ĮVADAS

Sprendžiant bet kokį matematinės statistikos uždavinį, svarbu yra tinkamai parinkti statistinį modelį, t.y. tikimybinių skirstinių šeimą, kurios rėmuose šis uždavinys sprendžiamas. Netinkamai parinkus statistinį modelį, galima gauti klaidingas išvadas. Statistinio modelio parinkimą palengvina profesinės žinios apie stebimą atsitiktinį dydį (toliau šiame darbe - a.d.) ir jo skirstinio charakteringas savybes, o taip pat vizualus empirinių ir teorinių charakteristikų grafikų palyginimas. Abejotinais atvejais taikomi statistiniai turimų duomenų ir bandomų teorinių modelių suderinamumo hipotezių tikrinimo kriteriai.

Vienas iš universalių, nepriklausančių nuo skirstinio pavidalo tokio tipo kriterijų yra χ^2 kriterijus suderinamumo hipoezėms tikrinti. Jis tampa nepriklausomas nuo stebimojo a.d. skirstinio, perėjus prie polinominio skirstinio: kriterijaus statistika priklauso tik nuo imties rezultatų patekimo į tam tikras aibes dažnių. Apie χ^2 kriterijaus svarbą liudija kad ir tai, jog 2001 metais Rusijoje įvestas valstybinis standartas „Empirinio skirstinio suderinamumo su teoriniu tikrinimo taisyklės. χ^2 tipo kriterijai“ (P 50.1.033 - 2001), kuris reglamentuoja χ^2 kriterijų taikymą.

Perėjus prie polinominio skirstinio suderinamumo hipotezė keičiama hipoteze apie gautojo polinominio skirstinio tikimybių vektoriaus lygybę duotam vektoriui (paprastoji suderinamumo hipotezė), arba apie galimybę tikimybių vektorių išreikšti duoto pavidalo funkcijomis nuo tam tikrų parametrų (sudėtinė suderinamumo hipotezė). Kriterijus grindžiamas tuo, kad tam tikra kvadratinė polinominio atsitiktinio vektoriaus forma X_n^2 asimptotiškai (augant imties didumui n) artėja į χ^2 skirstinį. Taigi χ^2 tipo kriterijai yra apytiksliai. Jų tikslumas priklauso nuo imties didumo n ir nuo statistikos X_n^2 konvergavimo į ribinį dėsnį greičio.

Yra daug mokslinių straipsnių ir knygų, kuriose nagrinėjama χ^2 tipo kriterijų savybės ir jų taikymo aspektai. Tiriama šių kriterijų tikslumo priklausomybė nuo imties didumo n, nuo stebimo dydžio reikšmių aibės padalijimo į intervalus taisyklių ir tokių intervalų skaičiaus, nuo parametrų vertinimo metodų ir kt. Susisteminta ir gana išsami bibliografija šiais klausimais yra pateikta anksčiau minėtame standarte. Literatūros sąrašas iš šio standarto pateikiamas 1 priede.

Tačiau nežinomi darbai, kuriuose χ^2 statistika būtų modifikuojama taip, kad aproksimacija būtų tikslesnė. Tada ir χ^2 kriterijus suderinamumo hipotezėms tikrinti būtų tikslesnis.

Šiame darbe daromas bandymas modifikuoti χ^2 statistiką taip, kad modifikuotos statistikos skirstinys būtų tiksliau aproksimuojamas ir remiantis tuo, sukonstruoti modifikuotą χ^2 kriterijų. Nors praktiškai svarbesnis yra sudėtinių suderinamumo hipotezių atvejis, tačiau šiame etape apsiribojame tik paprastosiomis suderinamumo hipotezėmis. Jeigu pateikiamas pasiūlymas pasiteisins, tai jį reikėtų pritaikyti su atitinkamais pakeitimais ir sudėtinių suderinamumo hipotezių atveju. Tikrinant paprastąją suderinamumo hipotezę, statistikos X_n^2 skirstinys aproksimuojamas χ^2 skirstiniu su k-1 laisvės laipsnių; čia k – intervalų, į kuriuos suskirstytos stebimo a.d. reikšmės, skaičius. Suderinamumo hipotezė atmetama α lygmens apytiksliu χ^2 kriterijumi, kai $X_n^2 > X_\alpha^2\left(k-1\right)$.

Siūlomas kriterijaus patikslinimo būdas, kurio esmė yra tokia: gauname statistikos X_n^2 trijų pirmųjų momentų išraiškas; atliekame transformaciją $Y = a + bX_n^2$ ir a.d. Y skirstinį aproksimuojame χ^2 skirstiniu su ω laisvės laipsnių; parametrai a, b, ω parenkami taip, kad statistikos Y vidurkis dispersija ir trečias centrinis momentas sutaptų su a.d. χ_ω^2 atitinkamais momentais, t.y. būtų lygūs ω, 2ω, 8ω. Modifikuotas α lygmens apytikslis kriterijus atmeta paprastąją suderinamumo hipotezę, kai $Y > \chi_\alpha^2\left(\omega\right)$.

Darbo tikslai:

- Pagal nurodytą literatūrą ([1], [2], [3]) išnagrinėti χ^2 tipo kriterijų konstravimo teorinius ir taikomuosius aspektus;
- Aptarti empirinių charakteristikų grafinio vaizdavimo metodus ir šių grafikų vizualinį palyginimą su atitinkamomis teorinėmis charakteristikomis; kartu aptarti intervalų, į kuriuos sugrupuojamos stebimojo atsitiktinio dydžio reikšmės, pavidalą ir skaičių;
- Gauti statistikos X_n^2 vidurkio, dispersijos ir trečiojo centrinio momento išraiškas; patyrinėti momentų savybes remiantis gautomis išraiškomis;
- Sukurti makroprogramą SAS paketo rėmuose pirmiesiems trims statistikos X_n^2 momentams skaičiuoti;
- Sukurti makroprogramą SAS paketo rėmuose a.d. X_n^2 reikšmėms modeliuoti; sukonstruoti kriterijų suderinamumo hipotezei tikrinti remiantis modeliuotomis reikšmėmis;
- Sukonstruoti modifikuotą apytikslį kriterijų, grindžiamą statistika V;
- Naudojant statistikos X_n^2 modeliuotas reikšmes, atlikti kriterijų, grindžiamų statistikomis X_n^2 ir V, palyginimą, kai tikrinamoji hipotezė teisinga ir kai ji neteisinga.

Diplominio darbo medžiaga suskirstyta į 5 skyrelius. Pirmajame aprašomas polinominis skirstinys bei jo pagrindu sudarytas χ^2 kriterijus suderinamumo hipotezėms tikrinti. Antrajame skyrelyje aptariamos suderinamumo hipotezių tikrinimo taisyklės, naudojant χ^2 kriterijų, bei optimalaus intervalų, į kuriuos suskaidomi turimi duomenys, skaičiaus parinkimo klausimas. Trečiajame skyrelyje apskaičiuojami pirmieji trys statistikos X_n^2 momentai, atliekama jų elgsenos analizė priklausomai nuo imties dydžio n ir įvedama statistikos X_n^2 korekcija, sutapatinant pirmuosius tris momentus su teorinėmis jų reikšmėmis. Ketvirtame skyrelyje pasiūlomas dar vienas kriterijus suderinamumo hipotezėms tikrinti. Jis remiasi SAS paketu modeliuotomis a.d. X_n^2 reikšmėmis ir nereikalauja sudarinėti statistikos bei skaičiuoti papildomų parametrų. Galiausiai penktajame skyrelyje atliekamas trijų pasiūlytų kriterijų tikslumo palyginimas. Yra lyginamas kriterijų reikšmingumo lygmuo ir kriterijaus galia, kai tikrinamoji hipotezė neteisinga.

1. Polinominis skirstinys ir χ^2 kriterijus

1.1. Polinominis skirstinys

Nagrinėkime nepriklausomų eksperimentų schemą, tarę, kad kiekviename bandyme gali įvykti nesutaikomi įvykiai A_1, ..., A_k, kurių skaičius k>2. Įvykio A_i tikimybė kiekviename bandyme pastovi ir lygi p_i, be to $p_1+...+p_k=1$ ir bandymai yra nepriklausomi. Tiksliau, su bet kuriuo natūraliuoju n ir i_j=1, ..., k, j=1, ..., n, įvykio $A_{i_1}\bigcap...\bigcap A_{i_n}$ tikimybė lygi įvykių A_{i_1},..., A_{i_n} tikimybių sandaugai:

$$\boldsymbol{P}\left\{A_{i_1}\bigcap...\bigcap A_{i_n}\right\}=\boldsymbol{P}\left\{A_{i_1}\right\}\cdot...\cdot\boldsymbol{P}\left\{A_{i_n}\right\}.$$

Tarkime, kad atlikta n tokių bandymų ir U_i žymi įvykio A_i, i=1, 2,...,k, įvykimų skaičių. Tada atsitiktinio vektoriaus (toliau žymėsime a.v.) $\boldsymbol{U}=(U_1,U_2,...,U_k)^T$ skirstinys vadinamas k-mačiu polinominiu su parametrais n ir $\boldsymbol{p}=\left(p_1,...,p_k\right)^T$. Sutrumpintai žymėsime $\boldsymbol{X}\sim P_k\ (n,\boldsymbol{p})$:

$$\boldsymbol{P}\left\{U_1=m_1,...,U_k=m_k\left|n,\boldsymbol{p}\right.\right\}=\frac{n!}{m_1!,...,m_k!}p_1^{m_1}\cdot...\cdot p_k^{m_k}; \tag{1.1}$$

čia $0\le m_j\le n$, j=1,…,k, $m_1+...+m_k=n$.

Generuojančiosios funkcijos daugiamatis analogas yra

$$\psi\left(\boldsymbol{s}\right)=\psi\left(s_1,...,s_k\right)=E\left(s_1^{U_1}\cdot...\cdot s_k^{U_k}\right)=\left(p_1s_1+...+p_ks_k\right)^n. \tag{1.2}$$

Matome, kad a.v. $\boldsymbol{U}=\boldsymbol{U}^{(1)}+...+\boldsymbol{U}^{(n)}$ yra suma vienodai pasiskirsčiusių nepriklausomų a.v.; čia $\boldsymbol{U}^{(i)}\sim P_k\left(1,\boldsymbol{p}\right)$, t.y. atsitiktinis vektorius, įgyjantis reikšmes $(1,0,...,0)^T$, $(0,1,...,0)^T$,..., $(0,0,...,1)^T$ atitinkamai su tikimybėmis $p_1,...,p_k$.

Akivaizdu, kad koordinatės U_i arba kelių koordinačių sumos marginalieji skirstiniai yra binominiai. A.v., sudaryto iš r<k atsitiktinio vektoriaus $\boldsymbol{U}$ koordinačių, skirstinys yra r-matis polinominis. Jeigu kelios a.v. $\boldsymbol{U}$ koordinatės yra fiksuotos, tai likusių koordinačių skirstinys taip pat polinominis.

Atsitiktinio dydžio U_i vidurkis ir dispersija atitinkamai yra

$$\boldsymbol{E}\boldsymbol{U}_i=np_i, \qquad \boldsymbol{V}\boldsymbol{U}_i=np_i\left(1-p_i\right),\ i=1,...,k,$$

o a.d. kovariaciją galime rasti, pavyzdžiui, iš (1.2) formulės:

$$\boldsymbol{Cov}\left(U_i,U_j\right)=\psi''_{s_i,s_j}\left(1,...,1\right)-\psi'_{s_i}\left(1,...,1\right)\psi'_{s_j}\left(1,...,1\right)=-np_ip_j;$$

čia $i\neq j$, $i,j=1,...,k$.

Taigi atsitiktinio vektoriaus $\boldsymbol{U}$ kovariacijų matrica $\boldsymbol{\Sigma}$ yra tokio pavidalo:

$$\boldsymbol{V}(\boldsymbol{U})=\boldsymbol{\Sigma}=n\left(\boldsymbol{P}-\boldsymbol{p}\boldsymbol{p}^T\right);$$

čia $\boldsymbol{P}$ – diagonalioji matrica, kurios elementai yra $p_1, p_2, ..., p_k$. Lengva patikrinti, kad Rang($\boldsymbol{\Sigma}$)=k-1, nes $p_1+...+p_k=1$.

1.2. χ^2 kriterijus

χ^2 kriterijus grindžiamas pateikiamomis teoremomis (žr. [2]):

1 T e o r e m a. Tegu p_i>0, i = 1,..., k. Apibrėžkime

$$X_n^2=\sum_{i=1}^{k}\frac{(U_i-np_i)^2}{np_i}=\sum_{i=1}^{k}\frac{U_i^2}{np_i}-n. \tag{1.3}$$

Tada atsitiktinio dydžio X_n^2 tikimybinis skirstinys silpnai konverguoja į chi-kvadrato skirstinį su k-1 laisvės laipsnių, kai $n\to\infty$.

2 T e o r e m a. Tarkime, kad tikimybės p_i yra tam tikros nežinomų parametrų $\theta_1, \theta_2, ..., \theta_s$ funkcijos $\left(p_i=p_i\left(\theta_1,...,\theta_s\right)\right)$, kurios kiekviename s-mačio neišsigimusio intervalo Θ taške tenkina salygas:

a) $\sum_{i=1}^{k} p_i\left(\theta_1,...,\theta_s\right)=1$;

b) $p_i\left(\theta_1,...,\theta_s\right)>c^2>0$, i=1,...,k;

c) egzistuoja tolydžios dalinės išvestinės $\frac{\partial p_i}{\partial\theta_j}$, $\frac{\partial p_i^2}{\partial\theta_j\partial\theta_l}$ (i=1,...,k; j,l=1,...,s);

d) matricos $\mathbf{D}=\left\|\frac{\partial p_i}{\partial\theta_j}\right\|$, i = 1,..., k, j = 1,..., s, rangas lygus s.

Tarkime, kad n kartų atlikome pirmame skyrelyje aprašytą bandymą, kuriame tikimybė, kad įvyks įvykis A_i, lygi $p_i^0=p_i\left(\theta_1^0,...,\theta_s^0\right)$, čia $\theta_0=\left(\theta_1^0,...,\theta_s^0\right)$ yra vidinis intervalo Θ taškas.

Parametro $\boldsymbol{\theta}$ ivertinį galima gauti modifikuotuoju χ^2 minimumo metodu, t.y. ivertinys yra lygčių sistemos

$$\sum_{i=1}^{k}\frac{V_i-np_i\left(\theta_1,...,\theta_s\right)}{p_i\left(\theta_1,...,\theta_s\right)}\cdot\frac{\partial p_i}{\partial\theta_j}=0,\ \text{j=1,...,s} \tag{1.4}$$

sprendinys.

Kai minėtos sąlygos patenkintos, lygčių sistema (1.4) turi vienintelį sprendinį $\hat{\boldsymbol{\theta}}$, kuris pagal tikimybę konverguoja į $\boldsymbol{\theta_0}$, kai $n \to \infty$. Statistika $X_1^2 = X_1^2\left(\hat{p}\right) = \sum_{i=1}^{k} \frac{\left(V_i - np_i\left(\hat{\theta}_1,...,\hat{\theta}_s\right)\right)^2}{np_i\left(\hat{\theta}_1,...,\hat{\theta}_s\right)}$ asimptotiškai pasiskirsčiusi pagal chi-kvadrato skirstinį su *k*-l-*s* laisvės laipsnių.

Tarkime, kad tikriname hipotezę H: $p_1 = p_{10}$, ..., $p_k = p_{k0}$ (čia $p_{10},...,p_{k0}$ yra duoti skaičiai, tokie, kad $p_{10} + ... + p_{k0} = 1$), kai alternatyva yra $\bar{H}$: $p_i \neq p_{p_{i0}}$ nors vienam $i = 1,...,k$. Jeigu $p_{i0} > 0$ ir hipotezė H teisinga, tai statistika

$$X_n^2 = \sum_{i=1}^{k} \frac{\left(U_i - np_{i0}\right)^2}{np_{i0}} \xrightarrow[n\to\infty]{d} \chi^2\left(k-1\right). \tag{1.5}$$

Hipotezė H atmetama apytiksliu reikšmingumo lygmens α kriterijumi, kai teisinga nelygybė

$$X_n^2 > \chi_\alpha^2\left(k\text{-}1\right), \tag{1.6}$$

čia $\chi_\alpha^2\left(k-1\right)$ yra χ^2 skirstinio su $k-1$ laisvės laispnių α kritinė reikšmė.

Jeigu hipotezė neteisinga, tai a.d. χ^2 skirstinys aproksimuojamas necentriniu χ^2 skirstiniu su k-1 laisvės laipsnių ir necentriškumo parametru

$$\delta = n\sum_{i=1}^{k} \frac{\left(p_i - p_{i0}\right)^2}{p_{i0}}. \tag{1.7}$$

Taigi apytikslio kriterijaus galios funkcija

$$\beta\left(\delta\right) = P\left\{\chi_{k-1,\delta}^2 > \chi_\alpha^2\left(k-1\right)\right\}, \tag{1.8}$$

čia $\chi_{k-1,\delta}^2$ yra a.d. turintis necentrinį χ^2 skirstinį su k-1 laisvės laipsnių ir necentriškumo parametru δ.

Analogiškai formuluojami kriterijai, tikrinant sudėtingąsias hipotezes H: $p_i = p_{i0}\left(\theta_1,...,\theta_s\right)$ apie tai, kad tikimybės p_i gali būti išreikštos funkcijomis $p_i\left(\theta_1,...,\theta_s\right)$ nuo nežinomų parametrų $\theta_1,...,\theta_s$, $s < k-1$.

Kadangi šiame darbe apsiribosime paprastosiomis hipotezėmis, tai kriterijaus formuluoti nepatariame.

2. PAPRASTOSIOS SUDERINAMUMO HIPOTEZĖS TIKRINIMAS χ^2 SUDERINAMUMO KRITERIJUMI

2.1. Paprastoji suderinamumo hipotezė

Tarkime $\boldsymbol{X}=(X_1,X_2,...,X_n)^T$ yra paprastoji imtis, gauta stebint a.d. $X \sim F \in G \subset \mathcal{F}$, kurio pasiskirstymo funkcija F priklauso tam tikrai pasiskirstymo funkcijų aibei $\mathcal{F}$.

Neparametrine hipoteze vadiname tvirtinimą

$$H_0 : X \sim F \in \mathcal{F}_0 \subset \mathcal{F},$$

kad a.d. X pasiskirstymo funkcija priklauso tam tikram funkcijų aibės $\mathcal{F}$ poaibiui $\mathcal{F}_0$.

Paprastąja suderinamumo hipoteze vadiname tvirtinimą

$$H_0 : X \sim F(x) \equiv F_0(x) \subset \mathcal{F}, \tag{2.1}$$

čia $F_0(x)$ yra pilnai nusakyta (žinoma) pasiskirstymo funkcija iš aibės $\mathcal{F}$.

2.2. χ^2 kriterijus

Tarkime, kad paprastoji imtis $\boldsymbol{X}=(X_1,X_2,...,X_n)^T$ gauta stebint a.d. X su pasiskirstymo funkcija $F(x)$.

Suskirstykime X galimų reikšmių sritį į k intervalų taškais $-\infty = a_0 < a_1 < ... < a_k = \infty$. Pažymėkime U_i stebėjimų, patekusių į intervalą $(a_{i-1}, a_i]$ skaičių:

$$U_i = \sum_{j=1}^{n} 1_{(a_{i-1},a_i]}(X_j),\ U_1 + ... + U_k = n,$$

o $p_i = \boldsymbol{P}\{X \in (a_{i-1}, a_i]\} = F(a_i) - F(a_{i-1})$ patekimo į i-tąjį intervalą tikimybę. Tada a.v. $\boldsymbol{U} = (U_1,...,U_k)^T$ turi polinominį skirstinį

$$\boldsymbol{U} = (U_1,...,U_k)^T \sim P_k(n, \boldsymbol{p}),\ \boldsymbol{p} = (p_1,...,p_k)^T. \tag{2.2}$$

Vietoje (2.1) hipotezės H_0 tikriname hipotezę

$$H_0' :\ p_i = p_{i0},\ i = 1,...,k, \tag{2.3}$$

kad polinominio skirstinio (2.2) tikimybės p_i yra lygios skaičiams p_{i0} , gautiems naudojant hipotetinę pasiskirstymo funkciją F_0:

$$p_{i0} = \boldsymbol{P}\left\{X \in \left(a_{i-1}, a_i\right] \middle| H_0\right\} = F_0\left(a_i\right) - F_0\left(a_{i-1}\right),\ i = 1,...,k\,.$$

Hipotezei H_0' tikrinti galime pritaikyti 1.2. skyrelyje pateiktą apytikslį kriterijų (1.6). Sudarome statistiką

$$X_n^2 = \sum_{i=1}^{k} \frac{(U_i - np_{i0})^2}{np_{i0}}. \tag{2.4}$$

Hipotezė H_0' atmetama apytiksliu reikšmingumo lygmens α kriterijumi, kai teisinga nelygybė

$$X_n^2 > \chi_\alpha^2\left(k\text{-}1\right). \tag{2.5}$$

Pastebėsime, kad hipotezės H_0 ir H_0' nėra ekvivalenčios. Hipotezė H_0' tvirtina tik, kad pasiskirstymo funkcijos pokytis i-tajame intervale lygus p_{i0}, tačiau nereglamentuoja jos elgesio intervalo viduje. Norint suartinti hipotezes H_0 ir H_0', reikėtų abscisių ašies padalijimą į intervalus padaryti kuo smulkesnį. Tačiau, iš kitos pusės, kriterijus (2.5) yra apytikslis ir jo tikslumas priklauso nuo to, kaip gerai a.d. X_n^2 skirstinys aproksimuojamas χ^2 skirstiniu. Jeigu imties didumas n yra mažas, o hipotetinės tikimybės p_{i0} mažos, tai ,imant smulkius intervalus, patekusių į juos stebėjimų skaičiai U_i įgis mažas reikšmes ir aproksimacija χ^2 skirstiniu bus netiksli.

Praktiškai taikant χ^2 kriterijų matematinės statistikos knygose (žr., pvz.,[1]), rekomenduojama, kad a.d. U_i realizacijos būtų ne mažesnės už 5 (kai kuriose knygose pvz. [4], ne mažesnės už 10).

Lieka atviras klausimas, kaip parinkti abscisių ašies suskaidymo į intervalus taškus a_i (tuo pačiu ir hipotetines tikimybes p_{i0}). Suprantama, kad nuo suskaidymo gali priklausyti gaunamos išvados: esant vienam suskaidymui hipotezė gali būti atmesta, o kitam priimta, remiantis tais pačiais statistiniais duomenimis.

2.3. Intervalų skaičiaus ir suskaidymo taškų parinkimas

Praktiškai statistinis modelis dažnai parenkamas vizualiai palyginant stebimo a.d. teorines charakteristikas (pasiskirstymo funkciją, tankį ir pan.) su jų empiriniais analogais. Aprašomojoje statistikoje yra pateikiami įvairūs duomenų grafinio vaizdavimo metodai, stengiantis juos pavaizduoti taip, kad vaizdžiau būtų galima palyginti su atitinkamomis teorinėmis charakteristikomis. Be to, tinkamo modelio parinkimą gali palengvinti profesinės žinios apie stebimą a.d. *X*. Abejotinais atvejais yra taikomi statistiniai turimų duomenų atitikimo hipotetiniams teoriniams modeliams kriterijai. Vieni iš universalių ir praktikoje dažnai naudojamų yra χ^2 tipo kriterijai. Pagrindinis jų trūkumas yra tai, kad jie reikalauja duomenų grupavimo, todėl imtis turėtų būti pakankamai didelė. Mažoms imtims naudojami Kolmogorovo ar ω^2 kriterijai, kurie taikomi negrupuotiems duomenims.

Trumpai aptarsime grafinius duomenų vaizdavimo metodus, kurie tam tikra prasme susiję su intervalų parinkimu taikant χ^2 kriterijų.

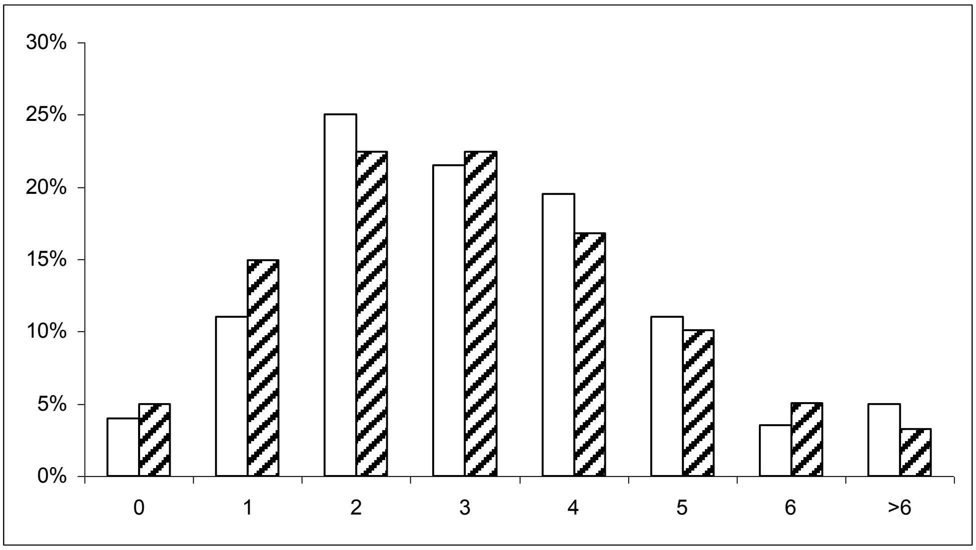
2.3.1 Diskretusis atvejis

Tarkime, kad paprastoji imtis $\boldsymbol{X} = (X_1, X_2, ..., X_n)^T$, gauta stebint diskretųjį a.d. *X*, kurio galimos reikšmės $b_1, b_2, ...$, o jų įgijimo tikimybės $p_i = \boldsymbol{P}\{X = b_i\}$, $i = 1, 2, ...$ Tikimybių p_i įvertiniai

$$\hat{p}_i = \frac{D_i}{n} = \frac{1}{n}\sum_{j=1}^{n} 1_{\{b_i\}}\left(X_j\right),\ i = 1, 2, \ldots .$$

Tikimybinio skirstinio empirinį analogą galima pavaizduoti stulpeline diagrama: aukščio d_i / n stulpeliai virš taškų b_i; čia d_i yra a.d. D_i realizacija. Jeigu hipotetinį skirstinį nusako tikimybės $p_{i0} = \boldsymbol{P}\{X = b_i | H_0\}$, tai greta stulpelių d_i / n galima pavaizduoti stulpelius aukščio p_i.

Pav. 2.1 yra pateikta stulpelinė diagrama, kai statistiniai duomenys gauti $n = 200$ modeliuojant Puasono a.d. $X \sim P\ (\lambda)$, Greta stulpelių aukščio d_i / n yra pavaizduotos Puasono skirstinio tikimybės $p_i = \dfrac{\lambda^m}{m!} \cdot e^{-\lambda}$, $i = 0, 1, 2, ...$ (šviesūs stulpeliai).



2.1 pav. Stulpelinė diagrama

Vizualiai palyginę d_i / n ir p_i, galime priimti sprendimą apie modelio tinkamumą. Abejotinu atveju tikriname sudarinamumo hipotezę, pavyzdžiui, taikydami χ^2 kriterijų.

Jeigu a.d. X įgyjamų reikšmių skaičius baigtinis ir lygus k, o hipotetinės tikimybės nėra mažos (t.y. $d_i \geq 5$), tai remdamiesi a.v. $\boldsymbol{U} = (U_1, ..., U_k)^T = (D_1, ..., D_k)^T$ tikriname hipotezę H_0': $p_i = p_{i0}$, $i = 1, ..., k$, kriterijumi (2.5).

Jeigu galimų reikšmių skaičius didelis ar begalinis (pvz. Puasono neigiamas binominis skirstinys ir pan.), tai kraštines reikšmes tenka apjungti. Tiksliau nagrinėjame vektorių $\boldsymbol{U} = (U_1, ..., U_k)^T$, kuriame $U_1 = \sum_{i \leq R_1} D_i$, $U_j = D_{R_1 + j}$, $j = 2, ..., k-1$, $U_k = \sum_{i \geq R_1 + k} D_i$. Remdamiesi vektoriumi $\boldsymbol{U}$ tikriname hipotezę H_0': $p_1 = \sum_{i \leq l} p_{i0}$, $p_j = p_{j0}$, $j = 2, ..., k-1$, $p_k = \sum_{i \geq l+k} p_{i0}$.

Taigi diskretaus skirstinio atveju intervalų skaičius k ir abscisių ašies padalijimas į intervalus nusakomas beveik vienareikšmiškai pagal aprašytąją procedūrą ir nesukelia sunkumų.

2.3.2. Tolydusis atvejis

Tarkime, kad paprastoji imtis $\boldsymbol{X} = (X_1, X_2, ..., X_n)^T$ yra gauta stebint a.d. $X \sim F \in \mathcal{F}$; čia $\mathcal{F}$ kokia nors absoliučiai tolydžių skirstinių pasiskirstymo funkcijų aibė. Tikriname hipotezę $H_0 : X \sim F(x) \equiv F_0(x)$ arba $X \sim f(x) \equiv f_0(x) = F_0'(x)$; čia $F_0(x)$ $(f_0(x))$ yra pilnai nusakyta (žinoma) absoliučiai tolydaus skirstinio pasiskirstymo funkcija (tankis).

Pradžioje aptarsime grafinius duomenų vaizdavimo būdus, kurie įgalina vizualiai palyginti empirines ir teorines charakteristikas. Dėl vaizdumo aptarimą pavaizduosime konkrečiu pavyzdžiu.

2 priede yra paprastosios didumo $n = 200$ imties, gautos modeliuojant a.d. $X \sim N\ (1,1)$, realizacija.

1) Empirinė pasiskirstymo funkcija. Tegu $\boldsymbol{x} = (x_1, ..., x_n)^T$ yra atsitiktinės imties $\boldsymbol{X} = (X_1, ..., X_n)^T$ realizacija. Priskirkime kiekvienai stebėtai reikšmei x_i tikimybę $1/n$ ir sudarykime gauto diskretaus skirstinio pasiskirstymo funkciją

$$F_n(x) = \frac{d_n(x)}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty, x]}(x_i), x \in \mathbf{R}. \tag{2.6}$$

Pasiskirstymo funkcija $F_n(x)$ yra atsitiktinės funkcijos

$$\hat{F}_n(x) = \frac{d_n(x)}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(-\infty,x]}(X_i), x \in \mathbf{R},$$

vadinamos empirine pasiskirsymo funkcija, realizacija.

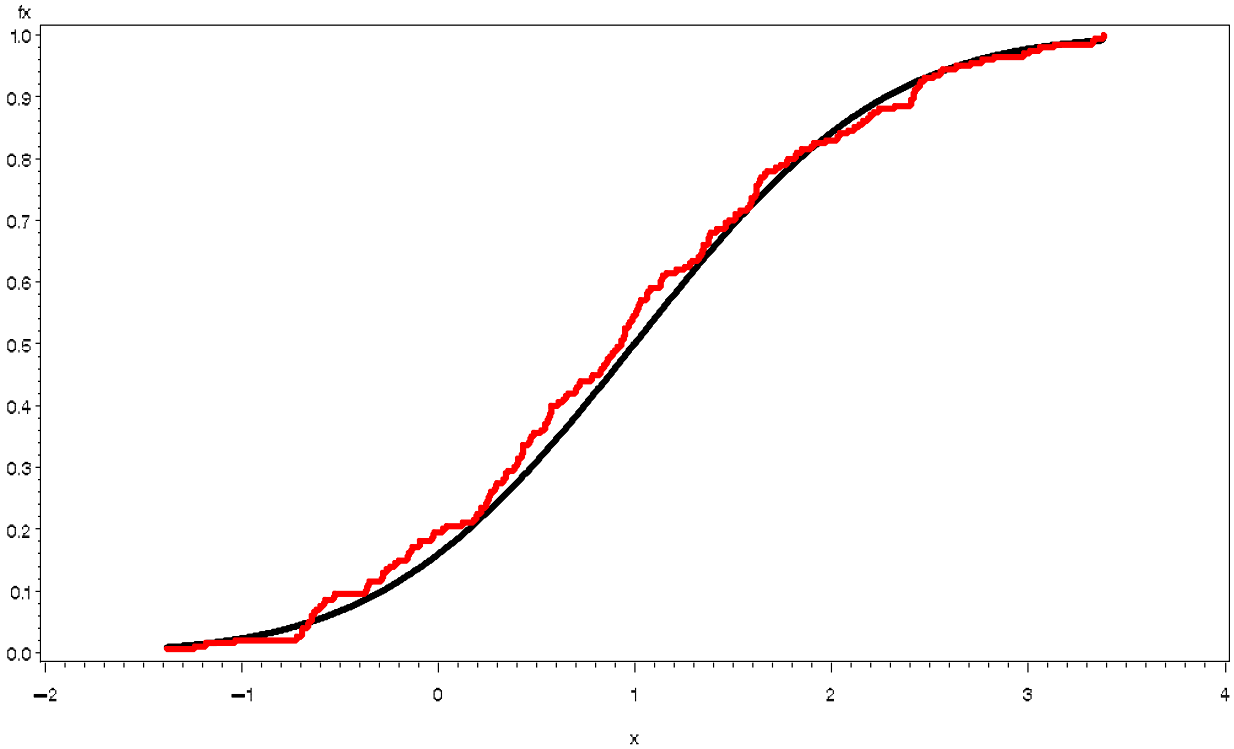
3 t e o r e m a (Glivenkos-Kantelio). Jei F yra bet kuri pasiskirstymo funkcija, tai

$$\sup_{x\in\mathbf{R}} \left|\hat{F}_n(x) - F(x)\right| \xrightarrow{b.v.} 0, \text{ kai } n \to \infty,$$

$$\hat{F}_n(x) \to F(x).$$

Iš šios teoremos išplaukia, kad empirinė pasiskirstymo funkcija yra geras teorinės pasiskirstymo funkcijos analogas. Todėl esant teisingai hipotezei H_0, jos realizacija $F_n(x)$ turėtų būti artima funkcijai $F_0(x)$. Pavaizdavę funkcijų $F_n(x)$ ir $F_0(x)$ grafikus, galima daryti preliminarią išvadą apie hipotezės H_0 teisingumą.

2.2 pav. yra pateiktas empirinės pasiskirstymo funkcijos realizacijos $F_n(x)$, sukonstruotos pagal 2 priedo duomenis, grafikas. Kartu yra pateiktas ir modeliuoto a.d. X pasiskirstymo funkcijos $F_0(x) = \Phi(x-1)$ grafikas.



2.2 pav. Empirinė ir teorinė pasiskirstymo funkcijos

2) Histograma. Charakteringąsias skirstinio savybes geriau atspindi tankio funkcija, todėl paprastai yra lyginamas empirinis tankio analogas (histograma) su teoriniu tankiu.

Histograma konstruojama tokiu būdu: a.d. X galimų reikšmių sritis sudalinama į vienodo ilgio h intervalus $\left(a_{i-1}, a_i\right]$. Jeigu $h = a_i - a_{i-1}$ yra mažas, tai su bet kuriuo $x \in \left(a_{i-1}, a_i\right]$

$$F\left(a_i\right) - F\left(a_{i-1}\right) = \int_{a_{i-1}}^{a_i} f\left(y\right) dy \approx f\left(x\right) h,$$

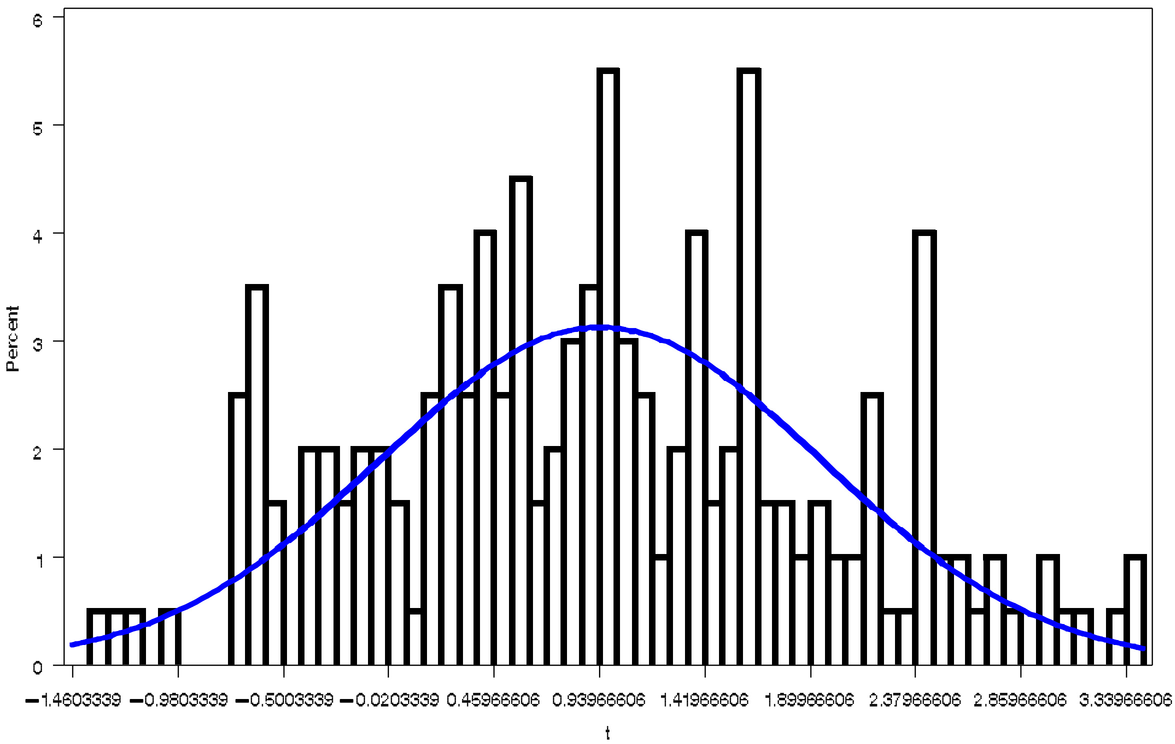
todėl tankio reikšmė nedaug skiriasi nuo

$$f_n\left(x\right) = \frac{1}{h}\left[F_n\left(a_i\right) - F_n\left(a_{i-1}\right)\right] = \frac{1}{nh} \sum_{j=1}^{n} 1_{\left(a_{i-1}, a_i\right]}\left(x_j\right) = \frac{n_i}{nh};$$

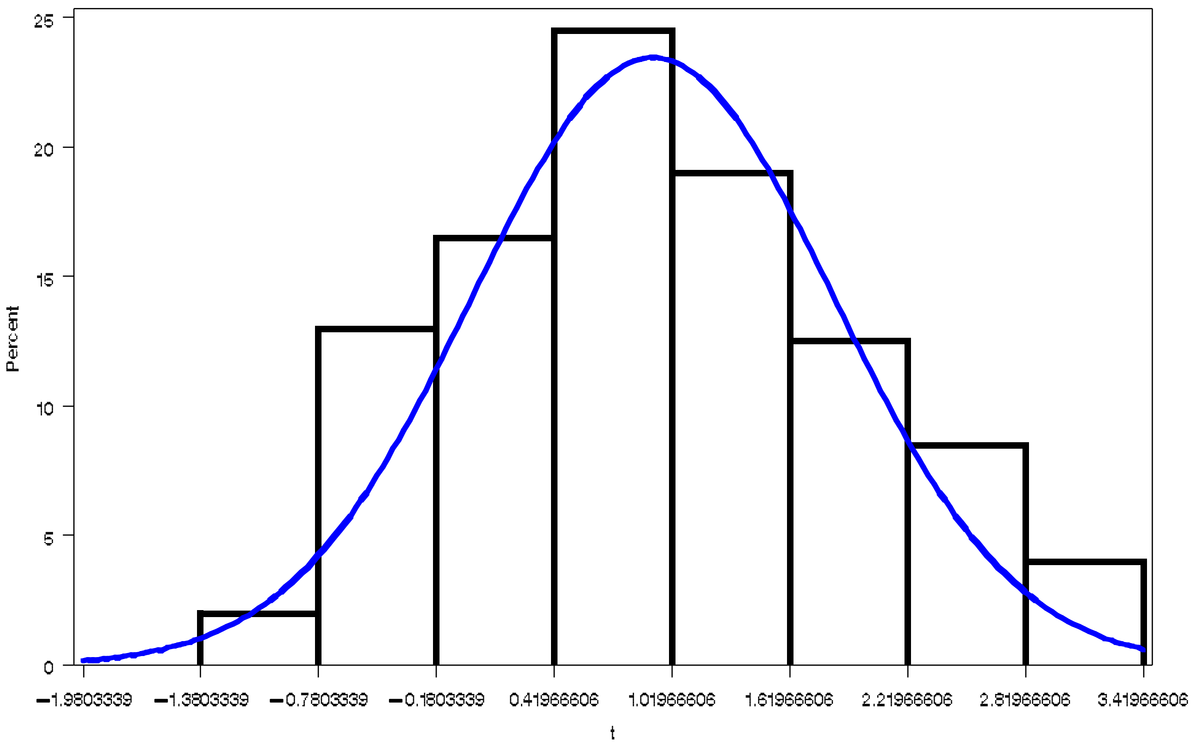
čia n_i yra skaičius tų x_j, kurie priklauso intervalui $\left(a_{i-1}, a_i\right]$.

Funkcijos $f_n\left(x\right)$ grafikas yra vadinamas histograma.

Jeigu n yra fiksuotas, tai svarbu tinkamai parinkti h. Jei h labai mažas, tai kiekviename intervale bus po 0 ar 1 iš x_j. Gauname aukštus stulpelius, atitinkančius 1 ir intervalus be stulpelių, atitinkančius 0 (žr. 2.3 pav.). Jei h didelis, tai turime du ar tris plačius stulpelius (žr. 2.4 pav.). Abiem atvejais sunku ką nors pasakyti apie tankio pavidalą.



2.3 pav. Histograma ir teorinis tankis, *h = 0,08*



2.4 pav. Histograma ir teorinis tankis, ***h = 0,6***

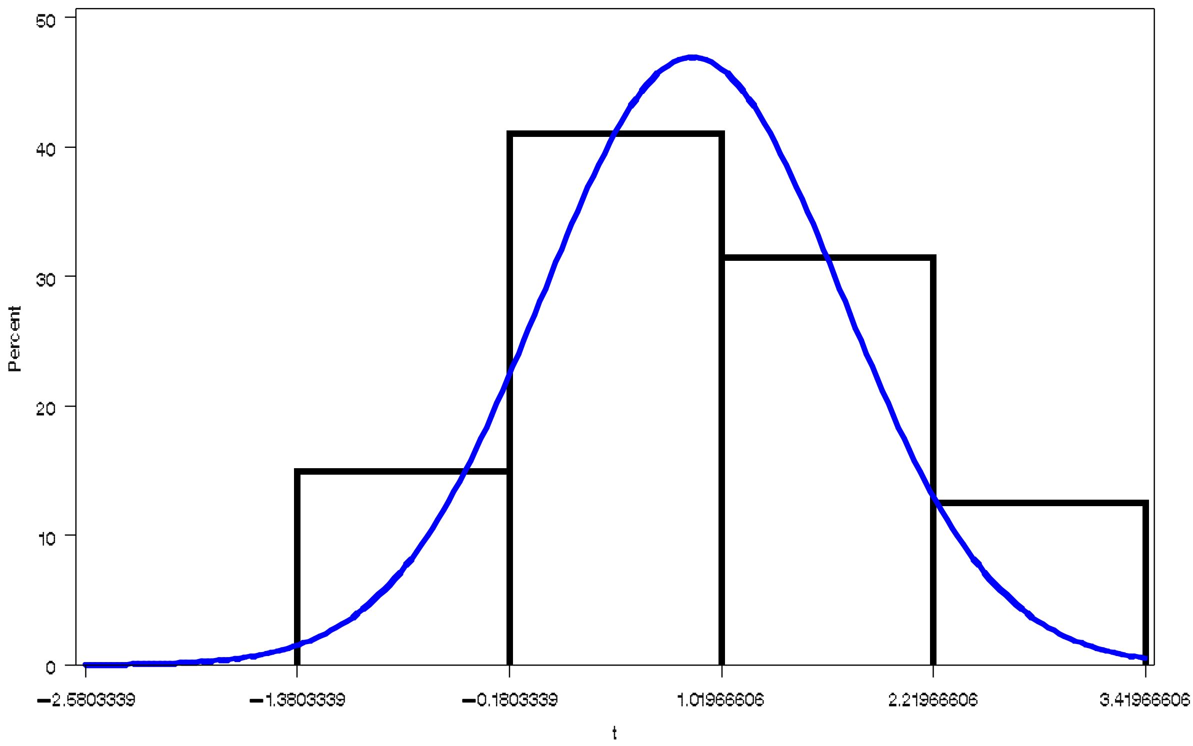
Kai kuriuose statistiniuose paketuose naudojama taisyklė (pagal nutylėjimą): jei $n \leq 25$, imami $k = 5$ intervalai. Jei $n > 25$, imama $k = [1 + \ln n]$, bet ne daugiau 21 intervalo.

Parinkus intervalų skaičių, grupavimo intervalų galais imama

$$a_0 = -\infty,\ \ a_1 = x_{\min} + h,\ a_2 = x_{\min} + 2h, \ldots,\ a_{k-1} = x_{\max} - h,\ a_k = \infty; \tag{2.7}$$

čia $x_{\min}$ ir $x_{\max}$ yra ektremaliosios imties reikšmės, o $h = (x_{\max} - x_{\min})/k$.

2.4 pav. yra pateiktas $f_n(x)$ grafikas pagal 2 priedo duomenis imant h=0,6



2.5 pav. Histograma ir teorinis tankis, $h = 1,2$

Dažnai sprendimas apie modelio tinkamumą priimamas vizualiai palyginus tankį $f_0\left(x\right)$ ir histogramą $f_n\left(x\right)$. Abejotinais atvejais nuspręsti apie modelio tinkamumą galime remdamiesi statistiniu suderinamumo kriterijumi.

χ^2 suderinamumo kriterijus konstruojamas tokiu būdu: apibrėžiame atsitiktinius dydžius

$$i = 1,\dots,k;\ U_1 = \sum_{i=1}^{n} 1_{(-\infty,a_1]}\left(X_j\right),\ U_i = \sum_{i=1}^{n} 1_{(a_{i-1},a_i]}\left(X_j\right),\ i = 2,\dots,k-1;\ U_k = \sum_{i=1}^{n} 1_{(a_{k-1},\infty]}\left(X_j\right).$$

Remdamiesi polinominiu vektoriumi $\boldsymbol{U} = (U_1,\dots,U_k)^T$ vietoje hipotezės H_0 tikriname hipotezę H_0': $p_i = p_{i0}$, $i = 1,\dots,k$;

$$p_{10} = \int_{-\infty}^{a_1} f_0\left(x\right)dx;\ p_{i0} = \int_{a_{i-1}}^{a_i} f_0\left(x\right)dx,\ i = 2,\dots,k-1,\ p_{k0} = \int_{a_{k-1}}^{\infty} f_0\left(x\right)dx.$$

Tuo tikslu apskaičiuojame statistikos (2.4) reikšmę ir patikriname nelygybę (2.5).

2.1 P a s t a b a. Aprašytas kriterijus nėra visai korektiškas. Intervalų parinkimas pagal (2.4) ar panašią taisyklę reiškia, kad intervalų galai a_i priklauso nuo stebėjimų $X_1, X_2, \dots, X_n$. Tai reiškia, kad ir hipotetinės tikimybės p_{i0} priklauso nuo imties. Tuo tarpu 1 teoremos tvirtinimas yra gautas, tariant, kad p_{i0} yra konstantos.

2.2 P a s t a b a. Jeigu tankis $f_0(x)$ yra unimodalus ir mažėja, kai argumentas x tolsta nuo modos į kairę ir dešinę, tai sudarant 2.3 pav. pavaizduotą histogramą gausime, kad vidiniuose intervaluose stebėjimų skaičius didelis, o tolstant nuo centro stebėjimų skaičiai bus maži. Taip grupuojant duomenis prarandama žymiai daugiau informacijos, negu tuo atveju, kai stebėjimų skaičiai intervaluose pasiskirstę tolygiau. Suprantama, kad tada intervalų ilgiai nebus vienodi.

3) Tikimybinė kreivė. Iš tikimybių teorijos kurso žinome, kad a.d. $Y = F_0(X) \sim U(0,1)$ turi tolygųjį intervale $(0,1)$ skirstinį, jeigu a.d. *X* pasiskirstymo funkcija $F_0(x)$ absoliučiai tolydi. Taigi , jei hipotezė H_0 teisinga, a.v. $\boldsymbol{Y} = (Y_1,...,Y_n)^T$, $Y = F_0(X_i)$, $i = 1,...,n$, galime traktuoti kaip paprastąją imtį, gautą stebint a.d. $Y \sim U(0,1)$. Atsitiktinio dydžio *Y* pasiskirstymo funkcija

$$G(y) = \begin{cases} 0, & kai \quad y < 0, \\ y, & kai \quad 0 \le y \le 1, \\ 1, & kai \quad y > 1; \end{cases} \tag{2.8}$$

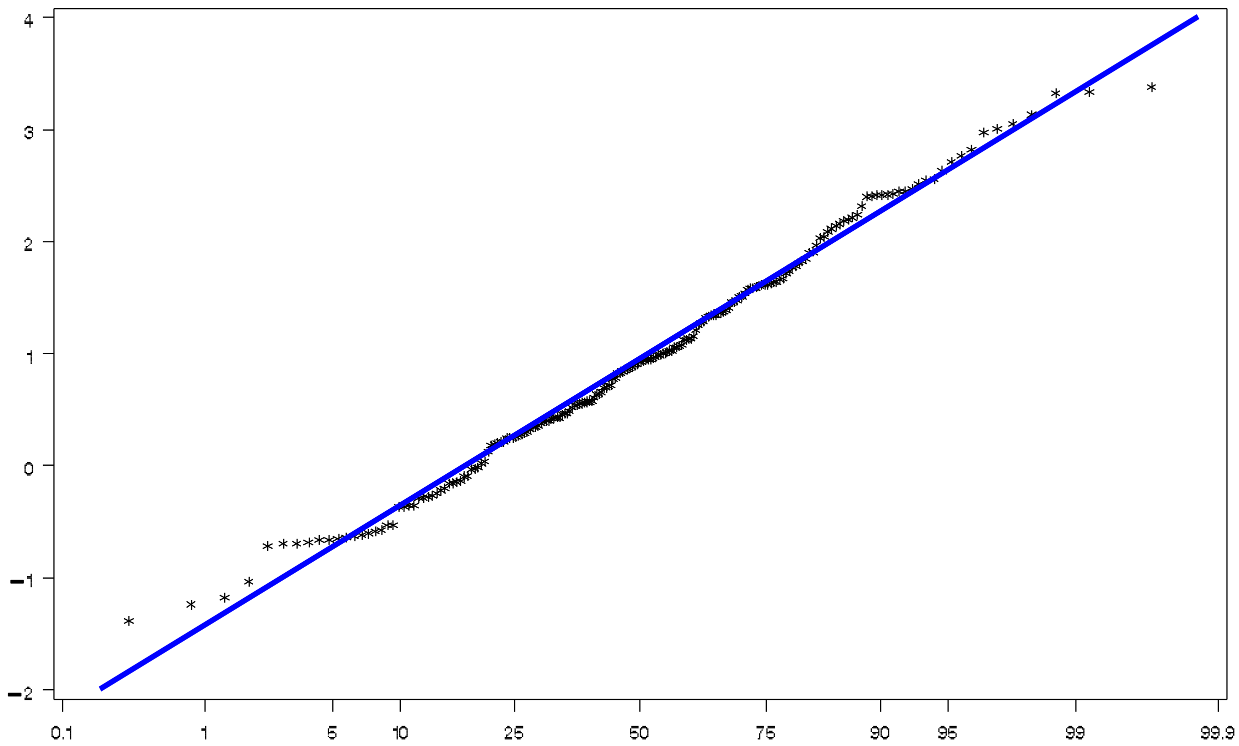
o tankio funkcija

$$g(y) = G'(y) = \begin{cases} 1, & kai \quad y \in (0,1), \\ 0, & kai \quad y \notin (0,1). \end{cases} \tag{2.9}$$

Jeigu remdamiesi imtimi $(Y_1,...,Y_n)^T$ sudarysime empirinę pasiskirstymo funkciją $\hat{G}_n(y)$, tai jos realizacijų palyginimas su $G(y)$ bus žymiai vaizdesnis, nei lyginant $F_n(x)$ ir $F_0(x)$.

Pavyzdžiui, galime atidėti taškus $(y_i, i/n)$, $i = 1,...,n$. Tada esant teisingai hipotezei, šie taškai bus išsidėstę apie tiesę iš (pav. 2.6). Tokio tipo grafikas vadinamas tikimybine kreive.

Pav. 2.6 yra pateikta tikimybinė kreivė, sudaryta remiantis 2 priedo duomenimis.



2.6 pav. Tikimybinė kreivė

4) Transformuotų duomenų histograma. Jeigu remdamiesi transformuotais duomenimis $Y_1,...,Y_n$ sudarysime histogramą, tai ją reikės lyginti su paprasto pavidalo tankiu $g(y)$. Toks palyginimas, matyt, bus žymiai vaizdesnis, negu lyginant histogramą $f_n(x)$ su tankiu $f_0(x)$ pagal 2.5 pav. Tada transformuotiems dydžiams grupavimo intervalų galai bus

$$c_0 = 0,\ c_i = \Phi(a_i - 1),\ i = 1,...,k-1;\ c_k = 1;$$

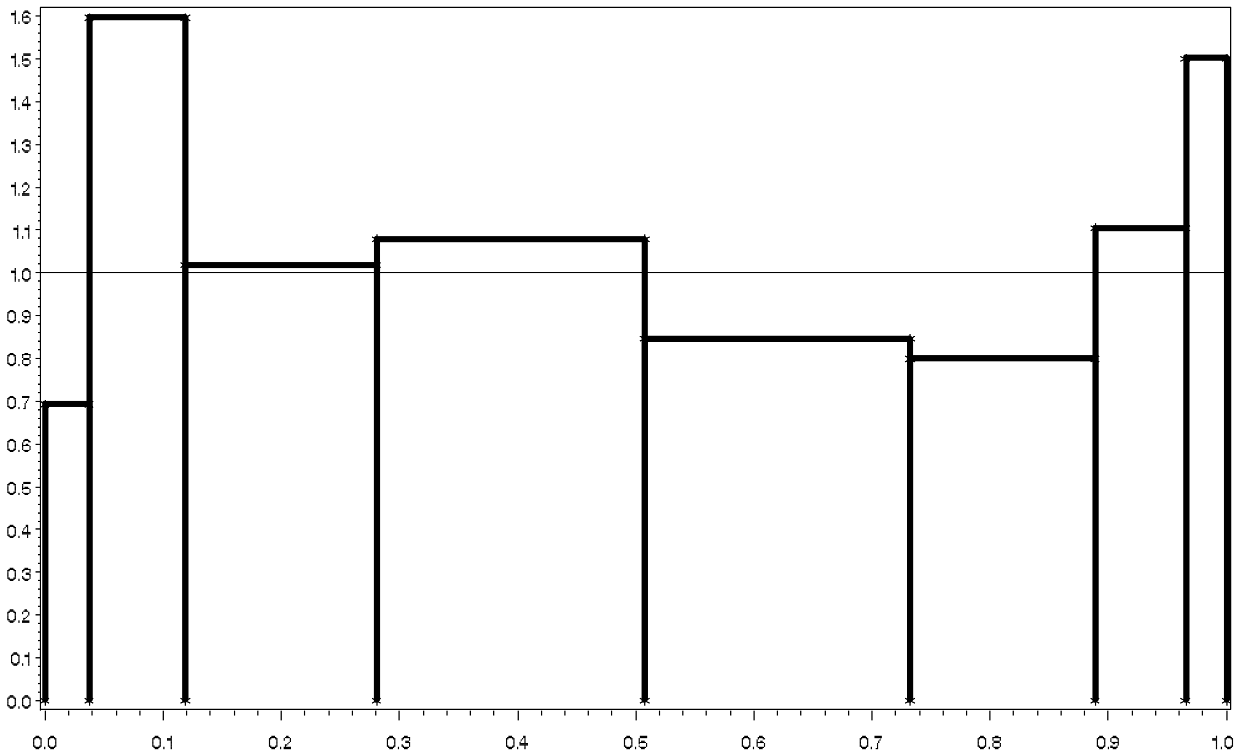
intervalų ilgiai

$$h_i = \Phi(a_i - 1) - \Phi(a_{i-1} - 1),$$

ir tansformuotų dydžių histograma

$$g_n(y) = \frac{n_i}{nh_i},\ \text{kai}\ y \in (c_{i-1}, c_i],\ i = 1,...,k. \tag{2.10}$$

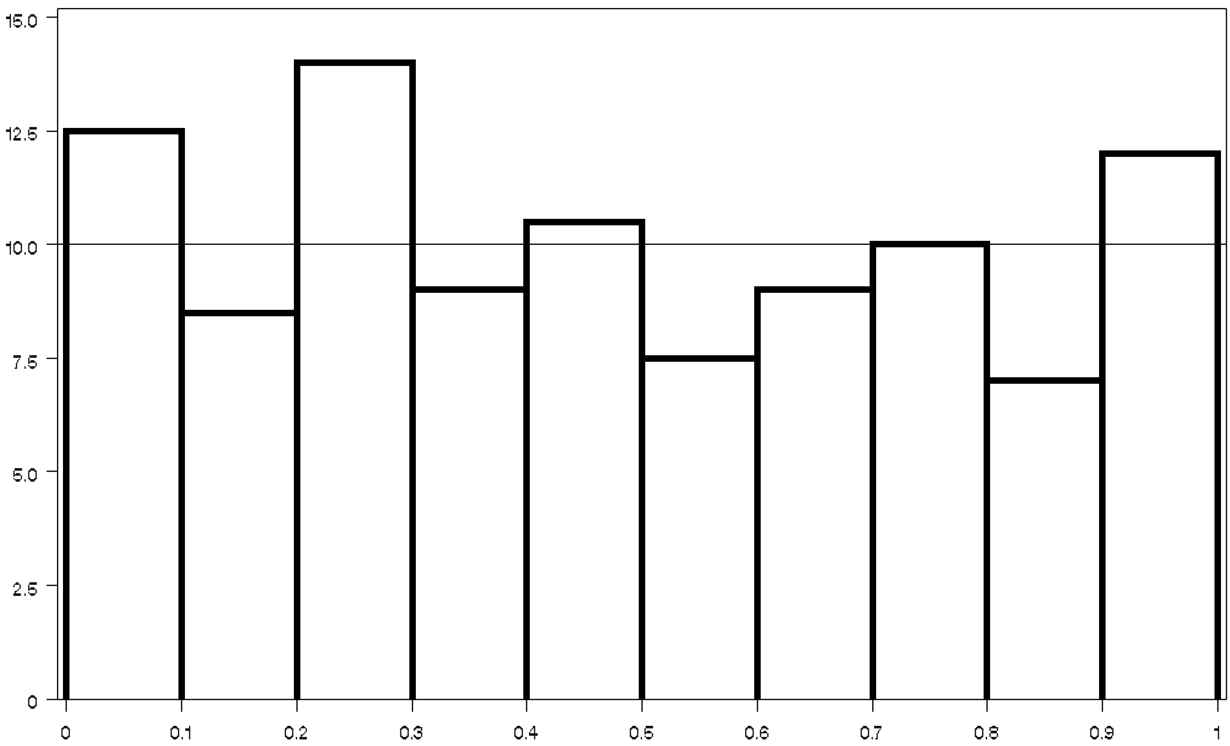
2.7 pav. yra pateiktas $g_n(y)$ grafikas kartu su teoriniu tankiu (2.9).



2.7 pav. Transformuotų dydžių histograma, kai intervalų dydžiai nevienodi

Iš šio paveiksliuko dar akivaizdžiau matome, kad statistiniai duomenys naudojami neracionaliai: vidiniai intervalai žymiai didesni už kraštinius. Jeigu iš pat pradžių būtumėm stebėję a.d. $Y \sim U\left(0,1\right)$, tai , suprantama, intervalą $\left(0,1\right)$ būtume suskaidę į vienodo ilgio intervalus.

2.8 pav. yra pateikta transformuotų dydžių histograma, kai imam $k = 10$ vienodo ilgio $h = 1/10$ intervalų pagal 2 priedo duomenis



2.8 pav. Transformuotų dydžių histograma, kai grupavimo intervalų ilgiai vienodi

Jeigu grįšime atgal prie a.d. $X_1,...,X_n$, tai vietoje 2.6 pav. gausime 2.3 tipo histogramą, kurioje intervalų ilgiai nėra vienodi (vienodų tikimybių intervalai). Grupavimo intervalų galais bus taškai

$$a_i = 1 + \Phi^{-1}\left(\frac{i}{k}\right) = 1 + z_{i/k}\,,\ i = 1,...,k-1\,; \tag{2.11}$$

čia $z_{i/k}$ yra standartinio normaliojo skirstinio α kritinė reikšmė.

Mūsų nuomone, histogramos ir teorinio tankio palyginimas 2.8 pav. yra vaizdesnis, negu pav. 2.7 ir tuo labiau negu pav. 2.5. Jeigu vizualiai palyginus $g_n(y)$ ir $g(y)$ kyla abejonės dėl modelio $f_0(x)$ tinkamumo, tai galima tikrinti suderinamumo hipotezę H_0': $p_i = \frac{1}{k}$, $i = 1,...,k$, kriterijumi (2.5). Statistikos (2.4) išraiška yra paprastesnė

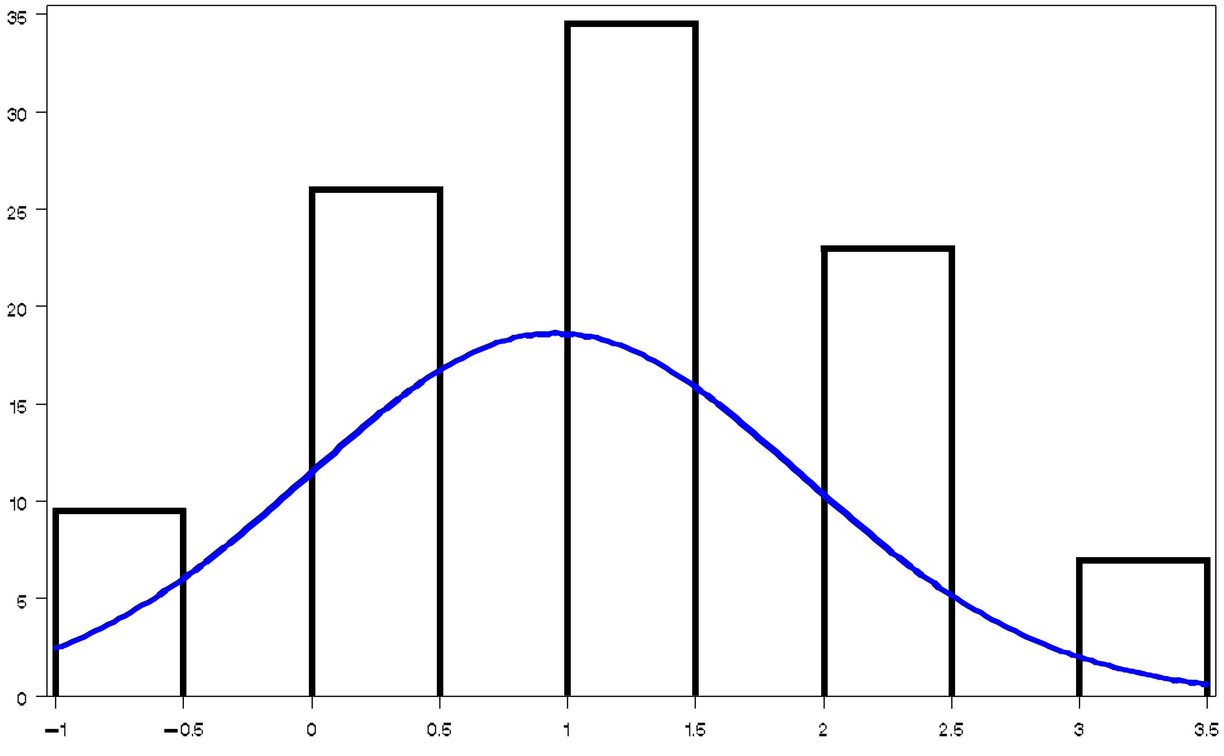
$$X_n^2 = \sum_{i=1}^{k} \frac{(U_i - n/k)^2}{n/k} = \frac{k}{n}\sum_{i=1}^{k} U_i^2 - n\,. \tag{2.12}$$

Duomenų klasifikavimas naudojant vienodų tikimybių intervalus turi privalumų. Akivaizdu, kad grupavimo intervalų galai skirtingai nuo (2.7) tipo jau nepriklauso nuo imties. Be to, intervalų parinkimas yra vienareikšmiškai nustatytas, kai tik parinktas jų skaičius k. Daugelyje darbų (žr. pvz. [1]) yra tvirtinama, kad vienodų tikimybių χ^2 kriterijus yra galingesnis už vienodo ilgio intervalų χ^2 kriterijų, jeigu alternatyvų klasė nėra apibrėžta. Be to, esant vienodoms tikimybėms, konvergavimas į ribinį χ^2 dėsnį yra greitesnis. Šį faktą iš dalies patvirtina ir rezultatai, gauti tolimesniuose skyreliuose.

2.3. Pastaba. Jeigu tikrinama sudėtingoji hipotezė $H : X \sim f\left(x;\boldsymbol{\theta}\right)$, tai 2.5 tipo paveikslėlyje yra lyginama histograma su tankiu $f\left(x;\hat{\boldsymbol{\theta}}\right)$; čia $\hat{\boldsymbol{\theta}}$ parametro $\boldsymbol{\theta}$ įvertinys. Manome, kad ir šiuo atveju palyginimas bus vaizdesnis sudarant 2.8 pav. pavaizduoto tipo histogramą, kuri gaunama atliekant transformaciją $Y_i = F_0\left(X_i,\hat{\boldsymbol{\theta}}\right)$, $i = 1,...,n$.

2.4. Pastaba. Kartais stebint tolydųjį a.d. statistiniai duomenys yra grubiai suapvalinti. Pavyzdžiui, matuojant gaminio parametrą, matavimo prietaisas įgalina užregistruoti jo reikšmes tik tam tikru stambiu žingsniu. Duomenų apvalinimas gali labai iškreipti histogramos pavidalą.

Pavyzdžiui, tarkime, kad 2 priedo duomenys yra suapvalinti iki sveikų vienetų. 2.7 pav. yra pateikta histograma, sudaryta pagal tokius suapvalintus duomenis (pagal nutylėjimą, naudojant SAS programų sistemą).



2.9 pav. Iki vienetų suapvalintų duomenų histograma

Lygindami su 2.3 pav. matome, kad histograma yra labai iškraipyta ir programų pakete SAS nėra numatytas histogramos koregavimas atsižvelgiant į duomenų apvalinimą. Koregavimas atsižvelgiant į duomenų apvalinimą aptartas [2].

Norint, kad histograma nebūtų iškraipoma, kai duomenys suapvalinti iki sveikų vienetų, intervalo ilgis h turi būti didesnis už 1, o intervalų galai a_i turi būti pavidalo $m \pm 0,5$; m – sveikas skaičius. Jeigu duomenys suapvalinti r ženklų po kablelio tikslumu, tai intervalo galai turėtų būti pavidalo $\left(m \pm 0,5\right) \cdot 10^{-r}$ pavidalo.

Tokiu atveju gauname, kad po transformavimo 2.8 pav. intervalo galai turėtų būti pavidalo $F\left(\left(m \pm 0,5\right) \cdot 10^{-r}\right)$ pavidalo. Todėl 2.8 pav. gautus intervalų galus reikėtų pastumti iki artimiausių tokio pavidalo skaičių. Suprantama, kad patekimo į intervalus tikimybės bus tik apytiksliai vienodos.

3. χ^2 KRITERIJAUS APROKSIMACIJŲ PATIKSLINIMAI

3.1. Aproksimacijos konstravimas

Kaip anksčiau minėjome, kriterijus (2.5), naudojamas suderinamumo hipotezei (2.3) tikrinti, yra tik apytikslis. Bandysime jį patikslinti, įvesdami keletą parametrų taip, kad gautos statitikos pirmieji trys momentai sutaptų su a.d. X_n^2 trimis momentais. Tegu:

$$V = a + bX^2. \tag{3.1}$$

čia a ir b yra parametrai, kuriuos reikia parinkti. Statistikos V skirstinį aproksimuosime χ^2 skirstiniu, kurio laisvės laipsnių skaičius ω irgi yra parenkamas parametras. Tada vietoje (2.5) gausime tokį kriterijų. Hipotezė H_0 atmetama, kai:

$$V > \chi_\alpha^2\left(\omega\right). \tag{3.2}$$

Norėdami parinkti tris parametrus a, b ir ω, sprendžiame lygčių sistemą:

$$\begin{cases} \boldsymbol{E}V = a + b\varsigma_1 = \omega \\ \boldsymbol{Var}\left(V\right) = b^2\varsigma_2 = 2\omega, \\ \boldsymbol{\mu_3}\left(V\right) = b^3\varsigma_3 = 8\omega \end{cases} \tag{3.3}$$

čia ς_1, ς_2, ir ς_3 yra statistikos X_n^2 pirmasis momentas, dispersija ir trečias centrinis momentas. Iš lygčių sistemos gauname tokias parametrų išraiškas:

$$a = \frac{8\varsigma_2^3}{\varsigma_3^2} - \frac{4\varsigma_1\varsigma_2}{\varsigma_3},\ b = \frac{4\varsigma_2}{\varsigma_3}\ \text{ir}\ \omega = \frac{8\varsigma_2^3}{\varsigma_3^2}. \tag{3.4}$$

3.2. Polinominio vektoriaus koordinačių momentai

3.2.1. Polinominio vektoriaus koordinačių pradiniai momentai

Statitikos X_n^2 momentams skaičiuoti naudosime sąryšius tarp pradinių ir faktorialinių momentų.

Kad skaičiavimai būtų paprastesni, šiek tiek supaprastinsime statistikos (2.4) išraišką užrašydami ją tokiu pavidalu:

$$X_n^2 = \frac{1}{n}\sum_{i=1}^{k}\frac{U_i^2}{p_{i0}} - n. \tag{3.5}$$

Pažymėkime $Z = \frac{1}{n}\sum_{i=1}^{k}\frac{U_i^2}{p_{i0}}$, tuomet $X_n^2 = Z - n$.

Tuomet statistikos X_n^2 vidurkis

$$\boldsymbol{E}X_n^2 = \boldsymbol{E}Z - n = \boldsymbol{E}\left[\sum_{i=1}^{k}\frac{U_i^2}{np_{i0}}\right] - n = \sum_{i=1}^{k}\frac{\alpha_2(i)}{np_{i0}} - n, \tag{3.6}$$

čia $\alpha_2(j)$ yra polinominio vektoriaus j-osios koordinatės antras pradinis momentas

$$\alpha_2(i) = EU_i^2.$$

Tuomet statistikos (3.5) dispersija

$$\boldsymbol{Var}(X^2) = \boldsymbol{Var}(Z-n) = \boldsymbol{Var}(Z) = \boldsymbol{E}Z^2 - (\boldsymbol{E}Z)^2, \tag{3.7}$$

o trečiasis centrinis momentas

$$\boldsymbol{\mu_3}(X^2) = \boldsymbol{\mu_3}(Z) = \boldsymbol{E}Z^3 - 3\boldsymbol{Var}(Z)\cdot\boldsymbol{E}Z^2 - (\boldsymbol{E}Z)^3. \tag{3.8}$$

Pagal (3.7) statistikos Z dispersijai apskaičiuoti reikalingos dvi išraiškos: $\boldsymbol{E}Z^2$ ir $(\boldsymbol{E}Z)^2$. Pradėsime nuo $\boldsymbol{E}Z^2$, kurio išraiška

$$\boldsymbol{E}Z^2 = \boldsymbol{E}\left[\sum_{i=1}^{k}\frac{U_i^2}{np_{i0}}\right]^2 = \boldsymbol{E}\left[\sum_{i=1}^{k}\frac{U_i^4}{n^2p_{i0}^2} + \sum_{i\neq j}\frac{U_i^2U_j^2}{n^2p_{i0}p_{j0}}\right] = \frac{1}{n^2}\sum_{i=1}^{k}\frac{\boldsymbol{E}(U_i^4)}{p_{i0}^2} + \frac{1}{n^2}\sum_{i\neq j}\frac{\boldsymbol{E}(U_i^2U_j^2)}{p_{i0}p_{j0}}. \tag{3.9}$$

Taigi, tam kad apskaičiuotume $\boldsymbol{E}Z^2$, pirmiausia turime rasti a.d. U_i, $i = 1,...,k$, ketvirtąjį pradinį momentą $\alpha_4(i)$ ir pradinį antros eilės mišrųjį momentą $\alpha_{22}(i,j)$

$$\boldsymbol{E}(U_i^4) = \alpha_4(i),$$

$$\boldsymbol{E}\left(U_i^2 U_j^2\right) = \alpha_{22}\left(i,j\right).$$

Trečiojo centrinio momento radimui, remiantis 3.8 išraiška, reikės papildomai apskaičiuoti EZ^3 ir $\left(EZ\right)^3$, Pradėsime nuo EZ^3, kurio išraiška:

$$\boldsymbol{E}Z^3 = \boldsymbol{E}\left[\sum_{i=1}^{k}\frac{U_i^2}{np_{i0}}\right]^3 = \sum_{i=1}^{k}\frac{\boldsymbol{E}U_i^6}{n^3 p_{i0}} + 3\sum_{i\neq j}\frac{\boldsymbol{E}\left(U_i^4 U_j^2\right)}{n^3 p_{i0}^2 p_{j0}} + \sum_{i\neq j\neq l}\frac{\boldsymbol{E}\left(U_i^2 U_j^2 U_l^2\right)}{n^3 p_{i0} p_{j0} p_{l0}}. \qquad (3.10)$$

Šiuo atveju mums reikalingi 6-tasis pradinis α_6 ir antros bei trečios eilės mišrieji pradiniai momentai $\alpha_{42}\left(i,j\right)$ ir $\alpha_{222}\left(i,j,l\right)$:

$$\boldsymbol{E}\left(U_i^6\right) = \alpha_6\left(i\right),$$

$$\boldsymbol{E}\left(U_i^4 U_j^2\right) = \alpha_{42}\left(i,j\right),$$

$$\boldsymbol{E}\left(U_i^2 U_j^2 U_l^2\right) = \alpha_{222}\left(i,j,l\right).$$

3.2.2. Polinominio vektoriaus koordinačių faktorialiniai momentai

Polinominio a.v. $\boldsymbol{U} = \left(U_1,...,U_k\right)^T$ eilės r mišriuoju faktorialiniu momentu vadiname vidurkį

$$\upsilon_{r_1,...,r_k} = \boldsymbol{E}\left[U_1\left(U-1_1\right)\cdot...\cdot\left(U_1-r_1+1\right)U_2\left(U_2-1\right)\cdot...\cdot\left(U_2-r_2+1\right)\cdot...\cdot U_k\left(U_k-1\right)\cdot...\cdot\left(U_k-r_k+1\right)\right] =$$
$$= \boldsymbol{E}\left(U_1^{[r_1]}\cdot...\cdot U_k^{[r_k]}\right),\ r_1+...r_k = r, \qquad (3.11)$$

čia $x^{[r]} = x\left(x-1\right)\cdot...\cdot\left(x-r+1\right)$

Remiantis generuojančios funkcijos išraiška (1.2) gauname, kad

$$\upsilon_{r_1,...,r_k} = \frac{\partial r_1+...+r_k}{\partial s_1^{r_1}\cdot...\cdot s_1^{r_k}}\psi\left(s_1,...,s_k\right)\Bigg|_{\vec{S}=\vec{1}} = n^{[r]}p_1^{r_1}\cdot...\cdot p_k^{r_k}. \qquad (3.12)$$

Taigi faktorialiniai momentai paprastai išreiškiami parametru n ir vektoriaus $\boldsymbol{p} = \left(p_1,...,p_k\right)^T$ koordinatėmis.

Išreiškę reikalingus pradinius momentus faktorialiniais momentais, gausime ir šių momentų išraiškas. Koordinatės U_i eilės r faktorialinius momentus žymėsime $\upsilon_r\left(i\right)$. Iš (3.11) gauname:

$$
\begin{cases}
\upsilon_1(i) = \boldsymbol{E}(U_i) = \alpha_1(i) \\
\upsilon_2(i) = \boldsymbol{E}\left[U_i(U_i-1)\right] = \boldsymbol{E}\left(U_i^2\right) - \boldsymbol{E}(U_i) = \alpha_2(i) - \alpha_1(i) \\
\upsilon_3(i) = \boldsymbol{E}\left[U_i(U_i-1)(U_i-2)\right] = \boldsymbol{E}\left(U_i^3\right) - 3\boldsymbol{E}\left(U_i^2\right) + 2\boldsymbol{E}(U_i) = \alpha_3(i) - 3\alpha_2(i) + 2\alpha_1(i) \\
\upsilon_4(i) = \boldsymbol{E}\left[U_i(U_i-1)(U_i-2)(U_i-3)\right] = \boldsymbol{E}\left(U_i^4\right) - 6\boldsymbol{E}\left(U_i^3\right) + 11\boldsymbol{E}\left(U_i^2\right) - 6\boldsymbol{E}(U_i) = \\
= \alpha_4(i) - 6\alpha_3(i) + 11\alpha_2(i) - 6\alpha_1(i) \\
\upsilon_5(i) = \boldsymbol{E}\left[U_i(U_i-1)(U_i-2)(U_i-3)(U_i-4)\right] = \boldsymbol{E}\left(U_i^5\right) - 10\boldsymbol{E}\left(U_i^4\right) + 35\boldsymbol{E}\left(U_i^3\right) - 50\boldsymbol{E}\left(U_i^2\right) + \\
+24\boldsymbol{E}(U_i) = \alpha_5(i) - 10\alpha_4(i) + 35\alpha_3(i) - 50\alpha_2(i) + 24\alpha_1(i) \\
\upsilon_6(i) = \boldsymbol{E}\left[U_i(U_i-1)(U_i-2)(U_i-3)(U_i-4)(U_i-5)\right] = \boldsymbol{E}\left(U_i^6\right) - 15\boldsymbol{E}\left(U_i^5\right) + 85\boldsymbol{E}\left(U_i^4\right) - \\
-225\boldsymbol{E}\left(U_i^3\right) + 274\boldsymbol{E}\left(U_i^2\right) - 120\boldsymbol{E}(U_i) = \alpha_6(i) - 15\alpha_5(i) + 85\alpha_4(i) - 225\alpha_3(i) + \\
+274\alpha_2(i) - 120\alpha_1(i)
\end{cases}
$$

Išsprendę šią lygčių sistemą, gauname:

$$\alpha_1(i) = \upsilon_1(i);$$

$$\alpha_2(i) = \upsilon_2(i) + \upsilon_1(i);$$

$$\alpha_3(i) = \upsilon_3(i) + 3\upsilon_2(i) + \upsilon_1(i);$$

$$\alpha_4(i) = \upsilon_4(i) + 6\upsilon_3(i) + 7\upsilon_2(i) + \upsilon_1(i);$$

$$\alpha_5(i) = \upsilon_5(i) + 10\upsilon_4(i) + 25\upsilon_3(i) + 15\upsilon_2(i) + \upsilon_1(i);$$

$$\alpha_6(i) = \upsilon_6(i) + 15\upsilon_5(i) + 65\upsilon_4(i) + 90\upsilon_3(i) + 31\upsilon_2(i) + \upsilon_1(i)$$

Pasinaudoję (3.12), gauname

$$
\begin{aligned}
&\alpha_1(i) = np_i; \\
&\alpha_2(i) = n^{[2]}p_i^2 + np_i; \\
&\alpha_3(i) = n^{[3]}p_i^3 + 3n^{[2]}p_i^2 + np_i; \\
&\alpha_4(i) = n^{[4]}p_i^4 + 6n^{[3]}p_i^3 + 7n^{[2]}p_i^2 + np_i; \\
&\alpha_5(i) = n^{[5]}p_i^5 + 10n^{[4]}p_i^4 + 25n^{[3]}p_i^3 + 15n^{[2]}p_i^2 + np_i; \\
&\alpha_6(i) = n^{[6]}p_i^6 + 15n^{[5]}p_i^5 + 65n^{[4]}p_i^4 + 90n^{[3]}p_i^3 + 31n^{[2]}p_i^2 + np_i.
\end{aligned}
\tag{3.13}
$$

Pereikime prie mišrių dviejų koordinačių momentų radimo. Iš (3.11) gauname:

$$\begin{cases} \upsilon_{11} = \alpha_{11} \\ \upsilon_{12}(i,j) = \boldsymbol{E}\left[U_i U_j \left(U_j - 1\right)\right] = \boldsymbol{E}\left[U_i U_j^2\right] - \boldsymbol{E}\left[U_i U_j\right] = \alpha_{12}(i,j) - \alpha_{11}(i,j) \\ \upsilon_{21}(i,j) = \boldsymbol{E}\left[U_i \left(U_i - 1\right) U_j\right] = \boldsymbol{E}\left[U_i^2 U_j\right] - \boldsymbol{E}\left[U_i U_j\right] = \alpha_{12}(i,j) - \alpha_{11}(i,j) \\ \upsilon_{22}(i,j) = \boldsymbol{E}\left[U_i \left(U_i - 1\right) U_j \left(U_j - 1\right)\right] = \boldsymbol{E}\left[U_i^2 U_j^2\right] - \boldsymbol{E}\left[U_i^2 U_j\right] \boldsymbol{-} \boldsymbol{E}\left[U_i U_j^2\right] + \boldsymbol{E}\left[U_i U_j\right] = \\ = \alpha_{22}(i,j) - \alpha_{21}(i,j) - \alpha_{12}(i,j) + \alpha_{11}(i,j) \\ \upsilon_{31}(i,j) = \boldsymbol{E}\left[U_i \left(U_i - 1\right)\left(U_i - 2\right) U_j\right] = \alpha_{31}(i,j) - 3\alpha_{21}(i,j) + 2\alpha_{11}(i,j) \\ \upsilon_{32}(i,j) = \boldsymbol{E}\left[U_i \left(U_i - 1\right)\left(U_i - 2\right) U_j \left(U_i - 1\right)\right] = \boldsymbol{E}\left[U_i^3 U_j^2\right] - \boldsymbol{E}\left[U_i^3 U_j\right] - 3\boldsymbol{E}\left[U_i^2 U_j^2\right] + \\ +3\boldsymbol{E}\left[U_i^2 U_j\right] + 2\boldsymbol{E}\left[U_i U_j^2\right] - 2\boldsymbol{E}\left[U_i U_j\right] = \alpha_{32}(i,j) - \alpha_{31}(i,j) - 3\alpha_{22}(i,j) + 3\alpha_{21}(i,j) + \\ +2\alpha_{12}(i,j) - 2\alpha_{11}(i,j) \\ \upsilon_{41}(i,j) = \boldsymbol{E}\left[U_i \left(U_i - 1\right)\left(U_i - 2\right)\left(U_i - 3\right) U_j\right] = \boldsymbol{E}\left[U_i^4 U_j\right] - 6\boldsymbol{E}\left[U_i^3 U_j\right] + 11\boldsymbol{E}\left[U_i^2 U_j\right] - \\ -6\boldsymbol{E}\left[U_i U_j\right] = \alpha_{41}(i,j) - 6\alpha_{31}(i,j) + 11\alpha_{21}(i,j) - 6\alpha_{11}(i,j) \\ \upsilon_{42}(i,j) = \boldsymbol{E}\left[U_i \left(U_i - 1\right)\left(U_i - 2\right)\left(U_i - 3\right) U_j \left(U_j - 1\right)\right] = \boldsymbol{E}\left[U_i^4 U_j^2\right] - \boldsymbol{E}\left[U_i^4 U_j\right] \boldsymbol{-6}\boldsymbol{E}\left[U_i^3 U_j^2\right] + \\ +6\boldsymbol{E}\left[U_i^3 U_j\right] + 11\boldsymbol{E}\left[U_i^2 U_j^2\right] \text{-} 11\boldsymbol{E}\left[U_i^2 U_j\right] \text{-} 6\boldsymbol{E}\left[U_i U_j^2\right] + 6\boldsymbol{E}\left[U_i U_j\right] = \alpha_{42}(i,j) - \alpha_{41}(i,j) - \\ -6\alpha_{32}(i,j) + 6\alpha_{31}(i,j) + 11\alpha_{22}(i,j) - 11\alpha_{21}(i,j) - 6\alpha_{12}(i,j) + 6\alpha_{11}(i,j) \end{cases}$$

ir išsprendę šią lygčių sistemą, gauname:

$$\alpha_{22}(i,j) = \nu_{22}(i,j) + \nu_{21}(i,j) + \nu_{12}(i,j) + \nu_{11}(i,j);$$

$$\alpha_{42}(i,j) = \nu_{42}(i,j) + \nu_{41}(i,j) + 6\nu_{31}(i,j) + 6\nu_{32}(i,j) + 7\nu_{22}(i,j) + 7\nu_{21}(i,j) + \nu_{12}(i,j) + \nu_{11}(i,j);$$

Pasinaudojame 3.12, gauname

$$\alpha_{22}(i,j) = n^{[4]} p_i^2 p_j^2 + n^{[3]} p_i^2 p_j + n^{[3]} p_i p_j^2 + n^{[2]} p_i p_j;$$

$$\alpha_{42}(i,j) = n^{[6]} p_i^4 p_j^2 + n^{[5]} p_i^4 p_j + 6n^{[4]} p_i^3 p_j + +6n^{[5]} p_i^3 p_j^2 + 7n^{[4]} p_i^2 p_j^2 + 7n^{[3]} p_i^2 p_j + n^{[3]} p_i p_j^2$$

$$+n^{[2]} p_i p_j.$$

Pagaliau, spręsdami lygčių sistemą

$$\begin{cases}
\upsilon_{111}(i,j,l)=\alpha_{111}(i,j,l)\\
\upsilon_{112}(i,j,l)=\boldsymbol{E}\left[U_iU_jU_l(U_l-1)\right]=\boldsymbol{E}\left[U_iU_jU_l^2\right]-\boldsymbol{E}\left[U_iU_jU_l\right]=\alpha_{112}(i,j,l)-\alpha_{111}(i,j,l)\\
\upsilon_{121}(i,j,l)=\boldsymbol{E}\left[U_iU_j(U_j-1)U_l\right]=\boldsymbol{E}\left[U_iU_j^2U_l\right]-\boldsymbol{E}\left[U_iU_jU_l\right]=\alpha_{121}(i,j,l)-\alpha_{111}(i,j,l)\\
\upsilon_{122}(i,j,l)=\boldsymbol{E}\left[U_iU_j(U_j-1)U_l(U_l-1)\right]=\boldsymbol{E}\left[U_iU_j^2U_l^2\right]-\boldsymbol{E}\left[U_iU_j^2U_l\right]-\boldsymbol{E}\left[U_iU_jU_l^2\right]+\\
+\boldsymbol{E}\left[U_iU_jU_l\right]=\alpha_{122}(i,j,l)-\alpha_{121}(i,j,l)-\alpha_{112}(i,j,l)+\alpha_{111}(i,j,l)\\
\upsilon_{211}(i,j,l)=\boldsymbol{E}\left[U_i(U_i-1)U_jU_l\right]=\alpha_{211}(i,j,l)-\alpha_{111}(i,j,l)\\
\upsilon_{212}(i,j,l)=\boldsymbol{E}\left[U_i(U_i-1)U_jU_l(U_l-1)\right]=\boldsymbol{E}\left[U_i^2U_jU_l^2\right]-\boldsymbol{E}\left[U_i^2U_jU_l\right]-\boldsymbol{E}\left[U_iU_jU_j^2\right]+\\
+\boldsymbol{E}\left[U_iU_jU_l\right]=\alpha_{212}(i,j,l)-\alpha_{211}(i,j,l)-\alpha_{112}(i,j,l)+\alpha_{111}(i,j,l)\\
\upsilon_{221}(i,j,l)=\boldsymbol{E}\left[U_i(U_i-1)U_j(U_j-1)U_l\right]=\boldsymbol{E}\left[U_i^2U_j^2U_l\right]-\boldsymbol{E}\left[U_i^2U_jU_l\right]-\boldsymbol{E}\left[U_iU_j^2U_l\right]+\\
+\boldsymbol{E}\left[U_iU_jU_l\right]=\alpha_{221}(i,j,l)-\alpha_{211}(i,j,l)-\alpha_{121}(i,j,l)+\alpha_{111}(i,j,l)\\
\upsilon_{222}(i,j,l)=\boldsymbol{E}\left[U_i(U_i-1)U_j(U_j-1)U_l(U_l-1)\right]=\boldsymbol{E}\left[U_i^2U_j^2U_l^2\right]-\boldsymbol{E}\left[U_i^2U_j^2U_l\right]-\boldsymbol{E}\left[U_i^2U_jU_l^2\right]+\\
+\boldsymbol{E}\left[U_i^2U_jU_l\right]-\boldsymbol{E}\left[U_iU_j^2U_l^2\right]+\boldsymbol{E}\left[U_iU_j^2U_l\right]+\boldsymbol{E}\left[U_iU_jU_l^2\right]-\boldsymbol{E}\left[U_iU_jU_l\right]=\alpha_{222}(i,j,l)-\\
-\alpha_{221}(i,j,l)-\alpha_{212}(i,j,l)+\alpha_{211}(i,j,l)-\alpha_{122}(i,j,l)+\alpha_{121}(i,j,l)+\alpha_{112}(i,j,l)-\alpha_{111}(i,j,l)
\end{cases}$$

randame

$$\alpha_{222}(i,j,l)=\nu_{222}(i,j,l)+\nu_{221}(i,j,l)+\nu_{211}(i,j,l)+\nu_{212}(i,j,l)+\nu_{122}(i,j,l)+\nu_{121}(i,j,l)+$$
$$\nu_{112}(i,j,l)+\nu_{111}(i,j,l)=n^{[6]}p_i^2p_j^2p_l^2+n^{[5]}p_i^2p_j^2p_l+n^{[4]}p_i^2p_jp_l+n^{[5]}p_i^2p_jp_l^2+n^{[5]}p_ip_j^2p_l^2+$$
$$+n^{[4]}p_ip_j^2p_l+n^{[4]}p_ip_jp_l^2+n^{[3]}p_ip_jp_l\,.$$

3.2.3. χ^2 statistikos vidurkis

Statistikos X_n^2 vidurkį rasime į išraišką (3.6) įsistatę gautas polinominio vektoriaus koordinačių antrų pradinių momentų reikšmes (3.13)

$$\boldsymbol{E}X_n^2=\sum_{i=1}^{k}\frac{n(n-1)p_i^2+np_i}{np_{i0}}-n=k-1+n\sum_{i=1}^{k}\frac{(p_i-p_{i0})^2}{p_{i0}}+\sum_{i=1}^{k}\frac{(p_i-p_{i0})(1-p_i)}{p_{i0}}.\quad(3.14).$$

3.2.4. χ^2 statistikos dispersija

Kaip jau minėjome 3.2.1. skyrelyje, statistikos X_n^2 dispersijai apskaičiuoti reikalingos $\boldsymbol{E}Z^2$ ir $(\boldsymbol{E}Z)^2$ reikšmės. Pradėsime nuo $\boldsymbol{E}Z^2$ ir įsistatę 3.2.2. skyrelyje apskaičiuotas $\alpha_4(i)$ ir $\alpha_{22}(i,j)$ reikšmes į (3.9) išraišką, gauname :

$$\boldsymbol{E}Z^2=\frac{1}{n^2}\sum_{i=1}^{k}\frac{n^{[4]}p_{i0}^4+6n^{[3]}p_{i0}^3+7n^{[2]}p_{i0}^2+np_{i0}}{p_{i0}^2}+\frac{1}{n^2}\sum_{i\neq j}\frac{n^{[4]}p_{i0}^2p_{j0}^2+n^{[3]}p_{i0}^2p_{j0}+n^{[3]}p_{i0}p_{j0}^2+n^{[2]}p_{i0}p_{j0}}{p_{i0}p_{j0}}$$

Iškeliame n prieš sumos ženklą:

$$\boldsymbol{E}Z^2=\frac{1}{n}\sum_{i=1}^{k}\frac{(n-1)^{[3]}p_i^4+6(n-1)^{[2]}p_i^3+7(n-1)p_i^2+p_i}{p_{i0}^2}+\frac{1}{n}\sum_{i\neq j}\frac{(n-1)^{[3]}p_i^2p_j^2+(n-1)^{[2]}p_i^2p_j}{p_{i0}p_{j0}}+$$

$$+\frac{1}{n}\sum_{i\neq j}\frac{(n-1)^{[2]}p_ip_j^2+(n-1)p_ip_j}{p_{i0}p_{j0}}. \tag{3.15}$$

Dėl paprastumo įveskime keletą pažymėjimų:

$$\frac{p_i}{p_{i0}}=\varepsilon_i,\ \frac{p_i^2}{p_{i0}}=\delta_i,\ \frac{p_i}{p_{i0}^2}=\sigma_i. \tag{3.16}$$

Remiantis (3.5) ir įvestu pažymėjimu $Z=\frac{1}{n}\sum_{i=1}^{k}\frac{U_i^2}{p_{i0}}$, turime $\boldsymbol{E}X_n^2=\boldsymbol{E}Z-n$, todėl iš (3.6) gauname, kad

$$\boldsymbol{E}Z=\sum_{i=1}^{k}\frac{(n-1)p_i^2+p_i}{p_{i0}}=(n-1)\sum_{i=1}^{k}\delta_i+\sum_{i=1}^{k}\varepsilon_i. \tag{3.17}$$

Pakelkime kvadratu:

$$(\boldsymbol{E}Z)^2=\left[\sum_{i=1}^{k}\left((n-1)\frac{p_i^2}{p_{i0}}+\frac{p_i}{p_{i0}}\right)\right]=\sum_{i=1}^{k}\left((n-1)\frac{p_i^2}{p_{i0}}+\frac{p_i}{p_{i0}}\right)^2+\sum_{i\neq j}\left((n-1)\frac{p_i^2}{p_{i0}}+\frac{p_i}{p_{i0}}\right)\times$$

$$\times\left((n-1)\frac{p_j^2}{p_{j0}}+\frac{p_j}{p_{j0}}\right)=(n-1)^2\sum_{i=1}^{k}\left(\frac{p_i^2}{p_{i0}}\right)^2+2(n-1)\sum_{i=1}^{k}\left(\frac{p_i}{p_{i0}}\right)\left(\frac{p_i^2}{p_{i0}}\right)+\sum_{i=1}^{k}\left(\frac{p_i}{p_{i0}}\right)^2+$$

$$+(n-1)^2\sum_{i\neq j}\left(\frac{p_i^2}{p_{i0}}\right)\left(\frac{p_j^2}{p_{j0}}\right)+2(n-1)\sum_{i\neq j}\left(\frac{p_i^2}{p_{i0}}\right)\left(\frac{p_i}{p_{i0}}\right)+\sum_{i\neq j}\left(\frac{p_i}{p_{i0}}\right)\left(\frac{p_j}{p_{j0}}\right). \tag{3.18}$$

Tuomet įsistatę gautas $\boldsymbol{E}Z^2$ ir $(\boldsymbol{E}Z)^2$ išraiškas (3.15) ir (3.18) į (3.7) ir gausime tokią a.d. Z dispersiją:

$$\boldsymbol{Var}(Z)=\frac{(n-1)^{[3]}}{n}\sum_{i=1}^{k}\delta_i^2+\frac{6(n-1)^{[2]}}{n}\sum_{i=1}^{k}\varepsilon_i\delta_i+\frac{7(n-1)}{n}\sum_{i=1}^{k}\varepsilon_i^2+\frac{1}{n}\sum_{i=1}^{k}\sigma_i+\frac{(n-1)^{[3]}}{n}\sum_{i\neq j}\delta_i\delta_j$$

$$+\frac{(n-1)^{[2]}}{n}\sum_{i\neq j}\delta_i\varepsilon_j+\frac{(n-1)^{[2]}}{n}\sum_{i\neq j}\varepsilon_i\delta_j+\frac{(n-1)}{n}\sum_{i\neq j}\varepsilon_i\varepsilon_j-(n-1)^2\sum_{i=1}^{k}\delta_i^2-2(n-1)\sum_{i=1}^{k}\varepsilon_i\delta_i-\sum_{i=1}^{k}\varepsilon_i^2-$$

$$-(n-1)^2\sum_{i\neq j}\delta_i\delta_j-2(n-1)\sum_{i\neq j}\varepsilon_i\delta_j-\sum_{i\neq j}\varepsilon_i\varepsilon_j.$$

Sutraukę kai kuriuos narius, gauname

$$\left[\frac{(n-1)^{[3]}}{n}-(n-1)^2\right]\left(\sum_{i=1}^{k}\delta_i\right)^2+\left[\frac{6(n-1)^{[2]}}{n}-2(n-1)\right]\sum_{i=1}^{k}\varepsilon_i\delta_i+\left[\frac{(n-1)}{n}-1\right]\left(\sum_{i=1}^{k}\varepsilon_i\right)^2+$$

$$+\left[\frac{2(n-1)^{[2]}}{n}-2(n-1)\right]\sum_{i\neq j}\varepsilon_i\delta_j+\frac{6(n-1)}{n}\sum_{i=1}^{k}\varepsilon_i^2+\frac{1}{n}\sum_{i=1}^{k}\sigma_i\ .$$

Kad nebeliktų nepilnų sumų, dar kartą sutraukiame narius ir tokią a.d. Z dispersiją

$$\boldsymbol{Var}(Z)=\frac{2(n-1)}{n}\left[2(n-2)\sum_{i=1}^{k}\varepsilon_i\delta_i-(2n-3)\left(\sum_{i=1}^{k}\delta_i\right)^2-2\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\right)+3\sum_{i=1}^{k}\varepsilon_i^2\right]-$$

$$-\frac{1}{n}\left[\left(\sum_{i=1}^{k}\varepsilon_i\right)^2-\sum_{i=1}^{k}\sigma_i\right]. \tag{3.19}$$

3.2.5. χ^2 statistikos trečiasis centrinis momentas

Į išraišką (3.10) įsistatome 3.2.2. skyrelyje gautas $\alpha_6(i)$, $\alpha_{42}(i,j)$ ir $\alpha_{222}(i,j,l)$ išraiškas:

$$\boldsymbol{E}Z^3=\frac{n^{[6]}}{n^3}\sum_{i=1}^{k}\left(\frac{p_i^2}{p_{i0}}\right)^3+\frac{15n^{[5]}}{n^3}\sum_{i=1}^{k}\left(\frac{p_i}{p_{i0}}\right)\left(\frac{p_i^2}{p_{i0}}\right)^2+\frac{65n^{[4]}}{n^3}\sum_{i=1}^{k}\left(\frac{p_i}{p_{i0}}\right)^2\left(\frac{p_i^2}{p_{i0}}\right)+\frac{90n^{[3]}}{n^3}\sum_{i=1}^{k}\left(\frac{p_i}{p_{i0}}\right)^3+$$

$$+\frac{31n^{[2]}}{n^3}\sum_{i=1}^{k}\left(\frac{p_i}{p_{i0}}\right)\left(\frac{p_i}{p_{i0}^2}\right)+\frac{1}{n^3}\sum_{i=1}^{k}\left(\frac{p_i}{p_{i0}^3}\right)+\frac{3n^{[6]}}{n^3}\sum_{i\neq j}\left(\frac{p_i^2}{p_{i0}}\right)^2\left(\frac{p_j^2}{p_{j0}}\right)+\frac{3n^{[5]}}{n^3}\sum_{i\neq j}\left(\frac{p_i^2}{p_{i0}}\right)^2\left(\frac{p_j}{p_{j0}}\right)+$$

$$+\frac{18n^{[4]}}{n^3}\sum_{i\neq j}\left(\frac{p_i^3}{p_{i0}^2}\right)\left(\frac{p_j}{p_{j0}}\right)+\frac{18n^{[5]}}{n^3}\sum_{i\neq j}\left(\frac{p_i^3}{p_{i0}^2}\right)\left(\frac{p_j^2}{p_{j0}}\right)+\frac{21n^{[4]}}{n^3}\sum_{i\neq j}\left(\frac{p_i}{p_{i0}}\right)^2\left(\frac{p_j^2}{p_{j0}}\right)+$$

$$+\frac{21n^{[3]}}{n^3}\sum_{i\neq j}\left(\frac{p_i}{p_{i0}}\right)^2\left(\frac{p_j}{p_{j0}}\right)+\frac{3n^{[3]}}{n^3}\sum_{i\neq j}\left(\frac{p_i}{p_{i0}^2}\right)\left(\frac{p_j^2}{p_{j0}}\right)+\frac{3n^{[2]}}{n^3}\sum_{i\neq j}\left(\frac{p_i}{p_{i0}^2}\right)\left(\frac{p_j}{p_{j0}}\right)+$$

$$+\frac{n^{[6]}}{n^3}\sum_{i\neq j\neq l}\left(\frac{p_i^2}{p_{i0}}\right)\left(\frac{p_j^2}{p_{j0}}\right)\left(\frac{p_l^2}{p_{l0}}\right)+\frac{n^{[5]}}{n^3}\sum_{i\neq j\neq l}\left(\frac{p_i^2}{p_{i0}}\right)\left(\frac{p_j^2}{p_{j0}}\right)\left(\frac{p_l}{p_{l0}}\right)+\frac{n^{[4]}}{n^3}\sum_{i\neq j\neq l}\left(\frac{p_i^2}{p_{i0}}\right)\left(\frac{p_j}{p_{j0}}\right)\left(\frac{p_l}{p_{l0}}\right)+$$

$$+\frac{n^{[4]}}{n^3}\sum_{i\neq j\neq l}\left(\frac{p_i}{p_{i0}}\right)\left(\frac{p_j^2}{p_{j0}}\right)\left(\frac{p_l}{p_{l0}}\right)+\frac{n^{[5]}}{n^3}\sum_{i\neq j\neq l}\left(\frac{p_i^2}{p_{i0}}\right)\left(\frac{p_j}{p_{j0}}\right)\left(\frac{p_l^2}{p_{l0}}\right)+\frac{n^{[5]}}{n^3}\sum_{i\neq j\neq l}\left(\frac{p_i}{p_{i0}}\right)\left(\frac{p_j^2}{p_{j0}}\right)\left(\frac{p_l^2}{p_{l0}}\right)+$$

$$+\frac{n^{[4]}}{n^3}\sum_{i\neq j\neq l}\left(\frac{p_i}{p_{i0}}\right)\left(\frac{p_j}{p_{j0}}\right)\left(\frac{p_l^2}{p_{l0}}\right)++\frac{n^{[3]}}{n^3}\sum_{i\neq j\neq l}\left(\frac{p_i}{p_{i0}}\right)\left(\frac{p_j}{p_{j0}}\right)\left(\frac{p_l}{p_{l0}}\right).$$

Naudodami ankstesnius pažymėjimus ir papildomai įvedę

$$\frac{p_i}{p_{i0}^3}=\xi_i \tag{3.20}$$

bei šiek tiek supaprastinę reiškinį, gauname:

$$\boldsymbol{E}Z^3=\frac{n^{[6]}}{n^3}\left(\sum_{i=1}^{k}\delta_i\right)^3+\frac{n^{[3]}}{n^3}\left(\sum_{i=1}^{k}\varepsilon_i\right)^3+\frac{89n^{[3]}}{n^3}\sum_{i=1}^{k}\varepsilon_i^3+\frac{18n^{[3]}}{n^3}\sum_{i\neq j}\varepsilon_i^2\varepsilon_j+\frac{15n^{[5]}}{n^3}\sum_{i=1}^{k}\delta_i^2\varepsilon_i+$$

$$+\frac{65n^{[4]}}{n^3}\sum_{i=1}^{k}\varepsilon_i^2\delta_i+\frac{31n^{[2]}}{n^3}\sum_{i=1}^{k}\varepsilon_i\sigma_i+\frac{1}{n^3}\sum_{i=1}^{k}\xi_i+\frac{3n^{[5]}}{n^3}\sum_{i\neq j}\delta_i^2\varepsilon_j+\frac{18n^{[4]}}{n^3}\sum_{i\neq j}\left(\varepsilon_i\delta_i\right)\varepsilon_j+$$

$$+\frac{18n^{[5]}}{n^3}\sum_{i\neq j}\left(\varepsilon_i\delta_i\right)\delta_j+\frac{21n^{[4]}}{n^3}\sum_{i\neq j}\varepsilon_i^2\delta_j+\frac{3n^{[3]}}{n^3}\sum_{i\neq j}\sigma_i\delta_j+\frac{3n^{[2]}}{n^3}\sum_{i\neq j}\sigma_i\varepsilon_j+\frac{3n^{[5]}}{n^3}\sum_{i\neq j\neq j}\delta_i\delta_j\varepsilon_l+$$

$$+\frac{3n^{[4]}}{n^3}\sum_{i\neq j\neq l}\delta_i\varepsilon_j\varepsilon_l\,.$$

Apskaičiuojame antrąjį trečiojo centrinio momento išraiškos (3.8) dėmenį sudaugindami turimas -3***Var(Z)*** (3.19) ir ***E****Z* (3.17) reikšmes

$$-3\boldsymbol{Var}\left(Z\right)\cdot\boldsymbol{E}Z=\left[-\frac{6\left(n-1\right)}{n}\left(2\left(n-2\right)\sum_{i=1}^{k}\varepsilon_i\delta_i-\left(2n-3\right)\left(\sum_{i=1}^{k}\delta_i\right)^2-2\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\right)+3\sum_{i=1}^{k}\varepsilon_i^2\right)+\right.$$

$$\left.+\frac{3}{n}\left(\left(\sum_{i=1}^{k}\varepsilon_i\right)^2-\sum_{i=1}^{k}\sigma_i\right)\right]\cdot\left[\sum_{i}^{k}\varepsilon_i+\left(n-1\right)\sum_{i=1}^{k}\delta_i\right]=-\frac{12\left(n-1\right)^{[2]}}{n}\left[\sum_{i=1}^{k}\delta_i\varepsilon_i^2+\sum_{i\neq j}\varepsilon_i\left(\varepsilon_j\delta_j\right)+\right.$$

$$\left.+\left(n-1\right)\left(\sum_{i=1}^{k}\delta_i^2\varepsilon_i+\sum_{i\neq j}\delta_i\left(\varepsilon_j\delta_j\right)\right)\right]+\frac{6\left(n-1\right)\left(2n-3\right)}{n}\left[\sum_{i=1}^{k}\delta_i^2\varepsilon_i+\sum_{i\neq j}\delta_i^2\varepsilon_j+2\sum_{i\neq j}\left(\varepsilon_i\delta_i\right)\delta_j+\right.$$

$$\left.+\sum_{i\neq j\neq l}\delta_i\delta_j\varepsilon_l+\left(n-1\right)\left(\sum_{i=1}^{k}\delta_i^3+\sum_{i\neq j}\delta_i\delta_j^2+2\sum_{i\neq j}\delta_i^2\delta_j+\sum_{i\neq j\neq l}\delta_i\delta_j\delta_l\right)\right]-\frac{18\left(n-1\right)}{n}\left[\sum_{i=1}^{k}\varepsilon_i^3+\sum_{i\neq j}\varepsilon_i^2\varepsilon_j+\right.$$

$$\left.+\left(n-1\right)\left(\sum_{i=1}^{k}\varepsilon_i^2\delta_i+\sum_{i\neq j}\delta_i\varepsilon_j^2\right)\right]+\frac{12\left(n-1\right)}{n}\left[\sum_{i=1}^{k}\varepsilon_i^2\delta_i+\sum_{i\neq j}\varepsilon_i^2\delta_j+2\sum_{i\neq j}\left(\varepsilon_i\delta_i\right)\varepsilon_j+\sum_{i\neq j\neq l}\varepsilon_i\varepsilon_j\delta_l+\right.$$

$$\left.+\left(n-1\right)\left(\sum_{i=1}^{k}\delta_i^2\varepsilon_i+\sum_{i\neq j}\delta_i\left(\varepsilon_j\delta_j\right)+\sum_{i\neq j}\delta_i^2\varepsilon_j+\sum_{i\neq j}\left(\varepsilon_i\delta_i\right)\delta_j+\sum_{i\neq j\neq l}\delta_i\varepsilon_j\delta_l\right)\right]+\frac{3}{n}\left[\left(\sum_{i=1}^{k}\varepsilon_i\right)^3+\right.$$

$$\left.+\left(n-1\right)\left(\sum_{i=1}^{k}\varepsilon_i^2\delta_i+\sum_{i\neq j}\varepsilon_i^2\delta_j+2\sum_{i\neq j}\varepsilon_i\left(\varepsilon_j\delta_j\right)+\sum_{i\neq j\neq l}\varepsilon_i\varepsilon_j\delta_l\right)\right]-\frac{3}{n}\left[\sum_{i=1}^{k}\varepsilon_i\sigma_i+\sum_{i\neq j}\varepsilon_i\sigma_j+\right.$$

$$\left.+\left(n-1\right)\left(\sum_{i=1}^{k}\varepsilon_i^3+\sum_{i\neq j}\delta_i\sigma_j\right)\right].$$

Galiausiai apskaičiuojame $-\left(\boldsymbol{E}Z\right)^3$:

$$-\left(\boldsymbol{E}Z\right)^3=-\sum_{i=1}^{k}\varepsilon_i^3-\sum_{i\neq j}\varepsilon_i\varepsilon_j^2-2\left(n-1\right)\sum_{i=1}^{k}\varepsilon_i^2\delta_i-2\left(n-1\right)\sum_{i\neq j}\varepsilon_i\left(\varepsilon_j\delta_j\right)-\left(n-1\right)^2\sum_{i=1}^{k}\delta_i^2\varepsilon_i-$$

$$-\left(n-1\right)^2\sum_{i\neq j}\varepsilon_i\delta_j^2-2\sum_{i\neq j}\varepsilon_i^2\varepsilon_j-\sum_{i\neq j\neq l}\varepsilon_i\varepsilon_j\varepsilon_l-2\left(n-1\right)\sum_{i\neq j}\varepsilon_i^2\delta_j-2\left(n-1\right)\sum_{i\neq j}\varepsilon_i\left(\varepsilon_j\delta_j\right)-$$

$$-2\left(n-1\right)\sum_{i\neq j\neq l}\varepsilon_i\delta_j\varepsilon_l-2\left(n-1\right)^2\sum_{i\neq j}\left(\varepsilon_i\delta_i\right)\delta_j-\left(n-1\right)^2\sum_{i\neq j\neq l}\varepsilon_i\delta_j\delta_l-\left(n-1\right)\sum_{i=1}^{k}\varepsilon_i^2\delta_i-$$

$$-\left(n-1\right)\sum_{i\neq j}\varepsilon_i^2\delta_j-2\left(n-1\right)^2\sum_{i=1}^{k}\varepsilon_i\delta_i^2-2\left(n-1\right)^2\sum_{i\neq j}\left(\varepsilon_i\delta_i\right)\delta_j-\left(n-1\right)^3\sum_{i=1}^{k}\delta_i^3-\left(n-1\right)^3\sum_{i\neq j}\delta_i^2\delta_j-$$

$$-2(n-1)\sum_{i\neq j}(\varepsilon_i\delta_i)\varepsilon_j-(n-1)\sum_{i\neq j\neq l}\varepsilon_i\varepsilon_j\delta_l-2(n-1)^2\sum_{i\neq j}(\varepsilon_i\delta_i)\delta_j-2(n-1)^2\sum_{i\neq j}\varepsilon_i\delta_i^2-$$
$$-2(n-1)^2\sum_{i\neq j\neq l}\varepsilon_i\delta_j\delta_l-2(n-1)^3\sum_{i\neq j}\delta_i^2\delta_i-(n-1)^3\sum_{i\neq j\neq l}\delta_i\delta_j\delta_l.$$

Gautas $\boldsymbol{E}Z^3$, $-3\boldsymbol{Var}(Z)\cdot\boldsymbol{E}Z$, $-(\boldsymbol{E}Z)^3$ išraiškas įsistatome į (3.8), ir pasinaudoję galiojančiu sąryšiu $\left(\sum_{i=1}^{k}\eta_i\right)^3=\sum_{i=1}^{k}\eta_i^3+3\sum_{i\neq j}\eta_i^2\eta_i+\sum_{i\neq j\neq l}\eta_i\eta_j\eta_l$, sutraukiame narius:

$$\boldsymbol{\mu_3}(Z)=\frac{n^{[6]}}{n^3}\left(\sum_{i=1}^{k}\delta_i\right)^3+\left(\sum_{i=1}^{k}\varepsilon_i\right)^3\left[\frac{n^{[3]}}{n^3}+\frac{3}{n}\right]+\left(\sum_{i=1}^{k}\varepsilon_i^3\right)\left[-\frac{18(n-1)}{n}-\frac{3(n-1)}{n}+\frac{89n^{[3]}}{n^3}-1\right]+$$
$$+\left(\sum_{i=1}^{k}\delta_i^3\right)\left[\frac{6(n-1)^2(2n-3)}{n}-(n-1)^3\right]+\left(\sum_{i=1}^{k}\varepsilon_i^2\delta_i\right)\left[-\frac{12(n-1)^{[2]}}{n}-\frac{18(n-1)^2}{n}+\frac{12(n-1)}{n}+\right.$$
$$\left.\frac{65n^{[4]}}{n^3}-2(n-1)-(n-1)+\frac{3(n-1)}{n}\right]+\left(\sum_{i=1}^{k}\varepsilon_i\delta_i^2\right)\left[-\frac{12(n-1)(n-1)^{[2]}}{n}+\frac{6(n-1)(2n-3)}{n}+\right.$$
$$\left.\frac{12(n-1)^2}{n}+\frac{15n^{[5]}}{n^3}-3(n-1)^2\right]+\left(\sum_{i=1}^{k}\varepsilon_i\sigma_i\right)\left[-\frac{3}{n}+\frac{31n^{[2]}}{n^3}\right]+\left(\sum_{i=1}^{k}\xi_i\right)\frac{1}{n^3}+\left(\sum_{i\neq j}\delta_i^2\varepsilon_j\right)\times$$
$$\left[\frac{6(n-1)(2n-3)}{n}+\frac{12(n-1)^2}{n}+\frac{3n^{[5]}}{n^3}-(n-1)^2-2(n-1)^2\right]+\left(\sum_{i\neq j}\varepsilon_i^2\delta_j\right)\left[-\frac{18(n-1)^2}{n}+\right.$$
$$\left.+\frac{12(n-1)}{n}+\frac{3(n-1)}{n}+\frac{21n^{[4]}}{n^3}-2(n-1)-(n-1)\right]+\left(\sum_{i\neq j}\varepsilon_i^2\varepsilon_j\right)\left[-\frac{18(n-1)}{n}+\frac{18n^{[3]}}{n^3}-1-2\right]+$$
$$\left(\sum_{i\neq j}\delta_i^2\delta_j\right)\left[\frac{6(n-1)^2(2n-3)}{n}+\frac{12(n-1)^2(2n-3)}{n}-(n-1)^3-2(n-1)^3\right]+$$
$$+\left(\sum_{i\neq j}\varepsilon_i\sigma_j\right)\left[-\frac{3}{n}+\frac{3n^{[2]}}{n^3}\right]+\left(\sum_{i\neq j}\delta_i\sigma_j\right)\left[-\frac{3(n-1)}{n}+\frac{3n^{[3]}}{n^3}\right]+\left(\sum_{i\neq j}\varepsilon_i\left(\varepsilon_j\delta_j\right)\right)\left[-\frac{12(n-1)^{[2]}}{n}+\right.$$
$$\left.+\frac{18n^{[4]}}{n^3}+\frac{24(n-1)}{n}+\frac{6(n-1)}{n}-2(n-1)-2(n-1)-2(n-1)\right]+\left(\sum_{i\neq j}\delta_i\left(\varepsilon_j\delta_j\right)\right)\times$$
$$\times\left[-\frac{12(n-1)(n-1)^{[2]}}{n}+\frac{12(n-1)(2n-3)}{n}+\frac{24(n-1)^2}{n}+\frac{18n^{[5]}}{n^3}-4(n-1)^2-2(n-1)^2\right]+$$
$$+\left(\sum_{i\neq j\neq l}\delta_i\delta_j\varepsilon_l\right)\left[\frac{6(n-1)(2n-3)}{n}+\frac{12(n-1)^2}{n}+\frac{3n^{[5]}}{n^3}-(n-1)^2-2(n-1)^2\right]+\left(\sum_{i\neq j\neq l}\varepsilon_i\varepsilon_j\delta_l\right)\times$$
$$\times\left[\frac{15(n-1)}{n}-2(n-1)-(n-1)+\frac{3n^{[4]}}{n^3}\right]+\left(\sum_{i\neq j\neq l}\delta_i\delta_j\delta_l\right)\left[\frac{6(n-1)^2(2n-3)}{n}-(n-1)^3\right]-\left(\sum_{i\neq j\neq l}\varepsilon_i\varepsilon_j\varepsilon_l\right).$$

Atliekame tarpinius skaičiavimus:

1) $\frac{n^{[3]}}{n^3}+\frac{3}{n}=\frac{n(n-1)(n-2)+3n^2}{n^3}=\frac{n^2+2}{n^2}$;

2) $-\frac{21(n-1)}{n}+\frac{89n(n-1)(n-2)}{n^3}-1=\frac{-21n^3+21n^2+89n^3-267n^2+178n-n^3}{n^3}=$

$=\frac{67n^2-246n+178}{n^2}$;

3) $\frac{6(n-1)^2(2n-3)}{n}-(n-1)^3=\frac{(n-1)^2(12n-18-n^2+n)}{n}=$

$=\frac{13n^3-n^4-18n^2-26n^2+2n^3+36n+13n-n^2-18}{n}=\frac{15n^3-n^4-45n^2+49n-18}{n}$;

4) $-\frac{12(n-1)^{[2]}}{n}-\frac{18(n-1)^2}{n}+\frac{12(n-1)}{n}+\frac{65n^{[4]}}{n^3}-3(n-1)+\frac{3(n-1)}{n}=\frac{-12n^2(n^2-3n+2)}{n^3}-$

$-\frac{18n^2(n^2-2n+1)+12n^3-12n^2+65n(n^2-3n+2)(n-3)-3n^4+3n^3+3n^2(n-1)}{n^3}=$

$=\frac{-12n^4+36n^3-24n^2-18n^4+36n^3-18n^2+12n^3-12n^2+65n^4-195n^3-195n^3+585n^2}{n^3}+$

$+\frac{130n^2-390n-3n^4+3n^3+3n^3-3n^2+32n^4-303n^3+661n^2-390n}{n^3}=$

$=\frac{32n^3-300n^2+658n-390}{n^2}$;

5) $-\frac{12(n-1)^2(n-2)}{n}+\frac{6(n-1)(2n-3)}{n}+\frac{12(n-1)^2}{n}+\frac{15n(n-1)(n-2)(n-3)(n-4)}{n}-$

$-3(n^2-2n+1)=\frac{-12n^2(n^3-4n^2+5n-2)+(12n^2-30n+18)n^2+(12n^2-24n+12)n^2}{n^3}+$

$+\frac{15n(n^2-3n+2)(n^2-7n^2+12)-3n^5+6n^4-3n^3}{n^3}=-\frac{72n^3-408n^2+696n-360}{n^3}$;

6) $-\frac{3}{n}+\frac{31n^{[2]}}{n^3}=-\frac{3n^2+31n(n-1)}{n^3}=\frac{28n-31}{n^2}$;

7) $\frac{6(n-1)(2n-3)}{n}+\frac{12(n-1)^2}{n}+\frac{3n^{[5]}}{n^3}-3(n-1)^2=\frac{6(n-1)(2n-3+2n-2)}{n}+$

$+\frac{3n(n-1)(n-2)(n-3)(n-4)}{n^3}-3(n-1)^2=\frac{6n^2(4n^2-9n+5)+3n(n^4-7n^3+12n^2-3n^3)}{n^3}+$

$+\frac{3n(21n^2-36n+2n^2-14n+24)-3n^3(n^2-2n+1)}{n^3}=\frac{48n^2-120n+72}{n^2}$;

$$8)\ -\frac{18(n-1)^2}{n}+\frac{12(n-1)}{n}+\frac{3(n-1)}{n}+\frac{21n^{[4]}}{n^3}-3(n-1)=\frac{-18n^2\left(n^2-2n+1\right)+15n^2(n-1)}{n^3}+$$

$$+\frac{21n(n-1)(n-2)(n-3)-3n^3(n-1)}{n^3}=\frac{-18n^4+36n^3-18n^2+15n^3-15n^2}{n^3}+$$

$$+\frac{\left(21n^2-21n\right)\left(n^2-5n+6\right)-3n^4+3n^3}{n^3}=-\frac{72n^2-198n+126}{n^2};$$

$$9)\ -\frac{18(n-1)}{n^2}+\frac{18n^{[3]}}{n^3}-3=\frac{-18n^3+18n^2+18n\left(n^2-3n+2\right)-3n^3}{n^3}=-\frac{3n^2+36n-36}{n^3};$$

$$10)\ \frac{6(n-1)^2(2n-3)}{n}+\frac{12(n-1)^2(2n-3)}{n}-3(n-1)^3=\frac{18\left(2n^3-3n^2-4n^2+6n+2n-3\right)}{n}-$$

$$-\frac{\left(3n^3-9n^2+9n-3\right)n}{n}=-\frac{3n^4-45n^3+135n^2-147n+54}{n};$$

$$11)\ -\frac{3}{n}+\frac{3n^{[2]}}{n^3}=\frac{-3n^2+3n^2-3n}{n^3}=-\frac{3}{n^2};$$

$$12)\ \frac{-3(n-1)}{n}+\frac{3n^{[3]}}{n^3}=\frac{-3n^3+3n^2+3n\left(n^2-3n+2\right)}{n^3}=-\frac{6n-6}{n^2};$$

$$13)\ -\frac{12(n-1)^{[2]}}{n}+\frac{18n^{[4]}}{n^3}+\frac{30(n-1)}{n}-6(n-1)=\frac{-12n^2\left(n^2-3n+2\right)}{n^3}+$$

$$+\frac{18n\left(n^2-3n+2\right)(n-3)+30n^3-30n^2-6n^4+6n^3}{n^3}=\frac{-12n^4+36n^3-24n^2+18n\left(n^3-3n^2\right)}{n^3}-$$

$$-\frac{18n\left(3n^2+9n+2n-6\right)+36n^3-30n^2-6n^4}{n^3}=-\frac{36n^2-144n+108}{n^2};$$

$$14)\ -\frac{12(n-1)(n-1)^{[2]}}{n}+\frac{12(n-1)(2n-3)}{n}+\frac{24(n-1)^2}{n}+\frac{18n^{[5]}}{n^3}-6(n-1)^2=$$

$$=\frac{-12n^2(n-1)\left(n^2-3n+2\right)+12n^2\left(2n^2-5n+3\right)+24n^2\left(n^2-2n+1\right)}{n^3}+$$

$$+\frac{+18n\left(n^2-3n+2\right)\left(n^2-7n+12\right)-6n^3\left(n^2-2n+1\right)}{n^3}=\frac{-12n^2\left(n^3-3n^2+2n-n^2+3n-2\right)}{n^3}+$$

$$+\frac{24n^4-60n^3+36n^2+24n^4-48n^3+24n^2+18n\left(n^4-7n^3+12n^2-3n^3+21n^2-36n+2n^2\right)}{n^3}-$$

$$-\frac{18n(14n-24)-6n^5+12n^4-6n^3}{n^3}=-\frac{72n^3-456n^2+816n-432}{n^2};$$

15) $\frac{6(n-1)(2n-3)}{n}+\frac{12(n-1)^2}{n}+\frac{3n^{[5]}}{n^3}-(n-1)^2-2(n-1)^2=\frac{6n^2\left(2n^2-5n+3\right)}{n^3}+$

$+\frac{12n^4-24n^3+12n^2+3n\left(n^2-3n+2\right)\left(n^2-7n+12\right)-3n^5+6n^4-3n^3}{n^3}=\frac{48n^2-120n+72}{n^2};$

16) $\frac{15(n-1)}{n}-3(n-1)+\frac{3n^{[4]}}{n^3}=\frac{15n^3-15n^2-3n^4+3n^3+3n\left(n^2-3n+2\right)(n-3)}{n^3}=$

$=\frac{15n^3-15n^2-3n^4+3n^3+3n\left(n^3-3n^2-3n^2+9n+2n-6\right)}{n^3}=\frac{18n-18}{n^3};$

17) $\frac{6(n-1)^2(2n-3)}{n}-(n-1)^3=\frac{(n-1)^2\left(12n-18-n^2+n\right)}{n}=\frac{\left(n^2-2n+1\right)\left(-n^2+13n-18\right)}{n}=$

$=-\frac{n^4-15n^3+45n^2-49n+18}{n};$

Įrašome gautas reikšmes:

$$\boldsymbol{\mu_3}(Z)=\frac{n^{[6]}}{n^3}\left(\sum_{i=1}^{k}\delta_i\right)^3+\frac{n^2+2}{n^2}\left(\sum_{i=1}^{k}\varepsilon_i\right)^3+\frac{67n^2-246n+178}{n^2}\sum_{i=1}^{k}\varepsilon_i^3+$$

$$\frac{-n^4+15n^3-45n^2+49n-18}{n}\sum_{i=1}^{k}\delta_i^3+\frac{32n^3-300n^2+658n-390}{n^2}\sum_{i=1}^{k}\varepsilon_i^2\delta_i+$$

$$\frac{-72n^3+408n^2-696n+360}{n^2}\sum_{i=1}^{k}\varepsilon_i\delta_i^2+\frac{28n-31}{n}\sum_{i=1}^{k}\varepsilon_i\sigma_i+\frac{1}{n^3}\sum_{i=1}^{k}\xi_i+\frac{48n^2-120n+72}{n^2}\sum_{i\neq j}\delta_i^2\varepsilon_j+$$

$$\frac{-72n^2+198n-126}{n^2}\sum_{i\neq j}\varepsilon_i^2\delta_j+\frac{-3n^2-36n+36}{n^2}\sum_{i\neq j}\varepsilon_i^2\varepsilon_j+\frac{-3n^4+45n^3-135n^2+147n-54}{n}\times$$

$$\times\sum_{i\neq j}\delta_i^2\delta_j-\frac{3}{n^2}\sum_{i\neq j}\varepsilon_i\sigma_j+\frac{-6n+6}{n^2}\sum_{i\neq j}\delta_i\sigma_j+\frac{-36n^2+144n-108}{n^2}\sum_{i\neq j}\varepsilon_i\left(\varepsilon_j\delta_j\right)+$$

$$\frac{-72n^3+456n^2-816n+432}{n^2}\sum_{i\neq j}\delta_i\left(\varepsilon_j\delta_j\right)+\frac{48n^2-120n+72}{n^2}\sum_{i\neq j\neq l}\delta_i\delta_j\varepsilon_l+\frac{18n-18}{n^2}\sum_{i\neq j\neq l}\varepsilon_i\varepsilon_j\delta_l+$$

$$+\frac{-n^4+15n^3-45n^2+49n-18}{n^2}\sum_{i\neq j\neq l}\delta_i\delta_j\delta_l-\sum_{i\neq j\neq l}\varepsilon_i\varepsilon_j\delta_l\,.$$

Pritaikome sąryšį $\left(\sum_{i=1}^{k}\eta_i\right)^2\sum_{i=1}^{k}\tau_i=\sum_{i=1}^{k}\eta_i^2\tau_i+\sum_{i\neq j}\eta_i^2\tau_j+2\sum_{i\neq j}\left(\eta_i\tau_i\right)\tau_j+\sum_{i\neq j\neq l}\eta_i\eta_j\tau_l$ ir dar labiau supaprastiname reiškinį:

$$\boldsymbol{\mu_3}(Z)=\frac{n^{[6]}}{n^3}\left(\sum_{i=1}^{k}\delta_i\right)^3+\frac{n^2+2}{n^2}\left(\sum_{i=1}^{k}\varepsilon_i\right)^3-\left(\sum_{i=1}^{k}\varepsilon_i\right)^3-\frac{6n-6}{n^2}\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\sigma_i\right)+$$

$$+\frac{68n^2-240n+172}{n^2}\sum_{i=1}^{k}\varepsilon_i^3-\frac{36n-36}{n^2}\sum_{i\neq j}\varepsilon_i^2\varepsilon_j-\frac{n^4-15n^3+45n^2-49n+18}{n}\left(\sum_{i=1}^{k}\delta_i\right)^3+$$

$$+\frac{48n^2-120n+72}{n^2}\left(\sum_{i=1}^{k}\delta_i\right)^2\sum_{i=1}^{k}\varepsilon_i-\frac{72n^3-360n^2+576n-288}{n^2}\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\delta_i\right)+$$

$$+\frac{18n-18}{n^2}\left(\sum_{i=1}^{k}\varepsilon_i\right)^2\sum_{i=1}^{k}\delta_i-\frac{36n^2-108n+72}{n^2}\left(\sum_{i=1}^{k}\varepsilon_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\delta_i\right)+\frac{32n^3-264n^2+532n-300}{n^2}\times$$

$$\times\sum_{i=1}^{k}\varepsilon_i^2\delta_i-\frac{72n^2-180n+108}{n^2}\sum_{i\neq j}\varepsilon_i^2\delta_j+\frac{1}{n^3}\sum_{i=1}^{k}\xi_i-\frac{3}{n^2}\left(\sum_{i=1}^{k}\varepsilon_i\right)\left(\sum_{i=1}^{k}\sigma_i\right)+\frac{28n-28}{n^2}\sum_{i=1}^{k}\varepsilon_i\sigma_i\ .$$

Apskaičiuojame koeficientą prie $\left(\sum_{i=1}^{k}\delta_i\right)^3$:

$$\frac{n^{[6]}}{n^3}-\frac{n^4-15n^3+45n^2-49n+18}{n}=\frac{n\left(n^2-3n+2\right)\left(n^2-7n+12\right)\left(n-5\right)-n^6+15n^5-45n^4}{n^3}+$$

$$+\frac{49n^3-18n^2}{n^3}=\frac{n\left(n^4-7n^3+12n^2-3n^3+21n^2-36n+2n^2-14n+24\right)\left(n-5\right)-n^6+15n^5}{n^3}-$$

$$-\frac{45n^4+49n^3-18n^2}{n^3}=\frac{n\left(n^4-10n^3+35n^2-50n+24\right)\left(n-5\right)-n^6+15n^5-45n^4+49n^3}{n^3}-$$

$$-\frac{18n^2}{n^3}=\frac{40n^3-176n^2+256n-120}{n^2}\ .$$

Dar kartą, naudodamiesi sąryšiu $\left(\sum_{i=1}^{k}\eta_i\tau_i\right)\sum_{i=1}^{k}\tau_i=\sum_{i=1}^{k}\eta_i\tau_i^2+\sum_{i\neq j}\left(\eta_i\tau_i\right)\tau_j$, supaprastiname reiškinį:

$$\boldsymbol{\mu_3}\left(Z\right)=\frac{40n^3-176n^2+256n-120}{n^2}\left(\sum_{i=1}^{k}\delta_i\right)^3+\frac{2}{n^2}\left(\sum_{i=1}^{k}\varepsilon_i\right)^3-\frac{6n-6}{n^2}\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\sigma_i\right)-\frac{36n-36}{n^2}\times$$

$$\times\left(\sum_{i=1}^{k}\varepsilon_i^2\right)\left(\sum_{i=1}^{k}\varepsilon_i\right)+\frac{68n^2-204n+136}{n^2}\sum_{i=1}^{k}\varepsilon_i^3+\frac{18n-18}{n^2}\left(\sum_{i=1}^{k}\varepsilon_i\right)^2\left(\sum_{i=1}^{k}\delta_i\right)-\frac{36n^2-108n+72}{n^2}\times$$

$$\times\left(\sum_{i=1}^{k}\varepsilon_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\delta_i\right)-\frac{72n^2-180n+108}{n^2}\left(\sum_{i=1}^{k}\varepsilon_i^2\right)\left(\sum_{i=1}^{k}\delta_i\right)+\frac{32n^3-192n^2+352n-192}{n^2}\sum_{i=1}^{k}\varepsilon_i^2\delta_i+$$

$$+\frac{1}{n^3}\left(\sum_{i=1}^{k}\xi_i\right)-\frac{3}{n^2}\left(\sum_{i=1}^{k}\varepsilon_i\right)\left(\sum_{i=1}^{k}\sigma_i\right)+\frac{28n-28}{n^2}\sum_{i=1}^{k}\varepsilon_i\sigma_i+\frac{48n^2-120n+72}{n^2}\left(\sum_{i=1}^{k}\delta_i\right)^2\left(\sum_{i=1}^{k}\varepsilon_i\right)-$$

$$-\frac{72n^3-360n^2+576n-288}{n^2}\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\delta_i\right).$$

Prieš skliaustus iškeliame $\frac{2\left(n-1\right)}{n^2}$, $\frac{1}{n^2}$ ir gauname tokią galutinę trečiojo centrinio momento išraišką:

$$\boldsymbol{\mu_3}(Z)=\frac{2(n-1)}{n^2}\Bigg[\left(20n^2-68n+60\right)\left(\sum_{i=1}^{k}\delta_i\right)^3-3\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\sigma_i\right)-18\left(\sum_{i=1}^{k}\varepsilon_i^2\right)\left(\sum_{i=1}^{k}\varepsilon_i\right)+$$

$$+\left(34n-68\right)\sum_{i=1}^{k}\varepsilon_i^3+9\left(\sum_{i=1}^{k}\varepsilon_i\right)^2\left(\sum_{i=1}^{k}\delta_i\right)-\left(18n-36\right)\left(\sum_{i=1}^{k}\varepsilon_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\delta_i\right)-\left(36n-54\right)\times$$

$$\times\left(\sum_{i=1}^{k}\varepsilon_i^2\right)\left(\sum_{i=1}^{k}\delta_i\right)+\left(16n^2-80n+96\right)\sum_{i=1}^{k}\varepsilon_i^2\delta_i+14\sum_{i=1}^{k}\varepsilon_i\sigma_i+\left(24n-36\right)\left(\sum_{i=1}^{k}\delta_i\right)^2\left(\sum_{i=1}^{k}\varepsilon_i\right)-$$

$$-\left(36n^2-144n+144\right)\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\delta_i\right)\Bigg]+\frac{1}{n^2}\left[2\left(\sum_{i=1}^{k}\varepsilon_i\right)^3+\frac{1}{n}\left(\sum_{i=1}^{k}\xi_i\right)-3\left(\sum_{i=1}^{k}\varepsilon_i\right)\left(\sum_{i=1}^{k}\sigma_i\right)\right]. \quad (3.21)$$

3.2.6. Programa χ^2 statistikos trims momentams skaičiuoti

Sudaryta programa, kuri įgalina skaičiuoti X_n^2 statistikos vidurkį, dispersiją ir trečiąjį centrinį momentus. Įėjimo parametrai: imties dydis n; U_i tikimybės $p_1,...,p_k$, hipotetinės tikimybės $p_{i0},...,p_{k0}$. Realizavus programą, gaunamos $\boldsymbol{E}X_n^2$, $\boldsymbol{Var}\left(X_n^2\right)$ ir $\boldsymbol{\mu_3}\left(X_n^2\right)$ reikšmės.

Kartu numatytas ir parametrų a, b ir ω, kurie yra reikalingi patikslintos aproksimacijos konstravimui, radimas.

Pavyzdžiui, įvedę $n=200$, $p=\left(0,25,\ 0,25,\ 0,25,\ 0,25\right)^T$ $p_{i0}=\left(0,25,\ 0,25,\ 0,25,\ 0,25\right)^T$, gausime tokias reikšmes:

$\boldsymbol{E}X_n^2=3$; $\boldsymbol{Var}\left(X_n^2\right)=5,97$; $\boldsymbol{\mu_3}\left(X_n^2\right)=24,12$;

$a=$ -0,044; $b=0,99$; $\omega=2,926$.

3.2.7. χ^2 statistikos momentų savybės, kai hipotezė teisinga

Tarkime, kad hipotezė $H: p_1=p_{10},\ ...,p_k=p_{k0}$ yra teisinga. Tada statistikos X_n^2 skirstinys aproksimuojamas χ^2 skirstiniu su k-1 laisvės laipsniu. Remiantis (3.14), (3.19) ir (3.21) išraiškomis, gauname, kad esant teisingai hipotezei

$$\boldsymbol{E}X_n^2=k-1;$$

$$\boldsymbol{Var}\left(X_n^2\right)=\frac{2(n-1)}{n}\left(2(n-2)-(2n-3)-2k+3k\right)-\frac{1}{n}\left(k^2-\sum_{i=1}^{k}\frac{1}{p_{i0}}\right)=$$

$$=2(k-1)+\frac{1}{n}\left(\sum_{i=1}^{k}\frac{1}{p_{i0}}-k^2-1\right)=2(k-1)+O\left(\frac{1}{n}\right);$$

$$\boldsymbol{\mu_3}\left(X_n^2\right)=\frac{2(n-1)}{n^2}\Big[\left(20n^2-68n+60\right)-3-18k^2+(34n-68)k+9k^2-(18n-36)k-(36n-54)k+$$

$$+\left(16n^2-80n+96\right)+14\cdot\sum_{i=1}^{k}\frac{1}{p_{i0}}+(24n-36)k-\left(36n^2-144n+144\right)\Big]+\frac{1}{n^2}\Bigg[2k^3+\sum_{i=1}^{k}\frac{1}{p_{i0}^2}-$$

$$-3k\sum_{i=1}^{k}\frac{1}{p_{i0}}\Bigg]=8(k-1)+\frac{1}{n^2}\Bigg[n\left(-18k^2-28k+17+28\sum_{i=1}^{k}\frac{1}{p_{i0}}\right)+2k^3+9k^2+14k-9+\sum_{i=1}^{k}\frac{1}{p_{i0}^2}-$$

$$-(3k+14)\sum_{i=1}^{k}\frac{1}{p_{i0}}\Bigg]=8(k-1)+O\left(\frac{1}{n}\right).$$

Matome, kad statistikos X_n^2 vidurkis sutampa su a.d. χ_{k-1}^2 vidurkiu $k-1$. Taigi vidurkio korekcija nereikalinga.

Jeigu statistikos X_n^2 skirstinys būtų χ_{k-1}^2, tai dispersija turėtų būti lygi $2(k-1)$. Matome, kad gauta dispersija skiriasi nuo $2(k-1)$. Jeigu tarsime, kad hipotetinės tikimybės yra vienodos ir lygios $1/k$, t.y. $p_{i0}=1/k$, $i=1,...,k$, tai $\frac{1}{n}\left(\sum_{i=1}^{k}\frac{1}{p_{i0}}-k^2\right)=0$ ir tuo atveju statistikos X_n^2 dispersija bus artimesnė $2(k-1)$. Taigi tikėtina, kad X_n^2 aproksimacija šiuo atveju bus tikslinga.

Taip pat matome, kad trečias centrinis momentas nesutampa su $8(k-1)$ (tai turėtų būti jo reikšmė, jeigu statistikos X_n^2 skirstinys sutaptų su skirstiniu χ_{k-1}^2). Jeigu vėl tartume, kad $p_{i0}=1/k$, $i=1,...,k$, gautume tokį statistikos X_n^2 trečiąjį centrinį momentą:

$$\boldsymbol{\mu_3}\left(X_n^2\right)=8(k-1)+\frac{\left(-10k^2-28k+17\right)}{n}+\left(9k^2+14k-9+\sum_{i=1}^{k}\frac{1}{p_{i0}^2}-14\sum_{i=1}^{k}\frac{1}{p_{i0}}\right)\frac{1}{n^2}.$$

Matome, kad tokiu atveju statistikos X_n^2 trečiasis centrinis momentas vis tiek pakankamai skiriasi nuo $8(k-1)$ ir, norint patikslinti kriterijų (2.5), vertėtų įvesti korekciją.

3.2.8. χ^2 statistikos momentų savybės, kai hipotezė neteisinga

Kai hipotezė (2.3) neteisinga, statistikos X_n^2 skirstinį paprastai aproksimuojame necentriniu χ^2 skirstiniu su $k-1$ laisvės laipsniu ir necentriškumo parametru

$$\delta_n = n\sum_{i=1}^{n}\frac{\left(p_i - p_{i0}\right)^2}{p_{i0}}.$$

Taigi X_n^2 vidurkis turėtų būti artimas a.d. $\chi^2_{k-1,\delta}$ vidurkiui, t.y.

$$\boldsymbol{E}\left(\chi^2_{k-1,\delta_n}\right) = k - 1 + \delta_n.$$

Iš išraiškos (3.14) gauname

$$\boldsymbol{E}X_n^2 = k - 1 + \delta_n + \sum_{i=1}^{k}\frac{\left(p_i - p_{i0}\right)\left(1 - p_i\right)}{p_{i0}}. \tag{3.22}$$

Matome, kad vidurkis $\boldsymbol{E}X_n^2$ skiriasi nuo $k-1+\delta_n$ dėmeniu $\sum_{i=1}^{k}\frac{\left(p_i - p_{i0}\right)\left(1 - p_i\right)}{p_{i0}}$.

Jeigu nors vienam $i \in \{1,...,k\}$ skirtumas $p_i - p_{i0} > 0$, tai $\delta_n \to \infty$, kai $n \to \infty$. Kadangi dispersija $\boldsymbol{Var}\left(X_n^2\right) = O(n)$, tai kriterijaus galia artėja į 1, kai $n \to \infty$. Todėl, norint palyginti galios ar $\boldsymbol{E}X_n^2$ elgesį, reikėtų nagrinėti artėjančias alternatyvas, kad δ_n būtų konstantos eilės.

Parinkime

$$p_i = p_{i0} + \frac{c_i}{\sqrt{n}}, \tag{3.23}$$

čia c_i yra konstantos, tenkinančios sąlygas:

$$\sum_{i=1}^{k} c_i = 0,\ p_{i0} < \frac{c_i}{\sqrt{n}} < 1 - p_{i0}.$$

Tada

$$\delta_n = n\sum_{i=1}^{k}\frac{c_i^2}{np_{i0}} = \sum_{i=1}^{k}\frac{c_i^2}{p_{i0}} = \Delta = const,$$

o trečiasis išraiškos (3.22) dėmuo bus lygus

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\frac{c_i\left(1 - p_{i0} - \frac{c_i}{\sqrt{n}}\right)}{p_{i0}} = O\left(\frac{1}{\sqrt{n}}\right),\ n \to \infty.$$

Dabar iš (3.22) turime tokią išraišką

$$\boldsymbol{E}X_n^2 = k - 1 + \Delta + O\left(\frac{1}{\sqrt{n}}\right).$$

ir matome, kad X_n^2 vidurkis artėja prie $k-1+\Delta$, kai $n \to \infty$.

Kai hipotezė (2.3) neteisinga, statistikos X_n^2 skirstinį aproksimuojame skirstiniu $\chi^2_{k-1,\Delta}$, todėl $\boldsymbol{Var}\left(X_n^2\right)$ turėtų būti artima $2\left(k - 1 + 2\Delta\right)$. Panagrinėkime dispersijos elgesį anksčiau įvestų artėjančių alternatyvų atveju.

Skyrelyje 3.2.4. dėl skaičiavimų paprastumo įsivedėme pažymėjimus (3.16), taigi į jų išraiškas įrašę (3.23), gauname

$$\varepsilon_i = \frac{p_{i0} + \frac{c_i}{\sqrt{n}}}{p_{i0}} = 1 + \frac{c_i}{p_{i0}\sqrt{n}};$$

$$\delta_i = \frac{p_{i0}^2 + \frac{2p_{i0}c_i}{\sqrt{n}} + \frac{c_i^2}{n}}{p_{i0}} = p_{i0} + \frac{2c_i}{\sqrt{n}} + \frac{c_i^2}{np_{i0}}; \tag{3.24}$$

$$\sigma_i = \frac{p_{i0} + \frac{c_i}{\sqrt{n}}}{p_{i0}^2} = \frac{1}{p_{i0}} + \frac{c_i}{p_{i0}^2\sqrt{n}}.$$

Šias išraiškas įsistatome į statistikos X_n^2 dispersijos išraišką (3.19) padėmeniui

$$\sum_{i=1}^{k}\varepsilon_i\delta_i = \sum_{i=1}^{k}\left(p_{i0} + \frac{2c_i}{\sqrt{n}} + \frac{c_i^2}{np_{i0}} + \frac{c_i}{\sqrt{n}} + \frac{2c_i^2}{np_{i0}} + \frac{c_i^3}{p_{i0}^2 n\sqrt{n}}\right) = 1 + \frac{3\Delta}{n} + O\left(\frac{1}{n\sqrt{n}}\right);$$

$$\left(\sum_{i=1}^{k}\delta_i\right)^2 = \left(1 + \frac{\Delta}{n}\right)^2 = 1 + \frac{2\Delta}{n} + \left(\frac{\Delta}{n}\right)^2;$$

$$\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\right) = \left(1 + \frac{\Delta}{n}\right)k;$$

$$\sum_{i=1}^{k}\varepsilon_i^2 = \sum_{i=1}^{k}\left(1 + \frac{2c_i}{p_{i0}\sqrt{n}} + \frac{c_i^2}{np_{i0}^2}\right) = k + O\left(\frac{1}{\sqrt{n}}\right);$$

$$\left(\sum_{i=1}^{k}\varepsilon_i\right)^2 = k^2 + O\left(\frac{1}{\sqrt{n}}\right);$$

$$\sum_{i=1}^{k}\sigma_i = \sum_{i=1}^{k}\frac{1}{p_{i0}} + \frac{1}{\sqrt{n}}\sum_{i=1}^{k}\frac{c_i}{p_{i0}^2}.$$

Taigi

$$\boldsymbol{Var}\left(X_n^2\right) = \frac{4(n-1)(n-2)}{n}\cdot\left(1 + \frac{3\Delta}{n}\right) - \frac{2(n-1)(2n-3)}{n}\left(1 + \frac{2\Delta}{n} + \left(\frac{\Delta}{n}\right)^2\right) - \frac{4(n-1)}{n}\left(1 + \frac{\Delta}{n}\right)k +$$

$$+\frac{6(n-1)k}{n} - \frac{1}{n}\left(k^2 + O\left(\frac{1}{\sqrt{n}}\right)\right) + \frac{1}{n}\sum_{i=1}^{k}\frac{1}{p_{i0}} + \frac{1}{n}\sum_{i=1}^{k}\frac{c_i}{p_{i0}^2} = \frac{2(n-1)}{n}(2n - 4 - 2n + 3) + \frac{2(n-1)}{n}\cdot k +$$

$$\frac{\Delta}{n^2}\left(12(n-1)(n-2) - 4(n-1)(2n-3) + \frac{2\Delta(n-1)(2n-3)}{n} - 4(n-1)k\right) - \frac{k^2}{n} +$$

$$+\frac{1}{n}\sum_{i=1}^{k}\frac{1}{p_{i0}} + \frac{1}{n}\sum_{i=1}^{k}\frac{c_i}{p_{i0}^2} = 2(k-1)\cdot\frac{n-1}{n} + \frac{4\Delta}{n^2}(n-1)(n-3) + O\left(\frac{1}{\sqrt{n}}\right) =$$

$$2(k-1)+4\Delta+O\left(\frac{1}{\sqrt{n}}\right),\ n\to\infty.$$

Vėl akivaizdu, kad, šiek tiek pakoregavus hipotetines tikimybes, dispersija artėja prie skirstinio $\chi^2_{k-1,\Delta}$ dispersijos $2(k-1)+4\Delta$.

Kai hipotezė (2.3) neteisinga, a.d. X_n^2 trečiasis centrinis momentas turėtų artėti į $8(k-1)+24\Delta$. Vėl nagrinėkime trečiojo centrinio momento elgesį, naudodami artėjančias alternatyvas.

Statistikos X_n^2 trečiojo centrinio momento skaičiavimui supaprastinti skyrelyje 3.2.5. įsivedėme dar vieną pažymėjimą (3.20), o, į šią išraišką įstatę (3.23), gauname

$$\xi_i=\frac{1}{p_{i0}^2}+\frac{c_i}{p_{i0}^3\sqrt{n}}. \tag{3.25}$$

Taigi remdamiesi (3.24), (3.25) ir anksčiau apskaičiuotomis kai kuriomis sumomis, apskaičiuojame likusius išraiškos (3.21) dėmenis:

$$\left(\sum_{i=1}^k\delta_i\right)^3=\left(1+\frac{\Delta}{n}\right)^3;$$

$$\left(\sum_{i=1}^k\delta_i\right)\left(\sum_{i=1}^k\sigma_i\right)=\left(1+\frac{\Delta}{n}\right)\sum_{i=1}^k\frac{1}{p_{i0}}+O\left(\frac{1}{\sqrt{n}}\right);$$

$$\left(\sum_{i=1}^k\varepsilon_i^2\right)\left(\sum_{i=1}^k\varepsilon_i\right)=\left(k+O\left(\frac{1}{\sqrt{n}}\right)\right)\left(k+O\left(\frac{1}{\sqrt{n}}\right)\right)=k^2+O\left(\frac{1}{\sqrt{n}}\right);$$

$$\sum_{i=1}^k\varepsilon_i'^3=\sum_{i=1}^k\left(1+\frac{3c_i}{p_{i0}\sqrt{n}}+\frac{3c_i^2}{p_{i0}^2n}+\frac{c_i^3}{p_{i0}^3n\sqrt{n}}\right)=k+O\left(\frac{1}{n\sqrt{n}}\right);$$

$$\left(\sum_{i=1}^k\varepsilon_i\right)^2\left(\sum_{i=1}^k\delta_i\right)=\left(k+O\left(\frac{1}{\sqrt{n}}\right)\right)^2\left(1+\frac{\Delta}{n}\right)=k^2\left(1+\frac{\Delta}{n}\right)+O\left(\frac{1}{\sqrt{n}}\right);$$

$$\left(\sum_{i=1}^k\varepsilon_i\right)\left(\sum_{i=1}^k\varepsilon_i\delta_i\right)=k\left(1+\frac{3\Delta}{n}+O\left(\frac{1}{n\sqrt{n}}\right)\right)=k+\frac{3k\Delta}{n}+O\left(\frac{1}{n\sqrt{n}}\right);$$

$$\left(\sum_{i=1}^k\varepsilon_i^2\right)\left(\sum_{i=1}^k\delta_i\right)=\left(k+O\left(\frac{1}{n}\right)\right)\left(1+\frac{\Delta}{n}\right)=k+\frac{k\Delta}{n}+O\left(\frac{1}{n}\right);$$

$$\sum_{i=1}^k\varepsilon_i^2\delta_i=\sum_{i=1}^k\left[\left(1+\frac{2c_i}{p_{i0}\sqrt{n}}+\frac{c_i^2}{np_{i0}^2}\right)\left(p_{i0}+\frac{2c_i}{\sqrt{n}}+\frac{c_i^2}{np_{i0}}\right)\right]=\sum_{i=1}^k\left(p_{i0}+\frac{2c_i}{\sqrt{n}}+\frac{c_i^2}{np_{i0}}+\frac{2c_i}{\sqrt{n}}+\frac{4c_i^2}{np_{i0}}+\right.$$

$$\left.+\frac{2c_i^3}{p_{i0}^2n\sqrt{n}}+\frac{c_i^2}{np_{i0}}+\frac{2c_i^3}{p_{i0}^2n\sqrt{n}}+\frac{c_i^4}{n^2p_{i0}^3}\right)=1+\frac{6\Delta}{n}+O\left(\frac{1}{n\sqrt{n}}\right);$$

$$\sum_{i=1}^{k}\varepsilon_i\delta_i=\sum_{i=1}^{k}\left[\left(1+\frac{c_i}{p_{i0}\sqrt{n}}\right)\left(\frac{1}{p_{i0}}+\frac{c_i}{p_{i0}^2\sqrt{n}}\right)\right]=\sum_{i=1}^{k}\left(\frac{1}{p_{i0}}+\frac{2c_i}{p_{i0}^2\sqrt{n}}+\frac{c_i^2}{p_{i0}^3 n}\right)=\sum_{i=1}^{k}\frac{1}{p_{i0}}+O\left(\frac{1}{\sqrt{n}}\right);$$

$$\left(\sum_{i=1}^{k}\delta_i\right)^2\left(\sum_{i=1}^{k}\varepsilon_i\right)=k\left(1+\frac{\Delta}{n}\right)^2;$$

$$\left(\sum_{i=1}^{k}\delta_i\right)\left(\sum_{i=1}^{k}\varepsilon_i\delta_i\right)=\left(1+\frac{\Delta}{n}\right)\left(1+\frac{3\Delta}{n}+O\left(\frac{1}{n\sqrt{n}}\right)\right)=1+\frac{4\Delta}{n}+\frac{3\Delta^2}{n^2}+O\left(\frac{1}{n\sqrt{n}}\right);$$

$$\left(\sum_{i=1}^{k}\varepsilon_i\right)^3=k^3;$$

$$\sum_{i=1}^{k}\xi_i=\sum_{i=1}^{k}\frac{1}{p_{i0}^2}+O\left(\frac{1}{\sqrt{n}}\right);$$

$$\left(\sum_{i=1}^{k}\varepsilon_i\right)\left(\sum_{i=1}^{k}\sigma_i\right)=k\sum_{i=1}^{k}\frac{1}{p_{i0}}+O\left(\frac{1}{\sqrt{n}}\right).$$

Įsistatome gautas reikšmes į (3.21)

$$\boldsymbol{\mu_3}\left(X_n^2\right)=\frac{2(n-1)}{n^2}\Bigg[\left(20n^2-68n+60\right)\left(1+\frac{\Delta}{n}\right)^3-3\left(\left(1+\frac{\Delta}{n}\right)\sum_{i=1}^{k}\frac{1}{p_{i0}}+O\left(\frac{1}{\sqrt{n}}\right)\right)-$$

$$-18\left(k^2+O\left(\frac{1}{n}\right)\right)+\left(34n-68\right)\left(k+O\left(\frac{1}{n\sqrt{n}}\right)\right)+9k^2\left(1+\frac{\Delta}{n}\right)-\left(18n-36\right)\times$$

$$\times\left(k+\frac{3k\Delta}{n}+O\left(\frac{1}{n\sqrt{n}}\right)\right)-\left(36n-54\right)\left(k+\frac{k\Delta}{n}+O\left(\frac{1}{n}\right)\right)+\left(16n^2-80n+96\right)\times$$

$$\times\left(1+\frac{6\Delta}{n}+O\left(\frac{1}{n\sqrt{n}}\right)\right)+14\left(\sum_{i=1}^{k}\frac{1}{p_{i0}}+O\left(\frac{1}{\sqrt{n}}\right)\right)+\left(24n-36\right)k\left(1+\frac{\Delta}{n}\right)^2-\left(36n^2-144n+144\right)\times$$

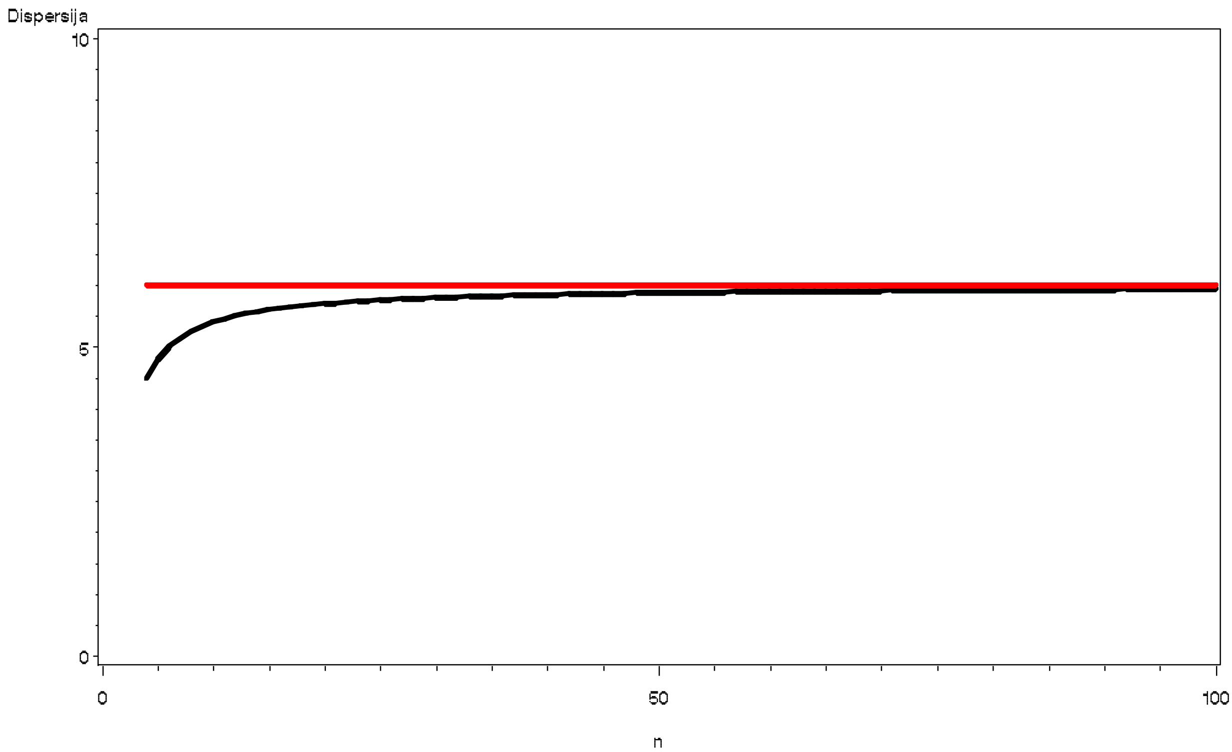
$$\times\left(1+\frac{4\Delta}{n}+\frac{3\Delta^2}{n^2}+O\left(\frac{1}{n\sqrt{n}}\right)\right)\Bigg]+\frac{1}{n^2}\left[2k^3+\sum_{i=1}^{k}\frac{1}{p_{i0}^2}-3k\sum_{i=1}^{k}\frac{1}{p_{i0}}+O\left(\frac{1}{\sqrt{n}}\right)\right]=$$

$$=8\left(k-1\right)\frac{n-1}{n}+24\Delta\frac{n-1}{n}+O\left(\frac{1}{\sqrt{n}}\right)=8\left(k-1\right)+24\Delta+O\left(\frac{1}{\sqrt{n}}\right),\ n\to\infty.$$

Taigi ir vėl galime įsitikinti, kad pasitelkiant artėjančias alternatyvas, galima priartinti statistikos X_n^2 pasiskirstymą prie necentrinio skirstinio χ^2 su k-1 laisvės laipsniais ir necentriškumo parametru δ_n.

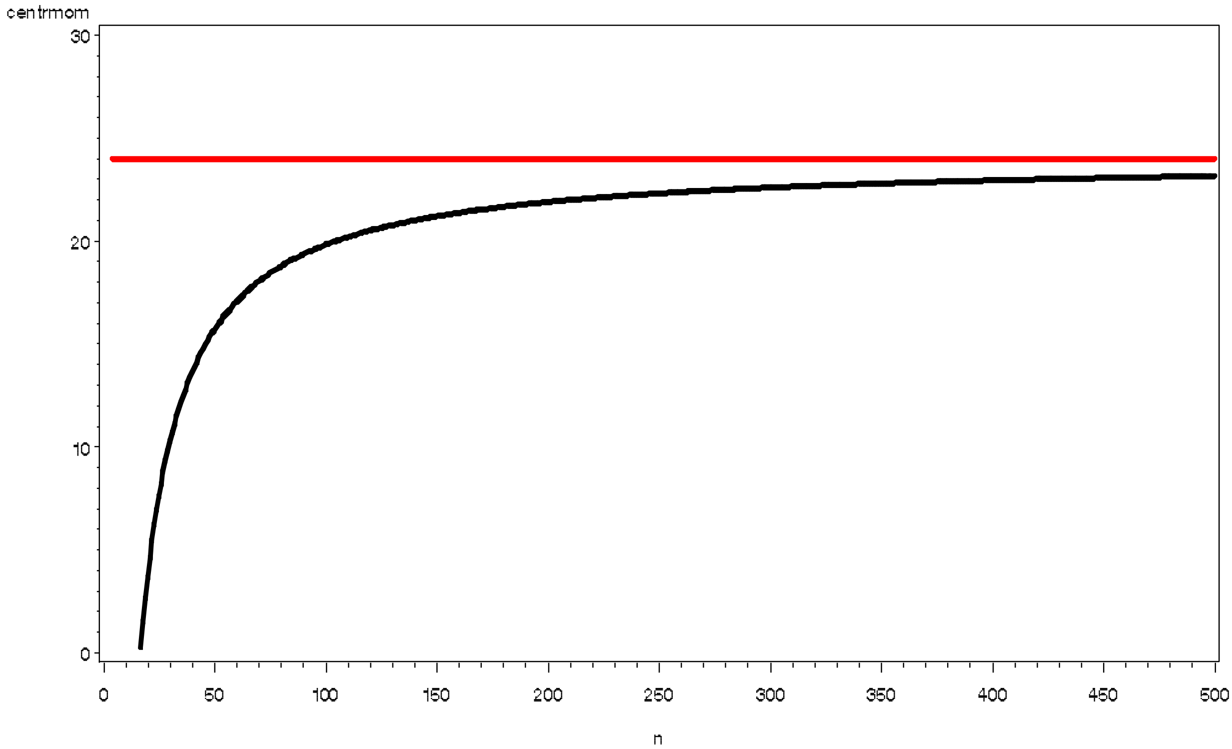
3.2.9. Grafinis χ^2 statistikos momentų elgsenos vaizdavimas

3.1. pavyzdys. Tarkime, kad $k=4$. Tikriname hipotezę $H: p_i = p_{i0} = \frac{1}{4}$, $i=1,...,4$. Jeigu hipotezė H teisinga, tai $\boldsymbol{E}X_n^2 = 3$, $\boldsymbol{Var}\left(X_n^2\right) = 6 + O\left(\frac{1}{n}\right) \neq 6$ $\boldsymbol{\mu_3}\left(X_n^2\right) = 24 + O\left(\frac{1}{n^2}\right) \neq 24$. Tačiau iš 3.1 ir 3.2 paveikslėlių matome, kad tiek a.d. X_n^2 dispersija, tiek trečiasis centrinis momentas artėja prie skirstinio χ_3^2 momentų reikšmių (horizontalios linijos), didėjant imties dydžiui n ir pakankamai greitai, kai yra parinktos vienodos hipotetinės tikimybės.



3.1 pav. Statistikos X_n^2 dispersijos elgesys, kintant imties dydžiui n ir esant parinktoms vienodoms hipotetinėms tikimybėms, kai hipotezė teisinga

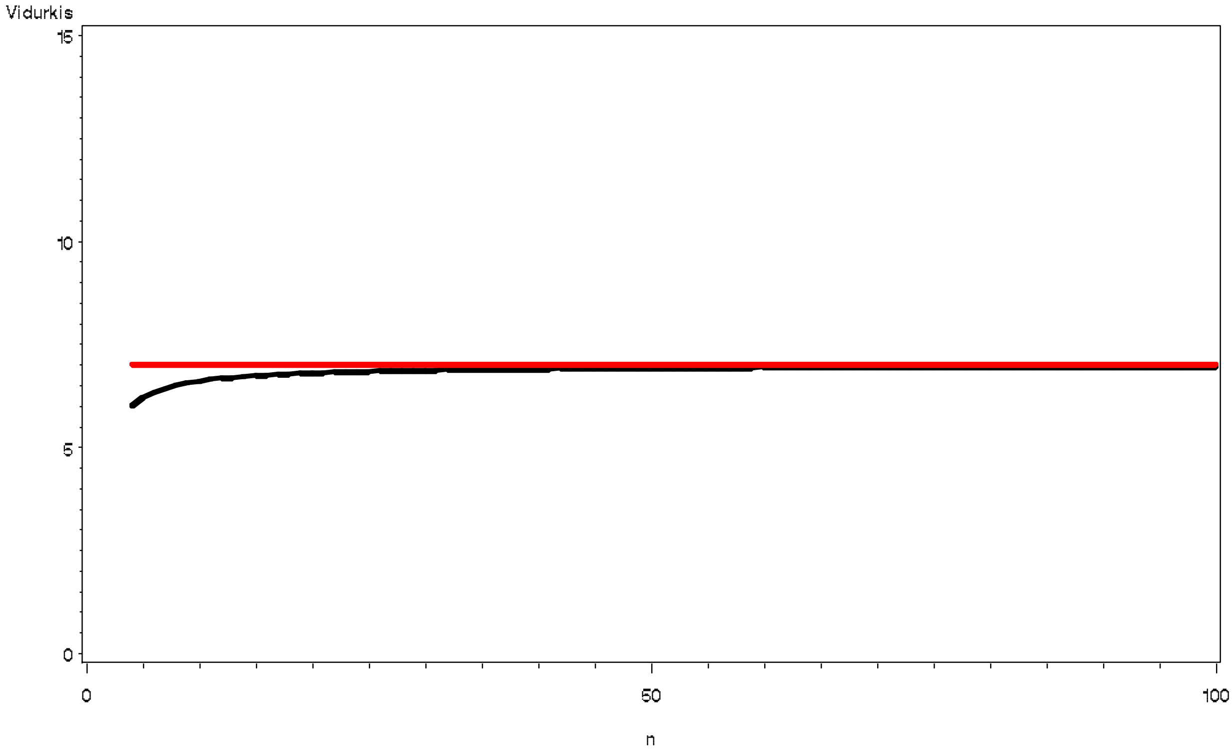
Iš grafiko pastebime, kad esant teisingai hipotezei, a.d. X_n^2 tikroji dispersija yra labai arti teorinės ir, maždaug kai $n \geq 60$, tarp jų beveik nebelieka skirtumo. Taigi šiuo atveju dispersijai korekcija nereikalinga. Esant nedideliems n, tikroji dispersija $\boldsymbol{Var}\left(X_n^2\right)$ yra mažesnė už $2(k-1)$.



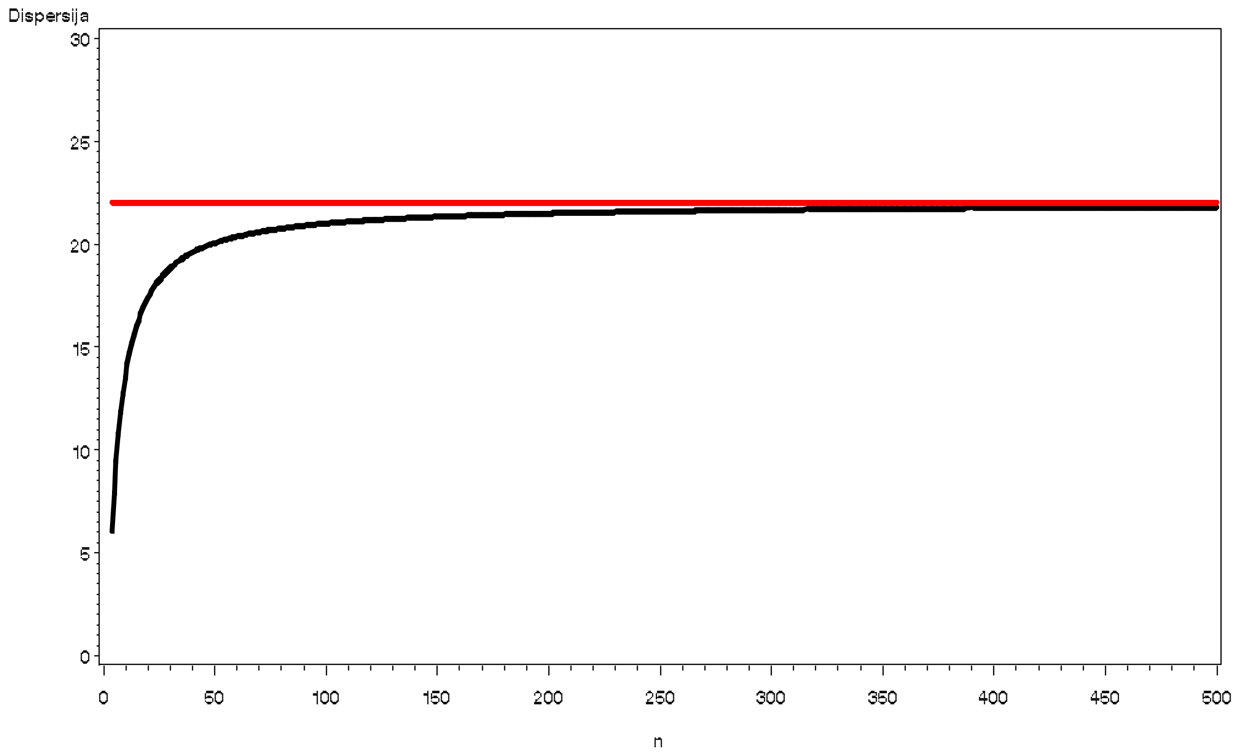
3.2 pav. Statistikos X_n^2 trečiojo centrinio momento elgesys, kintant imties dydžiui n ir esant parinktoms vienodoms hipotetinėms tikimybėms, kai hipotezė teisinga

Trečiasis centrinis momentas, kaip matome, iš pradžių labai skiriasi nuo teorinio ir tik nuo pakankamai didelio imties dydžio n pradeda artėti prie 24. Vadinasi, kai imtis yra mažesnė, įvesti korekciją būtų pravartu.

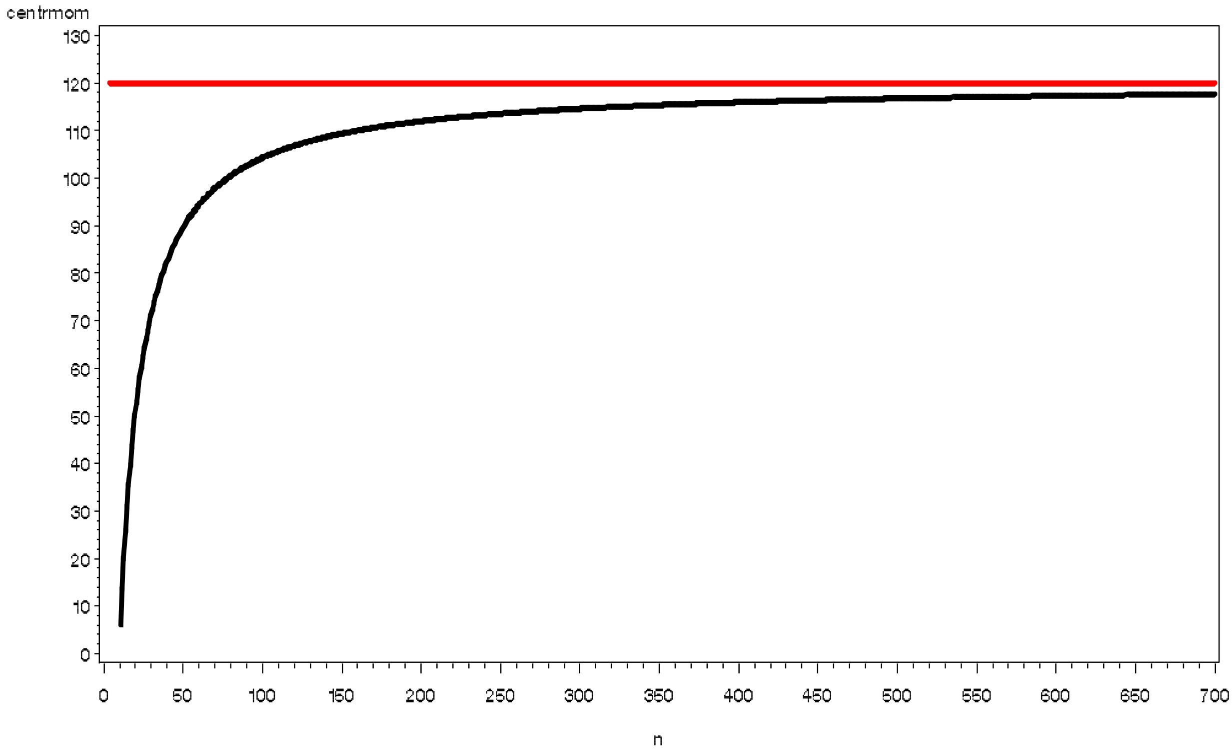
Tegu H neteisinga ir parenkame artėjančių alternatyvų seką: $p_1 = p_2 = \frac{1}{4} - \frac{c}{\sqrt{n}}$, $p_3 = p_4 = \frac{1}{4} + \frac{c}{\sqrt{n}}$. Tuomet $\Delta = \sum_{i=1}^{k} \frac{c^2}{1/4} = 16c^2$. Pasirinkime $\Delta = 4$, tuomet $c = \frac{1}{2}$. Šiuo atveju a.d. X_n^2 vidurkio, dispersijos ir trečiojo centrinio momento elgesys iliustruotas 3.3, 3.4 ir 3.5 paveikslėliuose.



3.3 pav. Statistikos X_n^2 vidurkio elgesys, kintant imties dydžiui n ir parinktoms artėjančioms alternatyvoms, kai hipotezė neteisinga



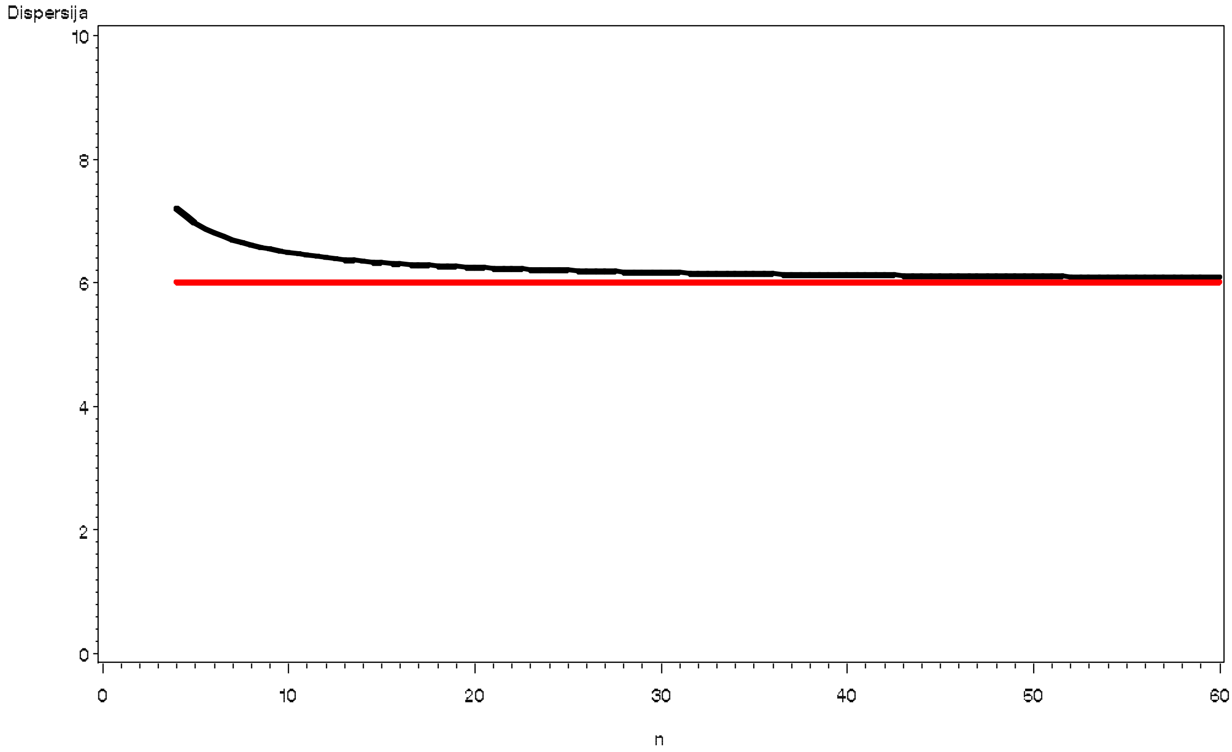
3.4 pav. Statistikos X_n^2 dispersijos elgesys, kintant imties dydžiui n ir parinktoms artėjančioms alternatyvoms, kai hipotezė neteisinga



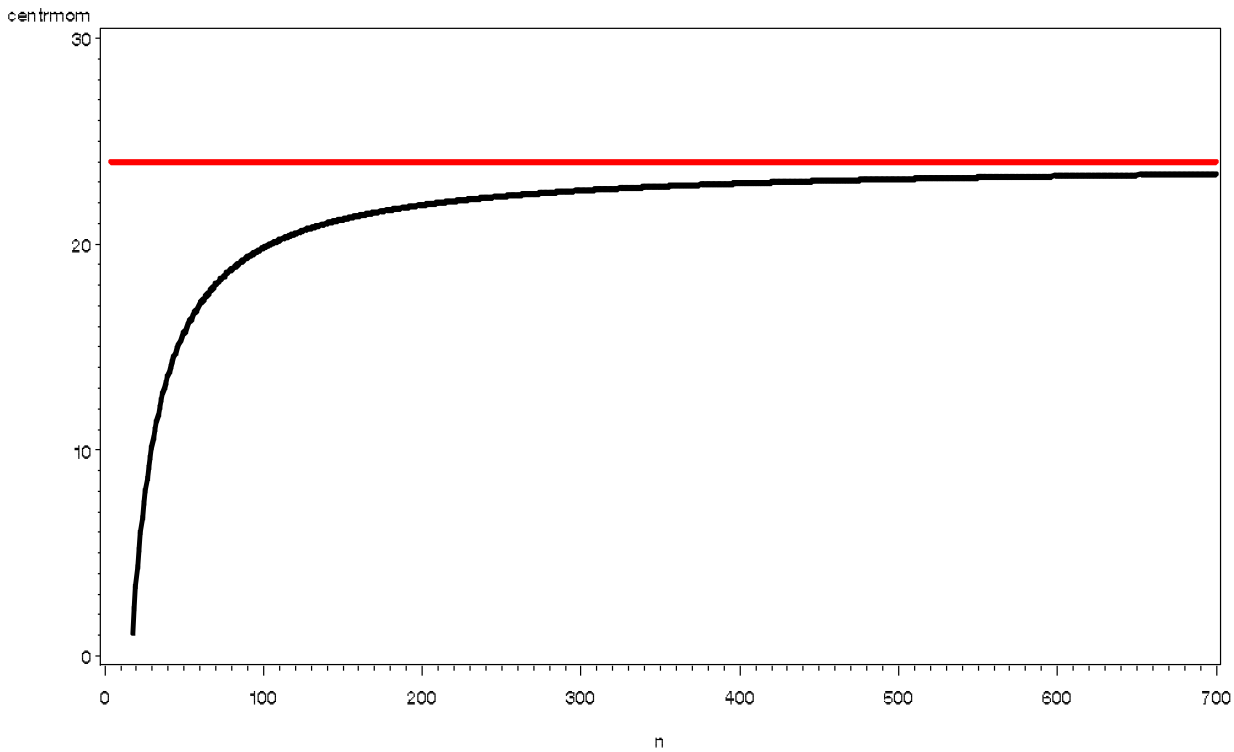
3.5 pav. Statistikos X_n^2 trečiojo centrinio momento elgesys, kintant imties dydžiui n ir parinktoms artėjančioms alternatyvoms, kai hipotezė neteisinga

Iš šių grafikų pastebime, kad, esant neteisingai hipotezei, a.d. X_n^2 momentai yra labiau nutolę nuo teorinių savo reikšmių negu, kai hipotezė yra teisinga. Vidurkis jau nebėra konstanta, nors vis dar nedaug skiriasi nuo teorinio savo atitikmens. Šiuo atveju dispersijos reikšmė, esant pakankamai dideliam imties dydžiui n, priartėja prie $2(k-1)+4\Delta = 22$ .Tačiau palyginę su pirmuoju dispersijos grafiku 3.1, galime pastebėti, kad artėjimas yra lėtesnis. Tą patį galima pasakyti ir apie trečiojo centrinio momento elgseną, kuris tik esant dideliam imties dydžiui reikšmingai priartėja prie $8(k-1)+24\Delta$.

3.2. pavyzdys. Šį kartą pateiksime pavyzdį, kai hipotetinės tikimybės yra nevienodos. Taigi, pasirinkę $k=4$, tikrinsime hipotezę H: $p_1 = p_{10} = \frac{1}{16}$, $p_2 = p_{20} = \frac{3}{16}$, $p_3 = p_{30} = \frac{5}{16}$, $p_4 = p_{40} = \frac{7}{16}$. Jeigu hipotezė teisinga, $\boldsymbol{E}X_3^2 = 3$, $\boldsymbol{Var}\left(X_n^2\right) \neq 6$, $\boldsymbol{\mu_3}\left(X_n^2\right) \neq 24$. Taigi statistikos X_n^2 momentai nevisiškai atitinka teorinio skirstinio χ_3^2 momentų reikšmes. Tačiau iš paveikslėlių 3.6 ir 3.7 matome, kad didinant imties dydį n, jos tampa beveik lygios.



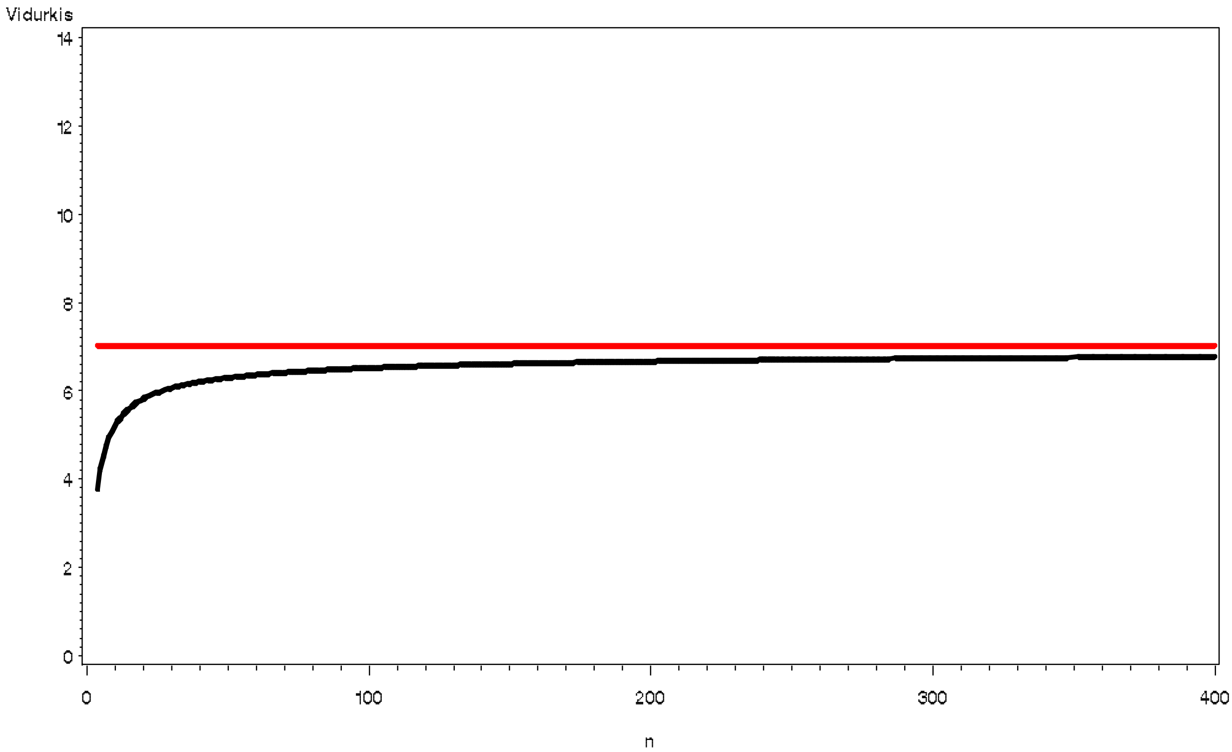
3.6 pav. Statistikos X_n^2 dispersijos elgesys, kintant imties dydžiui n ir parinktoms skirtingoms hipotetinėms tikimybėms, kai hipotezė teisinga



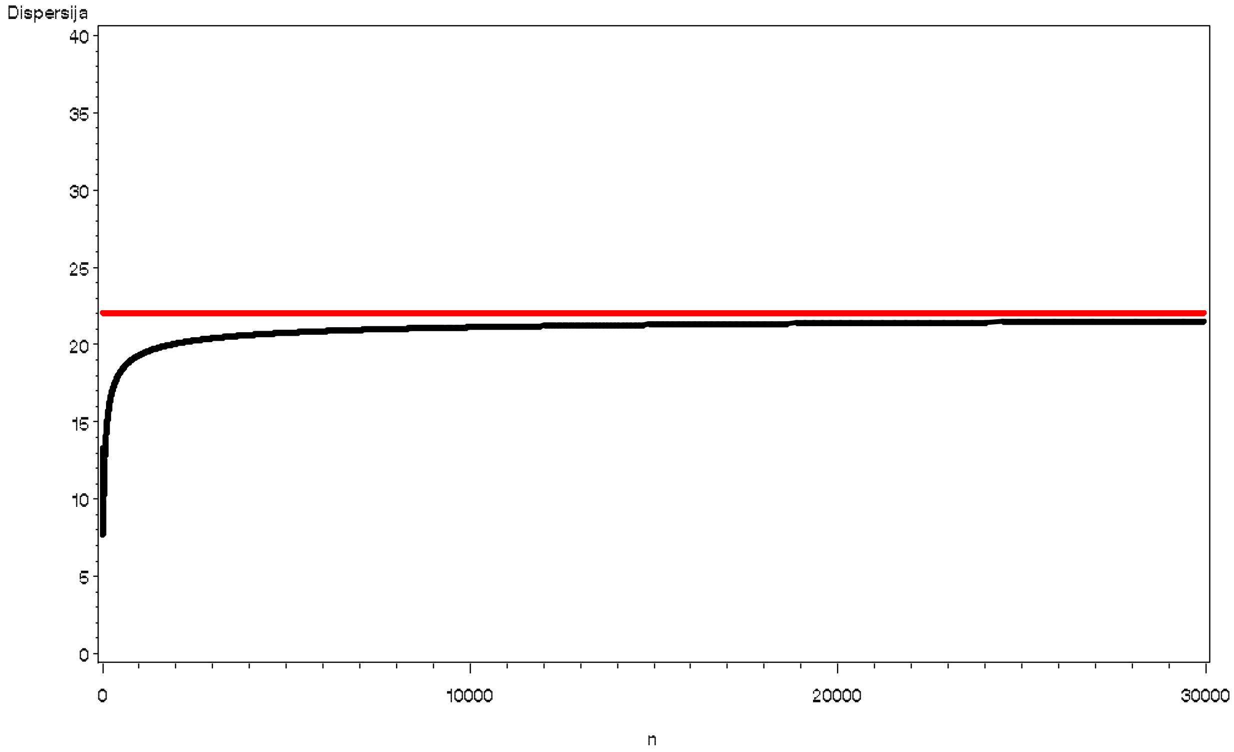
3.7 pav. Statistikos X_n^2 trečiojo centrinio momento elgesys, kintant imties dydžiui n ir parinktoms skirtingoms hipotetinėms tikimybėms, kai hipotezė teisinga

Matome, kad esant teisingai hipotezei, statistikos X_n^2 momentai elgiasi panašiai kaip ir pirmuoju atveju, kai hipotetinės tikimybės buvo vienodos.

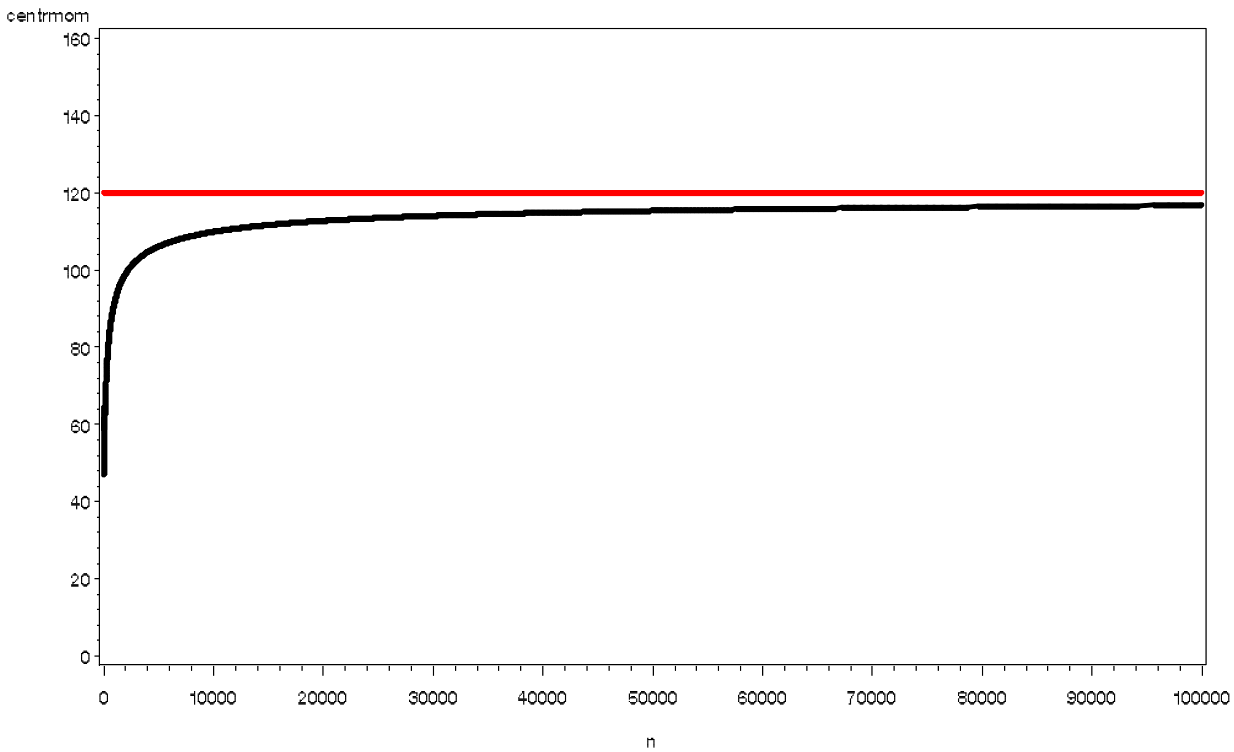
Jeigu hipotezė neteisinga, parenkame artėjančių alternatyvų seką $p_1 = p_{10} + \frac{c}{\sqrt{n}} = \frac{1}{16} + \frac{c}{\sqrt{n}}$, $p_2 = p_{20} + \frac{c}{\sqrt{n}} = \frac{3}{16} + \frac{c}{\sqrt{n}}$, $p_3 = p_{30} - \frac{c}{\sqrt{n}} = \frac{5}{16} - \frac{c}{\sqrt{n}}$, $p_4 = p_{40} - \frac{c}{\sqrt{n}} = \frac{7}{16} - \frac{c}{\sqrt{n}}$. Parametrą c parinksime tokį, kad $\Delta = \sum_{i=1}^{k} \frac{c^2}{p_{i0}} = 4$. Vadinasi $c = \sqrt{\frac{105}{704}} \approx 0,3862$. 2.16, 2.17 ir 2.18 pav. pavaizduota, kaip keičiasi a.d. X_n^2 trijų pirmųjų momentų reikšmės, esant pasirinktoms artėjančioms alternatyvoms bei didėjant imties dydžiui n.



3.8 pav. Statistikos X_n^2 vidurkio elgesys, kintant imties dydžiui n ir remiantis 3.2. pavyzdyje parinktomis artėjančiomis alternatyvomis, kai hipotezė neteisinga



3.9 pav. Statistikos X_n^2 dispersijos elgesys, kintant imties dydžiui n ir remiantis 3.2. pavyzdyje parinktomis artėjančiomis alternatyvomis, kai hipotezė neteisinga



3.10 pav. Statistikos X_n^2 trečiojo centrinio momento elgesys, kintant imties dydžiui n ir remiantis 3.2. pavyzdyje parinktomis artėjančiomis alternatyvomis, kai hipotezė neteisinga

Iš šių grafikų jau akivaizdžiai galime pastebėti, kaip svarbu, kad hipotetinės tikimybės būtų vienodos, nes tik esant labai dideliems imties dydžiams a.d. X_n^2 momentai gali priartėti prie savo teorinių atitikmenų. Taip pat nebkyla abejonių, jog statistikos X_n^2 korekcija yra reikalinga mažesnėms imtims.

Iš paveikslėlių galime pamatyti, kad antro pavyzdžio atveju, kai hipotetinės tikimybės yra nevienodos ir imtis nedidelė, statistikos X_n^2 pirmieji trys momentai labiau nutolę nuo teorinio skirstinio χ_3^2 momentų reikšmių ir kur kas lėčiau į jas artėja.

4. KRITERIJUS, KURIS REMIASI MODELIUOTOMIS REIKŠMĖMIS

Tarkime, kad tikriname hipotezę $H: \quad p_i = p_{i0}, \quad i = 1,...,k$, remdamiesi a.v. $U = \left(U_1,...,U_k\right)^T \sim \mathcal{P}_k\left(n, \boldsymbol{p}\right)$ realizacija $u = \left(u_1,...,u_k\right)^T$.

Pagal apytikslį χ^2 suderinamumo kriterijų, hipotezė H atmetama, kai teisinga nelygybė

$$X_n^2 = \sum_{i=1}^{k} \frac{U_i^2}{np_{i0}} - n > \chi_\alpha^2\left(k-1\right).$$

Kitaip tariant , H atmetama, kai realizacija x_n^2 patenka į kritinę sritį. Kaip žinome, atliekant skaičiavimus kompiuteriu, kritinė sritis nėra apibrėžiama. Vietoje to yra randama vadinamoji P-reikšmė (P-value), t.y. skaičiuojama tikimybė, kad a.d. X_n^2 pateks į dešinę nuo gautos realizacijos. Hipotezė yra atmetama tada, kai teisinga nelygybė

$$p_r < \alpha . \tag{4.1}$$

čia p_r yra p-reikšmė, o α - pasikliovimo lygmuo hipotezei tikrinti. Kadangi yra programos , kurios įgalina rasti a.d. X_n^2 realizacijas, tai galima įvertinti tikrąją P-reikšmę. Modeliuojame N a.d. X_n^2 realizacijų ir tikriname nelygybę $X_n^2 > x_n^2$, kur x_n^2 yra gautoji realizacija pagal turimus duomenis. Tarkime, kad iš N modeliuotų reikšmių į dešinę nuo turimos realizacijos x_n^2 pateko M reikšmių. Tada nežinomos P-reikšmės p_r įvertinys yra

$$\hat{p}_r = \frac{M}{N} . \tag{4.2}$$

Kriterijus pagal modeliuotas reikšmes atmeta tikrinamą hipotezę, kai teisinga nelygybė

$$\hat{p}_r < \alpha . \tag{4.3}$$

Šio kriterijaus tikslumas priklauso nuo N . Pavyzdžiui, nežinomai P-reikšmei galime sukonstruoti pasikliauties intervalą

$$\hat{p}_r - z_\alpha \sqrt{\frac{\hat{p}_r\left(1-\hat{p}_r\right)}{N}} < p_r < \hat{p}_r + z_\alpha \sqrt{\frac{\hat{p}_r\left(1-\hat{p}_r\right)}{N}} \ .$$

Taigi kriterijaus (4.2) tikslumas priklauso tik nuo N parinkimo.

Norint taikyti šį kriterijų, reikia įvesti imties didumą n, hipotetines tikimybes $p_{10},...,p_{k0}$, modeliuojamų a.d. X_n^2 skaičių N bei sprendžiant uždavinį gautą statistikos X_n^2 realizaciją.

5. KRITERIJŲ TIKSLUMO PALYGINIMAS

Ansktesniuose skyreliuose aprašėme tris galimus kriterijus suderinamumo hipotezėms tikrinti: (1.6), (3.2) ir (4.1). Kad įsitikintume, jog naudojant mūsų įvestą statistikos X_n^2 korekciją

$$V = aX_n^2 + b \ ,$$

sukonstruotas kriterijus yra tikslesnis, reikia atlikti palyginamąją analizę. Vienas iš galimų būdų tai padaryti yra palyginant tikimybes

$$\boldsymbol{P}\left\{X_n^2 > d\right\}$$

su įvairiomis d reikšmėmis (arba a.d. X_n^2 realizacijomis).

Tikrinant papratąją suderinamumo hipotezę ir taikant klasikinį χ^2 kriterijų (1.6), a.d. X_n^2 yra aproksimuojamas pagal χ^2 skirstinį su $k-1$ laisvės laipsniu, kur k yra tikimybių p_i skaičius, t.y. intervalų, į kuriuos yra suskaidyta duomenų aibė, skaičius. Taigi skaičiuosime tikimybę

$$\boldsymbol{P}\left\{\chi_{k-1}^2 > d\right\}. \tag{5.1}$$

Kai hipotezė neteisinga, a.d. X_n^2 skirstinys aproksimuojamas necentriniu χ^2 skirstiniu su necentriškumo parametru δ . Tada skaičiuojama tikimybė

$$\boldsymbol{P}\left\{\chi_{k-1,\delta}^2 > d\right\}, \tag{5.2}$$

kur δ yra necentriškumo parametras.

Taikant antrąjį kriterijų (3.2), statistikos skirstinys aproksimuojamas tuo pačiu χ^2 skirstiniu, bet su ω laisvės laipsnių, kur ω reikšmė yra gauta sprendžiant lygčių sistemą (3.3). Vadinasi skaičiuosime tikimybę

$$\boldsymbol{P}\{V > z\} = \boldsymbol{P}\left\{a + bX_n^2 > z\right\} = \boldsymbol{P}\left\{X_n^2 > \frac{z-a}{b}\right\}, \tag{5.3}$$

čia b ir a yra koeficientai, rasti iš lygčių sistemos (3.3), z_i - a.d. V realizacija. Kadangi yra lyginamas kriterijų tikslumas, tai turėtų būti tenkinama lygybė:

$$\frac{z-a}{b}=d\ , \tag{5.4}$$

čia d_i yra statistikos X_n^2 realizacija. Išsprendus (5.1) atžvilgiu z_i, gaunama

$$z=a+bd\ .$$

Tada (5.4) galima pakeisti į

$$P\{V>a+bd\}, \tag{5.5}$$

Trečiuoju atveju, taikant kriterijų (4.1), statistikos X_n^2 pasiskirstymo mums nereikia žinoti. Jis remiasi modeliuotomis reikšmėmis. Šiuo atveju papraščiausiai rasime P-reikšmės įvertinį (4.2).

5.1 -5.18 pav. yra pavaizduotos tikimybės (5.1), (5.3) ir (4.2) priklausomai nuo parinktų a.d. X_n^2 realizacijų, o taip pat kriterijaus galia pagal (5.2), (5.3) ir (4.2), kai hipotezė neteisinga ir nagrinėjame artėjančių alternatyvų seką, kad necentriškumo parametras δ yra lygus konstantai, pastaruoju atveju $\Delta=4$. Gautos tikimybių reikšmės yra pateiktos 4 priede.

Pereikime prie konkretesnių pavyzdžių.

1. *k = 4* , *p*$_{i0}$ *= 1/4* , hipotezė teisinga

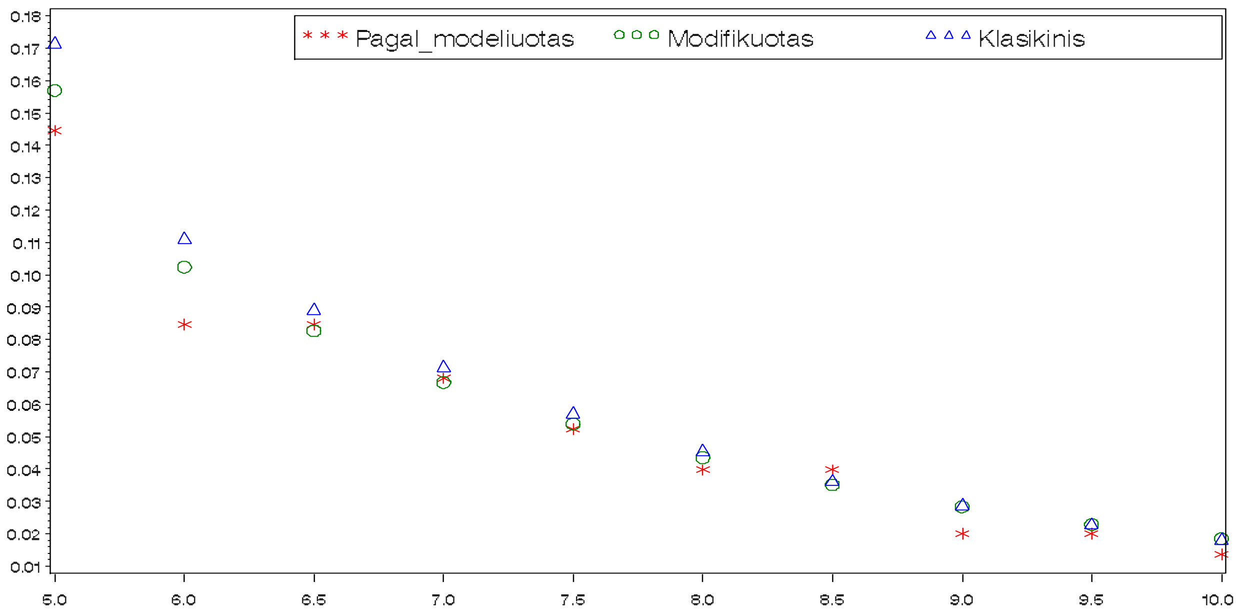
Tarkime, kad tikriname hipotezę:

$$H:\ p_i=p_{i0}=\frac{1}{4},\ i=1,...,4\ . \tag{5.6}$$

Parinkus 10 galimų statistikos X_n^2 realizacijų, pvz. 5; 6; 6,5; 7; 7,5; 8; 8,5; 9; 9,5;10, apskaičiuotos tikimybės

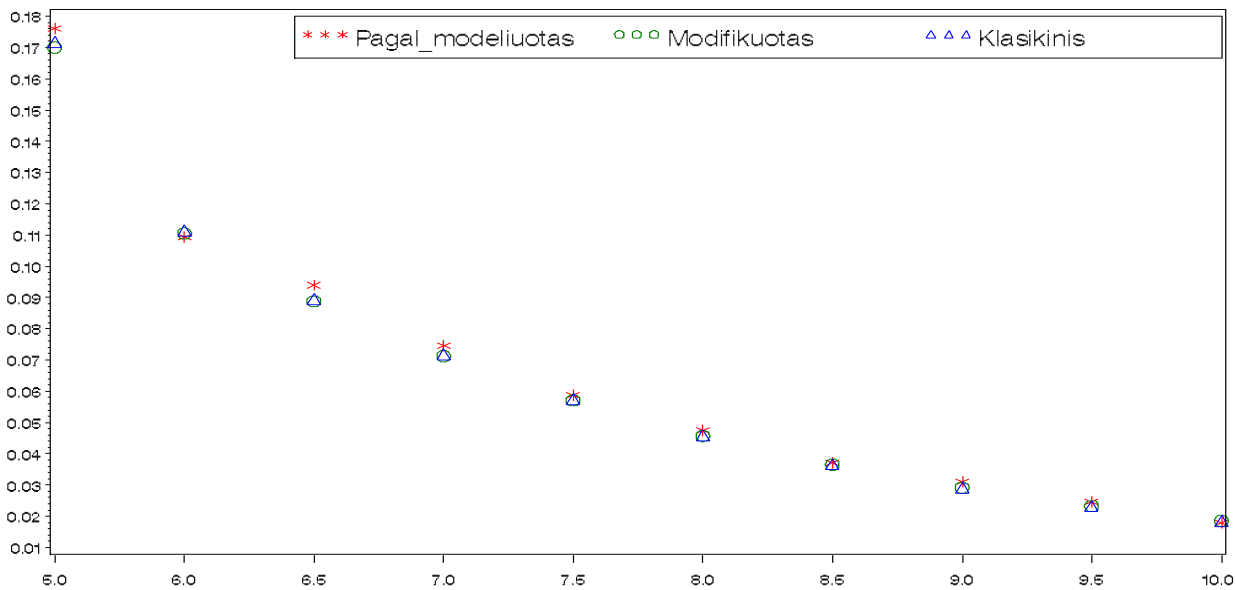
$$\boldsymbol{P}\left\{X_n^2>d_i\right\},\ i=1,...,10\ . \tag{5.7}$$

čia d_i, $i=1,...,10$ yra pasirinktos statistikos X_n^2 realizacijos.



5.1 pav. Kriterijų tikslumo palyginimas, esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, ***n = 12***

Iš paveikslėlio matyti, jog modifikuotąjį kriterijų atitinkančios tikimybės (rutuliukai) yra arčiau kriterijų, kuris remiasi vien modeliuotomis reikšmėmis, atitinkančių tikimybių (žvaigždutės). Taigi galima daryti prielaidą, kad, įvedus statistikos X_n^2 korekciją ir esant nedidelėms imtims, šiek tiek sumažėja tikimybė atmesti teisingą hipotezę. Imties dydžiui didėjant (žr. leteles 1-4 priede nr. 4), visus tris kriterijus atitinkančios tikimybės priartėja vienos prie kitų ir tampa beveik vienodos, todėl tampa nebesvarbu kurį kriterijų yra geriau rinktis, norint patikrinti hipotezę (5.6).

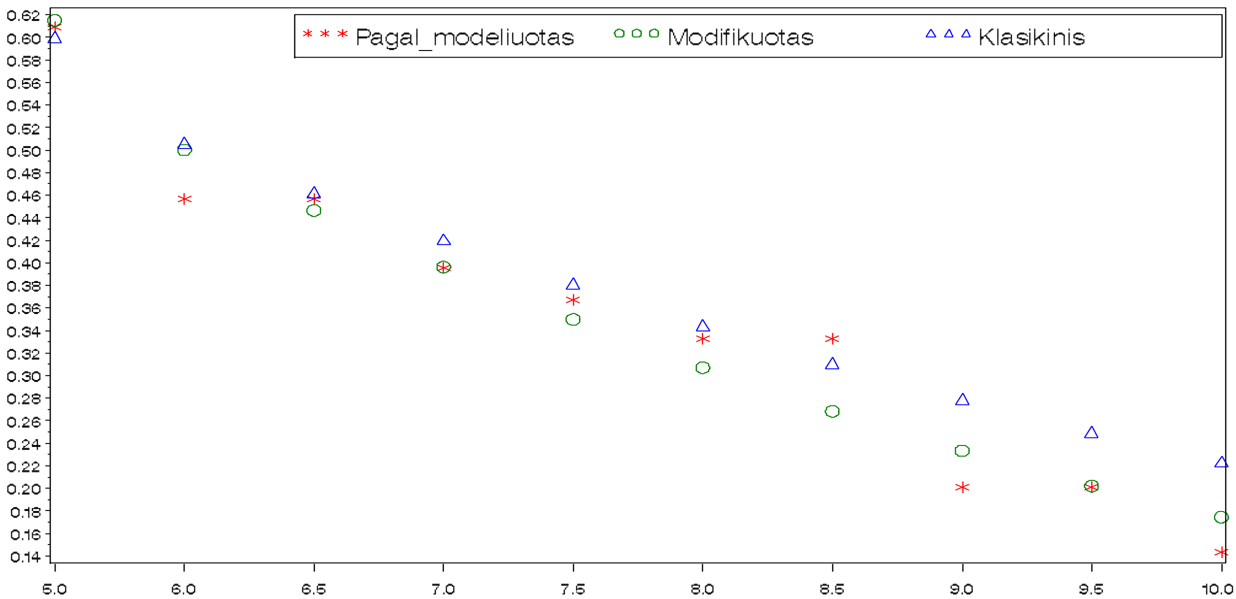


5.2 pav. Kriterijų tikslumo palyginimas, esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, ***n = 100***

2. $k = 4$, $p_{i0} = 1/4$, hipotezė neteisinga

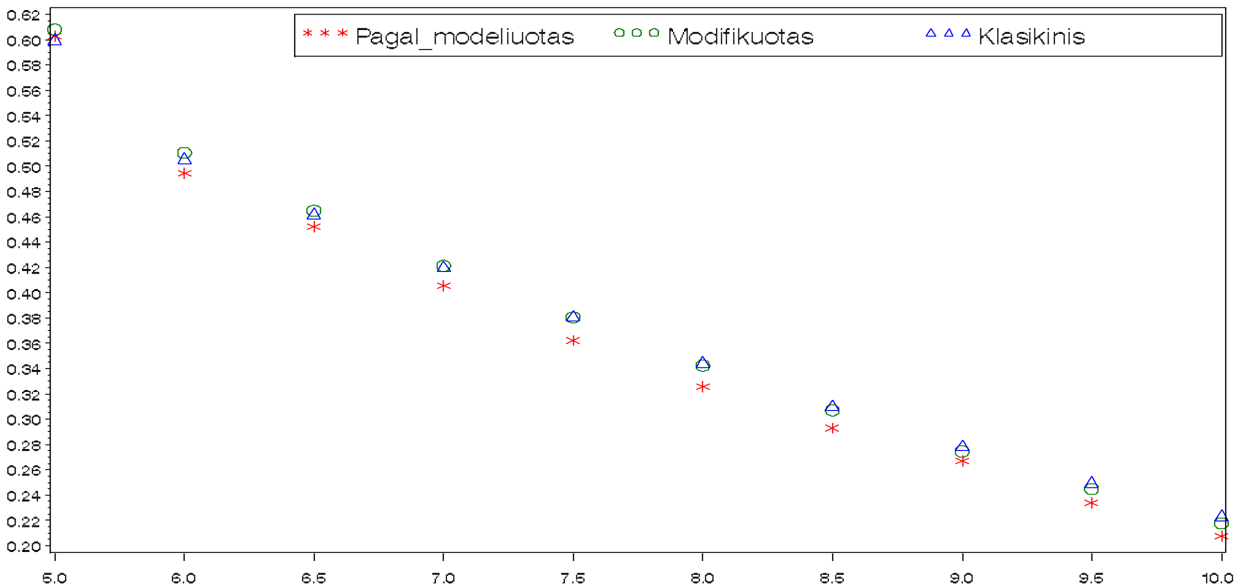
Tarkime, kad hipotezė (5.6) yra neteisinga. Tuomet jai patikrinti naudosime artėjančių alternatyvų seką: $p_i = p_{i0} - \frac{c}{\sqrt{n}} = \frac{1}{4} - \frac{c}{\sqrt{n}}$, $i = 1,2$ ir $p_i = p_{i0} + \frac{c}{\sqrt{n}} = \frac{1}{4} + \frac{c}{\sqrt{n}}$. Kadangi $\Delta = 4$, tai $c = 0,5$. Naudodamiesi tokiomis išraiškomis, apskaičiuojame tikimybes (5.2), (5.3) ir (4.2).

Iš 5.3 pav. galima pastebėti, kad visos tikimybės padidėjo, nes dėl Δ statistikos X_n^2 skirstinio tankis pasislinko į dešinę. Šį kartą nesimato tokios aiškios tendencijos kaip pirmuoju atveju. Kartais galia klasikiniu kriterijumi aproksimuojama geriau negu modifikuotu, kartais atvirkščiai. Galima daryti prielaidą, kad atstumai tarp tikimybių, esant nedidelėms imtims, yra didesni negu esant vienodoms hipotetinėms tikimybėms.



5.3 pav. Kriterijų tikslumo palyginimas, esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 12*

5.4 pav. Parodo, kad vėl, augant imties dydžiui, tikimybės priartėja vienos prie kitų, ypač atitinkančios klasikinį ir modifikuotą kriterijus. Tai tik patvirtina 3.2.7. skyrelyje padarytas išvadas, braižant statistikos X_n^2 vidurkio, dispersijos ir trečiojo centrinio momento elgsenos grafikus, kad korekcija yra labiau reikalinga esant mažesnėms imtims.

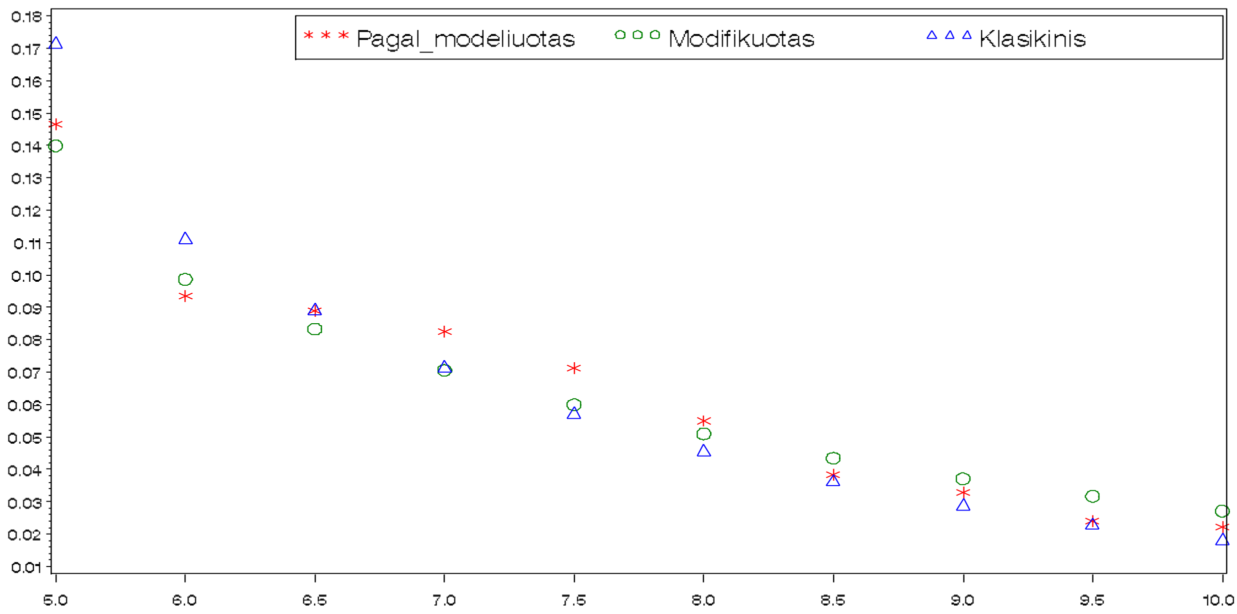


5.4 pav. Kriterijų tikslumo palyginimas, esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 100*

3. $k = 4$, hipotetinės tikimybės nevienodos, hipotezė teisinga. Svarbu pasižiūrėti, kaip atrodo tikimybės (5.1), (5.2), (5.3) ir (4.2), jeigu hipotetinės tikimybės yra nevienodos, pvz., kai tikrinama hipotezė

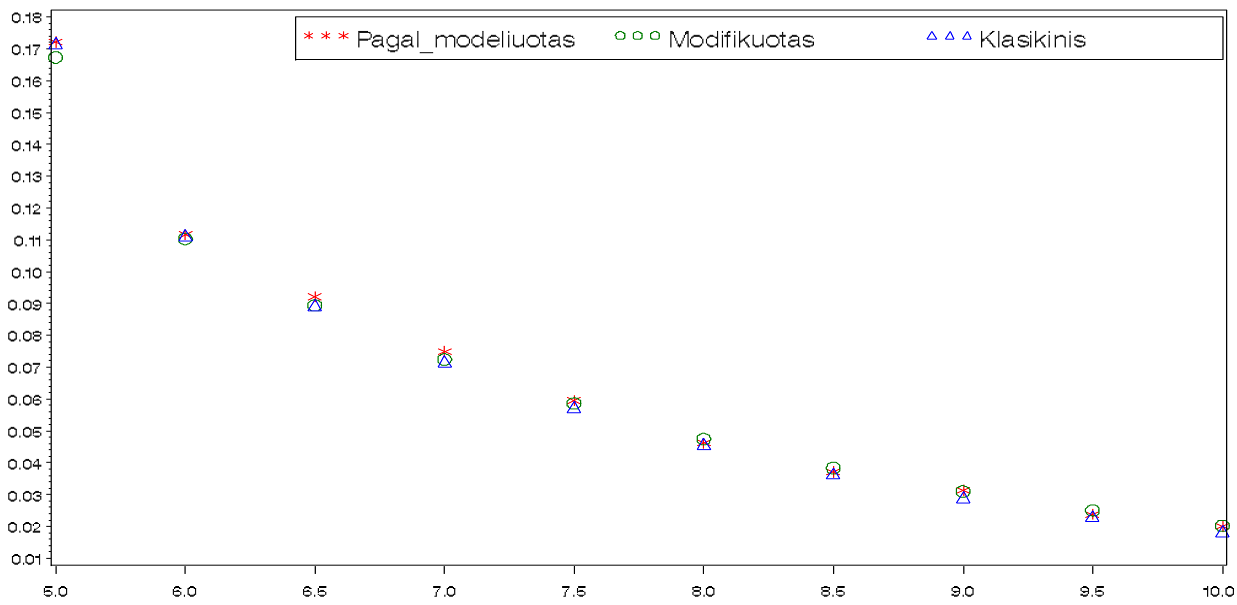
$$H: \; p_1 = p_{10} = \frac{1}{16}, \; p_2 = p_{20} = \frac{3}{16}, \; p_3 = p_{30} = \frac{5}{16}, \; p_4 = p_{40} = \frac{7}{16}. \qquad (5.8)$$

Tarkime, kad ši hipotezė yra teisinga. Tada iš 5.5 pav. pastebime, kad modifikuotą kriterijų atitinkančios tikimybės yra arčiau tikimybių, apskaičiuotų taikant kriterijų, kuris remiasi modeliuotomis reikšmėmis. Taigi net ir esant nevienodoms hipotetinėms tikimybėms, įvesta statistikos X_n^2 korekcija yra naudinga, esant nedidelėms imtims.



5.5. pav. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 12*

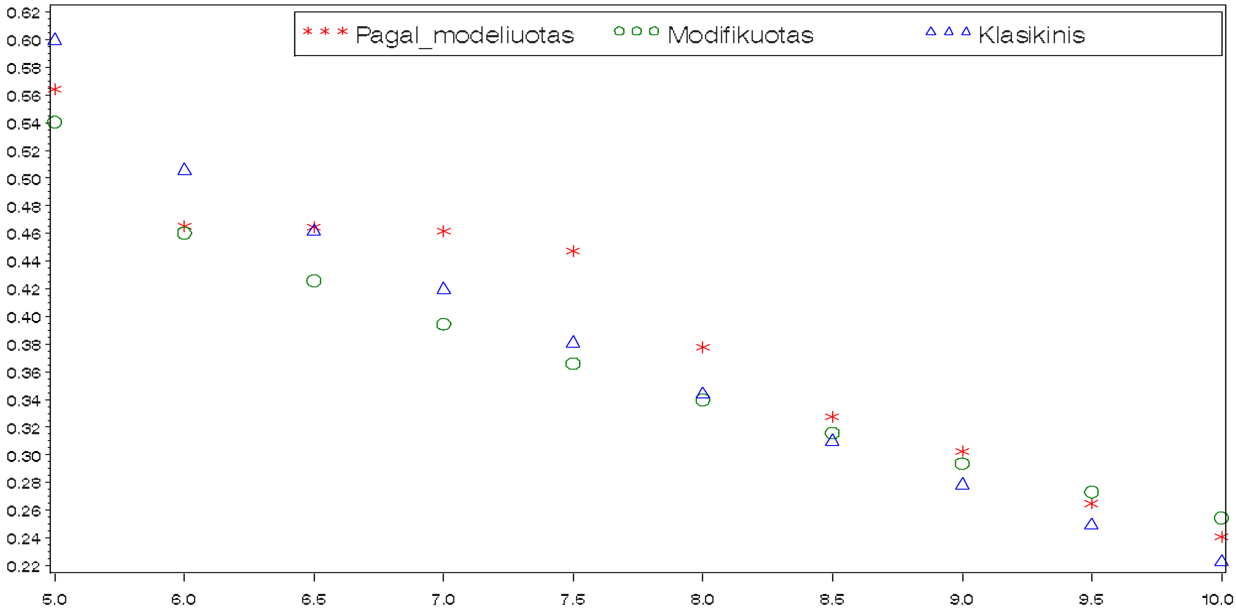
5.6 pav. tik patvrtina prielaidą apie anksčiau pastebėtą tendenciją, jog, didėjant imties dydžiui, tikimybės supanašėja ir tuomet galima taikyti bet kurį kriterijų hipotezei tikrinti, nes jų tikslumas yra vienodas.



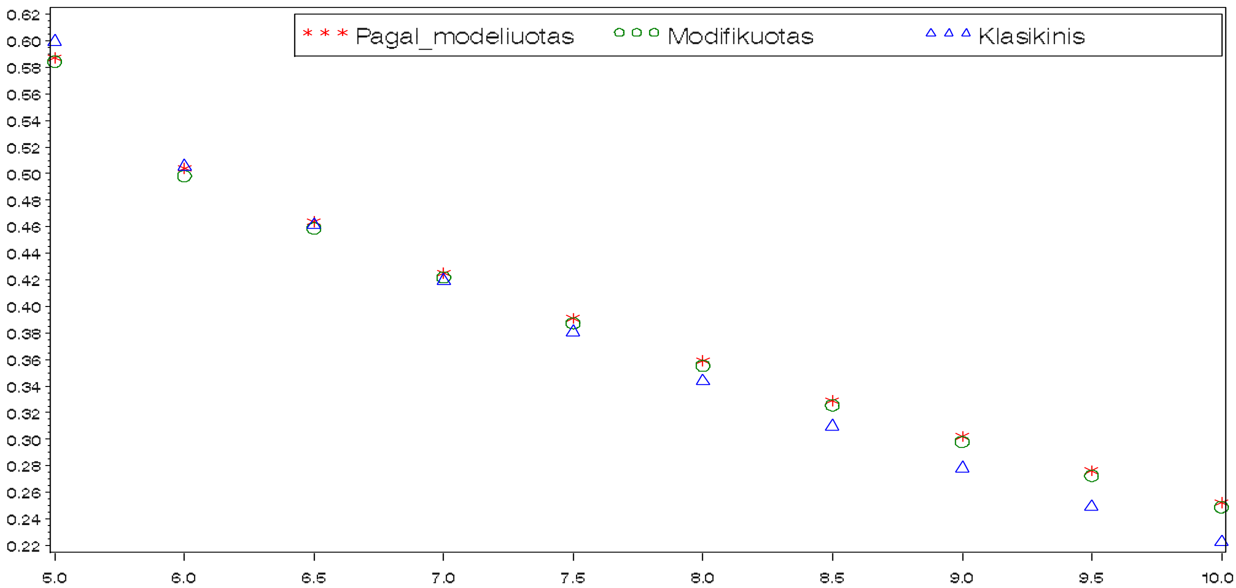
5.6. pav. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 100*

4. $k = 4$, hipotetinės tikimybės nevienodos, hipotezė neteisinga

Tarkime, kad hipotezė (5.8) yra neteisinga. Pažiūrėkime, kaip keičiasi tikimybės šiuo atveju. Matome, kad esant nevienodoms tikimybėms yra sunkiau padaryti išvadą, kuris kriterijus yra tikslesnis. Kai kurie rutuliukai atitinkantys modifikuotą kriterijų yra labiau priartėję prie žvaigždučių, atitinkančių kriterijų, kuris remiasi modeliuotomis reikšmėmis,o kartais trikampiukai, atitinkantys klasikinį kriterijų, yra artimesni žvaigždutėms (5.7 pav.).



5.7. pav. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, $n = 12$



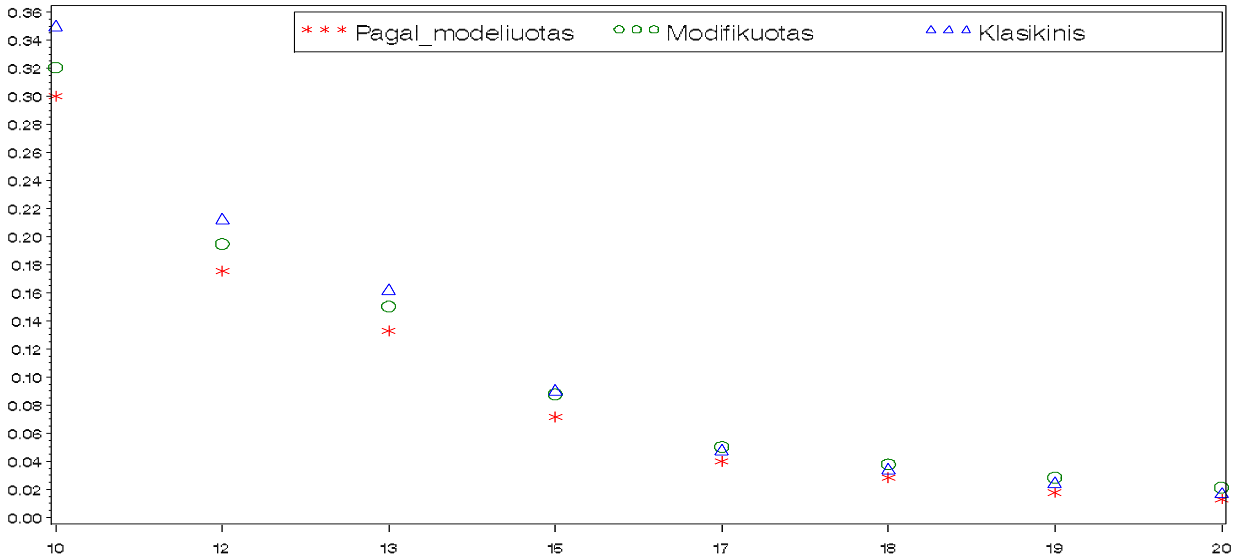
5.8. pav. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, $n = 100$

Tuo tarpu iš 5.8 pav. vėl matome, kad, esant pakankamai didelėms imtims, visos tikimybės tampa artimos.

5. $\boldsymbol{k = 10}$**,** $\boldsymbol{p_i = 1/10}$**, hipotezė teisinga.** Taip pat pravartu paanalizuoti, kaip keičiasi kriterijų tikslumas, keičiantis hipotetinių tikimybių skaičiui, t.y. k. Tarkime, kad tikrinama hipotezė:

$$H: \; p_i = p_{i0} = \frac{1}{10}, \; i = 1,\ldots,10 \,. \tag{5.9}$$

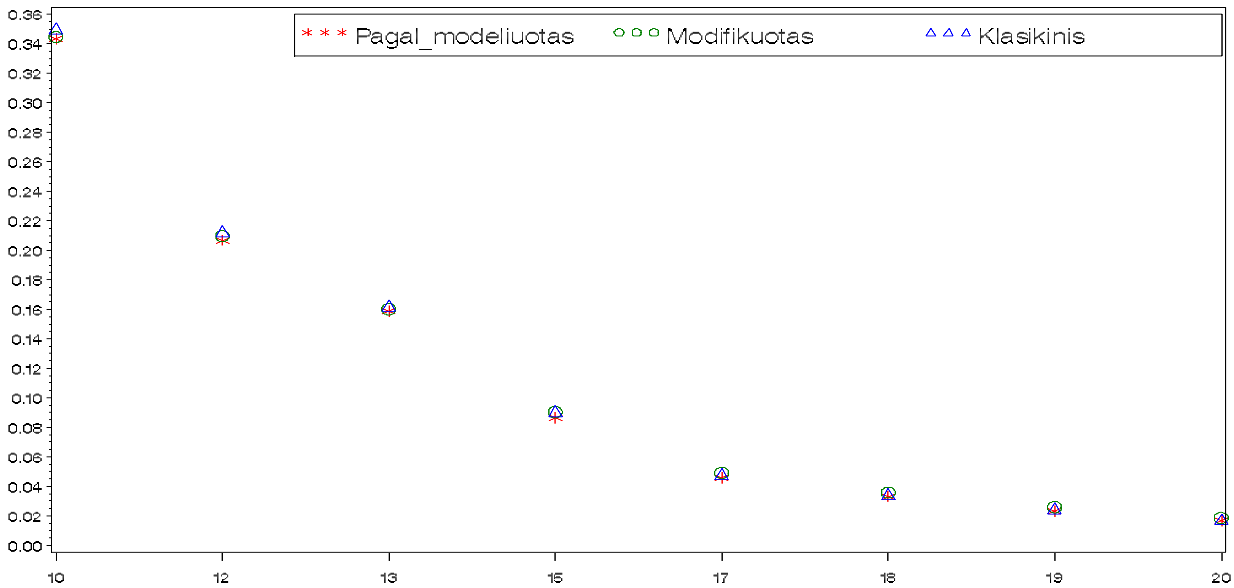
Šiuo atveju hipotetinės tikimybės yra vienodos ir imamos naujos realizacijos tikimybėms (5.1), (5.2), (5.3) ir (4.2) skaičiuoti, pvz. 10, 12, 13, 15, 17, 18, 19, 20.



5.9. pav. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, $\boldsymbol{n = 20}$

Kaip ir pirmuoju atveju, iš 5.9 pav. matome, kad įvesta statistikos X_n^2 korekcija yra naudinga esant nedidelėms imtims ir kai tikrinama hipotezė yra teisinga. Tuomet tikimybė atmesti teisingą hipotezę yra mažesnė.

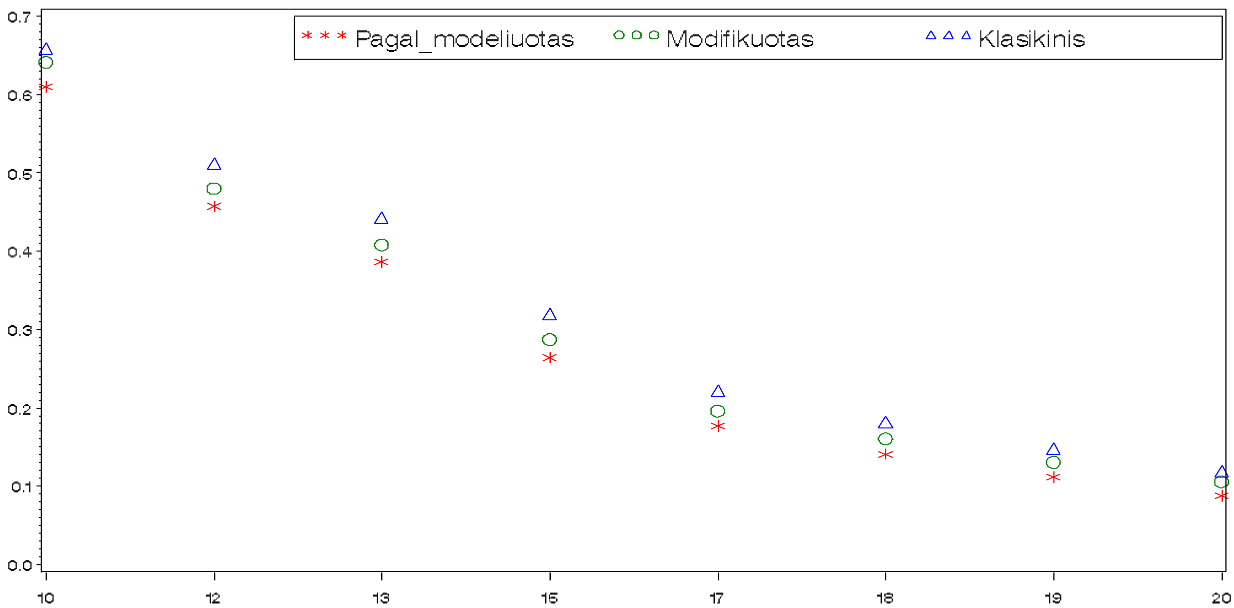
5.10 pav. vėl parodo, jog esant dideliai imčiai visos tikimybės suvienodėja. Galima pastebėti (žr. lenteles 17 – 20 priede nr. 4), kad kuo daugiau hipotetinių tikimybių parenkame, tuo greičiau vyksta artėjimas ir tiksliau aproksimuojamas kriterijaus reikšmingumo lygmuo.



5.10. pav. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 100*

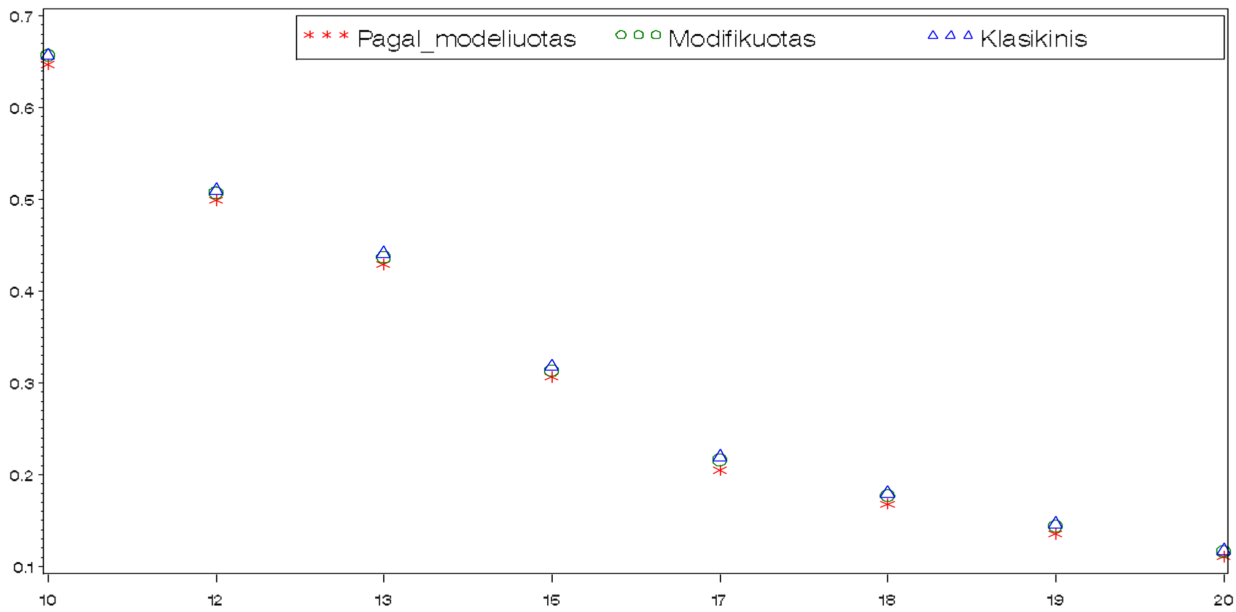
6. $k = 10$, $p_i = 1/10$, hipotezė neteisinga

Tarkime, kad hipotezė (5.9) yra neteisinga. 5.11 pav. labai aiškiai parodo, kad modifikuotu kriterijumi reikšmingumo lygmuo aproksimuojamas tiksliau negu klasikiniu χ^2 kriterijumi.



5.11. pav. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 20*

Net ir esant didelėms imtims, modifikuotas kriterijus yra tikslesnis (5.12 pav.). Taigi kuo daugiau parenkame vienodų hipotetinių tikimybių, tuo tikslesnius rezultatus gauname.

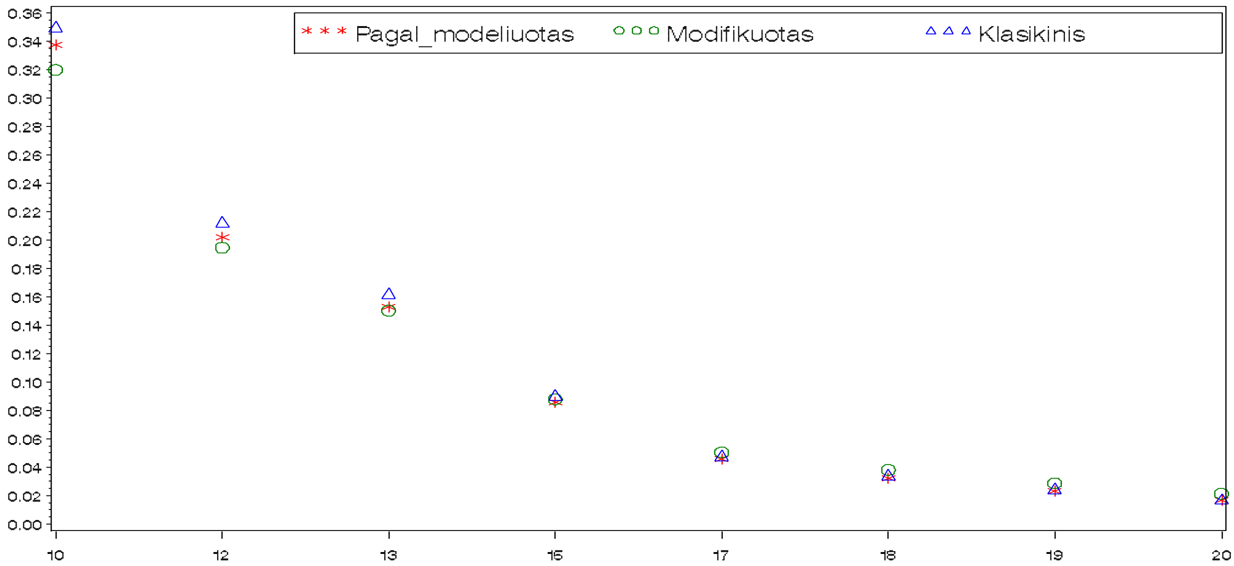


5.12. pav. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 100*

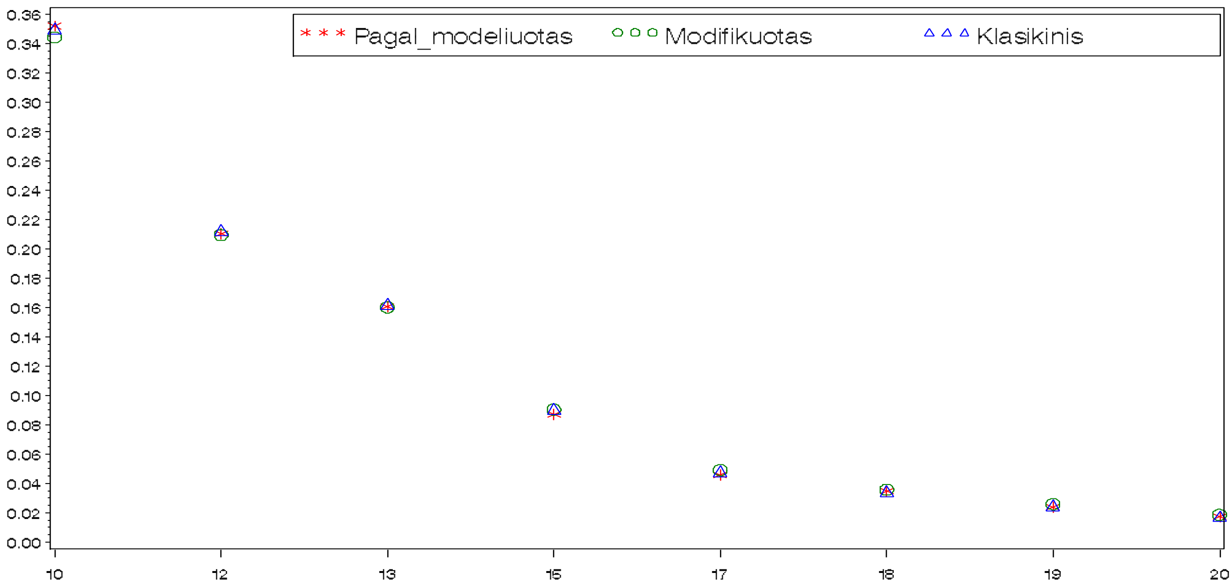
7. $k = 10$, hipotetinės tikimybės nevienodos, hipotezė teisinga Tarkime, kad tikrinama dar viena hipotezė:

$$H:\ p_i = p_{i0} = \frac{9}{100},\ \text{kai}\ i = 1,\dots,5;\ p_i = p_{i0} = \frac{11}{100},\ \text{kai}\ i = 6,\dots,10 \qquad (5.10).$$

Šiuo atveju hipotetinės tikimybės yra skirtingos, o realizacijos imamos tos pačios, kaip ankstesniame uždavinyje. Esant teisingai hipotezei, nesunku pamatyti, kad tikimybės yra labai artimos ir modifikuoto kriterijaus reikšmingumo lygmuo yra šiek tiek mažesnis negu klasikinio kriterijaus(5.13 pav.). Didėjant imties dydžiui n visų kriterijų tikslumas tampa beveik vienodas (žr. pav. 5.14).



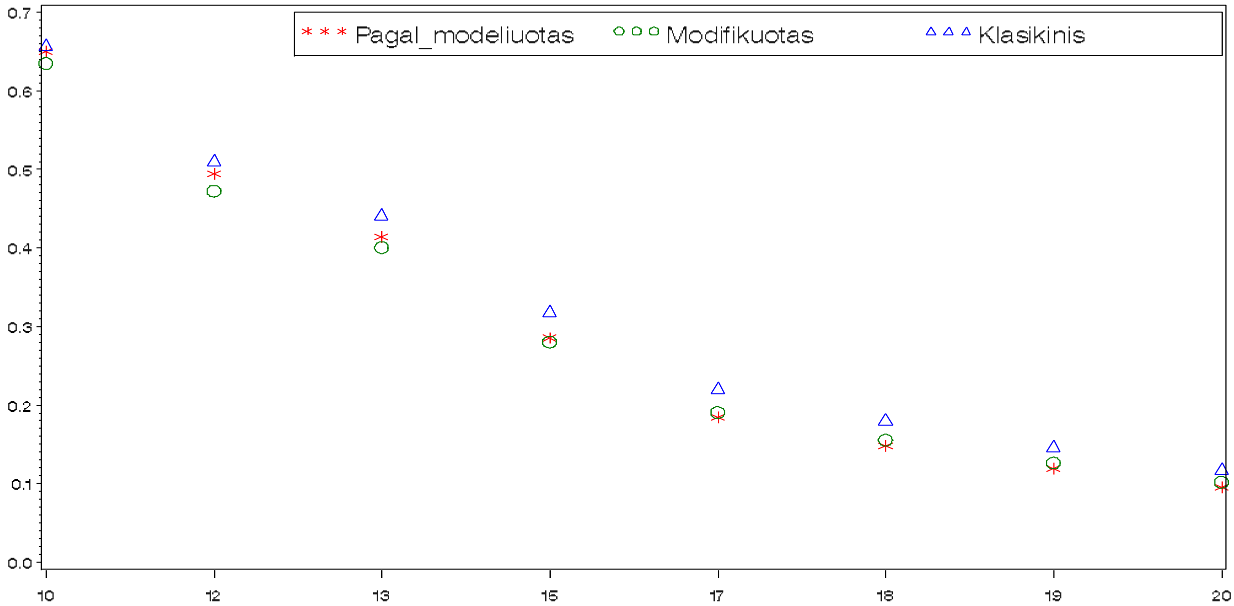
5.13 pav. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 20*



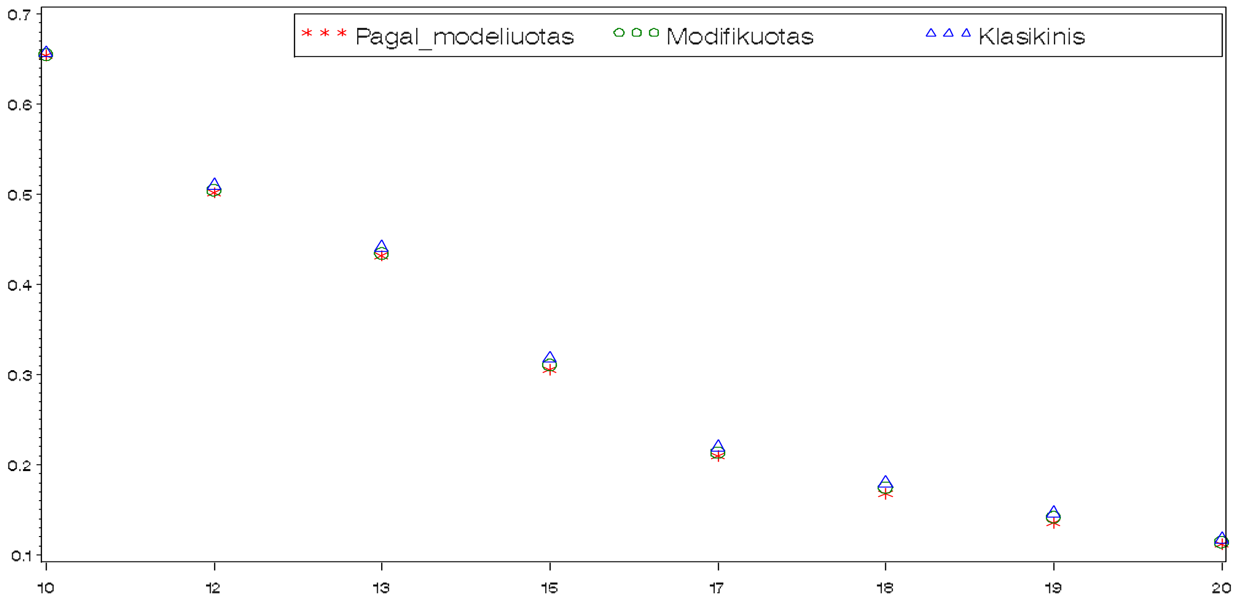
5.14. pav. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 100*

7. $k = 10$, hipotetinės tikimybės nevienodos, hipotezė neteisinga

Kai hipotezė neteisinga, net ir esant nevienodoms hipotetinėms tikimybėms, iš 5.15 ir 5.16 pav. matome, kad buvo naudinga įvesti statistikos X_n^2 korekciją. Modifikuotą kriterijų atitinkančios tikimybės yra beveik lygios kriterijų, kuris remiasi modeliuotomis reikšmėmis, atitinkančioms tikimybėms, kai imame įvairaus dydžio imtis. Taip pat pasitvirtina prielaida, kad, didinant hipotetinių tikimybių skaičių, didėja kriterijų tikslumas ir išryškėja statistikos X_n^2 korekcijos naudingumas.



5.15. pav. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, $n = 20$



5.16. pav. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, $n = 100$

IŠVADOS

Šiame magistro baigiamajame darbe aptarti χ^2 kriterijaus paprastajai suderinamumo hipotezei tikrinti ypatumai bei jo tikslumo priklausomybė nuo duomenų suskaidymo į intervalus taisyklių. Pastebėta, jog geriau yra, kai hipotetinės tikimybės parenkamos vienodos: skaičiavimai tuomet yra paprastesni, o kriterijaus statistikos X_n^2 vidurkis, dispersija ir tečiasis centrinis momentai yra artimesni teoriniams jų atitikmenims. Siekant patikslinti kriterijų, įvesta statistikos X_n^2 korekcija $V = a + bX_n^2$, (kur a ir b yra parametrai, išspręsti iš tam tikros lygčių sistemos), ir, remiantis ja, sukonstruotas naujas kriterijus.

Taip pat pasiūlytas papildomas kriterijus hipotezei tikrinti. Jis remiasi modeliuotomis statistikos X_n^2 reikšmėmis ir konstruojamas lyginant P-reikšmės įvertį su reikšmingumo lygmeniu α. Šio kriterijaus tikslumas priklauso tik nuo statistikos modeliuotų reikšmių skaičiaus. Beje, naudojant šiuolaikines technologijas, P-reikšmės įvertį yra gana paprasta apskaičiuoti realiomis kompiuterinio darbo laiko sąnaudomis.

Galiausiai atlikta palyginamoji trijų kriterijų analizė: įvestos tam tikros galimos realizacijos ir kiekvienai iš jų apskaičiuotas reikšmingumo lygmuo arba galia (priklausomai nuo to, ar tikrinama hipotezė teisinga ar neteisinga) kiekvieno kriterijaus atveju. Stebėta, kaip keičiasi rezultatai, naudojant vienodas ir nevienodas hipotetines tikimybes bei didinant imties dydį.

Atliktos analizės išvados:

- Kriterijus, kuris remiasi modeliuotomis reikšmėmis, yra pakankamai tikslus ir jo pakanka hipotezėms tikrinti, jeigu sumodeliuojama užtektinai reikšmių;
- Įvesta statistikos X_n^2 korekcija šiek tiek patikslina rezultatus tikrinant hipotezę, ir tuomet, kai ji yra teisinga, ir tuomet, kai neteisinga;
- Didėjant imties dydžiui, visi trys kriterijai tampa beveik visiškai vienodai tikslūs ir klausimas, kuris yra labiau tinkamas, nebetenka prasmės;
- Didinant hipotetinių tikimybių skaičių k ir imant jas vienodas, rezultatai gaunami tikslesni;

o Siekiant palengvinti darbą SAS paketo aplinkoje sukurtos programos vertinti modifikuoto kriterijaus parametrus bei modeliuoti statistikos X_n^2 reikšmes.

SANTRAUKA

Šiame baigiamajame magistro darbe analizuojamos χ^2 kriterijaus suderinamumo hipotezėms tikrinti savybės. Siekiama jį patikslinti, įvedant kriterijaus statistikos X_n^2 korekciją $V = a + bX_n^2$. Atlikus klasikinio ir modifikuoto kriterijų tikslumo palyginamąją analizę, naudojant modeliuotas X_n^2 reikšmes, padaryta keletas išvadų. Kriterijus, kuris remiasi statistika V, šiek tiek patikslina hipotezių tikrinimo rezultatus, tačiau, kai imtis yra pakankamai didelė, kriterijų tikslumas beveik nesiskiria. Be to, kriterijus, kuris remiasi modeliuotomis reikšmėmis taip pat yra pakankamai tikslus. Taigi, naudojantis šiuolaikinėmis pažangiomis technologijomis, ir, modeliuojant pakankamai didelę a.d. X_n^2 imtį, šio kriterijaus galėtų pakakti, kad patikrinti suderinamumo hipotezę. Skaičiavimų palengvinimui sukurtos makroprogramos a.d. X_n^2 momentams skaičiuoti ir a.d. X_n^2 realizacijoms modeliuoti.

MODIFIED CHI SQUARE GOODNESS OF FIT TEST

Paper for the Master‘s degree

Vilnius University, Faculty of Mathematics and Informatics

Supervisor doc. J. Kruopis

Vilnius 2009

SUMMARY

The main purpose of this master thesis is to improve the clasical chi square goodness of fit test so that it would depend less on the rule, according to which the data is divided into intervals, and on the hypothetical probabilities.

The tasks of the paper are the following:

- To examine (according to the literature ([1], [2], [3])) theoretical and applied aspects of constructing chi square test;
- To analyse the methods of graphic representation of the empirical characteristics of the data and graphic comparison with their theoretical equivalents. Altogether to look through the ways of dividing the data into intervals (determining their number and length);
- To calculate the mean, variance and third central moment of chi square test statistics X_n^2 and to analyse their properties;
- To create a macro for counting the previously mentioned three moments using SAS software;
- To create a macro for modeling statistics X_n^2 and to construct goodness of fit test according to the modeled values using SAS software;
- To construct an aproximate modified chi square goodness of fit test based on a new statistics;
- To perform comparison of the clasical and modified goodness of fit tests, having the modeled values;

Analysis of graphing the empirical data and comparison with the theoretical equivalents has showed that the distribution (density) is recognised easier when the probabilities of values to be prescribed to certain intervals are equal. This way the execution of computations is easier as well as the mean, variance and the third central moment of statistics X_n^2 are closer to their theoretical equivalents.

The analysis of mean, variance and the third central moment of statistics X_n^2 has been performed according to the changes of the sample size. It has showed that when a sample is small, there is a need for correction of statistics X_n^2 since the empirical values of the moments are rather far from their theoretical equivalents.

Referring to the delineated before, a modification V of statistics X_n^2 has been imposed: $V = a + bX_n^2$.

In addition, one more test has been proposed. It is based only on the modeled values of the statistics X_n^2 and the p-value.

Afterwords, the comparison of the three tests (clasical chi square, modified chi square goodness of fit tests and the test based on the modeled values) has been executed. The method used was comparing the probabilities $\boldsymbol{P}\left\{X_n^2 > d\right\}$ according to each of the tests. The results were there following:

- The test based on the modeled values is rather precise and it can be sufficient for testing hypothesis if there are enough modeled values;
- The correction of the statistics X_n^2 makes the goodness of fit test slightely more accurate in both cases when the tested hypothesis is correct and incorrect;
- While the sample size is increasing the accuracies of the tests become almost equal and afterwards the question which of the tests is better to be applied is not actual anymore;
- The more hypothetical probabilities are included in the test the more precise results are gained as well as in the case when the hypothetical probabilities are equal.

LITERATŪROS SĄRAŠAS:

1. Кендалл М. Г., Стюарт А. „Статистические выводы и связи“. Москва „Наука“,1972
2. J. Kruopis „Matematinė statistika“. Vilnius: „Mokslas“,1977.
3. V. Bagdonavičius, J. Kruopis „Matematinė statistika“. II dalis (rankraštis).
4. Крамер Г. „Математические методы статистики“. Москва „Мир“, 1975.
5. V. Bagdonavičius, J. Kruopis „Matematinė statistika“, I dalis. Vilnius: „TEV“, 2007.
6. Levulienė R. „Statistika su SAS“. http//www.mif.vu.lt/~rutal72, 2005.

Priedas 1

Literatūros, kurioje nagrinėjami χ^2 tipo kriterijai, sąrašas iš Rusijos valstybinio standarto „Empirinio skirstinio suderinamumo su teoriniu tikrinimo taisyklės. χ^2 tipo kriterijai (P50.1033-2001)“.

Библиография

[1] Кендалл М., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973. — 900 с.

[2] Никулин М. С. Критерий хи-квадрат для непрерывных распределений с параметрами сдвига и масштаба//Теория вероятностей и ее применение. — 1973. — Т.XVIII. — № 3. — С. 583—591

[3] Никулин М. С. О критерии хи-квадрат для непрерывных распределений//Теория вероятностей и ее применение. — 1973. — Т. XVIII. — № 3. — С. 675—676

[4] Мирвалиев М., Никулин М. С. Критерии согласия типа хи-квадрат//Заводская лаборатория. — 1992. — Т. 58. — № 3. — С. 52—58

[5] Aguirre N., Nikulin M. Chi-squared goodness-of-fit test for the family of logistic distributions//Kybernetika. — 1994. — V. 30. — № 3. — P. 214—222

[6] Chernoff H., Lehmann E. L. The use of maximum likelihood estimates in χ^2 test for goodness of fit//Ann. Math. Stat., 1954. — V. 25. — P. 579—586

[7] Чибисов Д. М. Некоторые критерии типа хи-квадрат для непрерывных распределений//Теория вероятностей и ее применение. — 1971. — Т. XVI. — № 1. — С. 3—20

[8] Рао С. Р. Линейные статистические методы и их применения. — М.: Наука, 1968. — 548 с.

[9] Кокс Д., Хинкли Д. Теоретическая статистика. — М.: Мир, 1978. — 560 с.

[10] Денисов В. И., Лемешко Б. Ю. Оптимальное группирование при обработке экспериментальных данных//Измерительные информационные системы. — Новосибирск, 1979. — С. 5—14

[11] Денисов В. И., Зачепа Г. Г., Лемешко Б. Ю. Об асимптотически оптимальном группировании при оценивании основного параметра гамма-распределения по группированным данным//Применение ЭВМ в оптимальном планировании и проектировании. — Новосибирск, 1974. — С. 50—53

[12] Лемешко Б. Ю. Об оценивании параметров распределений по группированным наблюдениям//Вопросы кибернетики. — М., 1977. — Вып. 30. — С. 80—96

[13] Лемешко Б. Ю. Об оценивании по частично группированным выборкам параметров распределения Эрланга//Применение ЭВМ в оптимальном планировании и проектировании. — Новосибирск. 1982. — С. 90—93

[14] Лемешко Б. Ю., Панфилов А. Г., Панфилова Л. В. Асимптотически оптимальное группирование данных при анализе наблюдений, распределенных по законам Лапласа и двойного показательного//Математическое обеспечение стохастических и детерминированных моделей. — Новосибирск. 1986. — С. 15—23

[15] Лемешко Б. Ю. К вопросу решения задачи асимптотически оптимального группирования данных при обработке наблюдений, подчиняющихся бета-распределению//Машинные методы оптимизации, моделирования и планирования эксперимента. — Новосибирск, 1988. — С.134—138

[16] Денисов В. И., Лемешко Б. Ю., Цой Е. Б. Оптимальное группирование, оценка параметров и планирование регрессионных экспериментов. В 2 ч. — Новосибирск: Изд-во НГТУ, 1993. — 347 с.

[17] Лемешко Б. Ю. Статистический анализ одномерных наблюдений случайных величин: Программная система. — Новосибирск: Изд-во НГТУ, 1995. — 125 с.

[18] Лемешко Б. Ю. Асимптотически оптимальное группирование наблюдений — это обеспечение максимальной мощности критериев//Надежность и контроль качества. — 1997. — № 8. — С. 3—14

[19] Лемешко Б. Ю. Асимптотически оптимальное группирование наблюдений в критериях согласия//Заводская лаборатория. — 1998. — Т. 64. — № 1 — С. 56—64

[20] Денисов В. И., Лемешко Б. Ю., Постовалов С. Н. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Методические рекомендации. Часть I. Критерии типа χ^2. — Новосибирск: Изд-во НГТУ, 1998. — 126 с.

[21] Mann H. B., Wald A. On the choice of the number of class intervals in the application of the chi square test//Ann. Math. Stat., 1942. V. 13. — P. 306—317.

[22] Лемешко Б. Ю. Постовалов С. Н. Прикладные аспекты использования критериев согласия в случае проверки сложных гипотез//Надежность и контроль качества. — 1997. — № 11. — С. 3—17

[23] Лемешко Б. Ю. Постовалов С. Н. О зависимости предельных распределений статистик χ^2 Пирсона и отношения правдоподобия от способа группирования данных//Заводская лаборатория. — 1998. — Т. 64. — № 5. — С. 56—63

[24] Лемешко Б. Ю. Постовалов С. Н. О распределениях статистик непараметрических критериев согласия при оценивании по выборкам параметров наблюдаемых законов//Заводская лаборатория. — 1998. — Т. 64. — № 3. — С. 61—72

[25] Новицкий П. В., Зограф И. А. Оценка погрешностей результатов измерений. — Л.: Энергоатомиздат. 1991. — 303 с.

[26] Cochran W. G. Some Methods of Strengthening the Common χ^2 Tests//Biometrics, 1954. — V. 10. — P. 417

[27] Mann H. B., Wald A. On the choice of the number of intervals in the application of the chi-square test//Ann. Math. Stat., 1942. — V. 13. — P. 478—479

Priedas 1

[28] Sturgess H. A. The choice of classic intervals//J. Am. Statist. Assoc. — March 1926. — 47 p.
[29] Шторм Р. Теория вероятностей. Математическая статистика. Статистический контроль качества. — М.: Мир, 1970. — 368 с.
[30] Heinhold I., Gaede K. W. Ingeniur statistic. — München; Wien, Springler Verlag, 1964. — 352 s.
[31] Таушанов З., Тонева Е., Пенова Р. Вычисление энтропийного коэффициента при малых выборках//Изобретательство, стандартизация и качество. — 1973. — № 5. — София
[32] Тонева Е. Аппроксимация распределений погрешности средств измерений//Измерительная техника. — 1981. — № 6. — С. 15—16
[33] Алексеева И. У. Теоретическое и экспериментальное исследование законов распределения погрешностей, их классификация и методы оценки их параметров: Автореф. дис. на соиск. учен. степени кан. техн. наук. — Л., 1975. — 20 с.

2 Priedas

Dydžio $n = 200$ normaliojo atsitiktinio dydžio modeliuota imtis

i	t	i	t	i	t	i	t
1	-1,38033	**51**	0,263282	**101**	0,931221	**151**	1,620111
2	-1,23674	**52**	0,268439	**102**	0,944852	**152**	1,626607
3	-1,17861	**53**	0,285822	**103**	0,947397	**153**	1,63845
4	-1,02991	**54**	0,292034	**104**	0,950201	**154**	1,645171
5	-0,71567	**55**	0,301982	**105**	0,95074	**155**	1,659458
6	-0,6919	**56**	0,331975	**106**	0,96694	**156**	1,675699
7	-0,69059	**57**	0,345369	**107**	0,972697	**157**	1,718866
8	-0,68621	**58**	0,345619	**108**	0,988954	**158**	1,739087
9	-0,66215	**59**	0,356336	**109**	0,989598	**159**	1,773428
10	-0,65889	**60**	0,38401	**110**	1,00284	**160**	1,779622
11	-0,64678	**61**	0,398331	**111**	1,009089	**161**	1,814679
12	-0,64183	**62**	0,403123	**112**	1,017674	**162**	1,823131
13	-0,63115	**63**	0,406127	**113**	1,024082	**163**	1,847633
14	-0,6186	**64**	0,421883	**114**	1,02827	**164**	1,900723
15	-0,60114	**65**	0,429699	**115**	1,059279	**165**	1,909382
16	-0,58209	**66**	0,430655	**116**	1,059537	**166**	1,972216
17	-0,5741	**67**	0,431258	**117**	1,069815	**167**	2,031813
18	-0,52974	**68**	0,46167	**118**	1,078057	**168**	2,043475
19	-0,52776	**69**	0,469545	**119**	1,12858	**169**	2,086171
20	-0,36792	**70**	0,471742	**120**	1,129046	**170**	2,118251
21	-0,36449	**71**	0,487062	**121**	1,137561	**171**	2,141879
22	-0,35613	**72**	0,524196	**122**	1,139834	**172**	2,159822
23	-0,34961	**73**	0,539951	**123**	1,162299	**173**	2,187452
24	-0,2849	**74**	0,541106	**124**	1,213353	**174**	2,199227
25	-0,28375	**75**	0,558218	**125**	1,255957	**175**	2,216325
26	-0,27892	**76**	0,561508	**126**	1,277496	**176**	2,244373
27	-0,26497	**77**	0,566991	**127**	1,293083	**177**	2,322075
28	-0,24641	**78**	0,572311	**128**	1,323378	**178**	2,4031
29	-0,2179	**79**	0,572629	**129**	1,336173	**179**	2,405428
30	-0,20181	**80**	0,575272	**130**	1,344407	**180**	2,416285
31	-0,15624	**81**	0,609365	**131**	1,346211	**181**	2,417716
32	-0,15127	**82**	0,638722	**132**	1,34785	**182**	2,423281
33	-0,14047	**83**	0,646232	**133**	1,371053	**183**	2,432626
34	-0,13359	**84**	0,659725	**134**	1,371363	**184**	2,448098
35	-0,09598	**85**	0,7018	**135**	1,377792	**185**	2,452837
36	-0,09139	**86**	0,703384	**136**	1,385702	**186**	2,469882
37	-0,02908	**87**	0,715199	**137**	1,414233	**187**	2,519816
38	-0,02231	**88**	0,723033	**138**	1,460113	**188**	2,55172
39	-0,01759	**89**	0,781236	**139**	1,462134	**189**	2,563901
40	0,023947	**90**	0,788326	**140**	1,47941	**190**	2,633169
41	0,043867	**91**	0,826614	**141**	1,508069	**191**	2,71138
42	0,126677	**92**	0,832956	**142**	1,513563	**192**	2,76831
43	0,183155	**93**	0,843276	**143**	1,538197	**193**	2,827768
44	0,192015	**94**	0,857966	**144**	1,573948	**194**	2,980797
45	0,2006	**95**	0,864331	**145**	1,583981	**195**	3,007404
46	0,216128	**96**	0,873789	**146**	1,589215	**196**	3,057473
47	0,217228	**97**	0,886007	**147**	1,592473	**197**	3,135568
48	0,245902	**98**	0,901183	**148**	1,610498	**198**	3,329717
49	0,249842	**99**	0,911131	**149**	1,616209	**199**	3,341454
50	0,253282	**100**	0,930324	**150**	1,619199	**200**	3,384354

Dydžio $n = 200$ puasono atsitiktinio dydžio imtis

i	p	i	p	i	p	i	p
1	5	**51**	4	**101**	2	**151**	3
2	4	**52**	4	**102**	3	**152**	0
3	9	**53**	0	**103**	4	**153**	4
4	3	**54**	2	**104**	4	**154**	3
5	2	**55**	3	**105**	1	**155**	2
6	4	**56**	8	**106**	7	**156**	3
7	3	**57**	4	**107**	3	**157**	0
8	2	**58**	2	**108**	4	**158**	5
9	3	**59**	2	**109**	3	**159**	2
10	4	**60**	3	**110**	1	**160**	1
11	1	**61**	2	**111**	3	**161**	0
12	2	**62**	3	**112**	2	**162**	2
13	3	**63**	2	**113**	4	**163**	2
14	4	**64**	1	**114**	3	**164**	1
15	3	**65**	2	**115**	3	**165**	5
16	7	**66**	3	**116**	4	**166**	2
17	4	**67**	3	**117**	4	**167**	2
18	5	**68**	1	**118**	3	**168**	2
19	4	**69**	0	**119**	3	**169**	6
20	2	**70**	4	**120**	2	**170**	1
21	5	**71**	4	**121**	1	**171**	5
22	1	**72**	1	**122**	3	**172**	5
23	3	**73**	2	**123**	2	**173**	2
24	5	**74**	4	**124**	3	**174**	1
25	4	**75**	3	**125**	3	**175**	2
26	3	**76**	1	**126**	4	**176**	6
27	2	**77**	5	**127**	1	**177**	2
28	5	**78**	3	**128**	1	**178**	3
29	5	**79**	3	**129**	4	**179**	5
30	4	**80**	4	**130**	2	**180**	3
31	2	**81**	2	**131**	5	**181**	1
32	1	**82**	4	**132**	3	**182**	3
33	2	**83**	3	**133**	5	**183**	2
34	2	**84**	2	**134**	4	**184**	4
35	5	**85**	1	**135**	3	**185**	2
36	7	**86**	2	**136**	3	**186**	4
37	5	**87**	2	**137**	3	**187**	2
38	4	**88**	1	**138**	6	**188**	7
39	4	**89**	4	**139**	2	**189**	3
40	4	**90**	5	**140**	4	**190**	0
41	4	**91**	2	**141**	5	**191**	1
42	4	**92**	8	**142**	5	**192**	6
43	7	**93**	2	**143**	2	**193**	3
44	2	**94**	4	**144**	6	**194**	1
45	0	**95**	4	**145**	6	**195**	5
46	4	**96**	3	**146**	2	**196**	3
47	3	**97**	2	**147**	5	**197**	2
48	2	**98**	2	**148**	5	**198**	7
49	1	**99**	4	**149**	6	**199**	0
50	2	**100**	2	**150**	2	**200**	3

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,1446	0,157036	0,171797
6	0,0846	0,102457	0,11161
6,5	0,0846	0,082751	0,089663
7	0,0681	0,066833	0,071898
7,5	0,0524	0,053974	0,057558
8	0,0398	0,043588	0,046012
8,5	0,0398	0,0352	0,036733
9	0,0202	0,028425	0,029291
9,5	0,0202	0,022953	0,023331
10	0,0136	0,018534	0,018566

1 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 12*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,1875	0,162984	0,171797
6	0,098	0,106159	0,11161
6,5	0,0907	0,085563	0,089663
7	0,0734	0,068912	0,071898
7,5	0,0635	0,055466	0,057558
8	0,0386	0,044618	0,046012
8,5	0,0321	0,035873	0,036733
9	0,0285	0,028829	0,029291
9,5	0,024	0,023159	0,023331
10	0,0166	0,018597	0,018566

2 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 20*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,1708	0,168288	0,171797
6	0,1106	0,109445	0,11161
6,5	0,0882	0,08804	0,089663
7	0,0748	0,070723	0,071898
7,5	0,0568	0,056743	0,057558
8	0,0514	0,045476	0,046012
8,5	0,0377	0,036411	0,036733
9	0,0327	0,029128	0,029291
9,5	0,025	0,023282	0,023331
10	0,018	0,018596	0,018566

3 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 50*

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,1761	0,170045	0,171797
6	0,1094	0,11053	0,11161
6,5	0,094	0,088854	0,089663
7	0,0747	0,071314	0,071898
7,5	0,0589	0,057154	0,057558
8	0,0476	0,045748	0,046012
8,5	0,0371	0,036576	0,036733
9	0,0312	0,029213	0,029291
9,5	0,0246	0,02331	0,023331
10	0,0179	0,018584	0,018566

4 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 100*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,6093	0,615219	0,600666
6	0,4569	0,500345	0,506838
6,5	0,4569	0,446704	0,462842
7	0,3958	0,396425	0,421149
7,5	0,3673	0,34986	0,38192
8	0,3327	0,307187	0,345245
8,5	0,3327	0,268443	0,311157
9	0,2011	0,233557	0,27964
9,5	0,2011	0,202376	0,250641
10	0,1439	0,174692	0,224078

5 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 12*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,6351	0,614514	0,600666
6	0,4636	0,508733	0,506838
6,5	0,4528	0,458889	0,462842
7	0,4174	0,411772	0,421149
7,5	0,3896	0,367703	0,38192
8	0,2942	0,32687	0,345245
8,5	0,2727	0,289351	0,311157
9	0,2532	0,255135	0,27964
9,5	0,2272	0,22414	0,250641
10	0,189	0,196236	0,224078

6 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 20*

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,6006	0,610205	0,600666
6	0,4976	0,510911	0,506838
6,5	0,4459	0,464118	0,462842
7	0,4107	0,419772	0,421149
7,5	0,3587	0,378125	0,38192
8	0,3394	0,339327	0,345245
8,5	0,2879	0,303441	0,311157
9	0,2684	0,270461	0,27964
9,5	0,2341	0,240328	0,250641
10	0,201	0,21294	0,224078

7 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 50*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,603	0,608161	0,600666
6	0,4943	0,510747	0,506838
6,5	0,4525	0,464891	0,462842
7	0,4055	0,421428	0,421149
7,5	0,3624	0,380586	0,38192
8	0,3261	0,342496	0,345245
8,5	0,2931	0,307212	0,311157
9	0,2674	0,274723	0,27964
9,5	0,234	0,244971	0,250641
10	0,2079	0,21786	0,224078

8 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 100*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,1467	0,13996	0,171797
6	0,0936	0,098717	0,11161
6,5	0,0889	0,083365	0,089663
7	0,0827	0,070605	0,071898
7,5	0,0714	0,059946	0,057558
8	0,055	0,051006	0,046012
8,5	0,0384	0,043481	0,036733
9	0,0328	0,037128	0,029291
9,5	0,0241	0,03175	0,023331
10	0,0222	0,027187	0,018566

9 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 12*

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,159	0,151175	0,171797
6	0,1113	0,104075	0,11161
6,5	0,0923	0,086647	0,089663
7	0,0693	0,072269	0,071898
7,5	0,0585	0,060371	0,057558
8	0,0475	0,050501	0,046012
8,5	0,0352	0,042296	0,036733
9	0,0325	0,035462	0,029291
9,5	0,0262	0,029761	0,023331
10	0,0211	0,024998	0,018566

10 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 20*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,1715	0,163084	0,171797
6	0,11	0,108909	0,11161
6,5	0,089	0,089015	0,089663
7	0,0714	0,072762	0,071898
7,5	0,0581	0,059481	0,057558
8	0,0489	0,048628	0,046012
8,5	0,0392	0,039757	0,036733
9	0,0311	0,032507	0,029291
9,5	0,0237	0,02658	0,023331
10	0,0191	0,021735	0,018566

11 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 50*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,1722	0,167381	0,171797
6	0,1117	0,110352	0,11161
6,5	0,0922	0,089477	0,089663
7	0,0749	0,072496	0,071898
7,5	0,0596	0,058698	0,057558
8	0,0462	0,047498	0,046012
8,5	0,0373	0,038414	0,036733
9	0,0313	0,031053	0,029291
9,5	0,0239	0,025091	0,023331
10	0,02	0,020266	0,018566

12 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 100*

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,5643	0,540563	0,600666
6	0,4652	0,46033	0,506838
6,5	0,4644	0,425921	0,462842
7	0,4615	0,394621	0,421149
7,5	0,4474	0,366044	0,38192
8	0,3777	0,339876	0,345245
8,5	0,3276	0,315852	0,311157
9	0,3029	0,293751	0,27964
9,5	0,265	0,273382	0,250641
10	0,241	0,254581	0,224078

13 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 12*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,5812	0,553878	0,600666
6	0,5118	0,472495	0,506838
6,5	0,4759	0,43687	0,462842
7	0,4233	0,404161	0,421149
7,5	0,3774	0,374083	0,38192
8	0,3562	0,346393	0,345245
8,5	0,3077	0,320873	0,311157
9	0,2911	0,297333	0,27964
9,5	0,2697	0,275604	0,250641
10	0,2472	0,255532	0,224078

14 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 20*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,5912	0,573759	0,600666
6	0,5054	0,489721	0,506838
6,5	0,4499	0,451866	0,462842
7	0,4216	0,416644	0,421149
7,5	0,3853	0,383927	0,38192
8	0,3681	0,353583	0,345245
8,5	0,3256	0,325474	0,311157
9	0,3013	0,299464	0,27964
9,5	0,2723	0,275419	0,250641
10	0,2452	0,25321	0,224078

15 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 50*

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
5	0,5872	0,584435	0,600666
6	0,5039	0,498232	0,506838
6,5	0,4639	0,45883	0,462842
7	0,425	0,421919	0,421149
7,5	0,3912	0,387462	0,38192
8	0,359	0,355392	0,345245
8,5	0,3289	0,32562	0,311157
9	0,3022	0,298045	0,27964
9,5	0,2764	0,272556	0,250641
10	0,2525	0,249037	0,224078

16 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 100*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,3	0,320574	0,350485
12	0,1757	0,194933	0,213309
13	0,1329	0,150302	0,162606
15	0,0717	0,08784	0,090936
17	0,04	0,050421	0,048716
18	0,0283	0,037995	0,035174
19	0,0179	0,028543	0,025193
20	0,0133	0,021383	0,017912

17 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 20*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,3271	0,330606	0,350485
12	0,1998	0,201095	0,213309
13	0,1551	0,154498	0,162606
15	0,0848	0,089082	0,090936
17	0,0488	0,05009	0,048716
18	0,0337	0,037273	0,035174
19	0,0266	0,027614	0,025193
20	0,0154	0,020376	0,017912

18 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 30*

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,3363	0,338591	0,350485
12	0,1976	0,206001	0,213309
13	0,159	0,15779	0,162606
15	0,0879	0,08993	0,090936
17	0,045	0,049659	0,048716
18	0,0319	0,036542	0,035174
19	0,0258	0,026737	0,025193
20	0,0175	0,019463	0,017912

19 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 50*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,3437	0,344551	0,350485
12	0,2069	0,209663	0,213309
13	0,1588	0,160217	0,162606
15	0,0864	0,090475	0,090936
17	0,0461	0,049234	0,048716
18	0,0333	0,0359	0,035174
19	0,0233	0,026	0,025193
20	0,0171	0,018715	0,017912

20 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 100*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,61	0,641165	0,658935
12	0,4576	0,48021	0,512518
13	0,3872	0,408451	0,443255
15	0,2644	0,287392	0,320171
17	0,1768	0,196161	0,221874
18	0,1413	0,160542	0,182163
19	0,1115	0,130676	0,148306
20	0,0879	0,105842	0,119793

21 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 20*

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,6421	0,648372	0,658935
12	0,4917	0,492293	0,512518
13	0,4327	0,421072	0,443255
15	0,3045	0,298646	0,320171
17	0,2054	0,20457	0,221874
18	0,1581	0,167451	0,182163
19	0,1387	0,13618	0,148306
20	0,1055	0,110095	0,119793

22 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 30*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,6408	0,653518	0,658935
12	0,493	0,501201	0,512518
13	0,4345	0,430548	0,443255
15	0,3045	0,307391	0,320171
17	0,2109	0,211307	0,221874
18	0,1669	0,173067	0,182163
19	0,1421	0,140721	0,148306
20	0,1115	0,113658	0,119793

23 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 50*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,6471	0,657016	0,658935
12	0,4996	0,507422	0,512518
13	0,4291	0,437263	0,443255
15	0,307	0,313762	0,320171
17	0,2051	0,216344	0,221874
18	0,1684	0,17732	0,182163
19	0,1354	0,144205	0,148306
20	0,1111	0,116431	0,119793

24 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir vienodoms hipotetinėms tikimybėms, *n = 100*

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,3377	0,319972	0,350485
12	0,2023	0,194783	0,213309
13	0,1533	0,150317	0,162606
15	0,086	0,088046	0,090936
17	0,0461	0,05068	0,048716
18	0,0324	0,03825	0,035174
19	0,0234	0,028782	0,025193
20	0,0171	0,0216	0,017912

25 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, $n = 20$

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,346	0,330174	0,350485
12	0,2069	0,200985	0,213309
13	0,1541	0,154509	0,162606
15	0,0841	0,089235	0,090936
17	0,0467	0,050282	0,048716
18	0,0346	0,037461	0,035174
19	0,0244	0,027789	0,025193
20	0,0187	0,020534	0,017912

26 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, $n = 30$

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,3502	0,338316	0,350485
12	0,2135	0,205929	0,213309
13	0,1576	0,157798	0,162606
15	0,091	0,090031	0,090936
17	0,047	0,049785	0,048716
18	0,0326	0,036664	0,035174
19	0,024	0,02685	0,025193
20	0,0181	0,019563	0,017912

27 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, $n = 50$

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,3522	0,344408	0,350485
12	0,2103	0,209626	0,213309
13	0,1609	0,160222	0,162606
15	0,0872	0,090529	0,090936
17	0,046	0,049301	0,048716
18	0,0349	0,035965	0,035174
19	0,024	0,02606	0,025193
20	0,0173	0,018767	0,017912

28 lentelė. Kriterijų tikslumo palyginimas esant teisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 100*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,6498	0,634817	0,658935
12	0,4949	0,472541	0,512518
13	0,4143	0,400772	0,443255
15	0,2858	0,28051	0,320171
17	0,1844	0,190597	0,221874
18	0,1487	0,155676	0,182163
19	0,1195	0,126483	0,148306
20	0,0954	0,102274	0,119793

29 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 20*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,65	0,64393	0,658935
12	0,4946	0,486602	0,512518
13	0,4251	0,415201	0,443255
15	0,2965	0,293087	0,320171
17	0,2008	0,199854	0,221874
18	0,1622	0,163236	0,182163
19	0,1321	0,132475	0,148306
20	0,1049	0,106883	0,119793

30 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 30*

4 Priedas

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,6508	0,650583	0,658935
12	0,4968	0,497235	0,512518
13	0,4253	0,426342	0,443255
15	0,3047	0,303194	0,320171
17	0,2073	0,207578	0,221874
18	0,1658	0,169664	0,182163
19	0,1345	0,13767	0,148306
20	0,1089	0,110965	0,119793

31 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 50*

Realizacija	Pagal modeliuotas reikšmes	Modifikuotas	Klasikinis
10	0,6544	0,65525	0,658935
12	0,5019	0,504911	0,512518
13	0,4321	0,434529	0,443255
15	0,3058	0,310896	0,320171
17	0,2098	0,213684	0,221874
18	0,1676	0,174845	0,182163
19	0,1361	0,141946	0,148306
20	0,112	0,114402	0,119793

32 lentelė. Kriterijų tikslumo palyginimas esant neteisingai hipotezei ir nevienodoms hipotetinėms tikimybėms, *n = 100*

5 Priedas

Makroprograma ***modchi*** apskaičiuoja statistikos X_n^2 vidurkį, dispersiją ir trečiąjį centrinį momentą bei modifikuotos statitikos parametrus ***a**, **b*** ir ***w*** ir juos atspausdina. Po to randa tikimybes, kad statistikos reikšmė pateks į sritį už tam tikrų įvestų realizacijų kiekvieno kriterijaus atveju.

Makroprograma galima naudotis, įvedus parametrus:

k – hipotetinių tikimybių skaičius;

n – imties dydis;

c – parametras, kurį naudojame apskaičiuoti tikrosioms tikimybėms, kai hipotezė neteisinga (esant teisingai hipotezei *c* priskiriamas 0);

delta – necentriškumo parametras, kai hipotezė neteisinga (esant teisingai hipotezei *delta* priskiriamas 0);

r – realizacijų, su kuriomis lyginamos statistikų reikšmės, skaičius;

real – realizacijų, su kuriomis lyginamos statistikų reikšmės, masyvas;

array1 – tikrųjų tikimybių masyvas;

array2 – hipotetinių tikimybių masyvas;

```
%macro modchi (k,n,c,delta,r,real,array1,array2);
Data tikimybes;
        Array apsk(&k)a1-a&k (&array1);/*tikrųjų tikimybiu masyvas*/
        Array hipo(&k)b1-b&k (&array2);/*hipotetiniu tikimybiu masyvas*/
                Do i=1 to &k;
                        pi=apsk[i]+((-1)**i)*&c/sqrt(&n);
                        pi0=hipo[i];
                        s1=(pi-pi0)*(1-pi)/pi0;
                        s2=(pi-pi0)**2/pi0;
                        s3=pi**3/pi0**2;
                        s4=pi**2/pi0;
                        s5=pi/pi0;
                        s6=pi**2/pi0**2;
                        s7=pi/pi0**2;
                        s8=pi**3/pi0**3;
                        s9=pi**4/pi0**3;
```

```
                    ss0=pi**2/pi0**3;
                    ss1=pi/pi0**3;
                    suma1+s1;
                    suma2+s2;
                    suma3+s3;
                    suma4+s4;
                    suma5+s5;
                    suma6+s6;
                    suma7+s7;
                    suma8+s8;
                    suma9+s9;
                    suma10+ss0;
                    suma11+ss1;
                    delta=&c**2/pi0;
                    sum_delta+delta;
                    output;
            end;
Keep i pi pi0 s1 s2 s3 s4 s5 s6 s7 s8 s9 ss0 ss1 suma1 suma2 suma3 suma4 suma5 suma6 suma7
suma8 suma9 suma10 suma11 sum_delta;
run;
data momentai;/*randa statistikos vidurki, dispersija ir treciaji centrini momenta*/
set tikimybes;
        where i=&k;
        vidurkis=&k-1+suma1+suma2*&n;

        dispersija=(2*(&n-1)/&n)*(2*(&n-2)*suma3-(2*&n-3)*(suma4)**2-
        2*suma4*suma5+3*suma6)-(1/&n)*((suma5)**2-suma7);

        centrmom=(2*(&n-1)/(&n)**2)*((20*(&n)**2-68*&n+60)*(suma4)**3-
        3*suma4*suma5-18*suma6*suma5+(34*&n-68)*suma8+9*(suma5)**2*suma4-
        (18*&n-36)*suma5*suma3-(36*&n-54)*suma6*suma4+(16*(&n)**2-
        80*&n+96)*suma9+14*suma10+(24*&n-36)*(suma4)**2*suma5-(36*(&n)**2-
        144*&n+144)*suma4*suma3)+(1/(&n)**2)*(2*(suma5)**3+suma11-
        3*suma5*suma7);
```

```
keep vidurkis dispersija centrmom;
run;
data mod;
SET momentai;/*Randa modifikuotos statistikos V=a+bX parametrus a ir b bei laisves laipsniu skaiciu w*/
        a=8*dispersija**3/centrmom**2-4*vidurkis*dispersija/centrmom;
        b=4*dispersija/centrmom;
        w=8*dispersija**3/centrmom**2;
run;
proc print data=mod;/*spausdinamos vidurkio, dispersijos ir treciojo centrinio momento reiksmes, a,b ir w*/
        Var vidurkis dispersija centrmom a b w;
run;
data mod; /* Konstruojami klasikinis ir modifikuotas kriterijai*/
Set mod;
        Array realizacija(&r)re1-re&r (&real);
        retain Realizacijos 0 d 0;
                Do while (Realizacijos<realizacija[&r]);
                Modifikuotas=1-cdf('Chisq',realizacija[d+1]*b+a,w);
                Klasikinis=1-cdf('Chisq',realizacija[d+1],&k-1,&delta);
                d=d+1;
                Realizacijos=realizacija[d];
                output;
        end;
keep Realizacijos Modifikuotas Klasikinis;
run;
%mend modchi;
```

6 Priedas

Makroprograma ***modeliavimas*** modeliuoja statistikos X_n^2 reikšmes, skaičiuoja kriterijaus (4.1) p-reikšmę ir konstruoja kriterijų hipotezei tikrinti.

Makroprograma galima naudotis, įvedus parametrus:

k – hipotetinių tikimybių skaičius;

n – duomenų imties dydis;

z – statistikos X_n^2 generuotų reikšmių imties dydis;

c – parametras, kurį naudojame apskaičiuoti tikrosioms tikimybėms, kai hipotezė neteisinga (esant teisingai hipotezei *c* priskiriamas 0);

real – realizacijų, su kuriomis lyginamos statistikų reikšmės, masyvas;

array1 – hipotetinių tikimybių masyvas;

```
%macro modeliavimas (k,n,z,c,real,array1);
Data tik_mod;
        Array hipo(&k)hipo1-hipo&k (&array1);/*tikruju tikimybiu masyvas*/
        Do i=1 to &k;
                pi0=hipo[i]+(-1)**i*&c/sqrt(&n);
                output;
        end;
Keep pi0;
run;
proc transpose data=tik_mod out=tik_mod (drop=_Name_) prefix=t;
run;
data imtis; /*modeliuoja a.d. X imti*/
set tik_mod;
        Array polinom(&k)t1-t&k;
        Array hip(&k)h1-h&k (&array1);
        Do i=1 to &z;
                p=0;d=0;paskutinis=&n;Suma_x=0;
                        Do j=1 to &k-1;
                        Retain p d Suma_x paskutinis;
                        if paskutinis=0 then do;
                        pol=0;
```

```
                X=(pol-&n*hip[j])**2/(&n*hip[j]);
                suma_x+X;
                p=p;
                d=d;
                paskutinis=paskutinis;
                Stat=Suma_x+(&n*hip[j+1])**2/(&n*hip[j+1]);
                output;
                end;
                else do;
                pol=ranbin(1,paskutinis,polinom[j]/(1-p));
                X=(pol-&n*hip[j])**2/(&n*hip[j]);
                Suma_x+X;
                p=p+polinom[j];
                d=d+pol;
                paskutinis=&n-d;
                Stat=Suma_x+(paskutinis-
                &n*hip[j+1])**2/(&n*hip[j+1]);
                output;
                end;
                end;
    end;
run;
data kriterijus;
set imtis;
    where j=&k-1;
    Array rel(8)a1-a8 (&real);
        if Stat>a1 then p1=1;else p1=0;
        if Stat>a2 then p2=1;else p2=0;
        if Stat>a3 then p3=1;else p3=0;
        if Stat>a4 then p4=1;else p4=0;
        if Stat>a5 then p5=1;else p5=0;
        if Stat>a6 then p6=1;else p6=0;
        if Stat>a7 then p7=1;else p7=0;
        if Stat>a8 then p8=1;else p8=0;
run;
```

```
proc means data=kriterijus sum noprint;
Var p1 p2 p3 p4 p5 p6 p7 p8;
Output out=sumos (drop=_TYPE_ _FREQ_) sum(p1:)=suma1 sum(p2:)=suma2
sum(p3:)=suma3 sum(p4:)=suma4 sum(p5:)=suma5 sum(p6:)=suma6 sum(p7:)=suma7
sum(p8:)=suma8;
run;
data tikimybes;
set sumos;
        Array suma(8)suma1-suma8;
        Array rl(8)rl1-rl8 (&real);
        retain Realizacijos 0 d 0;
                Do while (Realizacijos<rl[8]);
                Pagal_modeliuotas=suma[d+1]/&z;
                d=d+1;
                Realizacijos=rl[d];
        output;
        end;
keep Realizacijos Pagal_modeliuotas;
run;
%mend krit_palyginimas;
```