

Johan Pekka Kalervo Sihvonen

MCMC

SHV-työ
Kesäkuu 2007

Abstract

As an almost standard approach for modelling (for example rate making) non-life insurance mathematicians use so called generalized linear models (GLM). Generalized linear models enable flexible regression modelling for non-normal data. For example modelling of claim numbers can be achieved through choosing the distribution of response to be Poisson distribution. In case of generalized linear models plenty of software exists for the parameter estimation and significance testing.

In this work fitting of so called hierarchical generalized linear model for the claim number data is considered. In practice this modelling means deriving so called Prior Distributions for parameters. Prior Distribution describes the distribution of the parameter before exploring the value of the response. Through Bayesian analysis so called Posterior Distributions for the parameters can be derived. From the Posterior Distributions parameter estimates which take into account prior information can be calculated. Parameter estimation for these kind of models can be done through Markov Chain Monte Carlo simulation with so called Metropolis-Hastings-algorithms.

Sisältö

1 Johdanto	1
2 Bayesiläinen päättely ja hierarkkiset mallit	1
3 Markovin ketjuista	4
3.1 Määritelmä ja siirtymätodennäköisyydet	4
3.2 Tila-avaruuden luokittelu	4
3.3 Stationaarinen jakauma ja jaksollisuus	5
4 MCMC-menetelmistä	7
4.1 Metropolis-Hastings-algoritmit	8
4.2 Erityyppisiä ehdotussiirtymäfunktioita	10
4.2.1 Symmetriset ketjut	11
4.2.2 Satunnaiskulkuketjut	11
4.2.3 Riippumattomat ketjut	12
4.3 Komponenteittaiset Metropolis-Hastings-algoritmit	13
4.4 Konvergenssitarkastelut	17
4.4.1 Teoreettinen tapa	17
4.4.2 Tilastollinen tapa	19
4.5 Metropolis-Hastings-algoritmin ohjelmoimisesta	20
5 Käytännön sovellus	21
5.1 Tarkasteltavat mallit	21
5.2 Yhden parametrin tilanne	22
5.3 Usean parametrin tilanne	29
6 Yhteenveto	39
Lähteet	40
Liite 1. Yhden parametrin Metropolis-Hastings SAS-koodi	41
Liite 2. Usean parametrin Metropolis-Hastings SAS-koodi	42

1 Johdanto

Vakuutusmatemaattiseen käytännön työhön jo lähes vakiintunut työkalu esimerkiksi hinnoittelussa, mutta myös muunlaisessa mallintamisessa, ovat yleistetyt lineaariset mallit. Yleistetyt lineaariset mallit laajentavat tavanomaisen regressioanalyysin ja sallivat vasteen jakaumalle myös muita vaihtoehtoja kuin normaalijakauman. Esimerkiksi vahinkojen lukumäärien mallintaminen onnistuu yleistettyjen lineaaristen mallien avulla valitsemalla vasteen jakaumaksi Poisson-jakauman.

Parametrien estimointiin ja tilastollisen merkitsevyyden testaamiseen on yleistettyjen lineaaristen mallien osalta olemassa runsaasti valmiita ohjelmistoja. Esimerkkinä mainittakoon SAS[®]-ohjelman proseduri Proc GenMod.

Tässä tutkielmassa tutkitaan niin kutsutun hierarkkisen yleistetyn lineaarisen mallin sovittamista liikennevahinkojen vahinkojen lukumääräaineistoon. Käytännössä tämä mallinnus tarkoittaa niin kutsuttujen *priori*-jakaumien määräämistä parametreille. Priori-jakauma kuvaa parametrin jakaumaa ennen vasteen arvon havaitsemista. Bayesiläisen päättelyn avulla määrätään parametreille priorinformaation huomioivat *posteriori*-jakaumat. Posteriorijakaumista saadaan lopulta halutut priorinformaation huomioon ottavat estimaatit parametreille. Tämänkaltaisille malleille parametrien estimointiin voi suorittaa MCMC-menetelmiin kuuluvilla Metropolis-Hastings-algoritmeilla.

2 Bayesiläinen päättely ja hierarkkiset mallit

Tässä tutkielmassa kuvataan vasteen Y arvoihin liittyvää epävarmuutta vasteen tiheys- tai todennäköisyysfunktiolla, josta käytetään merkintää $f(y|\theta)$. Vasteen jakauman oletetaan kuuluvan johonkin tiettyyn jakaumien luokkaan. Parametri θ on tarkasteltavaan jakaumien luokkaan liittyvä indeksi.

Kanonisessa tilanteessa satunnaisotos on poimittu jakaumasta, jonka tiheysfunktio on $f(y|\theta)$. Yleensä tässä tapauksessa havainnot y_i ovat riippumattomia ja samoin jakautuneita.

Yleensä parametrin θ merkitys mallintamisessa on selkeä ja sen luonteesta

on olemassa tietoa. Tämä tieto tulisi ottaa huomioon mallinnusvaiheessa. Bayesiläisessä analyysissä parametriin θ liittyvä *prioritieto* mallinnetaan tiheysfunktioilla $p(\theta)$. Tiheysfunktion $p(\theta)$ määääjää jakaumaa kutsutaan priorijakaumaksi, sillä se kuvaa parametrin θ jakaumaa ennen kuin on havaittu vasteen $\mathbf{Y}$ arvo.

Bayesiläiseen päättelyyn liittyy priorijakauman $p(\theta)$ lisäksi vasteen havaittu jakauma $f(y|\theta)$. Kuten vasteen $\mathbf{Y}$ havaittua jakaumaa kuvataan vakion (tässä vaiheessa siis tunnettu) θ avulla, niin myös priorijakaumalle voidaan määrätä jakauman määrävät parametrit. Joskus nämä parametrit on järkevää erottaa parametrista θ . Näitä parametreja kutsutaan *hyperparametreiksi*, sillä ne ovat parametrien jakaumien parametreja. Hyperparametrit oletetaan tunnetuiksi.

Kun vasteen havaittu jakauma tulkitaan parametrin θ funktioksi päädytään uskottavuusfunktioon $L(\theta) = f(y|\theta)$. Bayesiläisessä päättelyssä tutkitaan mikä on parametrin θ jakauma, kun on havaittu vasteen arvo y . Tätä jakaumaa kutsutaan posteriorijakaumaksi. Posteriorijakauma $p(\theta|y)$ voidaan laskea *Bayesin kaavan* avulla

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{f(y)}, \quad (1)$$

missä

$$f(y) = \int f(y|\theta)p(\theta)d\theta.$$

Koska yhtälön (1) vasen puoli on parametrin θ tiheysfunktio, niin tässä vasteen arvo y tulkitaan havaituksi vakioksi. Täten kaavassa (1) myös tiheysfunktion arvo $f(y)$ on vakio. Jatkossa posteriorijakaumasta käytetään merkintää $\pi(\theta)$. Havaintoarvo y sisältyy edelliseen merkintään implisiittisesti. Tällöin Bayesin kaava voidaan kirjoittaa kompaktimmassa muodossa

$$\pi(\theta) \propto L(\theta)p(\theta). \quad (2)$$

Parametrivektorin $\theta = (\theta_1, \dots, \theta_d)^T$ tapauksessa parametrin θ_i sekä marginaalinen että ehdollinen jakauma saadaan yhteistiheysfunktioista $\pi(\theta_1, \dots, \theta_d)$. Parametrin θ_i marginaalinen posteriorijakauma on

$$\pi(\theta_i) = \int \pi(\theta_1, \dots, \theta_d)d\theta_{-i},$$

missä $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$. Parametrin θ_i kaikki mahdolliset ehdolliset jakaumat ovat

$$\pi(\theta_i | \theta_j, j \in C) = \pi(\theta_i, \theta_j, j \in C) / \pi(\theta_j, j \in C), \quad (3)$$

missä $C \subset \{1, \dots, i-1, i+1, \dots, d\}$. Ehdollisista jakaumista tärkein on täysin ehdollistettu jakauma $\pi(\theta_i | \theta_{-i})$, jonka tiheysfunktioista käytetään jatkossa merkintää $\pi_i(\theta_i)$.

Sijoittamalla kaavaan (3) $C = \{\theta_j | j \neq i\}$ saadaan tulos

$$\pi(\theta) = \pi_i(\theta_i) \pi(\theta_{-i}), \quad (4)$$

jota tarvitaan jatkossa tutkittaessa Markovin ketjujen simulointia.

Kun posteriori-jakauma on saatu selville, niin on mielenkiintoista tiivistää siihen liittyvä tieto muutamaan tunnuslukuun. Erityisesti kiinnostavia ovat sijaintiin ja hajontaan liittyvät tunnusluvut. Sijaintia voidaan kuvata esimerkiksi posteriori-jakauman odotusarvolla sekä hajontaa posteriori-jakauman keskihajonnalla.

Usein ongelmien rakenne edellyttää tarkasteltavalta mallilta kvalitatiivista tietoa parametrivektorista β . Tämä tieto tulee jotenkin sisällyttää malliin. Eräs ratkaisu tähän ongelmaan ovat *hierarkkiset mallit*. Näissä malleissa priorijakauma määritellään kahdessa vaiheessa. Ensimmäisessä vaiheessa määritellään parametrivektorin β ehdollinen jakauma ehtomuuttujien avulla ja toisessa vaiheessa ehtomuuttujien jakaumat. Esimerkiksi hierarkkinen yleistetty lineaarinen malli, jossa priorijakaumien oletetaan olevan normaali-jakaumia, voidaan tällöin kirjoittaa muodossa

$$\mathbf{Y}_i | \mu_i \sim EF(\mu_i) \quad , \quad i = 1, \dots, n$$

$$\eta = \mathbf{X}_1 \beta_1$$

$$\beta_1 | \beta_2 \sim N(\mathbf{X}_2 \beta_2, \mathbf{C})$$

$$\beta_2 \sim N(\mathbf{b}, \mathbf{B}),$$

missä $\eta = (\eta_1, \dots, \eta_n)^T$, $\eta_i = g(\mu_i)$ ja $EF(\cdot)$ on eksponenttijakaumien perheen jakauma.

3 Markovin ketjuista

Tässä pykälässä esitellään pääpiirteissään Markovin ketjujen teoria. Markovin ketjut ovat stokastisia prosesseja. Markovin ketjujen perusominaisuuksien ja tulosten esittelyn jälkeen näytetään miten Markovin ketjuja voidaan käyttää iteratiivisissa simulointimenetelmissä.

3.1 Määritelmä ja siirtymätodennäköisyydet

Markovin ketju on stokastinen prosessi, missä prosessin tulevaisuus on riippumaton sen menneisyydestä. Käytetään ketjun tila-avaruudesta merkintää S ja parametrijoukosta merkintää T . Tällöin formaalisesti ilmaistuna saadaan seuraava määritelmä.

Määritelmä 3.1. Stokastinen prosessi $\{\theta^{(t)} : t \in T\}$ on Markovin ketju mikäli

$$\begin{aligned} P(\theta^{(n+1)} \in A | \theta^{(n)} = x, \theta^{(n-1)} \in A_{n-1}, \dots, \theta^{(0)} \in A_0) \\ = P(\theta^{(n+1)} \in A | \theta^{(n)} = x) \end{aligned} \quad (5)$$

kaikilla joukoilla $A_0, \dots, A_{n-1}, A \subset S$ ja arvoilla $x \in S$. $\square$

Yhtälöä (5) kutsutaan *Markovin ominaisuudeksi*.

Yleisesti ottaen todennäköisyydet kaavassa (5) riippuvat tekijöistä x, A ja n , kun nämä todennäköisyydet eivät riipu ajasta n , niin Markovin ketjun sanotaan olevan *homogeeninen*. Mikäli Markovin ketju on homogeeninen, niin voidaan määritellä siirtymäfunktio $P(x, A)$ seuraavasti:

1. Kaikilla $x \in S$, $P(x, \cdot)$ on tila-avaruudessa S määritelty todennäköisyysjakauma.
2. Funktio $P(x, A)$ on määritelty kaikilla osajoukoilla $A \subset S$.

3.2 Tila-avaruuden luokittelu

Seuraavaksi määritellään muutamia Markovin ketjujen tilojen luokitteluun liittyviä käsitteitä. Merkitään

$$T_i = \inf\{n > 0 | \theta^{(n)} = i\}.$$

Tällöin T_i on ensimmäinen hetki, jolloin Markovin ketju käy tilassa i .

Todennäköisyys sille, että tilasta x alkava ketju käy jossakin vaiheessa tilassa y , on

$$\rho_{xy} = P(T_y < \infty).$$

Tilan $y \in S$ sanotaan olevan *toistuva*, mikäli Markovin ketju, joka alkaa tilasta y , palaa tilaan y todennäköisyydellä 1 ($\rho_{yy} = 1$). Vastaavasti tila y on *tilapäinen*, mikäli todennäköisyys sille, että tilaan y ei palata, on positiivinen ($\rho_{yy} < 1$).

Tutkitaan tila-avaruuden S jakamista osajoukkoihin, jotka koostuvat toistuvista ja tilapäisistä tiloista. Sanotaan, että tila x on *yhteydessä* tilaan y (merkintä $x \rightarrow y$), mikäli $\rho_{xy} > 0$. Tila-avaruuden S osajoukon C sanotaan olevan *suljettu*, jos

$$\rho_{xy} = 0 \quad \forall x \in C \text{ ja } y \notin C.$$

Osajoukko C on *redusoitumaton*, mikäli kaikilla pareilla $x, y \in C$ on voimassa $x \rightarrow y$. Markovin ketjun sanotaan olevan *redusoitumaton*, mikäli tila-avaruus S on *redusoitumaton*. Mikäli $\rho_{xy} > 0$, niin $P^n(x, y) > 0$ ja päinvastoin. Tämän ominaisuuden nojalla voidaan osoittaa, että jos $x \in S$ on toistuva ja $x \rightarrow y$, niin myös y on toistuva [8]. Jos on voimassa $x \rightarrow y$ ja $y \rightarrow x$, niin tilat x ja y ovat *keskenään yhteydessä*, ja merkitään $x \leftrightarrow y$. Relaatio $\leftrightarrow$ on ekvivalenssirelaatio [8].

3.3 Stationaarinen jakauma ja jaksollisuus

Markovin ketjujen simuloinnin kannalta on tärkeää tietää ketjujen asymp-toottisesta käyttäytymisestä.

Määritelmä 3.2. Jakauma π on Markovin ketjun (siirtymätodennäköisyydet $P(x, y)$) *stationaarinen jakauma*, mikäli

$$\sum_{x \in S} \pi(x) P(x, y) = \pi(y), \quad \forall y \in S. \quad \square \quad (6)$$

Yhtälö (6) voidaan kirjoittaa myös matriisimuodossa $\pi = \pi P$. Mikäli Markovin ketju kerran päätyy stationaariseen jakaumaan, niin se yhtälön (6) nojalla myös pysyy siinä tämän jälkeen.

Lähteessä [2] on osoitettu, että jos stationaarinen jakauma π on olemassa ja $\lim_{n \rightarrow \infty} P^n(x, y) = \pi(y)$, niin tällöin riippumatta ketjun alkutilan jakaumasta $(\pi^{(0)})$ $\pi^{(n)}$ lähestyy jakaumaa π , kun $n \rightarrow \infty$.

Määritelmä 3.3. Markovin ketjun tilan x jakso d_x on joukon

$$\{n \geq 1 : P^n(x, x) > 0\}$$

suurin yhteinen tekijä. □

Selvästi jos $P(x, x) > 0$, niin $d_x = 1$ ja jos $x \leftrightarrow y$, niin $d_x = d_y$. Täten redusoitumattoman ketjun tiloilla on sama jakso. Tila x on jaksoton, mikäli $d_x = 1$.

Käytetään tilaan i paluajan odotusarvosta merkintää

$$m_i = E(T_i | \theta^{(0)} = i).$$

Tila i on *positiivisesti toistuva*, mikäli $m_i < \infty$. Positiivisesti toistuva jaksoton tila on *ergodinen*. Markovin ketju on jaksollinen jaksona $d > 1$, mikäli sen tilat ovat d -jaksollisia. Vastaavasti ketju on jaksoton, jos kaikki ketjun tilat ovat jaksottomia. Markovin ketju on ergodinen, jos kaikki sen tilat ovat ergodisia.

Jaksottomuus on liiallinen ehto stationaarisen jakauman olemassaololle, mutta se takaa siirtymätodennäköisyyksien suppenemisen.

Ergodisille ketjuille on voimassa niin kutsuttu ergodinen teoreema (ergodic theorem), joka on suurten lukujen lain vastine Markovin ketjuille.

Teoreeman esittelyä varten määritellään ensin *ergodisen keskiarvon* käsite.

Määritelmä 3.4. Reaaliarvoisen funktion $t(\theta)$ ergodinen keskiarvo on

$$\bar{t}_n = (1/n) \sum_{i=1}^n t(\theta^{(i)}). \quad \square$$

Seuraavaksi esitetään ergodinen teoreema.

Lause 3.5. Ergodinen teoreema Jos Markovin ketju on ergodinen ja sillä on yksikäsitteinen stationaarinen jakauma π , jolle $E_\pi[t(\theta)] < \infty$, niin tällöin

$$\bar{t}_n \rightarrow E_\pi[t(\theta)] \quad \text{todennäköisyydellä } 1, \quad \text{kun } n \rightarrow \infty. \quad \square \quad (7)$$

Ergodisen teoreeman nojalla Markovin ketjun arvojen keskiarvot muodostavat riippuvuudestaan huolimatta vahvasti konsistentteja estimaatteja stationaarisen jakauman arvoille.

Jatkossa tutkittavien Metropolis-Hastings-algoritmien yhteydessä tärkeitä ovat niin kutsutut kääntyvät Markovin ketjut, jotka määritellään seuraavaksi.

Määritelmä 3.6. Markovin ketju on kääntyvä, mikäli sille on voimassa

$$\pi(x)P(x, y) = \pi(y)P(y, x), \quad \text{kaikilla } x, y \in S. \quad \square \quad (8)$$

Kääntövyysehto tarkoittaa sanallisesti sitä, että kun ketju on saavuttanut stationaarisen jakauman, niin sen siirtymäintensiteetti tilasta x tilaan y on sama kuin tilasta y tilaan x .

4 MCMC-menetelmistä

Tarkastellaan ergodista Markovin ketjua $(\theta^{(n)})_{n \geq 0}$, jolla on siirtymäfunktio $P(x, y)$, alkutilan jakauma $\pi^{(0)}$ ja jonka tila-avaruudelle S on voimassa $S \subset \mathbb{R}^p$. Ketjun arvon generointi alkaa muodostamalla realisaatio satunnaismuuttujalle $\theta^{(0)}$ alkutilan jakaumasta $\pi^{(0)}$. Tämän jälkeen satunnaismuuttujan $\theta^{(1)}$ realisaatio generoidaan jakaumasta $P(\theta^{(0)}, \cdot)$. Edelleen satunnaismuuttujalle $\theta^{(2)}$ jatketaan samalla tavalla ja generoidaan realisaatio jakaumasta $P(\theta^{(1)}, \cdot)$. Samalla tavalla jatkamalla vaiheessa n generoidaan satunnaismuuttujalle $\theta^{(n)}$ realisaatio jakaumasta $P(\theta^{(n-1)}, \cdot)$ ja kun n kasvaa, niin generoidut arvot lähestyvät rajajakaumasta π generoituja arvoja. Suurilla arvoilla n generoituja realisaatioita voidaan pitää approksimaatioina jakaumasta π . Stationaarisuuden nojalla kaikki konvergenssin jälkeen generoidut realisaatiot ovat arvoja rajajakaumasta π .

Halutaan generoida otos jakaumasta $\pi(\theta)$ ($\theta \in S$) ja oletetaan, että kyetään generoimaan Markovin ketju, jonka tila-avaruus on S ja jota pystytään simuloimaan. Oletetaan vielä, että tällä Markovin ketjulla on stationaarisena jakaumana $\pi(\theta)$. Mikäli nyt simuloimme generoitua Markovin ketjua kauan, niin ketjun arvot lähestyvät arvoja rajajakaumasta $\pi(\theta)$ ja täten ketjusta simuloituja arvoja voidaan käyttää otoksena rajajakaumasta.

Yllä kuvatulla menetelmällä pystytään simuloimaan myös muitakin kuin Markovin ketjuja. Tulokset antavat menetelmän, jonka avulla pystytään generoimaan arvoja mistä tahansa jakaumasta π . Näitä menetelmiä kutsutaan yhteisellä nimellä *MCMC* (Markov Chain Monte Carlo). Seuraavaksi tutkitaan erästä MCMC-menetelmää *Metropolis-Hastings-algoritmeja*.

4.1 Metropolis-Hastings-algoritmit

Tässä pykälässä oletetaan, että tutkittavaa posteriori-jakaumaa π ei kyetä tarkasti ratkaisemaan, jolloin joudutaan turvautumaan simulointiin pohjautuviin menetelmiin. Posteriori-jakauma ratkaistaan MCMC-menetelmiin kuuluvilla nk. Metropolis-Hastings-algoritmeilla. Perusidea on muodostaa siirtymätodennäköisyysfunktio $P(\theta, \phi)$, jolla on stationaarisena jakaumana posteriori-jakauma π . Tämä on yksinkertaisinta tehdä tutkimalla kääntyviä Markovin ketjuja, jolloin P toteuttaa ehdon

$$\pi(\theta)P(\theta, \phi) = \pi(\phi)P(\phi, \theta), \quad \text{kaikilla } (\theta, \phi). \quad (9)$$

Tämä on Markovin ketjun kääntyvyysehto. Yllä oleva ehto ei ole välttämätön Markovin ketjun konvergenssille, mutta se on riittävä ehto sille, että posteriori-jakauma π on Markovin ketjun stationaarinen jakauma.

Jatkossa siirtymätodennäköisyysfunktio $P(\theta, \phi)$ koostuu kahdesta osasta, *ehdotussiirtymäfunktiosta* $q(\theta, \phi)$ sekä *hyväksymistodennäköisyydestä* $\alpha(\theta, \phi)$, joille on voimassa

$$P(\theta, \phi) = q(\theta, \phi)\alpha(\theta, \phi), \quad \text{jos } \theta \neq \phi. \quad (10)$$

Tästä seuraa, että ketju pysyy tilassa θ (positiivisella) todennäköisyydellä

$$P(\theta, \theta) = 1 - \int q(\theta, \phi)\alpha(\theta, \phi)d\phi. \quad (11)$$

Yhtälöistä (10) ja (11) saadaan yhdistämällä siirtymätodennäköisyysfunktioiksi

$$P(\theta, A) = \int_A q(\theta, \phi)\alpha(\theta, \phi)d\phi + I(\theta \in A)[1 - \int q(\theta, \phi)\alpha(\theta, \phi)d\phi], \quad (12)$$

missä A on parametriavaruuden osajoukko ja $I(\cdot)$ indikaattorifunktio. Jakamalla on siis tiheysfunktio, kun $\theta \neq \phi$ ja todennäköisyysatomi, kun $\theta = \phi$.

Hastings [5] ehdottaa hyväksymistodennäköisyyden $\alpha(\theta, \phi)$ valitsemista niin, että muodostuvasta ketjusta tulee kääntyvä. Tämän periaatteen mukainen yleisimmin käytetty hyväksymistodennäköisyyden muoto on

$$\alpha(\theta, \phi) = \min\left\{1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)}\right\}. \quad (13)$$

MCMC-menetelmiä, joissa siirtymätodennäköisyysfunktio on muotoa (12) ja hyväksymistodennäköisyys muotoa (13) kutsutaan (jatkossa) Metropolis-Hastings-algoritmeiksi. Hastings [5] kutsuu kaavassa (13) esiintyvää termiä $\frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)}$ "testisuhteeksi" (test ratio). Testisuhde voidaan kirjoittaa muotoon

$$\frac{\pi(\phi)/q(\theta, \phi)}{\pi(\theta)/q(\phi, \theta)}, \quad (14)$$

tästä huomataan, että posteriori-jakaumaan liittyvä verrannollisuuskertoimen, joka usein on ongelmallinen, kumoutuu. Näiden valintojen optimaalisuutta voidaan perustella sillä, että ne minimoivat momenttiestimaattien asymp-toottiset varianssit. Peskun [9] osoittaa diskreetissä tapauksessa hyväksymistodennäköisyyden (13) optimaalisuuden suurelle luokalle eri siirtymätodennäköisyyksiä $q(\theta, \phi)$.

Todistus sille, että $P(\theta, \phi)$ toteuttaa kääntyvyyshdon löytyy esimerkiksi lähteestä [11]. Roberts and Smith [10] osoittavat, että jos q on redusoitumaton ja jaksoton ja $\alpha(\theta, \phi) > 0$ kaikilla arvopareilla (θ, ϕ) , niin Metropolis-Hastings-algoritmi määrittelee redusoitumattoman ja jaksottoman Markovin ketjun, jolla on siirtymätodennäköisyytenä P ja rajajakaumana posteriori-jakauma π .

Esimerkki 4.1. Metropolis-Hastings-algoritmi:

1. Alustetaan iteraatiolaskuri $j \leftarrow 1$ ja muodostetaan alkutila $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$.
2. Generoidaan ketjun uusi tila ϕ jakaumasta $q(\theta^{(j-1)}, \cdot)$.
3. Määritetään siirtymisen hyväksymistodennäköisyys $\alpha(\theta^{(j-1)}, \phi)$ kaavan (13) mukaan. Mikäli siirtyminen hyväksytään, niin $\theta^{(j)} = \phi$. Mikäli siirtymistä ei hyväksytä, niin $\theta^{(j)} = \theta^{(j-1)}$ eli ketju pysyy samassa tilassa.

4. Muutetaan laskurin arvoa $j \leftarrow j + 1$ ja palataan kohtaan 2 kunnes ketju konvergoi. $\square$

Ennen algoritmin 3. kohtaa generoidaan ketjun arvoista riippumaton arvo u tasaisesta jakaumasta väliltä $[0, 1]$. Mikäli $u \leq \alpha$, niin siirtyminen uuteen tilaan hyväksytään ja jos $u > \alpha$ ketju pysyy samassa tilassa. Ehdotussiirtymäfunktio q määrittelee siis ainoastaan mahdollisen tilasiirtymän, joka hyväksytään tai hylätään hyväksymistodennäköisyyden α arvosta riippuen.

Mikäli q on symmetrinen eli toteuttaa ehdon

$$q(\theta, \phi) = q(\phi, \theta), \quad (15)$$

niin hyväksymistodennäköisyys yksinkertaistuu muotoon

$$\alpha(\theta, \phi) = \min\left\{1, \frac{\pi(\phi)}{\pi(\theta)}\right\}. \quad (16)$$

Tämän lisäehdon toteuttavia Metropolis-Hastings-algoritmeja kutsutaan *Metropolis*-algoritmeiksi. Tässä tapauksessa hyväksymistodennäköisyys riippuu siis vain posteriori-jakauman arvoista ketjun nykyisessä tilassa θ ja ehdotustilassa ϕ .

Käytännössä ketju saattaa pysyä pitkään samassa tilassa ja tällöin konvergoiminen kohti rajajaukamana olevaa stationaarista jakaumaa on hidasta.

Ketju saattaa pysyä samassa tilassa useiden iteraatioiden ajan, mikä ei ole käytännön tilanteissa suotavaa. Ketjun siirtymistä voidaan tutkia laskemalla *hyväksymissuhdeluku*

$$hs = \frac{k}{m},$$

missä k on hyväksytyjen siirtymien lukumäärä ja m iteraatioiden kokonaislukumäärä.

4.2 Erityyppisiä ehdotussiirtymäfunktioita

Tähän mennessä ehdotussiirtymäfunktiolle q ei ole asetettu erityisiä muotovaatimuksia. Tässä alitykalässä tutkitaan Metropolis-Hastings-algoritmin erikoistapauksia, joita saadaan asettamalla erilaisia ehtoja ehdotussiirtymäfunktiolle q . Mainittakoon, että yleisesti ketjun määrää sen siirtymätodennäköisyysfunktio P eikä ehdotussiirtymäfunktio q , mutta tässä alitykalässä

käsitteiden nimet viittaavat ensisijaisesti ehdotussiirtymäfunktion q ominaisuuksiin.

4.2.1 Symmetriset ketjut

Ketjun sanotaan olevan symmetrinen, jos sen siirtymätodennäköisyysfunktio on symmetrinen parametriensa suhteen eli toteuttaa ehdon $P(\theta, \phi) = P(\phi, \theta)$ kaikilla arvopareilla (θ, ϕ) .

Metropolis-Hastings-algoritmeissa symmetriavaatimus asetetaan ehdotussiirtymäfunktiolle q . Esimerkki symmetrisestä ketjusta on Metropolis-algoritmi. Jos q riippuu arvoparista (θ, ϕ) vain itseisarvon $|\theta - \phi|$ nojalla, niin on voimassa $q(\theta, \phi) = q(\phi, \theta)$. Tässä tapauksessa hyväksymistodennäköisyysfunktio sievenee muotoon $\min\{1, \pi(\phi)/\pi(\theta)\}$, eikä täten riipu ehdotussiirtymäfunktiosta q . Tämä yksinkertaistaa laskentaa ja nopeuttaa täten käytännönsovelluksia.

4.2.2 Satunnaiskulkuketjut

Jälleen kyseessä on ehdotussiirtymäfunktiolle q asetettava lisäehto. Satunnaiskulku on Markovin ketju, jonka kehityksen määrää lauseke $\theta^{(j)} = \theta^{(j-1)} + w_j$, missä w_j on ketjusta riippumaton satunnaismuuttuja. Yleisesti häiriötekijät w_j ovat riippumattomia ja samoin jakautuneita tiheysfunktionaan f_w . Ehdotussiirtymäfunktio on muotoa $q(\theta, \phi) = f_w(\phi - \theta)$. Jos f_w on symmetrinen arvon 0 ympärillä, niin ketju on symmetrinen ja kaikki yllä mainitut symmetristen ketjun ominaisuudet ovat voimassa. Tästä huomataan, että Metropolis-algoritmi on satunnaiskulkuketjujen erikoistapaus.

Suurimmassa osassa Metropolis-Hastings-algoritmien käytännönsovelluksia käytetään satunnaiskulkuketjuja. Käytetyimmät valinnat ehdotussiirtymäfunktiolle q ovat (nollaan keskittynyt) normaali jakauma sekä Studenten t -jakauma. Tällöin ehdotetut arvot jakautuvat ketjun aikaisemman arvon ympärille. Tärkeä jäljelle jäävä valinta on tiheysfunktion f_w hajonnan valinta. Suuri varianssi mahdollistaa suuret siirtymät ketjun nykyisestä tilasta, mutta samalla todennäköisesti hyväksymistodennäköisyydet ovat todella pieniä. Pieni varianssi rajoittaa ketjun siirtymistä, mutta toisaalta, jos varianssi on

pieni, niin hyväksymistodennäköisyydet ovat yleensä suuria. Lähteessä [12] Tierney ehdottaa tiheysfunktion f_w varianssimatriisille muotoa cV , missä c on viritysvakio sekä V on jokin posteriorivarianssin approksimaatio. Viritysvakion valinta riippuu siitä, mitä halutaan optimoida (suuria hyväksymistodennäköisyyksiä vai suuria siirtymiä ketjun tiloissa). Gelman, Roberts ja Gilks [3] ehdottavat normaalijakaumasatunnaiskulkutapauksessa vakiolle c arvoja väliltä [2, 3].

4.2.3 Riippumattomat ketjut

Tässä tapauksessa ehdotussiirtymäfunktio q ei riipu ketjun nykyisestä tilasta eli q on siis muotoa $q(\theta, \phi) = f(\phi)$. Ensisilmäyksellä vaikuttaa siltä, että riippumattomuus aiemmasta tilasta on vastoin ketjun Markovin ominaisuutta. Tulee kuitenkin muistaa, että q on vain ehdotussiirtymäfunktio, joka yhdessä hyväksymistodennäköisyyden α kanssa määrää varsinaisen siirtymätodennäköisyysfunktion P . Täten P riippuu edelleen ketjun aiemmasta tilasta ja ketjulle on voimassa Markovin ominaisuus.

Käytetään muotoa (14) testisuurelle, jolloin se yksinkertaistuu muotoon $w(\phi)/w(\theta)$, missä $w = \pi/f$ on painofunktio. Yksi suosittu valinta funktiolle f on priorijakauman tiheysfunktio. Tässä tapauksessa $w = L$ eli painoksi tulee uskottavuusfunktion arvo ja hyväksymistodennäköisyys saa muodon $\min\{1, L(\phi)/L(\theta)\}$. Riippumattomat ketjut ovat laskennallisesti tehokkaita, sillä ne tuottavat yhden yksinkertaisimmista muodoista hyväksymistodennäköisyydelle α . Suurin ongelma riippumattomien ketjujen osalta syntyy, jos priorijakauma ja uskottavuusinformaatio ovat ristiriidassa keskenään. Menetelmä valitsee tällöin suurella todennäköisyydellä priorijakauman tukemia arvoja, jotka eivät luultavasti ole posteriorijakauman mukaisia.

Uskottavuusinformaation mukaanottaminen ehdotussiirtymäfunktioon auttaa yllä kuvattuun ongelmaan. Bennet, Racine-Poon ja Wakefield [1] muodostavat normaalijakaumanmukaisia riippumattomia ehdotussiirtymiä käyttämällä normaaliapproksimaatioita uskottavuusfunktioon.

Yleisesti riippumattomien ketjujen tapauksessa tulee välttää painojen suuria hajontaja, koska tämä voi johtaa ketjun pysymiseen kauan tilassa,

jolla on suuri paino. Tämän takia suositellaan lähes vakiopainot w antavan funktion f valintaa, joka tapauksessa funktion f tulisi ainakin olla rajoitettu. Sekä funktio f että posteriori-jakauma π ovat tiheysfunktioita. Tämän takia on luonnollista olettaa, että tiheysfunktioiden f ja π tulisi olla niin samantyyppisiä kuin mahdollista. Tällöin priorijakaumaa ei voida käyttää, jos se eroa merkittävästi uskottavuusfunktioista.

Tierney [12] ehdottaa kapeahäntäisten tiheysfunktioiden f , kuten normaalijakauman tiheysfunktion, välttämistä. Tällöin tiheysfunktion f häntä ei vaikuta niin paljon painofunktion arvoihin, jolloin painofunktion varianssi on pienempi.

4.3 Komponenteittaiset Metropolis-Hastings-algoritmit

Aikaisemmat tarkastelut ovat rajoittuneet yhden parametrin simulointitilanteisiin. (Oikeastaan useamman parametrin simulointiin myös, mutta olettaen yhteistiheysfunktioit tunnetuiksi, mikä ei käytännössä yleensä ole mahdollista). Oletetaan nyt, että tarkasteltavia parametreja on d kappaletta $\theta = (\theta_1, \dots, \theta_d)^T$. Tutkitaan näiden parametrien simulointia yksi kerrallaan. Tämä voidaan tehdä kahdella tavalla:

a) Jokaisella iteraatiokierroksella valitaan satunnaisesti yksi parametri päivitettäväksi.

b) Jokaisella iteraatiokierroksella päivitetään yksi parametri ja päivittäjäjärjestys on päätetty etukäteen. Esimerkiksi päivitetään järjestyksessä $1 \rightarrow 2 \rightarrow \dots \rightarrow d$.

Yllä olevat ovat esimerkkejä sekoitus- (a) sekä sykliisiirtymistä (b).

Tapauksessa a) määritellään siirtymäfunktiot P_m , joilla on yhteinen stationaarinen jakauma π ja todennäköisyydet tai painot w_m , $m = 1, \dots, r$, jotka toteuttavat ehdot $w_m \geq 0$ ja $\sum_{m=1}^r w_m = 1$. Sekoitussiirtymäfunktio P määräytyy summana $P = \sum_{m=1}^r w_m P_m$. Tapaus a) on tämän erikoistapaus, jossa $r = d$, $w_m = 1/d$ ja jokainen siirtymäfunktio P_m liikuttaa vain vastaavaa parametrivektorin θ parametria θ_m , $m = 1, \dots, d$.

Siirtymäfunktioiden P_m ominaisuudet siirtyvät sekoitussiirtymäfunktiolle P . Ensinnäkin sekoitussiirtymäfunktio määrittelee Markovin ketjun siirtymä-

funktion, jolla on stationaarisena jakaumana π . Lisäksi jos jokin siirtymä-funktioista P_m määrittelee redusoitumattoman ja jaksottoman Markovin ketjun, niin myös P määrittelee tällöin redusoitumattoman ja jaksottoman Markovin ketjun.

Merkitään syklisten siirtymien tapauksessa siirtymäfunktioita P_c , $c = 1, \dots, r$. Tällöin ketjun siirtymä tapahtuu vasta, kun kaikkia komponentteja on päivitetty. Tutkittaessa siirtymää tilasta θ tilaan ϕ , Markovin ketjun yksittäisen syklisiirtymän aikana, tulee käydä läpi kaikki mahdolliset välitilasiirtymät ψ_c joihin siirtymät P_c voivat johtaa $c = 1, \dots, r - 1$, joille lopulta siirtymä P_r johtaa tilaan $\psi_r = \phi$. Määrittemällä alkutilaksi $\psi_0 = \theta$ saadaan siirtymäfunktio

$$P(\theta, \phi) = \int \dots \int \prod_{c=1}^r P_c(\psi_{c-1}, \psi_c) d\psi_1 \dots d\psi_{r-1}. \quad (17)$$

Tapaus b) on tämän erikoistapaus, jossa $r = d$ ja jokainen siirtymäfunktio P_c liikuttaa vain vastaavaa parametrivektorin θ parametria θ_c , $c = 1, \dots, d$.

Useat siirtymäfunktioiden P_c ominaisuuksista siirtyvät syklifunktiolle P . Ensinnäkin P määrittelee Markovin ketjun siirtymätodennäköisyysfunktion, jolla on stationaarisena jakaumana π . Sen sijaan, toisin kuin sekoitus-siirtymien tilanteessa, nyt yhden siirtymäfunktion P_c redusoitumattomuudesta ja jaksottomuudesta ei suoraan seuraa siirtymäfunktion P redusoitumattomuus ja jaksottomuus. Kuitenkin, jos kaikki siirtymäfunktiot P_c ovat redusoitumattomia ja jaksottomia, niin myös P on redusoitumaton ja jaksoton [12].

Siirtymäfunktiot voidaan nyt yhdistää Metropolis-Hastings-algoritmiin. Jokainen siirtymäfunktio P_i voidaan kirjoittaa ehdotussiirtymäfunktionsa q_i sekä hyväksymistodennäköisyytensä α_i avulla muodossa $P_i = q_i \alpha_i$. Tutkitaan nyt parametrikohdaisia syklisiirtymiä eli ehdotussiirtymäfunktio $q_i(\theta, \phi)$ ehdottaa parametrivektorin θ i . parametrin tilasiirtymää, $i = 1, \dots, r$.

Luvun 2 yhtälön (4) mukaan on voimassa

$$\pi(\theta) = \pi_i(\theta_i) \pi(\theta_{-i}),$$

missä π_i on parametrin θ_i täysin ehdollistettu tiheysfunktio. Ehdotussiirtymän q_i ehdottoma siirtymä liikuttaa ainoastaan parametria θ_i ja täten on

voimassa

$$\theta_{-i} = \phi_{-i}. \quad (18)$$

Nyt yhtälöiden (4) ja (18) perusteella saadaan

$$\pi(\phi)/\pi(\theta) = \pi_i(\phi_i)\pi(\phi_{-i})/\pi_i(\theta_i)\pi(\theta_{-i}) = \pi_i(\phi_i)/\pi_i(\theta_i).$$

Edelleen, koska siirtymäfunktio P_i liikuttaa vain i . parametria, niin se määrittelee redusoituvan Markovin ketjun. Ehdotussiirtymäfunktio q_i voidaan täten kirjoittaa muodossa $q_i(\theta_i, \phi_i)$, vaikka se saattaa hyvinkin riippua muiden parametrien arvoista. Erityisesti huomataan, että hyväksymistodennäköisyys voidaan nyt kirjoittaa muodossa

$$\alpha_i(\theta_i, \phi_i) = \min\{1, \frac{\pi_i(\phi_i)q_i(\phi_i, \theta_i)}{\pi_i(\theta_i)q_i(\theta_i, \phi_i)}\} = \min\{1, \frac{\pi(\phi)q_i(\phi_i, \theta_i)}{\pi(\theta)q_i(\theta_i, \phi_i)}\}. \quad (19)$$

Jokainen siirtymäfunktio P_i määrittelee kääntyvän Markovin ketjun, jolla on stationaarisen jakaumana $\pi_i(\theta_i)$, $i = 1, \dots, d$.

Tutkitaan yksinkertaisuuden vuoksi kahden parametrin tilannetta $\theta = (\theta_1, \theta_2)$. Siirtyminen tilasta $\theta = (\theta_1, \theta_2)$ tilaan $\phi = (\phi_1, \phi_2)$ tapahtuu kun kumpikin parametri siirtyy oman siirtymäfunktionsa mukaan. Jos π on tämän syklisiirtymäfunktion stationaarinen jakauma, niin sen tulee toteuttaa ehto

$$\pi(\phi) = \int \int \pi(\theta)P_1(\theta_1, \phi_1)P_2(\theta_2, \phi_2)d\theta. \quad (20)$$

Kaavan (20) oikellisuus nähdään tutkimalla sen oikeaa puolta, joka voidaan kirjoittaa muotoon

$$\begin{aligned} & \int \int \pi(\theta_1|\theta_2)\pi(\theta_2)P_1(\theta_1, \phi_1)P_2(\theta_2, \phi_2)d\theta_1d\theta_2 \\ &= \int \pi(\theta_2)[\int \pi(\theta_1|\theta_2)P_1(\theta_1, \phi_1)d\theta_1]P_2(\theta_2, \phi_2)d\theta_2 \\ &= \int \pi(\theta_2)\pi(\phi_1|\theta_2)P_2(\theta_2, \phi_2)d\theta_2 \\ &= \int \pi(\phi_1)\pi(\theta_2|\phi_1)P_2(\theta_2, \phi_2)d\theta_2 \end{aligned}$$

$$\begin{aligned}
&= \pi(\phi_1) \int \pi(\theta_2|\phi_1) P_2(\theta_2, \phi_2) d\theta_2 \\
&= \pi(\phi_1) \pi(\phi_2|\phi_1) \\
&= \pi(\phi).
\end{aligned}$$

Yllä toinen ja viides yhtäsuuruus seuraavat ehdollisten tiheysfunktioiden stationaarisuudesta ja loput yhtäsuuruudet saadaan perustodennäköisyyslaskujen sekä integrointien avulla. Tulos voidaan yleistää matemaattisella induktiolla koskemaan useampidimensioisia tapauksia. Yllä olevassa on oletettu siirtymätodennäköisyysfunktioiden olevan jatkuvia. Tulokset voidaan pienin teknisin korjauksin saattaa voimaan myös sekoitus siirtymien tapauksessa. Voidaan myös osoittaa, että vaikka siirtymäfunktiot P_i olisivat redusoituvia, niin P on redusoitumaton ja jaksoton [12]. Täten rajajaukauma π on yksikäsitteinen.

Tämän nojalla saadaan uusi komponentteittainen versio aiemmin esitetystä Metropolis-Hasting-algoritmista.

Esimerkki 4.2. *Komponentteittainen Metropolis-Hastings-algoritmi:*

1. Alustetaan iteraatiolaskuri $j \leftarrow 1$ ja muodostetaan alkutila $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$.
2. Alustetaan komponenttilaskuri $i \leftarrow 1$.
3. Generoidaan i . parametrin uusi tila ϕ_i jakaumasta $q_i(\theta_i^{(j-1)}, \phi_i)$.
4. Määritetään siirtymisen hyväksymistodennäköisyys $\alpha_i(\theta_i^{(j-1)}, \phi_i)$ kaavan (19) mukaan. Mikäli siirtyminen hyväksytään, niin $\theta_i^{(j)} = \phi_i$. Mikäli siirtymistä ei hyväksytä, niin $\theta_i^{(j)} = \theta_i^{(j-1)}$ eli ketju pysyy samassa tilassa.
5. Muutetaan laskurin arvoa $i \leftarrow i + 1$ ja palataan kohtaan 3 kunnes $i = d$. Kun $i = d$ mennään kohtaan 6.
6. Muutetaan laskurin arvoa $j \leftarrow j + 1$ ja palataan kohtaan 2 kunnes ketju konvergoi. $\square$

Ennen algoritmin 4. kohtaa generoidaan ketjun arvoista riippumaton arvo u tasaisesta jakaumasta väliltä $[0, 1]$. Mikäli $u \leq \alpha_i$, niin siirtyminen uuteen tilaan hyväksytään ja jos $u > \alpha_i$ ketju pysyy samassa tilassa. Jälleen siirtymäfunktio q_i määrittelee siis ainoastaan mahdollisen tilasiirtymän, joka hyväksytään tai hylätään hyväksymistodennäköisyyden α_i arvosta riippuen.

4.4 Konvergenssitarkastelut

MCMC-menetelmien tapauksessa algoritmien konvergenssitarkastelut eivät ole aivan suoraviivaisia. Markovin ketju saavuttaa rajajakaumana olevan stationaarisen jakaumansa mahdollisesti vasta tuhansien iteraatiokierrosten jälkeen. Lisäksi on muistettava, että algoritmi ei suppene kohti tarkasteltavan parametrin piste-estimaattia vaan parametrin posteriori-jakaumaa. Siihen, kuinka monen iteraatiokierroksen jälkeen algoritmin tuottamia arvoja voidaan pitää arvoina posteriori-jakaumasta, ei ole olemassa yksiselitteistä vastausta.

Konvergenssitarkastelua varten on olemassa kaksi päävaihtoehtoa. Toinen on teoreettinen ja siinä pyritään mittaamaan kuinka kaukana algoritmin tuottamat arvot ovat stationaarisesta jakaumasta. Teoreettisessa lähestymistavassa tutkitaan *kokonaisvariaationormia* (total variation norm)

$$\|P(\theta^{(t)}, \theta) - \pi(\theta)\| = \frac{1}{2} \sup_{A \subset S} \int_A |P(\theta^{(t)}, \theta) - \pi(\theta)| d\theta. \quad (21)$$

Täten, jos $\|P(\theta^{(t)}, \theta) - \pi(\theta)\|$ lähestyy arvoa 0, kun t lähestyy ääretöntä, niin $\theta^{(t)}$ konvergoi kohti satunnaismuuttujaa jakaumasta $\pi(\theta)$.

Teoreettisen lähestymistavan hyöty on siinä, että sen avulla pystytään vakuuttumaan todellisesta konvergoimisesta (ainakin teoriassa). Teoreettisen lähestymistavan mukaisia konvergenssituloksia on kuitenkin vaikeaa tutkia käytännössä.

Toinen lähestymistapa konvergenssiongelmaan on käytännönläheisempi tilastollinen tapa. Tässä tavassa tutkitaan tilastollisesti algoritmin tuottamia tuloksia. Tilastollinen tapa on selkeästi käytännönläheisempi ja siten helpompi toteuttaa kuin teoreettinen. Tilastollisen tavan ongelmana verrattuna teoreettiseen on se, että tutkimalla pelkästään tilastollisesti algoritmin tuottamia tuloksia ei voida koskaan täysin vakuuttua algoritmin konvergoimisesta rajajakaumaan.

4.4.1 Teoreettinen tapa

Tähän mennessä menetelmien konvergenssiasteesta ei ole sanottu vielä mitään. Yleisesti ottaen Markovin ketjut joko ovat stationaarisessa jakaumassaan tai

eivät ole. Ergodinen teoreemakin (katso lause 3.5) kertoo vain konvergoimisesta rajatuloksena, joten siitä ei ole suoraan hyötyä käytännönsovelluksissa, joissa iteraatiokertojen lukumäärät väistämättä ovat äärellisiä.

Ergodisen teoreeman eräs parannus on niin kutsuttu *heikko ergodisuuslause*. Tätä varten olkoon $T(A)$ ensimmäinen hetki jolloin Markovin ketju käy tila-avaruuden S osajoukossa A toisin sanoen $T(A) = \inf\{t \geq 1 | \theta^{(t)} \in A\}$. Oletetaan, että Markovin ketjulla on stationaarisena jakaumana $\pi(\theta)$ ja että $\pi(A) > 0$, kaikilla tila-avaruuden S osajoukoilla A . Tällöin kyseinen Markovin ketju on *toisen asteen ergodinen*, jos

$$\int_A \pi(d\theta) E_\pi[T(A)^2] < \infty. \quad [7] \quad (22)$$

Toisin sanoen ehto toisen asteen ergodisuudelle on, että ensimmäisten hetkien, jolloin Markovin ketju käy tila-avaruuden osajoukossa A , toisen momentin tulee olla äärellinen.

Jos Markovin ketju on toisen asteen ergodinen ja $P(\theta^{(t)}, \theta)$ sekä $\pi(\theta)$ ovat tarpeeksi säännöllisiä, niin tällöin Markovin ketjun konvergenssiaste kohti stationaarista jakaumaa on luokkaa n^{-2} eli toisin sanoen on voimassa

$$\lim_{n \rightarrow \infty} n^2 \cdot \|P(\theta^{(n)}, \theta) - \pi(\theta)\| = 0. \quad (\text{Heikko ergodisuuslause})$$

Tämä on sinänsä hyödyllinen tulos, mutta valitettavasti käytännön ongelmien yhteydessä on usein hankalaa muodostaa toisen asteen ergodisia Markovin ketjuja.

Eräs hyödyllisempi ja voimakkaampi konvergenssitulos on niin kutsuttu *geometrisen ergodisuus*. Tätä varten tehdään samat oletukset kuin yllä, mutta korvataan yllä ensimmäisten hetkien, jolloin Markovin ketju käy tila-avaruuden S osajoukossa A , toisia momenteja koskenut ehto ehdolla

$$\|P(\theta^{(t)}, \theta) - \pi(\theta)\| \leq m(\theta)\rho^t, \quad \forall \theta, \quad 0 < \rho < 1, \quad (23)$$

missä $m(\theta)$ on äärellinen ja ei-negatiivinen funktio.

Näiden oletusten ollessa voimassa Markovin ketju konvergoi kohti stationaarista jakaumaa geometrisella konvergenssiasteella, joka voi olla todella nopea riippuen arvosta ρ .

Jos funktion $m(\theta)$ sijaan löydetään vakio m , jolle

$$\|P(\theta^{(t)}, \theta) - \pi(\theta)\| \leq m\rho^t, \quad \forall \theta, \quad 0 < \rho < 1,$$

niin Markovin ketju on *tasaisen ergodinen* ja tällöin ketju konvergoi vielä nopeammin.

Geometrisesta ja tasaisesta ergodisuudesta on kaksi hyötyä. Ensinnäkin näissä tapauksissa tiedetään, että Markovin ketju konvergoi käytännön simuloitien kannalta järkevässä ajassa stationaariseen jakaumaan (toki konvergenssin vaatima aika riippuu mallin kompleksisuudesta sekä datan rakenteesta). Toiseksi geometrisen ja tasaisen ergodisuuden tapauksissa on mahdollista määrätä rajat Markovin ketjun konvergoimiseen vaadittaville iteraatiokierrosten lukumäärille.

Esimerkki 4.3. Riippumattomien ketjujen tapauksessa (katso alipykälä 4.2.3) jos painofunktio, $w(\theta) = \pi(\theta)/f(\theta)$, on rajoitettu ja lisäksi $\pi(\theta)$ rajoitettu, niin ettei se voi saada arvoa 0, on Markovin ketju tasaisen ergodinen ja erityisesti $\rho \leq 1 - \sup(w(\theta))^{-1}$ (lähde [12]). $\square$

4.4.2 Tilastollinen tapa

Markovin ketjun konvergoimista voidaan tutkia myös algoritmin tuottamien tulosteiden pohjalta. Tilastollisessa tavassa tutkitaan ketjun tuottamia arvoja, joten tämä lähestymistapa on selkeästi empiirinen ja siten huomattavasti käytännönläheisempi kuin yllä tutkittu teoreettinen tapa. Tilastollisen konvergenssitarkastelun yhteydessä ei kuitenkaan koskaan voida olla varmoja siitä, että Markovin ketju on todella saavuttanut stationaarisen jakaumansa.

Konvergenssia voidaan tutkia graafisesti muodostamalla ketjun arvoista sopivia tulosteita. Oletetaan nyt, että käytössä on n rinnakkaista ketjua, joissa jokaisessa suoritetaan m iteraatiota. Menettelemällä tällä tavalla voidaan mistä tahansa parametrin θ funktiosta ψ piirtää histogrammeja iteraatioiden lukumäärän m funktiona. Tätä menetelmää toistetaan k iteraation välein, missä arvon k ei tarvitse olla suuri, jos on näyttöä konvergenssista m iteraation jälkeen. Käytetty arvo k ei saa kuitenkaan olla liian pieni, sillä tällöin ketjun korrelaattorakenne vaikuttaa menetelmän tuloksiin. Tyypillisesti

järkevä arvo on välillä $10 < k < 50$ [2]. Konvergenssi hyväksytään, mikäli histogrammeja ei voida erottaa toisistaan. Vastaavaa ideaa voidaan käyttää myös yhden ketjun tapauksissa, jolloin konvergenssi hyväksytään mikäli ketju alkaa toistaa kvalitatiivisesti samankaltaista rataa. Vielä eräs tilastollinen mahdollisuus konvergenssin tarkasteluun on ergodisten keskiarvojen (katso määritelmä 3.4) käyrän asettuminen tietyn vakioarvon lähelle.

Tilastollisten menetelmien käytössä tulee muistaa menetelmien empiirisyys. Markovin ketju saattaa tulosteiden valossa näyttää konvergoivan, mutta muutettaessa tulostusasteikon skaalausta tulkinta voi muuttua päinvastaiseksi. Useiden ketjujen tapauksessa saattaa näyttää siltä kuin ketju olisi konvergoinut stationaariseen jakaumaan, vaikka käytännössä näin ei olisikaan tapahtunut.

Tilastollisia konvergenssitarkasteluja varten määritellään vielä niin kutsuttu *burn in* -arvo.

Määritelmä 4.4. Metropolis-Hastings-algoritmin burn in -arvo on ensimmäinen iteraatiokierros, jolla algoritmin tuottamia arvoja pidetään havaintona stationaarisesta jakaumasta. $\square$

4.5 Metropolis-Hastings-algoritmin ohjelmoimisesta

Avain Metropolis-Hastings-algoritmin ohjelmoimiseen on oikeanlaisen otantastrategian muodostamisessa. Otantastrategian tulee olla sellainen, että Markovin ketjulla on sama todennäköisyys siirtyä tilasta x tilaan y kuin tilasta y tilaan x (vertaa määritelmä 3.6). Tämä otantastrategia saavutetaan muodostamalla sopivat ehdotussiirtymäfunktio $q(x, y)$ sekä hyväksymistodennäköisyys $\alpha(x, y)$.

Esimerkki 4.5. Sopivia muotoja ehdotussiirtymäfunktiolle ovat esimerkiksi symmetrinen tasainen jakauma, jolloin $y|x \sim U(x - \epsilon, x + \epsilon)$ tai ketjun nykyiseen arvoon keskittynyt normaalijakauma, jolloin $y|x \sim N(x, \epsilon)$. Avoimeksi tässä vaiheessa jää kysymys siitä miten arvo ϵ tulisi määrätä. $\square$

MCMC-menetelmien ohjelmoimisessa kannattaa kiinnittää huomiota tietokoneen aktiivisen muistin käyttämiseen. MCMC-simuloinnissa pyritään muodostamaan otos satunnaismuuttujien $x^{(1)}, x^{(2)}, \dots, x^{(N)}$ arvoista, mikä saattaa

huokutella käyttämään taulukkorakennetta $x[1 \dots N]$. MCMC-menetelmien hyöty on kuitenkin siinä, että iteraatiokierroksella t riittää tietää nykyinen parametrijoukko $x^{(t)}$ iteraatiokierroksen $t - 1$ parametrijoukon $x^{(t-1)}$ funktiona (vertaa Markovin ominaisuus). Täten tulostamalla jokaisella iteraatiokierroksella saadut parametrien arvot voidaan vähentää tietokoneen aktiivisen muistin käyttöä, sillä tällöin iteraatiokierroksella t riittää pitää aktiivisessa muistissa vain parametrijoukko $x^{(t)}$ ja ehdotussiirtymäfunktion ehdottama parametrijoukko.

5 Käytännön sovellus

Käytännön sovelluksen aineistona on tutkielman tekijän opinnäytetyössä [11] käytetty aineisto. Kokonaisuudessaan aineisto käsittää Vahinkovakuutusyhtiö Pohjolan päivittäisiä liikennevahinkojen lukumääriä. Vahinkoaineisto ulottuu vuoden 1994 tammikuusta vuoden 2004 lokakuuhun. Vahinkoaineisto on rajattu käsittämään pääkaupunkiseutu (Helsinki, Espoo sekä Vantaa). Vahinkoaineistoon on lisätty Ilmatieteen laitokselta saadut vastaavan ajanjakson säähavainnot Helsinki-Vantaan lentoaseman säähavaintopisteestä.

Metropolis-Hastings-algoritmin ominaisuuksien takia aineistoa on rajattu nyt siten, että tarkasteluun tulevat mukaan vain tapaukset, joissa vakuutetun kotipaikkakuntana on ollut Helsinki ja vahinko on sattunut aikavälillä 1. marraskuuta 2003 - 31. lokakuuta 2004.

Seuraavaksi esitetään liikennevakuutuksen vahinkojen lukumäärien arviointiin opinnäytetyössä tilastollisesti merkitseviksi todettujen selittävien muuttujien parametrien estimointi aikaisemmin esiteltyjen Metropolis-Hastings-algoritmien avulla.

5.1 Tarkasteltavat mallit

Seuraavassa muodostetaan vahinkojen lukumäärille hierarkkinen yleistetty lineaarinen malli, jossa vasteen eli vahinkojen lukumäärän oletetaan noudattavan Poisson-jakaumaa sekä linkkifunktiona käytetään logaritmifunktiota. Malli poikkeaa yleistetyistä lineaarisista malleista siinä, että nyt huomioidaan

parametrivektorin β priorijakauma $p(\beta)$.

Tutkittavat mallit ovat muotoa

$$\mathbf{Y}_i | \lambda_i \sim Po(\lambda_i)$$

$$\log(\lambda_i) = \eta_i = \mathbf{x}_i^T \beta$$

$$\beta | \mu, \sigma^2 \sim p(\beta),$$

missä parametrivektorin β priorijakaumaa ei ole vielä kiinnitetty.

Yksittäisen havainnon tiheysfunktio on tässä tapauksessa

$$f(y_i | \lambda_i) \propto \lambda_i^{y_i} e^{-\lambda_i} = e^{\eta_i y_i} e^{-e^{\eta_i}},$$

josta saadaan uskottavuusfunktioksi

$$L(\beta, \phi) = f(\mathbf{y} | \beta) \propto \prod_{i=1}^n e^{\mathbf{x}_i^T \beta y_i} e^{-e^{\mathbf{x}_i^T \beta}}. \quad (24)$$

Tämän nojalla tarkasteltavat posteriorijakaumat ovat muotoa

$$\pi(\beta) = L(\beta) p(\beta)$$

$$\propto \left(\prod_{i=1}^n e^{\mathbf{x}_i^T \beta y_i} e^{-e^{\mathbf{x}_i^T \beta}} \right) \cdot p(\beta).$$

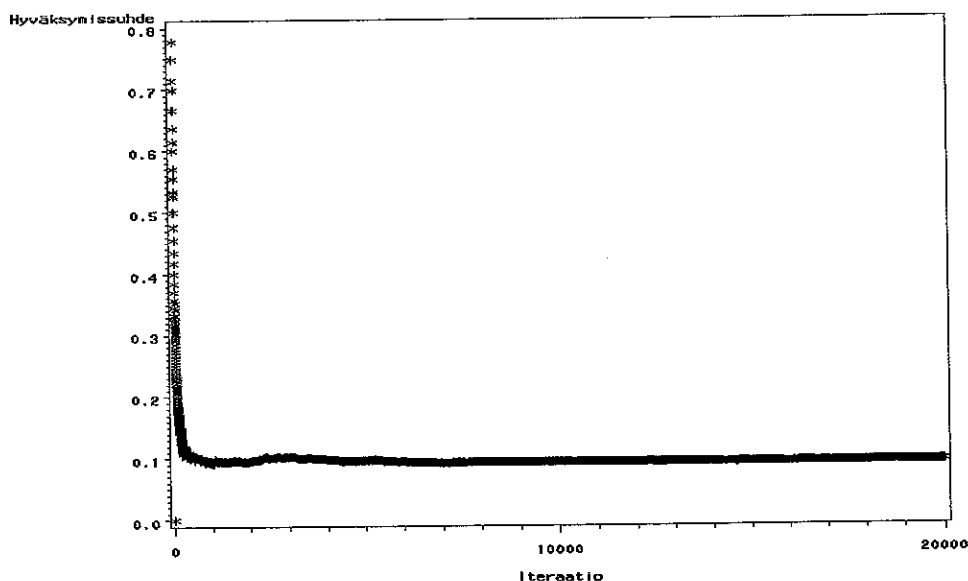
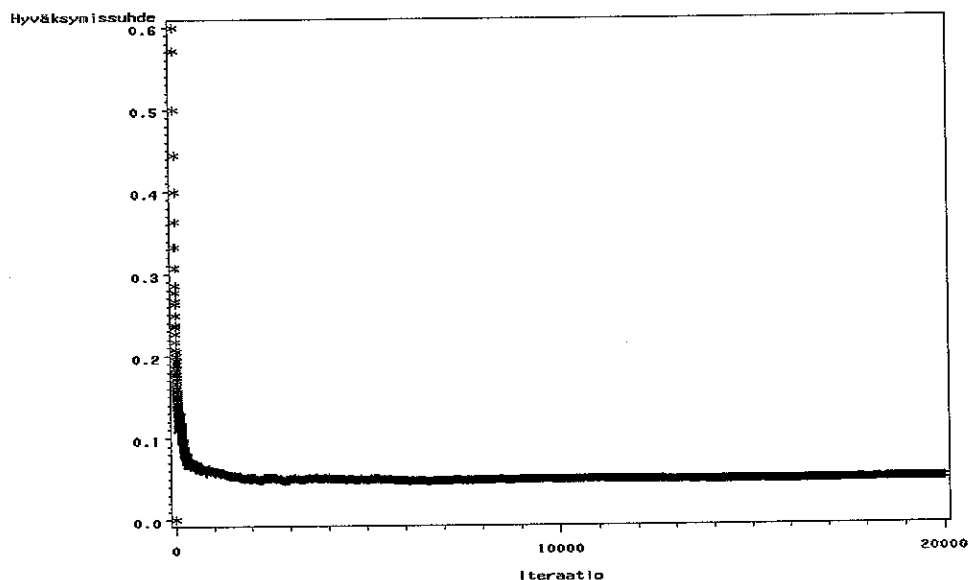
Tarkasteltavat mallit eivät yleisesti ratkea analyttisesti, vaan parametrien estimointi suoritetaan luvussa 4 esiteltujen Metropolis-Hastings -algoritmien avulla.

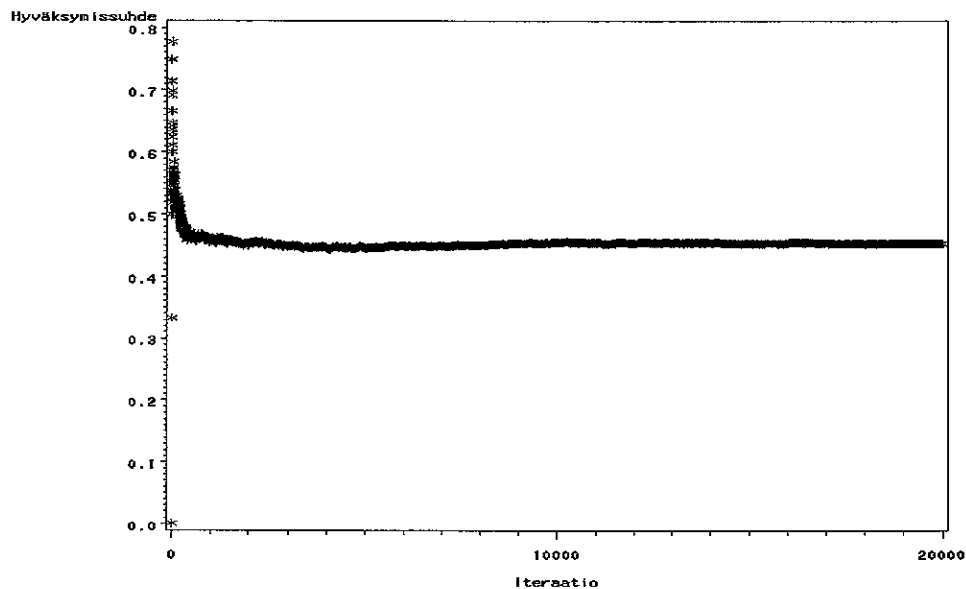
5.2 Yhden parametrin tilanne

Tutkitaan tutkielman tekijän opinnäytetyössä tilastollisesti merkitseväksi todetun vakuutusvuosien lukumäärän parametrin estimointia. Vakuutusvuosien lukumäärän parametrin priorijakaumaksi on valittu normaalijakauma $N(\mu, \sigma^2)$, missä hyperparametreiksi on valittu opinnäytetyön yleistetyn lineaarisen mallin mukaiset arvot $\mu = 0,7528$ (vakuutusvuosien lukumäärän parametrin estimaatti) sekä $\sigma^2 = 0,0069 * 0,0069$ (vakuutusvuosien lukumäärän parametrin estimaatin keskivirheen neliö). Ehdotussiirtymäfunktiona q käytetään tasaista jakaumaa $U(\beta - \epsilon, \beta + \epsilon)$. Iteraatioita suoritetaan

20 000 ja vain joka 25. havainto otetaan mukaan lopulliseen otokseen (vertaa alipykälä 4.4.2).

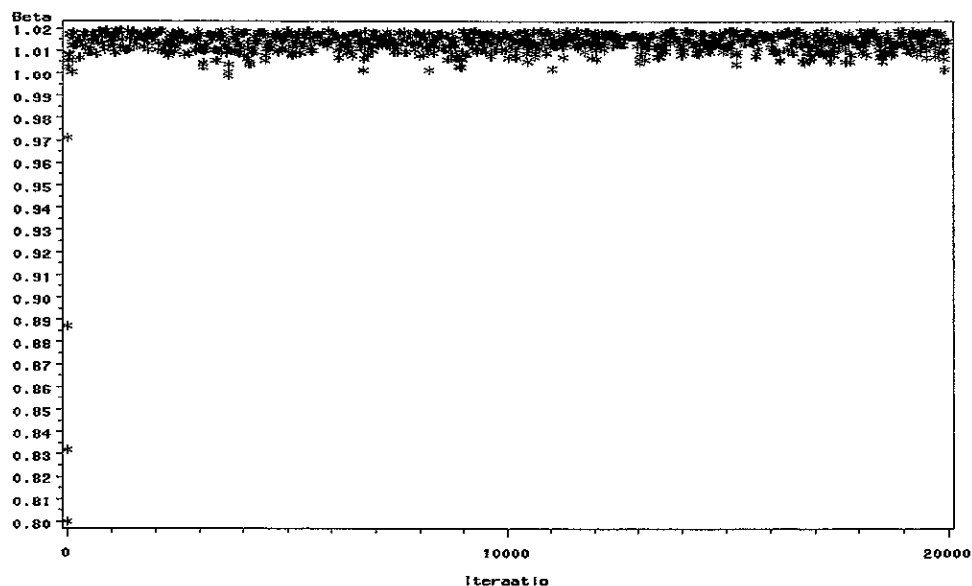
Seuraavista kuvista käy ilmi arvon ϵ vaikutus hyväksymissuhdelukuun. Kuva 5.1. Arvon ϵ vaikutus hyväksymissuhdelukuun. Ensimmäisessä kuvassa $\epsilon = 0,1$, toisessa $\epsilon = 0,05$ sekä kolmannessa $\epsilon = 0,01$. Hyväksymissuhdeluku paranee huomattavasti kun ϵ pienennetään arvosta 0,05 arvoon 0,01.

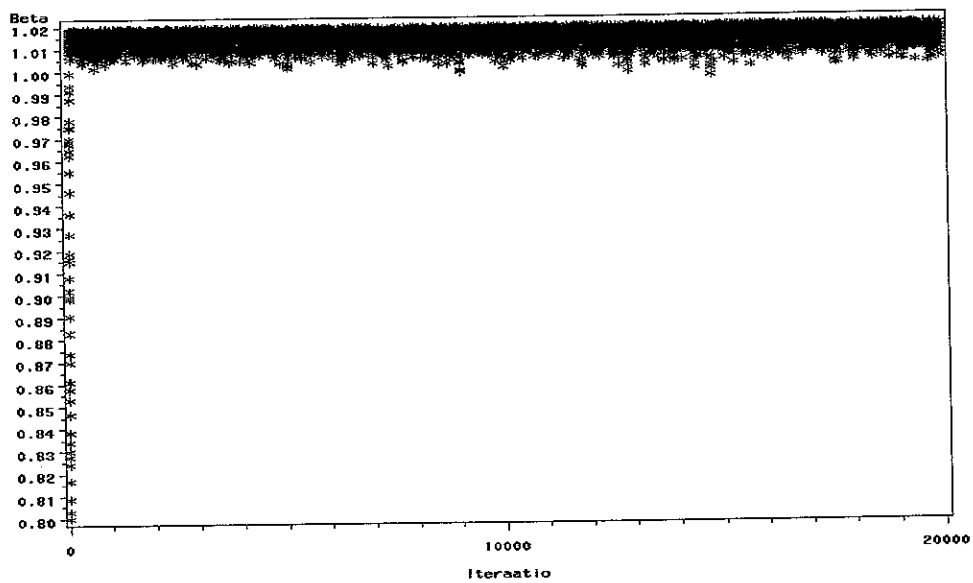
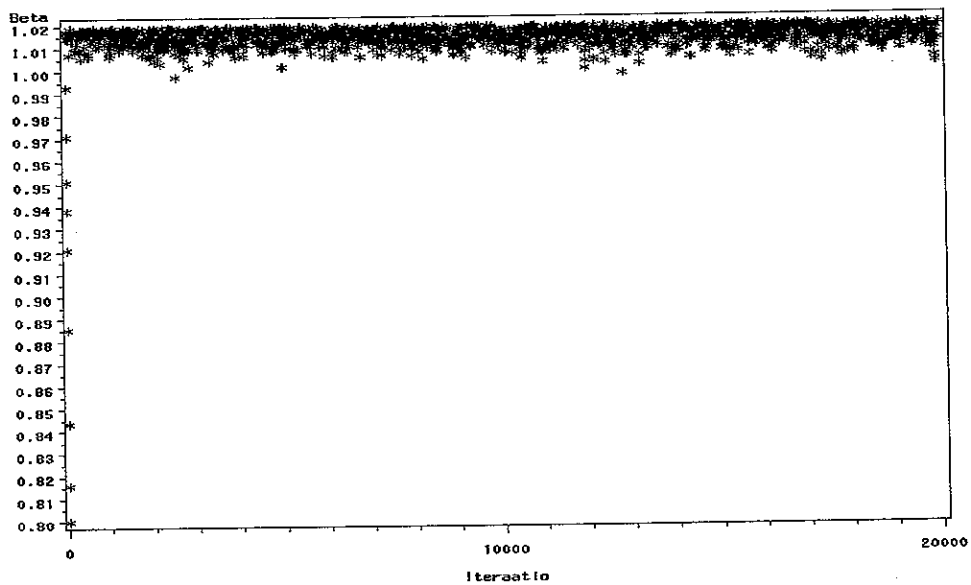




Tutkitaan arvon ϵ vaikutusta algoritmin konvergoimiseen simuloitujen parametrien arvojen käyrien nojalla.

Kuva 5.2. Arvon ϵ vaikutus algoritmin konvergoimiseen. Ensimmäisessä kuvassa $\epsilon = 0,1$, toisessa $\epsilon = 0,05$ sekä kolmannessa $\epsilon = 0,01$. Kuvien perusteella algoritmi hakeutuu sitä nopeammin stationaarisen jakauman lähelle mitä suurempi arvo ϵ on.

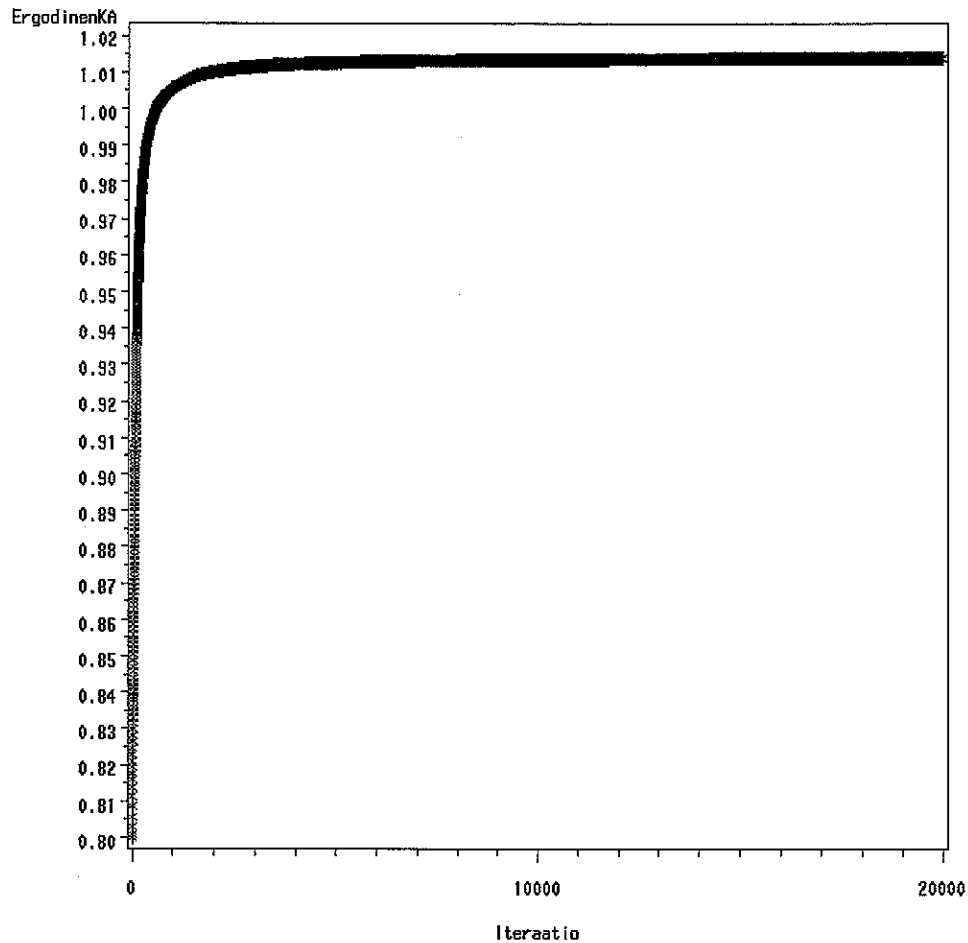




Kuvan 5.2 perusteella algoritmi näyttäisi löytävän melko nopeasti stationaarisen jakauman.

Algoritmin konvergoimisesta vakuuttumisen vuoksi tutkitaan vielä ergodisen keskiarvon käyrää.

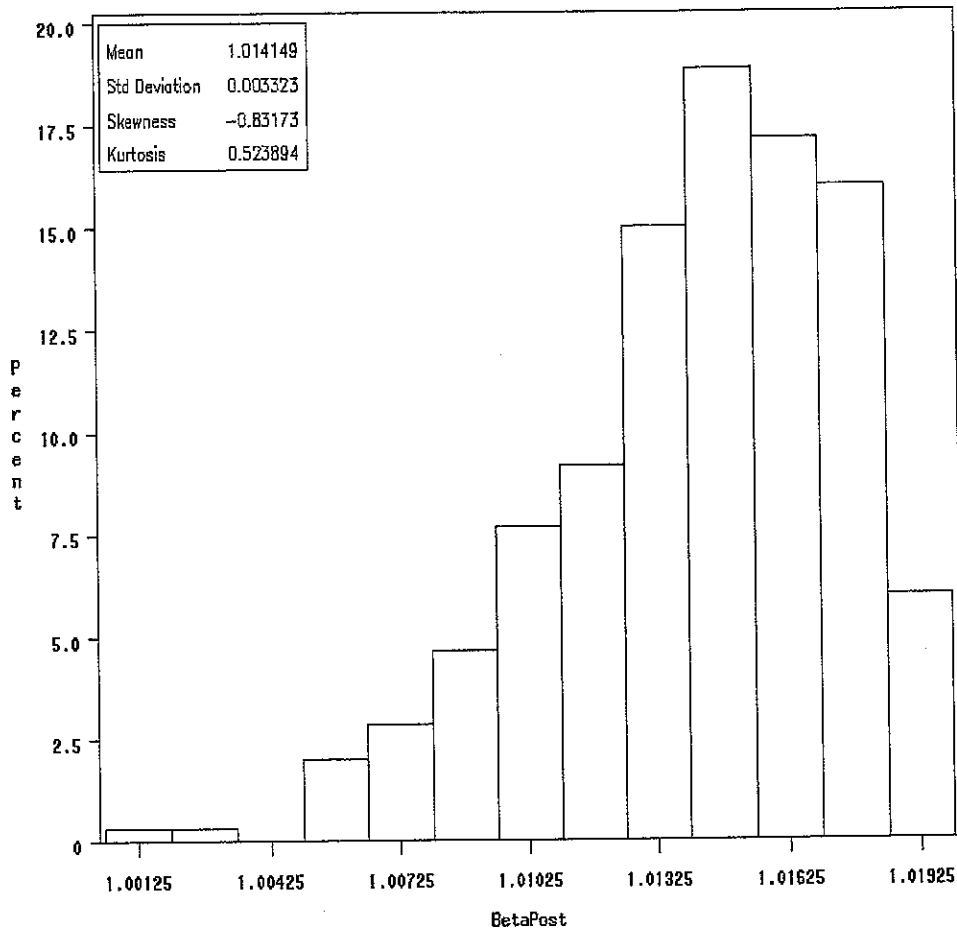
Kuva 5.3. Ergodisen keskiarvon käyrä arvolla $\epsilon = 0,01$ (funktio t on identiteettifunktio).



Ergodinen keskiarvo hakeutuu vakioarvon lähelle, mikä on merkki algoritmin konvergoimisesta.

Seuraavaksi esitetään burn in -arvolla 5 000 sekä arvolla $\epsilon = 0,01$ saatu vakuutusvuosien lukumäärän parametrin posteriori-jakauma.

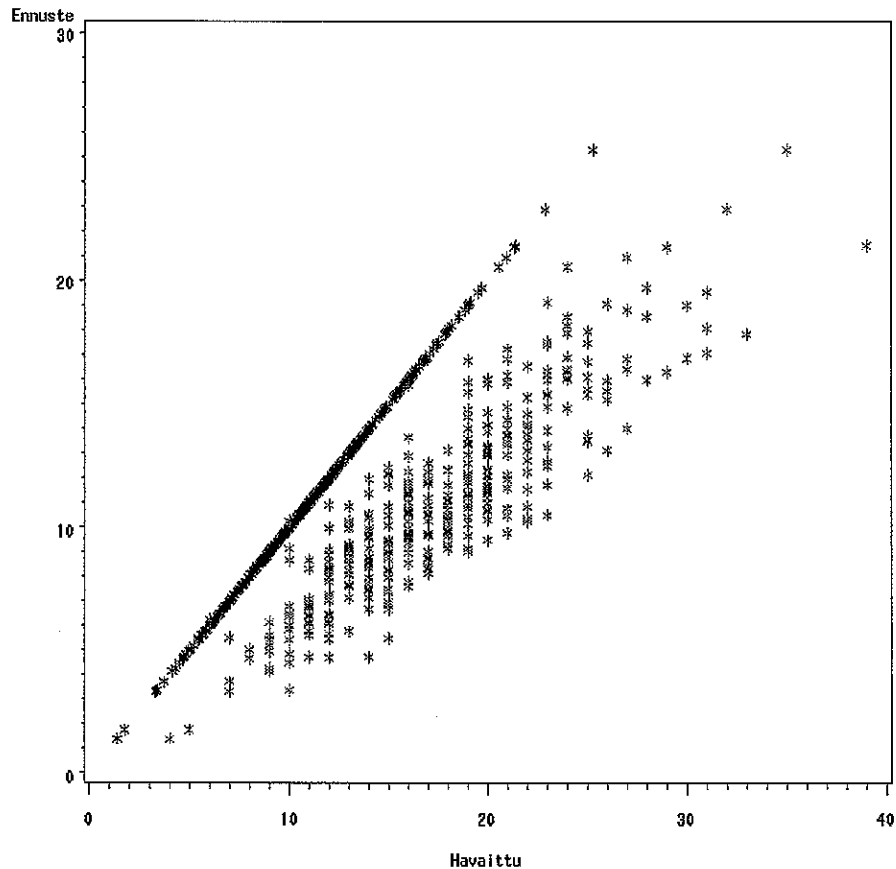
Kuva 5.4. Vakuutusvuosien lukumäärän parametrin posteriori-jakauma. Kuvasta havaitaan, että priori-jakaumasta poiketen posteriori-jakauma on vasemmalle vino.



Posteriori-jakauman keskiarvosta saadaan vakuutusvuosien lukumäärän parametrin estimaatiksi 1,014 sekä tämän keskihajonnaksi 0,003. Posteriori-jakauman vinous on -0,832 eli jakauma on selvästi vasemmalle vino. Lisäksi jakauman huipukkuus on 0,524.

Lopuksi tutkitaan vielä kuinka hyvin mallin antamat ennusteet mukailevat havaittuja vahinkojen lukumääriä.

Kuva 5.5. Yhden parametrin mallin ennuste vastaan havaittu kuvio.



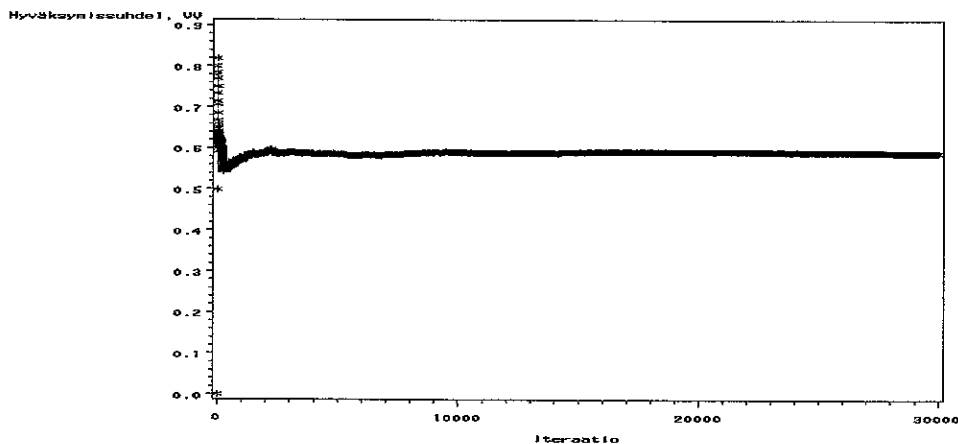
Kuvasta 5.5 nähdään, että yhden parametrin mallin antamat ennusteet ovat säännöllisesti pienempiä kuin todelliset havaitut arvot. Täten yhden parametrin malli on liian karkea vahinkojen lukumäärien ennustamiseen.

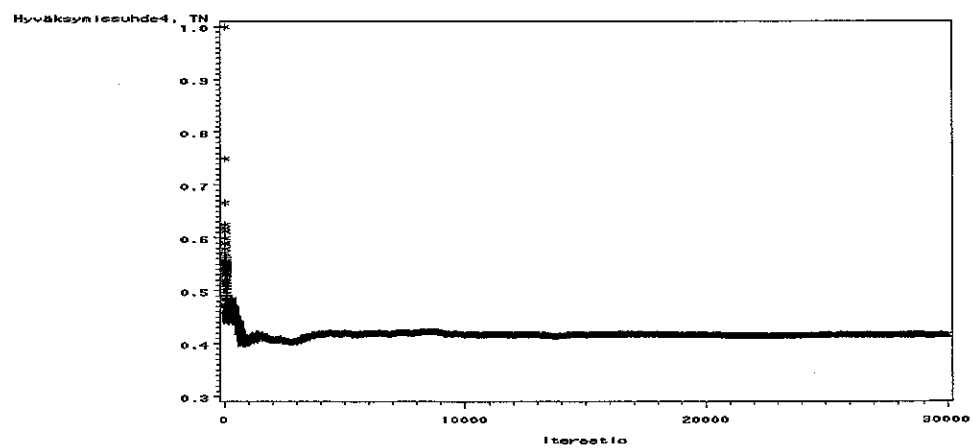
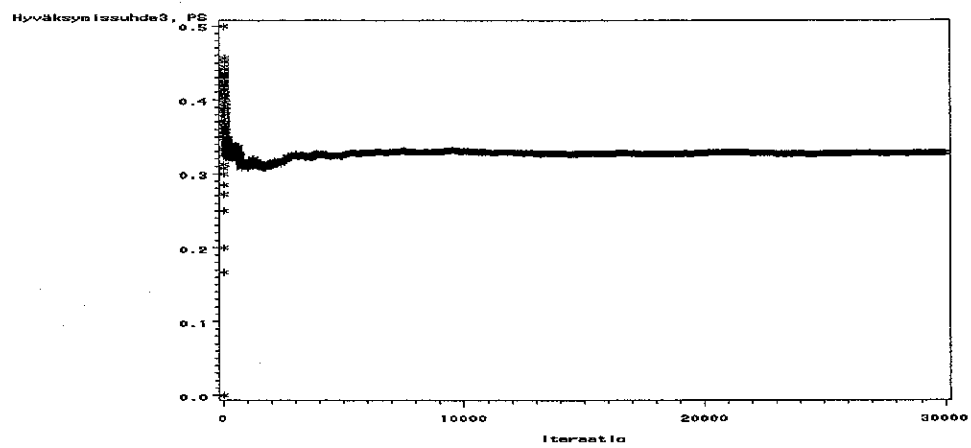
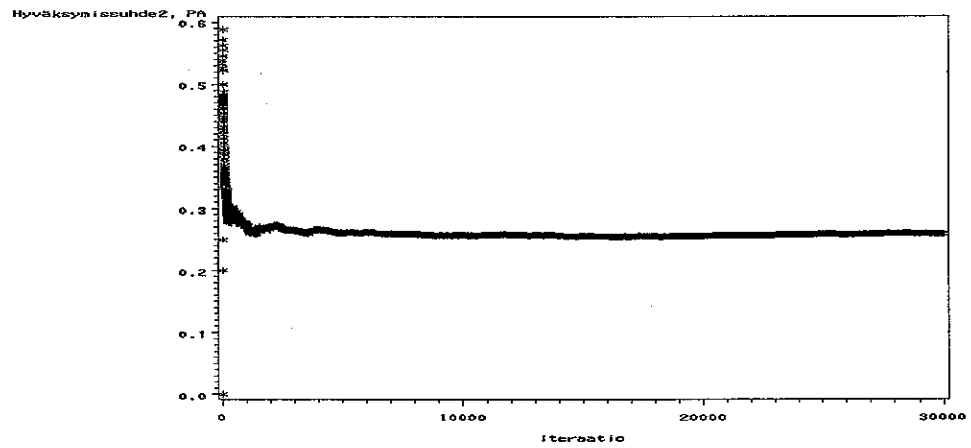
5.3 Usean parametrin tilanne

Tutkitaan parametrien estimointia mallissa, jossa ovat nyt mukana opinnäytetyössä [11] tilastollisesti merkitseviksi todetut tekijät: vakuutusvuosien lukumäärä, päivän alin lämpötila, päivän sademäärä sekä tuulen nopeus. Priori-jakaumien oletetaan olevan normaalijakaumia ja hyperparametrit on määrätty samalla tavalla kuin yhden parametrin tilanteessa. Hyperparametrien arvot ovat vakuutusvuosien lukumäärälle $\mu = 0,7528$ sekä $\sigma^2 = 0,0069 * 0,0069$ vastaavasti päivän alimmalle lämpötilalle $\mu = -0,0024$ sekä $\sigma^2 = 0,0006 * 0,0006$, päivän sademäärälle $\mu = 0,0020$ sekä $\sigma^2 = 0,0007 * 0,0007$ ja tuulen nopeudelle $\mu = 0,0037$ sekä $\sigma^2 = 0,0014 * 0,0014$. Ehdotussiirtymä-funktioina q_i käytetään tasaisia jakaumia $U_i(\beta_i - \epsilon_i, \beta_i + \epsilon_i)$, missä $\epsilon_i = 0,01$, $i = 1, 2, 3, 4$. Iteraatioita suoritetaan 30 000 ja vain joka 25. havainto otetaan mukaan lopulliseen otokseen. Algoritmina käytetään esimerkin 4.2 mukaista komponenteittaista Metropolis-Hastings-algoritmia.

Seuraavaksi tutkitaan hyväksymissuhdelukuja.

Kuva 5.6. Usean parametrin mallin hyväksymissuhdeluvut parametreittain. VV on vakuutusvuosien lukumäärä, PA päivän alin lämpötila, PS on päivän sademäärä sekä TN on tuulen nopeus.



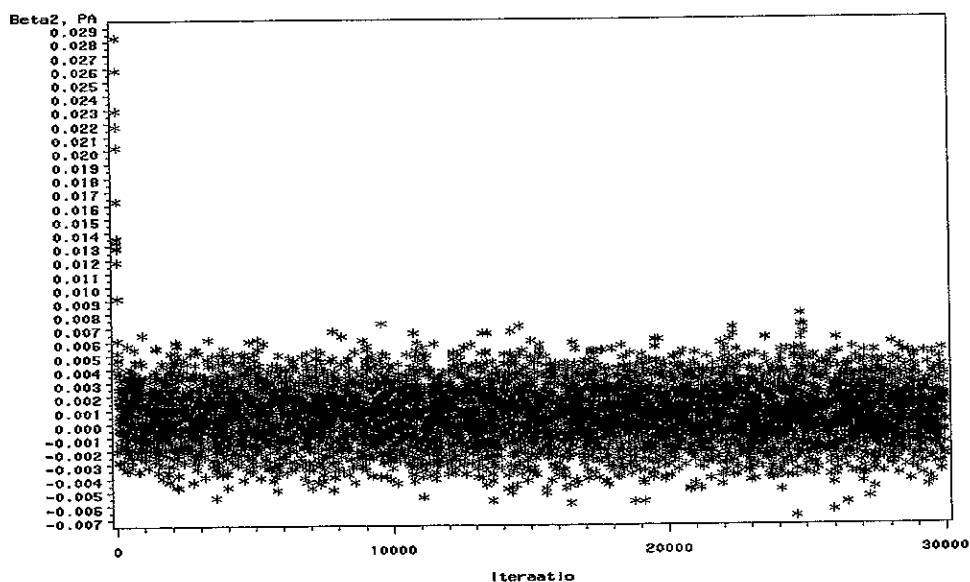
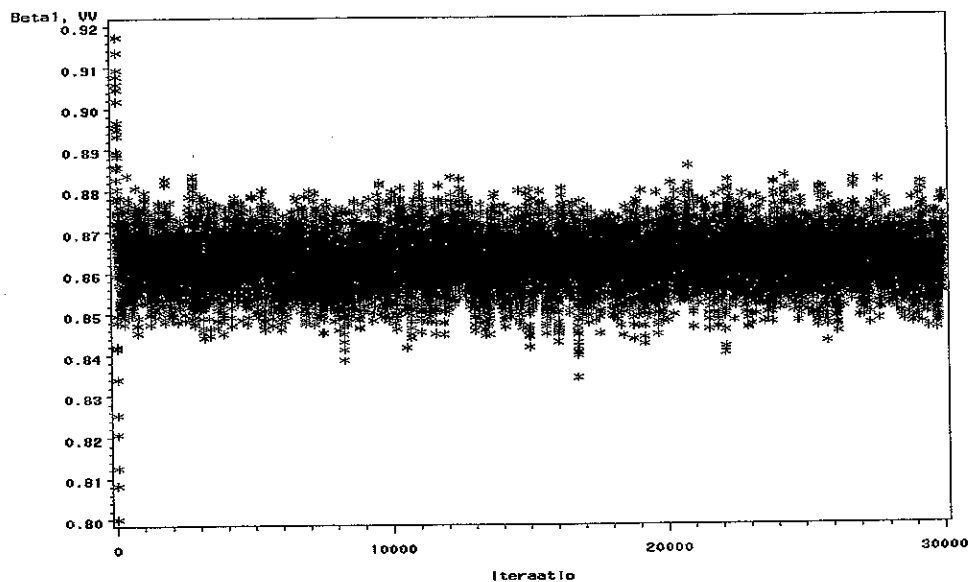


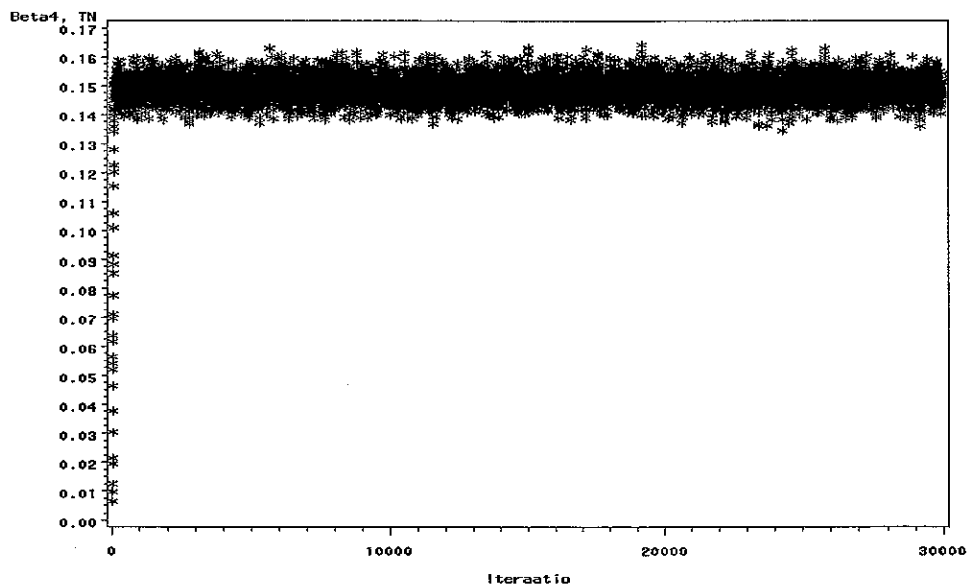
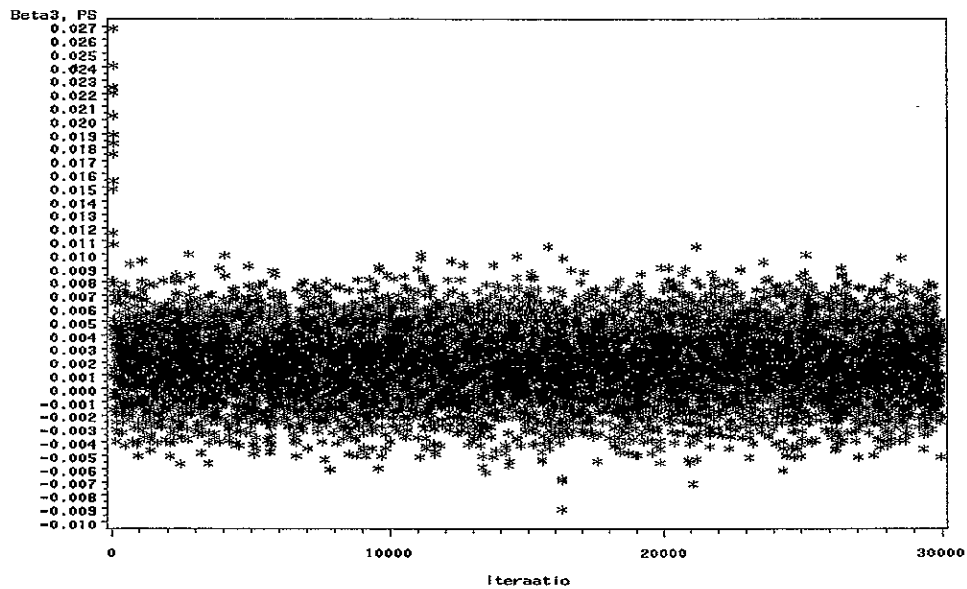
Hyväksymissuhdeluku vakiintuu kaikilla parametreilla nopeasti. Päivän alimman lämpötilan hyväksymissuhdeluku näyttää vakiintuvan alhaisimmalle

tasolle, mutta sekin on kuitenkin luokkaa 0,3 (eli noin kolme siirtymää kymmenestä hyväksytään).

Tutkitaan miltä simuloitujen parametrien arvojen käyrät näyttävät.

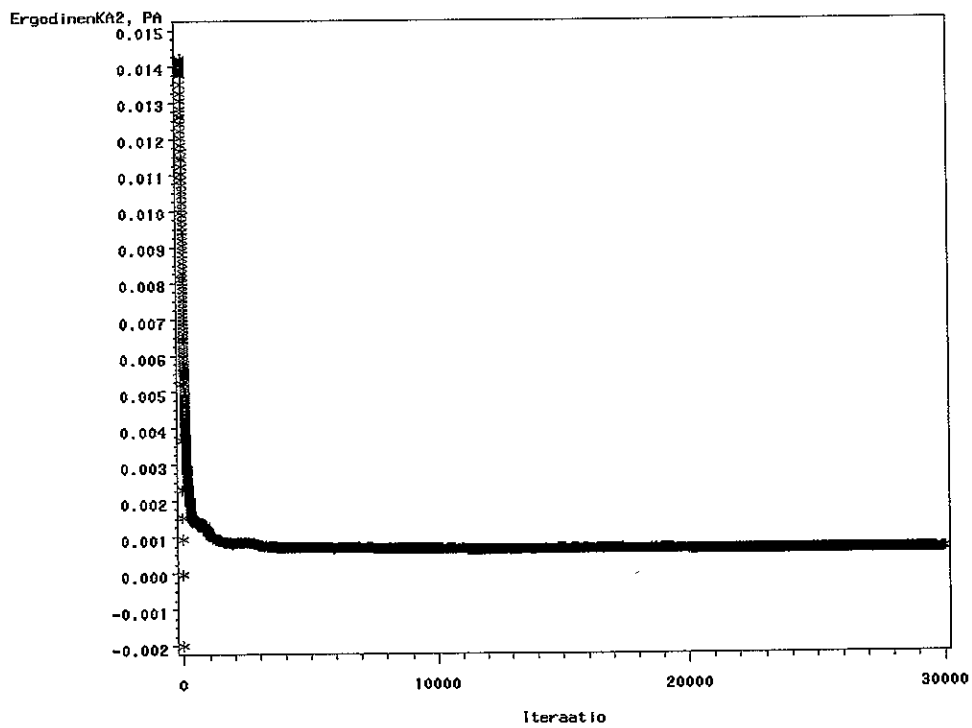
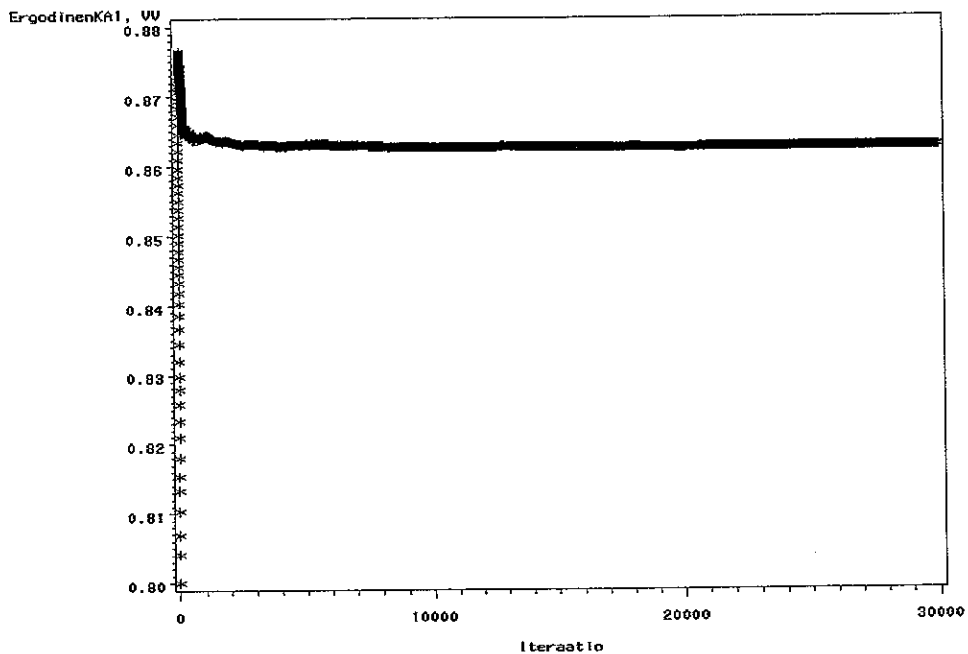
Kuva 5.7. Parametrien arvot iteraatiokierrosten funktiona.

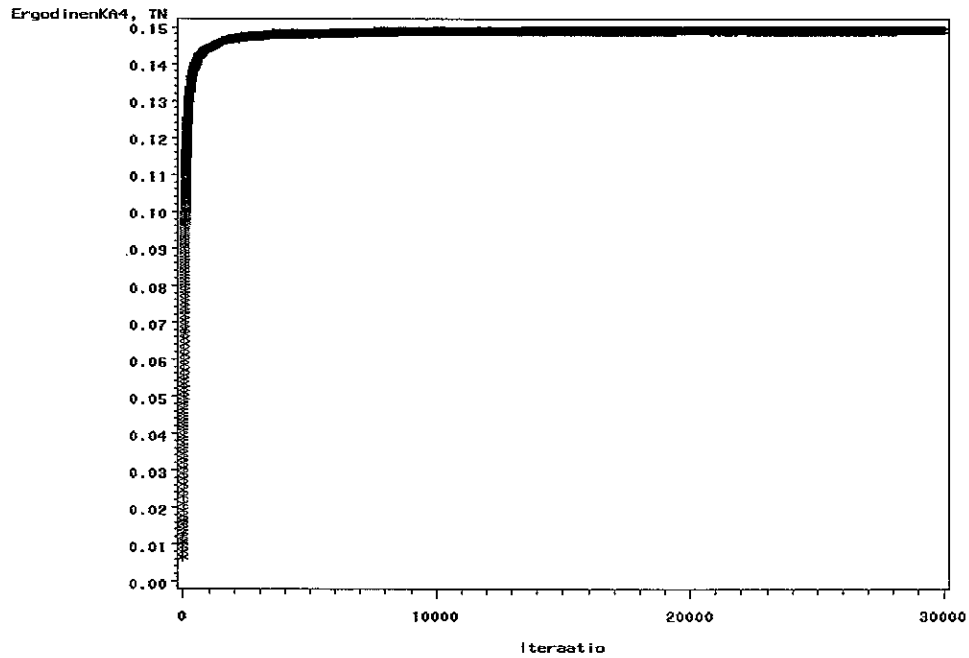
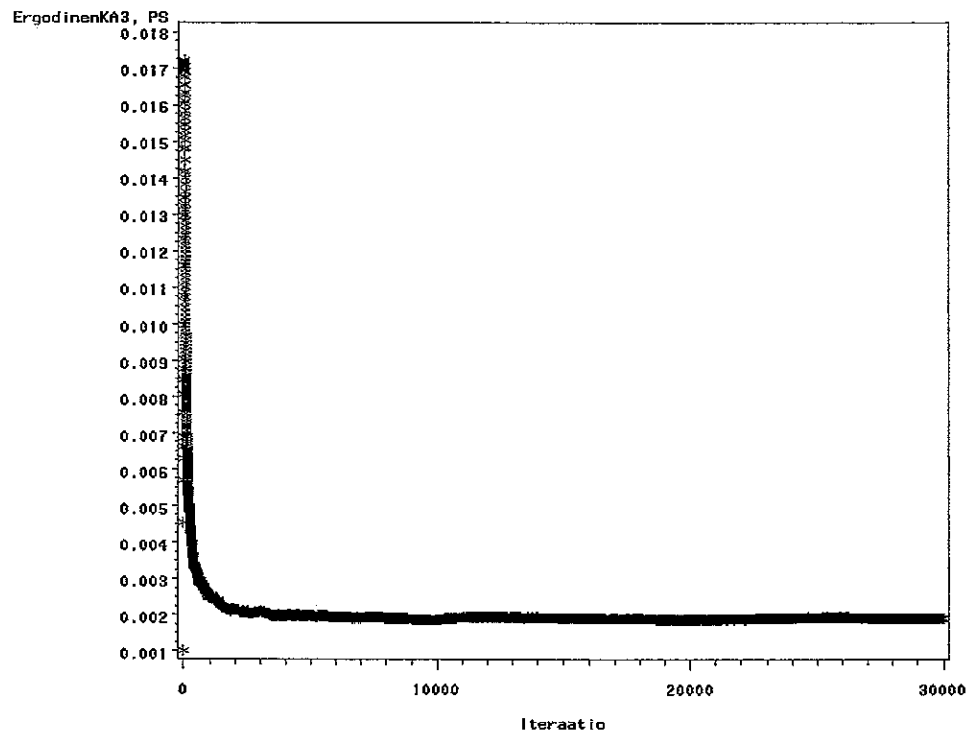




Simuloinnit näyttävät silmämääräisesti saavuttavan stationaarisen jakau-
man melko nopeasti. Konvergenssista varmistumisen vuoksi tutkitaan seu-
raavaksi ergodisten keskiarvojen käyrien suppenemista.

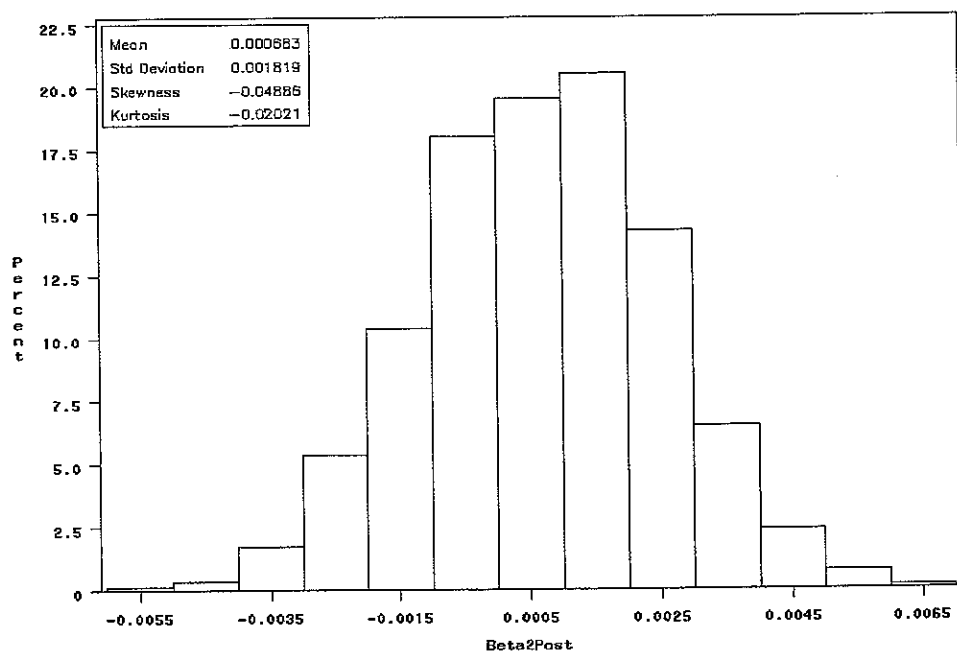
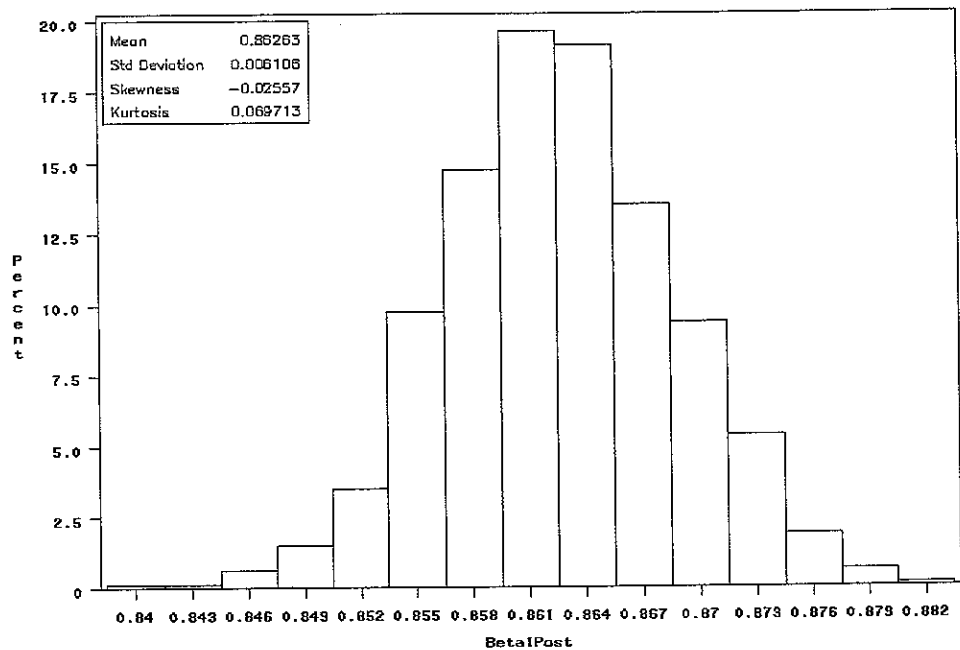
Kuva 5.8. Parametrien ergodisten keskiarvojen käyrät (funktio t on identiteettifunktio).

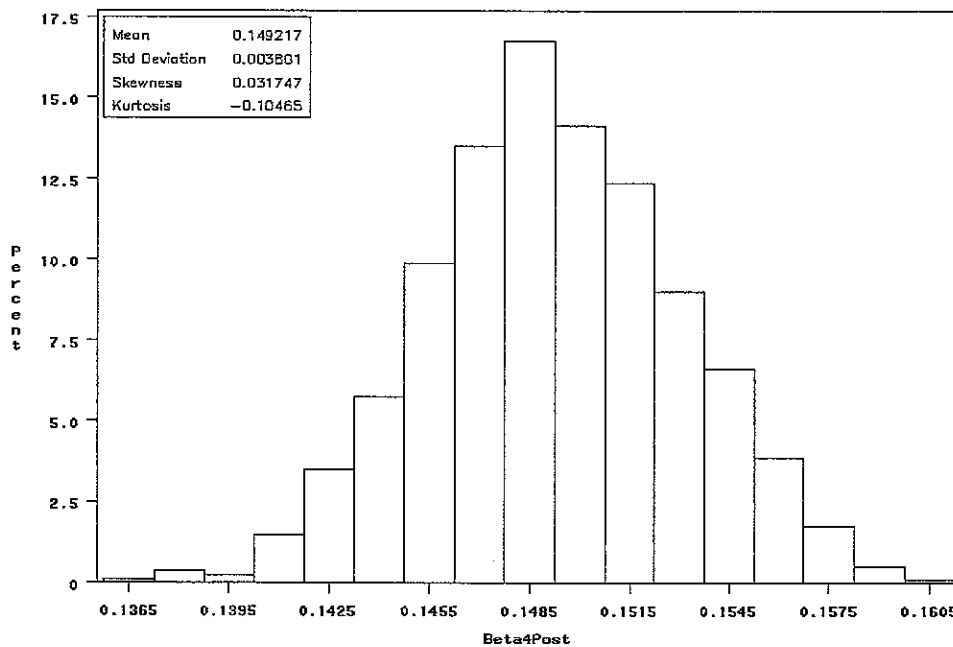
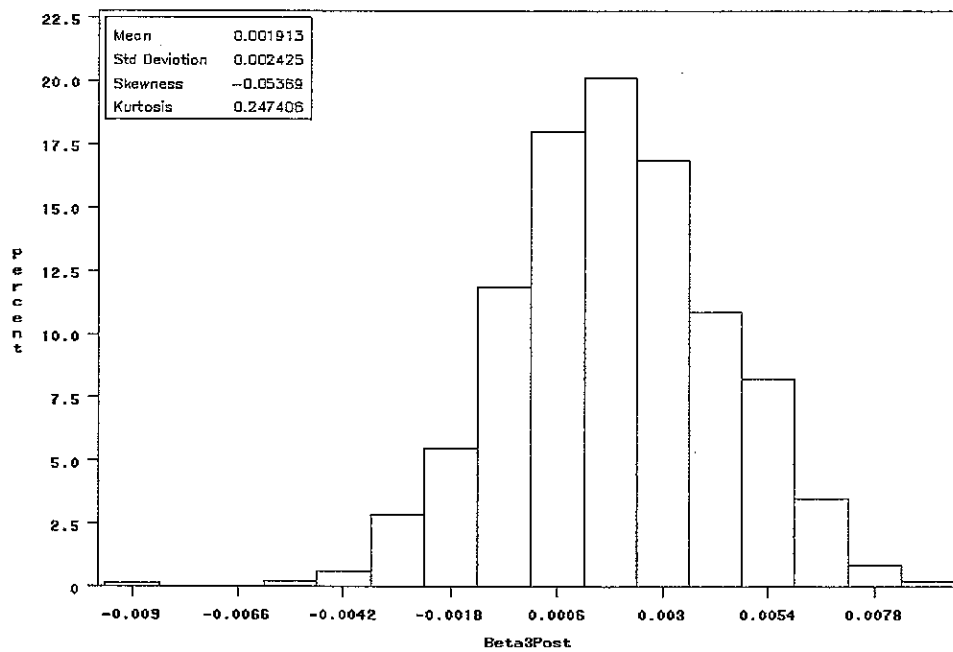




Kaikkien parametrien ergodisten keskiarvojen käyrät näyttävät suppenevan jonkin vakioarvon lähelle, joka tukee yllä esitettyä konvergenssi päätelmää.

Tutkitaan burn in -arvolla 10 000 saatuja parametrien posteriori-jakaumia.
Kuva 5.9. Parametrien posteriori-jakaumat sekä posteriori-jakaumien keskiarvot, keskihajonnat, vinoudet ja huipukkuudet.



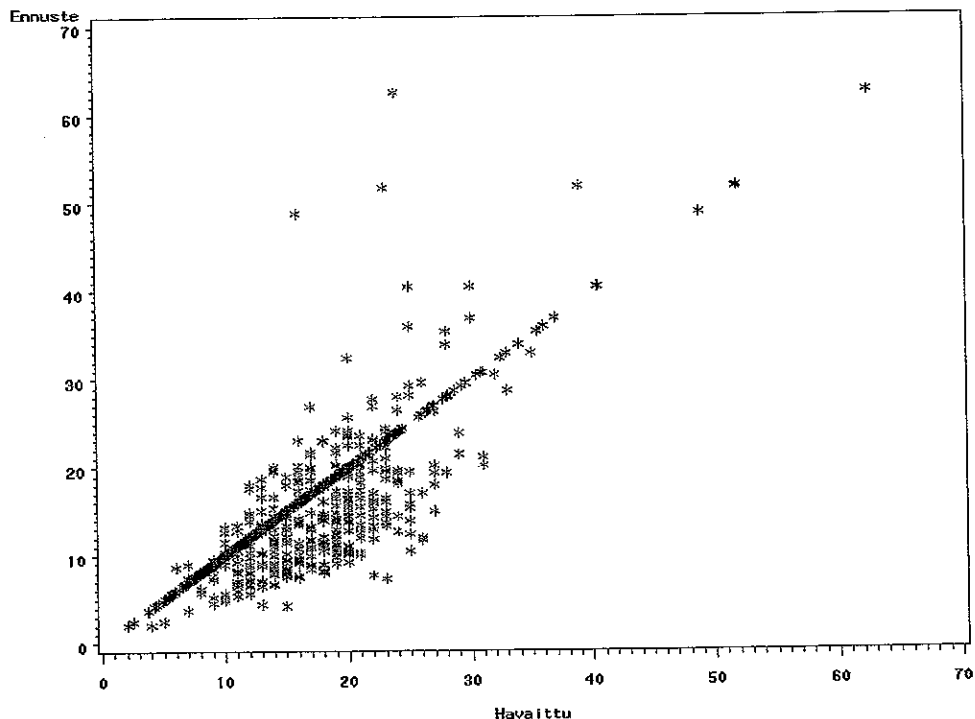


Huomataan, että parametrien posteriori-jakaumat eivät ole vinoja, vaan nättejä yksihiippuisia jakaumia (vertaa yhden parametrin tilanteeseen, jossa saatiin voimakkaasti vasemmalle vino jakauma). Parametrien estimaateiksi

saadaan vakuutusvuosien lukumäärälle 0,863, päivän alimmalle lämpötilalle 0,001, päivän sademäärälle 0,002 sekä tuulen nopeudelle 0,149.

Tutkitaan vielä kuinka hyvin mallin antamat ennusteet osuvat havaittuihin vahinkojen lukumääriin.

Kuva 5.10. Usean parametrin mallin ennuste vastaan havaittu kuvio.



Kuvasta 5.10 nähdään, että mallin antamat ennusteet osuvat paljon paremmin havaittuihin arvoihin kuin yhden parametrin tilanteessa. Kuvasta 5.10 havaitaan myös muutamia silmämääräisesti selvästi poikkeavia havaintoja. Useamman parametrin malli mallintaa selvästi paremmin liikennevakuutuksen vahinkojen lukumääriä kuin edellä tutkittu yhden parametrin malli.

Verrataan vielä lopuksi yllä Metropolis-Hastings-algoritmillä saatuja parametrien estimaatteja ja parametrien estimaattien keskihajontoja opinnäyte-työssä [11] yleistetyllä lineaarisella mallilla saatuihin parametrien estimaatteihin ja parametrien estimaattien keskivirheisiin.

Taulukko 5.1. Opinnäytetyössä [11] yleistetyllä lineaarisella mallilla saadut parametriestimaatit ja parametriestimaattien keskivirheet sekä Metropolis-Hastings-algoritmillä saadut parametriestimaatit ja parametriestimaattien keskihajonnat.

Parametri	GLM Estimaatti	GLM KV	M-H Estimaatti	M-H KH
VV	0,7528	0,0069	0,8626	0,0061
PA	-0,0024	0,0006	0,0007	0,0018
PS	0,0020	0,0007	0,0019	0,0024
TN	0,0037	0,0014	0,1492	0,0038

Taulukosta 5.1 nähdään, että Metropolis-Hastings-algoritmillä saadun vakuutusvuosien lukumäärän parametriestimaatin keskihajonta on pienempi kuin yleistetyllä lineaarisella mallilla saadun parametriestimaatin keskivirhe. Muiden parametriestimaattien osalta yleistetyn lineaarisen mallin keskivirheet ovat pienempiä kuin Metropolis-Hastings-algoritmillä saatujen parametriestimaattien keskihajonnat. Täten Metropolis-Hastings-algoritmillä saatiin tarkempi parametriestimaatti vakuutusvuosien lukumäärän parametrille.

Metropolis-Hastings-algoritmia käytettäessä jouduttiin havaintoaineistoa rajoittamaan, sillä Metropolis-Hastings-algoritmia sovellettaessa joudutaan (hyväksymistodennäköisyyttä α määrittäessä) jokaisella iteraatiokierroksella laskemaan havaintojen lukumäärän mukainen määrä uskottavuusfunktioiden tuloja. Käytännössä SAS[®]-ohjelmiston laskentatarkkuus ei olisi riittänyt hyväksymistodennäköisyyksien laskemiseen, mikäli havaintoaineistoa ei olisi rajoitettu.

Metropolis-Hastings-algoritmillä pystytään määrittämään vain numeeristen selittäjien parametriestimaatit. Luokittelevien selittäjien, kuten opinnäytetyössä [11] käytetyn vahingon sattumiskuukauden osalta, aineisto pitäisi Metropolis-Hastings-algoritmia sovellettaessa jakaa osiin ja määrätä parametrien estimaatit jokaiselle osa-aineistolle erikseen.

Vertaamalla Metropolis-Hastings-algoritmillä saatuja tuloksia opinnäytetyössä [11] yleistetyllä lineaarisella mallilla saatuihin tuloksiin huomataan, että yleistetty lineaarinen malli soveltuu paremmin vahinkojen lukumäärien arvioimiseen.

6 Yhteenveto

Tutkielmassa on esitetty ensin luvuissa 2 ja 3 pohjatietoja Bayesiläisestä analyysistä sekä Markovin ketjuista.

Pohjustuksen jälkeen luvussa 4 on esitelty Markov Chain Monte Carlo -menetelmien teoriaa sekä erityisesti perehdytty niin kutsutun Metropolis-Hastings-algoritmin ominaisuuksiin. Luvussa 4 on tutkittu muun muassa Metropolis-Hastings-algoritmin tarvitsemia ehdotussiirtymäfunktioita sekä algoritmin konvergenssia.

Tutkielmassa [11] on tutkittu parametrien estimoimista yleistettyjen lineaaristen mallien avulla. Tässä SHV-tutkielmassa esitetty parametrien estimointi Metropolis-Hastings-algoritmin avulla muodostaa toisen mahdollisuuden tämentyypiseen parametrien estimointiin.

Tutkielman viidennessä luvussa on esitetty käytännön sovelluksena liikenevakuutuksen vahinkojen lukumäärämalliin tutkielmassa [11] valittujen tekijöiden parametrien estimointi käyttäen luvun 4 Metropolis-Hastings-algoritmeja. Käytännön sovelluksissa parametrien priorijakaumiksi on valittu normaali-jakaumat sekä ehdotussiirtymäfunktioiksi tasaiset jakaumat.

Metropolis-Hastings-algoritmit yhdistettyinä hierarkkisiin yleistettyihin lineaarisiin malleihin mahdollistavat monipuoliset mallinnusmahdollisuudet.

Lähteet

- [1] Bennett J.E, Racine-Poon A., Wakefield J.C.: *MCMC for Nonlinear Hierarchical Models*. Markov Chain Monte Carlo in Practice, Chapman and Hall, Lontoo, 1996.
- [2] Gamerman D.: *Markov Chain Monte Carlo*. Chapman and Hall, New York, 1997.
- [3] Gelman A., Roberts G.O., Gilks W.R. *Efficient Metropolis Jumping Rules*. Bayesian Statistics 5, Oxford University press, s.599-607, 1996.
- [4] Gill J.: *Is Partial-Dimension Convergence a Problem for MCMC Algorithms?*. Artikkel, University of California-Davis, 2005.
- [5] Hastings W.K.: *Monte Carlo Sampling Methods Using Markov Chains and their Applications*. Biometrika **57**, s.97-109, 1970.
- [6] Krogstad F.: *Coding the Metropolis-Hastings Algorithm*. Artikkel, UW Forestry, 1999.
- [7] Nummelin E.: *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University press, Cambridge, 1984.
- [8] Parvinen K.: *Stokastiset prosessit*. Luentomoniste, Turun yliopisto, 2002.
- [9] Peskun P.H.: *Optimum Monte Carlo Sampling Using Markov Chains*. Biometrika **60**, s.607-612, 1973.
- [10] Roberts G.O., Smith A.F.M.: *Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms*. Stochastic Processes and their Applications **49**, s.207-216, 1994.
- [11] Sihvonen P.: *Vahinkojen lukumäärien arvioimisesta yleistetyillä lineaarisilla malleilla sekä Bayesiläisellä analyysillä*. Pro gradu -tutkielma, Turun yliopisto, 2005.
- [12] Tierney L.: *Markov Chains for Exploring Posterior Distributions (with discussion)*. Annals of Statistics **22**, s.1701-1762, 1994.

Liite 1. Yhden parametrin Metropolis – Hastings SAS – koodi

```
/*METROPOLIS-HASTINGS -ALGORITMI*/  
Data Metropolis;  
Set VahLKM; /*VAHINKOJEN LUKUMÄÄRÄT*/  
Set Riskimitta; /*VAKUUTUSVUOSIEN LUKUMÄÄRÄT*/  
Array VahLKM{ &HAV } VahLKM1-VahLKM&HAV; /*HAV ON 366KPL*/  
Array Riskimitta{ &HAV } Riskimitta1-Riskimitta&HAV;  
KA = 0.7528; /*PRIORI -JAKAUMAN KESKIARVO*/  
VA = 0.0069*0.0069; /*PRIORI -JAKAUMAN VARIANSSI*/  
Epsilon = 0.01; /*EHDOTUSSIIRTYMÄFUNKTION EPSILON*/  
Laskuri=0;  
BetaOld=0.8; /*ALKUARVO PARAMETRILLE BETA*/  
BetaNew=0;  
Seed = 8546867; /*SATUNNAISLUKUGENERAATTORIN SIEMENLUKU*/  
Do i = 1 To &AJOJA; /*AJOJA ON ITERAATIOIDEN LUKUMÄÄRÄ*/  
RU = ranuni(Seed);  
/*EHDOTUSSIIRTYMÄ*/  
BetaNew = BetaOld - Epsilon + 2*RU*Epsilon;  
/*HYVÄKSYMISTODENNÄKÖISYYDEN LASKENTA*/  
Alpha=(exp(-(BetaNew-KA)*(BetaNew-KA)/(2*VA)))  
/(exp(-(BetaOld-KA)*(BetaOld-KA)/(2*VA)));  
Do j = 1 to HAV; /*HAV ON HAVAINTOJEN LUKUMÄÄRÄ = 366*/  
Alpha = Alpha*(exp(Riskimitta(j)*VahLKM(j)*BetaNew - exp(Riskimitta(j)*  
BetaNew)))/(exp(Riskimitta(j)*VahLKM(j)*BetaOld - exp(Riskimitta(j)*BetaOld)));  
End;  
/*TESTATAAN HYVÄKSYTÄÄNKÖ EHDOTETTU SIIRTYMÄ*/  
Seedi = 1985367;  
TestU = ranuni(Seedi);  
If Alpha > TestU Then BetaOld = BetaNew;  
If Alpha > TestU Then Laskuri = Laskuri + 1;  
Output;  
End;  
Run;
```

Liite 2. Usean parametrin Metropolis – Hastings SAS – koodi
/*KOMPONENTEITTAINEN METROPOLIS-HASTINGS -ALGORITMI*/
Data Metropolis;
Set VahLKM; /*Vahinkojen lukumäärä*/
Set Riskimitta; /*Vakuutusvuosien lukumäärä*/
Set PA; /*Päivän alin lämpötila*/
Set PS; /*Päivän sademäärä*/
Set TN; /*Tuulennopeus*/
Array Y{&HAV} VahLKM1-VahLKM&HAV;
Array XRM{&HAV} Riskimitta1-Riskimitta&HAV;
Array XPA{&HAV} PA1-PA&HAV;
Array XPS{&HAV} PS1-PS&HAV;
Array XTN{&HAV} TN1-TN&HAV;
/*PARAMETRIEN PRIORI -JAKAUMIEN HYPERPARAMETRIT*/
/*KESKIARVOT*/
Array KA{4} KA1-KA4;
KA1 = 0.7528; /*VAKUUTUSVUOSIEN PRIORI -JAKAUMAN KESKIARVOT*/
KA2 = -0.0024; /*PÄIVÄN ALIMMAN LÄMPÖTILAN PRIORI -JAKAUMAN KESKIARVO*/
KA3 = 0.0020; /*PÄIVÄN SADEMÄÄRÄN PRIORI -JAKAUMAN KESKIARVOT*/
KA4 = 0.0037; /*TUULEN NOPEUDEN PRIORI -JAKAUMAN KESKIARVOT*/
Array VA{4} VA1-VA4;
VA1 = 0.0069*0.0069; /*VAKUUTUSVUOSIEN LUKUMÄÄRÄN PRIORI -JAKAUMAN VARIANSSI*/
VA2 = 0.0006*0.0006; /*PÄIVÄN ALIMMAN LÄMPÖTILAN PRIORI -JAKAUMAN VARIANSSI*/
VA3 = 0.0007*0.0007; /*PÄIVÄN SADEMÄÄRÄN PRIORI -JAKAUMAN VARIANSSI*/
VA4 = 0.0014*0.0014; /*TUULENNOPEUDEN PRIORI -JAKAUMAN VARIANSSI*/

```

/*PARAMETRIVEKTORIT*/
Array BetaOld{4} BetaOld1-BetaOld4;
BetaOld1=0.80; BetaOld2=-0.002; BetaOld3=0.001;BetaOld4=0.002;
Array BetaNew{4} BetaNew1-BetaNew4;
BetaNew1=0; BetaNew2=0; BetaNew3=0; BetaNew4=0;
/*EHDOTUSSIIRTYMÄFUNKTIOIDEN EPSILONIT*/
Epsilon1=0.01;
Epsilon2=0.01;
Epsilon3=0.01;
Epsilon4=0.01;
Laskuri1=0;
Laskuri2=0;
Laskuri3=0;
Laskuri4=0;
/*SATUNNAISLUKUGENERAATTORIEN SIEMENLUVUT*/
Seed1 = 8546867;
Seed2= 8957467;
Seed3= 336767;
Seed4= 1765746;
Do i = 1 To &AJOJA; /*AJOJA ON ITERAATIOIDEN LUKUMÄÄRÄ*/
RU1 = ranuni(Seed1);
RU2 = ranuni(Seed2);
RU3 = ranuni(Seed3);
RU4 = ranuni(Seed4);
/*EHDOTUSSIIRTYMÄT*/
BetaNew(1) = BetaOld(1) - Epsilon1 + 2*RU1*Epsilon1;
BetaNew(2) = BetaOld(2) - Epsilon2 + 2*RU2*Epsilon2;
BetaNew(3) = BetaOld(3) - Epsilon3 + 2*RU3*Epsilon3;
BetaNew(4) = BetaOld(4) - Epsilon4 + 2*RU4*Epsilon4;
/*HYVÄKSYMISTODENNÄKÖISYYKSIEN LASKENNAT*/
Alpha1=1;
Alpha2=1;
Alpha3=1;

```

```

Alpha4=1;
/*KOMONENTTEITTAINEN ALGORITMI, LIIKUTETAAN AINA VAIN
YHTÄ PARAMETRIA*/
Alpha1=Alpha1*(exp(-(BetaNew(1)-KA(1))*(BetaNew(1)-KA(1))/(2*VA(1))
-(BetaOld(2)-KA(2))*(BetaOld(2)*KA(2))/(2*VA(2))
-(BetaOld(3)-KA(3))*(BetaOld(3)*KA(3))/(2*VA(3))
-(BetaOld(4)-KA(4))*(BetaOld(4)*KA(4))/(2*VA(4))))
/(exp(-(BetaOld(1)-KA(1))*(BetaOld(1)-KA(1))/(2*VA(1))
-(BetaOld(2)-KA(2))*(BetaOld(2)*KA(2))/(2*VA(2))
-(BetaOld(3)-KA(3))*(BetaOld(3)*KA(3))/(2*VA(3))
-(BetaOld(4)-KA(4))*(BetaOld(4)*KA(4))/(2*VA(4))));
Alpha2=Alpha2*(exp(-(BetaOld(1)-KA(1))*(BetaOld(1)-KA(1))/(2*VA(1))
-(BetaNew(2)-KA(2))*(BetaNew(2)*KA(2))/(2*VA(2))
-(BetaOld(3)-KA(3))*(BetaOld(3)*KA(3))/(2*VA(3))
-(BetaOld(4)-KA(4))*(BetaOld(4)*KA(4))/(2*VA(4))))
/(exp(-(BetaOld(1)-KA(1))*(BetaOld(1)-KA(1))/(2*VA(1))
-(BetaOld(2)-KA(2))*(BetaOld(2)*KA(2))/(2*VA(2))
-(BetaOld(3)-KA(3))*(BetaOld(3)*KA(3))/(2*VA(3))
-(BetaOld(4)-KA(4))*(BetaOld(4)*KA(4))/(2*VA(4))));
Alpha3=Alpha3*(exp(-(BetaOld(1)-KA(1))*(BetaOld(1)-KA(1))/(2*VA(1))
-(BetaOld(2)-KA(2))*(BetaOld(2)*KA(2))/(2*VA(2))
-(BetaNew(3)-KA(3))*(BetaNew(3)*KA(3))/(2*VA(3))
-(BetaOld(4)-KA(4))*(BetaOld(4)*KA(4))/(2*VA(4))))
/(exp(-(BetaOld(1)-KA(1))*(BetaOld(1)-KA(1))/(2*VA(1))
-(BetaOld(2)-KA(2))*(BetaOld(2)*KA(2))/(2*VA(2))
-(BetaOld(3)-KA(3))*(BetaOld(3)*KA(3))/(2*VA(3))
-(BetaOld(4)-KA(4))*(BetaOld(4)*KA(4))/(2*VA(4))));
Alpha4=Alpha4*(exp(-(BetaOld(1)-KA(1))*(BetaOld(1)-KA(1))/(2*VA(1))
-(BetaOld(2)-KA(2))*(BetaOld(2)*KA(2))/(2*VA(2))
-(BetaOld(3)-KA(3))*(BetaOld(3)*KA(3))/(2*VA(3))
-(BetaNew(4)-KA(4))*(BetaNew(4)*KA(4))/(2*VA(4))))
/(exp(-(BetaOld(1)-KA(1))*(BetaOld(1)-KA(1))/(2*VA(1))
-(BetaOld(2)-KA(2))*(BetaOld(2)*KA(2))/(2*VA(2))

```


$-(\text{BetaOld}(3)-\text{KA}(3)) * (\text{BetaOld}(3) * \text{KA}(3)) / (2 * \text{VA}(3))$
 $-(\text{BetaOld}(4)-\text{KA}(4)) * (\text{BetaOld}(4) * \text{KA}(4)) / (2 * \text{VA}(4))));$
 Do k = 1 to &HAV;
 $\text{Alpha1} = \text{Alpha1} * (\exp((\text{XRM}(k) * \text{BetaNew}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)) * \text{Y}(k)$
 $- \exp(\text{XRM}(k) * \text{BetaNew}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)))) /$
 $(\exp((\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)) * \text{Y}(k)$
 $- \exp(\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)))));$
 $\text{Alpha2} = \text{Alpha2} * (\exp((\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaNew}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)) * \text{Y}(k)$
 $- \exp(\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaNew}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)))) /$
 $(\exp((\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)) * \text{Y}(k)$
 $- \exp(\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)))));$
 $\text{Alpha3} = \text{Alpha3} * (\exp((\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaNew}(3) + \text{XTN}(k) * \text{BetaOld}(4)) * \text{Y}(k)$
 $- \exp(\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaNew}(3) + \text{XTN}(k) * \text{BetaOld}(4)))) /$
 $(\exp((\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)) * \text{Y}(k)$
 $- \exp(\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)))));$
 $\text{Alpha4} = \text{Alpha4} * (\exp((\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaNew}(4)) * \text{Y}(k)$
 $- \exp(\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaNew}(4)))) /$
 $(\exp((\text{XRM}(k) * \text{BetaOld}(1) + \text{XPA}(k) * \text{BetaOld}(2) +$
 $\text{XPS}(k) * \text{BetaOld}(3) + \text{XTN}(k) * \text{BetaOld}(4)) * \text{Y}(k)$

```

- exp(XRM(k)*BetaOld(1)+XPA(k)*BetaOld(2)+
XPS(k)*BetaOld(3)+XTN(k)*BetaOld(4))));
End;
/*TESTATAAN HYVÄKSYTÄÄNKÖ SIIRTYMÄT*/
Seedi1 = 1985367;
Seedi2 = 878376;
Seedi3 = 735653;
Seedi4 = 835637;
TestU1 = ranuni(Seedi1);
TestU2 = ranuni(Seedi2);
TestU3 = ranuni(Seedi3);
TestU4 = ranuni(Seedi4);
If Alpha1 > TestU1 Then Do BetaOld1 = BetaNew1; Laskuri1 =
Laskuri1 + 1;End;
If Alpha2 > TestU2 Then Do BetaOld2 = BetaNew2; Laskuri2 =
Laskuri2 + 1;End;
If Alpha3 > TestU3 Then Do BetaOld3 = BetaNew3; Laskuri3 =
Laskuri3 + 1;End;
If Alpha4 > TestU4 Then Do BetaOld4 = BetaNew4; Laskuri4 =
Laskuri4 + 1;End;
Output;
End;
Run;

```