

Imbi Traat, Natalja Lepik (Tartu Ülikool), 2013



E-kursuse

Bayesi statistika Markovi ahelatega

materjalid

Aine maht 6 EAP

Imbi Traat, Natalja Lepik (Tartu Ülikool), 2013

Sisukord

1	Bayesi statistika	8
1.1	Bayesi teoreem	8
1.2	Vaatluste prognoosimine	13
1.3	Mitmemõõtmeline parameeter Bayesi analüüsis	15
1.4	Karakteristikute leidmine järeldaotusest	15
1.5	Kaasjaotus (<i>Conjugate distribution</i>)	18
1.5.1	Eksponeentsiaalpere kaasjaotused (ühemõõtmeline juht)	18
1.5.2	Tihedusfunktsiooni teisenemine juhusliku suuruse teisendamisel	20
1.5.3	Eksponeentsiaalpere kanooniline kuju	20
1.5.4	Näiteid eksponeentsiaalpere kaasjaotuste kohta	21
1.5.5	Mitmemõõtmelise eksponeentsiaalpere kaasjaotused	22
1.6	Veel kasulikke jaotusi	24
1.6.1	Gamma-, pöördgamma- ja normaalgamma jaotused	24
1.6.2	Normaaljaotuse marginaaljaotused ja tinglikud jaotused	25
1.7	Kaasjaotused regressioonimudeli korral	27
1.8	Tinglikud kaasjaotused (<i>conditional conjugacy</i>)	29
1.9	Hierarhilised mudelid	30
1.10	Mudeli sobivus	32
1.10.1	Mudeli sobivus prognoosiva jaotuse abil	32
1.10.2	Bayesi teststatistik	34
1.10.3	Hälbimus	35
1.11	Mitteinformatiivne eeljaotus	39
1.11.1	Korrektseid ja mittekorrektseid eeljaotused	39
1.11.2	Jeffrey invariantsuse omadus	41
2	Markovi ahelad	43
2.1	Markovi ahela üleminekutõenäosused	43
2.2	Markovi ahela marginaaljaotus	45
2.3	Seisundite klassifikatsioon	50
2.4	Piirteoreemid	52
2.5	Ergoodilise keskmise dispersioon ja tsentraalne piirteoreem	54
2.6	Pööratavad ahelad (<i>reversible chains</i>)	55
2.7	Pidevad seisundite ruumid	58
3	MCMC ja Bayesi statistika	61
3.1	Gibbsi valik	61
3.1.1	Definitsioon ja omadused	61
3.2	Praktilise rakendamise probleeme	65
3.2.1	Täistingliku tiheduse leidmine	65
3.3	Simuleerimine täistinglikust tihedusest	66
3.3.1	Valimi moodustamine	68
3.3.2	Uuendamise strateegiad	69
3.3.3	Valimi kasutamine	69

3.4	Praktilise rakendamise probleeme (jätkub)	72
3.4.1	Ümberparametriseerimine	73
3.4.2	Plokkimine	74
3.4.3	Koondumisdiagnostika	74
3.5	Metropolis-Hastingsi algoritmid	76
3.5.1	Definitsioon ja omadused	76
3.5.2	Ahelate erijuhud	79
3.6	Hübriidalgoritmid	81
3.6.1	Komponendiviisiline üleminek Metropolis-Hastingsi algoritmis	81
3.6.2	Metropolis + Gibbs	82
3.7	Veel andmeanalüüsi näiteid	86
	Kirjandus	93

Sissejuhatus

Antud konspekti koostamisel on kasutatud kirjanduse loetelus toodud algallikaid.

Statistikas on kaks suunda:

Sagedusstatistikud (*Frequentists*) – klassikaline statistika, statistilisteks otsustuseks kasutatakse ainult vaatlusandmete infot (sagedustabel, empiiriline jaotus jms.)

Bayesi statistikud (*Bayesians*) – statistiliste otsustuste tegemiseks kasutatakse vaatlusandmete kõrval ka aprioorseid ehk eelinfot.

Näide: Iganädalase arvamusküsitluse abil hinnatakse inimeste teadlikkuse sõltuvust reklaamikuludest:

y inimeste osakaal (%), kes panid tähele reklaami TV-s

x kulutused toote reklaamile

π teadlikkus, st tõenäosus, et inimene paneb tähele reklaami TV-s

Lihtsaimaks mudeliks on lineaarne seos: $\pi = \alpha + \beta x$. Kuna $\pi \in [0, 1]$, siis see pole alati sobiv mudel. Sageli kasutatakse logit teisendust,

$$\text{logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x.$$

Parameeter β on see, mis uurijat eriti huvitab. Mida suurem on β väärtus, seda kiiremini kasvab inimeste teadlikkus reklaamikulude kasvades.

Klassikaline statistika ei kasuta eelinfot, eeldatakse, et α ja β on konstandid (tundmatud), valimhinnangud $\hat{\alpha}$ ja $\hat{\beta}$ on juhuslikud. Bayesi statistikas saab mudelisse lisada eelteadmisi α ja β kohta. Need antakse jaotusega (st. parameetreid α ja β käsitletakse juhuslike suurustena). Ka tulemus α ja β kohta saadakse jaotusena. Bayesi hindamisülesande komponendid on:

$$\begin{array}{c} \alpha \text{ ja } \beta \text{ eeljaotus (aprioorne jaotus)} \\ + \text{ mudel (andmete ja parameetrite vaheline seos)} \\ + \text{ valimi andmed} \\ \Downarrow \\ \alpha \text{ ja } \beta \text{ järeljaotus (aposterioorne jaotus)} \end{array}$$

Järeljaotuse tuletamine on tähtis samm, aga mitte lõplik. Järeljaotus annab meile kogu info parameetri(te) kohta. Interpreteerimiseks kasutame aga järeljaotuse karakteristikuid, mis meie tundmatut parameetrit lihtsamalt kirjeldavad:

Keskväärts

Dispersioon

Mediaan, mood

Tõenäosusintervall

Näide (järg): Kuidas Bayesi statistik teeb oma järeldused? Reklaamiülesandes on põhiline huviobjekt β . Kui β jaotus paikneb 0 ümbruses, siis ei saa järeldada, et kultustel oleks teadlikkusele mõju, kui aga jaotus kontsentreerub mingite positiivsete väärtuste ümber, siis võib suure tõenäosusega öelda, et kulutused reklaamile tõstavad inimeste teadlikkust antud toote suhtes. Misa suuremate väärtuste ümber β jaotus kontsentreerub, seda tugevam on kulutuste mõju teadlikkusele.

Ülesanne. Tehkem mõned joonised võimalike järeljaotuste kohta parameetri β korral. Interpreteerigem.

Järeljaotuse teadmine võimaldab palju mitmekülgsemat infot parameetri kohta anda, kui seda saame klassikalises statistikas. Näiteks kui jaotuse määramispiirkond sisaldab 0, ei pea leppima järeldusega, et mõju teadlikkusele puudub, vaid saame leida arvuliselt $P(\beta > 0)$, ja näiteks veenduda, et $\beta > 0 = 0.92$. Seega tõenäosusega 0.92 on kulutustel siiski positiivne mõju teadlikkusele.

Järeljaotuse suur dispersioon viitab sellele, et meie teadmine parameetri kohta ei ole eriti täpne, on ebamäärane.

Probleemiks võib olla eeljaotuse andmine parameetritele, kui eelnev teadmine puudub. Ka siin on Bayesi statistikas lahendus olemas. Antakse lihtsalt üks hästi suure dispersiooniga jaotus, nn mitteinformatiivne eeljaotus. Varsti näeme, et sellisel juhul määravad järeljaotuse ainult andmed ja Bayesi tulemused langevad kokku klassikalise statistika tulemustega.

Ülesanne. Määrakem oma ettekujutust kasutades naiste keskmisele pikkusele eeljaotus. Olgu see normaaljaotuste perest. Millised valiksite jaotuse parameetrid?

Enamusel juhtudel on järeljaotuse leidmine keeruline ja analüütiliselt seda leida ei õnnestu:

- mudel on keeruline (palju parameetreid, keerulised funktsioonid);
- eelinfo esitatakse keeruliste jaotustega.

Näide (järg). Dünaamiline mudel:

$$\text{logit}(\pi_t) = \alpha_t + \beta_t x,$$

kus t – on nädal. Mudel muutub ajas: reklaaminäites mõjutab teadlikkust keskkond, so teiste suhtumine, konkurentide reklaamid jne. Loomulik on eeldada, et kõrvutiolevate ajamomentide mudelid on sarnasemed:

$$\begin{aligned}\alpha_t &= \alpha_{t-1} + \omega_{1t} \\ \beta_t &= \beta_{t-1} + \omega_{2t}\end{aligned}$$

Mudeli parameetrite arv on $2n$, kus n on vaadeldavate nädalate arv. Antud juhul on nii mudeli struktuur kui ka tundmatute parameetrite jaotus märksa keerulisem.

Bayesi statistikas ülesannete lahendamiseks kaks ligikaudsete meetodite klassi:

- determineeritud meetodid
 - analüütilised (piirjaotuste valemid),
 - numbrilised (keskväärtuste st integraalide leidmiseks),
- stohhastilised meetodid
 - keerulisest parameetrite ühisjaotusest on võimalik simuleerida väärtusi,
 - simuleeritud väärtuste abil saadakse valimijaotus, mis lähendab tundmatute parameetrite ühisjaotust,
 - valimikarakteristikud iseloomustavad parameetrite jaotuskarakteristikuid.

Stohhastilise simuleerimise meetodid on toetatud tõenäosusteooria tulemustega.

- Suurte arvude seadus järeljaotuse ja/või selle karakteristikute lähendamine muutub järjest paremaks, kui simuleeritud väärtuste arv kasvab.
- Tsentraalne piirteoreem – võimaldab lähendamise viga mõõta.

Bayesi statistika rakendusvaldkond on kasvanud tohutult viimase paarikümne aasta jooksul seoses spetsiaalsete simuleerimismeetodite kasutuselevõtuga – need on MCMC (Markov Chain Monte Carlo) meetodid.

Kursuse põhieesmärk: õppida tundma metoodikat Bayesi statistika tegemiseks stohhastilise simuleerimise abil, st MCMC meetodil:

- MCMC võimaldab simuleerida realisatsioone juhuslike suuruste ühisjaotusest, Bayesi rakenduses keeruliste mudelite tundmatute parameetrite ühisjaotusest.
- Realisatsioonid simuleeritakse Markovi ahelana, st järgmine väärtus sõltub ainult ahela eelmisest väärtusest.
- Markovi ahel koondub teatud tasakaaluolekusse, kus iga järgnev väärtus simuleeritakse juba ühest ja samast jaotusest. See jaotus on Markovi ahela piirjaotus.
- Tuleb konstrueerida selline Markovi ahel, et meid huvitav järeljaotus oleks Markovi ahela piirjaotus.
- Simuleerimine meie jaotusest tähendab siis Markovi ahela simuleerimist, kuni see on jõudnud tasakaaluolekusse.
- Statistiliste otsustused tehakse simuleeritud väärtustelt pärast tasakaaluolekusse jõudmist. Tuleb arvestada vaatluste korreleeritusega.

MCMC meetodeid on palju. Tähtsamad MCMC meetodid simuleerimiseks:

- Gibbsi valik (Gibbs sampling) – ühisjaotuse mitmemõõtmeline punkt simuleeritakse koordinaathaaval kasutades tinglikke jaotusi. Sõltuvus eelmisest elemendist Markovi ahelas on määratud tinglike jaotustega. Paljudel juhtudel on tinglikud jaotused suhteliselt lihtsad võrreldes ühisjaotusega.
- Metropolis-Hastingsi algoritm – kahesammuline algoritm (ettepanek, vastuvõtmine):
 - ettepanek soovib mingit järgmist väärtust trajektooris
 - vastuvõtmine kindlustab liikumise soovitud piirjaotuse suunas; ebasobiva liikumissuuna ettepanek lükatakse kõrvale. Otsustamine käib tõenäosuslikult.

Sõltuvus eelmisest elemendist Markovi ahelas tuleneb vastuvõtmistõenäosusest.

Metropolis-Hastingsi algoritm töötab ka siis, kui tinglikud jaotused on keerulised.

1 Bayesi statistika

Thomas Bayes (1701-1761) oli vaimuliku poeg Inglismaal. Ta õppis teoloogiat Edingurgh Ülikoolis. Töötas kirikuõpetajana (Presbyterian minister) Lõuna-Inglismaal. Ei abiellunud. Huvitus matemaatikast, kuid oma tulemusi ta ise ei publitseerinud. Ta pärandas 100 naela Richard Price'le, kes oli ka matemaatikahuviline kirikuõpetaja, kellelt Bayesi sugulased palusid, et ta Bayesi märkmed läbi uuriks. Tänu Price'le ilmusid publikatsioonid, mille läbi Bayesi tööd said tuntuks:

- Bayesi versioon tõenäosusteooriast.
- Binoomjaotuse parameetrite hindamisest - parameetrite võimalikke väärtusi võib kujutada pideva jaotuse abil (eel-, järeljaotus, tinglik tõenäosus).
- Beta jaotuse integraali leidmisest.

Üldkogumi ehk mudeli parameetrid ei ole konstandid. Neid ei saagi täpselt määrata. Parameetrite kohta saab anda üksnes tõenäosuslikke otsustusi. Parameetrid antakse jaotusega (selles mõttes juhuslikud).

- Info vaatlusandmetest kombineeritakse eelinfo parameetrite kohta (eeljaotus) läbi Bayesi teoreemi.
- Saadakse parameetrite järeljaotus.
- Interpretatsioonid parameetrite kohta (keskmine, vahemikud ...)
- Uute vaatluste tegemisel võib leitud järeljaotuse võtta parameetrite teadaolevaks eeljaotuseks.

1.1 Bayesi teoreem

Olgu $\mathbf{x} = (x_1, \dots, x_n)'$ juhuslik valim, vaatluste vektor. Tähistame vektori \mathbf{x} tihedusfunktsiooni (pideval juhul) ja tõenäosusfunktsiooni (diskreetsel juhul) $f(\mathbf{x}|\theta)$, kus θ on parameeter (võib olla ka vektor). Vektori \mathbf{x} komponendid x_i on sõltumatud.

Näide 1

*Kaks lähenemist θ hindamisel. θ - füüsikaline konstant
 ε_i - mõõtmisviga
 $\varepsilon_i \sim N(0, \sigma^2)$, kus σ^2 on teadaolev mõõtmistäpsus
 $x_i = \theta + \varepsilon_i$ on mõõdetud väärtus, kus $i = 1, \dots, n$.
 valimi tihedusfunktsioon on*

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^n f_N(x_i|\theta, \sigma^2) = \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - \theta)^2}{2\sigma^2}\right\} \end{aligned}$$

θ - ühelt poolt jaotuse parameeter, teiselt poolt meid huvitav füüsikaline konstant.

Eesmärgiks saada infot θ kohta

Sagedusstatistikud

Vaatlusandmed x_i on ainus allikas, mida kasutada statistiliste otsuste tegemisel θ kohta. θ suurima tõepära hinnangu saame maksimeerides valimi tihedusfunktsiooni θ suhtes. Selle tegevuse juures ei võta arvesse oma teadmist θ kohta.

Bayesi statistikud

Tuleb kasutada nii vaatlusandmeid x_i kui ka eelinfot θ kohta. On vaja mehhanismi, mis lubab eelinfo matemaatilisse avaldisse võtta. Selleks on Bayesi valem.

Ajalooliselt on olnud palju vaidlusi kahe lähenemisviisi pooldajate vahel. Taheti leida ainuõiget meetodit. Nüüd on leitud, et mõlemad lähenemisviisid on õigustatud ja kummalgi on oma koht probleemide lahendamisel.

Lõplik valik teooriate vahel sõltub nende rakendatavusest, Bayesi suund seni maha jäänud. MCMC tehnika lubab rakendada Bayesi statistikat küllalt keeruliste probleemidele.

Bayesi statistika plussid:

- Parameetrite hinnangud jaotusena võimaldab rohkem teada parameetrite kohta, ka täpsemalt teada (eelinfo);
- Võimaldab loomulikult prognoosida puuduvaid ja ka tulevikuväärtusi;
- Matemaatiliselt ilus ja ühtne lähenemisviis mudelite hindamiseks, valimiseks ja sobivuse kontrolliks;
- Paljude keeruliste statistika ülesannete puhul on Bayesi meetod ainuvõimalik ülesande matemaatilise esituse viis;
- Samuti paljude lihtsate, kuid eriliste statistika ülesannete puhul, näiteks harva esineva sündmuse tõenäosuse hindamine väikeselt valimilt (kõik 0-d);
- Kui eelinfot pole, siis ka selle saab Bayesi lähenemisel arvesse võtta ja saame sama tulemuse, mis klassikalises statistikas. Selles mõttes on klassikaline statistika Bayesi statistika erijuhuks.

Eelinfot θ kohta väljendame eeljaotuse abil:

$$p(\theta) - \theta \text{ eeljaotus,}$$

uurija subjektiivne eeldus parameetri kohta. Eeljaotus kirjeldab, millised θ väärtused on tõenäosemad ja millised vähem tõenäosed. Jaotus $p(\theta)$ võib omakorda sõltuda mingitest parameetritest, mida nimetatakse hüperparameetriteks.

Kodune ülesanne nr. 1.1 Pane kirja eeljaotus keskmisele peaümberrõõdule $\theta \sim N(\mu, \sigma^2)$, st määra hüperparameetrid μ ja σ^2 oma subjektiivse arvamus kohaselt. Mõttele, missugustes piirides Sinu arvates võiks keskmine olla suure tõenäosusega (näiteks 0.95), seejärel kasuta piiride jaoks teadmist $P(\mu - 2\sigma < \theta < \mu + 2\sigma) \approx 0.95$.

Hüperparameetrid on parameetrite jaotuse parameetrid. Hüperparameetrid on teadaolevad konstandid.

Bayesi statistikas on olemas kaks komponenti:

$f(\mathbf{x}|\theta)$ – vaatlusandmete jaotus,

$p(\theta)$ – parameetri eeljaotus.

Vaatluste vektor \mathbf{x} on fikseeritud konstant, θ on tundmatu.

Teame, et $f(\mathbf{x}|\theta)$ kui θ funktsioon on tõepärafunktsioon,

$$\ell(\theta) = f(\mathbf{x}|\theta).$$

Parameetri θ järeljaotuse leiame Bayesi teoreemi abil:

$$p(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)p(\theta)}{f(\mathbf{x})} \quad (1)$$

kus $f(\mathbf{x})$ on vektori \mathbf{x} marginaalne tihedusfunktsioon:

$$f(\mathbf{x}) = \int \underbrace{f(\mathbf{x}|\theta)p(\theta)}_{\mathbf{x} \text{ ja } \theta \text{ ühistihedusfunktsioon}} d\theta \quad (2)$$

$f(\mathbf{x})$ valemis (1) on normeeriv konstant, mis ei sõltu θ -st.

Tänu valemile (2) $\Rightarrow \int p(\theta|x) = 1$.

Järeljaotus $p(\theta|\mathbf{x})$ väljendab θ jaotust realiseerunud vaatluste \mathbf{x} korral.

Tähistame $\pi(\theta) \equiv p(\theta|\mathbf{x})$

Bayesi teoreemi alternatiivne kuju:

$$\pi(\theta) \propto \ell(\theta) \cdot p(\theta) \quad (3)$$

kus võrdeteguriks on normeeriv konstant $1/f(\mathbf{x})$.

NB! Simuleerimaks $\pi(\theta)$ -st ei ole vaja teada normeerivat konstanti, mis on suur pluss.

Bayesi teoreemi järjestikune rakendamine

Oletame, et \mathbf{x} on vaadeldud ja saadud järeljaotus $\pi_1(\theta)$. Oletame, et on saadud uued vaatlused \mathbf{y} , mis on ka seotud parameetriga θ . Seost kirjeldab oma tõepärafunktsioon $\ell_2(\theta)$. Võtame $\pi_1(\theta)$ eeljaotuseks ja rakendades Bayesi teoreemi, saame uue täpsustatud järeljaotuse:

$$\pi_2(\theta) = p(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi_1(\theta)}{f(\mathbf{y})} = \frac{f(\mathbf{y}|\theta)f(\mathbf{x}|\theta)p_1(\theta)}{f(\mathbf{x})f(\mathbf{y})} = \frac{\ell_2(\theta)\ell_1(\theta)p_1(\theta)}{f(\mathbf{x})f(\mathbf{y})}.$$

$$\pi_2(\theta) \propto \ell_2(\theta)\ell_1(\theta)p_1(\theta),$$

$$\pi_2(\theta) \propto \ell_2(\theta)\pi_1(\theta).$$

Näide 2

Järg Olgu parameetri θ (füüsikaline konstant) eeljaotus kujul:

$$p(\theta) = N(\mu, \tau^2),$$

kus μ ja τ^2 on eeljaotuse **hüperparameetrid**. Eeljaotuse interpretatsioon - 95%-lissetõenäosusega $\theta \in (\mu - 2\tau, \mu + 2\tau)$. Tõepärafunktsioon:

$$\begin{aligned} \ell(\theta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - \theta)^2}{2\sigma^2}\right\} \propto \\ &\propto \prod_{i=1}^n \exp\left\{-\frac{(x_i - \theta)^2}{2\sigma^2}\right\} = \\ &= \exp\left\{-\sum_{i=1}^n \frac{(x_i - \theta)^2}{2\sigma^2}\right\} = \\ &= \exp\left\{-\frac{\sum_{i=1}^n [(x_i - \bar{x})^2 - 2(x_i - \bar{x})(\bar{x} - \theta) + (\bar{x} - \theta)^2]}{2\sigma^2}\right\} = \\ &= \exp\left\{-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2} - \frac{n(\bar{x} - \theta)^2}{2\sigma^2}\right\} \end{aligned}$$

Kuna eksponentide korrutis $e^{-a-b} = e^{-a}e^{-b}$, siis viies θ -st mittesõltuv osa võrdetegurisse saame

$$\ell(\theta) \propto \exp\left\{-\frac{n(\bar{x} - \theta)^2}{2\sigma^2}\right\}.$$

Seega θ järeljaotus võtab kuju:

$$\pi(\theta) \propto \exp\left\{-\frac{n(\bar{x} - \theta)^2}{2\sigma^2}\right\} \exp\left\{-\frac{(\theta - \mu)^2}{2\tau^2}\right\}, \quad (4)$$

millest edasi teisendades,

$$\pi(\theta) \propto \exp\left\{-\frac{1}{2} \frac{(\theta - \mu_1)^2}{\tau_1^2}\right\} \quad (5)$$

kus

$$\tau_1^2 = \left[\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right]^{-1} \quad (6)$$

$$\mu_1 = \tau_1^2 \left[\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}\right]. \quad (7)$$

Järelikult θ järeljaotus on normaaljaotus

$$\pi(\theta) = N(\mu_1, \tau_1^2). \quad (8)$$

Järeljaotuse parameetrid μ_1 ja τ_1^2 sõltuvad vaatlustest ja eeljaotuse hüperperemeetritest. Juhul kui eeljaotuse dispersioon $\tau^2 \rightarrow \infty$, siis eeljaotus lameneb, $p(\theta) \propto k$ (konstant).

Eelinformatsioon θ kohta väheneb st. kõik θ sama tõenäosel.

Valemist (6) järeldub, et kui $\tau^2 \rightarrow \infty$ siis
$$\begin{cases} \tau_1^2 \rightarrow \frac{\sigma^2}{n} \\ \mu_1 \rightarrow \bar{x}. \end{cases}$$

Piiril on järeldaotus $\pi(\theta) = N(\bar{x}, \sigma^2/n)$. Võrdle klassikalise statistika tulemustega θ hindamisel! Seal on hinnanguks valimikeskmine \bar{x} ja selle dispersioon on σ^2/n .

Kodune ülesanne nr. 1.2 Tuleta (5)-(7). Alusta valemist (4), eralda θ -st mittesõltuv osa. See läheb võrdeteguriks.

Bayesi analüüsi esimesel sammul tuleb sageli eeldada mitteinformatiivset eeljaotust. Aga miks me ei tee siis klassikalist statistilist analüüsi?

- Keeruliste ülesannete korral pole meetodeid;
- Bayesi formuleerimine koos MCMC-simuleerimisega võimaldab seda ülesannet lahendada.

Näide 3

Järeldaotus Bernoulli/binoom mudelis.

Olgu vaatlus $x_i \sim B(1, \theta)$, $i = 1, \dots, n$. Siis $x = \sum_{i=1}^n x_i \sim B(n, \theta)$.

Näiteks, $x_i = 1$ (sünnib poisid), $x_i = 0$ (sünnib tüdruk), θ - poisi sündimise tõenäosus. Andmete tõepärafunktsioon:

$$\ell(\theta) \propto \theta^x (1 - \theta)^{n-x}.$$

Eeljaotus parameetrile (peab olema 0 ja 1 vahel):

$$p(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1} - \text{beeta}(\alpha, \beta).$$

Kirjuta tabelit kasutades välja beetajaotuse tihedusfunktsioon, keskvärtus ja dispersioon.

Lihtne on tuletada järeldaotust

$$\pi(\theta) \propto \theta^{x+\alpha-1} (1 - \theta)^{n-x+\beta-1} - \text{beeta}(x + \alpha, n - x + \beta) \text{ järeldaotus.}$$

Mis on järeldaotuse keskmine ja dispersioon?

Mis on mitteinformatiivne eeljaotus siin ja mis tuleb sel korral järeldaotuseks, keskmiseks ja dispersiooniks? Võrdle klassikalise statistika hinnanguga binoomjaotuse parameetrile.

1.2 Vaatluste prognoosimine

Lisaks järeldaotuse $\pi(\theta)$ leidmisele on Bayesi analüüsis teiseks huviojektiks vaatluste marginaaltihedus $f(x)$. Selle abil kontrollitakse Bayesi mudeli kooskõla reaalsusega. Seose (2) tõttu

$$f(x) = \int f(x|\theta)p(\theta)d\theta = E[f(x|\theta)],$$

mis on vaatluste oodatav jaotus, kus keskmine on võetud parameetrite eeljaotuse $p(\theta)$ suhtes. Sellisel viisil leitud $f(x)$ prognoosib vaatlusi arvestades meie eelteadmist parameetri kohta. Ta ei arvesta olemasolevaid vaatlusi. Vaatame, milline tihedusfunktsioon prognoosib uusi vaatlusi y , kui x on vaadeldud:

$$\begin{aligned} f(y|x) &= \int f(y, \theta|x)d\theta = \\ &= \int f(y|\theta, x)f(\theta|x)d\theta = \\ &=_{y \perp x | \theta} \int f(y|\theta)\pi(\theta)d\theta. \end{aligned}$$

Arvestasime, et y ja x on fikseeritud θ korral tinglikult sõltumatud.

Valimid samast jaotusest: $x = (x_1, \dots, x_n)'$ ja $y = (x_{n+1}, \dots, x_{n+m})'$ jaotusest $f(x|\theta)$ on tinglikult sõltumatud valimi eelduse tõttu. Prognoosiv tihedusfunktsioon on aluseks statistilisele prognoositeooriale (*predictive inference*), mis väidab:

- lõplik teooria õigsuse test on kooskõla reaalsusega;
- otsustused parameetrite kohta on kontrollimatud, sest parameetrid on ikkagi vaadeldamatud;
- aluseks peab olema vaatluste jaotus.

Näide 4

Järg. Prognoosiv jaotus normaalse mudeli korral.

Olgu $x_i|\theta \sim N(\theta, \sigma^2)$, $i = 1, \dots, n$, sõltumatud, θ tundmatu, σ^2 teada.

$p(\theta) = N(\mu, \tau^2)$ - eeljaotus

$\pi(\theta) = N(\mu_1, \tau_1^2)$ - järeldaotus, kus parameetrid μ_1 ja τ_1^2 on (7)ja (6).

Siis järgmise vaatluse prognoosiv jaotus on:

$y|x \sim f(y|x) = \int f(y|\theta)\pi(\theta)d\theta$, $x = (x_1, \dots, x_n)$. Kasutades siin $N(\theta, \sigma^2)$ ja $N(\mu_1, \tau_1^2)$ tihedusi, saame

$$f(y|x) = \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\theta)^2/2\sigma^2} \frac{1}{\sqrt{2\pi\tau_1^2}} e^{-(\theta-\mu_1)^2/2\tau_1^2} d\theta.$$

Paneme tähele, et integreeritav on 2-mõõtmeline tihedus $f(y, \theta)$ ja e-astmes on meil (y, θ) suhtes ruutfunktsioon, järelikult on integreeritav 2-mõõtmelise normaaljaotuse tihedus ning integreerides ühe argumendi välja, saame ühemõõtmelise normaaljaotuse tiheduse, järelikult $y|x \sim N(\cdot, \cdot)$.

Parameetrite leidmiseks kasutame seoseid:

$$\begin{aligned} E(u) &= E[E(u|v)], \\ Var(u) &= Var[E(u|v)] + E[Var(u|v)]. \end{aligned}$$

Väline operaator on v jaotuse suhtes. Meie juhul $u = y|x$, $v = \theta$. Tuletame:

Lõpuks saame, et

$$y|x \sim N(\mu_1, \sigma^2 + \tau_1^2)$$

on tulevikuvaatlusi prognoosiv jaotus:

- keskmine on sama, mis parameetri järeljaotusel $\pi(\theta) = N(\mu_1, \tau_1^2)$;
- dispersioon on esialgsete vaatluste omast suurem, sest θ on juhuslik, rohkem määramatust.
- Mitteinformatiivse eeljaotuse korral $\tau_1^2 = \sigma^2/n$ ja siis $y|x \sim N(\mu_1, \frac{n+1}{n}\sigma^2)$

Märkus: Prognoosiva jaotuse analüütilise kuju tuletamine on üldjuhul küllalt raske. Analüütilist kuju pole aga üldse vaja teada, et simuleerida prognoosivast jaotusest.

1. vaatenurk simuleerimaks tihedusest $f(y|x)$

Paneme tähele, et tegemist on jaotuste pideva seguga, mida väljendab integraal

$$y|x \sim f(y|x) = \int f(y|\theta)\pi(\theta)d\theta$$

(diskreetsel juhul, kus θ omandab lõpliku või loenduva arvu väärtusi, avaldub jaotuste segu summana, $\sum_i f(y|\theta_i)\alpha_i$, $\sum_i \alpha_i = 1$, $\alpha_i = \pi(\theta_i)$).

Esiteks simuleerime väärtuse θ jaotusest $N(\mu_1, \tau_1^2)$ ja siis y jaotusest $N(\theta, \sigma^2)$. Kordame vajalik arv kordi.

2. Vaatenurk

$$f(y|x) = \int f(y, \theta)d\theta = \int f(y|\theta)\pi(\theta)d\theta$$

Punkti (y, θ) simuleerimiseks ühisjaotusest

$$\begin{cases} \theta \leftarrow \pi(\theta) \\ y \leftarrow f(y|\theta) \end{cases}$$

integreerime üle θ annab y marginaaljaotuse. Simulatsioonid y marginaaljaotusest saame kui korjame kokku p -de $\binom{\theta}{y}$ simulatsioonidest y koordinaadid.

Näide 5

Järg Prognoosiv jaotus binoom-mudeli korral

Olgu $x = (x_1, \dots, x_n)$, $x_i|\theta \sim B(1, \theta)$, $i = 1, \dots, n$. Parameetri tähendus on tõenäosus saada väärtust 1. Olgu tulevikuvaatlus y . Kui θ on teada, siis $y|\theta \sim B(1, \theta)$. Kui θ ei ole teada, mis on siis $y|x$ prognoosiv jaotus?

Eeljaotus parameetrile $p(\theta) = \text{beta}(\alpha, \beta)$. Järeljaotus $\pi(\theta) = \text{beta}(\sum x_i + \alpha, n - \sum x_i + \beta)$.

Prognosis jaotus $y|x \sim f(y|x) = \int f(y|\theta)\pi(\theta)d\theta$. Siin $y \in \{0, 1\}$ ja $f(y|x) = P(\text{saada väärtus } y|x)$.

Leidkem $f(1|x)$ ja interpreteerige.

1.3 Mitmemõõtmeline parameeter Bayesi analüüsis

Mitmemõõtmelise parameetri $\theta = (\theta_1, \dots, \theta_d)$ korral saame rääkida

- ühisjäreljaotusest
 $\pi(\theta) = \pi(\theta_1, \dots, \theta_d)$
- parameetri marginaalsest jaotusest
 $\pi(\theta_i) = \int \pi(\theta_1, \dots, \theta_d)d\theta_{-i}$, kus $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$
- parameetri tinglikust jaotusest

$$\pi(\theta_i|\theta_j, j \in C) = \frac{\pi(\theta_i, \theta_j, j \in C)}{\pi(\theta_j, j \in C)},$$

$$C \subset \{1, \dots, i-1, i+1, \dots, d\}$$

- parameetri täistinglikust jaotusest
 $\pi(\theta_i|\theta_{-i}) \equiv \pi_i(\theta_i)$

Antud kursuse jaoks on olulisim parameetri θ_i täistinglik jaotus, mida tähistame $\pi_i(\theta_i)$.

1.4 Karakteristikute leidmine järeljaotusest

Parameetri θ järeljaotus $\pi(\theta)$ annab kogu info parameetri kohta. Sellest on võimalik tuletada mistahes karakteristikuid, mis kirjeldavad parameetri jaotust ja aitavad tulemust interpreteerida:

- Paiknemiskarakteristikud
 - keskväertus (aposterioorne keskväertus), $\pi(\theta)$ keskväertus
 - mood (aposterioorne mood)
 - mediaan (aposterioorne mediaan)
- Hajuvuskarakteristikud
 - dispersioon, θ posterioorne dispersioon ehk jaotuse $\pi(\theta)$ dispersioon
 - standardhälve
 - täpsus (precision), dispersiooni pöördväertus
 - kvartiilvahemik

Mitmemõõtmelisel juhul iseloomustab parameetervektorit $\theta = (\theta_1, \dots, \theta_d)$ mitmemõõtmeline jaotus. Olulist informatsiooni saadakse θ kovariatsioonimaatriksist

- maatriksi diagonaalelemendid on θ_i dispersioonid;
- diagonaalelementide ruutjuured on θ_i standardhälbed;
- kovariatsioonimaatriksi pöördmaatriks on täpsusemaatriks (precision matrix).

Paiknemis- ja hajuvuskarakteristikuid saab anda θ ühisjaotuse, marginaaljaotuste ja tinglike jaotuste kohta. Komponenti θ_i jaotus on θ marginaaljaotus, mis leitakse integreerimise abil, $\pi(\theta_i) = \int \pi(\theta_1, \dots, \theta_d) d\theta_{-i}$.

Marginaaljaotusest saadakse otsustused θ_i kohta, näiteks tõenäosusintervall:

C on $100(1 - \alpha)\%$ tõenäosusintervall θ_i -le, kui

$$\int_C \pi(\theta_i) d\theta_i = 1 - \alpha \quad (9)$$

Antud definitsioon ei määra intervalli üheselt. Seetõttu leitakse lühima pikkusega intervall C_0 so. intervall, mille punktid on suurema aposterioorse tihedusfunktsiooni väärtusega kui intervalli mittekuuluvad punktid.

Mitmemõõtmelise parameetri korral konstrueeritakse tõenäosusintervallid analoogiliselt valemile (9).

Enamus kirjeldavaid karakteristikuid saadakse $\pi(\theta)$ integreerimisel. Lihtsamatel juhtudel on see võimalik analüütiliselt. Praktilist huvi pakkuvatel keerukate mudelite korral on vaja ligikaudseid meetodeid. Siin aitavad simuleerimismeetodid, eriti MCMC.

NB! Mitmemõõtmelisejaotuse simulatsioonidest saadakse marginaaljaotuste simulatsioonid korjates kokku õiged koordinaadid. Seega integreerimist polegi vaja marginaaljaotuse leidmisel.

Märkus terminoloogia kohta. Klassikalises statistikas räägime usaldusintervallist (confidence interval) parameetrile, kusjuures parameeter on fikseeritud. Bayesi statistikas on meil parameetri jaotus ja räägime selle jaotuse intervallidest. Inglise keeles kasutatakse *credibility interval* (CI) või *probability interval*. Meie kasutame terminit tõenäosusintervall.

Näide 6

Järg. Järeldaotuse karakteristikud, mis on kasulikud tulemuse interpreteerimisel.

Leidsime, et järeldaotus $\pi(\theta) = N(\mu_1, \tau_1^2)$. Siit saame leida karakteristikud:

- θ keskväärts, mood, mediaan = μ_1 (Bayesi punkthinnang),
- θ standardhälve = τ_1
- θ täpsus = $1/\tau_1^2$
- $qnorm(0.025, \mu_1, \tau_1)$ ja $qnorm(0.975, \mu_1, \tau_1)$ annavad 95% tõenäosusintervalli.

Näide 7

Olgu andmed $x_i \sim N(\theta, \sigma^2), i = 1, \dots, n$. Olgu nüüd eeljaotuseks Cauchy jaotus samade hüperparameetritega, mis varasem normaaljaotus, μ ja τ^2 . Siis

$$p(\theta) \propto \frac{1}{\tau^2 + (\theta - \mu)^2}.$$

Järejaotus avaldub kujul:

$$\pi(\theta) \propto l(\theta)p(\theta) \propto \exp\left\{-\frac{1}{2} \frac{(\bar{x} - \theta)^2}{\sigma^2/n}\right\} \frac{1}{\tau^2 + (\theta - \mu)^2}$$

Saadud järeljaotust iseloomustab:

- Rohkem lihtsustada pole võimalik.
- Ühegi jaotuse analüütilist kuju pole võimalik ära tunda.
- Ühtki karakteristikut ei saa analüütiliselt leida.
- Küllaltki irregulaarne funktsioon, mis nõuab erilist hoolt numbriliste meetodite rakendamisel.

Üldiselt kasutatakse raskete sabadega eel-jaotust siis, kui eelinfo parameetrite kohta on ebakindel. Sel juhul saab suurema kaalu järeljaotuse määramisel andmete tõepära-funktsioon (vaatlusandmetel on suurem kaasarääkimisõigus).

1.5 Kaasjaotus (*Conjugate distribution*)

Eeljaotuse valik on Bayesi analüüsi tähtis komponent. Oma subjektiivset või varasemast teadaolevat infot parameetri kohta saab väljendada mitmesuguste jaotustega. Tuleb vaid hüperparameetrid määrata nii, et jaotuse põhimass asuks nende θ väärtuste kohal, mis meie arvates on kõige tõenäosemad. Osutub siiski, et mõned eeljaotused lihtsustavad oluliselt järeljaotuse tuletamist. Siit jõuame kaasjaotuse mõisteni.

Näites 2.1 oli parameetri eeljaotus $p(\theta) = N(\mu, \tau^2)$. Vaatluste $x = (x_1, \dots, x_n)'$ jaotus $x_i \sim N(\theta, \sigma^2)$. Parameetri järeljaotus $\pi(\theta) = N(\mu_1, \tau_1^2)$, kus

$$\tau_1^2 = \left[\frac{n}{\sigma^2} + \frac{1}{\tau^2} \right]^{-1} \quad \mu_1 = \tau_1^2 \left[\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2} \right]$$

Järelikult eeljaotus ja järeljaotus kuuluvad samasse klassi, normaaljaotuste peresse. Sel juhul statistiliste otsustuste tegemine lihtsustub tunduvalt. Üleminek eeljaotuselt järeljaotusele toob kaasa ainult hüperparameetrite muutuse ja mingeid täiendavaid ümberarvutusi pole vaja teha. Järeljaotust saab lihtsalt korrigeerida uute vaatluste lisandumisel.

Olgu saadud uued vaatlused $y = (y_1, \dots, y_m)'$ samast üldkogumist. Esimesel sammul saadud järeljaotuse võtame eeljaotuseks: $p_1(\theta) = N(\mu_1, \tau_1^2)$.

Uus järeljaotus on samuti normaaljaotus kujul: $\pi_1(\theta) = N(\mu_2, \tau_2^2)$, kus

$$\tau_2^2 = \left[\frac{m}{\sigma^2} + \frac{1}{\tau_1^2} \right]^{-1}, \quad \mu_2 = \tau_2^2 \left[\frac{m\bar{y}}{\sigma^2} + \frac{\mu_1}{\tau_1^2} \right]$$

DEFINITSIOON: Jaotuste pere P on **kaasjaotuste pere** vaatluste jaotusele F , kui iga eeljaotuse $p \in P$ ja iga vaatluste jaotuse $f \in F$ korral, järeljaotus $\pi \in P$.

Näiteks, normaaljaotuste pere on kaasjaotuste pere normaaljaotusega vaatlustele, aga beeta-jaotuste pere on kaasjaotuseks Bernoulli/binoomjaotusega vaatlustele.

1.5.1 Eksponentsiaalpere kaasjaotused (ühemõõtmeline juht)

Eksponentsiaaljaotuste pere tihedusfunktsiooni saab avaldada kujul:

$$f(x|\theta) = a(x) \exp\{\phi(\theta)t(x) + b(\theta)\} \quad (10)$$

Paljud enamkasutatavad jaotused kuuluvad eksponentsiaaljaotuste peresse:

a) Normaaljaotus $N(\theta, \sigma^2)$, kus σ^2 on teada

$$\phi(\theta) = \frac{\theta}{\sigma^2}, \quad t(x) = x, \quad b(\theta) = -\frac{\theta^2}{2\sigma^2}, \quad a(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$

b) Binoomjaotus $B(n, \theta)$; $f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$,

$$\phi(\theta) = \ln\left(\frac{\theta}{1-\theta}\right), \quad t(x) = x, \quad b(\theta) = n \ln(1-\theta), \quad a(x) = \binom{n}{x}$$

c) Eksponentjaotus $Exp(\theta)$, $f(x|\theta) = \theta e^{-x\theta}$

$$\phi(\theta) = -\theta, \quad t(x) = x, \quad b(\theta) = \ln(\theta), \quad a(x) = 1$$

d) Poissoni jaotus $Po(\theta)$, $f(x|\theta) = \exp(\theta) \frac{\theta^x}{x!}$

$$\phi(\theta) = \ln(\theta), \quad t(x) = x, \quad b(\theta) = -\theta, \quad a(x) = 1/x!$$

Tähtsad jaotuste pered, mis ei kuulu eksponentsiaaljaotuste peresse:

- ühtlane jaotus $U(0, \theta)$
- t-jaotus
- tiheduste segud

Juhusliku suuruse $t(X)$ keskväärtus ja dispersioon, kus X on jaotusega (10) avalduvad üldkujul järgmiselt:

$$E(t(X)) = -b'(\theta)/\phi'(\theta), \tag{11}$$

$$D(t(X)) = \frac{\phi''(\theta)b'(\theta) - \phi'(\theta)b''(\theta)}{(\phi'(\theta))^3}. \tag{12}$$

Kodune ülesanne nr. 1.3 Tuleta valemid (11)-(12) kasutades rekurrentset seost $E(t(X))^n$ jaoks. Eksponentsiaalpere korral on diferentseerimine parameetri järgi ja integreerimine vahetatavad operatsioonid.

Kodune ülesanne nr. 1.4 Leida (11)-(12) abil normaaljaotuse $E(X)$ ja $D(X)$.

Eksponentsiaalpere tihedusfunktsiooni (10) võib esitada erinevatel kujudel. Sageli antakse see kujul

$$f(x|\theta) = \exp\{\phi(\theta)t(x) + a(x) + b(\theta)\}.$$

Ka liikme $\phi(\theta)t(x)$ esitamiseks on palju võimalusi. Näiteks $\phi(\theta)t(x) = [c\phi(\theta)][(1/c)t(x)]$. Ühte esitust nimetatakse kanoonioliseks. Selle saamiseks tuleb teisendada nii parameetrit (ümberparametriseerimine) kui ka jaotuse väärtuspiirkonda (juhuslikku suurust ennast).

1.5.2 Tihedusfunktsiooni teisenemine juhusliku suuruse teisendamisel

Olgu vektori $x = (x_1, \dots, x_d)$ tihedusfunktsioon $f_x(x_1, \dots, x_d)$. Olgu $g(x_1, \dots, x_d) = (y_1, \dots, y_d)$ üksühene diferentseeruv teisendus, st leidub pöördteisendus $g^{-1}(y_1, \dots, y_d) = (x_1, \dots, x_d)$, siis vektori $y = (y_1, \dots, y_d)$, tihedusfunktsioon on

$$f_y(y_1, \dots, y_d) = f_x(g^{-1}(y_1, \dots, y_d))\mathbf{J}, \quad (13)$$

kus jakobiaan \mathbf{J} on osatuletiste matriksi determinandi absoluutväärtus:

$$\mathbf{J} = \text{abs} \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \dots & \frac{\partial x_d}{\partial y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_d} & \dots & \frac{\partial x_d}{\partial y_d} \end{vmatrix}$$

Sageli tuleb teisendusi teha ühemõõtmeliste juhuslike suuruste tihedusfunktsioonidega. Olgu vaatlus x jaotusega $f_x(x)$. Olgu teisendus $y = g(x)$, millel eksisteerib pöördteisendus $g^{-1}(y) = x$. Jakobiaan $\mathbf{J} = \text{abs} \left(\frac{dx}{dy} \right)$. Ühemõõtmelisel juhul saame valemi (13) erijuhu:

$$f_y(y) = f_x(g^{-1}(y)) \left| \frac{dx}{dy} \right|. \quad (14)$$

1.5.3 Eksponentsiaalpere kanooniline kuju

Olgu $\eta = \phi(\theta)$ uus parameeter. Olgu $y = t(x)$ 1 : 1 teisendus, siis $x = t^{-1}(y)$. Tihedus (10) teiseneb nüüd (14) abil kujule:

$$f(y|\eta) = a(t^{-1}(y)) \exp\left\{\eta y + b(\phi^{-1}(\eta))\right\} \left| \frac{dx}{dy} \right|$$

Jakobiaan on y -funktsioon ja läheb esimesse tegurisse. Saime eksponentsiaalpere kanoonilise kuju:

$$f(y|\eta) = a^*(y) \exp\left\{\eta y + b^*(\eta)\right\},$$

mis esialgsetes tähistes on:

$$f(x|\theta) = a(x) \exp\left\{\theta x + b(\theta)\right\}. \quad (15)$$

Kodune ülesanne nr. 1.5 Leida kanoonilise eksponentsiaalpere esindaja jaoks $E(X)$ ja $D(X)$.

Kaasjaotus. Olgu vaatluste jaotus $f(x|\theta)$ eksponentsiaalse pere jaotus kujul (15). Olgu parameetri θ eeljaotus kujul:

$$p(\theta) = k(\alpha, \beta) \exp\left\{\alpha\theta + \beta b(\theta)\right\}. \quad (16)$$

Leiame järeljaotuse pärast ühte vaatlust:

$$\begin{aligned}\pi(\theta) \propto f(x|\theta)p(\theta) &\propto \exp\{\theta x + b(\theta)\} \exp\{\alpha\theta + \beta b(\theta)\} = \\ &= \exp\{(\alpha + x)\theta + (\beta + 1)b(\theta)\} = \exp\{\alpha_1\theta + \beta_1 b(\theta)\},\end{aligned}$$

kus

$$\begin{aligned}\alpha_1 &= \alpha + x \\ \beta_1 &= \beta + 1\end{aligned}$$

Seega vaatluste jaotuse (15) korral on eeljaotus $p(\theta)$ ja järeljaotus $\pi(\theta)$ sama kujuga. Jaotus (16) on kaasjaotus eksponentsiaaljaotuste perele. Tähistame eeljaotuste klassi kujul (16) $CP(\alpha, \beta)$ (*conjugate prior*).

1.5.4 Näiteid eksponentsiaalpere kaasjaotuste kohta

Oleme juba näinud

- normaalsete vaatluste korral on keskväärtusparameetri hindamisel kaasjaotuste pereks normaaljaotuste pere;
- Bernoulli/binoom vaatluste korral on tõenäosusparameetri hindamisel kaasjaotuste pereks beeta-jaotuste pere.

Näide 8

Olgu vaatlus x Poissoni jaotusega, jaotuse parameeter λ aga gammajaotusega.

$$\begin{aligned}x &\sim Po(\lambda), & f(x|\lambda) &= e^{-\lambda} \frac{\lambda^x}{x!}, \\ \lambda &\sim G(\alpha, \beta), & p(\lambda) &\propto \lambda^{\alpha-1} e^{-\beta\lambda} \quad (= e^{-\beta\lambda + (\alpha-1)\ln\lambda})\end{aligned}$$

Järejaotuse leiame Bayesi teoreemist:

$$\begin{aligned}\pi(\lambda) &\propto e^{-\lambda} \lambda^x \lambda^{\alpha-1} e^{-\beta\lambda} = \\ &= \lambda^{\alpha+x-1} e^{-(\beta+1)\lambda} \\ \Rightarrow \pi(\lambda) &= G(\alpha + x, \beta + 1)\end{aligned}$$

Järeldus: Gamma-jaotuste pere on kaasjaotuseks Poissoni-jaotusega vaatlustele. Kui ühe vaatluse korral õnnestub näidata, et eeljaotus $p(\theta)$ ja järeljaotus $\pi(\theta)$ on samast klassist, siis Bayesi teoreemi rekursiivsel rakendamisel saab veenduda, et ka pärast n vaatlust on $p(\theta)$ ja $\pi(\theta)$ samast klassist.

Näide 9

1) x_1 vaatlemise järel

$$\begin{aligned}\pi(\lambda) &= G(\alpha_1, \beta_1) \\ \alpha_1 &= \alpha + x_1, \quad \beta_1 = \beta + 1\end{aligned}$$

2) x_2 vaatlemise järel

$$\pi(\lambda) = G(\alpha_2, \beta_2)$$

$$\alpha_2 = \alpha_1 + x_2 = \alpha + x_1 + x_2, \quad \beta_2 = \beta_1 + 1 = \beta + 2$$

...

n) x_n vaatlemise järel

$$\pi(\lambda) = G(\alpha_n, \beta_n)$$

$$\alpha_n = \alpha + \sum_{i=1}^n x_i, \quad \beta_n = \beta_1 + n$$

1.5.5 Mitmemõõtmelise eksponentsiaalpere kaasjaotused

Kui parameeter θ on vektor $\theta = (\theta_1, \dots, \theta_k)$ ja vaatlused moodustavad samuti vektori $x = (x_1, \dots, x_k)$, siis mitmemõõtmelise eksponentsiaalpere tihedusfunktsioon avaldub kujul:

$$f(x|\theta) = a(x) \exp\left\{\sum_{i=1}^k \theta_i x_i + b(\theta)\right\}. \quad (17)$$

Analoogiliselt ühemõõtmelisele juhule, otsime mitmemõõtmelisel juhul kaasjaotuste peret sarnasel kujul vaatluste jaotusega:

$$p(\theta) = k(\alpha, \beta) \exp\left\{\sum_{i=1}^k \alpha_i \theta_i + \beta b(\theta)\right\}, \quad (18)$$

kus normeeriv konstant k sõltub hüperparameetritest $\alpha = (\alpha_1, \dots, \alpha_k)$ ja β . Eeldame, et meil on mitmemõõtmeline vaatlus eksponentsiaalperest (17). Rakendades Bayesi teoreemi, saame:

$$\begin{aligned} \pi(\theta) &\propto f(x|\theta)p(\theta) \propto \exp\left\{\sum_{i=1}^k \theta_i x_i + b(\theta)\right\} \exp\left\{\sum_{i=1}^k \alpha_i \theta_i + \beta b(\theta)\right\} = \\ &= \exp\left\{\sum_{i=1}^k (x_i + \alpha_i) \theta_i + (\beta + 1)b(\theta)\right\} = \exp\left\{\sum_{i=1}^k \alpha_{1i} \theta_i + \beta_1 b(\theta)\right\} \end{aligned}$$

Järeldus: Jaotused (18) on kaasjaotuste pere mitmemõõtmelisele eksponentsiaaljaotusele (17).

Näide 10

Multinomiaaljaotus on mitmemõõtmelise eksponentsiaalpere esindaja ja tema kaasjaotus on Dirichlet jaotus. Olgu

$$x = (x_1, \dots, x_k) \sim M(n, \theta_1, \dots, \theta_k), \quad \sum_{i=1}^k \theta_i = 1, \quad \sum_{i=1}^k x_i = n.$$

Ühistihedusfunktsioon avaldub kujul:

$$f(x|\theta_1, \dots, \theta_k) = \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k \theta_i^{x_i}, \quad 0 \leq x_i \leq n.$$

Tihedusfunktsiooni saab esitada kujul (17). Tee seda!

Dirichlet jaotus:

$$f_D(\theta) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k \theta_i^{\alpha_i - 1}, \quad 0 < \theta_i < 1, \quad \sum_{i=1}^k = 1.$$

Bayesi teoreemist saame:

$$\pi(\theta_1, \dots, \theta_k) \propto \prod_{i=1}^k \theta_i^{x_i} \prod_{i=1}^k \theta_i^{\alpha_i - 1} = \prod_{i=1}^k \theta_i^{\alpha_i + x_i - 1}.$$

Seega ka järeljaotuseks tõenäosuste vektorile on Dirichlet jaotus.

Kodune ülesanne nr. 1.6 Veendu, et $k = 2$ korral taandub Dirilecht jaotus beeta jaotuseks. Leida marginaaljaotused $k = 3$ korral.

Näide 11

Eelnevalt näitasime, et normaaljaotuste pere on kaasjaotuseks normaaljaotusega vaatlustele. Vaatleme mitmemõõtmelist normaaljaotust. Olgu parameeter $\theta \sim N(\mu, B)$ ja vaatluste vektor $x|\theta \sim N(\theta, \Sigma)$. Saab näidata, et parameetri järeljaotus on kujul:

$$\pi(\theta) = N(\mu + B(\Sigma + B)^{-1}(x - \mu), B - B(\Sigma + B)^{-1}B)$$

Seega on ka mitmemõõtmelisel juhul normaaljaotuste pere kaasjaotuseks normaaljaotusega vaatlustele.

Kodune ülesanne nr. 1.7 Näidata, et ühemõõtmelisel juhul langeb kokku tulemusega (6)-(8).

1.6 Veel kasulikke jaotusi

1.6.1 Gamma-, pöördgamma- ja normaalgamma jaotused

Gammajaotus – Juhuslik suurus x on gammajaotusega $x \sim G(\alpha, \beta)$, kui tema tihedusfunktsioon avaldub kujul

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x \geq 0, \quad (19)$$

kus $\Gamma(\alpha)$ on gammafunktsioon.

Tingimusest $\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = 1$ saab tuletada gammafunktsiooni kuju:

$$\Rightarrow \Gamma(\alpha) = \int_0^\infty \beta^\alpha x^{\alpha-1} e^{-\beta x} dx.$$

Teeme muutuja vahetuse $\beta x = z \Rightarrow x = \frac{z}{\beta}, dx = \frac{dz}{\beta}$

$$\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} \beta e^{-z} \frac{dz}{\beta} = \int_0^\infty z^{\alpha-1} e^{-z} dz$$

Gammajaotuse omadused:

1. Kui $\alpha = \frac{n}{2}$ ja $\beta = \frac{1}{2}$, siis gammajaotus on χ^2 -jaotus vabadusastmete arvuga $n \Rightarrow$

$$G\left(\frac{n}{2}, \frac{1}{2}\right) = \chi_n^2$$

2. Kui $x \sim G(\alpha, \beta)$, siis $bx \sim G\left(\alpha, \frac{\beta}{b}\right)$

Kodune ülesanne nr. 1.8 Tõestada omadused 1. ja 2. Kasuta jakobiaanteisendust (14).

Pöördgammajaotus – Kui $x \sim G(\alpha, \beta)$, siis x^{-1} jaotust nimetatakse pöörgammajaotuseks:

$$\frac{1}{x} \sim IG(\alpha, \beta).$$

Tuletame tihedusfunktsiooni

$$y = \frac{1}{x} \Rightarrow x = \frac{1}{y}, \quad \frac{dx}{dy} = -\frac{1}{y^2}$$

$$f(y|\alpha, \beta) = f(x(y)|\alpha, \beta) \left| \frac{dx}{dy} \right| = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{y^{\alpha-1}} e^{-\frac{\beta}{y}} \frac{1}{y^2} = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{y^{\alpha+1}} e^{-\frac{\beta}{y}}$$

Normaalgamma jaotus – Läheb vaja Bayesi regressioonanalüüsis. Juhuslike suuruste (x_1, x_2) ühisjaotust nimetatakse normaalgamma jaotuseks parameetritega $\mu, \sigma^2, \alpha, \beta$, ja tähistatakse $(x_1, x_2) \sim NG(\mu, \sigma^2, \alpha, \beta)$, kui

$$\begin{aligned} x_1|x_2 &\sim N\left(\mu, \frac{\sigma^2}{x_2}\right) \\ x_2 &\sim G(\alpha, \beta) \end{aligned} \quad (20)$$

Normaalgammajaotuse tihedusfunktsioon tuleb üldtuntud seosest:

$$f(x_1, x_2) = f(x_1|x_2)f(x_2).$$

Kodune ülesanne nr. 1.9 Kirjuta välja $NG(\mu, \sigma^2, \alpha, \beta)$ tihedusfunktsioon.

Normaalgamma marginaaljaotused – Olgu $(x_1, x_2) \sim NG(\mu, \sigma^2, \alpha, \beta)$, kus

$$\begin{aligned} x_1|x_2 &\sim N\left(\mu, \frac{\sigma^2}{x_2}\right) \\ x_2 &\sim G(\alpha, \beta). \end{aligned}$$

Siin on x_2 marginaaljaotus teada aga x_1 oma mitte. Teame, et selliselt esitatud 2-mõõtmelist jaotust on lihtne simuleerida ja simulatsioonidelt saaksime aimu ka x_1 marginaaljaotusest (praktikumis).

Teoreetiliselt on suhteliselt lihtsalt leitavad x_1 marginaalne keskväärtus ja dispersioon:

$$E(x_1) = E_{x_2}(E(x_1|x_2)) \underset{const}{=} \mu \quad (\text{sama, mis tinglikel})$$

$$Var(x_1) = \underbrace{Var_{x_2}(E(x_1|x_2))}_{\mu} + E_{x_2}(Var(x_1|x_2)) = \sigma^2 E\left(\frac{1}{x_2}\right) = \sigma^2 \frac{\beta}{\alpha - 1}, \quad \alpha > 1$$

$$x_2 \sim G(\alpha, \beta)$$

$$\frac{1}{x_2} \sim IG(\alpha, \beta)$$

Seega x_1 jaotust lähendav normaaljaotus võiks olla

$$N\left(\mu, \frac{\sigma^2 \beta}{\alpha - 1}\right)$$

Kasutame praktikumis regressiooni juures.

1.6.2 Normaaljaotuse marginaaljaotused ja tinglikud jaotused

Olgu vektor x mitmemõõtmelise normaaljaotusega:

$$x = (x_1, \dots, x_d)' \sim N(\mu, \Sigma)$$

$$f(x|\mu, \Sigma) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x - \mu)' \Sigma^{-1} (x - \mu)\right\}$$

- standardnormaaljaotuse saame, kui $\mu = 0$ ja $\Sigma = I_d$, sel juhul komponendid x_i on sõltumatud standardse normaaljaotusega juhuslikud suurused;
- ühemõõtmelise normaaljaotuse saame, kui $d = 1$.

Lineaarteisendus – Olgu $x \sim N_d(\mu, \Sigma)$. Siis $r \times d$ maatriksi A ja r -mõõtmelise vektori b korral,

$$y = Ax + b \sim N_r(A\mu + b, A\Sigma A')$$

$$Ey = A\mu + b, \quad Dy = E(y - Ey)(y - Ey)' = A\Sigma A'$$

Kasutatakse sõltuvustega normaaljaotuse genereerimiseks. Olgu $x \sim N_d(0, I)$. Oskame genereerida. Tahame saada $y \sim N(m, B)$. Teame $Bx + m \sim N(m, BB')$. Tingimusest $BB' = \Sigma$ määrame B :

- $B = U\Lambda^{1/2}$, kus U ja Λ on vastavalt B omavektorid ja omaväärtused (diagonaalil);
- B on Choleky lahutuse kolmnurk maatriks.

Marginaaljaotus – Normaaljaotuse marginaaljaotused on normaaljaotused. Olgu vektor $\mathbf{x} = (x_1, \dots, x_d)' \sim N(\mu, \Sigma)$ jagatud kahte blokki \mathbf{x}_1 ja \mathbf{x}_2 vastavalt suurusega d_1 ja $d_2 = d - d_1$. Samuti jaotame parameetrid μ ja Σ

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{ja} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Siit marginaaljaotused: $\mathbf{x}_i \sim N_{d_i}(\mu_i, \Sigma_{ii})$, $i = 1, 2$.

Tinglik jaotus - Eeldades samasugust \mathbf{x} , μ ja Σ lahutust kaheks osaks, saame tingliku jaotuse kujul:

$$\mathbf{x}_1 | \mathbf{x}_2 \sim N_{d_1}(\mu_{1.2}, \Sigma_{11.2})$$

$$\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2)$$

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Kodune ülesanne nr. 1.10 Kirjutada siit kahemõõtmelise normaaljaotuse tinglik jaotus.

Analoogilise tulemuse saab välja kirjutada $\mathbf{x}_2 | \mathbf{x}_1$ jaoks, vahetades kõigis indeksites väärtused 1 ja 2.

Siin peavad maatriksid Σ_{11} ja Σ_{22} olema täisastakuga, muidu nad ei oleks pööratavad.

Ühisjaotuse konstrueerimine tinglike jaotuste abil - Ühisjaotus avaldub tingliku ja marginaaljaotuse kaudu:

$$f(\mathbf{x}_1, \mathbf{x}_2) = f(\mathbf{x}_1 | \mathbf{x}_2)f(\mathbf{x}_2).$$

Kui

$$\mathbf{x}_1 | \mathbf{x}_2 \sim N_{d_1}(\mu_1 + B_1(\mathbf{x}_2 - \mu_2), B_2),$$

$$\mathbf{x}_2 \sim N_{d_2}(\mu_2, \Sigma_{22}),$$

siis \mathbf{x} ühisjaotus avaldub:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim N_d \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right], \quad \Sigma_{11} = B_2 + B_1\Sigma_{22}B_1', \quad \Sigma'_{21} = \Sigma_{12} = B_1\Sigma_{22}$$

1.7 Kaasjaotused regressioonimudeli korral

Olgu $y = (y_1, \dots, y_n)$ vaatluste vektor, kus y_i on sõltumatud ja normaaljaotusega. Regressioonimudel on

$$y_i | \beta, \sigma^2 \sim N(x_{i1}\beta_1 + \dots + x_{id}\beta_d, \sigma^2), \quad i = 1, \dots, n,$$

kus x_{i1}, \dots, x_{id} on i -nda vaatluse seletavate tunnuste väärtused ja β_1, \dots, β_d on regressioonkordajad. Mudeli saab kirjutada maatrikskujul:

$$y | \beta, \sigma^2 \sim N(X\beta, \sigma^2 I_n), \quad (21)$$

kus

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{pmatrix} \quad \text{ja} \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix}.$$

Üldisemal juhul y_i dispersioonid erinevad $D(y) = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$. Meie vaatleme juhtu, kus σ^2 on konstantne. Eesmärgiks on otsustused (β, σ^2) kohta.

Dispersiooniparameetri σ^2 asemel vaadeldakse Bayesi analüüsis sageli tema pöördväärtust, mida nimetatakse täpsuse parameetriks, $\phi = 1/\sigma^2$. Täpsuse parameetri abil näeb regressioonimudel välja nii

$$y | \beta, \phi \sim N(X\beta, \phi^{-1} I_n), \quad (22)$$

Eeljaotuse $p(\beta, \phi)$ ja vaatluste jaotuse (22) abil tuleb leida järeljaotus $\pi(\beta, \phi)$.

Regressioonimudeli (22) korral eksisteerib kaasjaotuste pere (β, ϕ) jaotuse kirjeldamiseks, see on normaalgamma jaotuste pere:

$$\beta | \phi \sim N(b_0, \phi^{-1} B_0) \quad \text{ja} \quad \phi \sim G\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right).$$

Sellisel kujul antud gammajaotuse parameetrid lihtsustavad hilisemat tuletamist. Interpretamiseks pangem tähele, et S_0 aitab ϕ jaotuse keskpunkti määratleda $E\phi = \frac{1}{S_0}$. Suurus n_0 aitab reguleerida dispersiooni. Mida suurem on n_0 , seda väiksem ϕ dispersioon:

$$D\phi = \frac{n_0/2}{(n_0/2)^2 S_0^2} = \frac{2}{n_0} (E\phi)^2$$

Eeljaotuseks parameetrite vektorile on tihedusfunktsioonide korrutis:

$$\begin{aligned} p(\beta, \phi) &= p(\beta | \phi) p(\phi) \propto \\ &\propto |\phi B_0^{-1}|^{1/2} \exp\left\{-\frac{1}{2}(\beta - b_0)' \phi B_0^{-1}(\beta - b_0)\right\} \phi^{(n_0/2)-1} \exp\left\{-\frac{n_0 S_0}{2} \phi\right\} \propto \\ &\propto \phi^{d/2} \exp\left\{-\frac{\phi}{2}(\beta - b_0)' B_0^{-1}(\beta - b_0)\right\} \phi^{(n_0/2)-1} \exp\left\{-\frac{n_0 S_0}{2} \phi\right\} \end{aligned} \quad (23)$$

Tõepärafunktsioon andmete jaotusele tuleneb eeldusest (22):

$$\begin{aligned} \ell(\beta, \phi) &= f(y | \beta, \phi) \propto \\ &\propto |\phi I_n|^{1/2} \exp\left\{-\frac{1}{2}(y - X\beta)' \phi I_n (y - X\beta)\right\} = \phi^{n/2} \exp\left\{-\frac{\phi}{2}(y - X\beta)'(y - X\beta)\right\} \\ &= \phi^{n/2} \exp\left\{-\frac{\phi}{2}[Q(\beta) + S_e]\right\} \end{aligned} \quad (24)$$

kus

$$\begin{aligned} Q(\beta) &= (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \\ \hat{\beta} &= (X' X)^{-1} X' y \\ S_e &= (y - X \hat{\beta})' (y - X \hat{\beta}) = \sum_{i=1}^n e_i^2 \\ e_i &= y_i - (x_{i1} \hat{\beta}_1 + \dots + x_{id} \hat{\beta}_d) \end{aligned}$$

$\hat{\beta}$ on suurima tõepära hinnang β -le, maksimiseerib tõepärafunktsiooni.

Kodune ülesanne nr. 1.11 Tee läbi teisendused (24) ja (25) saamiseks (vihje valemi (24) saamiseks lisa $\mp X \hat{\beta}$).

Tulemuste (23) ja (24) abil saame leida järeldaotuse:

$$\begin{aligned} \pi(\beta, \phi) &\propto \ell(\beta, \phi) p(\beta, \phi) \propto \\ &\propto \phi^{d/2} \exp\left\{-\frac{\phi}{2}(\beta - b_1)' B_1^{-1}(\beta - b_1)\right\} \phi^{(n_1/2)-1} \exp\left\{-\frac{\phi}{2} n_1 S_1\right\} \end{aligned} \quad (25)$$

kus

$$\begin{aligned} n_1 &= n_0 + n \\ n_1 S_1 &= n_0 S_0 + (y - X b_1)' y + (b_0 - b_1)' B_0^{-1} b_0 \\ b_1 &= B_1 (B_0^{-1} b_0 + X' y) \\ B_1^{-1} &= B_0^{-1} + X' X \end{aligned}$$

Järeldaotus (25) ja eeldaotus (23) on sama kujuga, tegemist on normaalgamma jaotusega. See jaotuste pere on kaasjaotusteks regressioonimudelile (21) ehk andmete y_i jaotusele. Seos (25) teisiti:

$$\beta | \phi \sim N(b_1, \phi^{-1} B_1) \quad \text{ja} \quad \phi \sim G\left(\frac{n_1}{2}, \frac{n_1 S_1}{2}\right), \quad (26)$$

ehk σ^2 kaudu:

$$\beta | \sigma^2 \sim N(b_1, \sigma^2 B_1) \quad \text{ja} \quad \sigma^2 \sim IG\left(\frac{n_1}{2}, \frac{n_1 S_1}{2}\right),$$

Seos (25) annab meile parameetrite ühisjaotuse. Teame ϕ marginaaljaotust, kuid mitte vektori β oma. Eespool tuletasime marginaalse keskväärtuse ja dispersiooni 1-dimensionaalsete suuruste korral. Tulemus üldistub lihtsalt mitmemõõtmelise β jaoks.

$$E\beta = E[E(\beta | \phi)] = b_1$$

$$D\beta = E[D(\beta | \phi)] = E[\phi^{-1} B_1] = E(\phi^{-1}) B_1 = \frac{n_1 S_1}{n_1 - 2} B_1.$$

Keskväärtuse $E(\phi^{-1})$ leidmisel kasutasime teadmist, et ϕ^{-1} on pöördgamma jaotusega. Ka vektori β tinglikest jaotustest saab kasulikke statistilisi otsustusi teha vektori β kohta, näiteks fikseerides ϕ kõige tõepärasema väärtuse. Ei tea aga näiteks β -vektori ühe komponendi jaotust. Selleks oleks meil vaja seost (25) integreerida üle kõigi ülejäänud muutujate. Siin tuleb appi simuleerimistehnika:

1. simuleerime ϕ gammajaotusest;
2. simuleerime $\beta|\phi$ mitmemõõtmelisest normaaljaotusest;
3. kordame palju kordi, järeldused teeme simuleeritud arvude pealt (iga komponendi jaoks on oma simuleeritud jada)

Üldisematel juhtudel, kus klassikalise regressioonimudeli eeldused ei ole täidetud, ei leidu kaasjaotuste peret ja ainsaks võimaluseks jääb simuleerimine seosest $\pi(\beta, \phi) \propto \ell(\beta, \phi)p(\beta, \phi)$.

Ka **üldistatud lineaarsete** mudelite korral ei leidu kaasjaotuste peret. Andmemudel sel korral:

$$f(y_i|\theta_i) = a(y_i) \exp\{y_i\theta_i + b(\theta_i)\}, \text{ eksponentsiaalne pere}$$

$$E(y_i|\theta_i) = -b'(\theta_i) = \mu_i$$

$$g(\mu_i) = x_{i1}\beta_1, \dots, x_{id}\beta_d, \text{ } g \text{ on seose- ehk linkfunktsioon.}$$

Näide 12

Olgu vaatlused binoomjaotusega $y_i|n_i, \pi_i \sim B(n_i, \pi_i)$, $E(y_i|n_i, \pi_i) = n_i\pi_i$. Siis kasutatavad seosefunktsioonid on:

- $\pi_i = \Phi(\alpha + \beta x_i)$, $\Phi^{-1}(\pi_i) = \alpha + \beta x_i$, *probit*;
- $\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta x_i$, *logit*;
- $\ln(-\ln(1 - \pi_i)) = \alpha + \beta x_i$, *log-log*.

Mitme seletava tunnuse korral asendub $\alpha + \beta x_i$ summaga $\sum_{k=1}^d \beta_k x_{ik}$.

Olgu vaatlused Poissoni jaotusega $y_i|\lambda_i \sim Po(\lambda_i)$, $E(y_i|\lambda_i) = \lambda_i$, siis kasutatakse seosefunktsiooni:

$$\ln(\lambda_i) = \sum_{k=1}^d \beta_k x_{ik}$$

Bayesi statistika jaoks on probleem $\beta = (\beta_1, \dots, \beta_d)$ jaotus. Puudub kaasjaotuste klass, nii et $p(\beta)$ ja $\pi(\beta)$ oleksid samast jaotuste klassist. Pole võimalik saada analüütilisi tulemusi järeljaotuse $\pi(\beta)$ jaoks ja teha statistilisi otsustusi.

1.8 Tinglikud kaasjaotused (conditional conjugacy)

Keeruliste mudelite korral on raske leida kaasjaotust. Vahel õnnestub parameetervektor jagada osadeks:

$$\theta = \underbrace{(\theta_1, \dots, \theta_{\text{I}})}_{\theta_{\text{I}}} \underbrace{(\dots, \theta_d)}_{\theta_{\text{II}}}$$

nii et leidub kaasjaotuste pere tinglikele jaotustele $\theta_{\text{I}}|\theta_{\text{II}}$ ja $\theta_{\text{II}}|\theta_{\text{I}}$.

Vaatame regressioonimudelit

$$y \sim N(X\beta, \sigma^2 I_n)$$

Parameetrite vektor on (β, ϕ) , kus $\phi = 1/\sigma^2$. Nägime, et tinglik jaotus $\beta|\phi \sim N(.,.)$ on kaasjaotus. Vektori β tinglik eeljaotus ja tinglik järeljaotus on samas klassis. Normaalkaasjaotuste klass on tinglikus mõttes kaasjaotuste klassiks regressioonimudelid.

DEFINITSIOON: Parameetervektori $\theta = (\theta_1, \dots, \theta_d)$ komponendil θ_i (skalaar või vektor) on tingliku kaassuse omadus, kui täistinglik eeljaotus $p_i(\theta_i) = p(\theta_i|\theta_{-i})$ ja täistinglik järeljaotus $\pi_i(\theta_i) = \pi(\theta_i|\theta_{-i})$ kuuluvad samasse jaotuste peresse.

Tinglik kaassus on olemas paljude keeruliste mudelite korral. Samas on ka parameetrite tinglikke jaotusi uurijal lihtsam eeljaotustena ette anda kui kõigi parameetrite ühisjaotust. Tinglikku kaassust kasutatakse simuleerimisel (Gibbsi valik).

1.9 Hierarhilised mudelid

Olgu üldkogumis d gruppi. Grupi i keskmine olgu β_i . Igas grupis on n_i vaatlust:

$$y_{ij} \sim N(\beta_i, \sigma^2), \quad j = 1, \dots, n_i; \quad i = 1, \dots, d.$$

Selline mudel on regressioonimudeli erijuht vaatluste vektoriga

$$y = (y_{11}, \dots, y_{1n_1}, \dots, y_{d1}, \dots, y_{dn_d})'$$

ja plaanimaatriksiga $X = \text{diag}(1_{n_1}, \dots, 1_{n_d})$, kus 1_m on m -dimensionaalne 1-de vektor. Regressioonkordajate vektor on $\beta = (\beta_1, \dots, \beta_d)'$.

Mudeli $y \sim N(X\beta, \sigma^2 I_n)$ Bayesi analüüsiks on vaja määrata β eeljaotus. Üks võimalus on eeldada, et β_1, \dots, β_d on sõltumatud ja anname ette β_i jaotused. Puuduseks see, et liiga palju eeljaotusi ja sõltumatus ei pruugi ka kehtida.

Teine võimalus on defineerida eeljaotus hierarhiliselt:

I tase – eeldatakse, et β_1, \dots, β_d on normaaljaotusega $\beta_i \sim N(\mu, \tau^2)$;

II tase – fikseeritakse eeljaotus parameetritele (μ, τ^2) , kus μ on üldkeskmine ja τ iseloomustab varieeruvust üldkeskmise ümber. Nüüd on ainult kahele parameetrile vaja anda eeljaotus ja pealegi on olemas interpretatsioon.

I tasemel $\beta|\mu, \tau^2 \sim N(1_d\mu, \tau^2 I_d)$

II tasemel (eeljaotuste fix.) $\mu \sim N(b_0, B_0)$

$$\sigma^2 \sim F_\sigma, \quad \tau^2 \sim F_\tau, \quad \mu, \sigma^2, \tau^2 \text{ on sõltumatud.}$$

Parameetrite eeljaotus läbi kahe taseme:

$$\begin{aligned} p(\beta, \mu, \sigma^2, \tau^2) &= p(\beta|\mu, \sigma^2, \tau^2)p(\mu)p(\sigma^2)p(\tau^2) = \\ &= p(\beta|\mu, \tau^2)p(\mu)p(\sigma^2)p(\tau^2). \end{aligned}$$

Teise taseme võib omakorda jagada kaheks tasemeks, et anda ka eeljaotused hierarhiliselt. Näiteks üks võimalus on anda regressioonimudel $y \sim N(X\beta, \sigma^2 I_n)$ hierarhiliselt kolmel tasemel:

$$\begin{aligned} y|\beta_1, \phi &\sim N(X_1\beta_1, \phi^{-1}I_n) \\ \beta_1|\beta_2 &\sim N(X_2\beta_2, C) \\ \beta_2 &\sim N(b_0, B_0) \\ \phi &\sim G\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right) \end{aligned} \quad (27)$$

Siin X_1 on plaanimaatriks ja β_1 regressioonkordajate vektor, β_2 on regressioonkordajate vektor β_1 kirjeldamiseks ja X_2 vastav plaanimaatriks. Kõigi mudelis olevate juhuslike suuruste ühisjaotus:

$$p(y, \beta_1, \beta_2, \phi) = p(y|\beta_1, \phi)p(\beta_1|\beta_2)p(\beta_2)p(\phi)$$

β_1 ja ϕ järeljaotust pole analüütiliselt võimalik leida. Mudel (27) taandub regressioonimudeli Bayesi analüüsile, kui vaatame mudelit tinglikult fikseeritud β_2 korral. Siis kaob rida $\beta_2 \sim N(b_0, B_0)$ mudelist. Tulemused on samuti tõlgendatavad vaid tinglikult fikseeritud β_2 korral. Tähistades $X_2\beta_2 = b_0$ saame analoogilise mudeli sellega, mis on (21).

Saab näidata, et parameetrite täistinglikud järeljaotused on samas klassis eeljaotustega

$$\begin{aligned} \pi(\beta_1|\beta_2, \phi) &= N \\ \pi(\beta_2|\beta_1, \phi) &= N \\ \pi(\phi|\beta_1, \beta_2) &= G \end{aligned}$$

Need on analüütiliselt leitavad ja neile saab rajada statistilise otsustamise. Otsused on tinglikud. Marginaalsed jaotused saab leida MCMC-ga.

Üldistatud lineaarse mudeli saab ka anda hierarhiliselt:

$$\begin{aligned} y_i|\mu_i &\sim EF(\mu_i), \quad \mu_i = E(y_i|\mu_i), \\ \eta &= X_1\beta_1, \quad \eta = (\eta_1, \dots, \eta_n)', \quad i = 1, \dots, n, \quad \eta_i = g(\mu_i) \\ \beta_1|\beta_2 &\sim N(X_2\beta_2, C) \\ \beta_2 &\sim N(b, B) \end{aligned}$$

EF – eksponentsiaaljaotuste pere (*exponential family*).

Näide 13

Uuriti spetsiaalsete õppeprogrammide (*coaching programmes*) mõju testi tulemustele 8 koolis. Testiks oli õppesobivustest SAT-V (*Scholastic Aptitude Test-Verbal*), mida kasutavad kolledzid tudengite vastuvõtmisel. Saadavad punktid saavad varieeruda 200-800 vahel, keskmisega 500 ja st-hälbega umbes 100. Igas koolis osales umbes 30 õpilast.

μ – õppeprogrammi üldkeskmine mõju;

β_i – keskmine mõju i -ndas koolis (kooli mõju) $\beta \sim N(\mu, \tau^2)$;

τ^2 – kooli mõjude vaheline dispersioon.

1.10 Mudeli sobivus

Bayesi mudel on

$$f(y, \theta) - \text{andmete ja parameetrite ühisjaotus.}$$

Bayesi mudeli komponendid:

$$\begin{aligned} f(y|\theta) &- \text{andmete jaotus;} \\ p(\theta) &- \text{parameetrite eeljaotus.} \end{aligned}$$

Siit saadakse

$$\begin{aligned} \pi(\theta) &\propto \underbrace{f(y|\theta)}_{\ell(\theta)} p(\theta) - \text{parameetrite järeljaotus;} \\ f(y) &= E_{\pi}[f(y|\theta)] - \text{prognoosiv jaotus.} \end{aligned}$$

Kui hästi kirjeldab mudel andmeid ja meie eelteadmist parameetrite kohta? Need otsustused on rajatud põhiliselt prognoosivale jaotusele. Kui otsuseks on, et ei kirjelda hästi, siis ei sobi kas andmete mudel või eeljaotus vms.

1.10.1 Mudeli sobivus prognoosiva jaotuse abil

Eeldame mitteinformatiivset eeljaotust, st üksnes andmed määravad parameetrite järeljaotuse. Tahame kontrollida, kas prognoosiv jaotus prgnoosib olemasolevatega sarnaseid andmeid. Olgu $y = (y_1, \dots, y_n)$ andmestik ja tähistame

$$y^{rep} = (y_1^{rep}, \dots, y_n^{rep})$$

replikaat e. prognoositud andmestikku. Uut andmestikku kirjeldab prognoosiv tihedus:

$$f(y^{rep}|y) = \int \underbrace{f(y^{rep}|\theta)\pi(\theta)}_{f(y^{rep},\theta)} d\theta = E_{\pi}[f(y^{rep}|\theta)], \quad (28)$$

kus fikseeritud θ korral on tihedus $f(y^{rep}, \theta)$ sama, mis tihedus $f(y, \theta)$. Tihedusest (28) simuleerime andmestiku y^{rep} ja teeme neid mitu komplekti.

Bayesi mudel sobib, kui andmestikud y ja y^{rep} on sarnased. Andmestike võrdlemisel uuritakse erinevaid aspekte:

- vastavate empiiriliste jaotuste paiknemist;
- empiiriliste jaotuste hajuvust;
- sümmeetria omadusi;
- min, max elemente;
- ...

Kui mingi aspekti osas y^{rep} komplektid erinevad süstemaatiliselt, siis mudel ei sobi antud aspekti kirjeldamiseks. Lisaks visuaalsele vaatlusele (histogramm), võib võrrelda andmetelt ja replikaatidelt arvutatud suurusi – Bayesi teststatistik.

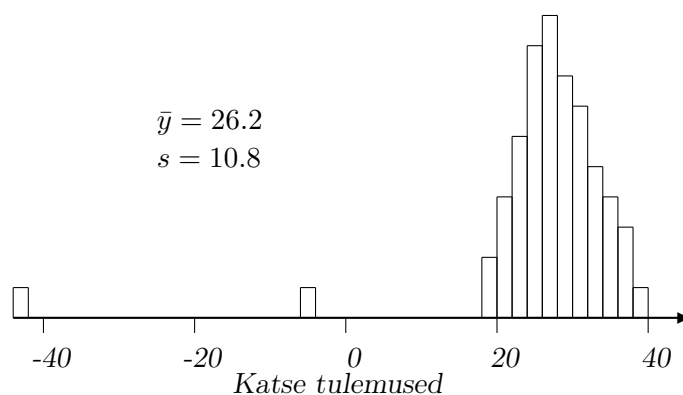
Näide 14

Simon Newcomb 1882 – valguse kiiruse mõõtmine.

Vahemaa, mis katse läbiviimiseks võeti, oli 7442m. Newcomb mõõtis aega, mis kulus valgusel antud vahemaa läbimiseks 66 korda. Ta tegi mõõtmistulemused (nanosekundites) käepärasemaks teisendusega

$$y_i = y_i^{mdetud} - 24800.$$

andmeid kirjeldas histogramm.



Eeldades mõõtmistulemustele tavapärasest normaaljaotust,

$$y_i \sim N(\mu, \sigma^2) \text{ ehk } y_i \sim N(\mu, \phi), \text{ kus } \phi = 1/\sigma^2,$$

tekib küsimus selle eelduse paikapidavusest. Kas prognoosivast jaotusest on võimalik saada selliseid andmeid?

Võttes (μ, ϕ) eeljaotuseks normaalgamma jaotuse, on ka järeljaotuseks normaalgamma,

$$\pi(\mu, \phi) = \pi(\mu|\phi)\pi(\phi),$$

kus $\pi(\mu|\phi)$ on teatav normaaljaotus ja $\pi(\phi)$ on teatav gammajaotus.

Replikaatandmestiku simuleerime kasutades avaldist (28):

1. $k = 1$ (k – korduse number);
2. Simuleerime $(\mu, \phi) \leftarrow \pi(\mu, \phi)$;
3. Saadud vektoriga (μ, ϕ) simuleerime vektori y^{rep} , st $y_i^{rep} \leftarrow N(\mu, \sigma^2)$, $i = 1, \dots, 66$;
4. $k = k + 1$ ja kordame 2. ja 3. sammu.

Paneme tähele, et prognoositud andmestikud erinevad reaalsest (Gelman jt 2004, lk 160-161). Uurime vähimat väärtust -44 (reaalne). Vähimate väärtuste jaotus 20 prognoositud andmestiku põhjal näitab, et need on kaugel -44-st ja normaalne mudel ei kirjelda seda varieeruvust, mis on andmetes. Seega ka meid huvitavad valguse kiiruse näitajad (μ, ϕ) ei ole usaldusväärsed.

1.10.2 Bayesi teststatistik

Klassikaline teststatistik tohib sõltuda vaid valimist ja mitte parameetritest. Bayesi teststatistik sõltub aga mõlemast.

DEFINITSIOON: Bayesi teststatistik on skalaarne suurus $T(y, \theta)$, kus y on andmed ja θ parameetrid).

Bayesi statistik konstrueeritakse nii, et see mõõdab mudeli sobivuse/sobimatuse mingit aspekti ja arvutatakse nii replikaat, kui reaalse andmestiku puhul. Kui $T(y^{rep}, \theta)$ ja $T(y, \theta)$ on lähedased, siis mudel sobib selles aspektis.

Kuid Bayesi teststatistik on juhuslik suurus, kuidas otsustame läheduse üle? Selleks on Bayesi p -value (olulisustõenäosus):

$$p_B = P[T(y^{rep}, \theta) \geq T(y, \theta) | y]. \quad (29)$$

Kui $p_B \approx 1$ või $p_B \approx 0$, siis on replikaatandmed selles aspektis ekstreemsemad, kui reaalsed \Rightarrow mudel ei sobi.

Tõenäosus P on võetud üle $(y^{rep}, \theta | y)$ ühisjoatuse $f(y^{rep}, \theta | y)$ fikseeritud andmete y korral:

$$p_B = \int I_{\{T(y^{rep}, \theta) \geq T(y, \theta)\}} f(y^{rep}, \theta | y) dy^{rep} d\theta \quad (30)$$

Integraali (30) arvutamine on keeruline analüütiliselt, Monte-Carlo integreerimine on aga lihtne. Paneme tähele, et (30) on $(0, 1]$ -juhusliku suuruse $I_{\{\cdot\}}$ keskväärtus tiheduse $f(y^{rep}, \theta | y)$ suhtes. Keskväärtushinnanguks on suuruse $I_{\{\cdot\}}$ valimikeskmine.

Selle leidmiseks viime läbi järgmise tsükli:

1. $i = 1$ (valimi indeks);
2. genereerime $(y_i^{rep}, \theta_i) \leftarrow f(y^{rep}, \theta | y)$, (y_i^{rep} on sama dimensiooniga, mis y);
3. arvutame $T(y_i^{rep}, \theta_i)$, $T(y, \theta_i)$;
4. loendame sündmuse $T(y_i^{rep}, \theta_i) \geq T(y, \theta_i)$
5. $i = i + 1$ ja kordame 2., 3. ja 4. sammu L korda.

Integraali hinnanguks on

$$\hat{p}_B = \frac{\#\{T(y_i^{rep}, \theta_i) \geq T(y, \theta_i), i = 1, \dots, L\}}{L}$$

(binaarse juhusliku suuruse valimikeskmine).

Genereerimissammu tegemiseks paneme tähele, et

$$f(y^{rep}, \theta | y) = f(y^{rep} | \theta) \underbrace{p(\theta | y)}_{\pi(\theta)},$$

kusjuures $f(y^{rep}|\theta)$ sõltub y -st ainult läbi θ . Seega punkti (y_i^{rep}, θ_i) saamiseks:

$$\begin{aligned}\theta_i &\leftarrow \pi(\theta), \\ y_i^{rep} &\leftarrow f(y^{rep}|\theta_i).\end{aligned}$$

Näide 15

Olgu $y_1, \dots, y_n \leftarrow B(1, \theta)$, sõltumatud Bernoulli andmed. $p(\theta) \sim U(0, 1)$, mitteinformatiivne beeta-jaotus. Siis $\pi(\theta) \sim \text{beeta}(s + 1, n - s + 1)$, $s = \sum_{i=1}^n y_i$.

Kontrollitav aspekt – kui sõltumatuse eeldus ei oleks täidetud, kas siis mudel sobiks? Olgu andmed 11 00000 11111 00000000, $n = 20$. Valime teststatistikuks üleminekute $0 \rightarrow 1$ ja $1 \rightarrow 0$ arvu. Andmetes

$$T(y, \theta) = 3.$$

$T(y^{rep}, \theta)$ leiame simuleerimise teel:

1. $\theta \leftarrow \text{beeta}(s + 1, n - s + 1)$;
2. $y^{rep} = (y_1^{rep}, \dots, y_{20}^{rep}) \leftarrow B(1, \theta)$, sõltumatud;
3. arvutame $T(y^{rep}, \theta)$ – üleminekute arv;
4. kordame samme 1-3 L korda.

$$\hat{p}_B = \frac{\#\{T(y_i^{rep}, \theta_i) \geq T(y, \theta_i), i = 1, \dots, L\}}{L} = \frac{9838}{10000} \approx 0,98$$

\Rightarrow Antud Bayesi mudel ei kirjelda selliste sõltuvustega andmeid.

Siin teststatistik $T(y, \theta)$ ei sõltunud θ -st, aga näiteks sümmeetria testimiseks võiks kasutada:

$$T(y, \theta) = |y_{[0.75n]} - \theta| - |y_{[0.25n]} - \theta|$$

1.10.3 Hälbimus

Mudeli sobivuse näitaja (üks Bayesi teststatistik) – mudelite võrdlemisel.

DEFINITSIOON: $D(y, \theta) = -2 \ln f(y, \theta)$

Hälbimus on arv, kui y ja θ on fikseeritud, vastasel juhul juhuslik suurus. Tavaliselt positiivne aga võib olla ka negatiivne.

Reegel: See mudel, millel on väiksem hälbimus, on parem antud andmete kirjeldamiseks. Kuidas aga mõistame "väiksem" juhuslike suuruste korral.

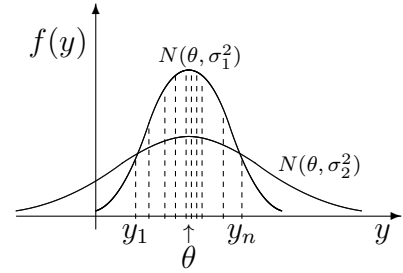
Näide Olgu andmed $y = (y_1, \dots, y_n)$. Olgu 2 mudelit (σ_1^2 ja σ_2^2 teada):

$$\begin{aligned}M1 : y_i|\theta &\sim N(\theta, \sigma_1^2), i = 1, \dots, n, \text{ sõltumatud;} \\ M2 : y_i|\theta &\sim N(\theta, \sigma_2^2), i = 1, \dots, n, \text{ sõltumatud.}\end{aligned}$$

$$f(y|\theta, \sigma_1^2) = \prod_{i=1}^n f(y_i|\theta, \sigma_1^2);$$

$$f(y|\theta, \sigma_2^2) = \prod_{i=1}^n f(y_i|\theta, \sigma_2^2);$$

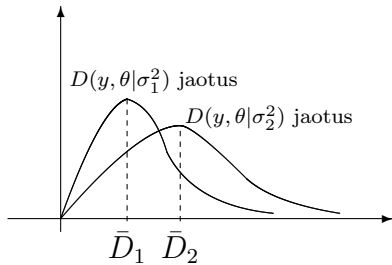
$$f(y|\theta, \sigma_1^2) > f(y|\theta, \sigma_2^2).$$



Fikseeritud y ja θ korral

$$D(y, \theta|\sigma_1^2) < D(y, \theta|\sigma_2^2). \quad (*)$$

Aga θ järeljaotus on määratud kitsas piirkonnas andmete keskvaartuse lähedal, seega



- (*) kehtib ka enamuse teiste θ -de korral järeljaotusest;
- $D(y, \theta|\sigma_1^2)$ jaotuse määramispiirkond paikneb vasakul $D(y, \theta|\sigma_2^2)$ omast ja on kitsam;
- $M1$ on parem mudel antud andmete kirjeldamiseks.

Näide 16

Olgu andmemudel $y_i|\theta \sim N(E(y_i|\theta), \sigma^2)$, $i = 1, \dots, n$, sõltumatud, σ^2 teada:

$$f(y|\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - E(y_i|\theta))^2}.$$

Siis

$$D(y, \theta) = -2 \left[-\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - E(y_i|\theta))^2 \right] =$$

$$= \underbrace{n \ln(2\pi\sigma^2)}_{\substack{\text{see liige võib olla} \\ \text{neg. kui } \sigma^2 \text{ väike}}} + \frac{1}{\sigma^2} \underbrace{\sum_{i=1}^n (y_i - E(y_i|\theta))^2}_{\text{see osa sõltub } \theta\text{-st}}$$

Fikseerime θ ning paneme tähele, et mida väiksem on θ -st sõltuv osa, seda paremini kirjeldab regressioonipind $E(y_i|\theta)$ andmeid y_i . \Rightarrow mida väiksem on $D(y, \theta)$, seda paremini sobib mudel andmetega.

Kuna θ on juhuslik, siis peame vaatama $D(y, \theta)$ jaotust võrreldavate mudelite korral. Sama otsustuskeem $D(y, \theta)$ abil sobib suvaliste mudelite korral.

Fikseeritud andmete korral on $D(y, \theta)$ juhuslik, sest θ on juhuslik. Teda kirjeldab jaotus. Jaotuse asemel vaadatakse harilikult selle mingeid karakteristikuid. Moodustame hällbimuse baasil 2 näitajat:

1. Hällbimuse punkthinnang – $\hat{D} = D(y, \hat{\theta})$, kus $\hat{\theta}$ on järeljaotuse $\pi(\theta)$ keskmine;

2. Keskmise hälbumus – $\bar{D} = E_{\pi}[D(y, \theta)|y]$, st keskmine on üle parameetri järeldaotuse $\bar{D} = \int D(y, \theta)\pi(\theta)d\theta$.

Parem mudel on see, mille mõlemad näitajad on väiksemad teise mudeli omadest.

\bar{D} arvutamine Monte-Carlo integreerimise abil (tegu on keskvaartuse leidmisega):

1. $\theta_l \leftarrow \pi(\theta)$;
2. arvutame $D(y, \theta_l)$;
3. kordame samme 1. ja 2. L korda;
4. integraali hindame valimikeskmisega

$$\hat{D} = \frac{1}{L} \sum_{l=1}^L D(y, \theta_l).$$

Samas võib teha näitajad \hat{D} ja \bar{D} väikeseks parameetrite arvu kasvatamisega (mudeli keerukuse suurendamisega), mis ei ole aga mõistlik idee.

Näiteks regressioonmudelis argumenttunnuste ja seega ka parameetrite lisamisel võime y -tunnuse hajuvuse regressioonipinna suhtes kuitahes väikeseks ajada.

Vajame näitajat, mis arvestaks mõlemat, parameetrite arvu ja hälbumust. Hea on väike parameetrite arv ja väike hälbumus.

DEFINITSIOON: Suurust $P_D = \bar{D} - \hat{D}$ (≥ 0) nimetatakse efektiivseks parameetrite arvuks.

Tüüpiliselt on see ligikaudu võrdne tegeliku parameetrite arvuga, aga ei pruugi alati olla. Bayesi mudeli korral

$$\hat{D} = D(y, \hat{\theta}) \leq D(y, \theta) \forall \theta \quad \Rightarrow \quad P_D \geq 0$$

Mitteinformatiivse eeljaotuse korral on järeldaotuse keskmine parameetri suurima tõepära hinnang, st. \hat{D} maksimiseeri tõepära ja seega minimiseerib hälbumuse.

Antud suurus ei tule aga ilma kriitikata, esiteks või P_D võib siiski mõnedel patoloogilistel juhtudel tulla negatiivne ja P_D ei ole invariantne ümberparametriseerimise suhtes.

Spiegelhalter (2002) soovib mudelite võrdlemisel kasutada suurust DIC – *Deviance Information Criterion*, mis summeerub kahest komponendist:

$$DIC = \text{mudeli sobivus} + \text{keerukus}$$

$$\begin{aligned} DIC &= \bar{D} + P_D \quad (= \bar{D} + \bar{D} - \hat{D} = 2\bar{D} - \hat{D} = \\ &= 2\bar{D} - 2\hat{D} + \hat{D} = 2P_D + \hat{D}) \end{aligned}$$

1. Mida väiksem DIC , seda parem mudel;

Mudel	\hat{D}	\bar{D}	P_D	DIC
M1, hierarhiline	57,8	60,6	2,8	63,4
M2, võrdsed	59,5	60,5	1,0	61,5
M3, eraldi grupid	54,9	62,5	7,7	70,3

2. Vähima DIC -ga mudel prognoosib esialgse andmestikuga kõige sarnasema replikaatandmestiku (seda mitteinformatiivse või ka tegelikkusega kooskõlas oleva informatiivse eeljaotuse korral).

Näide 17

Õppeprogrammid koolides (Gelman jt 2004, lk 140).

y_1, \dots, y_8 8 kooli, kus kasutatakse erinevaid õppeprogramme, y_i on i -nda kooli testitulemus (agregeeritud suurus).

Vaatame kolme mudelit:

Mudel 1 Hierarhiline mudel:

$$M1 : \begin{cases} y_j \sim N(\beta_j, \sigma_j^2), j = 1, \dots, 8, \sigma_j^2 \text{ on teada } (\beta_j - \text{kooli keskmine}); \\ \beta_j \sim N(\mu, \sigma_\beta^2) \forall j; \\ (\mu, \sigma_\beta^2) \sim p(\cdot) \text{ eeljaotus} \end{cases}$$

Mudel 2 Üldkeskmise mudel, kooli mõju puudub:

$$M2 : \begin{cases} y_j \sim N(\beta_j, \sigma_j^2), j = 1, \dots, 8, \beta_j \equiv \mu; \\ \mu \sim p(\cdot) \end{cases}$$

Mudel 3 Eraldi analüüs gruppides:

$$M3 : \begin{cases} y_j \sim N(\beta_j, \sigma_j^2); \\ (\beta_1, \dots, \beta_8) \sim N(\mu, \infty) \text{ mitteinformatiivne, sõltumatud.} \end{cases}$$

Siin β_j hinnatakse üksnes andmetelt, st. $E\beta_j = y_j$.

Hälbimus:

$$\begin{aligned} D(y, \theta) &= -\ln f(y|\theta) = \\ &= -2 \ln \left[\prod_{j=1}^J \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(y_j - \beta_j)^2}{2\sigma_j^2}} \right] = \\ &= \text{const} + \sum_{j=1}^J \frac{(y_j - \beta_j)^2}{2\sigma_j^2} \end{aligned}$$

Tõlgendus.

1.11 Mitteinformatiivne eeljaotus

Sageli pole võimalik parameetri eeljaotust defineerida, sest meil puudub informatsioon, tema kohta. Soovitakse eeljaotus defineerida nii, et rema osatähtsus parameetri järeljaotuses oleks minimaalne. Sellist eeljaotust nimetatakse kirjanduses:

- *vague* (nõrk);
- *flat* (lame);
- *diffuse* (hajunud);
- *noninformative* (mitteinformatiivne).

Eesmärgiks on lasta andmetel endil rääkida.

1.11.1 Korrektsed ja mittekorrigeeritud eeljaotused

Näide 18

Meenutame näidet:

$$\begin{aligned} y_i|\theta &\sim N(\theta, \sigma^2), \sigma^{teada}, \\ p(\theta) &= N(\mu_0, \tau_0^2); \\ \pi(\theta) &= N(\mu_1, \tau_1^2); \\ \mu_1 &= \tau_1^2 \left[\frac{n\bar{y}}{\sigma^2} + \frac{\mu}{\tau_0^2} \right] \\ \tau_1^2 &= \left[\frac{n}{\sigma^2} + \frac{1}{\tau_0^2} \right]^{-1} \end{aligned}$$

Kui τ_0^2 on suur, siis

$$\pi(\theta) \approx N(\bar{y}, \frac{\sigma^2}{n}). \quad (31)$$

Kuidas saime $\pi(\theta)$:

$$\pi(\theta) \propto p(\theta)\ell(\theta) \propto p(\theta) \exp\left\{-\frac{n(\bar{y} - \theta)^2}{2\sigma^2}\right\}$$

Tulemuse (31) saame siit ka siis, kui

$$p(\theta) \propto \text{const}, \theta \in (-\infty, \infty).$$

Selline eeljaotus ei ole rangelt võttes võimalik, ei integreeru väärtuseks 1 üle määramispiirkonna.

DEFINITSIOON: Eeljaotust $p(\theta)$ nimetame korrektseks, kui ta ei sõltu andmetest ja integreerub väärtuseks 1. Kui $p(\theta)$ integreerub suvaliseks positiivseks arvuks, nimetame teda normeerimata tiheduseks. Kui integraali väärtuseks on ∞ , siis nimetame eeljaotust mittekorrigeeritudseks.

Normeerimata tihedust saab alati normeerida konstandiga korrutamise teel.

Meie näites annab mittekorrektne eeljaotus $p(\theta) \propto \text{const}$ ometi korrektse järeljaotuse, eeldusel, et on vähemalt üks vaatlus ($n = 1$).

Näide 19

Siin veendume, et järeljaotus parameetrile σ^2 tuleb sama nii korrektsest kui ka ebakorrektselt mitteinformatiivsest eeljaotusest lähtudes.

Olgu $y_i|\theta, \sigma^2 \sim N(\theta, \sigma^2)$, θ on teada, σ^2 on tundmatu. Teame, et kui täpsuse parameetri $\phi = 1/\sigma^2$ eeljaotuseks valida gamma-jaotuse, siis on ka ϕ järeljaotus gamma (see pere on kaasjaotuste pere normaaljaotusega andmetele). Parameeter σ^2 järeljaotuseks on IG.

$$f(y_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\theta)^2} = \frac{\sqrt{\phi}}{\sqrt{2\pi}} e^{-\frac{\phi}{2}(y_i-\theta)^2}$$

$$p(\phi) = G(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \phi^{\alpha-1} e^{-\beta\phi}, \quad \phi > 0.$$

Kui α, β on väikesed, siis on see jaotus mitteinformatiivne: $E(\phi) = \frac{\alpha}{\beta}$, $D(\phi) = \frac{\alpha}{\beta^2}$

$$\ell(\phi) = \left(\sqrt{\frac{\phi}{2\pi}} \right)^n e^{-\frac{\phi}{2} \sum_i (y_i - \theta)^2}$$

$$\begin{aligned} \pi(\phi) &\propto \ell(\phi)p(\phi) \propto \phi^{\frac{n}{2}} e^{-\frac{\phi}{2} \sum_i (y_i - \theta)^2} \phi^{\alpha-1} e^{-\beta\phi} = \\ &= \phi^{\alpha-1+\frac{n}{2}} e^{-\phi(\frac{1}{2} \sum_i (y_i - \theta)^2 + \beta)} = G\left(\alpha + \frac{n}{2}, \frac{1}{2} \sum_i (y_i - \theta)^2 + \beta\right) \end{aligned}$$

ehk

$$\sigma^2 \sim IG\left(\alpha + \frac{n}{2}, \frac{1}{2} \sum_i (y_i - \theta)^2 + \beta\right)$$

Osutub aga, et σ^2 järeljaotuseks on ka siis IG, kui valida $p(\sigma^2) \propto \frac{1}{\sigma^2}$, kus $\sigma^2 \in (0, \infty)$. See on mittekorrektne jaotus (*improper*), sest $\int_0^\infty \frac{1}{x} dx = \ln(x)|_0^\infty = \infty$.

Kui $p(\sigma) \propto \frac{1}{\sigma^2}$, siis

$$\begin{aligned} \ell(\sigma^2) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_i (y_i - \theta)^2} \\ \pi(\sigma^2) &\propto \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_i (y_i - \theta)^2} \frac{1}{\sigma^2} \propto \\ &\propto \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} e^{-\frac{1}{\sigma^2} (\frac{1}{2} \sum_i (y_i - \theta)^2)} = \\ &= (\sigma^2)^{-(\frac{n}{2}+1)} e^{-\frac{1}{\sigma^2} (\frac{1}{2} \sum_i (y_i - \theta)^2)} = \\ &= IG\left(\frac{n}{2}, \frac{1}{2} \sum_i (y_i - \theta)^2\right) \end{aligned}$$

Võrdluseks varemsaaduga meenuta, et α ja β olid väikesed.

Bayesi analüüsis lähtutakse sageli mingist mittekorrektsest eeljaotusest, lihtsuse mõttes, ja saadakse järeldaotus.

$$\pi(\theta) \propto p(\theta)\ell(\theta),$$

mis on küll normeerimata, kuid mida on võimaliks normeerida ja saada täiesti õige jaotus. Järeldaotus, mis on saadud mittekorrektse eeljaotuse abil tuleb analüüsil kasutada suure etevaatlisusega:

- kas integraal on lõplik?
- normeerida;
- kas kuju on mõistlik (*sensible*).

1.11.2 Jeffrey invariantsuse omadus

Probleem mitteinformatiivse eeljaotusega – mitteinformatiivsus ei ole invariantne ümberparametriseerimise suhtes.

Näide 20

Olgu $\theta \in [0, 1]$, siis mitteinformatiivne eeljaotus on $p(\theta) = 1, \theta \in [0, 1]$. Vaatame 1 : 1 teisendust $\phi = \frac{1}{\theta}, (\theta = \frac{1}{\phi})$. Siis tihedusfunktsioon teiseneb Jakobiaanteisendusega:

$$p(\phi) = p(\theta) \left| \frac{1}{\phi^2} \right| = \frac{1}{\phi^2}, \quad \phi \in [1, \infty).$$

See pole mitteinformatiivne, annab väikestele väärtustele suurema tõenäosuse. Aga samas, kui θ kohta polnud eelinfot, kuidas siis ϕ kohta on? Peaks olema $p(\theta) \approx \text{const}, \phi \in [1, \infty)$.

Probleemi lahenduseks on Jeffrey eeljaotus, mis defineeritakse Fisher informatsiooni baasil.

Jeffrey eeljaotuseks on

$$p(\theta) \propto |I(\theta)|^{\frac{1}{2}},$$

kus $I(\theta)$ on Fisher informatsioonimaatriks:

$$I(\theta) = -E \left[\frac{\partial^2 \ln f(y|\theta)}{\partial \theta \partial \theta'} \middle| \theta \right].$$

Tegemist on maatriksiga, kus on leitud teist järku osatuletised vektori θ kõigi komponentide järgi.

Näide 21

Olgu 1 vaatlus binoomjaotusest $y \sim B(n, \theta)$ (n Bernoulli vaatlust, mille summa on y)

$$f(y|\theta) = C_n^y \theta^y (1 - \theta)^{n-y}$$

$$l = \ln f(y|\theta) \propto y \ln \theta + (n - y) \ln(1 - \theta)$$

$$\frac{\partial}{\partial \theta} l \propto \frac{y}{\theta} - \frac{n - y}{1 - \theta}$$

$$\frac{\partial^2}{\partial \theta^2} l \propto -\frac{y}{\theta^2} - \frac{n - y}{(1 - \theta)^2}$$

$$E \frac{\partial^2}{\partial \theta^2} l \propto -\frac{n\theta}{\theta^2} - \frac{n - n\theta}{(1 - \theta)^2} =$$

$$= -\frac{n}{\theta} - \frac{n}{1 - \theta} = -n \frac{1 - \theta + \theta}{\theta(1 - \theta)} = -\frac{n}{\theta(1 - \theta)}$$

$$I(\theta) = \frac{n}{\theta(1 - \theta)}$$

Jeffrey eeljaotus on seega,

$$p(\theta) \propto \theta^{-\frac{1}{2}} (1 - \theta)^{-\frac{1}{2}},$$

mis on beeta $(\frac{1}{2}, \frac{1}{2})$ jaotus. Seni kasutasime mitteinformatiivset jaotust beeta(1, 1). Mõlemal juhul on keskväärtaus $E(\theta) = \frac{1}{2}$. Jeffrey jaotus on aga invariantne ümberparametriseerimise suhtes.

Koduülesanne. Kumma beeta-jaotuse dispersioon on suurem.

2 Markovi ahelad

Markovi ahelate teooria rajas Andrei Andrejevitš Markov 20. sajandi alguses uurides vokaalide ja konsosnatide vaheldumist Puškini poemis Onegin. Ta arendas tõenäosusliku mudeli, et järgmine element (vokaal või konsonant) sõltub talle vahetult eelnevast.

Kuidas on Markovi ahelad seotud Bayesi statistikaga? Bayesi statistikas huvitume parameetrite järeljaotusest. Analüütiliselt on seda võimalik tuletada lihtsamatel juhtudel. Osutub aga, et simuleerida on võimalik järeljaotusest ka keerulistel juhtudel ja siis analüüsi teha simuleeritud väärtuste pealt. Simuleerimismeetodid, mis seda võimaldavad, genereerivad aga M -ahelaid. Valim on M -ahela trajektoor. Põhiküsimus: kas valim on ikka meie soovitud järeljaotusest. Siit M -ahela mõisted: statsionaarne jaotus, piirjaotus, koondumine, sissepõlemisperioon, algväärtused jm.

Antud kursuses vajame Markovi ahelatega seotud mõisteid seoses simuleerimismeetoditega mitmemõõtmelistest jaotustest – *Markov chain Monte Carlo methods* (McMC).

Olgu $\{x^t, t \geq 0\}$ \tilde{U} diskreetse ajaga juhuslik protsess:

- t -st sõltuv juhuslike suuruste (vektorite) jada;
- t on loenduv $0, 1, 2, \dots$

Eeldame, et x^t seisundite ruum Ω on lõplik

$$x^t \in \Omega, |\Omega| < \infty,$$

st. x^t on iga t korral diskreetne juhuslik suurus (vektor) jaotusega $x^t \sim \pi^t$, π^t on ahela marginaaljaotus.

DEFINITSIOON: Markovi ahel x^t on diskreetse ajaga juhuslik protsess, mille korral kehtib omadus:

$$P(x^t \in A | x^{t-1}, x^{t-2}, \dots, x^0) = P(x^t \in A | x^{t-1}), A \subset \Omega.$$

Definitsioon ütleb, et kui ajamomendil t_1 on x^{t_1} fikseeritud, siis tuleviku ahela väärtus ja mineviku protsess on sõltumatud. Üldjuhul on aga ahela elemendid sõltuvad, näiteks $\{x^1, x^2, \dots, x^t\}$ on mingi t -mõõtmelise jaotusega.

2.1 Markovi ahela üleminekutõenäosused

Üleminekutõenäosus seisundist i seisundisse j t sammu jooksul:

$$P_{ij}(t) = P(x^t = j | x^0 = i)$$

$$P_{ij}(1) = P_{ij}$$

$$P_{ij}(0) = I(i = j)$$

Üleminekutõenäosuste maatriks t sammu jooksul:

$$\mathbf{P}(t) = [P_{ij}(t)]$$

DEFINITSIOON: Markovi ahel on **monotoonne (homogeenne)**, kui üleminekutõenäosused ei sõltu ajast:

$$P(x^t = j | x^{t-1} = i) = P_{ij}, \forall t.$$

Kui seisundite ruumis Ω on n erinevat elementi, siis saame $n \times n$ üleminekutõenäosuste maatriksi:

$$\begin{array}{c} i \setminus j \\ \begin{bmatrix} 1 \\ \vdots \\ i \\ \vdots \\ n \end{bmatrix} \end{array} \begin{array}{c} \begin{bmatrix} 1 & \dots & j & \dots & n \end{bmatrix} \\ \underbrace{\begin{bmatrix} P_{11} & \dots & P_{1j} & \dots & P_{1n} \\ \vdots & & \vdots & & \vdots \\ P_{i1} & \dots & P_{ij} & \dots & P_{in} \\ \vdots & & \vdots & & \vdots \\ P_{n1} & \dots & P_{nj} & \dots & P_{nn} \end{bmatrix}}_P \end{array}$$

Üleminekutõenäosuste maatriksi ridade summa on 1. Sellist maatriksit nimetatakse stohhastiliseks.

Näide 22

Juhuslik ekslemine – osake võib liikuda sirgel ühe sammu paremale, vasakule või jääda paigale. Olgu x^t osakese positsioon ajahetkel t , $t > 0$:

$$x^t = x^{t-1} + \omega_t = \omega_1 + \omega_2 + \dots + \omega_t, \quad x^0 = 0,$$

kus $\omega_i \in \{-1, 0, 1\}$ on sammu pikkus ja $\omega_i \sim f$ on sõltumatud juhuslikud suurused:

$$\begin{cases} f(1) = p \\ f(-1) = q \\ f(0) = r, \quad p + q + r = 1 \end{cases}$$

Üleminekutõenäosuste maatriks (üleminek seisundist i seisundisse j):

$$P_{ij} = \begin{cases} p, & j = i + 1 \\ q, & j = i - 1 \\ r, & j = i \\ 0, & j \neq i - 1, i, i + 1 \end{cases}$$

Niiviisi defineeritud x^t on Markovi ahel. Seisund x^t sõltub ainult eelmisest seisundist x^{t-1} . Kui aga x^{t-1} pole fikseeritud, siis ahela seisund ajahetkel t on kirjeldatud $\omega_1 + \dots + \omega_t$ jaotusega. Ahela seisundite ruum Ω on kõigi täisarvude hulk. Maatriks P on lõpmatudimensionaalne, aga 3-diagonaalne.

Champman-Kolmogorovi seos

$$P_{ij}(t) = P(x^t = j | x^0 = i) = \sum_{s \in \Omega} P(x^{t-1} = s | x^0 = i) P_{sj} = \sum_{s \in \Omega} P_{is}(t-1) P_{sj}$$

Maatrikskujul:

$$\mathbf{P}(t) = \mathbf{P}(t-1)\mathbf{P}.$$

Siit, monotoonse ahela korral, kus üleminekutõenäosused ei sõltu t -st

$$\mathbf{P}(t) = \mathbf{P}^t.$$

Sama mõttekäiguga saab tuletada üldisema seose:

$$p_{ij}(t+k) = p(x^{t+k} = j | x^0 = i) = \sum_{s \in \Omega} P(x^t = s | x^0 = i) P_{sj}(k) = \sum_{s \in \Omega} P_{is}(t) P_{sj}(k).$$

Maatrikskujul:

$$\mathbf{P}(t+k) = \mathbf{P}(t)\mathbf{P}(k)$$

2.2 Markovi ahela marginaaljaotus

Ahela $\{x^t : t \geq 0\}$ marginaaljaotus on tema ühe komponendi x^t jaotus fikseeritud t korral.

$$\pi_t(j) = P(x^t = j), \quad j \in \Omega \quad (\text{diskreetne juht});$$

Kui $j = 1, \dots, n$ (ahela võimalikud seisundid), siis marginaaljaotus on n -dimensionaalne vektor.

$$\pi_t = (\pi_t(1), \dots, \pi_t(n)).$$

Ahela algjaotus on π_0 , st. kui $t = 0$.

Kodune ülesanne nr. 2.1 Leia juhusliku ekslemise ahela marginaaljaotused π_1 ja π_2 .

Vaatleme x^t jaotust:

$$\pi_t(j) = P(x^t = j) = \sum_{s \in \Omega} p(x^t = j | x^0 = s) P(x^0 = s) = \sum_{s \in \Omega} P_{sj}(t) \pi_0(s).$$

Ahela marginaaljaotus sõltub algjaotusest ja üleminekutõenäosustest. Maatrikskujul ahela marginaaljaotus sammul t on

$$\pi_t = \pi_0 \mathbf{P}(t) = \pi_0 \mathbf{P}(t-1) \mathbf{P} = \pi_{t-1} \mathbf{P}. \quad (32)$$

DEFINITSIOON: Olgu antud monotoonne Markovi ahel x^t üleminekutõenäosustega P_{ij} . Jaotust π nimetatakse ahela x^t statsionaarseks jaotuseks, kui ta ei sõltu ajamomendist t .

Statsionaarse jaotuse korral saab seos $\pi_t = \pi_{t-1} \mathbf{P}$ kuju $\pi = \pi P$ ehk:

$$\sum_i \pi(i) P_{ij} = \pi(j),$$

kus $\pi = (\pi(1), \dots, \pi(n))$ ja $P = (P_{ij})$, $i, j = 1, \dots, n$. Kui t kasvades x^t jaotus koondub ahela statsionaarseks jaotuseks π , siis jaotus jääb samaks kõigi järgmiste sammude korral.

Järgmises näites vaatame selliseid mõisteid nagu statsionaarne jaotus, selle olemasolu ja ühesus, koondumine statsionaarseks jaotuseks, lahutuv ahel, üleminekutõenäosuste maatriksi koondumine.

Näide 23

Olgu $\{x^t : t \geq 0\}$ Markovi ahel seisundite ruumiga $\Omega = \{0, 1\}$. Algjaotust tähistame $\pi_0 = (\pi_0(0), \pi_0(1))$, kus $\pi_0(0)$ - seisundi 0 ja $\pi_0(1)$ - seisundi 1 tõenäosus. Üleminekutõenäosuste maatriks:

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

kus p on seisundist 0 seisundisse 1 üleminekutõenäosus ja q seisundist 1 seisundisse 0 üleminekutõenäosus. Leiame t -nda elemendi marginaaljaotuse: $\pi_t(j) = P(x^t = j)$, $j = 0, 1$, st. seisundi j tõenäosuse t sammu pärast.

Väärtuse 0 tõenäosus:

$$\begin{aligned} \pi_t(0) &= P(x^t = 0) = \sum_{j=0}^1 P(x^t = 0 | x^{t-1} = j) P(x^{t-1} = j) = \\ &= (1-p) \underbrace{P(x^{t-1} = 0)}_{\pi_{t-1}(0)} + q \underbrace{P(x^{t-1} = 1)}_{\pi_{t-1}(1)} = (1-p-q)\pi_{t-1}(0) + q \end{aligned}$$

P_{j0} - seisundist j seisundisse 0 üleminekutõenäosus, $P_{00} = 1-p$, $P_{10} = q$.

$$\pi_{t-1}(0) = (1-p-q)\pi_{t-2}(0) + q.$$

Jätkates samm-sammult marginaaljaotuste leidmist, saame

$$\pi_t(0) = (1-p-q)^t \pi_0(0) + q \sum_{k=0}^{t-1} (1-p-q)^k.$$

Kasutades geomeetrilise progressiooni valemit:

$$\sum_{k=0}^r a_0 q^k = a_0 \frac{1-q^{r+1}}{1-q},$$

saame

$$\begin{aligned} \pi_t(0) &= (1-p-q)^t \pi_0(0) + q \frac{1-(1-p-q)^t}{p+q} = \\ &= \frac{q}{p+q} + (1-p-q)^t \left[\pi_0(0) - \frac{q}{p+q} \right]. \end{aligned} \tag{33}$$

Loomulikult $\pi_t(1) = 1 - \pi_t(0)$. Saime x^t marginaaljaotuse. See sõltub algjaotusest ja üleminekutõenäosusest.

Vaatame $p+q$ võimalikke väärtusi

- $p+q = 0$, siis $p = q = 0$.

Üleminekumaatriks $\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Ahel ei muuda seisundit, jääb konstantselt sellesse seisundisse, kuhu alguses sattus (lahutuv ahel).

Iga algjaotus on aga statsionaarne, x^t marginaaljaotus

$$\left. \begin{aligned} P(x^t = 0) &= \pi_0(0) \\ P(x^t = 1) &= \pi_0(1) \end{aligned} \right\} \Rightarrow \pi_t = \pi_0$$

- P on lahutuv;
- Statsionaarne jaotus pole üheselt määratud.
- $0 < p + q < 2$, st $0 < 1 - p - q < 1$, siis ahelal leidub piirjaotus, sõltumata algjaotusest. Piirjaotus on statsionaarne jaotus:

$$\pi_t(0) \xrightarrow{t \rightarrow \infty} \frac{q}{p+q},$$

sest $|1 - p - q| < 1$ ja $(1 - p - q)^t \xrightarrow{t \rightarrow \infty} 0$.

$$\pi_t(1) = [1 - \pi_t(0)] \xrightarrow{t \rightarrow \infty} \left[1 - \frac{q}{p+q}\right] = \frac{p}{p+q}.$$

Ahela piirjaotus ei sõltu algjaotusest.

- $p + q = 2$, siis pole ahelal statsionaarset jaotust:

$$(1 - p - q)^t = (-1)^t,$$

saame perioodilise jaotuse

$$\pi_t(0) = \frac{1}{2} + (-1)^t \left[\pi_0(0) - \frac{1}{2} \right],$$

ahela paaritute elementide jaotus on $\pi_t(0) = 1 - \pi_0(0)$, $\pi_t(1) = \pi_0(0)$ ja paariselementidel $\pi_t(0) = \pi_0(0)$, $\pi_t(1) = 1 - \pi_0(0)$.

Seega sõltub palju üleminekutõenäosuste maatriksist, kas ahelal on statsionaarne jaotus, kas on piirjaotus, kas sõltumata algjaotusest.

Vaatame üleminekutõenäosuste maatriksi muutumist läbi aja. Seosest $\pi_t = \pi_0 P(t)$ saame leida $P(t)$:

$$\begin{cases} \pi_t(0) = \sum_j P_{j0}(t) \pi_0(j), \\ \pi_t(1) = 1 - \pi_t(0). \end{cases}$$

$$\pi_t(0) = \sum_j P_{j0}(t) \pi_0(j) = P_{00}(t) \pi_0(0) + P_{10}(t) \pi_0(1).$$

Algjaotus võib olla suvaline, seega võtame $\pi_0(0) = 1 \Rightarrow P_{00}(t) = \pi_t(0)$. Leiame valemist (33) $P_{00}(t)$ kuju

$$P_{00}(t) = \frac{q}{p+q} + (1 - p - q)^t \left[\pi_0(0) - \frac{q}{p+q} \right] = \frac{q}{p+q} + (1 - p - q)^t \frac{p}{p+q}.$$

Analoogiliselt leiame $P_{10}(t)$, $P_{01}(t)$ ja $P_{11}(t)$. Üleminekutõenäosuste maatriks t sammu pärast:

$$\mathbf{P}(t) = \frac{1}{p+q} \begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1 - p - q)^t}{p+q} \begin{pmatrix} q & -p \\ -q & p \end{pmatrix}$$

Ahela statsionaarne jaotus on aluseks MCMC-meetodile. Saab näidata, et kui statsionaarne jaotus π leidub ja

$$\lim_{t \rightarrow \infty} P_{ij}(t) = \pi(j)$$

siis sõltumatu algjaotusest $\pi_t \xrightarrow[t \rightarrow \infty]{} \pi$. Seega ahela marginaaljaotus koondub tema statsionaarseks jaotuseks.

Näide 24

Leiame statsionaarse jaotuse π , mis on võrrandi $\pi = \pi P$ lahend.

$$\pi(0)P_{0j} + \pi(1)P_{1j} = \pi(j), \quad j = 0, 1.$$

Lahendiks on

$$\pi = \left(\frac{q}{p+q}, \frac{p}{p+q} \right).$$

See oli algjaotus, mis oli invariantne sammude suhtes.

Kui $p + q < 2$, siis

$$\lim_{t \rightarrow \infty} P(t) = \frac{1}{p+q} \begin{pmatrix} q & p \\ q & p \end{pmatrix}.$$

Piiril saavad read võrdseks, ridades on tõenäosusjaotused. Ükskõik millises seisundis oleme, järgmise hetke seisund omandatakse jaotuse $\left(\frac{q}{p+q}, \frac{p}{p+q} \right)$ kohaselt $x^t | x^{t-1} \sim x^t$. Seega üleminekutõenäosuste maatriksiga järgmist elementi genereerides genereerime tegelikult väärtuse x^t marginaaljaotusest. See jaotus ei muutu enam. Ahel koondub oma statsionaarseks jaotuseks, milleks on $\left(\frac{q}{p+q}, \frac{p}{p+q} \right)$ ja seda sõltumata algjaotusest.

Näide 25

Olgu meid huvitav jaotus

$$\pi(i) = \begin{cases} \frac{1}{3}, & i = 0, 1, 2; \\ 0, & \text{mujal} \end{cases}$$

Tahame, et see oleks mingi ahela statsionaarne jaotus. Statsionaarsuse tingimusest saab leida üleminekutõenäosuste maatriksi:

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

Kontrollime, kas $\pi(i)$ on statsionaarne selle ahela suhtes ehk kas $\pi = \pi P$

$$\begin{aligned} \pi &= \left(\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3} \right) \\ \pi P &= \left(\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3} \right) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \left(\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3} \right) \end{aligned}$$

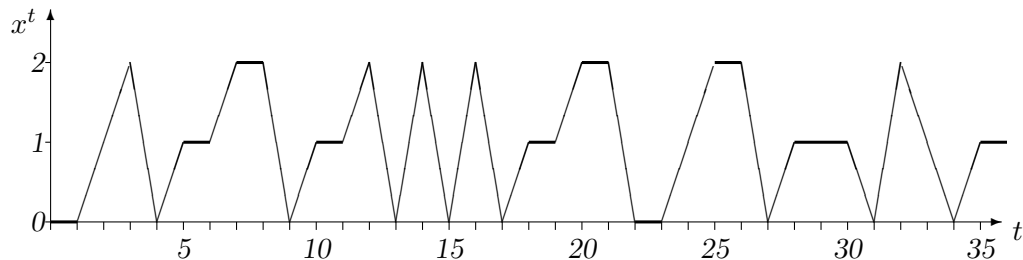
Järelikult on $\pi(i)$ selle ahela statsionaarne jaotus. NB! Samale jaotusele π võib vastata mitu erinevat maatriksit P , ehk erinevat Markovi ahelat. Siit see, et simuleerimiseks jaotusest π saab kasutada erinevaid M -ahelaid, erinevaid meetodeid.

Algoritm realisatsioonide saamiseks jaotusest $\pi(i)$ kasutades tema Markovi ahelat:

1. $x^0 = 0, t = 0$;
2. $u \sim U(0, 1), t = t + 1$;
3. Kui $u < \frac{1}{2}$, siis $x^t = (x^{t-1} + 1) \bmod 3$ (3-ga jagamisel tekivad jäägid), muidu $x^t = x^{t-1}$;
4. Korrata samme 2 ja 3.

Kolmandal sammul kasutame sisuliselt maatriksi P ($t-1$)-ndat rida x^t väärtuste tekitamiseks.

Lõpuks on valmis keskest läbi $\frac{1}{3}$ "0", $\frac{1}{3}$ "1" ja $\frac{1}{3}$ "2".



Simuleerimine kasutades Markovi ahelat on paljulubav, aga tekivad uued probleemid:

- üleminekumaatriksi P määramine;
- kui pikk on sissepõlemisperiood (koondumine statsionaarseks jaotuseks);
- sõltuvusega seotud probleemid.

Kuidas antud jaotuse π jaoks leida üleminekutõenäosuste maatriksi P ehk kuidas leida Markovi ahel nii, et π oleks selle ahela statsionaarseks jaotuseks.

Järgmises loengus anname teoreemid, et Markovi ahel koondub statsionaarseks jaotuseks, mis on just meie huvialune π .

2.3 Seisundite klassifikatsioon

Markovi ahelate Monte Carlo meetodid tuginevad piirteoreemidele, mis eeldavad teatud tüüpi ahelaid. Tähtsad eeldused on mittelahutuvus ja aperiodilisus. Nende mõistete selgitamiseks on vaja defineerida rida mõisteid ahela seisunditele.

Defineerime:

- $f_j(t) = P(x^t = j, x^{t-1} \neq j, \dots, x^1 \neq j | x^0 = j)$ – tõenäosus, et ahel väljub seisundist j jõuab sinna tagasi täpselt t sammuga. Kehtib: $f_j(t) \leq P_{jj}(t)$, kus $P_{jj}(t)$ on üleminekutõenäosus t sammuga.
- $F_j = \sum_{t=1}^{\infty} f_j(t)$ – tõenäosus, et ahel, mis alustab seisundist j jõuab sinna üldse tagasi.

Ilmselt $F_j \neq P_{jj}(t)$.

Tõenäosuse F_j väärtuse järgi jagunevad seisundid:

- $F_j = 1$ – seisund j on **korduv (rekurentne, ingl. k. recurrent)** (kindlasti pöördutakse seisundisse j tagasi);
- $F_j < 1$ – seisund j on **mööduv (mitterekurentne)** (ei pruugi seisundisse j tagasi pöörduda);
- $P_{jj}(t) \xrightarrow{t \rightarrow \infty} 0$ – seisund j on **nulliline** (vastasel juhul mitterekurentne).

Tähistame:

T_j – seisundisse j tagasipöördumise aeg (sammude arv) alustades seisundist j . T_j on juhuslik suurus. Üldjuhul, rekurentse seisundi j korral, $ET_j = \mu_j < \infty$ (tagasipöördumise aeg on lõplik).

Tähistame N_j seisundi j külastuste arv.

- Kui j on rekurentne, siis külastuste arv $N_j = \infty$;
- Kui j on mitterekurentne, siis külastuste arv $N_j < \infty$.

DEFINITSIOON: Markovi ahel on **mittelahutuv (irreducible)**, kui iga seisundite paari $i, j \in \Omega$ korral leidub sammude arv t , nii et $P_{ij}(t) > 0$, see tähendab, et iga seisund on saavutatav.

Mitelahutuvuse kontrollimiseks tuleks teha suunatud graaf ja veenduda, et leidub noolte jada igast seisundist igasse seisundisse.

DEFINITSIOON: Hulka C Markovi ahela seisundite ruumis nimetatakse suletud hulgaks, kui $i \in C, j \notin C$ korral $P_{ij}(t) = 0, \forall C$.

DEFINITSIOON: Seisund j on **perioodiline** perioodiga d_j , kui temasse tagasipöördumine on võimalik ainult sammude arvuga $d_j > 1$ või selle kordsega, kus $d_j = \dot{S}ÜT\{t : f_j(t) > 0\}$.

Kui $d_j = 1$, siis seisund on **aperioodiline** (mitteperioodiline).

DEFINTISOON: Markovi ahelat nimetatakse aperioodiliseks, kui kõik tema seisundid on aperioodilised; vastasel korral on ahel perioodiline.

Kui üleminekutõenäosus $P_{jj} > 0$ (seisundisse saab jääda positiivse tõenäosusega), siis seisund on mitteperioodiline.

Solidaarsusteoreem – Mittelahutuvas Markovi ahelas on kõik seisundid sama tüüpi. Kui üks seisund on rekurentne, siis on ka kõik teised seisundid rekurentsed; kui üks seisund on perioodiline perioodiga d , siis on ka teised seisundid perioodilised; kui üks seisund on nulliline, siis on ka kõik teised seisundid nullilised.

DEFINTISOON: Kui seisund j on aperioodiline ($d_j = 1$) ja rekurentne, siis see seisund on **ergoodiline**. Markovi ahel on **ergoodiline**, kui kõik tema seisundid on ergoodilised.

Näide 26

Olgu meil ahel x^t olekute ruumiga $\{1, 2, 3, 4\}$ ja olgu üleminekutõenäosuste maatriks:

$$P = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Ahel on mittelahutuv (kontrolli!). Klassifitseerida on vaja üks seisund. Teised on sama tüüpi. Mõttekas on alustada diagonaalelementidest.

$$P_{11} > 0 \Rightarrow \text{seisund 1 on mitteperioodiline.}$$

Kas on ka rekurentne?

$$f_1(1) = \frac{1}{4} - \text{seisundisse 1 tagasipöördumise tõenäosus 1 sammuga;}$$

$$f_1(2) = \frac{1}{4} - \text{seisundisse 1 tagasipöördumise tõenäosus 2 sammuga;}$$

$$f_1(3) = \frac{1}{4} - \text{seisundisse 1 tagasipöördumise tõenäosus 3 sammuga;}$$

$$f_1(4) = \frac{1}{4} - \text{seisundisse 1 tagasipöördumise tõenäosus 4 sammuga.}$$

Rohkem võimalusi ei ole, kui alustame seisundist 1. Kuna $F_1 = \sum_{t=1}^4 f_q(t) = 1$ - siis on seisund 1 rekurentne. Järelikult on kõik seisundid rekurentsed.

Keskmine tagasipöördumise aeg seisundisse 1 (alustades seisundist 1):

$$\mu_1 = \sum_{t=1}^4 t f_1(t) = \frac{1}{4}(1 + 2 + 3 + 4) = 2,5 \text{ sammu.}$$

Ahel on aperiodiline ja rekurentne \Rightarrow ahel on ergoodiline.

Üleminekutõenäosuste maatriksid $P^t, t = 4, 16, 64$ sammu:

0.489	0.098	0.176	0.238	0.400	0.100	0.200	0.300	0.4	0.1	0.2	0.3
0.250	0.250	0.250	0.250	0.398	0.100	0.201	0.301	0.4	0.1	0.2	0.3
0.313	0.063	0.313	0.313	0.400	0.100	0.200	0.300	0.4	0.1	0.2	0.3
0.391	0.078	0.141	0.391	0.400	0.100	0.200	0.300	0.4	0.1	0.2	0.3

64 sammuga on ahel koondunud tasakaaluolekusse. Piirjaotuseks on $(0.4, 0.1, 0.2, 0.3)$, mis on ka selle ahela statsionaarne jaotus. Kontrolli!

Kodune ülesanne nr. 2.2 Olgu Markovi ahela olekute ruum $\{1, 2, 3, 4\}$. Leia ahela a) periood ja suletud hulgad. Kas ahel on lahutuv? Joonista suunatud graaf. Klassifitseeri ahela b) olekud.

$$a) \quad P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad b) \quad P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

2.4 Piirteoreemid

TEOREEM: (olemasolu ja ühesus). Igal mittelahutuval aperiodilisel Markovi ahelal $\{x^t, t \geq 0\}$ leidub statsionaarne jaotus ja see on tema ainus statsionaarne jaotus.

TEOREEM: (koondumine sõltumata algjaotusest). Olgu $\{x^t, t \geq 0\}$ **mittelahutuv aperiodiline** Markovi ahel seisundite ruumiga D , üleminekutõenäosuste maatriksiga P , **statsionaarse** jaotusega π ja suvalise algjaotusega π^0 , siis

$$\pi^t \xrightarrow[t \rightarrow \infty]{} \pi$$

Teoreem ütleb meile seda, et kui laseme ahelal piisavalt kaua joosta (st $t = n$ suur), siis sammul n on x^n jaotuseks statsionaarne jaotus π , ja seda sõltumata algjaotusest. Ka iga $t > n$ korral on x^t jaotuseks π . Sellise koondumise kohta öeldakse sageli, et Markovi ahel koondub ekviliibrium- ehk tasakaaluolekusse.

Mida tähendab, et üks jaotus koondub teiseks? Mingil viisil defineeritud kaugusmõõt läheneb nulli. Defineerime jaotuste vahelise kauguse selliselt: kui ξ_1 ja ξ_2 on mingid jaotused, st $\xi_i(A)$ määrab hulga A tõenäosuse, siis nende jaotuste vaheliseks kauguseks on:

$$\|\xi_1 - \xi_2\| = \sup_{A \in \Omega} |\xi_1(A) - \xi_2(A)|$$

Seega koondumine $\pi^t \xrightarrow[t \rightarrow \infty]{} \pi$ tähendab $\|\pi^t - \pi\| = \sup_{A \in \Omega} |\pi^t(A) - \pi(A)| \xrightarrow[t \rightarrow \infty]{} 0$. Diskreetsel juhul tõenäosuste vektor π^t saab tõenäosuste vektoriks π .

Marginaaljaotuse koondumine piirjaotuseks tähendab ühtlasi üleminekutõenäosuste koondumist:

$$\lim_{t \rightarrow \infty} P_{ij}(t) = \pi(j), \quad \forall i, j \in \Omega.$$

Seega üleminekutõenäosuste maatriksi ridadeks saab tõenäosuste vektor π . Kõik read on võrdsed (st kõik tinglikud jaotused on võrdsed) ja kõigist seisunditest saadakse seisund j tõenäosusega $\pi(j)$ iga j .

Mittelahutuva ja aperioidilise ahela jaoks saab formuleerida tähtsad piirteoreemid statistiliste otsustuste tegemiseks.

DEFINTISOON: Olgu $f(x)$ reaalkäitumuseline funktsioon. Funktsiooni $f(x)$ **ergoodiliseks keskmiseks** nimetatakse suurust

$$\bar{f}_N = \frac{1}{N} \sum_{t=1}^N f(x^t),$$

kus x^t on Markovi ahel. Kõik tulemused kehtiksid ka siis, kui summeerimist alustada algväärtusest $t = 0$, aga kokkuleppeliselt alustatakse ahela väärtusest $t = 1$. Ergoodiline keskmine on lihtsalt juhuslike suuruste aritmeetiline keskmine. Sõna ergoodiline viitab siin liidetavate sõltuvusele (Markovi ahel).

Ergoodiline teoreem (Markovi ahelate SAS) – Kui ahel on ergoodiline piirjaotusega π ja kui $E_\pi[f(x)] < \infty$, siis ergoodiline keskmine koondub teoreetiliseks keskmiseks:

$$\bar{f}_N \xrightarrow{N \rightarrow \infty} E_\pi[f(x)] \text{ tõenäosusega } 1,$$

kus $E_\pi[f(x)] = \int f(x)d\pi$ on funktsiooni f teoreetiline keskmine jaotuse π suhtes (nn ÜK keskväärts).

Järeldus: Kui funktsioon $f(x) = x$, siis vastavalt teoreemile koondub Markovi ahela valimikeskmine jaotuse π keskväärtsuseks:

$$\bar{x}_N \xrightarrow{N \rightarrow \infty} \int x d\pi$$

Järeldus: Ahela väärtuste (ja väärtuste funktsioonide) keskmised annavad mõjusad hinnangud jaotuse π parameetritele hoolimata ahela sõltuvustest.

Järeldus: Kui $f(x) = \begin{cases} 1, & \text{kui } x = j, j \in \Omega; \\ 0, & \text{muidu,} \end{cases}$

siis ergoodiline keskmine lihtsalt loendab seisundi j suhtelist sagedust ahela realisatsioonis. Ergoodilise teoreemi järgi see suhteline sagedus koondub seisundi j tõenäosuseks piirjaotuse järgi, viimane väljendab ka seisundi keskmist külüstussagedust:

$$\bar{f}_N = \frac{\#j}{N} \xrightarrow{N \rightarrow \infty} \pi(j) = \frac{1}{\mu_j}$$

μ_j – seisundisse j tagasipöördumise keskmine aeg.

Analoogiliselt suurte arvude seadusele, eksisteerib ka tsentraalne piirteoreem Markovi ahelate jaoks. Piirteoreemi sõnastamiseks on vaja defineerida uued mõisted, eelkõige ahela elementide sõltuvuse kirjeldamiseks.

2.5 Ergoodilise keskmise dispersioon ja tsentraalne piirteoreem

Olgu π ahela x^t statsionaarne jaotus.

DEFINTISOON: Ahela $f^t = f(x^t)$ autokovariatsiooniks sammuga k nimetatakse suurust

$$\gamma_k = Cov(f^t, f^{t+k}) = E_\pi(f^t f^{t+k}) - E_\pi f^t E_\pi f^t E_\pi f^{t+k}.$$

Ahela f^t dispersioon: $\sigma^2 = \gamma_0 = E(f^t)^2 - (E f^t)^2$

Ahela f^t autokorrelatsioon $\rho_k = \frac{\gamma_k}{\sigma^2}$

Tahame leida ergoodilise keskmise dispersiooni $D_\pi(\bar{f}_N)$. Otsime seda kujul:

$$D_\pi(\bar{f}_N) = \frac{\tau_N^2}{N}.$$

Leiame τ_N^2 avaldise:

$$\begin{aligned} \tau_N^2 &= N D_\pi(\bar{f}_N) = N D_\pi\left(\frac{1}{N} \sum_{t=1}^N f^t\right) = \frac{1}{N} D_\pi\left(\sum_{t=1}^N f^t\right) = \\ &= \frac{1}{N} \left(\sum_{t=1}^N \underbrace{D_\pi(f^t)}_{\sigma^2} + 2 \sum_{t=1}^{N-1} \left[Cov(f^t, f^{t+1}) + Cov(f^t, f^{t+2}) + \dots + Cov(f^t, f^N) \right] \right) = \\ &= \frac{1}{N} \left(N\sigma^2 + 2 \left[(N-1)\gamma_1 + (N-2)\gamma_2 + \dots + \gamma_{N-1} \right] \right) = \\ &= \sigma^2 + 2 \sum_{k=1}^{N-1} \frac{N-k}{N} \gamma_k = \sigma^2 \left[1 + 2 \sum_{k=1}^{N-1} \frac{N-k}{N} \frac{\gamma_k}{\sigma^2} \right] = \sigma^2 \left[1 + 2 \sum_{k=1}^{N-1} \frac{N-k}{N} \rho_k \right] \end{aligned}$$

Saab näidata, et $\tau_N^2 \xrightarrow{N \rightarrow \infty} \tau^2$, kus $\tau^2 = \sigma^2 (1 + 2 \sum_{k=1}^{\infty} \rho_k)$, kui selline lõpmatu summa leidub (koondub absoluutselt).

Tuleb vahet teha ahela elemendi dispersioonil $\sigma^2 = D_\pi(f^t)$ ja ergoodilise keskmise dispersiooni iseloomustaval τ^2 -l. Sõltumatu valimi korral $\tau^2 = \sigma^2$. Sammu k suurenedes tavaliselt ahela elementidevahelised korrelatsioonid vähenevad ja τ^2 või τ_N^2 on leitav summa mõne esimese liikme abil.

Ergoodilise keskmise dispersioon on seega:

$$D_\pi(\bar{f}_N) = \frac{\sigma^2}{N} \left[1 + 2 \sum_{k=1}^{N-1} \frac{N-k}{N} \rho_k \right].$$

Paneme tähele, et sõltumatu ahela korral jääb avaldisest järele meile tuntud valimikeskmise dispersiooni avaldis.

Tsentraalne piirteoreem – Kui ahel on ühtlaselt geomeetriliselt ergoodiline ja $f^2(x)$ on integreeruv π suhtes, siis

$$\sqrt{N} \frac{\bar{f}_N - E_\pi[f(x)]}{\tau} \xrightarrow{N \rightarrow \infty} N(0, 1).$$

Tsentraalne piirteoreem annab teoreetilise aluse ligikaudsete usalduspiiride konstrueerimiseks jaotuse π parameetritele. Selleks on vaja hinnata tundmatu suurus τ^2 .

(Geomeetriliselt ergoodiline koondumine piirjaotuseks toimub geomeetrilise progressiooni kiirusega, ühtlaselt sama kiirusega iga hulga korral Ω -st.)

2.6 Pööratavad ahelad (*reversible chains*)

Pööratavad Markovi ahelad on teatud alamklass Markovi ahelaid, mis on eriti sobivad jaotustest genereerimise algoritmidega, sest nende korral on lihtne leida sellist Markovi ahelat, mille jaoks etteantud jaotus oleks statsionaarne jaotus.

Olgu $\{x^t, t \geq 0\}$ monotoonne Markovi ahel üleminekutõenäosustega P_{ij} ja statsionaarse jaotusega π . Vaatame Markovi ahelat tagurpidi: x^t, x^{t+1}, \dots . Saab näidata, et kehtib seos:

$$P(x^t = j | x^{t+1}, x^{t+2}, \dots) = P(x^t = j | x^{t+1} = i).$$

Seega on ka tagurpidi ahel Markovi ahel.

Mõtleme nii, et meil on juhuslike suuruste vektor, saame rääkida tinglikest jaotustest mõlemat pidi.

Tähistame $P_{ij}^*(t)$ tagurpidi Markovi ahela üleminekutõenäosusi $t + 1$ sammust t -sse.

$$\begin{aligned} P_{ij}^*(t) &= P(x^t = j | x^{t+1} = i) = \\ &= \frac{P(x^{t+1} = i | x^t = j) P(x^t = j)}{P(x^{t+1} = i)} = \\ &= \frac{\pi_t(j) P_{ij}}{\pi_{t+1}(i)} \end{aligned}$$

Üldjuhul pole tagurpidi ahel monotoonne. Juhul kui ahel saavutab statsionaarse jaotuse π , ($t \rightarrow \infty$) või ahela algjaotus $\pi_0 = \pi$, siis on tagurpidi ahel samuti monotoonne.

$$P_{ij}^*(t) = \frac{\pi(j) P_{ij}}{\pi(i)} = P_{ij}^*. \quad (34)$$

DEFINITSIOON: Markovi ahel on **pööratav**, kui $P_{ij}^* = P_{ij}, \forall i, j \in \Omega$ korral, st. tagurpidi Markovi ahel on samade üleminekutõenäosustega.

Pööratavuse tingimus (34) esitatakse tavaliselt kujul:

$$\pi(i)P_{ij} = \pi(j)P_{ji} \quad \forall i, j \in \Omega. \quad (35)$$

Pööratavuse tingimust võib interpreteerida järgmiselt – Tasakaaluolekus on kiirus, mille-ga süsteem liigub seisundist i seisundisse j sama, mis liikumisel seisundist j seisundisse i . Seepärast nimetatakse (35) **detailse tasakaalu võrranditeks** (*detailed balance equations*) ("tasakaal" – võrdsustab liikumiskiirused seisundite vahel; "detailne" – kõigi võimalike seisundite paaride kohta).

Pööratavad ahelad on kasulikud, sest kui mittelahutuva ahela korral eksisteerib jaotus π , mis rahuldab detailse tasakaalu võrrandeid (35), siis π on selle ahela statsionaarne jaotus. Temast simuleerimiseks võime simuleerida ahelat vastava üleminekutõenäosuste maatriksiga. Maatriks P leitakse seosest (35).

Statsionaarsuse näitamiseks summeerime (35) mõlemad pooled üle j :

$$\begin{aligned} \sum_j \pi(i)P_{ij} &= \sum_j \pi(j)P_{ji} \\ \pi(i) \underbrace{\sum_j P_{ij}}_{=1} &= \sum_j \pi(j)P_{ji} \\ \pi(i) &= \sum_j \pi(j)P_{ji}, \forall i \end{aligned}$$

$$\pi = \pi P, \text{ kus } \pi \text{ on reavektor}$$

Saime statsionaarse jaotuse tingimuse.

Olgu eesmärgiks konstrueerida ahel jaotusega π . Markovi ahela konstrueerimine antud statsionaarse jaotusega π taandub üleminekutõenäosuste P_{ij} leidmisele, mis rahuldavad detailse tasakaalu võrrandit (35). See on alati võimalik.

Metropolise algoritm (1953)

Olgu meid huvitav jaotus $p_j, j \in \Omega, \sum_j p_j = 1$. Metropolis püstitas ja lahendas ülesande, kuidas konstrueerida Markovi ahel statsionaarse jaotusega π , nii et $\pi(j) = p_j, j \in \Omega$. Olgu Q mingi mittelahutuva üleminekutõenäosuste maatriks seisundite ruumil Ω , mis rahuldab sümmeetria tingimust:

$$Q_{ij} = Q_{ji}, \quad i, j \in \Omega.$$

Metropolis defineeris Markovi ahela $\{x^t, t \geq 0\}$ järgmiselt:

- kui $x^t = i$, siis üleminek seisundisse $x^{t+1} = j$ soovitatakse vastavalt üleminekutõenäosusele Q_{ij} ;

- soovitatud väärtus võetakse x^{t+1} jaoks vastu tõenäosusega $\min \left\{ 1, \frac{p_j}{p_i} \right\}$, vastasel juhul jääb ahel seisundisse $x^{t+1} = i$.

Vaatame saadud ahela üleminekutõenäosusi:

1. $i \neq j$

$$\begin{aligned} P_{ij} &= P(x^{t+1} = j, VVV | x^t = i) = \\ &= P(x^{t+1} = j | x^t = i) P(VVV | x^t = i) = \\ &= Q_{ij} \min \left\{ 1, \frac{p_j}{p_i} \right\}. \end{aligned}$$

VVV – Väärtus Võetakse Vastu. Kui $x^t = i$ on fikseeritud, siis soovitus j ja vastuvõtmine on sõltumatud.

2. $i = j$

$$\begin{aligned} P_{ii} &= P(x^{t+1} = i, VVV | x^t = i) + P(x^{t+1} \neq i, \overline{VVV} | x^t = i) = \\ &= Q_{ii} \min \left\{ 1, \frac{p_i}{p_i} \right\} + \sum_{j \neq i} Q_{ij} \left(1 - \min \left\{ 1, \frac{p_j}{p_i} \right\} \right) = \\ &= Q_{ii} + \sum_{i \neq j} Q_{ij} \left(1 - \min \left\{ 1, \frac{p_j}{p_i} \right\} \right). \end{aligned}$$

\overline{VVV} – soovitatud väärtust vastu ei võeta.

Kontrollime, kas meie jaotus rahuldab detailse tasakaalu võrrandit. Kui $i = j$, siis (35) on triviaalne samasus.

Kui $i \neq j$

1. Olgu $p_j > p_i$

$$\begin{aligned} p_i P_{ij} &= p_i Q_{ij} \min \left\{ 1, \frac{p_j}{p_i} \right\} = p_i Q_{ij} 1 = \\ &= Q_{ji} \frac{p_i}{p_j} p_j = Q_{ji} \min \left\{ 1, \frac{p_i}{p_j} \right\} p_j = p_j P_{ji} \end{aligned}$$

2. Olgu $p_j < p_i$ Saab näidata analoogiliselt võrrandi kehtivust. Samuti saab seda teha $p_j = p_i$ korral.

Järelikult rahuldab Metropolis algoritmi abil konstrueeritud Markovi ahel detailse tasakaalu võrrandeid. Seega p_j , $j \in \Omega$, $\sum_j p_j = 1$ on ahela statsionaarne jaotus. Ahela tekitamisel kasutatakse üleminekutõenäosusi Q_{ij} . Kui maatriks Q on aperioidiline, siis on seda ka P ja statsionaarne jaotus on ühtlasi ka piirjaotus. See lubab alustamist suvalisest algväärtusest nii et teatud aja pärast saame algoritmiga genereerida meid huvitavat jaotust. Sümmeetrilist

üleminekutõenäosuste maatriksit pole raske konstrueerida. Näiteks juhuslik ekslemine loenduva seisundite ruumi korral, kus üleminekutõenäosuste maatriks Q on:

$$Q_{ij} = \begin{cases} p, & j = i + 1 \\ p, & j = i - 1 \\ r, & j = i \\ 0, & j \neq i - 1, i, i + 1 \end{cases}$$

2.7 Pidevad seisundite ruumid

Olgu seisundite ruum $\Omega \subset \mathbf{R}$ reaaltelje osa. Tähistame seisundite i, j asemel seisundid $x, y \in \Omega \subset \mathbf{R}$. Olgu x^t monotoonne ahel. Pideval juhul ei saa konstrueerida üleminekutõenäosuste maatriksit. Selle asemel defineeritakse tõenäosus:

$$P(x, y) = P(x^{t+1} \leq y | x^t = x) \underset{mon}{=} P(x^1 \leq y | x^0 = x), \quad x, y \in \Omega.$$

See on x^{t+1} tinglik jaotusfunktsioon. Tingliku tiheduse saame leida diferentseerides:

$$p(x, y) = \frac{\partial P(x, y)}{\partial y}, \quad x, y \in \Omega.$$

Suurust $p(x, y)$ (või tema argumendist y sõltuvat osa) nimetatakse **üleminekutuumaks** (*transition kernel*). Sisuliselt on tegu x^{t+1} tingliku tihedusega kohal y tingimusel, et $x^t = x$.

Tinglik üleminekutõenäosus m sammu järel alates sammust t on

$$P^m(x, y) = P(x^{t+m} \leq y | x^t = x), \quad x, y \in \Omega.$$

Üleminekutuum m sammu järel on

$$p^m(x, y) = \frac{\partial P^m(x, y)}{\partial y}, \quad x, y \in \Omega.$$

Chapman-Kolmogorovi võrrandid pideval juhul:

$$P^{t+m}(x, y) = \int_{-\infty}^{+\infty} P^m(z, y) p^t(x, z) dz$$

$$\text{Võrdle! } P_{ij}(t+m) = \sum_{s \in \Omega} P_{is}(t) P_{sj}(m)$$

x^t marginaaljaotus π_t saadakse x^{t-1} jaotusest üleminekutuumaa abil:

$$\pi_t(y) = \int_{-\infty}^{+\infty} p(x, y) \pi_{t-1}(x) dx.$$

$$\text{Võrdle! } \sum_i \pi_{t-1}(i) P_{ij} = \pi_t(j)$$

Statsionaarne jaotus pideval juhul. Olgu Markovi ahel üleminekutuumaga $p(x, y)$. Kui kehtib seos

$$\pi(y) = \int_{-\infty}^{+\infty} p(x, y)\pi(x)dx,$$

siis jaotus π on ahela statsionaarne jaotus.

Näide 27

Juhuslik ekslemine pideval juhul.

$$x^t = x^{t-1} + \omega_t$$

$$\omega_t \sim f(\omega).$$

Pideval juhul liigutakse mingi juhusliku pikkusega sammu ω_t võrra. Suurused ω_t on sõltumatud.

Ahela algjaotus $x^0 \sim \pi_0$ olgu mingi pidev jaotus. Ajahetkel t oleme reaalteljel

$$x^t = x^0 + \omega_1 + \omega_2 + \dots + \omega_t$$

Üleminekutõenäosused:

$$\begin{aligned} P(x, y) &= P(x^{t+1} \leq y | x^t = x) = P(x^t + \omega_{t+1} \leq y | x^t = x) = \\ &= P(\omega_{t+1} \leq y - x^t | x^t = x) = P(\omega_{t+1} \leq y - x) = \int_{-\infty}^{y-x} f(\omega)d\omega \end{aligned}$$

Üleminekutuum

$$p(x, y) = \frac{\partial P(x, y)}{\partial y} = \frac{\partial}{\partial y} \int_{-\infty}^{y-x} f(\omega)d\omega = f(y - x)$$

Ahela marginaaljaotus on juhuslike suuruste summa $x^t = x^0 + \omega_1 + \omega_2 + \dots + \omega_t$ jaotus. Kui $\omega_i \sim N(0, \sigma^2)$ ja algjaotus $\pi_0 \sim N(a, \sigma_0^2)$, siis ahela marginaaljaotus sammul t on

$$\pi_t \sim N(a, \sigma_0^2 + t\sigma^2).$$

Kas juhuslik ekslemine on homogeenne Markovi ahel? Kas π_t on statsionaarne jaotus?

Piirteoreemid pideva seisundite ruumiga Markovi ahelate korral on põhimõtteliselt samad. Mõned muudatused on seisundite klassifitseerimisel. Üksiku diskreetse seisundi asemel vaatame seisundite hulka $A \subset \Omega$

- tagasipöördumise tõenäosus hulka A ;
- tagasipöördumise aeg hulka A ;
- perioodilisus hulka A suhtes.

Kõik olulised piirteoreemid kehtivad ka pideva seisundite ruumiga Markovi ahelate korral, kusjuures x^t võib olla mitmemõõtmeline (d -dimensionaalne). Olgu Markovi ahel x^t seisundite ruumiga $\Omega \subset \mathbf{R}^d$ ja statsionaarse jaotusega π . Piirteoreemid:

- mittelahutuvus + aperioidilisus \equiv leidub ühene piirjaotus π .
- (Suurte arvude seadus) Reaalarvulise funktsiooni $f(x^t)$ ergoodiline keskmine

$$\bar{f} = \frac{1}{N} \sum_{t=1}^N f(x^t)$$

koondub peaaegu kindlalt keskmiseks piirjaotuse π suhtes $E_\pi f(x)$, kui viimane eksisteerib.

- Tsentraalne piirteoreem. Ergoodiline keskmine, mis on sobivalt normeeritud, koondub standardseks normaaljaotuseks $N(0, 1)$

Tierney (1994) tõestas, et enamuse tänapäeval simuleerimiseks kasutatavaid Markovi ahelaid on mittelahutuvad ja aperioidilised, st ülaltoodud tulemused on rakendatavad.

Analoogiliselt diskreetse juhuga saab esitada detailse tasakaalu võrrandid (pööratav ahel):

$$\pi(x)p(x, y) = \pi(y)p(y, x), \quad x, y \in \Omega.$$

Integreerides mõlemad pooli x järgi, saame statsionaarsuse tingimuse $\pi(y) = \int_{-\infty}^{+\infty} p(x, y)\pi(x)dx$.

Pööratavad ahelad on väga kasulikud soovitava piirjaotusega π Markovi ahela määramiseks. Kui detailse tasakaalu võrrandid on rahuldatud, on tegemist Markovi ahelaga, mille statsionaarne jaotus on π .

3 MCMC ja Bayesi statistika

MCMC – Markov Chain Monte Carlo (Markovi ahelate Monte Carlo). Need on simuleerimismeetodid, kus meie huvialusest mitmemõõtmelisest jaotusest simuleeritakse Markovi ahelate abil. Valimiks on Markovi ahela realistasioon. Neid meetodeid on palju. Selles peatükis vaatleme tähtsamaid.

3.1 Gibbsi valik

Geman&Geman (1984) artiklis vaadeldi simuleerimist mitmemõõtmelisest Gibbsi jaotusest nii, et simuleeriti järjestikku tema täistinglikest tihedustest. Gelfand&Smith (1990) märgivad, et meetod on üldkasutatav (st ka teiste jaotuste korral). Sealtpeale kasutatakse meetodit laialdaselt, aga nimi jäi tema esimese kasutamise järgi *Gibbs sampling* (tookord valik Gibbsi jaotusest, nüüd lihtsalt Gibbsi valik).

Gibbsi tihedus mehhaanilises statistikas (*mechanical statistics*):

$$f(x_1, \dots, x_d) \propto \exp \left\{ -\frac{1}{kT} E(x_1, \dots, x_d) \right\},$$

kus $k > 0$ konstant, T on süsteemi temperatuur, E on süsteemi energia (positiivne funktsioon), x_i i -nda osakese positsioon või kiirus, (kujundi tuvastamise ülesandes on x_i i -nda pikseli värv). Tahetakse simuleerida d -mõõtmelist punkti (x_1, \dots, x_d) .

3.1.1 Definiitsioon ja omadused

Olgu huvialune jaotus $\pi(x)$, kus $x = (x_1, \dots, x_d)$. Tahame simuleerida $\pi(x)$ -st, aga see võib olla

- traditsiooniliste meetoditega liiga keeruline, võimatu, aeganõudev, kallis;
- appi tulevad MCMC-meetodid.

DEFINITSIOON: Gibbsi valik on MCMC-skeem simuleerimaks jaotusest $\pi(x)$, kus üleminekutuum moodustatakse $\pi(x)$ täistinglike jaotuste abil.

Tuletame meelde, et üleminekutuum on tinglik jaotus, mis vastavalt käesoleva hetke seisundile annab tõenäosused järgmise hetke seisundiks: diskreetsel juhul näiteks üleminekutõenäosuste maatriks P ja pideval juhul tinglik tihedus (harilikult mitmemõõtmeline).

Olgu täistinglikud jaotused $\pi_i(x_i) = \pi(x_i|x_{-i}) \forall i = 1, \dots, d$ teada sellel määral, et neist on võimalik simuleerida.

Gibbsi generaator – algoritm Gibbsi valiku jaoks on järgmine:

- 1) $t = 1$, anneme algväärtused $x^0 = (x_1^0, \dots, x_d^0)$;
- 2) genereerime uue väärtuse $x^t = (x_1^t, \dots, x_d^t)$ järgmiste sammudega :

$$\begin{aligned} x_1^t &\sim \pi(x_1 | x_2^{t-1}, \dots, x_d^{t-1}) \\ x_2^t &\sim \pi(x_2 | x_1^t, x_3^{t-1}, \dots, x_d^{t-1}) \\ &\dots \\ x_d^t &\sim \pi(x_d | x_1^t, \dots, x_{d-1}^t) \end{aligned}$$

- 3) $t = t + 1$, korrata punkte 2. ja 3.

Sammu 2 nimetatakse uuendamise sammuks. Tegemist on Markovi ahelaga, sest x^t saadakse kasutades x^{t-1} . Kui koondumine on toimunud, siis x^t on jaotusest π genereeritud väärtus. Suurim probleem: täistinglike jaotuste kuju peab olema teada.

Bayesi statistikas on parameetrid juhuslikud ja meid huvitab nende järeljaotus teatav mitmemõõtmeline jaotus.

Näide 28

(Carlin, Gelfand ja Smith, 1992). Olgu andmed y_1, \dots, y_n Poissoni jaotusest, kus kahtlustatakse parameetri muutumist punktis m , $m = 1, \dots, n$. Kui m on teada, siis saab kirja panna järgmise mudeli:

$$\begin{aligned} y_i | \lambda &\sim Po(\lambda), \quad i = 1, \dots, m; \\ y_i | \phi &\sim Po(\phi), \quad i = m + 1, \dots, n. \end{aligned}$$

Bayesi statistikas tuleb fikseerida parameetrite eeljaotused. Poissoni jaotusega vaatlustele on kaasjaotuseks gammajaotus:

$$\begin{aligned} \lambda &\sim G(\alpha, \beta); \\ \phi &\sim G(\gamma, \delta). \end{aligned}$$

$$m \sim \text{ühtlane diskreetne jaotus hulgal } \{1, \dots, n\} \Rightarrow P(m = i) = \frac{1}{n}$$

Eeljaotuse andmisel eeldame, et λ , ϕ , m on sõltumatud, α , β , γ , δ on hüperparameetrid, st teadaolevad konstandid. Gammajaotuse tihedusfunktsioon

$$G(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x \geq 0, \quad EX = \frac{\alpha}{\beta}, \quad DX = \frac{\alpha}{\beta^2}.$$

Järeljaotus:

$$\begin{aligned} \pi(\lambda, \phi, m) &\propto f(y_1, \dots, y_n | \lambda, \phi, m) p(\lambda, \phi, m) = \\ &= \left(\prod_{i=1}^m f_{Po}(y_i | \lambda) \right) \left(\prod_{i=m+1}^n f_{Po}(y_i | \phi) \right) f_G(\lambda | \alpha, \beta) f_G(\phi | \gamma, \delta) \frac{1}{n} \propto \\ &\propto \left(\prod_{i=1}^m e^{-\lambda} \lambda^{y_i} \right) \left(\prod_{i=m+1}^n e^{-\phi} \phi^{y_i} \right) \lambda^{\alpha-1} e^{-\beta \lambda} \phi^{\gamma-1} e^{-\delta \phi} = \\ &= \lambda^{\alpha + \sum_{i=1}^m y_i - 1} e^{-(\beta+m)\lambda} \phi^{\gamma + \sum_{i=m+1}^n y_i - 1} e^{-(\delta+n-m)\phi} \end{aligned}$$

Saame välja kirjutada täistinglikud jaotused. Tähistame $\pi_\lambda(\lambda) - \lambda$ täistinglik jaotus, kui ϕ ja m on fikseeritud. Täistingliku jaotuse leidmiseks:

- korjata kokku kõik λ -t sisaldavad tegurid;
- ülejäänud osa läheb jaotuse konstanti.

Saame jaotused:

$$\begin{aligned}\pi_\lambda(\lambda) &= G\left(\alpha + \sum_{i=1}^m y_i, \beta + m\right) \\ \pi_\phi(\phi) &= G\left(\gamma + \sum_{i=m+1}^n y_i, \delta + n - m\right) \\ \pi_m(m) &= \pi(m|\lambda, \phi) = \frac{\pi(\lambda, \phi, m)}{\pi(\lambda, \phi)}, \text{ väärtuse tõenäosus, kus} \\ \pi(\lambda, \phi) &= \sum_{m=1}^n \pi(\lambda, \phi, m)\end{aligned}$$

Kodune ülesanne nr. 3.1 Korjates kokku m -ist sõltuvad liikmed näita, et

$$\pi(m|\lambda, \phi) \propto \left(\frac{\lambda}{\phi}\right)^{\sum_{i=1}^m y_i} e^{m(\phi-\lambda)}$$

Neist jaotustest on lihtne simuleerida ja teostada Gibbsi valiku sammud (1)–(3)

$$x^0 = (\lambda^0, \phi^0, m^0), \quad \text{näiteks } \lambda^0 = \frac{\alpha}{\beta}, \quad \phi^0 = \frac{\gamma}{\delta}, \quad m^0 = 1.$$

Saame realisatsioonid järeljaotusest: $x^t = (\lambda^t, \phi^t, m^t) \sim \pi(\lambda, \phi, m)$.

Gibbsi valiku korral toimub liikumine koordinaattelgede suunas, d sammuga on punkt uuendatud. Joonisel on näha ahela tüüpiline trajektoor Gibbsi valiku korral, kui on tegemist kahemõõtmelise jaotusega ($d = 2$).

Järeldused:

- Gibbsi valik defineerib Markovi ahela. Punkt x^t , mis tekib iteratsioonil t , on juhuslik ja sõltub ainult ahela väärtusest iteratsioonil $t - 1$.
- Liikumine jaotuse määramispiirkonnas toimub telgede suunas.

- Markovi ahel on monotoonne. Üleminekutuum ei sõltu t väärtusest. Skeemi samm 2 tehakse samade tihedusfunktsioonidega. Gibbsi valiku üleminekutuum:

$$p(x, y) = \pi(y_1|x_2, \dots, x_d)\pi(y_2|y_1, x_3, \dots, x_d) \dots \pi(y_d|y_1, y_2, \dots, y_{d-1}, y_d). \quad (36)$$

Üleminekutuum $p(x, y)$ ei sõltu t -st, vaid üksnes ahela väärtustest x ja y .

- Kui jaotuse määramispiirkond on sidus, siis Gibbsi valikuga saadud Markovi ahel on mittelahutuv. Mittelahutuvuse tingimus $P(x, A) > 0, \forall A \subset \Omega$. Põhiliste statistikas kasutatavate jaotuste jaoks on Gibbsi valik mittelahutuv.

Näide 29

Olgu $x = (x_1, x_2)$ ühtlase jaotusega üle kahe piirkonna $A = A_1 \times A_2$ ja $B = B_1 \times B_2$. Jaotus

$$\pi(x) = \begin{cases} p_A, & x \in A \\ p_B, & x \in B \end{cases}$$

$$p_A + p_B = 1$$

Paiknegu piirkonnad A_1, A_2, B_1 ja B_2 järgmiselt:

Täistinglikud jaotused on samuti ühtlased jaotused. Kui $x^0 \in A$, siis x_1 väärtusi valitakse ühtlaselt üle A_1 ja x_2 väärtusi ühtlaselt üle A_2 . Ahel ei saa kunagi sattuda regiooni B . Analoogiline on olukord, kui $x^0 \in B$. Gibbsi valik määrab siin lahutuva Markovi ahela. Ahela piirjaotus on ühtlane jaotus piirkonnas A või B sõltuvalt ahela alguspunktist. Seega jaotusest $\pi(x)$ ei saa Gibbsi generaatoriga genereerida.

Näitame, et Markovi ahela, mille üleminekutuum on kujul (36) piirjaotus on just meie π . Eespool oli teoreem, et mittelahutuval, aperiodilisel M -ahelal leidub ühene piirjaotus. Leidugu ahelal piirjaotus π^∞ . Peab olema rahuldatud statsionaarsuse tingimus:

$$\pi^\infty(y) = \int \pi^\infty(x)p(x, y)dx.$$

Vaatleme kahemõõtmelist juhuslikku vektorit $d = 2$, $x = (x_1, x_2)$. Üleminekutuum lihtsustub:

$$p(x, y) = \pi(y_1|x_2)\pi(y_2|y_1), \quad (37)$$

kus $y = (y_1, y_2)$ on ahela uus väärtus.

Tähistame piirjaotuse marginaaljaotused $\pi^\infty(x_1), \pi^\infty(x_2)$. Statsionaarsuse tingimus kahemõõtmelisel juhul:

$$\pi^\infty(y) = \iint \pi^\infty(x)p(x, y)dx_1dx_2. \quad (*)$$

Vaatleme nüüd integraali, kus esineb ka meie jaotus π , st. tema tinglik jaotus.

$$\iint \pi(y_2|y_1)\pi^\infty(x_1|x_2)dx_1dy_2 = \underbrace{\int \pi(y_2|y_1)dy_2}_{=1} \underbrace{\int \pi^\infty(x_1|x_2)dx_1}_{=1} = 1$$

Korrutades mõlemaid pooli $\pi^\infty(x_2)$ -ga, saame

$$\pi^\infty(x_2) = \iint \pi(y_2|y_1) \underbrace{\pi^\infty(x_1|x_2)\pi^\infty(x_2)}_{\pi^\infty(x)} dx_1dy_2 = \iint \pi(y_2|y_1)\pi^\infty(x)dx_1dy_2 \quad (38)$$

Statsionaarsuse tingimusest (*) saame marginaaltiheduse $\pi^\infty(y_1)$, integreerides y_2 järgi.

$$\begin{aligned} \pi^\infty(y_1) &= \iiint \pi^\infty(x)p(x,y)dx_1dx_2dy_2 = \\ &= \iiint \pi^\infty(x)\pi(y_1|x_2)\pi(y_2|y_1)dx_1dx_2dy_2 = \\ &= \int \pi(y_1|x_2)\pi^\infty(x_2)dx_2. \end{aligned}$$

Kuna $\pi^\infty(y_1) = \int \pi(y_1|x_2)\pi^\infty(x_2)dx_2$ kehtib ja teiseltpoolt marginaaljaotus $\pi^\infty(y_1)$ saadakse üheselt ühisjaotuse $\pi^\infty(y_1|x_2)\pi^\infty(x_2)$ integreerimisel, siis järelikult $\pi(y_1|x_2) = \pi^\infty(y_1|x_2)$. M -ahela piirjaotuse täistinglik jaotus on jaotuse π täistinglik jaotus. Sama saab näidata ka teise koordinaadi kohta. Seega piirjaotuse π^∞ , täistinglikud jaotused on jaotuse π täistinglikud jaotused.

Sama saab näidata ka $d > 2$ korral jagades parameetervektori x kahte ossa x_i ja x_{-i} , $i = 1, \dots, d$. Besag (1974) on näidanud, et teatud küllaltki nõrkadel tingimustel kõigi täistinglike jaotuste hulk määrab ära ühisjaotuse.

Seega Markovi ahel üleminekutuumaga (36) koondub meid huvitavaks jaotuseks π ja iteratiivne valikuskeem (1)–(3) genereerib teatud hetkest alates väärtusi sellest jaotusest.

3.2 Praktilise rakendamise probleeme

3.2.1 Täistingliku tiheduse leidmine

Gibbsi valiku teostamiseks on vaja teada mitmemõõtmelise jaotuse täistinglikke tihedusi. See on lihtne, kui normeerivat konstanti pole vaja teada.

Olgu $f(x_1, x_2, \dots, x_d)$ d -mõõtmeline tihedus (ka tõenäosusfunktsioon). Täistingliku tiheduse $f_i(x_i) = f(x_i|x_{-i})$ saamiseks korjatakse kokku ühistiheduse kõik liikmed (tegurid), mis sisaldavad argumenti x_i . Suurust x_i mittedisaldavad tegurid lähevad normeerivasse konstanti.

Näide 30

Vaatame mitmemõõtmelist normaaljaotust

$$f(x|\mu, \Sigma) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x - \mu)' \Sigma^{-1} (x - \mu)\right\}.$$

Olgu dimensioon $d = 2$. Tähistame $\mu = (\mu_1, \mu_2)$, $\Sigma^{-1} = (\sigma^{ij})$. Siis (veendu ise!)

$$f(x_1, x_2) \propto \exp\left\{-\frac{1}{2}[(x_1 - \mu_1)^2 \sigma^{11} + 2(x_1 - \mu_1)(x_2 - \mu_2) \sigma^{12} + (x_2 - \mu_2)^2 \sigma^{22}]\right\}$$

Siit saame x_1 tingliku tiheduse tuuma:

$$f(x_1|x_2) \propto \exp\left\{-\frac{1}{2}[(x_1 - \mu_1)^2 \sigma^{11} + 2(x_1 - \mu_1)(x_2 - \mu_2) \sigma^{12}]\right\}$$

Toodud avaldisest piisab simuleerimaks sellest tihedusest. Siin saame siiski jaotust täpsustada. Kuna e astmes on argumendi x_1 ruutfunktsioon, on tegu 1-mõõtmelise normaaljaotusega.

Kodune ülesanne nr. 3.2 Leia tinglik keskväärtus $\mu_{1,2}$ ja dispersioon $\sigma_{1,2}^2$. Selleks teisenda tingliku tiheduse avaldis argumendi x_1 suhtes normaaljaotuse kujule. Pane kirja jaotuskarakteristikud ka teise komponendi tingliku jaotuse jaoks.

Kodune ülesanne nr. 3.3 Tuletada Dirichlet' jaotuse täistinglikud tihedused normeeriva konstandi täpsuseni, kui

$$f_D(x_1, x_2, \dots, x_k) = \frac{\Gamma\left(\sum_{i=1}^k \alpha_i\right)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i-1}, \quad 0 < x_i < 1 (\alpha_i > 0), \quad \sum_{i=1}^k x_i = 1.$$

Pööra tähelepanu määramispiirkonnale.

Kodune ülesanne nr. 3.4 Tuletada 2-mõõtmelise tiheduse tinglikud tihedused:

$$f(x, y) = \frac{2(x+y)}{(x+1)^3(y+1)^3}, \quad x, y \in (0, \infty).$$

3.3 Simuleerimine täistinglikust tihedusest

Kasutada võib kõiki ühedimensionaalsete jaotuste jaoks leiduvaid simuleerimismeetodeid. Kuna aga sageli pole normeeriv konstant teada, siis on eriti sobilikuks meetodiks valikumee-
tod (*rejection sampling*) koos oma erijuhu Neumanni meetodiga.

Pöördjaotusfunktsiooni meetod – Olgu x pidev juhuslik suurus ja F tema jaotusfunktsioon.

Põhitulemus: Kui $x \sim F$ siis $u = F(x) \sim U(0, 1)$.

Kontrollime:

$$F_u(y) = P(u \leq y) = P(F(x) \leq y) = P(x \leq F^{-1}(y)) = F(F^{-1}(y)) = y, \quad \text{kui } 0 < y < 1;$$

$F_u(y) = 0$, kui $y < 0$ ja $F_u(y) = 1$, kui $y > 1$. Järelikult $F_u(y)$ on $U(0, 1)$ jaotusfunktsioon ja u on $U(0, 1)$ juhuslik suurus.

Simuleerimisel on vaja vastupidi: – saada $U(0, 1)$ juhuslikust suuruselt suvaline j.s. Kuna $x \sim F$ ja F pöördfunktsioon on F^{-1} , siis seosest $u = F(x) \sim U(0, 1)$ jäeldub

$$x = F^{-1}(u) \sim F.$$

Lõpuks, selleks, et genereerida realisatsiooni x jaotusest F , genereerime arvu $u \sim U(0, 1)$ ja arvutame $x = F^{-1}(u)$. NB! Meetod vajab jaotusfunktsiooni valemit.

Näide 31

EkspONENTJAOTUS $x \sim \text{Exp}(\lambda)$, $x > 0$, $\lambda > 0$, $f(x) = \lambda e^{-\lambda x}$ Leiname: $F(x) = 1 - e^{-\lambda x}$, võrdsustame $u = 1 - e^{-\lambda x}$, millest avaldame $x = -\frac{1}{\lambda} \log(1 - u)$. Kuna $u \sim 1 - u$ (sama $U(0, 1)$ jaotusega), siis saame $\text{Exp}(\lambda)$ realisatsioon, arvutades $x = -\frac{1}{\lambda} \log(u)$.

Kodune ülesanne nr. 3.5 Tuleta eeskiri Weibull'i jaotusest simuleerimiseks, kui

$$f(x) = \frac{cx^{c-1}}{b^c} e^{-(x/b)^c}, \quad x \in (0, \infty), \quad b, c > 0.$$

Valikumeetod – On üks kasulikumaid meetodeid keerulisest jaotusest simuleerimiseks. Teada peab olema tiheduse tuuma funktsionaalne kuju. Normeerivat konstanti pole vaja teada. Tahame simuleerida tihedusest $\pi(\cdot)$,

1. Leiname lihtsama tiheduse $q(\cdot)$ ja sellise konstandi c , et $\pi(x) \leq cq(x)$ iga x ;
2. Simuleerime väärtuse x tihedusest $q(\cdot)$ ja u tihedusest $U(0, 1)$;
3. Aktsepteerime x $\pi(\cdot)$ realisatsioonina, kui $cuq(x) \leq \pi(x)$, muidu samm 2.

Tihedust $q(x)$ nimetatakse ümbriseks, ümbris- ehk ettepanekutiheduseks.

Tõestamiseks, et x tihedus on π kasutame Bayesi teoreemi:

$$f(x|cuq(x) \leq \pi(x)) = \frac{\Pr(cuq(x) \leq \pi(x)|x)q(x)}{\int \Pr(cuq(x) \leq \pi(x)|x)q(x)dx} = \frac{[\pi(x)/cq(x)]q(x)}{\int [\pi(x)/cq(x)]q(x)dx} = \frac{\pi(x)}{\int \pi(x)dx}.$$

Näeme, et tinglik tihedus $f(\cdot)$ on normeeritud, isegi kui π seda pole. Meetodi efektiivsust iseloomustab vastuvõtmistõenäosus:

$$\Pr(cuq(x) \leq \pi(x)) = \int \Pr(cuq(x) \leq \pi(x)|x)q(x)dx = \frac{\pi(x)}{cq(x)}q(x)dx = \frac{1}{c} \int \pi(x)dx.$$

Kui π on normeeritud tihedus, siis on efektiivsus $1/c$. Mida suurem c , seda väiksem on meetodi efektiivsus.

Kui tiheduse määramispiirkond on lõplik, saab ümbrisjaotusena kasutada ühtlast jaotust (kiire, aga efektiivsus madal). Seda erijuhtu nimetatakse Neumanni meetodiks.

Kodune ülesanne nr. 3.6 Panna kirja algoritm simuleerimiseks tihedusest $\pi(x) = \text{const.} \cdot \sin^2(x)e^{-x}$, $x \in [0, \infty)$.

Kodune ülesanne nr. 3.7 Simuleeri Dirichlet jaotuse täistinglikust Neumanni meetodil.

3.3.1 Valimi moodustamine

Gibbsi valikuskeemi rakendamisel tekib mitmeid teisi praktilisi probleeme:

- kui pikk on sissepõlemisperiood?
- kas valim moodustada ühe või mitme ahela elementidest?
- kuidas muuta meetod efektiivsemaks?

Olgu eesmärgiks saada valim suurusega n jaotusest π . Kulugu ahela koondumiseks m iteratsiooni. Võimalused:

1. Käivitada n sõltumatut ahelat eri algväärtustest. Pärast koondumist (m iteratsiooni järel) võtta igast ahelast üks väärtus ja tulemuseks on n sõltumatut realisatsiooni jaotusest π .

Selle meetodi korral tuleb genereerida $m \times n$ väärtust. Kui ahelate algväärtused on sõltumatud, siis koosneb valim *sõltumatustest* realisatsioonidest jaotusest π . Sõltumatus on saavutatav, kui algväärtused on kõik erinevad ja soovitatavalt suurema dispersiooniga, kui järeljaotuses.

2. Genereerida üks Markovi ahel. Pärast koondumist on kõik ahela väärtused jaotusest π . Valim suurusega n saadakse, võttes pärast koondumist n järjestikust ahela realisatsiooni. Vaja genereerida $m + n$ väärtust. Puudused:
 - valimi elemendid ei ole sõltumatud;
 - kui ahela autokorrelatsioon on suur, siis võtab palju aega, kuni ahel täidab adekvaatselt kogu seisundite ruumi (n tuleb võtta väga suur);
 - sõltuvust tuleb arvestada tsentraalse piirteoreemi juures.
3. Käivitatakse üks ahel. Pärast koondumist võetakse valimisse iga k -s ahela element. Mida suurem on k seda väiksem on ahela elementide korrelatsioon. Piisavalt suure k väärtuse korral on saadud valimi elemendid suhteliselt sõltumatud. Vaja genereerida $m + k \times n$ väärtust. Efektiivsuses võitu pole ja hinnangud ei ole nii täpsed, kui saadakse kõigi ahela elementide kasutamisel.
4. Käivitada $l < 10$ sõltumatut ahelat ja pärast koondumist võtta igast ahelast n/l elementi. Vaja genereerida $l(m + n/l) = lm + n$ väärtust. Arvutuslikult kaotatakse efektiivsuses võrreldes ühe ahelaga ja võidetakse võrreldes sõltumatu valimiga.

Praktikas on leitud, et (1) meetod on liiga ebaefektiivne ja mittevajalik. Suurim debatt käib meetodi (2) ja (4) vahel.

3.3.2 Uuendamise strateegiad

Kirjelatud Gibbsi skeemi kohaselt uuendati vektori komponente alati ühes determineeritud järjekorras $1 \rightarrow 2 \rightarrow \dots \rightarrow d$ Selle asemel on palju teisi võimalikke uuendamise strateegiaid.

1. Pööratav Gibbsi generaator. Igas iteratsioonis uuendatakse komponendid algul $1 \rightarrow 2 \rightarrow \dots \rightarrow d$ ja seejärel vastupidises järjekorras $d \rightarrow \dots \rightarrow 2 \rightarrow 1$. Iga iteratsioon sisaldab $2d$ uuendamist.
2. Juhusliku järjekorraga Gibbsi generaator. Iga parameetervektori komponendi jaoks on fikseeritud positiivne tõenäosus, mille abil valitakse igas iteratsioonis üks komponent i vahemikust $\{1, \dots, d\}$ ja uuendatakse ainult seda komponenti. Võrreldavuse tagamiseks determineeritud järjekorraga, vaadeldakse ühe iteratsioonina d sellist uuendamist.
3. Juhuslikud permutatsioonid. Igal iteratsioonil valitakse $\{1, \dots, d\}$ juhuslik permutatsioon ja komponente külastatakse selles järjekorras.

Näidatud on, et kui $\pi = N_d(\mu, \Sigma)$, siis deterministlik järjekord tagab kiirema koondumise kui juhuslik järjekord, kui Σ kolm-diagonaalne.

3.3.3 Valimi kasutamine

Olgu saadud valim d -dimensionaalsest jaotusest π , $x = (x_1, \dots, x_d) \sim \pi(x)$. Olgu tegemist järjestikuste iteratsioonidega ühest Markovi ahelast.

Tähistame vektori x i -nda komponendi valimi: $x_i^t, t = 1, \dots, n$, s.o. valim x_i marginaaljaotusest $\pi(x_i)$. Olgu antud funktsioon $f(x)$, siis ergoodilise teoreemi järgi:

$$\bar{f} = \frac{1}{n} \sum_{t=1}^n f(x^t) \rightarrow E_{\pi} f(x) = \int f(x) d\pi$$

Tänu ergoodilisele teoreemile saame punkthinnangud teoreetilistele keskväärtustele :

$$f(x) = x, \quad \text{siis } \bar{x} = \frac{1}{n} \sum_{t=1}^n x_i^t = \hat{E}_{\pi}(x_i)$$

$$f(x) = x^2, \quad \text{siis } \frac{1}{n} \sum_{t=1}^n (x_i^t)^2 = \hat{E}_{\pi}(x_i^2)$$

Jaotuse π i -nda komponendi dispersiooni hinnang:

$$\frac{1}{n} \sum_{t=1}^n (x_i^t - \bar{x})^2 = \hat{E}_{\pi}(x_i - E_{\pi} x_i)^2 = \hat{\sigma}_{x_i}^2$$

Valimi keskmine ja valimi dispersioon on mõjusad hinnangud funktsiooni $f(x)$ keskväärtusele ja dispersioonile.

Ergoodilise keskmise dispersiooni $D_\pi \bar{f}$ hindamine on sõltuvuse tõttu keerulisem. Olgu esimese t vaatluse ergoodiline keskmine

$$\bar{f}^t = \frac{1}{t} \sum_{k=1}^t f^k \rightarrow E_\pi f(x) = \int f(x) d\pi.$$

Siin $f^k = f(x^k)$ ja x^k on ahela element. Tuletasime teoreetilise valemi:

$$D_\pi \bar{f}^t = \frac{\tau_t^2}{t}$$

$$\tau_t^2 = \sigma^2 \left(1 + s \sum_{k=1}^{t-1} \frac{t-k}{t} \right), \quad \tau_t^2 \xrightarrow{t \rightarrow \infty} \tau^2$$

$$\tau^2 = \sigma^2 \left(1 + 2 \sum_{k=1}^{\infty} \rho_k \right)$$

Autokorrelatsioon $\rho_k = \frac{\gamma_k}{\sigma^2}$.

Autokovariatsioon $\gamma_k = Cov(f^t, f^{t+k})$, $\sigma^2 = \gamma_0$.

Valemid sisaldavad tundmatuid suurusi γ_k , ρ_k . Kuidas neid suurusi hinnata genereeritud Markovi ahela väärtuste abil?

Otsene meetod – hinnata $D_\pi \bar{f}^t$ liikmeid valimimomentidega

$$\hat{\rho}_k = \begin{cases} \frac{\hat{\gamma}_k}{\hat{\sigma}^2}, & k \leq k^* \\ 0, & k > k^* \end{cases}$$

k^* – selline k väärtus, millest suuremate korral f^t , f^{t+k} pole enam korreleeritud.

$$\hat{\sigma}^2 = \hat{\gamma}_0 = \hat{D}_\pi f^t = \frac{1}{t} \sum_{k=1}^t (f^k - \bar{f}^t)^2$$

$$\hat{\gamma}_k = \frac{1}{t} \sum_{j=1}^t f^j f^{j+k} - (\bar{f}^t)^2$$

Saadud hinnang ei ole mõjus: $\hat{\tau}_t^2 \not\xrightarrow{t \rightarrow \infty} \tau^2$.

Esitatud ja uuritud on mitmeid erinevaid variante konstandi $1/t$ valikuks, mis muudaksid antud hinnangu mõjusaks. Mõjusus on tagatud kõrgemat järku autokorrelatsioonidele väiksema kaalu andmisega.

Kirjanduses on välja on pakutud mitmeid aegridade mudelitel (autoregressiivne mudel, libiseva keskmise mudel) baseeruvaid meetodeid. Järgmine lihtne meetod on muutnud selle populaarseks Markovi ahelate valimikeskmise dispersiooni hindamisel.

Plokkide meetod (*Batch means method*) – Markovi ahel x^t jagatakse m -elemendilisteks plokkideks (k plokki)

$$x^t \quad \underbrace{\underbrace{\dots}_m \quad \underbrace{\dots}_m \quad \underbrace{\dots}_m \quad \underbrace{\dots}_m}_{k} \quad t = 1 \dots mk$$

Idee seisneb selles, et üksteisest piisvalt kaugel seisvad elemendid on sõltumatud.

Leitakse iga ploki keskmised $\bar{f}_1, \bar{f}_2, \dots, \bar{f}_k$, mida vaadatakse kui sõltumatuid sama jaotusega juhuslikke suurusi. Kehtib,

$$\begin{aligned} E_\pi \bar{f}_i &= E_\pi f(x) \\ D_\pi \bar{f}_i &\sim \frac{\tau^2}{m} = \frac{k\tau^2}{km} = kD_\pi \bar{f}^t, \text{ kui } m \text{ on piisavalt suur.} \end{aligned}$$

Siin \bar{f}^t on kõigi t elemendi keskmine, mis on ühtlasi gruppide keskmiste keskmine:

$$\bar{f}^t = \frac{1}{k} \sum_{i=1}^k \bar{f}_i$$

Suuruste \bar{f}_i sõltumatuse tõttu ja kuna kõik on sama jaotusega, võib nende teoreetilist dispersiooni hinnata valimidispersiooniga:

$$\hat{D}_\pi \bar{f}_i = \frac{1}{k-1} \sum_{i=1}^k (\bar{f}_i - \bar{f}^t)^2.$$

Nüüd:

$$\hat{D}_\pi \bar{f}^t = \frac{1}{k} \hat{D}_\pi \bar{f}_i = \frac{1}{k(k-1)} \sum_{i=1}^k (\bar{f}_i - \bar{f}^t)^2.$$

Kuna $D_\pi \bar{f}^t = \frac{\tau^2}{t}$, siis saame hinnangu ka τ^2 -le.

Statistilised otsustused täpsuse kohta rajatakse tsentraalsele piirteoreemile. Et olla oma väidetes kindlamal poolel, kasutatakse t -jaotust:

$$\frac{\bar{f}^t - E_\pi f(x)}{\sqrt{\hat{D}_\pi \bar{f}^t}} = \frac{\sqrt{t}(\bar{f}^t - E_\pi f(x))}{\hat{\tau}} \sim t_{(k-1)}$$

Märkus Ergoodilise keskmise standardhälbe hinnangut $\sqrt{\hat{D}_\pi \bar{f}^t}$ nimetatakse WinBUGS'i väljundis Monte-Carlo veaks (MC-error). Ergoodiline keskmine hindab parameetri järeljaotuse keskmist ja MC-viga omakorda selle hinnangu viga. Simulatsioonide arvu suurendades võime teoreetiliselt MC-vea kuitahes väikeseks teha. Arvutustehnilised probleemid (arvu kujutamise lõplik täpsus, generaatorid on on lõpliku perioodiga jt.) panevad sellele siiski piiri.

stats produces summary statistics for the variable, pooling over the chains selected. The required percentiles can be selected using the percentile selection box. The quantity reported in the MC error column gives an estimate of τ/\sqrt{N} , the Monte Carlo standard error of the mean. The batch means method outlined by Roberts (1996; p.50) is used to estimate τ .

Tõenäosusintervalli (*Credibility interval*) hindamine

Sagedusstatistikas kasutame terminit usaldusvahemik (*confidence interval*) parameeter on konstant ja vahemik juhuslik. Bayesi statistikas kasutame terminit tõenäosusintervall (*credibility interval*) konstantne vahemik CI , millesse juhuslik parameeter θ kuulub.

$$P(\theta \in CI) = 1 - \alpha$$

Tõenäosusintervall konstrueeritakse valimi kvantiilide abil. Kui meil on genereeritud $n = 1000$ väärtust, siis moodustame variatsioonirea $f(x_i^t)$ väärtustest. Tõenäosusintervalli piirideks võtame 25-nda ja 975-nda variatsioonirea elemendi $(f^{(25)}, f^{(975)})$ on vahemik, kuhu $f(x_i)$ väärtused kuuluvad tõenäosusega 0.95. Üldjuhul $(1 - \alpha)$ -tõenäosusintervalli hinnanguks on valimikvantiilid $(f^{[(\alpha/2)n]}, f^{[(1-\alpha/2)n]})$.

Meenutagem, et Bayesi statistikas on x_i^t parameetri väärtuste jada θ_i^t , kus i viitab parameetervektori komponendile ja t ahela elemendile. Valimi kvantiilid on mõjusad hinnangud parameetri jaotuse kvantiilidele $q_{0.025}$ ja $q_{0.975}$.

3.4 Praktilise rakendamise probleeme (jätkub)

Olgu $x = (x_1, \dots, x_d) \sim \pi(x)$ simulatsioonid $x^t = (x_1^t, \dots, x_d^t)$. Vektori x marginaaltiheduse $\pi(x_i)$ hindamine valimilt:

- väärtuste x_i^t histogramm (korjame kokku igast vektorist x^t inda koordinaadi);
- x_i^t silutud histogramm;
- tinglike tiheduste abil:

Hindamine tinglike tiheduste abil on sobiv Gibbsi valiku jaoks, sest tinglikud tihedused on teada.

$$\pi(x_i) = \int \pi(x) dx_{-i} = \int \pi(x_i | x_{-i}) \pi(x_{-i}) dx_{-i}$$

Integraali Monte-Carlo hinnang:

$$\hat{\pi}(x_i) = \frac{1}{n} \sum_{t=1}^n \pi(x_i | x_{-i}^t), \quad (39)$$

kus x_{-i}^t , $t = 1, \dots, n$ on valim marginaaljaotusest $\pi(x_{-i})$. Vastavalt ergoodilisele teoreemile $\hat{\pi}(x_i) \rightarrow \pi(x_i)$. Hinnangud $\hat{\pi}(x_i)$ on pidevad pideva x_i korral. Seda meetodit nimetatakse tihedusfunktsiooni hinnanguks kernel- ehk tuumameetodil.

Näide 32

Marginaaljaotuse hindamine kahemõõtmelise normaaljaotuse korral tuumameetodil.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

kus $\rho = \text{Cor}(x_1, x_2)$. Leiti tinglikud jaotused:

$$\begin{aligned} x_1|x_2 &\sim \pi(x_1|x_2) = \pi(x_1|x_{-1}) = N(\rho x_2, 1 - \rho^2) \\ x_2|x_1 &\sim \pi(x_2|x_1) = \pi(x_2|x_{-2}) = N(\rho x_1, 1 - \rho^2) \end{aligned}$$

Leiame x_1 marginaaljaotuse hinnangu $\pi(x_1)$ tuumameetodil. Selleks kasutame realisatsioone Markovi ahelast:

$$x^t : \begin{pmatrix} x_1^1 \\ x_2^1 \end{pmatrix} \dots \begin{pmatrix} x_1^t \\ x_2^t \end{pmatrix}$$

$\pi(x_1)$ leidmiseks kasutame punkte $x_2^1 \dots x_2^t$. Rakendame valemit (39)

$$\begin{aligned} \hat{\pi}(x_1) &= \frac{1}{n} \sum_{t=1}^n \pi(x_1|x_{-1}^t) = \frac{1}{n} \sum_{t=1}^n \pi(x_1|x_2^t) = \\ &= \frac{1}{n} \sum_{t=1}^n \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp \left\{ -\frac{(x_1 - \rho x_2^t)^2}{2(1-\rho^2)} \right\} \end{aligned}$$

Hinnangu arvutamiseks peame teadma x_2^t valimiväärtusi.

3.4.1 Ümberparametriseerimine

Juhusliku vektori komponentide tugeva sõltuvuse korral, on Gibbsi valikuga saadud ahela liikumistee väike.

Probleemid:

- läheb kaua aega, enne kui jaotuse kandja läbitakse
- kui valim pole piisavalt suur, siis saadakse valim, kus mõned seisundite ruumi osad on ebaproportsionaalselt esindatud

Probleemi lahenduseks on **ümberparametriseerimine** – Selle meetodi korral teisendatakse juhuslikku vektorit x , nii et selle komponendid oleksid ligikaudu sõltumatud. Põhiline meetod on lineaarteisendus A , nii et $y = Ax$ kovariatsioonimaatriks oleks diagonaalmaatriks.

Näide 33

Regressioonimudel.

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

Regressioonimudelis on järeljaotuse $\pi(\alpha, \beta | \sigma^2)$ komponendid α ja β tugevalt sõltuvuses. Teisendame (tsentreerime):

$$y_i = \alpha^* + \beta(x_i - \bar{x}) + \varepsilon_i,$$

kus $\alpha^* = \alpha + \beta \bar{x}$. Osutub, et α^* ja β on sõltumatud tingimusel σ^2 fikseeritud. Saab näidata, et $\pi(\alpha^*, \beta | \sigma^2) = \pi(\alpha^* | \sigma^2) \pi(\beta | \sigma^2)$. Genereerides jaotusest $\pi(\alpha^*, \beta, \sigma^2)$ vektoreid $(\alpha^*, \beta, \sigma^2)$, siis koondumine toimub kiiremini ja avaldisest saame leida ka α hinnangu.

Kodune ülesanne nr. 3.8 Mis on lineaarteisenduse maatriksiks, mis teisendab vektori $(\alpha, \beta, \sigma^2)$ vektoriks $(\alpha^*, \beta, \sigma^2)$.

3.4.2 Plokkimine

Juhusliku vektori sõltuvad komponendid võetakse blokki, leitakse plokkide täistinglikud jaotused. Genereeritakse mitmemõõtmelisest jaotusest. Sel juhul koondumine piirjaotuseks kiireneb.

Kolmemõõtmelise vektori korral $x = (x_1, x_2, x_3)$ võtame komponendid x_1 ja x_2 ühte plokki ja leiame tihedused $\pi(x_1, x_2 | x_3)$ ja $\pi(x_3 | x_1, x_2)$. Genereerime tasandil, mis lõikab telge x_3 fikseeritud punktis x_3 . Seejärel toimub genereerimine telje x_3 suunas. Plokkimise tulemusena paranevad ka valimihinnangud.

3.4.3 Koondumisdiagnostika

Teame, et Markovi ahel vajab teatud sissepõlemisperioodi, enne kui toimub koondumine piirjaotuseks ja hakatakse simuleerima meie jaotusest π .

Vaatame WinBUGS-is realiseeritud koondumisdiagnostikat. Kasutatakse Brooks-Gelman-Rubin-i (BGR) koondumisstatistikut. Statistik \hat{R} , mis on intervallipikkuste suhe konstrueeritakse järgmiselt:

1. Pannakse jooksma m paralleelset ahelat erinevate algväärtustega ja tehakse n simulatsiooni igas. Iga individuaalse ahela jaoks arvutatakse empiirilise $(1 - \alpha)$ tõenäosusintervalli laius; st. $\alpha/2$ ja $(1 - \alpha/2)$ kvantiilide vahe n simuleeritud väärtuse hulgas. Saame m nn. ahelasisest intervalli laiust.
2. Kogu vaatluste hulga mn korral (pooled sequence) arvutatakse samuti $(1 - \alpha)$ intervalli laius.

3. \hat{R} defineeritakse seosega

$$\hat{R}_{intervall} = \frac{\text{ühise ahela intervalli laius}}{\text{ahelasiseste intervallilaiuste keskmine}}$$

$\hat{R}_{intervall}$ koondub 1-ks, kui ahel koondub.

BGR diag WinBUGS-is arvutab intervalli $\hat{R}_{intervall}$ $\alpha = 0.2$ jaoks. Ühise ahela 80% intervall on roheline, ahelasiseste 80% intervallide keskmine on sinine ja nende suhe $\hat{R}_{intervall}$ on punane. Paremaks graafikul kujutamisesks on ühise ja ahelasiseste intervallide pikkused normeeritud nii, et üleüldine maksimum on 1. Koondumine on saanud, kui $\hat{R}_{intervall}$ on 1 lähedal, ja nii ühine kui ka ahelasisesed intervallipikkused on stabiilsed.

Näide koondumise graafikust

Arvulised väärtused võib saada aknasse, kui topelt-klikkida joonisel, millele järgneb *ctrl*-vasakuhiirenupuklikk aknal.

Arvulise väljundi näide:

```
-----80% interval-----
                Unnormalized                Normalized as plotted
End iteration  of pooled mean within  of pooled mean within  BGR ratio
  of bin      chains    chain      chains    chain
51            2.519    2.926    0.8606    1.0        0.8606
101           2.533    2.652    0.8657    0.9061     0.9554
151           2.322    2.445    0.7935    0.8354     0.9498
201           2.399    2.416    0.8199    0.8254     0.9933
251           2.48     2.485    0.8474    0.849      0.9981
301           2.587    2.585    0.884     0.8832     1.001
351           2.542    2.595    0.8687    0.8866     0.9798
401           2.501    2.543    0.8546    0.8691     0.9833
```

Rohkem informatsiooni:

WinBUGS. User manual. The Inference Menu. Samples.

Brooks, S. P. and Gelman, A. (1998) Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*. 7, 434-455.

http://www.amstat.org/publications/jcgs/pdf_98/Brooks.pdf

3.5 Metropolis-Hastingsi algoritmid

Algoritmid põhinevad kahel artiklil:

- Metropolis, N. et al (1953) Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, 21, 1087–91
- Hastings, W. K. (1970) Monte-Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57, 97–109.

3.5.1 Definiitsioon ja omadused

Olgu π jaotus, millest tahame genereerida. Sageli on otsene *i.i.d.* genereerimine raske või võimatu. Osutub, et teades väärtust x , saame genereerida järgmise väärtuse ja nii üha edasi. Tekib Markovi ahela trajektoor. Ahela üleminekut $x \rightarrow y$ iseloomustab üleminekutuum $p(x, y)$, mis on tinglik tihedus või tinglike tõenäosuste vektor. Et Markovi ahel oleks seotud meie jaotusega π , leiame üleminekutuuma nii, et $\pi(\cdot)$ oleks ahela piirjaotus. Leidmisel on lihtsaim lähtuda pööratava ahela nõudest, st detailse tasakaalu võrranditest:

$$\pi(x)p(x, y) = \pi(y)p(y, x) \quad \forall x, y.$$

Detailse tasakaalu võrrandi rahuldamine on piisav tingimus, et $\pi(\cdot)$ oleks ahela, mille üleminekutuum on $p(x, y)$, piirjaotus. See ei ole tarvilik tingimus. On palju teisi Markovi ahelaid,

Metropolis-Hastingsi algoritmide üleminekutuum konstrueeritakse kujul:

$$p(x, y) = q(x, y)\alpha(x, y), \quad x \neq y, \quad (40)$$

kus

$q(x, y)$ – suvaline üleminekutuum, ettepanekutuum (tinglik tihedus või tinglike tõenäosuste vektor fikseeritud x korral),

$\alpha(x, y)$ – vastuvõtmise tõenäosus $\alpha(x, y) \in [0, 1]$.

Seos (40) ei integreeru üheks üle y . Kui $q(x, y)$ on tinglik tihedus, siis $\int_{-\infty}^{\infty} q(x, y)dy = 1$, aga korrutamine $\alpha(x, y)$ -ga teeb tulemust väiksemaks. Et $p(x, y)$ oleks y tinglik tihedus, peab ahel saama jääda samasse seisundisse positiivse tõenäosusega:

$$p(x, x) = 1 - \int q(x, y)\alpha(x, y)dy. \quad (41)$$

Üleminekutuum on nn segajaotus:

- kui $x \neq y$, siis y jaoks on tegemist normeerimata tihedusega kujul (40);
- punktis $x = y$ on tegemist aatomiga, mille tõenäosusmass on (41).

Hastings (1970) defineeris vastuvõtmistõenäosuse kujul:

$$\alpha(x, y) = \min \left[1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right] \quad (42)$$

DEFINITSIOON: Algoritmid üleminekutuumaga (40) ja (41) ning vastuvõtmistõenäosusega (42) on Metropolis-Hastingsi algoritmid.

Kontrollime, kas detailse tasakaalu võrrandid on rahuldatud jaotuse $\pi(\cdot)$ jaoks.

1. Kui $x = y$, siis on tegemist samasusega.
2. Kui $x \neq y$.

Vaatame juhtu, kus $\pi(y)q(y, x) < \pi(x)q(x, y)$

$$\begin{aligned} \pi(x)p(x, y) &= \pi(x)q(x, y)\alpha(x, y) \\ &= \pi(x)q(x, y) \min \left[1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right] = \pi(y)q(y, x) \\ &= \pi(y)q(y, x) \underbrace{\min \left[1, \frac{\pi(x)q(x, y)}{\pi(y)q(y, x)} \right]}_{\alpha(y, x)} \\ &= \underbrace{\pi(y)q(y, x)}_{p(y, x)} \end{aligned}$$

Kuna jaotuse π korral Metropolis-Hastingsi algoritmiga saadud Markovi ahel rahuldab detailse tasakaalu võrrandeid, siis järelikult on π selle ahela statsionaarne jaotus ja teatud tingimustel q -le ka piirjaotus suvalisest algväärtusest lähtudes. Märkused:

- π on suvaline;
- q on nn ettepanekutuum ja võib olla üsna suvaline (paindlik algoritm);
- q ei sõltu π -st;
- q peab siiski olema mittelahutuv ja aperioidiline. Samuti peab $\alpha(x, y) > 0, \quad \forall x, y$;
- kui π on d -dimensionaalne, siis on ka q d -dimensionaalne tinglik jaotus;
- detailse tasakaalu võrrandid on rahuldatud $p(\cdot, \cdot)$, mitte $q(\cdot, \cdot)$ jaoks;

Metropolis-Hastingsi generaator:

1. Ahela algväärtustamine $t = 0, x^0$;
2. $t = t + 1$;
3. Genereerida seisund y tihedusest $y \sim q(x^{t-1}, \cdot)$;
4. Leida vastuvõtmistõenäosus $\alpha(x^{t-1}, y)$ avaldisest (42);
5. Otsustada vastuvõtmine, selleks genereerida $u \sim U(0, 1)$.
 Kui $u \leq \alpha(x^{t-1}, y)$, siis uus seisund võetakse vastu $x^t = y$,
 Kui $u > \alpha(x^{t-1}, y)$, siis uut seisundit vastu ei võeta $x^t = x^{t-1}$.

6. Tagasi sammu 2.

Erijuhul, kui ettepanekutuum on **sümmeetriline** $q(y, x) = q(x, y)$, siis vastuvõtmise tõenäosus on kujul

$$\alpha(x, y) = \min \left[1, \frac{\pi(y)}{\pi(x)} \right]$$

ja meil on tegemist Metropolisise algoritmiga.

Algoritmi omadusi

Kuidas valida q ? Kas ta peaks soovitama pikki või lühikesi liikumisi?

1. Kui on väikesed liikumissammud, st. x ja y on lähedased, siis on ka $\pi(x)$ ja $\pi(y)$ ning samuti $q(x, y)$, $q(y, x)$ lähedased. Järelikult $\alpha(x, y) \approx 1$.
Enamus liikumissettepanekuid võetakse vastu, aga sammude väiksuse tõttu läheb π kandja läbimiseks palju aega. Sissepõlemisperiood on suhteliselt pikk.
2. Kui on suured liikumissammud, st. x ja y kauged, siis satub y jaotuse $\pi(y)$ saba piirkonda, kus $\alpha(x, y) \ll 1$. Ettepanekuid ei võeta vastu ja koondumine on aeglane. Ahelas esinevad pikad sama väärtuse blokid.

Üleminekutuum $q(x, y)$ peab olema selline, et

- tagab ahela piisava liikumise;
- tagab hea vastuvõtmismäära;
- temast on lihtne genereerida.

Seni puuduvad üldised kriteeriumid q valikuks. Kui vastuvõtmise tõenäosus α on väike, siis võib ahel jääda samasse seisundisse pikaks ajaks.

Ahela headust näitab vastuvõtmismäär

$$\text{Vastuvõtmismäär} = \frac{\text{Vastuvõtmiste arv}}{\text{Iteratsioonide arv}}$$

Vastuvõtmismäär vahemikus 20% – 50% on sobiv ja selline q on küllalt hea.

Urime vastuvõtmistõenäosust $\alpha(x, y)$:

- Jaotus π esineb vastuvõtmistõenäosuses suhtena $\frac{\pi(y)}{\pi(x)}$;
- $\alpha(x, y)$ konstrueerimiseks pole vaja teada jaotuse normeerivat konstanti;
- Sobib eriti Bayesi järeljaotustest genereerimiseks, kus me sageli ei oska normeerivat konstanti analüütiliselt välja kirjutada $\pi(\theta) \propto \ell(\theta)p(\theta)$.

Metropolis-Hastingsi algoritmi puhul sobib sama koondumisdiagnostika, mis Gibbsi valiku puhul.

Näide 34

Normaaljaotuse $N(0, 1)$ genereerimine Metropolis-Hastingsi algoritmiga erineva dispersiooniga ettepanekutuumadega (vt joonist):

a) $q(x, y) = N(x, 0.5)$;

b) $q(x, y) = N(x, 0.1)$;

c) $q(x, y) = N(x, 10)$.

3.5.2 Ahelate erijuhud

Metropolis-Hastingsi ahelaid klassifitseeritakse ettepanekutuuma q järgi.

1. **Sümmeetriline** $q(x, y) \Rightarrow$ Metropolise algoritm.

Näiteks $q(x, y) = f(|x - y|)$ on sümmeetriline. Nüüd $\alpha(x, y) = \min \left[1, \frac{\pi(y)}{\pi(x)} \right]$ ei sõltu q -st. Võidame arvutuskiiruses.

2. **Juhuslik ekslemine** (ka mitmemõõtmeline)

$$x^t = x^{t-1} + \omega^t,$$

kus ω^t (sammu pikkus) on sõltumatud sama jaotusega juhuslikud suurused $\omega^t \sim f_\omega$. Juhusliku ekslemise tüüpi ettepanekutuuma on

$$q(x, y) = f(x - y)$$

Kui f on sümmeetriline 0-punkti suhtes, siis $q(x, y)$ on sümmeetriline. Paljud Metropolis-Hastingsi algoritmid ongi juhuslikud ekslemised. Tavaliselt $f_\omega = \begin{cases} N(0, \sigma^2) \\ t - \text{jaotus} \end{cases}$ σ^2 või t -jaotuse vabadusastmete arvu valik määrab liikumise ulatuse. Kui σ^2 on suur, siis on liikumine suur, aga $\alpha(x, y)$ väike ja vastupidi.

Soovitatakse f_ω dispersiooni kujul $c\hat{\sigma}^2$, kus c on häälestuskonstant ja $\hat{\sigma}^2$ on jaotuse π dispersiooni jäme hinnang (mitmemõõtmelisel juhul jaotuse π kovariatsioonimaatriksi hinnang). Kui ettepanekutuumaks on normaalne juhuslik ekslemine, on soovitatud $c = 2$ või 3 .

3. **Sõltumatu ahel**

Üleminekuettepanek on sõltumatu antud seisundist x

$$q(x, y) = f(y)$$

Üleminekutõenäosused:

$$\alpha(x, y) = \min \left[1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right] = \min \left[1, \frac{\pi(y)/f(y)}{\pi(x)/f(x)} \right] = \min \left[1, \frac{w(y)}{w(x)} \right]$$

Bayesi kontekstis on π parameetrite jaotus ja x, y , asemel on loomulikum kasutada θ, ϕ . Populaarseks valikuks funktsiooni f jaoks on parameetri eeljaotus $f(\theta) = p(\theta)$. Sel juhul

$$w(\theta) = \frac{\pi(\theta)}{p(\theta)} = \ell(\theta)$$

ja vastuvõtmistõenäosus rajaneb tõepärade suhtele:

$$\alpha(\theta, \phi) = \min \left[1, \frac{\ell(\phi)}{\ell(\theta)} \right]$$

Sellise valiku eeliseks on lihtne $\alpha(\theta, \phi)$ avaldis. Puudus, kui on konflikt tõepära ja eeljaotuse vahel, siis väärtused, mis soovitatakse eeljaotusest suure tõenäosusega, lükatakse tagasi väikese vastuvõtmistõenäosuse tõttu. Väärtusi, mis annaksid suure vastuvõtmistõenäosuse, valitakse harva ettepanekutuumaks oleva eeljaotuse poolt.

Üldine eeskiri sõltumatus ahela korral on vältida funktsiooni $w(\cdot)$ suurt varieeruvust, sest see suurendab võimalusi, et ahel satub paljudeks sammudeks suure $w(\cdot)$ väärtusega seisunditesse. Soovitav on valida $w(\cdot)$ võimalikult konstantne või vähemalt tõkestatud. Kuna f ja π on mõlemad tihedused, siis on soovitav, et f ja π oleksid võimalikult lähedased. Kui tõepära ja eeljaotus ei ole piisavalt lähedased, siis ei saa kasutada eeljaotust ettepanekutuumana.

Tierney (1994) soovitas mitte kasutada ettepanekutiheduseks kergete sabadega jaotusi nagu näiteks normaaljaotus, vaid kasutada näiteks t -jaotust väikese vabadusastmete arvuga. Sel juhul ei ole funktsioon $w(\cdot)$ nii tugevalt mõjutatud f sabadest ja tema varieeruvus on väiksem.

4. Teisi ettepanekutuumi

- f kui ümbrisjaotus π jaoks (valikumeetod Metropolis-Hastings'i kontekstis);
- autoregressioon-ettepanek $x^t = a + bx^{t-1} + w^t$. Erijuhud: $b = 1$, vabaliikmega juhuslik ekslemine;
- $b = -1$, vahelduv ahel, tekitab negatiivseid korrelatsioone ahela elementide vahel ja selleläbi ergoodiliste keskmiste väiksemaid dispersioone;
- palju teisi võimalusi, sõltuvalt probleemist – uurimisvaldkond;

Kodune ülesanne nr. 3.9 Panna kirja Metropolise algoritm regressiooniparameetrite järeljaotusest genereerimiseks: järeljaotus, ettepanekujaotus, vastuvõtmistõenäosus. Regressioonimudeliks on:

$$y_i \sim N(\mu, \sigma^2),$$

$$\mu_i = \beta_1 + \beta_2(x_i - \bar{x}).$$

Lahendame praktikumis konkreetsete andmetega.

3.6 Hübriidalgoritmid

3.6.1 Komponendiviisiline üleminek Metropolis-Hastingsi algoritmis

Metropolis-Hastingsi algoritmi korral genereeritakse üldjuhul uus punkt ettepanekutuuma $q(x, y)$ abil mitmemõõtmelisest jaotusest. Käesolevas peatükis vaatame üleminekut juhusliku vektori $x = (x_1, \dots, x_d)$ komponentide kaupa. Selleks on kaks võimalust:

- uuendatakse üks juhuslikult valitud komponent, üleminekutuuma on jaotuste segu;
- uuendatakse ettemääratud järjekorras $1 \rightarrow 2 \rightarrow \dots \rightarrow d$, üleminekutuuma on tsükkel.

Üleminekutuumade omadused:

1. Següleminekutuuma

$$p(x, y) = \sum_{i=1}^r w_i p_i, \quad \sum_{i=1}^r w_i = 1, \quad r \leq d,$$

p_i – i -nda komponendi üleminekutuuma;

w_i – i -nda komponendi tõenäosus e. jaotuste segu kaal.

x ja y erinevad maksimaalselt ühe komponendi võrra ühel iteratsioonil. Kui üks p_i -dest määrab mitteperioodilise ja mittelahutava ahela, siis teeb seda ka $p(x, y)$.

2. Tsükliline üleminekutuuma

p_i – komponendi uuendamise tuuma, $i = 1, \dots, d$. Üleminek $x \rightarrow y$ toimub d sammuga. Vahepealsed seisundid tähistame z -ga

$$\begin{aligned} x &= (x_1, x_2, \dots, x_{d-1}, x_d) \xrightarrow{p_1} z_1 = (y_1, x_2, \dots, x_{d-1}, x_d) \\ z_1 &= (y_1, x_2, \dots, x_{d-1}, x_d) \xrightarrow{p_2} z_2 = (y_1, y_2, \dots, x_{d-1}, x_d) \\ &\dots \\ z_{d-1} &= (y_1, y_2, \dots, y_{d-1}, x_d) \xrightarrow{p_d} y = (y_1, y_2, \dots, y_{d-1}, y_d) \end{aligned}$$

Üleminekutuuma ühistihedusfunktsioon $\prod_{i=1}^d p_i$.

Tähistame $x \equiv z_0$ ja $y \equiv z_d$, siis $p_i = p(z_{i-1}, z_i)$, $i = 1, \dots, d$. Meid huvitab y tihedusfunktsioon tingimusel x .

$$p(x, y) = \int \dots \int \prod_{i=1}^d p_i dz_1 \dots dz_{d-1}.$$

Metropolis-Hastingsi algoritmi jaoks tuleb iga üleminekutuuma esitada kujul:

$$p_i = q_i \alpha_i$$

q_i on vektori x i -nda komponendi ettepanekutuum (ühemõõtmeline). Vektori ühisjaotus avaldub kujul:

$$\pi(x) = \pi_i(x_i)\pi(x_{-i})$$

$\pi_i(x_i)$ – i -nda komponendi täistinglik jaotus;

$\pi(x_{-i})$ – ülejäänud komponentide marginaaljaotus.

Metropolis-Hastingsi algoritmis on suhe:

$$\frac{\pi(y)}{\pi(x)} = \frac{\pi_i(y_i)\pi(y_{-i})}{\pi_i(x_i)\pi(x_{-i})} = \frac{\pi_i(y_i)}{\pi_i(x_i)},$$

sest muutub ainult i -s komponent, ülejäänud komponendid jäävad samaks $\pi(y_{-i}) = \pi(x_{-i})$. Võime välja kirjutada vastuvõtmistõenäosuse avaldise:

$$\alpha_i(x_i, y_i) = \min \left[1, \frac{\pi_i(y_i)q_i(y_i, x_i)}{\pi_i(x_i)q_i(x_i, y_i)} \right]$$

Komponendiviisiline üleminekutuum defineerib pööratava ahela piirjaotusega $\pi_i(x_i)$, $i = 1, \dots, d$. Täistinglikud on meie π omad. Üldjuhul on piirjaotuseks ka meie ühisjaotus $\pi(\cdot)$.

Algoritm tsüklilise ülemineku jaoks:

1. Ahela ja iteratsioonide loendaja algväärtustamine $j = 1, x^0$;
2. Komponendi loendur $i = 1$;
3. Genereerida i -s komponent y_i tihedusest $q_i(x_i^{j-1}, y_i)$;
4. Arvutada vastuvõtmistõenäosus $\alpha_i(x_i^{j-1}, y_i)$, genereerida $u \sim U(0, 1)$
 kui $u \leq \alpha_i \rightarrow x_i^j = y_i$ uus väärtus võetakse vastu;
 kui $u > \alpha_i \rightarrow x_i^j = x_i^{j-1}$;
5. $i = i + 1$
 kui $i \leq d$, siis korrata alates 3. sammust;
6. $j = j + 1$ ja korrata alates 2. sammust.

See algoritm on Metropolise (1953) esialgne algoritm. Tema teostas seisundite komponendiviisilist üleminekut. Ta simuleeris molekulide asukohta. Ettepanekutuumaks oli kahedimensionaalne ühtlane jaotus keskpunktiga molekuli eelmises asukohas. Jaotuseks π oli d -dimensionaalne Gibbsi jaotus.

3.6.2 Metropolis + Gibbs

Komponendiviisilise uuendamise korral on vastuvõtmistõenäosus kujul

$$\alpha_i(x_i, y_i) = \min \left[1, \frac{\pi_i(y_i)q_i(y_i, x_i)}{\pi_i(x_i)q_i(x_i, y_i)} \right]$$

- Kui ettepanekutuum on lähedane täistinglikule tihedusele $q_i(x_i, y_i) \approx \pi_i(y_i)$, siis on vastuvõtutõenäosus $\alpha_i(\cdot) \approx 1$. See hoiab kokku arvutusteks kuluvat aega.
- Kui jaotuse π mõni täistinglik on lihtsalt simuleeritav, siis võtame selle vastava komponendi ettepanekutuumaks. Vastuvõtmine toimub siis tõenäosusega 1 ja pole üldse vaja arvutada $\alpha_i(\cdot)$, jälle arvutuslik kokkuhoid (Gibbsi Metropolise sees).
- Kui kõigi komponentide uuendamisel on ettepanekutuumaks on täpselt $q_i(x_i, y_i) = \pi_i(y_i)$, siis on tegemist p uhta Gibbsi valikuga.
- Samas võib Gibbsi valiku korral täistinglikust simuleerimiseks kasutada M-H algoritmi (1-dimensionaalset). See on Metropolis Gibbsi sees. Siin meie muutumispunktiga Poissoni vaatluste näide. Meil olid 3 parameetri λ, ϕ, m täistinglikud jaotused. Nendest kordamööda simuleerimine on Gibbsi valik Kui aga m simuleerimiseks kasutada M-H algoritmi, on tegu metropolisega Gibbsi sees.

Näide 35

Gibbs koos Metropolis-Hastingsiga. Blokiviisiline uuendamine.

Olgu x_1, \dots, x_n sõltumatud sama jaotusega juhuslikud suurused (ssjjs), jaotusega $f(x)$. Eesmärgiks on genereerida vektor $x = (x_1, \dots, x_n)$, kus komponentide summa on ette antud

$$\sum_{i=1}^n x_i = s,$$

st genereerida valim valimiruumi kitsendatud osast. Võimatu kasutada valikumeetodit, sest aeg väga pikk.

Rakendame Gibbsi valikut juhuslikele kahestele blokkidele. Vt näiteks blokki (x_4, x_8) . Uuendame (x_4, x_8) tingimusel, et ülejäänud x_i -d on fikseeritud ja $\sum_{i=1}^n x_i = s$, seega on fikseeritud ka $x_4 + x_8 = s_{48}$. Teeme koordinaatide ortogonaalse teisenduse:

$$\begin{cases} y_1 = \frac{(x_4 + x_8)}{\sqrt{2}} \\ y_2 = \frac{(x_4 - x_8)}{\sqrt{2}} \end{cases} \implies \begin{cases} x_4 = \frac{(y_1 + y_2)}{\sqrt{2}} \\ x_8 = \frac{(y_1 - y_2)}{\sqrt{2}} \end{cases}$$

Tingimuse $x_4 + x_8 = s_{48}$ tõttu on ainult üks komponent y_2 juhuslik. Leiame tema tingliku tiheduse: $f_{y_2|y_1}$

Ühistihedusfunktsioon

$$f_{y_1, y_2}(y_1, y_2) = f\left(\frac{y_1 + y_2}{\sqrt{2}}\right) f\left(\frac{y_1 - y_2}{\sqrt{2}}\right) \left| \frac{dx}{dy} \right|.$$

Tinglik tihedusfunktsioon

$$f_{y_2|y_1} = \frac{f_{y_1, y_2}(y_1, y_2)}{f_{y_1}(y_1)} = \frac{f_{y_1, y_2}(y_1, y_2)}{\int f_{y_1, y_2}(y_1, y_2) dy_2}.$$

Jaotuse normeerivat konstanti

$$\frac{1}{\int f_{y_1, y_2}(y_1, y_2) dy_2}$$

on raske leida. Otseselt ei saa Gibbsi valikut rakendada, sest täistinglikke jaotusi on raske leida. Kasutame Metropolis-Hastingsi algoritmi:

1. Algväärtustame vektori $x = (x_1, \dots, x_n)$ lubatavate väärtustega nii, et $\sum_{i=1}^n x_i = s$;
2. Valime ettepanekutuuma q , mis on sõltumatu käesolevast väärtusest, näiteks võiks selleks olla üldkogumijaotus $f(x)$;
3. Valime juhuslikult bloki;
4. Olgu käesoleval hetkel blokis teise komponendi väärtus y_2 ;
5. Genereerime ettepanekuga uue väärtuse y'_2 ;
6. y_2 võetakse vastu tõenäosusega

$$\alpha(y_2, y'_2) = \min \left[1, \frac{f_{y_2|y_1}(y'_2)q(y_2)}{f_{y_2|y_1}(y_2)q(y'_2)} \right] = \min \left[1, \frac{f_{y_1, y_2}(y_1, y'_2)q(y_2)}{f_{y_1, y_2}(y_1, y_2)q(y'_2)} \right];$$

7. kui võtame vastu y'_2 , arvutame selle bloki uued x_4, x_8 ; Kui vastu ei võta jääb blokk muutmata;
8. Tagasi bloki valikusse;

Näide 36

Tugevad kitsendused. Olgu x_1, \dots, x_n ssjj suurused tihedusega $f(x)$. Tahame simuleerida x_1, \dots, x_n tinglikust jaotusest tingimusel

$$x_1 - \bar{x} = y_1, \quad x_2 - \bar{x} = y_2, \quad \dots, \quad x_n - \bar{x} = y_n,$$

kus y_i -d on fikseeritud. Näeme, et

$$(x_1, \dots, x_n) = (y_1, \dots, y_n) + t(1, \dots, 1),$$

kus $t = \bar{x}$ on juhuslik. Metropolis-Hastings on siin loomulik:

1. Algväärtus t -le, näiteks $t_0 = 0$.
2. Leiame vektori $(x_{1,t_0}, \dots, x_{n,t_0}) = (y_1, \dots, y_n) + t_0(1, \dots, 1)$.
3. Teeme uue väärtuse ettepaneku t jaoks kasutades mingit üleminekutuuma $q(t_0, t)$. Olgu ettepanekud väärtus t_1 . Vastuvõtmise korral muutuvad x -id: $x_{i,t_0} \rightarrow x_{i,t_1}, \forall i$.
4. Ettepanek võetakse vastu tõenäosusega

$$\alpha(x_{t_0}, x_{t_1}) = \min \left[1, \frac{\prod_{i=1}^n f(x_{i,t_0}) q(t_1, t_0)}{\prod_{i=1}^n f(x_{i,t_1}) q(t_0, t_1)} \right].$$

Antud ülesandel on rakendus **Pitman**'i hinnangu arvutamisel. Olgu X_1, \dots, X_n ssjj vaatlused mudelist $X = \theta + W$, kus $W \sim f(x)$. Tähistame

$$Y_1 = X_1 - \bar{X}, \quad Y_2 = X_2 - \bar{X}, \quad \dots, \quad Y_n = X_n - \bar{X}.$$

Suuruse θ parimaks hinnanguks ruutkeskmise vea suhtes on Pitmani hinnang:

$$\hat{\theta}_P = \bar{x} + E_{\theta=0}(\bar{X} | Y_i = y_i, \forall i),$$

kus y_i, \bar{x} on valimist leitud. Teisiti väljendades on Pitmani hinnang tõepärafunktsiooni, kui tihedusfunktsiooni keskvväärtus. Pannes tähele, et

$$X_i \sim f(x_i - \theta),$$

saame

$$\hat{\theta}_P = \frac{\int \theta f(x_1 - \theta) \cdots f(x_n - \theta) d\theta}{\int f(x_1 - \theta) \cdots f(x_n - \theta) d\theta}.$$

Hinnangut on raske välja arvutada, aga simuleerides palju kordi \bar{X} kitsenduste $Y_i = y_i, \forall i$ korral saame seda teha.

3.7 Veel andmeanalüüsi näiteid

Näide 37

Hierarhiline regressioon. Rasedate kaalu juurdekasv.

Uuriti 68 raseda kaalu juurdekasvu 5-7 arstivisiidi ajal. y_{ij} – i -nda naise kaal j -ndal visiidil arsti juurde ($i = 1, \dots, 68$; $j = 1, \dots, n_i$). t_{ij} – i -nda naise j -nda visiidi ajahetk rasedusnädalates. Kokku oli mõõtmistulemusi $\sum_{i=1}^{68} n_i = 427$

Joonisel 5.4 (Gamerman and Lopes, 2006) on kõigi naiste kaalude juurdekasvu andmed (juurdekasv normaalkaalu suhtes).

Mudel: lihtne hierarhiline regressioon aja järgi.

$$y_{ij} | \alpha_i, \beta_i, \sigma^2 \sim N(\alpha_i + \beta_i t_{ij}, \sigma^2), \quad i = 1, \dots, 68.$$

Parameetrid α_i ja β_i sõltuvad naisest. Kõigile tuleb määrata eeljaotused. Seda tehakse hierarhiliselt:

$$\begin{aligned} \alpha_i | \alpha &\sim N(\alpha, \tau_\alpha^{-1}) \\ \beta_i | \beta &\sim N(\beta, \tau_\beta^{-1}) \\ (\alpha, \beta)' &\sim N((0, 0)', 10^3 I_2) \\ \sigma^{-2}, \tau_\alpha, \tau_\beta &\sim G(0.001, 0.001) \end{aligned}$$

Hierarhiline esitus ja kaasjaotuste kasutamine võimaldab tuletada täistinglikud jaotused ühisjäreljaotusest

$$\pi(\theta) \propto \ell(\theta) p(\theta), \quad \text{kus } \theta = (\alpha_1, \alpha_2, \dots, \alpha_{68}, \beta_1, \beta_2, \dots, \beta_{68}, \alpha, \beta, \sigma^2, \tau_\alpha, \tau_\beta).$$

Need on kõik normaal- ja gamma jaotused, millest on lihtne simuleerida. Saab kasutada Gibbsi valikut ja genereerida järeljaotusest π .

Ülesanne. Pane kirja tõepärafunktsioon ja parameetrite eeljaotus $p(\theta)$

Vt. jooniseid 5.5-5.8 (Gamerman and Lopes, 2006).

Näide 38

Allikad. I. Ntzoufras (2009),

www.stat-athens.aueb.gr/~jbn/winbugs_book

Ühefaktoriline dispersioonanalüüs. Erakooli disrektor tahab tööle võtta uue matemaatikaõpetaja. Nelja kandidaadi võimeid hinnatakse järgmises väikeses katses. grupp õpilasi (25 tükki) jagati juhuslikult nelja rühma. Igas rühmas õpetati sama matemaatilist teemat 2 tundi päevas ühe nädala jooksul. Lõpus pidid õpilased tegema testi, mis andis neile punkte (0-100). Direktor võtab kooli õpetaja, kelle rühm sooritas testi kõige paremini.

```

model{
  # model's likelihood
  for (i in 1:n){
    mu[i] <- m + alpha[ class[i] ]
    grade[i] ~ dnorm( mu[i], tau )
  }
  ##### stz constraints
  alpha[1] <- -sum(alpha[2:TUTORS])
  ##### CR Constraints
  # alpha[1] <- 0.0

  # priors
  m~dnorm( 0.0, 1.0E-04)
  for (i in 2:TUTORS){ alpha[i]~dnorm(0.0, 1.0E-04)}
  tau ~dgamma( 0.01, 0.01)
  s <- sqrt(1/tau) # precision
}

INITS list( m=1.0, alpha=c(NA, 0,0,0), tau=1.0 )

DATA (LIST) list( n=25, TUTORS=4,
  grade=c(84, 58, 100, 51, 28, 89, 97, 50, 76, 83, 45, 42, 83,
    64, 47, 83, 81, 83, 34, 61, 77, 69, 94, 80, 55, 79),
  class=c(1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2,
    3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4) )

```

Vt WinBUGS lahendust lisamaterjalist.

Näide 39

Poissoni regressioon. Lennuki vigastused 30 rünnakus Vietnami sõjas:
damage – vigastuste arv lennuki keres saadud ühes rünnakus;
type – lennuki tüüp (0-A4, 1-A6);
bombload – pommilaadung (tonnides);
airexp – meeskonna summaarne kogemus (kuudes).

Obs	y	x1	x2	x3
1	0	0	4	91.5
2	1	0	4	84
.....				
28	5	1	14	88.9
29	5	1	14	73.7
30	7	1	14	57.8

Millest sõltub vigastuste arv?

```

model{
  # Poisson model likelihood
  for (i in 1:30){

```

```

    damage[i] ~ dpois( lambda[i] )
    log(lambda[i]) <- beta[1] + beta[2] * type[i] + beta[3] * bombload[i] +
    beta[4] * airexp[i]
  }

  #
  # prior
  for (j in 1:4){
    beta[j]~dnorm( 0.0, 0.001 )
    B[j] <- exp( beta[j] )
  }
  #
  # profiles
  # values for bombload
  profiles[1,1] <- ranked( bombload[], 1 ) # minimum of bombload
  profiles[2,1] <- mean(bombload[]) # mean of bombload
  profiles[3,1] <- 0.5*( ranked( bombload[], 15 )+ranked( bombload[], 16 )) #median
  profiles[4,1] <- ranked( bombload[], 30 ) #max
  # values for airexp
  profiles[1,2] <- ranked( airexp[], 30 ) #max experience
  profiles[2,2] <- mean(airexp[]) #mean
  profiles[3,2] <- 0.5*( ranked( airexp[], 15 )+ranked( airexp[], 16 )) #median
  profiles[4,2] <- ranked( airexp[], 1 ) #min experience

  for (k in 1:4){
    a4.profile[k] <- exp( beta[1] + beta[3]*profiles[k,1] + beta[4]*profiles[k,2] )
    a6.profile[k] <- a4.profile[k]*B[2]
    # this is equivalent to setting exp( beta[1] + beta[3]*profile[k,1] +
    # beta[4]*profile[k,2] )
  }
}

INITS list( beta=c(0,0,0,0) )

DATA (LIST) list( damage = c(0, 1, 0, 0, 0, 0, 1, 0, 0, 2, 1, 1, 1,
1, 2, 3, 1, 1, 1, 2, 0, 1, 1, 2, 5, 1, 1, 5, 5, 7), type = c(0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1), bombload = c(4, 4, 4, 5, 5, 5, 6, 6, 6, 7, 7, 7, 8,
8, 8, 7, 7, 7, 10, 10, 10, 12, 12, 12, 8, 8, 8, 14, 14, 14), airexp
= c(91.5, 84, 76.5, 69, 61.5, 80, 72.5, 65, 57.5, 50, 103, 95.5, 88,
80.5, 73, 116.1, 100.6, 85, 69.4, 53.9, 112.3, 96.7, 81.1, 65.6, 50,
120, 104.4, 88.9, 73.7, 57.8) )

```

Näide 40

Dünaamiline-hierarhiline segamudel.

Uuriti 20 lapse hamba kõrgust millimeetrites neljal ajahetkel (8-, 8.5-, 9-, ja 9.5-aastatel), y_{ti}

– lapse lõikehamba kõrgus ajahetkel t

$$\begin{aligned}y_{ti} &\sim N(\theta_{ti}, \sigma^2) \\ \theta_{ti} &\sim (1-p)N(\theta_{t-1,i} + \lambda_{ti}, W_1) + pN(\mu_t, V_1) \\ \lambda_{ti} &\sim (1-p)N(\lambda_{t-1,i}, W_2) + pN(\gamma_t, V_2) \\ \mu_t &\sim N(\mu_{t-1} + \gamma_t, W_1) \\ \gamma_t &\sim N(\gamma_{t-1}, W_2)\end{aligned}$$

Lõpuks tuleb anda eeljaotused esimese ajahetke parameetritele:

$$\begin{aligned}\mu_1, \gamma_1 &\sim N(0, 10^3) \\ \sigma^2, V_1, V_2, W_1, W_2 &\sim IG(0.1, 0.1)\end{aligned}$$

Näide 41

Siin on $M-H$ parem, kui Gibbs. Mittelineaarne regressioonimudel kestusandmetele:

$$y_{ij} = f(\theta_i, t_{ij}) + \varepsilon_{ij}$$

y_{ij} – indiviidi i mõõdetud väärtus hetkel t_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n_i$.
 θ_i – mittelineaarse regressiooni parameeter

$$\varepsilon_{ij} \sim N(0, \sigma^2).$$

Indiviidide regressiooniparameetrid antakse hierarhiliselt

$$\theta_i \sim N(\mu, W).$$

Hüperparameetritele määratakse sõltumatud eeljaotused:

$$\mu \sim N(b, B), \quad W \sim IG\left(\frac{n_w}{2}, \frac{n_w S_w}{2}\right) \quad \text{ja} \quad \sigma^2 \sim IG\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right)$$

Näiteks farmakoloogilistes uuringutes on tähtis koht järgmisel mittelineaarsel seosefunktsioonil:

$$f(\theta_1, \theta_2, \theta_3; t) = \theta_1 + \frac{\theta_2 t}{\theta_3 + t}. \quad (43)$$

Funktsioon kirjeldab aine kontsentratsiooni taset hetkel t . Pole kaasjaotuste peret parameetrite θ_i jaoks. Samuti on parameetrite täistinglikud jaotused rasked simuleerimiseks. Siin tuleb appi Metropolis-Hastingsi algoritm. Soovitus: alati kui täistinglikud on liiga keerulised kasuta teisi MCMC skeeme, näiteks $M-H$.

Ülesanne. Olgu σ^2 teada. Eeldades normaalseid eeljaotusi parameetritele, pane kirja mudeli (43) järeljaotus ja täistinglik jaotus θ_1 jaoks.

Näide 42

Reageerimisaja hierarhiline segamudel.

Kirjeldatud näide illustreerib, kuidas mudeli kontrolli ja laiendamist kasutada koos, et saavutada andmetega paremini sobiv mudel.

Näide käsitleb psühholoogilist katset, kus mõõdeti 17-1 inimesel 30 korda reageerimisaja. Katsealustest 11 olid terved ja 6 skisofreenikud. Andmetele sobitati mitmeid mudeleid kasutades suurima tõepära hinnangut. Andmed on esitatud joonistel 11.1 ja 11.2. Skisofreenikute reageerimisaja on keskmiselt pikem. Mõnede skisofreenikute reageerimisaja varieerub tunduvalt suuremas ulatuses, kui tervetel inimestel. Praegune psühholoogia teooria väidab, et skisofreenikud kannatavad tähelepanuhäirete ja üldise motoorse pidurduse all. Mõlemad pikendavad reageerimisaja: tähelepanuhäired mõnedes katsetes ja pidurdus kõikides katsetes.

Andmetele sobitati järgmine lihtne mudel. Tervete katsealuste logaritmilised reageerimisajad on normaaljaotusega keskmisega α_i , $i = 1, \dots, 11$ ja konstantse dispersiooniga σ_t^2 . Skisofreenikute reageerimisajad on kahekomponendilise segujaotusega: tõenäosusega $(1-\lambda)$ puudub tähelepanuhäire ja ajad on normaaljaotusega keskmisega α_i , $i = 12, \dots, 17$ ja konstantse dispersiooniga σ_y^2 ; tõenäosusega λ on reageerimisaja pikem tähelepanuhäire tõttu ja on normaaljaotusega keskvaertusega $\alpha_i + \tau$ ja sama dispersiooniga.

Reageerimisajad on kõik positiivsed. Ebasümmeetrilise jaotuse tõttu kasutati modelleerimisel mõõdetud aja logaritmi. Mudelisse lisati veel hierarhiline parameetri β , mis mõõdab motoorset pidurdust. Hierarhilise mudeli saab esitada järgmiselt:

$$\begin{aligned} y_{ij} | \alpha_i, z_{ij}, \phi &\sim N(\alpha_i + \tau z_{ij}, \sigma_t^2) \\ \alpha_i | z, \phi &\sim N(v + \beta S_i, \sigma_\alpha^2) \\ z_{ij} | \phi &\sim Be(\lambda S_i) \end{aligned}$$

y_{ij} – i -ndal isikul mõõdetud j -s reageerimisaja;

v – tervete katsealuste reageerimisaja üldkeskmine;

S_i – indikaator, mis on 1, kui katsealune on skisofreenik ja 0 muidu;

z_{ij} – mittevaadeldud indikaator, mis on 1, kui i -nda katsealuse j -s mõõtmise pikenes mingi mõjutuse tõttu ja 0 muidu (pole modelleerimisel vaja, aga lihtsustavad arvutamist);

$\phi = (\sigma_\alpha^2, \beta, \lambda, \tau, v, \sigma_t^2)$

Parameetrite σ_α^2 , β , λ , τ , v ja σ_t^2 eeljaotuseks on ühtlane jaotus, λ on määratud lõigul $[0.001, 0.999]$, σ_α^2 , τ ja σ_t^2 peavad olema positiivsed.

Mudeli parameetrite leidmiseks käivitati 10 paralleelset ahelat Gibbsi valikus, mis koondusid pärast 200 iteratsiooni. Valimist jäeti välja 100 esimest iteratsiooni igast ahelast. Järgi jäi 1000 mudeli parameetrite vektori valimit.

Selle mudeli analüüsi on tehtud raamatus Gilks, Richardson, Spiegelhalter (1996).

Mudeli kontroll

Valitud mudel arvestas keskmiste erinevust kahe katsealuste grupi vahel, aga tekkis küsimus mudeli sobivusest üksikutele katsealustele. Kahe esimese skisofreeniku mõõtmistulemused varieeruvad tunduvalt vähem kui ülejäänud selles grupis. Mõõtmistulemuste võrdlemiseks mudeliga arvutati si, skisofreenikute mõõtmistulemuste standardhälve ($i = 12, \dots, 17$). Defineeriti kolm test-statistikut: s_i kuue väärtuse väiksem $T_{\min}(y)$, suurim $T_{\max}(y)$ ja keskmine $T_{\text{avg}}(y)$. Joonise 11.2 järgi võib arvata, et T_{\min} on väiksem ja T_{\max} suurem, kui mudeli prognoositud väärtused. Keskmine avg T on lisatud kontrolliks, selle väärtus peaks tulema lähedane mudeli poolt prognoositavale, kuna see on põhiliselt prognoositud mudeli parameetrite λ , τ ja σ_t^2 abil.

Vaadeldud andmete põhjal (joonis 11.2) on teststatistiku väärtused:

$$T_{\min}(y) = 0.11; \quad T_{\max}(y) = 0.58; \quad T_{\text{avg}}(y) = 0.30$$

Mudeli kontrollimiseks simuleeriti prognoosiv andmestik iga simuleeritud parameetrite komplekti kohta (1000 andmestikku) mudeli järelejaotusest. Iga simuleeritud andmestiku jaoks arvutati $T_{\min}(y^*)$, $T_{\max}(y^*)$ ja $T_{\text{avg}}(y^*)$. Joonisel 11.3 on näha vastavad histogrammid.

Mudeli laiendamine

Mudeli parandamiseks lisatakse mudelisse kaks uut parameetrit, mis lubavad mõnedel skisofreenikutel reageerimisaegu ilma tähelepanuhäireteta ja võimaldavad suuremat varieeruvust reageerimisaegadele, mis on mõjutatud tähelepanuhäiretest. Uued parameetrid on:
 ω – tõenäosus, et skisofreenik kannatab tähelepanuhäirete all;
 σ_{t2}^2 – tähelepanuhäiretest mõjutatud mõõtmiste dispersioon.

Mõlema parameetri eeljaotus on ühtlane jaotus. Arvutuslikel eesmärkidel lisatakse mudelisse indikaator W_i , mis on 1, kui isik võib kannatada tähelepanuhäirete all ja 0 muidu. Indikaator W_i on automaatselt 0 tervete katsealuste korral ja skisofreenikute korral omandab indikaator W_i väärtuse 1 tõenäosusega ω .

Esialgne mudel on uue mudeli erijuht tingimusel, et $\omega = 1$ ja $\sigma_{t2}^2 = \sigma_t^2$. Uue mudeli sobitamine on küllaltki lihtne. Gibbsi valikusse tuleb lisada kolm uut sammu parameetrite ω , σ_{t2}^2 ja W_i uuendamiseks.

Kasutati kümme sõltumatut ahelat, genereeriti 500 väärtust ja neist kasutati 250 väärtust. Enne mudeli kontrolli prognoosiva simuleerimise abil on huvitav võrrelda esialgse ja laiendatud mudeli psühholoogide jaoks olulisemate parameetrite hinnanguid. Psühholooge huvitavad järgmised parameetrid:

λ – tõenäosus, et reageerimisaeg pikeneb tähelepanuhäirete tõttu

ω – tõenäosus, et skisofreenik kannatab tähelepanuhäirete all;

τ – tähelepanuhäirest sõltuv reageerimisaja pikenemine (logaritmilises skaalas);

β – skisofreeniku tähelepanuhäirest mõjutamata keskmine logaritmiline reageerimisaeg minus tervete katsealuste keskmine logaritmiline reageerimisaeg.

Tabel 1: Esialgse ja uue mudeli võrdlus

Parameeter	Esialgne mudel			Laiendatud mudel		
	2.5%	mediaan	97.5%	2.5%	mediaan	97.5%
λ	0.07	0.12	0.18	0.46	0.64	0.88
ω		1		0.24	0.56	0.84
τ	0.74	0.85	0.96	0.21	0.42	0.60
β	0.17	0.32	0.48	0.07	0.24	0.43

Tabel näitab olulist erinevust parameetrite hinnangutes.

Laiendatud mudeli kontroll

Kui hästi sobib uus mudel andmetega? Eeldatakse suuremat sobivust võrreldes esialgse mudeliga, sest uued parameetrid lisati just eesmärgiga parandada sobivust. Kontrollimiseks kasutatakse samu teststatistikuid nagu esialgse mudeli puhul. Tulemusi näitab joonis 11.4.

Kirjandus

Albert, J. (2009). *Bayesian Computation with R*, 2nd ed., Springer.

Congdon, P. (2003). *Applied Bayesian Modelling*. Wiley.

Gamerman, D., Lopes, H.F. (2006). *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference* 2nd ed. Chapman&Hall.

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

Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman&Hall.

Häggström, O. (2002). *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press.

Hoff, P.D. (2009). *A First Course in Bayesian Statistical Methods*. Springer.

Lunn, D., Jackson, Ch, Best, N., Thomas, A., Spiegelhalter, D. (2013). *The BUGS Book. A Practical Introduction to Bayesian Analysis*. CRC Press.

Ntzoufras, Ioannis (2009). *Bayesian Modeling Using WinBUGS*. Wiley.

Press, S. J. (2003). *Subjective and Objective Bayesian Statistics. Principles, Models and Applications*, 2nd ed., Wiley.

Pärna, K. (2013). it Tõenäosusteooria algkursus. Tartu Ülikooli kirjastus.

Rubinstein, R.Y., Kroese, D.P. (2008). *Simulation and the Monte Carlo Method*. Wiley.

Traat, I. (2006). *Matemaatilise statistika põhikursus*. Tartu Ülikooli kirjastus.

Interneti materjalid

Tarkvaraga WinBUGS seonduv

<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/remote14.shtml>

Näited raamatu "Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference"juurde

<http://www.dme.ufrj.br/mcmc/>

Näited raamatu "Bayesian Modeling Using WinBUGS"juurde

www.stat-athens.aueb.gr/~jbn/winbugs_book