Applied Mathematical Sciences, Vol. 5, 2011, no. 4, 181 - 192

Probability Metrics and their Applications

Hamidreza Mostafaei[1]

Department of Statistics, The Islamic Azad University
North Tehran Branch, Iran
hrmostafaei@yahoo.com

Shaghayegh Kordnourie

Department of Statistics
The Islamic Azad University
North Tehran Branch, Iran
shaghayeghkordnourie@yahoo.com

Abstract

In this article, we introduce the definitions and characteristics of some important probability metrics (distances); furthermore, we examine some applications of these metrics. The relationships among these metrics are evaluated. Finally, we study the convexity property of metrics and this property is investigated for the Birnbaum-Orlicz average distance.

Keywords: Sprobability metrics, Relative entropy ,Wasserstein metric, Birnbaum-Orlicz average distance, convexity

1 INTRODUCTION

A metric or a distance is a function that determines distances between points in a space .A functional distribution $d : \Omega \to R_+$ is called a probability metric if it satisfies:

1)Identity: $d(X, Y) \geq 0 \; and \; d(X, Y) = 0 \Leftrightarrow P(X = Y) = 1 \;\; \forall X, Y \in \Omega$

2)Symmetry:

$$d(X, Y) = d(Y, X) \;\; \forall X, Y \in \Omega \tag{1}$$

[1]Corresponding author: Department of statistics Islamic Azad University North Tehran Branch, Iran, Tehran, Ghods sq., Darband st., Faculty of Basic Sciences, Post Box:19585/936

3)Triangle inequality: $d(X,Y) \leq d(X,Z) + d(Z,Y) \ \ \forall X,Y,Z \in \Omega$
If a probability metric identifies the distribution,($p_x = p_y \iff dis(X,Y) = 0$),then the metric is called simple; Else, it is called compound. The ideal probability metrics should have two properties:
1)The homogeneity property of order $r \in R : d(kX,kY) = |k|^r d(X,Y) \ \forall X,Y \in \Omega$
2) The regularity property: $d(X+Z,Y+Z) \leq d(X,Y) \ \forall X,Y,Z \in \Omega$

Stoyanov and et al.[7],are studied the connections between the theory of probability metrics and financial economics, especially portfolio theory. Gibbs (2000) [5], gave precise bounds on the convergence time of Gibbs sampler used in the bayesian restoration of degraded images by the choice of metrics Gibbs and Su (2002) [4], illustrated the ways in which the choice of metric can affect the rates of convergence. Also, they gave several examples of random walks whose qualitive convergence behavior depends strongly on the metric chosen. In our article, first we introduce some simple and compound metrics (distances) and mention some applications of them. Second, we review the relationships among probability metrics, briefly. Finally, we study the convexity property of metrics and give an example for evaluating the convexity of the Birnbaum-Orlicz average distance.

2 METRICS ON PROBABILITY MEASURES

2.1 Wasserstein Metric (Kantorovich metric):

For probability measures μ, ν on a metric space, which can be R or any metric space, the Wasserstein metric is defined by:

$$l_p^p(\mu,\nu) = \int_{-\infty}^{\infty} |F(x) - G(x)| d_x = \int_0^1 |F^{-1}(t) - G^{-1}(t)| d_t \qquad p \geq 1 \qquad (2)$$

Where F, G are the distribution functions of μ, ν respectively and F^{-1}, G^{-1} are their inverses. It may be equivalently defined by:

$$l_p^p(\mu,\nu) = infE[|X-Y|^p] \ \ (i.e\ all\ random\ variables\ with\ distributions\ \mu,\nu) \qquad (3)$$

Ferns and et al.[3], showed that if integration reduces to summation, the Kantorovich metric applied to state distribution μ, ν by:

$$max_{u_i} \sum_{i-1}^{|s|} (\mu(s_i) - \nu(s_i)) u_i \ where\ \forall i,j\ u_i - u_j \leq d_w(s_i,s_j) \ , \forall 0 \leq u_i \leq 1 \quad (4)$$

Huber(1981)[6],(who named this metric as the Bounded Lipschitz metric),showed that:

$$l(\mu,\nu) = d_{BL}(\mu,\nu) = SUP|\int \Psi d_{\mu} - \int \Psi d_{\nu}| \tag{5}$$

Where function Φ ,satisfying the Lipschitz condition:

$$|\Phi(x) - \Phi(y)| \leq d(x,y) \tag{6}$$

The following two statements are equivalent: 1) $d_{BL}(\mu,\nu) \leq \epsilon$, 2)there are random variables with $\xi(x) = \mu$ and $\xi(y) = \nu$ such that $Ed(X,Y) \leq \epsilon$ (See Huber (1981) for proof). Chan and et al.[2]],used the Wasserstein distance to compare two normalized image histograms. The linear Wasserstein distance between two normalized histograms $p_a(y)$ and $p_b(y)$ was defined by:

$$l(p_a, p_b) = \int_0^1 |F_a(y) - F_b(y)| d(y) \tag{7}$$

Where $F_a(y)$ and $F_b(y)$ are the corresponding cumulative distributions of $p_a(y)$ and $p_b(y)$, respectively. Also, they suggested a new nonparametric region-based active contour model for segmenting clutter images based on the Wasserstein metrics in comparing histograms of different areas in the image. The Wasserstein model can be used to define the transformation cost as follow:

$$T_c(\mu,\nu) = \int_0^1 c(F^{-1}(t), G^{-1}(t)) d_t, \tag{8}$$

Where $c(X,Y)$ is a convex cost function and if it is linear, $c(X,Y) = |X - Y|$,the transformation cost is defined: $T_1(\mu,\nu) = \int_0^1 (F^{-1}(t) - G^{-1}(t)) d_t$, and with Fubini's theorem: $T(\mu,\nu) = \int_0^1 |F(t) - G(t)| d_t$, .
The markov decision processes (MDP) are the model of choice for decision making that provide a standard formalism for describing multi-stage decision making in probabilistic environments under uncertainty .Ferns [3] argued the methods for computing state similarity in MDP and investigated ways of obtaining useful metrics through efficient computation and approximation of the Kantorovich metric. Stoyanov and et al.[7] showed that for X and Y as the return distributions of two portfolios, the absolute difference between VARs(the values at risk) at any tail probability level ϵ ,can be defined as:$|VAR(X) - VAR(Y)| \leq l_{\infty}(X,Y)$.

2.2 Hellinger distance:

For probability measures μ, ν on a metric space:

$$d_H^2(\mu,\nu) = \frac{1}{2}\int(\sqrt{\frac{d_\mu}{d_\lambda}} - \sqrt{\frac{d_\nu}{d_\lambda}})^2 d_\lambda \tag{9}$$

Where $\frac{d_\mu}{d_\lambda}$ and $\frac{d_\nu}{d_\lambda}$ are the random-nikodym derivatives of μ, ν respectively.

$$d_H(\mu,\nu) = (\frac{1}{2}\int(\sqrt{f} - \sqrt{g})^2 d_\lambda)^{1/2}, \quad 0 \leq d_H(\mu,\nu) \leq 1, \tag{10}$$

Where f, g are densities of the measures μ, ν with respect to dominating measure λ (some authors omit the factor $\frac{1}{2}$). The Hellinger distance can be defined by the Bhattacharyya distance $BC(\mu,\nu)$ as follow:

$$d_H(\mu,\nu) = \{\frac{1}{2}(2(1 - BC(\mu,\nu)))\}^{1/2} \tag{11}$$

For countable state space:

$$d_H(\mu,\nu) = [\sum_{w\in\Omega}(\sqrt{\mu(w)} - \sqrt{\nu(w)})^2]^{1/2} \tag{12}$$

Su and White(2005)[8],suggested a nonparametric test of conditional independence based on the weighted Hellinger between the two conditional densities, $f(y|x,z)$ and $f(y|x)$, which is identically zero under the null.Let $f(.|.)$ be the conditional density of one random vector given another and x, y, z are $d1-, d2-, d3-$ vectors ,respectively.

$$H_0 : Pr\{f(y|x,z) = f(y|x)\} = 1 \ \ \forall\ y \in R^{d2} \tag{13}$$

$$H_1 : Pr\{f(y|x,z) = f(y|x)\} < 1 for some\ y \in R^{d2}$$

The test statistics is based on the weighted Hellinger distance between $f(x,y,z)f(x)$ and $f(x,y)f(x,z)$:

$$\Gamma(f,F) = \int\{1 - \sqrt{\frac{f(x,y)f(x,z)}{f(x,y,z)f(x)}}\}^2 a(x,y,z) d_{F(x,y,z)} \tag{14}$$

Where $a(.)$ is a nonnegative weighting function with support $A \subset R^d$ $(d = d1 + d2 + d3)$ .Also, Hellinger distance can be used by choosing $(a \equiv 1)$.

2.3 Relative Entropy (Kullback-leibler divergence; Information divergence)

If μ ,with support $\delta(\mu)$,is absolutely continuous with respect to ν ,then $\frac{d_\mu}{d_\nu} = \frac{f}{g}$ is the Radom nikodym derivative of μ with respect to ν ,then:

$$d_I(\mu,\nu) = \int_{\delta_\mu} log\frac{d_\mu}{d\nu} d_\mu = \int_{\delta_\mu} \frac{d_\mu}{d_\nu} log\frac{d\mu}{d\nu} d_\nu. \tag{15}$$

Which we can consider it as the entropy of μ relative to ν. For a countable space:

$$d_I(\mu,\nu) = \sum_{w\in\Omega} \mu(\nu) log\frac{\mu(w)}{\nu(w)}, \quad 0 \leq d_I(\mu,\nu) \leq \infty \tag{16}$$

We mention the applications of the Kullback L divergence in information theory:

1) The Shannon entropy:

$$H(X) = E_x I(X) = logN - d_I(p(X), P_U(x)) \tag{17}$$

Where $p_U(x)$ is the uniform distribution and $p(x)$ is the real distribution.

2) The Cross entropy:

$$H(p,q) = E_p[-logq] = H(P) + d_I(p,q) \tag{18}$$

Where p and q are two distributions.

3) The conditional Entropy:

$$\begin{aligned} H(X|Y) &= logN - d_I(p(X,Y), p_U(X)p(Y)) \qquad (19)\\ &= logN - E_y\{d_I(p(X|Y), p_U(X)\} \end{aligned}$$

4) The self information:

$$I(m) = d_I(\delta_{im}, p_i) \tag{20}$$

Where δ_{im} is a kronecker delta and, p_i is the probability distribution.

5) The mutual information:

$$I(X,Y) = E_y(d_I(p(X|Y), p(X)) \tag{21}$$

6) In the Bayesian statistics:

$$d_I(p(X|Y), p(X|I)) = \sum p(X|Y) log\frac{p(X|Y)}{p(X|I)} \tag{22}$$

The Kullback divergence in Bayesian statistics is used as a measure of information which is obtained by moving from a prior distribution to a posterior one.

2.4 Prokhorov (Levy-Prokhorov) metric:

The Prokhorov metric is on the collection of probability measures on a given metric space .For any subsets $A \subset \Omega$,the closed δ-neighborhood of A is defined as: $A^{\delta} = \{x \in \Omega | inf d(x, y) \leq \delta \; ; y \in A\}$ .For β as the Borel-σ -algebra and g_{η} as the space of all probability measures an (Ω, β) and μ, ν as the probability measures:

$$d_p(\mu, \nu) = inf\{\delta > 0 | \mu(A) \leq \nu(A^{\delta}) \; \forall A \in \beta\} \tag{23}$$

Huber (1981), proved that this metric is symmetric in μ, ν and it metrizes the weak topology in g_{η} .

2.5 Levy metric:

The Levy metric is a special case of metric Levy -Prokhorov metric and it is on the space of cumulative distribution functions of one dimensional random variables.

$$d_L(F, G) = inf\{\varepsilon : \forall x F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \;\; (0 \leq d_L(F, G) \leq 1)\} \tag{24}$$

It metrizes the weak topology (see Huber (1981) for proof).also, it can metrizes the weak convergence. As Huber(1981) showed, let the observations be independent, with common distribution F and let $T_n = T_n(x_1, x_2, ..., x_n)$ be a sequence of estimates or test statistics with values in R .This sequence is called **Robust** at $F = F_0$ if the sequence of maps of distributions $F \longrightarrow \zeta_F(T_n)$ is equicontinuous at F_0 ,that is ,if ,for every $\epsilon > 0$,there is a $\delta > 0$ and n_0 such that $\forall F, \forall n \geq n_0$:

$$d_*(F_0, F) = \leq \delta \Rightarrow d_*(\zeta_{F_0}(T_n), \zeta_F(T_n)) \leq \varepsilon \tag{25}$$

Where, d_* is any metric generating the weak topology. Huber(1981),worked on the Levy metric for F and the Prokhorov metric for $\zeta(T_n)$ .Suppose that $T_n = T(F_n)$,derives from a functional T which is defined on some weakly open subset of the space of all the probability measures; If T is weakly continuous at F ,then $\{T_n\}$ is consistent at F ,in the sense that $T_n \longrightarrow T(F)$ in probability and almost surely; Moreover, assume that $\{T_n\}$ is consistent in neighborhood of F_0 ,then T is continuous at F_0 iff $\{T_n\}$ is robust at F_0 .(See Huber(1981) for proof).

2.6 Kolmogrov (Uniform) metric:

Kolmogrov metric is a distance between distribution functions of probability measures.

$$d_K(F,G) = SUP|F(x) - G(x)| \quad x \in R, \quad (0 \leq d_K \leq 1) \tag{26}$$

As Huber (1981) showed, it doesn't generate the weak topology. The Kolmogrov metric is applied in central limit theorem in probability theory and it is completely insensitive to the tails of the distribution which describe the probabilities of extreme events. Boratynska and Zielinski (1993)[1],proposed an upper bound for the Kolmogrov distance between the posterior distributions in terms of that between the prior distributions is given. Let F be a given prior distribution and $x \in \Omega$ be a fixed point in the sample space. For every likelihood function $l_x(.)$ and $l_x^* = l_x^+(.) + l_x^-(.)$:

$$d_k(F_x, G_x) \leq \frac{d_k(F,G)}{max\{m_x(F), m_x(G)\}}(s(x) + u(x)) \tag{27}$$

Where F and G are the continuous distribution functions, $u(x) = \int d_{l^*}(\theta)$, Θ is our parametric space, F_x and G_x are the cdf of the corresponding posterior distributions ,$s(x) = sup_{\theta \in \Theta l_x(\theta)}$ and given a cdf H ,let $m_x(H) = \int l_x(\theta) dH(\theta)$. By this inequality the bayes robustness can be investigated (See Boratynska and Zielinski (1993)).

2.7 Total Variation:

The standard metric for measuring the distance between probability function is the total variation metric.

$$d_{TV}(\mu,\nu) = SUP|\mu(A) - \nu(A)| = \frac{1}{2} Max_{|h| \leq 1} |\int h(d_\mu - d_\nu)|, (0 \leq d_{TV}(\mu,\nu) \leq 1) \tag{28}$$

Where $h : \Omega \longrightarrow R$ and μ, ν are probability measures. For countable state space which is half the L^1 norm between the two measures:

$$d_{TV}(\mu,\nu) = \frac{1}{2} \sum_{s \in \Omega} |\mu(A) - \nu(A)| \tag{29}$$

Gibbs (2000)[5], showed that the Total variation is a strong measure of convergence especially in assessing convergence of MCMC algorithm. For a markov chain with probability transition matrix P, stationary distribution π and countable space $\aleph$,and initial configuration $x^0 \in \aleph$,the Total variation distance at time t is :

$$tvd_{x^0}(t) = d_{TV}(P^t(x^0,0),\pi(0)) = Sup_{A\subset\aleph}|P^t(x^0,A)-\pi(A)| = \frac{1}{2}\sum_{x\in\aleph}|P^t(x^0,x)-\pi(x)| \quad (30)$$

Where $P^t(x^0,x)$ is the probability that the markov chain with initial state x^0 is in state x at iteration t and A is any set. Also he defined the convergence time of the markov chain used by the Gibbs sampler as:

$$\tau(\varepsilon) = maxmin\{t : tvd_{x^0}(t') \leq \varepsilon \;\; \forall t' \geq t\} \quad (31)$$

Where ε is a pre-specified error tolerance.

2.8 χ^2 distance:

$$d_{\chi^2}(\mu,\nu) = \int_{s(\mu)\cup s(\nu)} \frac{(f-g)^2}{g} d_\lambda, \quad (0 \leq d_{\chi^2}(\mu,\nu) \leq \infty) \quad (32)$$

Where f, g are densities of measures μ, ν with respect to a dominating measures λ and $s(\mu), s(\nu)$ are their supports on Ω .For a countable space Ω :

$$d_{\chi^2}(\mu,\nu) = \sum_{w\in s(\mu)\cup s(\nu)} \frac{(\mu(w)-\nu(w))^2}{\nu(w)} \quad (33)$$

2.9 Separation distance:

$$d_s(\mu,\nu) = max(1 - \frac{\mu(i)}{\nu(i)}) \quad (34)$$

Where μ, ν are probability measures and the state space is a countable space.

2.10 Discrepancy metric:

$$d_D(\mu,\nu) = Sup|\mu(B) - \nu(B)|, \quad (B \textit{ is all closed balls}) \quad (35)$$

This metric is scale-invariant.

2.11 The Birnbaum-Orlicz average distance:

$$\theta_H(X,Y) = \int_{\Re} H(|F_X(x) - F_Y(x)|) d_x \quad H \in \mathbf{H} \quad (36)$$

Where X and Y are the random variables that $E|X| < \infty$, $E|Y| < \infty$ and all functions satisfying Orlicz's condition are denoted by **H**. By choosing $H(t) = t^p$, $p \geq 1$:

$$\theta_p(X,Y) = (\int_{-\infty}^{\infty} |F_X(t) - F_Y(t)|^p d_t)^{\frac{1}{min(1,\frac{1}{p})}} \qquad p > 0 \tag{37}$$

If $p \to \infty$ then the Kolmogrov metric is obtained.

2.12 The p-average compound metric:

It is a compound metric and defined as:

$$\zeta_p(X,Y) = \sqrt[p]{E|X-Y|^p} \qquad p \geq 1 \tag{38}$$

Where X and Y are r.v.s with $E|X|^p < \infty and E|Y|^p < \infty$.

2.13 The Ky Fan metric:

It is a compound metric and defined as:

$$K(X,Y) = inf\{\varepsilon > 0 : P(|X-Y| > \varepsilon) < \varepsilon\} \tag{39}$$

Where X and Y are real-valued r.v.s .It also metrizes convergence in probability of real-valued random variables.

2.14 The Birnbaum -Orlicz compound metric:

$$\Theta_p(X,Y) = [\int_{-\infty}^{\infty} \tau^p(t;x,y)d_t]^{1/p} \tag{40}$$

Where $\tau(t;x,y) = P(X \leq t < Y) + P(Y \leq t < X)$.

3 PROBABILITY METRICS AND RELATIONSHIPS AMONG THEM

We summarize the relationships between metrics (distances) as follow:

$d_L(F,G) \leq d_K(F,G)$	$d_D(\mu,\nu) \leq x + \varphi(x)$ $where\ x = d_p(\mu,\nu)$
$d_K(f,g) \leq (1 + SUP\lvert G'(x)\rvert d_L^{(F,G)})$ $If\ G(x)\ is\ absolutely\ continuous.$	$(d_p)^2 \leq d_w \leq (diam(\Omega)+1)d_p$ $where\ diam(\Omega) = SUP\{d(X,Y) : X, Y \in \Omega\}$
$d_K \leq d_D \leq 2d_K$	$d_{min}.d_D \leq d_w$ $where\ d_{min} = min_{X\neq Y} d(X,Y)$
$d_L \leq d_P$	$d_D \leq d_{TV}$
$(d_P)^2 \leq d_w \leq 2d_p$	$d_P \leq d_{TV}$
$d_w \leq diam(\Omega).d_{TV}$ $Where\ diam(\Omega)\ = Sup\{d(X,Y) : X, Y \in \Omega\}$	$d_H(\mu,\nu) \leq \sqrt{2}(d_{\chi^2(\mu,\nu)})^{1/4}$
$d_{min}.d_{TV} \leq d_w\ where\ d_{min} = min_{X\neq Y} d(X,Y)$	$(d_H)^2 \leq d_I$
$\frac{(d_H)^2}{2} \leq d_{TV} \leq d_H$	$d_H(\mu,\nu) \leq (d_{\chi^2}(\mu,\nu))^{1/2}$
$d_{TV} \leq d_S$	$d_{TV}(\mu,\nu) \leq \frac{1}{2}\sqrt{d_{\chi^2}(\mu,\nu)}$
$2(d_{TV})^2 \leq d_I$	$d_I(\mu,\nu) \leq log[1 + d_{\chi^2}(\mu,\nu)]$ $specifically\ d_I(\mu,\nu) \leq d_{\chi^2}(\mu,\nu)$

Table 1: Relationships among metrics (distances) (see proofs in Huber (1981) and Gibbs and Su (2002))

4 Convexity

For a given compound metric μ ,the function $\hat{\mu}$ defined by:

$$\hat{\mu}(X,Y) = inf\{\mu(\widetilde{X},\widetilde{Y}) : \widetilde{X} =^d X, \widetilde{Y} =^d Y)\} \quad (41)$$

is said to be a minimal metric with respect to μ . The convexity property is defined as : $\mu(aX + (1-a)Y, Z) \leq a\mu(X,Z) + (1-a)\mu(Y,Z)$; As Stoyanov and et al.[7] showed, in addition to the Cambanis-Simons-Stout theorem, there is another method of obtaining forms of minimal and maximal functions. By applying the Frechet-Hoeffding inequality between distribution functions, we have:

$$max(F_X(x) + F_Y(y) - 1, 0) \leq P(X \leq x, Y \leq y) \leq min(F_X(x), F_Y(y)) \quad (42)$$

They also obtained:

$$1)\theta_p^*(X,Y) = [\int_{-\infty}^{\infty}(max(F_Y(t) - F_X(t), 0))^p d_t]^{\frac{1}{p}} \qquad p \geq 1 \quad (43)$$

Where X and Y are zero-mean r.v.s.

$$2)\theta_p^*(X,Y) = \hat{\Theta}_p^*(x,y) \quad (44)$$

Where $\hat{\Theta}_p^*(X,Y)$ is the asymmetric version of Birnbaum-Orlicz metric.We give an example on the convexity of r.d metrics. It is based on the functional $\mu(X,Y) = \Theta_2(X,Y)$ .We show that the minimal metric $\theta_2(X,Y) = \hat{\Theta}_2(X,Y)$ doesn't satisfy the convexity property. Suppose that $X \in N(0,\sigma_X^2), Y \in N(0,\sigma_Y^2), Z \in N(o,\sigma_Z^2)$,then:

$$\begin{aligned}
\theta_2(X,Y) &= (\int_{-\infty}^{\infty} |F_X(t) - F_Y(t)|^2 d_t)^{\frac{1}{2}} \\
&= (\int_{-\infty}^{\infty} |\sigma_X F_X(t) - \sigma_Y F_Y(t)|^2 d_t)^{\frac{1}{2}} \\
&= (\int_{-\infty}^{\infty} |(\sigma_X - \sigma_Y) F(t)| d_t)^{\frac{1}{2}} \\
&= |\sigma_x - \sigma_Y| (\int_{-\infty}^{\infty} (F(t))^2 d_t)^{\frac{1}{2}} \\
&= \infty \qquad (45)
\end{aligned}$$

Where F is the cumulative distribution function of the standard normal distribution. This means that the Birnbaum-Orlicz metric doesn't have a convexity property for p=2; Moreover, for p=1, there is a lack of convexity for Birnbaum-Orlicz metric through the Kantorovich metric since they are coincided.

5 Conclusions

In this paper, we provided a summary of some important simple and compound metrics; Furthermore, we discussed the applications of probability metrics in several fields. The relationships among probability metrics were evaluated. Finally we showed that the Birnbaum-Orlicz metric doesn't have a convexity property.

References

[1] Bortynska A. and Zielinki R.,, Bayes robustness via the Kolmogrov metric, Application Mathematicae,22(1993),pp.139-143 .

[2] Chan T.,Esedoglu S. ,Ni k. ,Histogram based segmentation using Wasserstein distances.

[3] Ferns N., Castro p., Precup D. ,Panangaden P.,Methods for computing state similarity in markov decision processes.

[4] Gibbs A.L. and Su.F.E,On choosing and bounding probability metrics,(2002).

[5] Gibbs A.L. ,Convergence of markov chain monte carlo algorithms with applications to image restoration, PhD thesis, university Toronto(2000).

[6] Huber P.J., Robust statistics, John Wiley and Sons, New York(1981).

[7] Stoyanov, S., S. Rachev, S. Ortobelli and F. Fabozzi , 'Relative deviation metrics with applications in finance', Department of Probability and Applied Statistics, University of California, Santa Barbara, USA Technical report (2006).

[8] Su L., White H., A nonparametric Hellinger Metric test for conditional Independence(2005).

Received: May, 2009