Tartu Ülikool

Loodus- ja täppisteaduste valdkond

Matemaatika ja statistika instituut

Kristjan Sinikas

Kindlustuslepingute hinnastamine Bayesi meetodi abil

Kindlustus- ja finantsmatemaatika

Magistritöö (30 EAP)

Juhendaja: PhD Märt Möls

Tartu 2021

KINDLUSTUSLEPINGUTE HINNASTAMINE BAYESI MEETODI ABIL

Magistritöö

Kristjan Sinikas

Lühikokkuvõte

Käesoleva magistritöö eesmärk on uurida Bayesi statistika kasutatavust kahjukindlustuse hinnastamisel. Antud töös on kasutatud Bayesi statistika meetodit andmaks hinnanguid kindlustuslepingute kahjusageduse ja keskmise kahju suuruse osas ning seda on võrreldud suurima tõepära meetodi hinnangutega kasutades erinevaid kaofunktsioone. Leiti, et Bayesi meetod on hea andmete vähesuse korral.

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

Märksõnad: Bayesi printsiip, kindlustusmatemaatika, kahjukindlustus.

INSURANCE PREMIUM PRICING WITH BAYESIAN METHOD

Master thesis

Kristjan Sinikas

Abstract

The purpose of this thesis is to study Bayesian statistics usability in non-life insurance pricing. In this thesis Bayesian statistics method is used to estimate the claim frequency and average claim size and it is compared to maximum likelihood estimation using different loss functions. It is found that Bayesian method is good with low amount of data.

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

Key Words: Bayesian analysis, actuarial mathematics, non-life insurance.

Sisukord

Sissejuhatus

Kindlustusselts on ettevõte, mille eesmärgiks on teenida kasumit. Kasumivõimalusteks on investeerimisest ja rahavoost tulenev tulu, millest viimane on põhisissetulek püsikulude ja tegevuskulude katteks. Uue toote turule tulekuga on tarvis hinnastada see optimaalselt, st. maksimiseerida hind turul, võttes arvesse, et konkureerivatest seltsidest oleks see väiksem (eeldusel, et kindlustustoode on samasugune ja klient teeb oma otsused ainult hinna põhiselt). Üheks võimalikuks lahenduseks esmase hinna leidmisel ja jätkuval parendamisel on Bayesi statistika.

Magistritöö eesmärk on demonstreerida Bayesi statistika võimalikku kasutust kaskokindlustuse hinnastamisel, simuleerida võimalikke kaskokindlustuse hindu ja uurida kuidas neid edaspidi muuta ning kui suur rahaline kahju võib tulla ekspedi eksimustest ning kuidas seda minimiseerida.

Magistritöö on jagatud seitsmeks peatükiks. Probleemipüstituse peatükis kirjeldatakse töös lahendatavat probleemi. Bayesi statistika peatükis tutvustatakse meetodit, millega eelnevas peatükis välja toodud probleemi hakatakse lahendama. Peatükis töös kasutatavad jaotused antakse ülevaade jaotustest, mida kasutatakse järgnevate peatükkide simulatsioonides. Bayesi statistika meetodi koondumine ja tehtud vea maksumuse peatükis kasutatakse Bayesi statistikat ja antakse ülevaade võimalike vigade suurusest. Kahju suuruse peatükis kirjeldatakse eeljaotuse valiku mõju kahju suuruse hinnangule. Peatükis kindlustuslepingu hind kirjeldatakse kuidas päriseluliselt läheneda töös kirjeldatud probleemile. Peatükis päriseluline näide on toodud näide Bayesi statistika kasutamise kohta tegelike andmete peal.

Töö vormistamiseks on kasutatud tekstitöötlusprogrammi LaTeX. Töös on kasutatud statistika tarkvara R (R Core Team, 2018).

Autor tänab juhendajat Märt Mölsi tema kiirete vastuste ja aja leidmise eest.

1 Probleemipüstitus

Kindlustusselts müüb klientidele toodet, mille ostnud klient saab õnnetuse korral temale tekkinud materiaalse (või vaimse) kahju tagajärjel tekkinud finantskulude eest hüvitist. Kindlustusseltsi eesmärgiks on teenida kasumit ja seetõttu on kindlustuslepingute hinnad vaja võimalikult täpselt hinnastada. Kui selts hindab müüdavate lepingute kulusid väiksemana kui need tegelikkuses tulevad, siis väheneb kasumimarginaal või tekib kahjum. Vastupidisel korral hinnatakse riskid suuremaks kui need tegelikult on, mistõttu võib klientidest ilma jääda, kuna kindlustusturul pakuvad teised seltsid madalamat hinda ning seega võib kokkuvõttes vähem raha teenida.

Selleks, et kasum oleks võimalikult suur ja riskid täpselt hinnatud tuleb kindlustusseltsidel klientidele leida õiglased hinnad. Lahenduseks on erinevaid võimalusi, näiteks saab seda teha seltsi kogunenud ajalooliste andmete analüüsi või turu-uuringu põhjal. Turu-uurimine on ajakulukas ning annab teada konkureerivate seltside hinnad, mis ei anna mingisugust informatsiooni selle kohta kui tihedalt ja kui suured kahjud tekivad. Parem ja usaldusväärsem informatsioon peitub ajaloolistes ehk eelnevate müüdud kindlustuslepingutega seotud informatsioonis, kus on kliendi kohta teada erinevat kindlustatud eset (või elusolendit) puudutavat informatsiooni. Mainitud andmete analüüsimiseks on olemas palju artikleid ja raamatuid, kuid suurem probleem tekib siis, kui seltsil puuduvad ajaloolised andmed.

Toote informatsiooni puudulikkuse korral tuleb teha toote (ja tema hindade) kohta taustauuring ja määrata esialgsed hinnad ning pärast müüdud lepingute kahjude tekkimist (eeldadavasti tekivad kahjud) peab võimalikult kiiresti korrigeerima hindu nende andmete põhiselt. Kuidas määrata ja jooksvalt uuendada kindlustustoodete hindu? Sellele küsimusele hakkabki autor töös vastust leidma.

Kindlustusseltsi sissetulevaks rahavooks on müüdud kindlustuslepingutest saadav raha. Selleks, et riske hajutada, on mõistlik pakkuda erinevaid tooteid, kuna on võimalus, et turuolukord on selline, et müüakse ühele kindlustustootele kahjumlik-

ke lepinguid ja ei ole võimalik tulu teenida.
Uueks tooteks on valitud kaskokindlustus, mille kohta on adekvaatsete hindade modelleerimiseks seltsil vähe infot.

Kaskokindlustust nimetatakse sõiduki kindlustuseks ja õnnetuse korral kaetakse tekkinud finantsilised kulud kindlustusseltsi poolt, st. finantseeritakse, et remondikulud või teisel juhul, kui on vajadus sõiduk liiklusregistrist maha kanda, makstakse kinni sõiduki väärtus enne õnnetust.
Kaskokindlustuse kahjude andmed on igal kindlustusseltsil ettevõttesisene salajane informatsioon ning ei ole Eesti turul saadaval. Kättesaadavad on summaarsed kahjusummad füüsiliste ja juriidiliste isikute lõikes, kuid see ei anna piisavalt informatsiooni tegelike hindade arvutamiseks ning seetõttu on vaja leida moodus, kuidas poliise hinnastada võimalikult täpselt ning kahjude tekkimise korral hindu jooksvalt uuendada.

2 Bayesi statistika

Enne kui Bayesi statistikat tutvustada, siis on tarvilik selgitada, miks on Bayesi statistika meetodid üldse vajalikud ja miks mitte kasutada klassikalise statistika lähenemist?

Klassikaline statistika (*frequentism*) ja Bayesi statistika probleemile lähenemise erinevused on järgnevad:

- Klassikaline statistika väidab, et igal parameetril on üks kindel ja fikseeritud väärtus. Enamasti pole teada, milline see parameetri väärtus on, kuid seda saab hinnata kasutades valimit;
- Bayesi statistikas ei ole parameetrite väärtused enamasti fikseeritud, vaid parameetri tegelikku väärtust käsitletakse kui juhuslikku suurust. Mõnda parameetri võimalikke väärtuseid võidakse lugeda tõenäolisemateks kui teisi — lähtudes näiteks eelnevatest kogemustest.

Seega Bayesi statistika võimaldab kasutada eelnevaid teadmisi parameetrite hindamisel, mis on hea valikuvõimalus olukorras, kus andmeid on vähe, sest sel juhul võib ainult vaatlusandmetele tuginev hinnang parameetrile olla väga ebatäpne.

Bayesi statistika aluseks on tõenäosusteoorias ja statistikas tuntud Bayesi valem (Pärna, 2013, lk 41):

$$\mathrm{P}(B_i|A) = \frac{\mathrm{P}(A|B_i)\mathrm{P}(B_i)}{\sum_{i=1}^{n} \mathrm{P}(A|B_i)\mathrm{P}(B_i)} \tag{1}$$

Valem (1) kirjeldab, mis on sündmuse B_i toimumise tõenäosus, kui on teada, et toimus sündmus A. Murrulugeja on sündmuse A toimumise tõenäosus tingimusel, et toimub sündmus B_i korrutatud B_i toimumise tõenäosusega ja murrunimetaja on nende sündmuste tõenäosuse summa iga erineva i korral.

Tuues paralleeli parameetrite hindamisele, siis andmed on kui realiseerunud sündmus A (teada olev info), mille toimumisele saab arvutada tõenäosuse iga erineva parameetri B_i korral. Modifitseerides Bayesi valemit (1) saame võrduse (2) (Gray ja Pitts, 2012, lk 158-159):

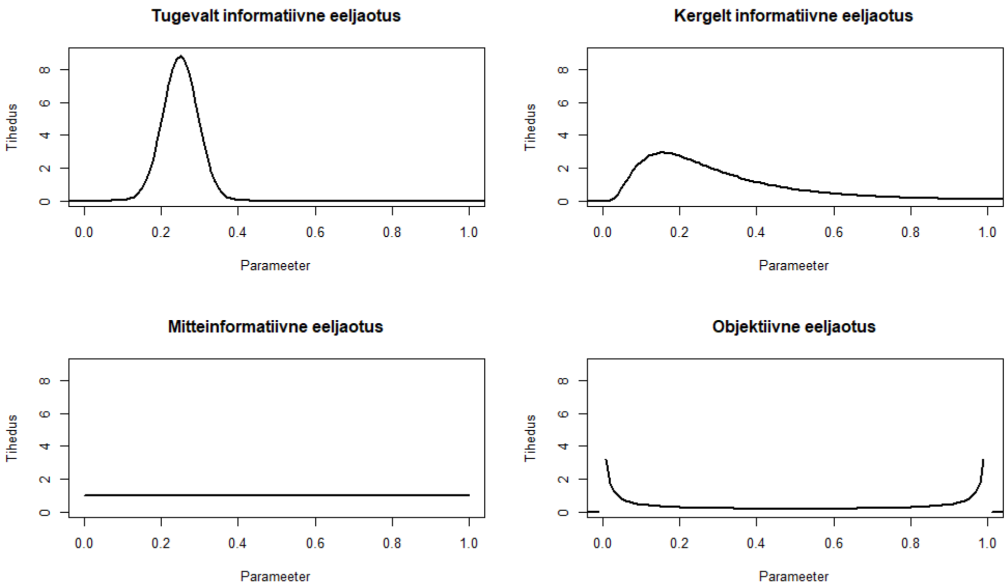
$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} \tag{2}$$

- $f(\theta|x)$ - järeljaotus parameetritele antud andmete kohaselt;
- $f(x|\theta) = \mathrm{L}(x|\theta)$ - tõepärafunktsioon;
- $f(\theta)$ - eeljaotus parameetritele (teadmised parameetrite kohta enne andmestiku vaatamist);
- $f(x) = \int f(x|\theta)f(\theta)d\theta$ - normaliseerimise konstant, et järeljaotuse alune pindala oleks 1 ja rahuldaks tihedusfunktsiooni tingimusi.

Bayesi statistika meetodil on kolm põhilist komponenti: eeljaotus, tõepärafunktsioon ja järeljaotus. Seega kui leida andmete kohaselt tõepärafunktsioon ning valida kogemuse põhjal sobilik eeljaotus, siis järeljaotuseks saadakse tõenäosusfunktsioon hinnatavale parameetrile.

Eeljaotus on eelnevast kogemusest oletatav jaotus parameetrile, mis on leitud enne kui andmeid on nähtud. See on funktsioon, mis annab juhuslikud väärtused parameetrile (eksperdi valitud, on võimalik valida ka funktsioone, mis annavad ühtlaseid väärtuseid kogu reaal-teljel, mis kaotaks ära meetodi idee kasutada eelteadmisi, kuna siis järeljaotus oleks proportsionaalne tõepärafunktsiooniga ja näiteks järeljaotuse maksimum saavutatakse sama parameetri väärtuse korral, mis maksimiseerib ka tõepärafunktsiooni väärtust).

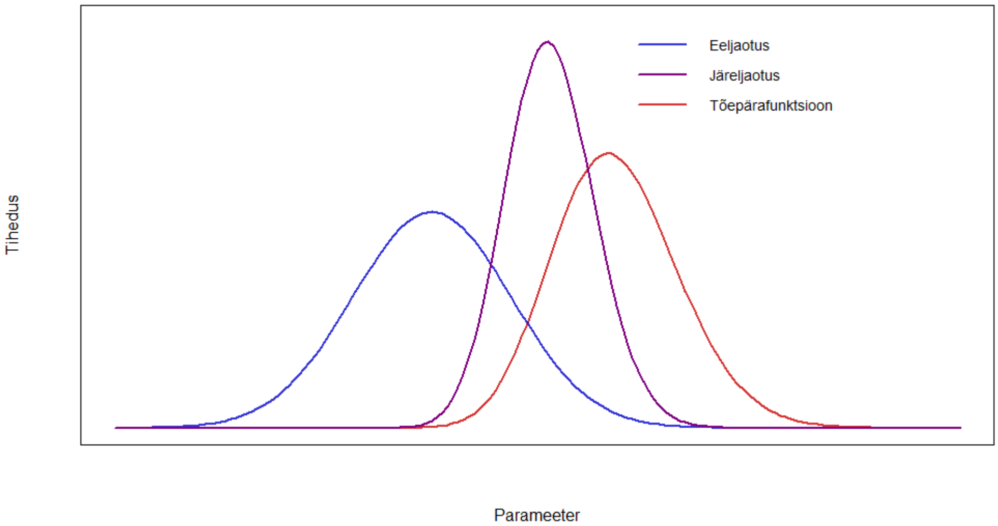
Joonisel 1 on näha erinevat eeljaotuse valiku võimalust, kus tugevalt informatiivne eeljaotus on hea valik, kui eksperdil on vajalikud kogemused ja teadmised valitava parameetri kohta. Kergelt informatiivse eeljaotuse valik sobib antud juhul siis, kui on teadmised, et tegeliku parameetri väärtus on nulli ja ühe vahel ning jääb nullile ligemale kui ühele. Mitteinformatiivset eeljaotust võiks kasutada siis, kui ekspert ei oska arvata midagi parameetri tegeliku väärtuse kohta. Objektiivne eeljaotus valitakse siis, kui on teada, mis piirkonda parameeter jääb (nt. tõenäosus $(0;1)$ ja dispersioon (>0)).



Joonis 1: Neli erinevat eeljaotust

Tõepärafunktsioon on funktsioon, mis arvutab jaotusfunktsiooni võimalike parameetrite väärtuste kohta tõepära, et valim on saadud neist parameetritest.

Järeljaotus on eeljaotusest ja tõepärafunktsioonist tingitud funktsioon, mis nende funktsioonide kokku korrutamisel annab tõenäosused funktsiooni parameetritele, mis kirjeldavad vaadeldud andmeid nagu seda on visualiseeritud joonisel 2.



Joonis 2: Bayesi meetodi visualiseering

Selleks, et lihtsamini mõista Bayesi statistika meetodit, siis saab võrduse 2 kirjutada mitte-formaalselt (Gray ja Pitts, 2012, lk 159):

$$\text{järeljaotus} = \frac{\text{eeljaotus} \times \text{tõepärafunktsioon}}{\text{normeeriv konstant}} \tag{3}$$

Kuna normeeriv konstant ei sõltu parameetri θ väärtusest, siis võrduse (2) kirjutada proportsionaalses vormis:

$$f(\theta|x) \propto f(x|\theta)f(\theta) \tag{4}$$

See tähendab, et järeljaotus on proportsionaalne eeljaotuse ja tõepärafunktsiooni korrutisega.

Nüüd tekib aga küsimus, kui tihti järeljaotust arvutada, kui andmeid tekib jooksvalt juurde? Järeljaotust võib arvutada pärast iga andmestikku lisandunud objekti või pärast n lisandunud andmerida, lõpptulemus jääb ikka samaks ja ei sõltu sellest, kas arvutati igal sammul või lõpus ühe korra.

Bayesi statistikas tuleb otsuseid parameetri tegeliku väärtuse kohta teha järeljaotuse põhjal. Järeljaotus võimaldab meil otsustada kui tõenäoliselt on parameetri väärtus mingis kindlas vahemikus kuid ei anna otseselt tulemuseks ühte kindlat väärtust või hinnangut tundmatule parameetrile.

Kindluslepingute hinnastamiseks on tarvis leida siiski üks konkreetne hinnang, punkthinnang. Selleks punkthinnanguks võib valida kas järeljaotuse keskmise, mediaani või moodi. Juhul kui eeljaotus ei välista parameetri tegelikku väärtust siis vastavalt Bernstein-von Mises'e teoreemile koondub järeljaotuse põhjal leitud hinnang (ükskõik kas kasutame järeljaotuse keskmist, mediaani või moodi) suurima tõepära hinnnaguks. Punkthinnangu valikul on enam levinud järeljaotuse keskmine või selle maksimumpunkt. Praktikas kasutatakse sageli punkthinnanguks maksimumi, kuna siis ei ole tarvis arvutada valemist (2) normeerivat konstanti $f(x)$, kuna maksimumpunkt asub samas kohas. Järeljaotuse keskmise valikul hajutab see rohkem riske ja on teoorias parem valik, aga kuna keskmine sõltub normaliseerimise konstandist, siis on tarvis see välja arvutada, mis aga praktikas tihti ei õigusta ennast.

Eksperdiks sobib spetsialist, kes on kokku puutunud sarnaste andmetega. Esmase eeljaotuse parameetrite valikul on vaja kätte saada eksperdi tunnetuslik arvamus, mis eeljaotuse keskväärtusele punkthinnanguna ei pruugigi nii raske olla, kui tegemist on selles vallas tegutsenud professionaaliga, kust andmed pärinevad. Kuna ekspert tihti ei oskagi öelda, milline on täpne punkthinnang, oskab ta siiski sageli pakkuda võimalike väärtuste vahemikku. Kui näiteks kasutada eeljaotusena normaaljaotust siis võib eeljaotuse keskväärtuseks valida eksperdi arvates kõige tõenäolisema parameetri väärtuse ja eeljaotuse hajuvuse valida selliselt, et eksperdi arvates tõenäoliste parameetri väärtuste piirkonda sattumise tõenäosus oleks suur — näiteks 95%. Vaatluste hajuvuse (või ülehajuvusparameetri) hindamisel ei pruugi aga eksperdi sisehääl osata meile suurt midagi öelda. Sellisel juhul võib kasutada sarnaste andmestike abi — näiteks võime hinnata hajuvus- või ülehajuvusparameetri väärtuse mõne lähedase kindlustustoote portfelli kasutades ja kasutada saadud hinnangut kui eeljaotuse keskväärtust.

Näide.

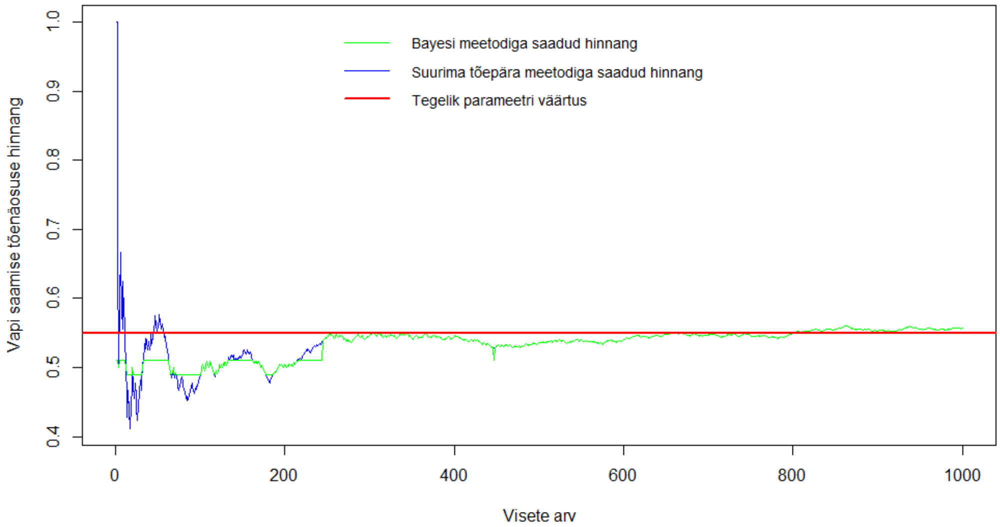
Visatakse münti, mille tõenäosus vapi ja kirja saamisel on teadmata. Esimesel viskel tuleb vapp, mis võiks olla vapi saamise tõenäosus?

Intuitiivselt võiks arvata, et kui esimene vise on vapp, siis see ei tähenda, et ka järgmistel visetel tuleks vapp. Mille tõttu Bayesi statistikaga saab kasutada optimistlikkuma hinnangut parameetrile valides näiteks eeljaotuseks kahe ühtlase jaotuse segu: 0,49 U(0,49; 0,51) + 0,51 U(0; 1). Parameetrite hinnangud pärast igat viset lähenevad õige parameetri punkthinnangu suunas ja ei muutu volatiilselt nagu seda teevad suurima tõepära (ST) hinnangud vapi saamise tõenäosusele, pärast kahte viset $\hat{p} \in \{0{,}5; 1\}$, mille ulatus on pool võimalikest tõenäosustest $(1 - 0{,}5 = 0{,}5)$.

Vapi saamine ST: $\hat{p} = 1$ — Vapi saamine Bayes: $\hat{p} = 0{,}51$

$\hat{p} = 1$ — $\hat{p} = 0{,}5$ — $\hat{p} = 0{,}51$ — $\hat{p} = 0{,}5$

Pärast tuhandet viset on siiski mõlemal sama hinnang 0,557, kus suurima ja väikseima hinnangu erinevus suurima tõepära meetodil leitud hinnangute vahel on 0,588, aga Bayesi meetodil leitud hinnangute maksimaalne erinevus on kõigest 0,070. Tegelik vapi saamise tõenäosus ehk parameetri väärtus oli $p = 0{,}55$ ning parameetrite hinnanguid pärast igat viset on näha joonisel 3.



Joonis 3: Hinnangud pärast tuhandet viset

Antud juhul eeljaotuse valik osutus küllaltki heaks ja kui on teada, et sobiliku eeljaotuse valikul koonduvad Bayesi meetodiga saadud hinnangud suurima tõepära hinnanguteks. Väheste andmete korral — kui on toimunud vaid mõned mündivisked — on Bayesi meetodil leitud hinnangud vähem volatiilsed.

3 Töös kasutatavad jaotused

Poliisi hinnastamine koosneb antud töös kahest komponendist, milleks on kahjusagedus ja kahju suurus. Portfelli simuleerimisel kasutati mõlema komponendi jaoks eri jaotusi.

3.1 Kahjude arvu jaotus

Kahjude arvu jaotust on tarvis selleks, et hinnata kindlustusportfelli kahjusagedust.

Kahjusageduseks loetakse portfellis jagatise $\frac{\text{portfelli kahjude arv}}{\text{portfelli kindlustusaastad}}$ keskväärtuseks, kus portfelli kahjude arv on portfellis olevate poliiside kogu kahjude summa ja portfelli kindlustusaastad on portfellis olevate kogu poliiside kestvused kokku summeeritud ja teisendatud aastatesse, st. et kahjusagedus on positiivne suurus ja jääb üldjuhul alla ühe. Kahjude arvude kirjeldamiseks on sobilikud diskreetsed jaotused. Antud töös on valitud jaotuseks negatiivne binoomjaotus, kus on kasutatud alternatiivset parameetrite määramist.

Negatiivsel binoomjaotusel on kaks parameetrit $\mu \in \mathbb{R}^+$ ja $\phi \in \mathbb{R}^+$ (Stan Development Team, 2021).

Tähistus $N \sim \mathrm{nb}(\mu, \phi)$.

Tõenäosusfunktsioon:

$$P(N = n) = C^n_{n+\phi-1} \cdot \left(\frac{\mu}{\mu + \phi}\right)^n \cdot \left(\frac{\phi}{\mu + \phi}\right)^\phi, \quad n = 0, 1, 2 \dots . \tag{5}$$

$E[N] = \mu$, $D[N] = \mu + \frac{\mu^2}{\phi}$, kus μ kirjeldab keskväärtust ja ϕ on ülehajuvust iseloomustav parameeter (kirjeldab liigset hajuvust võrreldes Poissoni jaotusega juhuslike suuruste hajuvusega). Negatiivset binoomjaotust saab parametriseerida ka teisiti, aga siia töösse on valitud bioloogide poolt eelistatud parametriseering, sest üks parameetritest langeb kokku kahjusagedusega. Negatiivne binoomjaotus on diskreetne jaotus, mis tekib siis, kui uuritava tunnuse jaotuseks on Poissoni jaotuste segu (mõnede lepingute korral on kahjude tekkimise sagedus veidi suurem kui teiste lepingute korral, nn Gamma-Poissoni segujaotus).

3.2 Kahju suuruste jaotus

Kahju suurusteks nimetatakse kahju realiseerumise korral seltsile tekkinud finantsilise väljamineku suurust eurodes. Kahju suuruste kirjeldamiseks sobivad pidevad jaotused ning antud töös on selleks valitud lognormaaljaotus.

Lognormaaljaotusel on kaks parameetrit: μ ja σ (> 0), mis esindavad jaotuse keskmist ja standardhälvet (Gray ja Pitts, 2012, lk 35-36).

Tähistus $X \sim \text{lognormal}(\mu, \sigma)$.

Tihedusfunktsioon:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\Big\{ -\frac{1}{2}\Big(\frac{\log x - \mu}{\sigma}\Big)^2 \Big\},\ x > 0. \tag{6}$$

Nimetus tuleneb sellest, et

$$X \sim \text{lognormal}(\mu, \sigma) \Leftrightarrow Y = \log X \sim \mathbf{N}(\mu, \sigma^2) \tag{7}$$

Lognormaaljaotuse keskväärtus ja dispersioon avalduvad kujul:

$E[X] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$, $D[X] = \exp(2\mu + \sigma^2) \cdot (e^{\sigma^2} - 1)$

Kus dispersioon ja keskväärtus on omavahel seotud:

$D[X] = (E[X])^2 \cdot (e^{\sigma^2} - 1)$

Parameetrite avaldamine keskväärtuse ja dispersiooni kaudu:

$$\sigma = \sqrt{\log(\frac{D[X]}{(E[X])^2 + 1})} \tag{8}$$

$$\mu = \log(E[X]) - \frac{\sigma^2}{2} \tag{9}$$

3.3 Eeljaotused

Kahjusageduse keskväärtuse eeljaotusena on kasutatud normaaljaotust:

$$\mu \sim N(\text{eksperdi hinnang kahjusagedusele}; 0{,}01^2).$$

Kahjusageduse ülehajuvusparameetri kirjeldamiseks on kasutatud ühtlast jaotust vahemikus (0;4), $\phi \sim U(0; 4)$.

Tegelik kahjusageduse keskväärtus jääb 95% tõenäosusega eksperthinnangule lähemale kui kahekordne eeljaotuse standardhälve. Kuna on usutav, et eksperthinnangu viga kahjusagedusele on väiksem kui 0,02, siis valisime eeljaotuse standardhälbeks 0,01.

Paljud autorid (David ja Jemna, 2015; Valecký, 2016) rõhutavad, et kahjusageduste dispersioon on suurem keskväärtusest, seega oleme valinud sellise eeljaotuse ülehajuvusparameetrile, mis vastab suurenenud dispersioonile.

Kahju suurust iseloomustava parameetri μ eeljaotusena on kasutatud normaaljaotust:

$$\mu \sim N(\mu_{ekspert}; 0{,}1^2).$$

Kahju suurust iseloomustava parameetri σ eeljaotusena on kasutatud lognormaaljaotust:

$$\sigma \sim lognormal(\sigma_{ekspert}; 0{,}2)$$

Eksperdil on lihtsam hinnata kahjude keskväärtust ja dispersiooni ning nende saadud hinnangute põhjal kasutades teisendusvalemeid (8, 9) on saadud $\sigma_{ekspert}, \mu_{ekspert}$ lognormaaljaotuse parameetritele.

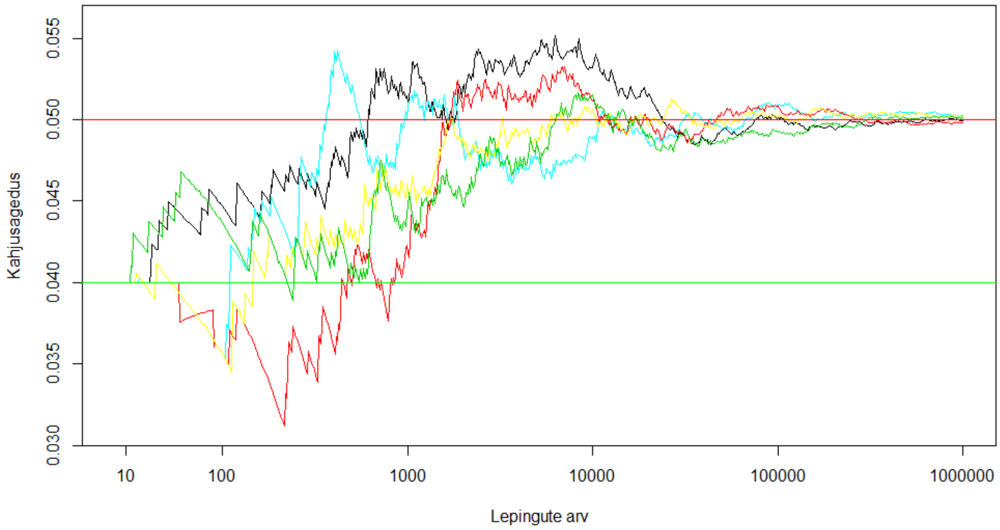
3.4 Mitmemõõtmeline eeljaotus

Kahe parameetri hindamisel on tarvis mitmemõõtmelist eeljaotust. Siin töös on modelleeritud kahjusagedust negatiivse binoomjaotuse abil, seega peame hindama keskväärtuse (μ) ja ülehajuvusparameetri (ϕ) väärtuseid ja kahju suurust on modelleeritud kasutades lognormaaljaotust, mille puhul tuleb samuti hinnata kahe parameetri väärtuseid (μ, σ). Lihtsuse mõttes on töösse valitud, et hinnatavad parameetrid on omavahel sõltumatud — mitmemõõtmeline eeljaotus saadakse mõlema parameetri eeljaotuste korrutamise tulemusena (Dekking *et al.*, 2005, lk 13).

4 Bayesi statistika meetodi koondumine ja tehtud vea maksumus

Kindlustusseltsis on hädavajalik täpne hinnastamine, et maksimeerida kasumit ja minimeerida riske. Preemiate hinnastamine toimub enamjaolt ajalooliste andmete põhjal, mis on eelnevate lepingute näol on müüdud. Väheste andmete korral võib olla hinnastamine ekslik ja seetõttu on tarvis võimalikult kiiresti jõuda hindadeni, kus risk on õigesti hinnastatud.

Simuleeritud on viis potentsiaalset kindlustusportfelli eeldades, et tegelikud kahjude sagedused käituvad kui negatiivse binoomjaotusega juhuslikud suurused. Kõigi nende portfellide jaoks on leitud Bayesi statistika abil hinnangud kahjusagedustele pärast igat lepingut. Saadud trajektorid on kujutatud joonisel 4. Rohelise vertikaalse joonega on toodud eksperdi hinnang kahjusagedusele, milleks on 0,04 ja punase vertikaalse joonega on toodud tegelik andmete genereerimisel kasutatud kahjusagedus (0,05). Visuaalselt on näha, et Bayesi meetodi abil leitud kahjusageduse hinnang koondub tegelikuks kahjusageduseks - nii nagu see teoreetiliselt peakski aset leidma (kui eeljaotus ei välista tegelikku kahjusagedust).



Joonis 4: Bayesi meetodi koondumine viie portfelli trajektoori põhjal

Jooniselt 4 on näha, et kahjusageduse hinnang muutub vähem volatiilseks kui lepingute arv suureneb ja hetkeline kahjusageduse hinnang ei jää eksperdi hinnangu ja tegeliku hinnangu vahele, vaid võib liikuda mõlemas suunas, seda seetõttu, et kahjude arvud on juhuslikud suurused.

Teada on, et ka suurima tõepära meetod koondub tegelikuks kahjusageduseks, kuid miks on siis tarvilik Bayesi statistikat kasutada? Sellele küsimusele leitakse vastus portfelli kahjude suuruste näol käesolevas peatükis.

Hindamaks võimalikku tehtavat viga, siis on mõistlik see ära siduda rahalise väärtusega ehk mida suurem viga, seda suurem rahaline kaotus. Siit tekib probleem, kuidas hinnata tehtud viga. Kahjusageduse tegelikkusest kõrgemaks hindamisel on ka poliisi hind selle võrra kõrgem. Kas see tähendab, et kindlustuselts teenib selle võrra ka rohkem kasumit?

Tegelikult see nii ei ole. Kui poliisi hinna tõstmine suurendaks kindlustusseltsi kasumeid, siis tõstaksid kindlustusseltsid kasumi maksimiseerimise eesmärgil poliisi hinnad lõpmatult suureks. Praktikas muutub liiga kõrgelt hinnastatud poliiside müük raskeks ja kulukaks. Kuidas siis arvestada ülehinnatud poliiside kulu kindlustusseltsile?

Üks võimalus hinnastada hinnangu viga on järgmine: kui kahjusagedus hinnatakse tegelikust madalamaks, siis loeme, et kindlustusselts peab oma vahenditest korvama tekkiva keskmise kahju. Kui aga kahjusagedus hinnatakse tegelikust suuremaks, siis arvestame, et kindlustusselts ei võida ega kaota (kahjusageduste ülehindamisest tekkinud potentsiaalse võidu tasakaalustavad tõusvad turustuskulud). Sellisel juhul tekitavad kahjusageduste alahinnangud kindlustusseltsile täiendavaid kulusid, kuid ülehinnangud ei ole kindlustusseltsi vaatenurgast ei kahjulikud ega ka kasulikud. Teisisõnu, kasutaksime järgmisel kujul olevat kaofunktsiooni (*loss function*) (Zacks, Lakshmikantham ja Tsokos, 1981, lk 301):

Ühepoolne kaofunktsioon:

$$\text{kadu} = \sum_j \max(p - h_j; 0) \tag{10}$$

Kus summeerimine toimub üle kõigi portfellis olevate lepingute, p näitab tegelikku kahjusagedust h_j näitab j. lepingu sõlmimise ajal leitud hinnangut kahjusagedusele.

Teiseks võimaluseks oleks lugeda ka kahjusageduse ülehinnastamine kindlustusseltsile kahjulikuks. Näiteks võime vea kulukuse kirjeldamisel kasutada kaofunktsioonina absoluutset viga ehk lugeda ülehinnangut kindlustusseltsile sama kahjulikuks kui alahinnangut:

Absoluutsel veal põhinev kaofunktsioon:

$$\text{kadu} = \sum_j |p - h_j| \tag{11}$$

Kumbki väljapakutud kaofunktsioon ei kirjelda täpselt ebatäpsete hinnangute rahalist mõju kindlustusseltsile, küll aga võiksid toodud kaofunktsioonid mingis mõttes välja tuua kaks kõige ekstreemsemat praktikas esineda võivat situatsiooni - tegelik kaofunktsioon võiks ehk olla midagi nende kahe funktsiooni vahepealset.

Iga simulatsiooni korral saame leida keskmise valehinnangutest tingitud kahju ehk kaofunktsiooni väärtuse. Kuna aga hinnangud on juhuslikud siis on ka kaofunktsiooni väärtused ehk poliiside valest hinnnastamisest tingitud kindlustusseltsi kulud juhuslikud.
Järgnevalt toome välja Bayesi statistika meetodi ja suurima tõepära meetodi erinevused portfelli summaarsete vigade näol.

4.1 Ühepoolne viga

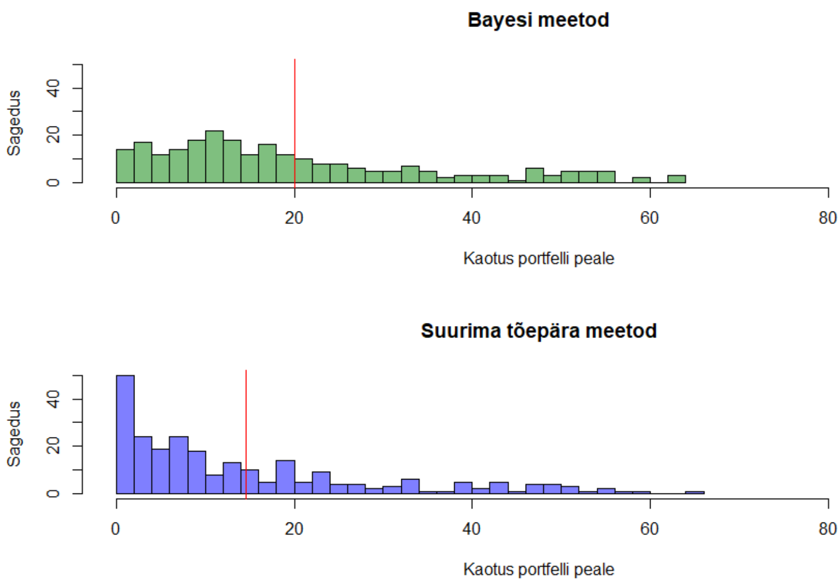
Ühepoolseks veaks loetakse, kui lepingu kahjusageduse hinnang on madalam kui tegelik kahjusagedus, kui kahjusagedus hinnang on kõrgem tegelikust kahjusagedusest, siis viga ei teki.

Portfellide ühepoolsete vigade arvutamisel kasutame valemit:

$$k_i = \sum_j \max(p - h_j; 0) \cdot m, \quad i = 1, \dots, 250, \ j = 1, \dots, n. \tag{12}$$

- k_i - vea suurus i-ndas simuleeritud portfellis;
- $m = 100$ - lepingute arv ühes sammus;
- $n = 100$ - sammude arv;
- $p = 0{,}05$ - kahjusageduse keskväärtus;
- h_j - hinnang kahjusagedusele.

Joonisel 5 on välja toodud histogrammid, kus kaotus portfelli peale on arvutatud võrdusega (12), punane joon märgib keskmist portfelli kahju. Jooniselt 5 on näha, et suurima tõepära meetodiga on keskmine kahju väiksem, kui Bayesi statistika meetodiga.



Joonis 5: Ühepoolse vea kahjud 250 erineva portfelli peale

Suurima tõepära meetodiga on nullile lähedaste kahjudega portfelle palju, kuna antud meetod on väga volatiilne kui andmeid on vähe ja kuna kahjude arv on genereeritud kahju tekkimise tõenäosusega 0,05, siis koondumine võtab aega ning kõik kahjusagedused, mis on ülehinnastatud, kajastatakse kui õigesti hinnastatud, tingitud võrdusest (12).

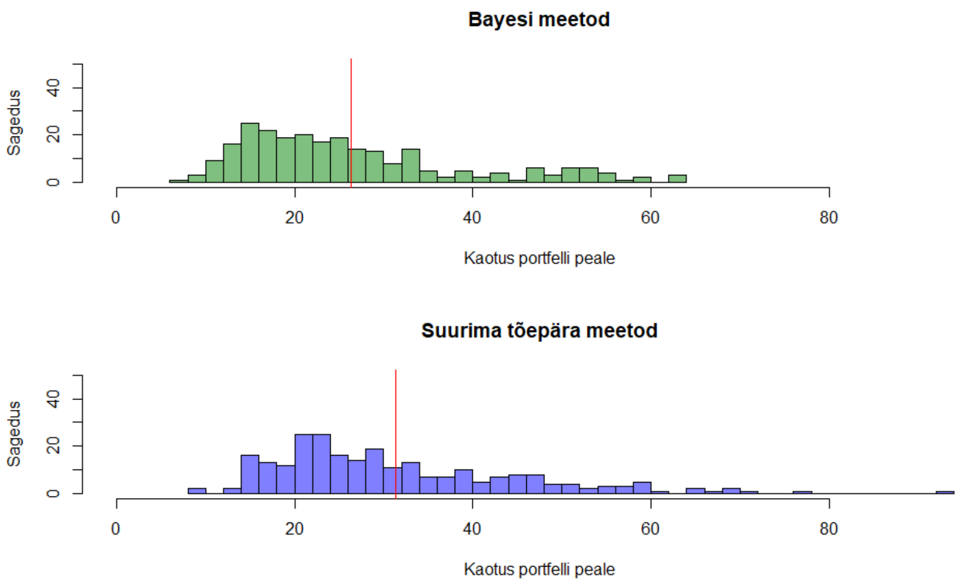
4.2 Absoluutne viga

Absoluutseks veaks loetakse, kui lepingu hinnatav kahjusagedus on madalam või kõrgem kui tegelik kahjusagedus.
Portfellide absoluutsete vigade arvutamisel kasutame valemit:

$$k_i = \sum_j |p - h_j| \cdot m, \quad i = 1, \ldots, 250, \; j = 1, \ldots, n. \tag{13}$$

Tähistused on samad, mis eelnevas alampeatükis.

Joonisel 6 on näidatud histogrammidena, kus simulatsioonid on arvutatud, kasutades võrdust (13).

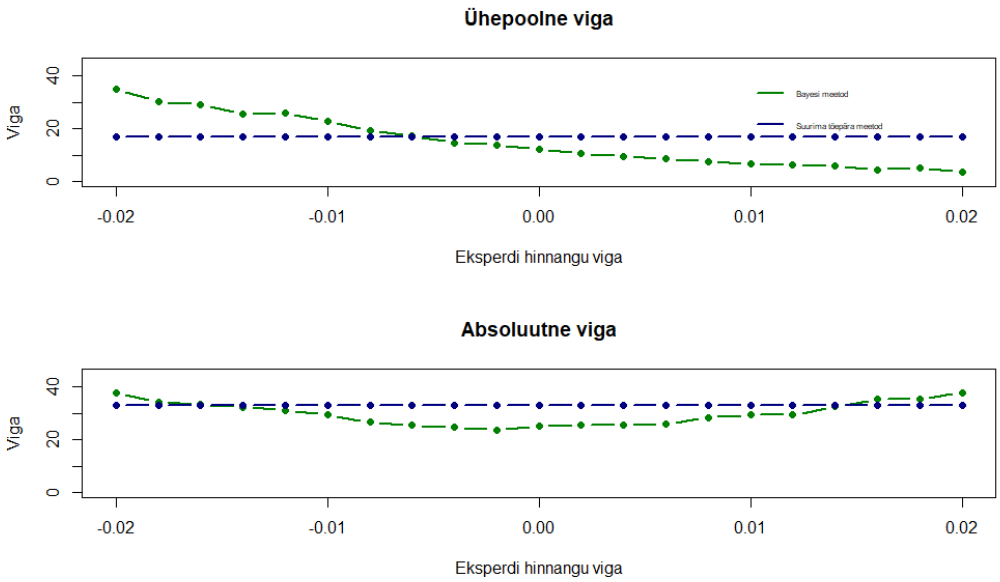


Joonis 6: Absoluutse vea kahjud 250 erineva portfelli peale

Jooniselt 6 on näha, et Bayesi statistika meetodi keskmine kaotus porfellide peale ja ka hajuvus on väiksem kui suurima tõepära meetodil. Keskmine portfelli kahju on absoluutse veaga suurem, kuna lisanduvad ka ülehinnastamise kahjud. Hajuvus on Bayesi statistika meetodiga väiksem, kuna võetakse arvesse eksperdi hinnangut ja see ei lase kahjusageduse hinnangul teha igal sammul suuri muutuseid.

4.3 Eksperdist tulenev viga

Ekspert võib eeljaotuse valikul kahjusageduse hinnanguga eksida ja kokkuvõttes rohkem kahju teha kui klassikalise lähenemisega saadud kahjusageduse hinnang, seda viga iseloomustab joonis 7.



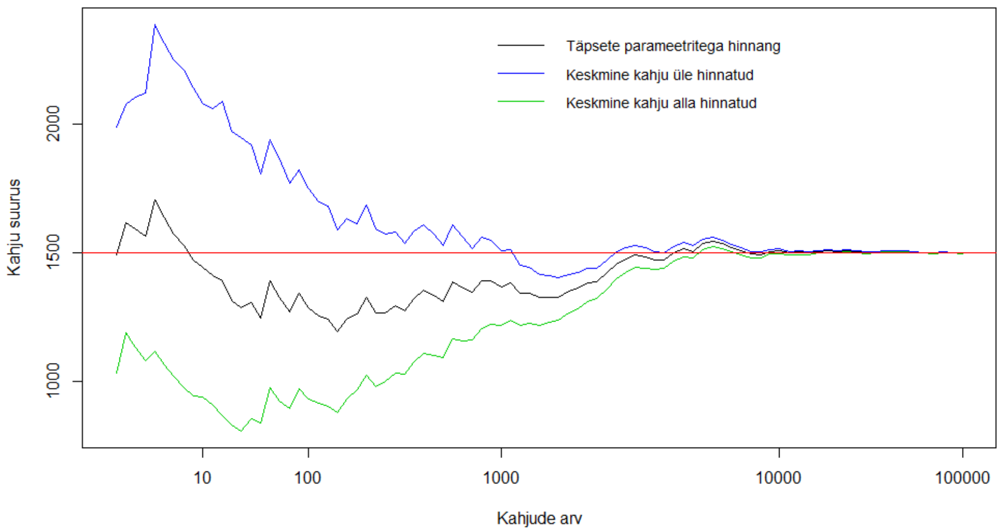
Joonis 7: Ühepoolne viga ja absoluutne viga, kus iga parameetri kohta on 250 trajektoori, igas on 10 000 lepingut 100 sammu tagant

Mida rohkem eksperdi kahjusageduse hinnang on mööda tegelikust kahjusagedust, seda suuremaks kujuneb viga. Joonisel 7 ühepoolse veaga muutub portfelli kahju Bayesi statistika meetodiga suuremaks suurima tõepära meetodiga saadud hinnangust, kui eksperdi hinnang kahjusagedusele on 12% ja rohkem madalam kui tegelik hinnang. Absoluutse vea puhul on Bayesi statistika meetodi kasutamine kuni 24%-se eksimuse korral kindlustusseltsile vähem kulukas, kui seda on suurima tõepära hinnang.

5 Kahju suurus

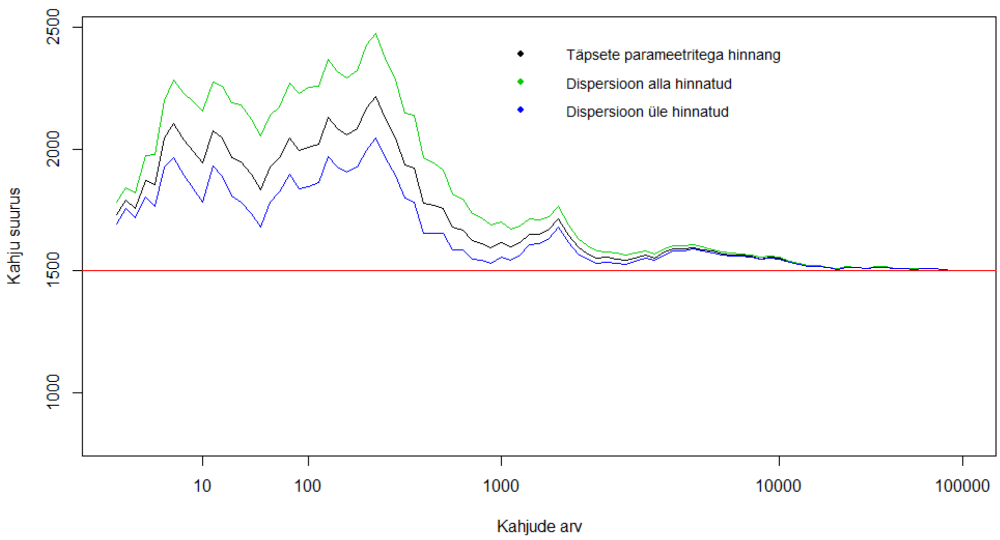
Kindlustuslepingu hinnaarvutusel mängib olulist rolli lisaks kahjusagedusele, mis annab informatsiooni kui tihti antud grupis leping põhjustab kahju, ka kahjude suurused. Kahjude suuruste simuleerimisprotsessis on kasutatud lognormaaljaotust. Kuna lognormaaljaotusel on kaks parameetrit, siis antakse ülevaade kahju suuruse hinnangute kohta Bayesi meetodi abil, kus ühel juhul on muudetud eksperdi hinnanguid keskmise kahju keskväärtusele ja teisel juhul keskmise kahju dispersioonile.

Jooniselt 8 on näha, et kui keskmine kahju on üle hinnatud ja hajuvuse parameeter on täpselt hinnatud, siis keskmise kahju suuruse hinnang on suurem kui tegelik kahju keskmine. Keskmise kahju alla hindamise korral on see väiksem, kui tegelik keskmine kahju. Hajuvuse parameetrid ei mõjuta nii tugevalt hinnanguid keskmisele kahjule.



Joonis 8: Kahju suuruse hindamine muutes lognormaalsejaotuse parameetreid

Joonistelt 8 ja 9 on näha, et Bayesi statistikaga leitud hinnanguid mõjutab jaotuse keskmise muutmine rohkem, kui hajuvusparameetri muutmine. Joonisel 9 hinnang kahju suuruse keskmisele on kõrgem kui tegelik kahju suurus ning on näha, et dispersiooni hinnang mõjutab hinnanguid, kuid mitte nii kardinaalselt kui keskmiste hinnangud.



Joonis 9: Kahju suuruse hindamine muutes lognormaalsejaotuse hajuvusparameetrit

5.1 Kahju suuruste hinnangute vead

Kahju suurusel on simuleeritud andmed lognormaalsest jaotusest keskmise kahju suurusega 1 500 ja dispersiooniga 7 000 000. Eksperdi antud hinnang kahju suurusele on 20% suurem ehk 1 800, kus dispersioon on õigesti hinnatud, kuna eelnevalt on selge, et dispersiooni mõjutus on lõpliku hinnangu saamisel väiksema tähtsusega.

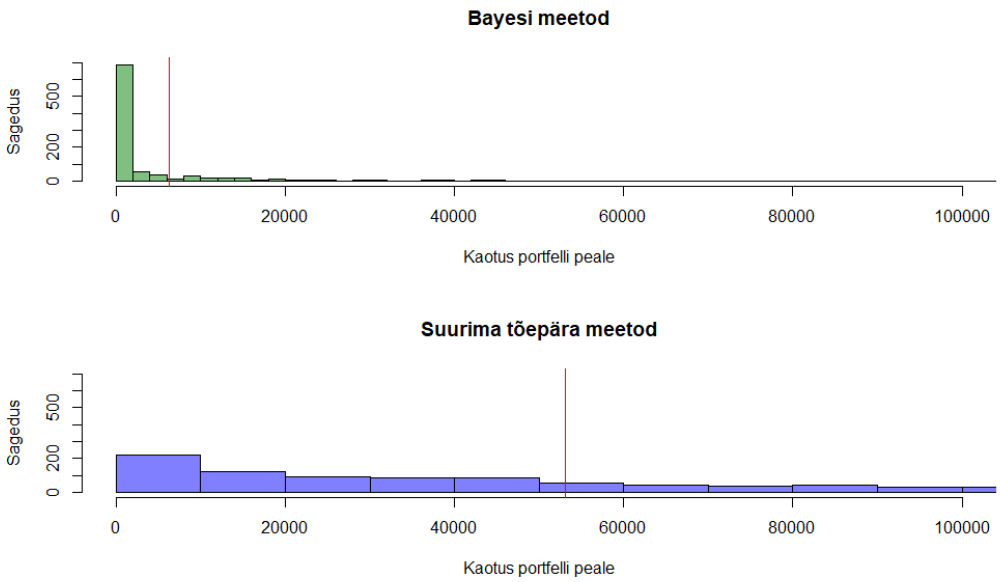
Vigade arvutus on sarnaselt valemitele (12) ja (13), kus tähistus on järgnev:

- k_i - vea suurus i-ndas simuleeritud portfellis;

- $i = 1, 2, \ldots, 1000$ - portfellid;
- $m = 10$ - kahjude arv ühes sammus;
- $n = 100$ - sammude arv;
- $p = 1\ 500$ - kahju suuruste keskväärtus;
- h_j - hinnang kahju suuruse keskväärtusele.

5.1.1 Ühepoolne viga

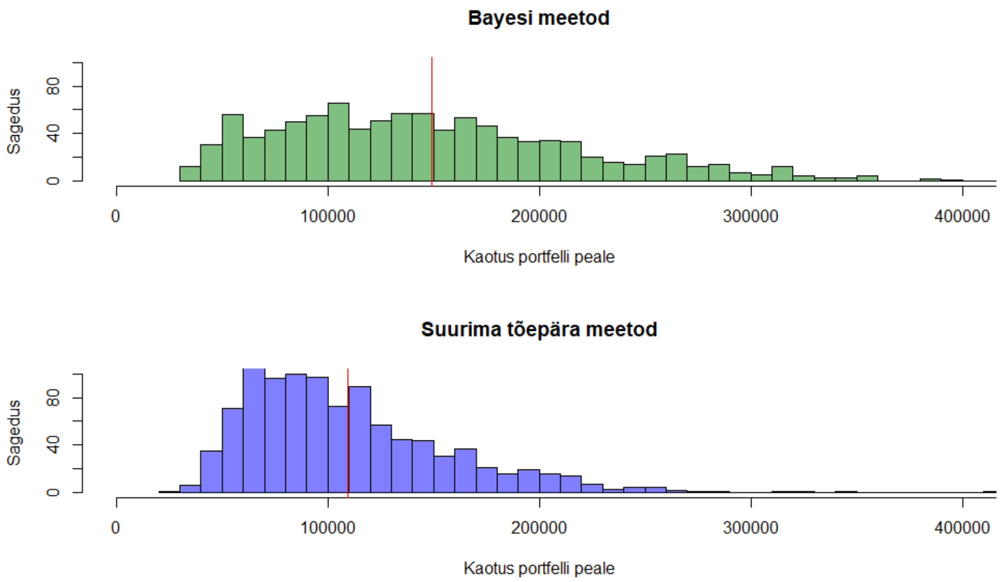
Joonisel 10 Bayesi statistikaga saadud hinnang on nullile lähedal, sest antud hinnang kahju suurusele on suurem, kui tegelik keskmine kahju ja seetõttu vea valem annulleerib need vead ära, kus hinnang on kõrgem kui keskmine kahju.



Joonis 10: Ühepoolsed vead keskmise kahju suurusele

5.1.2 Absoluutne viga

Absoluutse vea puhul on jooniselt 11 näha, et suurima tõepära hinnang keskmisele kahjule osutub paremaks, kui seda on Bayesi statistikaga saadud hinnang.



Joonis 11: Absoluutsed vead keskmise kahju suurusele

Kahjusageduse hindamise korral oli Bayesi statistikat kasutades kuni 20%-se tehtud eksimuse korral absoluutne viga väiksem suurima tõepära meetodi veast, kui seda on kahju suuruse hindamisel. Selleks on mitmeid põhjuseid, näiteks kahjude arvude korral on tegemist diskreetsete väärtustega andmetega, mis on nullidest ülepaisutatud ja seega koondumine õigeks väärtuseks suurima tõepära meetodi abil võtab kauem aega ning nende sammude jooksul tehtud viga suureneb, kuna on tegelikust väärtusest pikemalt möödas. Teiseks võimaluseks on valitud kehvem eeljaotus, mis ei õigusta ennast ära, sest see ei lase koonduda hinnangutel piisavalt täpseks, et vead kiiremini miinimumini viia.

6 Kindlustuslepingu hind

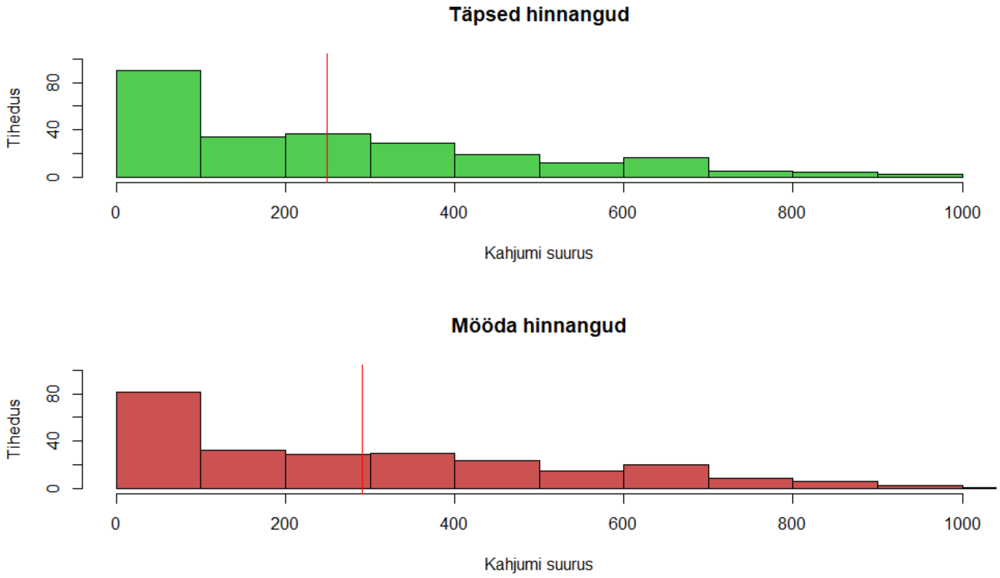
Pärast kahjusageduse ja kahju suuruse hindamist on vajalik teada mis võiks olla müüdava lepingu hind (lisaks tulevad ka kindlustusseltsi püsikulud ja kasumimarginaal, mis antud kontekstis ei ole oluline):

$$\text{kahju} = \text{kahjusagedus} \times \text{keskmine kahju suurus} \tag{14}$$

Klassikalises riskimudelis eeldatakse, et kahjusagedused ja kahju suurused on sõltumatud (Grandell, 1991, lk 1). Lepingu hinnastame keskmise kahju (14) järgi, mis on neile klientidele sama, kes kuuluvad saadaval olevate tunnuste poolest homogeensesse gruppi.

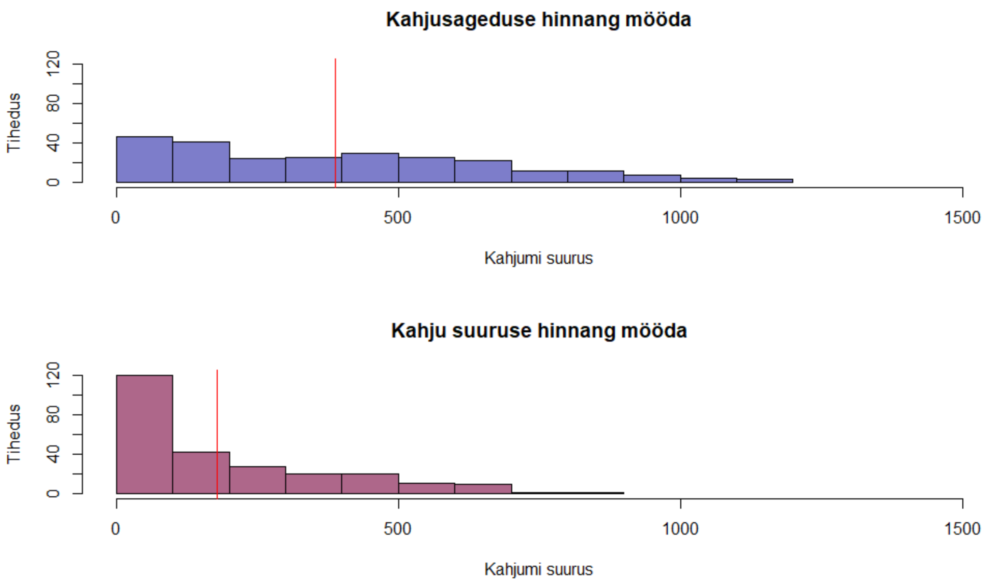
Joonistel 12 - 15 on kasutatud 250 erinevat trajektoori, kus on 10 000 saja sammu tagant, kus igal sammul arvutati uued hinnangud kahjusagedusele ja kahju suurusele (kõigi 250 trajektoori kahjude arvu ja suuruste andmed on erinevate hinnangute arvutamistel sama).

Joonistel 12 ja 13, saadud kahjumi suurused on saadud kaofunktsiooni (10) põhjal. Joonisel 12 on näha, kus nii kahjusagedus kui ka kahju suurus on eksperdi poolt mööda hinnatud, aga seda erinevates suundades (kahjusagedusele on antud madalam hinnang ja kahju suurusele kõrgem hinnang).



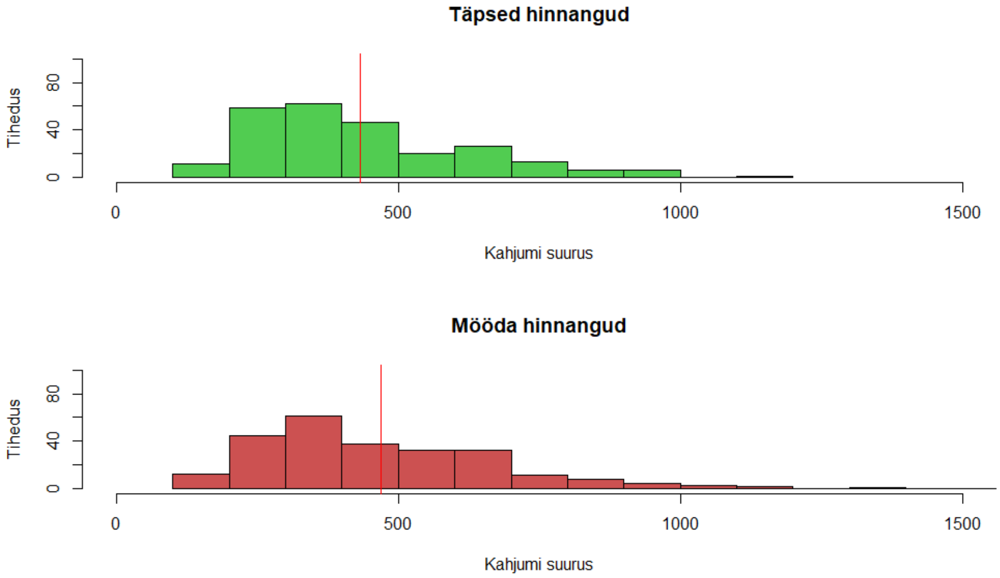
Joonis 12: Ühepoolne viga portfellide kahjude suurusestele (1)

Joonisel 13 kahju suuruse hinnang on hinnatud üle, seega on mõistetav, et ühepoolne viga on portfelli rahalistes kaotustes väiksem, kui mööda hinnatud kahjusageduse korral, mis on hinnatud alla tegeliku kahjusagedust.



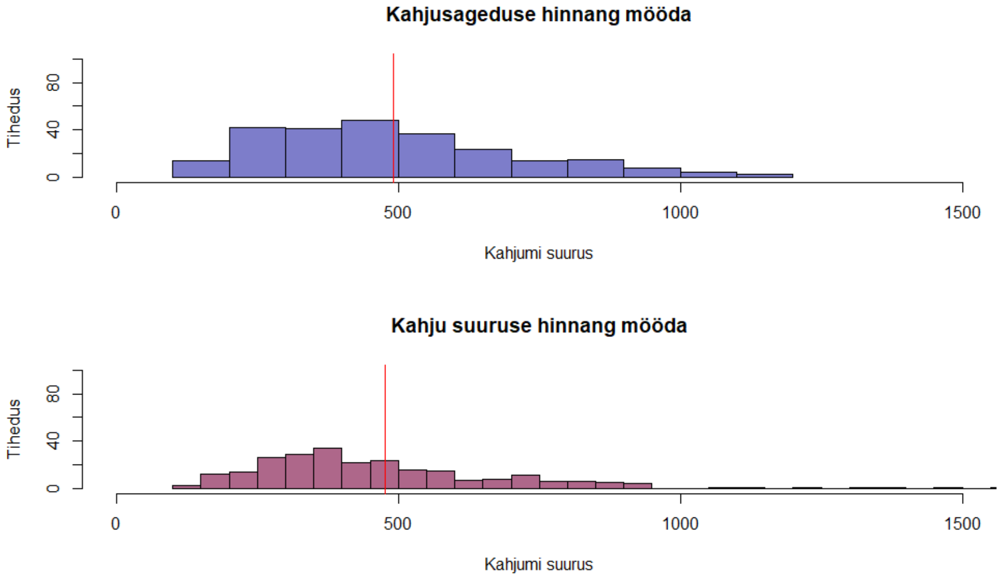
Joonis 13: Ühepoolne viga portfellide kahjude suurusestele (2)

Joonistel 14 ja 15 saadud kahjumi suurused on saadud kaofunktsiooni (11) põhjal. Ühepoolsetest vigadest saadud keskmine kahjumi suurus portfellidele, mis on märgitud punase vertikaalse joonega joonistel 12 ja 13 on väiksemad, kui absoluutsetest vigadest saadud keskmise kahjumi suurused portfellidele, mis on märgitud joonistele 14 ja 15, samuti punase vertikaalse joonega.



Joonis 14: Absoluutne viga portfellide kahjude suurustele (1)

Joonisel 14 on näha, et täpne hinnang on parem, kui üle hinnatud kahju suurus ja alla hinnatud kahjusagedus, kuid erisuunalised vead tasakaalustavad teineteist siiski suuresti ja kindlustuse kahjumi suuruste jaotus on väga lähedal sellele, mida võiksime näha õigete eksperthinnangute korral. See viitab sellele, et üllatavalt täpne tuleb tegelik kindlustuslepingu hind, kui on tehtud kaks vastassuunalist viga võrreldes joonisega 15, kus ühe eksperdi hinnangu vea korral on müüdavatest lepingutest tulenevad kahjud suuremad.

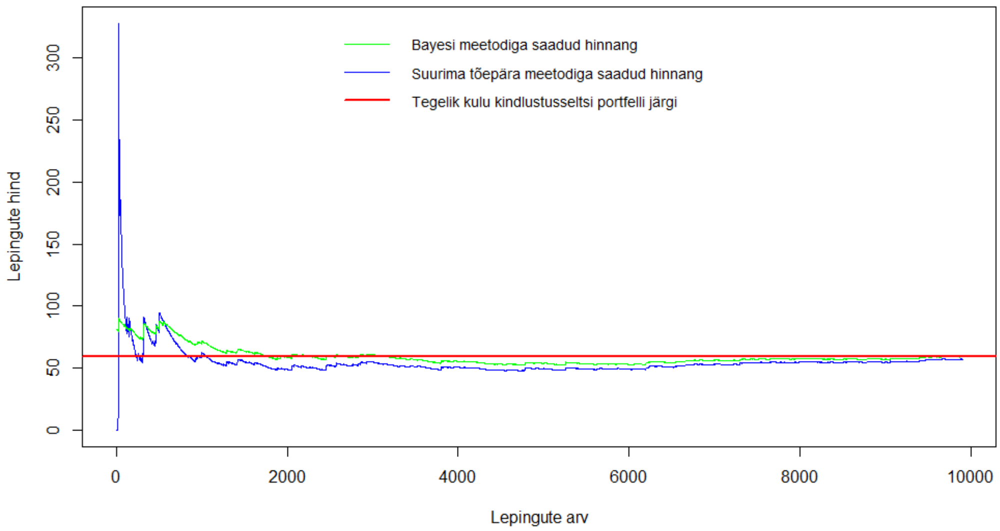


Joonis 15: Absoluutne viga portfellide kahjude suurustele (2)

Joonisel 15 on kahjusageduse mööda hindamine kahjumit väiksemana hoidnud, see tähendab, et kui punkthinnangud on sama palju erinevad mõlemal hinnangul, siis olulist rolli mängib eeljaotuse valik, mistõttu kahju suurus eeljaotuse valikule oleks tulnud paremini määrata. Juhul kui tehakse eksperdi hinnanguga viga mõlema parameetri puhul ja need on vastassuunalised, siis need tasakaalustavad teineteist, kuid mida täpsemad on eeljaotused kahjusagedusele ja kahju suurusele, seda väiksem on potentsiaalne kahjum suurus portfellis.

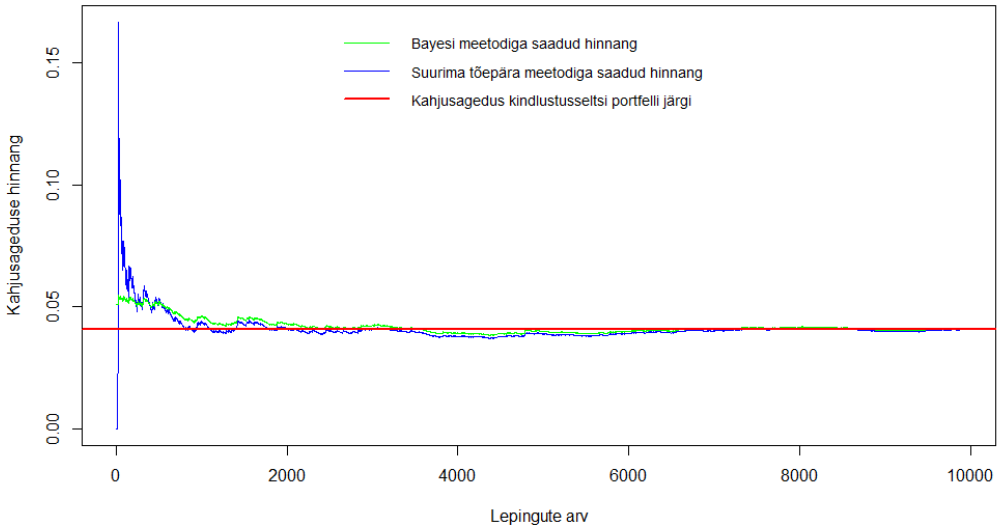
7 Päriseluline näide

Päriseluliseks näiteks toome liikluskindlustuse andmed, kus on teada 2018-2019. aasta sõidukite ühe aasta pikkuste poliiside arv ja tehtud kahjude arv koos kahju suurusega. Andmetes on kahte marki autod: BMW (1371 lepingut) ja Volkswagen (9900 lepingut). Demonstreerimise võtmes võetakse teatavaks BMW andmed ja hakatakse nende andmete põhjal hindama Volkswageni kindlustuslepingu hindu. Eeljaotuse tarvis on valitud BMW põhjal hinnang kahjusagedusele 0,0511 ja hinnang keskmisele kahjule 1619,16EURi.

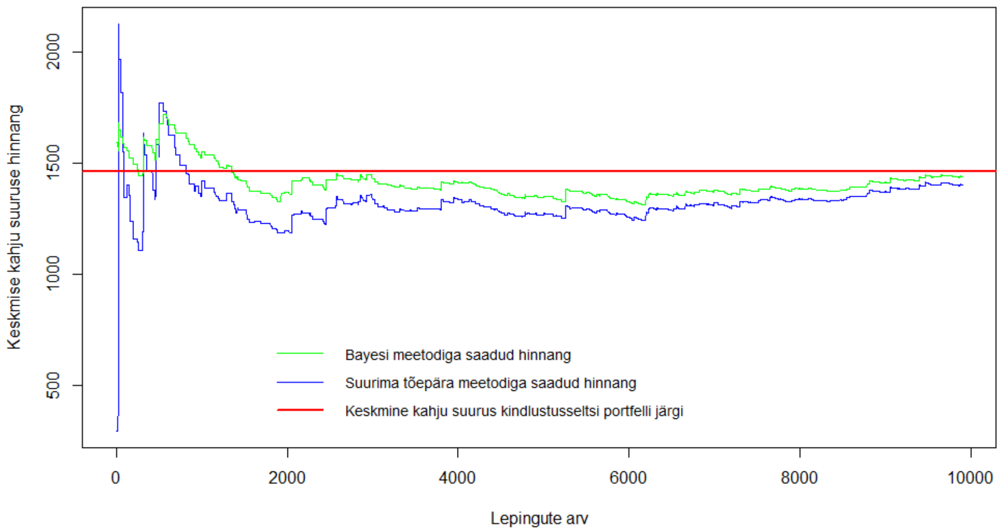


Joonis 16: Lepingute hinna muutus kahe erineva meetoditega

Joonisel 16 on näha, et suurima tõepära hinnang on väga volatiilne, kui lepinguid on vähe. Lepingu hindade komponendid on eraldi toodud joonistel 17 ja 18, kus suurima tõepära meetodil saadud hinnangud on mõlemal joonisel väga muutlikud.



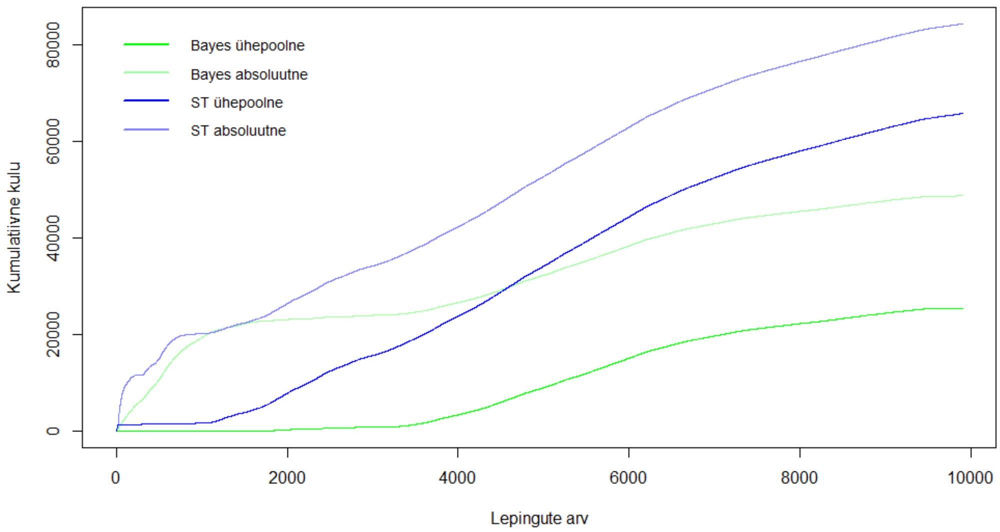
Joonis 17: Kahjusageduse muutus kahe erineva meetoditega



Joonis 18: Kahju suuruse muutus kahe erineva meetoditega

Joonisel 19 ühepoolse vea põhjal on Bayesi meetodiga saadud hinnangute põhjal arvutatud hindade kumulatiivne kahju väiksem kui suurima tõepära meetodi hinnangute põhjal. Absoluutse vea põhjal on suurima tõepära meetodil saadud kumulatiivse kulu summana suurem, võrreldes Bayesi meetodiga.

Portfelli täpseks maksumuseks on antud andmete põhjal 589 161 EURi. Ühepoolse vea õppimise kuludeks saadakse Bayesi meetodi abil 25 483 EURi (see on 4% kogu portfelli hinnast) ja suurima tõepära meetodiga 65 799 EURi (see on 11% kogu portfelli hinnast). Absoluutse vea maksumusteks oleks vastavalt 48 764 EURi (see on 8% kogu portfelli hinnast) ja 84 383 EURi (see on 14% kogu portfelli hinnast). See tähendab, et antud näite põhjal on Bayesi meetod ca 6% parem kui suurima tõepära meetod, st. Bayesi meetodi valikul võrreldes suurima tõepära meetodiga jääks alles 35 367 EURi.



Joonis 19: Hinnastamisest tulenev kumulatiivne kulu

Bayesi meetodi kasuks räägib antud juhul andmete vähesus ja kahju suuruse kirjeldamiseks valitud eeljaotus ning tegelikust keskmisest kahjust kõrgem valitud kahju suurus. Suurima tõepära meetodil on lepingu hinnaks algusjärgus 0 ja tõuseb kuni 328 EURini, mis on väheusutav stsenaarium praktikas tekkima.

Kokkuvõte

Antud töös on vaadeldud, millist kasu võiks saada Bayesi meetodi rakendamisest kindlustuses esilekerkivate andmete analüüsimisel — ja milliseid probleeme antud meetodi rakendamisel võib esineda.

Kui kasutada Bayesi statistika abil leitud hinnangud suurima tõepära meetodil leitud hinnangute asemel siis võiksid kindlustusseltsi kulud ja võimalikud kaotused väheneda — vähemalt uute kindlustustoodete puhul, kus sõlmitud lepingute ja esinenud kahjude arv on väike. Kui kasutatava informatsiooni hulk suureneb, siis Bayesi statistika abil leitud hinnangud lähenevad suurima tõepära meetodil leitud hinnangutele. Kuna Bayesi meetodi kasutamine on sageli arvutusmahukam ja keerulisem kui suurima tõepära meetodi kasutamine, siis hiljem on mõistlik lepingute hinnastamisel üle minna klassikalise statistika töövõtete peale.

Bayesi statistika on võrdlemisi hea meetod, kui on eelnevaid teadmisi sarnaste andmete kohta. Oluliseks osaks on oskus valida eeljaotus, sest vale eeljaotuse ja ka parameetrite valik võib teha lõppkokkuvõttes rohkem kahju kui kasu.
Kasutades paremuse hindamiseks kahte erinevat kaofunktsiooni, töötas Bayesi meetod paremini nii simulatsioonide põhjal otsustades kui ka pärisandmete näite põhjal (mille puhul mitmed töös tehtud eeldused päris täpselt paika ei pidanud), kui suurima tõepära meetod.

Kasutatud allikad

David, Mihaela ja Danut Jemna (juuli 2015). "Modeling the Frequency of Auto Insurance Claims by Means of Poisson and Negative Binomial Models". *Analele Stiintifice ale Universitatii Al I Cuza din Iasi - Sectiunea Stiinte Economice* 62, lk. 151–168. DOI: 10.1515/aicue-2015-0011.

Dekking, F.M., C. Kraaikamp, H.P. Lopuhaä ja L.E. Meester (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Texts in Statistics. Springer.

Grandell, Jan (1991). *Aspects of Risk Theory*. 1. väljaanne. Springer Series in Statistics. Springer-Verlag New York.

Gray, Roger J. ja Susan M. Pitts (2012). *Risk Modelling in General Insurance: From Principles to Practice*. 1. väljaanne. International Series on Actuarial Science. Cambridge University Press.

Pärna, Kalev (2013). *Tõenäosusteooria algkursus*. 1. väljaanne. Tartu Ülikooli Kirjastus.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Stan Development Team (2021). *Stan Modeling Language Users Guide and Reference Manual*. Stan Development Team. URL: https://mc-stan.org/docs/2_26/functions-reference/nbalt.html.

Valecký, Jiří (mai 2016). "Modelling Claim Frequency in Vehicle Insurance". *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* 64, lk. 683–689. DOI: 10.11118/actaun201664020683.

Zacks, Shelemyahu, V. Lakshmikantham ja C. P. Tsokos (1981). *Parametric Statistical Inference. Basic Theory and Modern Approaches*. International series in nonlinear mathematics. Elsevier Ltd, Pergamon Press.

LISA1

```
#Kristjan Sinikas
#Kindlustuslepingute hinnastamine Bayesi meetodi abil
#
#funktsioonid:
#sisendid:
#ise valitud parameetrid vektorina
#andmed (positiivne täisarv)
#annab andmetele tõenäosuse, et tuli antud parameetriga (log tõepära)
lnegbin <- function(par, andmed=andmed){
  par1=exp(par[1])
  par2=exp(par[2])
  tul=sum(log(dnbinom(andmed, mu=par1, size = par2)))
  return(tul)
}
#funktsioon tõepära maksimiseerimiseks
l_max <- function(par0, andmed){
  ajut=optim(par0, lnegbin, andmed=andmed, control=list(fnscale=-1))
  return(ajut$par)
}
#eeljaotus võtan eeljaotuseks kahjusagedus 0.04 ja size parameetriks teadmata hajuvuse
log_pr <- function(par){
  par1=exp(par[1])
  par2=exp(par[2])
  log(dnorm(par1, mean=0.04, sd=0.01))+log(dunif(par2, 0,4))}
#kuna log, siis eeljaotus + tõepära
log_koos <- function(par, andmed){
  log_pr(par)+lnegbin(par, andmed)
}
# Leiame järeljaotuse maksimumi:
#väljastab log väärtused, seega hiljem peab exp() võtma
l_max_jareljaotus <- function(par0, andmed){
  ajut=optim(par0, log_koos, andmed=andmed, control=list(fnscale=-1))
  return(ajut$par)
}
#funktsioon näitamaks bayesi statistika meetodi koondumist visualiseerimiseks
viis_trajektori <- function(ss){
  set.seed(ss)
  #et saada rohkem lepinguid joonisele korraga ehk vähem samme arvutamiseks
  asi=unique(round(exp(seq(0,log(1000000),length=1000))))
  n=1000000 #mitu kordust
  m=1 #mitu poliisi
  #andmete arvutamine
  andmed1 <- rnbinom(n*m,mu=0.05,size=2)
  #algväärtused optimiseerimiseks
  init1 <- log(0.04)
  init2 <- log(2)
  proov <- data.frame(matrix(nrow=length(asi),ncol = 2))
  start_time <- Sys.time()
  for (i in 1:length(asi)) {
    proov[i,] = l_max_jareljaotus(c(init1, init2), andmed=andmed1[1:(asi[i])])
    init1 <- proov[i,1]
    init2 <- proov[i,2]
  }
```

```
  end_time <- Sys.time()
  print(end_time - start_time)
  return(proov)
}
########################
#leiame viis trajektori#
########################
viistr1 <- viis_trajektori(450)  #Time difference of 8.110002 mins
viistr2 <- viis_trajektori(250)  #Time difference of 8.061486 mins
viistr3 <- viis_trajektori(777)  #Time difference of 8.186005 mins
viistr4 <- viis_trajektori(2020) #Time difference of 8.280056 mins
viistr5 <- viis_trajektori(10001)#Time difference of 10.60317 mins
#kanname viis trajektooor joonisele
par(mfrow=c(1,1))
plot(exp(viistr1[,1]), type="l",ylim=c(0.031,0.056), xlab="Lepingute arv", ylab="Kahjusagedus",
     xaxt="n")
axis(1,at=c(10,96,263,429,596,762),labels=c("10","100","1000","10000","100000","1000000"))
lines(exp(viistr2[,1]), type="l", col=5)
lines(exp(viistr3[,1]), type="l", col=10)
lines(exp(viistr4[,1]), type="l", col=7)
lines(exp(viistr5[,1]), type="l", col=3)
abline(h=0.04, col="green") #ekspert hinnang
abline(h=0.05, col="red") #tegelik
########################
# vead                 #
########################
####funktsioon, mis väljastab kaks tulpa
#esimeses on bayesi hinnang, teises on suurima tõepära
kahjum <- function(ss=1, n=100, m=1) {
  set.seed(ss)
  andmed1 <- rnbinom(n*m,mu=0.05,size=2)
  andmed <- andmed1[1:m]
  proov1 <- data.frame(matrix(nrow=n,ncol = 2))
  proov2 <- data.frame(matrix(nrow=n,ncol = 2))
  start_time <- Sys.time()
  for (i in 1:n) {
    proov1[i,] = l_max_jareljaotus(c(log(0.04), log(2)), andmed=andmed1[1:(i*m)])
    proov2[i,] = l_max(c(log(0.04), log(2)), andmed=andmed1[1:(i*m)])
  }
  end_time <- Sys.time()
  print(end_time - start_time)
  matr <- data.frame(matrix(nrow=n,ncol = 2))
  matr[,1]<-proov1[,1]
  matr[,2]<-proov2[,1]
  return(matr)
}
k = 250
n=100
m=100
#bayesiga
M <- matrix(nrow = n, ncol= k)
#suurim tõepära
N <- matrix(nrow = n, ncol= k)
for (i in 1:k) {
  abi = kahjum(ss=i,n=n,m=m)
  M[,i] = exp(abi[,1])
```

```
  N[,i] = exp(abi[,2])
  print(i)
}
#bayesi meetodiga ühepoolne viga
vek=M
vek[vek>0.05]<- 0.05
#potentsiaalne preemia
preemia_pot <- n*m*0.05
#preemia kätte
preemia_kat <- colSums(m*vek)
#kaotus
kaotus = preemia_kat-preemia_pot
#tõepära meetodiga ühepoolne viga
vek2=N
vek2[vek2>0.05]<- 0.05
#potentsiaalne preemia
preemia_pot2 <- n*m*0.05
#preemia kätte
preemia_kat2 <- colSums(m*vek2)
#kaotus
kaotus2 = preemia_kat2-preemia_pot2
par(mfrow=2:1)
h <- hist(-kaotus, breaks = 25,
          xlim = c(0,95),ylim=c(0,50),
          main="Bayesi meetod", ylab="Sagedus",xlab="Kaotus portfelli peale",
          col = rgb(0,0.5,0,0.5))
abline(v=mean(-kaotus), col="red")
h2 <- hist(-kaotus2, breaks = 25,
           xlim = c(0,95),ylim=c(0,50),
           main="Suurima tõepära meetod", ylab="Sagedus",xlab="Kaotus portfelli peale",
           col = rgb(0,0,1,0.5))
abline(v=mean(-kaotus2), col="red")

#bayesi keskmine absoluutne viga
vek=M
kaotus_matr <- abs(vek-0.05)
#kaotus
kaotus = colSums(m*kaotus_matr)
#suurima tõepära keskmine absoluutne viga
vek2=N
kaotus_matr2 <- abs(vek2-0.05)
#kaotus
kaotus2 = colSums(m*kaotus_matr2)
par(mfrow=2:1)
h <- hist(kaotus, breaks = 25,
          xlim = c(0,95),ylim=c(0,50),
          main="Bayesi meetod", ylab="Sagedus",xlab="Kaotus portfelli peale",
          col = rgb(0,0.5,0,0.5))
abline(v=mean(kaotus), col="red")
h2 <- hist(kaotus2, breaks = 35,
           xlim = c(0,95),ylim=c(0,50),
           main="Suurima tõepära meetod", ylab="Sagedus",xlab="Kaotus portfelli peale",
           col = rgb(0,0,1,0.5))
abline(v=mean(kaotus2), col="red")
##########################
#eksperdist sõltuvad vead#
```

```
########################
#värskendamaks eksperdi arvamust
log_pr1 <- function(par){
  par1=exp(par[1])
  par2=exp(par[2])
  log(dnorm(par1, mean=antud, sd=0.01))+log(dunif(par2, 0,4))
}
log_koos1 <- function(par, andmed){
  log_pr1(par)+lnegbin(par, andmed)
}
# Leiame järeljaotuse maksimumi:
l_max_jareljaotus1 <- function(par0, andmed){
  ajut=optim(par0, log_koos1, andmed=andmed, control=list(fnscale=-1))
  return(ajut$par)
}
kahjum1 <- function(n=100, m=1, andmed1, par1) {
  proov1 <- data.frame(matrix(nrow=n,ncol = 2))
  proov2 <- data.frame(matrix(nrow=n,ncol = 2))
  start_time <- Sys.time()
  for (i in 1:n) {
    proov1[i,] = l_max_jareljaotus1(c(log(par1), log(2)), andmed=andmed1[1:(i*m)])
    proov2[i,] = l_max(c(log(par1), log(2)), andmed=andmed1[1:(i*m)])
    print(i)
  }
  end_time <- Sys.time()
  print(end_time - start_time)
  return(cbind(proov1[,1],proov2[,1]))
}
trajed <- function(k=2,n=10,m=10, antud){
  print(antud)
  M1 <- matrix(nrow = n, ncol= k)
  N1 <- matrix(nrow = n, ncol= k)
  matria <- data.frame(matrix(nrow=1, ncol=2))
  for (i in 1:k) {
    antud <- antud
    andmed1 <- rnbinom(n*m,mu=0.05,size=2)
    abi = kahjum1(n=n,m=m,andmed1=andmed1,par1 = antud)
    M1[,i] = exp(abi[,1])
    N1[,i] = exp(abi[,2])
    print(i)
  }
  return(cbind(M1, N1)) #esimesed k tulpa on bayes, järgmised k tulpa on ST
}
set.seed(101)
k = 250
n=100
m=100
start_time<- Sys.time()
antud=0.03
korra1 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.032
korra2 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.034
korra3 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.036
korra4 <- trajed(k=k,n=n, m=m, antud=antud)
```

```
antud=0.038
korra5 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.04
korra6 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.042
korra7 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.044
korra8 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.046
korra9 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.048
korra10 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.05
korra11 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.052
korra12 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.054
korra13 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.056
korra14 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.058
korra15 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.06
korra16 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.062
korra17 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.064
korra18 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.066
korra19 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.068
korra20 <- trajed(k=k,n=n, m=m, antud=antud)
antud=0.07
korra21 <- trajed(k=k,n=n, m=m, antud=antud)
end_time<- Sys.time()
end_time-start_time
par(mfrow=c(1,1))
plot(korra1[,1], type="l", ylim=c(0,0.1))
lines(korra2[,1], col=2)
lines(korra3[,1], col=3)
lines(korra4[,1], col=4)
lines(korra5[,1], col=5)
lines(korra6[,1], col=6)
lines(korra7[,1], col=7)
lines(korra8[,1], col=8)
lines(korra9[,1], col=9)
lines(korra10[,1], col=10)
lines(korra11[,1], col=11)
lines(korra12[,1], col=12)
lines(korra13[,1], col=13)
lines(korra14[,1], col=14)
lines(korra15[,1], col=15)
lines(korra16[,1], col=16)
lines(korra17[,1], col=17)
lines(korra18[,1], col=18)
lines(korra19[,1], col=19)
```

```
lines(korra20[,1], col=20)
lines(korra21[,1], col=21)
k1 <- korra1[,1:k]
k2 <- korra2[,1:k]
k3 <- korra3[,1:k]
k4 <- korra4[,1:k]
k5 <- korra5[,1:k]
k6 <- korra6[,1:k]
k7 <- korra7[,1:k]
k8 <- korra8[,1:k]
k9 <- korra9[,1:k]
k10 <- korra10[,1:k]
k11 <- korra11[,1:k]
k12 <- korra12[,1:k]
k13 <- korra13[,1:k]
k14 <- korra14[,1:k]
k15 <- korra15[,1:k]
k16 <- korra16[,1:k]
k17 <- korra17[,1:k]
k18 <- korra18[,1:k]
k19 <- korra19[,1:k]
k20 <- korra20[,1:k]
k21 <- korra21[,1:k]
#ühepoolne viga
k1[k1>0.05]<- 0.05
k2[k2>0.05]<- 0.05
k3[k3>0.05]<- 0.05
k4[k4>0.05]<- 0.05
k5[k5>0.05]<- 0.05
k6[k6>0.05]<- 0.05
k7[k7>0.05]<- 0.05
k8[k8>0.05]<- 0.05
k9[k9>0.05]<- 0.05
k10[k10>0.05]<- 0.05
k11[k11>0.05]<- 0.05
k12[k12>0.05]<- 0.05
k13[k13>0.05]<- 0.05
k14[k14>0.05]<- 0.05
k15[k15>0.05]<- 0.05
k16[k16>0.05]<- 0.05
k17[k17>0.05]<- 0.05
k18[k18>0.05]<- 0.05
k19[k19>0.05]<- 0.05
k20[k20>0.05]<- 0.05
k21[k21>0.05]<- 0.05
#potentsiaalne preemia
preemia_pot <- n*m*0.05
#preemia kätte
preemia_kat1 <- colSums(m*k1)
preemia_kat2 <- colSums(m*k2)
preemia_kat3 <- colSums(m*k3)
preemia_kat4 <- colSums(m*k4)
preemia_kat5 <- colSums(m*k5)
preemia_kat6 <- colSums(m*k6)
preemia_kat7 <- colSums(m*k7)
preemia_kat8 <- colSums(m*k8)
```

```
preemia_kat9 <- colSums(m*k9)
preemia_kat10 <- colSums(m*k10)
preemia_kat11 <- colSums(m*k11)
preemia_kat12 <- colSums(m*k12)
preemia_kat13 <- colSums(m*k13)
preemia_kat14 <- colSums(m*k14)
preemia_kat15 <- colSums(m*k15)
preemia_kat16 <- colSums(m*k16)
preemia_kat17 <- colSums(m*k17)
preemia_kat18 <- colSums(m*k18)
preemia_kat19 <- colSums(m*k19)
preemia_kat20 <- colSums(m*k20)
preemia_kat21 <- colSums(m*k21)
#kaotus
kaotus1 = preemia_kat1-preemia_pot
kaotus2 = preemia_kat2-preemia_pot
kaotus3 = preemia_kat3-preemia_pot
kaotus4 = preemia_kat4-preemia_pot
kaotus5 = preemia_kat5-preemia_pot
kaotus6 = preemia_kat6-preemia_pot
kaotus7 = preemia_kat7-preemia_pot
kaotus8 = preemia_kat8-preemia_pot
kaotus9 = preemia_kat9-preemia_pot
kaotus10 = preemia_kat10-preemia_pot
kaotus11 = preemia_kat11-preemia_pot
kaotus12 = preemia_kat12-preemia_pot
kaotus13 = preemia_kat13-preemia_pot
kaotus14 = preemia_kat14-preemia_pot
kaotus15 = preemia_kat15-preemia_pot
kaotus16 = preemia_kat16-preemia_pot
kaotus17 = preemia_kat17-preemia_pot
kaotus18 = preemia_kat18-preemia_pot
kaotus19 = preemia_kat19-preemia_pot
kaotus20 = preemia_kat20-preemia_pot
kaotus21 = preemia_kat21-preemia_pot
kahjumidyp <- c(mean(-kaotus1),
                mean(-kaotus2),
                mean(-kaotus3),
                mean(-kaotus4),
                mean(-kaotus5),
                mean(-kaotus6),
                mean(-kaotus7),
                mean(-kaotus8),
                mean(-kaotus9),
                mean(-kaotus10),
                mean(-kaotus11),
                mean(-kaotus12),
                mean(-kaotus13),
                mean(-kaotus14),
                mean(-kaotus15),
                mean(-kaotus16),
                mean(-kaotus17),
                mean(-kaotus18),
                mean(-kaotus19),
                mean(-kaotus20),
                mean(-kaotus21))
```

```
#absoluutviga
k1 <- korra1[,1:k]
k2 <- korra2[,1:k]
k3 <- korra3[,1:k]
k4 <- korra4[,1:k]
k5 <- korra5[,1:k]
k6 <- korra6[,1:k]
k7 <- korra7[,1:k]
k8 <- korra8[,1:k]
k9 <- korra9[,1:k]
k10 <- korra10[,1:k]
k11 <- korra11[,1:k]
k12 <- korra12[,1:k]
k13 <- korra13[,1:k]
k14 <- korra14[,1:k]
k15 <- korra15[,1:k]
k16 <- korra16[,1:k]
k17 <- korra17[,1:k]
k18 <- korra18[,1:k]
k19 <- korra19[,1:k]
k20 <- korra20[,1:k]
k21 <- korra21[,1:k]
#kaotus
kaotus1 = colSums(m*abs(k1-0.05))
kaotus2 = colSums(m*abs(k2-0.05))
kaotus3 = colSums(m*abs(k3-0.05))
kaotus4 = colSums(m*abs(k4-0.05))
kaotus5 = colSums(m*abs(k5-0.05))
kaotus6 = colSums(m*abs(k6-0.05))
kaotus7 = colSums(m*abs(k7-0.05))
kaotus8 = colSums(m*abs(k8-0.05))
kaotus9 = colSums(m*abs(k9-0.05))
kaotus10 = colSums(m*abs(k10-0.05))
kaotus11 = colSums(m*abs(k11-0.05))
kaotus12 = colSums(m*abs(k12-0.05))
kaotus13 = colSums(m*abs(k13-0.05))
kaotus14 = colSums(m*abs(k14-0.05))
kaotus15 = colSums(m*abs(k15-0.05))
kaotus16 = colSums(m*abs(k16-0.05))
kaotus17 = colSums(m*abs(k17-0.05))
kaotus18 = colSums(m*abs(k18-0.05))
kaotus19 = colSums(m*abs(k19-0.05))
kaotus20 = colSums(m*abs(k20-0.05))
kaotus21 = colSums(m*abs(k21-0.05))
kahjumidabs <- -c(mean(-kaotus1),
                  mean(-kaotus2),
                  mean(-kaotus3),
                  mean(-kaotus4),
                  mean(-kaotus5),
                  mean(-kaotus6),
                  mean(-kaotus7),
                  mean(-kaotus8),
                  mean(-kaotus9),
                  mean(-kaotus10),
                  mean(-kaotus11),
                  mean(-kaotus12),
```

```
                  mean(-kaotus13),
                  mean(-kaotus14),
                  mean(-kaotus15),
                  mean(-kaotus16),
                  mean(-kaotus17),
                  mean(-kaotus18),
                  mean(-kaotus19),
                  mean(-kaotus20),
                  mean(-kaotus21))
#siit algab suurima tõepära
k1 <- korra1[,(k+1):(2*k)]
k2 <- korra2[,(k+1):(2*k)]
k3 <- korra3[,(k+1):(2*k)]
k4 <- korra4[,(k+1):(2*k)]
k5 <- korra5[,(k+1):(2*k)]
k6 <- korra6[,(k+1):(2*k)]
k7 <- korra7[,(k+1):(2*k)]
k8 <- korra8[,(k+1):(2*k)]
k9 <- korra9[,(k+1):(2*k)]
k10 <- korra10[,(k+1):(2*k)]
k11 <- korra11[,(k+1):(2*k)]
k12 <- korra12[,(k+1):(2*k)]
k13 <- korra13[,(k+1):(2*k)]
k14 <- korra14[,(k+1):(2*k)]
k15 <- korra15[,(k+1):(2*k)]
k16 <- korra16[,(k+1):(2*k)]
k17 <- korra17[,(k+1):(2*k)]
k18 <- korra18[,(k+1):(2*k)]
k19 <- korra19[,(k+1):(2*k)]
k20 <- korra20[,(k+1):(2*k)]
k21 <- korra21[,(k+1):(2*k)]
#ühepoolne viga
k1[k1>0.05]<- 0.05
k2[k2>0.05]<- 0.05
k3[k3>0.05]<- 0.05
k4[k4>0.05]<- 0.05
k5[k5>0.05]<- 0.05
k6[k6>0.05]<- 0.05
k7[k7>0.05]<- 0.05
k8[k8>0.05]<- 0.05
k9[k9>0.05]<- 0.05
k10[k10>0.05]<- 0.05
k11[k11>0.05]<- 0.05
k12[k12>0.05]<- 0.05
k13[k13>0.05]<- 0.05
k14[k14>0.05]<- 0.05
k15[k15>0.05]<- 0.05
k16[k16>0.05]<- 0.05
k17[k17>0.05]<- 0.05
k18[k18>0.05]<- 0.05
k19[k19>0.05]<- 0.05
k20[k20>0.05]<- 0.05
k21[k21>0.05]<- 0.05
#potentsiaalne preemia
preemia_pot <- n*m*0.05
#preemia kätte
```

```
preemia_kat1 <- colSums(m*k1)
preemia_kat2 <- colSums(m*k2)
preemia_kat3 <- colSums(m*k3)
preemia_kat4 <- colSums(m*k4)
preemia_kat5 <- colSums(m*k5)
preemia_kat6 <- colSums(m*k6)
preemia_kat7 <- colSums(m*k7)
preemia_kat8 <- colSums(m*k8)
preemia_kat9 <- colSums(m*k9)
preemia_kat10 <- colSums(m*k10)
preemia_kat11 <- colSums(m*k11)
preemia_kat12 <- colSums(m*k12)
preemia_kat13 <- colSums(m*k13)
preemia_kat14 <- colSums(m*k14)
preemia_kat15 <- colSums(m*k15)
preemia_kat16 <- colSums(m*k16)
preemia_kat17 <- colSums(m*k17)
preemia_kat18 <- colSums(m*k18)
preemia_kat19 <- colSums(m*k19)
preemia_kat20 <- colSums(m*k20)
preemia_kat21 <- colSums(m*k21)
#kaotus
kaotus1 = preemia_kat1-preemia_pot
kaotus2 = preemia_kat2-preemia_pot
kaotus3 = preemia_kat3-preemia_pot
kaotus4 = preemia_kat4-preemia_pot
kaotus5 = preemia_kat5-preemia_pot
kaotus6 = preemia_kat6-preemia_pot
kaotus7 = preemia_kat7-preemia_pot
kaotus8 = preemia_kat8-preemia_pot
kaotus9 = preemia_kat9-preemia_pot
kaotus10 = preemia_kat10-preemia_pot
kaotus11 = preemia_kat11-preemia_pot
kaotus12 = preemia_kat12-preemia_pot
kaotus13 = preemia_kat13-preemia_pot
kaotus14 = preemia_kat14-preemia_pot
kaotus15 = preemia_kat15-preemia_pot
kaotus16 = preemia_kat16-preemia_pot
kaotus17 = preemia_kat17-preemia_pot
kaotus18 = preemia_kat18-preemia_pot
kaotus19 = preemia_kat19-preemia_pot
kaotus20 = preemia_kat20-preemia_pot
kaotus21 = preemia_kat21-preemia_pot
kahjumidypst <- c(mean(-kaotus1),
                  mean(-kaotus2),
                  mean(-kaotus3),
                  mean(-kaotus4),
                  mean(-kaotus5),
                  mean(-kaotus6),
                  mean(-kaotus7),
                  mean(-kaotus8),
                  mean(-kaotus9),
                  mean(-kaotus10),
                  mean(-kaotus11),
                  mean(-kaotus12),
                  mean(-kaotus13),
```

```
                  mean(-kaotus14),
                  mean(-kaotus15),
                  mean(-kaotus16),
                  mean(-kaotus17),
                  mean(-kaotus18),
                  mean(-kaotus19),
                  mean(-kaotus20),
                  mean(-kaotus21))
#keskmine absoluutviga
k1 <- korra1[,(k+1):(2*k)]
k2 <- korra2[,(k+1):(2*k)]
k3 <- korra3[,(k+1):(2*k)]
k4 <- korra4[,(k+1):(2*k)]
k5 <- korra5[,(k+1):(2*k)]
k6 <- korra6[,(k+1):(2*k)]
k7 <- korra7[,(k+1):(2*k)]
k8 <- korra8[,(k+1):(2*k)]
k9 <- korra9[,(k+1):(2*k)]
k10 <- korra10[,(k+1):(2*k)]
k11 <- korra11[,(k+1):(2*k)]
k12 <- korra12[,(k+1):(2*k)]
k13 <- korra13[,(k+1):(2*k)]
k14 <- korra14[,(k+1):(2*k)]
k15 <- korra15[,(k+1):(2*k)]
k16 <- korra16[,(k+1):(2*k)]
k17 <- korra17[,(k+1):(2*k)]
k18 <- korra18[,(k+1):(2*k)]
k19 <- korra19[,(k+1):(2*k)]
k20 <- korra20[,(k+1):(2*k)]
k21 <- korra21[,(k+1):(2*k)]
#kaotus
kaotus1 = colSums(m*abs(k1-0.05))
kaotus2 = colSums(m*abs(k2-0.05))
kaotus3 = colSums(m*abs(k3-0.05))
kaotus4 = colSums(m*abs(k4-0.05))
kaotus5 = colSums(m*abs(k5-0.05))
kaotus6 = colSums(m*abs(k6-0.05))
kaotus7 = colSums(m*abs(k7-0.05))
kaotus8 = colSums(m*abs(k8-0.05))
kaotus9 = colSums(m*abs(k9-0.05))
kaotus10 = colSums(m*abs(k10-0.05))
kaotus11 = colSums(m*abs(k11-0.05))
kaotus12 = colSums(m*abs(k12-0.05))
kaotus13 = colSums(m*abs(k13-0.05))
kaotus14 = colSums(m*abs(k14-0.05))
kaotus15 = colSums(m*abs(k15-0.05))
kaotus16 = colSums(m*abs(k16-0.05))
kaotus17 = colSums(m*abs(k17-0.05))
kaotus18 = colSums(m*abs(k18-0.05))
kaotus19 = colSums(m*abs(k19-0.05))
kaotus20 = colSums(m*abs(k20-0.05))
kaotus21 = colSums(m*abs(k21-0.05))
kahjumidabsst <- -c(mean(-kaotus1),
                    mean(-kaotus2),
                    mean(-kaotus3),
                    mean(-kaotus4),
```

```
                    mean(-kaotus5),
                    mean(-kaotus6),
                    mean(-kaotus7),
                    mean(-kaotus8),
                    mean(-kaotus9),
                    mean(-kaotus10),
                    mean(-kaotus11),
                    mean(-kaotus12),
                    mean(-kaotus13),
                    mean(-kaotus14),
                    mean(-kaotus15),
                    mean(-kaotus16),
                    mean(-kaotus17),
                    mean(-kaotus18),
                    mean(-kaotus19),
                    mean(-kaotus20),
                    mean(-kaotus21))
par(mfrow=c(2,1))
plot(seq(0.03,0.07,0.002)-0.05,kahjumidyp,
     type="b",pch=16,
     lwd=2,col=rgb(0,0.5,0,1),
     ylab="Viga", xlab="Eksperdi hinnangu viga", ylim = c(0,45), main="Ühepoolne viga")
legend("topright", col = c(rgb(0,0.5,0,1),rgb(0,0,0.5,1)),
       legend=c("Bayesi meetod","Suurima tõepära meetod"),
       lty=1,lwd=2.5,cex = 0.5,box.lty=0, inset = 0.02)
lines(seq(0.03,0.07,0.002)-0.05,rep(mean(kahjumidypst),21),
      type="b",pch=16,
      lwd=2,col=rgb(0,0,0.5,1),
      ylab="Viga", xlab="Eksperdi hinnangu viga", main="Ühepoolne viga")
#absoluutne viga
plot(seq(0.03,0.07,0.002)-0.05,kahjumidabs,
     type="b",pch=16,
     lwd=2,col=rgb(0,0.5,0,1),
     ylab="Viga", xlab="Eksperdi hinnangu viga", ylim=c(0,45), main="Absoluutne viga")
lines(seq(0.03,0.07,0.002)-0.05,rep(mean(kahju

                                          midabsst),21),
      type="b",pch=16,
      lwd=2,col=rgb(0,0,0.5,1),
      ylab="Viga", xlab="Eksperdi hinnangu viga", main="Keskmine absoluutne viga")
##arvutamaks parameetreid
hin_param <- function(kk,ha){#keskmine kahju ja hajuvus
  #sdlog parameeter
  sigma <- sqrt(log(ha/(kk**2)+1))
  #meanlogparameeter
  mu=log(kk)-sigma**2/2
  return(c(mu, sigma))
}
param_hin <- function(mu, sigma){
  kesk <- exp(mu+sigma**2/2)
  haj <- kesk**2*(exp(sigma**2)-1)
  return(c(kesk, haj))
}
llognorm <- function(par, andmed=andmed){
  par1=exp(par[1])
  par2=exp(par[2])
```

```
  tul=sum(log(dlnorm(andmed, meanlog=par1, sdlog = par2)))
  return(tul)
}
l_maxi <- function(par0, andmed){
  ajut=optim(par0, llognorm, andmed=andmed, control=list(fnscale=-1))
  return(ajut$par)
}
log_prior <- function(par){
  par1=exp(par[1])
  par2=exp(par[2])
  #siin peab olema kahe parameetri summa, kus on kaetud parameetrid antud tihedusfunktsioonidega
  log(dnorm(par1, mean=mu1,sd=0.1))+log(dlnorm(par2,hin_param(sigma1,0)[1],0.2))
}
log_koosi <- function(par, andmed){
  log_prior(par)+llognorm(par, andmed)
}
# Leiame järeljaotuse maksimumi:
l_max_jareljaotusi <- function(par0, andmed){
  ajut=optim(par0, log_koosi, andmed=andmed, control=list(fnscale=-1))
  return(ajut$par)
}
kahjum <- function(n=100, m=1) {
  andmed1 <- rlnorm(n*m,meanlog=param[1],sdlog=param[2])
  init1 <- log(param[1])
  init2 <- log(param[1])
  prooviks1 <- data.frame(matrix(nrow=n,ncol = 2))
  prooviks2 <- data.frame(matrix(nrow=n,ncol = 2))
  start_time <- Sys.time()
  for (i in 1:n) {
    prooviks1[i,] = l_max_jareljaotusi(c(init1, init2), andmed=andmed1[1:(i*m)])
    prooviks2[i,] = l_maxi(c(init1, init2), andmed=andmed1[1:(i*m)])
    init1 <- prooviks1[i,1]
    init2 <- prooviks1[i,2]
  }
  end_time <- Sys.time()
  print(end_time - start_time)
  matr <- data.frame(matrix(nrow=n,ncol = 4))
  matr[,1]<-prooviks1[,1]#bay meetod mu
  matr[,2]<-prooviks1[,2]#bay sigma
  matr[,3]<-prooviks2[,1]#tõepära mu
  matr[,4]<-prooviks2[,2]#tõepära sigma
  return(matr)
}
kahjum_lisa <- function(ss=1, n=100, m=1) {
  set.seed(ss)
  andmed1 <- rlnorm(n*m,meanlog=param[1],sdlog=param[2])
  init1 <- log(param[1])
  init2 <- log(param[1])
  prooviks1 <- data.frame(matrix(nrow=length(asi),ncol = 2))
  prooviks2 <- data.frame(matrix(nrow=length(asi),ncol = 2))
  start_time <- Sys.time()
  for (i in 1:length(asi)) {
    prooviks1[i,] = l_max_jareljaotusi(c(init1, init2), andmed=andmed1[1:asi[i]])
    prooviks2[i,] = l_maxi(c(init1, init2), andmed=andmed1[1:asi[i]])
    init1 <- prooviks1[i,1]
    init2 <- prooviks1[i,2]
```

```
    print(i)
  }
  end_time <- Sys.time()
  print(end_time - start_time)
  matr <- data.frame(matrix(nrow=length(asi),ncol = 4))
  matr[,1]<-prooviks1[,1]#bay meetod mu
  matr[,2]<-prooviks1[,2]#bay sigma
  matr[,3]<-prooviks2[,1]#tõepära mu
  matr[,4]<-prooviks2[,2]#tõepära sigma
  return(matr)
}
######################
#muudetud parameetrid#
######################
#võtame arvesse, et meil on keskmine kahju 1500 ja dispersioon 7* 10^6
set.seed(420)
n=100000 #mitu kordust
m=1 #mitu kahjuga poliisi
k=3#mitu trajektori
ekk <- c(1500*1,1500*0.75,1500*1.25)
eha <- c(7000000*1,7000000*1,7000000*1)
asi = unique(round(exp(seq(0,log(n),length=100))))
prooviks <- matrix(ncol=k,nrow=length(asi))
for (i in 1:k) { #igas trajektooris on eri hinnangud
  kk=ekk[i]
  ha=eha[i]
  mu1 <- hin_param(kk,ha)[1]
  sigma1 <- hin_param(kk,ha)[2]
  param = hin_param(1500,7000000)
  abi = kahjum_lisa(ss=12,n=n,m=m)
  kes <- exp(abi[,1])
  sig <- exp(abi[,2])
  print(dim(abi))
  for (j in 1:length(asi)) {
    prooviks[j,i] = exp(kes[j]+sig[j]**2/2)
  }
}
par(mfrow=c(1,1))
plot(prooviks[,1],type="l",
     ylim=c(min(prooviks),max(prooviks)),
     xaxt="n",
     xlab="Kahjude arv",
     ylab="Kahju suurus"
)
legend("topright",
       legend=c("Täpsete parameetritega hinnang","Keskmine kahju üle hinnatud","Keskmine kahju alla hinnatud"
       ), inset=0.02, lty=1,
       col=c(col=1,col=4,col=3), box.lty=0, lwd=1, cex=0.85)#,pch=16)
axis(1,at=c(10,21,41,70,89),labels=c("10","100","1000","10000","100000"))
for (m in 2:3) {
  lines(prooviks[,m], col=m+1)
}
abline(h=1500,col="red")
#dispersioonide tarvis
set.seed(420)
n=100000 #mitu kordust
```

```
m=1 #mitu kahjuga poliisi
k=3#mitu trajektori
ekk <- c(1500*1.25,1500*1.25,1500*1.25)
eha <- c(7000000*1,7000000*0.5,7000000*1.5)
asi = unique(round(exp(seq(0,log(n),length=100))))
prooviksa <- matrix(ncol=k,nrow=length(asi))
#siin on see programm
for (i in 1:k) { #igas trajektooris on eri hinnangud
  kk=ekk[i]
  ha=eha[i]
  mu1 <- hin_param(kk,ha)[1]
  sigma1 <- hin_param(kk,ha)[2]
  param = hin_param(1500,7000000)
  abi = kahjum_lisa(ss=113,n=n,m=m)
  kes <- exp(abi[,1])
  sig <- exp(abi[,2])
  #print(dim(abi))
  for (j in 1:length(asi)) {
    prooviksa[j,i] = exp(kes[j]+sig[j]**2/2)
  }
}
plot(prooviksa[,1],type="l",
     ylim=c(min(prooviks),max(prooviksa)),
     xaxt="n",
     xlab="Kahjude arv",
     ylab="Kahju suurus"
)
legend("topright",
       legend=c("Täpsete parameetritega hinnang","Dispersioon alla hinnatud","Dispersioon üle hinnatud"
       ), inset=0.02, lty=1,
       col=c(col=1,col=3,col=4), box.lty=0, lwd=1, cex=0.85)
axis(1,at=c(10,21,41,70,89),labels=c("10","100","1000","10000","100000"))
for (m in 2:3) {
  lines(prooviksa[,m], col=m+1)
}
abline(h=1500,col="red")
###########
#vead     #
###########
set.seed(908765)
n=100 #mitu kordust
m=10 #mitu kahjuga poliisi
k=1000#mitu trajektori
proovi1 <- matrix(ncol=k,nrow=n)
proovi2 <- matrix(ncol=k,nrow=n)
for (i in 1:k) { #igas trajektooris on eri hinnangud
  kk=1800
  ha=7000000
  mu1 <- hin_param(kk,ha)[1]
  sigma1 <- hin_param(kk,ha)[2]
  param = hin_param(1500,7000000)#tegelik
  abi = kahjum(n=n,m=m)
  kes <- exp(abi[,1])
  sig <- exp(abi[,2])
  kes2 <- exp(abi[,3])
  sig2 <- exp(abi[,4])
```

```
  for (j in 1:n) {
    proovi1[j,i] = exp(kes[j]+sig[j]**2/2)
    proovi2[j,i] = exp(kes2[j]+sig2[j]**2/2)
  }

}
M <- proovi1
N <- proovi2
#################
#bayesi meetodiga#
#################
vek=M
vek[vek>1500]<- 1500
#potentsiaalne preemia
preemia_pot <- n*m*1500
#preemia kätte
preemia_kat <- colSums(m*vek)
#kaotus
kaotus = preemia_kat-preemia_pot
##################
#tõepära meetodiga#
##################
vek2=N
vek2[vek2>1500]<- 1500
#potentsiaalne preemia
preemia_pot2 <- n*m*1500
#preemia kätte
preemia_kat2 <- colSums(m*vek2)
#kaotus
kaotus2 = preemia_kat2-preemia_pot2
#seda regulli pärast
par(mfrow=2:1)
h <- hist(-kaotus, breaks = 45,
          xlim = c(0,100000),ylim=c(0,700),
          main="Bayesi meetod", ylab="Sagedus",xlab="Kaotus portfelli peale",
          col = rgb(0,0.5,0,0.5)
          ,xaxt="n"
)
axis(1,at=c(0,20000,40000,60000,80000,100000),
     labels =c("0","20000","40000","60000","80000","100000"))
abline(v=mean(-kaotus), col="red")
h2 <- hist(-kaotus2, breaks = 35,
           xlim = c(0,100000),ylim=c(0,700),
           main="Suurima tõepära meetod", ylab="Sagedus",xlab="Kaotus portfelli peale",
           col = rgb(0,0,1,0.5)
           ,xaxt="n"
)
axis(1,at=c(0,20000,40000,60000,80000,100000),
     labels =c("0","20000","40000","60000","80000","100000"))
abline(v=mean(-kaotus2), col="red")
#nüüd abs
#bayes
vek=M
kaotus_matr <- abs(vek-1500)
#kaotus
kaotus = colSums(m*kaotus_matr)
```

```
#ST
vek2=N
kaotus_matr2 <- abs(vek2-1500)
#kaotus
kaotus2 = colSums(m*kaotus_matr2)
par(mfrow=2:1)
h <- hist(kaotus, breaks = 45,
          xlim = c(0,400000),ylim=c(0,100),
          main="Bayesi meetod", ylab="Sagedus",xlab="Kaotus portfelli peale",
          col = rgb(0,0.5,0,0.5)
          ,xaxt="n"
)
axis(1,at=c(0,100000,200000,300000,400000,500000),
     labels =c("0","100000","200000","300000","400000","500000" ))
abline(v=mean(kaotus), col="red")
h2 <- hist(kaotus2, breaks = 35,
           xlim = c(0,400000),ylim=c(0,100),
           main="Suurima tõepära meetod", ylab="Sagedus",xlab="Kaotus portfelli peale",
           col = rgb(0,0,1,0.5)
           ,xaxt="n"
)
axis(1,at=c(0,100000,200000,300000,400000,500000),
     labels =c("0","100000","200000","300000","400000","500000" ))
abline(v=mean(kaotus2), col="red")
#######
#kokku#
#######
trajed2 <- function(k=2,n=10,m=10, antud){
  print(antud)
  M1 <- matrix(nrow = n, ncol= k)
  N1 <- matrix(nrow = n, ncol= k)
  ant <- c()
  for (i in 1:k) {
    antud <- antud
    andmed1 <- rnbinom(n*m,mu=0.05,size=2)
    ant[i]<- sum(andmed1)
    #print(ant)
    abi = kahjum1(n=n,m=m,andmed1=andmed1,par1 = antud)
    M1[,i] = exp(abi[,1])
    N1[,i] = exp(abi[,2])
    print(i)
  }
  return(rbind(cbind(M1, N1),ant)) #esimesed k tulpa on bayes, järgmised k tulpa on ST
}
set.seed(101)
k = 250
n=100
m=100
antud=0.04
vigad <- trajed2(k=k,n=n, m=m, antud=antud)
viga1<- vigad[1:n,]
ant1 <- vigad[n+1,]
set.seed(101)
antud=0.05
vigad <- trajed2(k=k,n=n, m=m, antud=antud)
viga2<- vigad[1:n,]
```

```
ant2 <- vigad[n+1,]
viga11 <- viga1[,1:k] #mööda
viga21 <- viga2[,1:k] #bay pihta
viga3 <- viga1[,(k+1):(2*k)]
plot(viga21[,2], type="l")
lines((viga11[,2]))
kahjum5 <- function(ss=1,n=100, m=1) {
  set.seed(ss)
  andmed1 <- rlnorm(n*m,meanlog=param[1],sdlog=param[2])
  init1 <- log(param[1])
  init2 <- log(param[1])
  prooviks1 <- data.frame(matrix(nrow=n,ncol = 2))
  prooviks2 <- data.frame(matrix(nrow=n,ncol = 2))
  start_time <- Sys.time()
  for (i in 1:n) {
    prooviks1[i,] = l_max_jareljaotusi(c(init1, init2), andmed=andmed1[1:(i*m)])
    prooviks2[i,] = l_maxi(c(init1, init2), andmed=andmed1[1:(i*m)])
    init1 <- prooviks1[i,1]
    init2 <- prooviks1[i,2]
  }
  end_time <- Sys.time()
  print(end_time - start_time)
  matr <- data.frame(matrix(nrow=n,ncol = 4))
  matr[,1]<-prooviks1[,1]#bay meetod mu
  matr[,2]<-prooviks1[,2]#bay sigma
  matr[,3]<-prooviks2[,1]#tõepära mu
  matr[,4]<-prooviks2[,2]#tõepära sigma
  return(matr)
}
proovi1 <- matrix(ncol=k,nrow=n)
proovi2 <- matrix(ncol=k,nrow=n)
proovi3 <- matrix(ncol=k,nrow=n)
proovi4 <- matrix(ncol=k,nrow=n)
for (i in 1:k) { #igas trajektooris on eri hinnangud
  kk=1500
  ha=7000000
  mu1 <- hin_param(kk,ha)[1]
  sigma1 <- hin_param(kk,ha)[2]
  param = hin_param(1500,7000000)#tegelik
  m=i
  abi = kahjum5(ss=i,n=n,m=m)
  kes <- exp(abi[,1])
  sig <- exp(abi[,2])
  kes2 <- exp(abi[,3])
  sig2 <- exp(abi[,4])
  kk=1800
  ha=7000000
  mu1 <- hin_param(kk,ha)[1]
  sigma1 <- hin_param(kk,ha)[2]
  param = hin_param(1500,7000000)#tegelik
  abi = kahjum5(ss=i,n=n,m=m)
  kes3 <- exp(abi[,1])
  sig3 <- exp(abi[,2])
  kes4 <- exp(abi[,3])
  sig4 <- exp(abi[,4])
  for (j in 1:n) {
```

```
    proovi1[j,i] = exp(kes[j]+sig[j]**2/2)
    proovi2[j,i] = exp(kes2[j]+sig2[j]**2/2)
    proovi3[j,i] = exp(kes3[j]+sig3[j]**2/2)
    proovi4[j,i] = exp(kes4[j]+sig4[j]**2/2)
  }
}
proovi1 #õiged bay
proovi3 #mööda bay
hind1 <- matrix(nrow=n,ncol = k)
hind2 <- matrix(nrow=n,ncol = k)
hind3 <- matrix(nrow=n,ncol = k)
hind4 <- matrix(nrow=n,ncol = k)
for (i in 1:n) {
  for (j in 1:k) {
    hind1[i,j] = viga11[i,j]*proovi3[i,j] #sagedus mööda, kahju mööda
    hind2[i,j] = viga11[i,j]*proovi1[i,j] #sagedus mööda, kahju õige
    hind3[i,j] = viga21[i,j]*proovi3[i,j] #sagedus õige, kahju mööda
    hind4[i,j] = viga21[i,j]*proovi1[i,j] #sagedus õige, kahju õige
  }
}
```

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Kristjan Sinikas**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose
   Kindlustuslepingute hinnastamine Bayesi meetodi abil,
   mille juhendaja on **Märt Möls**,
   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 Commonsi 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.

Kristjan Sinikas
25.05.2021