Hubera lokācijas parametra novērtējums ar empīriskās ticamības metodi

Māra Vēliņa

LU 69. konference

17.03.2011

Saturs

M-novērtējums

Pieņemsim, ka $X_1, X_2, \ldots, X_n \sim$ iid ar sadalījuma funkciju F. M-novērtējums ir statistiskais funkcionālis T_n, kas noteiktai funkcijai ρ minimizē

$$\sum_{i=1}^{n} \rho(X_i, t) = \int \rho(x, t) dF_n(x),$$

kur F_n - sadalījuma funkcija.
Ja ρ ir diferencējama pēc t, tad $\sum_{i=1}^{n} \rho(X_i, t)$ minimizē vienādojuma $\sum_{i=1}^{n} \psi(X_i, t) = 0$ sakne, kur $\psi(x, t) = \frac{\partial}{\partial t} \rho(x, t)$.

Piemēri

- Ja $\psi(x, t) = x - t$, tad $T_n = \bar{X}$.
- Ja $\psi(x, \theta) = -\frac{d}{d\theta} \log f(x, \theta)$ kādai noteiktai blīvuma funkciju klasei $f(x, \theta)$, tad T_n ir sakne ticamības vienādojumam

$$\frac{d}{d\theta} \log \left(\prod_{i=1}^{n} f(X_i, \theta) \right) = 0.$$

- Mediāna. $\psi(x, t) = \psi_0(x - t)$, kur $\psi_0(z) = k \operatorname{sgn}(z)$, $k > 0$.

Hubera lokācijas parametra M-novērtējums

Hubers (1964) ieviesa $\psi(x,t) = \psi_0(x-t)$, kur

$$\psi_0(z) = \begin{cases} c, & z \geq c \\ z, & |z| < c \\ -c, & z \leq -c. \end{cases} \quad (1)$$

Motivācija:

- Neierobežotām ψ funkcijām piemīt nevēlamas īpašības (piemēram, izlēcēju ietekme;
- Ievēro, ψ robežgadījumiem atbilstošos M-novērtējumus:
  - Ja $c \to \infty$, tad iegūst vidējo vērtību;
  - ja $c \to 0$, tad iegūst mediānu.

Skalētais lokācijas parametra novērtējums

Nepieciešams robusts dispersijas novērtējums! Aplūko

$$\psi(x,t) = \psi_0((x-t)/S_n),$$

kur

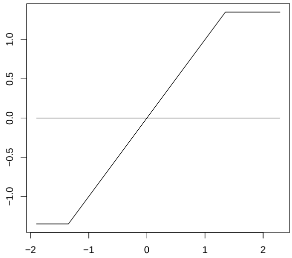
$$S_n = MAD = \mathrm{median}(|X_i - \mathrm{median}(X_i)|).$$

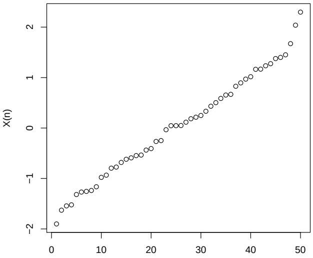
Atbilstošais M-novērtējums

$$\sum_{i=1}^{n} \psi\left(\frac{X_i - t}{S_n}\right) = 0.$$

Vislielākās ticamības novērtējums, kur blīvuma funkcijas "vidum" ir normālais, bet astēm - dubultais eksponenciālais sadalījums. Robusts novērtējums, pat ja dati satur augstu procentu izlēcēju (līdz 50 %).

Piemērs. Ģenerēti dati no $\mathrm{N}(0,1)$ sadalījuma, $n = 50$.
Hubera novērtējums $t = 0.030$, $\bar{X} = 0.194$, mediāna$= 0.048$.





EL metode

- $X_1, X_2, \ldots, X_n \sim \text{iid}$, ar nezināmu sadalījuma funkciju F un parametru θ.
- Funkcija $g(X, \theta)$, $E\{g(X, \theta)\} = 0$
- EL vidējai vērtībai: $g(X_i, \theta) = X_i - \mu$;
- Funkciju $L(F) = \prod_{i=1}^{n} \omega_i$ maksimizē pie ierobežojumiem

$$\omega_i \geq 0, \sum \omega_i = 1, \sum \omega_i g(X_i, \theta) = 0.$$

- Empīriskā ticamības funkcija parametram θ:

$$L(\theta) = \prod_{i=1}^{n} \omega_i = \prod_{i=1}^{n} n^{-1}\{1 + \lambda(\theta) g(X_i, \theta)\}^{-1}$$

- Empīriskā ticamības attiecība parametram θ:

$$L(\theta_0) = \frac{L(\theta)}{L(\theta_0)} = \prod_{i=1}^{n}\{1 + \lambda(\theta) g(X_i, \theta)\}^{-1}.$$

- Logaritmiskā empīriskās ticamības attiecība (Neparametriskā Vilksa teorēma)

$$W(\theta_0) = -2 \log R(\theta_0) \to_d \chi^2_1,$$

kad $n \to \infty$ un spēkā $H_0 : \theta = \theta_0$.

- Owen (1988) ieguva EL ticamības intervālus plašai M-novērtējumu klasei, tai skaitā Hubera novērtējumam.
- Pie noteiktiem regularitātes nosacījumiem, M-novērtējumiem ir spēkā Vilksa teorēma.
- Tsao, Zhu (2001) parādīja, ka EL ticamības intervāli Hubera novērtējumam ir asimptotiski robusti.

EL ticamības intervāli Hubera novērtējumam

Empīriskā ticamības attiecību parametram t

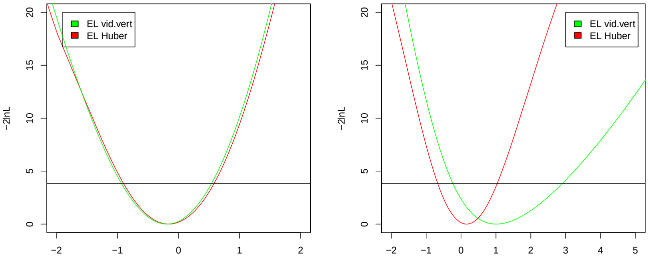
$$R(t) = \sup\{\prod_{i=1}^{n} \omega_i \sum_{i=1}^{n} \omega_i \psi(X_i, t) = 0, \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1\}$$

maksimizē $\prod \omega_i(\lambda)$, kur

$$\omega_i(\lambda) = \{n(1 + \lambda Z_i)\}^{-1},$$

kur $Z_i = \psi(X_i, t)$ un λ iegūst no

$$n^{-1} \sum Z_i/(1 + \lambda Z_i) = 0.$$



att.: EL -2*ln ticamības līknes
(a) $\mathrm{N}(0,3)$ (b) $0.95 * \mathrm{N}(0,3) + 0.05 * \mathrm{N}(20,3)$

tabula: Hubera un vidējās vērtības novērtējumi un ticamības intervāli, alpha=0.05

	$N(0,3)$				0.95*N(0,3)+0.05*N(20,3)			
Izlase	Intervāls		Novērtējums		Intervāls		Novērtējums	
n=50	EL(huber)	0.494	EL(huber)	-0.055	EL(huber)	1.706	EL(huber)	0.159
	EL(vid.vert)	0.492	EL(vid.vert)	-0.064	EL(vid.vert)	3.140	EL(vid.vert)	1.008
	t-test	0.506	vid.vert	-0.064	t-test	3.117	vid.vert	1.008
	z-test	0.554	huber	-0.076	z-test	0.554	huber	0.159
	Butstraps	0.497			Butstraps	3.057		
n=20	EL(huber)	0.667	EL(huber)	-0.167	EL(huber)	2.478	EL(huber)	-0.441
	EL(vid.vert)	0.667	EL(vid.vert)	-0.167	EL(vid.vert)	4.894	EL(vid.vert)	0.498
	t-test	0.732	vid.vert	-0.167	t-test	4.938	vid.vert	0.498
	z-test	0.877	huber	-0.643	z-test	0.877	huber	-0.441
	Butstraps	0.699			Butstraps	4.583		
n=10	EL(huber)	1.001	EL(huber)	-0.067	EL(huber)	4.303	EL(huber)	-0.189
	EL(vid.vert)	1.001	EL(vid.vert)	-0.067	EL(vid.vert)	9.680	EL(vid.vert)	1.008
	t-test	1.239	vid.vert	-0.067	t-test	11.494	vid.vert	1.799
	z-test	1.240	huber	-0.201	z-test	1.240	huber	-0.189
	Butstraps	1.039			Butstraps	9.740		