XXXV Konferencja
Statystyka Matematyczna Wisła 2009

NIERÓWNOŚCI DLA KWANTYLI ROZKŁADU CHI-KWADRAT

Tadeusz Inglot

Instytut Matematyki i Informatyki
Politechniki Wrocławskiej

PROBLEM

$\Phi(z)$ – dystrybuanta rozkładu $N(0,1)$,

$z(\alpha)$ – kwantyl rzędu $1-\alpha$ tzn.

$$\Phi(z(\alpha)) = 1 - \alpha.$$

χ^2_k – zmienna losowa o rozkładzie chi-kwadrat z k stopniami swobody,

$u(\alpha, k)$ – kwantyl rzędu $1-\alpha$ dla χ^2_k tzn.

$$P(\chi^2_k \geq u(\alpha, k)) = \alpha.$$

Jaka jest asymptotyka $z(\alpha)$ **oraz** $u(\alpha, k)$, **gdy** $\alpha \to 0$ **i/lub** $k \to \infty$**?**

- ważne dla pewnych rozważań asymptotycznych, typowa sytuacja: $k = c_1 \log n$, $\alpha = e^{-n^{c_2}}$ czyli $\alpha = \exp\{-e^{c_2 k/c_1}\}$.
- inna asymptotyka dla ustalonego α, a inna dla ustalonego k.

cel: możliwie precyzyjne oszacowania, cytowania w MR i ZBL.

KWANTYLE ROZKŁADU NORMALNEGO

Twierdzenie 1.

$$z(\alpha) \geq \sqrt{2\log(1/\alpha)} - \frac{\log(4\log(1/\alpha)) + 2}{2\sqrt{\log(1/\alpha)}},$$

$$z(\alpha) \leq \sqrt{2\log(1/\alpha)} - \frac{\log(2\log(1/\alpha)) + 3/2}{2\sqrt{\log(1/\alpha)}}$$

dla każdego $\alpha \leq 0.1$.

Formuła przybliżona

$$z(\alpha) \approx \hat{z}(\alpha) = \sqrt{2\log(1/\alpha)} - \frac{A\log(2\log(1/\alpha)) + B}{2\sqrt{2\log(1/\alpha)}},$$

$A = 0.867, \quad B = 2.382.$

α	$z(\alpha)$	$\hat{z}(\alpha)$	błąd
0.1	1.2816	1.2825	+0.07%
0.05	1.6449	1.6441	-0.05%
0.01	2.3263	2.3253	-0.04%
0.005	2.5758	2.5750	-0.03%
0.001	3.0902	3.0902	0
0.0001	3.7190	3.7202	+0.03%
0.00001	4.2649	4.2670	+0.05%
0.000001	4.7534	4.7562	+0.06%

Dla $0.000001 \leq \alpha \leq 0.1$ błąd nie przekracza 0.07%.

OGON ROZKŁADU CHI-KWADRAT

Całkowanie przez części daje

$$\int_u^\infty x^{(k-2)/2} e^{-x/2} dx =$$

$$2u^{(k-2)/2} e^{-u/2} + (k-2) \int_u^\infty \frac{1}{x} x^{(k-2)/2} e^{-x/2} dx.$$

Stąd, dla $k \geq 2, \quad u > k-2$

$$P(\chi_k^2 \geq u) \geq \frac{2}{2^{k/2}\Gamma(k/2)} \exp\left\{-\frac{u}{2} + \frac{k-2}{2} \log u\right\}$$

oraz

$$P(\chi_k^2 \geq u) \leq$$
$$\frac{2}{2^{k/2}\Gamma(k/2)} \frac{u}{u-k+2} \exp\left\{-\frac{u}{2} + \frac{k-2}{2} \log u\right\}.$$

Po przekształceniach

$$\frac{1}{2}\,\mathcal{E}_k(u) \le P(\chi_k^2 \ge u) \le \frac{1}{\sqrt{\pi}}\frac{u}{u-k+2}\,\mathcal{E}_k(u),$$
$$\forall k \ge 2 \quad \forall u > k-2,$$

gdzie

$$\mathcal{E}_k(u) = \exp\left\{-\frac{u}{2} + \frac{k}{2} + \frac{k-2}{2}\log\frac{u}{k} - \frac{1}{2}\log k\right\}.$$

Po dokładniejszych rachunkach oszacowanie z dołu można poprawić

$$(*) \qquad P(\chi_k^2 \ge u) \ge \frac{1-e^{-2}}{2}\frac{u}{u-k+2\sqrt{k}}\,\mathcal{E}_k(u),$$
$$\forall k \ge 2 \quad \forall u \ge k.$$

OSZACOWANIA Z GÓRY

Nierówność globalna

Laurent & Massart (2000)

$$u(\alpha,k) \le k + 2\log\frac{1}{\alpha} + 2\sqrt{k\log\frac{1}{\alpha}}, \quad \forall k \ge 1 \quad \forall \alpha \in (0,1).$$

Nierówności w dolnym zakresie

Twierdzenie 2. $\forall c \in (0,2) \;\; \exists a = a(c) > 0$

$$u(\alpha,k) \le k + 2\log\frac{1}{\alpha} + c\sqrt{k\log\frac{1}{\alpha}}, \quad \forall k \ge 1 \quad \forall \alpha < e^{-ak}.$$

Twierdzenie 3.

$$u(\alpha,k) \le k + 2\log\frac{1}{\alpha} + 2\sqrt{k}\log^{5/4}\left(\log\frac{1}{\alpha}\right),$$
$$\forall k \ge 2 \quad \forall \alpha < \exp\{-e^{(k-2)^4/16k^2}\}.$$

- wykładnik 5/4 można dowolnie zmniejszyć, ograniczając odpowiednio zakres α.

OSZACOWANIA Z DOŁU

Nierówności globalne

Brain & Mi (2001) – oparta na CLT

$u(\alpha, k) > k + \sqrt{k}, \qquad \forall k \geq 1 \ \ \forall \alpha \leq 0.15.$

Inglot & Ledwina (2006)

$u(\alpha, k) \geq k + 2\log\frac{1}{\alpha} - 3, \qquad \forall k \geq 3 \ \ \forall \alpha \leq \frac{1}{k}.$

Stosując (*) mamy

Twierdzenie 4.

$u(\alpha, k) \geq k + 2\log\frac{1}{\alpha} - 5/2, \quad \forall k \geq 2 \ \ \forall \alpha \leq 0.17,$

$u(\alpha, k) \geq k + 2\log\frac{1}{\alpha}, \qquad \forall k \geq 9 \ \ \forall \alpha \leq \frac{1}{17}.$

OSZACOWANIA Z DOŁU

Nierówności w dolnym i środkowym zakresie

Inglot & Ledwina (2006)

$$u(\alpha,k) \geq k + 2\log\frac{1}{\alpha} + \frac{1}{3}\sqrt{k}\log\log\frac{1}{\alpha} - 3.1,$$
$$\forall k \geq 3 \quad \forall \alpha \leq e^{-k/3},$$

$$u(\alpha,k) \geq k + 2\log\frac{1}{\alpha} + \frac{1}{4}\sqrt{k\log\frac{1}{\alpha}},$$
$$\forall k \geq 32 \quad \forall \alpha \in [e^{-k/3}, 1/k].$$

Stosując (*) mamy

Twierdzenie 5.

$$u(\alpha,k) \geq k + 2\log\frac{1}{\alpha} + \sqrt{k}\log\log\frac{1}{\alpha},$$
$$\forall k \geq 32 \quad \forall \alpha \leq e^{-k/8},$$

$$u(\alpha,k) \geq k + 2\log\frac{1}{\alpha} + \frac{1}{4}\sqrt{k\log\frac{1}{\alpha}},$$
$$\forall k \geq 17 \quad \forall \alpha \in [e^{-560k}, 1/k].$$

APROKSYMACJE KWANTYLI

Niejawne

Fisher (1925)

Z CLT $\sqrt{2\chi_k^2} - \sqrt{2k-1} \to N(0,1),\ k \to \infty.$

Stąd $u(\alpha, k) \approx \frac{1}{2}(z(\alpha) + \sqrt{2k-1})^2.$

Wilson-Hilferty (1931)

$(\chi_k^2/k)^{1/3}$ ma rozkład asymptotyczny $N\left(1 - \frac{2}{9k}, \sqrt{\frac{2}{9k}}\right).$

Stąd $u(\alpha, k) \approx k\left(1 - \frac{2}{9k} + z(\alpha)\sqrt{\frac{2}{9k}}\right)^3.$

Różne ulepszenia W-H:

Haldane (1937), Goldberg & Levine (1946),

Severo & Zelen (1960), Zar (1974, 1978).

Hoaglin (1977)

$$u(\alpha, k) \approx \left(1.06807\sqrt{k} + 2.13161\sqrt{\log_{10}\frac{1}{\alpha}} \right.$$
$$\left. -0.04589\sqrt{k\log_{10}\frac{1}{\alpha}} - 1.37266\right)^2 .$$

Gilbert (1977).

NOWA APROKSYMACJA

Postać

$$u(\alpha,k) \approx k + 2\log\frac{1}{\alpha} + C_1\sqrt{k\log\frac{1}{\alpha}} + C_2\sqrt{k}\log\log\frac{1}{\alpha}$$

$$+C_3\sqrt{k} + +C_4\sqrt{\log\frac{1}{\alpha}} + C_5\log\log\frac{1}{\alpha} + C_6.$$

Dla porównania

Fisher

$$u(\alpha,k) \approx k + \log\frac{1}{\alpha} + 2\sqrt{k\log\frac{1}{\alpha}} + \text{ reszta.}$$

Wilson-Hilferty

$$u(\alpha,k) \approx k + \frac{4}{3}\log\frac{1}{\alpha} + 2\sqrt{k\log\frac{1}{\alpha}} + \text{ reszta.}$$

Propozycja wyboru stałych

$$u(\alpha,k) \approx k+2\log\frac{1}{\alpha}+1.62\sqrt{k\log\frac{1}{\alpha}}+0.63012\sqrt{k}\log\log\frac{1}{\alpha}$$
$$-1.12032\sqrt{k}-2.48\sqrt{\log\frac{1}{\alpha}}-0.65381\log\log\frac{1}{\alpha}-0.22872.$$

Dla $0.0005 \le \alpha \le 0.4$, $4 \le k \le 100$, błąd $< 0.3\%$.
Dla $0.0001 \le \alpha \le 0.4$, $3 \le k \le 100$, błąd $< 0.71\%$.

k	α			
	0.0005	0.05	0.2	0.3
4	+0.29%	+0.16%	+0.23%	+0.14%
6	+0.14%	−0.17%	−0.01%	+0.03%
9	+0.05%	−0.24%	−0.04%	+0.03%
25	−0.03%	−0.07%	+0.10%	+0.13%
100	−0.08%	+0.13%	+0.19%	+0.16%
400		+0.14%	+0.15%	+0.12%