

Copula goodness-of-fit testing

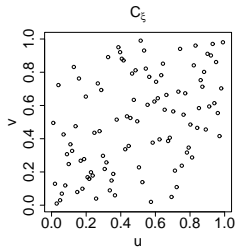
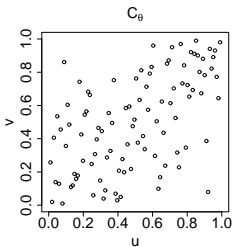
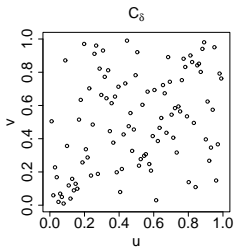
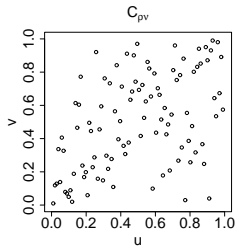
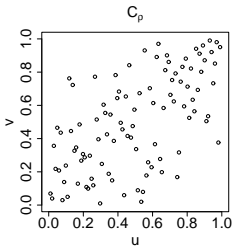
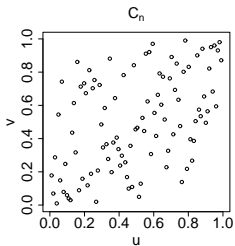
Daniel Berg

Universitetet i Oslo & Norsk Regnesentral

DET 14. NORSKE STATISTIKERMØTET
Sommarøya 19. -21. Juni 2007

Outline

- ▷ 1. Introduksjon
- ▷ 2. Copula GoF testing
 - 2.1 Lovende tester
 - 2.2 Cpit2-testen
- ▷ 3. MC simuleringer
- ▷ 4. Idéer til videre arbeid
- ▷ 5. Oppsummering



1. Introduksjon

- ▷ Passer vår modell til dataene?
- ▷ Univariat \Rightarrow Anderson-Darling eller QQ-plot
- ▷ Multivariat \Rightarrow færre alternativer
- ▷ Grensefordelingen til en copula GoF test avhenger av parameteren
- ▷ p-verdi estimeres via tunge parametriske bootstrap prosedyrer
- ▷ Fokus i litteraturen er nesten utelukkende bivariat

2. Copula GoF testing

- ▷ Flere teknikker har blitt foreslått:
 - Diskretisering av sannsynlighets rommet (binning)
 - Multivariat glatting
 - **Dimensjonsreduksjon**
 - Konstruksjoner for å teste for en spesifikk copula
 - **Generelle tester for å teste enhver copula**

2.1 Lovende tester

- ▶ $\sqrt{n}\{C_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u})\}$, $\mathbf{u} = u_1, \dots, u_d$,
- ▶ $\sqrt{n}\{K_n(\mathbf{u}) - K_{\theta_n}(\mathbf{u})\}$, $K(t) = P(C(\mathbf{u}) \leq t)$,
- ▶ $\sqrt{n}\{L_n(\mathbf{u}) - L_{\theta_n}(\mathbf{u})\}$, $L(t) = P(\prod_{i=1}^d u_i \leq t)$,
- ▶ Rosenblatt's transformasjon \Rightarrow test for uavhengighet.

2.1.1 Eksempel på K_n -plott

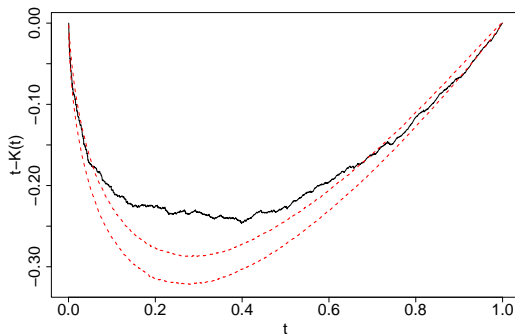


Figure: Eksempel på K -plot ($t - K(t)$, 95% ci)

2.2 Cpit2-testen (Berg and Bakken, 2007)

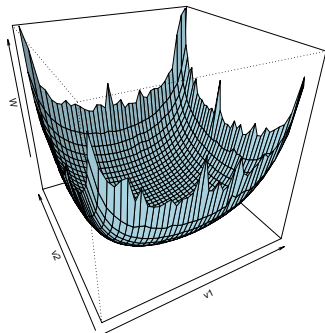
- ▷ n observasjoner av en d -dimensjonal vektor \mathbf{X}
- ▷ Pseudo-observasjoner $\mathbf{Z} \sim U(0, 1)^d$
- ▷ Rosenblatt's transformasjon (cpit) basert på $\mathbf{Z} \Rightarrow \mathbf{V}$
- ▷ Teste \mathbf{V} for uavhengighet

2.2 Cpit2-testen

- ▶ Breymann et al. (2003): $W_G = \sum_{i=1}^d \Phi^{-1}(V_i)^2$
- ▶ Two problems:
 1. Vekt funksjonen $\Phi^{-1}(\cdot)$ for ekstrem - problem for liten n
 2. Oppdager ikke radiell asymmetri ($\Phi^{-1}(0.2)^2 = \Phi^{-1}(0.8)^2$)
- ▶ Rosenblatt's transformasjon på $\mathbf{V} \Rightarrow \mathbf{H}$
- ▶ $W_B = \sum_{i=1}^d \Gamma_V(V_{(i)}; \alpha) \cdot \Gamma_H(H_i; \alpha)$
- ▶ Eksempler på $\Gamma(X)$: $\Phi^{-1}(X)^2$, $|X - 0.5|$,
 $(X - 0.5)^\alpha$, $\alpha = (2, 4, \dots)$
- ▶ Spesialtilfelle: $\Gamma_V(X) = \Phi^{-1}(X)^2$ og $\Gamma_H(X) = 1$

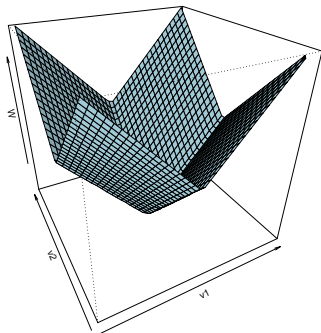
2.2 Cpit2-testen

$$W_B = \sum_{i=1}^d \Phi^{-1}(V_{(i)}):$$



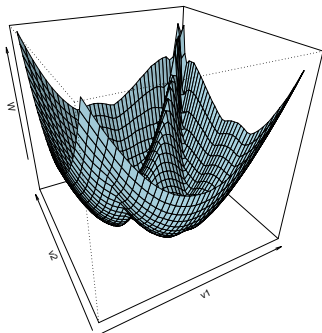
2.2 Cpit2-testen

$$W_B = \sum_{i=1}^d |V_{(i)} - 0.5|:$$



2.2 Cpit2-testen

$$W_B = \sum_{i=1}^d |V_{(i)} - 0.5| \cdot |H_i - 0.5|:$$



3. MC simuleringer (1/3)

\mathcal{H}_0	\mathcal{H}_1	$B_{2,0}$	C_n	K_n	L_n
Gauss	Gauss	0.048	0.051	0.046	0.049
	Student t4	0.022	0.075	0.055	0.046
	Clayton	0.011	0.728	0.638	0.405
	Gumbel	0.016	0.221	0.204	0.078
Student t4	Gauss	0.049	0.049	0.063	0.054
	Student t4	0.051	0.049	0.054	0.055
	Clayton	0.070	0.705	0.682	0.532
	Gumbel	0.081	0.184	0.210	0.093
Clayton	Gauss	0.075	0.854	0.691	0.605
	Student t4	0.205	0.820	0.566	0.534
	Clayton	0.051	0.048	0.051	0.046
	Gumbel	0.238	0.990	0.922	0.911
Gumbel	Gauss	0.047	0.287	0.372	0.175
	Student t4	0.058	0.248	0.305	0.115
	Clayton	0.057	0.988	0.998	0.850
	Gumbel	0.048	0.057	0.050	0.054

Table: Styrke (%) til copula GoF tester, $d = 2, n = 100, \tau = 0.5$.

3. MC simuleringer (2/3)

\mathcal{H}_0	\mathcal{H}_1	$B_{2,0}$	C_n	K_n	L_n
Gauss	Gauss	0.058	0.053	0.051	0.053
	Student t4	0.642	0.205	0.147	0.116
	Clayton	0.046	0.943	0.924	0.570
	Gumbel	0.024	0.660	0.903	0.962
Student t4	Gauss	0.049	0.052	0.052	0.046
	Student t4	0.054	0.053	0.053	0.055
	Clayton	0.624	0.964	0.875	0.690
	Gumbel	0.486	0.372	0.882	0.974
Clayton	Gauss	0.188	0.999	0.448	0.975
	Student t4	0.932	0.987	0.157	0.879
	Clayton	0.049	0.049	0.052	0.051
	Gumbel	0.444	0.996	0.915	1.000
Gumbel	Gauss	0.055	0.999	0.983	0.737
	Student t4	0.764	0.983	0.982	0.644
	Clayton	0.175	1.000	1.000	1.000
	Gumbel	0.027	0.052	0.039	0.049

Table: Styrke (%) til copula GoF tester, $d = 5$, $n = 200$, $\tau = 0.25$.

3. MC simuleringer (3/3)

\mathcal{H}_0	\mathcal{H}_1	$B_{2,0}$	C_n	K_n	L_n
Gauss	Gauss	0.052	0.057	0.053	0.050
	Student t4	1.000	0.668	0.383	0.411
	Clayton	0.159	0.998	0.995	0.957
	Gumbel	0.098	0.979	1.000	1.000
Student t4	Gauss	0.044	0.053	0.051	0.051
	Student t4	0.059	0.052	0.053	0.047
	Clayton	1.000	1.000	0.964	0.988
	Gumbel	1.000	0.932	1.000	1.000
Clayton	Gauss	0.236	1.000	0.858	0.998
	Student t4	1.000	1.000	0.277	0.950
	Clayton	0.042	0.052	0.051	0.051
	Gumbel	0.642	1.000	1.000	1.000
Gumbel	Gauss	0.155	1.000	1.000	1.000
	Student t4	1.000	1.000	1.000	0.998
	Clayton	0.568	1.000	1.000	1.000
	Gumbel	0.044	0.052	0.045	0.052

Table: Styrke (%) til copula GoF tester, $d = 5$, $n = 1000$, $\tau = 0.15$.

4. Idéer til videre arbeid

- ▶ GoF test basert på to estimatorer av θ , f.eks.

$$\sqrt{n}\{\rho^{-1}(\rho_n) - \tau^{-1}(\tau_n)\}$$

- ▶ Hvilken effekt har parameter estimator på styrken til en test:

$$\sqrt{n}\{C_n - C_{\rho^{-1}(\rho_n)}\} \text{ vs. } \sqrt{n}\{C_n - C_{\tau^{-1}(\tau_n)}\}$$

- ▶ Lokal styrke ved små pertubasjoner fra \mathcal{H}_0 :

$$Q_{\delta_n} = (1 - \delta_n)C + \delta_n D, \quad \delta_n = \delta/\sqrt{n}, \quad \delta \ll n$$

- ▶ Hvilken effekt har bootstrap parametre på styrken til en test.
Kan vi finne optimale valg (tommelfinger regler)?
- ▶ FGoF?

n	$\sqrt{n}\{C_n - C_\theta\}$	τ	$\mathcal{H}_1(\mathcal{H}_0 = \text{Clayton})$			
			Clayton	Frank	Gumbel-Hougaard	Gaussian
250	$\sqrt{n}\{C_n - C_0\}$	0.1	0.0530	0.1517	0.9971	0.0636
		0.2	0.0533	0.6316	1.0000	0.2536
		0.3	0.0514	0.9697	1.0000	0.7100
	$\sqrt{n}\{C_n - C_\rho\}$	0.1	0.0487	0.2295	0.1212	0.1094
		0.2	0.0493	0.6874	0.3730	0.3175
		0.3	0.0483	0.9425	0.6746	0.6076
	$\sqrt{n}\{C_n - C_\tau\}$	0.1	0.0484	0.2339	0.1169	0.1081
		0.2	0.0513	0.6846	0.3652	0.3125
		0.3	0.0531	0.9487	0.6677	0.5870
	$\sqrt{n}\{C_n - C_{PL}\}$	0.1	0.0438	0.0860	0.0619	0.0466
		0.2	0.0509	0.5161	0.3366	0.2010
		0.3	0.0536	0.8717	0.7240	0.4988
2500	$\sqrt{n}\{C_n - C_0\}$	0.1	0.0526	0.6669	1.0000	0.2606
		0.2	0.0499	0.9992	1.0000	0.8772
		0.3	0.0480	1.0000	1.0000	0.9997
	$\sqrt{n}\{C_n - C_\rho\}$	0.1	0.0515	0.5541	0.2282	0.1868
		0.2	0.0512	0.9956	0.8133	0.7390
		0.3	0.0506	1.0000	0.9941	0.9858
	$\sqrt{n}\{C_n - C_\tau\}$	0.1	0.0512	0.5495	0.2238	0.1962
		0.2	0.0492	0.9967	0.8147	0.7369
		0.3	0.0505	1.0000	0.9934	0.9853
	$\sqrt{n}\{C_n - C_{PL}\}$	0.1	0.0477	0.6234	0.4573	0.3298
		0.2	0.0500	0.9967	0.9461	0.8339
		0.3	0.0501	1.0000	0.9996	0.9927

5. Oppsummering

- ▷ Flere gode tester, dimensjonsreduksjon mest lovende, spesielt for høyere dimensjoner
- ▷ Cpit2-testen, basert på Rosenblatt's transformasjon
- ▷ Styrke sammenligning: $\sqrt{n}\{C_n - C_{\theta_n}\}$ svært god
- ▷ Mye spennende å jobbe videre med
- ▷ TITAN - 800 cpus - ikke mulig foruten!

Referanser

- Berg, D. and H. Bakken (2007). A copula goodness-of-fit test based on the probability integral transform. Submitted.
- Breymann, W., A. Dias, and P. Embrechts (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance* 1, 1–14.

Contact details

Daniel Berg,
Universitetet i Oslo & Norsk Regnesentral,
daniel@danielberg.no
www.danielberg.no