

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Mikael Raihhelgauz
Bayesi segumudelid tiheduse hindamiseks

Matemaatilise statistika eriala

Magistritöö (30 EAP)

Juhendaja: prof. Jüri Lember

TARTU 2022

BAYESI SEGUMUDELID TIHEDUSE HINDAMISEKS

Magistritöö

Mikael Raihhelgauz

Lühikokkuvõte

Magistritöö eesmärk on tutvustada levinumaid Bayesi segumudeleid ja vastavaid Gibbsi valikul põhinevaid meetodeid tiheduse ligikaudseks hindamiseks. Samuti illustreeritakse meetodite tööd arvutisimulatsioonide abil.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

Märksõnad: Bayesi statistika, Bayesi segumudelid, Gibbsi valik, tiheduse hindamine.

BAYESIAN MIXTURE MODELS FOR DENSITY ESTIMATION

Master's thesis

Author

Abstract

The purpose of this master's thesis is to give an overview of the most common Bayesian mixture models as well as the corresponding Gibbs sampling-based methods for approximate density estimation. The performance of these methods is exemplified by computer simulations.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics.

Key Words: Bayesian statistics, Bayesian mixture models, Gibbs sampling, density estimation.

Sisukord

Sissejuhatus	3
1 Parameetriline Bayesi mudel	4
1.1 Viitade ja parameetrite väärtuste genereerimine	10
1.2 Tiheduse hindamine	17
2 Dirichlet' protsess: mõningad eelteadmised	21
3 Mitteparameetriline Bayesi mudel	35
4 Bayesi segumudelid Gaussi tuumaga	53
4.1 Eeljaotuse valik	54
4.2 Hinnangu mõjususe	57
4.3 Simulatsioonid	62
Kasutatud allikad	77
Lisa 1. Lõpliku normaaljaotuste segu tiheduse hinnangud	79
Lisa 2. Üldise segu tiheduse hinnangud	80
Lisa 3. Ühtlaste jaotuste segu tiheduse hinnangud	81

Sissejuhatus

Üks tüüpilisi statistilisi ülesandeid on andmeid x_1, \dots, x_n genereerinud tiheduse ψ_0 hindamine (peetakse silmas tihedust Lebesgue'i mõõdu suhtes). Juhul, kui andmete tegeliku jaotuse kohta on vähe eelteadmisi (nt ei soovita eeldada kindlat jaotuste peret), võib vaadelda hinnatavat tihedust ψ_0 segujaotuse tihedusena

$$\psi_0(x) = \int \psi(x; \theta) p_0(d\theta),$$

kus ψ on tuum ja p mingi tõenäosusmõõt. Bayesi raamistikus vaadeldakse mõõtu p_0 juhusliku suurusena P ja postuleeritakse oma esialgseid aimdusi selle kohta eeljaotusena. Niisugust konfiguratsiooni võib nimetada “Bayesi segumudeliks”.

1973. aastal kirjeldas Ferguson [5] Dirichlet' protsessi jaotust. Selle seadmine juhusliku mõõdu P eeljaotuseks on oluliselt lihtsustanud tööd Bayesi segumudelitega. Hilisem arvutitehnika areng ja MCMC simulatsioonimeetodite väljatöötamine on teinud Bayesi segumudelid hõlpsa vahendi tiheduse ψ_0 ligikaudseks hindamiseks [10]. Siinse magistritöö eesmärk on tutvustada levinumaid Bayesi segumodeleid, anda ülevaade vastavatest Gibbsi valikul põhinevatest tiheduse hindamise meetoditest ja illustreerida nende tööd arvutisimulatsioonide abil.

Töö esimeses peatükis käsitleme segujaotuse tiheduse hindamist juhul, kui p_0 on diskreetne tõenäosusmõõt. Teises peatükis tutvustame Dirichlet' protsessi defineerimiseks tarvilikku mõõduteoreetilist aparatuuri ja kirjeldame selle olulisemaid omadusi. Kolmandas peatükis üldistame esimeses peatükis esitatud tiheduse hindamise meetodeid olukorrale, kus enam ei eeldata mõõdu p_0 diskreetsust. Neljandas peatükis vaatleme lähemalt Bayesi segumodeleid Gaussi tuumaga ja viime läbi arvutisimulatsioonid.

Suurem osa töös esitatud tulemustest põhineb Subhashis Ghosali ja Aad van der Vaarti [7] raamatul “*Fundamentals of Nonparametric Bayesian Inference*”. Simulatsioonide läbiviimisel kasutati rakendustarkvara R.

1 Parameetriline Bayesi mudel

Enne siinse peatüki probleemistiku tutvustamist esitame tuuma ja segujaotuse tiheduse mõiste, mida kasutame läbivalt kogu magistritöö ulatuses.

Definitsioon 1.1. Tuum. Olgu (Θ, Σ) mõõtv ruum. Funktsiooni

$$\psi(x; \theta) : \mathbb{R} \times \Theta \rightarrow [0, \infty)$$

nimetatakse tuumaks, kui

1. ψ on $\mathcal{B}(\mathbb{R}) \otimes \Sigma$ -mõõtv;
2. leidub Boreli σ -algebral $\mathcal{B}(\mathbb{R})$ defineeritud mõõt λ nii, et iga $\theta \in \Theta$ korral kujutus $\psi_\theta(x) := \psi(x; \theta)$ on mingi tõenäosusmõõdu ν_θ tihedus λ suhtes.

Märkus 1.1.1. Esmapilgul võib tunduda, et $\mathcal{B}(\mathbb{R}) \otimes \Sigma$ -mõõtuvuse nõue on liialt abstraktne ja praktikas raskesti kontrollitav. Tegelikuses on tingimus rahuldatud siis, kui iga $\theta \in \Theta$ korral $\psi_\theta(x) := \psi(x; \theta)$ on pidev ja iga $x \in \mathbb{R}$ korral $\psi_x(\theta) := \psi(x; \theta)$ on mõõtv. Veendume selles.

Kui $n \in \mathbb{N}$, siis iga $k \in \mathbb{Z}$ korral

$$\frac{\lfloor nx \rfloor}{n} = \frac{k}{n} \Leftrightarrow x \in \left[\frac{k}{n}, \frac{k+1}{n} \right)$$

ja seega

$$\begin{aligned} \psi\left(\frac{\lfloor nx \rfloor}{n}; \theta\right) &= I_{\left[0, \frac{1}{n}\right)}(x) \cdot \psi(0; \theta) + \\ &+ \sum_{k=1}^{\infty} I_{\left[\frac{k}{n}, \frac{k+1}{n}\right)}(x) \cdot \psi\left(\frac{k}{n}; \theta\right) + \sum_{k=1}^{\infty} I_{\left[-\frac{k}{n}, -\frac{k+1}{n}\right)}(x) \cdot \psi\left(-\frac{k}{n}; \theta\right). \end{aligned}$$

Tähistame

$$\phi_{k,n}(x; \theta) = I_{\left[\frac{k}{n}, \frac{k+1}{n}\right)}(x) \cdot \psi\left(\frac{k}{n}; \theta\right) = \begin{cases} \psi_{\frac{k}{n}}(\theta), & \text{kui } x \in \left[\frac{k}{n}, \frac{k+1}{n}\right), \\ 0, & \text{mujal.} \end{cases}$$

Olgu $B \in \mathcal{B}(\mathbb{R})$ suvaline. Kui $0 \notin B$, siis

$$\phi_{k,n}^{-1}(B) = \left[\frac{k}{n}, \frac{k+1}{n} \right) \times \psi_{\frac{k}{n}}^{-1}(B) \in \mathcal{B}(\mathbb{R}) \otimes \Sigma,$$

kuna $\left[\frac{k}{n}, \frac{k+1}{n} \right) \in \mathcal{B}(\mathbb{R})$ ja $\psi_{\frac{k}{n}}$ on eelduse kohaselt Σ -mõõtuv. Juhul, kui $0 \in B$,

$$\phi_{k,n}^{-1}(B) = \left[\frac{k}{n}, \frac{k+1}{n} \right) \times \psi_{\frac{k}{n}}^{-1}(B) \cup \left[\frac{k}{n}, \frac{k+1}{n} \right)^c \times \Theta \in \mathcal{B}(\mathbb{R}) \otimes \Sigma.$$

Kuna iga $k \in \mathbb{Z}$ korral funktsioon $\phi_{k,n}(x; \theta)$ on mõõtuv, siis ka nende funktsioonide summana esituv

$$\psi_n(x; \theta) := \psi\left(\frac{\lfloor nx \rfloor}{n}; \theta\right)$$

on $\mathcal{B}(\mathbb{R}) \otimes \Sigma$ -mõõtuv.

Fikseerime $x \in \mathbb{R}$ ja $\theta \in \Theta$. Paneme tähele, et

$$\frac{\lfloor nx \rfloor}{n} \rightarrow x.$$

Pidevuse Heine kriteeriumi tõttu

$$\psi(x; \theta) = \lim_n \psi\left(\frac{\lfloor nx \rfloor}{n}; \theta\right) = \lim_n \psi_n(x; \theta).$$

Kuna $\psi(x; \theta)$ esitub mõõtuvate funktsioonide jada piirväärtusena, on see ka ise $\mathcal{B}(\mathbb{R}) \otimes \Sigma$ -mõõtuv.

Definitsioon 1.2. Segujaotuse tihedus. Olgu p tõenäosusmõõt ruumil (Θ, Σ) ja funktsioon $\psi(x; \theta) : \mathbb{R} \times \Theta \rightarrow [0, \infty)$ olgu tuum. Seosega

$$\psi_p : \mathbb{R} \rightarrow [0, \infty), \quad \psi_p(x) := \int \psi(x; \theta) p(d\theta)$$

määratud kujutust ψ_p nimetatakse segujaotuse tiheduseks.

Märkus 1.2.1. Veendume, et kujutuse ψ_p nimetamine tiheduseks on õigustatud, s.t tegemist on mittenegatiivse $\mathcal{B}(\mathbb{R})$ -mõõtuva funktsiooniga. Olgu $\theta \in \Theta$ ja $x \in$

\mathbb{R} suvalised. Kuna funktsioon ψ_θ on tihedus, kehtib võrratus $\psi(x, \theta) \geq 0$. Nüüd Lebesgue'i integraali monotoonsuse tõttu

$$\psi_p(x) = \int \psi(x; \theta) p(d\theta) \geq 0.$$

Et ψ on mittenegatiivne ja $\mathcal{B}(\mathbb{R}) \otimes \Sigma$ -mõõtuv, siis Tonelli teoreemi põhjal on kujutus $\psi_p(x)$ $\mathcal{B}(\mathbb{R})$ -mõõtuv. Tonelli teoreemist järeldeb ka kujutuse

$$\psi_\lambda : \Theta \rightarrow \mathbb{R}, \quad \psi_\lambda(\theta) = \int \psi(x; \theta) \lambda(dx)$$

mõõtuvus, kusjuures suvalise $\theta \in \Theta$ korral

$$\psi_\lambda(\theta) = \int \psi(x; \theta) \lambda(dx) = \nu_\theta(\mathbb{R}) = 1.$$

Järelikult

$$\begin{aligned} \int_{\mathbb{R}} \psi_p(x) \lambda(dx) &= \int_{\mathbb{R}} \left(\int_{\Theta} \psi(x; \theta) p(d\theta) \right) \lambda(dx) = \\ &= \int_{\Theta} \left(\int_{\mathbb{R}} \psi(x; \theta) \lambda(dx) \right) p(d\theta) = \\ &= \int_{\Theta} \psi_\lambda(\theta) p(d\theta) = \int_{\Theta} p(d\theta) = p(\Theta) = 1. \end{aligned}$$

Olles esitanud tuuma ja segujaotuse tiheduse mõisted, oleme valmis tutvustama siinses peatükis käsitletavat probleemi. Olgu meil andmed x_1, \dots, x_n , mis pärinevad tõenäosusjaotusest Ψ_0 , kusjuures eeldame, et jaotusel Ψ_0 leidub tihedus Lebesgue'i mõõdu suhtes ψ_0 . Huvipakkuv ülesanne on tiheduse ψ_0 hindamine.

Eeldame, et tihedus ψ_0 esitub segujaotuse tihedusena, s.t leidub tuum ψ ja (mõõtuval) parameeterruumil (Θ, Σ) määratud tõenäosusmõõt p_0 , et iga $x \in \mathbb{R}$ korral

$$\psi_0(x) = \int \psi(x; \theta) p_0(d\theta).$$

Tuum ψ käsitleme läbivalt statistikule teadaoleva funktsioonina. Praktikas on sa-

geli võimalik valida tuum ψ nii, et tegelikule tihedusele ψ_0 seatavad lisaeldused poleks kuigi kitsendavad. Tuuma valiku kriteeriumitel peatume veidi pikemalt 4. peatükis.

Antud peatükis seame veel ühe kitsendava tingimuse katõenäosusmõõdule p_0 . Nii-
melt eeldame, et tegemist on diskreetse tõenäosusmõõduga

$$p_0 = \sum_{j=1}^k p_j^0 \delta_{\theta_j^0}, \quad (p_1^0, \dots, p_k^0) \in \mathbb{S}^k,$$

kus \mathbb{S}^k tähistab simpleksit $\{(p_1, \dots, p_k) : \min_i p_i \geq 0, \sum_{i=1}^k p_i = 1\}$, s.t tegelik tihedus esitub kujul

$$\psi_0(x) = \sum_{j=1}^k p_j^0 \psi(x; \theta_j^0),$$

kus arv $k \in \mathbb{N}$ on taas statistikule teada, kuid tõelised parameetrid

$$p_1^0, \dots, p_k^0, \theta_1^0, \dots, \theta_k^0$$

on tundmatud.

Hindamaks segujaotuse tihedust Bayesi meetodil käsitleme parameetreid juhuslike suurustena ja võtame kasutusele allpool kirjeldatud mudeli, milles postuleerime nende eeljaotused. Kõigepealt aga toome Dirichlet' jaotuse definitsiooni ja tutvustame selle kolme omadust, mis on edasise arutelu seisukohalt olulised. Lausete 2.7, 1.5 ja 1.6 võib leida nt autori bakalaureusetööst [14, ptk 2].

Definitsioon 1.3. Dirichlet' jaotus. Öeldakse, et juhuslik vektor

$$\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_k) : \Omega \rightarrow \mathbb{R}^k$$

on Dirichlet' jaotusega parameetritega $k \in \mathbb{N} \setminus \{1\}$ ja $\alpha_1, \dots, \alpha_k \in (0, \infty)$ (lühemalt

$\mathbf{p} \sim \text{Dir}(\alpha_1, \dots, \alpha_k)$, kui \mathbf{p} tihedusfunktsioon avaldub kujul

$$f(p_1, \dots, p_k) = \begin{cases} \frac{\Gamma(\alpha_1 + \dots + \alpha_k)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)} \prod_{i=1}^k p_i^{\alpha_i - 1}, & (p_1, \dots, p_k) \in \mathbb{S}^k, \\ 0, & \text{mujal.} \end{cases}$$

Lause 1.4. Olgu $\mathbf{p} \sim \text{Dir}(\alpha_1, \dots, \alpha_k)$ ja juhusliku vektori \mathbf{p} fikseeritud väärtuse $p = (p_1, \dots, p_k)$ korral olgu $X_1, \dots, X_n \stackrel{i.i.d}{\sim} p$. Siis juhusliku vektori \mathbf{p} järeljaotus on $\text{Dir}(\alpha_1 + n(1), \dots, \alpha_k + n(k))$ ehk

$$\mathbf{p} | X_1 = x_1, \dots, X_n = x_n \sim \text{Dir}(\alpha_1 + n(1), \dots, \alpha_k + n(k)),$$

kus $n(j)$, $j = 1, \dots, k$, tähistab, mitu korda väärtus j esineb vaatluste x_1, \dots, x_n seas.

Lause 1.5. Olgu \mathbf{p} juhuslik vektor jaotusega $\text{Dir}(\alpha_1, \dots, \alpha_k)$ ja tähistagu $|\alpha|$ summat $\sum_{i=1}^k \alpha_i$. Siis \mathbf{p} keskväärtus on

$$E(\mathbf{p}) = \left(\frac{\alpha_1}{|\alpha|}, \dots, \frac{\alpha_k}{|\alpha|} \right).$$

Lause 1.6. Olgu \mathbf{p} juhuslik vektor jaotusega $\text{Dir}(\alpha_1, \dots, \alpha_k)$. Siis iga $j \in \{1, \dots, k\}$ korral juhusliku suuruse \mathbf{p}_j dispersioon on

$$D(\mathbf{p}_j) = \frac{\alpha_j}{|\alpha|^2} \left(\frac{|\alpha| - \alpha_j}{|\alpha| + 1} \right).$$

Definitsioon 1.7. Parameetriline segumudel. Olgu $\mathcal{Y} = \{1, 2, \dots, k\}$ ja c loendav mõõt kogumil $2^{\mathcal{Y}}$. Tähega Θ tähistame eukleidilist parameetrite ruumi \mathbb{R}^m ja tähega $\mathcal{B}(\Theta)$ tähistame sellele vastavat Boreli σ -algebrat. Lisaks sellele vaatleme mõõtuvaid ruume $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ ja $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ning tõenäosusruumi $(\Omega, \mathcal{F}, \mathbf{P})$. Kuivõrd lähte- ja sihthulgale vastavad σ -algebrad on nüüd paika pandud, saame

käsitleda juhuslikke suurusi

$$\begin{aligned}
 (\mathbf{p}_1, \dots, \mathbf{p}_k) &= \mathbf{p} : \Omega \rightarrow \mathbb{R}^k, \\
 \boldsymbol{\theta}_i &: \Omega \rightarrow \Theta, \quad i = 1, \dots, k, \\
 Y_i &: \Omega \rightarrow \mathcal{Y}, \quad i = 1, \dots, n, \\
 X_i &: \Omega \rightarrow \mathbb{R}, \quad i = 1, \dots, n.
 \end{aligned}$$

Olgu \mathbf{p} juhuslik vektor jaotusega $Dir(\alpha_1, \dots, \alpha_k)$ ja iga $i = 1, \dots, k$ korral olgu π juhusliku suuruse $\boldsymbol{\theta}_i$ tihedus Lebesgue'i mõõdu suhtes. Olgu $\mathbf{p}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k$ omavahel sõltumatud.

Juhusliku vektori \mathbf{p} fikseeritud väärtuse $p = (p_1, \dots, p_k)$ korral olgu viidad Y_1, \dots, Y_n sõltumatud juhuslikud suurused jaotusega $\sum_{i=1}^k p_i \delta_i$. Kui Y_1, \dots, Y_n ja $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k)$ väärtused on omakorda fikseeritud – tähistame neid vastavalt kui y_1, \dots, y_n ja $\theta = (\theta_1, \dots, \theta_k)$ –, siis X_1, \dots, X_n on omavahel sõltumatud juhuslikud suurused, kusjuures igaüks neist on tihedusele $\psi(\cdot, \theta_{y_i})$ vastava jaotusega.

Eelkirjeldatud mudelit võime tähistada järgnevalt:

$$\begin{aligned}
 (\mathbf{p}_1, \dots, \mathbf{p}_k) &= \mathbf{p} \sim Dir(\alpha_1, \dots, \alpha_k), \\
 \boldsymbol{\theta}_i &\stackrel{\text{ind}}{\sim} \pi, \quad i = 1, \dots, k \\
 Y_1, \dots, Y_n | \mathbf{p} &= p \stackrel{\text{i.i.d}}{\sim} \sum_{i=1}^k p_i \delta_i, \\
 X_i | Y_i = y_i, \boldsymbol{\theta} &= \theta \stackrel{\text{i.i.d}}{\sim} \psi(\cdot; \theta_{y_i}), \quad i = 1, \dots, n.
 \end{aligned}$$

Märkus 1.7.1. Siinse peatüki ulatuses kasutame tähte ψ tähistamiseks tuuma, aga erinevaid ühistihedusi ja tinglikke tihedusi. Kirjapildi lihtsustamiseks kasutame eri jaotustele vastavate tiheduste eristamiseks üksnes väikseid tähti p, θ, x_j ja y_j funktsiooni $\psi(\cdot)$ argumentide kohal. Näiteks

$$\mathbf{p}, \boldsymbol{\theta}, Y_1, \dots, Y_n, X_1, \dots, X_n$$

ühistihedust korrutismõõdu $\text{Leb}^k \times \text{Leb}^{m \cdot k} \times c^n \times \text{Leb}^n$ suhtes tähistame kui

$$\psi(p, \theta, y_{1:n}, x_{1:n}) = \psi(p) \cdot \pi(\theta) \cdot \psi(y_{1:n}|p) \cdot \psi(x_{1:n}|y_{1:n}, \theta),$$

kus $\pi(\theta) = \prod_{i=1}^k \pi(\theta_i)$, $y_{1:n} = (y_1, \dots, y_n)$ ja $x_{1:n} = (x_1, \dots, x_n)$.

Vahel kasutame juhuslike suuruste väärtuste tähistamiseks ka teisi sümboleid. Näiteks kirjutame tavapärasest notatsiooni väänares

$$\mathbf{P}(Y_t = j | Y_1 = y_1, \dots, Y_{t-1} = y_{t-1}, Y_{t+1} = y_{t+1}, \dots, Y_n = y_n) = \psi(y_t = j | y_{-t}).$$

Siinjuures märgime, et väikese ψ kasutamine on jätkuvalt õigustatud, kuna formaalselt on eeltoodud tinglik tõenäosus tihedus loendava mõõdu c suhtes.

1.1 Viitade ja parameetrite väärtuste genereerimine

Suvalise $x \in \mathbb{R}$ korral tahame hinnata $\psi_0(x)$ ehk tegeliku tiheduse väärtust selles punktis. Standardne võte Bayesi lähenemise puhul on valida teatud kaofunktsioon $L(\cdot; \cdot)$ ja leida tiheduse hinnang $\hat{\psi}_0(x)$ nii, et keskvärtus (üle θ ja \mathbf{p})

$$E\left(L\left(\sum_{i=1}^k \mathbf{p}_i \psi(x, \theta_i), \hat{\psi}_0(x)\right)\right)$$

oleks minimaalne.

Levinuim kaofunktsiooni valik on seosega

$$L : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, \quad L(z; a) = |z - a|^2$$

määratud ruutkaofunktsioon. Kui mingil juhuslikul suurusel Z on lõplik keskvärtus ja teine moment, siis keskmist ruutkadu $E(L(Z; a)) = E|Z - a|^2$ minimiseerib $\hat{a} = EZ$. Niisiis, etteantud vaatluste $X_1 = x_1, \dots, X_n = x_n$ korral on parim tiheduse hinnang $E\left(\sum_{i=1}^k \mathbf{p}_i p(x, \theta_i) | x_1, \dots, x_n\right)$, s.t keskvärtus juhusliku vektori

$(\boldsymbol{\theta}, \mathbf{p})$ järeljaotuse järgi.

Paraku osutub eelneva keskväärtuse täpne rehkendamine ka võrdlemisi väikeste valimimahtude korral tülikaks, kuna $\boldsymbol{\theta}, \mathbf{p} | X_{1:n} = x_{1:n}$ järeljaotuse avaldamiseks tuleb leida $\boldsymbol{\theta}, \mathbf{p}, Y_{1:n} | X_{1:n} = x_{1:n}$ jaotus ja seejärel $Y_{1:n}$ välja integreerida, s.t summeerida üle kõikvõimalike realisatsioonide $y_{1:n} \in \{1, \dots, k\}^n$.

Alternatiivina võime hinnata huvipakkuvat keskväärtust ligikaudselt. Selleks aga vajame meetodit, mis võimaldaks genereerida Y_1, \dots, Y_n väärtusi etteantud vaatluste $x_{1:n}$ korral.

Praktikas on vahetult $Y_{1:n} | X_{1:n} = x_{1:n}$ jaotusest valimite genereerimine enamasti liiga arvutusmahuks, kuna vastav tihedus sisaldab taas väga suure liidetavate arvuga summasid. Viimast probleemi on aga võimalik vältida, rakendades Gibbsi valikut. Järgnevalt tuletamegi algoritmi genereerimaks viitade Y_1, \dots, Y_n väärtusi Gibbsi valiku meetodil. Iga $t = 1, \dots, n$ korral tähistagu Y_{-t} juhuslikku vektorit $(Y_1, \dots, Y_{t-1}, Y_{t+1}, \dots, Y_n)$ ning y_{-t} tähistagu selle vektori realisatsiooni. Sümbolite X_{-t} ja x_{-t} tähendus on sama.

Lause 1.8 on tuletatud raamatus "*Fundamentals of Nonparametric Bayesian Inference*" esitatud jaotuse [7, lk 106] analoogina parameetrilise mudeli korral.

Lause 1.8. *Etteantud Y_{-t} ja $X_{1:n}$ väärtuste korral avaldub Y_t tihedus kujul*

$$\psi(y_t = j | y_{-t}, x_{1:n}) = \frac{b_j}{C} \left(\frac{\int \psi(x_t, \theta_j) \prod_{\substack{i: y_i = j \\ i \neq t}} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j}{\int \prod_{\substack{i: y_i = j \\ i \neq t}} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j} \right), \quad (1.1)$$

kus $b_j = \alpha_j + n_{-t}(j)$, $n_{-t}(j)$ tähistab, mitu korda j esineb vektoris y_{-t} ja

$$C = \sum_{r=1}^k \frac{(\alpha_r + n_{-t}(r)) \int \psi(x_t, \theta_r) \prod_{\substack{i: y_i = r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r) d\theta_r}{\int \prod_{\substack{i: y_i = r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r) d\theta_r}.$$

Tõestus. Kõigepealt leiame tõenäosuse, et $Y_t = j$ vektori y_{-t} fikseerimisel. Paneme tähele, et

$$\psi(y_t = j | y_{-t}) = \int_{\mathbb{S}^k} \psi(y_t = j | p, y_{-t}) \psi(p | y_{-t}) dp.$$

Kuivõrd \mathbf{p} fikseeritud väärtuse korral on Y_1, \dots, Y_n sõltumatud, saame kirjutada

$$\begin{aligned} \int_{\mathbb{S}^k} \psi(y_t = j | p, y_{-t}) \psi(p | y_{-t}) dp &= \int_{\mathbb{S}^k} \psi(y_t = j | p) \psi(p | y_{-t}) dp = \\ &= \int_{\mathbb{S}^k} p_j \psi(p | y_{-t}) dp = E(\mathbf{p}_j | y_{-t}). \end{aligned}$$

Lause 2.7 põhjal

$$\mathbf{p} | Y_{-t} = y_{-t} \sim \text{Dir}(\alpha_1 + n_{-t}(1), \dots, \alpha_k + n_{-t}(k))$$

ja seega lause 1.5 tõttu

$$\psi(y_t = j | y_{-t}) = E(\mathbf{p}_j | y_{-t}) = \frac{\alpha_j + n_{-t}(j)}{|\alpha| + n - 1}.$$

Kuna Y_t ja $\boldsymbol{\theta}$ väärtuste fikseerimisel ei sõltu X_t jaotus enam vektoritest Y_{-t} ja X_{-t} , võime kirjutada

$$\psi(x_t | x_{-t}, y_{-t}, y_t = j, \theta_1, \dots, \theta_k) = \psi(x_t, \theta_j).$$

Samamoodi ei sõltu Y_t jaotus vektorist $\boldsymbol{\theta}$ ja Y_{-t} väärtuse fikseerimise järel ka vektorist X_{-t} . Formaaalselt tähendab see, et

$$\psi(y_t | x_{-t}, y_{-t}, \boldsymbol{\theta}) = \psi(y_t = j | y_{-t}).$$

Eelnevaid fakte kokku võttes saame

$$\begin{aligned} \psi(y_t = j, x_t | x_{-t}, y_{-t}, \theta_1, \dots, \theta_k) &= \psi(y_t = j | y_{-t}) \psi(x_t, \theta_j) = \\ &= \frac{\alpha_j + n_{-t}(j)}{|\alpha| + n - 1} \psi(x_t, \theta_j). \end{aligned} \quad (1.2)$$

Järgnevalt paneme tähele, et

$$\begin{aligned}\psi(\theta_1, \dots, \theta_k | x_{-t}, y_{-t}) &= \frac{\psi(\theta_1, \dots, \theta_k, x_{-t} | y_{-t})}{\psi(x_{-t} | y_{-t})} = \\ &= \frac{\psi(x_{-t} | \theta_1, \dots, \theta_k, y_{-t}) \psi(\theta_1, \dots, \theta_k | y_{-t})}{\psi(x_{-t} | y_{-t})}.\end{aligned}$$

Kuna vektor θ on vektorist Y_{-t} sõltumatu ja ka $\theta_1, \dots, \theta_k$ on omavahel sõltumatud, siis võime kirjutada

$$\psi(\theta_1, \dots, \theta_k | y_{-t}) = \psi(\theta_1, \dots, \theta_k) = \prod_{r=1}^k \pi(\theta_r).$$

Kui Y_{-t} ja θ väärtused on fikseeritud, siis vektori X_{-t} komponendid on omavahel sõltumatud ja nende ühistihedus avaldub lihtsa korrutisena:

$$\psi(x_{-t} | \theta_1, \dots, \theta_k, y_{-t}) = \prod_{\substack{i=1 \\ i \neq t}}^n \psi(x_i, \theta_{y_i}).$$

Siit saame võrduse

$$\psi(\theta_1, \dots, \theta_k, x_{-t} | y_{-t}) = \left(\prod_{\substack{i=1 \\ i \neq t}}^n \psi(x_i, \theta_{y_i}) \right) \left(\prod_{r=1}^k \pi(\theta_r) \right) = \prod_{r=1}^k \prod_{\substack{i: y_i=r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r).$$

Integreerides $\theta_1, \dots, \theta_k$ välja, leiame tiheduse

$$\begin{aligned}\psi(x_{-t} | y_{-t}) &= \int \dots \int \left(\prod_{r=1}^k \prod_{\substack{i: y_i=r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r) \right) d\theta_1 \dots d\theta_k = \\ &= \prod_{r=1}^k \int \prod_{\substack{i: y_i=r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r) d\theta_r.\end{aligned}$$

Järelikult

$$\psi(\theta_1, \dots, \theta_k | x_{-t}, y_{-t}) = \frac{\prod_{r=1}^k \prod_{\substack{i:y_i=r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r)}{\prod_{r=1}^k \int \prod_{\substack{i:y_i=r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r) d\theta_r}. \quad (1.3)$$

Tegemaks kirjepilti veidi kompaksemaks tähistame vektorit $(\theta_1, \dots, \theta_k)$ sümboliga $\theta_{1:k}$. Toetudes võrdusele (1.3) saame, et

$$\begin{aligned} \psi(y_t = j, x_t | y_{-t}, x_{-t}) &= \int \dots \int \psi(y_t = j, x_t | y_{-t}, x_{-t}, \theta_{1:k}) \psi(\theta_{1:k} | x_{-t}, y_{-t}) d\theta_1 \dots d\theta_k = \\ &= \int \dots \int \frac{\alpha_j + n_{-t}(j)}{|\alpha| + n - 1} \psi(x_t, \theta_j) \times \\ &\quad \times \frac{\prod_{r=1}^k \prod_{\substack{i:y_i=r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r)}{\prod_{r=1}^k \int \prod_{\substack{i:y_i=r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r) d\theta_r} d\theta_1 \dots d\theta_k = \\ &= \frac{\alpha_j + n_{-t}(j)}{|\alpha| + n - 1} \left(\prod_{\substack{r=1 \\ r \neq j}}^k \frac{\int \prod_{\substack{i:y_i=r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r) d\theta_r}{\int \prod_{\substack{i:y_i=r \\ i \neq t}} \psi(x_i, \theta_r) \pi(\theta_r) d\theta_r} \right) \times \\ &\quad \times \left(\frac{\int \psi(x_t, \theta_j) \prod_{\substack{i:y_i=j \\ i \neq t}} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j}{\int \prod_{\substack{i:y_i=j \\ i \neq t}} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j} \right) = \\ &= \frac{\alpha_j + n_{-t}(j)}{|\alpha| + n - 1} \left(\frac{\int p(x_t, \theta_j) \prod_{\substack{i:y_i=j \\ i \neq t}} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j}{\int \prod_{\substack{i:y_i=j \\ i \neq t}} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j} \right) \end{aligned}$$

Lõpuks jõuamegi soovitud tiheduseni:

$$\begin{aligned} \psi(y_t = j | y_{-t}, x_{1:n}) &= \frac{\psi(y_t = j, x_t | y_{-t}, x_{-t})}{\sum_{r=1}^k \psi(y_t = r, x_t | y_{-t}, x_{-t})} = \\ &= \frac{\alpha_j + n_{-t}(j)}{C} \left(\frac{\int \psi(x_t, \theta_j) \prod_{\substack{i:y_i=j \\ i \neq t}} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j}{\int \prod_{\substack{i:y_i=j \\ i \neq t}} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j} \right). \end{aligned}$$

□

Toetudes lausele 1.8 saame formuleerida järgneva algoritmi 1 $Y_{1:n}$ väärtuste gene-

reerimiseks Gibbsi valiku meetodil tingimusel, et on antud vaatlused $X_{1:n} = x_{1:n}$. Siinne algoritm on parameetiline analoog raamatus "*Fundamentals of Nonparametric Bayesian Inference*" esitatud algoritmile 3 [7, lk 107].

Algoritm 1 Juhuslike suuruste Y_1, \dots, Y_n väärtuste genereerimine Gibbsi valiku meetodil fikseeritud $X_{1:n} = x_{1:n}$ korral

Sisend: andmed $x_{1:n}$, algväärtused $y_{1:n}^{(0)} = (y_1^{(0)}, \dots, y_n^{(0)})$, soovitud valimite arv m

- 1: **iga** $j = 1, \dots, m$ korral
- 2: **iga** $t = 1, \dots, n$ korral
- 3: fikseeritud $y_{-t} = (y_1^{(j)}, \dots, y_{t-1}^{(j)}, y_{t+1}^{(j-1)}, \dots, y_n^{(j-1)})$ korral genereeri $y_t^{(j)}$ vastavalt tihedusele (1.1)
- 4: salvesta $y_{1:n}^{(j)} = (y_1^{(j)}, \dots, y_n^{(j)})$

Väljund: vektorid $y_{1:n}^{(j)}$, $j = 1, \dots, m$

Siinjuures tuleb tähele panna, et algoritmis 1 vektori θ väärtusi ei simuleerita. Y_t väärtuste genereerimisel integreeritakse $\theta_1, \dots, \theta_k$ lihtsalt välja. Praktikas ei pruugi olla algoritmis 1 esinevad integraalid analüütiliselt leitavad. Viimasel juhul tuleb kasutada alternatiivset algoritmi (edaspidi algoritm 2), kus parameetrite θ ja viitade Y_1, \dots, Y_n väärtusi genereeritakse eraldi. Selleks peame tuletama viisi θ_j , $j = 1, \dots, k$, väärtuste genereerimiseks juhul, kui on antud andmed $X_{1:n} = x_{1:n}$ ja vektori $Y_{1:n}$ väärtused $y_{1:n}$ on fikseeritud.

Lause 1.9. *Juhusliku parameetri $\theta_j | X_{1:n} = x_{1:n}, Y_{1:n} = y_{1:n}$ tinglik tihedus on*

$$\psi(\theta_j | x_{1:n}, y_{1:n}) = \frac{\prod_{i:y_i=j} \psi(x_i, \theta_j) \pi(\theta_j)}{\int \prod_{i:y_i=j} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j}. \quad (1.4)$$

Tõestus. Paneme tähele, et θ_j jaotus sõltub juhuslikust suurusel X_i vaid siis, kui $y_i = j$. Fikseeritud $y_{1:n}$ korral vaatleme hulki $J = \{i : y_i = j\}$ ja vektoreid $x_J = (x_i)_{i \in J}$. Siis

$$\psi(\theta_j | x_{1:n}, y_{1:n}) = \frac{\psi(\theta_j, x_J | y_{1:n})}{\int \psi(\theta_j, x_J | y_{1:n}) d\theta_j} = \frac{\prod_{i:y_i=j} \psi(x_i, \theta_j) \pi(\theta_j)}{\int \prod_{i:y_i=j} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j}.$$

□

Kuna algoritmis 2 uuendatakse kordamööda vektoreid θ ja $y_{1:n}$, peame tuletama ka juhusliku viida $Y_t|Y_{-t} = y_{-t}, X_{1:n} = x_{1:n}, \theta = \theta$ järeljaotusele vastava tiheduse.

Lause 1.10. *Juhusliku viida $Y_t|Y_{-t} = y_{-t}, X_{1:n} = x_{1:n}, \theta = \theta$ järeljaotusele vastav tihedus on*

$$\psi(y_t = j|y_{-t}, x_{1:n}, \theta_1, \dots, \theta_k) = \frac{(\alpha_j + n_{-t}(j))\psi(x_t, \theta_j)}{\sum_{r=1}^k (\alpha_r + n_{-t}(r))\psi(x_t, \theta_r)}. \quad (1.5)$$

Tõestus. Paneme tähele, et

$$\psi(y_t = j|y_{-t}, x_{1:n}, \theta_1, \dots, \theta_k) = \frac{\psi(x_t, y_t = j|x_{-t}, y_{-t}, \theta_1, \dots, \theta_k)}{\sum_{r=1}^k \psi(x_t, y_t = r|x_{-t}, y_{-t}, \theta_1, \dots, \theta_k)}.$$

Asendades nüüd lugejas ja nimetajas olevad elemendid lause 1.8 tõestuses leitud tihedusega (1.2) saamegi soovitud funktsiooni. \square

Toetudes lausetele 1.9 ja 1.10, võime sõnastada soovitud algoritmi 2. Märgime, et tegemist on raamatus "*Fundamentals of Nonparametric Bayesian Inference*" esitatud algoritmi 2 [7, lk 107] parameetrilise analoogiga.

Algoritm 2 Y_1, \dots, Y_n ja θ väärtuste iteratiivne genereerimine Gibbsi valiku meetodil fikseeritud $X_{1:n} = x_{1:n}$ korral

Sisend: andmed $x_{1:n}$, algväärtused $y_{1:n}^{(0)} = (y_1^{(0)}, \dots, y_n^{(0)})$ ja $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$, soovitud valimite arv m

- 1: **iga** $j = 1, \dots, m$ korral
- 2: **iga** $t = 1, \dots, n$ korral
- 3: fikseeritud $y_{-t} = (y_1^{(j)}, \dots, y_{t-1}^{(j)}, y_{t+1}^{(j-1)}, \dots, y_n^{(j-1)})$, $\theta^{(j-1)}$ ja vaatluste $x_{1:n}$ korral genereeri $y_t^{(j)}$ vastavalt tihedusele (1.5)
- 4: salvesta $y_{1:n}^{(j)} = (y_1^{(j)}, \dots, y_n^{(j)})$
- 5: **iga** $r = 1, \dots, k$ korral
- 6: genereeri $\theta_r^{(j)}$ vastavalt tihedusele (1.4), kui on antud $x_{1:n}$ ja $y_{1:n}^{(j)}$
- 7: salvesta $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_k^{(j)})$

Väljund: vektorid $y_{1:n}^{(j)}$ ja $\theta^{(j)}$, $j = 1, \dots, m$

1.2 Tiheduse hindamine

Tuletame meelde, et siinse peatüki eesmärk on esitada eeskiri segujaotuse tiheduse

$$\psi_0(x) = \sum_{j=1}^k p_j^0 \psi(x; \theta_j^0)$$

hindamiseks Bayesi meetodil. Nagu varemgi öeldud, etteantud andmete $X_{1:n} = x_{1:n}$ korral esitub parim tiheduse hinnang kujul

$$\hat{\psi}_0(x) = E\left(\sum_{j=1}^k \mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | x_{1:n}\right) = \sum_{j=1}^k E(\mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | x_{1:n}).$$

Paraku osutub tinglike keskväärtuste $E(\mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | x_{1:n})$, $j = 1, \dots, k$, rehkendamine keeruliseks ka võrdlemisi lihtsa struktuuiga mudelite ja väikeste valimimahtude korral. Nende arvutamiseks peame teadma tihedust $\psi(p, \theta | x_{1:n})$. Viimasel aga puudub lihtne analüütiline kuju, s.t tuleb kasutada tihedust $\psi(p, \theta, y_{1:n} | x_{1:n})$ ja sealt kõikvõimalikud vektori $Y_{1:n}$ realisatsioonid välja summeerida. Selle operatsiooni arvutusmahukust illustreerib järgmine näide.

Näide 1.1. Olgu meil andmed $x_1, \dots, x_{100} \in \mathbb{R}$. Eeldame, et need pärinevad kahe võrdse dispersiooniga normaaljaotuse segust, kusjuures meil on head põhjused arvata, et eri komponentidesse kuulumise tõenäosused on enam-vähem võrdsed. Oma eelteadmised postuleerime järgmise mudeli abil:

$$\begin{aligned} \mathbf{p}_1 &\sim \text{Beta}(5; 5), \quad \mathbf{p}_2 = 1 - \mathbf{p}_1, \\ \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 &\stackrel{i.i.d.}{\sim} \mathcal{N}(\mu_0, \sigma_0), \\ Y_1, \dots, Y_{100} | \mathbf{p} &= (p_1, p_2) \stackrel{i.i.d.}{\sim} p_1 \delta_1 + p_2 \delta_2, \\ X_i | Y_i = y_i, (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) &= (\mu_1, \mu_2) \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu_{y_i}, \sigma). \end{aligned}$$

Antud juhul tuum $\psi(\cdot; \mu_{y_i})$ vastab normaaljaotuse $\mathcal{N}(\mu_{y_i}, \sigma)$ tihedusele ja π on $\mathcal{N}(\mu_0, \sigma_0)$ tihedus. Tingliku keskväärtuste $E(\sum_{j=1}^2 \mathbf{p}_j \psi(x; \boldsymbol{\mu}_j) | x_{1:100})$ leidmiseks

vajaminev tihedus avaldub kujul

$$\begin{aligned}\psi(p, \mu | x_{1:100}) &= \frac{\psi(p, \mu, x_{1:100})}{\psi(x_{1:100})} = \\ &= \frac{\sum_{y_{1:100} \in \{1,2\}^{100}} \psi(p, \mu, x_{1:100}, y_{1:100})}{\psi(x_{1:100})}\end{aligned}$$

Seega otsitav keskvärtus avaldub summana

$$\sum_{y_{1:100} \in \{1,2\}^{100}} \int_{\mathbb{S}^k} \int_{\mathbb{R}} \left(\sum_{j=1}^2 p_j \psi(x; \mu_j) \right) \frac{\psi(p, \mu, x_{1:100}, y_{1:100})}{\psi(x_{1:100})} d\mu dp.$$

Kokkuvõttes näeme, et tingliku keskvärtuse arvutamisel peame leidma summamärgi all oleva integraali kõigi 2^{100} võimaliku $Y_{1:100}$ realisatsiooni korral. Võrdluseks, tänapäeva võimsaimad arvutid suudavad teostada umbes $10^{22} < 2^{77}$ arvutustunnis.

Eelkirjeldatud probleemi vältimiseks võime hinnata tinglikku keskvärtust ligikaudselt. Selleks paneme tähele, et

$$E\left(\sum_{j=1}^k \mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | x_{1:n}\right) = E\left[E\left(\sum_{j=1}^k \mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | Y_{1:n}, x_{1:n}\right) \middle| x_{1:n}\right], \quad (1.6)$$

kusjuures võrduse (1.6) paremal pool olev keskvärtus esitub summana

$$\sum_{y_{1:n} \in \{1, \dots, k\}^n} E\left(\sum_{j=1}^k \mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | y_{1:n}, x_{1:n}\right) \cdot \psi(y_{1:n} | x_{1:n}). \quad (1.7)$$

Osutub, et fikseeritud $y_{1:n}$ korral on $E\left(\sum_{j=1}^k \mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | y_{1:n}, x_{1:n}\right)$ rehendamise sageli teostatav. Niisiis, summa (1.7) ligikaudselt hindamiseks võime genereerida valimeid $y_{1:n}^{(m)}$, $m = 1, \dots, q$, jaotusest tihedusega $\psi(y_{1:n} | x_{1:n})$, arvutada iga $m = 1, \dots, q$ korral $E\left(\sum_{j=1}^k \mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | y_{1:n}^{(m)}, x_{1:n}\right)$ ja seejärel võtta ψ_0 hinnanguks

empiiirilise keskmise

$$\hat{\psi}_0 = \frac{1}{q} \sum_{m=1}^q E \left(\sum_{j=1}^k \mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | y_{1:n}^{(m)}, x_{1:n} \right).$$

Siinjuures paneme tähele, et vektori $Y_{1:n}$ väärtuse fikseerimisel ei sõltu \mathbf{p} jaotus juhuslikust vektorist $X_{1:n}$ ega juhuslikest suurustest $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k$. Formaalselt

$$\begin{aligned} \psi(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k, p | x_{1:n}, y_{1:n}) &= \psi(p | \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k, x_{1:n}, y_{1:n}) \psi(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k | x_{1:n}, y_{1:n}) = \\ &= \psi(p | y_{1:n}) \psi(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k | x_{1:n}, y_{1:n}). \end{aligned}$$

Siit järeldub omakorda, et iga $j = 1, \dots, k$, korral

$$\begin{aligned} E(\mathbf{p}_j \psi(x, \boldsymbol{\theta}_j) | x_{1:n}, y_{1:n}) &= E(\mathbf{p}_j | y_{1:n}) E(\psi(x, \boldsymbol{\theta}_j) | x_{1:n}, y_{1:n}) = \\ &= \frac{\alpha_j + n(j)}{|\alpha| + n} E(\psi(x, \boldsymbol{\theta}_j) | x_{1:n}, y_{1:n}). \end{aligned}$$

Kui on võimalik rakendada algoritmi 1, toimime järgnevalt.

- (i) Genereerime valimid $y_{1:n}^{(m)}$, $m = 1, \dots, q$, vastavalt algoritmile 1.
- (ii) Iga valimi $y_{1:n}^{(m)}$, $m = 1, \dots, q$, ja iga $j = 1, \dots, k$ korral leiame $n^{(m)}(j)$ (mitu korda j esineb vektoris $y_{1:n}^{(m)}$).
- (iii) Iga $m = 1, \dots, q$ ja iga $j = 1, \dots, k$ korral leiame tingliku keskvaertuse

$$\begin{aligned} E(\psi(x, \boldsymbol{\theta}_j) | x_{1:n}, y_{1:n}^{(m)}) &= \int \psi(x, \boldsymbol{\theta}_j) \psi(\boldsymbol{\theta}_j | x_{1:n}, y_{1:n}^{(m)}) d\boldsymbol{\theta}_j = \\ &= \frac{\int \psi(x, \boldsymbol{\theta}_j) \prod_{i: y_i^{(m)}=j} \psi(x_i, \boldsymbol{\theta}_j) \pi(\boldsymbol{\theta}_j) d\boldsymbol{\theta}_j}{\int \prod_{i: y_i^{(m)}=j} \psi(x_i, \boldsymbol{\theta}_j) \pi(\boldsymbol{\theta}_j) d\boldsymbol{\theta}_j}. \end{aligned}$$

(iv) Lõpuks keskmistame üle valimite $y_{1:n}^{(m)}$ ja leiame soovitud hinnangu

$$\hat{\psi}_0(x) = \frac{1}{q} \sum_{m=1}^q \sum_{j=1}^k \frac{\alpha_j + n^{(m)}(j)}{|\alpha| + n} \cdot \frac{\int \psi(x, \theta_j) \prod_{i: y_i^{(m)}=j} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j}{\int \prod_{i: y_i^{(m)}=j} \psi(x_i, \theta_j) \pi(\theta_j) d\theta_j}.$$

Nagu eelnevalt mainitud, ei ole mõnikord võimalik rakendada algoritmi 1 või leida punktis (iii) toodud integraali. Sellisel juhul tuleb hinnata tihedust toetudes iteratiivsele algoritmile 2, mille puhul genereeritakse ka vektori θ väärtusi. Paneme tähele, et fikseeritud $\theta^{(m)}$ korral

$$E(\psi(x, \theta_j^{(m)}) | x_{1:n}, y_{1:n}) = \psi(x, \theta_j^{(m)}),$$

kuna $\psi(x, \theta_j^{(m)})$ on nüüd konstant. Seguajaotuse tiheduse hindamise protsessi saab seega formuleerida altoodud viisil.

- (i) Iga $m = 1, \dots, q$ korral genereerime $\theta^{(m)}$ ja $y_{1:n}^{(m)}$ vastavalt algoritmile 2.
- (ii) Fikseeritud $\theta^{(m)}$ ja $y_{1:n}^{(m)}$ korral leiame hinnangu

$$\begin{aligned} \hat{\psi}_0^{(m)}(x) &= E\left(\sum_{j=1}^k \mathbf{P}_j \psi(x, \theta_j^{(m)}) | x_{1:n}, y_{1:n}^{(m)}\right) = \\ &= \sum_{j=1}^k \left(\frac{\alpha_j + n^{(m)}(j)}{|\alpha| + n}\right) E(\psi(x, \theta_j^{(m)}) | x_{1:n}, y_{1:n}^{(m)}) = \\ &= \sum_{j=1}^k \left(\frac{\alpha_j + n^{(m)}(j)}{|\alpha| + n}\right) \psi(x, \theta_j^{(m)}). \end{aligned}$$

- (iii) Nüüd keskmistades üle $m = 1, \dots, q$ leiame soovitud seguajaotuse tiheduse hinnangu:

$$\hat{\psi}_0(x) = \frac{1}{q} \sum_{m=1}^q \hat{\psi}_0^{(m)}(x).$$

2 Dirichlet' protsess: mõningad eelteadmised

Eelmises peatükis eeldasime, et hinnatav tihedus ψ_0 on segujaotuse tihedus

$$\psi_0(x) = \int \psi(x; \theta) p_0(d\theta),$$

kus p_0 on parameetritega $p_1^0, \dots, p_k^0, \theta_1^0, \dots, \theta_k^0$ üheselt määratud diskreetne tõenäosusmõõt. Järgnevalt soovime üldistada mudeli konfiguratsiooni juhule, kus p_0 võib olla suvaline tõenäosusmõõt ruumil $(\Theta, \mathcal{B}(\Theta))$. Bayesi raamistikus nõuab niisugune probleemipüstitus seda, et käsitleksime mõõtu p_0 juhusliku suurusena P ja annaksime juhuslikule mõõdule P mingi eeljaotuse. Selles peatükis kirjeldame juhusliku tõenäosusmõõdu defineerimiseks tarvilikku mõõduteoreetilist aparatuuri, seejärel esitame Dirichlet' protsessi jaotuse – juhusliku mõõdu P loomuliku eeljaotuse – definitsiooni ja vaatleme selle olulisemaid omadusi. Definitsioonid ja tõestuseta esitatud tulemused on võetud vahetute tsitaatidena autori bakalaureusetööst [14]. Käesoleva töö ulatuses eeldame, et parameeterruum Θ on eukleidiline ruum \mathbb{R}^m ja vastav σ -algebra $\mathcal{B}(\Theta)$ on vastav Boreli σ -algebra $\mathcal{B}(\mathbb{R}^m)$. Antud peatüki raames esitatud tulemused kehtivad aga ka üldisema separaabli ja täieliku meetrilise ruumi (S, d) korral. Olgu $\mathcal{B}(S)$ sellele ruumile vastav Boreli σ -algebra ja \mathcal{P} olgu kõigi mõõtuval ruumil $(S, \mathcal{B}(S))$ määratud tõenäosusmõõtude hulk. Juhusliku tõenäosusmõõdu P defineerimiseks peame kõigepealt täpsustama, millised \mathcal{P} alamhulgad on mõõtuvad.

Selleks defineerime Prohhorovi meetrika ρ , mis on kaugus ruumil \mathcal{P} . Kaugus määrab teadaolevalt lahtiste hulkade kogumi τ_ρ ja see annab meile omakorda Boreli σ -algebra $\mathcal{B}(\mathcal{P})$, mis on vähim lahtiseid hulki sisaldav σ -algebra. Märgime ette-ruttavalt, et Dirichlet' protsessi jaotus ongi mõõtuval ruumil $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ määratud tõenäosusmõõt.

Definitsioon 2.1. Prohhorovi meetrika. Prohhorovi meetrikaks nimetatakse funktsiooni $\rho : \mathcal{P} \times \mathcal{P} \rightarrow [0, \infty)$, mis on defineeritud seosega:

$$\rho(p, q) = \inf\{\epsilon > 0 : p(A) \leq q(A^\epsilon) + \epsilon \text{ ja } q(A) \leq p(A^\epsilon) + \epsilon, \forall A \in \mathcal{B}(S)\},$$

kus $A^\epsilon = \{x \in S : d(x, A) < \epsilon\}$.

Definitsioon 2.2. Juhuslik tõenäosusmõõt. Olgu $(\Omega, \Sigma, \mathbf{P})$ mingi tõenäosusruum ja $\mathcal{B}(\mathcal{P})$ Prohhorovi meetrika indutseeritud Boreli σ -algebra. $(\Sigma, \mathcal{B}(\mathcal{P}))$ -mõõtuvat kujutust $P : \Omega \rightarrow \mathcal{P}$ nimetatakse juhuslikuks tõenäosusmõõduks ruumil $(S, \mathcal{B}(S))$.

Iga $A \in \mathcal{B}(S)$ korral võime vaadelda kujutust

$$T_A : \mathcal{P} \rightarrow [0, 1], \quad T_A(p) = p(A).$$

Kui funktsioon $f : S \rightarrow \mathbb{R}$ on integreeruv, siis omab tähendust ka kujutus

$$T_f : \mathcal{P} \rightarrow [0, 1], \quad T_f(p) = \int_S f dp.$$

Seda laadi kujutused on kasulikud nii Dirichlet' protsessi defineerimisel kui mõningate edaspidise arutelu seisukohalt oluliste omaduste tõestamisel. Nende kohta kehtib järgmine kasulik lause, mille tõestuse leiab [14, lause 1.16].

Lause 2.3. (a) Kui $f : S \rightarrow \mathbb{R}$ on pidev ja tõkestatud funktsioon, siis kujutus T_f on $\mathcal{B}(\mathcal{P})$ -mõõtuv.

(b) Iga $A \in \mathcal{B}(S)$ korral kujutus T_A on $\mathcal{B}(\mathcal{P})$ -mõõtuv.

Tähistades kompositsiooni $T_A \circ P$ kui $P(A)$, oleme valmis defineerima Dirichlet' protsessi.

Definitsioon 2.4. Dirichlet' protsess. Olgu α lõplik mõõt ruumil $(S, \mathcal{B}(S))$. Dirichlet' protsessiks baasimõõduga α nimetatakse juhuslikku tõenäosusmõõtu P

ruumil $(S, \mathcal{B}(S))$, mis rahuldab ruumi S iga lõpliku mõõtuva tükelduse A_1, \dots, A_k korral omadust

$$(P(A_1), \dots, P(A_k)) \sim \text{Dir}(k; \alpha(A_1), \dots, \alpha(A_k)).$$

Märkus 2.4.1. Kui P on Dirichlet' protsess baasimõõduga α , siis see defineerib ruumil $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ tõenäosusmõõdu $\mathbf{P}P^{-1}$, mida nimetame Dirichlet' protsessi jaotuseks ja tähistame sümboliga Π_α .

Mõningate allpool toodud tulemuste tõestamisel tuleb leida $P(A)$ tüüpi kompositsioonide keskväärtsi. Baasimõõduga α Dirichlet' protsessi korral on nende arvutamine lihtne. Kui $\bar{\alpha} = \frac{\alpha}{\alpha(S)}$, kehtib järgmine lause, mille tõestus on toodud autori bakalaureusetöös [14, lause 3.11].

Lause 2.5. *Olgu P Dirichlet' protsess baasimõõduga α . Siis suvalise $A \in \mathcal{B}(S)$ korral*

$$E(P(A)) = \bar{\alpha}(A).$$

Dirichlet' protsessi lõplikku baasimõõtu α võib esitada ka kujul $\alpha = M \cdot \bar{\alpha}$, kus nn kontsentratsiooniparameeter M on konstant $\alpha(S)$ ja $\bar{\alpha}$ seosega $\bar{\alpha}(A) = \frac{\alpha(A)}{M}$, $A \in \mathcal{B}(S)$ määratud tõenäosusmõõt. Järgmine lause avab kontsentratsiooniparameetri M praktilist tähendust. Nimelt kompositsiooni $P(A)$ dispersioon on seda väiksem, mida suurem on M väärtus. Teisipidi võib öelda, et M suurenedes kontsentreerub Dirichlet' protsessi jaotuse tõenäosusmass tõenäosusmõõdu $\bar{\alpha}$ ümbruses. Lause 2.6 tõestus on esitatud [14, lause 3.12].

Lause 2.6. *Olgu P Dirichlet' protsess baasimõõduga α . Siis suvalise $A \in \mathcal{B}(S)$ korral*

$$D(P(A)) = \frac{\bar{\alpha}(A) \cdot (1 - \bar{\alpha}(A))}{M + 1}.$$

Üks Dirichlet' protsessi suurimaid voorusi seisneb selles, et tema järeljaotus on samuti Dirichlet' protsessi jaotus. Formaalselt võtab omaduse kokku lause 2.7. Selle tõestuse leiab [14, teoreem 3.15].

Lause 2.7. Olgu P Dirichlet' protsess baasimõõduga α ja fikseeritud $P = p$ korral olgu X_1, \dots, X_n sõltumatud juhuslikud suurused jaotusega p . Siis peaaegu iga X_1, \dots, X_n realisatsiooni x_1, \dots, x_n puhul

$$P|X_1 = x_1, \dots, X_n = x_n \sim \Pi_{\alpha + \sum_{i=1}^n \delta_{x_i}}.$$

Lause 2.7 teeb Dirichlet' protsessi jaotusest loomuliku eeljaotuse valiku situatsioonis, kus puudub kindlus andmeid genereerinud mehhanismi osas (kasvõi juhul, kui ei soovita eeldada, et vaatlused pärinevad ühest konkreetsest jaotuste perest). Näiteks kui arvame, et olemasolevad andmed x_1, \dots, x_n on sõltumatud juhuslikud vaatlused jaotusest q , aga meil on ka kahtlusi oma intuitsioonis, võime võtta kasutusele mudeli

$$\begin{aligned} P &\sim \Pi_{Mq}, \\ X_1, \dots, X_n | P = p &\stackrel{i.i.d.}{\sim} p, \end{aligned}$$

kus konstant M väljendab, millisel määral me usaldame oma hüpoteesi, et andmeid genereerinud jaotus on q .

Siinkohal kerkib esile loomulik küsimus: milleks eeldada, et hinnatav tihedus ψ_0 esitub segujaotuse tihedusena ja seada eeljaotust mõõdule p_0 selle asemel, et vaadelda andmeid x_1, \dots, x_n vaatlustena mudelist

$$\begin{aligned} P &\sim \Pi_{M\Psi'}, \\ X_1, \dots, X_n | P = p &\stackrel{i.i.d.}{\sim} p, \end{aligned} \tag{2.1}$$

kus Ψ' on tegelikule tihedusele vastava tõenäosusmõõdu Ψ_0 aprioorne hinnang?

Järgnevalt näitame, miks mudel (2.1) ei ole siiski tiheduse ψ_0 hindamiseks sobilik ja segujaotuse eeldusest ei ole võimalik loobuda. Nimelt osutub (vt teoremi 2.11), et Dirichlet' protsessi realisatsioon on peaaegu kindlasti diskreetne tõenäosusmõõt.

Viimane fakt tähendab omakorda seda, et mudeli (2.1) eeldustel ei saa tegelikul jaotusel Ψ_0 olla tihedust Lebesgue'i mõõdu suhtes – meie aga tahame hinnata just sellist tihedust. Vastava tõestuskäigu hõlbustamiseks esitame kõigepealt lemmad 2.8 – 2.10.

Lemma 2.8 tõestus tugineb raamatus [7, Proposition A5] esitatud tõestuse skeemile.

Lemma 2.8. *Olgu $A \in \mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$. Siis kujutus*

$$f_A : \mathcal{P} \rightarrow \mathbb{R}, \quad f_A(p) = \int I_A(p, x)p(dx),$$

on $\mathcal{B}(\mathcal{P})$ -mõõtuv.

Tõestus. Paneme tähele, et $I_A(p, x) = 1$ parajasti siis, kui $x \in A_p = \{x : (p, x) \in A\}$. Järelikult suvalise $p \in \mathcal{P}$ korral

$$f_A(p) = \int I_A(p, x)p(dx) = \int I_{A_p}(x)p(dx) = p(A_p).$$

Nüüd vaatleme kogumit $\mathcal{L} = \{A \in \mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S) : f_A \text{ on } \mathcal{B}(\mathcal{P})\text{-mõõtuv}\}$. Veendume, et tegemist on λ -süsteemiga hulgal $\mathcal{P} \times S$.

1. Olgu $A = \mathcal{P} \times S$. Siis suvalise $p \in \mathcal{P}$ korral $A_p = S$ ja $f_A(p) = p(S) = 1$. Fikseerime vabalt $B \in \mathcal{B}(\mathbb{R})$. Kui $1 \in B$, siis $f_A^{-1}(B) = \mathcal{P} \in \mathcal{B}(\mathcal{P})$. Vastasel juhul $f_A^{-1}(B) = \emptyset \in \mathcal{B}(\mathcal{P})$. Järelikult $\mathcal{P} \times S \in \mathcal{L}$.
2. Olgu $A \in \mathcal{L}$ ja $p \in \mathcal{P}$ suvalised. Siis

$$f_{A^c}(p) = p((A^c)_p) = p((A_p)^c) = 1 - p(A_p) = 1 - f_A(p).$$

Kuna f_{A^c} esitub f_A pideva teisendusena ja f_A on $\mathcal{B}(\mathcal{P})$ -mõõtuv, siis ka f_{A^c} on $\mathcal{B}(\mathcal{P})$ -mõõtuv ning $A^c \in \mathcal{L}$.

3. Olgu $A_1, A_2, \dots \in \mathcal{L}$ lõikumatud. Defineerime funktsioonid $g_n(p) = \sum_{i=1}^n f_{A_i}(p)$, $n \in \mathbb{N}$. Et iga $n \in \mathbb{N}$ korral g_n on mõõtuvate funktsioonide lõplik summa,

siis iga g_n on mõõtv. Paneme tähele

$$f_{\bigcup_i A_i}(p) = p\left(\left(\bigcup_i A_i\right)_p\right) = p\left(\bigcup_i (A_i)_p\right) = \sum_i p((A_i)_p) = \sum_i f_{A_i}(p) = \lim_n g_n(p).$$

Kuna $f_{\bigcup_i A_i}$ esitub mõõtuvate funktsioonide jada piirväärtusena, on see samuti mõõtv. Järelikult $\bigcup_i A_i \in \mathcal{L}$.

Nüüd paneme tähele, et mõõtuvate ristkülikute π -süsteem $\mathcal{M} = \{C \times D : C \in \mathcal{B}(\mathcal{P}), D \in \mathcal{B}(S)\}$ on kogumi \mathcal{L} alamhulk. Tõepoolest, suvaliste $p \in \mathcal{P}$ ja $C \times D \in \mathcal{M}$ korral

$$f_{C \times D}(p) = p((C \times D)_p) = \begin{cases} p(D), & \text{kui } p \in C, \\ p(\emptyset), & \text{kui } p \notin C. \end{cases}$$

See tähendab, et $f_{C \times D}(p) = T_D(p) \cdot I_C(p)$. Kuna T_D ja I_C on mõlemad mõõtvad lause 2.3 põhjal, siis ka nende korrutis on mõõtv. Niisiis, $C \times D \in \mathcal{L}$.

Kuna $\mathcal{M} \subset \mathcal{L}$, siis Dynkini $\pi - \lambda$ teoreemi põhjal $\sigma(\mathcal{M}) \subset \mathcal{L}$. Meenutame, et korrutis- σ -algebra $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ on defineeritud kui $\sigma(\mathcal{M})$. Ühtlasi kehtib sisalduvus $\mathcal{L} \subset \mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$. Kokkuvõttes saame seega, et $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S) = \mathcal{L}$. \square

Lemma 2.9 tõestus tugineb raamatus [7, Proposition A8] esitatud tõestuse skeemile.

Lemma 2.9. *Hulk $E = \{(p, x) \in \mathcal{P} \times S : p(\{x\}) > 0\}$ on $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ -mõõtv.*

Tõestus. Näitamaks, et $E \in \mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$, piisab veenduda, et kujutus

$$\xi : \mathcal{P} \times S \rightarrow [0, 1], \quad \xi(p, x) = p(\{x\})$$

on $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ -mõõtv, sest $E = \xi^{-1}((0, 1])$.

Iga $F \in \mathcal{B}(S) \otimes \mathcal{B}(S)$ korral defineerime funktsiooni

$$g_F : \mathcal{P} \times S, \quad g_F(p, x) = p(F_x),$$

kus $F_x = \{y : (x, y) \in F\}$. Vaatleme kogumit

$$\mathcal{N} = \{F \in \mathcal{B}(S) \otimes \mathcal{B}(S) : g_F \text{ on } \mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)\text{-mõõtuv}\}.$$

Veendume, et tegemist on λ -süsteemiga hulgal $S \times S$.

1. Olgu $(p, x) \in \mathcal{P} \times S$ suvaline. Siis

$$g_{S \times S}(p, x) = p((S \times S)_x) = p(S) = 1.$$

Järelikult iga $B \in \mathcal{B}(\mathbb{R})$ korral $g_{S \times S}^{-1}(B) = \mathcal{P} \times S \in \mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$, kui $1 \in B$.

Vastasel juhul $g_{S \times S}^{-1}(B) = \emptyset \in \mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$. Niisiis, $S \times S \in \mathcal{N}$.

2. Olgu $F \in \mathcal{N}$ ja $(p, x) \in \mathcal{P} \times S$ suvalised. Siis

$$g_{F^c}(p, x) = p((F^c)_x) = p((F_x)^c) = 1 - p(F_x) = 1 - g_F(p, x).$$

Kuna g_{F^c} esitub funktsiooni g_F pideva teisendusena ja g_F on $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ -mõõtuv, siis ka g_{F^c} on $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ -mõõtuv ja $F^c \in \mathcal{N}$.

3. Olgu $F_1, F_2, \dots \in \mathcal{N}$ lõikumatud. Defineerime iga $n \in \mathbb{N}$ korral mõõtuvad funktsioonid

$$g_n : \mathcal{P} \times S, \quad g_n(p, x) = \sum_i^n g_{F_i}(p, x).$$

Märkame, et

$$\begin{aligned} g_{\bigcup_i F_i}(p, x) &= p\left(\left(\bigcup_i F_i\right)_x\right) = p\left(\bigcup_i (F_i)_x\right) = \sum_i p((F_i)_x) = \\ &= \sum_i g_{F_i}(p, x) = \lim_n g_n(p, x). \end{aligned}$$

Et $g_{\bigcup_i F_i}(p, x)$ on mõõtuvate funktsioonide jada piirväärtus, on see samuti $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ -mõõtuv. Järelikult $\bigcup_i F_i \in \mathcal{N}$.

Olgu \mathcal{G} hulga $S \times S$ kõigi mõõtuvate ristkülikute hulk, s.t $\mathcal{G} = \{G \times H : G, H \in$

$\mathcal{B}(S)\}$. Fikseerime vabalt $G \times H \in \mathcal{G}$. Siis

$$\begin{aligned} g_{G \times H}(p, x) &= p((G \times H)_x) = p(H) \cdot I_G(x) = \\ &= T_H(p) \cdot I_G(x) = T_H(\pi_1(p, x)) \cdot I_G(\pi_2(p, x)), \end{aligned}$$

kus $\pi_1 : (p, x) \mapsto p$ ja $\pi_2 : (p, x) \mapsto x$ on koordinaatfunktsioonid. Korrutis- σ -algebra definitsiooni kohaselt on π_1 ja π_2 vastavalt $(\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S), \mathcal{B}(\mathcal{P}))$ - ja $(\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S), \mathcal{B}(S))$ -mõõtuvad. Teame ka, et T_H on $(\mathcal{B}(\mathcal{P}), \mathcal{B}(\mathbb{R}))$ -mõõtuv lause 2.3 põhjal ja I_G on $(\mathcal{B}(S), \mathcal{B}(\mathbb{R}))$ -mõõtuv. Järelikult kompositsioonid $T_H \circ \pi_1$ ja $I_G \circ \pi_2$ on vastavalt $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ -mõõtuvad. Niisiis ka nende korrutisena esitatav funktsioon $g_{G \times H}$ on $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ -mõõtuv. Sellega oleme tõestanud, et $\mathcal{G} \subset \mathcal{N}$.

On ilmne, et \mathcal{G} on π -süsteem. Nüüd Dynkini $\pi - \lambda$ teoreemi kohaselt $\mathcal{B}(S) \otimes \mathcal{B}(S) = \sigma(\mathcal{G}) \subset \mathcal{N}$. See tähendab, et $\mathcal{N} = \mathcal{B}(S) \otimes \mathcal{B}(S)$.

Lõpetuseks veendume, et diagonaal $D = \{(x, x) : x \in S\}$ on $\mathcal{B}(S) \otimes \mathcal{B}(S)$ -mõõtuv. Et S on meetriline ruum, siis see on ka Hausdorffi ruum. Kuna Hausdorffi ruumi korral on diagonaal D hulga $S \times S$ kinnine alamhulk [8, lause 2.6], siis $D \in \mathcal{B}(S \times S)$. Ruumi S separaalusel järeldeb $S \times S$ separaablus, mis omakorda tähendab, et $\mathcal{B}(S \times S) = \mathcal{B}(S) \otimes \mathcal{B}(S)$ [2, lk 244]. Niisiis, $D \in \mathcal{B}(S) \otimes \mathcal{B}(S)$.

Kuna suvalise $(p, x) \in \mathcal{P} \times S$ korral $\xi(p, x) = g_D(p, x)$, oleme tõestanud, et kujutus ξ on $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ -mõõtuv ja seega $E \in \mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$. \square

Lemma 2.10 tõestus tugineb raamatus [7, Proposition A7] esitatud tõestuse skeemile.

Lemma 2.10. *Kõigi ruumil $(S, \mathcal{B}(S))$ määratud diskreetsete tönäosusmõõtude hulk*

$$D = \{p \in \mathcal{P} : p(\{x : p(\{x\}) > 0\}) = 1\}$$

on $\mathcal{B}(\mathcal{P})$ -mõõtuv.

Tõestus. Lemma 2.9 põhjal hulk $E = \{(p, x) \in \mathcal{P} \times S : p(\{x\}) > 0\}$ on $\mathcal{B}(\mathcal{P}) \otimes \mathcal{B}(S)$ -

mõõtuv. Seega lemma 2.8 põhjal kujutus f_E on $\mathcal{B}(\mathcal{P})$ -mõõtuv.

Paneme tähele, et $I_E(p, x) = 1$ parajasti siis, kui $(p, x) \in E$ ehk $x \in \{x : p(\{x\}) > 0\}$. Seega suvalise $p \in \mathcal{P}$ korral

$$f_E(p) = \int I_E(p, x)p(dx) = \int I_{\{x:p(\{x\})>0\}}(x)p(dx) = p(\{x : p(\{x\}) > 0\}).$$

Näeme, et $D = f_E^{-1}(\{1\})$. Kuna $\{1\} \in \mathcal{B}(\mathbb{R})$ ja f_E on $\mathcal{B}(\mathcal{P})$ -mõõtuv, siis $D \in \mathcal{B}(\mathcal{P})$. \square

Teoreemi 2.11 tõestus tugineb raamatus [7, Theorem 4.14] esitatud tõestuse skeemile.

Teoreem 2.11. *Peaaegu iga Dirichlet protsessi realisatsioon on diskreetne tõenäosusmõõt. S.t*

$$\Pi_\alpha(\{p \in \mathcal{P} : p(\{x : p(\{x\}) > 0\}) = 1\}) = 1.$$

Tõestus. Olgu juhuslik mõõt P Dirichlet' protsess baasimõõduga α ja juhuslik suurus X olgu vaatlus Dirichlet' protsessist, s.t $P \sim \Pi_\alpha$ ja $X|P = p \sim p$. Tõenäosusmõõt $p \in \mathcal{P}$ on diskreetne parajasti siis, kui $p(\{x : p(\{x\}) > 0\}) = 1$. Hulk $\{\omega : P(\omega)(\{X(\omega)\}) > 0\}$ on \mathcal{F} -mõõtuv, kuna seda saab esitada kujul $(\xi(P, X))^{-1}((0, 1])$, kus ξ on lemmas defineeritud 2.9 kujutus. Kuna lemmas 2.9 on näidatud, et ξ on mõõtuv, siis ka eeltoodud originaal on ilmselt mõõtuv. Seejuures paneme tähele, et

$$\begin{aligned} \mathbf{P}(P(\{X\}) > 0) &= \int \mathbf{P}(P(\{X\}) > 0 | P = p) \Pi_\alpha(dp) = \\ &= \int \mathbf{P}(\{\omega : p(\{X(\omega)\}) > 0\}) \Pi_\alpha(dp) = \\ &= \int \mathbf{P}(X(\omega) \in \{x : p(\{x\}) > 0\}) \Pi_\alpha(dp) = \\ &= \int \mathbf{P}X^{-1}(\{x : p(\{x\}) > 0\}) \Pi_\alpha(dp) = \\ &= \int p(\{x : p(\{x\}) > 0\}) \Pi_\alpha(dp). \end{aligned}$$

Siit järeldub, et peaaegu iga P realisatsioon on diskreetne parajasti siis, kui $\mathbf{P}(P(\{X\}) > 0) = 1$. Tõepoolest, oletame, et $p(\{x : p(\{x\}) > 0\}) = 1$ Π_α -p.k. (eeldus on tähenduslik, kuna lemma 2.10 põhjal on kõigi diskreetsete jaotuste hulk mõõtuv). Siis

$$\mathbf{P}(P(\{X\}) > 0) = \int p(\{x : p(\{x\}) > 0\})\Pi_\alpha(dp) = \int 1\Pi_\alpha(dp) = 1.$$

Nüüd oletame, et $p(\{x : p(\{x\}) > 0\}) \neq 1$ Π_α -p.k. See tähendab, et leidub mingi hulk $A \in \mathcal{B}(\mathcal{P})$ nii, et $\Pi_\alpha(A) > 0$ ja iga $p \in A$ korral $p(\{x : p(\{x\}) > 0\}) < 1$. Nüüd Lebesgue'i integraali monotoonsuse tõttu

$$\begin{aligned} \mathbf{P}(P(\{X\}) > 0) &= \int_A p(\{x : p(\{x\}) > 0\})\Pi_\alpha(dp) + \\ &+ \int_{A^c} p(\{x : p(\{x\}) > 0\})\Pi_\alpha(dp) < \Pi_\alpha(A) + \Pi_\alpha(A^c) = 1. \end{aligned}$$

Ehk kui $\mathbf{P}(P(\{X\}) > 0) = 1$, siis peaaegu iga P realisatsioon on diskreetne.

Olgu tehtud vaatlus $X = x$. Lause 2.7 põhjal on P järeljaotus Dirichlet' protsessi jaotus baasimõõduga $\alpha + \delta_x$, s.t $P|X = x \sim \Pi_{\alpha+\delta_x}$. Vaatleme nüüd ruumi S mõõtuvat tükeldust $\{x\}, \{x\}^c$. Dirichlet' protsessi definitsiooni kohaselt

$$(P(\{x\}), P(\{x\}^c))|X = x \sim Dir(\alpha(\{x\}) + 1, \alpha(\{x\}^c)),$$

mis on samaväärne asjaoluga $P(\{x\})|X = x \sim Be(\alpha(\{x\}) + 1, \alpha(\{x\}^c))$. Et beeta-jaotuse realisatsioon on peaaegu kindlasti positiivne, siis

$$\mathbf{P}(P(\{X\}) > 0|X = x) = \mathbf{P}(P(\{x\}) > 0) = 1.$$

Kui p_X on jaotus, millest pärineb meie vaatlus (teatud P realisatsioon), siis

$$\mathbf{P}(P(\{X\}) > 0) = \int \mathbf{P}(P(\{X\}) > 0|X = x)p_X(dx) = \int 1p_X(dx) = p_X(S) = 1.$$

□

Tõestame veel ühe tulemuse, mida rakendame 3. peatükis. Lause 2.12 tõestus põhineb raamatus [7, Proposition 4.3] esitatud tõestuse ideel.

Lause 2.12. *Olgu $f : S \rightarrow \mathbb{R}$ $\mathcal{B}(S)$ -mõõtuv integreeruv funktsioon ja kujutus T_f määratud seosega*

$$T_f : \mathcal{P} \rightarrow \mathbb{R}, \quad p \mapsto \int f(x)p(dx).$$

Kui P on Dirichlet' protsess baasimõõduga α , siis

$$E(T_f(P)) = \int f(x)\bar{\alpha}(dx).$$

Tõestus. 1) Kõigepealt vaatleme juhtu, kus $A \in \mathcal{B}(S)$ ja $f = I_A$. Siis

$$\begin{aligned} E(T_f(P)) &= \int_{\Omega} T_f(P(\omega))\mathbf{P}(d\omega) = \int_{\mathcal{P}} T_f(p)\Pi_{\alpha}(dp) = \int_{\mathcal{P}} \left(\int_S I_A p(dx) \right) \Pi_{\alpha}(dp) = \\ &= \int_{\mathcal{P}} p(A)\Pi_{\alpha}(dp) = \int_{\mathcal{P}} T_A(p)\Pi_{\alpha}(dp) = \int_{\Omega} T_A(P(\omega))\mathbf{P}(d\omega) = \\ &= E(T_A(P)) = E(P(A)) = \bar{\alpha}(A) = \int I_A(x)\bar{\alpha}(dx), \end{aligned}$$

kus eelviimane võrdus kehtib lause 2.5 tõttu.

2) Nüüd oletame, et f on lihtne mõõtuv funktsioon, s.t $f = \sum_{i=1}^n \alpha_i I_{A_i}$, kus $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ ja $A_1, \dots, A_n \in \mathcal{B}(S)$. Siis

$$T_f(p) = \int \sum_{i=1}^n \alpha_i I_{A_i} p(dx) = \sum_{i=1}^n \alpha_i \int I_{A_i} p(dx) = \sum_{i=1}^n \alpha_i T_{I_{A_i}}(p).$$

Keskvärtuse lineaarsuse tõttu

$$\begin{aligned} E(T_f(P)) &= E\left(\sum_{i=1}^n \alpha_i T_{I_{A_i}}(P) \right) = \sum_{i=1}^n \alpha_i E(T_{I_{A_i}}(P)) = \sum_{i=1}^n \alpha_i \int I_{A_i}(x)\bar{\alpha}(dx) = \\ &= \int \sum_{i=1}^n \alpha_i I_{A_i}(x)\bar{\alpha}(dx) = \int f(x)\bar{\alpha}(dx). \end{aligned}$$

3) Kui f on mittenegatiivne $\mathcal{B}(S)$ -mõõtuv funktsioon, siis leidub selline lihtsate mõõtuvate funktsioonide jada f_n , et $f_n \nearrow f$. Monotoonse koondumise teoreemi tõttu kehtib ka koondumine $T_{f_n} \nearrow T_f$. Tõepoolest, suvalise $p \in \mathcal{P}$ korral

$$\lim_n T_{f_n}(p) = \lim_n \int f_n p(dx) = \int f p(dx) = T_f(p).$$

Kasutades veel kord monotoonse koondumise teoreemi ja eelmises punktis tõestatud lihtsate mõõtuvate funktsioonide omadust, saame, et

$$E(T_f(P)) = \lim_n E(T_{f_n}(P)) = \lim_n \int f_n \bar{\alpha}(dx) = \int f \bar{\alpha}(dx).$$

4) Lõpetuseks paneme tähele, et integreeruv funktsioon f esitub kujul $f = f^+ - f^-$, kus f^+ ja f^- on mittenegatiivsed mõõtuvad funktsioonid. Lebesgue'i integraali lineaarsuse tõttu kehtib suvalise $p \in \mathcal{P}$ korral

$$T_f(p) = \int f p(dx) = \int (f^+ - f^-) p(dx) = \int f^+ p(dx) - \int f^- p(dx) = T_{f^+}(p) - T_{f^-}(p).$$

Nüüd rakendame taas lineaarsuse omadust ja punkti (3):

$$\begin{aligned} E(T_f(P)) &= E(T_{f^+}(P) - T_{f^-}(P)) = E(T_{f^+}(P)) - E(T_{f^-}(P)) = \\ &= \int f^+ \bar{\alpha}(dx) - \int f^- \bar{\alpha}(dx) = \int (f^+ - f^-) \bar{\alpha}(dx) = \int f \bar{\alpha}(dx). \end{aligned}$$

□

Eelnevalt juba tõime näitena mudelit

$$\begin{aligned} P &\sim \Pi_\alpha, \\ X_1, \dots, X_n | P &= p \stackrel{i.i.d}{\sim} p, \end{aligned}$$

mida võib nimetada "vaatlusteks Dirichlet' protsessist". Juhuslike suuruste X_1, \dots, X_n ühisjaotus on sageli keeruline, $X_1, X_2 | X_1, \dots, X_n | X_1, \dots, X_n$ tinglikud jaotused

aga omavad palju lihtsamat kuju. Nende kirjeldamiseks esitame üldistatud Polya urni mõiste.

Definitsioon 2.13. Olgu α lõplik mõõt ruumil $(S, \mathcal{B}(S))$, $M = \alpha(S)$ ja $\bar{\alpha}$ seosega

$$\bar{\alpha}(B) = \frac{\alpha(B)}{M}, \quad B \in \mathcal{B}(S)$$

määratud tõenäosusmõõt. Juhuslike suuruste jada X_1, X_2, \dots nimetatakse üldistatud Polya urni protsessiks, kui $X_1 \sim \bar{\alpha}$ ja iga $n \in \mathbb{N}$ korral

$$X_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1} \sim \begin{cases} \delta_{x_1} & \text{tõenäosusega } \frac{1}{M+n-1}, \\ \vdots & \\ \delta_{x_{n-1}} & \text{tõenäosusega } \frac{1}{M+n-1}, \\ \bar{\alpha} & \text{tõenäosusega } \frac{M}{M+n-1}. \end{cases}$$

Osutub, et vaatlused Dirichlet' protsessist ongi Polya urni protsess [14, teoreem 3.17]. Näeme, et Dirichlet' protsessit pärit vaatluste korral väheneb n kasvades uute väärtuste esinemise tõenäosus ja kasvab vanade väärtuste "kordamise" tõenäosus.

Kirjeldatud tendentsi illustreerimiseks räägitakse sageli nn Hiina restorani protsessist. Kujutame ette Hiina restorani, kus on loenduv arv laudu, kusjuures ühegi laua juures ei saa istekoht kunagi otsa (toole võib alati juurde tuua). Järjestikuseid realisatsioone Polya urni protsessist võib vaadelda Hiina restorani saabuvate klientidena. Esimene klient saabub söögikohta ja istub suvalisse lauda (genereeritakse väärtus x_1 jaotusest $\bar{\alpha}$). Järgmine klient valib uue laua tõenäosusega $\frac{1}{M+1}$ (väärtus x_2 genereeritakse samuti jaotusest $\bar{\alpha}$) või istub esimese külalise juurde tõenäosusega $\frac{M}{M+1}$ ($x_2 = x_1$). Laud muutub seda populaarsemaks, mida rohkem inimesi istub selle taga. Kui m -ndas lauas istub k inimest, istub n -is klient nende juurde tõenäosusega $\frac{k}{M+n-1}$ või valib mõne tühja laua tõenäosusega $\frac{M}{M+n-1}$. Antud seaduspära olulisus avaldub järgmises peatükis, kuna Dirichlet protsessist

pärit vaatluste kordumisega arvestamine võimaldab tuletada efektiivsemaid algoritme tiheduse hindamiseks.

3 Mitteparameetriline Bayesi mudel

Tuletame meelde, et siinse magistritöö eesmärk on hinnata andmeid x_1, \dots, x_n genereerinud tihedust ψ_0 . Esimeses peatükis eeldasime, et huvipakkuv tihedus esitub kujul

$$\psi_0(x) = \sum_{j=1}^k p_j^0 \psi(x; \theta_j^0),$$

s.t tegemist on lõpliku seguga, mille komponentide arv on meile teada. Viimane eeldus taandas tiheduse hindamise probleemi parameetrite

$$p_1^0, \dots, p_k^0, \theta_1^0, \dots, \theta_k^0$$

hindamisele. Bayesi raamistikus sai vaadelda eelmainitud parameetreid juhuslike suurustena ja postuleerida oma aimdusi nende väärtustest vastavate eeljaotuste kaudu.

Olgu mainitud, et kirjeldatud eeldused tegeliku tiheduse kohta on siiski küllaltki kitsendavad. Iga tihedus ei ole esitatav lõpliku seguna ja ka lõplike segude korral ei ole täpne komponentide arv k sageli teada. Selles peatükis vaatleme üldisemat juhtu, kus ψ_0 on segujaotuse tihedus

$$\psi_0 = \int \psi(x; \theta) p_0(d\theta),$$

kusjuures p_0 võib olla suvaline tõenäosusmõõt ruumil $(\Theta, \mathcal{B}(\Theta))$.

Bayesi raamistikus vaatleme tõenäosusmõõtu p_0 juhusliku mõõduna P ja postuleerime oma apriorseid teadmisi selle kohta eeljaotuse kaudu. Juhul, kui me ei soovi teha tugevaid eeldusi mõõdu p_0 kohta (nt me ei tea kindlalt, mis jaotuste perega on tegemist), üks loomulikke eeljaotuse valikuid on Dirichlet' protsessi jaotus. Niisuguse valiku kasuks räägib ennekõike kolm asjaolu:

1. Saame postuleerida oma eelarusaama tõelisest jaotusest p_0 läbi baasimõõdu α . Tuletame meelde näite leheküljelt 23. Kui arvame, et $\theta_1, \dots, \theta_n$ on valim

jaotusest q , siis võtame Dirichlet' protsessi baasimõõduks

$$\alpha = M \cdot q,$$

kus M on konstant, mis väljendab meie usaldusastet oma eelduse suhtes.

2. Vaatluste arvu n kasvades koondub $P|\boldsymbol{\theta}_1 = \theta_1, \dots, \boldsymbol{\theta}_n = \theta_n$ järeljaotuse tõenäosusmass tõelise jaotuse q ümber. Piisava hulga andmete korral peaksime jõudma tõelise jaotuse p_0 lähedale sõltumata sellest, kui halb oli meie esmane hinnang q [7, Corollary 4.17].
3. Kui P eeljaotus on Dirichlet' protsessi jaotus baasimõõduga α , siis $P|\boldsymbol{\theta}_1 = \theta_1, \dots, \boldsymbol{\theta}_n = \theta_n$ järeljaotus on Dirichlet protsessi jaotus baasimõõduga $\alpha + \sum_{i=1}^n \delta_{\theta_i}$ – s.t järeljaotuse "kuju" on meile teada.

Mingis mõttes kandub 1. ja 2. omadus ka (juhuslikule) tiheduse hinnangule

$$\psi_P(x) = \int \psi(x; \theta) P(d\theta).$$

1. Kui arvame, et tegelik tihedus ψ_0 esitub kujul

$$\psi_0 = \int \psi(x; \theta) q(d\theta),$$

saame postuleerida oma eelarusaama, võttes $P \sim \Pi_{Mq}$. Tõepoolest, lause 2.12 põhjal

$$E(\psi_P(x)) = \int \psi(x) q(d\theta).$$

2. Teatud tingimustel koondub $\psi_P|X_1 = x_1, \dots, X_n = x_n$ järeljaotuse tõenäosusmass vaatluste arvu n kasvades tegeliku tiheduse ψ_0 ümber ka siis, kui esmane hinnang q on halb. Mõjusust tagavaid lisaeldusi käsitleme lähemalt magistritöö 4. peatükis.

Dirichlet' protsessi 3. hea omadus ei kandu vahetult hinnangule ψ_P (s.t ψ_P järeljao-

tus ei ole üldiselt lihtsasti esitatav), kuid osutub sellegipoolest kasulikuks tiheduse hindamisel.

Eeltoodud kaalutlustel käsitleme andmeid x_1, \dots, x_n "vaatlustena Dirichlet' protsessi segust". Antud peatükis esitame vastava mudeli definitsiooni ja seejärel tuleme eeskirja tiheduse ψ_0 hindamiseks selle mudeli abil.

Definitsioon 3.1. Vaatlused Dirichlet' protsessi segust. Olgu $(\Omega, \mathcal{F}, \mathbf{P})$ tõenäosusruum. Parameetrite väärtuste hulga eristamiseks vaatluste väärtuste hulgast kasutame tähistust $\Theta = \mathbb{R}$. Kujutus $\psi : \mathbb{R} \times \Theta \rightarrow [0, \infty)$ olgu tuum. Vaatleme mõõtuvast ruumi $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$, kus \mathcal{P} on kõigi mõõtuval ruumil $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ määratud tõenäosusmõõtude hulk ja $\mathcal{B}(\mathcal{P})$ Prohhorovi meetrika indutseeritud σ -algebra (vt definitsiooni 2.1 koos kaasneva aruteluga). Samuti vaatleme mõõtuvaid ruume $(\Theta, \mathcal{B}(\Theta))$ ja $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ning juhuslikke suursusi

$$\begin{aligned} P &: \Omega \rightarrow \mathcal{P}, \\ \theta_i &: \Omega \rightarrow \Theta, \quad i = 1, \dots, n, \\ X_i &: \Omega \rightarrow \mathbb{R}, \quad i = 1, \dots, n, \end{aligned}$$

kus P on Dirichlet' protsess mitteatomaarse lõpliku baasimõõduga α , selle fikseeritud realisatsiooni p korral θ_i aga sõltumatud juhuslikud suurused jaotusega p . Vektori $(\theta_1, \dots, \theta_n)$ fikseeritud väärtuse $(\theta_1, \dots, \theta_n)$ korral olgu X_i üksteisest sõltumatud juhuslikud suurused tihedusega $\psi(\cdot; \theta_i)$ (Lebesgue'i mõõdu suhtes). Eelkirjeldatud mudelit nimetame vaatlusteks Dirichlet' protsessi segust ja tähistame järgnevalt:

$$\begin{aligned} P &\sim \Pi_\alpha, \\ \theta_1, \dots, \theta_n | P &= p \stackrel{\text{i.i.d.}}{\sim} p, \\ X_i | \theta_i &= \theta_i \stackrel{\text{ind}}{\sim} \psi(\cdot; \theta_i), \quad i = 1, \dots, n. \end{aligned} \tag{3.1}$$

Parim hinnang tihedusfunktsiooni väärtusele (ruutkaofunktsiooni mõttes) punktis $x \in \mathbb{R}$ etteantud vaatluste x_1, \dots, x_n korral on tinglik keskväärtus

$$E(\psi_P(x)|x_{1:n}) = E\left(\int \psi(x; \theta)P(d\theta)|x_{1:n}\right). \quad (3.2)$$

Paneme tähele, et tingliku keskväärtuse omaduse tõttu

$$E\left(\psi_P(x)|X_{1:n}\right) = E\left[E\left(\psi_P(x)|\boldsymbol{\theta}_{1:n}, X_{1:n}\right)|X_{1:n}\right] \quad \text{p.k.}$$

Parameetrite $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$ väärtuste fikseerimisel on P ja X_1, \dots, X_n omavahel tinglikult sõltumatud ning seega

$$E\left[E\left(\psi_P(x)|\boldsymbol{\theta}_{1:n}, X_{1:n}\right)|X_{1:n}\right] = E\left[E\left(\psi_P(x)|\boldsymbol{\theta}_{1:n}\right)|X_{1:n}\right] \quad \text{p.k.}$$

Viimane esitus on kasulik, kuna fikseeritud $\boldsymbol{\theta}_{1:n}$ korral on sisemise keskväärtuse leidmine pole keeruline. Nimelt lause 2.7 põhjal on

$$P|\boldsymbol{\theta}_1 = \theta_1, \dots, \boldsymbol{\theta}_n = \theta_n$$

järeljaotus Dirichlet' protsessi jaotus baasimõõduga $\alpha + \sum_{i=1}^n \delta_{\theta_i}$. Niisiis, lausest 2.12 järeldub, et

$$\begin{aligned} E\left(\int \psi(x; \theta)P(d\theta)|\boldsymbol{\theta}_{1:n}\right) &= \int \psi(x; \theta)\left(\frac{\alpha + \sum_{i=1}^n \delta_{\theta_i}}{M + n}\right)(d\theta) = \\ &= \frac{1}{M + n}\left[\int \psi(x; \theta)\alpha(d\theta) + \sum_{i=1}^n \psi(x; \theta_i)\right]. \end{aligned}$$

Märgime, et praktikas enamasti valitakse mõõt α ja tuum ψ nii, et

$$g : \mathbb{R} \rightarrow \mathbb{R}, \quad g(x) = \int \psi(x; \theta)\alpha(d\theta),$$

oleks proportsionaalne mingi tuntud tihedusfunktsiooniga. Näiteks kui $\alpha = M \cdot \bar{\alpha}$,

kus M on konstant ja $\bar{\alpha}$ mingile normaaljaotusele vastav tõenäosusmõõt, ning $\psi_\theta(x) = \psi(x; \theta)$ on fikseeritud standardhälbega σ jaotuse $\mathcal{N}(\mu = \theta; \sigma)$ tihedusfunktsioon, siis ka $g(x)$ on proportsionaalne normaaljaotuse tihedusfunktsiooniga. Kokkuvõttes näeme, et fikseeritud parameetrite $\theta_1, \dots, \theta_n$ korral ei ole segujaotuse tiheduse hindamine kuigi keeruline. Huvipakkuv keskväärtus (3.2) aga avaldub kujul

$$\frac{1}{M+n} \left[\int \psi(x; \theta) \alpha(d\theta) + E \left(\sum_{i=1}^n \psi(x; \theta_i) \middle| X_{1:n} = x_{1:n} \right) \right]. \quad (3.3)$$

Paraku $\theta_{1:n} | X_{1:n} = x_{1:n}$ tinglikul jaotusel üldjuhul puudub lihtsa analüütilise kujuga tihedus, mistõttu osutub (3.3) teise liidetava täpne rehkendamine ka võrdlemisi väikeste valimimahtude keeruliseks [7, lk 103–104]. Seepärast soovime tuletada viisi hindamiseks keskväärtust (3.2) ligikaudselt. Idee on genereerida parameetrite valimeid $\theta_{1:n}^{(m)} = (\theta_1^{(m)}, \dots, \theta_n^{(m)})$, $m = 1, \dots, q$, vastavalt $\theta_{1:n} | X_{1:n} = x_{1:n}$ järeljaotusele ja seejärel keskmistada nende järgi:

$$\hat{E} \left(\sum_{i=1}^n \psi(x; \theta_i) \middle| X_{1:n} = x_{1:n} \right) = \frac{1}{q} \sum_{m=1}^q \left(\sum_{i=1}^n \psi(x; \theta_i^{(m)}) \right)$$

Valimite genereerimine vahetult $\theta_{1:n} | X_{1:n} = x_{1:n}$ tinglikust jaotusest on üldiselt keeruline – taas põhjusel, et vastaval tihedusel enamasti puudub analüütiline kuju. See-eest osutub parameetrite valimite genereerimine Gibbsi valiku meetodil märksa lihtsamaks.

Järgnevalt rakendame juba esimesest peatükist tuttavat kirjaviisi. Eri jaotustele vastavate tiheduste eristamiseks kasutame üksnes väikseid tähti p , θ_i ja x_i funktsiooni $\psi(\cdot)$ argumentide kohal. Näiteks

$$P, \theta_1, \dots, \theta_n, X_1, \dots, X_n$$

ühistihedust korrutismõõdu $\Pi_\alpha \times \text{Leb}^n \times \text{Leb}^n$ suhtes tähistame kui

$$\psi(p, \theta_1, \dots, \theta_n, x_1, \dots, x_n) = \psi(p) \cdot \psi(\theta_{1:n}|p) \cdot \psi(x_{1:n}|\theta_{1:n}).$$

Iga $t = 1, \dots, n$ korral tähistagu θ_{-t} juhuslikku vektorit

$$(\theta_1, \dots, \theta_{t-1}, \theta_{t+1}, \dots, \theta_n)$$

ning θ_{-t} selle vektori realisatsiooni. Sümbolite X_{-t} ja x_{-t} tähendus olgu jätkuvalt analoogne. Lause 3.2 annab meile jaotuse, millest Gibbsi valiku algoritmis juhuslikke vaatlusi genereerida. Selle tõestus põhineb [7, Theorem 5.3] esitatud skeemil.

Teoreem 3.2. *Vastaku $P, \theta_1, \dots, \theta_n, X_1, \dots, X_n$ mudelile 3.1. Siis*

$$\theta_t | \theta_{-t} = \theta_{-t}, X_{1:n} = x_{1:n} \sim \frac{q_t G_t + \sum_{i \neq t} \psi(x_t; \theta_i) \delta_{\theta_i}}{q_t + \sum_{i \neq t} \psi(x_t; \theta_i)}, \quad (3.4)$$

kus $q_t = \int \psi(x_t; \theta) \alpha(d\theta)$ ja G_t on seosega

$$G_t(A) = \frac{1}{q_t} \int_A \psi(x_t; \theta) \alpha(d\theta), \quad A \in \mathcal{B}(\Theta) \quad (3.5)$$

määratud tõenäosumõõd.

Tõestus. Meie eesmärk on näidata, et suvalise hulga $A \in \mathcal{B}(\Theta)$ ja vektorite $\theta_{-t}, x_{1:n}$ korral

$$\mathbf{P}(\theta_t \in A | \theta_{-t} = \theta_{-t}, X_{1:n} = x_{1:n}) = \frac{q_t G_t(A) + \sum_{i \neq t} \psi(x_t; \theta_i) \delta_{\theta_i}(A)}{q_t + \sum_{i \neq t} \psi(x_t; \theta_i)}. \quad (3.6)$$

Etteruttavalt märgime, et tinglik tõenäosus (3.6) avaldub järgmise integraalina:

$$\begin{aligned} & \mathbf{P}(\theta_t \in A | \theta_{-t} = \theta_{-t}, X_{1:n} = x_{1:n}) = \\ & = \int \int I_A(x, \theta) \psi(\theta_t | \theta_{-t}, x_{1:n}) \text{Leb}(dx) \left(\alpha + \sum_{i \neq t} \delta_{\theta_i} \right) (d\theta). \end{aligned} \quad (3.7)$$

Bayesi teoreemi tõttu omakorda kehtib võrdus

$$\psi(\theta_t | \theta_{-t}, x_{1:n}) = \frac{\psi(\theta_t, x_t | \theta_{-t}, x_{-t})}{\int \psi(\theta_t, x_t | \theta_{-t}, x_{-t}) \left(\alpha + \sum_{i \neq t} \delta_{\theta_i} \right) (d\theta)}. \quad (3.8)$$

Märkus: Praegusel etapil ei pruugi olla ilmne, miks (3.7) ja (3.8) puhul integreeritakse moodsu $\alpha + \sum_{i \neq t} \delta_{\theta_i}$ järgi. Seda asjaolu põhjendame allpool. Esitasime eeltoodud võrdused juba siin tõestuse loetavuse huvides. Nimelt näitlikustavad need, kuidas seostub tihedus $\psi(\theta_t, x_t | \theta_{-t}, x_{-t})$ väitega, mida püüame tõestada.

Veendume, et kehtib võrdus

$$\psi(\theta_t, x_t | \theta_{-t}, x_{-t}) = \frac{\psi(x_t; \theta_t)}{M + n - 1}.$$

Selleks fikseerime vabalt hulga $B \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\Theta)$ ja vaatleme tinglikku tõenäosust

$$\mathbf{P}((X_t, \theta_t) \in B | \theta_{-t} = \theta_{-t}, X_{-t} = x_{-t}).$$

Kõigepealt paneme tähele, et tingliku keskväärtuse omaduste tõttu

$$\begin{aligned} \mathbf{P}((X_t, \theta_t) \in B | \theta_{-t} = \theta_{-t}, X_{-t} = x_{-t}) &= \\ &= E(I_B(X_t, \theta_t) | \theta_{-t} = \theta_{-t}, X_{-t} = x_{-t}) = \\ &= E\left(E(I_B(X_t, \theta_t) | P, \theta_{-t}, X_{-t}) | \theta_{-t} = \theta_{-t}, X_{-t} = x_{-t}\right). \end{aligned} \quad (3.9)$$

Kuna fikseeritud P realisatsiooni p korral on vektorid (X_i, θ_i) omavahel sõltumatud, siis

$$E(I_B(X_t, \theta_t) | P = p, \theta_{-t}, X_{-t}) = E(I_B(X_t, \theta_t) | P = p) \text{ p.k.}$$

ja (3.9) tõttu

$$\begin{aligned} & \mathbf{P}((X_t, \boldsymbol{\theta}_t) \in B | \boldsymbol{\theta}_{-t} = \theta_{-t}, X_{-t} = x_{-t}) = \\ & = E\left(E(I_B(X_t, \boldsymbol{\theta}_t) | P) | \boldsymbol{\theta}_{-t} = \theta_{-t}, X_{-t} = x_{-t}\right). \end{aligned}$$

Juhusliku vektori $\boldsymbol{\theta}_{-t}$ väärtuse θ_{-t} fikseerimisel on aga P ja X_{-t} omakorda tinglikult sõltumatud. Niisiis,

$$E\left(E(I_B(X_t, \boldsymbol{\theta}_t) | P) | \boldsymbol{\theta}_{-t} = \theta_{-t}, X_{-t} = x_{-t}\right) = E\left(E(I_B(X_t, \boldsymbol{\theta}_t) | P) | \boldsymbol{\theta}_{-t} = \theta_{-t}\right).$$

Kui $\boldsymbol{\theta}_{-t} = \theta_{-t}$, siis ühtlasi on lause 2.7 põhjal juhusliku mõõdu P järeljaotus $\Pi(\alpha + \sum_{i \neq t} \delta_{\theta_i})$ ja kehtib

$$E\left(E(I_B(X_t, \boldsymbol{\theta}_t) | P) | \boldsymbol{\theta}_{-t} = \theta_{-t}\right) = \int E(I_B(X_t, \boldsymbol{\theta}_t) | P = p) \Pi\left(\alpha + \sum_{i \neq t} \delta_{\theta_i}\right)(dp). \quad (3.10)$$

Integraali (3.10) leidmiseks defineerime kujutused

$$g : \Theta \rightarrow \mathbb{R}, \quad g(\theta) = \int I_B(x, \theta) \psi(x; \theta) \text{Leb}(dx),$$

$$T_g : \mathcal{P} \rightarrow \mathbb{R}, \quad T_g(p) = \int g(\theta) p(d\theta).$$

Kuna B on mõõtuv hulk ning ψ on tuum (s.t mõõtuv ja mittenegatiivne), siis ka korrutis $I_B \psi$ on $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\Theta)$ -mõõtuv ja mittenegatiivne. Sellest järeldub, et kujutus g on $\mathcal{B}(\Theta)$ -mõõtuv [3, Theorem 18.3]. Et $\psi(x; \theta)$ on definitsiooni 3.1 kohaselt $X_t | \boldsymbol{\theta}_t = \theta$ tinglik tihedus Lebesgue'i mõõdu suhtes, siis on selge, et $g(\theta)$ on tingliku keskvaertuse $E(I_B(X_t, \boldsymbol{\theta}_t) | \boldsymbol{\theta}_t = \theta)$ variant.

Kuivõrd $\boldsymbol{\theta}_t$ väärtuse fikseerimisel $(X_t, \boldsymbol{\theta}_t)$ ei sõltu enam juhuslikust mõõdust P , siis formaalselt

$$E(I_B(X_t, \boldsymbol{\theta}_t) | \boldsymbol{\theta}_t = \theta_t) = E(I_B(X_t, \boldsymbol{\theta}_t) | \boldsymbol{\theta}_t = \theta_t, P = p) \text{ p.k.}$$

Viimast fakti silmas pidades, paneme tähele, et $T_g(p)$ on omakorda

$$E(I_B(X_t, \boldsymbol{\theta}_t) | P = p)$$

variant. Tõepoolest, suvalise $C \in \mathcal{B}(\mathcal{P})$ korral

$$\begin{aligned} \int_C T_g(p) \Pi_\alpha(dp) &= \int I_C(p) \left(\int g(\theta) p(d\theta) \right) \Pi_\alpha(dp) = \\ &= \int I_C(p) \left(\int E(I_B(X_t, \boldsymbol{\theta}_t) | \boldsymbol{\theta}_t = \theta) p(d\theta) \right) \Pi_\alpha(dp) = \\ &= \int I_C(p) \left(\int E(I_B(X_t, \boldsymbol{\theta}_t) | \boldsymbol{\theta}_t = \theta, P = p) p(d\theta) \right) \Pi_\alpha(dp) = \\ &= \int I_C(p) E(I_B(X_t, \boldsymbol{\theta}_t) | P = p) \Pi_\alpha(dp) = \\ &= E(I_C(P) \cdot E(I_B(X_t, \boldsymbol{\theta}_t) | P)) = \\ &= E(E(I_C(P) \cdot I_B(X_t, \boldsymbol{\theta}_t) | P)) = \\ &= E(I_C(P) \cdot I_B(X_t, \boldsymbol{\theta}_t)). \end{aligned}$$

Niisiis, (3.10) tõttu

$$\begin{aligned} \mathbf{P}((X_t, \boldsymbol{\theta}_t) \in B | \theta_{-t}, x_{-t}) &= \int E(I_B(X_t, \boldsymbol{\theta}_t) | P = p) \Pi\left(\alpha + \sum_{i \neq t} \delta_{\theta_i}\right)(dp) = \\ &= \int T_g(p) \Pi\left(\alpha + \sum_{i \neq t} \delta_{\theta_i}\right)(dp). \end{aligned}$$

Rakendades nüüd lauset 2.12, saame

$$\begin{aligned} \int T_g(p) \Pi\left(\alpha + \sum_{i \neq t} \delta_{\theta_i}\right)(dp) &= \int g(\theta) \left(\frac{\alpha + \sum_{i \neq t} \delta_{\theta_i}}{M + n - 1} \right) (d\theta) = \\ &= \frac{1}{M + n - 1} \int \int I_B(x, \theta) \psi(x; \theta) \text{Leb}(dx) \left(\alpha + \sum_{i \neq t} \delta_{\theta_i} \right) (d\theta). \end{aligned}$$

Sellega oleme näidanud, et

$$\psi(\theta_t, x_t | \theta_{-t}, x_{-t}) = \frac{\psi(x_t; \theta_t)}{M + n - 1}$$

on $X_t, \boldsymbol{\theta}_t | X_{-t}, \boldsymbol{\theta}_{-t}$ tihedus korrutismõõdu $\text{Leb} \times (\alpha + \sum_{i \neq t} \delta_{\theta_i})$ suhtes. Järelikult

$$\psi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{-t}, x_{1:n}) = \frac{\psi(x_t; \boldsymbol{\theta}_t)}{\int_{\Theta} \psi(x_t; \theta) (\alpha + \sum_{i \neq t} \delta_{\theta_i}) (d\theta)}. \quad (3.11)$$

Kui $A \in \mathcal{B}(\Theta)$, siis ei ole raske näha, et

$$\begin{aligned} \mathbf{P}(\boldsymbol{\theta}_t \in A | \boldsymbol{\theta}_{-t} = \theta_{-t}, X_{1:n} = x_{1:n}) &= \int_A \psi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{-t}, x_{1:n}) (\alpha + \sum_{i \neq t} \delta_{\theta_i}) (d\theta) = \\ &= \frac{\int_A \psi(x_t; \theta) (\alpha + \sum_{i \neq t} \delta_{\theta_i}) (d\theta)}{\int_{\Theta} \psi(x_t; \theta) (\alpha + \sum_{i \neq t} \delta_{\theta_i}) (d\theta)} = \\ &= \frac{\int_A \psi(x_t; \theta) \alpha (d\theta) + \sum_{i \neq t} \psi(x_t; \theta_i) \delta_{\theta_i}(A)}{\int_{\Theta} \psi(x_t; \theta) \alpha (d\theta) + \sum_{i \neq t} \psi(x_t; \theta_i)} = \\ &= \frac{q_t G(A) + \sum_{i \neq t} \psi(x_t; \theta_i) \delta_{\theta_i}(A)}{q_t + \sum_{i \neq t} \psi(x_t; \theta_i)}. \end{aligned}$$

□

Märkus 3.2.1. Parameetrite vektoris $\boldsymbol{\theta}_{-t}$ on kokku $k \leq n - 1$ unikaaslet koordinaatide väärtust ϕ_1, \dots, ϕ_k . Iga $j = 1, \dots, k$ korral tähistagu $n_{-t}(j)$, mitu korda väärtus μ_j esineb vektoris $\boldsymbol{\theta}_{-t}$. Siis jaotus (3.4) esitub ka kujul

$$\boldsymbol{\theta}_t | \boldsymbol{\theta}_{-t}, x_{1:n} \sim \begin{cases} G_t & \text{tõenäosusega } \frac{q_t}{q_t + \sum_{i \neq t} \psi(x_t; \theta_i)}, \\ \delta_{\phi_j} & \text{tõenäosusega } \frac{n_{-t}(j) \cdot \psi(x_t; \phi_j)}{q_t + \sum_{i \neq t} \psi(x_t; \theta_i)}. \end{cases} \quad (3.12)$$

Teoreemis 3.2 ei tehta mingeid lisaeldsui mõõdu α kohta, s.t tegemist võib olla mistahes lõpliku mõõduga ruumil $(\Theta, \mathcal{B}(\Theta))$. Niisugusel mõõdul ei pruugi leiduda tihedust Lebesgue'i mõõdu suhtes. Viimasel juhul puudub ühene eeskiri juhuslike arvude genereerimiseks jaotusest G_t .

Praktikas valitakse mõõt α nii, et sel leiduks tihedus Lebesgue'i mõõdu suhtes h_α .

Siis suvalise $A \in \mathcal{B}(\Theta)$ korral

$$G_t(A) = \frac{1}{q_t} \int_A \psi(x_t; \theta) \alpha (d\theta) = \frac{1}{q_t} \int_A \psi(x_t; \theta) \cdot h_\alpha(\theta) d\theta$$

ehk

$$g : \Theta \rightarrow \mathbb{R}, \quad g(\theta) = \psi(x_t; \theta) \cdot h_\alpha(\theta)$$

on mõõdule G_t vastav tihedus (Lebesgue'i mõõdu suhtes).

Toetudes teoreemile 3.2, saame sõnastada järgneva algoritmi soovitud valimite $\theta_{1:n}^{(m)}$, $m = 1, \dots, q$, genereerimiseks Gibbsi valiku meetodil. Algoritm on kohandatud raamatust [7, Theorem 5.3].

Algoritm 3 Parameetrite $\theta_1, \dots, \theta_n$ väärtuste genereerimine Gibbsi valiku meetodil

Sisend: andmed $x_{1:n}$, parameetrite algväärtused $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$, soovitud valimite arv m

- 1: **iga** $j = 1, \dots, m$ korral
- 2: **iga** $t = 1, \dots, n$ korral
- 3: fikseeritud $\theta_{-t} = (\theta_1^{(j)}, \dots, \theta_{t-1}^{(j)}, \theta_{t+1}^{(j-1)}, \theta_n^{(j-1)})$ ja vaatluste $x_{1:n}$ korral genereeri $\theta_t^{(j)}$ vastavalt jaotusele (3.12)

Väljund: parameetrite vektorid $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_n^{(j)})$, $j = 1, \dots, m$

Olgu $J \subset \{1, \dots, n\}$. Vaatluste $\{x_i\}_{i \in J}$ kohta võib öelda, et need kuuluvad samasse klastrisse, kui mingi $\theta \in \Theta$ korral tegemist on realisatsioonidega samast tihedusest $\psi_\theta(x) = \psi(x; \theta)$. Esitusest (3.12) on selge, et algoritm 3 määrab eri vaatlused suurematesse klastritesse, s.t sammul $j \in \{1, \dots, m\}$ parameetri θ_t uueks väärtuseks seatakse teatud tõenäosusega mingi θ_{-t} koordinaat.

Algoritm 3 tekitab uusi parameetrite väärtusi vaid siis, kui θ_t määratakse täiesti uude klastrisse. Kontsentratsiooniparameetri M madalate väärtuste korral on selle tõenäosus väike, mistõttu on ka valimite $\theta_{1:n}^{(j)}$, $j = 1, \dots, m$, empiirilise jaotuse koondumine soovitud $\theta_1, \dots, \theta_n | X_{1:n} = x_{1:n}$ jaotuseks aeglane [12]. Probleemi illustreerimiseks toome järgmise näite.

Näide 3.1. Olgu meil vaatlused

$x_{1:10}$: 144.77; 158.80; 156.51; 147.74; 171.58; 177.51; 194.59; 205.75,

mis pärinevad jaotusest

$$0.5 \cdot \mathcal{N}(\mu_1 = 160, \sigma = 15) + 0.5 \cdot \mathcal{N}(\mu_2 = 180, \sigma = 15).$$

Eeldame, et tegelik $\sigma = 15$ on teada ja võtame genereeritavate parameetrite rolli komponentide keskväärtused ehk $\theta = \mu$. Mudelis kasutatav tuum vastab normaaljaotuse tihedusele

$$\psi(x; \mu) = \frac{1}{15 \cdot \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{225}\right)$$

ja P eeljaotuseks valime Dirichlet' protsessi baasimõõduga $\alpha = M\bar{\alpha}$, kus $M = 1$ ja $\bar{\alpha}$ on jaotusele $\mathcal{N}(\mu_0 = 170; \sigma_0 = 30)$ vastav tõenäosusmõõt. Parameetrite vektori alglähendiks võtame vektori

$\mu_{1:8}^{(0)}$: 172.66; 172.66; 172.66; 172.66; 172.66; 172.66; 172.66; 172.66.

Genereerime algoritmi 3 abil viis parameetrite valimit:

$\mu_{1:8}^{(1)}$: 172.66; 172.66; 172.66; 172.66; 172.66; 172.66; 172.66; 172.66;

$\mu_{1:8}^{(2)}$: 172.66; 172.66; 172.66; 170.66; 172.66; 172.66; 201.06; 189.45;

$\mu_{1:8}^{(3)}$: 172.66; 172.66; 172.66; 172.66; 189.45; 172.66; 172.66; 189.45;

$\mu_{1:8}^{(4)}$: 172.66; 172.66; 172.66; 136.39; 172.66; 172.66; 189.45; 172.66;

$\mu_{1:8}^{(5)}$: 136.39; 172.66; 172.66; 172.66; 172.66; 189.45; 189.45; 189.45.

Näitest 3.1 näeme, et uued parameetrite väärtused tekivad harva (viie iteratsiooni käigus genereeriti vaid kolm uut unikaalset väärtust – need on tähistatud kastidega). Enamik uuendusi taandub sisuliselt vaatluse ümberpaigutamisele ühest ole-

masolevast klastrist teise.

Gibbsi valiku algoritm muutub efektiivsemaks, kui simuleerime parameetrite unikaalseid väärtusi ja klastrikuuluvust eraldi. Nii on võimalik igal iteratsioonil uuendada korraga kõiki sama klatri parameetrite väärtusi, mis tagab ka valimite empiirilise jaotuse kiirema koondumise soovitud järeljaotuseks. Algoritmi tuletuskäigu hõlbustamiseks võtame kasutusele mõningad uued matemaatilised objektid.

Etteantud parameetrite $\theta_1, \dots, \theta_n$ korral võime rääkida nende seas esinevatest unikaalsetest väärtustest ϕ_1, \dots, ϕ_k ja viitadest $y_1, \dots, y_n \in \{1, \dots, k\}$, mille puhul kehtib seos

$$(\theta_1, \dots, \theta_n) = (\phi_{y_1}, \dots, \phi_{y_n}).$$

Märgime, et mudelis (3.1) on klastrate arv $K : \Omega \rightarrow \{1, \dots, n\}$ juhuslik. Gibbsi valiku meetodil juhuslikke viitasid või parameetrite väärtuseid genereerides aga tinglikustame alati teatud fikseeritud klasterduse järgi, s.t algoritmi igal sammul on klastrate arv teada (ehkki see võib uue valimi genereerimisel ka kasvada või kahaneda). Seetõttu võime algoritmi implementeerimiseks tarvilike tiheduste ja tõenäosusfunktsioonide tuletamisel läbivalt eeldada, et $K = k$ on fikseeritud.

Etteantud $K = k$ korral võime vaadelda parameetrite unikaalseid väärtuseid juhuslikena, kusjuures Polya urni esitusest on ilmne, et tegemist on juhusliku valimiga baasijaotusest $\bar{\alpha}$ (Dirichlet' protsessit pärinevate vaatluste uued väärtused võetakse teistest sõltumatult jaotusest $\bar{\alpha}$):

$$\phi_i : \Omega \rightarrow \Theta, \quad \phi_i \stackrel{i.i.d}{\sim} \bar{\alpha}, \quad i = 1, \dots, k.$$

Juhuslikena võime vaadelda ka viitasid

$$Y_1, \dots, Y_n : \Omega \rightarrow \mathbb{N}, \quad Y_i = j \leftrightarrow X_i \sim \psi(\cdot; \phi_j), \quad j \in \mathbb{N}.$$

Järgnevalt tuletame viisi μ_1, \dots, μ_k ja Y_1, \dots, Y_n iteratiivseks genereerimiseks Gibbsi valiku meetodil. Algoritmi esimeses faasis uuendame viitasid y_1, \dots, y_n . Selleks

peame teadma $Y_t|Y_{-t} = y_{-t}, X_{1:n} = x_{1:n}, \boldsymbol{\mu}_{1:k} = \mu_{1:k}$ jaotust, mille avaldame lauses 3.3. Kirjapaneku hõlbustamiseks eeldame vaikumisi, et vektoris y_{-t} on esindatud kõik klastrid $1, \dots, k$. Viimast asjaolu on oluline pidada silmas algoritmi implementeerimisel. Näiteks, kui meil on vektor

$$y_{1:5} = (1, 1, 3, \boxed{2}, 4)$$

ja soovime uuendada koordinaati y_4 , siis tinglikustamisel kasutatav viitade vektor peab olema kujul

$$y_{-4} = (1, 1, \boxed{2}, \cancel{3}, \boxed{3}).$$

Analoogiliselt visatakse parameetrite unikaalsete väärtuste vektori $\phi_{1:k}$ hetkeseisust välja koordinaat ϕ_2 ja määratakse $\phi_2^{uus} := \phi_3$ ning $\phi_3^{uus} := \phi_4$, s.t tinglikustamisel kasutatav parameetrite vektor on

$$\phi_{-2} = (\phi_1, \cancel{\phi_2}, \phi_2^{uus}, \phi_3^{uus}).$$

Lause 3.3 tõestus põhineb artiklis [12] esitatud tõestuse skeemil.

Lause 3.3. *Olgu k vektori y_{-t} koordinaatide suurim väärtus. Iga $j \in \mathbb{N}$ korral tähistagu $n_{-t}(j)$, mitu korda väärtus j esineb vektoris y_{-t} . Siis*

$$\mathbf{P}(Y_t = j | y_{-t}, \phi_{1:k}, x_{1:n}) = \begin{cases} \frac{n_{-t}(j)\psi(x_t; \phi_j)}{q_t + \sum_{l=1}^k n_{-t}(l)\psi(x_t; \phi_l)}, & \text{kui } j \text{ on } y_{-t} \text{ koordinaat,} \\ \frac{q_t}{q_t + \sum_{l=1}^k n_{-t}(l)\psi(x_t; \phi_l)}, & \text{kui } j = k + 1. \end{cases} \quad (3.13)$$

Tõestus. Kõigepealt vaatleme juhtu, kus j on vektoris y_{-t} esinev väärtus. Paneme tähele, et $Y_t = j$ parajasti siis, kui $\boldsymbol{\theta}_t = \phi_j$. Kuna α on mitteatomaarne mõõt, siis

$$G_t(\{\phi_j\}) = \frac{1}{q_t} \int_{\{\mu_j\}} \psi(x_t; \theta) \alpha(d\theta) = 0$$

ja teoreemi 3.2 põhjal

$$\begin{aligned}
\mathbf{P}(Y_t = j | y_{-t}, \phi_{1:k}, x_{1:n}) &= \mathbf{P}(\boldsymbol{\theta}_t = \phi_j | \theta_{-t}, x_{1:n}) = \\
&= \frac{q_t G_t + \sum_{i \neq t} \psi(x_t; \theta_i) \delta_{\theta_i}(\{\phi_j\})}{q_t + \sum_{i \neq t} \psi(x_t; \theta_i)} = \\
&= \frac{\sum_{i: \theta_i = \phi_j, i \neq t} \psi(x_t; \phi_j)}{q_t + \sum_{i \neq t} \psi(x_t; \theta_i)} = \\
&= \frac{n_{-t}(j) \psi(x_t; \phi_j)}{q_t + \sum_{l=1}^k n_{-t}(l) \psi(x_t; \phi_l)}.
\end{aligned}$$

Nüüd vaatleme juhtu, kus Y_t omandab uue väärtuse. Selge, et see on samaväärne sündmusega, kus $\boldsymbol{\theta}_t$ väärtus ei ole ϕ_1, \dots, ϕ_k seas. Rakendades taas teoreemi 3.2, saame, et

$$\begin{aligned}
\mathbf{P}(Y_t = k + 1 | y_{-t}, \phi_{1:k}, x_{1:n}) &= \mathbf{P}(\boldsymbol{\theta}_t \in \Theta \setminus \{\phi_1, \dots, \phi_k\} | \theta_{-t}, x_{1:n}) = \\
&= \frac{q_t G_t + \sum_{i \neq t} \psi(x_t; \theta_i) \delta_{\theta_i}(\Theta \setminus \{\phi_1, \dots, \phi_k\})}{q_t + \sum_{i \neq t} \psi(x_t; \theta_i)} = \\
&= \frac{q_t G_t(\Theta \setminus \{\phi_1, \dots, \phi_k\})}{q_t + \sum_{i \neq t} \psi(x_t; \theta_i)} = \\
&= \frac{q_t}{q_t + \sum_{l=1}^k n_{-t}(l) \psi(x_t; \phi_l)}.
\end{aligned}$$

□

Algoritmi teises faasis genereeritakse kõik parameetrite unikaalsed väärtused. Rõhutame taas, et parameetrite väärtusi uuendatakse, tinglikustades fikseeritud viitade y_1, \dots, y_n järgi, seega ka klastrite arv k on teada. Selles kontekstis on tähenduslik rääkida juhuslikest suurustest ϕ_1, \dots, ϕ_k .

Et parameetrite unikaalsed väärtused genereeritakse üksteisest sõltumatult, peame teadma $\phi_j | Y_{1:n} = y_{1:n}, X_{1:n} = x_{1:n}$ jaotust iga $j = 1, \dots, k$ korral. Tuletame selle lauses 3.4, mille tõestus tugineb [7, lk 105-106] esitatud mõttekäigul.

Lause 3.4. *Etteantud viitade $y_{1:n}$ korral olgu $k = \max\{y_1, \dots, y_n\}$, $j \in \{1, \dots, k\}$ ja $J = \{i : y_i = j\}$. Kasutame sümbolit X_J tähistamaks vaatluste alamhulga ele-*

mentidest moodustatud vektorit $(X_i)_{i \in J}$. Kui $X_{1:n} = x_{1:n}$ ja $Y_{1:n} = y_{1:n}$, siis ϕ_j tinglik tihedus (mõõdu $\bar{\alpha}$ suhtes) on

$$\psi(\phi_j | y_{1:n}, x_{1:n}) = \frac{\prod_{i:y_i=j} \psi(x_i; \phi_j)}{\int \prod_{i:y_i=j} \psi(x_i; \phi_j) \bar{\alpha}(d\phi_j)}. \quad (3.14)$$

Tõestus. Kuna fikseeritud klastrite arvu k korral viidatakse $Y_{1:n}$ ei sõltu juhuslikust suurusel ϕ_j , siis $\psi(y_{1:n} | \phi_j) = \psi(y_{1:n})$. Ühtlasi märkame, et viitade $y_{1:n}$ fikseerimisel sõltub ϕ_j üksnes vaatlustest X_J ehk

$$\psi(\phi_j | y_{1:n}, x_{1:n}) = \psi(\phi_j | y_{1:n}, x_J)$$

ja

$$\psi(\phi_j | y_{1:n}, x_J) = \frac{\psi(x_J | y_{1:n}, \phi_j) \cdot \psi(y_{1:n} | \phi_j) \cdot \psi(\phi_j)}{\psi(x_J, y_{1:n})} = \frac{\psi(x_J | y_{1:n}, \phi_j) \cdot \psi(\phi_j)}{\psi(x_J | y_{1:n})}.$$

Teame, et parameetrite θ_i , $i \in J$, väärtuste fikseerimisel on X_i , $i \in J$, sõltumatud juhuslikud suurused tihedustega $\psi(\cdot; \theta_i)$. Kuivõrd ϕ_j ja $Y_{1:n}$ väärtuste fikseerimine on samaväärne θ_i , $i \in J$, fikseerimisega, kehtib

$$\psi(x_J | y_{1:n}, \phi_j) = \prod_{i:y_i=j} \psi(x_i; \phi_j).$$

Et ϕ_j marginaaljaotus on $\bar{\alpha}$, siis ϕ_j tihedus $\psi(\phi_j)$ mõõdu $\bar{\alpha}$ suhtes on lihtsalt indikaatorfunktsioon $I_{\Theta}(\phi_j)$, mille väärtus on alati 1. Kokkuvõttes

$$\psi(\phi_j | y_{1:n}, x_J) = \frac{\prod_{i:y_i=j} \psi(x_i; \phi_j)}{\int \psi(x_J, \phi_j | y_{1:n}) \bar{\alpha}(d\phi_j)} = \frac{\prod_{i:y_i=j} \psi(x_i; \phi_j)}{\int \prod_{i:y_i=j} \psi(x_i; \phi_j) \bar{\alpha}(d\phi_j)}.$$

□

Märkus 3.4.1. Lauses 3.4 toodud tihedus (3.14) on tihedus mõõdu $\bar{\alpha}$ suhtes. Nagu varemgi mainitud, valitakse mõõt α enamasti nii, et sel oleks tihedus Lebesgue'i mõõdu suhtes h_α . Mõõdule $\bar{\alpha}$ vastav tihedusfunktsioon on siis lihtsalt $\frac{1}{|\alpha|} h_\alpha$, kus

$|\alpha| = \alpha(\Theta)$ Tõepoolest, suvalise $B \in \mathcal{B}(\Theta)$ korral

$$\bar{\alpha}(B) = \frac{1}{|\alpha|} \alpha(B) = \frac{1}{|\alpha|} \int_B h_\alpha(\theta) d\theta.$$

Kui h_α eksisteerib, siis $\phi_j | Y_{1:n} = y_{1:n}, X_{1:n} = x_{1:n}$ tihedusfunktsioon on

$$\psi(\phi_j | y_{1:n}, x_{1:n}) = \frac{(\prod_{i:y_i=j} \psi(x_i; \phi_j)) \cdot h_\alpha(\phi_j)}{\int \prod_{i:y_i=j} \psi(x_i; \phi_j) \cdot h_\alpha(\phi_j) d\phi_j}. \quad (3.15)$$

Toetudes nüüd lausetele 3.3 ja 3.4, saame formuleerida algoritmi 4 parameetrite unikaalsete väärtuste μ_1, \dots, μ_k ja viitade y_1, \dots, y_n kordamööda genereerimiseks Gibbsi valiku meetodil. Algoritm 4 pärineb raamatust [7, Algorithm 2].

Algoritm 4 Parameetrite unikaalsete väärtuste μ_1, \dots, μ_k ja viitade y_1, \dots, y_n genereerimine Gibbsi valiku meetodil

Sisend: andmed $x_{1:n}$, parameetrite algväärtused $\mu^{(0)} = (\mu_1^{(0)}, \dots, \mu_k^{(0)})$, parameetrite unikaalsete väärtuste arv k , viitade algväärtused $y^{(0)} = (y_1^{(0)}, \dots, y_n^{(0)})$, soovitud valimite arv m

- 1: **iga** $j = 1, \dots, m$ korral
- 2: **iga** $t = 1, \dots, n$ korral
- 3: fikseeritud $y_{-t} = (y_1^{(j)}, \dots, y_{t-1}^{(j)}, y_{t+1}^{(j-1)}, y_n^{(j-1)})$, $\mu^{(j-1)}$ korral genereeri $y_t^{(j)}$ vastavalt jaotusele (3.13)
- 4: **kui** $y_t^{(j)} = k + 1$, **siis**
- 5: genereeri uus parameetri väärtus $\mu_{k+1}^{(j-1)}$ vastavalt jaotusele (3.5)
- 6: uuenda $k = k + 1$
- 7: salvesta $y_{1:n}^{(j)} = (y_1^{(j)}, \dots, y_n^{(j)})$
- 8: **iga** $l=1, \dots, k$
- 9: fikseeritud $y_{1:n}^{(j)}$ korral genereeri $\mu_l^{(j)}$ vastavalt tihedusele (3.15)
- 10: salvesta $\mu^{(j)} = (\mu_1^{(j)}, \dots, \mu_k^{(j)})$

Väljund: parameetrite vektorid $\mu^{(j)}$, $j = 1, \dots, m$, ning viitade vektorid $y_{1:n}^{(j)}$, $j = 1, \dots, m$.

Võrdlemaks algoritmi 4 algoritmiga 3, genereerime viis parameetrite valimit, kasutades näites 3.1 toodud vaatlusi ja samasugust mudeli konfiguratsiooni.

Näide 3.2. Algoritmi aglähenditeks võtame

$$y_{1:8}^{(0)} = (1, 1, 1, 1, 1, 1, 1, 1), \quad \phi_1^{(0)} = 172.66.$$

Saame valimid

$$\begin{aligned} \mu_{1:8}^{(1)} &: 178.88; \boxed{184.07}; 178.88; 178.88; 178.88; 178.88; 178.88; 178.88; \\ \mu_{1:8}^{(2)} &: 176.86; 176.86; 176.86; 176.86; 176.86; 176.86; 176.86; \boxed{197.86}; \\ \mu_{1:8}^{(3)} &: 165.44; 165.44; 165.44; 165.44; 165.44; 165.44; 165.44; 221.05; \\ \mu_{1:8}^{(4)} &: 169.26; 169.26; 169.26; 169.26; 169.26; 169.26; 169.26; 169.26; \\ \mu_{1:8}^{(5)} &: 170.49; 170.49; 170.49; 170.49; 170.49; 170.49; 170.49; 170.49. \end{aligned}$$

Nagu näha, uusi klastreid tekitatakse vaid kahel juhul (vastavad parameetrite väärtused on tähistatud kastidega). Sellegipoolest genereerib algoritm uusi parameetrite väärtusi igal sammul – ka siis, kui kõik vaatlused jäävad seotuks sama klastriga.

4 Bayesi segumudelid Gaussi tuumaga

Esimeses peatükis tuletasime eeskirja segujaotuse tiheduse ψ_0 hindamiseks eeldusel, et parameetri θ tegelik jaotus p_0 on lõplik diskreetne tõenäosusmõõt, s.t

$$p_0 = \sum_{j=1}^k p_j^0 \delta_{\theta_j^0} \quad \text{ja} \quad \psi_0(x) = \sum_{j=1}^k p_j^0 \psi(x; \theta_j^0).$$

Kolmandas peatükis vaatlesime selle konfiguratsiooni üldistust, kus p_0 võib olla suvaline tõenäosusmõõt ruumil $(\Theta, \mathcal{B}(\Theta))$ ja tegelik tihedus esitub Lebesgue'i integraalina

$$\psi_0(x) = \int \psi(x; \theta) p_0(d\theta).$$

Mõlemal juhul eeldasime, et tuum $\psi(x; \theta)$ on teadaolev funktsioon. Praktikas ei ole "õige" tuuma valik kuigi suur probleem ja ψ määramisel tehakse vaid väga üldisi eeldusi tegeliku tiheduse ψ_0 kohta. Näiteks kui tahetakse hinnata tihedust, mille kandja on kogu reaaltelg \mathbb{R} , siis loomulik tuuma valik on normaaljaotuse tihedusfunktsioon parameetriga $\theta = (\mu, \sigma)$ [7, lk 115]

$$\psi(x; \theta) := \psi(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right). \quad (4.1)$$

Niisugune valik ei ole kuigi kitsendav, sest igat tihedusfunktsiooni võib lähendada kuitahes hästi (L_1 normi mõttes) normaaljaotuste seguga. Eeldusel, et hinnatav tihedusfunktsioon koondub oma otstes nulli, kehtib viimane väide ka ühtlase koondumise mõttes [13].

Siinse magistritöö ulatuses vaatame Gaussi tuuma lihtsamat varianti, kus standardhälve on fikseeritud väärtus $\sigma = 1$ ja parameeter on $\theta = \mu$:

$$\psi(x; \mu) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - \mu)^2\right). \quad (4.2)$$

Selles peatükis anname ülevaate eeljaotuse valikust parameetrilises ja mittepara-

meetrilises segumudelis Gaussi tuuma $\psi(x; \mu)$ korral, seejärel anname ülevaate hinnangu ψ_P mõjusust tagavatest tingimustest ja lõpuks võrdleme esimeses ning kolmandas peatükis kirjeldatud tiheduse hindamise meetodeid arvutisimulatsioonide abil.

4.1 Eeljaotuse valik

Siinses magistritöös käsitlesime parameetrilist Bayesi mudelit

$$\begin{aligned}
 (\mathbf{p}_1, \dots, \mathbf{p}_k) = \mathbf{p} &\sim \text{Dir}(\alpha_1, \dots, \alpha_k), \\
 \boldsymbol{\theta}_i &\stackrel{\text{i.i.d.}}{\sim} \pi, \quad i = 1, \dots, k \\
 Y_1, \dots, Y_n | \mathbf{p} &= p \stackrel{\text{i.i.d.}}{\sim} \sum_{i=1}^k p_i \delta_i, \\
 X_i | Y_i = y_i, \boldsymbol{\theta} &= \theta \stackrel{\text{i.i.d.}}{\sim} \psi(\cdot; \theta_{y_i}), \quad i = 1, \dots, n.
 \end{aligned} \tag{4.3}$$

ja mitteparameetrilist mudelit

$$\begin{aligned}
 P &\sim \Pi_\alpha, \\
 \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n | P &= p \stackrel{\text{i.i.d.}}{\sim} p, \\
 X_i | \boldsymbol{\theta}_i &= \theta_i \stackrel{\text{i.i.d.}}{\sim} \psi(\cdot; \theta_i), \quad i = 1, \dots, n.
 \end{aligned} \tag{4.4}$$

Tiheduse ψ_0 hindamiseks tuleb mudelis (4.3) valida parameetrid $\alpha_1, \dots, \alpha_k \in (0, \infty)$ ja tihedusfunktsioon π . Mitteparameetrilise mudeli (4.4) korral peame valida mitteatomaarse lõpilku moodsu α . Peatume põgusalt kriteeriumitel nende valikute langetamiseks.

Parameetrilise mudeli korral väljendavad parameetrid $\alpha_1, \dots, \alpha_k$ meie eelarusaama jaotuse komponentide kaalude p_1^0, \dots, p_k^0 kohta. Paneme tähele, et juhul, kui \mathbf{p} on

juhuslik vektor jaotusega $\text{Dir}(\alpha_1, \dots, \alpha_k)$, siis lause 1.5 põhjal

$$E(\mathbf{p}) = \left(\frac{\alpha_1}{|\alpha|}, \dots, \frac{\alpha_k}{|\alpha|} \right), \quad |\alpha| = \sum_{j=1}^k \alpha_j.$$

Kui $k = 3$ ja on põhjust arvata, et tegelikud komponentide kaalud võiksid olla näiteks

$$p_1 = \frac{1}{5}, p_2 = \frac{3}{4}, p_3 = \frac{1}{20},$$

tasub valida

$$\frac{\alpha_1}{|\alpha|} = \frac{1}{5}, \frac{\alpha_2}{|\alpha|} = \frac{3}{4}, \frac{\alpha_3}{|\alpha|} = \frac{1}{20}.$$

Suurus $|\alpha|$ väljendab omakorda meie usaldusastet oma eelarusaama suhtes. Tõepoolest, lause 1.6 põhjal iga $j = 1, \dots, k$ korral

$$D(\mathbf{p}_j) = \frac{1}{|\alpha| + 1} \cdot \frac{\alpha_j}{|\alpha|} \left(1 - \frac{\alpha_j}{|\alpha|} \right) \rightarrow 0, \quad \text{kui } |\alpha| \rightarrow \infty,$$

ehk suure $|\alpha|$ väärtuse korral kontsentreerub \mathbf{p} eeljaotuse tõenäosusmass postuleeritud eelarusaama ümbruses. Juhul, kui puuduvad põhjused mingite konkreetsete p_1, \dots, p_k väärtuste eelistamiseks, võetakse enamasti

$$\alpha_1 = \dots = \alpha_k = 1,$$

sest siis vastab $\text{Dir}(\alpha_1, \dots, \alpha_k)$ ühtlasele jaotusele $(k - 1)$ -ühiksimpleksil.

Tihedusfunktsiooni $\pi(\mu)$ valikul nõutakse, et tegelikud parameetrite väärtused μ_1, \dots, μ_k kuuluksid funktsiooni π kandjasse ja ψ_0 hindamisel esinevad integraalid (nt integraalid tiheduses (1.1)) oleksid kergesti rehkendatavad. Esimene omadus on täidetud siis, kui π kandja on kogu reaaltelg \mathbb{R} . Teise omaduse rahuldamiseks valitakse π niimoodi, et see vastaks $\psi(x; \mu)$ kaasjaotusele ehk iga $j = 1, \dots, k$ korral kuuluks $\mu_j | X_1 = x_1, \dots, X_n = x_n$ järeljaotus μ_j eeljaotusega samasse jaotuste perre. Siinses peatükis vaadeldava Gaussi tuuma $\psi(x; \mu)$ korral on mõlemad tingimused täidetud, kui π on mingi normaaljaotuse $\mathcal{N}(\mu_0, \sigma_0)$ tihedusfunktsioon [11].

Hüperparameetrite μ_0 ja σ_0 konkreetsete väärtuste valik on märksa keerulisem ülesanne, mille puhul puuduvad mingid universaalsed rusikareeglid. Enamasti valitakse need lähtuvalt kindlast andmestikust x_1, \dots, x_n . Sisuliselt peaks μ_0 paiknema nii ja σ_0 olema piisavalt suur, et π tõenäosusmass "katakts" võimalikult hästi tegelike parameetrite μ_1, \dots, μ_k tõenäolise paiknemise vahemikku [11].

Lõplikku mitteatomaarset baasimõõtu α võib esitada ka kujul

$$\alpha = M \cdot \bar{\alpha},$$

kus $\bar{\alpha}$ on tõenäosusmõõt ja $M = \alpha(\Theta)$. Loomulik on nõuda, et tõenäosusmõõdul $\bar{\alpha}$ leiduks tihedus Lebesgue'i mõõdu suhtes $h_{\bar{\alpha}}$ (vt nt märkust 3.4.1), mis rahuldaks tihedusfunktsioonile π seatud nõudmisi. Siin vaadeldava Gaussi tuuma korral tuleks seega valida $\bar{\alpha}$ rolli taas mingile normaaljaotusele $\mathcal{N}(\mu_0, \sigma_0)$ vastav tõenäosusmõõt [4].

Tuletame nüüd meelde 2. peatükis esitatud tulemusi. Lause 2.5 kohaselt

$$E(P(A)) = \bar{\alpha}(A), \quad \forall A \in \mathcal{B}(\Theta),$$

ehk tõenäosusmõõt $\bar{\alpha}$ väljendab meie aprioorseid teadmisi mõõdu p_0 kohta. Teisalt lause 2.6 väidab, et

$$D(P(A)) = \frac{\bar{\alpha}(A) \cdot (1 - \bar{\alpha}(A))}{M + 1}, \quad A \in \mathcal{B}(\Theta).$$

Mida suurem on kontsentratsiooniparameeter M , seda tugevamini kontsentreerub juhusliku mõõdu P eeljaotuse tõenäosusmass esmase hinnangu $\bar{\alpha}$ ümber. Võib öelda, et kontsentratsiooniparameeter M postuleerib meie usaldusastet sellise eeljaotuse valiku suhtes.

Siinkohal tekib küsimus, kuidas niisugust usalduse mõistet kvantifitseerida ja millest lähtuvalt valida konkreetset M väärtust mingi etteantud andmeanalüüsi ülesan-

de korral. Teatud mõttes peegeldab kontsentratsiooniparameeter M meie eelarusaama sellest, kui hästi jaotuvad andmed x_1, \dots, x_n selgetesse klastritesse. Tuletame meelde, et mitteparameetrilise mudeli (4.4) korral on klastrite (s.t parameetrite unikaalsete väärtuste) arv K juhuslik, kusjuures Charles E. Antoniak on näidanud [1], et

$$E(K) = \sum_{i=1}^n \frac{M}{M+i-1}.$$

Olgu n tiheduse ψ_0 hindamiseks kasutatava valimi maht. Kui $M = \frac{1}{n}$, siis $E(K) \approx 1$. Kui $M = n^2$, siis $E(K) \approx n$. Parameetri M madalad väärtused väljendavad seega arusaama, et olemasolevad andmed on hästi kirjeldatavad ka mingi parameetrilise mudeliga (4.3). Suur M jällegi väljendab tugevat usku, et tihedus ψ_0 ei ole kuigi hästi lähendatav lõpliku komponentide seguga [4].

4.2 Hinnangu mõjus

Üks loomulikke nõudmisi, mida seatakse tiheduse hinnangule on mõjus. Intuitiivselt tähendab mõjus seda, et vaatluste arvu n kasvades jõuab hinnang tegelikule tihedusele aina lähemale. Bayesi terminoloogiat kasutades võib öelda ka, et vaatluste arvu kasvades koondub hinnangu järeljaotuse tõenäosusmass tegeliku tiheduse ümber.

Siinkohal tuleb täpsustada, mida peetakse silmas, kui öeldakse, et tiheduse hinnang on tegeliku tiheduse "lähedal". Tiheduste vahelist kaugust võib defineerida paljudel eri viisidel. Siinse magistritöö raames võtame kasutusele L_1 -normi poolt indutseeritud meetrika

$$d : L_1 \times L_1 \rightarrow \mathbb{R}, \quad d(f_1, f_2) := \|f_1 - f_2\|_{L_1} = \int |f_1 - f_2| \text{Leb}(dx).$$

Märgime, et juhul, kui f_1 ja f_2 on tihedused (Lebesgue'i mõõdu suhtes), ei saa nende kaugus olla suurem kui 2. Tõepoolest,

$$d(f_1, f_2) = \|f_1 - f_2\|_{L_1} \leq \|f_1\|_{L_1} + \|f_2\|_{L_1} = \int |f_1| dx + \int |f_2| dx \leq 1 + 1 = 2.$$

Tuletame meelde, et siinse töö raames käsitletud Bayesi hinnangud tihedusele esituvad kujul

$$\psi_P(\cdot) = \int \psi(\cdot; \theta) P(d\theta),$$

kus ψ on tuum ja P juhuslik tõenäosusmõõt ruumil $(\Theta, \mathcal{B}(\Theta)) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Niisiis, definitsioon 4.1 sätestab formaalselt, mida edaspidi mõjususe all silmas peetakse.

Definitsioon 4.1. Olgu ψ_0 tegelik tihedus. Ütleme, et juhuslik hinnang ψ_P on mõjus, kui iga $\epsilon > 0$ korral

$$\mathbf{P}(d(\psi_P, \psi_0) < \epsilon | X_1, \dots, X_n) \xrightarrow{n} 1$$

peaaegu iga $(X_i)_{i=1}^\infty$ realisatsiooni korral.

Definitsioon 4.1 ei sea kitsendusi hinnangus ψ_P kasutatavale tuumale ψ , kuid meie vaatame üksnes kitsast juhtu, kus ψ on fikseeritud standardhälbega Gaussi tuum (4.2). Meile pakub huvi, milliseid tingimusi peab rahuldama tegelik tihedus ψ_0 ja juhuslik mõõt P , et hinnang ψ_P oleks mõjus. Vastamaks sellele küsimusele, esitame veel tiheduse Kullback-Leibleri omaduse mõiste.

Definitsioon 4.2. Öeldakse, et tihedusel f (Lebesgue'i mõõdu suhtes) on Kullback-Leibleri omadus, kui iga $\epsilon > 0$ korral

$$\mathbf{P}\left(P \in \left\{P : \int f(x) \log\left(\frac{f(x)}{\psi_P(x)}\right) dx < \epsilon\right\}\right) > 0.$$

Antonio Lioji *et al.* [9] on näidanud, et Kullback-Leibleri omadusega ψ_0 ja Gaussi tuuma korral on tiheduse hinnang ψ_P mõjus üsna tagasihoidlikel eeldustel. Nende

tõestatud teoreem käsitleb üldisemat tuuma (4.1), kus standardhälve σ pole fikseeritud ja seda vaadeldakse juhusliku parameetrina. Kuivõrd meie vaatleme kitsamat konfiguratsiooni (tuum $\psi(x; \mu)$), mil normaaljaotuse standardhälve on $\sigma = 1$, võtab tulemuse kokku järgmine teoreem.

Teoreem 4.3. *Olgu Π juhusliku moodsu P eeljaotus ja β seosega*

$$\beta(A) = E(P(A)), \quad A \in \mathcal{B}(\Theta),$$

määratud lõplik moods. Tiheduse hinnang ψ_P on mõjus, kui tegelikul tihedusel ψ_0 on Kullback-Leibleri omadus ja

$$\int |\mu| \beta(d\mu) < \infty.$$

Kullback-Leibleri omaduse olemasolu tegelikul tihedusel ψ_0 ei ole üldiselt triviaalne tingimus. Ghosal, Ghosh ja Ramamoorthi [6] on aga näidanud, et juhul, kui tegelik tihedus esitub Gaussi tuuma seguna (s.t rahuldab siin käsitletavate mudelite eeldusi), tagab Kullback-Leibleri omaduse järgnevas teoreemis sõnastatud kriteerium.

Teoreem 4.4. *Olgu $p_0 \in \mathcal{P}$ ja kehtigu*

$$\psi_0(\cdot) = \int \psi(\cdot; \mu) p_0(d\mu).$$

Kui tõenäosusmoodsu p_0 kandja on kompaktne ja p_0 ise kuulub juhusliku moodsu P eeljaotuse Π kandjasse, siis tihedusel ψ_0 on Kullback-Leibleri omadus.

Kui tegelik tihedus ψ_0 esitub lõpliku seguna, annab mudel (4.3) loomuliku eeljaotuse valiku korral (tihedus ψ vastab jaotusele $\mathcal{N}(\mu_0, \sigma_0)$) alati mõjusa hinnangu. Kui ψ_0 on hoopis üldine segu, tagab loomulik eeljaotuse valik – s.t P eeljaotus on $\Pi_{\bar{\alpha}}$, kus $\bar{\alpha}$ vastab mingile normaaljaotusele – mudeli (4.4) antava hinnangu mõjususe juhul, kui tegeliku segaja p_0 kandja on kompaktne. Veendume, et need faktid järelduvad teoreemidest 4.3 ja 4.4.

Järeldus 4.4.1. *Esitugu tegelik tihedus lõpliku normaaljaotuste seguna, s.t*

$$\psi_0(x) = \sum_{j=1}^k p_j \psi(x; \mu_j), \quad (p_1, \dots, p_k) \in \mathbb{S}^k. \quad (4.5)$$

Kui mudelis (4.3) tihedus π vastab jaotusele $\mathcal{N}(\mu_0, \sigma_0)$ ja ψ on fikseeritud standardhällbega Gaussi tuum $\psi(x; \mu)$, siis vastav tiheduse hinnang ψ_P on mõjus.

Tõestus. Märgime, et tihedust ψ_0 võib esitada ka Lebesgue'i integraalina

$$\psi_0(x) = \int \psi(x; \mu) p_0(d\mu), \quad p_0 = \sum_{j=1}^k p_j \delta_{\mu_j}.$$

Kuivõrd p_0 on diskreetne tõenäosusmõõt, siis selle kandja on lõplik (ja järelikult ka kompaktne) hulk $\{\mu_1, \dots, \mu_k\}$. Juhuslik mõõt P esitub vaadeldava mudeli korral kujul

$$P = \sum_{j=1}^k \mathbf{p}_j \delta_{\mu_j}.$$

Et $(\mathbf{p}_1, \dots, \mathbf{p}_k)$ eeljaotus on $\text{Dir}(\alpha_1, \dots, \alpha_k)$ ja $\mu_j, j = 1, \dots, k$, eeljaotus on $\mathcal{N}(\mu_0, \sigma_0)$, siis on selge, et P eeljaotuse kandja on

$$S = \left\{ \sum_{j=1}^k p_j \delta_{\mu_j} : (p_1, \dots, p_k) \in \mathbb{S}^k, (\mu_1, \dots, \mu_k) \in \mathbb{R}^k \right\}$$

ja $p_0 \in S$. Teoreemi 4.4 põhjal on tihedusel ψ_0 seega Kullback-Leibleri omadus.

Fikseerime nüüd $A \in \mathcal{B}(\Theta)$. Kuivõrd $(\mathbf{p}_1, \dots, \mathbf{p}_k)$ ja (μ_1, \dots, μ_k) on omavahel sõltumatud, siis

$$\begin{aligned} \beta(A) &= E(P(A)) = E\left(\sum_{j=1}^k \mathbf{p}_j \delta_{\mu_j}(A)\right) = \\ &= \sum_{j=1}^k E(\mathbf{p}_j \delta_{\mu_j}(A)) = \sum_{j=1}^k E(\mathbf{p}_j) \cdot E(\delta_{\mu_j}(A)). \end{aligned}$$

Teoreemi 1.5 põhjal $E(\mathbf{p}_j) = \frac{\alpha_j}{|\alpha|}$. Teisalt

$$\begin{aligned} E(\delta_{\mu_j}(A)) &= \int E(\delta_{\mu_j}(A) | \mu_j = \mu) \pi(\mu) d\mu = \\ &= \int \delta_{\mu}(A) \pi(\mu) d\mu = \int I_A(\mu) \pi(\mu) d\mu = \int_A \pi(\mu) d\mu. \end{aligned}$$

Niisiis,

$$\beta(A) = \sum_{j=1}^k \frac{\alpha_j}{|\alpha|} \int_A \pi(\mu) d\mu = \int_A \pi(\mu) d\mu$$

ehk β on jaotusele $\mathcal{N}(\mu_0, \sigma_0)$ vastav tõenäosusmõõt ja

$$\int |\mu| \beta(d\mu) < \infty.$$

Teoreemi 4.3 põhjal on hinnang ψ_P mõjus. □

Järeldus 4.4.2. *Esitugu tegelik tihedus normaaljaotuste üldise seguna, s.t*

$$\psi_0(x) = \int \psi(x; \mu) p_0(d\mu),$$

kus $p_0 \in \mathcal{P}$. Kui mudelis (4.4) $\bar{\alpha}$ on jaotusele $\mathcal{N}(\mu_0, \sigma_0)$ vastav tõenäosusmõõt, p_0 kandja on kompaktne ja ψ on fikseeritud standardhällbega Gaussi tuum $\psi(x; \mu)$, siis vastav tiheduse hinnang ψ_P on mõjus.

Tõestus. Et $\bar{\alpha}$ vastab normaaljaotusele, siis $\alpha = M\bar{\alpha}$ kandja on kogu reaaltelg ja järelikult ka Dirichlet' protsessi jaotuse Π_α kandja on kogu tõenäosusmõõtude ruum \mathcal{P} [7, Theorem 4.15]. Eelduse kohaselt on p_0 kandja kompaktne ja seega teoreemi 4.4 põhjal on tihedusel ψ_0 Kullback-Leibleri omadus.

Kuivõrd P on Dirichlet' protsess baasimõõduga α järeldub teoreemist 2.5, et suvalise $A \in \mathcal{B}(\Theta)$ korral

$$\beta(A) = E(P(A)) = \bar{\alpha}(A)$$

ehk tõenäosusmõõt β vastab jaotusele $\mathcal{N}(\mu_0, \sigma_0)$. See tähendab, et

$$\int |\mu| \beta(d\mu) < \infty.$$

Teoreemi 4.3 põhjal on hinnang ψ_P mõjus.

□

4.3 Simulatsioonid

Normaaljaotuste lõplik segu

Genereerime vaatlused kolme komponendiga normaaljaotuste segust

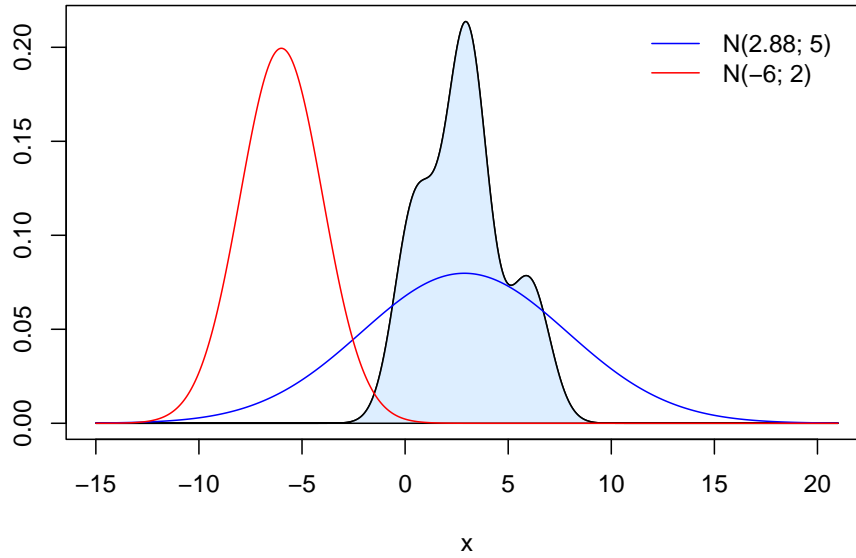
$$0,29 \cdot \mathcal{N}(0,5; 1) + 0,52 \cdot \mathcal{N}(3; 1) + 0,19 \cdot \mathcal{N}(6; 1). \quad (4.6)$$

Joonisel 1 on kujutatud andmete tegelikule jaotusele 4.6 vastav tihedusfunktsioon (see on värvitud helesiniseks). Parameetrilise Bayesi mudeli implementeerimisel peame valima parameetrite eeljaotusele vastava tihedusfunktsiooni π . Mitteparameetrilise mudeli korral tuleb valida kontsentratsiooniparameeter M ja tõenäosusmõõdule $\bar{\alpha}$ vastav tihedus $h_{\bar{\alpha}}$. Mõlema mudeli korral eksperimenteerime kahe π ja $h_{\bar{\alpha}}$ variandiga.

Kummalegi variandile vastavate tihedusfunktsioonide graafikud on samuti esitatud joonisel 4.6. Sinisega on tähistatud n-ö “hea” π või $h_{\bar{\alpha}}$ valik $\mathcal{N}(2.88; 5)$: keskväärtuseks on võetud 10000 vaatluse valimikeskmise ja standardhälve on piisavalt suur, et tihedusfunktsioon “haaraks endasse” võimalikult suure vahemiku tegelikust tihedusest genereeritavatest väärtustest. Punane joon vastab “halvale” tiheduse valikule $\mathcal{N}(-6; 2)$. Selle keskväärtus on kaugel valimikeskmisest ja väike standardhälve kontsentreerib tõenäosusmassi “halva” keskväärtuse ümber.

Kokku saame 4 erinevat Bayesi mudeli konfiguratsiooni, mis on toodud tabelis 1.

Märgime, et parameetrilise mudeli korral tuleb valida ka segu komponentide arv k



Joonis 1: Lõpliku normaaljaotuste segu tihedus ja tiheduste π või $h_{\bar{\alpha}}$ varian-
did

ja Dirichlet' jaotuse hüperparameetrid $\alpha_1, \dots, \alpha_k$. Simulatsioonides proovime väärtusi $k = 3$ ja $k = 10$. Mõlemal juhul valime mitteinformatiivse vektori \mathbf{p} eeljaotuse, võttes

$$\alpha_j = 1, \quad j = 1, \dots, k.$$

Mitteparameetrilise mudeli puhul tuleb veel valida täpsuskordaja M . Simulatsioonides proovime iga valimimahu n korral väärtusi

$$M \in \{1, \sqrt{n}, n\}.$$

Kõigi Bayesi mudelite puhul kasutame Gaussi tuuma (4.2), s.t $\psi(x; \mu_i)$ vastab jaotuse $\mathcal{N}(\mu_i; 1)$ tihedusfunktsioonile.

Lisaks sellele hindame tihedust (4.6) EM algoritmiga lõplike normaaljaotuste segu-

MP hea (M)	MP halb (M)
$P \sim \Pi(M \cdot \mathcal{N}(2,88; 5))$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n P = p \stackrel{i.i.d}{\sim} p$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$	$P \sim \Pi(M \cdot \mathcal{N}(-6; 2))$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n P = p \stackrel{i.i.d}{\sim} p$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$
P hea (k)	P halb (k)
$\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_k) \sim \text{Dir}(1, \dots, 1)$ $Y_1, \dots, Y_n \mathbf{p} = p \stackrel{i.i.d}{\sim} \sum_{i=1}^k p_i \delta_i$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k \stackrel{i.i.d}{\sim} \mathcal{N}(2,88; 5)$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$	$\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_k) \sim \text{Dir}(1, \dots, 1)$ $Y_1, \dots, Y_n \mathbf{p} = p \stackrel{i.i.d}{\sim} \sum_{i=1}^k p_i \delta_i$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k \stackrel{i.i.d}{\sim} \mathcal{N}(-6; 2)$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$

Tabel 1: Bayesi mudelite konfiguratsioonid

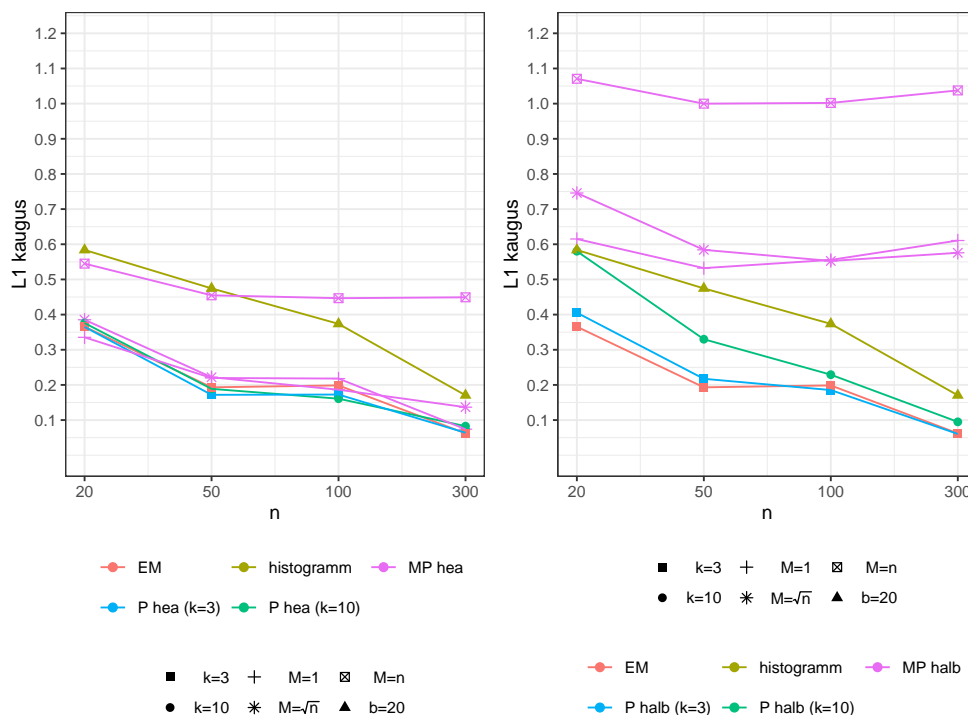
de jaoks (funktsioon `normalmixEM()` R-i paketist `mixtools`) vastavalt $k = 3$ ning $k = 10$ komponendi korral, eeldusel, et iga komponendi standardhälve on $\sigma = 1$. Lõpuks kasutame tiheduse hindamiseks ka tihedushistorgammi (funktsioon `hist()` R-i paketist `graphics`), mis jaotab lõigu $[-3, 9]$ kahekümneks ($b = 20$) võrdseks osaks.

Siin ja järgnevates simulatsioonides kasutame parameetrilise Bayesi mudeli implementeerimisel 1. peatükis kirjeldatud algoritmi 1. Mitteparameetrilise mudeli jaoks kasutame igal pool 3. peatükis kirjeldatud algoritmi 4.

Tiheduse hindamisel vaatleme valimimahtusid $n = 20, 50, 100, 300$. Iga valimimahtu korral arvutame välja hinnatud tiheduse L_1 -kauguse tegelikust tihedusest (4.7). Kauguse hinnangud on esitatud tabelis 4 ning joonisel 4.

Simulatsioonide tulemused võib võtta kokku järgmiselt:

- Bayesi mudelitest andsid parameetrilised mudelid parimaid hinnanguid kummagi tiheduse π valiku korral. Tulemus on ootuspärane, kuivõrd 3 komponendiga parameetiline mudel on õige mudel ja 10 komponendiga parameetiline mudel saab anda häid hinnanguid, seades 7 üleliigse komponendi kaalud vä-



Joonis 2: Tiheduse hinnangute L_1 -kaugus tegelikust tihedusest. Vasakpoolsel joonisel on esitatud Bayesi mudelite “headele” konfiguratsioonidele vastavate hinnangute kaugused tegelikust tihedusest, paremal pool on “halvad” konfiguratsioonid. EM algoritmile ja histogrammile vastavad jooned on mõlemal pool samad.

ga madalaks. Väiksematel valimimahtudel andis P hea ($k=3$) paremad hinnangud kui P halb ($k=3$), sama kehtis ka 10 komponendiga mudeli korral. Suurematel valimitel läksid erinevused väiksemaks – see on loogiline, sest valimimahu kasvades väheneb hinnangu sõltuvus eeljaotusest. Parameetriliste mudelite korral täheldame ka L_1 -kauguse vähenemist, mis on kooskõlas nende mõjususega.

- EM algoritm andis kõigil valimimahtudel 3-komponendilise Bayesi mudeliga väga sarnased hinnangud. Sisuliselt eeldavad mõlemad meetodid, et tegelik tihedus ψ_0 esitub kolme normaalkomponendi seguna.

- Üldiselt andsid mitteparameetrilised mudelid parameetristest mudelistest halvemaid hinnanguid. See on ootuspärane, kuivõrd MP mudelid püüavad hinnata tihedust üldisematel eeldustel, mistõttu on ka hinnangute koondumine tegelikuks tiheduseks aeglasem. Märgime, et $M = n$ korral oli isegi MP hea sooritus vilets. Nimelt postuleerib kõrge M väärtus, et tegelik tihedus on halvasti lähendatav lõpliku seguga. Antud juhul on aga tegelik tihedus parajsti selline lõplik segu, mistõttu ka suure M valik vale.

Mudel	$n = 20$	$n = 50$	$n = 100$	$n = 300$
MP hea ($M=1$)	0.335 (0.007)	0.22 (0.005)	0.218 (0.006)	0.074 (0.002)
MP hea ($M=\sqrt{n}$)	0.385 (0.008)	0.222 (0.004)	0.186 (0.004)	0.137 (0.003)
MP hea ($M=n$)	0.545 (0.01)	0.455 (0.008)	0.447 (0.008)	0.449 (0.008)
MP halb ($M=1$)	0.615 (0.014)	0.532 (0.014)	0.556 (0.015)	0.611 (0.017)
MP halb ($M=\sqrt{n}$)	0.746 (0.014)	0.584 (0.012)	0.553 (0.012)	0.576 (0.015)
MP halb ($M=n$)	1.071 (0.018)	1 (0.016)	1.002 (0.017)	1.038 (0.017)
P hea ($k=3$)	0.365 (0.008)	0.172 (0.005)	0.173 (0.005)	0.064 (0.002)
P hea ($k=10$)	0.376 (0.008)	0.189 (0.004)	0.161 (0.004)	0.083 (0.003)
P halb ($k=3$)	0.405 (0.012)	0.217 (0.006)	0.185 (0.005)	0.061 (0.002)
P halb ($k=10$)	0.58 (0.012)	0.33 (0.006)	0.229 (0.004)	0.095 (0.002)
EM ($k=3$)	0.366 (0.008)	0.193 (0.005)	0.199 (0.006)	0.062 (0.002)
histogramm ($b=20$)	0.584 (0.017)	0.475 (0.013)	0.374 (0.01)	0.17 (0.005)

Tabel 2: Tiheduse hinnangute L_1 -kaugus tegelikust tihedusest. Sulgudes toodud arvud tähistavad tõenäosusega 0,95 kehtivat viga

Simulatsioonide tulemuste illustreerimiseks vaatame ka, millised “MP hea”, “MP halb”, “P hea” ja “P halb” mudelid andsid parimad tiheduse hinnangud valimil mahuga $n = 300$ ja joonistame välja vastavad tihedusfunktsioonid. Võrdluse mõttes esitame ka EM algoritmi ($k = 10, n = 300$) ja histogrammi ($n = 300, b = 20$) abil leitud tiheduse hinnangute graafikud. Kõik need graafikud on koondatud lisas 1.

Üldine segu

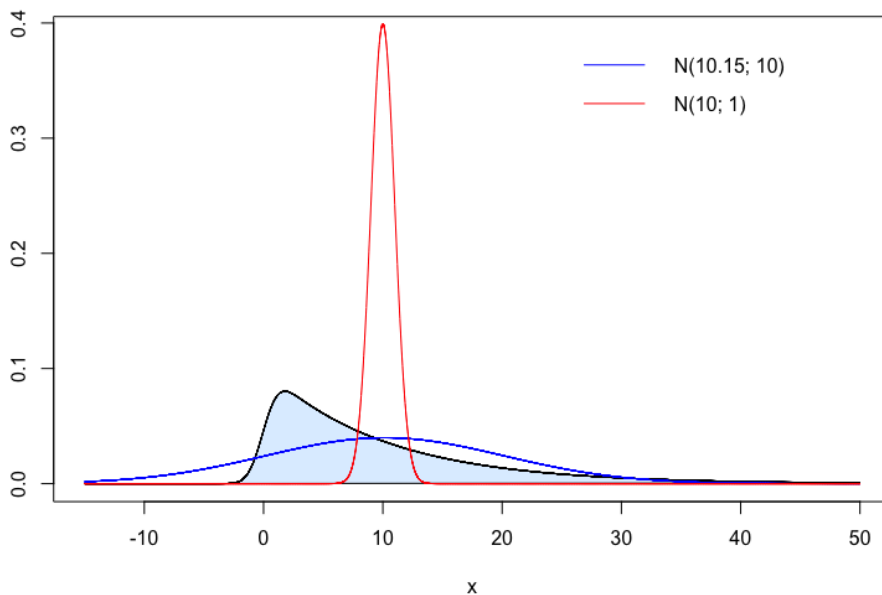
Genereerime vaatlused mudelist

$$\begin{aligned}\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n &\stackrel{i.i.d}{\sim} \text{Exp}(0.1), \\ X_i | \boldsymbol{\mu}_i = \mu_i &\stackrel{ind}{\sim} \mathcal{N}(\mu_i, 1), \quad i = 1, \dots, n.\end{aligned}$$

See tähendab, et tegelik tihedus on

$$\psi_p(x) = \int_0^\infty \psi(x; \mu) \cdot 0.1 \exp(-0.1\mu) d\mu = 0.1 \exp\left(-0.1x + \frac{0.1^2}{2}\right) \cdot \Phi(x-0.1), \quad (4.7)$$

kus Φ on standardsele normaaljaotusele vastav jaotusfunktsioon.



Joonis 3: Üldise segujaotuse tihedus ja tiheduste π või $h_{\bar{\alpha}}$ variandid

Joonisel 3 on kujutatud tihedusfunktsioon (4.7) ja kaks varianti tihedusfunktsioonile π või $h_{\bar{\alpha}}$, mida peame valima Bayesi mudeli implementeerimisel. Sinisega on kujutatud n-ö hea tihedusfunktsiooni valikut, mis vastab jaotusele $\mathcal{N}(10,15;10)$:

keskväärtuseks on võetud 10000 vaatluse valimikeskmine ja standardhälve on piisavalt suur, et tihedusfunktsioon “haaraks endasse” võimalikult laia vahemiku tegelikust tihedusest genereeritavatest väärtustest. Punane joon kujutab “halba” valikut, mis vastab jaotusele $\mathcal{N}(\mu = 10, 15; 1)$. Selle keskväärtus on küll lähedal valimikeskmisele, ent väike standardhälve kontsentreerib tõenäosumassi keskväärtuse ümber, mistõttu genereeritakse “kaugemaid” parameetrite väärtusi väga väikese tõenäosusega.

Kirjeldatud parameetrite eeljaotused annavad meile kokku neli Bayesi mudeli konfiguratsiooni, mis on toodud tabelis 3. Parameetrilise mudeli korral tuleb taas valida segu komponentide arv k ja Dirichlet’ jaotuse hüperparameetrid $\alpha_1, \dots, \alpha_k$. Nagu ka eelmise simulatsiooni korral, proovime väärtusi $k = 3$ ja $k = 10$. Mõlemal juhul valime mitteinformatiivse vektori \mathbf{p} eeljaotuse, võttes

$$\alpha_j = 1, \quad j = 1, \dots, k.$$

Mitteparameetrilise mudeli korral proovime simulatsioonides proovime iga valimimahu n korral kontsentratsiooniparameetri väärtusi

$$M \in \{1, \sqrt{n}, n\}.$$

Kõigi Bayesi mudelite puhul kasutame Gaussi tuuma (4.2), s.t $\psi(x; \mu_i)$ vastab jaotuse $\mathcal{N}(\mu_i; 1)$ tihedusfunktsioonile.

Lisaks sellele hindame tihedust (4.7) EM algoritmiga lõplike normaaljaotuste segude jaoks (funktsioon `normalmixEM()` R-i paketist `mixtools`) vastavalt $k = 3$ ning $k = 10$ komponendi korral, eeldusel, et iga komponendi standardhälve on $\sigma = 1$. Lõpuks kasutame tiheduse hindamiseks ka tihedushistorgammi (funktsioon `hist()` R-i paketist `graphics`), mis jaotab lõigu $[-100, 100]$ tuhandeks ($b = 1000$) võrdseks osaks.

Tiheduse hindamisel kasutame valimimahtusid $n = 20, 50, 100, 300$. Iga valimima-

MP hea (M)	MP halb (M)
$P \sim \Pi(M \cdot \mathcal{N}(10,15; 10))$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n P = p \stackrel{i.i.d.}{\sim} p$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$	$P \sim \Pi(M \cdot \mathcal{N}(10; 1))$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n P = p \stackrel{i.i.d.}{\sim} p$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$
P hea (k)	P halb (k)
$\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_k) \sim \text{Dir}(1, \dots, 1)$ $Y_1, \dots, Y_n \mathbf{p} = p \stackrel{i.i.d.}{\sim} \sum_{i=1}^k p_i \delta_i$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k \stackrel{i.i.d.}{\sim} \mathcal{N}(10,15; 10)$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$	$\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_k) \sim \text{Dir}(1, \dots, 1)$ $Y_1, \dots, Y_n \mathbf{p} = p \stackrel{i.i.d.}{\sim} \sum_{i=1}^k p_i \delta_i$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k \stackrel{i.i.d.}{\sim} \mathcal{N}(10; 1)$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$

Tabel 3: Bayesi mudelite konfiguratsioonid

hu korral arvutame välja hinnatud tiheduse L_1 -kauguse tegelikust tihedusest (4.7).

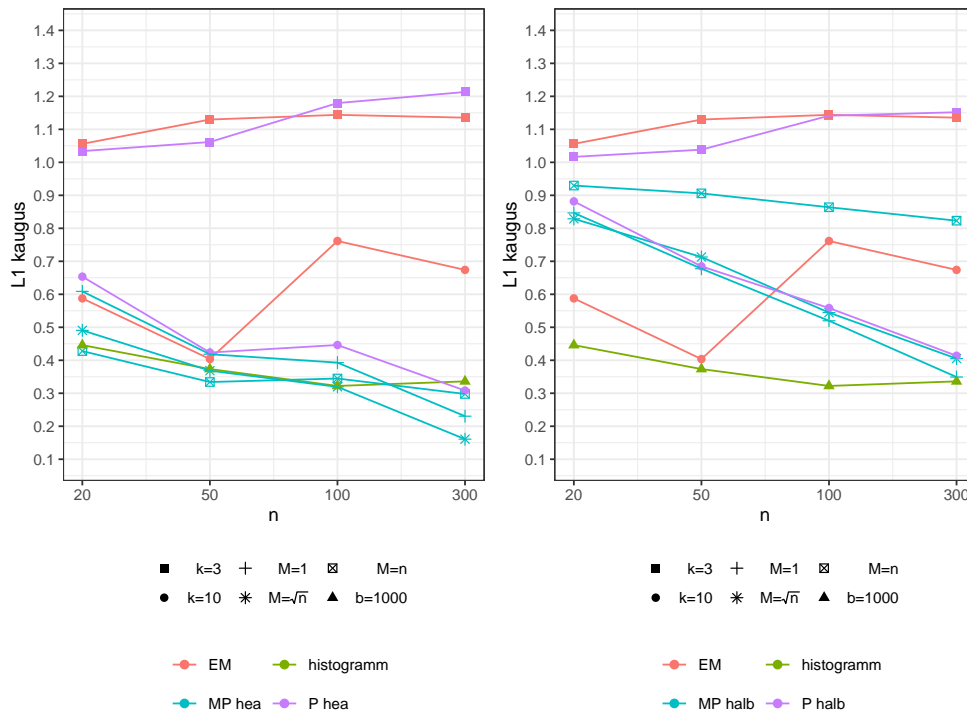
Kauguse hinnangud on esitatud tabelis 4 ning joonisel 4.

Simulatsioonide tulemused võib võtta kokku järgmiselt:

- Mõistliku eeljaotuse valiku korral (Dirichlet' protsessi jaotus Π_α , kus $\alpha = M \cdot \mathcal{N}(10,15; 10)$) andsid parima tiheduse hinnangu mitteparameetrilised mudelid. Tulemus on ootuspärane, kuna tegemist on õigete mudelitega – s.t tegelik tihedus esitub $\mathcal{N}(\mu; 1)$ üldise seguna. Valimimahu kasvades muutusid tiheduse hinnangud paremaks.

Võib tekkida küsimus, miks selline paranemine leidis aset, kui järelduses 4.4.2 toodud mõjususe kriteerium on rikutud (nimelt eksponentjaotuse kandja pole kompaktne). Viimane asjaolu aga ei pea meid heidutama, kuna järelduses 4.4.2 toodud tingimused on mõjususe tagamiseks piisavad, ent mitte tarvilikud, s.t hinnang võib osutuda mõjusaks ka siis, kui need tingimused on rikutud. Teisipidi olgu mainitud, et siinse simulatsiooni tulemusi ei saa ka kuidagiviisi mõjususe tõestuseks pidada.

- Viletsa eeljaotuse valiku (Dirichlet' protsessi jaotus Π_α , kus $\alpha = M \cdot \mathcal{N}(10; 1)$)



Joonis 4: Tiheduse hinnangute L_1 -kaugus tegelikust tihedusest. Vasakpoolsel joonisel on esitatud Bayesi mudelite “headele” konfiguratsioonidele vastavate hinnangute kaugused tegelikust tihedusest, paremal pool on “halvad” konfiguratsioonid. EM algoritmile ja histogrammile vastavad jooned on mõlemal pool samad.

korral andsid mitteparameetriselised Bayesi mudelid üldiselt viletsamaid hinnanguid kui histogramm, $n = 300$ korral jäi histogrammi ja MP halva ($M = 1$) hinnangu headuse erinevus statistilise vea piiridesse. MP halb ($M = n$) jäi oma sooritusest alla ka 10 komponendiga Bayesi mudelile ja oli iga valimimahu korral oluliselt halvem kui teised mitteparameetriselised mudelid. Niisugune tulemus on ootuspärane, kuna kõrge usaldusaste ($M = n$) viletsa eeljaotuse suhtes aeglustab koondumist tegelikuks jaotuseks. Sellegipoolest kasvas valimimahu kasvades kõigi mitteparameetriseliste Bayesi mudelite abil leitud hinnangute täpsus.

- 3 komponendiga parameetriselise Bayesi mudel andis väga halvad hinnangud

($\|\hat{\psi}_0 - \psi_0\|_{L_1} > 1$) nii viletsa kui ka mõistliku parameetrite eeljaotuse korral. Tulemus on ootuspärane, kuna tegemist on teoreetiliselt valede mudelitega – s.t tegelik tihedus ei esitu ühikstandardhälbega normaaljaotuste lõpliku seguna. Ka *ad hoc* vaatevinklist on tegelik tihedus (4.7) halvasti lähendatav kõigest kolme normaalkomponendiga. Erinevalt teistest Bayesi mudelitest läksid hinnangud valimimahu kasvades halvemaks. Viimane asjaolu pole üllatav, sest parameetriliste Bayesi mudelite antava hinnangu mõjususe kriteerium ei ole täidetud, kui tegelik tihedus on üldine segu.

- 10 komponendiga parameetiline Bayesi mudel andis oluliselt paremad tiheduse hinnangud kui 3 komponendiga parameetiline mudel nii hea kui ka halva eeljaotuse valiku korral. Üldiselt käitus see sarnaselt madala täpsusparameetriga ($M = 1$) mitteparameetrilisele Bayesi mudeliga, mis on kooskõlas üldise intuitsiooniga M ja k valiku kohta: madal M väärtus mitteparameetrilises mudelis väljendab aprioorset arusaama, et hinnatav tihedus on hästi lähendatav lõpliku normaalkomponentide hulgaga, teisalt muudab kõrge komponentide arv k parameetrilist mudelit “paindlikumaks” ja teatud mõttes sarnasemaks mitteparameetrilisele mudelile.
- Suuremate valimimahtude korral ($n = 100, 300$) on EM algoritmi abil saadud tiheduse hinnangud on halvad nii 3- kui ka 10-komponendilise mudeli eeldustel, jäädes täpsuselt alla 10 komponendiga Bayesi mudelile. Tulemus ei ole üllatav, kuivõrd tegeliku tiheduse (4.7) korral tegeleb EM algoritm sisuliselt vale tõepärafunktsiooni numbrilise maksimeerimisega.

Simulatsioonide tulemuste illustreerimiseks vaatame ka, millised “MP hea”, “MP halb”, “P hea” ja “P halb” mudelid andsid parimad tiheduse hinnangud valimil mahuga $n = 300$ ja joonistame välja vastavad tihedusfunktsioonid. Võrdluse mõttes esitame ka EM algoritmi ($k = 10, n = 300$) ja histogrammi ($n = 300, b = 1000$) abil leitud tiheduse hinnangute graafikud. Kõik need graafikud on koondatud lisas 2.

Mudel	$n = 20$	$n = 50$	$n = 100$	$n = 300$
MP hea (M=1)	0.614 (0.01)	0.413 (0.007)	0.357 (0.007)	0.26 (0.004)
MP hea (M= \sqrt{n})	0.52 (0.008)	0.366 (0.005)	0.304 (0.005)	0.18 (0.002)
MP hea (M=n)	0.411 (0.006)	0.328 (0.005)	0.342 (0.005)	0.301 (0.005)
MP halb (M=1)	0.923 (0.022)	0.704 (0.015)	0.845 (0.019)	0.813 (0.022)
MP halb (M=n)	0.899 (0.022)	0.924 (0.023)	0.928 (0.024)	0.839 (0.023)
MP halb (M= \sqrt{n})	0.889 (0.019)	0.821 (0.017)	0.69 (0.014)	0.651 (0.017)
P hea (k=3)	1.034 (0.023)	1.062 (0.023)	1.18 (0.026)	1.213 (0.028)
P hea (k=10)	0.654 (0.011)	0.423 (0.007)	0.446 (0.008)	0.308 (0.005)
P halb (k=3)	1.017 (0.024)	1.038 (0.023)	1.141 (0.025)	1.152 (0.026)
P halb (k=10)	0.882 (0.02)	0.685 (0.015)	0.558 (0.009)	0.414 (0.008)
EM (k=3)	1.049 (0.024)	1.133 (0.028)	1.174 (0.029)	1.194 (0.028)
EM (k=10)	0.595 (0.011)	0.438 (0.008)	0.629 (0.014)	0.805 (0.017)
histogramm (b=1001)	0.458 (0.01)	0.382 (0.006)	0.323 (0.006)	0.335 (0.006)

Tabel 4: Tiheduse hinnangute L_1 -kaugus tegelikust tihedusest. Sulgudes toodud arvud tähistavad tõenäosusega 0,95 kehtivat viga.

Ühtlaste jaotuste lõplik segu

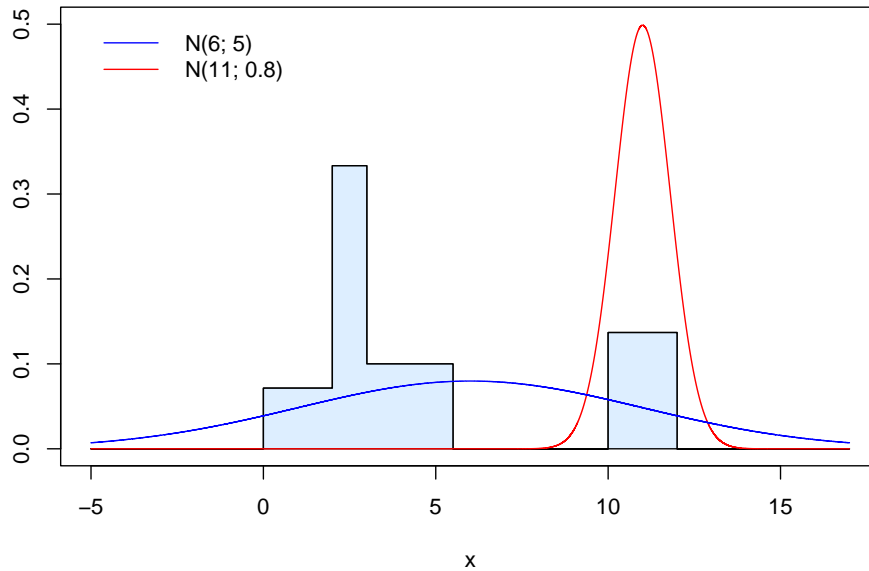
Genereerime vaatlused ühtlaste jaotuste segust

$$\frac{1}{7} \cdot \mathcal{U}(0; 2) + \frac{1}{3} \cdot \mathcal{U}(2; 3) + \frac{1}{4} \cdot \mathcal{U}(3; 5, 5) + \frac{23}{84} \cdot \mathcal{U}(10; 12). \quad (4.8)$$

Märgime, et erinevalt eelmistes simulatsioonides hinnatavatest jaotustest ei esitu (4.8) normaaljaotuste seguna. Nimelt on normaaljaotuste segu kandja alati kogu reaaltelg \mathbb{R} , aga siin vaadeldava ühtlaste jaotuste segu kandja on lõikude ühend $[0, 5, 5] \cup [10, 12]$.

Joonisel 5 on kujutatud tihedusfunktsioon (4.8) ja kaks varianti tihedusfunktsioonile π või $h_{\bar{\alpha}}$, mida peame valima Bayesi mudeli implementeerimisel. Sinisega on kujutatud hea tihedusfunktsiooni valikut, mis vastab jaotusele $\mathcal{N}(6; 5)$: keskväärtus ja standardhälve on valitud nii, et tihedusfunktsioon “haaraks endasse” võimalikult suure vahemiku tegelikust tihedusest genereeritavatest väärtustest. Punane joon kujutab “halba” valikut, mis vastab jaotusele $\mathcal{N}(11; 0,8)$. Selle tiheduse tõenäosusmass kontsentreerub vaid üksiku tegeliku jaotuse komponendi ümber.

Kirjeldatud tihedused annavad meile kokku neli Bayesi mudeli konfiguratsiooni, mis



Joonis 5: Ühtlaste jaotuste lõpliku segu tihedus ja tiheduste π või $h_{\bar{\alpha}}$ variandid

on toodud tabelis 5. Parameetrilise mudeli korral tuleb jällegi valida segu komponentide arv k ja Dirichlet' jaotuse hüperparameetrid $\alpha_1, \dots, \alpha_k$. Simulatsioonides proovime väärtusi $k = 3$ ja $k = 10$. Mõlemal juhul valime mitteinformatiivse vektori \mathbf{p} eeljaotuse, võttes

$$\alpha_j = 1, \quad j = 1, \dots, k.$$

Mitteparameetrilise mudeli puhul tuleb veel valida täpsuskordaja M . Simulatsioonides proovime iga valimimahu n korral väärtusi

$$M \in \{1, \sqrt{n}, n\}.$$

Kõigi Bayesi mudelite puhul kasutame Gaussi tuuma (4.2), s.t $\psi(x; \mu_i)$ vastab jaotuse $\mathcal{N}(\mu_i; 1)$ tihedusfunktsioonile.

Lisaks sellele hindame tihedust (4.7) EM algoritmiga lõplike normaaljaotuste segu-

MP hea (M)	MP halb (M)
$P \sim \Pi(M \cdot \mathcal{N}(6; 5))$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n P = p \stackrel{i.i.d}{\sim} p$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$	$P \sim \Pi(M \cdot \mathcal{N}(11; 0, 8))$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n P = p \stackrel{i.i.d}{\sim} p$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$
P hea (k)	P halb (k)
$\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_k) \sim \text{Dir}(1, \dots, 1)$ $Y_1, \dots, Y_n \mathbf{p} = p \stackrel{i.i.d}{\sim} \sum_{i=1}^k p_i \delta_i$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k \stackrel{i.i.d}{\sim} \mathcal{N}(6; 5)$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$	$\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_k) \sim \text{Dir}(1, \dots, 1)$ $Y_1, \dots, Y_n \mathbf{p} = p \stackrel{i.i.d}{\sim} \sum_{i=1}^k p_i \delta_i$ $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k \stackrel{i.i.d}{\sim} \mathcal{N}(11; 0, 8)$ $X_i \boldsymbol{\mu}_i = \mu_i \stackrel{ind}{\sim} \psi(x; \mu_i)$

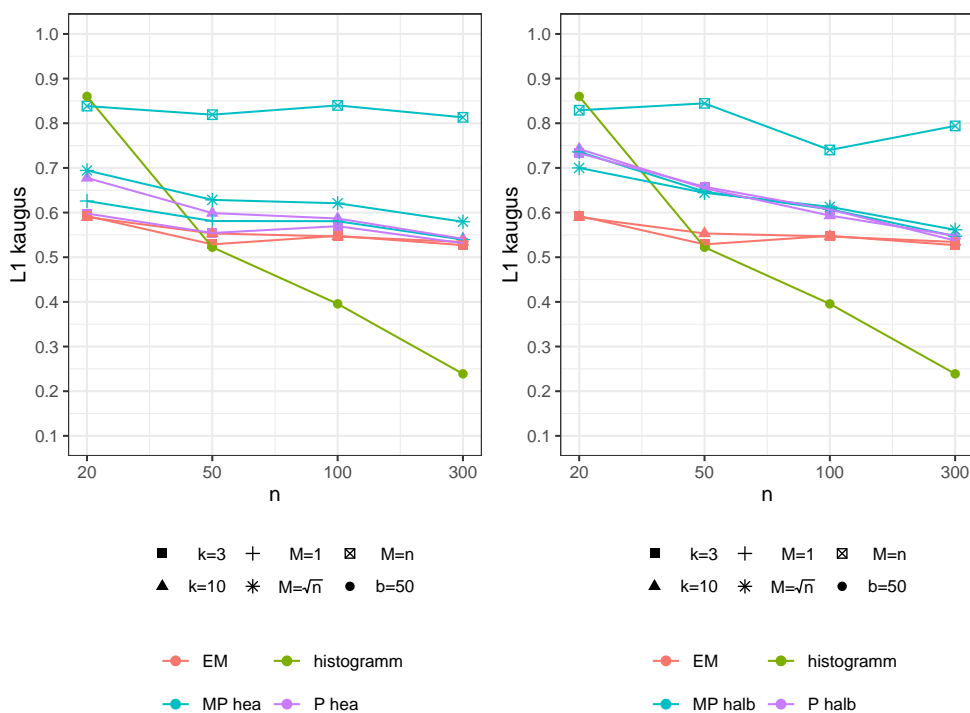
Tabel 5: Bayesi mudelite konfiguratsioonid

de jaoks (funktsioon `normalmixEM()` R-i paketist `mixtools`) vastavalt $k = 3$ ning $k = 10$ komponendi korral, eeldusel, et iga komponendi standardhälve on $\sigma = 1$. Lõpuks kasutame tiheduse hindamiseks ka tihedushistorgammi (funktsioon `hist()` R-i paketist `graphics`), mis jaotab lõigu $[-3, 15]$ viiekümneks ($b = 50$) võrdseks osaks.

Tiheduse hindamisel kasutame valimimahtusid $n = 20, 50, 100, 300$. Iga valimimahtu korral arvutame välja hinnatud tiheduse L_1 -kauguse tegelikust tihedusest (4.8). Kauguse hinnangud on esitatud tabelis 4.3 ning joonisel 6.

Simulatsioonide tulemused illustreerivad hästi siinses töös käsitletud Bayesi mudelite piiranguid. See, et parameetrilised mudelid annavad viletsaid tiheduse hinnanguid ($\|\psi_0 - \hat{\psi}_0\|_{L_1} > 0,5$), on igati oodatav – intuitsiooni tasemel on selge, et ühtlaste jaotuste segu ei olegi hästi lähendatav 3 või 10 komponentide normaaljao-tuste seguga. Samal põhjusel on ka EM-algoritmi abil saadud tiheduse hinnangud halvad.

Ka mitteparameetrilised mudelid kukuvad siinkohal läbi. Ehkki Gaussi tuuma va-



Joonis 6: Tiheduse hinnangute L_1 -kaugus tegelikust tihedusest. Vasakpoolsel joonisel on esitatud Bayesi mudelite “headele” konfiguratsioonidele vastavate hinnangute kaugused tegelikust tihedusest, paremal pool on “halvad” konfiguratsioonid. EM algoritmile ja histogrammile vastavad jooned on mõlemal pool samad.

likul seatakse üsna leebed eeldused tegelikule tihedusele ψ_0 (sisuliselt eeldatakse vaid seda, et tegeliku jaotuse kandja on \mathbb{R}), ei toimi meetod enam kuigi hästi, kui need on rikutud. Olukordades, kus on põhjust arvata, et tegeliku jaotuse kandja ei ole \mathbb{R} , valitakse enamasti teistsugused tuumad (vt näiteid [7, lk 116–121]).

Valimimahul $n = 100, 300$ andis teistest meetoditest oluliselt paremad tiheduse hinnangu ootuspäraselt hoopis histogramm. Ehkki tegemist on väga robustse mudeliga, osutusid selle eeldused antud olukorras kõige mõistlikumaks – sisuliselt hindab b lõiguga histogramm teatud b -komponendilist ühtlaste jaotuste segu.

Simulatsioonide tulemuste illustreerimiseks vaatame ka, millised "MP hea", "MP halb", "P hea" ja "P halb" mudelid andsid parimad tiheduse hinnangud valimil ma-

Mudel	$n = 20$	$n = 50$	$n = 100$	$n = 300$
MP hea ($M=1$)	0.626 (0.014)	0.581 (0.014)	0.581 (0.013)	0.539 (0.013)
MP hea ($M=\sqrt{n}$)	0.694 (0.015)	0.628 (0.014)	0.621 (0.013)	0.579 (0.013)
MP hea ($M=n$)	0.838 (0.018)	0.819 (0.018)	0.84 (0.017)	0.813 (0.017)
MP halb ($M=1$)	0.736 (0.019)	0.647 (0.017)	0.607 (0.015)	0.547 (0.013)
MP halb ($M=\sqrt{n}$)	0.7 (0.019)	0.644 (0.017)	0.613 (0.014)	0.561 (0.013)
MP halb ($M=n$)	0.829 (0.021)	0.845 (0.021)	0.74 (0.018)	0.794 (0.019)
P hea ($k=3$)	0.598 (0.014)	0.554 (0.013)	0.57 (0.013)	0.531 (0.013)
P hea ($k=10$)	0.678 (0.015)	0.599 (0.014)	0.587 (0.013)	0.541 (0.013)
P halb ($k=3$)	0.733 (0.019)	0.658 (0.017)	0.607 (0.015)	0.537 (0.013)
P halb ($k=10$)	0.742 (0.019)	0.654 (0.017)	0.593 (0.014)	0.548 (0.013)
EM ($k=3$)	0.592 (0.015)	0.529 (0.013)	0.548 (0.013)	0.527 (0.013)
EM ($k=19$)	0.59 (0.015)	0.553 (0.014)	0.547 (0.013)	0.534 (0.013)
hist ($b=50$)	0.86 (0.028)	0.522 (0.016)	0.396 (0.014)	0.239 (0.01)

Tabel 6: Tiheduse hinnangute L_1 -kaugus tegelikust tihedusest. Sulgudes toodud arvud tähistavad tõenäosusega 0,95 kehtivat viga.

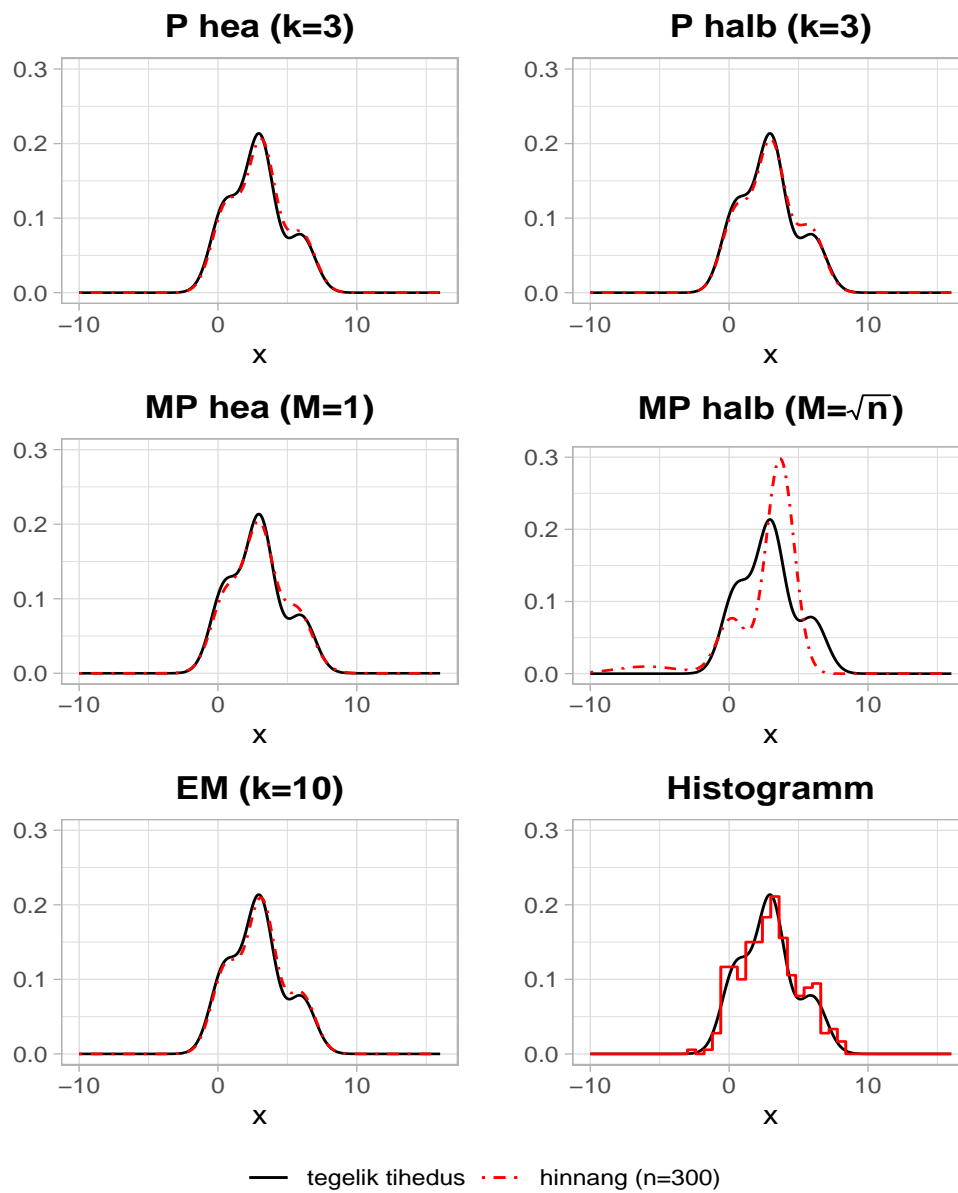
huga $n = 300$ ja joonistame välja vastavad tihedusfunktsioonid. Võrdluse mõttes esitame ka EM algoritmi ($k = 3, n = 300$) ja histogrammi ($n = 300, b = 50$) abil leitud tiheduse hinnangute graafikud. Kõik need joonised on koondatud lisa 3.

Kasutatud allikad

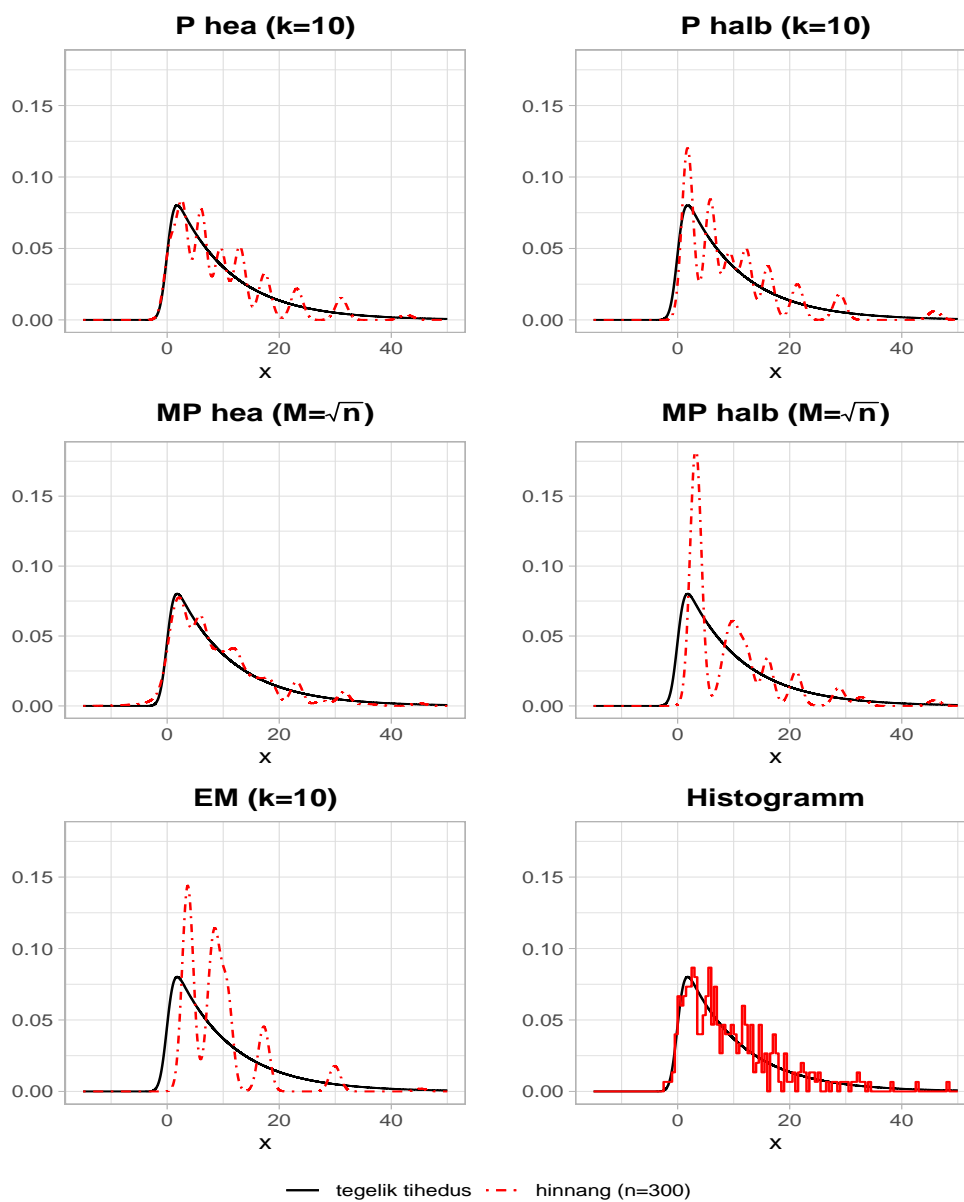
- [1] Charles E. Antoniak. “Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems”. *The Annals of Statistics* 2.6 (1974), lk. 1152–1174. URL: <https://doi.org/10.1214/aos/1176342871> (vaadatud 16.08.2022).
- [2] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 1999.
- [3] Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, 1986.
- [4] Michael D. Escobar. “Estimating Normal Means with a Dirichlet Process Prior”. *Journal of the American Statistical Association* 89.425 (1994), lk. 268–277. URL: <http://www.jstor.org/stable/2291223> (vaadatud 08.08.2022).
- [5] Thomas S. Ferguson. “A Bayesian Analysis of Some Nonparametric Problems”. *The Annals of Statistics* 1.2 (1973), lk. 209–230. URL: <http://www.jstor.org/stable/2958008> (vaadatud 18.08.2022).
- [6] S. Ghosal, J. K. Ghosh ja R. V. Ramamoorthi. “Posterior consistency of Dirichlet mixtures in density estimation”. *The Annals of Statistics* 27.1 (1999), lk. 143–158. URL: <https://doi.org/10.1214/aos/1018031105>.
- [7] Subhashis Ghosal ja Aad van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, 2017.
- [8] Rainis Haller. *Üldine topoloogia I (loengukonspekt)*. Tartu Ülikool, 2015.
- [9] Antonio Lijoi, Igor Prünster ja Stephen G. Walker. “On Consistency of Nonparametric Normal Mixtures for Bayesian Density Estimation”. *Journal of the American Statistical Association* 100.472 (2005), lk. 1292–

1296. URL: <http://www.jstor.org/stable/27590672> (vaadatud 15.08.2022).
- [10] Steven N. MacEachern. “Computational Methods for Mixture of Dirichlet Process Models”. Teoses: *Practical Nonparametric and Semiparametric Bayesian Statistics*. Toim. Dipak Dey, Peter Müller ja Debajyoti Sinha. New York, NY: Springer New York, 1998, lk. 23–43. ISBN: 978-1-4612-1732-9. URL: https://doi.org/10.1007/978-1-4612-1732-9_2.
- [11] Kevin Murphy. “Conjugate Bayesian analysis of the Gaussian distribution” (2007). URL: <https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf> (vaadatud 04.08.2022).
- [12] Radford M. Neal. “Markov Chain Sampling Methods for Dirichlet Process Mixture Models”. *Journal of Computational and Graphical Statistics* 9.2 (2000), lk. 249–265. URL: <http://www.jstor.org/stable/1390653> (vaadatud 17.08.2022).
- [13] T. Tin Nguyen *et al.* “Approximation by finite mixtures of continuous density functions that vanish at infinity”. *Cogent Mathematics & Statistics* 7.1 (2020). URL: <https://doi.org/10.1080/25742558.2020.1750861> (vaadatud 11.08.2022).
- [14] Mikael Raihhelgauz. *Dirichlet’ protsess (bakalaureusetöö)*. Tartu Ülikool, 2020.

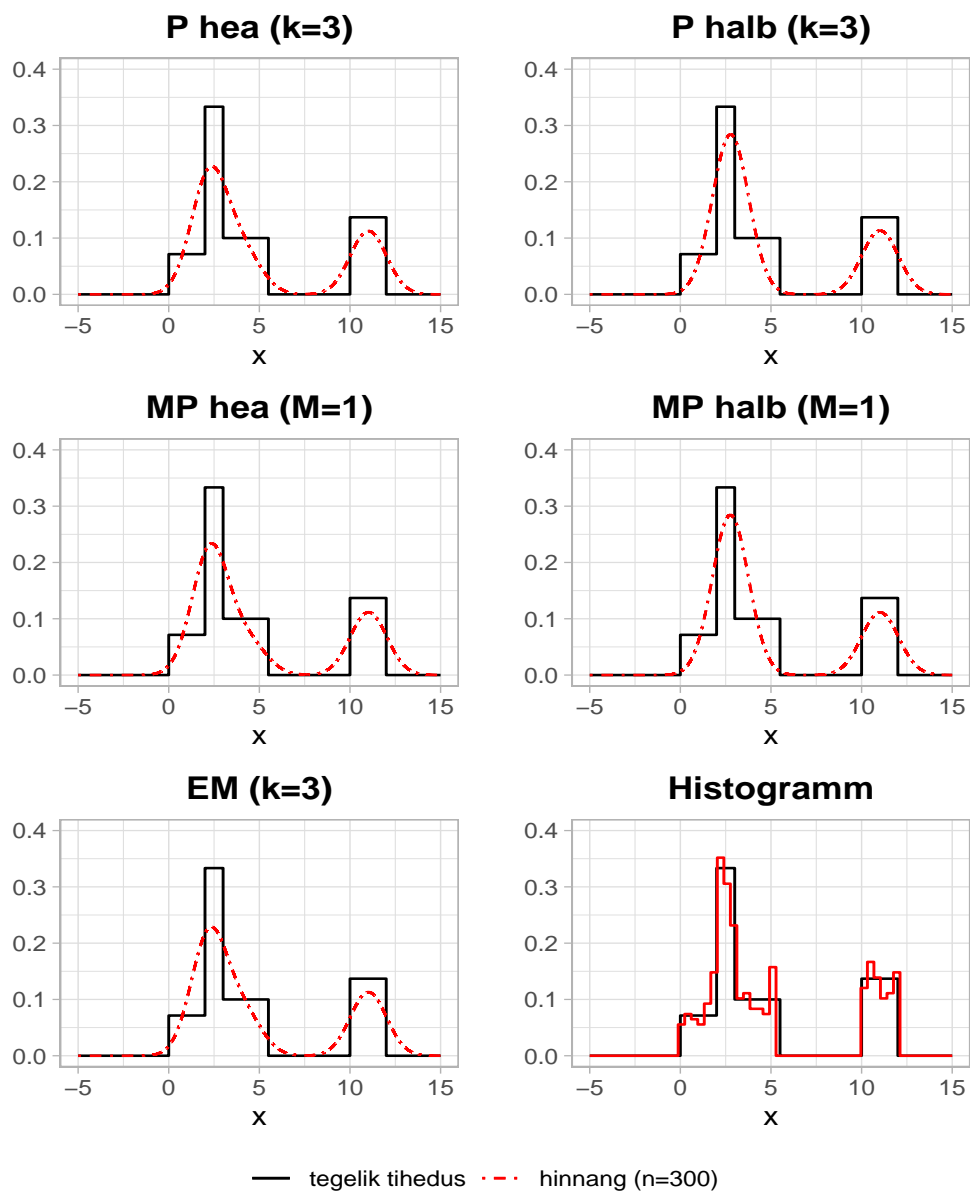
Lisa 1. Lõpliku normaaljaotuste segu tiheduse hinnangud



Lisa 2. Üldise segu tiheduse hinnangud



Lisa 3. Ühtlaste jaotuste segu tiheduse hinnangud



Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Mikael Raihhelgauz,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Bayesi segumudelid tiheduse hindamiseks", mille juhendaja on Jüri Lember, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Mikael Raihhelgauz

18.08.2022