Bikela–Rozenblata tests

Audris Ločmelis

Latvijas Universitāte

03.11.2010.

Bikela–Rozenblata testa aktualitāte

- Bikels un Rozenblats šo testu ieviesa 1973. gada publikācijā.
- Tests paredzēts gan vienkāršām, gan saliktām hipotēzēm.
- Lee un Na 2000.g. un 2001.g. publikācijās parāda, ka tests bez transformācijām strādā arī atkarīgiem datiem, kā izņēmumu var minēt eksplozīvas (explosive) laikrindas.

Apskatām vienkāršu hipotēzi

$$H_0 : f = f_0$$
$$H_1 : f \neq f_0$$

pie nozīmības līmeņa α un zināmas blīvuma funkcijas f_0.

Bikela–Rozenblata testa statistika

$$T_n = n h_n^{1/2} \int \left[f_n(x) - (K_{h_n} * f_0)(x)\right]^2 a(x)dx,$$

Blīvuma funkcijas kodolu novērtējums

No h_n atkarīgs blīvuma funkcijas novērtējums

$$\begin{aligned} f_n(x) &= \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \\ &= \frac{1}{h_n} \int K\left(\frac{x - t}{h_n}\right) dF_n(t). \end{aligned}$$

Konvolūcija

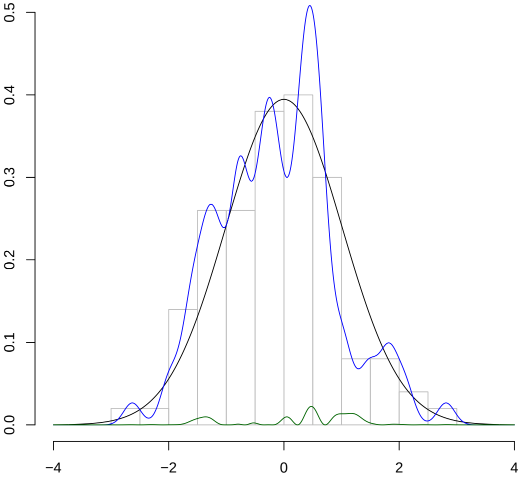
Nogludināts f_0

$$(K_{h_n} * f_0)(x) = \mathbb{E}_0 f_n(x)$$

Konvolūcijas operators $*$

$$(K_h * g)(\cdot) = \int h^{-1} K\left(\frac{\cdot - z}{h}\right) g(z)dz.$$

Testa statistika grafiski



att.: Konvolūcija, f_n un kvadrātiskā kļūda

Teorēma

Teorēma

Izpildoties H_0

$$(T_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

μ un σ^2

$$\mu = h_n^{-1/2} \int K^2(u)du \times \int f_0(x)a(x)dx$$

$$\sigma^2 = 2 \int f^2(x)dx \times \int \left[\int K(u)K(u+v)du \right]^2 dv.$$

Divas piezīmes

Netipiski, ka

šajā testā nav vienalga, kādu kodolu izvēlēties. Gosh, Huang (1991) parādīja, ka vislabāk tests strādā ar vienmērīgo kodolu un svaru funkciju $a(x) = 1$.

Tipiski, ka

joslas platuma h izvēlei ir ļoti liela nozīme. Bikela – Rozenblata testam nestrādā krosvalidācija. Apskatītajos gadījumos tā izvēl pārāk lielu h, kā rezultātā tiek pārgludināts blīvuma funkcijas novērtējums un konvolūcija. Statistikas vērtības kļūst tuvas nullei.

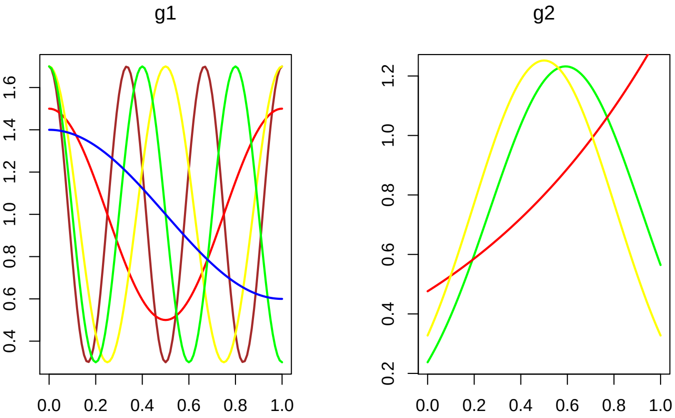
Kursa darbā izmantotās alternatīvas

No Kalenberg un Ledwina 1995. gada publikācijas

$$\begin{aligned} g_1(x) &= 1 + \rho\cos(j\pi x), \\ g_2(x) &= \exp\left(\sum_{j=1}^{k} \theta_j \pi_j(x) - \psi_k(\theta)\right), \end{aligned}$$

kur $\{\pi_j\}$ ir ortonormāli Ležandra polinomi intervālā $[0, 1]$, $\theta \in \mathbb{R}^k$, $\psi_k(\theta) = \log \int_0^1 \exp(\theta \circ \phi(x))dx$ un $\circ$ apzīmē skalāro reizinājumu.

Tuvas v.s.$[0, 1]$



g1
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0.0
0.2
0.4
0.6
0.8
1.0
g2
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.2
0.4
0.6
0.8
1.0

Joslas platuma izvēle neparametriskiem kodolu testiem

JASA tiks publicēts Gao un Gijbels raksts par šo tēmu

Tajā tiek apskatīts neparametrisks regresijas modelis

$$Y_i = m(X_i) + e_i, \; i = 1, 2, \dots, n,$$

kur $\{X_i\}$ ir stacionāru laikrindas mainīgo virkne, bet $\{e_i\}$ ir neatkarīgu, vienādi sadalītu kļūdu virkne ar $\mathbb{E}\left[e_i\right] = 0$ un $0 < \left[e_i^2\right] = \sigma^2 < \infty$. Turklāt $\forall\, 0 \le i \le j \le n$ X_i un e_i neatkarīgi.

Hipotēze

Rakstā tiek testēta parametriskā nulles hipotēze

$$\begin{aligned} H_0 :& m(x) = m_{\theta_0}(x) \\ H_1 :& m(x) = m_{\theta_1}(x) + \Delta_n(x), \end{aligned}$$

kur $\{\Delta_n(\cdot)\}$ ir lokālu alternatīvu virkne, $\lim_{n\to\infty} \Delta_n(x) = 0$

Divas interesējošās funkcijas

Var definēt nozīmīguma funkciju

$$\alpha_{\mathrm{n}}(\mathrm{h}) = \mathrm{P}(\hat{\mathrm{T}}_{\mathrm{n}}(\mathrm{h}) > \mathrm{l}_{\alpha} | \mathrm{H}_0)$$

un jaudas funkciju

$$\beta_{\mathrm{n}}(\mathrm{h}) = \mathrm{P}(\hat{\mathrm{T}}_{\mathrm{n}}(\mathrm{h}) > \mathrm{l}_{\alpha} | \mathrm{H}_1)$$

Šeit $\hat{\mathrm{T}}_{\mathrm{n}}(\mathrm{h})$ apzīmē neparametriska kodolu testa statistiku, bet l_{α} ir šīs statistikas sadalījuma $1 - \alpha$ kvantiles bootstrap aproksimācija.

Funkciju Edgeworth izvirzījumi

$$\alpha_n(x) = 1 - \Phi(l_\alpha - s_n) - k_n(1 - (l_\alpha - s_n)^2)\phi l_\alpha - s_n + o(\sqrt{h^d})$$
$$\beta_n(x) = 1 - \Phi(l_\alpha - r_n) - k_n(1 - (l_\alpha - r_n)^2)\phi l_\alpha - r_n + o(\sqrt{h^d})$$

Šeit Φ un ϕ ir attiecīgi sadalījuma un blīvuma funkcijas, s_n, k_n, r_n atkarīgi no h un $\sigma_n^2 = \int \Delta_n^2(x)\pi^2(x)dx$.

Mērķis

ir izvēlēties joslas platumu h_{ew} tādu, ka
$\beta_n(h_{ew}) = \max_{h \in H_n(\alpha)} \beta_n(h)$,
$H_n(\alpha) = \{h : \alpha - c_{min} < \alpha_n(h) < \alpha + c_{min}\}$ mazam
$0 < c_{min} < \alpha$. Jeb maksimizēt jaudu pie aptuveni nemainīga

Paldies par uzmanību!