Copula Entropy

Theory and Applications

MA Jian
majian.2003@tsinghua.org.cn

Contents

Lecture 6: Evaluation

Independence Measures : Implementations

Table: Independence Measures and their implementations.

Package	Measure	University
`copent`	CE	Tsinghua
`stats`	Ktau	Oxford / Cambridge
`energy`	dCor	NSF
`dHSIC`	dHSIC	ETHZ / MPI / Copenhagen
`HHG`	HHG.chisq, HHG.lr	Tel Aviv
`independence`	Hoeff, BDtau	Berlin / LSE
`Ball`	Ball	SYSU / Yale / Texas
`qad`	QAD	Salzburg
`BET`	BET	North Carolina
`MixedIndTests`	Mixed	McGill / HEC Montréal
`subcopem2D`	subcopula	UNAM
`EDMeasure`	MDM	UIUC / NWU
`FOCI`	CODEC	Stanford
`NNS`	NNS	Villanova

Independence Measures: Experiments I

- **Bivariate normal distribution** $\mathbf{X} \sim N(\mathbf{u}, \rho)$, whose covariance ρ changes from 0 to 0.9 by step 0.1;
- **Bivariate normal copula** $C_N(u, v; \rho)$ with margins as normal distribution $u \sim N(0, 1)$ and exponential distribution $v \sim E(\lambda = 2)$, the parameter ρ of normal copula changes from 0 to 0.9 by step 0.1;

Independence Measures: Experiments II

- **Archimedean copulas**, including Clayton copula, Frank copula, and Gumbel copula as follows:

$$C_{\alpha}^{Clayton}(u, v) = max\left([u^{\alpha} + v^{\alpha} - 1]^{-\frac{1}{\alpha}}, 0\right), \tag{1}$$

$$C_{\alpha}^{Frank}(u, v) = -\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1}\right), \tag{2}$$

$$C_{\alpha}^{Gumbel}(u, v) = \exp\left\{-[(-\ln u)^{\alpha} + (-\ln v)^{\alpha}]^{\frac{1}{\alpha}}\right\}, \tag{3}$$

margins are same as above, the parameter α of the copulas change from 1 to 10;

Independence Measures: Experiments III

- **Trivariate normal distribution $\mathbf{X} \sim N(\mathbf{u}, \Sigma)$**, where covariance matrix Σ is

$$\begin{vmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{vmatrix}, \tag{4}$$

  ρ changes from 0 to 0.9 by step 0.1;

- **Trivariate Gumbel copulas** as follows:

$$C_{\alpha}^{Gumbel}(u, v, w) = \exp\left\{-[(-\ln u)^{\alpha} + (-\ln v)^{\alpha} + (-\ln w)^{\alpha}]^{\frac{1}{\alpha}}\right\}, \tag{5}$$

  the parameter α change from 1 to 10, margins are a normal distribution $u \sim N(0, 2)$ and two exponential distributions $v \sim E(\lambda = 2), w \sim E(\lambda = 0.5)$;

Independence Measures: Experiments IV

- **Quadvariate normal distribution** $\mathbf{X} \sim N(\mathbf{u}, \Sigma)$ for simulating dependence relations between two bivariate random vectors, the covariance matrix Σ is

$$\begin{vmatrix} 1 & \rho_{12} & \rho & \rho \\ \rho_{12} & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho_{34} \\ \rho & \rho & \rho_{34} & 1 \end{vmatrix}, \tag{6}$$

$\rho_{12} = 0.8, \rho_{34} = 0.75$, and ρ changes from 0 to 0.8 by step 0.1.

Independence Measures : Results I

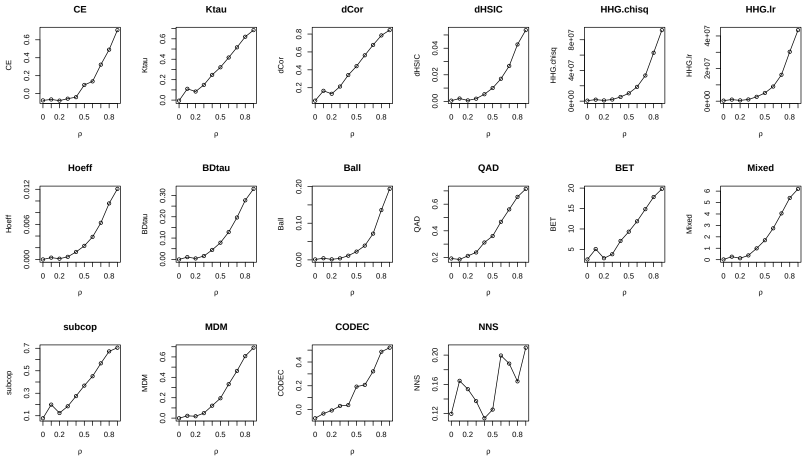


Figure: Experiments with bivariate normal distributions.

Independence Measures : Results II

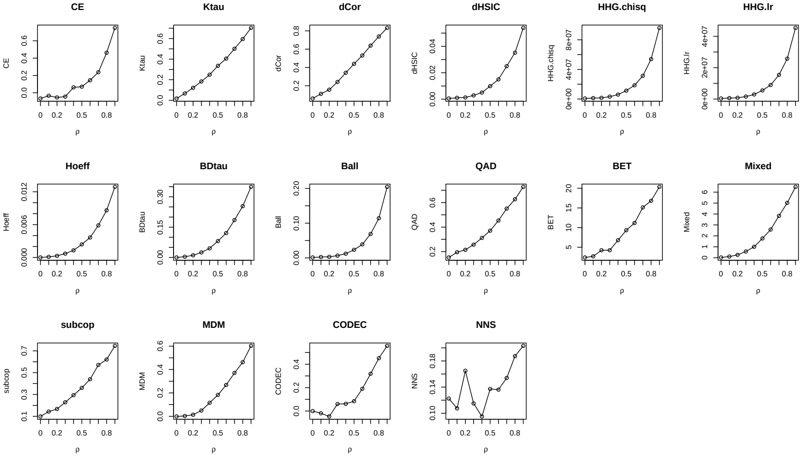
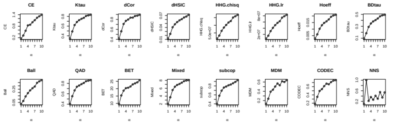
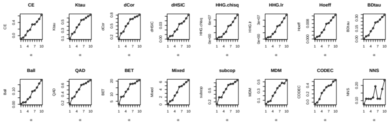


Figure: Experiments with bivariate normal copula functions.

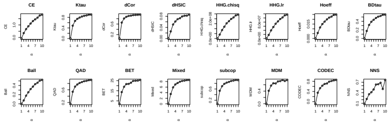
Independence Measures : Results III



(a) Clayton copula



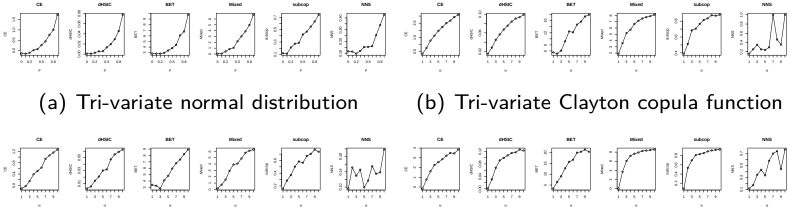
(b) Frank copula



(c) Gumbel copula

Figure: Experiments with bivariate Archimedean copula functions.

Independence Measures : Results IV



(a) Tri-variate normal distribution (b) Tri-variate Clayton copula function

(c) Tri-variate Frank copula function (d) Tri-variate Gumbel copula function

Figure: Experiments on multivariate measures.

Independence Measures : Results V

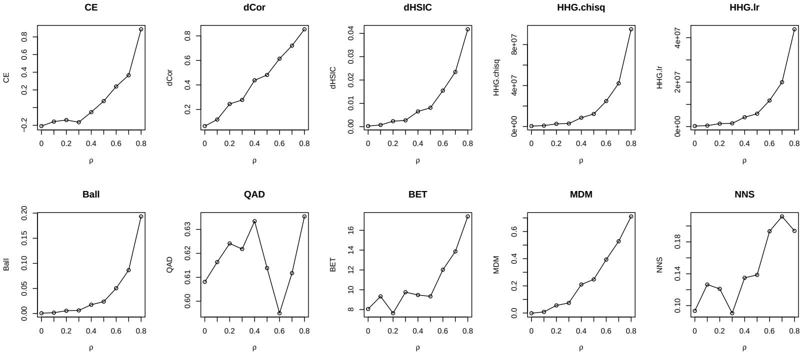


Figure: Experiments on independence between random vectors with quad-variate normal distributions.

Conditional Independence Measures : Implementations

Table: Conditional independence measures and their implementations.

Package	Measure	University
`copent`	CE	Tsinghua
`EDMeasure`	CMDM	UIUC / NWU
`FOCI`	CODEC	Stanford
`RCIT`	RCoT	Pittsburgh / CMU
`cdcsis`	CDC	SYSU / North Carolina / Yale
`GeneralisedCovarianceMeasure`	GCM	Cambridge / Copenhagen
`weightedGCM`	wGCM	ETHZ
`KPC`	KPC	Columbia
`ppcor`	pcor	-
`parCopCITest`	pcop	Copenhagen
`causallearn`	KCI	MPI
`knncmi`	CMI1	CMU
`pycit`	CMI2	DRL
`fcit`	FCIT	CalTech
`CCIT`	CCIT	UTAustin / Google / IBM
`pcit`	PCIT	UCL

Conditional Independence Measures: Experiments

1. random variables (X, Y, Z) governed by **trivariate normal distribution** $N(\mathbf{u}, \Sigma)$ with covariance matrix Σ as

$$\begin{vmatrix} 1 & \rho_{xy} & \rho_{xz} \\ \rho_{xy} & 1 & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & 1 \end{vmatrix}, \tag{7}$$

   where $\rho_{xy} = 0.7, \rho_{yz} = 0.6$, ρ_{xz} changes from 0 to 0.9 by step 0.1;

2. random variables (X, Y, Z) governed by **trivariate normal copula** $C_N(u_x, u_y, u_z; \Sigma)$ with covariance matrix Σ same as above, ρ_{xz} changes from 0 to 0.9 by step 0.1.

Conditional Independence Measures: Experiments

True value of CMI of normal distribution / copula

When the data are generated from trivariate normal distributions, the true value of CMI can be calculated as

$$CMI = \frac{1}{2}\log|\Sigma_{xz}| + \frac{1}{2}\log|\Sigma_{yz}| - \frac{1}{2}\log|\Sigma_{xyz}|, \tag{8}$$

where $\Sigma_{xyz}, \Sigma_{xz}, \Sigma_{yz}$ are correlation matrices accordingly.

Conditional Independence Measures : Results I

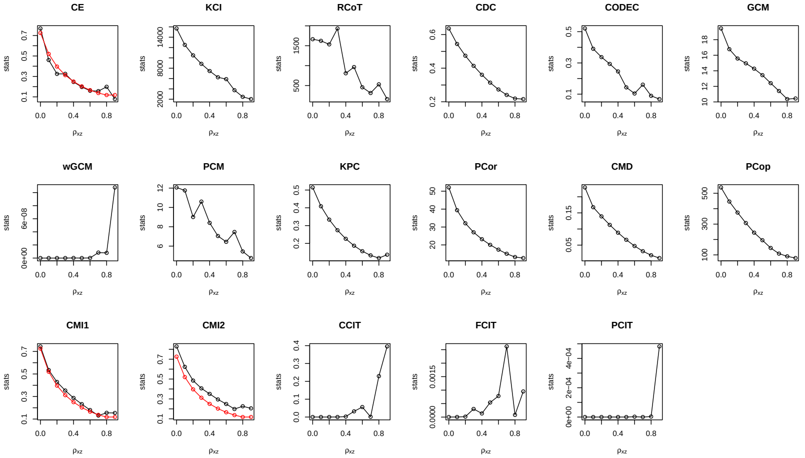


Figure: Experiments on conditional independence with tri-variate normal distributions.

Conditional Independence Measures : Results II

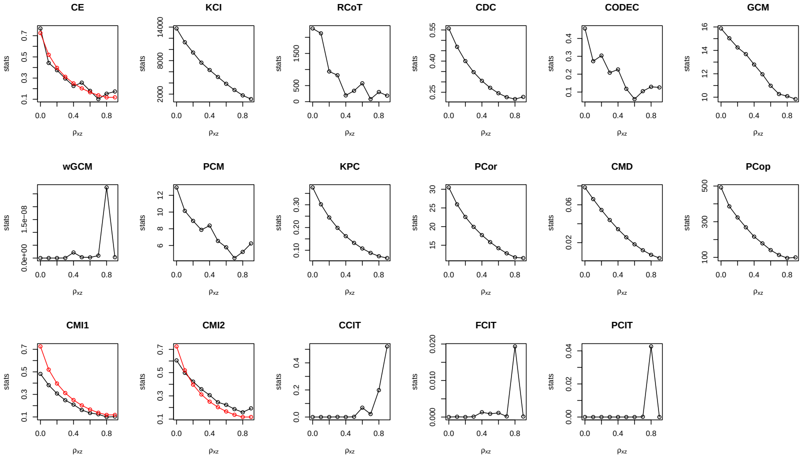


Figure: Experiments on conditional independence with tri-variate normal copula functions.

Multivariate Normality Tests : Implementations

Table: Multivariate normality tests and their implementations in R.

Package	Test	University
copent	CE	Tsinghua
MVN	Mardia	Leeds / Oxford
	Royston	MRC
	Henze-Zirker	Hannover
	Dornik-Haansen	Oxford / Copenhagen
	energy distance	NSF
mvnTest	Anderson-Darling	Columbia / Michigan
	Cramer-von Mises	Stockholm / Berlin
	McCulloch	Cornell
	Nikulin-Rao-Robson	Purdue
	McCulloch	OhioSU
	Dzhaparidze-Nikulin	Moscow
mnt	BHEP	Hannover
	Cox-Small	Imperial
	DEHT	KIT
	DEHU	Duisburg / KIT/ Hannover
	EHS	KIT
	HJG	KIT / Sevilla
	HV	NWU(SA)
	HZ	Hannover
	KKurt	Scripps
	MAKurt,MASkew	CaliforniaSU
	MKurt	Leeds / Oxford
	MQ1, MQ2	Simón Bolívar
	MRSSkew	UCSB / Trento / Debrecen
	MSkew	Lees / Oxford
	PU	CracowUT
	SR	NSF
mvnormtest	Shapiro-Wilk	MRC

Multivariate Normality Tests: Experiments

1. **Bivariate normal copula** $C_N(u, v)$ with margins as normal distribution $u \sim N(0, 1)$ and exponential distribution $v \sim E(\lambda)$. The parameter of exponential distribution $\lambda = 1, \ldots, 10$, when $\lambda = 2$, the simulated distribution is close to normal distribution and become non-normal as λ increases;
2. **Bivariate** t **distribution** $\mathbf{X} \sim t_\nu(\mathbf{0}, \Sigma)$, the non-diagonal element of Σ $\rho = 0.5$, non-normality is gradually weakened as $\nu = 1, \ldots, 10$;
3. **Mixture of two bivariate normal distributions** $X_1 \sim N_1(\mu_1, \rho_1)$ and $X_2 \sim N_2(\mu_2, \rho_2)$ with $\mu_1 = \mathbf{0}, \mu_2 = \mathbf{3}, \rho_1 = 0.5, \rho_2 = 0.8$and simulated data is derived from two samples from these two normal distributions as $\{(\beta - 1)X_1 + (10 - \beta)X_2\}/9$, where $\beta = 1, \ldots, 10$ so non-normality becomes stronger first and then weakened.

Multivariate Normality Tests : Results I

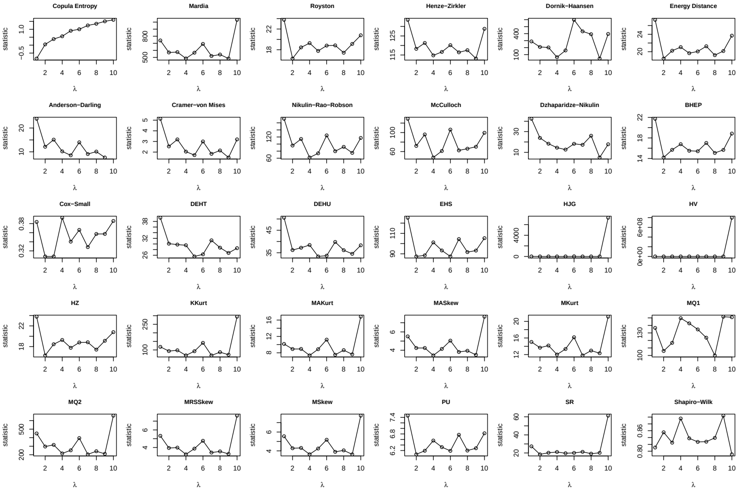


Figure: Experiments with bivariate normal distributions.

Multivariate Normality Tests : Results II

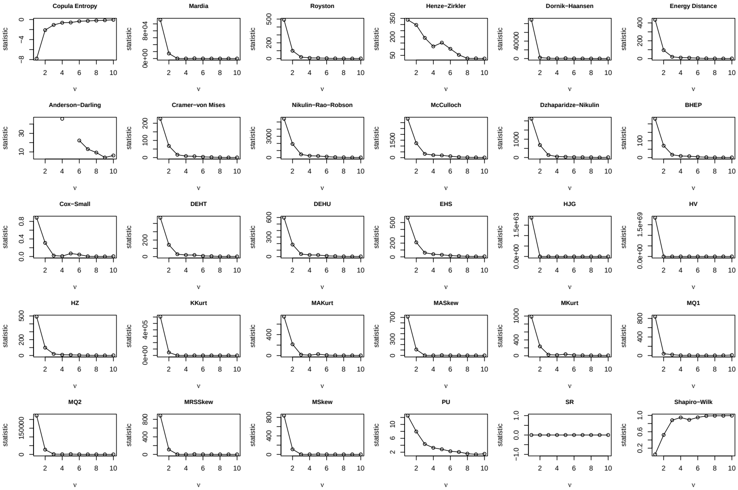


Figure: Experiments with bivariate t distributions.

Multivariate Normality Tests : Results III

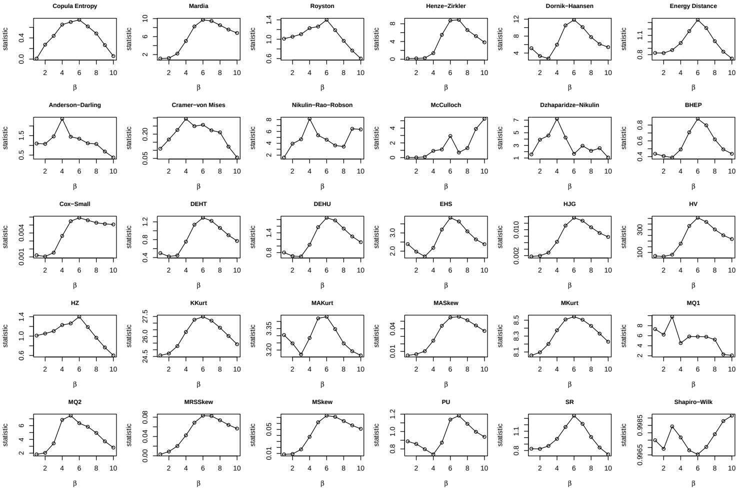


Figure: Experiments with mixture of two bivariate normal distributions.

Univariate Two-Sample Tests : Implementations

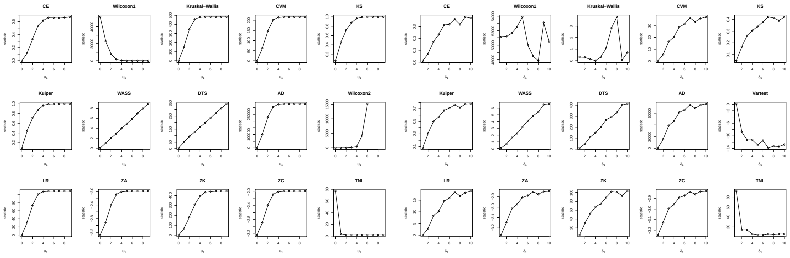
Table: Univariate two-sample tests and their implementations in R.

Package	Test	University
`copent`	CE	Tsinghua
`stat`	Wilcoxon1	OhioSU
	Kruskal-Wallis	Chicago
`twosamples`	CVM	Stockholm / Berlin
	KS	Moscow
	Kuiper	St Andrews
	WASS	GIT / UTAustin
	DTS	Chicago
	AD	Columbia / Michigan
`robusTest`	Wilcoxon2	OhioSU
	Vartest	UCL
`R2sample`	LR	Berkeley / UCSD
	ZA,ZK,ZC	Manitoba
`tnl.Test`	TNL	VCU / Ankara / Selcuk / Bentley / OhioSU

Univariate Two-Sample Tests: Experiments

1. **Two normal distributions with different mean**
   The first simulation includes a random variable governed by normal distribution $x_0 \sim N(\mu_0, \delta_0)$ with $\mu_0 = 0$ and $\delta_0 = 1$ and another random variable governed by normal distribution $x_2 \sim N(\mu_1, \delta_1)$ with μ_1 changes from 0 to 9 and $\delta_1 = 1$;
2. **Two normal distributions with different variances**
   The second simulation includes the first random variable x_0 same to the first simulation, and the second random variable $x_1 \sim N(\mu_1, \delta_1)$ with δ_1 changes from 1 to 10 and $\mu_1 = 0$.

Univariate Two-Sample Tests : Results



(a) Mean (b) Variance

Figure: Experiments on univariate two-sample tests.

Multivariate Two-Sample Tests : Implementations

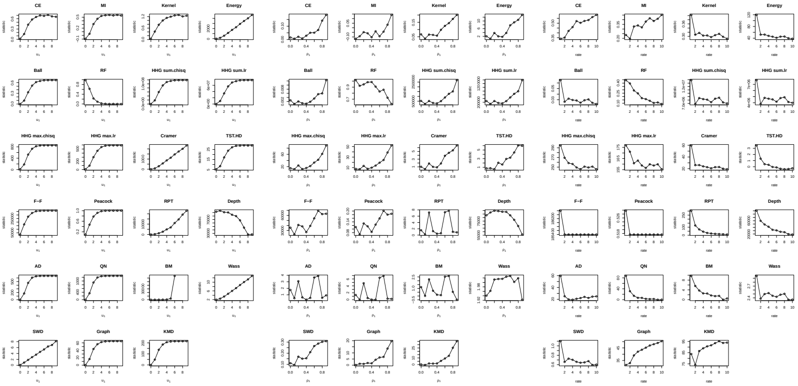
Table: Multivariate two-sample tests and their implementations in R.

Package	Test	University
`copent`	CE MI	Tsinghua Xavier / Birmingham
`kernlab`	Kernel	MPI / BNU / Yahoo
`energy`	Energy statistics	NSF
`Ball`	Ball divergence	SYSU / Yale
`hypoRF`	Random Forest	ETHZ / Zürich / Inria
`HHG`	HHG	Tel Aviv
`cramer`	Cramer	Hannover
`TwoSampleTest.HD`	TST.HD	Vigo
`fasano.franceschini.test`	F-F	Padova
`Peacock.test`	Peacock	Edinburgh
`RandomProjectionTest`	RPT	UCDavis / Sorbonne / MIT
`DepthProc`	Depth	Rutgers
`kSamples`	AD QN	Queensland Chicago
`lawstat`	BM	Göttingen
`T4transport`	WASS SWD	ENS / Google / ENASE ENS / Paris9 / Télécom Paris / Thales
`rgTest`	Graph	IowaSU
`KMD`	KMD	Columbia

Multivariate Two-Sample Tests: Experiments

1. **Two normal distributions with different mean**
   $\mathbf{x}_0 \sim N(u_0, \rho_0)$ and $\mathbf{x}_1 \sim N(u_1, \rho_1)$ with $u_0 = \mathbf{0}$ and $u_1 = (i, i)$, i changes from 0 to 9 by step 1, and $\rho_0, \rho_1 = 0.5$;
2. **Two normal distribution with different covariance**
   $\mathbf{x}_0 \sim N(u_0, \rho_0)$ with $u_0 = \mathbf{0}, \rho_0 = 0$ and another random variable $\mathbf{x}_1$ governed by normal distribution $N(u_1, \rho_1)$ with ρ_1 changes from 0 to 0.9 by step 0.1 and $u_1 = \mathbf{0}$;
3. **Normal distribution vs Normal copula**
   bivariate normal distribution $\mathbf{x}_0 \sim N(u_0, \rho_0)$ with $\rho_0 = 0$ and another random variable $\mathbf{x}_1$ governed by bivariate normal copula $c(u, v)$ with margins as normal distribution $u \sim N(0, 2)$ and exponential distribution $v \sim E(\lambda)$ with λ changes from 1 to 10.

Multivariate Two-Sample Tests : Results



(a) Mean (b) Covariance (c) Normal copula

Figure: Experiments on multivariate two-sample tests.

Change Point Detection : Implementations

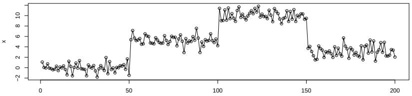
Table: Methods for change-point detection and their implementations in R.

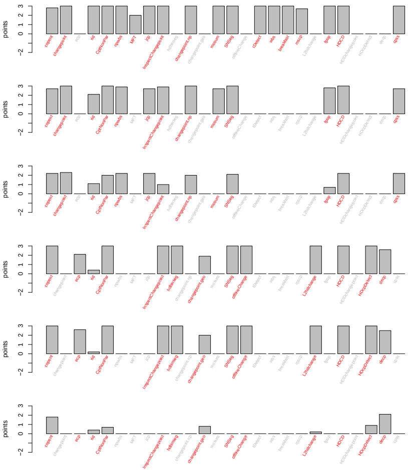
Package	Univariate			Multivariate			University
	Mean	MV	Var	Mean	MV	Var	
`copent`	✓	✓	✓	✓	✓	✓	Tsinghua
`changepoint`	✓	✓	✓				Lancester
`ecp`				✓	✓	✓	Cornell
`rid`	✓	✓	✓	✓	✓	✓	Tsinghua
`CptNonPar`	✓	✓	✓	✓	✓	✓	Bristol / Southampton
`npwbs`	✓	✓	✓				Edinburgh
`MFT`	✓						Frankfurt
`jcp`	✓	✓					TUWien
`InspectChangepoint`	✓	✓	✓	✓	✓	✓	Cambridge
`hdbinseg`				✓	✓	✓	Bristol / LSE
`changepoint.np`	✓	✓	✓				Lancester
`changepoint.geo`				✓	✓	✓	Lancester
`mosum`	✓	✓	✓				Magdeburg / Bristol
`SNSeg`	✓	✓	✓	✓	✓	✓	Miami / Notre Dame / Fudan / UIUC
`offlineChange`				✓	✓	✓	Harvard
`IDetect`	✓						LSE
`wbs`	✓						LSE
`breakfast`	✓						Cyprus / Bristol / LSE
`mscp`	✓						TUWien
`L2hdchange`				✓	✓	✓	WUSTL / York / Chicago
`gfpop`	✓	✓	✓				Évry / Arizona / Lancester / Clemson
`HDCD`	✓	✓	✓	✓	✓	✓	Oslo
`HDDchangepoint`				✓	✓	✓	Durham / GWU
`HDcpDetect`				✓	✓	✓	KentSU / PKU / MichiganSU
`decp`				✓	✓	✓	Lancester

Change Point Detection: Results

Simulated change points:

- Univariate cases
  - Mean
  - Variance
  - Mean-Variance
- Multivariate cases
  - Mean
  - Variance
  - Mean-Variance





Symmetry Test: Implementation

Table: Methods implemented in the `symmetry` package in R.

Method	Statistic	University
MI	The Mira test statistic	USI
CM	The Cabilio–Masaro test statistic	Acadia
MGG	The Miao, Gel and Gastwirth test statistic	Macalester / Waterloo / GWU
B1	The $\sqrt{b_1}$ test statistic	Belgrade
KS	The Kolmogorov–Smirnov test statistic	Belgrade
SGN	The Sign test statistic	Belgrade
WCX	The Wilcoxon test statistic	Belgrade
FM	The characterization based test	Alberta / Toronto
RW	The Rothman-Woodroofe test statistic	Hannover
BHI	The Litvinova test statistic	St Petersburg
BHK	The Baringhaus and Henze supremum-type test statistic	Hannover / KIT
BH2	The Baringhaus-Henze test statistic	Hannover / KIT
MOI,MOK	The Milošević and Obradović test statistics	St Petersburg / Rider
NAI,NAK	The Nikitin and Ahsanullah test statistics	St Petersburg / Rider
K2,K2U	The Božin, Milošević, Nikitin and Obradović Kolmogorov type statistics	Belgrade / St Petersburg
NAC1,NAC2,BHC1,BHC2	The Allison and Pretorius test statistics	NWU(SA)

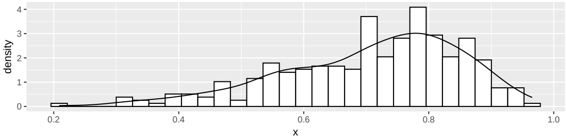
Symmetry Tests: Experiments

Beta distribution

Beta distributions are the continuous distribution on $[0, 1]$ or $(0, 1)$ with two parameters $a, b > 0$ for symmetry control, defined as

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \tag{9}$$

where Γ is Gamma function.



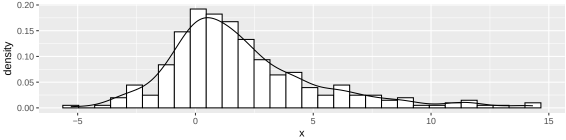
Symmetry Tests: Experiments

Asymmetric Laplace distribution

Asymmetric Laplace distributions generalize Laplace distribution and are defined as

$$f(x; \mu, \delta, k) = \frac{1}{\delta(k + k^{-1})} e^{-(x-\mu)k^s s/\delta}, \tag{10}$$

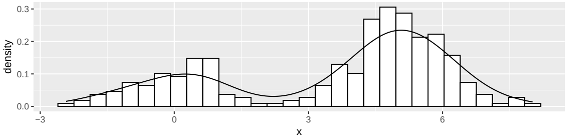
where $s = sgn(x - \mu)$, μ, δ is for location and scale, and k is the parameter for symmetry control.



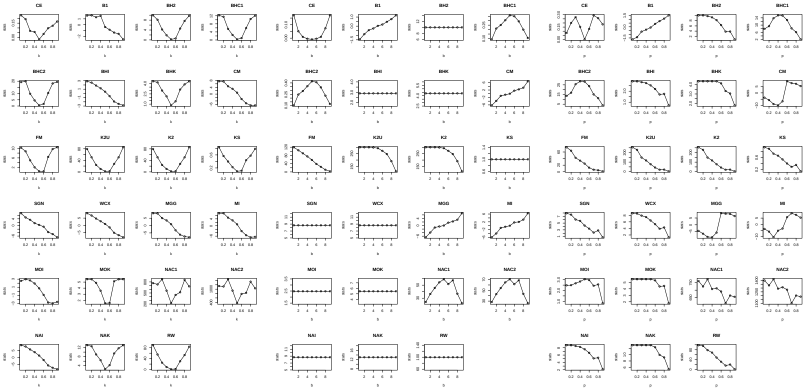
Symmetry Tests: Experiments

Bimodal normal distribution

Bimodal normal distribution is derived by mixture of two normal distributions $X_1 \sim N(\mu_1, \delta_1)$ and $X_2 \sim N(\mu_2, \delta_2)$ with a proportion parameter p for symmetry control.



Symmetry Tests: Results



(a) Asymmetric Laplace distributions

(b) Beta distributions

(c) Bimodal normal distributions

Figure: Simulation results on three types of distributions.

Copula Hypothesis Test: Implementation

Table: Methods implemented in the `gofCopula` package in R.

Method	**Statistic**	**University**
CvM	Cramér-von Mises statistic	Laval / HEC Montréal
KS	Kolmogorov-Smirnov statistic	Laval / HEC Montréal
KendallCvM	Kendall's process based CvM statistic	Laval / HEC Montréal
KendallKS	Kendall's process based KS statistic	Laval / HEC Montréal
PIOSTn,PIOSRn	PIOS tests	SWFEU / TUDresden / Simon Fraser / MississippiSU / Michigan
Rosenblatt*	Rosenblatt transformation based tests	Laval / HEC Montréal
White	White's information matrix equality based test	TUM

Copula Hypothesis Tests: Experiments

- Simulated copulas:
  - Gaussian copula
  - Archimedean copulas (Gumbel,Frank,Clayton)
- Hypotheses of copulas:
  - Gaussian copula
  - Archimedean copulas (Gumbel,Frank,Clayton)

Copula Hypothesis Tests: Results

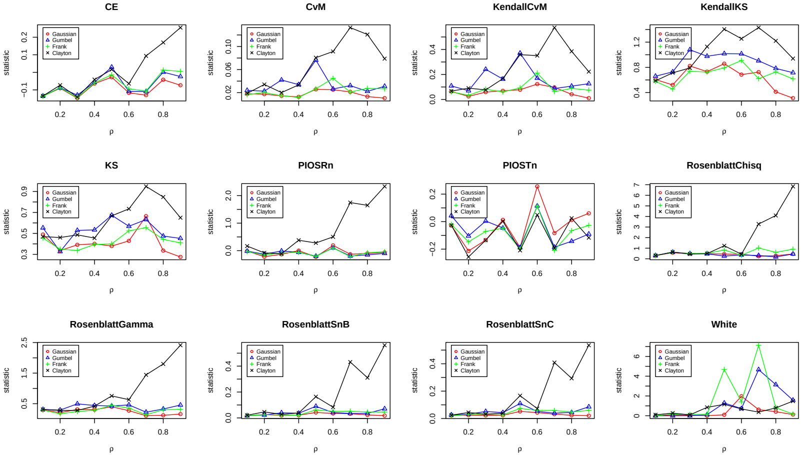


Figure: Simulation results on Gaussian copulas.

Copula Hypothesis Tests: Results

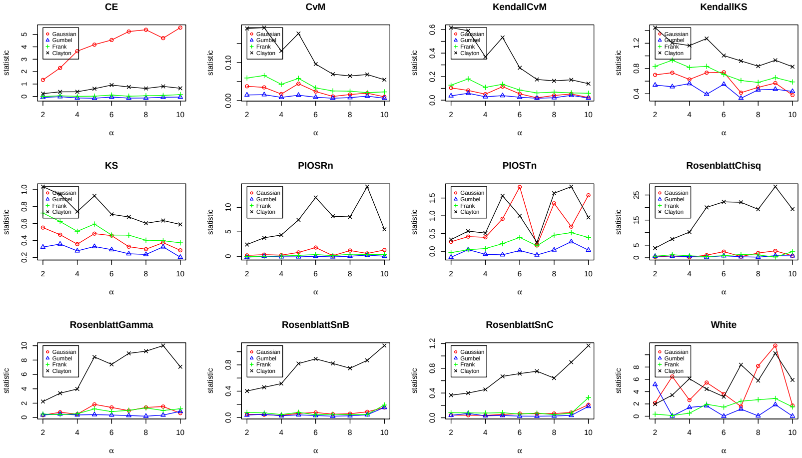


Figure: Simulation results on Gumbel copulas.

Copula Hypothesis Tests: Results

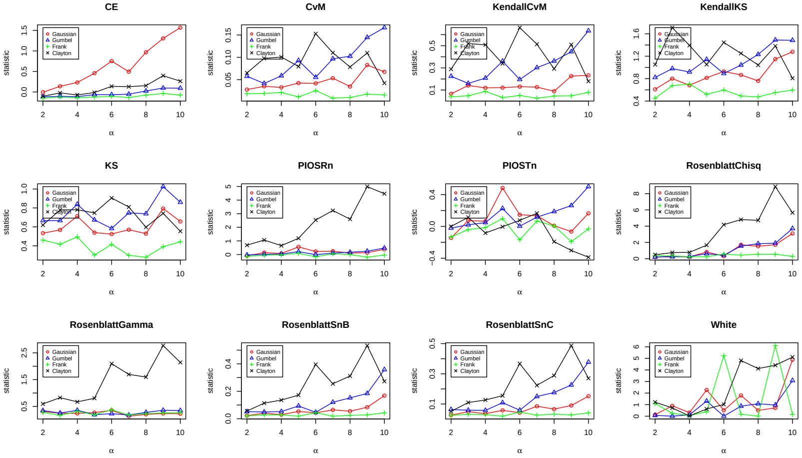


Figure: Simulation results on Frank copulas.

Copula Hypothesis Tests: Results

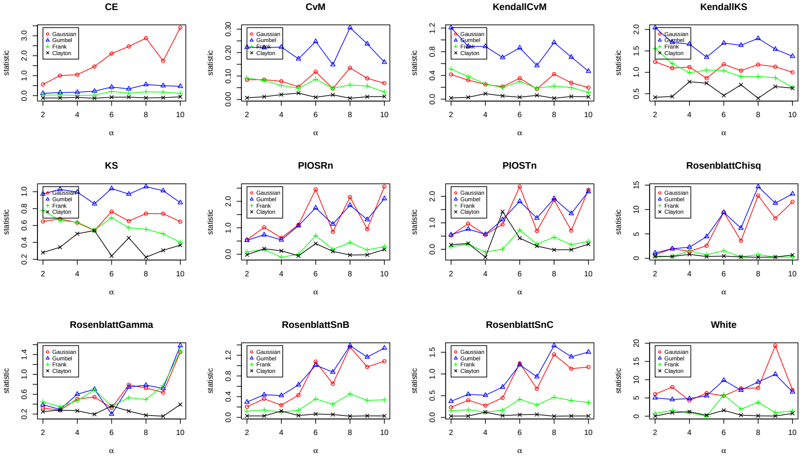


Figure: Simulation results on Clayton copulas.

Monograph

Jian Ma. "Copula Entropy: Theory and Applications". In: *arXiv preprint arXiv:2025.18168* (2025)

Thanks!

Q&A