APROXIMACE ROZDĚLENÍ LINEÁRNÍ KOMBINACE χ^2

Daniela JARUŠKOVÁ

Czech Technical University of Prague, Dept. of Mathematics

28. ledna 2010

$$Y = \sum_{j=1}^{n} c_j \, \chi_j^2(1), \quad c_1 \geq c_2 \geq \cdots \geq c_n > 0.$$

$$\mathbf{Y} = (Y_1, \ldots, Y_n)^T \sim N(0,\ R)$$

$$||\mathbf{Y}||^2 = \sum_{j=1}^{n} \lambda_j X_j^2,$$

kde $\{X_j\}$ i.i.d. $N(0,1)$ a $\{\lambda_j\}$ vlast. čísla matice R.

Johnson – Kotz: Continuous univariate distributions, vol.1, chap. 18 (8), Wiley & Sons (1994)

Charakteristická funkce

$$\varphi(t) = \prod_{j=1}^{n} \frac{1}{(1 - 2ic_j t)^{1/2}}$$

Imhof (1961) ... numericky invertovat
Můžeme spočítat přesně, pokud se jedná o lineární kombinaci exponenciálně rozdělených náhodných veličin.

$$P\Big(\sum_{j=1}^{p} \tilde{c}_j \chi_2^2(2\nu_j) > x\Big) = \sum_{j=1}^{p} \frac{1}{(\nu_j - 1)!} \Big(\frac{\partial^{\nu_j - 1}}{\partial c^{\nu_j - 1}} F_j(c, x) \Big)_{c = \tilde{c}_j}$$

$$F_j(c, x) = c^{q-1} \exp\Big\{ -\frac{x}{2c} \Big\} \frac{1}{\prod_{r \neq j}^{p} (c - \tilde{c}_r)^{\nu_r}}, \quad q = \sum_{j=1}^{p} \nu_j$$

Fatalov-Richter (1992) - Metoda sedlového bodu pro Laplaceovu transformaci:

$c_1 = \cdots = c_m > c_{m+1} \geq \cdots \geq c_n$

$$Q(x) = P\Big(\sum_{j=1}^{n} c_j\big(\chi_j^2(1)\big) > x\Big) \sim \frac{\exp\{-\frac{x}{2c_1}\}\big(\frac{x}{c_1}\big)^{m/2-1}}{2^{m/2-1}\Gamma(m/2)} \frac{1}{\prod_{j=m+1}^{n}\big(1-\frac{c_j}{c_1}\big)}$$

pro $x \to \infty$.

Poznámka.
Pro funkci přežití $G(x;m)$ a hustotu $g(x;m)$ rozdělení $\chi^2(m)$ platí

$$G(x;m) \sim 2\,g(x;m) \quad \text{pro} \quad x \to \infty.$$

Intuitivně bude fungovat, jestliže
1) $c_{m+1}, \ldots, c_n$ relativně malé vzhledem k c_1,
2) m blízko 2.

Field (1993) - Metoda sedlového bodu v inverzi charakteristické funkce

$$\varphi(it) = \exp\{K(t)\}; \quad K(t) = \sum_{j=1}^{n} -\frac{1}{2}\log(1 - 2c_j t)$$

Nechť x je fixní a $\widehat{T}$ je řešení rovnice:

$$K'(t) = x, \quad \text{tj.} \sum_{j=1}^{n} \frac{c_j}{1 - 2c_j t} = x.$$

Nechť $\widehat{W} = \sqrt{-2\big(K(\widehat{T}) - \widehat{T}x\big)}$,
pak

$$Q(x) \approx 1 - \Phi(\widehat{W}) + \phi(\widehat{W})\Big(\frac{1}{\widehat{T}\sqrt{K''(\widehat{T})}} - \frac{1}{\widehat{W}}\Big).$$

Daniels (1987) - Metoda sedlového bodu pro inverzi charakt.fce pro průměr
$X_1, \ldots, X_n$ i.i.d. s $k(t) = \log(\varphi(it))$.
Nechť $\widehat{t}$ je řešení $k'(t) = \bar{x}$, pak

$$Q_n(\bar{x}) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{n(k(t)-t\bar{x})} \frac{1}{t}\, dt = \frac{1}{2\pi i} \int_{\widehat{t}-i\infty}^{\widehat{t}+i\infty} e^{n(k(t)-t\bar{x})} \frac{1}{t}\, dt$$

Aproximace

$$e^{n(k(t)-t\bar{x})} = e^{n(k(\hat{t})-\hat{t}x)-n(k^{''}(\hat{t})/2)(t-\hat{t})^2}(1 + nC_3(t-\hat{t})^3 + \ldots)$$

vede na

$$Q_n(\bar{x}) \approx \frac{e^{n\left(k(\hat{t})-\hat{t}\,\bar{x}\right)}}{\sqrt{2\pi n k^{''}(\hat{t})}} \frac{1}{\hat{t}}.$$

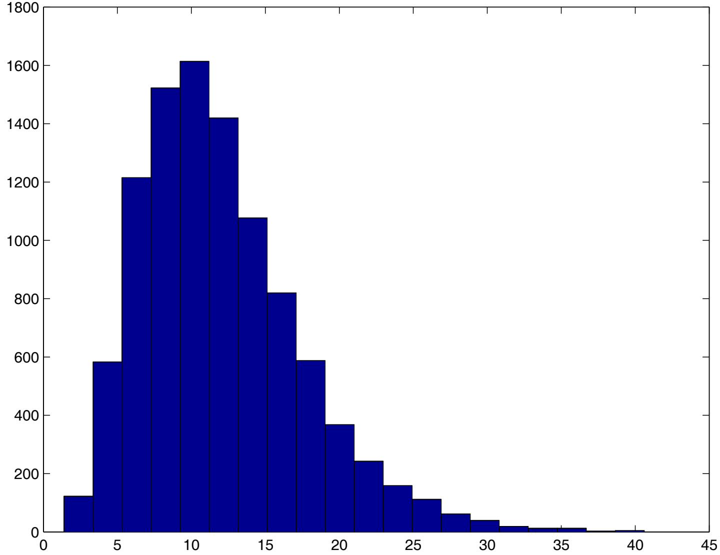
Pro fixní $\bar{x}$ definujme $\widehat{w} = \sqrt{-2\big(k(\hat{t}) - \hat{t}\,\bar{x}\big)}$. Použijme substituci $(1/2)w^2 - \widehat{w}w = k(t) - tk'(\hat{t})$. Potom

$$Q_n(\bar{x}) = \frac{1}{2\pi i}\int_{\widehat{w}-i\infty}^{\widehat{w}+i\infty} e^{n((1/2)w^2 - \widehat{w}w)}\frac{1}{w}dw +$$

$$+\frac{1}{2\pi i}e^{-(1/2)\widehat{w}^2 n}\int_{\widehat{w}-i\infty}^{\widehat{w}+i\infty} e^{(1/2)n(w-\widehat{w})^2}\left(\frac{1}{t}\frac{dt}{dw} - \frac{1}{w}\right)dw$$

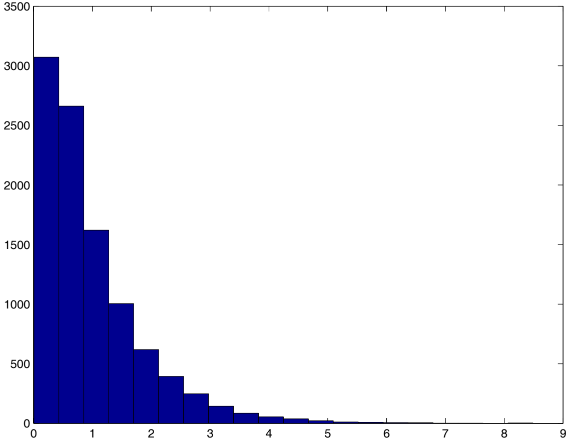
$$\approx 1 - \Phi(n\widehat{w}) + \phi(n\widehat{w})\left(\frac{1}{\hat{t}\sqrt{k''(\hat{t})}} - \frac{1}{\widehat{w}}\right)\frac{1}{\sqrt{n}}.$$

Fieldova aproximace bude fungovat dobře, pokud budou včechny konstanty $\{c_j\}$ přibližně stejně velké. Překvapivě funguje dobře i v dost obecných případech.
Moje konstanty (1.9104, 1.7706, 1.5648, 1.3266, 1.0904, 0.8841, 0.7245, 0.6161, 0.5534, 0.5249, 0.5177, 0.5165).


1800
1600
1400
1200
1000
800
600
400
200
0
0
5
10
15
20
25
30
35
40
45

pst	hodnota	Field	Fat.-Richter
0.60	12.445	0.60254	
0.70	14.027	0.70241	
0.80	16.059	0.80150	0.60321
0.90	19.265	0.90103	0.84349
0.95	22.265	0.95063	0.93360
0.96	23.207	0.96058	0.94917
0.97	24.412	0.97055	0.96385
0.98	26.065	0.98038	0.97730
0.99	28.829	0.99017	0.98953
0.995	31.621	0.99517	0.99519
0.999	37.788	0.99903	0.99912
0.9995	40.527	0.99953	0.99959

Konstanty (0.45,0.45,0.05,0.05)



pst	hodnota	Field	Fat.-Richter
0.95	5.6043	0.950257	0.94998

Reference:

Imhof J.P. : computing the distribution of quadratic forms in normal variables, Biometrika (1961), 48, 419 − 426

Fatalov V. and Richter W.D. : Gaussian probabilities of large deviation for fixed or increasing dimensions, Izvestiya Akademii Nauk Armenii. Matematika (1992), 27, 1 − 16.

Field Ch. : Tail areas of linear combinations of chi-squares and non-central chi-squares, J. Statis. Comput. Simul. (1993), 45, 143 − 248.