VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Andrius Čiginas

Dispersijos minimizavimas renkant nelygių tikimybių imtį

Magistro darbas

VILNIUS 2006

Matematinės analizės katedra

Darbo vadovas **prof. habil. dr. M. Bloznelis** ____________________
(parašas)

Darbas apgintas ____*2006 m. birželio mėn. 1 d.*____
Gynimo posėdžio protokolo Nr.________________

Registravimo Nr._____________
2006-06-20____________________

Turinys

1. Įvadas

Nagrinėkime populiaciją $\mathcal{U} = \{u_1, \ldots, u_N\}$ ir tarkime, kad norime įvertinti šios populiacijos parametrą $t_y = \sum_{1 \le i \le N} y_i$, čia $y_i = y(u_i)$ žymi tyrimo kintamojo y reikšmę populiacijos elementui u_i. Tuo tikslu iš populiacijos $\mathcal{U}$ rinksime imtį s. Tarkime, kad apie visus populiacijos elementus turime papildomos informacijos, kurią pažymėkime kaip vektorių $x = (x_1, \ldots, x_N)$ su teigiamomis koordinatėmis. Koordinatę x_i interpretuokime kaip elemento u_i dydį. Tuo atveju, kuomet kintamieji x ir y yra stipriai koreliuoti, imties išrinkimo plane gali būti tikslinga pasinaudoti santykiniais svoriais

$$p_i = x_i/t_x, \qquad t_x = \sum_{i=1}^{N} x_i. \tag{1}$$

Pavyzdžiui, priklausymo imčiai tikimybes $\pi_i = P(u_i \in s)$ galima imti proporcingas svoriui p_i, t.y.

$$\pi_i \approx cp_i, \qquad 1 \le i \le N. \tag{2}$$

Kröger, Särndal ir Teikari (2003) yra pateikę asimetrinių populiacijų ir ėmimo planų pavyzdžių, kuomet priklausymo imčiai tikimybės yra artimos (2) bei parodę, kad tokiu atveju kai kurių dažnai naudojamų populiacijos sumos t_y įvertinių $\hat{t}_y$ dispersijos yra pastebimai didesnės nei tų pačių įvertinių dispersijos, kai priklausymo imčiai s tikimybėmis imama $\pi_i \approx cp_i(h)$, kai čia

$$p_i(h) = (1-h)p_i + h/N, \qquad 1 \le i \le N. \tag{3}$$

Čia $h \in [0,1]$. Jie nagrinėjo Horvitz-Thompson'o, regresinį ir apibendrintą regresinį (GREG) įvertinius ir ėmimo planus, kuomet fiksuoto dydžio imtys yra renkamos be grąžinimo. Kompiuterinis modeliavimas parodė, kad pateiktiems pavyzdžiams dispersija minimizuojama parekant $h \in (0.2; 0.5)$.

Įvertinio dispersiją minimizuojančią parametro h reikšmę pažymėkime h^*. Nežinant tyrimo kintamojo y reikšmių visiems populiacijos elementams, neįmanoma sužinoti ir šio h^*. Paprasčiau būtų atsakyti į klausimą - ar $h^* > 0$ (t.y. kada į priklausymo imčiai tikimybes yra prasminga įtraukti pastovią komponentę) vienokiems ar kitokiems ėmimo planams, populiacijų modeliams ir įvertiniams $\hat{t}_y$? Kitas įdomus klausimas - kaip naudojantis papildoma informacija sukonstruoti reikšmės h^* įvertį? Pasirenkant gana paprastas aplinkybes, šiame darbe ir yra mėginama atsakyti į pastaruosius klausimus.

Sukonkretinkime situaciją. Tarkime, kad populiacijos taškai $\{(y_i,x_i) : 1 \le i \le N\}$ yra išsibarstę taip, tarytumei jie butų sugeneruoti naudojant tikimybinį modelį, kuriame $y_1, \ldots, y_N$ yra nepriklausomų atsitiktinių dydžių

$Y_1, \ldots, Y_N$ realizacijos. Jei $\hat{t}_y$ yra įvertinys, nusakytas imtimi s, bei imties elementų priklausymo tai imčiai tikimybėmis $\pi_i \approx cp_i(h)$, tai tegu $D_h^* = D_h^*(\hat{t}_y)$ žymi $\hat{t}_y$ sąlyginę dispersiją esant sąlygai $Y_1 = y_1, \ldots, Y_N = y_N$. Be to, tegu D_h žymi šios dispersijos vidurkį, t.y. $D_h = \mathbf{E}D_h^*$. Kol kas sakykime, kad intervale $0 \leq h \leq 1$ funkcija $h \to D_h$ įgyja vienintelę mažiausią reikšmę. Pažymėkime

$$h_0 = argmin(D_h). \tag{4}$$

Tada, remiantis didžiųjų skaičių dėsniu, galima tikėtis, kad dideliems N, skaičius h_0 yra artimas funkcijos $h \to D_h^*(\hat{t}_y)$ minimumo taškui. Dėl to į h_0 gali būti žiūrima kaip į nežinomo atsitiktinio dydžio h^* artinį. Tam, kad įsitikinti tokios aproksimacijos kokybe, galima pamėginti įvertinti vidutinę kvadratinę paklaidą $\mathbf{E}(h^* - h_0)^2$ ir palyginti funkcijų $D_{h^*}^*$, $D_{h_0}^*$, D_0^* ir D_1^* vidurkius.

Darbe nagrinėjamas paprastas Poisson'o imties, renkamos iš populiacijos, kuri yra generuota naudojant tiesinės regresijos modelį (žr. Särndal, Swensson ir Wretman (1997), 226 p.) atvejis. Modeliavimo pavyzdžiu buvo pasirinktas būtent Poisson'o ėmimas, nes tokiam imties išrinkimo planui minimizavimo problemos išsprendžiamos vienareikšmiškai, o ir pati analizė yra sąlyginai paprasta. Nagrinėjant du populiarius įvertinius - Horvitz-Thompson'o ir regresinį - parodoma, kad gana dažnai turime $h_0 > 0$. Nustatomas apytikslis skirtumo $h^* - h_0$ skirstinys, pasiūlomos išreikštinės dydžio h^* aproksimacijos. Taip pat pateikiama empirinė aproksimacijos $h^* \approx h_0$ tikslumo ir efektyvumo analizė.

Didelę dalį idėjų ir rezultatų, pateiktų šiame darbe, galima rasti VU MIF preprinte 2005 – 11 M. Bloznelis, A. Čiginas, On Variance minimization for unequal probability sampling.

Padėka. Šis magistro darbas buvo rengiamas dalyvaujant Šiaurės Ministrų Tarybos projekte Nordplus Neighbour. Dėkoju Šiaurės Ministrų Tarybos, Švedijos Umeå universiteto Matematikos ir matematinės statistikos katedros darbuotojams ir asmeniškai prof. G. Kulldorffui už paramą.

Acknowledgements. The master thesis has been prepared taking part in the program Nordplus Neighbour of the Nordic Council of Ministers. I would like to thank the staff of the Nordic Council of Ministers, department of Mathematics and Mathematical Statistics of the Umeå University and prof. G. Kulldorff personally for the support.

2. Populiacijos modelis

Tarkime, kad $y_1, \ldots y_N$ yra nepriklausomų atsitiktinių dydžių $Y_1, \ldots, Y_N$ realizacijos, tokių, kad kiekvienam k,

$$\mathbf{E}(Y_k) = \beta_1 + \beta_2 x_k, \qquad \mathbf{V}(Y_k) = \sigma_k^2. \tag{5}$$

Čia $\sigma_1, \ldots \sigma_N$ ir $x_1, \ldots, x_N$ nėra atsitiktiniai skaičiai, ir $x_k > 0$ kiekvienam k. Taip pat toliau visur laikysime, kad $\beta_2 \neq 0$.

Įveskime du populiacijos parametrus

$$M_1(\gamma) = \sum_{i=1}^N p_i^{\gamma-1} - \frac{1}{N}\sum_{i=1}^N p_i^{\gamma-2}, \qquad M_2(\gamma) = \sum_{i=1}^N p_i^{\gamma+1} - \frac{1}{N}\sum_{i=1}^N p_i^{\gamma},$$

kai čia $\gamma \in [0{,}2]$. Pastebėkime, kad $M_1(2) = M_2(0) = 0$, ir, kad $M_1(\gamma) = M_2(\gamma) = 0$, kai tik $p_i = N^{-1}$, $1 \leq i \leq N$, kiekvienam $\gamma \in [0{,}2]$.

Lema. *Tarkime, kad mažiausiai dvi iš tikimybių $\{p_i\}$ yra skirtingos.*
(i) $M_1(\gamma) < 0$, kai $\gamma \in [0{,}2)$. (ii) $M_2(\gamma) > 0$, kai $\gamma \in (0{,}2]$.

Įrodymas. Tvirtinimui (i) įrodyti sudarykime Lagrange'o funkciją

$$F(p_1, \ldots, p_N) = \sum_{i=1}^N p_i^{\gamma-1} - \frac{1}{N}\sum_{i=1}^N p_i^{\gamma-2} - \lambda\Big(\sum_{i=1}^N p_i - 1\Big).$$

Diferencijuodami ją ir išvestines prilygindami nuliui turime

$$\frac{\partial F}{\partial p_i} = (\gamma-1)p_i^{\gamma-2} - \frac{1}{N}(\gamma-2)p_i^{\gamma-3} - \lambda = 0, \quad 1 \leq i \leq N, \tag{6}$$

bei dar nepamirštame sąlygos $p_1 + \ldots + p_N = 1$. Lengva matyti, kad taškas $(\tilde{p}_1, \ldots, \tilde{p}_N) = (\frac{1}{N}, \ldots, \frac{1}{N})$ yra (6) sistemos sprendinys ir tai yra funkcijos $M_1(\gamma) = M_1(\gamma)(p_1, \ldots, p_N)$ griežtojo sąlyginio lokaliojo maksimumo taškas, nes šiame taške $\frac{\partial^2 F}{\partial p_i^2} = 2N^{3-\gamma}(\gamma-2) < 0$, kai $1 \leq i \leq N$ ir $\frac{\partial^2 F}{\partial p_i \partial p_j} = 0$, kai $1 \leq i < j \leq N$, ir tada $d^2F(\tilde{p}_1, \ldots, \tilde{p}_N) = 2N^{3-\gamma}(\gamma-2)\sum_{i=1}^N dp_i^2 < 0$. Kaip žinia, taške $(\tilde{p}_1, \ldots, \tilde{p}_N) = (\frac{1}{N}, \ldots, \frac{1}{N})$ turime $M_1(\gamma) = 0$.

Įsitikinsime, kad (6) sistema kitų sprendinių neturi. Jei tarsime, kad be rastojo egzistuoja dar vienas (6) sistemos sprendinys, tarkime $(\dot{p}_1, \ldots, \dot{p}_1)$, tai egzistuoja tokie indeksai $i_0, j_0 \in \{1, \ldots, N\}$, $i_0 \neq j_0$, kad $\dot{p}_{i_0}^{-1} < N$ ir $\dot{p}_{j_0}^{-1} > N$. Šiuos indeksus atitinkančias (6) sistemos lygtis atėmę vieną iš kitos, pažymėkime

$$\Upsilon = (\gamma-1)(\dot{p}_{i_0}^{\gamma-2} - \dot{p}_{j_0}^{\gamma-2}) - \frac{1}{N}(\gamma-2)(\dot{p}_{i_0}^{\gamma-3} - \dot{p}_{j_0}^{\gamma-3}) = 0. \tag{7}$$

Aišku, kad neva egzistuojantis sprendinys $(\dot{p}_1, \ldots, \dot{p}_1)$ turi tenkinti ir šią lygtį. Tačiau yra akivaizdu, kad $\Upsilon < 0$, kai $\gamma \in [1,2)$, o kai $\gamma \in [0,1)$, tai turime $\Upsilon = (\gamma-1)(\dot{p}_{i_0}^{\gamma-2} - \dot{p}_{j_0}^{\gamma-2}) - N^{-1}(\gamma-2)(\dot{p}_{i_0}^{-1}\dot{p}_{i_0}^{\gamma-2} - \dot{p}_{j_0}^{-1}\dot{p}_{j_0}^{\gamma-2}) < \dot{p}_{i_0}^{\gamma-2} - \dot{p}_{j_0}^{\gamma-2} < 0$. Taigi, prielaida apie dar vieną kritinį tašką yra neteisinga ir tvirtinimas (i) įrodytas.
Tvirtinimas (ii) įrodomas analogiškai. □

3. Poisson'o ėmimas

Renkant Poisson'o imtį populiacijos elementai u_k patenka į aibę s su atitinkamomis tikimybėmis π_k nepriklausomai vienas nuo kito. T.y. atsitiktiniai dydžiai $\mathbb{I}_k := \mathbb{I}_{\{u_k \in s\}}$ yra nepriklausomi. Pasirinkus skaičius $n < N$ ir $h \in [0,1]$ priklausymo imčiai tikimybėmis imama

$$\pi_k = \pi_k(h) = np_k(h), \quad 1 \leq k \leq N. \tag{8}$$

Tada tikėtinas renkamos imties dydis yra

$$\mathbf{E}(\mathbb{I}_1 + \cdots + \mathbb{I}_N) = \pi_1(h) + \cdots + \pi_N(h) = n. \tag{9}$$

Paprastumo dėlei, iš karto susitarkime nagrinėti atvejį

$$\pi_k(0) < 1, \qquad \text{kiekvienam} \qquad k = 1, \ldots, N. \tag{10}$$

Aišku, kad tada ir $\pi_k(h) < 1$ visiems $h \in [0,1]$ ir $k = 1, \ldots, N$.

Mėginsime parodyti, kad aprašyto ėmimo atveju funkcijos $h \to D_h^*$ ir $h \to D_h$ yra iškilos Horvitz-Thompson'o įvertinio ir regresinio įvertinio atvejais. Tai reikš, kad skaičiai $h_0 = argminD_h$ ir $h^* = argminD_h^*$ apibrėžti korektiškai.

4. Horvitz-Thompson'o įvertinys

Horvitz-Thompson'o įvertinys (toliau jį vadinsime HT įvertiniu)

$$\hat{t}_{yHT} = \sum_{i=1}^{N} \mathbb{I}_i y_i \pi_i^{-1} \tag{11}$$

yra nepaslinktasis, o jo dispersija yra

$$D_h^* = \sum_{i=1}^{N} y_i^2 (1 - \pi_i)\pi_i^{-1}. \tag{12}$$

Teiginys 1. *Funkcijos $h \to D_h$ ir $h \to D_h^*$ yra iškilos. Šios funkcijos yra konstantos, kai tik $p_i = N^{-1}$ kiekvienam $i = 1, \dots, N$.*

Įrodymas. Paprasta įsitikinti, kad

$$D_h = \sum_{k=1}^{N} \Big(\frac{1}{n} \frac{1}{p_k(h)} - 1 \Big) (\beta_1^2 + 2\beta_1\beta_2 x_k + \beta_2^2 x_k^2 + \sigma_k^2). \tag{13}$$

Pastebėkime, kad $\frac{\partial^2}{\partial h^2} \frac{1}{p_i(h)} > 0$, kai $p_i \neq N^{-1}$. Tuo atveju, kai $p_i = N^{-1}$, turime, kad $\frac{\partial}{\partial h} \frac{1}{p_i(h)} = 0$. Todėl funkcija $h \to D_h$ yra iškila ir D_h yra konstanta tik tada, kai $p_i = N^{-1}$ kiekvienam $i = 1, \dots, N$. Visiškai taip pat mąstytume ir apie funkciją $h \to D_h^*$. □

Sekantis teiginys įrodo, kad labai dažnai turime $h_0 > 0$. Vadinasi, priklausymo imčiai tikimybės (8), į kurias su svoriu $h_0 > 0$ yra įtraukta visiems populiacijos elementams vienoda ėmimo komponentė, lems mažesnę HT įvertinio dispersiją, nei klasikinis imties plano tikimybių (2) pasirinkimas.

Teiginys 2. *Sakykime, kad $\sigma_k^2 = \sigma^2 x_k^\gamma$, kur $\gamma \in [0{,}2]$. Tarkime, kad mažiausiai dvi iš tikimybių $\{p_i\}$ yra skirtingos.*
(i) Sakykime, kad $\gamma \in [0{,}2)$. Jei $\beta_1\beta_2 > 0$, tai $0 < h_0 < 1$. Jei $\beta_1 = 0, \beta_2 \neq 0$, tai tada turime $0 < h_0 < 1$, kai $\sigma^2 > 0$ ir turime $h_0 = 0$, kai $\sigma^2 = 0$.
(ii) Sakykime, kad $\gamma = 2$. Jei $\beta_1\beta_2 > 0$, tai $0 < h_0 < 1$. Jei $\beta_1 = 0, \beta_2 \neq 0$, tai $h_0 = 0$.

Įrodymas. Tegu D'_h reiškia iškilos funkcijos $h \to D_h$ išvestinę. Tvirtinimas (i) seka iš sąlygos $D'_0 < 0$, kurią atitinka (14) ir sąlygos $D'_1 > 0$, kurią atitinka (15). Čia

$$\beta_1^2 M_1(0) + 2\beta_1\beta_2 t_x M_1(1) + \sigma^2 t_x^\gamma M_1(\gamma) < 0, \tag{14}$$

$$2\beta_1\beta_2 M_2(1) + \beta_2^2 t_x M_2(2) + \sigma^2 t_x^{\gamma-1} M_2(\gamma) > 0. \tag{15}$$

Tvirtinimas (ii) seka iš sąlygos $D'_0 < 0$, kurią atitinka (16) ir sąlygos $D'_1 > 0$, kurią atitinka (17). Čia

$$\beta_1^2 M_1(0) + 2\beta_1\beta_2 t_x M_1(1) < 0, \tag{16}$$

$$2\beta_1\beta_2 M_2(1) + (\beta_2^2 + \sigma^2) t_x M_2(2) > 0. \tag{17}$$

□

8-ame skyrelyje empiriškai palyginamos dispersijos D_h^* reikšmės taškuose $h = h^*$, $h = h_0$, $h = 0$ ir $h = 1$.

5. Regresinis įvertinys

Atvejus $\beta_1 = 0$ ir $\beta_1 \neq 0$ nagrinėkime atskirai.

5.1. $\beta_1 = 0$ atvejis

Tarkime, kad $\beta_1 = 0$. Šiuo atveju regresinis įvertinys gali būti užrašytas pavidalu (žr. Särndal, Swensson ir Wretman (1997))

$$\hat{t}_{yr} = \hat{t}_{yHT} + \hat{B}(t_x - \hat{t}_{xHT}), \tag{18}$$

kur

$$\hat{t}_{xHT} = \sum_{k=1}^{N} \mathbb{I}_k x_k \pi_k^{-1}, \qquad \hat{B} = \Big(\sum_{k=1}^{N} \mathbb{I}_k \frac{x_k^2}{\sigma_k^2 \pi_k}\Big)^{-1} \sum_{k=1}^{N} \mathbb{I}_k \frac{x_k y_k}{\sigma_k^2 \pi_k}.$$

Šio įvertinio dispersijos išraiška yra ganėtinai sudėtinga, todėl paprastai nagrinėjama apytikslė dispersija

$$D_h^* = \sum_{k=1}^{N} (y_k - Bx_k)^2 \big(1 - \pi_k\big) \pi_k^{-1}, \quad \text{kur} \quad B = D^{-1} \sum_{k=1}^{N} \frac{x_k y_k}{\sigma_k^2}$$

ir $D = \sum_{k=1}^{N} \sigma_k^{-2} x_k^2$. Atlikę paprasčiausius skaičiavimus gauname, kad vidurkis $D_h = \mathbf{E} D_h^*$ gali būti užrašytas pavidalu

$$D_h = \sum_{k=1}^{N} \Big(\frac{1}{n} \frac{1}{p_k(h)} - 1\Big) (\sigma_k^2 - D^{-1} x_k^2). \tag{19}$$

Visiškai taip pat kaip ir HT įvertiniui parodoma, kad funkcijos $h \to D_h$ ir $h \to D_h^*$ yra iškilos.

Teiginys 3. *Sakykime, kad $\sigma_k^2 = \sigma^2 x_k^{\gamma}$, kur $\gamma \in [0,2]$. Tarkime, kad mažiausiai dvi iš tikimybių $\{p_i\}$ yra skirtingos. Tegu $\sigma^2 > 0$.*
(i) Sakykime, kad $\gamma = 0$. Tada $h_0 = 1$.
(ii) Sakykime, kad $\gamma \in (0,2)$. Tada $h_0 > 0$. Jei $\sigma^2 t_x^{\gamma-2} M_2(\gamma) > D^{-1} M_2(2)$, tai $0 < h_0 < 1$. Jei $\sigma^2 t_x^{\gamma-2} M_2(\gamma) \leq D^{-1} M_2(2)$, tai $h_0 = 1$.
(iii) Sakykime, kad $\gamma = 2$. Tada $h_0 = 0$.

Įrodymas. Įrodymo principas toks pats kaip ir Teiginio 2. Tvirtinimas (i) seka iš sąlygos $D_1' < 0$, kurią atitinka nelygybė $D^{-1} M_2(2) > 0$.
Tvirtinimas (ii) seka iš sąlygos $D_0' < 0$, kurią atitinka (20) ir sąlygos $D_1' > 0$, kurią atitinka (21). Čia

$$\sigma^2 t_x^{\gamma} M_1(\gamma) < 0, \tag{20}$$

$$\sigma^2 t_x^{\gamma-2} M_2(\gamma) - D^{-1} M_2(2) > 0. \tag{21}$$

Tvirtinimas (iii) seka iš to, kad $D_0' = 0$ ir iš sąlygos $D_1' > 0$, kurią atitinka nelygybė $(\sigma^2 - D^{-1}) M_2(2) > 0$. $\square$

5.2. $\beta_1 \neq 0$ atvejis

Tarkime, kad $\beta_1 \neq 0$. Tokiu atveju populiacijos dydis N gali būti traktuojamas kaip papildoma informacija ir tada mes turime regresinį įvertinį (žr. Särndal, Swensson ir Wretman (1997))

$$\hat{t}_{yr} = \hat{t}_{yHT} + \hat{B}_1(N - \hat{t}_{1HT}) + \hat{B}_2(t_x - \hat{t}_{xHT}). \tag{22}$$

Čia $\hat{t}_{1HT} = \sum_{i=1}^{N} \mathbb{I}_i \pi_i^{-1}$. Koeficientai

$$\begin{pmatrix} \hat{B}_1 \\ \hat{B}_2 \end{pmatrix} = \Big(\sum_{i=1}^{N} \mathbb{I}_i X_i X_i'/\sigma_i^2 \pi_i\Big)^{-1} \sum_{i=1}^{N} \mathbb{I}_i X_i y_i/\sigma_i^2 \pi_i,$$

kur $X_i = \binom{1}{x_i}$. Šio įvertinio dispersijos išraiška yra ganėtinai sudėtinga, tad nagrinėsime apytikslę dispersiją

$$D_h^* = \sum_{k=1}^{N} (\pi_k^{-1} - 1)(y_k - B_1 - x_k B_2)^2, \tag{23}$$

kur

$$\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \Big(\sum_{i=1}^{N} X_i X_i'/\sigma_i^2\Big)^{-1} \sum_{i=1}^{N} X_i y_i/\sigma_i^2.$$

Funkciją $D_h = \mathbf{E} D_h^*$ galima pertvarkyti į pavidalą

$$D_h = \sum_{k=1}^{N} \Big(\frac{1}{n}\frac{1}{p_k(h)} - 1\Big)\Big(\sigma_k^2 - \frac{1}{W}(D - 2Gx_k + Hx_k^2)\Big), \tag{24}$$

kur pažymėjome

$$D = \sum_{k=1}^{N} \frac{x_k^2}{\sigma_k^2}, \qquad G = \sum_{k=1}^{N} \frac{x_k}{\sigma_k^2}, \qquad H = \sum_{k=1}^{N} \frac{1}{\sigma_k^2}, \qquad W = DH - G^2.$$

Aišku, kad ir dabar nagrinėjamam įvertiniui funkcijos $h \to D_h$ ir $h \to D_h^*$ yra iškilos.

Turima funkcijos D_h išraiška tolimesnei analizei yra per daug sudėtinga, todėl imkime jai apytikslę (žr. Särndal, Swensson ir Wretman (1997))

$$D_h \simeq \sum_{k=1}^{N} \Big(\frac{1}{n}\frac{1}{p_k(h)} - 1\Big)\sigma_k^2. \tag{25}$$

Tokia funkcija taip pat yra iškila.

Teiginys 4. *Sakykime, kad* $\sigma_k^2 = \sigma^2 x_k^\gamma$*, kur* $\gamma \in [0{,}2]$*. Tarkime, kad mažiausiai dvi iš tikimybių* $\{p_i\}$ *yra skirtingos. Tarkime, kad funkcija* D_h *nesiskiria nuo tos, kuri yra apytikslės lygybės (25) dešinėje. Tegu* $\sigma^2 > 0$*.*
(i) Sakykime, kad $\gamma = 0$*. Tada* $h_0 = 1$*.*
(ii) Sakykime, kad $\gamma \in (0{,}2)$*. Tada* $0 < h_0 < 1$*.*
(iii) Sakykime, kad $\gamma = 2$*. Tada* $h_0 = 0$*.*

Įrodymas. Žinome, kad funkcija D_h yra iškila. Tvirtinimas (i) seka iš to, kad $D_1' = 0$ ir iš sąlygos $D_0' < 0$, kurią atitinka nelygybė $\sigma^2 M_1(0) < 0$.
Tvirtinimas (ii) seka iš sąlygos $D_0' < 0$, kurią atitinka nelygybė $\sigma^2 t_x^\gamma M_1(\gamma) < 0$ ir iš sąlygos $D_1' > 0$, kurią atitinka nelygybė $\sigma^2 t_x^\gamma M_2(\gamma) > 0$.
Tvirtinimas (iii) seka iš to, kad $D_0' = 0$ ir iš sąlygos $D_1' > 0$, kurią atitinka nelygybė $\sigma^2 t_x^2 M_2(2) > 0$. □

6. Apytikslis skirtumo $h^* - h_0$ skirstinys

Funkcijos $h \to D_h^*$ išvestinę pažymėkime d_h^*. Iš funkcijos $h \to D_h^*$ iškilumo seka, kad įvykiai $\{h^* \le x\}$ ir $\{d_x^* \ge 0\}$ yra tapatūs. Todėl

$$\mathbf{P}\{h^* \le x\} = \mathbf{P}\{d_x^* \ge 0\}. \tag{26}$$

Užrašykime funkciją d_x^* pavidalu

$$d_x^* = \frac{1}{n} \sum_{k=1}^{N} \frac{p_k - N^{-1}}{p_k^2(x)} Z_k, \tag{27}$$

kur Z_k yra nepriklausomi atsitiktiniai dydžiai. HT įvertiniui (11) turime, kad $Z_k = Y_k^2$, o regresiniams įvertiniams (18) ir (22) turime atitinkamai $Z_k = (Y_k - x_k B)^2$ ir $Z_k = (Y_k - B_1 - x_k B_2)^2$.

Trumpumo dėlei pažymėkime $v_k(x) = n^{-1}(p_k - N^{-1}) p_k^{-2}(x)$. Kadangi $\mathbf{E} d_x^* = D_x'$, tai, naudodamiesi lygybėmis (26) ir (27), gauname, kad

$$\mathbf{P}\{h^* \le x\} = \mathbf{P}\{\sum_{k=1}^{N} v_k(x)(Z_k - \mathbf{E} Z_k) \ge -D_x'\}.$$

Atsitiktinio dydžio d_x^* dispersiją pažymėkime $V_x^2 = \sum_{k=1}^N v_k^2(x) \mathbf{V}(Z_k)$. Pasinaudoję skleidiniais $D_{h_0+u}' = u D''(h_0) + o(u^2)$ ir $V_{h_0+u} = V_{h_0} + o(u)$, kai čia $u \to 0$, dideliems N turime normaliąją skirtumo $h^* - h_0$ skirstinio aproksimaciją

$$\mathbf{P}\{h^* \le h_0 + u\} \approx \Phi\Big(\frac{D_{h_0+u}'}{V_{h_0+u}}\Big) \approx \Phi\Big(u \frac{D_{h_0}''}{V_{h_0}}\Big).$$

7. Išreikštinės dydžio h^* aproksimacijos

Optimalios priklausymo imčiai tikimybės, t.y. tikimybės minimizuojančios įvertinio dispersiją, gali būti ieškomos ir kitokiais būdais. HT įvertiniui, atsisakę lygybėje (13) priklausomybės nuo parametro h, pažymėkime

$$D_\pi = \sum_{k=1}^{N} \Big(\frac{1}{\pi_k} - 1\Big) a_k^2, \tag{28}$$

kur $a_k^2 = \beta_1^2 + 2\beta_1\beta_2 x_k + \beta_2^2 x_k^2 + \sigma_k^2$. Kadangi, naudojantis Cauchy-Schwartz'o nelygybe, yra teisinga

$$\Big(\sum_{k=1}^{N} a_k\Big)^2 \le \Big(\sum_{k=1}^{N} \frac{a_k^2}{\pi_k}\Big)\Big(\sum_{k=1}^{N} \pi_k\Big),$$

tai turime, kad

$$D_\pi = \sum_{k=1}^{N} \frac{a_k^2}{\pi_k} - \sum_{k=1}^{N} a_k^2 \ge \frac{\Big(\sum_{k=1}^{N} a_k\Big)^2}{\sum_{k=1}^{N} \pi_k} - \sum_{k=1}^{N} a_k^2.$$

Šioje nelygybėje lygybė yra pasiekiama tada ir tik tada, kai $\pi_i \approx c a_i$ kiekvienam $i = 1, \ldots, N$. Tada, pasinaudoję sąlygos (9) analogu $\sum_{k=1}^{N} \pi_k = n$, turime, kad optimalios priklausymo imčiai tikimybės yra

$$\pi_i = \pi_i(a) = \frac{n a_i}{\sum_{i=1}^{N} a_i}, \qquad 1 \le i \le N, \tag{29}$$

kai čia a tegu reiškia vektorių $(a_1, \ldots, a_N)$.

Apibendrintam regresiniam įvertiniui (GREG) yra parodyta (žr. Särndal, Swensson ir Wretman (1997)), kad optimalios priklausymo imčiai tikimybės yra

$$\pi_i = \pi_i(\sigma) = \frac{n \sigma_i}{\sum_{i=1}^{N} \sigma_i}, \qquad 1 \le i \le N, \tag{30}$$

kai čia σ tegu reiškia vektorių $(\sigma_1, \ldots, \sigma_N)$.

Kadangi kiekviena iš tikimybių (29) ar (30) tiesiogiai priklauso nuo ją atitinkančio populiacijos elemento, tai tokių ėmimo tikimybių konstravimas pareikalautų ganėtinai tiksliai įvertinti kiekvieną a_i ar σ_i. Tuo tarpu, šiame darbe nagrinėjamos tikimybės (8) tiesiogiai priklauso tik nuo vieno parametro.

Funkcijų (13), (19) ir (24) minimizavimo uždavinį paprasčiausia spręsti skaitiniais metodais. Visgi, pasinaudodami tikimybėmis (29) ir (30), pasiūlysime išreikštines atsitiktinio dydžio h^* aproksimacijas HT ir regresiniam įvertiniams, neatskirdami abiejų nagrinėtų regresinio įvertinio atvejų. Pažymėkime atsitiktinio dydžio h^* artinius HT ir regresiniam įvertiniams atitinkamai h_{HT} ir h_R.

HT įvertiniui, esant pakankamai dideliems N, egzistuoja bent vienas populiacijos $\mathcal{U}$ elementas, sakykime, u_i, kuriam jo patekimo į imtį tikimybės $\pi_i(h_{HT})$ ir $\pi_i(a)$ yra gana artimos, t.y. $\pi_i(h_{HT}) \approx \pi_i(a)$. Iš šios apytikslės lygybės išsprendę parametrą h_{HT} turime

$$h_{HT} \approx \frac{\frac{a_i}{\sum_{i=1}^N a_i} - p_i}{\frac{1}{N} - p_i}, \qquad \text{kažkuriam indeksui} \qquad i. \tag{31}$$

Funkciją $f(\sigma_i) = a_i = \sqrt{(\beta_1 + \beta_2 x_i)^2 + \sigma_i^2}$ skleisdami MacLoren'o eilute ir imdami tris tokio skleidinio narius turime

$$f(\sigma_i) \approx \beta_1 + \beta_2 x_i + \frac{\sigma_i^2}{2(\beta_1 + \beta_2 x_i)} = \beta_1 + \beta_2 x_i + \frac{cv(y_i)}{2}\sigma_i, \tag{32}$$

čia $cv(y_i)$ žymi elemento u_i kitimo koeficientą. Sakykime, kad $cv(y_i) \approx cv(y)$, kur $cv(y)$ yra visos populiacijos $\mathcal{U}$ kitimo koeficientas. Naudodamiesi pastarąja prielaida, įrašę apytikslę lygybę (32) į (31), bei atlikę paprastus skaičiavimus gauname, kad

$$h_{HT} \approx \frac{\beta_1 + \frac{cv(y)}{2}\xi_i}{\beta_1 + \beta_2\mu_x + \frac{cv(y)}{2}\mu_\sigma}, \qquad \text{kažkuriam indeksui} \qquad i, \tag{33}$$

kai čia $\mu_x = t_x/N$, $\mu_\sigma = \frac{1}{N}\sum_{i=1}^N \sigma_i$ ir

$$\xi_i = \frac{\mu_x\sigma_i - \mu_\sigma x_i}{\mu_x - x_i}.$$

Sakykime, kad $\sigma_k^2 = \sigma^2 x_k^\gamma$, kur $\gamma \in [0{,}2]$. Tada, kai $\gamma = 0$, turime $\xi_i = \sigma$ o, kai $\gamma = 2$, turime $\xi_i = 0$. Naudodamiesi šiais kraštiniais atvejais bei remdamiesi empiriniais stebėjimais, tiesiog tariame, kad $\xi_i \approx (1 - \frac{\gamma}{2})\sigma$. Tokiu būdu gauname, kad

$$h_{HT} \approx \frac{\beta_1 + \frac{cv(y)}{2}(1 - \frac{\gamma}{2})\sigma}{\beta_1 + \beta_2\mu_x + \frac{cv(y)}{2}\mu_\sigma}. \tag{34}$$

Panašiai elgiamės ir regresinio įvertinio atveju. Iš apytikslės kuriam nors elementui $u_i \in \mathcal{U}$ lygybės $\pi_i(h_R) \approx \pi_i(\sigma)$ iš karto išsprendžiame

$$h_R \approx \frac{\xi_i}{\mu_\sigma}, \qquad \text{kažkuriam indeksui} \qquad i. \tag{35}$$

Tarę, kad $\sigma_k^2 = \sigma^2 x_k^\gamma$, kur $\gamma \in [0{,}2]$, ir tada, manydami, kad $\xi_i \approx (1-\frac{\gamma}{2})\sigma$, turime

$$h_R \approx \frac{(1-\frac{\gamma}{2})\sigma}{\mu_\sigma}. \tag{36}$$

Gautų atsitiktinio dydžio h^* artinių h_{HT} ir h_R tikslumą patikrinsime sekančiame skyrelyje.

8. Kompiuterinis modeliavimas

Fiksuokime populiacijos dydį $N = 1000$ ir imties dydį $n = 100$. Imkime tris skirtingus papildomos informacijos vektorius $\tilde{x} = (x_1, \ldots, x_N)$.
1. Pirmojo vektoriaus $\tilde{x}_E$ koordinatės yra

$$x_i = \Big| \ln(1 - \frac{i-0.5}{N}) \Big|, \qquad 1 \le i \le N. \tag{37}$$

2. Antrojo vektoriaus $\tilde{x}_L$ koordinatės yra

$$x_i = i^{0.25}, \qquad 1 \le i \le N. \tag{38}$$

3. Trečiojo vektoriaus $\tilde{x}_G$ koordinatės yra

$$x_i = (1.002)^i, \qquad 1 \le i \le N. \tag{39}$$

Pastebėsime, kad papildomos informacijos vektoriai (37), (38) ir (39) tenkina (10) sąlygą.

Turėdami papildomos informacijos vektorių $\tilde{x}$, imkime populiacijos modelį $y_i = 2 + x_i + \sigma_i \eta_i$, kur $\sigma_i^2 = \sigma^2 x_i^\gamma$, $1 \le i \le N$. Domimės atvejais, kai čia $\gamma \in \{0; 0.5; 1; 1.5; 2\}$, o σ yra parenkamas taip, kad koreliacijos koeficientas tarp kintamųjų x ir y yra kuo labiau artimas 0.9. Laikome, kad $\eta_1, \eta_2, \ldots$ yra nepriklausomų standartinių normaliųjų atsitiktinių dydžių seka.
Lentelėse T1.1-T1.3 yra pateikiami HT įvertinio dispersijos (12) empirinio tyrinėjimo rezultatai.
Lentelėse T2.1-T2.3 yra pateikiami regresinio įvertinio dispersijos (23) empirinio tyrinėjimo rezultatai.
Lentelėms įvedėme tokius pažymėjimus: E_1-E_4 atitinkamai reiškia santykių $\frac{D^*_{h_0}}{D^*_0}$, $\frac{D^*_{h_0}}{D^*_1}$, $\frac{D^*_{h^*}}{D^*_{h_0}}$, $\frac{D^*_{h^*}}{D^*_{h_{HT}}}$ (arba $\frac{D^*_{h^*}}{D^*_{h_R}}$) vidurkius, o V_1-V_4 - jų dispersijas. Dydžiai E_1-E_4 ir V_1-V_4, bei $\mathbf{E}(h^* - h_0)^2$ buvo įvertinti pasigaminus 50 nepriklausomų reikiamos populiacijos kopijų, o dydis $cv(y)$ buvo įvertintas imant pirmą pasitaikiusią populiacijos kopiją.

Lentelė T1.1 HT įvertinys, $x = \tilde{x}_E$

γ	h_0	$\frac{D_{h_0}}{D_0}$	$\frac{D_{h_0}}{D_1}$	E_1	E_2	E_3	E_4
0.0	0.675	0.2106	0.8940	0.2112	0.8937	0.9999	0.9999
0.5	0.668	0.2178	0.8905	0.2176	0.8895	0.9999	0.9999
1.0	0.663	0.2185	0.8877	0.2189	0.8889	0.9999	0.9999
1.5	0.660	0.2183	0.8856	0.2182	0.8865	0.9999	0.9999
2.0	0.658	0.2182	0.8836	0.2184	0.8836	0.9999	0.9999
	$\mathbf{E}(h^* - h_0)^2$	$cv(y)$	h_{HT}	V_1	V_2	V_3	V_4
0.0	3.37E-05	0.372	0.676	4.29E-04	8.33E-06	1.42E-09	1.60E-09
0.5	2.93E-05	0.361	0.671	2.73E-05	1.31E-05	1.27E-09	2.75E-09
1.0	3.07E-05	0.363	0.664	3.49E-06	2.02E-05	1.19E-09	1.10E-09
1.5	4.96E-05	0.356	0.658	2.91E-06	3.03E-05	3.56E-09	4.63E-09
2.0	6.22E-05	0.390	0.652	2.70E-06	4.76E-05	8.00E-09	1.61E-08

Lentelė T1.2 HT įvertinys, $x = \tilde{x}_L$

γ	h_0	$\frac{D_{h_0}}{D_0}$	$\frac{D_{h_0}}{D_1}$	E_1	E_2	E_3	E_4
0.0	0.311	0.9931	0.9787	0.9932	0.9787	0.9999	0.9999
0.5	0.310	0.9932	0.9786	0.9933	0.9786	0.9999	0.9999
1.0	0.308	0.9933	0.9785	0.9933	0.9785	0.9999	0.9999
1.5	0.307	0.9933	0.9784	0.9932	0.9786	0.9999	0.9999
2.0	0.306	0.9934	0.9784	0.9934	0.9784	0.9999	0.9999
	$\mathbf{E}(h^* - h_0)^2$	$cv(y)$	h_{HT}	V_1	V_2	V_3	V_4
0.0	9.65E-05	0.156	0.311	2.31E-07	3.12E-07	7.37E-11	7.50E-11
0.5	1.35E-04	0.154	0.309	2.84E-07	4.91E-07	6.86E-11	6.86E-11
1.0	6.94E-05	0.152	0.307	1.29E-07	2.90E-07	3.88E-11	4.00E-11
1.5	6.82E-05	0.155	0.306	1.04E-07	2.95E-07	3.67E-11	3.97E-11
2.0	7.19E-05	0.155	0.306	1.11E-07	3.47E-07	5.15E-11	5.12E-11

Lentelė T1.3 HT įvertinys, $x = \tilde{x}_G$

γ	h_0	$\frac{D_{h_0}}{D_0}$	$\frac{D_{h_0}}{D_1}$	E_1	E_2	E_3	E_4
0.0	0.403	0.9350	0.8883	0.9350	0.8882	0.9999	0.9999
0.5	0.394	0.9377	0.8858	0.9376	0.8860	0.9999	0.9999
1.0	0.387	0.9398	0.8835	0.9400	0.8833	0.9999	0.9999
1.5	0.382	0.9413	0.8815	0.9408	0.8822	0.9999	0.9999
2.0	0.378	0.9424	0.8800	0.9421	0.8803	0.9999	0.9999
	$\mathbf{E}(h^* - h_0)^2$	$cv(y)$	h_{HT}	V_1	V_2	V_3	V_4
0.0	8.81E-05	0.380	0.404	9.28E-06	9.64E-06	1.62E-09	1.65E-09
0.5	7.49E-05	0.384	0.391	7.02E-06	8.83E-06	1.13E-09	1.81E-09
1.0	5.88E-05	0.380	0.382	4.75E-06	8.63E-06	7.91E-10	1.08E-09
1.5	6.72E-05	0.379	0.377	4.64E-06	1.02E-05	8.33E-10	1.61E-09
2.0	7.68E-05	0.384	0.375	5.04E-06	1.36E-05	1.89E-09	2.71E-09

Lentelė T2.1 Regresinis įvertinys, $x = \tilde{x}_E$

γ	h_0	$\frac{D_{h_0}}{D_0}$	$\frac{D_{h_0}}{D_1}$	E_1	E_2	E_3	E_4
0.0	1.000	0.1099	1.0000	0.1227	1.0000	0.9998	0.9998
0.5	0.716	0.4496	0.9362	0.4460	0.9345	0.9994	0.9903
1.0	0.419	0.7831	0.7832	0.7814	0.7857	0.9995	0.9866
1.5	0.161	0.9587	0.6045	0.9582	0.6048	0.9994	0.9895
2.0	0.000	1.0000	0.4458	1.0000	0.4434	0.9999	0.9999
	$\mathbf{E}(h^* - h_0)^2$	$cv(y)$	h_R	V_1	V_2	V_3	V_4
0.0	1.98E-04	0.372	1.000	1.18E-03	4.09E-31	2.21E-07	2.21E-07
0.5	8.33E-04	0.361	0.827	1.06E-03	2.05E-04	3.96E-07	2.36E-05
1.0	8.33E-04	0.363	0.564	4.47E-04	4.86E-04	5.44E-07	2.73E-05
1.5	6.41E-04	0.356	0.272	8.58E-05	1.01E-03	6.45E-07	2.47E-05
2.0	1.65E-05	0.390	0.000	2.09E-31	6.04E-04	2.65E-08	2.65E-08

Lentelė T2.2 Regresinis įvertinys, $x = \tilde{x}_L$

γ	h_0	$\frac{D_{h_0}}{D_0}$	$\frac{D_{h_0}}{D_1}$	E_1	E_2	E_3	E_4
0.0	0.997	0.9340	0.9999	0.9352	0.9999	0.9997	0.9997
0.5	0.710	0.9668	0.9962	0.9655	0.9966	0.9995	0.9974
1.0	0.446	0.9869	0.9865	0.9861	0.9871	0.9993	0.9965
1.5	0.206	0.9971	0.9725	0.9968	0.9736	0.9995	0.9983
2.0	0.000	1.0000	0.9558	1.0000	0.9550	0.9999	0.9999
	$\mathbf{E}(h^* - h_0)^2$	$cv(y)$	h_R	V_1	V_2	V_3	V_4
0.0	5.35E-03	0.156	1.000	1.93E-04	8.16E-10	3.34E-07	3.56E-07
0.5	9.67E-03	0.154	0.517	8.22E-05	6.78E-06	4.35E-07	5.39E-06
1.0	1.28E-02	0.152	0.237	4.39E-05	3.21E-05	1.10E-06	9.95E-06
1.5	8.50E-03	0.155	0.081	5.56E-06	4.79E-05	3.42E-07	2.82E-06
2.0	1.47E-03	0.155	0.000	2.08E-31	4.42E-05	6.50E-08	6.51E-08

Lentelė T2.3 Regresinis įvertinys, $x = \tilde{x}_G$

γ	h_0	$\frac{D_{h_0}}{D_0}$	$\frac{D_{h_0}}{D_1}$	E_1	E_2	E_3	E_4
0.0	1.000	0.7029	1.0000	0.7032	1.0000	0.9998	0.9998
0.5	0.742	0.8191	0.9783	0.8170	0.9784	0.9995	0.9909
1.0	0.477	0.9172	0.9174	0.9159	0.9188	0.9993	0.9875
1.5	0.224	0.9799	0.8309	0.9791	0.8326	0.9994	0.9939
2.0	0.000	1.0000	0.7355	1.0000	0.7342	0.9998	0.9998
	$\mathbf{E}(h^* - h_0)^2$	$cv(y)$	h_R	V_1	V_2	V_3	V_4
0.0	6.05E-04	0.380	1.000	3.17E-04	1.93E-31	2.09E-07	2.09E-07
0.5	1.62E-03	0.384	0.578	2.76E-04	4.77E-05	3.40E-07	1.81E-05
1.0	2.10E-03	0.380	0.291	2.16E-04	1.92E-04	8.90E-07	3.23E-05
1.5	1.85E-03	0.379	0.108	5.15E-05	3.82E-04	7.48E-07	1.58E-05
2.0	3.63E-04	0.384	0.000	2.18E-31	2.37E-04	1.67E-07	1.67E-07

9. Išvados

Matėme, kad dažnai į priklausymo imčiai tikimybes yra prasminga įtraukti visiems populiacijos elementams vienodą ėmimo komponentę. Kai turime papildomos informacijos apie tiriamą populiaciją, o taip paprastai būna, kai tyrimas yra periodinis, tai, prieš rinkdami imtį, naudodamiesi teiginiais 2-4, galime nustatyti, ar verta priklausymo imčiai tikimybes konstruoti šiame darbe nagrinėjamu būdu. Tam tereikia įvertinti populiacijos parametrus β_1, β_2, σ ir γ.

Darbe nagrinėjamas Poisson'o ėmimas yra retas šių dienų praktikoje, tačiau tikėtina, kad panašūs į nustatytuosius dėsningumai galios ir sudėtingesniems ėmimo planams. Nagrinėto imties išrinkimo metodo trūkumu galima laikyti tai, kad turime iš anksto nuspręsti, kokį įvertinį (HT ar regresinį) naudosime vertinimo etape.

Kompiuterinio modeliavimo pavyzdžiai rodo, kad parametras h_0 gana tiksliai įvertina iš anksto nežinomą h^*. Be to, matome, kad reikšmės $D^*_{h_0}$ ir $D^*_{h^*}$ yra labai artimos. Todėl beveik nieko neprarandame priklausymo imčiai tikimybes (3) imdami su $h = h_0$. Net ir tuo atveju, kai vietoje skaitiniais metodais rastojo h_0 naudojame h_{HT} (arba h_R), matome, kad net ir šiems dydžiams, kurie, lyginant su h_0, nebūtinai yra tikslūs (tai pasakytina apie h_R), dispersijos $D^*_{h_{HT}}$ (arba $D^*_{h_R}$) ir $D^*_{h^*}$ yra kur kas labiau artimos nei $D^*_{h^*}$ ir D^*_0, ir $D^*_{h^*}$ ir D^*_1. Taip yra, nes taško h^* aplinkoje funkcija $h \to D^*_h$ vidutiniškai kinta lėtai.

Literatūra

1. C.E. Särndal, B. Swensson, J. Wretman, *Model assisted survey sampling*, (Springer series in Statistics) Springer-Verlag Berlin, Heidelberg, New York (1997).

2. H. Kröger, C.E. Särndal, I. Teikari, Poisson mixture sampling combined with order sampling, *Journal of Official Statistics*, 19, 59–70 (2003).

3. M. Bloznelis, A. Čiginas, On Variance minimization for unequal probability sampling, *VU MIF preprintas 2005–11* (2005).

Santrauka

Darbe nagrinėjame ėmimo planus, kurių priklausymo imčiai tikimybės yra dviejų komponenčių mišiniai. Pirmoji komponentė yra proporcinga papildoma informacija nusakytam populiacijos elemento dydžiui, o antroji yra vienoda visiems elementams. Ieškome tokių mišinių, kurie minimizuoja įvairių populiacijos sumos įvertinių dispersijas ir parodome kaip, naudojantis papildoma informacija, apytiksliai nustatyti optimalų mišinį. Pateikiame teorinius ir kompiuterinio modeliavimo rezultatus Poisson'o imtims, renkamoms iš populiacijų, kurios yra generuotos naudojant tiesinės regresijos modelį.

Summary

We consider sampling designs, where inclusion (to sample) probabilities are mixtures of two components. The first component is proportional to the size of a population unit (described by means of an auxiliary information available). The second component is the same for every unit. We look for mixtures that minimize variances of various estimators of the population total and show how auxiliary information could help to find an approximate location of such mixtures. We report theoretical and simulation results in the case of Poisson samples drawn from populations which are generated by a linear regression model.