LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMĀTIKAS NODAĻA

BEIJESA METODES MATEMĀTISKAJĀ STATISTIKĀ

MAĢISTRA DARBS

Autors: **Baiba Buceniece**

Stud. apl. bb08025

Darba vadītājs: doc. Dr.math. Jānis Valeinis

RĪGA 2010

Anotācija

Darbā ir apskatītas Beijesa metodes galveno statistisko uzdevumu (punktveida novērtējumu, intervālu novērtējumu, hipotēžu pārbaudes) risināšanā un salīdzinātas ar klasiskajām metodēm. Aprakstīta iespējamo aprioro sadalījumu konstruēšana. Analizēta saistīto un neinformatīvo aprioro sadalījumu izmantošana binomiālā un normālā sadalījuma, un lineārās regresijas parametru novērtēšanai. Veiktas simulācijas ticamības intervālu pārklājuma precizitātes analīzei.

Atslēgas vārdi: Beijesa teorēma, apriorais sadalījums, aposteriorais sadalījums

Abstract

In this work Bayesian methods for statistical inference (as point estimation, interval estimation, hypothesis testing) have been considered and compared with frequentist methods. Possible prior choice has been described. Applications of cojugate and noninformative priors for the estimation of binomial and normal distrubution parameters, and for linear regression parameters have been analized. Simulations have been made for analysis of coverage precision of credible intervals.

Keywords: Bayes theorem, prior, posterior

Saturs

Apzīmējumi

$Bi(n, p)$ binomiālais varbūtību sadalījums,

$U(a, b)$ vienmērīgais varbūtību sadalījums,

$Beta(a, b)$ beta varbūtību sadalījums,

$Puas(\mu)$ Puasona varbūtību sadalījums,

$N(\mu, \sigma^2)$ normālais varbūtību sadalījums,

$\Gamma(x)$ gamma funkcija

$i.i.d.$ neatkarīgi, vienādi sadalīti,

$Neg(m, \theta)$ negatīvais binomiālais varbūtību sadalījums,

$G(\alpha, \beta)$ gamma varbūtību sadalījums.

Ievads

Beijesa statistika savu nosaukumu ieguvusi no 18.gadsimta angļu garīdznieka *Thomas Bayes* (1702-1761). *Bayes* aplūkotā problēma (mūsdienu terminoloģijā) bija binomiālā sadalījuma $Bi(n, \theta)$ parametra θ novērtēšana un viņš izstrādāja problēmas atrisinājumu pie pieņēmuma, ka θ ir ar $U(0, 1)$ aprioro blīvumu. Taču *Bayes* publikācijai bija maza ietekme un lielu daļa no tā, kas tagad tiek saukts par Beijesa statistiku, izstrādāja franču matemātiķis *Laplace* (1749-1827) neatkarīgi no *Bayes*. *Laplace* plaši izmantoja vienmērīgo aprioro sadalījumu.

Beijesa principa izmantošana kļuva izplatīta 19.gadsimtā, bet 19.gadsimta beigās to sāka kritizēt, piemēram, *Venn* (1886) vai *Bertrand* (1889).

Šajā laikā, kad tika attīstītas modernākas statistikas teorijas (sākot ar *Francis Galton* (1822-1911) un *Karl Pearson* (1857-1936) darbu), Beijesa idejas nonāca grūtā stāvoklī. Lielākais moderno statistikas metožu veicinātājs *R. A. Fisher* (1890-1962) visā savas zinātniskās darbības laikā bija pret Beijesa idejām.

Atgriešanās pie Beijesa statistikas sākās ar *Jeffreys* grāmatas "Theory of probability" (1939) izdošanu. *Jeffreys* uzskatīja, ka apriorajam sadalījumam ir jābūt pēc iespējas neinformatīvam. Lai to iegūtu, viņš piedāvāja vispārīgu formulu, kas tagad ir pazīstama kā *Jeffreys* apriorais sadalījums. Tomēr viņa argumenti nepārliecināja skeptiķus.

Kamēr *Jeffreys* izstrādāja savu teoriju, *Neyman* (1894-1981) un *Egon Pearson* (1895-1980) (Karla dēls) publicēja savu hipotēžu pārbaudes teoriju (1933), kura arī izvairījās no jebkādas atsaukšanās uz Beijesa idejām.

Alternatīvas Beijesa statistikai kļuva par standartu 1930-tajos gados ar maksimālās ticamības novērtētāju formulēšanu un matemātiskās statistikas formalizētās teorijas attīstīšanu, kur apriorie sadalījumi parādījās tikai kā formāli optimālu novērtētāju konstruēšanas veids, kā aprakstījis *Wald* (1950) vai *Ibragimov* un *Has'minskii* (1981). Turpmākie *Gini* vai *de Finetti* mēģinājumi formalizēt Beijesa pieeju statistikā no 1930-tajiem līdz 1970-tajiem gadiem nenesa vairāk popularitātes pretēji tajā laikā dominējošajai *Neyman-Pearson* paradigmai, kaut arī Beijesa kopiena auga un radīja zinātniskus darbus, tādus kā *Savage* (1954) un *Lindley* (1965, 1971).

Tikai nesen Beijesa statistika guva jaunu stimulu, pateicoties jaunu skaitļošanas rīku izstrādāšanai - kas vienmēr ir bijis centrālais Beijesa paradigmā - un strauji augošā interese par šo pieeju statistiskajā modelēšanā. Beijesa statistikas dzīvotspēju var redzēt Beijesa publikāciju procentuālajā attiecībā statistikas žurnālos. Izskatās, ka šajā gadsimtā Beijesa statistikai tiks veltīts daudz vairāk uzmanības nekā 20.gadsimtā.

Šajā darbā ir apskatītas Beijesa metodes, sākot ar nosacītajām varbūtībām un pārejot uz varbūtību sadalījumiem. Ja ir doti dati, tad Beijesa teorēma ļauj izlabot aprioros uzskatus par parametriem. Tāpēc, pirmkārt, ir jābūt apriorajam sadalījumam, kurš iespējamajām parametru vērtībām piešķir relatīvus svarus. Ir vairākas, atšķirīgas iespējas, kā iegūt apriorо sadalījumu. Visērtāk ir izmantot saistītos sadalījumus, jo tad aposteriorais sadalījums būs tās pašas sadalījumu saimes pārstāvis, un tā parametrus būs viegli aprēķināt. Bet, ja saistītā apriorā sadalījuma forma nereprezentē aprioros uzskatus pietiekami labi, tad var formulēt diskrētu apriorо sadalījumu un izmantot interpolēšanu, lai iegūtu nepārtrauktu sadalījumu.

Beijesa pieeja ļoti daudzos gadījumos ir labāka par klasisko pieeju, pat ja tās abas tiek vērtētas pēc klasiskās statistikas kritērijiem. Klasiskās statistikas "objektivitāte" tiek panākta, pilnīgi ignorējot jebkādu apriorо informāciju par novērtējamo procesu. Taču zinātnē parasti ir kaut kāda apriorā informācija par pētāmo procesu. Šī informācijas atmešana ir informācijas (kura bieži pārvēršas naudā) izšķērdēšana. Beijesa statistika izmanto abus informācijas avotus: apriorо informāciju un informāfiju no iegūtajiem datiem.

Beijesa statistika sniedz tiešu parametru varbūtību formulējumu, kas zinātniekiem ir daudz noderīgāk nekā klasiskās statistikas sniegtie ticamības formulējumi. Piemēram, klasiskie ticamības intervāli bieži tiek nepareizi izskaidroti kā varbūtību intervāli. Statistiķi zina, ka šāds skaidrojums ir kļūdains, bet zinātniekiem arī nav sevišķi noderīgi, ja ticamība tiek saistīta ar visu iespējamo (bet nerealizējušos) datu kopu varbūtību.

Beijesa statistika izmanto vienu rīku - Beijesa teorēmu, kura tiek izmantota visās situācijās. Tas ir pretstats klasiskajai statistikai, kurai nepieciešami daudz dažādi rīki.

Beijesa teorēma dod veidu, kā atrast nākamo novērojumu prognozējamo sadalījumu. Taču ar klasiskajām metodēm ne vienmēr tas ir tik viegli izdarāms.

Šīs Beijesa statistikas priekšrocības statistiķiem bija zināmas jau labu laiku, bet bija grūtības tās pielietot praktiski. Ja ir viegli uzrakstīt aposteriorā sadalījuma formulu

$$g(\theta|dati) = \frac{g(\theta) \times f(dati|\theta)}{\int g(\theta) \times f(dati|\theta) d\theta},$$

tad slēgtā forma eksistē tikai dažos vienkāršos gadījumos (piemēram, normālai izlasei ar normālu apriorо sadalījumu). Pārējos gadījumos ir nepieciešams integrēt skaitliski. Tikai nesen ir izstrādāti datoru algoritmi (piem., *Gibbs sampler* un *Metropolis-hasting* algoritms), lai no aposteriorajiem sadalījumiem ģenerētu gadījuma izlases bez pilnīgas pašu sadalījumu aprēķināšanas. Ņemot pietiekami lielas izlases no aposteriorā sadalījuma, mēs to varam aproksimēt līdz vēlamajai precizitātei. Patlaban to var praktiski lietot problēmām ar daudziem parametriem. Bet šīs metodes jau ir pietiekami sarežģītas un šajās darbā netiks apskatītas.

Darbs sastāv no 8 nodaļām. Pirmajā nodaļā ir formulēta Beijesa teorēma nosacītajām varbūtībām. Otrajā nodaļā ir aprakstīti dažādi apriorie sadalījumi. Trešajā nodaļā apskatīta Beijesa teorēmas izmatošana diskrētiem gadījuma lielumiem ar diskrētiem aprioraјiem sadalījumiem. Ceturtajā nodaļā aprakstīta vienmērīgā un *Beta* apriorā sadalījuma izmantošana binomiālās proporcijas novērtēšanai. Piektajā nodaļā ir salīdzinātas Beijesa un klasiskās metode binomiālās proporcijas novērtēšanai. Sestajā nodaļā aprakstīta Beijesa pieeja normālā sadalījuma vidējās vērtības novērtēšanā. Septītajā nodaļā ir salīdzinātas Beijesa un klasiskā metodes normālā sadalījuma vidējās vērtības novērtēšanai. Astotajā nodaļā tiek analizēta *Beta* apriorā sadalījuma izmantošana binomiālās proporcijas novērtēšanai, dažādu aprioro sadalījumu izmantošana normālā sadalījuma vidējās vērtības novērtēšanai, aprioro sadalījumu izmantošana regresijas koeficientu novērtēšanai un ar simulāciju palīdzību analizēta Beijesa ticamības intervālu pārklājuma precizitāte. Pēdējā nodaļā ir apkopoti secinājumi par apskatītajām Beijesa metodēm. Pielikumā ir iekļauts Beijesa metožu teorētiskais apraksts vienkāršās lineārās regresijas parametru novērtēšanai, kā arī robusto Beijesa metožu apraksts gadījumiem, ja tiek pieļauta varbūtība, ka ir izvēlēts nepareizs apriorais sadalījums. Praktiskie uzdevumi tika veikti programmā R un izveidotie kodi ir iekļauti pielikumā.

Šī darba teorijas izklāsts tiks ņemts pamatā no [1]. Praktiskā daļa tiks veikta ar programmas R palīdzību. Idejas par dažādu specifisku paketēs *LearnBayes* un *Bolstad* iebūvētu funkciju izmantošanu, kuras ir izstrādātas tieši Beijesa metožu pielietošanai, ir ņemtas no [2], [3] un [4].

1. Beijesa teorēma

Beijesa teorēma tiek formulēta sekojoši:

$$P(B_i|A) = \frac{P(B_i) \times P(A|B_i)}{\sum_{j=1}^{n} P(B_j) \times P(A|B_j)},$$

kur $B_1, \ldots, B_n$ ir tādi notikumi, ka:

- $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$ ir visa notikumu telpa;
- $B_i \cap B_j = \varnothing$, ja $i = 1, \ldots, n$, $j = 1, \ldots, n$ un $i \neq j$.

Marginālās varbūtības $P(B_i)$, $i = 1, \ldots, n$ ir zināmas pirms eksperimenta sākuma un tiek sauktas par *apriorajām* varbūtībām.

Nenovēroto notikumu $B_1, \ldots, B_n$ *ticamība* ir nosacītā varbūtība $P(A|B_i)$ $i = 1, \ldots, n$. $P(B_i|A)$, $i = 1, \ldots, n$ ir notikuma B_i *aposteriorā* varbūtība. Tāpēc Beijesa teorēmu var pārrakstīt kā

$$\textit{aposteriorā varbūtība} = \frac{\textit{apriorā varbūtība} \times \textit{ticamība}}{\sum \textit{apriorā varbūtība} \times \textit{ticamība}}. \tag{1.1}$$

Ja katru ticamību reizinātu ar konstanti, tad arī saucējs tiktu reizināts ar to pašu konstanti, kura dalīšanā tiktu saīsināta un tiktu iegūtas tās pašas aposteriorās varbūtības. Tādēļ ir nepieciešams zināt ticamību tikai līdz proporcionalitātes konstantei. Ar *ticamību* katrai iespējamai vērtībai piešķirtais *relatīvais* svars ir viss, ko nepieciešams zināt. Līdzīgi, katru aprioro varbūtību varētu reizināt ar konstanti. Saucējs atkal tiktu reizināts ar to pašu konstanti, kura dalīšanā tiktu saīsināta, un tiktu iegūtas tās pašas aposteriorās varbūtības. Beijesa teorēma bieži tiek rakstīta tās proporcionālajā formā

$$\textit{aposteriorā varbūtība} \propto \textit{apriorā varbūtība} \times \textit{ticamība}.$$

Katram notikumam B_i, $i = 1, \ldots, n$ tiek piekārtots relatīvais svars pēc tam, kad A ir noticis. Dalīšana ar relatīvo svaru summu normē relatīvos svarus tā, lai to summa būtu vienāda ar 1. Šī dalīšana relatīvos svarus padara par varbūtību sadalījumu. [5]

2. Apriorie sadalījumi

2.1. Apriorā sadalījuma noteikšanas grūtības

Apriorā sadalījuma noteikšana ir svarīgākais solis Beijesa analīzē. Zināmā mērā tas ir arī visgrūtākais, jo praksē pieejamā apriorā informācija tikai retos gadījumos ir pietiekami precīza, lai aprioro sadalījumu varētu noteikt precīzi. Vairāki varbūtību sadalījumi var būt savienojami ar aprioro informāciju.

Visbiežāk aprioro sadalījumu ir nepieciešams izvēlēties (daļēji) patvaļīgi. Sistemātiska parametrisko (piemēram, normālā, gamma, beta u.c.) sadalījumu izmantošana un tālāka reducēšana uz *saistītajiem* sadalījumiem nav pamatota visos gadījumos, jo parametriskie sadalījumi var ignorēt daļu apriorās informācijas. Dažās situācijās aprioro sadalījumu jāizvēlas daļēji automatizēti, piemēram, ja nav nekādas apriorās informācijas.

Nav viena vienīga vieda, kā izvēlēties aprioro sadalījumu. Nepamatoti apriorie sadalījumi noved pie nepamatotiem aposterioriem rezultātiem. Ne-informatīvie apriorie sadalījumi tiek iegūti tieši no izlašu sadalījumiem, kaut gan daži Beijesa statistiķi nepiekrīt šādām automatizētām metodēm. Pavisam nesen *robustuma* un *jutības analīzes* teorētiska izstrādāšana ir nodrošinājusi vēl labāku pamatu Beijesa analīzei, kad tā sastopas ar nepilnīgu aprioro informāciju. Hierarhiskā modelēšana ir ļāvusi samazināt apriorā sadalījuma ietekmi uz rezultātiem.

2.2. Neīsts apriorais sadalījums

Daudzos gadījumos apriorais sadalījums tik noteikts uz subjektīva vai teorētiska pamata, kas varbūtību mēra vietā paredz $\sigma-$galīgu mēru parametru telpā Θ, t.i., mēru g tādu, ka

$$\int_{\Theta} g(\theta) d\theta = +\infty.$$

Šādos gadījumos saka, ka apriorais sadalījums ir *neīsts* (*improper*) (vai *vispārināts*).

Šī pieeja parasti ir vienīgais veids kā iegūt aprioro sadalījumu gadījumos, kad vienīgā pieejamā informācija ir zināšanas par izlases sadalījumu $f(x|\theta)$. No vispārinātajiem sadalījumiem iegūtie novērtējumi, parasti ir pietiekami labi, lai pamatotu šos sadalījumus. Vāju īsto (*proper*) aprioro sadalījumu (piemēram, $N(0, 100^2)$) vietā vajadzētu izvēlēties neīstos aprioros sadalījumus, jo vājiem īstajiem apriorajiem sadalījumiem trūkst robustuma.

2.3. Saistītie apriorie sadalījumi

Ja ir maz informācijas par modeli, vai tā nav droša, tad nav iespējams aprioro sadalījumu noteikt tikai subjektīvi. Ja aposteriorais sadalījums $g(\theta|x)$ un apriorais sadalījums $g(\theta)$ ir no vienas varbūtību sadalījumu saimes, tad tie tiek saukti par *saistītajiem* sadalījumiem un apriorais sadalījums tiek saukts par saistīto sadalījumu ticamībai.

Ja varbūtību sadalījumu saime ir parametriska, tad pārslēgšanās no apriorā sadalījuma uz aposterioro sadalījumu notiek, precizējot atbilstošos sadalījumu parametrus. Šīs īpašības dēļ saistītie apriorie sadalījumi ir ļoti populāri. Aposteriorais sadalījums šajā gadījumā vienmēr ir aprēķināms.

Saistītie apriorie sadalījumi reizēm tiek saukti par *objektīvajiem* apriorajiem sadalījumiem.

Eksponenciālās varbūtību sadalījumu saimes.

Definīcija 1. *μ ir $\sigma-$galīgs mērs uz X un Θ ir parametru telpa. C un h ir funkcijas attiecīgi no X un Θ uz R_+, un S un T ir funkcijas no Θ un X uz R^k. Sadalījumu saime ar blīvumiem*

$$f(x|\theta) = C(\theta)h(x)e^{\{S(\theta)\cdot T(x)\}}$$

tiek saukta par eksponenciālo saimi ar dimensiju k. Speciālā gadījumā, ja $\Theta \subset R^k$, $X \subset R^k$ un

$$f(x|\theta) = C(\theta)h(x)e^{\{\theta\cdot x\}},$$

saime tiek saukta par dabisku.

Teorēma 1. *$f(x|\theta)$ saistītā saime tiek uzdota ar*

$$g(\theta|\mu,\lambda) = K(\mu,\lambda)e^{\theta\cdot\mu-\lambda\psi(\theta)},$$

kur $K(\mu,\lambda)$ ir blīvumu normējošā konstante. Atbilstošais aposteriorais sadalījums $g(\theta|\mu+x,\lambda+1)$.

1.tabulā parādīti eksponenciālo saimju izplatītāko sadalījumu saistītie sadalījumi.

Apriorā sadalījuma modelēšana ir svarīga maziem izlašu apjomiem, bet pieaugot izlašu apjomiem, apriorā sadalījuma nozīme mazinās. Ja izlases apjoms tiecas uz bezgalību, tad lielākā daļa aprioro sadalījumu dos līdzīgus rezultātus, kuri būs līdzīgi rezultātiem, kuri balstīti tikai uz ticamības funkciju.

1. tabula Dabiskie saistītie apriorie sadalījumi dažām eksponenciālajām saimēm

$f(x\|\theta)$	$g(\theta)$	$g(\theta\|x)$
Normālais $N(\theta, \sigma^2)$	Normālais $N(\mu, \tau^2)$	$N(\varrho(\sigma^2\mu + \tau^2 x), \varrho\sigma^2\tau^2)$ $\varrho^{-1} = \sigma^2 + \tau^2$
Puasona $Puas(\theta)$	Gamma $G(\alpha, \beta)$	$G(\alpha + x, \beta + 1)$
Gamma $G(\nu, \theta)$	Gamma $G(\alpha, \beta)$	$G(\alpha + \nu, \beta + x)$
Binomiālais $Bi(n, \theta)$	Beta $Beta(\alpha, \beta)$	$Beta(\alpha + x, \beta + n - x)$
Negatīvais binomiālais $Neg(m, \theta)$	Beta $Beta(\alpha, \beta)$	$Beta(\alpha + m, \beta + x)$
Normālais $N(\mu, 1/\theta)$	Gamma $G(\alpha, \beta)$	$G(\alpha + 0.5, \beta + (\mu - x)^2/2)$

2.4. Neinformatīvie apriorie sadalījumi

Ja apriorā informācija nav pieejama, tad nav iespējams subjektīvi izvēlēties aprioro sadalījumu un nevar noteikt saistīto aprioro sadalījumu parametrus. Apriorais sadalījums jāiegūst no izlases sadalījuma, jo tā ir vienīgā pieejamā informācija. Tāpēc šādā veidā iegūtie apriorie sadalījumi tiek saukti par *neinformatīviem.*

Laplasa apriorais sadalījums. Laplass bija pirmais, kurš izmantoja neinformatīvo metodi. Viņš izmantoja *vienmērīgo aprioro sadalījumu.*

Jeffreys apriorais sadalījums. *Jeffreys neinformatīvie apriorie sadalījumi* ir balstīti uz *Fišera informāciju*, kura viendimensionālā gadījumā ir uzdota ar

$$I(\theta) = E_\theta\left[\left(\frac{\partial \log f(X|\theta)}{\partial\theta}\right)^2\right].$$

Šo informāciju var pārrakstīt arī kā

$$I(\theta) = -E_\theta\left[\frac{\partial^2 \log f(X|\theta)}{\partial\theta^2}\right]. \tag{2.1}$$

Jeffreys apriorais sadalījums ir

$$g^*(\theta) \propto \sqrt{I(\theta)},$$

kurš ir definēts līdz normalizācijas koeficientam, ja g^* ir īsts sadalījums. Bet parasti Jeffreys apriorie sadalījumi ir neīsti.

Ja θ ir daudzdimensionāls parametrs, tad Fišera informācijas matrica ir definēta kā (2.1) vispārinājums. Ja $\theta \in R^k$, tad $I(\theta)$ ir ar elementiem

$$I_{ij}(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(x|\theta)\right] \quad (i,j=1,\dots,k),$$

un Jeffreys neinformatīvais apriorais sadalījums ir definēts ar

$$g^*(\theta) \propto \sqrt{\det(I(\theta))}.$$

Jeffreys metode nodrošina vienu no labākajiem automatizētajiem neinformatīvo apriororo sadalījumu iegūšanas viediem.

References apriorie sadalījumi. Šeit pamatideja ir atrast tādu aprioro sadalījumu, kurš maksimizē aposterioro informāciju. Ja $x_{1:n}$ apzīmē izlasi $(x_1,\dots,x_n)$ un $K_n(g)$ ir Kullback-Leibler diverģence starp aprioro sadalījumu g un atbilstošo aposterioro sadalījumu

$$K_n(g) = \int g(\theta|x_{1:n})\log(g(\theta|x_{1:n})/g(\theta))d\theta,$$

tad tiek izmantots $E[K_n(g)]$, kur matemātiskā cerība ir ņemta no $x_{1:n}$ marginālā sadalījuma kā *trūkstošās informācijas* mērs. References apriorais sadalījums tiek definēts, maksimizējot

$$K^*(g) = \lim_{n\to\infty} E[K_n(g)].$$

[6]

3. Beijesa problemātika diskrētiem gadījuma lielumiem

Uzskatīsim, ka parametrs ir gadījuma lielums X ar iespējamām vērtībām $x_1, \ldots, x_I$. Gadījuma lielums Y, kurš ir atkarīgs no parametra, pieņem vērtības $y_1, \ldots, y_J$. Lietojot Beijesa teorēmu, izdarīsim secinājumus par parametru (gadījuma lielumu) X, pie dotas novērotās $Y = y_j$ vērtības.

Beijesa notikumu telpa sastāv no visiem iespējamiem sakārtotiem pāriem (x_i, y_j), $i = 1, \ldots, I$, $j = 1, \ldots, J$. Notikumi $(X = x_1), \ldots, (X = x_I)$ sadala notikumu telpu nešķeļošās daļās, bet nav zināms, kurš no tiem ir noticis. Tiek novērots notikums $(Y = y_j)$.

Beijesa notikumu telpu var uzskatīt par divdimensionālu, kur horizontālā dimensija ir novērota un vertikālā dimensija nav novērota. Beijesa notikumu telpa diskrētiem gadījuma lielumiem ir parādīta 2.tabulā. Varbūtības ir definētas visos Beijesa notikumu telpas punktos.

2. tabula Beijesa notikumu telpa

(x_1, y_1)	(x_1, y_2)	. . .	(x_1, y_j)	. . .	(x_1, y_J)
. . .	. . .	. . . .	. . .	. . . .	. . .
(x_i, y_1)	(x_i, y_2)	. . .	(x_i, y_j)	. . .	(x_i, y_J)
. . .	. . .	. . . .	. . .	. . . .	. . .
(x_I, y_1)	(x_I, y_2)	. . .	(x_I, y_j)	. . .	(x_I, y_J)

Ar $f()$ apzīmēsim varbūtību sadalījumu (nosacīto vai nenosacīto), kurš satur novēroto gadījuma lielumu Y, un ar $g()$ apzīmēsim varbūtību sadalījumu (nosacīto vai nenosacīto), kurš satur tikai (nenovēroto) parametru (gadījuma lielumu) X. Katra notikumu kopējā varbūtība Beijesa notikumu telpā tiek atrasta, lietojot reizināšanas likumu

$$f(x_i, y_j) = g(x_i) \times f(y_j | x_i). \tag{3.1}$$

Y marginālais sadalījums tiek atrasts summējot varbūtības pa kolonnām. Notikumu kopējās un marginālās varbūtības ir parādītas 3.tabulā.

Ja tiek novērots $Y = y_j$, tad reducētā Beijesa notikumu telpa ir sakārtotu pāru kopa j-tajā kolonnā. Tas parādīts 4.tabulā. X aposterioro varbūtību funkcija, ja dots $Y = y_j$,

3. tabula Notikumu X un Y kopējās un marginālās varbūtības

	apriorā	y_1	$\cdots$	y_j	$\cdots$	y_J
x_1	$g(x_1)$	$f(x_1, y_1)$	$\cdots$	$f(x_1, y_j)$	$\cdots$	$f(x_1, y_J)$
.	.	.	$\cdots$	.	$\cdots$	.
.	.	.	$\cdots$	.	$\cdots$	.
.	.	.	$\cdots$	.	$\cdots$	.
x_i	$g(x_i)$	$f(x_i, y_1)$	$\cdots$	$f(x_i, y_j)$	$\cdots$	$f(x_i, y_J)$
.	.	.	$\cdots$	.	$\cdots$	.
.	.	.	$\cdots$	.	$\cdots$	.
.	.	.	$\cdots$	.	$\cdots$	.
x_I	$g(x_I)$	$f(x_I, y_1)$	$\cdots$	$f(x_I, y_j)$	$\cdots$	$f(x_I, y_J)$
		$f(y_1)$	$\cdots$	$f(y_j)$	$\cdots$	$f(y_J)$

tiek uzdota ar

$$g(x_i|y_j) = \frac{g(x_i) \times f(y_j|x_i)}{\sum_{i=1}^{n} g(x_i) \times f(y_j|x_i)}. \tag{3.2}$$

Parametra X apriorais sadalījums tiek uzdots ar *aprioro* varbūtību funkciju $g(x_i)$, $i = 1, \ldots, n$. Mēs ticam, ka tāda ir katra x_i varbūtība pirms esam redzējuši datus.

Ja ir novērots $Y = y_j$, tad X *ticamība* tiek uzdota ar *ticamības funkciju* $f(y_j|x_i)$, $i = 1, \ldots, n$. Tā ir Y nosacītās varbūtības funkcija, ja dots $X = x_i$, kurš ir novērtēts pie vērtības y_j, kura ir realizējusies, un kur X ir ļauts mainīties pa visām tā iespējamajām vērtībām $x_1, \ldots, x_n$. Ir nepieciešams zināt nosacītā novērojumu sadalījuma formu, jo tā rāda, kā novērojuma Y sadalījums ir atkarīgs no X vērtības. Bet to ir nepieciešams novērtēt tikai pie tās vērtības y_j, kura faktiski ir realizējusies.

Parametra X aposterioro varbūtību sadalījums tiek uzdots ar *aposterioro* varbūtību funkciju $g(x_i|y_j)$, kura tiek novērtēta pie x_i, $i = 1, \ldots, n$, ja dots $Y = y_j$.

4. tabula Reducētā Beijesa notikumu telpa, ja dots $Y = y_j$

.	.	.	.	(x_1, y_j)	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	(x_i, y_j)	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	(x_I, y_j)	.	.	.	.

Piemērs 1. *Kastē atrodas 5 bumbiņas, no kurām dažas var būt sarkanas, bet pārējās ir*

zaļas. Nav zināms sarkano bumbiņu skaits. Ar gadījuma lielumu X apzīmēsim sarkano bumbiņu skaitu kastē. Iespējamās X vērtības ir x_i, $i = 0, \ldots, 5$. Tā kā mums nav ne jausmas par sarkano bumbiņu skaitu, tad pieņemsim, ka visas X vērtības ir vienādi iespējamas. Tad X apriorais sadalījums ir $g(0) = g(1) = g(2) = g(3) = g(4) = g(5) = \frac{1}{6}$.

Uz labu laimi no kastes izvelk bumbiņu. Gadījuma lielums $Y = 1$, ja izvilktā bumbiņa ir sarkana, 0 - ja zaļa. $Y|X$ nosacītais novērojumu sadalījums ir $P(Y = 1|X = x_i) = \frac{i}{5}$ un $P(Y = 0|X = x_i) = \frac{5-i}{5}$. Notikumu kopējās varbūtības tiek atrastas, reizinot apriorās varbūtības ar nosacītajām novērojumu varbūtībām. Y marginālās varbūtības tiek aprēķinātas, summējot pa kolonnām notikumu kopējās varbūtības. Tās parādītas 5.tabulā.

5. tabula Notikumu kopējo un marginālo varbūtību sadalījumi

x_i	*apriorā varbūtība*	$y_j = 0$	$y_j = 1$
0	1/6	$\frac{1}{6} \times \frac{5}{5} = \frac{5}{30}$	$\frac{1}{6} \times \mathbf{\frac{0}{5}} = 0$
1	1/6	$\frac{1}{6} \times \frac{4}{5} = \frac{4}{30}$	$\frac{1}{6} \times \mathbf{\frac{1}{5}} = \frac{1}{30}$
2	1/6	$\frac{1}{6} \times \frac{3}{5} = \frac{3}{30}$	$\frac{1}{6} \times \mathbf{\frac{2}{5}} = \frac{2}{30}$
3	1/6	$\frac{1}{6} \times \frac{2}{5} = \frac{2}{30}$	$\frac{1}{6} \times \mathbf{\frac{3}{5}} = \frac{3}{30}$
4	1/6	$\frac{1}{6} \times \frac{1}{5} = \frac{1}{30}$	$\frac{1}{6} \times \mathbf{\frac{4}{5}} = \frac{4}{30}$
5	1/6	$\frac{1}{6} \times \frac{0}{5} = \frac{0}{30}$	$\frac{1}{6} \times \mathbf{\frac{5}{5}} = \frac{5}{30}$
$f(y_j)$		$\frac{15}{30}$	$\mathbf{\frac{15}{30} = \frac{1}{2}}$

Pieņemot, ka no kastes izvilktā bumbiņa ir sarkana, reducētā notikumu telpa ir kolonnā ($y_j = 1$). Nosacītās novērojumu varbūtības šajā kolonnā ir izceltas. Tās veido ticamības funkciju. 6.tabulā tiek atrastas X aposteriorās varbūtības, ja dots $Y = 1$.

6. tabula $X|Y = 1$ aposterioro varbūtību atrašana

x_i	*apriorā varbūtība*	$y_j = 0$	$y_j = 1$	*aposteriorā varbūtība*
0	1/6	$\frac{1}{6} \times \frac{5}{5} = \frac{5}{30}$	$\frac{1}{6} \times \mathbf{\frac{0}{5}} = 0$	0
1	1/6	$\frac{1}{6} \times \frac{4}{5} = \frac{4}{30}$	$\frac{1}{6} \times \mathbf{\frac{1}{5}} = \frac{1}{30}$	$\frac{1}{30}/\frac{1}{2} = \frac{1}{15}$
2	1/6	$\frac{1}{6} \times \frac{3}{5} = \frac{3}{30}$	$\frac{1}{6} \times \mathbf{\frac{2}{5}} = \frac{2}{30}$	$\frac{2}{30}/\frac{1}{2} = \frac{2}{15}$
3	1/6	$\frac{1}{6} \times \frac{2}{5} = \frac{2}{30}$	$\frac{1}{6} \times \mathbf{\frac{3}{5}} = \frac{3}{30}$	$\frac{3}{30}/\frac{1}{2} = \frac{3}{15}$
4	1/6	$\frac{1}{6} \times \frac{1}{5} = \frac{1}{30}$	$\frac{1}{6} \times \mathbf{\frac{4}{5}} = \frac{4}{30}$	$\frac{4}{30}/\frac{1}{2} = \frac{4}{15}$
5	1/6	$\frac{1}{6} \times \frac{0}{5} = \frac{0}{30}$	$\frac{1}{6} \times \mathbf{\frac{5}{5}} = \frac{5}{30}$	$\frac{5}{30}/\frac{1}{2} = \frac{5}{15}$
$f(y_j)$		$\frac{15}{30}$	$\mathbf{\frac{15}{30} = \frac{1}{2}}$	

Vienīgā kolonna, kura tika izmantota aposterioro varbūtību sadalījuma atrašanai, bija kolonna $Y = 1$ reducētajā notikumu telpā. Notikumu kopējā varbūtība tika iegūta, reizinot *apriorās varbūtības ar ticamības funkciju. Aposteriorās varbūtības tiek aprēķinātas ar vienādojumu* (3.2)*. Tādējādi vienkāršāks veids, kā iegūt aposteriorās varbūtības, ir izmantot tikai reducētās notikumu telpas kolonnu.*

3.1. Divi ekvivalenti Beijesa teorēmas lietošanas veidi

Novērojumu analizēšana pēc kārtas pa vienam. Pieņemsim, ka uz labu laimi no kastes tiek izvilkta otra bumbiņa, pirmo neliekot atpakaļ kastē. Pieņemsim, ka otrā izvilktā bumbiņa ir zaļa, tāpēc $Y = 0$. Vēlamies atrast X aposteriorās varbūtības, ja doti divu novērojumu rezultāti (pirmā - sarkana, otrā - zaļa). Analizēsim novērojumus pa vienam, katru reizi izmantojot Beijesa teorēmu. Izmantosim tās pašas apriorās varbūtības, kuras bija zināmas pirms eksperimenta. Taču otrajam mēģinājumam, kā apriorās varbūtības izmantosim pirmā mēģinājuma aposteriorās varbūtības. Rezultāti parādīti 7.tabulā.

7. tabula Aposterioro varbūtību sadalījums pēc otrā novērojuma

x_i	*apriorā varbūtība*	*ticamība* (pie $y_j = 0$)	*apriorā varbūtība* $\times$ *ticamība*	*aposteriorā varbūtība*
0	0	0	0	$0/\frac{1}{3} = 0$
1	1/15	$\frac{4}{4}$	$\frac{1}{15}$	$\frac{1}{15}/\frac{1}{3} = \frac{1}{5}$
2	2/15	$\frac{3}{4}$	$\frac{1}{10}$	$\frac{1}{10}/\frac{1}{3} = \frac{6}{20}$
3	3/15	$\frac{2}{4}$	$\frac{1}{10}$	$\frac{1}{10}/\frac{1}{3} = \frac{6}{20}$
4	4/15	$\frac{1}{4}$	$\frac{1}{15}$	$\frac{1}{15}/\frac{1}{3} = \frac{1}{5}$
5	5/15	$\frac{0}{4}$	0	$0/\frac{1}{3} = 0$
			$\frac{1}{3}$	1.00

Visu novērojumu analizēšana reizē. Abus mēģinājumus var apskatīt kopā, lietojot Beijesa teorēmu tikai vienu reizi. Ņemam tās pašas apriorās varbūtības, kuras izmantojām pirmajam mēģinājumam, kad analizējām novērojumus pa vienam. Aprioro varbūtību funkcija ir $g(x) = \frac{1}{6}$, $x = 0, \ldots, 5$.

Pieņemsim, ka Y_1 un Y_2 ir attiecīgi pirmā un otrā mēģinājuma iznākumi. Otrā mēģinājuma varbūtības ir atkarīgas no tā, kādas bumbiņas palika kastē pēc pirmā mēģinājuma. Novērojumu varbūtība, ja noticis X, ir

$$f(y_1, y_2|x) = f(y_1|x) \times f(y_2|y_1, x).$$

X un Y_1, Y_2 kopējais sadalījums ir dots 8.tabulā. Pirmā bumbiņa bija sarkana, otrā - zaļa, tāpēc reducētās notikumu telpas varbūtības ir kolonnā $y_{j_1}, y_{j_2} = 1, 0$. Šajā kolonnā ir izceltas ticamības funkcijas vērtības, kuras uzdotas ar nosacītajām novērojumu varbūtībām.

8. tabula X, Y_1, Y_2 kopējais sadalījums un Y_1, Y_2 marginālais sadalījums

x_i	*apriorā varbūtība*	y_{j_1}, y_{j_2} 0,0	y_{j_1}, y_{j_2} 0,1	y_{j_1}, y_{j_2} 1,0	y_{j_1}, y_{j_2} 1,1
0	1/6	$\frac{1}{6} \times \frac{5}{5} \times \frac{4}{4}$	$\frac{1}{6} \times \frac{5}{5} \times \frac{0}{4}$	$\frac{1}{6} \times \frac{\mathbf{0}}{\mathbf{5}} \times \frac{\mathbf{4}}{\mathbf{4}}$	$\frac{1}{6} \times \frac{0}{5} \times \frac{0}{4}$
1	1/6	$\frac{1}{6} \times \frac{4}{5} \times \frac{3}{4}$	$\frac{1}{6} \times \frac{4}{5} \times \frac{1}{4}$	$\frac{1}{6} \times \frac{\mathbf{1}}{\mathbf{5}} \times \frac{\mathbf{4}}{\mathbf{4}}$	$\frac{1}{6} \times \frac{1}{5} \times \frac{0}{4}$
2	1/6	$\frac{1}{6} \times \frac{3}{5} \times \frac{2}{4}$	$\frac{1}{6} \times \frac{3}{5} \times \frac{2}{4}$	$\frac{1}{6} \times \frac{\mathbf{2}}{\mathbf{5}} \times \frac{\mathbf{3}}{\mathbf{4}}$	$\frac{1}{6} \times \frac{2}{5} \times \frac{1}{4}$
3	1/6	$\frac{1}{6} \times \frac{2}{5} \times \frac{1}{4}$	$\frac{1}{6} \times \frac{2}{5} \times \frac{3}{4}$	$\frac{1}{6} \times \frac{\mathbf{3}}{\mathbf{5}} \times \frac{\mathbf{2}}{\mathbf{4}}$	$\frac{1}{6} \times \frac{3}{5} \times \frac{2}{4}$
4	1/6	$\frac{1}{6} \times \frac{1}{5} \times \frac{0}{4}$	$\frac{1}{6} \times \frac{1}{5} \times \frac{4}{4}$	$\frac{1}{6} \times \frac{\mathbf{4}}{\mathbf{5}} \times \frac{\mathbf{1}}{\mathbf{4}}$	$\frac{1}{6} \times \frac{4}{5} \times \frac{3}{4}$
5	1/6	$\frac{1}{6} \times \frac{0}{5} \times \frac{0}{4}$	$\frac{1}{6} \times \frac{0}{5} \times \frac{4}{4}$	$\frac{1}{6} \times \frac{\mathbf{5}}{\mathbf{5}} \times \frac{\mathbf{0}}{\mathbf{4}}$	$\frac{1}{6} \times \frac{5}{5} \times \frac{4}{4}$
	$f(y_{j_1}, y_{j_2})$	40/120	20/120	20/120	40/120

X aposteriorā varbūtība, ja $Y_1 = 1$ un $Y_2 = 0$, tiek atrasta, normējot varbūtības reducētajā notikumu telpā tā, ka to summa ir vienāda ar 1. Tas ir parādīts 9.tabulā.

9. tabula Aposterioro varbūtību sadalījums, ja $Y_1 = 1$ un $Y_2 = 0$

x_i	*apriorā varbūtība*	y_{j_1}, y_{j_2} 0,0	y_{j_1}, y_{j_2} 0,1	y_{j_1}, y_{j_2} 1,0	y_{j_1}, y_{j_2} 1,1	*aposteriorā varbūtība*
0	1/6	$\frac{20}{120}$	0	**0**	0	$0 \quad = 0$
1	1/6	$\frac{12}{120}$	$\frac{4}{120}$	$\frac{\mathbf{4}}{\mathbf{120}}$	0	$\frac{4}{120} / \frac{20}{120} = \frac{1}{5}$
2	1/6	$\frac{6}{120}$	$\frac{6}{120}$	$\frac{\mathbf{6}}{\mathbf{120}}$	$\frac{2}{120}$	$\frac{6}{120} / \frac{20}{120} = \frac{3}{10}$
3	1/6	$\frac{2}{120}$	$\frac{6}{120}$	$\frac{\mathbf{6}}{\mathbf{120}}$	$\frac{6}{120}$	$\frac{6}{120} / \frac{20}{120} = \frac{3}{10}$
4	1/6	0	$\frac{4}{120}$	$\frac{\mathbf{4}}{\mathbf{120}}$	$\frac{12}{120}$	$\frac{4}{120} / \frac{20}{120} = \frac{1}{5}$
5	1/6	0	0	**0**	$\frac{20}{120}$	$0 \quad = 0$
	$f(y_{j_1}, y_{j_2})$			20/120		1.00

Aposteriorās varbūtības ir tādas pašas, kādas tika iegūtas, analizējot novērojumus pa vienam, izmantojot pirmā mēģinājuma aposteriorās varbūtības par otrā mēģinājuma apriorajām varbūtībām.

Tā kā tiek izmantota tikai tā kolonna, kura atbilst reducētajai notikumu telpai, tad ir vienkāršāk atrast aposterioro varbūtības, reizinot apriorās varbūtības ar ticamību un normējot, padarot to par varbūtību sadalījumu. Tas parādīts 10.tabulā.

10. tabula Aposterioro varbūtību sadalījums pēc abiem novērojumiem

x_i	*apriorā varbūtība*	*ticamība*	*apriorā varbūtība* $\times$ *ticamība*	*aposteriorā varbūtība*
0	1/6	$\frac{0}{20}$	$\frac{0}{120}$	$\frac{0}{120}/\frac{1}{6} = 0$
1	1/6	$\frac{4}{20}$	$\frac{4}{120}$	$\frac{4}{120}/\frac{1}{6} = \frac{1}{5}$
2	1/6	$\frac{6}{20}$	$\frac{6}{120}$	$\frac{6}{120}/\frac{1}{6} = \frac{3}{10}$
3	1/6	$\frac{6}{20}$	$\frac{6}{120}$	$\frac{6}{120}/\frac{1}{6} = \frac{3}{10}$
4	1/6	$\frac{4}{20}$	$\frac{4}{120}$	$\frac{4}{120}/\frac{1}{6} = \frac{1}{5}$
5	1/6	$\frac{0}{20}$	$\frac{0}{120}$	$\frac{0}{120}/\frac{1}{6} = 0$
			$\frac{1}{6}$	1.00

3.2. Beijesa teorēma binomiālajam sadalījumam ar diskrētu apրioro sadalījumu

Apskatīsim Beijesa teorēmas lietošanu, kad novērojumiem ir binomiāls sadalījums un parametram ir iespējamas tikai dažas vērtības. $Y|\pi$ ir ar $Bi(n,\pi)$ sadalījumu. (Ir n neatkarīgi mēģinājumi, kuru iznākums var būt "veiksme" vai "neveiksme", un "veiksmes" varbūtība visiem mēģinājumiem ir π. Y ir "veiksmju" skaits n mēģinājumos.) Ir iespējamas I diskrētas vērtības $\pi_1,\dots,\pi_I$.

Piemērs 2. $(Y|\pi) \sim Bi(n=4,\pi)$. *Pieņemsim, ka π var pieņemt tikai trīs vērtības* $0.4, 0.5$ *un* 0.6*, kuras ir vienādi iespējamas. π apriorais sadalījums, π un Y kopējais sadalījums doti 11.tabulā. Kopējais varbūtību sadalījums $f(\pi_i, y_j)$ tiek atrasts ar vienādojumu* (3.1)*. Pieņemsim, ka tika novērots $Y=3$. Reducētā notikumu telpa ir kolonna, kurā $Y=3$. Nosacītās novērojumu varbūtības (ticamība) šajā kolonnā ir izceltas.*

Binomiālā sadalījuma ticamības funkcija ir

$$f(y|\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y}, \quad 0 \le \pi \le 1,$$

kur $\binom{n}{y} = \frac{n!}{y!(n-y)!}$.

π apriorais sadalījums, (π, Y) kopējais varbūtību sadalījums un Y marginālais varbūtību sadalījums ir parādīti 12.tabulā.

Tā kā ir novērots $Y=3$, tad būtiska ir tikai ar "3" apzīmētā kolonna. π apriorais sadalījums, (π, Y) kopējais varbūtību sadalījums, Y marginālais varbūtību sadalījums un $(\pi|Y=3)$ aposterioro varbūtību sadalījums ir parādīti 12.tabulā.

11. tabula Kopējais varbūtību sadalījums

π	*apriorā varbūtība*	0	1	2	3	4
0.4	$\frac{1}{3}$	$\frac{1}{3} \times 0.1296$	$\frac{1}{3} \times 0.3456$	$\frac{1}{3} \times 0.3456$	$\frac{1}{3} \times \mathbf{0.1536}$	$\frac{1}{3} \times 0.0256$
0.5	$\frac{1}{3}$	$\frac{1}{3} \times 0.0625$	$\frac{1}{3} \times 0.2500$	$\frac{1}{3} \times 0.3750$	$\frac{1}{3} \times \mathbf{0.2500}$	$\frac{1}{3} \times 0.0625$
0.6	$\frac{1}{3}$	$\frac{1}{3} \times 0.0256$	$\frac{1}{3} \times 0.1536$	$\frac{1}{3} \times 0.3456$	$\frac{1}{3} \times \mathbf{0.3456}$	$\frac{1}{3} \times 0.1296$

12. tabula π kopējais, marginālais un aposterioro varbūtību sadalījums, ja $Y = 3$

π	*apriorā varbūtība*	0	1	2	**3**	4	*aposteriorā varbūtība*
0.4	$\frac{1}{3}$	0.0432	0.1152	0.1152	0.0512	0.0085	$\frac{0.0512}{0.2497} = 0.205$
0.5	$\frac{1}{3}$	0.0208	0.0833	0.1250	0.0833	0.0208	$\frac{0.0833}{0.2497} = 0.334$
0.6	$\frac{1}{3}$	0.0085	0.0512	0.1152	0.1152	0.0432	$\frac{0.1152}{0.2497} = 0.461$
marginālā varbūtība		0.0725	0.2497	0.3554	0.2497	0.0725	1.000

Nebija nepieciešams sastādīt visu varbūtību tabulu. Vienkāršāk ir aplūkot tikai reducētās notikumu telpas kolonnu. Tas parādīts 13.tabulā.

13. tabula Vienkāršota tabula aposteriorā sadalījuma atrašanai, ja $Y = 3$

π	*apriorā varbūtība*	*ticamība*	*apriorā varbūtība* $\times$ *ticamība*	*aposteriorā varbūtība*
0.4	$\frac{1}{3}$	0.1536	0.0512	$\frac{0.0512}{0.2497} = 0.205$
0.5	$\frac{1}{3}$	0.2500	0.0833	$\frac{0.0833}{0.2497} = 0.334$
0.6	$\frac{1}{3}$	0.3456	0.1152	$\frac{0.1152}{0.2497} = 0.461$
marginālā $P(Y = 3)$			0.2497	1.000

Visu aprioro varbūtību reizināšana ar konstanti nemaina Beijesa teorēmas rezultātu. Ja katra *apriorā varbūtība* $\times$ *ticamība* tiktu reizināta ar konstanti, tad marginālais elements, kurš tiek atrasts summējot šos reizinājumus, arī tiktu reizināts ar to pašu konstanti. Tādēļ aposteriorās varbūtības būtu tādas pašas kā iepriekš, jo konstante dalījumā tiktu saīsināta. Ja ir dota apriorā sadalījuma formula, tad jebkura tās daļa, kura nesatur parametru, var tikt iekļauta konstantē. Tas vienkāršo aprēķinus!

Ticamības reizināšana ar konstanti nemaina Beijesa teorēmas rezultātu. *Apriorā varbūtība* $\times$ *ticamība* vērtības arī tiktu reizinātas ar to pašu konstanti, kura tiktu saīsināta, aprēķinot aposteriorās varbūtības. Ja ir dota formula ticamības aprēķināšanai, tad jebkura tās daļa, kura nesatur parametru, var tikt iekļauta konstantē, vienkāršojot aprēķinus! [7].

Piemērs 2. *(turpinājums) Mēs izmantojām aprioro sadalījumu, kurš katrai π vērtībai piešķīra vienādu aprioro varbūtību. Šajā piemērā ir trīs iespējamās vērtības, tāpēc katrai no tām apriorā varbūtība ir vienāda ar $\frac{1}{3}$. Pareizināsim katru no trim apriorajām varbūtībām ar konstanti 3, lai iegūtu aprioros svarus vienādus ar 1. Tas vienkāršos aprēķinus. Novērojumiem ir $Bi(n=4,\pi)$ sadalījums un ir novērots $Y=3$. Formula binomiālajai ticamībai ir*

$$f(Y|\pi)=\binom{4}{3}\pi^3(1-\pi)^1.$$

Binomiālais koeficients $\binom{4}{3}$ nesatur parametru, tāpēc tā ir konstante ticamības kolonnā. Lai vienkāršotu aprēķinus, iekļausim to konstantē un izmantosim tikai to ticamības daļu, kura satur parametru. 14.tabulā redzam, ka tas dod tādu pašu rezultātu, kāds tika iegūts 13.tabulā.

14. tabula Vienkāršota tabula aposteriorā sadalījuma atrašanai, ja dots $Y=3$

π	*apriorā varbūtība (proporcionālā)*	*ticamība (proporcionālā)*	*apriorā varbūtība* $\times$ *ticamība*	*aposteriorā varbūtība*
0.4	1	$0.4^3\times 0.6^1=0.0384$	0.0384	$\frac{0.0384}{0.1873}=0.205$
0.5	1	$0.5^3\times 0.5^1=0.0625$	0.0625	$\frac{0.0625}{0.1873}=0.334$
0.6	1	$0.6^3\times 0.4^1=0.0864$	0.0864	$\frac{0.0864}{0.1873}=0.461$
marginālā summa			0.1873	1.000

4. Beijesa problemātika binomiālai proporcijai

Novērojuma Y nosacītais sadalījums (kopējais veiksmju skaits n mēģinājumos, ja dots parametrs π) ir $Bi(n, \pi)$. y nosacītās varbūtības funkcija pie dotā π ir

$$f(y|\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \quad y = 1, \ldots, n.$$

Ja aplūko šo pašu sakarību starp π un y, bet saglabā y fiksētu un ļauj π mainīties pa tā iespējamajām vērtībām, tad iegūst ticamības funkciju

$$f(y|\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \quad 0 \le \pi \le 1.$$

Lai izmantotu Beijesa teorēmu, ir nepieciešams apriorais sadalījums $g(\pi)$. Apriroro sadalījumu nedrīkst konstruēt no datiem. Reizināšana Beijesa teorēmā ir pamatota tikai tad, ja apriorais sadalījums ir *neatkarīgs* no ticamības! Aposteriorais sadalījums ir proporcionāls apriorā sadalījuma un ticamības reizinājumam. To pieraksta sekojoši

$$g(\pi|y) \propto g(\pi) \times f(y|\pi). \tag{4.1}$$

Tas dod aposteriorā blīvuma funkcijas formu, bet ne tieši pašu aposteriorā blīvuma funkciju. Lai iegūtu faktisko aposterioro sadalījumu, (4.1) nepieciešams dalīt ar kādu konstanti k, lai pārliecinātos, ka tas ir varbūtību sadalījums (laukumam zem aposteriorā sadalījuma funkcijas jābūt vienādam ar 1). k atrod integrējot $g(\pi) \times f(y|\pi)$ pa visu apgabalu. Tāpēc vispārīgā gadījumā

$$g(\pi|y) = \frac{g(\pi) \times f(y|\pi)}{\int_0^1 g(\pi) \times f(y|\pi)\, d\pi}. \tag{4.2}$$

Apskatīsim dažus iespējamos aprioros sadalījumus.

4.1. Vienmērīgais apriorais sadalījums binomiālai proporcijai

Ja iepriekš nav ne jausmas par to, kāda ir proporcija π, tad var izvēlēties apriroro sadalījumu, kurā visas vērtības ir "vienlīdzīgas". Vienmērīgais apriorais sadalījums visām iespējamajām π vērtībām piešķir vienādu svaru. Piemēram,

$$g(\pi) = 1, \quad 0 \le \pi \le 1.$$

Šajā gadījumā aposteriorais blīvums ir proporcionāls ticamībai:

$$g(\pi|y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \quad 0 \le \pi \le 1.$$

Varam ignorēt to daļu, kura nav atkarīga no π. Tā ir konstanta visām π vērtībām, tāpēc tā neietekmē aposteriorā sadalījuma formu. Apskatot to formulas daļu, kura nosaka aposteriorā sadalījuma formu, kā funkciju no π, atpazīstam, ka tas ir $Beta(a, b)$ sadalījums, kur $a = y + 1$ un $b = n - y + 1$.

4.2. Beta apriorais sadalījums binomiālai proporcijai

Pieņemsim, ka priekš π tiek izmantots $Beta(a, b)$ apriorais blīvums:

$$g(\pi; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1}(1-\pi)^{b-1}, \quad 0 \le \pi \le 1,$$

kur $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt, \quad \alpha > 0$. Varam ignorēt konstantes, kuras ietilpst apriorajā sadalījumā un ticamībā, un kuras nav atkarīgas no parametra. Tas dod

$$g(\pi|y) \propto \pi^{a+y-1}(1-\pi)^{b+n-y-1}, \quad 0 \le \pi \le 1.$$

Tas ir $Beta$ sadalījums ar parametriem

$$a' = a + y \quad \text{un} \quad b' = b + n - y. \tag{4.3}$$

Tas ir, pie a tiek pieskaitīts "veiksmju" skaits un pie b tiek pieskaitīts "neveiksmju" skaits:

$$g(\pi|y) = \frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)} \pi^{y+a-1}(1-\pi)^{n-y+b-1}, \quad 0 \le \pi \le 1.$$

π aposteriorais blīvums tika iegūts vienkārši, neizmantojot integrēšanu.

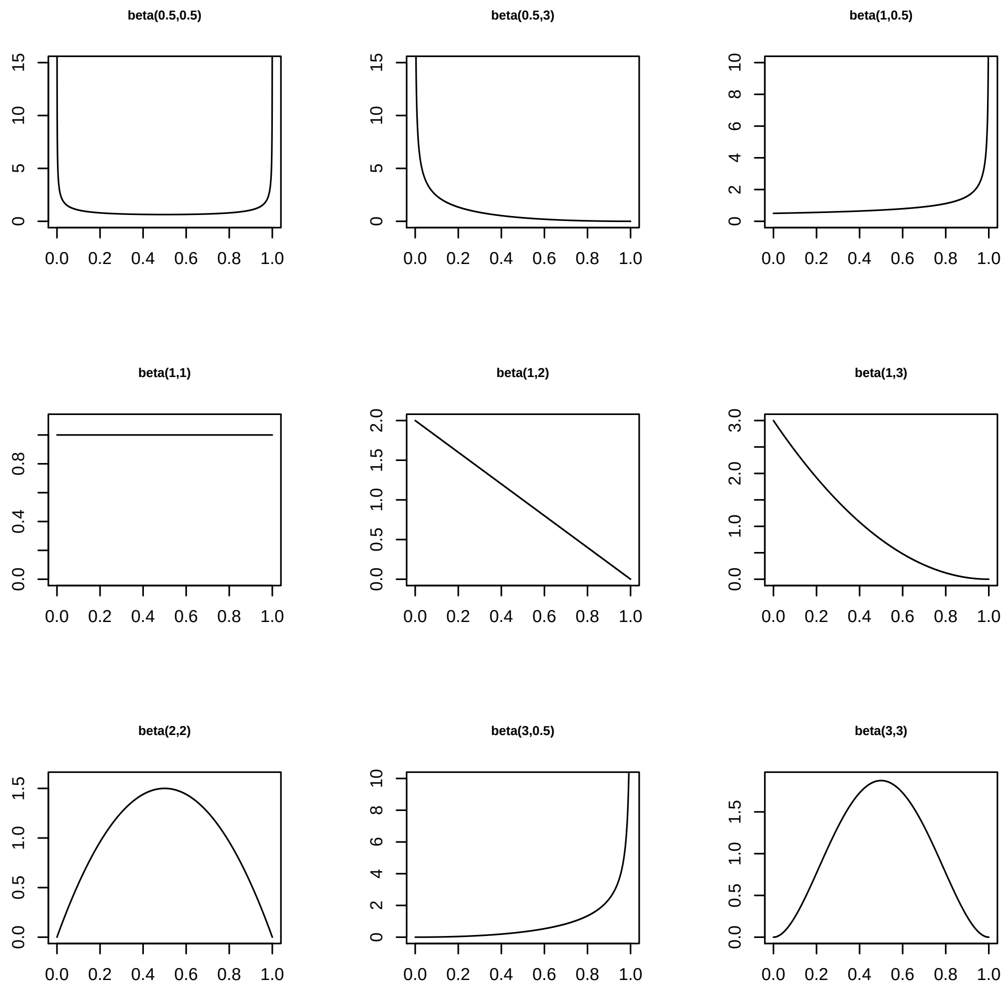
1. attēlā parādītas $Beta(a, b)$ blīvuma funkciju formas vērtībām $a = 0.5, 1, 2, 3$ un $b = 0.5, 1, 2, 3$.

Binomiālā novērojuma saistītā aprioro sadalījumu saime ir $Beta$ saime. Ja tiek izmantots apriorais sadalījums no saistītās saimes, tad aposteriorā sadalījuma atrašanai nav jāizmanto integrēšana. Ir nepieciešams izmantot tikai novērojumus, lai precizētu saistītās saimes apriorā sadalījuma parametrus, lai atrastu saistītās saimes aposterioro sadalījumu.

4.3. Apriorā sadalījuma izvēle

Saistītā apriorā sadalījuma izvēle, ja ir mazas apriorās zināšanas

Ja apriorās zināšanas ir nelielas, tad kāds no 1. attēlā redzamajiem $Beta(a, b)$ apriorajiem sadalījumiem varētu būt īsts apriorais sadalījums. Piemēram, ja apriorās zināšanas



1. att. Daži beta sadalījumi.

par π ir tādas, ka π ir ļoti mazs, tad $Beta(0.5, 1)$, $Beta(0.5, 2)$, $Beta(0.5, 3)$, $Beta(1, 2)$ vai $Beta(1, 3)$ visi varētu būt apmierinoši apriorie sadalījumi. Nav nozīmes tam, kurš no šiem apriorajiem sadalījumiem tiek izvēlēts, jo rezultējošie aposteriorie sadalījumi būs ļoti līdzīgi.

Saistītā apriorā sadalījuma izvēle, ja ir patiesas apriorās zināšanas par vidējo vērtību un standartnovirzi

Aprioro sadalījumu $Beta(a, b)$ jāizvēlas tādu, kurš saskan ar apriorajiem uzskatiem par vidējo vērtību un standartnovirzi. Ar π_0 apzīmēsim aprioro vidējo vērtību proporcijai, un ar σ_0 apzīmēsim aprioro standartnovirzi proporcijai.

$Beta(a, b)$ sadalījuma vidējā vērtība ir $\frac{a}{a+b}$. To pielīdzina apriorajai vidējai vērtībai

$$\pi_0 = \frac{a}{a+b}. \tag{4.4}$$

Beta sadalījuma standartnovirze ir $\sqrt{\frac{ab}{(a+b)^2(a+b+1)}}$. To pielīdzina apriorajai standartnovirzei. Ievērojot, ka $\frac{a}{a+b} = \pi_0$ un $\frac{b}{a+b} = 1 - \pi_0$, iegūst

$$\sigma_0 = \sqrt{\frac{\pi_0(1-\pi_0)}{a+b+1}}. \tag{4.5}$$

Vienādojumu (4.4) un (4.5) atrisināšana priekš a un b dos meklēto $Beta(a,b)$ apriororo sadalījumu.

Piesardzība pirms saistītā apriorā sadalījuma izmantošanas

1. $Beta(a,b)$ apriorais sadalījums jāattēlo grafiski. Ja forma izskatās pieņemami tuva apriorajiem uzskatiem, tad to var izmantot. Pretējā gadījumā var koriģēt π_0 un σ_0, kamēr atrod aprioro sadalījumu, kura grafiks aptuveni atbilst apriorajiem uzskatiem.

2. Jāaprēķina *ekvivalento aprioro izlases apjomu.* $Bi(n,\pi)$ sadalījuma izlases proporcijas $\hat{\pi} = \frac{y}{n}$ dispersija ir vienāda ar $\frac{\pi(1-\pi)}{n}$. Pielīdzinām šo dispersiju (ja apriorā vidējā vērtība ir π_0) apriorajai dispersijai.

$$\frac{\pi_0(1-\pi_0)}{n_{ekv}} = \frac{ab}{(a+b)^2 \times (a+b+1)}. \tag{4.6}$$

   Tā kā $\pi_0 = \frac{a}{a+b}$ un $(1-\pi_0) = \frac{b}{a+b}$, tad ekvivalentais izlases apjoms ir $n_{ekv} = a+b+1$. Tas nozīmē, ka informācijas daudzums par parametru no apriorā sadalījuma ir ekvivalents informācijas daudzumam no gadījuma izlases ar apjomu n_{ekv}.

Vispārīga nepārtraukta apriorā sadalījuma konstruēšana

Apriorā sadalījuma forma var nesakrist ar *Beta* formu, tāpēc var konstruēt diskrētu aprioro sadalījumu, kurš atbilst iedomātajiem svariem atsevišķām vērtībām no apgabala, kurš tiek uzskatīts par iespējamu, un pēc tam interpolēt starp tām, lai izveidotu nepārtrauktu aprioro sadalījumu. Ja apriorais sadalījums tiek konstruēts šādā veidā, tad, lai iegūtu aposteriororo sadalījumu, būs skaitliski jānovērtē apriorā sadalījuma un ticamības reizinājuma integrālis.

Ja ir pietiekami daudz datu, tad izvēlētā apriorā sadalījuma ietekme uz rezultātu būs maza, salīdzinot ar datu ietekmi. Pārliecināsimies, ka šajā gadījumā var iegūt ļoti līdzīgus aposterioros sadalījumus, lai gan tiek sākts ar dažādiem apriorajiem sadalījumiem.

4.4. Aposteriorā sadalījuma rezumēšana

Sadalījumu raksturošanai izmantosim izvietojuma (location) mērus, izkliedes mēru un sadalījuma procentīles.

Izvietojuma mēri

Aposteriorā moda. Ja aposteriorais sadalījums ir nepārtraukts, tad tā modu var atrast, pielīdzinot nullei aposteriorā blīvuma atvasinājumu. Ja aposteriorais sadalījums $g(\pi|y)$ ir $Beta(a', b')$, tad tā atvasinājums ir

$$g'(\pi|y) = (a'-1)\pi^{a'-2} \times (1-\pi)^{b'-1} + \pi^{a'-1} \times (-1)(b'-1)(1-\pi)^{b'-2}.$$

(Piezīme: šajā vienādojumā prim$'$ ir divas nozīmes: $g'(\pi|y)$ ir aposteriorā sadalījuma atvasinājums, bet a' un b' ir $Beta$ aposteriorā sadalījuma konstantes, kuras tiek atrastas, izmantojot precizēšanas likumus, kas uzdoti ar vienādojumiem (4.3).) Pielīdzinot $g'(\pi|y)$ nullei un atrisinot, iegūst

$$moda = \frac{a'-1}{a'+b'-2}.$$

Aposteriorajai modai kā izvietojuma mēram ir daži trūkumi. Pirmkārt, tā var atrasties sadalījuma beigās vai tuvu tām un tādējādi nereprezentēt visu sadalījumu. Otrkārt, var būt vairāki lokālie maksimumi.

Aposteriorā mediāna. Ja $g(\pi|y) \sim Beta(a', b')$, tad mediāna ir integrāļa

$$\int_0^{mediāna} g(\pi|y)\, d\pi = 0.5$$

atrisinājums. Vienīgais aposteriorās mediānas trūkums ir tas, ka jāatrod tās skaitliskā vērtība. Mediāna ir ļoti labs izvietojuma mērs.

Aposteriorā vidējā vērtība ir ļoti bieži lietots izvietojuma mērs

$$m' = \int_0^1 \pi g(\pi|y)\, d\pi. \tag{4.7}$$

Aposteriorā vidējā vērtība tiek spēcīgi ietekmēta, ja sadalījumam ir smaga aste. Šķībam (skewed) sadalījumam ar vienu smagu asti aposteriorā vidējā vērtība var būt samērā tālu no lielākajām varbūtībām. Ja $g(\pi|y) \sim Beta(a', b')$, tad aposteriorā vidējā vērtība ir vienāda ar

$$m' = \frac{a'}{a'+b'}. \tag{4.8}$$

$Beta(a, b)$ sadalījums ir ierobežots starp 0 un 1, tāpēc tam nav smagu astu. Aposteriorā vidējā vērtība ir labs $Beta$ aposteriorā sadalījuma izvietojuma mērs.

Izkliedes mēri

Ja aposteriorajam sadalījumam ir liela izkliede, tad mūsu zināšanas par parametru pat pēc novēroto datu analīzes joprojām būs neprecīzas.

Aposteriorā dispersija. Aposteriorā sadalījuma dispersija ir

$$D(\pi|y) = \int_0^1 (\pi - m')^2 g(\pi|y)\, d\pi. \tag{4.9}$$

Ja mums ir $Beta(a', b')$ aposteriorais sadalījums, tad aposteriorā dispersija ir

$$D(\pi|y) = \frac{a' \times b'}{(a' + b')^2 \times (a' + b' + 1)}. \tag{4.10}$$

Aposteriorā dispersija ir stipri ietekmēta sadalījumiem ar smagām astēm. Sadalījumiem ar smagām astēm dispersija būs ļoti liela, neskatoties uz to, ka lielākās varbūtības ir koncentrētas sadalījuma vidū. Turklāt tā ir kvadrātiskās vienībās, kas apgrūtina tās lieluma izskaidrošanu attiecībā pret vidējās vērtības lielumu. Šos aposteriorās dispersijas trūkumus pārvar, lietojot aposterioro standartnovirzi.

Aposteriorā standartnovirze nav kvadrātiskās vienībās, tāpēc tās lielums var tikt salīdzināts ar vidējās vērtības lielumu, un tā būs mazāk ietekmēta ar smagām astēm.

Aposteriorā sadalījuma procentīles. Aposteriorā sadalījuma k-tā procentīle tiek atrasta, skaitliski atrisinot vienādojumu

$$k = 100 \times \int_{-\infty}^{\pi_k} g(\pi|y)\, d\pi.$$

Starpkvartiļu intervāls. Starpkvartiļu intervāls

$$IQR = Q_3 - Q_1,$$

kur Q_3 ir 75-tā procentīle un Q_1 ir 25-tā procentīle, ir noderīgs izkliedes mērs, kuru neietekmē smagas astes.

4.5. Proporcijas novērtēšana

Punktveida novērtējums $\hat{\pi}$ ir no datiem aprēķināta statistika, kura tiek izmantota kā parametra π novērtējums. Piemēroti Beijesa punktveida novērtējumi ir, piemēram, izvietojuma mēri, kas aprēķināti no aposteriorā sadalījuma.

Novērtējuma aposteriorā vidējā kvadrātiskā kļūda. Proporcijas π novērtējuma $\hat{\pi}$ aposteriorā vidējā kvadrātiskā kļūda ir

$$PMS(\hat{\pi}) = \int_0^1 (\pi - \hat{\pi})^2 g(\pi|y)\, d\pi.$$

Pieskaitot un atņemot aposterioro vidējo vērtību m', iegūst

$$PMS(\hat{\pi}) = \int_0^1 (\pi - m' + m' - \hat{\pi})^2 g(\pi|y)\, d\pi.$$

Kāpinot kvadrātā, iegūst

$$PMS(\hat{\pi}) = \int_0^1 [(\pi - m')^2 + 2(\pi - m')(m' - \hat{\pi}) + (m' - \hat{\pi})^2] g(\pi|y)\, d\pi.$$

Sadala integrāli trīs integrāļos. Tā kā gan m', gan $\hat{\pi}$ ir konstantes attiecībā pret aposterioro sadalījumu, tad novērtējot integrāļus, iegūst

$$PMS(\hat{\pi}) = D(\pi|y) + 0 + (m' - \hat{\pi})^2.$$

Pēdējais saskaitāmais ir kvadrāts, tāpēc vienmēr ≥ 0. Vidēji kvadrātiskais attālums, kādā no aposteriorās vidējās vērtības m' atrodas īstā vērtība, ir mazāks nekā jebkuram citam novērtējumam $\hat{\pi}$, pie dotajiem apriorajiem uzskatiem un novērotajiem datiem. Aposteriorā vidējā vērtība ir optimālais *post-datu* novērtējums. Tāpēc aposteriorā vidējā vērtība ir visplašāk izmantotais Beijesa novērtējums.

4.6. Beijesa ticamības intervāls

Ir iespējami vairāki intervāli ar vienādu (aposterioro) varbūtību. Priekšroka tiek dota šaurākajam no tiem. Tā galapunkti būs tur, kur aposteriorā blīvuma funkcijai ir vienādi augstumi un laukums zem blīvuma funkcijas šajā intervālā ir $1 - \alpha$. Laukumiem zem kreisās un labās blīvuma funkcijas "astēm" nav obligāti jābūt vienādiem. Tomēr bieži vieglāk ir kopējo laukumu zem blīvuma funkcijas sadalīt divās vienādās daļās un atrast divpusējo intervālu. [8]

Beijesa ticamības intervāls priekš π

Ja tika izmantots $Beta(a, b)$ apriorais sadalījums, tad $\pi|y$ aposteriorais sadalījums ir $Beta(a', b')$. Divpusējo 95% Beijesa ticamības intervālu priekš π var atrast, aprēķinot starpību starp 97.5-to un 2.5-to procentīli. $Beta(a', b')$ aposterioro sadalījumu aproksimē ar normālo sadalījumu, kuram ir tāda pati vidējā vērtība un dispersija:

$$(\pi|y) \text{ ir aptuveni } N(m'; (s')^2),$$

kur aposteriorā vidējā vērtība ir uzdota ar vienādojumu (4.8) un aposteriorā dispersija ir izteikta ar vienādojumu (4.10). $(1 - \alpha) \times 100\%$ Beijesa ticamības apgabals priekš π ir aptuveni

$$m' \pm z_{\frac{\alpha}{2}} \times s',$$

kur $z_{\frac{\alpha}{2}}$ vērtība tiek atrasta no standartnormālā sadalījuma. Aproksimācija strādā ļoti labi, ja gan $a' \geq 10$, gan $b' \geq 10$.

5. Beijesa un klasiskās proporcijas problemātiku salīdzinājums

5.1. Punktveida novērtējumi

Statistiku, kas tiek aprēķināta no izlases datiem un tiek izmantota, lai novērtētu nezināmo parametru sauksim par parametra *novērtētāju*, un vērtību, kuru tā pieņem pie faktiskajiem izlases datiem - par *novērtējumu*. Aposteriorā sadalījuma rezumēšanai izmantosim aposterioro vidējo vērtību, jo tā minimizē aposterioro vidējo kvadrātisko kļūdu.

Nenovirzīti novērtētāji

Novērtētājs $\hat{\theta}$ tiek saukts par *nenovirzītu*, ja tā izlases sadalījuma vidējā vērtība ir patiesā parametra vērtība, t.i.,

$$E(\hat{\theta}) = \int \hat{\theta} f(\hat{\theta}|\theta) d\hat{\theta} = \theta,$$

kur $f(\hat{\theta}|\theta)$ ir novērtētāja $\hat{\theta}$ izlases sadalījums, ja dots parametrs θ. Klasiskā statistika liek uzsvaru uz nenovirzītiem novērtētājiem, jo visu iespējamo gadījuma izlašu vidējais nenovirzītais novērtētājs dod patieso vērtību. Novērtētāja $\hat{\theta}$ novirze ir

$$bias(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

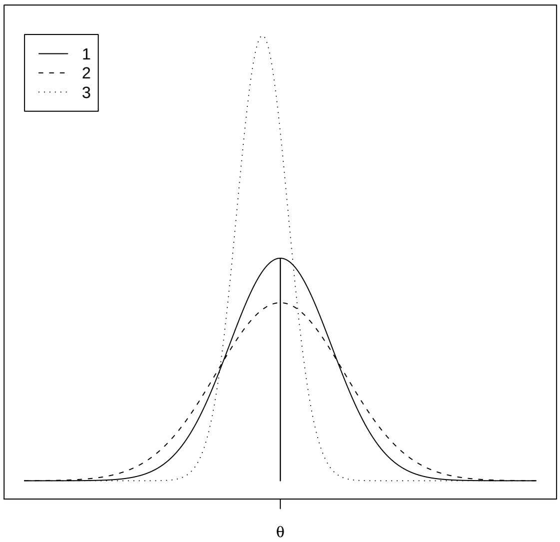
Turpretim Beijesa statistika neliek uzsvaru uz nenovirzītiem novērtētājiem. Faktiski Beijesa novērtētāji parasti ir novirzīti.

Minimālās dispersijas nenovirzīts novērtētājs

Klasiskajā statistikā minimālās dispersijas nenovirzīti novērtētāji bieži tiek uzskatīti par *labākajiem* novērtētājiem.

Tomēr ir iespējami novirzīti novērtētāji, kuri vidēji ir tuvāki patiesajai vērtībai nekā vislabākais nenovirzītais novērtētājs. Jāapskata iespējamais kompromiss starp novirzi un dispersiju. 2. attēls rāda trīs iespējamo θ novērtētāju izlases sadalījumus.

Novērtētājs 1 un novērtētājs 2 tiek uzskatīti par nenovirzītiem novērtētājiem. Novērtētājs 1 ir *labākais nenovirzītais* novērtētājs, jo starp nenovirzītajiem novērtētājiem tam ir vismazākā dispersija. Novērtētājs 3 tiek uzskatīts par novirzītu novērtētāju, bet tam ir



2. att. Trīs novērtētāju izlases sadalījumi.

mazāka dispersija nekā novērtētājam 1. Nepieciešama kaut kāda veida salīdzināšana, kura iekļauj novirzītus novērtētājus, lai uzzinātu, kurš vidēji būs tuvākais parametra vērtībai.

Novērtētāja vidējā kvadrātiskā kļūda

(Klasiskajā statistikā) Novērtētāja $\hat{\theta}$ vidējā kvadrātiskā kļūda ir

$$MS(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = \int (\hat{\theta} - \theta)^2 f(\hat{\theta}|\theta) d\hat{\theta}.$$

Tā tiek aprēķināta no novērtētāja izlases sadalījuma, kas nozīmē, ka vidējā kļūda tiek aprēķināta no visām iespējamajām izlasēm pie dotās fiksētās parametra vērtības. Tā *nav* aposteriorā vidējā kvadrātiskā kļūda, kas aprēķināta no aposteriorā sadalījuma. Novērtētāja vidējā kvadrātiskā kļūda ir novirzes kvadrāta un novērtētāja dispersijas summa:

$$MS(\hat{\theta}) = bias(\hat{\theta})^2 + D(\hat{\theta}). \tag{5.1}$$

Tas ir labāks (klasiskais) kritērijs spriešanai par novērtētāju nekā novirze vai dispersija vieni paši. Novērtētājs, kuram ir mazāka vidējā kvadrātiskā kļūda, ir tuvāk patiesajai vērtībai vidēji pa visām iespējamajām izlasēm.

5.2. Proporcijas novērtētāju salīdzinājums

Beijesa novērtētājiem bieži ir mazākas vidējās kvadrātiskās kļūdas nekā klasiskajiem novērtētājiem, t.i., vidēji tie ir tuvāki patiesajai vērtībai. Proporcijas π klasiskais novērtētājs ir

$$\hat{\pi}_f = \frac{y}{n},$$

kur y, veiksmju skaitam n mēģinājumos, ir $Bi(n, \pi)$ sadalījums. $\hat{\pi}_f$ ir nenovirzīts un $D(\hat{\pi}_f) = \frac{\pi \times (1-\pi)}{n}$. Tādēļ $\hat{\pi}_f$ vidējā kvadrātiskā kļūda ir vienāda ar

$$MS(\hat{\pi}_f) = 0^2 + D(\hat{\pi}_f) = \frac{\pi \times (1-\pi)}{n}.$$

Pieņemsim, ka par proporcijas π Beijesa novērtējumu tiek izmantota aposteriorā vidējā vērtība, kur tiek izmantots $Beta(1, 1)$ apriorais sadalījums (vienmērīgais apriorais sadalījums). Novērtētājs ir aposteriorā vidējā vērtība

$$\hat{\pi}_B = m' = \frac{a'}{a' + b'},$$

kur $a' = 1 + y$ un $b' = 1 + n - y$. To var pārrakstīt kā lineāru funkciju no y:

$$\hat{\pi}_B = \frac{y+1}{n+2} = \frac{y}{n+2} + \frac{1}{n+2}.$$

$\hat{\pi}_B$ izlases sadalījuma vidējā vērtība ir

$$\frac{n\pi}{n+2} + \frac{1}{n+2}$$

un $\hat{\pi}_B$ izlases sadalījuma dispersija ir

$$\left[\frac{1}{n+2}\right]^2 \times n\pi(1-\pi).$$

Tādēļ no vienādojuma (5.1) vidējā kvadrātiskā kļūda ir

$$\begin{aligned} MS(\hat{\pi}_B) &= \left[\frac{n\pi}{n+2} + \frac{1}{n+2} - \pi\right]^2 + \left[\frac{1}{n+2}\right]^2 \times n\pi(1-\pi) \\ &= \left[\frac{1-2\pi}{n+2}\right]^2 + \left[\frac{1}{n+2}\right]^2 \times n\pi(1-\pi). \end{aligned}$$

Piemēram, pieņemsim, ka $\pi = 0.4$ un izlases apjoms ir $n = 10$. Tad

$$MS(\hat{\pi}_f) = \frac{0.4 \times 0.6}{10} = 0.024$$

un

$$MS(\hat{\pi}_B) = \left[\frac{1 - 2 \times 0.4}{12}\right]^2 + \left[\frac{1}{12}\right]^2 \times 10 \times 0.4 \times 0.6 = 0.0169.$$

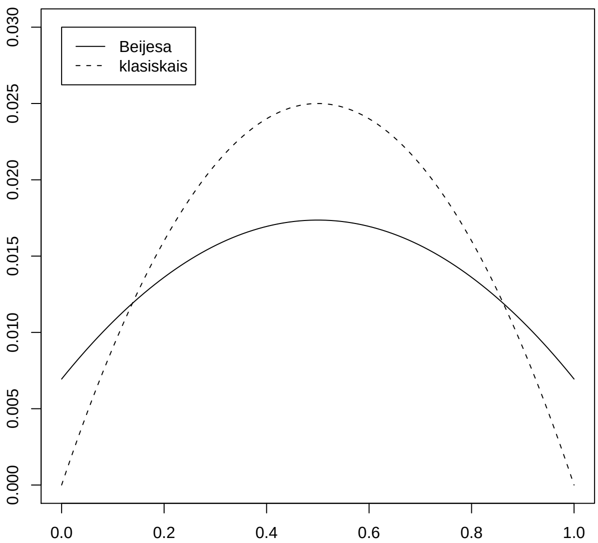
Ja $\pi = 0.5$ un $n = 10$, tad

$$MS(\hat{\pi}_f) = \frac{0.5 \times 0.5}{10} = 0.025$$

un

$$MS(\hat{\pi}_B) = \left[\frac{1 - 2 \times 0.5}{12}\right]^2 + \left[\frac{1}{12}\right]^2 \times 10 \times 0.5 \times 0.5 = 0.01736.$$

Tātad vidēji (šīm divām π vērtībām) Beijesa aposteriorais novērtētājs ir tuvāks patiesajai vērtībai nekā klasiskais novērtētājs. 3. attēls rāda vidējo kvadrātisko kļūdu Beijesa novērtētājam un klasiskajam novērtētājam kā funkciju no π. Gandrīz visā (bet ne pilnīgi visā) apgabalā Beijesa novērtētājs (izmantojot vienmērīgo aprioro sadalījumu) ir labāks nekā klasiskais novērtētājs.



3. att. Vidējā kvadrātiskā kļūda diviem novērtētājiem.

5.3. Ticamības intervālu novērtēšana

Klasiskie ticamības intervāli

Klasiskais $(1 - \alpha) \times 100\%$ ticamības intervāls parametram θ ir intervāls (l, u) tāds, ka

$$P(l \leq \theta \leq u) = 1 - \alpha.$$

Šī varbūtība tiek atrasta izmantojot parametra novērtētāja izlases sadalījumu.

Parametrs θ tiek uzskatīts par fiksētu, bet nezināmu konstanti. Galapunkti l un u ir gadījuma lielumi, jo tie ir atkarīgi no gadījuma izlases. Ja tiek ievietoti faktiskie izlases dati, kuri tika novēroti gadījuma izlasei, un tiek aprēķinātas l un u vērtības, tad neatliek nekas, kam būtu gadījuma raksturs. Intervāls vai nu satur, vai nesatur nezināmo fiksēto parametru, un nav zināms, kurš no šiem abiem gadījumiem izpildās. Intervāls vairs nevar tikt uzskatīts par varbūtību intervālu.

Pareizā interpretācija ir, ka $(1-\alpha)\times 100\%$ no šādā veidā aprēķinātajiem *gadījuma* intervāliem saturēs patieso vērtību. Tāpēc mums ir $(1-\alpha)\times 100\%$ *pārliecība*, ka mūsu intervāls satur patieso parametra vērtību.

Bieži izmantotā novērtētāja izlases sadalījums ir aptuveni normāls ar vidējo vērtību, kas ir vienāda ar patieso vērtību. Šajā gadījumā klasiskais ticamības intervāls ir formā

$$\textit{novērtētājs} \pm \textit{kritiskā vērtība} \times \textit{novērtētāja standartnovirze},$$

kur kritiskā vērtība nāk no *standartnormālā sadalījuma*. Piemēram, ja n ir liels, tad izlases proporcija

$$\hat{\pi}_f = \frac{y}{n}$$

ir aptuveni normāla ar vidējo vērtību π un standartnovirzi $\sqrt{\frac{\pi(1-\pi)}{n}}$. Tas dod aptuveno $(1-\alpha)\times 100\%$ divpusējo klasisko ticamības intervālu priekš π:

$$\hat{\pi}_f \pm z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{\pi}_f(1-\hat{\pi}_f)}{n}}. \tag{5.2}$$

Klasisko un Beijesa ticamības intervālu priekš π salīdzinājums

Varbūtību aprēķini klasiskajam ticamības intervālam balstās uz statistikas izlases sadalījumu. Tādēļ varbūtības ir *pirms-datu*. Tās nav atkarīgas no konkrētās novērotās izlases. Tas ir pretstats no aposteriorā sadalījuma aprēķinātajam Beijesa ticamības intervālam, kuram ir tieša varbūtības interpretācija pie novērotajiem izlases datiem. Citiem vārdiem, tās ir *pēc-datu* varbūtības.

5.4. Vienpusējās hipotēzes pārbaude

Klasiskā vienpusējās hipotēzes pārbaude

Piemērs 3. *Pieņemsim, ka jānosaka, vai pacientiem sniegtā jaunā ārstēšana ir labāka par standarta ārstēšanu. Ja tā ir, tad pacientu daļai π, kuriem kļūst labāk pēc jaunās ārstēšanas, jābūt lielākai par pacientu daļu π_0, kurai kļūst labāk pēc standarta ārstēšanas. No vēsturiskajiem ierakstiem ir zināms, ka $\pi_0 = 0.6$. 10 pacietu gadījuma grupai tiek sniegta jaunā ārstēšana. Pacientu skaits Y, kuriem kļūst labāk pēc jaunās ārstēšanas,* *būs sadalīts $Bi(n, \pi)$. Tiek novērots, ka labāk kļūst 8 pacientiem. Tas ir labāks rezultāts nekā tika sagaidīts, ja $\pi = 0.6$. Bet vai var secināt, ka $\pi > 0.6$ 5% nozīmības līmenī?*

Klasiskā vienpusējās hipotēzes pārbaude:

1. *Tiek izvirzīta hipotēze par (fiksētu, bet nezināmu) parametru. Piemēram, $H_0 : \pi \leq 0.6$.*
2. *Alternatīvā hipotēze ir $H_1 : \pi > 0.6$.*
3. *Testa statistikas nulles sadalījums ir testa statistikas izlases sadalījums, ja dots, ka H_0 ir patiesa. Šajā gadījumā tas būs $Bi(n, 0.6)$, kur $n = 10$ ir pacientu skaits, kuriem sniegta jaunā ārstēšana.*
4. *Tiek izvēlēts nozīmības līmenis $\alpha = 5\%$.*
5. *Noraidīšanas apgabals ir izvēlēts tā, lai tam zem nulles sadalījuma būtu varbūtība α. Ja izvēlamies noraidīšanas apgabalu $y \geq 9$, tad $\alpha = 0.0463$. Nulles sadalījums ar noraidīšanas apgabalu vienpusējās hipotēzes pārbaudei ir parādīts 15.tabulā.*
6. *Ja testa statistikas vērtība dotajai izlasei atrodas noraidīšanas apgabalā, tad līmenī α nulles hipotēze H_0 tiek noraidīta. Pretējā gadījumā nevar noraidīt H_0. Šajā gadījumā tika novērtos $y = 8$. Tas atrodas pieņemšanas apgabalā.*
7. *p-vērtība$= 0.1672$.*
8. *Ja p-vērtība$< \alpha$, tad testa statistika atrodas noraidīšanas apgabalā, un otrādi. Tāpēc ekvivalents hipotēžu pārbaudes veids ir noraidīt H_0, ja* p-vērtība$< \alpha$. *Apskatot abus veidus, nevar noraidīt nulles hipotēzi $H_0 : \pi \leq 0.6$. $y = 8$ atrodas pieņemšanas apgabalā un p-vērtība> 0.05.*

Testa *p-vērtība nav* aposteriorā varbūtība, ka nulles hipotēze ir patiesa pie dotajiem datiem. Tā ir "astes" varbūtība, kura aprēķināta izmantojot nulles sadalījumu. Binomiālā gadījumā

$$\textit{p-vērtība} = \sum_{y_{nov}}^{n} f(y|\pi_0),$$

kur y_{nov} ir novērotā y vērtība. Klasiskā hipotēžu pārbaude izmanto varbūtību, kas aprēķināta no visām iespējamajām datu kopām, kuras varētu realizēties (pie fiksētas parametra vērtības), bet hipotēze ir par to, ka parametra vērtība atrodas kādā vērtību apgabalā.

15. tabula Y nulles sadalījums ar noraidīšanas apgabalu vienpusējās hipotēzes pārbaudei

Vērtība	$f(y\|\pi = 0.6)$	Apgabals
0	0.0001	pieņemšanas
1	0.0016	pieņemšanas
2	0.0106	pieņemšanas
3	0.0425	pieņemšanas
4	0.1115	pieņemšanas
5	0.2007	pieņemšanas
6	0.2508	pieņemšanas
7	0.2150	pieņemšanas
8	0.1209	pieņemšanas
9	0.0403	noraidīšanas
10	0.0060	noraidīšanas

Beijesa vienpusējās hipotēzes pārbaude

Pārbaudīsim

$$H_0 : \pi \leq \pi_0 \quad \text{pret} \quad H_1 : \pi > \pi_0$$

nozīmības līmenī α, izmantojot Beijesa metodi. Aprēķināsim aposterioro varbūtību, ka nulles hipotēze ir patiesa, integrējot aposterioro blīvumu:

$$P(H_0 : \pi \leq \pi_0 | y) = \int_0^{\pi_0} g(\pi|y)\, d\pi.$$

Nulles hipotēze tiek noraidīta, ja šī aposteriorā varbūtība ir mazāka nekā nozīmības līmenis α.

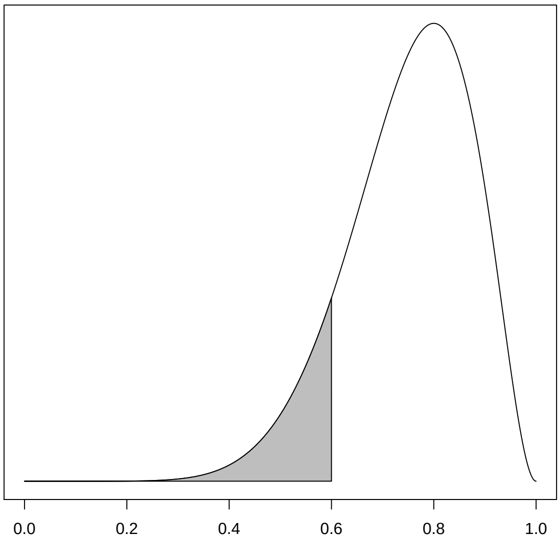
Piemērs 3. *(turpinājums) Pieņemsim, ka priekš π izmantojam $Beta(1,1)$ apriora sadalījumu. Ja dots $y = 8$, tad aposteriorais blīvums ir $Beta(9,3)$. Nulles hipotēzes aposteriorā varbūtība ir*

$$P(\pi \leq 0.6 | y = 8) = \int_0^{0.6} \frac{\Gamma(12)}{\Gamma(3)\Gamma(9)} \pi^8 (1-\pi)^2\, d\pi = 0.1189.$$

Tā nav mazāka par 0.05, tāpēc nevar noraidīt nulles hipotēzi 5% nozīmības līmenī. 4. attēlā parādīts aposteriorais blīvums. Nulles hipotēzes varbūtība ir laukums zem līknes pa kreisi no $\pi = 0.6$.

5.5. Divpusējās hipotēzes pārbaude

Klasiskā divpusējās hipotēzes pārbaude



4. att. Nulles hipotēzes aposteriorā varbūtība, $H_0 : \pi \leq 0.6$ ir pelēkais laukums.

Nulles sadalījums tiek novērtēts pie π_0 un noraidīšanas apgabals ir divpusējs. Šai pārbaudei aprēķinātās *p-vērtības* arī ir divpusējas.

Piemērs 4. *Monēta tiek mesta 15 reizes un 10 reizes uzkrīt ģērbonis. Vai ar 10 ģērboņa uzkrišanas reizēm no 15 metieniem ir pietiekami, lai noteiktu, ka monēta nav simetriska? Citiem vārdiem, vai varbūtība π, ka uzkritīs ģērbonis, atšķiras no $\frac{1}{2}$?*

Klasiskā divpusējās hipotēzes pārbaude:

1. *Tiek izvirzīta hipotēze par fiksētu, bet nezināmu parametru π. $H_0 : \pi = 0.5$*

2. *Alternatīvā hipotēze ir $H_1 : \pi \neq 0.5$. Noraidīšanas apgabals būs divpusējs.*

3. *Nulles sadalījums ir Y izlases sadalījums, ja nulles hipotēze ir patiesa. Tas ir $Bi(n = 15, \pi = 0.5)$.*

4. *Tiek izvēlēts 5% nozīmības līmenis.*

5. *Noraidīšanas apgabals ir izvēlēts tā, ka tam ir varbūtība α zem nulles sadalījuma. Ja noraidīšanas apgabalu izvēlas $Y \leq 3 \cup Y \geq 12$, tad $\alpha = 0.0352$. Nulles sadalījums un noraidīšanas apgabals divpusējai hipotēzei ir parādīti 16.tabulā.*

6. *Ja testa statistikas vērtība atrodas noraidīšanas apgabalā, tad līmenī α nulles hipotēze H_0 tiek noraidīta. Pretējā gadījumā H_0 nevar noraidīt. Šajā gadījumā ir novērots $y = 10$. Tas atrodas apgabalā, kurā nevar noraidīt nulles hipotēzi.*

7. *p-vērtība*$= P(Y \geq 10) + P(Y \leq 5) = 0.302$. *Tā ir lielāka par* α*, tāpēc nevar noraidīt* H_0.

16. tabula Y nulles sadalījums ar noraidīšanas apgabalu divpusējās hipotēzes pārbaudei

Vērtība	$f(y\|\pi = 0.5)$	Apgabals
0	0.0000	noraidīšanas
1	0.0005	noraidīšanas
2	0.0032	noraidīšanas
3	0.0139	noraidīšanas
4	0.0417	pieņemšanas
5	0.0916	pieņemšanas
6	0.1527	pieņemšanas
7	0.1964	pieņemšanas
8	0.1964	pieņemšanas
9	0.1527	pieņemšanas
10	0.0916	pieņemšanas
11	0.0417	pieņemšanas
12	0.0139	noraidīšanas
13	0.0032	noraidīšanas
14	0.0005	noraidīšanas
15	0.0000	noraidīšanas

Starp divpusējo hipotēžu pārbaudi un klasiskajiem ticamības intervāliem pastāv cieša saistība. Ja tiek pārbaudīta divpusējā hipotēze pie līmeņa α, tad tam ir atbilstošs parametra $(1 - \alpha) \times 100\%$ klasiskais ticamības intervāls. Ja nulles hipotēze $H_0 : \pi = \pi_0$ tiek noraidīta, tad vērtība π_0 atrodas ārpus klasiskā ticamības intervāla, un otrādi. Klasiskais ticamības intervāls "apkopo" visas iespējamās nulles hipotēzes, kuras tiktu akceptētas, ja tās tiktu pārbaudītas.

Beijesa divpusējās hipotēzes pārbaude

Ja izmantojam nepārtrauktu aprioro sadalījumu, tad iegūsim nepārtrauktu aposterioro sadalījumu. Precīzās parametra vērtības varbūtība, kuru reprezentē punktveida nulles hipotēze, būs vienāda ar nulli. Tāpēc hipotēzes pārbaudei nevar izmantot aposterioro varbūtību. Tās vietā izmantosim atbilstību starp hipotēžu pārbaudi un klasiskajiem ticamības intervāliem, klasisko ticamības intervālu vietā lietojot Beijesa ticamības intervālus.

Aprēķinām $(1-\alpha)\times 100\%$ Beijesa ticamības intervālu priekš π. Ja π_0 atrodas Beijesa ticamības intervālā, tad pieņemam nulles hipotēzi $H_0 : \pi = \pi_0$, un ja π_0 atrodas ārpus Beijesa ticamības intervāla, tad H_0 noraidām.

Piemērs 4. *(turpinājums) Ja izmantojam vienmērīgo apriori sadalījumu, tad aposteriorais sadalījums ir* $Beta(10+1, 5+1)$ *sadalījums.* 95% *Beijesa ticamības intervāls priekš* π*, kurš tiek atrasts izmantojot normālo aproksimāciju, ir*

$$\frac{11}{17} + 1.96 \times \sqrt{\frac{11\times 6}{((11+6)^2 \times (11+6+1))}} = 0.647 \pm 0.221 = (0.426, 0.868).$$

$\pi = 0.5$ *atrodas Beijesa ticamības intervālā, tāpēc nevar noraidīt* H_0*.*

6. Beijesa problemātika normālā sadalījuma vidējai vērtībai

6.1. Beijesa teorēma normālā sadalījuma vidējai vērtībai ar diskrētu aprioro sadalījumu

Vienam novērojumam no normālā sadalījuma

Ņemsim vienu novērojumu no nosacītā blīvuma $f(y|\mu)$, kurš ir normāli sadalīts ar zināmu dispersiju σ^2. Pieņemsim, ka vidējai vērībai ir iespējamas tikai m vērtības $\mu_1, \dots, \mu_m$. Šīm vērtībām izvēlamies diskrētu aprioro varbūtību sadalījumu. Ja nav nekādas apriorās informācijas, tad visām vērtībām piešķirsim vienādas apriorās varbūtības.

Aposteriorais sadalījums tiek definēts ar vienādojumu (1.1) (7.lpp.).

Atsevišķa novērojuma ticamība

Nosacītais $y|\mu$ novērojumu sadalījums ir normālais ar vidējo vērtību μ un zināmu dispersiju σ^2. Tā blīvums ir

$$f(y|\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}.$$

Katras parametra vērtības ticamība ir novērojumu sadalījuma vērtība pie novērotās vērtības. Formulas daļa, kas nav atkarīga no parametra μ, ir vienāda visām parametra vērtībām, tāpēc to iekļauj proporcionalitātes konstantē. Tādēļ ticamība ir formā

$$f(y|\mu) \propto e^{-\frac{1}{2\sigma^2}(y-\mu)^2}, \tag{6.1}$$

kur y ir konstants (novērotā vērtība) un μ mainās pa visām tā iespējamām vērtībām.

Ticamību aposteriorā sadalījuma aprēķināšanai var iegūt no "normālā sadalījuma ordinātu" tabulas vai no normālās blīvuma funkcijas, kas uzdota ar vienādojumu (6.1), saglabājot y fiksētu (novērotā vērtība) un mainot μ pa visām iespējamajām vērtībām.

Piemērs 5. *Pieņemsim, ka $(y|\mu) \sim N(\mu, \sigma^2 = 1)$, un ka μ var pieņemt tikai piecas vērtības: 2, 2.5, 3, 3.5 un 4. Pieņemsim, ka tās ir vienādi iespējamas. Ņemam vienu y novērojumu un iegūstam vērtību $y = 3.2$. Ja*

$$z = \frac{y-\mu}{\sigma},$$

tad ticamības $f(z)$ vērtības tiek atrastas no "normālā sadalījuma ordinātu" tabulas. Rezultāti ir parādīti 17.tabulā.

17. tabula: Metode 1: aposteriorā sadalījuma atrašana, izmantojot "normālā sadalījuma ordinātes"

μ	Apriorā varbūtība	z	Ticamība	Apriorā varbūtība × Ticamība	Aposteriorā varbūtība
2.0	0.2	-1.2	0.1942	0.03884	0.1238
2.5	0.2	-0.7	0.3123	0.06246	0.1991
3.0	0.2	-0.2	0.3910	0.07820	0.2493
3.5	0.2	0.3	0.3814	0.07628	0.2431
4.0	0.2	0.8	0.2897	0.05794	0.1847
				0.31372	1.0000

Ja ticamība tiek aprēķināta, izmantojot normālā blīvuma formulu, tad ticamība ir formā (6.1)*, kur $y = 3.2$ un μ mainās pa visām iespējamajām vērtībām. Rezultāti ir parādīti 18.tabulā. Šie rezultāti saskan ar to, kas tika aprēķināts izmantojot 1. metodi, neņemot vērā mazas noapaļošanas kļūdas.*

18. tabula: Metode 2: aposteriorā sadalījuma atrašana, izmantojot ticamību no normālā blīvuma formulas

μ	Apriorā varbūtība	Ticamība *(ignorējot konstanti)*	Apriorā varbūtība × Ticamība	Aposteriorā varbūtība
2.0	0.2	$e^{-\frac{1}{2}(3.2-2.0)^2} = 0.4868$	0.0974	0.1239
2.5	0.2	$e^{-\frac{1}{2}(3.2-2.5)^2} = 0.7827$	0.1565	0.1990
3.0	0.2	$e^{-\frac{1}{2}(3.2-3.0)^2} = 0.9802$	0.1960	0.2493
3.5	0.2	$e^{-\frac{1}{2}(3.2-3.5)^2} = 0.9560$	0.1912	0.2432
4.0	0.2	$e^{-\frac{1}{2}(3.2-4.0)^2} = 0.7261$	0.1452	0.1846
			0.7863	1.0000

Normālo novērojumu gadījuma izlase

Parasti viena novērojuma vietā ir gadījuma izlase $y_1, \dots, y_n$. Gadījuma izlasē novērojumi ir neatkarīgi viens no otra, tāpēc izlases kopējā ticamība ir

$$f(y_1, \dots, y_n|\mu) = f(y_1|\mu) \times f(y_2|\mu) \times \dots f(y_n|\mu).$$

Tāpēc Beijesa teorēma ar diskrētu aprioro sadalījumu tiek uzdota ar vienādojumu

$$g(\mu|y_1, \dots, y_n) \propto g(\mu) \times f(y_1|\mu) \times \dots f(y_n|\mu).$$

Apskatīsim gadījumu, kur katra novērojuma $y_j|\mu$ sadalījums ir normālais ar vidējo vērtību μ un zināmu dispersiju σ^2.

Aposterioro varbūtību atrašana, analizējot novērojumus pēc kārtas pa vienam. Iepriekšējā novērojuma aposteriorās varbūtības tiks izmantotas par nākamā novērojuma apriorajām varbūtībām.

Piemērs 6. *No $N(\mu, \sigma^2 = 1)$ sadalījuma tiek ņemta gadījuma izlase ar četriem novērojumiem. Novērojumi ir 3.2, 2.2, 3.6 un 4.1. Iespējamās μ vērtības ir 2, 2.5, 3, 3.5 un 4. Izmantosim aprioro sadalījumu, kurš tām piešķir vienādus svarus. Izmantosim Beijesa teorēmu, lai atrastu aposterioro sadalījumu, ja dota visa izlase. Rezultāti, kas iegūti analizējot novērojumus pa vienam, parādīti 19.tabulā. Lielas izlases gadījumā tas sagādā daudz darba. Daudz vieglāk ir izmantot visu izlasi uzreiz.*

Aposterioro varbūtību atrašana, analizējot visus novērojumus reizē. Visu novērojumu kopējā ticamība ir

$$f(y_1, \ldots, y_n|\mu) \propto e^{-\frac{1}{2\sigma^2}(y_1-\mu)^2} \times e^{-\frac{1}{2\sigma^2}(y_2-\mu)^2} \times \cdots \times e^{-\frac{1}{2\sigma^2}(y_n-\mu)^2}.$$

Saskaitot kāpinātājus, iegūst

$$f(y_1, \ldots, y_n|\mu) \propto e^{-\frac{1}{2\sigma^2}\left[(y_1-\mu)^2+(y_2-\mu)^2+\cdots+(y_n-\mu)^2\right]}.$$

Apskatām iekavas

$$\left[(y_1-\mu)^2 + \cdots + (y_n-\mu)^2\right] = y_1^2 - 2y_1\mu + \mu^2 + \cdots + y_n^2 - 2y_n\mu + \mu^2$$

un savelkam līdzīgos locekļus, lai iegūtu

$$= (y_1^2 + \cdots + y_n^2) - 2\mu(y_1 + \cdots + y_n) + n\mu^2.$$

Ja to ievieto atpakaļ, iznes n pirms iekavām un atdala pilnos kvadrātus, tad iegūst

$$\begin{aligned} f(y_1, \ldots, y_n|\mu) &\propto e^{-\frac{n}{2\sigma^2}\left[\mu^2 - 2\mu\bar{y} + \bar{y}^2 - \bar{y}^2 + \frac{y_1^2+\cdots+y_n^2}{n}\right]} \\ &\propto e^{-\frac{n}{2\sigma^2}\left[\mu^2 - 2\mu\bar{y} + \bar{y}^2\right]} \times e^{-\frac{n}{2\sigma^2}\left[\frac{y_1^2+\cdots+y_n^2}{n} - \bar{y}^2\right]}. \end{aligned}$$

Normālās gadījuma izlases $y_1, \ldots, y_n$ ticamība ir proporcionāla izlases vidējās vērtības $\bar{y}$ ticamībai. Ja ticamības funkcijas daļu, kura nesatur μ, iekļauj proporcionalitātes konstantē, tad iegūst

$$f(y_1, \ldots, y_n|\mu) \propto e^{-\frac{1}{2\sigma^2/n}(\bar{y}-\mu)^2}.$$

19. tabula Novērojumu analizēšana pa vienam

μ	Apriorā varbūtība$_1$	Ticamība$_1$ (*ignorējot konstanti*)	Apriorā varbūtība$_1$ $\times$Ticamība$_1$	Aposteriorā varbūtība$_1$
2.0	0.2	$e^{-\frac{1}{2}(3.2-2.0)^2} = 0.4868$	0.0974	0.1239
2.5	0.2	$e^{-\frac{1}{2}(3.2-2.5)^2} = 0.7827$	0.1565	0.1990
3.0	0.2	$e^{-\frac{1}{2}(3.2-3.0)^2} = 0.9802$	0.1960	0.2493
3.5	0.2	$e^{-\frac{1}{2}(3.2-3.5)^2} = 0.9560$	0.1912	0.2432
4.0	0.2	$e^{-\frac{1}{2}(3.2-4.0)^2} = 0.7261$	0.1452	0.1846
			0.7863	1.0000
μ	Apriorā varbūtība$_2$	Ticamība$_2$ (*ignorējot konstanti*)	Apriorā varbūtība$_2$ $\times$Ticamība$_2$	Aposteriorā varbūtība$_2$
2.0	0.1239	$e^{-\frac{1}{2}(2.2-2.0)^2} = 0.9802$	0.1214	0.1916
2.5	0.1990	$e^{-\frac{1}{2}(2.2-2.5)^2} = 0.9560$	0.1902	0.3002
3.0	0.2493	$e^{-\frac{1}{2}(2.2-3.0)^2} = 0.7261$	0.1810	0.2857
3.5	0.2432	$e^{-\frac{1}{2}(2.2-3.5)^2} = 0.4296$	0.1045	0.1649
4.0	0.1846	$e^{-\frac{1}{2}(2.2-4.0)^2} = 0.1979$	0.0365	0.0576
			0.6336	1.0000
μ	Apriorā varbūtība$_3$	Ticamība$_3$ (*ignorējot konstanti*)	Apriorā varbūtība$_3$ $\times$Ticamība$_3$	Aposteriorā varbūtība$_3$
2.0	0.1916	$e^{-\frac{1}{2}(3.6-2.0)^2} = 0.2780$	0.0533	0.0792
2.5	0.3002	$e^{-\frac{1}{2}(3.6-2.5)^2} = 0.5461$	0.1639	0.2573
3.0	0.2857	$e^{-\frac{1}{2}(3.6-3.0)^2} = 0.8353$	0.2386	0.3745
3.5	0.1649	$e^{-\frac{1}{2}(3.6-3.5)^2} = 0.9950$	0.1641	0.2576
4.0	0.0576	$e^{-\frac{1}{2}(3.6-4.0)^2} = 0.9231$	0.0532	0.0835
			0.6731	1.0000
μ	Apriorā varbūtība$_4$	Ticamība$_4$ (*ignorējot konstanti*)	Apriorā varbūtība$_4$ $\times$Ticamība$_4$	Aposteriorā varbūtība$_4$
2.0	0.0792	$e^{-\frac{1}{2}(4.1-2.0)^2} = 0.1103$	0.0087	0.0149
2.5	0.2573	$e^{-\frac{1}{2}(4.1-2.5)^2} = 0.2780$	0.0715	0.1226
3.0	0.3745	$e^{-\frac{1}{2}(4.1-3.0)^2} = 0.5461$	0.2045	0.3508
3.5	0.2576	$e^{-\frac{1}{2}(4.1-3.5)^2} = 0.8352$	0.2152	0.3691
4.0	0.0835	$e^{-\frac{1}{2}(4.1-4.0)^2} = 0.9950$	0.0838	0.1425
			0.5830	1.0000

Šai ticamībai ir $N(\mu, \frac{\sigma^2}{n})$ sadalījuma forma. Izlases vidējā vērtība $\bar{y}$ ir normāli sadalīta ar vidējo vērtību μ un dispersiju $\frac{\sigma^2}{n}$. Tāpēc izlases kopējā ticamība ir proporcionāla izlases vidējās vērtības ticamībai

$$f(\bar{y}|\mu) \propto e^{-\frac{1}{2\sigma^2/n}(\bar{y}-\mu)^2}.$$

Piemērs 6. *(turpinājums) Iepriekšējā izlasē vidējā vērtība $\bar{y} = 3.275$. Izmantosim $\bar{y}$ ticamību, kura ir proporcionāla visas izlases ticamībai. Rezultāti ir parādīti 20.tabulā. Trīs* *zīmēs rezultāti sakrīt ar iepriekšējiem rezultātiem. Nelielā atšķirība ceturtajā decimālajā zīmē ir uzkrāto noapaļošanas kļūdu dēļ, kad novērojumi tika analizēti pa vienam. Skaidrs, ka vieglāk ir izmantot $\bar{y}$, lai veiktu Beijesa teorēmas aprēķinus tikai vienreiz.*

20. tabula Analizējot visus novērojumus reizē, izmantojot izlases vidējās vērtības ticamību

μ	Apriorā varbūtība$_1$	Ticamība$_{\bar{y}}$ *(ignorējot konstanti)*	Apriorā varbūtība$_1$ $\times$Ticamība$_{\bar{y}}$	Aposteriorā varbūtība$_{\bar{y}}$
2.0	0.2	$e^{-\frac{1}{2\times 1/4}(3.275-2.0)^2} = 0.0387$	0.0077	0.0157
2.5	0.2	$e^{-\frac{1}{2\times 1/4}(3.275-2.5)^2} = 0.3008$	0.0602	0.1228
3.0	0.2	$e^{-\frac{1}{2\times 1/4}(3.275-3.0)^2} = 0.8596$	0.1719	0.3505
3.5	0.2	$e^{-\frac{1}{2\times 1/4}(3.275-3.5)^2} = 0.9037$	0.1807	0.3685
4.0	0.2	$e^{-\frac{1}{2\times 1/4}(3.275-4.0)^2} = 0.3495$	0.0699	0.1425
			0.4904	1.0000

6.2. Beijesa teorēma normālajai vidējai vērtībai ar nepārtrauktu aprioro sadalījumu

Dota gadījuma izlase $y_1, \ldots, y_n$ no $N(\mu, \sigma^2)$ sadalījuma. Ir daudz reālāk ticēt, ka ir iespējamas visas μ vērtības no kāda intervāla. Tas nozīmē, ka ir jāizmanto nepārtraukts apriorais sadalījums. Tādēļ aposteriorais sadalījums būs formā

$$g(\mu|y_1, \ldots, y_n) = \frac{g(\mu) \times f(y_1, \ldots, y_n|\mu)}{\int g(\mu) \times f(y_1, \ldots, y_n|\mu) d\mu}. \tag{6.2}$$

Normālajam sadalījumam gadījuma izlases ticamība ir proporcionāla izlases vidējās vērtības $\bar{y}$ ticamībai, tāpēc

$$g(\mu|y_1, \ldots, y_n) = \frac{g(\mu) \times e^{-\frac{1}{2\sigma^2/n}(\bar{y}-\mu)^2}}{\int g(\mu) \times e^{-\frac{1}{2\sigma^2/n}(\bar{y}-\mu)^2} d\mu}.$$

Tas ir spēkā jebkuram nepārtrauktam apriorajam blīvumam $g(\mu)$. Tomēr tas prasa integrēšanu, kuru varētu būt nepieciešams veikt skaitliski. Apskatīsim dažus īpašus gadījumus, kur aposterioro sadalījumu var atrast bez integrēšanas.

Plakans apriorais blīvums priekš μ

Plakans apriorais sadalījums katrai iespējamai μ vērtībai piešķir vienādu svaru, $g(\mu) = 1$. Plakans apriorais sadalījums nav īsts apriorais sadalījums, jo $-\infty < \mu < \infty$, tāpēc tā integrālis nevar būt 1. Tomēr šis *neīstais* apriorais sadalījums strādā labi. Kaut arī apriorais sadalījums ir neīsts, bet gadījumā, ja integrālis no aposteriorā sadalījuma ir vienāds ar 1, izvēlētais apriorais sadalījums būs īsts.

Tā kā apriorais sadalījums ir vienāds ar 1, tad aposteriorais sadalījums priekš atsevišķa normālā novērojuma y ir

$$g(\mu|y) \propto e^{-\frac{1}{2\sigma^2}(\mu-y)^2}$$

un priekš normālās gadījuma izlases $y_1, \ldots, y_n$ ir

$$g(\mu|\bar{y}) \propto e^{-\frac{1}{2\sigma^2/n}(\mu-\bar{y})^2}.$$

Normālais apriorais blīvums priekš μ

1. Atsevišķam novērojumam. Novērojums y ir gadījuma lielums, kas ņemts no normālā sadalījuma ar vidējo vērtību μ un zināmu dispersiju σ^2. Mums ir normālais apriorais sadalījums ar vidējo vērtību m un dispersiju s^2. Apriorais blīvums ir

$$g(\mu) \propto e^{-\frac{1}{2s^2}(\mu-m)^2}.$$

Ticamība ir

$$f(y|\mu) \propto e^{-\frac{1}{2\sigma^2}(y-\mu)^2}.$$

Apriorā sadalījuma reizinājums ar ticamību ir

$$g(\mu) \times f(y|\mu) \propto e^{-\frac{1}{2}\left[\frac{(\mu-m)^2}{s^2}+\frac{(y-\mu)^2}{\sigma^2}\right]}.$$

Vienādojot saucējus kāpinātājā, iegūst

$$\propto e^{-\frac{1}{2}\left[\frac{\sigma^2(\mu^2-2\mu m+m^2)+s^2(y^2-2y\mu+\mu^2)}{\sigma^2 s^2}\right]}.$$

Savelkot līdzīgos locekļus,

$$\propto e^{-\frac{1}{2}\left[\frac{(\sigma^2+s^2)\mu^2-2(\sigma^2 m+s^2 y)\mu+m^2\sigma^2+y^2 s^2}{\sigma^2 s^2}\right]}$$

un iznes pirms iekavām $(\sigma^2+s^2)/(\sigma^2 s^2)$. Papildinot līdz pilnam kvadrātam un iekļaujot proporcionalitātes konstantē to daļu, kas nav atkarīga no μ, iegūst

$$\propto e^{-\frac{1}{2\sigma^2 s^2/(\sigma^2+s^2)}\left[\mu^2-2\frac{(\sigma^2 m+s^2 y)}{\sigma^2+s^2}\mu+(\frac{(\sigma^2 m+s^2 y)}{\sigma^2+s^2})^2\right]}$$

$$\propto e^{-\frac{1}{2\sigma^2 s^2/(\sigma^2+s^2)}\left[\mu-\frac{(\sigma^2 m+s^2 y)}{\sigma^2+s^2}\right]^2}.$$

Redzam, ka aposteriorais sadalījums ir normālais, kura vidējā vērtība un dispersija ir uzdotas attiecīgi ar

$$m' = \frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2} \quad \text{un} \quad (s')^2 = \frac{\sigma^2 s^2}{(\sigma^2 + s^2)}. \tag{6.3}$$

Mēs sākām ar $N(m, s^2)$ apriororo sadalījumu un beidzām ar $N[m', (s')^2]$ aposterioro sadalījumu. Tas rāda, ka $N(m, s^2)$ sadalījums ir saistītā saime normālajam novērojumu sadalījumam ar zināmu dispersiju. Tādēļ nav nepieciešama integrēšana, lai aprēķinātu aposterioro sadalījumu. Ir nepieciešams tikai noteikt parametru precizēšanas likumus.

Parametru precizēšanas likums normālo sadalījumu saimei. Vienādojumā (6.3) dotos precizēšanas likumus var vienkāršot. Vispirms formulējam sadalījuma *precizitāti*, kas ir dispersijas apgrieztais lielums. Aposteriorā precizitāte ir

$$\frac{1}{(s')^2} = \left(\frac{\sigma^2 s^2}{\sigma^2 + s^2}\right)^{-1} = \frac{\sigma^2 + s^2}{\sigma^2 s^2} = \frac{1}{s^2} + \frac{1}{\sigma^2}.$$

Tādējādi aposteriorā precizitāte ir vienāda ar apriorās precizitātes un novērojuma precizitātes summu. Aposteriorā vidējā vērtība ir uzdota ar

$$m' = \frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2} = \frac{\sigma^2}{\sigma^2 + s^2} \times m + \frac{s^2}{\sigma^2 + s^2} \times y.$$

To var vienkāršot uz

$$m' = \frac{1/s^2}{1/\sigma^2 + 1/s^2} \times m + \frac{1/\sigma^2}{1/\sigma^2 + 1/s^2} \times y.$$

Šis precizēšanas likums ir spēkā arī plakanam apriorajam sadalījumam. Plakanam apriorajam sadalījumam ir bezgalīga dispersija, tāpēc tam ir nulles precizitāte. Aposteriorā precizitāte būs vienāda ar novērojumu precizitāti

$$1/\sigma^2 = 0 + 1/\sigma^2$$

un aposteriorā dispersija būs vienāda ar novērojuma dispersiju σ^2. Plakanam apriorajam sadalījumam nav labi definēta apriorā vidējā vērtība. Tā var būt jebkāda.

$$\frac{0}{1/\sigma^2} \times \text{jebkas} + \frac{1/\sigma^2}{1/\sigma^2} \times y = y,$$

tāpēc aposteriorā vidējā vērtība, izmantojot plakanu aprioro sadalījumu, ir vienāda ar novērojumu y.

2. Gadījuma izlasei $y_1, \ldots, y_n$. Gadījuma izlase $y_1, \ldots, y_n$ ir ņemta no normālā sadalījuma ar vidējo vērtību μ un zināmu dispersiju σ^2. Apriorais sadalījums ir $N(m, s^2)$, kas uzdots ar

$$g(\mu) \propto e^{\frac{1}{2s^2}(\mu-m)^2}.$$

Izmantosim izlases vidējās vērtības $\bar{y}$, kura ir sadalīta $N(\mu, \frac{\sigma^2}{n})$, ticamību. $\bar{y}$ precizitāte ir $\frac{n}{\sigma^2}$. Tā ir visu gadījuma izlases novērojumu precizitāšu summa.

Problēmu esam reducējuši uz precizēšanu, ja dots atsevišķs novērojums $\bar{y}$, kuru atrisinājām iepriekš. Aposteriorā precizitāte ir

$$\frac{1}{(s')^2} = \frac{1}{s^2} + \frac{n}{\sigma^2} = \frac{\sigma^2 + ns^2}{\sigma^2 s^2}. \tag{6.4}$$

Aposteriorā dispersija ir vienāda ar aposteriorās precizitātes apgriezto lielumu. Aposteriorā vidējā vērtība ir vienāda ar:

$$m' = \frac{1/s^2}{n/\sigma^2 + 1/s^2} \times m + \frac{n/\sigma^2}{n/\sigma^2 + 1/s^2} \times \bar{y}. \tag{6.5}$$

6.3. Normālā apriorā sadalījuma izvēle

Ja novērojums ir no normālā sadalījuma ar zināmu dispersiju, tad saistītā aprioro sadalījumu saime priekš μ ir $N(m, s^2)$. Ja var atrast šīs saimes pārstāvi, kurš saskan ar apriorajiem uzskatiem, tad aposteriorā sadalījuma atrašana, izmantojot Beijesa teorēmu, būs ļoti viegla. Aposteriorais sadalījums arī būs tās pašas saimes pārstāvis un parametri tiks precizēti ar vienkāršiem likumiem, kas uzdoti ar vienādojumiem (6.4) un (6.5). Nav nepieciešams veikt skaitlisku integrēšanu.

Pirmkārt, jāizlemj par aprioro vidējo vērtību m. Pēc tam jāizlemj par aprioro standartnovirzi s. Attālumu starp iespējamo μ lielāko un mazāko vērtību dala ar 6, lai iegūtu aprioro standartnovirzi s.

Apriorā sadalījuma lietderīga pārbaude ir "ekvivalentā izlases apjoma" aplūkošana. Aprioro dispersiju pielīdzina $s^2 = \sigma^2/n_{ekv}$ un atrisina priekš n_{ekv}. Tas saista aprioro precizitāti ar precizitāti no izlases. Ja n_{ekv} ir liels, tad ir ļoti stingri apriorie uzskati par μ. Būs nepieciešams daudz izlases datu, lai aposterioros uzskatus novirzītu tālu prom no apriorajiem uzskatiem. Ja n_{ekv} ir mazs, apriorie uzskati nav stingri un aposteriorie uzskati būs viegli ietekmējami ar daudz mazāku izlases datu apjomu.

Ja no saistītās saimes nevar atrast aprioro sadalījumu, kurš atbilst apriorajiem uzskatiem, tad vajadzētu noteikt aprioros uzskatus punktu izlasei no apgabala, kurš ir iespējams, un lineāri interpolēt starp tiem. Tad aposterioro sadalījumu var noteikt ar vienādojumu (6.2).

6.4. Beijesa ticamības intervāls normālā sadalījuma vidējai vērtībai

Ja dispersija ir zināma

Ja $y_1, \ldots, y_n$ ir gadījuma izlase no $N(\mu, \sigma^2)$ sadalījuma, tad izlases vidējās vērtības $\bar{y}$ sadalījums ir $N(\mu, \sigma^2/n)$. Izmantojot vai nu "plakanu", vai $N(m, s^2)$ aprioro sadalījumu, μ aposteriorais sadalījums pie dotā $\bar{y}$ ir $N[m', (s')^2]$, kurš tiek precizēts saskaņā ar likumiem (6.4) un (6.5).

$(1-\alpha) \times 100\%$ Beijesa ticamības intervāls priekš μ ir

$$m' \pm z_{\frac{\alpha}{2}} \times s'. \tag{6.6}$$

Ja dispersija nav zināma

Ja nav zināma dispersija, tad nav zināma precizitāte, tāpēc nevar tieši izmantot precizēšanas likumus. Tāpēc no datiem aprēķina izlases dispersiju

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

Pēc tam izmanto vienādojumus (6.4) un (6.5), lai atrastu $(s')^2$ un m', kur nezināmās dispersijas σ^2 vietā izmanto izlases dispersiju $\hat{\sigma}^2$.

Lai ņemtu vērā nenoteiktību σ^2 novērtēšanā, Beijesa ticamības intervāls tiek paplašināts, ņemot vērtības nevis no *standartnormālā* sadalījuma, bet gan no *Stjudenta t* sadalījuma . Pareizais Beijesa ticamības intervāls ir

$$m' \pm t_{\frac{\alpha}{2}} \times s'.$$

t vērtība tiek ņemta no rindas $df = n - 1$ (brīvības pakāpju skaits vienāds ar novērojumu skaits mīnus 1).

Ne-normālais apriorais sadalījums

Ja sāk ar ne-normālo aprioro sadalījumu, tad atrod aposterioro sadalījumu priekš μ, izmantojot Beijesa teorēmu, kur ir jāintegrē skaitliski. Aposteriorais sadalījums nebūs normālais. Var atrast $(1-\alpha)\times 100\%$ Beijesa ticamības intervālu, atrodot apakšējo vērtību μ_l un augšējo vērtību μ_u tādas, ka

$$\int_{\mu_l}^{\mu_u} g(\mu|y_1,\ldots,y_n)d\mu = 1-\alpha.$$

Ir daudz šādu vērtību. Labākā μ_l un μ_u izvēle dos īsāko iespējamo Beijesa ticamības intervālu. Šīs vērtības arī apmierina

$$g(\mu_l|y_1,\ldots,y_n) = g(\mu_u|y_1,\ldots,y_n).$$

7. Beijesa un klasiskās vidējās vērtības problemātiku salīdzinājums

7.1. Klasiskā un Beijesa punktveida novērtētāju salīdzinājums

Lai novērtētu populācijas vidējo vērtību μ, tiek izmantota izlases vidējā vērtība $\bar{y}$.

Klasiskie novērtētāji nezināmiem parametriem tiek novērtēti, ņemot vērā parametru izlases sadalījumus. Parasti lietotais kritērijs ir, ka novērtētājam ir jābūt *nenovirzītam*. Otrais kritērijs ir *minimālās dispersijas nenovirzīts novērtētājs*, kuram no klasiskā viedokļa vispārīgā gadījumā tiek dota priekšroka.

Ja ir dota gadījuma izlase no normālā sadalījuma, tad $\bar{y}$ izlases sadalījums ir $N(\mu, \frac{\sigma^2}{n})$. Izlases vidējā vērtība $\bar{y}$ ir μ *minimālās dispersijas nenovirzīts novērtētājs*.

Pieņemam aposteriorā sadalījuma vidējo vērtību par μ Beijesa novērtētāju:

$$\hat{\mu}_B = E(\mu|y_1, \ldots, y_n) = \frac{1/s^2}{n/\sigma^2 + 1/s^2} \times m + \frac{n/\sigma^2}{n/\sigma^2 + 1/s^2} \times \bar{y}.$$

Aposteriorā vidējā vērtība minimizē aposterioro vidējo kvadrātisko kļūdu. Tas nozīmē, ka $\hat{\mu}_B$ ir optimālais *pēc-datu* novērtētājs.

Salīdzināsim $\hat{\mu}_B$ ar $\hat{\mu}_f = \bar{y}$ pie klasiskā pieņēmuma, ka patiesā vidējā vērtība μ ir fiksēta, bet nezināma konstante. Varbūtības tiks aprēķinātas no $\bar{y}$ izlases sadalījuma. T.i., salīdzinām divus μ *pirms-datu* novērtētājus.

Aposteriorās vidējās vērtības sagaidāmā vērtība ir

$$E(\hat{\mu}_B) = \frac{1/s^2}{n/\sigma^2 + 1/s^2} \times m + \frac{n/\sigma^2}{n/\sigma^2 + 1/s^2} \times \mu.$$

Aposteriorās vidējās vērtības novirze ir vienāda ar tās sagaidāmās vērtības un īstās parametra vērtības starpību, kas vienkāršojas uz

$$\frac{\sigma^2}{ns^2 + \sigma^2}(m - \mu).$$

Aposteriorā vidējā vērtība ir novirzīts μ novērtējums. Novirze ir 0 tikai tad, ja apriorā vidējā vērtība sakrīt ar nezināmo patieso parametra vērtību. Varbūtība, ka tas varētu notikt, ir vienāda ar 0. Aposteriorā vidējās vērtības dispersija ir

$$\left[\frac{n/\sigma^2}{n/\sigma^2 + 1/s^2}\right]^2 \times \frac{\sigma^2}{n} = \left(\frac{ns^2}{ns^2 + \sigma^2}\right)^2 \times \frac{\sigma^2}{n}$$

un tā ir mazāka par $\frac{\sigma^2}{n}$, kas ir klasiskā novērtētāja $\hat{\mu}_f = \bar{y}$ dispersija. Novērtētāja vidējā kvadrātiskā kļūda apvieno novirzi un dispersiju vienā mērā:

$$MS(\hat{\mu}_B) = bias^2 + D(\hat{\mu}).$$

Klasiskais novērtētājs $\hat{\mu}_f = \bar{y}$ ir nenovirzīts μ novērtētājs, tāpēc tā vidējā kvadrātiskā kļūda ir vienāda ar dispersiju:

$$MS(\hat{\mu}_f) = \frac{\sigma^2}{n}.$$

Ja ir pieejama apriorā informācija, tad Beijesa novērtētājam ir mazākā vidējā kvadrātiskā kļūda visā μ iespējamo vērtību apgabalā.

7.2. Vidējās vērtības klasiskā un Beijesa ticamības intervālu salīdzinājums

Klasiskajā statistikā tiek formulēta varbūtība:

$$P\Big(\mu - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} < \bar{y} < \mu + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\Big) = 1 - \alpha,$$

kur $z_{\frac{\alpha}{2}}$ ir vērtība no divpusējā standartnormālā sadalījuma. Pārkārtojam šo varbūtības formulējumu tā, lai μ būtu vidū.

$$P\Big(\bar{y} - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{y} + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\Big) = 1 - \alpha.$$

Intervāla galapunkti ir gadījuma, jo tie ir atkarīgi no $\bar{y}$, kurš šajā interpretācijā ir gadījuma lielums. Tāpēc pareizā ticamības intervāla interpretācija ir, ka $(1-\alpha) \times 100\%$ no intervāliem, kas aprēķināti šādā veidā, saturēs patieso vērtību. Ja ņem gadījuma izlasi un aprēķina $\bar{y}$, tad neatliek nekas, kam būtu gadījuma raksturs un kam varētu piešķirt varbūtību. Intervāls, kuru aprēķinām, vai nu satur, vai nesatur patieso vērtību. Tāpēc sakām, ka mēs esam $(1-\alpha) \times 100\%$ *pārliecināti*, ka izmantojot novēroto $\bar{y}$ vērtību, aprēķinātais intervāls

$$\bar{y} \pm z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \tag{7.1}$$

satur patieso vērtību. Reizēm intervālu pieraksta, kā

$$\Big(\bar{y} - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\Big). \tag{7.2}$$

Tas kontrastē ar Beijesa ticamības intervālu priekš μ, kuru aprēķinājām iepriekšējā nodaļā. Varbūtības formulējums, kuru izveidojam, ir no parametra μ aposteriorā sadalījuma, ja doti izlases dati $y_1, \ldots, y_n$. Tas ir pamatots varbūtības formulējums, jo μ tiek uzskatīts par gadījuma lielumu. Bet tas ir *subjektīvs*, jo tas tiek konstruēts, izmantojot savu *subjektīvo* apriorio sadalījumu. Ja sāktu ar atšķirīgu apriorio sadalījumu, tad iegūtu (nedaudz) atšķirīgu Beijesa ticamības intervālu.

Ja priekš μ tiek ņemts "plakans" apriorais sadalījums, tad aposteriorā vidējā vērtība ir vienāda ar $m' = \bar{y}$ un aposteriorā dispersija ir vienāda ar $(s')^2 = \sigma^2/n$. Tāpēc šajā gadījumā, gan Beijesa ticamības intervāls, gan klasiskais ticamības intervāls būs formā (7.2).

7.3. Vienpusējās hipotēzes pārbaude par normālā sadalījuma vidējo vērtību

Pieņemsim, ka novērojumi ir *i.i.d.* $N(\mu, \sigma^2)$, kur σ^2 ir zināma.

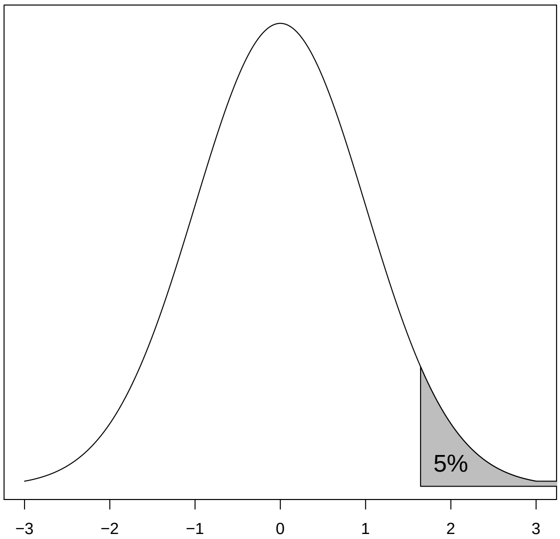
Klasiskā vienpusējās hipotēzes pārbaude par μ

Klasiskā pārbaude balstās uz statistikas izlases sadalījumu.

1. Tiek noformulēta nulles un alternatīvā hipotēze $H_0 : \mu \leq \mu_0$ pret $H_1 : \mu > \mu_0$.
2. $\bar{y}$ nulles sadalījums ir $N(\mu_0, \frac{\sigma^2}{n})$. Tas ir $\bar{y}$ sadalījums, ja nulles hipotēze ir patiesa. Tādējādi standartizētā mainīgā $z = \frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}$ nulles sadalījums būs $N(0, 1)$.
3. Tiek izvēlēts nozīmības līmenis α.
4. Tiek noteikts noraidīšanas apgabals, kuram ir varbūtība α, ja nulles hipotēze ir patiesa. Ja $\alpha = 0.05$, tad noraidīšanas apgabals ir $z > 1.645$. Tas parādīts 5. attēlā.
5. Tiek ņemti izlases dati un aprēķināts $\bar{y}$. Ja $\bar{y}$ vērtība atrodas noraidīšanas apgabalā, tad noraida nulles hipotēzi pie nozīmības līmeņa $\alpha = 0.05$; pretējā gadījumā nevar noraidīt nulles hipotēzi.
6. Cits veids, kā veikt pārbaudi, ir aprēķināt *p-vērtību*:

$$p\text{-}v\bar{e}rt\bar{\imath}ba = P\Big(Z \geq \frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}\Big).$$

Ja *p-vērtība*$\leq \alpha$, tad noraida nulles hipotēzi; pretējā gadījumā, to nevar noraidīt.



5. att.: $z = \frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}$ nulles sadalījums ar noraidīšanas apgabalu vienpusējās hipotēzes klasiskajai pārbaudei 5% nozīmības līmenī.

Beijesa vienpusējās hipotēzes pārbaude par μ

Beijesa statistikā vienpusējās hipotēzes pārbaude tiek veikta, aprēķinot nulles hipotēzes aposterioro varbūtību:

$$P(H_0 : \mu \leq \mu_0 | y_1, \ldots, y_n) = \int_{-\infty}^{\mu_0} g(\mu | y_1, \ldots, y_n)\, d\mu. \tag{7.3}$$

Ja aposteriorais sadalījums $g(\mu|y_1, \ldots, y_n)$ ir $N(m', (s')^2)$, tad to ir vienkārši atrast no standartnormālā sadalījuma tabulām.

$$\begin{aligned} P(H_0 : \mu \leq \mu_0 | y_1, \ldots, y_n) &= P\Big(\frac{\mu - m'}{s'} \leq \frac{\mu_0 - m'}{s'}\Big) \\ &= P\Big(Z \leq \frac{\mu_0 - m'}{s'}\Big), \end{aligned} \tag{7.4}$$

kur Z ir standartnormālais gadījuma lielums. Ja varbūtība ir mazāka par izvēlēto α, tad noraida nulles hipotēzi un var secināt, ka $\mu > \mu_0$.

7.4. Divpusējās hipotēzes pārbaude par normālā sadalījuma vidējo vērtību

Nulles hipotēze *(punktveida hipotēze)* ir patiesa tikai precīzai μ_0 vērtībai. Tas ir tikai viens punkts uz skaitļu taisnes. Pie visām pārējām parametru telpas vērtībām nulles hipotēze ir aplama. Tāpēc drīzāk vajadzētu pārbaudīt, vai nulles hipotēze ir apgabalā, kurš varētu būt patiess, nekā pārbaudīt, vai nulles hipotēze faktiski ir patiesa.

Klasiskā divpusējās hipotēzes pārbaude par μ

1. Tiek formulēta nulles un alternatīvā hipotēze

$$H_0 : \mu = \mu_0 \quad \text{pret} \quad H_1 : \mu \neq \mu_0. \tag{7.5}$$

2. Standartizētā mainīgā $z = \frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}$ nulles sadalījums būs $N(0, 1)$.

3. Tiek izvēlēts nozīmības līmenis α.

4. Tiek noteikts noraidīšanas apgabals, kuram ir varbūtība α, ja nulles hipotēze ir patiesa. Divpusējās hipotēzes pārbaudei ir divpusējs noraidīšanas apgabals.

5. Tiek ņemti izlases dati un aprēķināts $z = \frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}$. Ja vērtība atrodas noraidīšanas apgabalā, tad noraida hipotēzi pie nozīmības līmeņa α; pretējā gadījumā nevar noraidīt nulles hipotēzi.

6. Cits veids, kā veikt pārbaudi, ir aprēķināt *p-vērtību*. *p-vērtība* iekļauj divpusējā apgabala varbūtības:

$$p\text{-}v\bar{e}rt\bar{\imath}ba = P\Big(Z < -\Big|\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}\Big|\Big) + P\Big(Z > \Big|\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}\Big|\Big).$$

   Ja *p-vērtība*$\leq \alpha$, tad var noraidīt nulles hipotēzi; pretējā gadījumā, to nevar noraidīt.

Saistība starp divpusējo hipotēžu pārbaudi un klasisko ticamības intervālu. Noraidīšanas apgabals divpusējai pārbaudei pie nozīmības līmeņa α ir

$$z = \Big|\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}\Big| > z_{\frac{\alpha}{2}}$$

un tas var tikt pārveidots, lai iegūtu, vai nu

$$\mu_0 < \bar{y} - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}, \quad \text{vai} \quad \mu_0 > \bar{y} + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}.$$

Ja pie līmeņa α noraida $H_0 : \mu = \mu_0$, tad μ_0 atrodas ārpus $(1-\alpha)\times 100\%$ klasiskā ticamības intervāla priekš μ. Līdzīgi var parādīt, ka ja pieņem $H_0 : \mu = \mu_0$ pie līmeņa α, tad μ_0 atrodas $(1-\alpha)\times 100\%$ klasiskajā ticamības intervālā priekš μ. Tāpēc klasiskais ticamības intervāls satur visas tās μ_0 vērtības, kuras tiktu akceptētas, ja tās tiktu pārbaudītas.

Beijesa divpusējās hipotēzes pārbaude par μ

Ja divpusējo hipotēzi

$$H_0 : \mu = \mu_0 \quad \text{pret} \quad H_1 : \mu \neq \mu_0$$

pārbauda ar Beijesa metodi un ir nepārtraukts apriorais sadalījums, tad nevar aprēķināt nulles hipotēzes aposterioro varbūtību, kā tika darīts priekš vienpusējās hipotēzes. Ja ir nepārtraukts apriorais sadalījums, tad ir nepārtraukts aposteriorais sadalījums. Nepārtraukta gadījuma lieluma specifiskas vērtības varbūtība vienmēr ir vienāda ar 0. Nulles hipotēzes $H_0 : \mu = \mu_0$ aposteriorā varbūtība būs vienāda ar nulli.

Tā vietā aprēķina $(1 - \alpha) \times 100\%$ Beijesa ticamības intervālu priekš μ, izmantojot aposterioro sadalījumu. Ja μ_0 atrodas Beijesa ticamības intervālā, tad nenoraida nulles hipotēzi $H_0 : \mu = \mu_0$. (Tomēr tai ir nulles varbūtība būt tieši patiesai, ja apskata pietiekami daudz decimālo zīmju.) Ja μ_0 atrodas ārpus Beijesa ticamības intervāla, tad nulles hipotēze tiek noraidīta.

8. Praktiskā daļa

8.1. Beta apriorā sadalījuma izmantošana binomiālās proporcijas novērtēšanā

Pieņemsim, ka trīs studentiem ir dots uzdevums novērtēt pilsētas iedzīvotāju daļu π, kuri atbalsta kazino celtniecību pilsētā.

Anna uzskata, ka apriorā vidējā vērtība varētu būt 0.2 un apriorā standartnovirze 0.08. $Beta(a, b)$ apriorais sadalījums, kurš saskan ar viņas apriorajiem uzskatiem tiek atrasts izmantojot vienādojumu (4.5):

$$\frac{0.2 \times 0.8}{a+b+1} = 0.08^2.$$

Tādējādi viņas ekvivalentais izlases apjoms ir $a + b + 1 = 25$. Izmantojot vienādojumu (4.4) Anna atrod $a = 4.8$ un $b = 19.2$ savam apriorajam sadalījumam.

Atis ir jaunatnācējs pilsētā, tāpēc viņš nav zinošs par lokālo noskaņu "par" vai "pret" piedāvāto kazino. Viņš izlemj lietot vienmērīgo aprioro sadalījumu. Viņam $a = b = 1$. Viņa ekvivalentais izlases apjoms ir $a + b + 1 = 3$.

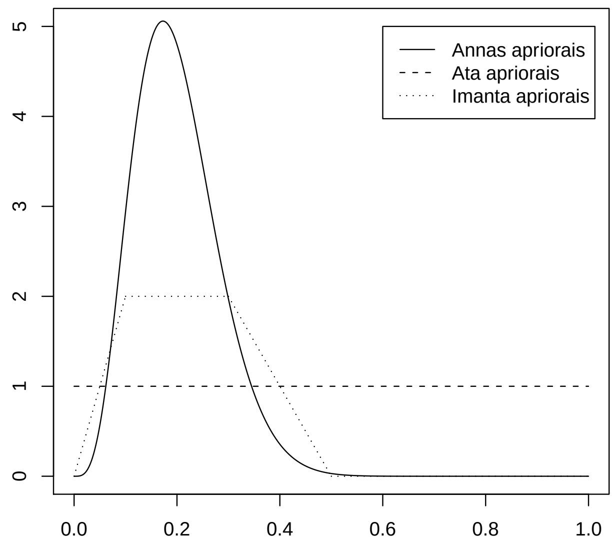
Imants nevar pielāgot $Beta(a, b)$ aprioro sadalījumu saviem uzskatiem. Viņš tic, ka apriorais sadalījums ir ar trapeces formu. Svari priekš viņa apriorā sadalījuma ir doti 21.tabulā un viņš lineāri interpolē starp tiem, lai iegūtu savu nepārtraukto aprioro sadalījumu. Ja interpolē starp šiem punktiem, tad Imanta apriorais sadalījums ir uzdots ar

$$g(\pi) = \begin{cases} 20\pi, & \text{ja} \quad 0 \le \pi \le 0.1 \\ 2, & \text{ja} \quad 0.1 \le \pi \le 0.3 \\ 5 - 10\pi, & \text{ja} \quad 0.3 \le \pi \le 0.5 \end{cases}$$

Šie trīs apriorie sadalījumi ir parādīti 6. attēlā. Imanta apriorais sadalījums patiesībā nav īsts, jo laukums zem tā nav vienāds ar 1. Taču tas nerada problēmas.

21. tabula Imanta apriorie svari

Vērtība	Svars
0	0
0.05	1
0.1	2
0.3	2
0.4	1
0.5	0

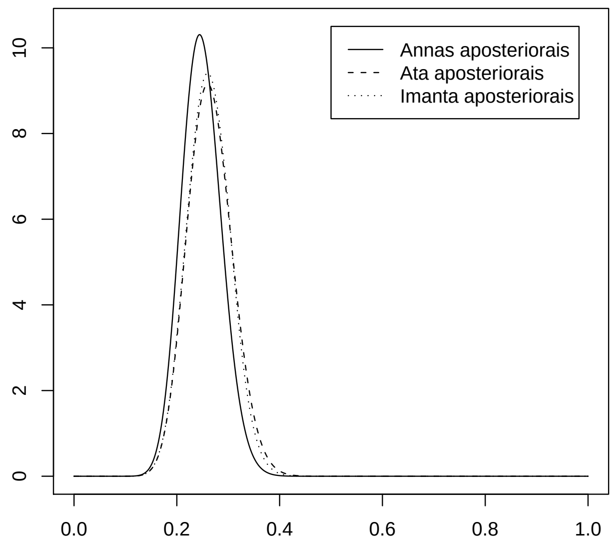


6. att. Annas, Ata un Imanta apriorie sadalījumi.

Tika aptaujāti $n = 100$ pilsētas iedzīvotāji un $y = 26$ no tiem atbalstīja kazino celtniecību pilsētā. Annas un Ata aposteriorie *Beta* sadalījumu parametri tiek noteikti ar vienādojuma (4.3) palīdzību. Annas aposteriorais sadalījums ir $Beta(4.8 + 26, 19.2 + 74)$. Ata aposteriorais sadalījums ir $Beta(1 + 26, 1 + 74)$. Imanta aposteriorais sadalījums tiek atrasts izmantojot vienādojumu (4.2). Lai skaitliski novērtētu Imanta aposterioro sadalījumu, tiek integrēts Imanta *apriorais sadalījums* $\times$ *ticamība*. Trīs aposteriorie sadalījumi ir parādīti 7. attēlā. Visi trīs studenti iegūst ļoti līdzīgus aposterioros sadalījumus, lai gan sāka ar samērā atšķirīgu formu apriorajiem sadalījumiem.

Lai atrastu Annas un Ata aposteriorās vidējās vērtības un dispersijas, ir jāizmanto vienādojumi (4.8) un (4.10). Imanta aposteriorās vidējās vērtības un dispersijas atrašanai jāizmanto vienādojumi (4.7) un (4.9). Aprēķinātās aposteriorās vidējās vērtības, mediānas, standartnovirzes un starpkvartiļu intervāli ir parādīti 22.tabulā. Aposteriorajiem sadalījumiem ir līdzīgas aprakstošās statistikas, neraugoties uz to, ka tika izmantoti dažādi apriorie sadalījumi.

Divpusējos 95% Beijesa ticamības intervālus priekš π, var aprēķināt divos veidos: izmantojot precīzo (Beta) blīvuma funkciju vai izmantojot normālo aproksimāciju. Aprēķinātie intervāli parādīti 23.tabulā. Annai, Atim un Imantam ir nedaudz atšķirīgi Beijesa ticamības intervāli, jo viņi sāka ar atšķirīgiem apriorajiem uzskatiem. Bet datu ietekme bija daudz lielāka nekā viņu aprioro sadalījumu ietekme un viņu aprēķinātie Beijesa tica-



7. att. Annas, Ata un Imanta aposteriorie sadalījumi.

22. tabula Aposterioro sadalījumu izvietojuma un izkliedes mēri

Persona	Aposteriorais sadalījums	Vidējā vērtība	Mediāna	Standart-novirze	Starpkvartiļu intervāls
Anna	$Beta(30.8, 93.2)$	0.248	0.247	0.039	0.053
Atis	$Beta(27, 75)$	0.265	0.263	0.043	0.059
Imants	*skaitlisks*	0.262	0.260	0.042	0.057

mības intervāli ir ļoti līdzīgi. Visos trīs gadījumos 95% Beijesa ticamības intervāls priekš π, kurš aprēķināts izmantojot normālo aproksimāciju ir gandrīz identisks atbilstošajam precīzajam 95% Beijesa ticamības intervālam.

23. tabula Precīzie un aptuvenie 95% Beijesa ticamības intervāli

Persona	Aposteriorais sadalījums	Precīzā ticamības interv.		Norm. aproks. ticamības interv.	
		Apakš. rob.	Augš. rob.	Apakš. rob.	Augš. rob.
Anna	$Beta(30.8, 93.2)$	0.177	0.328	0.172	0.324
Atis	$Beta(27, 75)$	0.184	0.354	0.183	0.355
Imants	*skaitlisks*	0.182	0.345	0.181	0.344

Klasiskais ticamības intervāls priekš π tiek aprēķināts izmantojot formulu (5.2).

$$0.26 \pm 1.96\sqrt{\frac{0.26 \times 0.74}{100}} = (0.174, 0.346).$$

Šo rezultātu var salīdzināt ar 23.tabulā parādītajiem Beijesa ticamības intervāliem.

8.2. Dažādu aprioro sadalījumu izmantošana normālā sadalījuma vidējās vērtības novērtēšanā

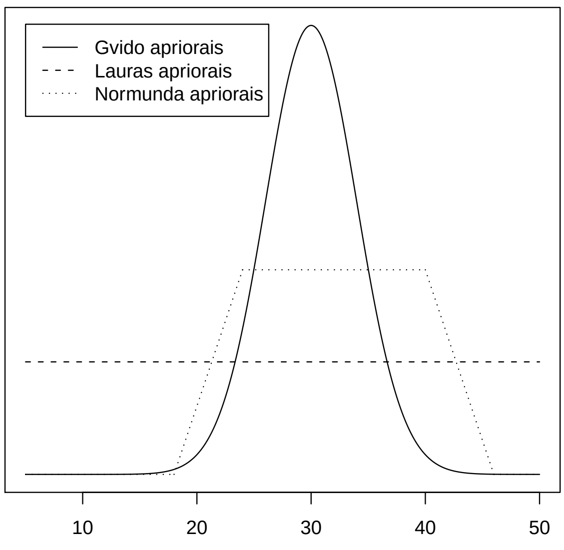
Pieņemsim, ka trīs studentiem ir dots uzdevums novērtēt vienu gadu vecu foreļu vidējo garumu strautā. Iepriekšējie pētījumi citos strautos ir parādījuši, ka gadu vecu foreļu garums ir normāli sadalīts ar zināmu standartnovirzi 2 cm.

Gvido izlemj, ka viņa apriorā vidējā vērtība būs 30 cm. Viņš nolemj neticēt, ka gadu veca varavīksnes forele var būt īsāka par 18 cm vai garāka par 42 cm. Tādēļ viņa apriorā standartnovirze ir 4 cm (kā aprakstīts 6.3. apakšnodaļā). Tāpēc viņš izmantos $N(30, 4^2)$ aprioro sadalījumu.

Laura neko nezina par forelēm, tāpēc viņa izlemj izmantot "plakanu" aprioro sadalījumu.

Normunds izlemj, ka viņa apriorajiem uzskatiem nepiemīt normālais sadalījums. Viņa apriorais sadalījums ir ar trapeces formu. Tas piešķir nulles svaru garumam līdz 18 cm, svaru 1 - garumam no 24 cm līdz 40 cm, un svaru 0 - garumam sākot ar 46 cm. Viņš lineāri interpolē starp šīm vērtībām.

Šo trīs aprioro sadalījumu formas ir parādītas 8. attēlā.



8. att. Gvido, Lauras un Normunda aprioro sadalījumu formas.

Tiek ņemta gadījuma izlase ar 12 vienu gadu vecām forelēm no strauta un atrasta izlases vidējā vērtība $\bar{y} = 32$cm. Gvido un Laura atrod savus aposterioros sadalījumus, izmantojot vienkāršos precizēšanas likumus normālajai saistītajai sadalījumu saimei, kas uzdoti ar vienādojumiem (6.4) un (6.5). Priekš Gvido

$$\frac{1}{(s')^2} = \frac{1}{4^2} + \frac{12}{2^2}.$$

Atrisinot to, tiek iegūta viņa aposteriorā dispersija $(s')^2 = 0.3265$. Viņa aposteriorā standartnovirze ir $s' = 0.5714$. Viņa aposteriorā vidējā vērtība ir

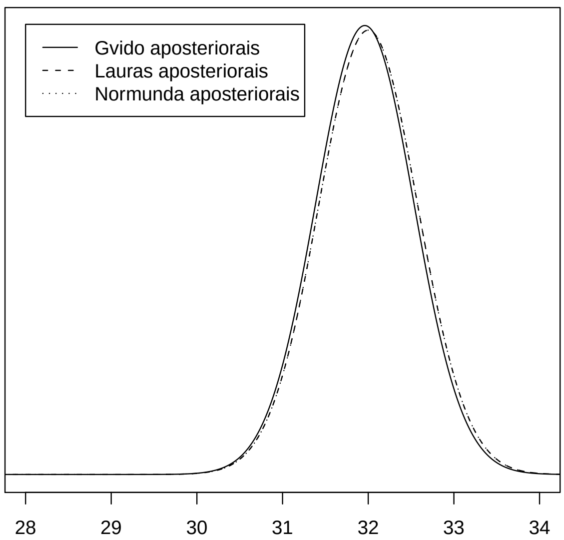
$$m' = \frac{\frac{1}{4^2}}{\frac{1}{0.5714^2}} \times 30 + \frac{\frac{12}{2^2}}{\frac{1}{0.5714^2}} \times 32 = 31.96.$$

Laura izmanto "plakanu" aprioro sadalījumu tāpēc viņas aposteriorā dispersija ir

$$(s')^2 = \frac{2^2}{12} = 0.3333$$

un viņas aposteriorā standartnovirze ir $s' = 0.5774$. Viņas aposteriorā vidējā vērtība $m' = 32$ ir vienāda ar izlases vidējo vērtību. Gan Gvido, gan Laurai ir normālais aposteriorais sadalījums.

Normunds atrod savu aposterioro sadalījumu, izmantojot vienādojumu (6.2), kas prasa skaitlisku integrēšanu. Trīs aposteriorie sadalījumi ir parādīti 9. attēlā. Aposteriorie sadalījumi ir līdzīgi, neraugoties uz to, ka tika sākts ar diezgan atšķirīgiem apriorajiem sadalījumiem.



9. att.: Gvido, Lauras un Normunda aposteriorie sadalījumi. (Lauras un Normunda aposteriorie sadalījumi ir gandrīz identiski.)

No Gvido un Lauras aposteriorajiem sadalījumiem tika aprēķināti 95% Beijesa ticamības intervāli, izmantojot vienādojumu (6.6). Normundam tas bija jāaprēķina skaitliski no sava skaitliskā aposteriorā sadalījuma. Beijesa ticamības intervāli parādīti 24.tabulā. Gvido, Laura un Normunds ieguva nedaudz atšķirīgus Beijesa ticamības intervālus, jo viņi sāka ar atšķirīgiem apriorajiem sadalījumiem. Bet datu ietekme bija daudz lielāka nekā viņu aprioro sadalījumu ietekme un viņu Beijesa ticamības intervāli ir diezgan līdzīgi.

24. tabula 95% Beijesa ticamības intervāli

Persona	Aposteriorais sadalījums	Ticamības intervāls	
		Apakš. rob.	Augš. rob.
Gvido	$N(31.96, 0.3265)$	30.84	33.08
Laura	$N(32, 0.3333)$	30.87	33.13
Normunds	*skaitlisks*	30.86	33.12

95% klasiskais ticamības intervāls priekš μ tiek aprēķināts izmantojot formulu (7.1)

$$\bar{y} \pm z_{0.025} \times \frac{\sigma}{\sqrt{n}} = 32 \pm 1.96 \times \frac{2}{\sqrt{12}} = (30.87, 33.13).$$

Salīdzinot to ar 95% Beijesa ticamības intervāliem 24.tabulā, redzam, ka klasiskais ticamības intervāls ir vienāds ar Lauras atrasto Beijesa ticamības intervālu, jo viņa izmantoja "plakanu" aprioro sadalījumu.

Izlasot publikāciju žurnālā, studenti uzzināja, ka gadu vecas foreles vidējais garums tipiskā izplatības vietā ir 31cm. Katrs no viņiem nolēma noteikt, vai foreļu vidējais garums strautā, kuru viņi pēta, ir lielāks par 31cm, pārbaudot hipotēzi

$$H_0 : \mu \leq 31 \quad \text{pret} \quad H_1 : \mu > 31$$

$\alpha = 5\%$ nozīmības līmenī. Beijesa vienpusējās hipotēzes pārbaudei viņi aprēķina nulles hipotēzes aposterioro varbūtību. Gvido un Laurai ir normālie aposteriorie sadalījumi, tāpēc viņi izmanto vienādojumu (7.4). Normundam nav normālais aposteriorais sadalījums, viņš to aprēķināja skaitliski. Viņš nulles hipotēzes aposterioro varbūtību aprēķina, skaitliski novērtējot vienādojumu (7.3). Beijesa vienpusējās hipotēzes pārbaudes rezultāti parādīti 25.tabulā.

Salīdzināsim to ar klasisko vienpusējās hipotēzes pārbaudi. $z = \frac{\bar{y}-31}{\sigma/\sqrt{n}}$ nulles sadalījums un pareizais noraidīšanas apgabals ir doti 5. attēlā. Šiem datiem $z = \frac{32-31}{2/\sqrt{12}} = 1.732$. Tas atrodas noraidīšanas apgabalā, tāpēc 5% līmenī nulles hipotēze tiek noraidīta. Cits veids, kā var veikt klasisko hipotēžu pārbaudi ir aprēķināt *p-vērtību*.

$$p\text{-}v\bar{e}rt\bar{\imath}ba = P\left(Z > \frac{32-31}{2/\sqrt{12}}\right) = P(Z > 1.732) = 0.0416.$$

25. tabula Beijesa vienpusējās hipotēzes pārbaudes rezultāti

Persona	Aposteriorais sadalījums	$P(\mu \leq 31 \mid y_1, \ldots, y_n)$		
Gvido	$N(31.96, 0.5714^2)$	$P(Z \leq \frac{31-31.96}{0.5714})$	=0.0465	noraida
Laura	$N(32, 0.5774^2)$	$P(Z \leq \frac{31-32}{0.5774})$	=0.0416	noraida
Normunds	*skaitlisks*	$\int_{-\infty}^{31} g(\mu \mid y_1, \ldots, y_n)\, d\mu$	=0.0489	noraida

p-vērtība ir mazāka par nozīmības līmeni α, tāpēc tāpat kā iepriekš nulles hipotēze tiek noraidīta.

8.3. Vienkāršā lineārā regresija

26.tabulā doti dati regresijas analīzei. x ir neatkarīgais mainīgais un y ir atkarīgais mainīgais. Datus aprakstošās statistikas ir: $\bar{x} = 14.3888$, $\bar{y} = 14.2208$, $\overline{x^2} = 207.0703$, $\overline{y^2} = 202.3186$ un $\overline{xy} = 204.6628$.

Regresijas taisnes virziena koeficients tiek aprēķināts izmantojot vienādojumu (1.3):

$$B = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{204.6628 - 14.3888 \times 14.2208}{207.0703 - (14.3888)^2} = \frac{0.04256896}{0.03275456} = 1.29963.$$

Mazāko kvadrātu taisnes vienādojums, kas uzdots ar vienādojumu (1.4) ir

$$y = 14.2208 + 1.29963 \times (x - 14.3888).$$

Datu (x un y) izkliedes diagramma kopā ar mazāko kvadrātu taisni ir parādīta 10. attēlā.

Mazāko kvadrātu novērtētās vērtības $\hat{y}_i = \bar{y} + B(x_i - \bar{x})$, atlikumi un atlikumu kvadrāti parādīti 27.tabulā. Mazāko kvadrātu taisnes novērtētā dispersija (uzdota ar vienādojumu (1.5)) ir

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = \frac{0.801882}{23} = 0.0348644.$$

Novērtētā standartnovirze ir $\hat{\sigma} = \sqrt{0.0348644} = 0.18672$.

Izmantosim $N(1, (0.3)^2)$ apriorо sadalījumu priekš β un $N(15, 1^2)$ aprioro sadalījumu priekš $\alpha_{\bar{x}}$. Tā kā nav zināma patiesā dispersija, tad jāizmanto novērtētā dispersija $\widehat{\sigma^2} = 0.0348644$.

$$SS_x = \sum_{i=1}^{n}(x_i - \bar{x})^2 = n(\overline{x^2} - \bar{x}^2) = 25 \times (207.0703 - 14.3888^2) = 0.81886$$

β aposteriorā precizitāte tiek aprēķināta ar vienādojumu (1.6)

$$\frac{1}{(s'_\beta)^2} = \frac{1}{0.3^2} + \frac{0.81886}{0.0348644} = 34.5981,$$

26. tabula Dati regresijas analīzei

Novērojuma numurs	x	y
1	14.36	13.84
2	14.48	14.41
3	14.53	14.22
4	14.52	14.63
5	14.35	13.95
6	14.31	14.37
7	14.44	14.41
8	14.23	13.99
9	14.32	13.89
10	14.57	14.59
11	14.28	14.32
12	14.36	14.31
13	14.50	14.43
14	14.52	14.44
15	14.28	14.14
16	14.13	13.90
17	14.54	14.37
18	14.60	14.34
19	14.86	14.78
20	14.28	13.76
21	14.09	13.85
22	14.20	13.89
23	14.50	14.22
24	14.02	13.80
25	14.45	14.67
Vidējā vērtība	14.3888	14.2208

tāpēc β aposteriorā standartnovirze ir

$$s'_\beta = 34.5981^{-\frac{1}{2}} = 0.17001.$$

β aposteriorā vidējā vērtība tiek aprēķināta ar vienādojumu (1.7)

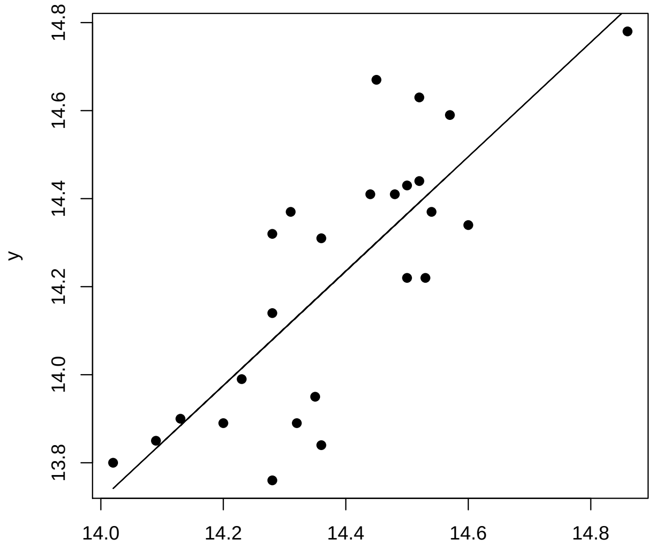
$$m'_\beta = \frac{\frac{1}{0.3^2}}{34.5981} \times 1 + \frac{\frac{0.81886}{0.0348644}}{34.5981} \times 1.29963 = 1.2034.$$

$\alpha_{\bar{x}}$ aposteriorā precizitāte tiek aprēķināta ar vienādojumu (1.8)

$$\frac{1}{(s'_{\alpha_{\bar{x}}})^2} = \frac{1}{1^2} + \frac{25}{0.0348644} = 718.064,$$

tāpēc aposteriorā standartnovirze ir

$$s'_{\alpha_{\bar{x}}} = 718.064^{-\frac{1}{2}} = 0.037318.$$

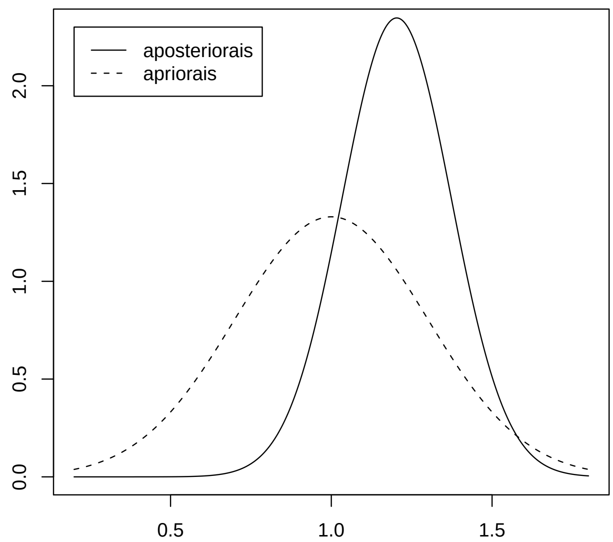


10. att. Datu $\left(x \text{ un } y\right)$ izkliedes diagramma un mazāko kvadrātu taisne

$\alpha_{\bar{x}}$ aposteriorā vidējā vērtība tiek aprēķināta ar vienādojumu (1.9)

$$m'_{\alpha_{\bar{x}}} = \frac{\frac{1}{1^2}}{718.064} \times 15 + \frac{\frac{25}{0.0348644}}{718.064} \times 14.2208 = 14.2219.$$

Virziena koeficienta apriorais un aposteriorais sadalījums ir parādīti 11. attēlā.



11. att. Virziena koeficienta apriorais un aposteriorais sadalījums.

Lai veiktu hipotēžu pārbaudi par to, vai x vispār ir atkarīgs no y, jāpārbauda hipotēze $H_0 : \beta = 0$ pret alternatīvu $H_1 : \beta \neq 0$ pie nozīmības līmeņa $\alpha = 0.05$. Lai pārbaudītu

27. tabula Mazāko kvadrātu novērtētās vērtības, regresijas atlikumi un atlikumu kvadrāti

Novērojuma numurs	x	y	MK novērtējums $\hat{y} = A_0 + Bx$	Atlikums $y - \hat{y}$	Atlikums^2 $(y - \hat{y})^2$
1	14.36	13.84	14.1834	-0.343371	0.117903
2	14.48	14.41	14.3393	0.070673	0.004995
3	14.53	14.22	14.4043	-0.184308	0.033970
4	14.52	14.63	14.3913	0.238688	0.056972
5	14.35	13.95	14.1704	-0.220374	0.048565
6	14.31	14.37	14.1184	0.251611	0.063308
7	14.44	14.41	14.2873	0.122659	0.015045
8	14.23	13.99	14.0144	-0.024418	0.000596
9	14.32	13.89	14.1314	-0.241385	0.058267
10	14.57	14.59	14.4563	0.133706	0.017877
11	14.28	14.32	14.0794	0.240600	0.057888
12	14.36	14.31	14.1834	0.126629	0.016035
13	14.50	14.43	14.3653	0.064681	0.004184
14	14.52	14.44	14.3913	0.048688	0.002371
15	14.28	14.14	14.0794	0.060600	0.003672
16	14.13	13.90	13.8845	0.015545	0.000242
17	14.54	14.37	14.4173	-0.047305	0.002238
18	14.60	14.34	14.4953	-0.155283	0.024113
19	14.86	14.78	14.8332	-0.053188	0.002829
20	14.28	13.76	14.0794	-0.319400	0.102016
21	14.09	13.85	13.8325	0.017531	0.000307
22	14.20	13.89	13.9754	-0.085429	0.007298
23	14.50	14.22	14.3653	-0.145319	0.021118
24	14.02	13.80	13.7415	0.058505	0.003423
25	14.45	14.67	14.3003	0.369662	0.136650
Vidējais	14.3888	14.2208			

hipotēzi ar Beijesa metodi, aprēķinām Beijesa ticamības intervālu. Tā kā nezināmās īstās dispersijas vietā tika izmantota novērtētā dispersija, tad 95% Beijesa ticamības intervāla aprēķināšanai jāizmanto vienādojums (1.10), kur $t_{\frac{\alpha}{2}}$ ir *Stjudenta* sadalījuma kritiskā vērtība ar $n-2=23$ brīvības pakāpēm. Aprēķinātais intervāls ir $(0.852, 1.555)$. Šis intervāls nesatur 0, tāpēc H_0 var noraidīt un secināt, ka y var novērtēt izmantojot x.

Atbilstošais klasiskais ticamības intervāls priekš β jāaprēķina, izmantojot vienādojumu (1.11). Aprēķinātais intervāls ir $(0.873, 1.726)$. Redzam, ka ar Beijesa metodi novērtētais ticamības intervāls ir īsāks nekā klasiskais ticamības intervāls.

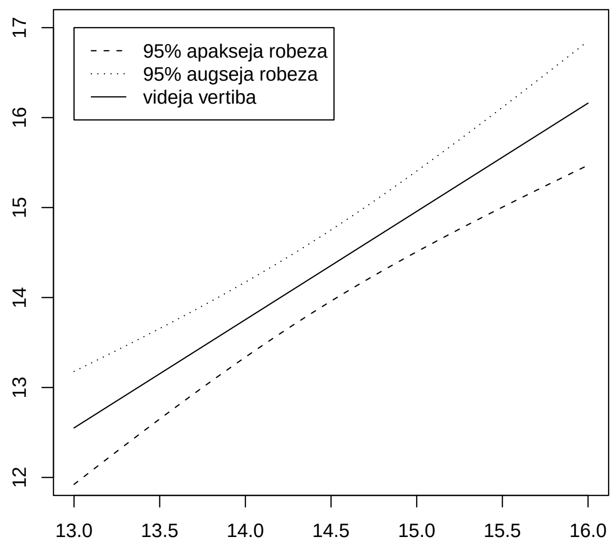
Aprēķināsim (y) prognozējamo sadalījumu kā funkciju no (x). Prognozējamā sadalījuma vidējā vērtība ir

$$m'_y = m'_{\alpha_{\bar{x}}} + m'_\beta \times (x - \bar{x}) = 14.2219 + 1.2034 \times (x - 14.3888)$$

un prognozējamā sadalījuma dispersija ir

$$(s'_y)^2 = \hat{\sigma^2} + s'^2_{\alpha_{\bar{x}}} + s'^2_\beta (x - \bar{x})^2 = 0.0348644 + 0.037318^2 + 0.17001^2 (x - 14.3888)^2.$$

95% prognozes intervāls ir formā $(m'_y - t_{0.025} \times s'_y, m'_y + t_{0.025} \times s'_y)$. Prognozējamās vidējās vērtības grafiks kopā ar 95% prognozējamām robežām ir parādīts attēlā 12.

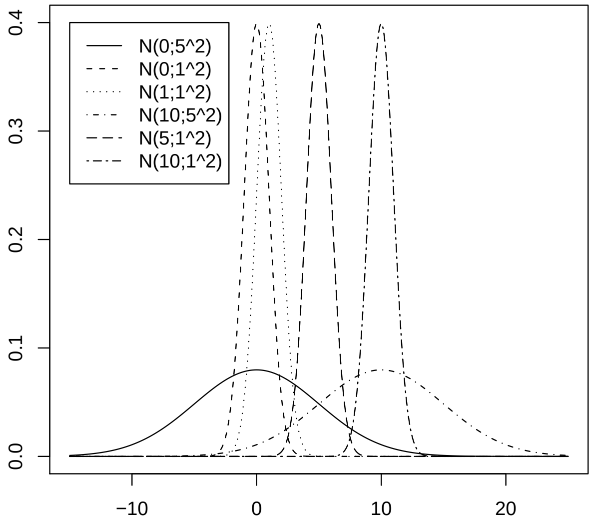


12. att. Prognozējamā vidējā vērtība ar 95% prognozējamām robežām.

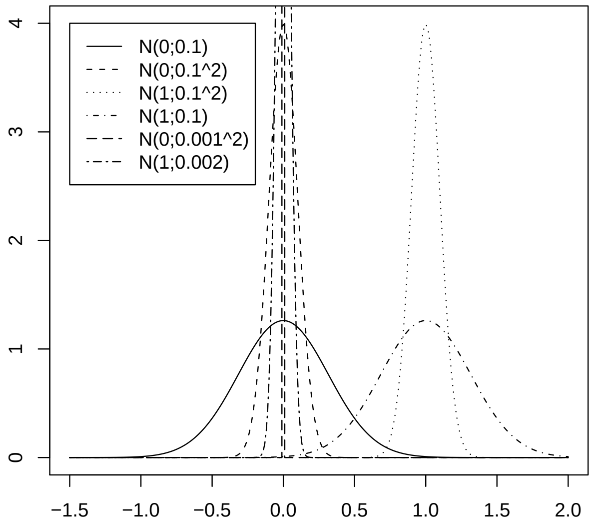
8.4. Normālā sadalījuma vidējās vērtības ticamības intervālu pārklājuma precizitāte

Lai noskaidrotu, cik labi strādā Beijesa metode salīdzinājumā ar klasisko *t-testu* [9], izmantosim ticamības intervālu pārklājuma precizitāti. 10000 reižu tika ģenerēti $N(0, 1)$ sadalījuma dati pie atšķirīgiem izlašu apjomiem, un izmantojot dažādus aprioros sadalījumus parametram μ, ar Beijesa metodi tika konstruēti ticamības intervāli pie nozīmības līmeņa $\alpha = 5\%$ un pārbaudīta to pārklājuma precizitāte, t.i., cik bieži īstā parametra vērtība iekrīt konstruētajā ticamības intervālā.

13. un 14. attēlos parādīti pārklājuma analīzei izmantotie normālie apriorie sadalījumi.



13. att. Daži normālie apriorie sadalījumi.



14. att. Daži normālie apriorie sadalījumi.

28.tabulā redzams, ka t-tests strādā labi pat pie nelieliem izlašu apjomiem.

Beijesa ticamības intervālu pārklājuma precizitāte, protams, ir atkarīga no izvēlētā apriorā sadalījuma. *Jeffreys* apriorais sadalījums priekš normālā sadalījuma vidējās vērtības ir plakanais apriorais sadalījums. Izmantojot šo aprioro sadalījumu konstruētais Beijesa ticamības intervāls sakrīt ar klasisko ticamības intervālu. Un tāpat kā klasisko ticamības intervālu pārklājuma precizitāte, arī Beijesa ticamības intervāliem ar plakano aprioro sadalījumu tā ir ļoti laba gan pie lieliem, gan pie maziem izlašu apjomiem.

28. tabula: Beijesa ticamības intervālu pārklājuma precizitāte $N(0,1)$ sadalījuma vidējai vērtībai ar dažādiem apriorajiem sadalījumiem

n	*t-tests*	$N(0,5^2)$	$N(0,1^2)$	$N\big(0,\big(\frac{1}{\sqrt{10}}\big)^2\big)$	$N\big(0,\big(\frac{1}{10}\big)^2\big)$
10	0.9513	0.9537	0.9559	0.9813	0.9979
20	0.9498	0.9493	0.9541	0.9772	0.9984
50	0.9501	0.9485	0.9526	0.9677	0.9984
100	0.9506	0.9496	0.9517	0.9560	0.9924
200	0.9534	0.9499	0.9521	0.9585	0.9838
500	0.9505	0.9489	0.9555	0.9516	0.9673
1000	0.9476	0.9475	0.9495	0.9485	0.9576
n	*Jeffreys*	$N\big(0,\big(\frac{1}{\sqrt{500}}\big)^2\big)$	$N\big(0,\big(\frac{1}{\sqrt{1000}}\big)^2\big)$	$N(1,1^2)$	$N(10,5^2)$
10	0.9494	0.9998	0.9999	0.9542	0.9490
20	0.9501	1.0000	1.0000	0.9470	0.9500
50	0.9506	1.0000	1.0000	0.9544	0.9554
100	0.9495	1.0000	1.0000	0.9496	0.9482
200	0.9494	0.9994	1.0000	0.9489	0.9515
500	0.9492	0.9945	0.9994	0.9513	0.9490
1000	0.9501	0.9841	0.9940	0.9504	0.9520
n	$N(5,1^2)$	$N\big(1,\big(\frac{1}{\sqrt{10}}\big)^2\big)$	$N(10,1^2)$	$N\big(1,\big(\frac{1}{10}\big)^2\big)$	
10	0.7626	0.5109	0.2702	0.0000	
20	0.8431	0.6190	0.4569	0.0000	
50	0.8981	0.7875	0.7183	0.0000	
100	0.9209	0.8562	0.8366	0.0000	
200	0.9371	0.9054	0.8908	0.0000	
500	0.9418	0.9272	0.9266	0.0104	
1000	0.9441	0.9365	0.9361	0.1359	

Arī, ja izvēlētā apriorā sadalījuma vidējā vērtība ir tuva patiesajai vidējai vērtībai, vai arī apriorajam sadalījumam ir pietiekami liela dispersija, tad pārklājuma precizitāte ir ļoti laba.

Taču, ja izvēlētā apriorā sadalījuma vidējā vērtība ir tālu no patiesās (un dispersija nav pārāk liela), tad pārklājuma precizitāte ir laba tikai pie lieliem izlašu apjomiem.

Secinājumi

Darbā tika apskatītas Beijesa metodes dažādu matemātiskās statistikas problemātiku risināšanai. Pielietojot Beijesa metodes, ļoti svarīga ir apriorā sadalījuma izvēle. Ir iespējams izvēlēties ļoti dažādus aprioros sadalījumus novērtējamajiem parametriem. Atkarībā no tā, cik daudz apriorās informācijas ir pieejams un cik daudz to vēlamies iekļaut apriorajā sadalījumā, mēs varam izvēlēties vai nu informatīvu vai neinformatīvu apriorо sadalījumu. Neinformatīvie apriorie sadalījumi var būt pat tādi sadalījumi, kas nav īsti varbūtību sadalījumi, t.i., integrālis no to blīvuma funkcijām nav vienāds ar 1. Taču tas nesagādā problēmas, ja pareizinot aposterioro sadalījumu ar piemērotu konstanti, integrālis no aposteriorā blīvuma tomēr ir 1. Viens no visbiežāk izmantotajiem neinformatīvajiem apriorajiem sadalījumiem ir *Jeffreys* apriorais sadalījums, kurš tiek iegūts no Fišera informācijas matricas.

Praksē, ja vien ir iespējams, tad visērtāk ir izmantot saistītos aprioros sadalījumus. Iegūtie aposteriorie sadalījumi tādā gadījumā būs tās pašas sadalījumu saimes pārstāvji, no kuras ir ņemts apriorais sadalījums, un aposteriorā sadalījuma parametri būs samērā vienkārši aprēķināmi.

Ar Beijesa metodi iegūtie punktveida novērtējumi bieži vien ir labāki par klasiskajiem novērtējumiem, ja tie tiek salīdzināti izmantojot klasiskos kritērijus, kā piemēram, vidējo kvadrātisko kļūdu. Tas nozīmē, ka pat tad, ja nav nekādas apriorās informācijas, tomēr labāk izmantot Beijesa metodes nekā atgriezties pie klasiskajām metodēm.

Beijesa ticamības intervāli parasti šaurāki par atbilstošajiem klasiskajiem ticamības intervāliem, jo Beijesa metode izmanto gan aprioro informāciju, gan no datiem iegūto informāciju, bet novērtējot ticamības intervālus ar klasisko metodi, apriorā informācija netiek ņemta vērā.

Izmantotā literatūra un avoti

[1] William M. Bolstad. *Introduction to Bayesian Statistics.* John Wiley and Sons, New Jersey, 2007.

[2] Jim Albert. *Bayesian Computation with R.* Springer Science + Business Media, LLC, New York, 2007.

[3] Jean-Michel Marin and Christian P.Robert. *Bayesian Core: A Practical Approach to Computational Bayesian Statistics.* Springer Science + Business Media, LLC, New York, 2007.

[4] Christian P.Robert and Jean-Michel Marin. *Bayesian Core: The Complete Solution Manual.* Springer-Verlag, Berlin, 2009.

[5] Karl-Rudolf Koch. *Introduction to Bayesian Analysis.* Springer-Verlag, Berlin, 2007.

[6] Christian P.Robert. *The Bayesian Choice.* Springer Science + Business Media, LLC, New York, 2007.

[7] Scott M. Lynch. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* Springer Science + Business Media, LLC, New York, 2007.

[8] Jayanta K.Ghosh, Mohan Delampady, and Tapas Samanta. *An Introduction to Bayesian Analysis. Theory and Methods.* Springer Science + Business Media, LLC, New York, 2006.

[9] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaa, and L.E. Meester. *A Modern Introduction to Probability and Statistics.* Springer Science + Business Media, London, 2005.

1. Beijesa problemātika vienkāršai lineārai regresijai

Dati sastāv no n sakārtotiem punktu pāriem (x_i, y_i) $i = 1, \ldots, n$. x ir prediktors (neatkarīgais mainīgais) bez kļūdas. y ir atbildes mainīgais, kurš kaut kādā nezināmā veidā ir atkarīgs no x. Katrs novērotais y ir ar kļūdu.

1.1. Mazāko kvadrātu regresija

Regresijas taisnes vienādojums tiek noteikts ar tās virziena koeficientu β un brīvo locekli α_0.

Normālie vienādojumi un mazāko kvadrātu taisne

Taisnes $y = \alpha_0 + \beta x$ atlikumu kvadrātu summa ir

$$SS_{res} = \sum_{i=1}^{n} [y_i - (\alpha_0 + \beta x_i)]^2.$$

Lai izmantojot aprēķinus atrastu α_0 un β vērtības, kuras minimizē SS_{res}, parciāli atvasina SS_{res} pēc α_0 un pēc β. Atvasinājumus pielīdzina nullei un atrisina šos vienādojumus. Vispirms ņem atvasinājumu pēc α_0. Tas dod vienādojumu

$$\frac{\partial SS}{\partial \alpha_0} = \sum_{i=1}^{n} 2 \times [y_i - (\alpha_0 + \beta x_i)]^1 \times (-1) = 0,$$

kas vienkāršojas uz

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \alpha_0 - \sum_{i=1}^{n} \beta x_i = 0$$

un tālāk uz

$$\bar{y} - \alpha_0 - \beta \bar{x} = 0. \tag{1.1}$$

Ņemot parciālo atvasinājumu pēc β, iegūst

$$\frac{\partial SS}{\partial \beta} = \sum_{i=1}^{n} 2 \times [y_i - (\alpha_0 + \beta x_i)]^1 \times (-x_i) = 0,$$

kas vienkāršojas uz

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \alpha_0 x_i - \sum_{i=1}^{n} \beta x_i^2 = 0$$

un tālāk uz

$$\overline{xy} - \alpha_0 \bar{x} - \beta \overline{x^2} = 0. \tag{1.2}$$

Vienādojumi (1.1) un (1.2) ir pazīstami kā *normālie* vienādojumi. Šeit *normalitāte* attiecas uz taisnajiem projekciju leņķiem[1] un tai nav nekāda sakara ar normālo sadalījumu. No vienādojuma (1.1) izsaka α_0, ievieto to vienādojumā (1.2) un izsaka β

$$\overline{xy} - (\bar{y} - \beta\bar{x})\bar{x} - \beta\overline{x^2} = 0.$$

Atrisinājums ir mazāko kvadrātu virziena koeficients[2]

$$B = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}. \tag{1.3}$$

Ļoti svarīgi neveikt noapaļošanu, lai neiegūtu būtiskas kļūdas slīpuma novērtējumā. B ievieto atpakaļ vienādojumā (1.1) un izsaka brīvo locekli

$$A_0 = \bar{y} - B\bar{x}.$$

Mazāko kvadrātu taisnes vienādojums ir

$$y = A_0 + Bx.$$

Mazāko kvadrātu taisnes alternatīvā forma. Taisni nosaka visi tie punkti

$$A_{\bar{x}} = A_0 + B\bar{x} = \bar{y},$$

kuru abscisas ir $\bar{x}$. Tāpēc mazāko kvadrātu taisne iet caur punktiem $(\bar{x}, \bar{y})$. Alternatīvs vienādojums mazāko kvadrātu taisnei ir

$$y = A_{\bar{x}} + B(x - \bar{x}) = \bar{y} + B(x - \bar{x}), \tag{1.4}$$

kurš ir īpaši noderīgs.

Dispersijas novērtēšana ap mazāko kvadrātu taisni

Dispersijas novērtējums ap mazāko kvadrātu taisni ir

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}[y_i - (A_{\bar{x}} + B(x_i - \bar{x}))]^2}{n-2}, \tag{1.5}$$

[1]Mazākie kvadrāti atrod (n-dimensionāla) novērojumu vektora projekciju plaknē, kura satur visas iespējamās (α_0, β) vērtības.

[2]Mazāko kvadrātu virziena koeficientam ir vairākas atšķirīgas formulas. Taču var parādīt, ka tās visas ir ekvivalentas. Konkrēti šī tiek izmantota, jo to ir viegli atcerēties.

kas ir atlikumu kvadrātu summas dalījums ar $n-2$. Atlikumu kvadrātu summa tiek dalīta ar $n-2$, jo atlikumu kvadrātu summas[3] aprēķināšanai tika izmantoti divi novērtējumi $A_{\bar{x}}$ un B.

1.2. Vienkāršās lineārās regresijas pieņēmumi

1. *Pieņēmums par vidējo vērtību.* y nosacītā vidējā vērtība, jo dots x, ir nezināma lineāra funkcija no x.

$$\mu_{y|x} = \alpha_0 + \beta x.$$

   Alternatīvā parametrizācijā

$$\mu_{y|x} = \alpha_{\bar{x}} + \beta(x - \bar{x}).$$

2. *Pieņēmums par kļūdu.* Novērojums ir vienāds ar vidējās vērtības un kļūdas summu, kur kļūdai ir $normal(0, \sigma^2)$ sadalījums ar *zināmu* σ^2.

3. *Pieņēmums par neatkarību.* Kļūdas visiem novērojumiem ir neatkarīgas viena no otras.

Izmantojot alternatīvo parametrizāciju, iegūstam

$$y_i = \alpha_{\bar{x}} + \beta \times (x_i - \bar{x}) + e_i,$$

kur $\alpha_{\bar{x}}$ ir y vidējā vērtība, ja dots $x = \bar{x}$, un β ir virziena koeficients. Katrs e_i ir i.i.d. $N(0, \sigma^2)$, kur σ^2 ir zināma. Tādējādi $y_i|x_i$ ir $N(\alpha_{\bar{x}} + \beta(x_i - \bar{x}), \sigma^2)$ sadalīti un neatkarīgi viens no otra.

1.3. Beijesa teorēma regresijas modelim

β un $\alpha_{\bar{x}}$ kopējā ticamība

i-tā novērojuma kopējā ticamība ir varbūtību blīvuma funkcija no diviem parametriem $\alpha_{\bar{x}}$ un β, kur (x_i, y_i) ir fiksētas novērojuma vērtības. Šī funkcija piešķir relatīvos svarus

[3] Dispersijas nenovirzīta novērtējuma iegūšanas vispārīgais likums ir, ka atlikumu kvadrātu summa jādala ar brīvības pakāpju skaitu. Pie katra novērtētā parametra atlikumu kvadrātu summas formulā tiek zaudēta viena brīvības pakāpe.

visām iespējamajām abu parametru $\alpha_{\bar{x}}$ un β vērtībām no novērojuma. Ja ignorējam to daļu, kura nesatur parametrus, tad novērojuma i ticamība ir

$$ticamība_i(\alpha_{\bar{x}}, \beta) \propto e^{-\frac{1}{2\sigma^2}[y_i - (\alpha_{\bar{x}} + \beta(x_i - \bar{x}))]^2}.$$

Tā kā visi novērojumi ir neatkarīgi, tad visas izlases novērojumu ticamība ir atsevišķo ticamību reizinājums:

$$ticamība_{izlasei}(\alpha_{\bar{x}}, \beta) \propto \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}[y_i - (\alpha_{\bar{x}} + \beta(x_i - \bar{x}))]^2}.$$

Eksponenšu reizinājums tiek atrasts, summējot kāpinātājus

$$ticamība_{izlasei}(\alpha_{\bar{x}}, \beta) \propto e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}[y_i - (\alpha_{\bar{x}} + \beta(x_i - \bar{x}))]^2\right]}.$$

Kāpinātāja daļu

$$\left[\sum_{i=1}^{n}[y_i - \bar{y} + \bar{y} - (\alpha_{\bar{x}} + \beta(x_i - \bar{x}))]^2\right]$$

sadalot trīs summās, iegūstam

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \bar{y})(\bar{y} - (\alpha_{\bar{x}} + \beta(x_i - \bar{x})))$$

$$+\sum_{i=1}^{n}(\bar{y} - (\alpha_{\bar{x}} + \beta(x_i - \bar{x})))^2.$$

Tas vienkāršojas uz

$$SS_y - 2\beta SS_{xy} + \beta^2 SS_x + n(\alpha_{\bar{x}} - \bar{y})^2,$$

kur $SS_y = \sum_{i=1}^{n}(y_i - \bar{y})^2$, $SS_{xy} = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})$ un $SS_x = \sum_{i=1}^{n}(x_i - \bar{x})^2$. Tāpēc kopējo ticamību var pierakstīt kā

$$ticamība_{izlasei}(\alpha_{\bar{x}}, \beta) \propto e^{-\frac{1}{2\sigma^2}[SS_y - 2\beta SS_{xy} + \beta^2 SS_x + n(\alpha_{\bar{x}} - \bar{y})^2]}.$$

Pārrakstot to kā divu eksponenšu reizinājumu,

$$\propto e^{-\frac{1}{2\sigma^2}[SS_y - 2\beta SS_{xy} + \beta^2 SS_x]} \times e^{-\frac{1}{2\sigma^2}[n(\alpha_{\bar{x}} - \bar{y})^2]}.$$

Pirmās eksponentes kāpinātājā iznesam SS_x pirms iekavām, papildinām līdz pilnam kvadrātam un proporcionalitātes konstantē iekļaujam daļu, kura nav atkarīga ne no viena parametra. Tas dod

$$ticamība_{izlasei}(\alpha_{\bar{x}}, \beta) \propto e^{-\frac{1}{2\sigma^2/SS_x}[\beta - \frac{SS_{xy}}{SS_x}]} \times e^{-\frac{1}{2\sigma^2/n}[(\alpha_{\bar{x}} - \bar{y})^2]}.$$

Ievērojam, ka $\frac{SS_{xy}}{SS_x} = B$ ir mazāko kvadrātu virziena koeficients un $\bar{y} = A_{\bar{x}}$ ir vertikālās taisnes $x = \bar{x}$ brīvā locekļa mazāko kvadrātu novērtējums. Kopējā ticamība ir izteikta kā divu atsevišķu ticamību reizinājums

$$ticam\bar{\imath}ba_{izlasei}(\alpha_{\bar{x}}, \beta) \propto ticam\bar{\imath}ba_{izlasei}(\alpha_{\bar{x}}) \times ticam\bar{\imath}ba_{izlasei}(\beta),$$

kur

$$ticam\bar{\imath}ba_{izlasei}(\beta) \propto e^{-\frac{1}{2\sigma^2/SS_x}(\beta - B)^2}$$

un

$$ticam\bar{\imath}ba_{izlasei}(\alpha_{\bar{x}}) \propto e^{-\frac{1}{2\sigma^2/n}(\alpha_{\bar{x}} - A_{\bar{x}})^2}.$$

Tā kā kopējā ticamība ir vienāda ar atsevišķu ticamību reizinājumu, tad atsevišķās ticamības ir neatkarīgas. Virziena koeficienta β ticamībai ir $N(B, \frac{\sigma^2}{SS_x})$ sadalījuma forma un $\alpha_{\bar{x}}$ ticamībai ir $N(A_{\bar{x}}, \frac{\sigma^2}{n})$ sadalījuma forma.

β un $\alpha_{\bar{x}}$ kopējais apriorais sadalījums

Katram parametram izmantosim neatkarīgu apriorо sadalījumu. Divu parametru kopējais apriorais sadalījums ir atsevišķo apriorо sadalījumu reizinājums:

$$g(\alpha_{\bar{x}}, \beta) \propto g(\alpha_{\bar{x}}) \times g(\beta).$$

Var izmantot vai nu *normālos* vai *plakanos* apriorоs sadalījumus.

Normālo apriorо sadalījumu izvēle priekš β un $\alpha_{\bar{x}}$. Cita šīs parametrizācijas izmantošanas priekšrocība ir tāda, ka par $\alpha_{\bar{x}}$ (taisnes $x = \bar{x}$ brīvo locekli) ir vairāk intuitīvo zināšanu nekā par α_0 (par y ass brīvo locekli). Jāizlemj, kāda varētu būt $a_{\bar{x}}$ apriorā vidējā vērtība $m_{a_{\bar{x}}}$. Lai iegūtu $a_{\bar{x}}$ apriorо standartnovirzi $s_{a_{\bar{x}}}$, starpību starp iespējamo y vērtību augšējo un apakšējo robežu dala ar 6.

Parasti mūs vairāk interesē virziena koeficients β. Reizēm vēlamies noteikt, vai tas varētu būt 0. Tādēļ mēs varam izvēlēties $m_\beta = 0$ kā β apriorо vidējo vērtību. Pēc tam ir jādomā par x pieauguma par 1 vienību ietekmi uz y augšējo un apakšējo robežu. Starpību dala ar 6, lai iegūtu s_β, kas ir β apriorā standartnovirze. Citos gadījumos, mums ir apriorie uzskati par virziena koeficientu no iepriekšējiem datiem. Tad jāizmanto $N(m_\beta, (s_\beta)^2)$, kurš atbilst apriorajiem uzskatiem.

β un $\alpha_{\bar{x}}$ kopējais aposteriorais sadalījums

Kopējais aposteriorais sadalījums ir

$$g(\alpha_{\bar{x}}, \beta | dati) \propto g(\alpha_{\bar{x}}, \beta) \times ticam\bar{\imath}ba_{izlasei}(\alpha_{\bar{x}}, \beta),$$

kur *dati* ir sakārtotu pāru $(x_1, y_1), \ldots, (x_n, y_n)$ kopa. Gan kopējais apriorais sadalījums, gan kopējā ticamība ir divu daļu reizinājums, kur viena daļa ir atkarīga no $\alpha_{\bar{x}}$, bet otra daļa ir atkarīga no β. Kopējā apriorā sadalījuma un kopējās ticamības pārkārtošana dod kopējo aposterioro sadalījumu, kurš ir izteikts kā marginālo aposterioro sadalījumu reizinājums

$$g(\alpha_{\bar{x}}, \beta) \propto g(\alpha_{\bar{x}} | dati) \times g(\beta | dati).$$

Tā kā kopējais aposteriorais sadalījums ir marginālo aposterioro sadalījumu reizinājums, tad tie ir neatkarīgi. Katru no šiem marginālajiem aposteriorajiem sadalījumiem var atrast, izmantojot vienkāršos precizēšanas likumus normālajam sadalījumam, kuri ir spēkā *normālajiem* un *plakanajiem* apriorajiem sadalījumiem. Piemēram, ja izmantojam $N(m_\beta, s_\beta^2)$ apriororo sadalījumu priekš β, tad iegūstam $N(m'_\beta, (s'_\beta)^2)$ aposterioro sadalījumu, kur

$$\frac{1}{(s'_\beta)^2} = \frac{1}{s_\beta^2} + \frac{SS_x}{\sigma^2} \tag{1.6}$$

un

$$m'_\beta = \frac{\frac{1}{s_\beta^2}}{\frac{1}{(s'_\beta)^2}} \times m_\beta + \frac{\frac{SS_x}{\sigma^2}}{\frac{1}{(s'_\beta)^2}} \times B. \tag{1.7}$$

Līdzīgi, ja izmantojam $N(m_{\alpha_{\bar{x}}}, s_{\alpha_{\bar{x}}}^2)$ apriororo sadalījumu priekš $\alpha_{\bar{x}}$, tad iegūstam $N(m'_{\alpha_{\bar{x}}}, (s'_{\alpha_{\bar{x}}})^2)$ aposterioro, kur

$$\frac{1}{(s'_{\alpha_{\bar{x}}})^2} = \frac{1}{s_{\alpha_{\bar{x}}}^2} + \frac{n}{\sigma^2} \tag{1.8}$$

un

$$m'_{\alpha_{\bar{x}}} = \frac{\frac{1}{s_{\alpha_{\bar{x}}}^2}}{\frac{1}{(s'_{\alpha_{\bar{x}}})^2}} \times m_{\alpha_{\bar{x}}} + \frac{\frac{n}{\sigma^2}}{\frac{1}{(s'_{\alpha_{\bar{x}}})^2}} \times A_{\bar{x}}. \tag{1.9}$$

Beijesa ticamības intervāls virziena koeficientam

$(1-\alpha) \times 100\%$ Beijesa ticamības intervāls virziena koeficientam β ir

$$m'_\beta \pm z_{\frac{\alpha}{2}} \times \sqrt{(s'_\beta)^2}.$$

Bet σ^2 vairumā gadījumu nav zināma. Saprātīgāka pieeja tādā gadījumā ir izmantot novērtējumu, kas aprēķināts no atlikumiem

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - (A_{\bar{x}} + B(x_i - \bar{x})))^2}{n-2}.$$

Nezināmās σ^2 dēļ radušās papildus nenoteiktības dēļ ir jāpaplašina ticamības intervāls. To dara standartnormālās kritiskās vērtības vietā izmantojot *Stjudenta t* kritisko vērtību ar $n-2$ brīvības pakāpēm. Tad Beijesa ticamības intervāls būs

$$m'_\beta \pm t_{\frac{\alpha}{2}} \times \sqrt{(s'_\beta)^2}. \tag{1.10}$$

Klasiskais ticamības intervāls virziena koeficientam

Ja dispersija σ^2 nav zināma, tad $(1-\alpha) \times 100\%$ klasiskais ticamības intervāls virziena koeficientam β ir

$$B \pm t_{\frac{\alpha}{2}} \times \frac{\hat{\sigma}}{\sqrt{SS_x}}, \tag{1.11}$$

kur $\hat{\sigma}^2$ ir dispersijas novērtējums, kas aprēķināts no mazāko kvadrātu taisnes atlikumiem. Klasiskajam ticamības intervālam ir tāda pati forma kā Beijesa ticamības intervālam, ja izmanto *plakanos* aprioros sadalījumus priekš β un $\alpha_{\bar{x}}$. Bet interpretācija, protams, ir atšķirīga. Pie klasiskajiem pieņēmumiem mēs esam $(1-\alpha) \times 100\%$ pārliecināti, ka intervāls satur īsto, nezināmo parametra vērtību.

Vienpusējās hipotēzes pārbaude par virziena koeficientu

Bieži mēs vēlamies noteikt, vai y pieauguma apjoms, kas saistīts ar x pieaugumu par 1 vienību, ir lielāks par kādu vērtību β_0. To var izdarīt pārbaudot

$$H_0 : \beta \leq \beta_0 \quad \text{pret} \quad H_1 : \beta > \beta_0$$

ar Beijesa metodi pie nozīmības līmeņa α. Tāpēc aprēķinām nulles hipotēzes aposterioro varbūtību. Tā ir

$$\begin{aligned} P(\beta \leq \beta_0 | dati) &= \int_{-\infty}^{\beta_0} g(\beta | dati) d\beta \\ &= P\left(Z \leq \frac{\beta_0 - m'_\beta}{s'_\beta}\right). \end{aligned}$$

Ja šī varbūtība ir mazāka par α, tad mēs noraidām H_0 un secinām, ka virziena koeficients β ir lielāks par β_0. (Ja tika izmantots dispersijas novērtējums, tad standartnormālā Z vietā jāizmanto *Stjudenta* t ar $n-2$ brīvības pakāpēm.)

Divpusējās hipotēzes pārbaude par virziena koeficientu

Ja $\beta = 0$, tad y vidējā vērtība vispār nav atkarīga no x. Pārbaudīsim $H_0 : \beta = 0$ pret $H_1 : \beta \neq 0$ pie nozīmības līmeņa α ar Beijesa metodi pirms regresijas modeļa izmantošanas prognozēšanai. Ja 0 atradīsies ārpus Beijesa ticamības intervāla, tad H_0 jānoraida. Pretējā gadījumā nulles hipotēzi nevar noraidīt un regresijas modeli nevar izmantot prognozēšanai.

1.4. Nākamā novērojuma prognozējamais sadalījums

Labākā y_{n+1} prognoze, ja dots x_{n+1}, būs

$$\bar{y}_{n+1} = \hat{\alpha}_{\bar{x}} + \hat{\beta} \times (x_{n+1} - \bar{x}),$$

kur $\hat{\beta}$ ir virziena koeficienta novērtējums un $\hat{\alpha}_{\bar{x}}$ ir taisnes $x = \bar{x}$ brīvā locekļa novērtējums.

Cik laba ir šī prognoze? Šeit nenoteiktībai ir divi cēloņi. Pirmkārt, prognozē tiek izmantotas novērtētās parametru vērtības nevis īstās, nezināmās parametru vērtības. Otrkārt, jaunais novērojums y_{n+1} satur pats savu novērojuma kļūdu e_{n+1}, kura būs neatkarīga no visām iepriekšējām novērojumu kļūdām. Nākamā novērojuma y_{n+1} *prognozējamais sadalījums*, ja dota vērtība x_{n+1} un *dati*, ietver abus nenoteiktības avotus. Tas tiek apzīmēts ar $f(y_{n+1}|x_{n+1}, dati)$ un tiek atrasts ar Beijesa teorēmas palīdzību.

Prognozējamā sadalījuma atrašana

Prognozējamais sadalījums tiek atrasts, integrējot nākamā novērojuma un parametru kopējo nosacīto aposterioro sadalījumu, ja doti *dati* (iepriekšējie novērojumi $(x_1, y_1), \ldots, (x_n, y_n)$ no modeļa) un nākamā x_{n+1} vērtība

$$f(y_{n+1}|x_{n+1}, dati) = \iint f(y_{n+1}, \alpha_{\bar{x}}, \beta | x_{n+1}, dati) d\alpha_{\bar{x}} d\beta.$$

Kopējā aposteriorā sadalījuma integrēšana pēc traucējošajiem parametriem tiek saukta par *marginalizāciju*. Tā ir viena no Beijesa statistikas priekšrocībām, jo šī metode vienmēr strādā. Aprēķinot prognozējamo sadalījumu, visi parametri tiek uzskatīti par traucējošiem.

Tāpēc, pirmkārt, ir ir jānosaka nākamā novērojuma un parametru kopējais aposteriorais sadalījums, ja dota x_{n+1} vērtība un *dati*:

$$\begin{aligned} f(y_{n+1}, \alpha_{\bar{x}}, \beta | x_{n+1}, dati) =& f(y_{n+1} | \alpha_{\bar{x}}, \beta, x_{n+1}, dati) \\ & \times g(\alpha_{\bar{x}}, \beta | x_{n+1}, dati). \end{aligned}$$

Nākamais novērojums y_{n+1}, ja doti parametri $\alpha_{\bar{x}}$ un β un ir zināma nākamā $x_{n+1} vrtba$, ir tikai vēl viens gadījuma novērojums no regresijas modeļa. Ja doti parametri $\alpha_{\bar{x}}$ un β, tad visi novērojumi ir neatkarīgi viens no otra. Tas nozīmē, ka, ja ir doti parametri, tad jaunais novērojums y_{n+1} nav atkarīgs no *datiem* (no iepriekšējiem regresijas novērojumiem). Tā kā aposteriorais sadalījums priekš $\alpha_{\bar{x}}, \beta$ tika aprēķināts tikai no *datiem* un nav atkarīgs no x_{n+1}, tad

$$f(y_{n+1}, \alpha_{\bar{x}}, \beta | x_{n+1}, dati) = f(y_{n+1} | \alpha_{\bar{x}}, \beta, x_{n+1}) \times g(\alpha_{\bar{x}}, \beta | dati).$$

$y_{n+1} | \alpha_{\bar{x}}, \beta, x_{n+1}$ ir sadalīts normāli ar vidējo vērtību, kura uzdota ar lineāru funkciju no parametriem $\mu_{n+1} = \alpha_{\bar{x}} + \beta(x_{n+1} - \bar{x})$ un zināmu dispersiju σ^2.

Parametru aposteriorie sadalījumi, kurus atradām iepriekš, izmantojot precizēšanas likumus, ir neatkarīgi $N(m'_{\alpha_{\bar{x}}}, (s'_{\alpha_{\bar{x}}})^2)$ un $N(m'_{\beta}, (s'_{\beta})^2)$. Tā kā nākamais novērojums ir lineāri atkarīgs no parametriem

$$\mu_{n+1} = \alpha_{\bar{x}} + \beta(x_{n+1} - \bar{x}),$$

tad problēmu vienkāršosim, ļaujot μ_{n+1} būt vienīgajam parametram. Komponentes $\alpha_{\bar{x}}$ un β ir neatkarīgas, tāpēc μ_{n+1} aposteriorais sadalījums būs *normālais* ar vidējo vērtību $m'_{\mu} = m'_{\alpha_{\bar{x}}} + (x_{n+1} - \bar{x}) \times m'_{\beta}$ un dispersiju $(s'_{\mu})^2 = (s'_{\alpha_{\bar{x}}})^2 + (x_{n+1} - \bar{x})^2 \times (s'_{\beta})^2$.

Prognozējamo sadalījumu atradīsim integrējot y_{n+1} un x_{n+1} kopējo aposterioro sadalījumu pēc μ_{n+1}.

$$\begin{aligned} f(y_{n+1} | x_{n+1}, dati) &= \int f(y_{n+1}, \mu_{n+1} | x_{n+1}, dati) d\mu_{n+1} \\ &= \int f(y_{n+1} | \mu_{n+1}, x_{n+1}, dati) \times g(\mu_{n+1} | x_{n+1}, dati) d\mu_{n+1} \\ &= \int f(y_{n+1} | \mu_{n+1}) \times g(\mu_{n+1} | x_{n+1}, dati) d\mu_{n+1} \\ &\propto \int e^{-\frac{1}{2\sigma^2}(y_{n+1} - \mu_{n+1})^2} \times e^{-\frac{1}{2(s'_{\mu})^2}(\mu_{n+1} - m'_{\mu})^2} d\mu_{n+1} \\ &\propto \int e^{-\frac{1}{2\sigma^2 (s'_{\mu})^2 / (\sigma^2 + (s'_{\mu})^2)}\left(\mu_{n+1} - \frac{y_{n+1}(s'_{\mu})^2 + m'_{\mu}\sigma^2}{(s'_{\mu})^2 + \sigma^2}\right)} \times e^{-\frac{1}{2((s'_{\mu})^2 + \sigma^2)}(y_{n+1} - m'_{\mu})^2} d\mu_{n+1}. \end{aligned}$$

Otrais reizinātājs nav atkarīgs no μ_{n+1}, tāpēc to var iznest pirms integrāļa. Atliek

$$f(y_{n+1}|x_{n+1}, dati) \propto e^{-\frac{1}{2((s'_\mu)^2+\sigma^2)}(y_{n+1}-m'_\mu)^2}.$$

Tas ir $N(m'_y, (s'_y)^2)$, kur $m'_y = m'_\mu$ un $(s'_y)^2 = (s'_\mu)^2 + \sigma^2$. Tādēļ novērojuma y_{n+1} prognozējamā vidējā vērtība pie x_{n+1} ir $\mu_{n+1} = \alpha_{\bar{x}} + \beta(x_{n+1} - \bar{x})$ aposteriorā vidējā vērtība plus novērojuma dispersija σ^2.

Beijesa ticamības intervāls prognozei. Vēlamies atrasts intervālu, kura aposteriorā varbūtība saturēt nākamo vērtību y_{n+1}, kura tiks novērota pie vērtības x_{n+1}, ir vienāda ar $1 - \alpha$. Tas būs $(1 - \alpha) \times 100\%$ Beijesa ticamības intervāls prognozei. Prognozējamā sadalījuma vidējā vērtība un dispersija ir attiecīgi m'_y un $(s'_y)^2$. Beijesa ticamības intervāls prognozei tiek uzdots ar

$$\begin{aligned} & m'_y \pm z_{\frac{\alpha}{2}} \times s'_y \\ &= m'_\mu \pm z_{\frac{\alpha}{2}} \times \sqrt{(s'_\mu)^2 + \sigma^2} \\ &= m'_{\alpha_{\bar{x}}} + m'_\beta(x_{n+1} - \bar{x}) \pm z_{\frac{\alpha}{2}} \times \sqrt{(s'_{\alpha_{\bar{x}}})^2 + (s'_\beta)^2(x_{n+1} - \bar{x})^2 + \sigma^2}, \end{aligned}$$

ja zināma novērojumu dispersija σ^2. Ja novērojumu dispersija nav zināma un tās vietā izmanto dispersijas novērtējumu, kas aprēķināts no atlikumiem, tad Beijesa ticamības intervāls ir

$$\begin{aligned} & m'_y \pm t_{\frac{\alpha}{2}} \times s'_y \\ &= m'_\mu \pm t_{\frac{\alpha}{2}} \times \sqrt{(s'_\mu)^2 + \hat{\sigma}^2} \\ &= m'_{\alpha_{\bar{x}}} + m'_\beta(x_{n+1} - \bar{x}) \pm t_{\frac{\alpha}{2}} \times \sqrt{(s'_{\alpha_{\bar{x}}})^2 + (s'_\beta)^2(x_{n+1} - \bar{x})^2 + \hat{\sigma}^2}, \end{aligned}$$

kur kritiskā vērtība tiek ņemta no *Stjudenta t* sadalījuma ar $n - 2$ brīvības pakāpēm. Šie prognožu Beijesa ticamības intervāli ir analogi klasiskajiem prognožu intervāliem. Beijesa ticamības intervāli prognozei vispārīgā gadījumā būs īsāki nekā atbilstošie klasiskie prognožu intervāli, jo Beijesa intervālos tiek izmantota informācija no apriorā sadalījuma un no datiem. Beijesa intervāli dos tieši tādu pašu rezultātu kā klasiskie intervāli, ja gan virziena koeficientam, gan brīvajam loceklim tiks izmantoti plakanie apriorie sadalījumi.

2. Robustās Beijesa metodes

Visi apriorie sadalījumi, kuri iespējamajām parametra vērtībām piešķir pieņemamas varbūtības, rezultātā dod līdzīgus (bet ne vienādus) aposterioros sadalījumus. Beijesa teorēma, izmantojot apriorο informāciju, dod labāku rezultātu nekā klasiskā pieeja, kura ignorē apriorο informāciju.

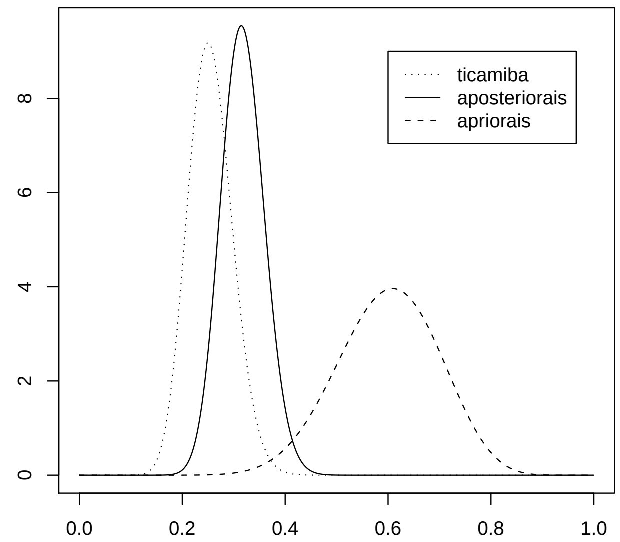
Tomēr ir iespējams kļūdīties apriorā sadalījuma izvēlē. Ja datu ticamība ļoti atšķiras no izvēlētā apriorā sadalījuma, tad apriorais sadalījums spēcīgi ietekmēs aposterioro sadalījumu.

Apskatīsim, kā varam padarīt Beijesa pieeju robustāku pret trūcīgi aprakstītu apriorο sadalījumu. Saistīto apriorο sadalījumu *salikums* dod iespēju to paveikt. Tiks pieļauta maza varbūtība, ka ir izvēlēts nepareizs apriorais sadalījums. Ja ticamība būs ļoti atšķirīga no apriorā sadalījuma, tad aposteriorā varbūtība, ka apriorais sadalījums ir noteikts nepareizi, būs liela un aposteriorais sadalījums vairāk būs atkarīgs no ticamības nekā no apriorā sadalījuma.

2.1. Nepareizi noteikta apriorā sadalījuma ietekme

Ja apriorais sadalījums piešķir lielu varbūtību vērtībām, kurām ir maza ticamība, un mazu varbūtību vērtībām, kurām ir liela ticamība, tad aposteriorais sadalījums piešķirs lielas varbūtības vērtībām, kuras nav stingri atbalstītas ne ar apriorο sadalījumu, ne ar ticamību. Tas var notikt, ja tiek nepareizi noteikts apriorais sadalījums, balstot to uz pagātnes datiem, kuri ir radušies pie savādākiem nosacījumiem nekā jaunie dati.

Piemērs 7. *Artūrs veiks pētījumu par to, cik daudzi pilsētas iedzīvotāji apmeklētu kazino, ja tāds pilsētā tiktu atvērts. Viņš savu apriorο sadalījumu balstīs uz savu draugu viedokļiem. No 25 aptaujātajiem draugiem, 15 apgalvoja, ka apmeklētu kazino. Tāpēc viņš izvēlas* $Beta(a, b)$ *apriorο sadalījumu, kurš saskan ar šiem viedokļiem. Apriorā vidējā vērtība ir* 0.6 *un ekvivalentais izlases apjoms ir* 25*. Tādēļ* $a+b+1 = 25$ *un* $\frac{a}{a+b} = 0.6$*. Tātad* $a = 14.4$ *un* $b = 9.6$*. Pēc tam viņš ņem gadījuma izlasi ar 100 pilsētas iedzīvotājiem un uzzina, ka kazino apmeklētu 25. Viņa aposteriorais sadalījums ir* $Beta(39.4, 84.6)$*. Artūra apriorais sadalījums, ticamība un aposteriorais sadalījums ir parādīti attēlā 15. Apriorais sadalījums un ticamība ir diezgan atšķirīgi. Aposteriorais sadalījums ir starp tiem. Tas piešķir lielu aposterioro varbūtību vērtībām, kuras nav stingri atbalstītas ne ar* *datiem (ticamību), ne ar apriori sadalījumu. Tas nav apmierinošs rezultāts.*

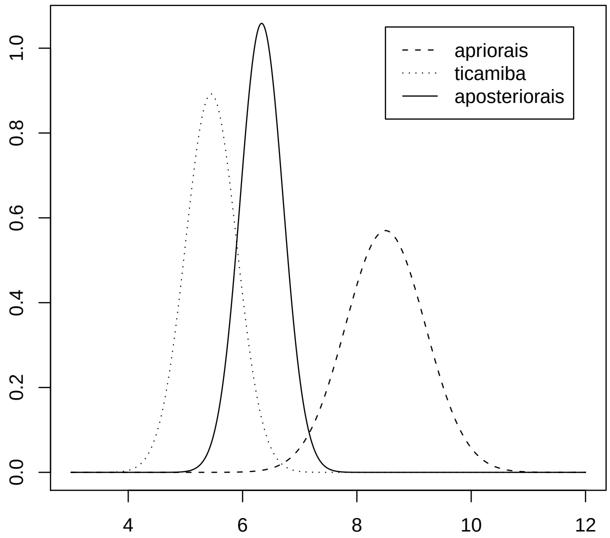


15. att. Artūra apriorais sadalījums, ticamība un aposteriorais sadalījums.

Piemērs 8. *Anda ņems izlasi no vasaras laikā ezerā izšķīdušā skābekļa līmeņa mērījumiem. Viņa pieņem, ka izšķīdušā skābekļa līmenis ir aptuveni normāli sadalīts ar vidējo vērtību* μ *un zināmu dispersiju* $\sigma^2 = 1$. *Līdzīgu eksperimentu viņa iepriekš veica upei, kura ietek ezerā. Viņa nolemj priekš* μ *izmantot* $N(8.5, 0.7^2)$ *apriori sadalījumu, kas ir līdzīgs upes pētījuma rezultātiem. Viņa ņem gadījuma izlasi ar 5 novērojumiem, kuru vidējā vērtība ir* 5.45. *Aposteriorā sadalījuma parametri tiek atrasti, izmantojot vienkāršos normālā sadalījuma precizēšanas likumus* (6.4) *un* (6.5). *Aposteriorais sadalījums ir* $N(6.334, 0.3769^2)$. *Apriorais sadalījums, ticamība un aposteriorais sadalījums ir parādīti attēlā 16. Aposteriorais blīvums ir starp apriori sadalījumu un ticamību, un piešķir lielu varbūtību vērtībām, kuras nav stingri atbalstītas ne ar datiem (ticamību), ne ar apriori sadalījumu. Tas ir ļoti neapmierinošs rezultāts.*

2.2. Beijesa teorēma ar saliktiem apriorajiem sadalījumiem

Pieņemsim, ka apriorais blīvums ir $g_0(\theta)$ un tas ir diezgan precīzs, jo mums ir nozīmīgas apriorās zināšanas. Tomēr mēs vēlamies pasargāt sevi no iespējas, ka nepareizi nosakām apriori sadalījumu, izmantojot nepareizas apriorās zināšanas. Ja apriorais sadalījums ir noteikts nepareizi, tad apriorais sadalījums priekš θ ir $g_1(\theta)$, kas ir vai nu ļoti "vājš"



16. att. Andas apriorais sadalījums, ticamība un aposteriorais sadalījums.

saistītais, vai "plakans" apriorais sadalījums. $g_0(\theta|y_1,\ldots,y_n)$ ir θ aposteriorais sadalījums, ja mēs sākam ar $g_0(\theta)$ kā aprioro sadalījumu. Līdzīgi $g_1(\theta|y_1,\ldots,y_n)$ ir θ aposteriorais sadalījums, ja mēs sākam ar $g_1(\theta)$ kā aprioro sadalījumu:

$$g_i(\theta|y_1,\ldots,y_n) \propto g_i(\theta) f(y_1,\ldots,y_n|\theta).$$

Saliktais apriorais sadalījums

Ievedam jaunu parametru I, kuram ir iespējamas divas vērtības. θ nosacītā apriorā varbūtība ir:

$$g(\theta|i) = \begin{cases} g_0(\theta), & \text{ja } i = 0, \\ g_1(\theta), & \text{ja } i = 1. \end{cases}$$

I apriorais varbūtību sadalījums ir $P(I = 0) = p_0$, kur p_0 ir kāda liela vērtība, piemēram, 0.9, 0.95 vai 0.99, jo mēs domājam, ka mūsu apriorais sadalījums $g_0(\theta)$ ir pareizs. Apriorā varbūtība, ka mūsu apriorais sadalījums ir noteikts nepareizi ir $p_1 = 1 - p_0$. θ un I kopējais apriorais sadalījums ir

$$g(\theta, i) = p_i \times (1 - i) \times g_0(\theta) + (1 - p_i) \times (i) \times g_1(\theta).$$

Šis kopējais sadalījums ir nepārtraukts pēc parametra θ un diskrēts pēc parametra I. Gadījuma lieluma θ marginālais apriorais blīvums tiek atrasts summējot I pa visām iespējamām vērtībām. Tam ir *salikts* apriorais sadalījums, jo tā blīvums

$$g(\theta) = \sum_{0}^{1} p_i g_i(\theta)$$

ir salikts no diviem apriorajiem blīvumiem.

Kopējais aposteriorais sadalījums

θ un I kopējais aposteriorais sadalījums, ja ir doti novērojumi $y_1, \ldots, y_n$, ir

$$g(\theta, i|y_1, \ldots, y_n) = c \times g(\theta, i) \times f(y_1, \ldots, y_n|\theta, i) \quad \text{priekš} \quad i = 0, 1$$

kādai konstantei c. Bet izlase ir atkarīga tikai no θ nevis no i, tāpēc kopējais aposteriorais sadalījums ir

$$\begin{aligned} g(\theta, i|y_1, \ldots, y_n) &= c \times p_i g_i(\theta) f(y_1, \ldots, y_n|\theta) \quad \text{priekš} \quad i = 0, 1 \\ &= c \times p_i h_i(\theta, y_1, \ldots, y_n) \quad \text{priekš} \quad i = 0, 1, \end{aligned}$$

kur $h_i(\theta, y_1, \ldots, y_n) = g_i(\theta) f(y_1, \ldots, y_n|\theta)$ ir parametra un datu kopējais sadalījums, ja $g_i(\theta)$ ir pareizais apriorais sadalījums. Marginālā aposteriorā varbūtība tiek atrasta:

$$\begin{aligned} P(I = i|y_1, \ldots, y_n) &= \int g(\theta, i|y_1, \ldots, y_n) d\theta \\ &= c \times p_i \int h_i(\theta, y_1, \ldots, y_n) d\theta \\ &= c \times p_i f_i(y_1, \ldots, y_n) \end{aligned}$$

priekš $i = 0, 1$, kur $f_i(y_1, \ldots, y_n)$ ir datu marginālā varbūtība (vai varbūtību blīvums), ja $g_i(\theta)$ ir pareizais apriorais sadalījums. Aposterioro varbūtību summa ir 1 un konstante c tiek saīsināta, tāpēc

$$P(I = i|y_1, \ldots, y_n) = \frac{p_i f_i(y_1, \ldots, y_n)}{\sum_{i=0}^{1} p_i f_i(y_1, \ldots, y_n)}.$$

Saliktais aposteriorais sadalījums

θ marginālo aposterioro sadalījumu atrod, summējot visas iespējamās i vērtības no kopējā aposteriorā sadalījuma:

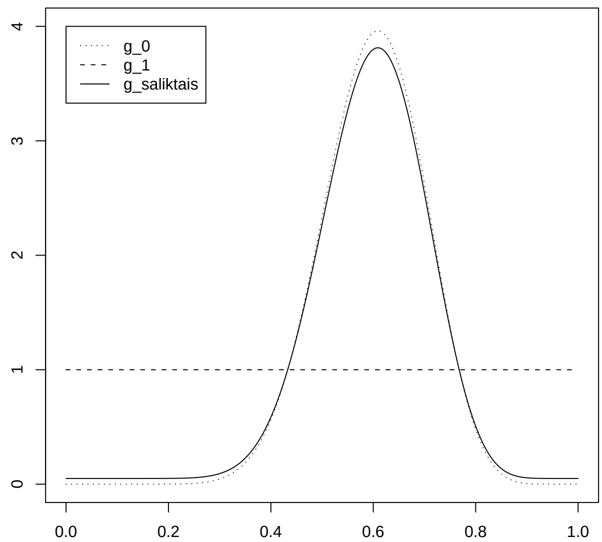
$$g(\theta|y_i, \ldots, y_n) = \sum_{i=0}^{1} g(\theta, i|y_i, \ldots, y_n).$$

Bet ir arī cits veids, kā iegūt kopējo aposterioro sadalījumu no nosacītajām varbūtībām:

$$g(\theta, i|y_i, \ldots, y_n) = g(\theta|i, y_i, \ldots, y_n) \times P(I = i|y_i, \ldots, y_n),$$

kur $g(\theta|i, y_i, \ldots, y_n) = g(\theta|y_i, \ldots, y_n)$ ir aposteriorais sadalījums, ja sāk ar aprioro sadalījumu $g_i(\theta)$. Tāpēc θ marginālais aposteriorais sadalījums ir

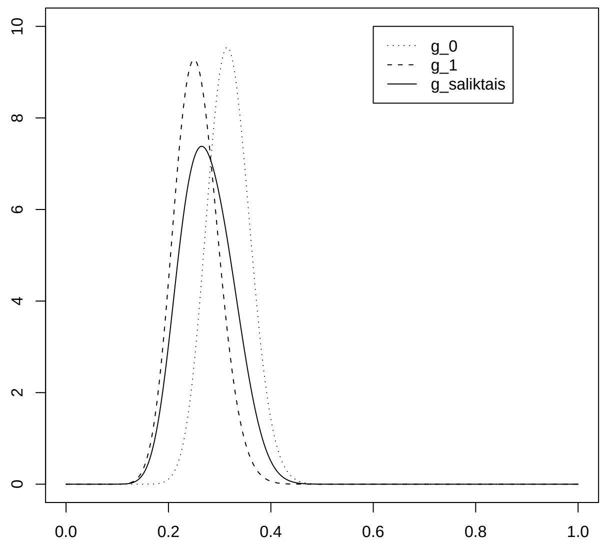
$$g(\theta|y_i, \ldots, y_n) = \sum_{i=0}^{1} g_i(\theta|y_i, \ldots, y_n) \times P(I = i|y_i, \ldots, y_n).$$



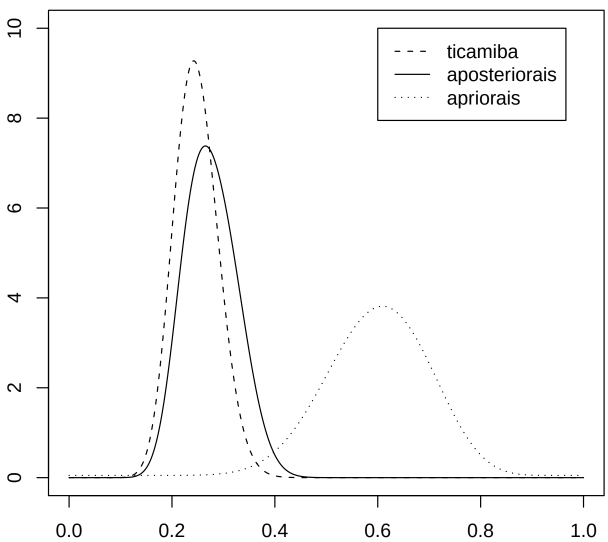
17. att. Viktora saliktais apriorais sadalījums un tā komponentes.

Piemērs 7. *(turpinājums) Artūra draugs Viktors nolemj analizēt Artūra datus ar saliktu aprioro sadalījumu.* g_0 *ir tas pats* $Beta(14.4, 9.6)$ *apriorais sadalījums, kādu izmantoja Artūrs. Par* g_1 *viņš pieņem (vienmērīgo)* $Beta(1, 1)$ *aprioro sadalījumu. Viktors pieņem, ka apriorā varbūtība* $p_0 = 0.95$. *Viktora saliktais apriorais sadalījums un tā komponentes parādītas attēlā 17. Viņa saliktais apriorais sadalījums ir diezgan līdzīgs Artūra apriorajam sadalījumam. Taču tam ir lielāki svari "astēs". Tas viņa aprioro sadalījumu padara robustāku pret nepareizu apriorā sadalījuma noteikšanu. Mūs interesē tikai novērotā vērtība* $y = 25$*:*

$$\begin{aligned} h_0(\pi, y = 25) &= \frac{\Gamma(24)}{\Gamma(14.1)\Gamma(9.6)} \pi^{13.4}(1-\pi)^{8.6} \times \left(\frac{100!}{25!75!}\right) \pi^{25}(1-\pi)^{75} \\ &= \frac{\Gamma(24)}{\Gamma(14.1)\Gamma(9.6)} \times \left(\frac{100!}{25!75!}\right) \times \pi^{38.4}(1-\pi)^{83.6} \end{aligned}$$



18. att. Viktora saliktais aposteriorais sadalījums un tā divas komponentes.



19. att. Viktora saliktais apriorais sadalījums, ticamība un saliktais aposteriorais sadalījums.

un

$$
\begin{aligned}
h_1(\pi, y=25) &= \pi^0(1-\pi)^0 \times \left(\frac{100!}{25!75!}\right)\pi^{25}(1-\pi)^{75} \\
&= \left(\frac{100!}{25!75!}\right)\pi^{25}(1-\pi)^{75}.
\end{aligned}
$$

Katrs no tiem ir Beta sadalījuma reizinājums ar konstanti, tāpēc

$$\int_0^1 h_0(\pi, y=25)d\pi = \frac{\Gamma(24)}{\Gamma(14.1)\Gamma(9.6)} \times \left(\frac{100!}{25!75!}\right) \times \int_0^1 \pi^{38.4}(1-\pi)^{83.6}d\pi$$
$$= \frac{\Gamma(24)}{\Gamma(14.1)\Gamma(9.6)} \times \left(\frac{100!}{25!75!}\right) \times \frac{\Gamma(39.4)\Gamma(84.6)}{\Gamma(124)}$$

un

$$\int_0^1 h_1(\pi, y=25)d\pi = \left(\frac{100!}{25!75!}\right) \times \int_0^1 \pi^{25}(1-\pi)^{75}d\pi$$
$$= \left(\frac{100!}{25!75!}\right) \times \frac{\Gamma(26)\Gamma(76)}{\Gamma(102)}.$$

Atceroties, ka $\Gamma(a) = (a-1) \times \Gamma(a-1)$ *un ja* a *ir vesels skaitlis, tad* $\Gamma(a) = (a-1)!$. *No otrā integrāļa iegūstam*

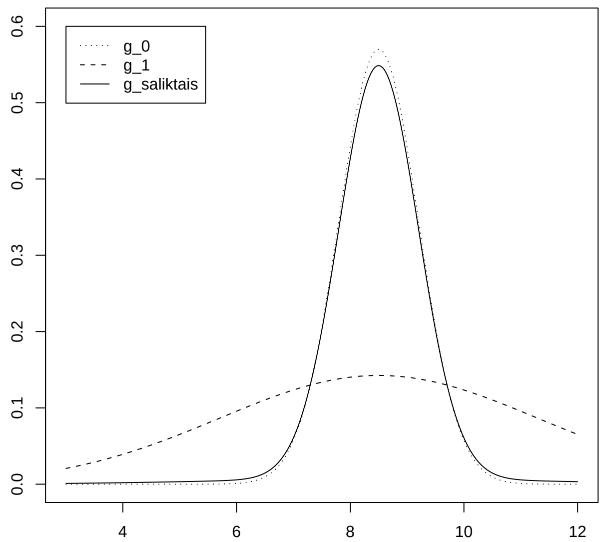
$$f_1(y=25) = \int_0^1 h_1(\pi, y=25)d\pi = \frac{1}{101} = 9.90099 \times 10^{-3}.$$

Pirmo integrāli var novērtēt skaitliski:

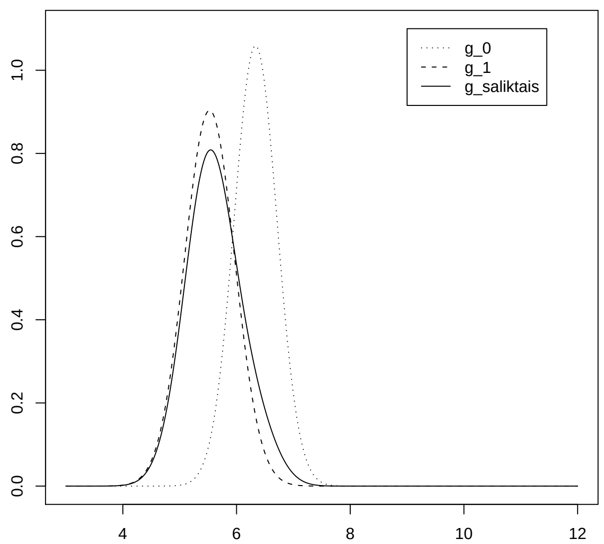
$$f_0(y=25) = \int_0^1 h_0(\pi, y=25)d\pi = 2.484 \times 10^{-4}.$$

Tāpēc aposteriorās varbūtības ir $P(I=0|25) = 0.323$ *un* $P(I=1|25) = 0.677$. *Aposteriorais sadalījums ir salikums* $g(\pi|25) = 0.323 \times g_0(\pi|25) + 0.677 \times g_1(\pi|25)$, *kur* $g_0(\pi|y)$ *un* $g_1(\pi|y)$ *ir saistītie aposteriorie sadalījumi, kas atrasti izmantojot atbilstošos aprioros sadalījumus* g_0 *un* g_1. *Viktora saliktais aposteriorais sadalījums un tā divas komponentes parādītas attēlā 18. Viktora apriorais un aposteriorais sadalījums kopā ar ticamību ir parādīti attēlā 19. Ja apriorais sadalījums un ticamība nesaskan, tad būtu jādarbojas ar ticamību, jo tā ir no datiem. Virspusēji Viktora apriorais sadalījums izskatās ļoti līdzīgs Artūra apriorajam sadalījumam. Taču tam ir smagākas astes un tas ļauj viņa aposteriorajam sadalījumam būtu ļoti tuvam ticamībai. Tas ir daudz apmierinošāk nekā Artūra analīze, kas parādīta attēlā 15.*

Piemērs 8. *(turpinājums) Andas draudzene Katrīna veiks analīzi, izmantojot normālo aprioro sadalījumu salikumu.* $g_0(\theta)$ *būs tas pats* $N(8.5, 0.7^2)$ *un* $g_1(\theta)$ *būs* $N(8.5, (4 \times 0.7)^2)$, *kuram ir tāda pati vidējā vērtība kā Andas apriorajam sadalījumam, bet ar 4 reizes lielāku standartnovirzi. Viņa pieļauj* 0.05 *varbūtību, ka Andas apriorais sadalījums bija noteikts nepareizi. Katrīnas saliktais apriorais sadalījums un tā komponentes ir parādītas attēlā 20. Viņas saliktais apriorais sadalījums ir ļoti līdzīgs Andas apriorajam*



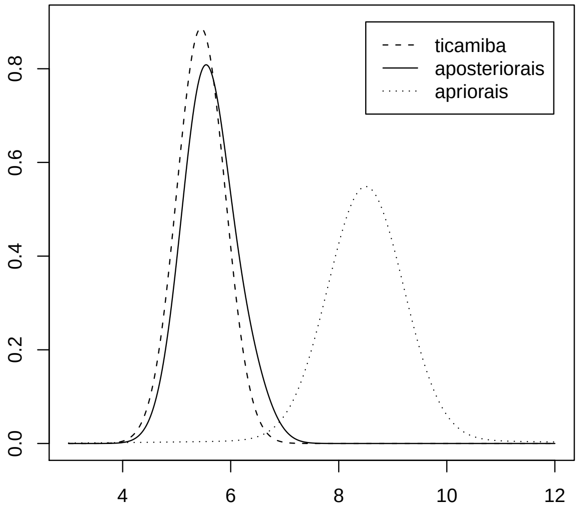
20. att. Katrīnas saliktais apriorais sadalījums un tā komponentes.



21. att. Katrīnas saliktais aposteriorais sadalījums un tā divas komponentes.

sadalījumam. Katrīnas aposteriorais sadalījums $g_0(\theta|\bar{y})$ *ir* $N(6.334, 0.3769^2)$*, kurš ir tāds pats kā Andai. Katrīnas aposteriorais sadalījums* $g_1(\theta|\bar{y})$*, ja ir nepareizi noteikts apriorais sadalījums, ir* $N(5.526, 0.4416^2)$*, kur parametri tiek atrasti ar vienkāršajiem normālā sadalījuma precizēšanas likumiem* (6.4) *un* (6.5)*. Normālā sadalījuma gadījumā*

$$\begin{aligned} h_i(\mu, y_1, \ldots, y_n) &\propto g_i(\mu) \times f(\bar{y}|\mu) \\ &\propto e^{-\frac{1}{s_i^2}(\mu - m_i)^2} \times e^{-\frac{1}{\sigma^2/n}(\bar{y}-\mu)^2}, \end{aligned}$$



22. att. Katrīnas saliktais apriorais sadalījums, ticamība un saliktais aposteriorais sadalījums.

kur m_i *un* s_i^2 *ir apriorā sadalījuma* $g_i(\mu)$ *vidējā vērtība un dispersija. Integrālis*

$$\int h_i(\mu, y_1, \ldots, y_n) d\mu$$

dod izlases nenosacīto varbūtību, ja g_i *ir pareizs apriorais sadalījums. Sareizinot eksponentes un integrējot pēc* μ*, iegūstam*

$$\begin{aligned} f_i(\bar{y}) &= \int h_i(\mu, \bar{y}) d\mu \\ &\propto \frac{1}{\sqrt{s_i^2 + \sigma^2/n}} \times e^{-\frac{1}{2(s_i^2+\sigma^2/n)}(\bar{y}-m_i)^2}, \end{aligned}$$

kas ir normālais blīvums ar vidējo vērtību m_i *un dispersiju* $\frac{\sigma^2}{n} + s_i^2$*. Šajā piemērā* $m_0 = 8.5$*,* $s_0^2 = 0.7^2$*,* $m_1 = 8.5$*,* $s_1^2 = (4 \times 0.7)^2$*,* $\sigma^2 = 1$ *un* $n = 5$*. Dati tiek apkopoti ar vērtību* $\bar{y} = 5.45$*, kura iegūta no izlases. Ievietojot šīs vērtības, iegūstam* $P(I = 0|\bar{y} = 5.45) = 0.12$ *un* $P(I = 1|\bar{y} = 5.45) = 0.88$*. Tāpēc Katrīnas aposteriorais sadalījums ir salikums* $0.12 \times g_0(\mu|\bar{y}) + 0.88 \times g_1(\mu|\bar{y})$*. Katrīnas saliktais aposteriorais sadalījums un tā komponentes ir parādītas attēlā 21. Katrīnas apriorais sadalījums, ticamība un aposteriorais sadalījums ir parādīti attēlā 22. Salīdzinot to ar Andas analīzi attēlā 16., redzams, ka salikumu izmantošana ir devusi aposterioro sadalījumu, kurš ir daudz tuvāks ticamībai, nekā aposteriorais sadalījums, kas iegūts ar sākotnējo nepareizi noteikto apriorо sadalījumu. Tas ir daudz apmierinošāks rezultāts.*

3. Izveidoto programmu kodi

Binomiālajai proporcijai

```
pi<-seq(0,1,length=10000)
anna<-dbeta(pi,4.8,19.2)
plot(pi,anna,type="l",ylim=c(0,5),ann=FALSE)
atis<-dbeta(pi,1,1)
lines(pi,atis,lty=2,xlab=NULL,ylab=NULL)
pi<-seq(0,0.1,length=10000)
imants<-20*pi
lines(pi,imants,lty=3)
pi<-seq(0.1,0.3,length=10000)
imants<-2*(pi/pi)
lines(pi,imants,lty=3)
pi<-seq(0.3,0.5,length=10000)
imants<-5-10*pi
lines(pi,imants,lty=3)
pi<-seq(0.5,1,length=10000)
imants<-0*pi/pi
lines(pi,imants,lty=3)
legend(0.6,5,legend=c("Annas apriorais","Ata apriorais","Imanta apriorais"),
 +lty=c(1,2,3))

library(Bolstad)
pi<-seq(0,1,by=0.001)
pi.prior<-rep(0,length(pi))
pi.prior[pi<=0.1]<-20*pi[pi<=0.1]
pi.prior[0.1<=pi]<-2*(pi/pi)[0.1<=pi]
pi.prior[0.3<=pi]<-5-10*pi[0.3<=pi]
pi.prior[0.5<pi]<-0*pi[0.5<pi]
results<-binogcp(26,100,"user",pi=pi,pi.prior=pi.prior,ret=TRUE)
anna<-dbeta(pi,30.8,93.2)
plot(pi,anna,type="l",ylim=c(0,10.5),ann=FALSE)
```

```
atis<-dbeta(pi,27,75)
lines(pi,atis,lty=2,xlab=NULL,ylab=NULL)
lines(pi,results$posterior,lty=3)
legend(0.5,10.5,legend=c("Annas aposteriorais","Ata aposteriorais",
 +"Imanta aposteriorais"),lty=c(1,2,3))

dens<-pi*results$posterior
post.mean<-sintegral(pi,dens)
dens<-(pi-post.mean)^2*results$posterior
post.var<-sintegral(pi,dens)
post.sd<-sqrt(post.var)
cdf<-sintegral(pi,results$posterior,n.pts=length(pi),ret=TRUE)
post.med<-cdf$x[with(cdf,which.max(x[y<=0.5]))]
post.q1<-cdf$x[with(cdf,which.max(x[y<=0.25]))]
post.q3<-cdf$x[with(cdf,which.max(x[y<=0.75]))]
post.iqr<-post.q3-post.q1
lb<-cdf$x[with(cdf,which.max(x[y<=0.025]))]
ub<-cdf$x[with(cdf,which.max(x[y<=0.975]))]
lb<-post.mean-qnorm(0.975)*post.sd
ub<-post.mean+qnorm(0.975)*post.sd
```

Normālā sadalījuma vidējai vērtībai

```
x<-seq(5,50,length=10000)
gvido<-dnorm(x,mean=30,sd=4)
plot(x,gvido,type="l",ann=FALSE,yaxt='n')
laura<-0.025*x/x
lines(x,laura,lty=2,xlab=NULL,ylab=NULL)
x<-seq(5,18,length=10000)
normunds<-0*x/x
lines(x,normunds,lty=3)
x<-seq(18,24,length=10000)
normunds<-(x-18)/132
lines(x,normunds,lty=3)
```

```
x<-seq(24,40,length=10000)
normunds<-(1/22)*(x/x)
lines(x,normunds,lty=3)
x<-seq(40,46,length=10000)
normunds<-(46-x)/132
lines(x,normunds,lty=3)
x<-seq(46,50,length=10000)
normunds<-0*x/x
lines(x,normunds,lty=3)
legend(5,0.1,legend=c("Gvido apriorais","Lauras apriorais",
 +"Normunda apriorais"),lty=c(1,2,3))

library(Bolstad)
y<-c(36,29,29,34,34,31,31,32,32,32,32,32)
mu<-seq(10,50,by=0.01)
mu.prior<-rep(0,length(mu))
mu.prior[mu<=18]<-0*mu[mu<=18]
mu.prior[18<=mu]<-(mu[18<=mu]-18)/132
mu.prior[24<=mu]<-(1/22)*(mu/mu)[24<=mu]
mu.prior[40<=mu]<-(46-mu[40<=mu])/132
mu.prior[46<=mu]<-0*mu[46<=mu]
results<-normgcp(y,2,"user",mu=mu,mu.prior=mu.prior,ret=T)
gvido<-dnorm(mu,mean=31.96,sd=0.5714)
plot(mu,gvido,type="l",xlim=c(28,34),ann=FALSE,yaxt='n')
laura<-dnorm(mu,mean=32,sd=0.5774)
lines(mu,laura,lty=2)
lines(mu,results$posterior,lty=3)
legend(28,0.7,legend=c("Gvido aposteriorais","Lauras aposteriorais",
 +"Normunda aposteriorais"),lty=c(1,2,3))

dens<-mu*results$posterior
post.mean<-sintegral(mu,dens)
```

```
dens<-(mu-post.mean)^2*results$posterior
post.var<-sintegral(mu,dens)
post.sd<-sqrt(post.var)

cdf<-sintegral(mu,results$posterior,n.pts=length(mu),ret=TRUE)
post.med<-cdf$x[with(cdf,which.max(x[y<=0.5]))]
post.q1<-cdf$x[with(cdf,which.max(x[y<=0.25]))]
post.q3<-cdf$x[with(cdf,which.max(x[y<=0.75]))]
post.iqr<-post.q3-post.q1
lb<-cdf$x[with(cdf,which.max(x[y<=0.025]))]
ub<-cdf$x[with(cdf,which.max(x[y<=0.975]))]
lb<-post.mean-qnorm(0.975)*post.sd
ub<-post.mean+qnorm(0.975)*post.sd
```

Vienkāršai lineārai regresijai

```
n<-25
x<-c(14.36,14.48,14.53,14.52,14.35,14.31,14.44,14.23,14.32,14.57,
 +14.28,14.36,14.5,14.52,14.28,14.13,14.54,14.6,
  +14.86,14.28,14.09,14.2,14.5,14.02,14.45)
x_vid<-mean(x)
y<-c(13.84,14.41,14.22,14.63,13.95,14.37,14.41,13.99,13.89,14.59,
 +14.32,14.31,14.43,14.44,14.14,13.9,14.37,
  +14.34,14.78,13.76,13.85,13.89,14.22,13.8,14.67)
y_vid<-mean(y)
xy<-x*y
xy_vid<-mean(xy)
x2<-x^2
x2_vid<-mean(x2)
y2<-y^2
y2_vid<-mean(y2)
B<-(xy_vid-x_vid*y_vid)/(x2_vid-(x_vid^2))
y_hat<-y_vid+B*(x-x_vid)
A0<-y_vid-B*x_vid
```

```
Ax_<-A0+B*x_vid
sigma2_nov<-sum((y-(Ax_+B*(x-x_vid)))^2)/(n-2)
sigma_nov<-sqrt(sigma2_nov)
SSx<-sum((x-x_vid)^2)
klas_int_lb<-(B-((qt(0.975,n-2)*sigma_nov)/sqrt(SSx)))
klas_int_ub<-(B+((qt(0.975,n-2)*sigma_nov)/sqrt(SSx)))
plot(x,y,pch=19)
lines(x,y_hat,lty=1)

x<-seq(0.2,1.8,length=10000)
poster<-dnorm(x,mean=1.2034,sd=0.17001)
plot(x,poster,type="l",ann=FALSE,ylim=c(0,2.3))
pri<-dnorm(x,mean=1,sd=0.3)
lines(x,pri,lty=2)
legend(0.2,2.3,legend=c("aposteriorais","apriorais"),lty=c(1,2))

x<-seq(13,16,length=10000)
my<-14.2219+1.2034*(x-14.3888)
plot(x,my,type="l",ylim=c(12,17),ann=F)
sy2<-0.0348644+(0.037318^2)+((0.17001^2)*((x-14.3888)^2))
sy<-sqrt(sy2)
low<-my-(2.069*sy)
lines(x,low,lty=2)
up<-my+(2.069*sy)
lines(x,up,lty=3)
legend(13,17,legend=c("95% apakseja robeza","95% augseja robeza",
 +"videja vertiba"),lty=c(2,3,1))
```

Ticamības intervālu pārklājuma precizitātei

```
library(Bolstad)
mu<-0
sd<-1
mu_prior<-0
```

```
sd_prior<-1
fun<-function(n){
dati<-rnorm(n,mu,sd)
pos<-normnp(dati,m.x=mu_prior,s.x=sd_prior,sigma.x=sd,ret=TRUE)
c.i.<-qnorm(c(0.025,0.975),pos$mean,pos$sd)
if ((mu>=c.i.[1]) & (mu<=c.i.[2])) i=1 else i=0}
N<-10000
n<-100
rez<-replicate(N,fun(n))
length(rez[rez>0])/N
```

Maģistra darbs “Beijesa metodes matemātiskajā statistikā” izstrādāts LU Fizikas un Matemātikas fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Baiba Buceniece

____________________ (paraksts) ________________ (datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: doc. Dr.math. Jānis Valeinis

____________________ (paraksts) ________________ (datums)

Recenzents: lektors Jānis Smotrovs

____________________ (paraksts) ________________ (datums)

Darbs iesniegts Matemātikas nodaļā ________________ (datums)

____ (darbu pieņēma)

Darbs aizstāvēts maģistra gala pārbaudījuma komisijas sēdē

________________ (datums) prot. Nr. ________, vērtējums________

Komisijas sekretārs/-e: ________________ (Vārds, Uzvārds) ________________ (paraksts)