

Empīrisko procesu pielietojums strukturālo attiecību modeļos

Juris Cielēns

Latvijas Universitāte

11.06.2010

Darba mērķi

- Konstruēt empīriskos procesus dažādiem strukturālo attiecību modeļiem;
- Pierādīt šo empīrisko procesu asimptotisko sadalījumu;
- Ar simulāciju palīdzību iegūt kritiskās vērtības pie dažādiem nozīmības līmeņiem un veikt pārklājumu precizitātes analīzi.

Glivenko-Kantelli teorēma

$$\sup_{-\infty < x < \infty} (|F_n(x) - F(x)|) \xrightarrow[n \rightarrow \infty]{a.s.} 0$$

Kolmogorova-Smirnova statistika

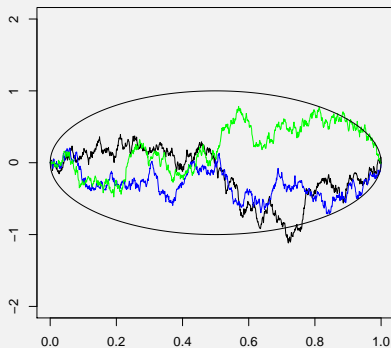
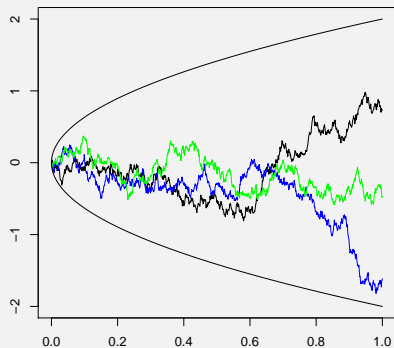
$$\sup_{-\infty < x < \infty} |\sqrt{n}(F_n(x) - F(x))| \xrightarrow{a.s.} \sup_{-\infty < x < \infty} |B(F(x))|$$

Brauna tilts

Par Brauna tiltu sauc Gausa procesu $\{B(t) : t \in [0, 1]\}$, kura kovariāciju struktūra ir $\text{cov}(B(s), B(t)) = s(1 - t)$, kur $s < t$. Šī procesa sadalījums ir tāds pats kā ar $W(t) - tW(1)$ sadalījums, kur W - standarta Brauna kustība.

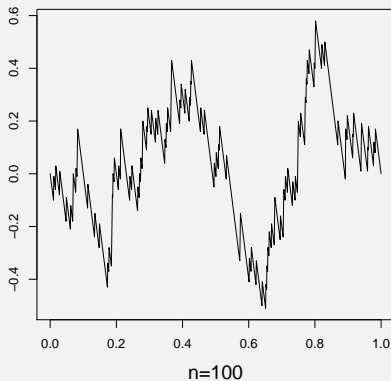
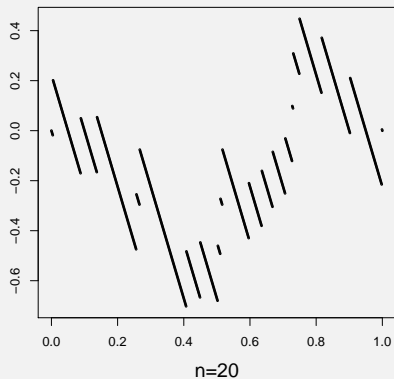
Empīriskais process

Brauna kustības un Brauna tilta piemēri



Empīriskais process

Empīriskā procesa konverģence uz Brauna tiltu



Tiek aplūkots process $\sqrt{n}(F_n(x) - F(x))$

Empīriskais process

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(X_i), \quad Pf = \int f dP.$$

Vispārināts empīriskais process

$$\sqrt{n}(\mathbb{P}_n f - Pf)$$

Strukturālie modeļi

Lokācijas-skalēšanas modelis

$$F_1(x) = F_2\left(\frac{x - \mu}{\sigma}\right) := F_2(x, h), x \in \mathbb{R}$$

$$F_1^{-1}(t) = \sigma F_2^{-1}(t) + \mu, t \in [0, 1]$$

Lēmaņa alternatīvu modelis

$$F_1(x) = 1 - (1 - F_2(x))^{1/h} := F_2(x, h), x \in \mathbb{R}$$

$$F_1^{-1}(t) = F_2^{-1}(1 - (1 - t)^h), t \in [0, 1]$$

Strukturālie modeļi

Vispārīgā forma (Freitag, Munk 2005)

Pieņemsim, ka $\mathcal{H} \subseteq \mathbb{R}^1$ un $\phi_1 : \mathbb{R} \times \mathcal{H} \rightarrow \mathbb{R}$, $\phi_2 : [0, 1] \times \mathcal{H} \rightarrow [0, 1]$. Funkcijas F_1 un F_2 ir saistītas ar strukturālu attiecību, kuru veido ϕ_1 un ϕ_2 , ja $(F_1, F_2) \in \mathcal{U}_{\phi_1, \phi_2} =: \mathcal{U}$, kur modeļu klase \mathcal{U} tiek definēta kā

$$\mathcal{U} := \{(F_1, F_2) \in \mathcal{F}^2 \times \mathcal{F}^2 \mid \text{h} \in \mathcal{H} \text{ tāds, ka} \\ F_1^{-1}(t) = \phi_1(F_2^{-1}(\phi_2(t, \text{h}))), t \in [0, 1]\}$$

Strukturālie modeļi

Mallova attālums lokācijas-skalēšanas modelim

$$M(F_1, F_2) := \int_0^1 (F_1^{-1}(t) - \sigma F_2^{-1}(t) - \mu)^2 dt$$

Mallova attālums Lēmaņa alternatīvu modelim

$$M(F_1, F_2) = \int_0^1 (F_1^{-1} - F_2^{-1}(1 - (1 - t)^h))^2 dt.$$

Lokācijas-skalēšanas modelis

Lokācijas-skalēšanas empīriskais process

$$PP_{LS}(t) := \sqrt{n}(F_{1n}(\hat{\sigma}F_{2m}^{-1}(t) + \hat{\mu}) - F_1(\sigma F_2^{-1}(t) + \mu)).$$

Empīriskā procesa asimptotiskais sadalījums

$$\forall \epsilon > 0 \quad \lim_{n,m \rightarrow \infty} P\left(\sup_{0 < t < 1} \left| PP_{LS}(t) - (B^{(n)}(t) + \sqrt{\frac{n}{m}}B^{(m)}(t) + \sqrt{nf_2}(F_2^{-1}(t))((\hat{\mu} - \mu) + F_2^{-1}(t)(\hat{\sigma} - \sigma))) \right| > \epsilon\right) = 0 \text{ g.d.}$$

Lokācijas-skalēšanas modelis

Kritiskās vērtības (izlases apjoms $n=500$)

		Parametri tiek novērtēti			Parametri zināmi		
1. sad.	2. sad.	90%	95%	99%	90%	95%	99%
$N(0,1)$	$N(2,16)$	1.159	1.258	1.456	1.697	1.880	2.234
$N(1,4)$	$N(-1,9)$	1.159	1.258	1.456	1.697	1.880	2.234
$U[0,1]$	$U[-2,4]$	1.173	1.286	1.499	1.711	1.895	2.262
$U[3,5]$	$U[4,9]$	1.173	1.286	1.499	1.711	1.895	2.262
$\text{Exp}(0.5)$	$\text{Exp}(0.25)$	2.333	2.743	3.521	1.682	1.866	2.276
$\text{Exp}(1)$	$\text{Exp}(2)$	2.333	2.743	3.521	1.682	1.866	2.276

Lokācijas-skalēšanas modelis

Pārklājumu precizitāte

n	Fiksēti parametri			N(0,1) un N(15,9)		
	90%	95%	99%	90%	95%	99%
20	0.924	0.970	0.999	0.970	0.970	0.996
50	0.938	0.968	0.993	0.953	0.953	0.998
100	0.931	0.961	0.991	0.915	0.960	0.994
200	0.906	0.948	0.983	0.908	0.942	0.985
500	0.904	0.955	0.997	0.877	0.955	0.984
1000	0.900	0.950	0.989	0.896	0.938	0.989
5000	0.884	0.939	0.989	0.894	0.943	0.987
10000	0.903	0.950	0.992	0.884	0.938	0.991

Lēmaņa alternatīvu modelis

Empīriskais process

$$PP_h(t) := \sqrt{n}(F_{1n}(F_{2m}^{-1}(1 - (1 - t)^{\hat{h}})) - F_1(F_2^{-1}(1 - (1 - t)^h)))$$

Empīriskā procesa asimptotiskais sadalījums

$$\forall \epsilon > 0 \quad \lim_{n,m \rightarrow \infty} P\left(\sup_{0 < t < 1} |PP_h(t) - B^{(n)}(F_1(F_2^{-1}(1 - (1 - t)^h)))\right. \\ \left. + \sqrt{\frac{n}{m}}B^{(m)}(1 - (1 - t)^h) - \sqrt{n}((1 - t)^h - (1 - t)^{\hat{h}}))| > \epsilon\right) = 0 \text{ g.d.,}$$

Lēmaņa alternatīvu modelis

Kritiskās vērtības (izlases apjoms $n=500$)

Sadalījums	h	90%	95%	99%
$N(0,1)$	0.5	1.9453	2.1689	2.6162
$N(0,1)$	1.5	1.6546	1.8335	2.2137
$N(1,4)$	0.5	1.9453	2.1689	2.6162
$N(1,4)$	1.5	1.6546	1.8335	2.2137
$\text{Exp}(0.5)$	0.5	1.9453	2.1466	2.5938
$\text{Exp}(1)$	1.5	1.6546	1.8559	2.2137
$U[-2,3]$	0.5	1.9230	2.1466	2.5714
χ_4^2	1.5	1.6546	1.8335	2.2137

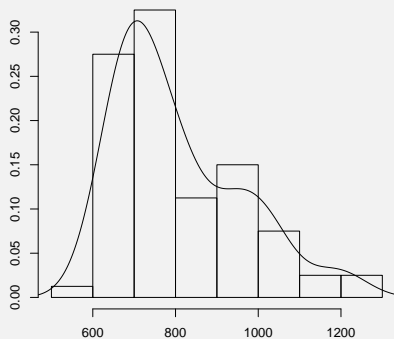
Lēmaņa alternatīvu modelis

Pārklājumu precizitāte

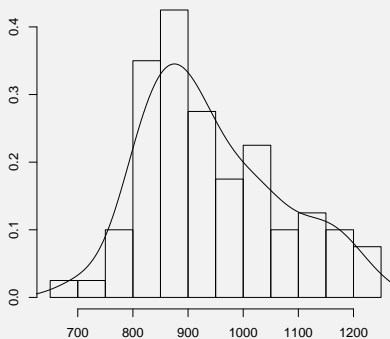
n	$N(2,9), h=0.5$			$\text{Exp}(2), h=1.5$		
	90%	95%	99%	90%	95%	99%
20	0.923	0.966	0.997	0.905	0.969	0.990
50	0.916	0.966	0.996	0.908	0.954	0.994
100	0.914	0.953	0.989	0.908	0.950	0.986
200	0.907	0.961	0.990	0.890	0.937	0.994
500	0.891	0.945	0.994	0.890	0.951	0.987
1000	0.893	0.947	0.990	0.917	0.95	0.993
5000	0.901	0.950	0.990	0.894	0.942	0.988
10000	0.910	0.956	0.991	0.889	0.952	0.993

Pielietojums reāliem datiem

Izlašu histogrammas



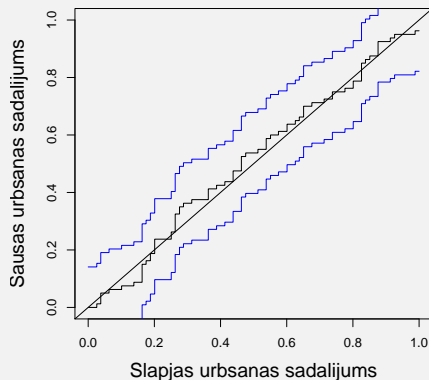
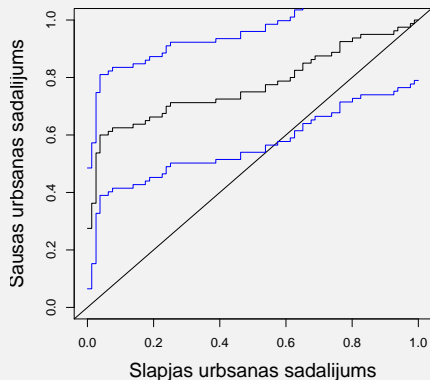
Sausas urbsanas histogramma



Slapjas urbsanas histogramma

Pielietojums reāliem datiem

Hipotēžu pārbaude



Secinājumi

- Tika iegūti asimptotiskie sadalījumi Lokācijas-skalēšanas un Lēmaņa alternatīvu modeļiem;
- Pārklājumu precizitātes pārbaude uzrāda labus rezultātus;
- Metode veiksmīgi pielietota reāliem datiem;

Atklāti jautājumi

- Mallova attāluma asimptotiskais sadalījums;
- Empīrisku procesu butstraps;
- Empīrisku procesu asimptotika citās funkciju klasēs.