ISSN 1648-8776
JAUNŲJŲ MOKSLININKŲ DARBAI. NR. 1 (41). 2014

SUBGAUSINIO VEKTORIAUS SKIRSTINIO PARAMETRŲ VERTINIMAS MONTEKARLO MARKOVO GRANDINĖS METODU

Leonidas Sakalauskas, Ingrida Vaičiulytė
Vilniaus universitetas, Matematikos ir informatikos institutas

Įvadas

Normalusis skirstinys – vienas iš plačiausiai matematinėje statistikoje naudojamų skirstinių, dažnai taikomas atliekant verslo ir ekonomikos duomenų analizę. Tačiau finansiniai duomenys neretai pasižymi asimetrija, ekscesu ir dideliais nuokrypiais, todėl normalusis skirstinys nėra geriausias variantas – pavyzdžiui, akcijų grąža ar rizikos veiksniai gali nebūti pasiskirstę pagal Gauso dėsnį (Kabasinskas ir kt., 2009). Todėl Gauso modeliai keičiami bendresniais – stabiliaisiais modeliais, kurie vis plačiau taikomi modeliuojant verslo, draudimo, finansų rinkas. Tačiau taikymą praktikoje riboja tai, jog stabiliųjų skirstinių pasiskirstymo ir tankio funkcijos negali būti išreiškiamos per elementariąsias funkcijas, išskyrus keletą pavienių atvejų, pvz., Koši ir Levi dėsnius. Taip pat stabilieji tankiai, kai $a=\frac{1}{3}$ ir $\beta=-1$, gali būti išreiškiami per modifikuotą Beselio funkciją $K_{\frac{1}{3}}(x)$, kai $a=\frac{2}{3}$, $a=\frac{3}{2}$ – per Vitakerio funkciją $W_{\frac{1}{2},\frac{1}{6}}(x)$, arba hipergeometrines funkcijas. Daugiau informacijos galima rasti D. Holt ir E. Crow (1973) darbe. Beje, stabilieji skirstiniai neturi baigtinės dispersijos (išskyrus normalųjį atvejį, kai stabilumo parametras $a=2$). Daugiamačių stabiliųjų dėsnių parametrų vertinimas kai kuriais statistiniais metodais ir su tuo susijusios problemos aptarti Press (1972), Rachev ir Xin (1993), Cheng ir Rachev (1995), Ravishanker ir Qiou (1999), Nolan (2001), Nolan ir kt. (2001) bei kitų autorių. Šiame straipsnyje nagrinėjamas subgausinio vektoriaus, išreiškiamo per Gauso vektorių ir a stabiliuosius dydžius, skirstinio parametrų vertinimas didžiausio tikėtinumo metodu, taikant Montekarlo Markovo grandinės algoritmą.

Gerai žinoma, kad vienmatis stabilusis dydis yra aprašomas keturiais parametrais: stabilumo $a\in(0;2]$, asimetrijos $\beta\in[-1;1]$, mastelio $\sigma>0$, poslinkio $\mu\in R$. Stabilumo indeksas a yra esminis modeliuojant finansines sekas, o mastelio (sklaidos) parametras σ stabiliuoju atveju gali būti taikomas rizikai matuoti. Atsitiktiniai dydžiai, kurie yra fiksuoto skaičiaus dedamųjų sudėties atžvilgiu, vadinami a stabiliaisiais.

Nagrinėsime stabilaus dydžio apibendrinimą daugiamačiu atveju. Vienmačiu atveju yra žinoma, kad

$$s=s_1\cdot s_2, \tag{1}$$

čia

$s_1^{a_2}$ – stabilusis atsitiktinis dydis, kurio asimetrijos parametras $\beta=1$, o formos parametras $a_1<1$;

s_2 – kitas stabilusis atsitiktinis dydis, kurio asimetrijos parametras $\beta=0$, o formos parametras a_2;

s – stabilusis atsitiktinis dydis, kurio asimetrijos parametras $\beta=0$, o formos parametras $a=a_1\cdot a_2$.

Taikant šį būdą paprastai parenkama, kad s_2 yra Gauso atsitiktinis dydis, nes šie dydžiai yra stabilūs, kai $a_1=\frac{a}{2}$, ir $a_2=2$. Atsitiktinis stabilusis dydis, kai $a_1<1$, o $\beta=1$, yra vadinamas subordinatoriumi ir visada įgyja tik teigiamas reikšmes (Rachev ir Mittnik, 1993). Pritaikius šį metodą daugiamačiu atveju gaunamas vektorius su priklausomomis komponentėmis, pagal kurį galima modeliuoti duomenis su „sunkiomis uodegomis".

Darbe naudojama Zolotariovo stabiliojo skirstinio tankio išraiška (Золотарев, 1983). Atsitiktinis vektorius X pasiskirstęs pagal stabilųjį dėsnį $S_a(\sigma,\beta,\mu)$, jei jo charakteringoji funkcija:

$$p(x,a,\sigma,\beta,\mu)=\begin{cases} ① \\ ② \end{cases}$$

$$① \exp\left\{-|\sigma\cdot\theta|^a\cdot\left(1-i\beta\,\mathrm{sgn}(\theta)\tan\left(\frac{\pi\alpha}{2}\right)\right)+i\mu\theta\right\},\ \text{jeigu } a\neq 1$$

$$② \exp\left\{-|\sigma\cdot\theta|\cdot\left(1+\frac{2i\beta}{\pi}\ln|\theta|\,\mathrm{sgn}(\theta)\right)+i\mu\theta\right\},\quad \text{jeigu } a=1 \tag{2}$$

čia $\theta\in R$ ir

$$\mathrm{sgn}(\theta)=\begin{cases}1, & \text{jei } \theta>0\\ 0, & \text{jei } \theta=0\\ -1, & \text{jei } \theta<0\end{cases}. \tag{3}$$

Taikant šį metodą galima gauti daugiamatį subgausinį vektorių:

$$X=\delta+\sqrt{s_1}s_2, \tag{4}$$

čia s_1 yra subordinatorius su parametru a, s_2 yra Gauso vektorius su nuliniu vidurkiu ir kovariacijų matrica Ω, o δ yra vidurkio vektorius.

Didžiausio tikėtinumo metodo įverčiai

Stabiliųjų vektorių (4) apibūdina šie parametrai: stabilumo indeksas a, vidurkio vektorius δ ir kovariacijų matrica Ω. Nagrinėsime šių parametrų vertinimą didžiausio tikėtinumo metodu.

Tegul $x=(x_1,x_2,\ldots,x_K)$ yra nepriklausomų p-mačių stabiliųjų vektorių imtis. Šios imties tikėtinumo funkciją galima užrašyti taip (Ravishanker, Qiou, 1999):

$$\widetilde{L}(x,\delta,\Omega,a)=\frac{\left(\frac{a}{2-a}\right)^K}{(2\pi)^{\frac{K}{2}}\cdot|\Omega|^{\frac{K}{2}}}\cdot\prod_{j=1}^{K}\int_0^{\infty}\int_1^{1}\exp\left[-\left|\frac{t_a(y_j)}{s_j}\right|^{\frac{a}{2-a}}-\right.$$

$$\left.-\frac{1}{2}(x_j-\delta)^T\frac{\Omega^{-1}}{s_j}(x_j-\delta)\right]\cdot\left|\frac{t_a(y_j)}{s_j}\right|^{\frac{a}{2-a}}\cdot\frac{1}{s_j^{\frac{p}{2}+1}}dy_j ds_j, \quad (5)$$

čia

$$t_a(y)=\frac{\sin\left(\frac{\pi\cdot a\cdot y}{2}\right)\cdot\sin\left(\frac{\pi\cdot(2-a)\cdot y}{2}\right)^{\frac{2-a}{a}}}{(\sin\pi\cdot y)^{\frac{2}{a}}\cdot\left(\cos\frac{\pi\cdot a}{4}\right)^{\frac{2}{a}}}. \quad (6)$$

Padarius keitinį

$$z_j=\left|\frac{t_a(y_j)}{s_j}\right|^{\frac{a}{2-a}}, \quad (7)$$

logaritminė tikėtinumo funkcija įgyja tokį pavidalą:

$$L(x,\delta,\Omega,a)=\frac{K}{2}\ln|\Omega|-$$

$$-\sum_{j=1}^{K}\ln\left(\int_0^{\infty}\int_0^{1}D(x_j,y_j,z_j,\delta,\Omega,a)\cdot\exp\{-z_j\}dy_j dz_j\right), \quad (8)$$

čia

$$D(x_j,y_j,z_j,\delta,\Omega,a)=$$

$$=\exp\left\{-\frac{z_j^{\frac{2-a}{a}}(x_j-\delta)^T\Omega^{-1}(x_j-\delta)}{2\cdot t_a(y_j)}\right\}\cdot\frac{z_j^{\frac{2-a}{a}\cdot\frac{p}{2}}}{t_a(y_j)^{\frac{p}{2}}}. \quad (9)$$

Diferencijuojant (8) išraišką pagal parametrus δ,Ω, prilyginant nuliui atitinkamas išvestines yra gaunamos lygtys, kurias turi tenkinti šių parametrų didžiausio tikėtinumo įverčiai:

$$\hat{\delta}=\frac{\sum_{j=1}^{K}\frac{x_j\cdot g_j}{f_j}}{\sum_{j=1}^{K}\frac{g_j}{f_j}}, \quad (10)$$

$$\hat{\Omega}=\frac{1}{K}\cdot\sum_{j=1}^{K}\frac{(x_j-\delta)(x_j-\delta)^T g_j}{f_j}, \quad (11)$$

čia

$$g_j=\int_0^{\infty}\int_0^{1}\frac{z_j^{\frac{2-a}{a}\cdot\frac{p}{2}}}{t_a(y_j)^{\frac{p}{2}}}\cdot D(x_j,y_j,z_j,\delta,\Omega,a)\cdot\exp\{-z_j\}dy_j dz_j, \quad (12)$$

$$f_j=\int_0^{\infty}\int_0^{1}D(x_j,y_j,z_j,\delta,\Omega,a)\cdot\exp\{-z_j\}dy_j dz_j. \quad (13)$$

Montekarlo Markovo grandinės algoritmas

Daugiamačio subgausinio vektoriaus parametrų įverčiams gauti gali būti pritaikomas EM algoritmas, Montekarlo metodu apskaičiuojant integralus (8), (12), (13), įeinančius į įverčių išraiškas.

Tarkime, sugeneruota k Montekarlo Markovo grandžių ir kiekvienoje grandyje apskaičiuoti δ_k, Ω_k, a_k įverčiai, paėmus kokias nors pradines reikšmes δ_0,Ω_0,a_0. Tarkime, kad $Y_i\sim U(0,1)$, $Z_i\sim-\ln(Y_i)$, čia $i=1,\ 2,\ldots,N^k$, N^k – yra Montekarlo imties tūris k grandyje.

Toliau apskaičiuojamos sumos:

$$P_{j,k}=\frac{1}{N^k}\sum_{i=1}^{N^k}D(X^j,Y_i,Z_i,\delta_k,\Omega_k,a_k), \quad (14)$$

$$PP_{j,k}=\frac{1}{N^k}\sum_{i=1}^{N^k}\left(D(X^j,Y_i,Z_i,\delta_k,\Omega_k,a_k)\right)^2, \quad (15)$$

$$V_{j,k}=\frac{1}{N^k}\sum_{i=1}^{N^k}\frac{Z_i^{\frac{2-a}{a}\cdot\frac{p}{2}}}{t_a(Y_i)^{\frac{p}{2}}}D(X^j,Y_i,Z_i,\delta_k,\Omega_k,a_k), \quad (16)$$

$$VV_{j,k}=\frac{1}{N^k}\sum_{i=1}^{N^k}\left(\frac{Z_i^{\frac{2-a}{a}\cdot\frac{p}{2}}}{t_a(Y_i)^{\frac{p}{2}}}D(X^j,Y_i,Z_i,\delta_k,\Omega_k,a_k)\right)^2, \quad (17)$$

kurios yra panaudojamos kitos iteracijos įverčiams apskaičiuoti:

$$\delta_{k+1}=\frac{\sum_{j=1}^{K}X^j\frac{V_{j,k}}{P_{j,k}}}{\sum_{j=1}^{K}\frac{V_{j,k}}{P_{j,k}}}, \quad (18)$$

$$\Omega_{k+1}=\frac{1}{K}\cdot\sum_{j=1}^{K}(X^j-\delta_k)(X^j-\delta_k)^T\frac{V_{j,k}}{P_{j,k}}. \quad (19)$$

Naudojant gautus įverčius, galima užrašyti logaritminės tikėtinumo funkcijos Montekarlo įvertį:

$$L_k=\frac{K}{2}\ln|\Omega_k|-\sum_{j=1}^{K}\ln\left(\frac{P_{j,k}}{N^k}\right). \quad (20)$$

Parametro a_{k+1} reikšmė yra gaunama sprendžiant vienmatį tikėtinumo funkcijos (20) minimizavimo pagal a uždavinį, įrašius (18) ir (19) reikšmes.

Montekarlo Markovo grandinės generuojamos pagal (18), (19) formules, kol pasikliautinumo intervalas

$$\left[L_k-\frac{2}{\sqrt{N^k}}\sqrt{N^k\cdot\sum_{j=1}^{K}\frac{PP_{j,k}}{P_{j,k}^2}-K},L_k+\right.$$

$$\left.+\frac{2}{\sqrt{N^k}}\sqrt{N^k\cdot\sum_{j=1}^{K}\frac{PP_{j,k}}{P_{j,k}^2}-K}\right] \quad (21)$$

tampa mažas ir neatmetama statistinė lygybių (10), (11) hipotezė. Pastarajai hipotezei tikrinti apskaičiuojamas kriterijus

$$H^k = \frac{K}{\frac{1}{K}\sum_{j=1}^{K}\frac{VV_{j,k}}{P_{j,k}^2}} \cdot \left(-\ln\frac{|\Omega_{k+1}|}{|\Omega_k|} + \left(\delta_{k+1}-\delta_k\right)^T \cdot \right.$$

$$\left. \cdot\, \Omega_{k+1}^{-1} \cdot \left(\delta_{k+1}-\delta_k\right) + \sum_{j=1}^{p}\left(\Omega_{k+1}\cdot\Omega_k^{-1}\right)_{j,j} - p \right), \qquad (22)$$

naudojant žinomą statistinį kriterijų, apibūdinantį imties vidurkio vektoriaus ir imties kovariacijų matricos sutapimą su žinomu vektoriumi ir žinoma matrica (Anderson, 1958). Gerai žinoma, kad imties tūriui didėjant kriterijaus (22) skirstinys artėja prie χ^2 skirstinio su $\upsilon = \frac{1}{2}p(p+3)$ laisvės laipsniu (Krishnaiah, 1984). Todėl statistinė hipotezė pagal kriterijų (22) yra atmetama, jei

$$H^k > T_{\alpha,\upsilon}, \qquad (23)$$

čia $T_{\alpha,\upsilon}$ – yra χ^2 skirstinio kvantilis, α – pasikliautinumo lygmuo. Jei ši hipotezė neatmetama ir pasikliautinumo intervalas (21) mažas, Montekarlo grandinės generavimą galima nutraukti ir priimti paskutinėje iteracijoje gautus įverčius.

Jei nors viena stabdymo sąlyga netenkinama, generuojama nauja Montekarlo imtis. Imties tūrį galima reguliuoti pagal taisyklę (Sakalauskas, 2000):

$$N^{k+1} \geq \frac{N^k}{H^k} \cdot T_{\gamma,\nu}, \qquad (24)$$

čia kai kuriais atvejais γ gali sutapti su α. Taikant šią taisyklę galima racionaliai parinkti imčių tūrį Montekarlo Markovo grandinėje ir kartu užtikrinti konvergavimą į tikėtinumo funkcijos maksimumą.

Kompiuterinis modeliavimas

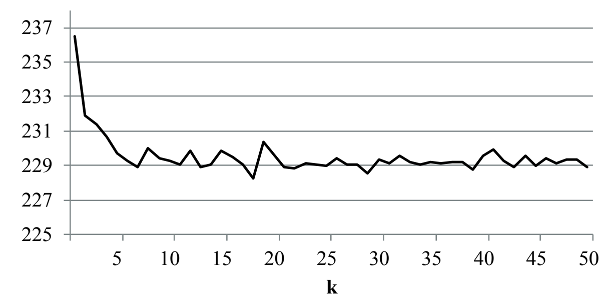
Tiriant sudarytą algoritmą buvo atliktas skaitinis eksperimentas. Sugeneruota $K = 50$ dvimačių vektorių su pradiniais duomenimis:

$$a = 1{,}25,\ \delta = \begin{pmatrix}1\\2\end{pmatrix},\ \Omega = \begin{pmatrix}3{,}67 & 0{,}86\\ 0{,}86 & 2{,}55\end{pmatrix},\ p = 2. \qquad (25)$$

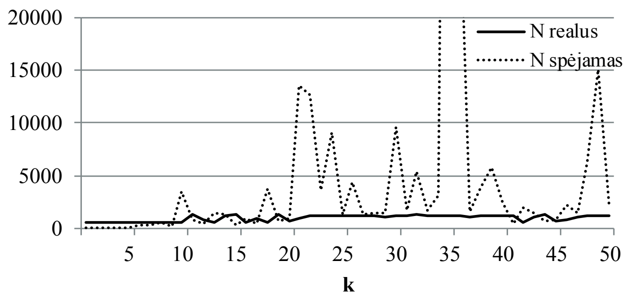
Toliau sugeneruota $k = 50$ Montekarlo grandžių pagal (18), (19), (24) išraiškas, paėmus pradines reikšmes $\delta_0 = 1{,}1\cdot\delta,\ \ \Omega_0 = 0{,}9\cdot\Omega$. Siekiant išvengti labai mažų arba labai didelių imties tūrio reikšmių, buvo taikoma riba: $N^k \geq 500$, ir generavimas buvo nutraukiamas, kai tikėtinumo funkcijos pasikliautinumo intervalas (21) tapdavo mažesnis už 0,5. Stabdymo taisyklės buvo pasiektos po $k = 11$ iteracijų. Gauti tokie įverčiai:

$$\delta_{11} = \begin{pmatrix}1{,}20\\2{,}55\end{pmatrix},\ \Omega_{11} = \begin{pmatrix}7{,}54 & 4{,}01\\ 4{,}01 & 5{,}66\end{pmatrix},\ L_{11} = 229{,}06,\ H^{11} = 22{,}25. \qquad (26)$$

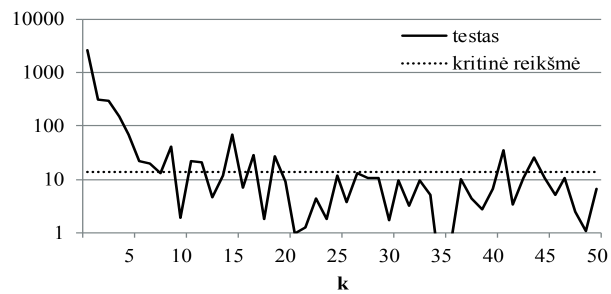
1 pav. pateikta tikėtinumo funkcijos diagrama, 2 pav. pateikta imties tūrio, reguliuojamo pagal (24) taisyklę, diagrama (žym. *N spėjamas*). Šiame paveiksle pavaizduotas *N realus* yra apskaičiuotas taip, kad pasikliautinumo intervalas (21) neviršytų pasirinktos kritinės reikšmės $\varepsilon = 0{,}5$. 3 pav. testas parodo statistikos (22), 4 pav. pasikliautinumo intervalo kitimo priklausomybes. 3 pav. kritinė reikšmė yra χ^2 skirstinio kvantilio reikšmė. Šios priklausomybės iliustruoja sudaryto algoritmo konvergavimą į uždavinio (8) sprendinį.



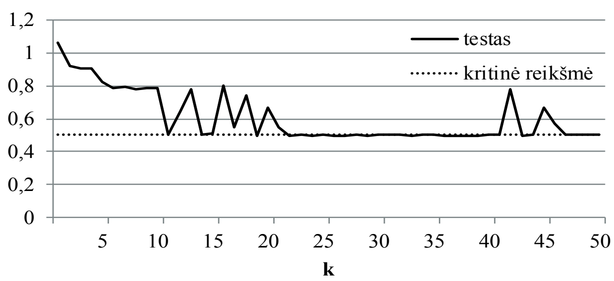
1 pav. Tikėtinumo funkcija



2 pav. Imties tūris



3 pav. Stabdymo statistika



4 pav. Pasikliautinumo intervalas

Išvados

Pagal sudarytą statistinį subgausinių sekų tyrimo algoritmą galima apskaičiuoti skirstinio tankio funkciją, generuoti atsitiktinių dydžių sekas, įvertinti šių sekų parametrus Montekarlo Markovo grandinės metodu, naudojant didžiausio tikėtinumo metodą.

Literatūra

1. Anderson T. W., 1958, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.
2. Cheng B. N., Rachev S. T., 1995, Multivariate Stable Future Prices. *Mathematical finance*. Vol. 5. P. 133–153.
3. Holt D., Crow E., 1973, Tables and Graphs of the Stable Probability Density Functions. *Journal of Research of the National Bureau of Standards*. Vol. 77B. P. 143–198.
4. Kabasinskas A., Rachev S., Sakalauskas L., Sun W., Belovas I., 2009, Stable Paradigm in Financial Markets. *Journal of Computational Analysis and Applications*. Vol. 11. No. 3. P. 642–688.
5. Krishnaiah P. R., 1984, *Handbook of Statistics 1: Analysis of Variance*. New York: Elsevier Science & Technology Books.
6. Nolan J. P., 2001, Maximum Likelihood Estimation and Diagnostics for Stable Distributions. *Lévy Processes: Theory and Applications*. P. 379–400.
7. Nolan J. P., Panorska A. K., McCulloch J. H., 2001, Estimation of Stable Spectral Measures. *Mathematical and Computer Modelling*. Vol. 34. P. 1113–1122.
8. Press S. J., 1972, Estimation in Univariate and Multivariate Stable Distributions. *Journal of the American Statistical Association*. Vol. 67. No. 340. P. 842–846.
9. Rachev S. T., Mittnik S., 1993, Modeling Asset Returns with Alternative Stable Distributions. *Econometric Reviews*. Vol. 12. No. 3. P. 261–330.
10. Rachev S. T., Xin H., 1993, Test for Association of Random Variables in the Domain of Attraction of Multivariate Stable Law. *Probability and Mathematical Statistics*. Vol. 14. No. 1. P. 125–141.
11. Ravishanker N., Qiou Z., 1999, *Monte Carlo* EM Estimation for Multivariate Stable Distributions. *Statistics & Probability Letters*. Vol. 45. No. 4. P. 335–340.
12. Sakalauskas L., 2000, Nonlinear Stochastic Optimization by Monte-Carlo Estimators. *Informatica*. Vol. 11. No. 4. P. 455–468.
13. Золотарев В. М., 1983, *Одномерные устойчивые распределения*. Москва: Наука.

ESTIMATION OF A SUB-GAUSSIAN VECTOR DISTRIBUTION BY THE MONTE-CARLO MARKOV CHAIN APPROACH

Leonidas Sakalauskas, Ingrida Vaičiulytė

Summary

The Monte-Carlo Markov chain procedure for estimation of a sub-Gaussian vector distribution by the maximal likelihood method, where the Monte-Carlo sample size is regulated to ensure the convergence, is constructed in the paper. The termination rule is also implemented by testing statistical hypotheses on an insignificant change of estimates in two steps of the procedure. The termination statistic is approximated with a χ^2 distribution. Computer simulation confirmed the applicability of the Monte-Carlo Markov chain approach.

Keywords: Monte-Carlo Markov chain, likelihood method, stable distribution, statistical modeling.

SUBGAUSINIO VEKTORIAUS SKIRSTINIO PARAMETRŲ VERTINIMAS MONTEKARLO MARKOVO GRANDINĖS METODU

Leonidas Sakalauskas, Ingrida Vaičiulytė

Santrauka

Šiame straipsnyje aprašomas subgausinio vektoriaus, išreiškiamo per Gauso vektorių ir a stabiliuosius dydžius, skirstinio parametrų vertinimas didžiausio tikėtinumo metodu, taikant Montekarlo Markovo grandinės algoritmą, naudojant statistinį stabdymo kriterijų ir imties tūrio reguliavimą. Stabdant algoritmą, tikrinama statistinė hipotezė apie nykstamai mažą skirtumą tarp dviejų gretimų Montekarlo grandžių įverčių. Hipotezei patikrinti naudojama statistinė χ^2 skirstinio aproksimacija. Atlikus skaičiavimus, gaunami nežinomų parametrų įverčiai.

Prasminiai žodžiai: Montekarlo Markovo grandinė, tikėtinumo metodas, stabilusis skirstinys, statistinis modeliavimas.

Įteikta 2014-01-12