

Semi-parametric estimation of mutual information ; optimal test of independence

Philippe Regnault¹
joint work with Amor Keziou¹

¹ Laboratoire de Mathématiques de Reims, EA 4535,
Université de Reims Champagne-Ardenne, France

2nd conference on Geometric Science of Information - 30/10/2015

Testing independence by means of mutual informations (MIs) thresholding

- Classical independence measures and related tests

- φ -mutual informations

- Mutual information test for real-valued r.v.

Semi-parametric modeling of the joint distribution

- The semi-parametric model

- Some examples

Dual estimates of mutual informations

- Dual representation of φ -MI

- Estimating φ -MI via the duality technique

- Asymptotic properties of the dual estimates

Bahadur asymptotic efficiency of φ -MI based tests

- KL-MI test maximizes Bahadur slope

Simulations

- For finite random variables

- For Gaussian couples

Classical χ^2 and MI independence tests

Notations :

- ▶ (X, Y) , couple of r.v. taking values in a finite space $\mathcal{X} \times \mathcal{Y}$;
- ▶ \mathbb{P} joint distribution of (X, Y) ;
- ▶ $\mathbb{P}^\perp := \mathbb{P}_1 \otimes \mathbb{P}_2$, product distribution of margins \mathbb{P}_1 and \mathbb{P}_2 of (X, Y) .

Dependence measures :

- ▶ χ^2 -dependence measure (\mathcal{X} and \mathcal{Y} finite sets, $\mathbb{P} = (p_{x,y})_{(x,y)}$) :

$$\chi^2(\mathbb{P}, \mathbb{P}^\perp) := \frac{1}{2} \int_{\mathcal{X} \times \mathcal{Y}} \left(\frac{d\mathbb{P}}{d\mathbb{P}^\perp}(x, y) - 1 \right)^2 d\mathbb{P}^\perp(x, y) = \frac{1}{2} \sum_{(x,y)} \frac{(p_{x,y} - p_x p_y)^2}{p_x p_y},$$

where $d\mathbb{P}/d\mathbb{P}^\perp$ denotes the density of \mathbb{P} with respect to \mathbb{P}^\perp .

- ▶ Mutual information (MI) :

$$I_{KL}(\mathbb{P}) := \mathbb{K}(\mathbb{P}, \mathbb{P}^\perp) = \int \frac{d\mathbb{P}}{d\mathbb{P}^\perp} \log \frac{d\mathbb{P}}{d\mathbb{P}^\perp} d\mathbb{P}^\perp = \sum_{(x,y)} p_{x,y} \log \frac{p_{x,y}}{p_x p_y},$$

where $\mathbb{K}(Q, P) := \int \frac{dQ}{dP} \log \frac{dQ}{dP} dP$, for P, Q distr., with Q a.c.w.r.t. P .

χ^2 and mutual information independence tests

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be an n -sample of $(X, Y) \sim \mathbb{P} = (p_{x,y})_{(x,y)}$. For testing

\mathcal{H}_0 : X and Y are independent, against $\mathcal{H}_1 := \overline{\mathcal{H}_0}$,

the **test statistics for χ^2 and MI independence tests** are :

$$2n\chi^2(\hat{\mathbb{P}}, \hat{\mathbb{P}}^\perp) = n \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \frac{(\hat{p}_{x,y} - \hat{p}_x \hat{p}_y)^2}{\hat{p}_x \hat{p}_y},$$

$$2n I_{KL}(\hat{\mathbb{P}}) = 2n \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \hat{p}_{x,y} \log \frac{\hat{p}_{x,y}}{\hat{p}_x \hat{p}_y},$$

where $\hat{\mathbb{P}} := (\hat{p}_{x,y})_{(x,y)}$ and $\hat{\mathbb{P}}^\perp := (\hat{p}_x \hat{p}_y)_{(x,y)}$ are, respectively, the empirical versions of $\mathbb{P} = (p_{x,y})_{(x,y)}$ and $\mathbb{P}^\perp = (p_x p_y)_{(x,y)}$.

Both $2n\chi^2(\hat{\mathbb{P}}, \hat{\mathbb{P}}^\perp)$ and $2n I_{KL}(\hat{\mathbb{P}})$ are asymptotically chi-squared distributed, with $(|\mathcal{X}| - 1)(|\mathcal{Y}| - 1)$ degrees of freedom under \mathcal{H}_0 .

φ -mutual informations

(X, Y) taking its values in a general measurable space $(\mathcal{X} \times \mathcal{Y}, \mathcal{A}_X \otimes \mathcal{A}_Y)$, with joint distribution $\mathbb{P} \in \mathcal{M}_1(\mathcal{X} \times \mathcal{Y})$.

Let $\varphi : \mathbb{R} \rightarrow [0, +\infty]$ be some non-negative closed proper convex function such that $\varphi(1) = 0$, $\varphi'(1) = 0$ and $\varphi''(1) = 1$.

φ -mutual information (φ -MI) of (X, Y) :

$$I_\varphi(X, Y) := I_\varphi(\mathbb{P}) := D_\varphi(\mathbb{P}, \mathbb{P}^\perp) := \int_{\mathcal{X} \times \mathcal{Y}} \varphi \left(\frac{d\mathbb{P}}{d\mathbb{P}^\perp}(x, y) \right) d\mathbb{P}^\perp(x, y).$$

- ▶ for $\varphi_1(x) = x \log x - x + 1$, $I_{\varphi_1}(\mathbb{P}) = I_{KL}(\mathbb{P})$ (classic MI) ;
- ▶ for $\varphi_2(x) = \frac{1}{2}(x - 1)^2$, $I_{\varphi_2}(\mathbb{P}) = \chi^2(\mathbb{P}, \mathbb{P}^\perp)$ (χ^2 measure of dependence).

φ -mutual informations

(X, Y) taking its values in a general measurable space $(\mathcal{X} \times \mathcal{Y}, \mathcal{A}_X \otimes \mathcal{A}_Y)$, with joint distribution $\mathbb{P} \in \mathcal{M}_1(\mathcal{X} \times \mathcal{Y})$.

Let $\varphi : \mathbb{R} \rightarrow [0, +\infty]$ be some non-negative closed proper convex function such that $\varphi(1) = 0$, $\varphi'(1) = 0$ and $\varphi''(1) = 1$.

φ -mutual information (φ -MI) of (X, Y) :

$$I_\varphi(X, Y) := I_\varphi(\mathbb{P}) := D_\varphi(\mathbb{P}, \mathbb{P}^\perp) := \int_{\mathcal{X} \times \mathcal{Y}} \varphi \left(\frac{d\mathbb{P}}{d\mathbb{P}^\perp}(x, y) \right) d\mathbb{P}^\perp(x, y).$$

- ▶ for $\varphi_1(x) = x \log x - x + 1$, $I_{\varphi_1}(\mathbb{P}) = I_{KL}(\mathbb{P})$ (classic MI) ;
- ▶ for $\varphi_2(x) = \frac{1}{2}(x - 1)^2$, $I_{\varphi_2}(\mathbb{P}) = \chi^2(\mathbb{P}, \mathbb{P}^\perp)$ (χ^2 measure of dependence).

φ -mutual informations

(X, Y) taking its values in a general measurable space $(\mathcal{X} \times \mathcal{Y}, \mathcal{A}_X \otimes \mathcal{A}_Y)$, with joint distribution $\mathbb{P} \in \mathcal{M}_1(\mathcal{X} \times \mathcal{Y})$.

Let $\varphi : \mathbb{R} \rightarrow [0, +\infty]$ be some non-negative closed proper convex function such that $\varphi(1) = 0$, $\varphi'(1) = 0$ and $\varphi''(1) = 1$.

φ -mutual information (φ -MI) of (X, Y) :

$$I_\varphi(X, Y) := I_\varphi(\mathbb{P}) := D_\varphi(\mathbb{P}, \mathbb{P}^\perp) := \int_{\mathcal{X} \times \mathcal{Y}} \varphi \left(\frac{d\mathbb{P}}{d\mathbb{P}^\perp}(x, y) \right) d\mathbb{P}^\perp(x, y).$$

- ▶ for $\varphi_1(x) = x \log x - x + 1$, $I_{\varphi_1}(\mathbb{P}) = I_{KL}(\mathbb{P})$ (classic MI);
- ▶ for $\varphi_2(x) = \frac{1}{2}(x - 1)^2$, $I_{\varphi_2}(\mathbb{P}) = \chi^2(\mathbb{P}, \mathbb{P}^\perp)$ (χ^2 measure of dependence).

Fundamental property :

$I_\varphi(\mathbb{P}) \geq 0$ and $I_\varphi(\mathbb{P}) = 0$ iff $\mathbb{P} = \mathbb{P}^\perp$ iff X and Y are independent.

Hence, the independence test problematic can be reformulated in

$$\mathcal{H}_0 : I_\varphi(\mathbb{P}) = 0 \quad \text{against} \quad \mathcal{H}_1 : I_\varphi(\mathbb{P}) > 0.$$

Dealing with continuous r.v. – classical approaches

Main estimation methods in the literature :

- ▶ **Plug-in** : replace \mathbb{P} by its empirical counterpart in the definition of I_φ .
Widely used and studied when \mathcal{X}, \mathcal{Y} finite sets ; see e.g., Pardo (2006).
Yields poor efficiency (power) for the related tests when dealing with continuous r.v., since it requires gathering data into classes (choice of the (number of) classes, etc).
- ▶ **Kernel estimation** : plug kernel estimates of the joint density and marginals into I_φ .
Leads to the difficulty of choosing the kernel and the bandwidth ; see e.g. Khan (2007).
Unkown (asymptotic) distributions of estimates.

Modeling the ratio $d\mathbb{P}/d\mathbb{P}^\perp$

Assume that the joint distribution \mathbb{P} of the random vector (X, Y) belongs to the semi-parametric model

$$\mathcal{M}_\Theta := \left\{ P \in \mathcal{M}_1(\mathcal{X} \times \mathcal{Y}); \frac{dP}{dP^\perp}(\cdot, \cdot) =: h_\theta(\cdot, \cdot); \theta = (\alpha, \beta^\top)^\top \in \Theta \right\},$$

where $\Theta \subseteq \mathbb{R}^{1+d}$ is the parameter space, and

$h_\theta(\cdot, \cdot) : (x, y) \in \mathcal{X} \times \mathcal{Y} \mapsto h_\theta(x, y) \in \mathbb{R}$ is some specified real-valued function, indexed by the parameter θ .

Main assumptions on the model :

- A.1** (identifiability) $(h_\theta(x, y) = h_{\theta'}(x, y), \forall (x, y) \in \mathcal{X} \times \mathcal{Y}) \Rightarrow (\theta = \theta')$;
- A.2** there exists (a unique) $\theta_0 \in \text{int}(\Theta)$ such that $h_{\theta_0}(x, y) = 1$, for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

The Gaussian model

If (X, Y) is Gaussian with unknown mean $\mu := (\mu_1, \mu_2)^\top$ and unknown variance matrix Γ , then

$$d\mathbb{P}/d\mathbb{P}^\perp =: h_\theta(x, y) = \exp \{ \alpha + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy \},$$

and $\theta := (\alpha, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\top$.

Note that the number of free parameters in θ_T is d , and that α_T is considered as a normalizing parameter due to the constraint

$$\int_{\mathcal{X} \times \mathcal{Y}} h_{\theta_T}(x, y) d\mathbb{P}^\perp(x, y) = \int_{\mathcal{X} \times \mathcal{Y}} d\mathbb{P}(x, y) = 1.$$

Moreover, the independence parameter is $\theta_0 := (0, \dots, 0)^\top \in \mathbb{R}^6$.

The finite support model

Assume that the support of \mathbb{P} , $\text{supp}(\mathbb{P}) = \mathcal{X} \times \mathcal{Y}$, is a finite-discrete set of size $K_1 \times K_2$; denote by $(\mathbb{P}(x, y))_{(x, y) \in \mathcal{X} \times \mathcal{Y}} := (p_{x, y})_{(x, y) \in \mathcal{X} \times \mathcal{Y}}$ the density of \mathbb{P} with respect to the counting measure on $\mathcal{X} \times \mathcal{Y}$.

We have

$$\frac{d\mathbb{P}}{d\mathbb{P}^\perp}(x, y) = \exp \left(\sum_{(a, b) \in \mathcal{X} \times \mathcal{Y}} \theta_{a, b} \delta_{(a, b)}(x, y) \right),$$

where

$$\theta_{a, b} = \log \left[\frac{p_{a, b}}{p_a p_b} \right], \quad (a, b) \in \mathcal{X} \times \mathcal{Y}.$$

Here, $\theta = (\theta_{a, b})_{(a, b) \in \mathcal{X} \times \mathcal{Y}} \in \Theta \subset \mathbb{R}^{K_1 K_2}$. The number of free-parameters in θ_T is then equal to $(K_1 - 1)(K_2 - 1)$.

The independence parameter is $\theta_0 := (0, \dots, 0)^T \in \mathbb{R}^{K_1 K_2}$.

Parametric copula models

Assume that \mathbb{P} has continuous c.d.f. F on \mathbb{R}^2 , with marginal c.d.f. F_1 and F_2 . The copula $C(\cdot, \cdot)$ of the vector (X, Y) , see e.g. Nelsen (2006), is defined, $\forall (u, v) \in]0, 1[^2$, by

$$C(u, v) := F(F_1^{-1}(u), F_2^{-1}(v)).$$

If $F(\cdot, \cdot)$ is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^2 , then we have the relation

$$\frac{d\mathbb{P}}{d\mathbb{P}^\perp}(x, y) = \frac{f(x, y)}{f_1(x)f_2(y)} = c(F_1(x), F_2(y)),$$

where $f(\cdot, \cdot)$ is the joint density of (X, Y) , f_1 and f_2 are the marginal densities, of X and Y , and $c(\cdot, \cdot)$ the copula density.

Numerous examples of semi-parametric models can be obtained by considering

$$h_\theta(x, y) = c_\theta(F_1(x), F_2(y)), \quad \theta \in \mathbb{R}^d$$

where $\{c_\theta, \theta \in \mathbb{R}^d\}$ is any parametric copula model.

Dual representation of φ -divergences

Notations :

\mathcal{F} : class of measurable real function defined on (Ω, \mathcal{A}) ;

$\langle \mathcal{F} \cup \mathcal{B} \rangle$: linear subspace spanned by \mathcal{F} and bounded functions ;

$\mathcal{M}_{\mathcal{F}}$: set of signed measures such that $|f|$ is integrable, $f \in \mathcal{F}$;

$\tau_{\mathcal{F}}$: weakest topology on $\mathcal{M}_{\mathcal{F}}$ for which applications $Q \in \mathcal{M}_{\mathcal{F}} \mapsto \int f dQ$, $f \in \langle \mathcal{F} \cup \mathcal{B} \rangle$ are continuous ;

$\tau_{\mathcal{M}}$: weakest topology on $\langle \mathcal{F} \cup \mathcal{B} \rangle$ for which applications $f \in \mathcal{M}_{\mathcal{F}} \mapsto \int f dQ$, $Q \in \mathcal{M}_{\mathcal{F}}$ are continuous.

(Browniatowski & Keziou, 2003, 2006)

For any signed measure P on (Ω, \mathcal{A}) , the application $Q \mapsto D_{\varphi}(Q, P)$ is convex and l.s.c. on $\mathcal{M}_{\mathcal{F}}$.

In addition, if Q is a.c.w.r.t. P , and $\varphi' \left(\frac{dQ}{dP} \right) \in \mathcal{F}$, the following dual representation holds :

$$D_{\varphi}(Q, P) = \sup_{f \in \mathcal{F}} \left\{ \int f dQ - \int \varphi^*(f) dP \right\},$$

where φ^* is the convex conjugate of φ .

Moreover, the supremum is achieved uniquely for $f = \varphi' \left(\frac{dQ}{dP} \right)$.

Dual representation of φ -MI in the semi-parametric context

Assumptions on the model :

A.3 $I_\varphi(\mathbb{P}) < \infty$;

A.4 for all $\theta \in \Theta$, $\int |\varphi'(h_\theta)| d\mathbb{P} < \infty$.

Specifying the dual representation of $D\varphi$ for $Q = \mathbb{P}$, $P = \mathbb{P}^\perp$ and $\mathcal{F} = \{\varphi'(h_\theta), \theta \in \Theta\}$, we obtain

Dual representation of φ -MI

$$I_\varphi(\mathbb{P}) = \sup_{\theta \in \Theta} \left\{ \int \varphi'(h_\theta) d\mathbb{P} - \int \varphi^*(\varphi'(h_\theta)) d\mathbb{P}^\perp \right\}.$$

The supremum is achieved uniquely for $\theta = \theta_T$ (the true parameter).

Dual estimates

Given $(X_1, Y_1), \dots, (X_n, Y_n)$ an n -sample of $(X, Y) \sim \mathbb{P}$, with \mathbb{P} such that $\frac{d\mathbb{P}}{d\mathbb{P}^\perp} = h_{\theta_T}$, we define the dual estimators of $I_\varphi(\mathbb{P})$ and θ_T respectively as

$$\begin{aligned} \hat{I}_\varphi &:= \sup_{\theta \in \Theta} \left\{ \int \varphi'(h_\theta) d\hat{\mathbb{P}} - \int \varphi^*(\varphi'(h_\theta)) d\hat{\mathbb{P}}_1 \otimes \hat{\mathbb{P}}_2 \right\} \\ &= \sup_{\theta \in \Theta} \left\{ \frac{1}{n} \sum_{i=1}^n \varphi'(h_\theta(X_i, Y_i)) - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \varphi^*(\varphi'(h_\theta(X_i, Y_j))) \right\}, \\ \hat{\theta}_\varphi &:= \arg \sup_{\theta \in \Theta} \left\{ \int \varphi'(h_\theta) d\hat{\mathbb{P}} - \int \varphi^*(\varphi'(h_\theta)) d\hat{\mathbb{P}}_1 \otimes \hat{\mathbb{P}}_2 \right\} \\ &= \arg \sup_{\theta \in \Theta} \left\{ \frac{1}{n} \sum_{i=1}^n \varphi'(h_\theta(X_i, Y_i)) - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \varphi^*(\varphi'(h_\theta(X_i, Y_j))) \right\}, \end{aligned}$$

where $\hat{\mathbb{P}}$ denotes the empirical distribution of the sample and $\hat{\mathbb{P}}^\perp = \hat{\mathbb{P}}_1 \otimes \hat{\mathbb{P}}_2$ the product of its marginals.

Example : finite distributions

In the context of finite distributions, dual and empirical estimates are equal :

$$\hat{I}_\varphi = \sum_{x,y} \varphi \left(\frac{\hat{p}_{x,y}}{\hat{p}_x \hat{p}_y} \right) \hat{p}_x \hat{p}_y.$$

Particularly, for $\varphi = \varphi_2$, the duality technique recovers the classic χ^2 estimate.

In addition,

$$\hat{\theta}_{x,y} = \log \frac{\hat{p}_{x,y}}{\hat{p}_x \hat{p}_y}, \quad (x,y) \in \mathcal{X} \times \mathcal{Y},$$

do not depend on the choice of φ .

Dual estimates for semi-parametric copula models

When dealing with semi-parametric copula models

$$h_{\theta}(x, y) = c_{\theta}(F_1(x), F_2(y)),$$

with unknown non-parametric cumulative distribution functions F_1 and F_2 , it is necessary to estimate them, for example, empirically.

Hence, the dual estimates of $I_{\varphi}(\mathbb{P})$ and $\theta_{\mathcal{T}}$ become

$$\widehat{I}_{\varphi} := \sup_{\theta \in \Theta} \left\{ \frac{1}{n} \sum_i \varphi' \circ c_{\theta} \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) - \frac{1}{n^2} \sum_{i,j} \varphi^* \circ \varphi' \circ c_{\theta} \left(\frac{R_i}{n+1}, \frac{S_j}{n+1} \right) \right\},$$

$$\widehat{\theta}_{\varphi} := \arg \sup_{\theta \in \Theta} \left\{ \frac{1}{n} \sum_i \varphi' \circ c_{\theta} \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) - \frac{1}{n^2} \sum_{i,j} \varphi^* \circ \varphi' \circ c_{\theta} \left(\frac{R_i}{n+1}, \frac{S_j}{n+1} \right) \right\}$$

where R_i is the rank of X_i in the sample X_1, \dots, X_n and S_j is the rank of Y_j in the sample Y_1, \dots, Y_n .

Consistency of dual estimates

Assumptions on the model :

- A.5** the parameter space is compact ;
- A.6** $\theta \mapsto \varphi'(h_\theta(x, y))$ is continuous for all (x, y) and $\int_{\mathcal{X} \times \mathcal{Y}} \sup_{\theta \in \Theta} |f_\theta(x, y)| \, d\mathbb{P}(x, y) < \infty$;
- A.7** $\theta \mapsto \varphi^*(\varphi'(h_\theta(x, y)))$ is continuous for all (x, y) and $\int_{\mathcal{X} \times \mathcal{Y}} \sup_{\theta \in \Theta} g_\theta(x, y)^2 \, d\mathbb{P}^\perp(x, y) < \infty$.

Proposition

Assume A.1-2, 5-7.

The dual estimates \widehat{I}_φ of $I_\varphi(\mathbb{P})$ and $\widehat{\theta}_\varphi$ of θ_T are consistent. Precisely, as $n \rightarrow \infty$, the following convergences in probability hold

$$\widehat{I}_\varphi \rightarrow I_\varphi(\mathbb{P}) \quad \text{and} \quad \widehat{\theta}_\varphi \rightarrow \theta_T.$$

Asymptotic distribution of \widehat{I}_{KL}

We assume the following specific form for the density ratio $h_\theta = \frac{d\mathbb{P}}{d\mathbb{P}^\perp}$:

$$h_\theta(x, y) = \exp(\alpha + m_\beta(x, y)) \quad \text{with} \quad m_\beta(x, y) := \sum_{k=1}^d \beta_k \xi_k(x) \zeta_k(y),$$

for some specified measurable real valued functions ξ_k and ζ_k , $k = 1, \dots, d$, defined, respectively, on \mathcal{X} and \mathcal{Y} .

Assumptions on the model :

- A.8** there exists a neighborhood $N(\theta_T)$ of θ_T such that $\{(x, y) \mapsto (\partial^3 / \partial^3 \theta) \varphi'(h_\theta(x, y)); \theta \in N(\theta_T)\}$ (resp. $\{(x, y) \mapsto (\partial^3 / \partial^3 \theta) \varphi^*(\varphi'(h_\theta(x, y))); \theta \in N(\theta_T)\}$) are dominated by some \mathbb{P} -integrable functions (resp. some \mathbb{P}^\perp -square-integrable functions) ;
- A.9** the integrals $\mathbb{P} \|f'_{\theta_T}\|^2$, $\mathbb{P}^\perp \|g'_{\theta_T}\|^2$, $\mathbb{P} \|f''_{\theta_T}\|$, $\mathbb{P}^\perp \|g''_{\theta_T}\|^2$ exist, and the matrix $\Sigma_1 := -(\mathbb{P} f''_{\theta_T} - \mathbb{P}^\perp g''_{\theta_T})$ is nonsingular.

Asymptotic distribution of \hat{I}_{KL}

Proposition

Assume A.1-2, 5-9. Under the null hypothesis ($\mathbb{P} = \mathbb{P}^\perp$), we have :

- ▶ $\sqrt{n}\hat{\theta}_{\varphi_1}$ converges in distribution to a centered multivariate normal random variable with covariance matrix $\Sigma = \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$, where Σ_2 is explicit ;
- ▶ $2n\hat{l}_{\varphi_1}$ converges in distribution to the random variable $Z^\top Z$, where Z is a centered multivariate normal random variable with covariance matrix

$$C = \Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}.$$

The most efficient test (among φ -MI tests) is...

Additional assumption

A.10 for all $f \in \varphi'(\mathcal{M}_\Theta)$, for all $a > 0$, $\int \exp(a|f|)d\mathbb{P}^\perp < \infty$ and $\int \exp(a|\varphi * \circ f|)d\mathbb{P}^\perp < \infty$.

Theorem (KL-MI test is the most efficient)

Let (X, Y) a couple of random variables with joint distribution $\mathbb{P} \in \mathcal{M}_\Theta$. Suppose that conditions (A.1-2, 5, 10) are fulfilled.

For the test problem

$$H_0 : \mathbb{P} = \mathbb{P}^\perp \quad \text{against} \quad H_1 : \mathbb{P} \neq \mathbb{P}^\perp,$$

the test based on the estimate \hat{I}_{KL} of the Kullback-Leibler mutual information is uniformly (i.e., whatever be the alternative $\mathbb{P} \in \mathcal{M}_\Theta$ satisfying conditions (A.1-2, 5, 10)) the most efficient test (in Bahadur's sense) among all \hat{I}_φ -based tests, including the classical χ^2 -independence one.

Testing independence of finite random variables

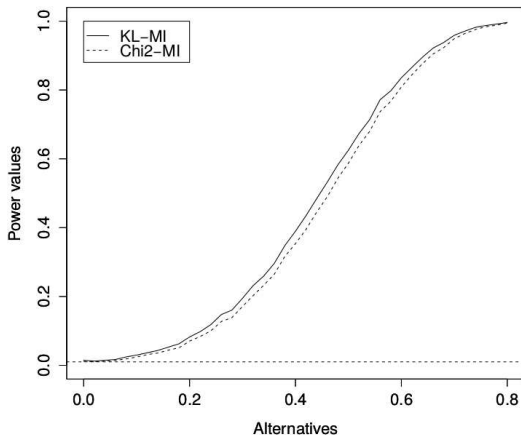
$$\mathcal{X} = \mathcal{Y} = \{0, 1\}.$$

Alternatives : $P_\theta = (p_{x,y;\theta})$
with

$$p_{x,y;\theta} := \frac{1-\theta}{4} + \frac{\theta}{2} \mathbb{1}_{\{y=x\}}.$$

Signif. level : 0.01.

Power of KL-MI and χ^2 tests estimated by Monte-Carlo from 10,000 samples of size 30.



Testing indep. of the components of a Gaussian couple

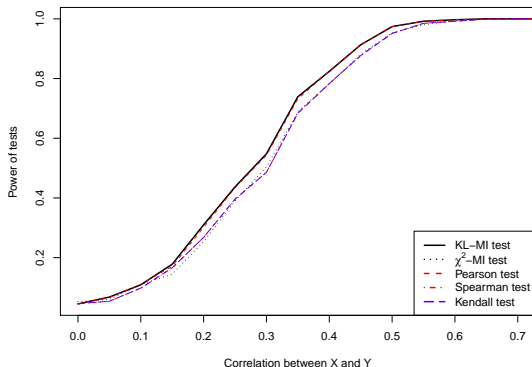
(X, Y) centred Gaussian couple with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

with $\rho \in [0, 1]$.

Signif. level : 0.05.

Power of KL-MI, χ^2 and non-correlation tests estimated by Monte-Carlo from 10,000 samples of size 50.



Bibliography

- ▶ Keziou, A., Regnault P. (2015). *Semiparametric estimation of mutual information and related criteria : optimal test of independence*. arXiv :1508.04671
- ▶ Pardo, L. (2006). *Statistical inference based on divergence measures*, volume 185 of *Statistics : Textbooks and Monographs*. Chapman & Hall/CRC, Boca Raton, FL.
- ▶ Khan, S., Bandyopadhyay, S., Ganguly, A.R., Saigal, S., Erickson, D.J., Protopopescu, V., Ostrouchov, G. (2007). Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. *Physical Review E*, 76, 026209.
- ▶ Nelsen, R.B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.
- ▶ Keziou, A. (2003). Dual representation of φ -divergences and applications. *C. R. Math. Acad. Sci. Paris*, 336(10), 857-862.
- ▶ Browniatowski, M, Keziou, A. (2006). Minimization of φ -divergences on sets of signed measures. *Studia Sci. Math. Hungar.*, 43(4), 403-442.
- ▶ Eichelsbacher, P., Schmock, U. (2002). Large deviations of U -empirical measures on strong topologies and applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5), 779-797.