viitenä eri kertana. Mittaukset on suoritettu heinä- ja elokuussa vuonna 1999, sekä kesä-, heinä- ja elokuussa vuonna 2000. Tarkasteltavaksi valitaan laji *Colias philodice*. Selittäjinä muuttujina ovat muuttujat `habitat`, `building` ja `urbanveg`. Muuttuja `habitat` kuvaa ympäristöä ja kertoo ruohomaan tyyppin ja laadun. Se saa arvoja 1=`hayfield`, 2=`mixed`, 3=`short` ja 4=`tall`. Arvot samassa järjestyksessä merkitsevät kylvettyä heinäpeltoa, syntyperäistä ruohotyypin sekoitusta, syntyperäistä lyhyttä ja pitkää ruohoa. Käsitellään muuttuja `habitat` faktorina ja sen arvoja omina tasoinaan. Tasoon `short` verrataan muita tasoja. Muuttuja `building` kertoo rakennusten maa-alan osuuden ympäristössä prosentteina ja `urbanveg` ympäröivän urbanisaation osuuden prosentteina.

Sovitetaan aineistoon Poisson-sekamalli

$$\eta_{ij} = \exp(\beta_0 + \beta_1 \cdot \text{mixed}_i + \beta_2 \cdot \text{tall}_i + \beta_3 \cdot \text{hayfield}_i + \beta_4 \cdot \text{building}_i + \beta_5 \cdot \text{urbanveg}_i + u_i),$$

missä η_{ij} on perhosten lukumäärän odotusarvo, $i = 1, \dots, 66$ ja $j = 1, \dots, 5$. Alkuarvoina ovat Laplace-estimaatit $\boldsymbol{\beta}^0 = \hat{\boldsymbol{\beta}}_L$, $(\sigma^2)^0 = \hat{\sigma}_L^2$, $\boldsymbol{\mu}^0 = \hat{\boldsymbol{u}}_L$ ja vakiovektori $\boldsymbol{\lambda}^0 = (2, \dots, 2)'$. Estimaattien variaatioaprosimaatioiden avulla lasketaan muuttujille riskisuhde $\exp(\hat{\boldsymbol{\beta}})$.

Taulukko 2: Variaatiomenetelmän avulla estimoidut riskisuhteet $\exp(\hat{\boldsymbol{\beta}})$ perhoslukumäärän odotusarvolle ja näiden approksimatiiviset 95%:n luottamusvälit.

	riskisuhde	95%:n luottamusväli
<code>mixed</code>	2.982	(1.424, 6.243)
<code>tall</code>	3.380	(1.745, 6.547)
<code>hayfield</code>	10.901	(4.948, 24.015)
<code>building</code>	0.953	(0.886, 1.024)
<code>urbanveg</code>	1.006	(0.978, 1.035)

Estimaatit $\exp(\hat{\boldsymbol{\beta}})$ ja niiden approksimatiiviset 95%:n luottamusvälit ovat taulukossa 2. Kun kaavan 2 mukaista alarajaa käsitellään kuten logaritmista uskottavuutta, voidaan luottamusvälinä käyttää väliä $\exp\left(\hat{\beta}_l \pm 1.96 \widehat{\text{s.e.}}(\hat{\beta}_l)\right)$, missä $\widehat{\text{s.e.}}(\hat{\beta}_l)$,

$l = 1, \dots, 5$, lasketaan kuten aliluvussa 2.2. Lisäksi variaatioapproksimaatio satunnaisvaikutusten varianssille on $\hat{\sigma}^2 = 0.685$ ja sen 95%:n luottamusväli on (0.358, 1.012). Vertailun vuoksi Laplace-menetelmällä saatu estimaatti satunnaisvaikutusten varianssille on $\hat{\sigma}_L^2 = 0.606$ ja sen 95%:n luottamusväli on (0.395, 0.957).

Selittävään muuttujaan **habitat** liittyvät kertoimet ovat kaikki positiivisia ja kertoimien suuruusjärjestys pienimmästä suurimpaan on **mixed**, **tall** ja **hayfield**. Syntyperäisellä lyhyellä ruoholla on siis odotetusti vähiten perhosia eri ympäristöistä. Muuttujan **mixed** estimaatti kertoo että syntyperäisellä ruohotyypin sekoituksella perhosten määrän ennustettu suhde verrattuna syntyperäiseen lyhyeen ruohon on 2.982, kun muut muuttujat ja satunnaisvaikutus pysyvät vakioina. Toisin sanoen perhosten odotettu lukumäärä syntyperäisellä ruohotyypin sekoituksella on 2.982-kertainen odotettuun perhosten määrään syntyperäisellä lyhyellä ruoholla. Samoin muuttujat **tall** ja **hayfield** vertaavat odotetua perhosten lukumäärää syntyperäisen lyhyen ruohon odotettuun perhosten määrään. Perhosten lukumäärän ennustettu suhde syntyperäisellä pitkällä ruoholla verrattuna syntyperäiseen lyhyeen ruohon on 3.380 ja perhosten lukumäärän ennustettu suhde kylvetyllä heinäpellolla verrattuna syntyperäiseen lyhyeen ruohon on 10.901. Näissäkin tulkinnoissa oletetaan, että muiden selittävien muuttujien arvot ja satunnaisvaikutus pysyvät vakioina.

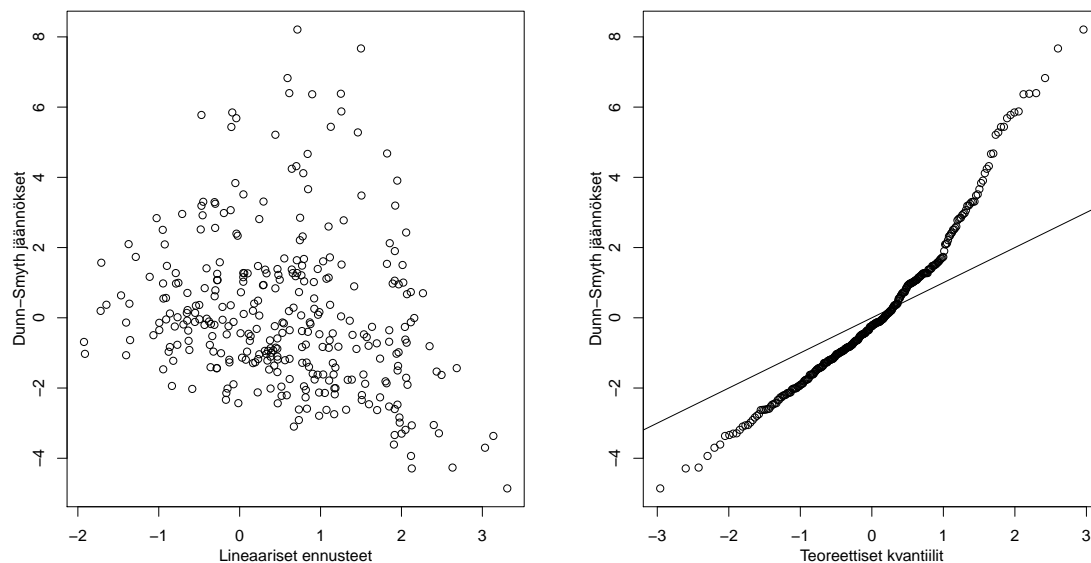
Ympäristössä olevien rakennusten maa-alan osuuden kasvu yhdellä prosenttiyksiköllä pienentää perhosten määrän ennusteen 0.953-kertaiseksi, kun muiden selittäjien arvot eivät muutu ja satunnaisvaikutus pysyy vakioina. Ympäröivän urbanisaation osuuden kasvu yhdellä prosenttiyksiköllä kasvattaa perhosten määrän ennustetta 1.006-kertaiseksi muiden olosuhteiden ja satunnaisvaikutuksen pysyessä vakioina. Nämä kaksi riskisuhdetta eivät ole tilastollisesti merkitseviä.

Taulukko 3: Laplace-menetelmän avulla estimoidut riskisuhteet $\exp(\hat{\beta})$ perhoslukumäärän odotusarvolle ja näiden 95%:n luottamusvälit.

	riskisuhde	95%:n luottamusväli
mixed	3.101	(1.542, 6.236)
tall	3.573	(1.913, 6.677)
hayfield	17.209	(8.267, 35.823)
building	0.949	(0.886, 1.016)
urbanveg	1.007	(0.981, 1.035)

Vertailun vuoksi lasketaan estimaatit riskisuhteille $\exp(\hat{\beta})$ ja niiden 95%:n luottamusvälit Laplace-approksimaatioiden avulla (Taulukko 3). Nähdään, että Laplace-approksimaatio muuttujien **mixed**, **tall** ja **urbanveg** estimaateille on hieman suurempi kuin vastaava variaatioapproksimaatio ja muuttujalle **hayfield** huomattavasti suurempi kuin vastaava variaatioapproksimaatio. Muuttujan **building** estimaatti laskettuna Laplace-menetelmällä taas on hieman pienempi kuin laskettuna variaatiomenetelmällä.

Jäännöstarkasteluiden avulla voidaan tutkia mallin sopivuutta aineistoon. Lukumääräaineistolle jäännösanalyysi on aineiston diskreettisyysden vuoksi haastavaa,



Kuva 2: Dunn-Smyth jäännökset normaali jakauman kvantiilin ja lineaaristen ennusteiden suhteen perhosaineistossa.

joten käytetään Dunn-Smyth -jäännöksiä (Dunn ja Smyth, 1996). Dunn-Smyth -jäännöksissä käytetään satunnaistamista tuottamaan jatkuvasti jakautuneita jäännöksiä, kun vaste on diskreetti, esimerkiksi Poisson-jakautunut. Dunn-Smyth jäännös havainnolle y_{ij} on

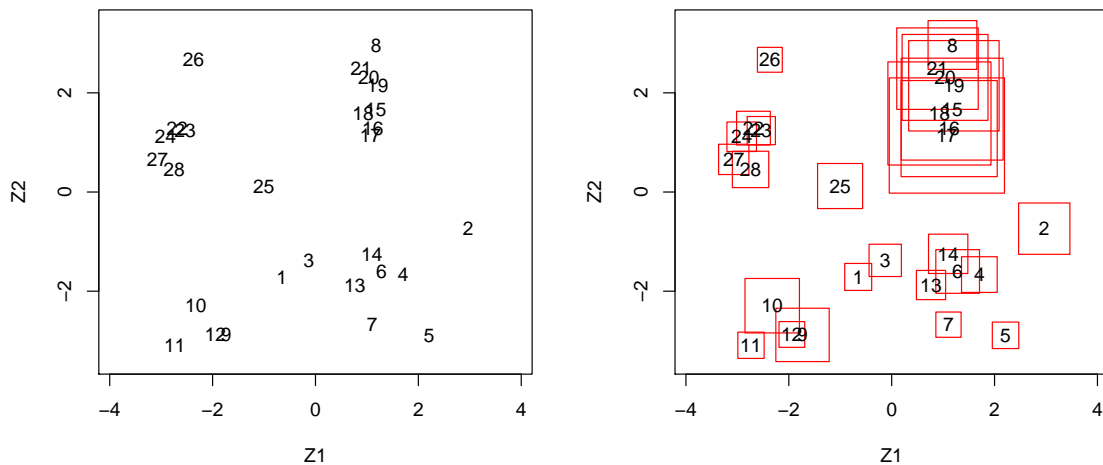
$$r_{ij} = \Phi^{-1}(v_{ij}F_{ij}(y_{ij}) + (1 - v_{ij})F_{ij}^-(y_{ij})),$$

missä $\Phi(\cdot)$ on standardinormaalijakauman kertymäfunktio ja $F_{ij}(\cdot)$ on havainnon y_{ij} kertymäfunktio, $F^-(y)$ on kertymäfunktion $F_{ij}(y)$ raja-arvo lähettyttäessä negatiiviselta puolelta ja v_{ij} on arvottu satunnaisesti tasajakaumasta. Jäännökset ovat normaali jakautuneita.

Kuvaan 2 on piirretty Dunn-Smyth -jäännösten kvantiilikuva sekä jäännösten ja lineaaristen ennusteiden yhteisjakauma. Jäännöskuvista (Kuva 2) nähdään, ettei malli sovi hyvin perhosaineistoon. Dunn-Smyth jäännökset kuvattuna teoreettisten kvantiilien suhteen näyttää, etteivät jäännökset ole normaali jakautuneet. Dunn-Smyth jäännökset kuvattuna lineaaristen ennusteiden suhteen paljastaa jäännösten vaihtelun kasvavan ennusteiden kasvaessa. Poisson-jakauma ei siis ole paras vaihtoehto kuvaamaan perhosten lukumäärää. Aineistoon voisi mahdollisesti sopia paremmin negatiivinen binomijakauma, joka sallisi myös ylihajontaa.

5.2 Hämähäkkiaineisto

Aineistossa on laskettu 12 eri hämähäkkilajin yksilöiden määrät näytteissä 28 mitauspaikalta (Van der Aart ja Smeenk-Enserink, 1974). Aineisto on tarjolla R-paketissa



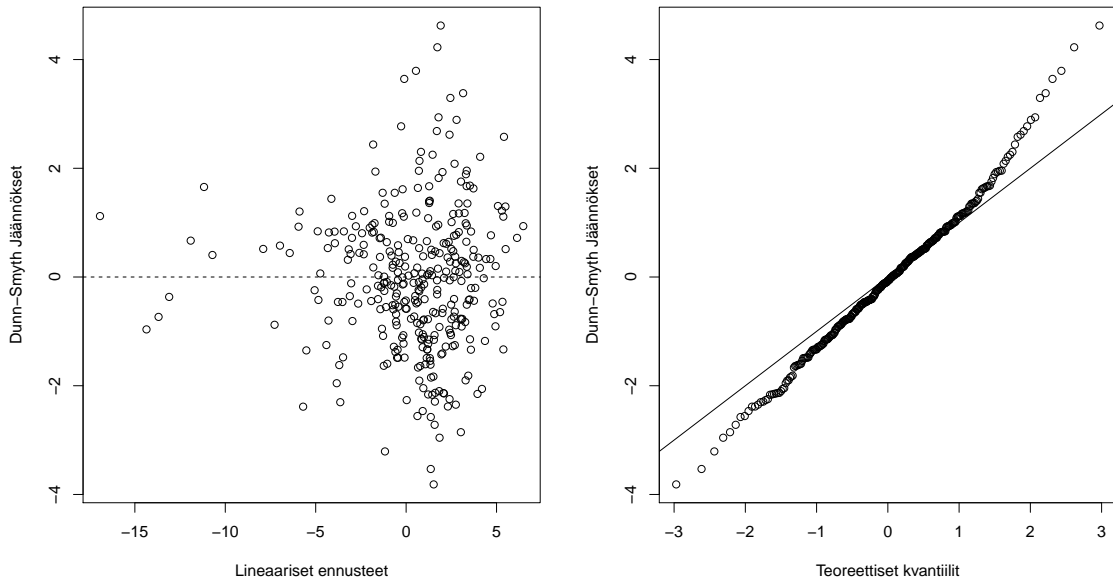
Kuva 3: Ordinaatiokuva hämähäkkiaineistosta ja variaatioparametrin $\hat{\lambda}_i$ avulla laskettua ordinaatiopisteiden vaihtelua.

`mvabund` (Wang et al., 2016) nimellä `spider`. Sovitetaan aineistoon latentti muuttujamalli olettaen, että yksilömäärä on Poisson-jakautunut. Tarkoituksena on ordinaatiokuvien avulla tarkastella, mitkä paikat ovat hämähäkkilajistoltaan lähellä toisiaan.

Aliluvussa 3.2 määritellyille variaatioparametreille oletetaan rajoite $\Lambda_i = \lambda_i \mathbf{I}$. Alkuarvoina malliparametreille ovat $\alpha^0 = (0, \dots, 0)'$, sekä `start.values` -funktion antamat alkuarvot $\beta^0 = \beta_{stv}$ ja $\Gamma^0 = \Gamma_{stv}$. Variaatioparametreille alkuarvoina ovat NMDS-menetelmän antamat arvot $\mu^0 = \mathbf{z}_{NMDS}$ ja $\lambda^0 = (1, \dots, 1)'$.

Ordinaatiokuva muodostetaan ennusteiden $\hat{\mathbf{u}} := \underline{\text{BP}}(\mathbf{u}) = \hat{\boldsymbol{\mu}}$ avulla. Ordinaatiokuvasta (Kuva 3) on havaittavissa kolme tai neljä eri ryhmää, joissa mittauspaikat ovat keskenään lajistoltaan samanlaisia. Havaintopaikat $\{1-7, 13-14\}$, paikat $\{9-12\}$, paikat $\{8, 15-21\}$ ja paikat $\{22-24, 26-28\}$. Havaintopaikka 25 jää näiden kaikkien keskelle. Satunnaisvaikutusten ennusteiden vaihtelua on arvioitu kuvassa 3 oikealla puolella. Ordinaatiopisteiden ympärille on piirretty vaihtelua kuvaavat laatikot, joiden sivujen pituudet ovat $2 * 1.96 \sqrt{\hat{\lambda}_i/n}$. Kuvasta nähdään, että paikkojen $\{15-21\}$ sijainnilla on muita suurempi epävarmuus, mutta ne kuitenkin selvästi muodostavat oman ryhmänsä havaintopaikan 8 kanssa.

Kuten sekamallien tapauksessa, mallikritiikkiin käytetään Dunn-Smyth jäännöksiä. Jäännöskuvista ilmenee (Kuva 4), ettei malli sovi hyvin aineistoon. Kun kuvataan Dunn-Smyth jäännökset lineaaristen ennusteiden suhteen (Kuva 4), nähdään, että ennusteet kasvavat lineaaristen jäännösten kasvaessa. Tämä kieli ylihajonnasta. Dunn-Smyth jäännökset kuvattuna teoreettisten kvantiilien suhteen näyttää (Kuva 4), etteivät jäännökset ole täysin normaalijakautuneet. Aiempien tutkimusten (Hui et al., 2015) nojalla on syytä olettaa, että aineistoon sopisi paremmin latentti muuttujamalli, joka sallii ylihajonnan.



Kuva 4: Dunn-Smyth jäännökset lineaaristen ennusteiden ja normaalijakauman kvanttiilin suhteen hämähäkkiaineistossa.

6 Pohdinta

Tässä työssä sovellettiin gaussista variaatiomenetelmää Poisson-sekamallin ja -latentin muuttujamallin parametrien estimoimiseksi. Lisäksi osoitettiin, että Poisson-jakauman tilanteessa uskottavuusfunktion alaraja saadaan kirjoitettua suljetussa muodossa ja parametrien estimointi on yksinkertaista esimerkiksi Kvasi-Newton -menetelmällä. Menetelmien teoreettisia ominaisuuksia tutkittiin simulointikokeiden avulla ja menetelmää sovellettiin kahteen aineistoon.

Simulointikokeiden perusteella lajien määrän kasvattaminen Poisson-sekamallin tapauksessa antaa tarkemman variaatioapproksimaation. Latentissa muuttujamallissa ordinaatiomenetelmälle gaussinen variaatiomenetelmä antaa hyvillä alkuarvoilla tarkemman tuloksen kuin perinteisesti käytetty NMDS-menetelmä. Latentin muuttujamallin kohdalla variaatiomenetelmä näyttäisi reagoivan herkästi huonoihin alkuarvoihin.

Hui et al. (2015) estimoivat latenttia muuttujamallia Monte Carlo EM-algoritmilla. He toteavat menetelmän olevan hidas pienilläkin aineistoilla. Variaatiomenetelmä tarjoaa nopeamman tavan laskea estimaatit. Variaatiomenetelmässä myös ennuste latenteille muuttujille saadaan kätevästi optimoinnin sivutuotteena.

Ormerod ja Wand (2012) yleistävät variaatiomenetelmän teorian yleistetyille lineaarisille sekamalleille. Variaatiomenetelmän mukaisia tuloksia on myös laajennettu eksponentiaaliselle perheelle latentin muuttujamallin tapauksessa (Hui et al., 2017). Tulevaisuuden haasteisiin lukeutuu variaatiomenetelmän tulosten yleistäminen koskemaan pitkittäisaineistoja, eli tapauksia, joissa mittauksia on samoilta paikoilta useam-

pina vuosina. Tuloksia voisi myös laajentaa koskemaan malleja, joissa on spatiaalista riippuvuutta, eli tilanteita, joissa mittauspaikkojen ei oleteta olevan riippumattomia toisistaan.

Viitteet

- D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.
- C.G. Broyden. Quasi-newton methods and their application to function minimisation. *Mathematics of Computation*, 21(99):368–381, 1967.
- P.K. Dunn and G.K. Smyth. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5:236–244, 1996.
- P. Gilbert and R. Varadhan. *numDeriv: Accurate Numerical Derivatives*, 2016. URL <https://CRAN.R-project.org/package=numDeriv>. R package version 2016.8-1.
- P. Hall, J.T. Ormerod, and M.P. Wand. Theory of gaussian variational approximation for a poisson mixed model. *Statistica Sinica*, 21:369–389, 2011.
- P. Huber, E. Ronchetti, and M.P. Victoria-Feser. Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66:893–908, 2004.
- F.K.C. Hui, S. Taskinen, S. Pledger, S.D. Foster, and D.I. Warton. Model-based approaches to unconstrained ordination. *Methods in Ecology and Evolution*, 6(4): 399–411, 2015.
- F.K.C. Hui, D.I. Warton, J.T. Ormerod, V. Haapaniemi, and S. Taskinen. Variational approximations for generalized linear latent variable models. *Journal of Computational and Graphical Statistics*, 26(1):35–43, 2017.
- M.I. Jordan. Graphical models. *Statistical Science*, 19:140–155, 2004.
- J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964a.
- J. B. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2):115–129, 1964b.
- S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.
- C.E. McCullouch and S.R. Searle. *Generalized, Linear, and Mixed Models*. John Wiley & Sons, New York, 2001.

- C.E. McCullouch, S.R. Searle, and J.M. Neuhaus. *Generalized, Linear, and Mixed Models*. Wiley, New York, 2nd edition, 2008.
- J. Oksanen, F.G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P.R. Minchin, R.B. O’Hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, E. Szoecs, and H. Wagner. *vegan: Community Ecology Package*, 2016. URL <https://CRAN.R-project.org/package=vegan>. R package version 2.4-1.
- J.C. Oliver, K.L. Prudic, and S.K. Collinge. Boulder county open space butterfly diversity and abundance. *Ecology*, 87:1066, 2006.
- J.T. Ormerod and M.P. Wand. Explaining variational approximations. *American Statistician*, 64:140–153, 2010.
- J.T. Ormerod and M.P. Wand. Gaussian variational approximate inference for generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 21:2–17, 2012.
- D.M. Titterington. Bayesian methods for neural networks and related models. *Statistical Science*, 19:128–139, 2004.
- P.J.M. Van der Aart and N. Smeenk-Enserink. Correlations between distributions of hunting spiders (lycosidae, ctenidae) and environmental characteristics in a dune area. *Netherlands Journal of Zoology*, 25:1–45, 1974.
- Y. Wang, U. Naumann, S. Wright, D. Eddelbuettel, and D. Warton. *mvabund: Statistical Methods for Analysing Multivariate Abundance Data*, 2016. URL <https://CRAN.R-project.org/package=mvabund>. R package version 3.11.9.

A Kaavan 2 todistus

Yksinkertaisen Poisson-sekamallin tapauksessa logaritmisen uskottavuusfunktion alaraja voidaan kirjoittaa (Ormerod ja Wand, 2010)

$$\begin{aligned} & \underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}; q) \\ &= \mathbb{E}_{q(\mathbf{u})} \left[\sum_{i=1}^m \left[\sum_{j=1}^{n_i} \left\{ y_{ij} (\mathbf{x}'_{ij} \boldsymbol{\beta} + u_i) - \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + u_i) - \log(y_{ij}!) \right\} - \frac{u_i^2}{2\sigma^2} \right] \right. \\ & \quad \left. - \frac{m}{2} \log(2\pi\sigma^2) - \log(q(u_1, \dots, u_m)) \right]. \end{aligned}$$

Odotusarvon lineaarisuuden perusteella

$$\begin{aligned} & \underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}; q) \\ &= \sum_{i=1}^m \left[\sum_{j=1}^{n_i} \left\{ y_{ij} (\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbb{E}_{q(\mathbf{u})} [u_i]) - \mathbb{E}_{q(\mathbf{u})} [\exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + u_i)] - \log(y_{ij}!) \right\} \right. \\ & \quad \left. - \frac{1}{2\sigma^2} \mathbb{E}_{q(\mathbf{u})} [u_i^2] \right] - \frac{m}{2} \log(2\pi\sigma^2) - \mathbb{E}_{q(\mathbf{u})} [\log(q(u_1, \dots, u_m))]. \end{aligned}$$

Koska tiheydeksi valittiin q tulo riippumattomista normaalijakauman tiheyksistä, eli $q(u_1, \dots, u_m) = q_1(u_1) \times \dots \times q_m(u_m)$, missä $u_i \sim N(\mu_i, \lambda_i)$, niin $\mathbb{E}_{q(\mathbf{u})} [u_i] = \mu_i$ kaikilla $i = 1, 2, \dots, m$. Samoin muuttujan u_i momentit generoiva funktio on tällöin

$$M_{u_i}(t) = \mathbb{E}_{q(\mathbf{u})} [\exp(tu_i)] = \exp\left(\mu_i t + \frac{1}{2} \lambda_i t^2\right).$$

Siispä $M_{u_i}(1) = \mathbb{E}_{q(\mathbf{u})} [\exp(u_i)] = \exp(\mu_i + \frac{1}{2} \lambda_i)$. Nyt

$$\begin{aligned} \mathbb{E}_{q(\mathbf{u})} [\exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + u_i)] &= \mathbb{E}_{q(\mathbf{u})} [\exp(\mathbf{x}'_{ij} \boldsymbol{\beta}) \exp(u_i)] = \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}) \mathbb{E}_{q(\mathbf{u})} [\exp(u_i)] \\ &= \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}) \exp\left(\mu_i + \frac{1}{2} \lambda_i\right) = \exp\left(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mu_i + \frac{1}{2} \lambda_i\right). \end{aligned}$$

Koska muuttujan u_i varianssi voidaan kirjoittaa muodossa

$$\text{Var}(u_i) = \mathbb{E}_{q(\mathbf{u})} [u_i^2] - (\mathbb{E}_{q(\mathbf{u})} [u_i])^2,$$

niin voidaan kirjoittaa myös

$$\mathbb{E}_{q(\mathbf{u})} [u_i^2] = \text{Var}(u_i) + (\mathbb{E}_{q(\mathbf{u})} [u_i])^2 = \lambda_i + \mu_i^2.$$

Lasketaan vielä $\mathbb{E}_{q(\mathbf{u})} [\log(q(u_1, \dots, u_m))] = \mathbb{E}_{q(\mathbf{u})} [\log(q_1(u_1) \times \dots \times q_m(u_m))]$. Logaritmin laskusääntöjen nojalla

$$\mathbb{E}_{q(\mathbf{u})} [\log(q_1(u_1) \times \dots \times q_m(u_m))] = \mathbb{E}_{q(\mathbf{u})} \left[\sum_{i=1}^m \log q(u_i) \right] = \sum_{i=1}^m \mathbb{E}_{q(\mathbf{u})} [\log q(u_i)],$$

missä

$$\log q(u_i) = \log \left((2\pi\lambda_i)^{-\frac{1}{2}} \exp \left(-\frac{(u_i - \mu_i)^2}{2\lambda_i} \right) \right) = -\frac{1}{2} \log(2\pi\lambda_i) - \frac{(u_i - \mu_i)^2}{2\lambda_i}$$

kaikilla $i = 1, 2, \dots, m$. Nyt

$$\mathbb{E}_{q(\mathbf{u})} [\log q(u_i)] = -\frac{1}{2} \log(2\pi\lambda_i) - \frac{\mathbb{E}_{q(\mathbf{u})} [(u_i - \mu_i)^2]}{2\lambda_i},$$

missä $\mathbb{E}_{q(\mathbf{u})} [(u_i - \mu_i)^2] = \text{Var}(u_i) = \lambda_i$. Saadaan siis

$$\mathbb{E}_{q(\mathbf{u})} [\log q(u_i)] = -\frac{1}{2} \log(2\pi\lambda_i) - \frac{\lambda_i}{2\lambda_i} = -\frac{1}{2} (\log(2\pi) + \log \lambda_i + 1).$$

Kun kaikki edellä oleva yhdistetään, saadaan logaritmissen uskottavuusfunktion alarajaksi

$$\begin{aligned} & \underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}; q) \\ &= \sum_{i=1}^m \left[\sum_{j=1}^{n_i} \left\{ y_{ij} (\mathbf{x}'_{ij}\boldsymbol{\beta} + \mu_i) - \exp \left(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mu_i + \frac{1}{2}\lambda_i \right) - \log(y_{ij}!) \right\} \right. \\ & \quad \left. - \frac{1}{2\sigma^2} (\lambda_i + \mu_i^2) \right] - \frac{m}{2} \log(2\pi\sigma^2) - \sum_{i=1}^m -\frac{1}{2} (\log(2\pi) + \log \lambda_i + 1) \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} \left[y_{ij} (\mathbf{x}'_{ij}\boldsymbol{\beta} + \mu_i) - \exp \left(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mu_i + \frac{1}{2}\lambda_i \right) - \log(y_{ij}!) \right] \\ & \quad - \sum_{i=1}^m \left[\frac{1}{2\sigma^2} (\lambda_i + \mu_i^2) \right] - \frac{m}{2} (\log(2\pi) + \log \sigma^2) + \frac{m}{2} (\log(2\pi) + 1) + \sum_{i=1}^m \frac{1}{2} \log \lambda_i \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} \left[y_{ij} (\mathbf{x}'_{ij}\boldsymbol{\beta} + \mu_i) - \exp \left(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mu_i + \frac{1}{2}\lambda_i \right) - \log(y_{ij}!) \right] \\ & \quad + \frac{m}{2} (1 - \log \sigma^2) + \frac{1}{2} \sum_{i=1}^m \left[\log \lambda_i - \frac{\lambda_i + \mu_i^2}{\sigma^2} \right]. \end{aligned}$$

B Log-uskottavuuden alarajan derivaatat Poisson-sekamallille

Variaatiomenetelmän mukaisen logaritmissen uskottavuusfunktion alarajan derivaatat malliparametrien suhteen ovat

$$\frac{\partial}{\partial \boldsymbol{\beta}} \underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}; q) = \sum_{i=1}^m \sum_{j=1}^{n_i} \left(y_{ij} - \exp\left(\mu_i + \frac{1}{2}\lambda_i + \mathbf{x}'_{ij}\boldsymbol{\beta}\right) \right) \mathbf{x}_{ij}$$

ja

$$\frac{\partial}{\partial \sigma^2} \underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}; q) = -\frac{m}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^m (\mu_i^2 + \lambda_i).$$

Variaatioparametrien suhteen derivaatat ovat (Hall et al., 2011)

$$\frac{\partial}{\partial \mu_i} \underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}; q) = \sum_{j=1}^{n_i} y_{ij} - \exp\left(\mu_i + \frac{1}{2}\lambda_i\right) \cdot \sum_{j=1}^{n_i} \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}) - \frac{\mu_i}{\sigma^2}$$

ja

$$\frac{\partial}{\partial \lambda_i} \underline{l}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}; q) = -\frac{1}{2} \exp\left(\mu_i + \frac{1}{2}\lambda_i\right) \cdot \sum_{j=1}^{n_i} \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}) + \frac{1}{2\lambda_i} - \frac{1}{2\sigma^2}$$

kaikilla $i = 1, 2, \dots, m$. Yleisemmässä muodossa derivaatat tarjoavat Ormerod ja Wand (2012).

C Kaavan 3 todistus

Variaatiomenetelmän mukainen alaraja logaritmiselle uskottavuusfunktiolle on

$$\underline{l}(\boldsymbol{\theta}; q) = \int_{\mathbb{R}^{n \cdot m}} q(\mathbf{u}) \cdot \log \left\{ \frac{p(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta})}{q(\mathbf{u})} \right\} d\mathbf{u} = \mathbb{E}_{q(\mathbf{u})} \left[\log \left\{ \frac{p(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta})}{q(\mathbf{u})} \right\} \right],$$

sillä q on tiheys. Sijoitetaan $p(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{u}; \boldsymbol{\theta}) p(\mathbf{u}; \boldsymbol{\theta})$. Logaritmin laskusääntöjen nojalla saadaan

$$\begin{aligned} \underline{l}(\boldsymbol{\theta}; q) &= \mathbb{E}_{q(\mathbf{u})} \left[\log \left\{ \frac{p(\mathbf{y}|\mathbf{u}; \boldsymbol{\theta}) p(\mathbf{u}; \boldsymbol{\theta})}{q(\mathbf{u})} \right\} \right] \\ &= \mathbb{E}_{q(\mathbf{u})} [\log(p(\mathbf{y}|\mathbf{u}; \boldsymbol{\theta})) + \log(p(\mathbf{u}; \boldsymbol{\theta})) - \log(q(\mathbf{u}))] \\ &= \mathbb{E}_{q(\mathbf{u})} \left[\log \prod_{i=1}^n p_i(\mathbf{y}_i|\mathbf{u}_i; \boldsymbol{\theta}) + \log \prod_{i=1}^n (p(\mathbf{u}_i; \boldsymbol{\theta})) - \log(q(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)) \right]. \end{aligned}$$

Koska y_{ij} ehdolla \mathbf{u}_i noudattaa Poisson-jakaumaa, saadaan

$$\begin{aligned} \log \prod_{i=1}^n p_i(\mathbf{y}_i|\mathbf{u}_i; \boldsymbol{\theta}) &= \sum_{i=1}^n \log(p_i(\mathbf{y}_i|\mathbf{u}_i; \boldsymbol{\theta})) = \sum_{i=1}^n \log \prod_{j=1}^p p_{ij}(y_{ij}|\mathbf{u}_i; \boldsymbol{\theta}) \\ &= \sum_{i=1}^n \sum_{j=1}^p \log(p_{ij}(y_{ij}|\mathbf{u}_i; \boldsymbol{\theta})) = \sum_{i=1}^n \sum_{j=1}^p \log \left(\frac{(\mu_{ij})^{y_{ij}}}{y_{ij}!} \cdot \exp(-\mu_{ij}) \right) \\ &= \sum_{i=1}^n \sum_{j=1}^p [y_{ij}(\alpha_i + \beta_j + \mathbf{u}'_i \boldsymbol{\gamma}_j) - \log(y_{ij}!) - \exp(\alpha_i + \beta_j + \mathbf{u}'_i \boldsymbol{\gamma}_j)]. \end{aligned}$$

Aiemmin todettiin, että

$$p(\mathbf{u}_i; \boldsymbol{\theta}) = p(\mathbf{u}_i) = (2\pi)^{-\frac{m}{2}} \exp\left(-\frac{1}{2} \mathbf{u}'_i \mathbf{u}_i\right).$$

Nyt

$$\begin{aligned} \log \prod_{i=1}^n p(\mathbf{u}_i; \boldsymbol{\theta}) &= \log \prod_{i=1}^n (2\pi)^{-\frac{m}{2}} \exp\left(-\frac{1}{2} \mathbf{u}'_i \mathbf{u}_i\right) = \sum_{i=1}^n \log \left((2\pi)^{-\frac{m}{2}} \exp\left(-\frac{1}{2} \mathbf{u}'_i \mathbf{u}_i\right) \right) \\ &= \sum_{i=1}^n \left[-\frac{m}{2} \log(2\pi) - \frac{1}{2} \mathbf{u}'_i \mathbf{u}_i \right] = -\frac{n \cdot m}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \mathbf{u}'_i \mathbf{u}_i. \end{aligned}$$

Edelliset yhdistämällä saadaan alaraja muotoon

$$\begin{aligned} \underline{l}(\boldsymbol{\theta}; q) &= \mathbb{E}_{q(\mathbf{u})} \left[\sum_{i=1}^n \sum_{j=1}^p [y_{ij}(\alpha_i + \beta_j + \mathbf{u}'_i \boldsymbol{\gamma}_j) - \log(y_{ij}!) - \exp(\alpha_i + \beta_j + \mathbf{u}'_i \boldsymbol{\gamma}_j)] \right. \\ &\quad \left. - \frac{n \cdot m}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \mathbf{u}'_i \mathbf{u}_i - \log(q(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)) \right]. \end{aligned}$$

Odotusarvon lineaarisuuden perusteella

$$l(\boldsymbol{\theta}; q) = \sum_{i=1}^n \sum_{j=1}^p \left[y_{ij} (\alpha_i + \beta_j + \mathbb{E}_{q(\mathbf{u})} [\mathbf{u}'_i \boldsymbol{\gamma}_j]) - \log(y_{ij}!) - \mathbb{E}_{q(\mathbf{u})} [\exp(\alpha_i + \beta_j + \mathbf{u}'_i \boldsymbol{\gamma}_j)] \right] \\ - \frac{n \cdot m}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \mathbb{E}_{q(\mathbf{u})} [\mathbf{u}'_i \mathbf{u}_i] - \mathbb{E}_{q(\mathbf{u})} [\log(q(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n))].$$

Valitaan

$$q(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = q_1(\mathbf{u}_1) \times q_2(\mathbf{u}_2) \times \dots \times q_n(\mathbf{u}_n)$$

siten, että jokainen $\mathbf{u}_i \sim N_m(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)$, missä $\boldsymbol{\mu}'_i = (\mu_{i1}, \dots, \mu_{im})$ ja $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{im})$ kaikilla $i = 1, 2, \dots, n$. Tällöin

$$\mathbf{u}'_i \boldsymbol{\gamma}_j = \boldsymbol{\gamma}'_j \mathbf{u}_i \sim N(\boldsymbol{\gamma}'_j \boldsymbol{\mu}_i, \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j),$$

jolloin siis $\mathbb{E}_{q(\mathbf{u})} [\mathbf{u}'_i \boldsymbol{\gamma}_j] = \boldsymbol{\gamma}'_j \boldsymbol{\mu}_i$.

Koska \mathbf{u}_i noudattaa normaalijakaumaa, niin sen momentit generoiva funktio on

$$m_{\mathbf{u}_i}(\mathbf{t}) := \mathbb{E} [\exp(\mathbf{t}' \mathbf{u}_i)] = \exp\left(\boldsymbol{\mu}'_i \mathbf{t} + \frac{1}{2} \mathbf{t}' \boldsymbol{\Lambda}_i \mathbf{t}\right).$$

Nyt voidaan kirjoittaa

$$\begin{aligned} \mathbb{E}_{q(\mathbf{u})} [\exp(\alpha_i + \beta_j + \mathbf{u}'_i \boldsymbol{\gamma}_j)] &= \mathbb{E}_{q(\mathbf{u})} [\exp(\alpha_i + \beta_j) \cdot \exp(\mathbf{u}'_i \boldsymbol{\gamma}_j)] \\ &= \exp(\alpha_i + \beta_j) \cdot \mathbb{E}_{q(\mathbf{u})} [\exp(\mathbf{u}'_i \boldsymbol{\gamma}_j)] = \exp(\alpha_i + \beta_j) \cdot m_{\mathbf{u}_i}(\boldsymbol{\gamma}_j) \\ &= \exp(\alpha_i + \beta_j) \cdot \exp\left(\boldsymbol{\mu}'_i \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right) = \exp\left(\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right). \end{aligned}$$

Lasketaan seuraavaksi $\mathbb{E}_{q(\mathbf{u})} [\mathbf{u}'_i \mathbf{u}_i]$. Koska valittiin $q(\mathbf{u}_i)$ siten, että $\mathbf{u}_i \sim N_m(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)$, niin

$$\mathbf{z}_i := \boldsymbol{\Lambda}_i^{-\frac{1}{2}} (\mathbf{u}_i - \boldsymbol{\mu}_i) \sim N_m(\mathbf{0}, \mathbf{I}_m),$$

kaikilla $i = 1, 2, \dots, n$. Muuttujan \mathbf{z}_i avulla ilmaistuna $\mathbf{u}_i = \boldsymbol{\Lambda}_i^{\frac{1}{2}} \mathbf{z}_i + \boldsymbol{\mu}_i$. Nyt

$$\mathbb{E}_{q(\mathbf{u})} [\mathbf{u}'_i \mathbf{u}_i] = \mathbb{E}_{\mathbf{z}} \left[\left(\boldsymbol{\Lambda}_i^{\frac{1}{2}} \mathbf{z}_i + \boldsymbol{\mu}_i \right)' \left(\boldsymbol{\Lambda}_i^{\frac{1}{2}} \mathbf{z}_i + \boldsymbol{\mu}_i \right) \right].$$

Koska $\boldsymbol{\Lambda}_i$ on diagonaalimatriisi, niin $\boldsymbol{\Lambda}_i^{\frac{1}{2}}$ on diagonaalimatriisi ja siten symmetrinen, joten $\left(\boldsymbol{\Lambda}_i^{\frac{1}{2}} \right)' = \boldsymbol{\Lambda}_i^{\frac{1}{2}}$. Tällöin $\left(\boldsymbol{\Lambda}_i^{\frac{1}{2}} \mathbf{z}_i + \boldsymbol{\mu}_i \right)' = \mathbf{z}'_i \boldsymbol{\Lambda}_i^{\frac{1}{2}} + \boldsymbol{\mu}'_i$. Sijoitetaan edellinen odotusarvon lausekkeeseen. Saadaan

$$\begin{aligned} \mathbb{E}_{q(\mathbf{u})} [\mathbf{u}'_i \mathbf{u}_i] &= \mathbb{E}_{\mathbf{z}} \left[\left(\mathbf{z}'_i \boldsymbol{\Lambda}_i^{\frac{1}{2}} + \boldsymbol{\mu}'_i \right) \left(\boldsymbol{\Lambda}_i^{\frac{1}{2}} \mathbf{z}_i + \boldsymbol{\mu}_i \right) \right] \\ &= \mathbb{E}_{\mathbf{z}} \left[\mathbf{z}'_i \boldsymbol{\Lambda}_i^{\frac{1}{2}} \boldsymbol{\Lambda}_i^{\frac{1}{2}} \mathbf{z}_i + \mathbf{z}'_i \boldsymbol{\Lambda}_i^{\frac{1}{2}} \boldsymbol{\mu}_i + \boldsymbol{\mu}'_i \boldsymbol{\Lambda}_i^{\frac{1}{2}} \mathbf{z}_i + \boldsymbol{\mu}'_i \boldsymbol{\mu}_i \right] \\ &= \mathbb{E}_{\mathbf{z}} \left[\mathbf{z}'_i \boldsymbol{\Lambda}_i^{\frac{1}{2}} \boldsymbol{\Lambda}_i^{\frac{1}{2}} \mathbf{z}_i \right] + \mathbb{E}_{\mathbf{z}} [\mathbf{z}'_i] \boldsymbol{\Lambda}_i^{\frac{1}{2}} \boldsymbol{\mu}_i + \boldsymbol{\mu}'_i \boldsymbol{\Lambda}_i^{\frac{1}{2}} \mathbb{E}_{\mathbf{z}} [\mathbf{z}_i] + \boldsymbol{\mu}'_i \boldsymbol{\mu}_i. \end{aligned}$$

Koska $\mathbb{E}_{\mathbf{z}} [\mathbf{z}_i] = \mathbf{0}$, niin

$$\mathbb{E}_{q(\mathbf{u})} [\mathbf{u}'_i \mathbf{u}_i] = \mathbb{E}_{\mathbf{z}} [\mathbf{z}'_i \mathbf{\Lambda}_i \mathbf{z}_i] + \boldsymbol{\mu}'_i \boldsymbol{\mu}_i.$$

Osoitetaan vielä, että $\mathbb{E}_{\mathbf{z}} [\mathbf{z}'_i \mathbf{\Lambda}_i \mathbf{z}_i] = \text{trace}(\mathbf{\Lambda}_i)$. Koska $\mathbf{z}_i \sim N_m(\mathbf{0}, \mathbf{I}_m)$ ja matriisi $\mathbf{\Lambda}_i$ on symmetrinen, niin $\mathbf{z}'_i \mathbf{\Lambda}_i \mathbf{z}_i$ on jakautunut kuten

$$a_1 X_1^2 + \dots + a_m X_m^2,$$

missä a_1, \dots, a_m ovat matriisin $\mathbf{\Lambda}_i$ ominaisarvot, satunnaismuuttujat X_1, \dots, X_m ovat riippumattomia ja kukin niistä noudattaa $N(0, 1)$ -jakaumaa. Koska matriisi $\mathbf{\Lambda}_i$ on diagonaalimatriisi, sen ominaisarvot ovat sen diagonaalialkioiden arvot $\lambda_{i1}, \dots, \lambda_{im}$. Nyt saadaan

$$\mathbb{E}_{\mathbf{z}} [\mathbf{z}'_i \mathbf{\Lambda}_i \mathbf{z}_i] = \mathbb{E} [\lambda_{i1} X_1^2 + \dots + \lambda_{im} X_m^2] = \lambda_{i1} \mathbb{E} [X_1^2] + \dots + \lambda_{im} \mathbb{E} [X_m^2].$$

Koska $X_k \sim N(0, 1)$, niin $X_k^2 \sim \chi_1^2$ kaikilla $k = 1, 2, \dots, m$. Tällöin $\mathbb{E} [X_k^2] = 1$ kaikilla $k = 1, 2, \dots, m$. Saadaan

$$\mathbb{E}_{\mathbf{z}} [\mathbf{z}'_i \mathbf{\Lambda}_i \mathbf{z}_i] = \lambda_{i1} + \dots + \lambda_{im} = \text{trace}(\mathbf{\Lambda}_i).$$

Siispä

$$\mathbb{E}_{q(\mathbf{u})} [\mathbf{u}'_i \mathbf{u}_i] = \text{trace}(\mathbf{\Lambda}_i) + \boldsymbol{\mu}'_i \boldsymbol{\mu}_i.$$

Lasketaan vielä

$$\begin{aligned} -\mathbb{E}_{q(\mathbf{u})} [\log(q(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n))] &= -\mathbb{E}_{q(\mathbf{u})} \left[\log \left(\prod_{i=1}^n q_i(\mathbf{u}_i) \right) \right] \\ &= -\mathbb{E}_{q(\mathbf{u})} \left[\sum_{i=1}^n \log(q_i(\mathbf{u}_i)) \right] = -\sum_{i=1}^n \mathbb{E}_{q(\mathbf{u})} [\log(q_i(\mathbf{u}_i))]. \end{aligned}$$

Tässä

$$\begin{aligned} &\mathbb{E}_{q(\mathbf{u})} [\log(q_i(\mathbf{u}_i))] \\ &= \mathbb{E}_{q(\mathbf{u})} \left[\log \left((2\pi)^{-\frac{m}{2}} \det(\mathbf{\Lambda}_i)^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{u}_i - \boldsymbol{\mu}_i)' \mathbf{\Lambda}_i^{-1} (\mathbf{u}_i - \boldsymbol{\mu}_i) \right) \right) \right]. \end{aligned}$$

Odotusarvon lineaarisuuden ja logaritmin laskusääntöjen nojalla

$$\begin{aligned} \mathbb{E}_{q(\mathbf{u})} [\log(q_i(\mathbf{u}_i))] &= -\frac{m}{2} \log(2\pi) - \frac{1}{2} \log(\det(\mathbf{\Lambda}_i)) \\ &\quad - \frac{1}{2} \mathbb{E}_{q(\mathbf{u})} [(\mathbf{u}_i - \boldsymbol{\mu}_i)' \mathbf{\Lambda}_i^{-1} (\mathbf{u}_i - \boldsymbol{\mu}_i)]. \end{aligned}$$

Jos oletuksen $\mathbf{u}_i \sim N_m(\boldsymbol{\mu}_i, \mathbf{\Lambda}_i)$ lisäksi $\mathbf{\Lambda}_i$ on positiivisesti definiitti, niin

$$(\mathbf{u}_i - \boldsymbol{\mu}_i)' (\mathbf{\Lambda}_i^{-1}) (\mathbf{u}_i - \boldsymbol{\mu}_i) \sim \chi_m^2.$$

Tällöin $\mathbb{E}_{q(\mathbf{u})}[(\mathbf{u}_i - \boldsymbol{\mu}_i)' \boldsymbol{\Lambda}_i^{-1} (\mathbf{u}_i - \boldsymbol{\mu}_i)] = m$. Nyt

$$\begin{aligned} -\mathbb{E}_{q(\mathbf{u})}[\log(q(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n))] &= -\sum_{i=1}^n \left[-\frac{m}{2} \log(2\pi) - \frac{1}{2} \log(\det(\boldsymbol{\Lambda}_i)) - \frac{m}{2} \right] \\ &= \frac{n \cdot m}{2} (\log(2\pi) + 1) + \frac{1}{2} \sum_{i=1}^n \log(\det(\boldsymbol{\Lambda}_i)). \end{aligned}$$

Kun kaikki edellämäinnittu yhdistetään, saadaan logaritmissen uskottavuuden alarajaksi

$$\begin{aligned} \underline{l}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}; q) &= \sum_{i=1}^n \sum_{j=1}^p \left[y_{ij} (\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j) - \log(y_{ij}!) - \exp\left(\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right) \right] \\ &\quad - \frac{n \cdot m}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n [\text{trace}(\boldsymbol{\Lambda}_i) + \boldsymbol{\mu}'_i \boldsymbol{\mu}_i] + \frac{n \cdot m}{2} (\log(2\pi) + 1) \\ &\quad + \frac{1}{2} \sum_{i=1}^n \log(\det(\boldsymbol{\Lambda}_i)) \\ &= \sum_{i=1}^n \sum_{j=1}^p \left[y_{ij} (\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j) - \log(y_{ij}!) - \exp\left(\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right) \right] \\ &\quad + \frac{m \cdot n}{2} + \frac{1}{2} \sum_{i=1}^n [\log(\det(\boldsymbol{\Lambda}_i)) - \text{trace}(\boldsymbol{\Lambda}_i) - \boldsymbol{\mu}'_i \boldsymbol{\mu}_i]. \end{aligned}$$

D Alarajan derivaatat latentissa muuttujamallissa

Variaatiomenetelmän mukaisen logaritmissen uskottavuusfunktion alarajan derivaatat malliparametrien suhteen ovat

$$\begin{aligned}\frac{\partial}{\partial \alpha_t} \underline{l}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}; q) &= \sum_{i=1}^n \sum_{j=1}^p \frac{\partial}{\partial \alpha_t} \left[y_{ij} \alpha_i - \exp\left(\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right) \right] \\ &= \sum_{i=1}^n \left[y_{tj} - \exp\left(\alpha_t + \beta_j + \boldsymbol{\mu}'_t \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_t \boldsymbol{\gamma}_j\right) \right]\end{aligned}$$

kaikilla $t = 1, 2, \dots, n$ ja

$$\begin{aligned}\frac{\partial}{\partial \beta_s} \underline{l}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}; q) &= \sum_{i=1}^n \sum_{j=1}^p \frac{\partial}{\partial \beta_s} \left[y_{ij} \beta_j - \exp\left(\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right) \right] \\ &= \sum_{i=1}^n \left[y_{is} - \exp\left(\alpha_i + \beta_s + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_s + \frac{1}{2} \boldsymbol{\gamma}'_s \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_s\right) \right]\end{aligned}$$

kaikilla $s = 1, 2, \dots, p$. Lisäksi

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\gamma}_s} \underline{l}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}; q) &= \sum_{i=1}^n \sum_{j=1}^p \frac{\partial}{\partial \boldsymbol{\gamma}_s} \left[y_{ij} (\boldsymbol{\mu}'_i \boldsymbol{\gamma}_j) - \exp\left(\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right) \right] \\ &= \sum_{i=1}^n \left[y_{is} \boldsymbol{\mu}_i - \exp\left(\alpha_i + \beta_s + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_s + \frac{1}{2} \boldsymbol{\gamma}'_s \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_s\right) \frac{\partial}{\partial \boldsymbol{\gamma}_s} \left(\boldsymbol{\mu}'_i \boldsymbol{\gamma}_s + \frac{1}{2} \boldsymbol{\gamma}'_s \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_s\right) \right].\end{aligned}$$

Koska $\boldsymbol{\Lambda}_i$ on kovarianssimatriisina symmetrinen, pätee

$$\frac{\partial}{\partial \boldsymbol{\gamma}_s} \boldsymbol{\gamma}'_s \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_s = 2 \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_s.$$

Nyt

$$\frac{\partial}{\partial \boldsymbol{\gamma}_s} \underline{l}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}; q) = \sum_{i=1}^n \left[y_{is} \boldsymbol{\mu}_i - (\boldsymbol{\mu}_i + \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_s) \exp\left(\alpha_i + \beta_s + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_s + \frac{1}{2} \boldsymbol{\gamma}'_s \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_s\right) \right].$$

kaikilla $s = 1, 2, \dots, p$.

Variaatiomenetelmän mukaisen logaritmissen uskottavuusfunktion alarajan derivaatat variaatioparametrien suhteen ovat

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\mu}_t} \underline{l}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}; q) &= \sum_{i=1}^n \sum_{j=1}^p \frac{\partial}{\partial \boldsymbol{\mu}_t} \left[y_{ij} \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j - \exp\left(\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right) \right] - \frac{1}{2} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\mu}_t} \boldsymbol{\mu}'_i \boldsymbol{\mu}_i \\ &= \sum_{j=1}^p \left[y_{tj} \boldsymbol{\gamma}_j - \exp\left(\alpha_t + \beta_j + \boldsymbol{\mu}'_t \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_t \boldsymbol{\gamma}_j\right) \frac{\partial}{\partial \boldsymbol{\mu}_t} \boldsymbol{\mu}'_t \boldsymbol{\gamma}_j \right] - \frac{2 \boldsymbol{\mu}_t}{2} \\ &= \sum_{j=1}^p \left[\left(y_{tj} - \exp\left(\alpha_t + \beta_j + \boldsymbol{\mu}'_t \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_t \boldsymbol{\gamma}_j\right) \right) \boldsymbol{\gamma}_j \right] - \boldsymbol{\mu}_t\end{aligned}$$

ja

$$\begin{aligned}
\frac{\partial}{\partial \lambda_{tk}} l(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}; q) &= \sum_{i=1}^n \sum_{j=1}^p \frac{\partial}{\partial \lambda_{tk}} \left[-\exp\left(\alpha_i + \beta_j + \boldsymbol{\mu}'_i \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_i \boldsymbol{\gamma}_j\right) \right] \\
&+ \frac{1}{2} \sum_{i=1}^n \frac{\partial}{\partial \lambda_{tk}} [\log(\det(\boldsymbol{\Lambda}_i)) - \text{trace}(\boldsymbol{\Lambda}_i)] \\
&= \sum_{j=1}^p \left[-\exp\left(\alpha_t + \beta_j + \boldsymbol{\mu}'_t \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_t \boldsymbol{\gamma}_j\right) \frac{\partial}{\partial \lambda_{tk}} \left(\frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_t \boldsymbol{\gamma}_j\right) \right] \\
&+ \frac{1}{2} \frac{\partial}{\partial \lambda_{tk}} [\log(\det(\boldsymbol{\Lambda}_t)) - \text{trace}(\boldsymbol{\Lambda}_t)] \\
&= \sum_{j=1}^p \left[-\exp\left(\alpha_t + \beta_j + \boldsymbol{\mu}'_t \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_t \boldsymbol{\gamma}_j\right) \frac{\partial}{\partial \lambda_{tk}} \frac{1}{2} \sum_{l=1}^m \gamma_{jl}^2 \lambda_{il} \right] \\
&+ \frac{1}{2} \frac{\partial}{\partial \lambda_{tk}} \left[\sum_{l=1}^m \log \lambda_{il} - \sum_{l=1}^m \lambda_{il} \right] \\
&= \sum_{j=1}^p \left[-\frac{1}{2} \gamma_{jk}^2 \exp\left(\alpha_t + \beta_j + \boldsymbol{\mu}'_t \boldsymbol{\gamma}_j + \frac{1}{2} \boldsymbol{\gamma}'_j \boldsymbol{\Lambda}_t \boldsymbol{\gamma}_j\right) \right] + \frac{1}{2} \left(\frac{1}{\lambda_{ik}} - 1 \right)
\end{aligned}$$

kaikilla $t = 1, 2, \dots, n$ ja $k = 1, \dots, m$.

E R-koodi

```
#####  
# Gaussian variational approximation for Poisson mixed model  
#####
```

```
GVAforPMM.6 <- function(y,X,ni ,p,max.iter ,par0 ,parV0)  
{  
  # m = number of site  
  # ni = number of measurements in site i  
  # p-1 = number of covariates  
  m<-dim(y)[1]  
  
  # Note. In count data covariate values at site are  
  # often same for different measurements  
  X2<-matrix(nrow=m, ncol=p);  
  for(i in 1:m) {X2[i ,]<-X[i*ni ,]}  
  
  # Initial values  
  new.mu <- mu <- parV0[1:m]  
  new.lambda <- lambda <- parV0[(m+1):length(parV0)]  
  new.beta <- beta <- par0[1:p]  
  new.s2 <- s2 <- par0[p+1]  
  par <- c(new.beta,new.s2)  
  parV <- c(new.mu,new.lambda)  
  current.loglik <- -1e6;  
  iter <- 1;  
  err <- 10;  
  talku<-Sys.time(); # timing  
  
  # Estimation  
  while((err >1.0001 || err <0.9999) && iter <= max.iter)  
  {  
    # Lowerbound for log-likelihood (without the constant)  
    ll <- function(x,v)  
    {  
      new.beta <- x[1:p]  
      new.s2 <- x[p+1]  
      new.mu <- v[1:m]  
      new.lambda <- v[(m+1):length(v)]  
      llsum <- 0  
      for(i in 1:m)  
      {  
        lleq <- 0
```



```

    for(j in 1:ni)
    {
      eta<-(t(new.beta)%*%X2[i,])[1,1]+new.mu[i]
      lleq<-lleq+y[i,j]*eta
      -exp(eta+new.lambda[i]/2)
    }
    llsum<- llsum + lleq + log(new.lambda[i])/2
      -(new.mu[i]^2+new.lambda[i])/new.s2/2
  }
  llsum + m*(1-log(new.s2))/2
}

```

Gradient for beta ja s2

```

grad.beta.s <- function(v,x)
{
  new.beta <- x[1:p]
  new.s2 <- x[p+1]
  new.mu <- v[1:m]
  new.lambda <- v[(m+1):length(v)]
  betaosa <- rep(0,p)
  for (i in 1:m)
  {
    h.sum <- rep(0,p)
    for(j in 1:ni)
    {
      eta<-(t(new.beta)%*%X2[i,])[1,1]+new.mu[i]
      h.sum<-h.sum+(y[i,j]-exp(eta
      +new.lambda[i]/2))*X2[i,]
    }
    betaosa <- betaosa + h.sum
  }
  s2osa<- -0.5*m/new.s2+0.5*sum(new.lambda)/
    (new.s2)^2+0.5*sum((new.mu)^2)/(new.s2)^2
  gradientti<-c(betaosa ,s2osa)
  return(gradientti)
}

```

Estimation , modelparameters

```

ptm <- proc.time()
q<-try(optim(par ,fn=ll ,gr=grad.beta.s ,v=parV ,
method="BFGS" ,control=list(fnscale=-1)),silent=TRUE)
new.beta <- q$par[1:p]
new.s2 <- q$par[p+1]
par <- c(new.beta ,new.s2)
cat("Parameters_calculated_in_time:\n")

```

```

print((proc.time()-ptm)[3])
print(par)

# Gradient for mu ja lambda
grad.mu.lam <- function(v,x)
{
  new.beta <- x[1:p]
  new.s2 <- x[p+1]
  new.mu <- v[1:m]
  new.lambda <- v[(m+1):length(v)]
  muosa<-rowSums(y)-ni*exp(X2%*%new.beta)*exp(
    new.mu+0.5*new.lambda)-new.mu/new.s2
  lambdaosa<- -(ni/2)*exp(X2%*%new.beta)*exp(new.mu
    +0.5*new.lambda)+0.5*(1/new.lambda)
    -0.5*(1/rep(new.s2 ,m))
  gradientti<-c(muosa,lambdaosa)
  return(gradientti)
}

# Estimation, variational parameters
ptm <- proc.time()
q<-try(optim(parV,fn=ll ,gr=grad.mu.lam, x=par,
  method="BFGS",control=list(fnscale=-1,maxit=150)),
  silent=TRUE)
new.mu <- q$par[1:m]
new.lambda <- q$par[(m+1):length(q$par)]
parV <- c(new.mu,new.lambda)
cat("Parameters_calculated_in_time:\n")
print((proc.time()-ptm)[3])
print(parV)

new.loglik <- q$value
err<-abs(new.loglik/current.loglik);
cat("New_Loglik:", new.loglik, "Current_Loglik:",
  current.loglik, "Ratio", err, "\n")
current.loglik <- new.loglik;
iter = iter + 1
}

tloppu<-Sys.time();aika<-difftime(tloppu,talku)
return(list(MALPAR=par,VARPAR=parV,VALUE=q$value,
  ITER=(iter-1),AIKA=aika))
);

```

```
#####
```

```

# Approximate standard errors for model parameters
# install.packages("numDeriv")
library(numDeriv)
seGVA <- function( f, fitGVA, ... )
{
  par <- c(fitGVA$MALPAR, fitGVA$VARPAR)
  H<-hessian(f, par, ...)
  FishIn<-solve(-H)
  se<-sqrt(diag(FishIn))#[1:5]
  return(se)
}
#####
# Diagnostics: Dunn-Smyth residuals
pit.site.mixed <- function(y, X, fitGVA, family)
{
  beta<-fitGVA$MALPAR[1:(length(fitGVA$MALPAR)-1)]
  U<-fitGVA$VARPAR[1:(length(fitGVA$VARPAR)/2)]
  N <- length(y)
  eta <- exp(X%*%beta+U)
  if(family == "poisson") {a<-ppois(y-1, eta);
  b<-ppois(y, eta)}
  u <- runif(n = length(y), min = a, max = b)
  qnorm(u)
}

# Dunn-Smyth residuals vs. linear predictor
# Dunn-Smyth residuals vs. theoretical quantiles
pit.res.mixed <- function(data, X, fitGVA, family)
{
  beta<-fitGVA$MALPAR[1:(length(fitGVA$MALPAR)-1)]
  U<-fitGVA$VARPAR[1:(length(fitGVA$VARPAR)/2)]

  ## Dunn-Smyth residuals
  pitres <- pit.site.mixed(y=data, X=X, fitGVA, family)
  ind<-as.vector(unlist(pitres))<100000

  ## linear predictor
  lin.pred.mixed <- X%*%beta+U

  par(mfrow=c(1,2), mgp=c(2,1,0))
  qqnorm(as.vector(unlist(pitres))[ind], main = "",
        xlab="Teoreettiset kvantiilit",
        ylab="Dunn-Smyth jaannokset");
  abline(0,1)
  plot(lin.pred.mixed, pitres,

```

```

        ylab="Dunn-Smyth_ jaannokset ",
        xlab="Lineariset_ennusteet")

    return(list(res = pitres , linpred = lin.pred.mixed))
};
#####
# Simulation for Poisson mixed model
# m=200, Niter=500 (m=50,100)
library(lme4)
X<-matrix(rep(c(rep(c(1,-1),5),rep(c(1,1),5))),100),
          ncol=2,byrow=T)
id<-sort(rep((1:200),5))
beta<-c(-1,1); sigma<-1
m<-200; ni<-5; p<-length(beta)
X2<-matrix(nrow=m,ncol=p); for(i in 1:m) {X2[i,]<-X[i*ni,]}
niter<-500
simLap<-NULL; simGVA1<-NULL; simGVA2<-NULL; simGVA3<-NULL

t1<-Sys.time(); for(k in 1:niter)
{
  cat("Kierros_",k,"\n")

  # Data simulation
  U <- rnorm(m,mean=0,sd=sigma)
  Ysim <-matrix(ncol=ni, nrow=m)
  for(j in 1:ni){ Ysim[,j]<- rpois(m, exp(X2%*%beta+U))}
  Ys<-as.vector(t(Ysim))

  # Laplace estimates
  ptm <- proc.time()
  fitsim0<-glmer(Ys~X[,2]+(1|id),family=poisson)
  simLap <- rbind( simLap, c(fitsim0@beta ,fitsim0@theta^2))
  uL <- fitsim0@u
  cat("Laplace_approximation_computed_in_time:\n")
  print((proc.time()-ptm)[3])

  # GVA, different initial values
  fitsim1 <- GVAforPMM.6(y=Ysim,X=X,ni=ni,p=p,max.iter=20,
  par0=c(rep(0,p),1), parV0=c(rep(0,m),rep(1,m)))
  simGVA1 <-rbind( simGVA1, fitsim1$MALPAR)

  fitsim2 <- GVAforPMM.6(y=Ysim,X=X,ni=ni,p=p,max.iter=20,
  par0=c(fitsim0@beta ,fitsim0@theta^2),
  parV0=c(uL,rep(1,m)))
  simGVA2 <-rbind( simGVA2, fitsim2$MALPAR)

```

```

ran.par <- c(rnorm(p,0,sd=1),abs(rnorm(1,1,sd=1)));
ran.parV <- c(rnorm(m,0,sd=1),abs(rnorm(m,1,sd=1)))
fitsim3 <- GVAforPMM.6(y=Ysim,X=X,ni=ni,p=p,max.iter=30,
par0=ran.par, parV0=ran.parV )
simGVA3 <- rbind( simGVA3,fitsim3$MALPAR)
}; t2<-Sys.time(); aika.koko<-difftime(t2,t1)

```

```

list (SIMLAP=simLap ,SIMGVA1=simGVA1 ,SIMGVA2=simGVA2,
SIMGVA3=simGVA3)

```

```

#####
# GLVM - estimation for Poisson data using (Gaussian)
# variational approximation
#####

```

```

QVApoisLVM.3<-function(data,att.dist="p",num.lv,max.iter=50,
start.index=NULL)

```

```
{
```

```

talku<-Sys.time();
n<-dim(data)[1] # number of sites
p<-dim(data)[2] # number of spieces
ptm <- proc.time()

```

```
## Starting values obtained by fitting marginal GLMs
```

```

res <- start.values(data,att.dist="p",num.lv);
new.b <- b <- res$params[,1]; new.g <- g
<- res$params[,2:3]; new.mu <- mu <- res$index;
new.a <-a <-rep(0,n); new.lambda<-lambda <- rep(1,n)

```

```

if(length(start.index)>0) { new.mu <- mu <- start.index}
cat("Starting_values_computed_in_time:\n")
print((proc.time()-ptm)[3])
current.loglik <- -1e6;
iter <- 1;
err <- 10;

```

```

while((err > 1.0005 || err < 0.9995) && iter <= max.iter)
{

```

```

## Update model params by maximizing
## lower bound for loglik
l10 <- function(x,v)
{
new.g <- c(x[1:p],0,x[(p+1):(2*p-1)])
new.g <- matrix(new.g,p,2)

```

```

new.b <- x[(2*p):(2*p+p-1)]
new.a <- x[(2*p+p):(2*p+p+n-1)]
new.mu <- v[(n+1):length(v)]
new.mu <- matrix(new.mu, n, 2)
new.lambda <- v[1:n]
llsum <- 0
for (i in 1:n)
{
  lleq <- 0
  for(j in 1:p)
  {
    eta <- new.a[i]+new.b[j]+(t(new.mu[i,])
      %*% new.g[j,])[1,1]
    lleq <- lleq+data[i,j]*eta-exp(eta
      +new.lambda[i]+new.lambda[i]*
      (t(new.g[j,])%*%(new.g[j,]))[1,1]/2))
    llsum <- llsum+lleq+log(new.lambda[i])
      -((t(new.mu[i,])%*% new.mu[i,])[1,1]
      +2*new.lambda[i])/2
  }
  llsum
}

```

Gradient of g, a and b (model parameters)

```

grad.mod <- function(x,v)
{
  new.g <- c(x[1:p],0,x[(p+1):(2*p-1)])
  new.g <- matrix(new.g,p,2)
  new.b <- x[(2*p):(2*p+p-1)]
  new.a <- x[(2*p+p):(2*p+p+n-1)]
  new.mu <- v[(n+1):length(v)]
  new.mu <- matrix(new.mu,n,2)
  new.lambda <- v[1:n]
  grad.g<-matrix(0,p,2)
  grad.b<-vector(length=p)
  grad.a<-vector(length=n)
  # for a -----
  for(i in 1:n)
  {
    apu<-0
    for(j in 1:p)
    {
      phi <- new.a[i]+new.b[j]+(t(new.mu[i,])%*%
        new.g[j,])[1,1]+new.lambda[i]*
        (t(new.g[j,])%*%(new.g[j,]))[1,1]/2
    }
  }
}

```

```

        apu <- apu + exp(phi)
      }
      grad.a[i] <- sum(data[i,]) - apu
    }
  # for g and b -----
  for(j in 1:p)
  {
    apub<-0
    apug<-c(0,0)
    for(i in 1:n)
    {
      nu <- new.a[i]+new.b[j]+(t(new.mu[i,])%*%
        new.g[j,])[1,1] + new.lambda[i]*
        (t(new.g[j,])%*%(new.g[j,]))[1,1]/2
      apub <- apub + exp(nu)
      apug <- apug+data[i,j]*new.mu[i,]-(new.mu[i,]
        + new.lambda[i]*new.g[j,])*exp(nu)
    }
    grad.b[j] <- sum(data[,j]) - apub
    grad.g[j,]<- apug
  }
  grad.g<-as.vector(grad.g)
  return(c(grad.g[-(p+1)], grad.b ,grad.a))
};

ptm <- proc.time()
q<-try(optim(c(g,b,a)[-(p+1)],method="BFGS",fn=l10 ,
  gr=grad.mod,v=c(new.lambda,new.mu),
  control=lisT(trace=0,fnscale=-1), silent=T)
if(!inherits(q, "try-error"))
{
  new.g <- matrix(c(q$par[1:p],0 ,
    q$par[(p+1):(2*p-1)]),p,2);
  new.b <- q$par[(2*p):(2*p+p-1)];
  new.a <- q$par[(2*p+p):(2*p+p+n-1)]
} else { new.g <- g; new.b <- b; new.a <- a }

cat("Parameters_calculated_in_time:\n")
print((proc.time()-ptm)[3])
print(new.g)

## Update all variational params by maximizing
## lower bound for loglik

# gradient for variational parameters

```

```

grad.var <- function(x, v)
{
  new.g <- c(x[1:p], 0, x[(p+1):(2*p-1)])
  new.g <- matrix(new.g, p, 2)
  new.b <- x[(2*p):(2*p+p-1)]
  new.a <- x[(2*p+p):(2*p+p+n-1)]
  new.mu <- v[(n+1):length(v)]
  new.mu <- matrix(new.mu, n, 2)
  new.lambda <- v[1:n]
  grad.mu <- matrix(0, n, 2)
  grad.lambda <- vector(length=n)

  # for mu and lambda—————
  for(i in 1:n)
  {
    apum <- c(0, 0)
    apul <- 0
    for(j in 1:p)
    {
      phi <- new.a[i] + new.b[j] + (t(new.mu[i,])%*%
        new.g[j,])[1, 1] + new.lambda[i] *
        (t(new.g[j,])%*%(new.g[j,]))[1, 1] / 2
      apum <- apum + (data[i, j] - exp(phi)) * new.g[j, ]
      apul <- apul - exp(phi) * (t(new.g[j,])%*%
        (new.g[j,]))[1, 1] / 2
    }
    grad.mu[i, ] <- apum - new.mu[i, ]
    grad.lambda[i] <- apul + 1/new.lambda[i] - 1
  }
grad.mu <- as.vector(grad.mu)
return(c(grad.lambda, grad.mu))
};

ptm <- proc.time()
q <- try(optim(c(lambda, mu), method="BFGS", fn=l10,
  gr=grad.var, x=c(new.g, new.b, new.a)[-(p+1)],
  control=list(trace=0, fnscale=-1)), silent=T)
if(!inherits(q, "try-error"))
{
  new.lambda <- q$par[1:n]; new.mu <-
  q$par[(n+1):(length(q$par))]; new.mu <-
  matrix(new.mu, n, 2)
} else { new.lambda <- lambda; new.mu <- mu }

  cat("Parameters_calculated_in_time:\n")

```



```

        print((proc.time()-ptm)[3])
        print(new.lambda)

    ## Take values of loglik from optim -function
    ## to define stopping rule
    new.loglik <- q$value
    err <- abs(new.loglik/current.loglik);
    cat("New_Loglik:",new.loglik,"Current_Loglik:",
        current.loglik,"Ratio",err,"\n")
    current.loglik <- new.loglik;

    b <- new.b
    a <- new.a
    g <- new.g
    mu <- new.mu
    lambda <- new.lambda
    iter = iter + 1
} # end of iteration

## Plot ordination points.
index<-mu
plot(index,type="n")
text(index,labels=seq(1,n))
tloppu<-Sys.time(); aika<-difftime(tloppu,talku)
return(list(site.params=a, spp.params=b, gamma=g,
index=index, mu=mu, lambda=lambda, iter=iter,
value=q$value, time=aika))
};
#####
# Starting values
# install.packages("mvabund")
# install.packages("mvtnorm")
library(mvtnorm)
library(mvabund)
start.values <- function(data, att.dist="p", num.lv = 2)
{
  N <- nrow(data)
  p <- ncol(data)
  data <- as.matrix(data)
  allb<-NULL
  par <- array(0, dim=c(p,num.lv+2,10))
  for(j in 1:10){
    allb<-cbind(allb ,rmvnorm(N,rep(0,num.lv)))}
  AIC <- NULL
  if(att.dist=="b")

```

```

{
  for (i in 1:10) {
    fit <- manyglm(data~allb[, (2*i-1):(2*i)],
                  family="binomial")
    par[, , i] <- cbind(t(fit$coef), rep(NA, p))
    AIC <- c(AIC, sum(fit$AIC)) }
  m <- (which(AIC == min(AIC)))[1]
  params <- par[, , m]
  index <- allb[, (2*m-1):(2*m)]
}
if(att.dist=="p")
{
  for (i in 1:10) {
    fit <- manyglm(data~allb[, (2*i-1):(2*i)],
                  family="poisson")
    par[, , i] <- cbind(t(fit$coef), rep(NA, p))
    AIC <- c(AIC, sum(fit$AIC)) }
  m <- (which(AIC == min(AIC)))[1]
  params <- par[, , m]
  index <- allb[, (2*m-1):(2*m)]
}
if(att.dist=="n")
{
  for (i in 1:10) {
    fit <- manyglm(data~allb[, (2*i-1):(2*i)],
                  family="normal")
    get.phi <- fit$phi
    par[, , i] <- cbind(t(fit$coef), get.phi)
    AIC <- c(AIC, sum(fit$AIC)) }
  m <- (which(AIC == min(AIC)))[1]
  params <- par[, , m]
  index <- allb[, (2*m-1):(2*m)]
}
if(att.dist=="nb")
{
  for (i in 1:10) {
    fit <- manyglm(data~allb[, (2*i-1):(2*i)],
                  family="negative.binomial")
    get.phi <- fit$phi; get.phi[get.phi == 0] <- 1e-5
    par[, , i] <- cbind(t(fit$coef), get.phi)
    AIC <- c(AIC, sum(fit$AIC)) }
  m <- (which(AIC == min(AIC)))[1]
  params <- par[, , m]
  index <- allb[, (2*m-1):(2*m)]
}

```

```

    list (params=params , index=index)
};

#####
### Hui et al. 2014:
# Basic residual analysis for LVMs
# Uses the pit.site function to calculate Dunn-Smyth
# residuals

pit.res<-function(data,index,params,att.dist,site.params)
{
  pitres <- data
  d <- dim(params)[2]; p <- dim(data)[2]; n <- dim(data)[1]
  if(length(att.dist)==1) {
    att.dist<-rep(att.dist[1],ncol(data))}
  palette(rainbow(p))
  for (a in 1:p) { pitres[,a] <- pit.site(data[,a],
    index,params[a,],site.params,family = att.dist[a]) }

  ## linear predictor
  lin.pred <- cbind(rep(1,dim(data)[1]),index) %*%
    t(params[,1:(d-1)])
  par(mfrow=c(1,2), cex=1)
  matplot(lin.pred, pitres, ylab="Dunn-Smyth_Jaannokset",
    xlab="Lineaariset_ennusteet", type="n")

  for(i in 1:p) { points(lin.pred[,i],pitres[,i]) }
  abline(0,0,lty=2)
  ind<-as.vector(unlist(pitres))<100000
  qqnorm(as.vector(unlist(pitres))[ind], main = "",
    ylab="Dunn-Smyth_Jaannokset",
    xlab="Teoreettiset_kvantiilit");
  abline(0,1)
  palette("default")
  return(list(res = pitres, linpred = lin.pred))
}

pit.site <- function(y, index, params, site.params, family)
{
  N <- length(y)
  Z <- cbind(rep(1,N),index)
  d <- length(params)
  mu <- exp(Z%*%params[1:(d-1)]+site.params)
  if(family == "poisson") {
    a<-ppois(y-1, mu);b<-ppois(y, mu)}

```

```

    if(family == "negative.binomial") {
      a <- pnbinom(y-1, size=1/params[d], mu=mu);
      b <- pnbinom(y, size=1/params[d], mu=mu) }
    if(family == "binomial") {
      p <- exp(Z%*%params[1:(d-1)]+site.params)/(1
              +exp(Z%*%params[1:(d-1)]+site.params))
      a <- pbinom(y - 1, 1, p); b <- pbinom(y, 1, p) }
    u <- runif(n = length(y), min = a, max = b)
    qnorm(u)
  }
#####
# Confidence interval
index <- indexspider
lambda <- spiderlambda2
n <- 28
ci <- 1.96 * sqrt(lambda/n)
symbols(index[,1], index[,2], xlab="Z1", ylab="Z2",
        squares=ci, fg="red")
text(index, labels=seq(1,28))

#####
# LVM-based simulations to check the performance of NMDS
# and LVM

# install.packages("vegan")
# install.packages("MASS")
library(vegan)
library(MASS)

# Fit LVM model once and use estimates to generate the data

index <- indexspider2
index.scaled <- scale(index, scale = T)
params <- spiderparams2
site.params <- spidersitepar2$x

sim.data <- function(index, params, att.dist, site.params)
{
  d <- dim(params)[2]
  N <- dim(index)[1]
  X <- cbind(rep(1,N), index)
  sigma <- NULL
  y <- NULL
  for (j in 1:length(att.dist)) {
    ally1 <- NULL

```

```

        nul <- X %*% params[j,]
        for (i in 1:N) {
            allmu <- exp(nul[i]+site.params[i])
            ally1 <- c(ally1, rpois(1, allmu)) }
        y <- cbind(y, ally1) }
    list(y=y)
}

dist1 <- NULL; dist2<-NULL; dist3<-NULL
ind1<-NULL; ind2<-NULL; ind3<-NULL
corr1 <-NULL; corr2 <-NULL; corr3 <-NULL

rep<-2500
talku<-Sys.time()
for (i in 1:rep)
{
    cat("Iteration_", i, "\n")
    # Do not allow datasets with rows/columns full of zeros
    nonz<-1
    while (nonz > 0)
    {
        dat<-sim.data(index, params, att.dist=rep("p", 12),
            site.params)
        y <- dat$y
        if ((sum(apply(y, 1, sum)==0)+sum(apply(y, 2, sum)==0))>0)
            { nonz<-1 }
        if ((sum(apply(y, 1, sum)==0)+sum(apply(y, 2, sum)==0))==0)
            { nonz<-0 }
    }
    bray <- metaMDS(y, k=2)
    ind1<-cbind(ind1, bray$points)
    pro1 <- procrustes(index.scaled, bray$points)
    dist1<-rbind(dist1, apply((pro1$Yrot-pro1$X)^2, 1, sum))
    corr1 <- c(corr1, protest(index.scaled, bray$points)$t0)
    fit<-try(QVApoisLVM.3(y, att.dist="p", num.lv=2,
        max.iter=20, silent=F) # Ordination given by new method
    apu.index<-cbind(bray$points[, 1], bray$points[, 2])
    fit2<-try(QVApoisLVM.3(y, att.dist="p", num.lv=2,
        max.iter=20, start.index=apu.index), silent=F)

    # Scale the new points; not needed maybe?
    if(!inherits(fit, "try-error")) {
        fit$index.std<-fit$index
        fit$index <- scale(fit$index.std, scale = T)
        ind2<-cbind(ind2, fit$index)

```

```

pro2 <- procrustes(index.scaled, fit$index)
dist2<-rbind(dist2, apply((pro2$Yrot-pro2$X)^2,1,sum))
corr2 <- c(corr2, protest(index.scaled, fit$index)$t0)}

if(!inherits(fit2, "try-error")) {
  fit2$index.std<-fit2$index
  fit2$index <- scale(fit2$index.std, scale = T)
  ind3<-cbind(ind3, fit2$index)
  pro3<-procrustes(index.scaled, fit2$index)
  dist3<-rbind(dist3, apply((pro3$Yrot-pro3$X)^2,1,sum))
  corr3<-c(corr3, protest(index.scaled, fit2$index)$t0)}
};tloppu<-Sys.time();aika<-difftime(tloppu, talku);

```