Economics Bulletin

Volume 36, Issue 2

Density estimation based on pointwise mutual information.

Akimitsu Inoue
Center of Advanced Research and Education, Graduate School of Business, Osaka City University

Abstract

The purpose of this article is to develop a new bivariate density estimation method based on the decomposition of joint density into pointwise mutual information and marginal densities. The pointwise mutual information and product of marginal densities are estimated by bivariate kernel density estimators with shuffled data. Our method is defined as a product of the marginal densities and pointwise mutual information. Monte-Carlo simulations indicate that this estimation method provides good finite sample performance for weak dependent data.

Citation: Akimitsu Inoue, (2016) "Density estimation based on pointwise mutual information.", *Economics Bulletin*, Volume 36, Issue 2, pages 1138-1148
Contact: Akimitsu Inoue - akimitsu.inoue@gmail.com.
Submitted: March 08, 2015. **Published:** June 11, 2016.

1. Introduction

Recently, the analysis of nonlinear dependence in multivariate data has become important. Because nonlinear dependence cannot be detected using a correlation coefficient, other measures have been proposed (Dionisio *et al.* 2004, and Tjøstheim and Hufthammer 2013). A previous study has used pointwise mutual information as a method for measuring nonlinear dependence (Takada 2012). Because pointwise mutual information is uniquely determined by the joint density function, a method for estimating pointwise mutual information indirectly through density estimation has been proposed (Moon *et al*. 1995, and Takada 2012). In contrast with this previous approach, we propose a method for estimating the joint density from an estimate of pointwise mutual information. Pointwise mutual information is a specific type of density ratio. Previous studies have proposed several techniques for density ratio estimation (Kelsall and Diggle 1995, Hazelton and Davies 2009, Davies *et al*. 2011 and Fernando *et al*. 2014). The error of the density ratio estimator can be decreased by using these techniques when the density ratio function is flat (Kelsall and Diggle 1995). Therefore, our approach may estimate the density function more efficiently than does the previous method when the shape of the target density function is complex and the shape of the corresponding pointwise mutual information function is relatively simple. However, our approach has not yet been studied.

The purpose of the present study is to develop a new bivariate density estimation method based on the decomposition of joint density into pointwise mutual information and marginal densities.

The proposed method is calculated as follows. First, pointwise mutual information is estimated by two bivariate kernel density estimators with shuffled data. Second, the marginal densities are estimated by two univariate kernel density estimators or one bivariate kernel density estimator using shuffled data. The proposed method is thus defined as a product of the marginal densities and pointwise mutual information. Additionally, this modified estimation method may not yield extremely high error in specific cases.

Monte Carlo simulations indicate that this method provides good finite sample performance for weak dependent data. Therefore the proposed method, is efficient for the analysis of noisy, weak dependent bivariate data.

The remainder of this paper is organized as follows. Section 2 reviews the existing methods for estimating pointwise mutual information. Because pointwise mutual information is defined as the ratio of two bivariate density functions, this section also reviews the conventional techniques for density ratio estimation. Section 3 describes our proposed approach. Section 4 compares the performance of the proposed method with that of ordinal kernel density estimation. Finally, section 5 provides some concluding remarks.

2. Pointwise Mutual Information

Our approach is based on the concept of pointwise mutual information. In this section, we review the definition of pointwise mutual information and the existing methods for estimating it.

Let X and Y be random variables of a probability space. The pointwise mutual information of X, Y at (X, Y) is defined as follows:

$$PMI(x,y) = \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)}, \tag{1}$$

where $f_{X,Y}(x,y)$ is the joint density function of X and Y at the point (x,y), and $f_X(x)$ and $f_Y(y)$ are the marginal densities of X at x and Y at y, respectively. Pointwise mutual information is a method for measuring the dependence of variables. This method is used for collocation extraction in natural language processing (Bouma 2009), as well as for tail dependence analysis of multivariate time-series data in finance (Takada 2012).

Several methods for estimating the pointwise mutual information of continuous random variables have been considered. Takada (2012) compared three estimators of pointwise mutual information: the products of the univariate density estimate (UD), marginal density estimate (MD), and bivariate density estimate in shuffled sequence (SD). According to Takada (2012), the UD and MD are both natural estimators directly following from the definition of pointwise mutual information. The UD method is described below. First, the joint density $f_{X,Y}(x,y)$ is estimated by a bivariate kernel density estimator. Second, the marginal densities $f_X(x)$, $f_Y(y)$ are estimated by univariate density estimators. Third, the ratio between these densities is computed. Hence, the UD estimator is defined as follows:

$$\widehat{PMI}_{UD}(x,y) = \log \frac{\hat{f}_{X,Y}(x,y)}{\hat{f}_X(x) \hat{f}_Y(y)}. \tag{2}$$

The MD method uses a numerical integration of the joint density estimate $\hat{f}_{X,Y}(x,y)$ to obtain the marginal densities $\hat{f}_X(x) = \int \hat{f}_{X,Y}(x,y)\, dx$ and $\hat{f}_Y(y) = \int \hat{f}_{X,Y}(x,y)\, dx$. Therefore, the MD method is defined as follows:

$$\widehat{PMI}_{MD}(x,y) = \log \frac{\hat{f}_{X,Y}(x,y)}{\int \hat{f}_{X,Y}(x,y)\, dy \int \hat{f}_{X,Y}(x,y) dx}. \tag{3}$$

UD has been indirectly used for average mutual information estimation (Moon *et al.* 1995). However, when estimating densities for which two random variables are independent, these methods tend to produce biases which mistakenly detect nonexistent dependence (Takada 2012). The SD method, has been proposed as an alternative to UD and MD (Takada, 2012). This method shuffles the sequence of data to avoid bias. The SD method replaces the product of marginal densities $\hat{f}_X(x)\hat{f}_Y(y)$ with $\hat{f}_{X_{shuffle},Y}(x,y)$, where $\hat{f}_{X_{shuffle},Y}(x,y)$ is obtained by randomly shuffling the sequence of the X_i to make $X_{shuffle}, Y$ be independent. Hence, the SD estimator is defined as follows:

$$\widehat{PMI}_{SD} = \frac{\hat{f}_{X,Y}(x,y)}{\hat{f}_{X_{shuffle},Y}(x,y)}. \tag{4}$$

The SD method is based on the definition of independence. A previous study has reported that this method can avoid bias when the shuffled density is estimated by an adaptive kernel density estimator.

3. Pointwise mutual information estimation as a density ratio estimation

In this section, we review the previous research on density ratio estimation. Density ratio estimation has previously been used for the estimation of relative risk in geographical epidemiology or spatial statistics (Silverman 1978, Kelsall and Diggle 1995, Kelsall and Diggle 1998, Clark and Lawson 2004, Hazelton and Davies 2009, Davies *et al.* 2011, and Fernando *et al*. 2014). The estimation of pointwise mutual information is a type of (logarithm of) density ratio estimation. Therefore, we can apply several techniques from this field to pointwise mutual information estimation.

The literature we review here concerns two bandwidth selection approaches for density ratio estimation, especially as they relate to our proposed estimator. The first approach is the cross-validation approach, and the second approach involves using a common bandwidth for the numerator and denominator. These approaches can be applied simultaneously.

Kelsall and Diggle (1995) proposed a bandwidth selection method using least-square cross-validation with the Taylor series expansion of the logarithm of density ratio estimation. Kelsall and Diggle (1998) then examined other types of cross-validation methods: likelihood cross-validation, least-square cross-validation, and weighted least-square cross-validation. Kelsall and Diggle (1998) reported that likelihood cross-validation is the best approach.

Using a common bandwidth for the numerator and denominator of the density ratio helps decrease variance (Kelsall and Diggle 1995). When the shapes of the densities of the numerator and denominator are similar and a common bandwidth is used, the biases of the estimators are cancelled out. Therefore, a larger bandwidth can be used to decrease the variance of the density ratio estimator.

However, these results have not yet been applied for pointwise mutual information estimation. Because pointwise mutual information estimation is a variant of density ratio estimation, we apply these findings to our proposed estimation method.

4. Proposed Estimation Method

Our proposed method is based on the following decomposition formula. The bivariate density function is decomposed into three terms: two marginal density functions and an exponential of the pointwise mutual information function.

$$f(x,y) = f(x)f(y)\exp\big(PMI(x,y)\big). \tag{5}$$

These terms are estimated separately and merged as a product, which may theoretically be used to construct an estimation method based on Eq. (5). However, this estimation method tends to be unstable in practice. When the target joint density is dissimilar to the product of the marginal densities, the consequential ratio estimate tends to be unstable, because the denominator part of the pointwise mutual information tends to be very small. This problem was observed in a simple numerical experiment.

To avoid this issue, we modify Eq. (5) to set an upper bound for its ratio term. The joint density of X and Y can be expressed as follows:

$$f_{X,Y}(x,y) = \alpha^{-1}\left[\left\{\alpha f_X(x)f_Y(y) + (1-\alpha)f_{X,Y}(x,y)\right\}\frac{f_{X,Y}(x,y)}{\alpha f_X(x)f_Y(y) + (1-\alpha)f_{X,Y}(x,y)} - (1-\alpha)f_{X,Y}(x,y)\right], \tag{6}$$

where α is a parameter that takes a value between 0 and 1. Based on Eq. (6), we obtained the proposed estimator, defined as follows:

$$\tilde{f}_{X,Y} = \hat{g}_{X,Y}(x,y)\left\{\widehat{f_Xf_Y}(x,y) + \hat{f}_{X,Y_{h_1}}(x,y)\right\} - \hat{f}_{X,Y_{h_1}}(x,y). \tag{7}$$

where $\hat{g}_{X,Y}(x,y)$ is the density ratio estimate of

$$\frac{f_{X,Y}(x,y)}{\alpha f_X(x)f_Y(y) + (1-\alpha)f_{X,Y}(x,y)}, \tag{8}$$

The input data for $\hat{g}_{X,Y}$ is the original data X, Y for the numerator. For the denominator, if $\alpha = 1/2$, a combination of the original data and the shuffled data $(X, X_{shuffle}),\ (Y, Y)$ may be used.

Two types of estimator for the term $\widehat{f_Xf_Y}(x,y)$ are available. The first is the "product method," defined as follows:

$$\widehat{f_Xf_Y}(x,y) = \hat{f}_X(x)f_Y(y). \tag{9}$$

The second is the "shuffle method," defined as follows:

$$\widehat{f_Xf_Y}(x,y) = \hat{f}_{X_{shuffle},Y}(x,y)\,. \tag{10}$$

The proposed estimator is defined as follows:

$$\tilde{f}_{X,Y}(x,y) = \alpha^{-1}\left[\left\{\alpha\widehat{f_Xf_Y}(x,y) + (1-\alpha)f_{X,Y_{h_1}}(x,y)\right\}g_{X,Y_{h_2}}(x,y) - (1-\alpha)f_{X,Y_{h_1}}(x,y)\right]. \tag{11}$$

The term $f_{X,Y_{h_1}}$ is computed by bivariate density estimation using the original data. The smoothing parameter for this estimator is selected by the direct plugin method discussed by Sheather and Jones [1991].

Three types bandwidth selection method are employed for $g_{X,Y}(x,y)$: LikCV, LSCV, and WLSCV. The loss functions are defined as follows:

$$CV_{Lik} = \frac{1}{n}\sum_{i=1}^{n} \log \hat{p}_{nu,i} - \frac{1}{2n}\sum_{i=1}^{2n} \log \hat{p}_{de,i}, \tag{12}$$

$$CV_{LS} = \frac{1}{n}\sum_{i=1}^{n} \left(1 - \hat{p}_{nu,i}\right)^2 - \frac{1}{2n}\sum_{i=1}^{2n} \hat{p}_{de,i}^2, \tag{13}$$

$$CV_{WLS} = \frac{1}{n}\sum_{i=1}^{n} \frac{\left(1 - \hat{p}_{nu,i}\right)^2}{\hat{p}_{nu,i}(1 - \hat{p}_{nu,i})} - \frac{1}{2n}\sum_{i=1}^{2n} \frac{\hat{p}_{de,i}^2}{\hat{p}_{de,i}\left(1 - \hat{p}_{de,i}\right)}, \tag{14}$$

where $\hat{p}_{nu,i}$ is the cross-validated estimate of the probability that the i-th sample was generated from the density of the numerator side, defined as follows:

$$\hat{p}_{nu,i} = \frac{\hat{f}_{X,Y_{-i}}(X_i, Y_i)}{\hat{f}_{X',Y'}(X_i, Y_i) + \hat{f}_{X,Y_{-i}}(X_i, Y_i)}. \tag{15}$$

For the denominator, the estimate is defined as follows:

$$\hat{p}_{de,i} = \frac{\hat{f}_{X',Y'}(X_i', Y_i')}{\hat{f}_{X'_{-i},Y'_{-i}}(X_i', Y_i') + \hat{f}_{X',Y'}(X_i', Y_i')}, \tag{16}$$

where X' and Y' are defined as follows:

$$X' = \left(X, X_{shuffle}\right), \;\; Y' = (Y, Y). \tag{17}$$

As previously noted, we use a common bandwidth for the numerator and denominator in the $g_{X,Y}(x,y)$ parts.

We can select the degree of freedom of the bandwidth matrix using some bandwidth selection method for bivariate density estimation. In this paper, we use the diagonal bandwidth matrix, which is defined as follows:

$$h_X = h_{SJ_X}\lambda_{CV}, \tag{18}$$

$$h_Y = h_{SJ_Y}\lambda_{CV}.$$

where h_{SJ_X} and h_{SJ_Y} are the bandwidths selected by the Sheather and Jones [1991] direct plugin method using the sample X, Y.

5. Simulation Study

5.1 Data

Randomly generated data are used for Monte Carlo simulation. We adopt mixed normal distributions containing one to four elements. The distributions are defined as follows:

$$\sum_{k=1}^{m} p_k \phi(\mu_{x,k}, \mu_{y,k}, \sigma^2). \tag{19}$$

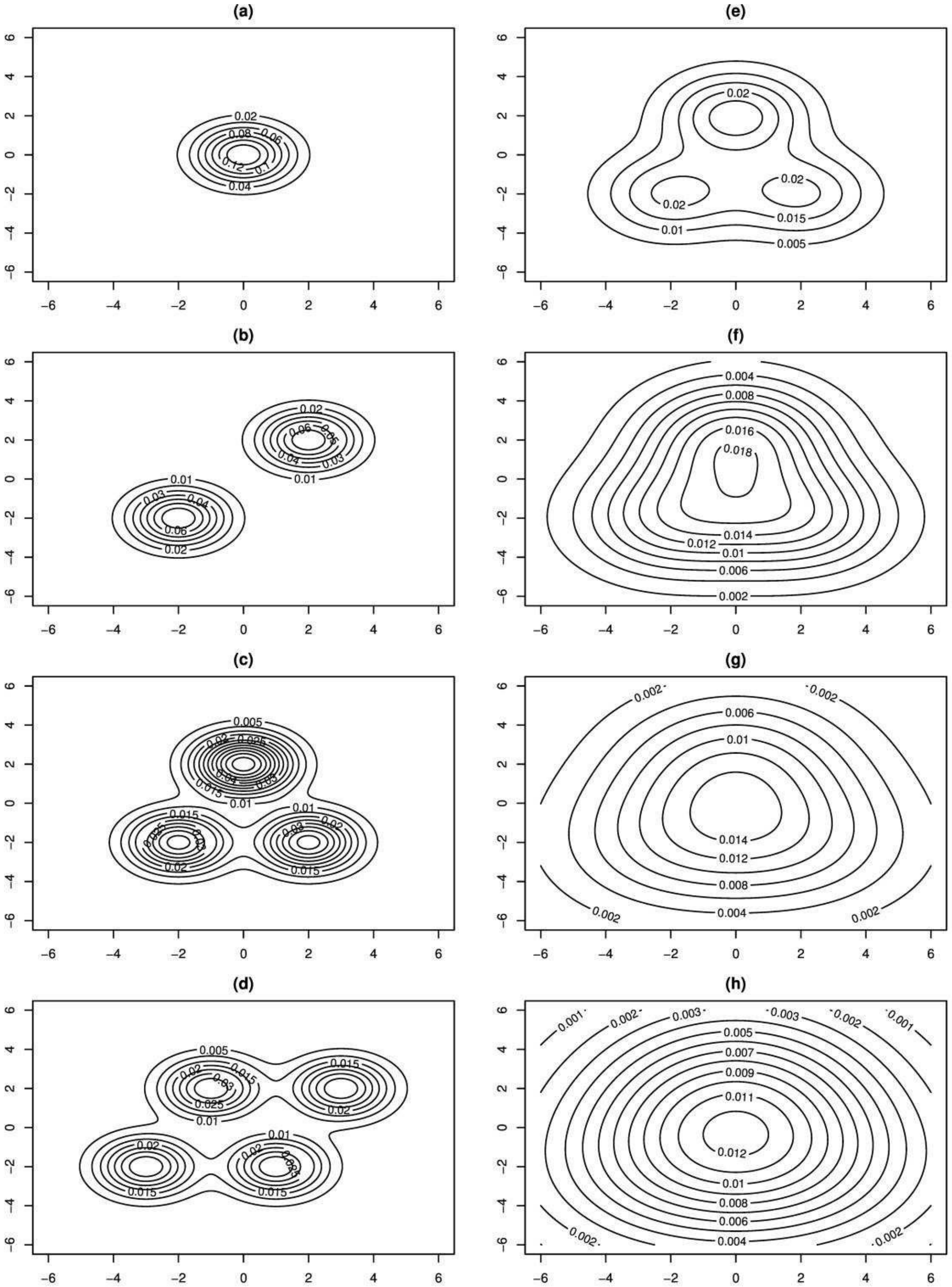
where $\phi(\mu_x, \mu_y, \sigma^2)$ is the bivariate normal density function. The location parameters and mixing ratios of the distributions used here are shown in Table 1.

The parameters for adjusting the degree of overlap between the elements are selected as follows. When the number of elements is one, two, or four, σ is set to 1.00. In addition, when the number of elements

Table 1. Location parameters and mixture ratios of each factor of the target mixture densities

m	$\mu_{x,1}$	$\mu_{x,2}$	$\mu_{x,3}$	$\mu_{x,4}$	$\mu_{y,1}$	$\mu_{y,2}$	$\mu_{y,3}$	$\mu_{y,4}$	σ	p_1	p_2	p_3	p_4
1	0				0				1.00	1.0			
2	2	-2			2	-2			1.00	0.5	0.5		
3	2	-2	0		-2	-2	2		1.00	0.3	0.3	0.4	
4	-3	-1	1	3	-2	2	-2	2	1.00	0.25	0.25	0.25	0.25
3	2	-2	0		-2	-2	2		1.25	0.3	0.3	0.4	
3	2	-2	0		-2	-2	2		1.50	0.3	0.3	0.4	
3	2	-2	0		-2	-2	2		1.75	0.3	0.3	0.4	
3	2	-2	0		-2	-2	2		2.00	0.3	0.3	0.4	
3	2	-2	0		-2	-2	2		2.25	0.3	0.3	0.4	
3	2	-2	0		-2	-2	2		2.50	0.3	0.3	0.4	
3	2	-2	0		-2	-2	2		2.75	0.3	0.3	0.4	
3	2	-2	0		-2	-2	2		3.00	0.3	0.3	0.4	

Figure 1. Shapes of target densities



is three, σ varies every 0.25 from 1.00 to 3.00. The shapes of the distributions determined by these parameter settings are shown in Fig. 1. Figs. 1(a–d) correspond to the four distributions of the different numbers of mixtures. In addition, Figs. 1(e–h) correspond to the four distributions of $\sigma = 1.5, 2.0, 2.5,$ and 3.0 with three elements. For each distribution setting, the experiment were ran with the sample sizes at 50, 100, 200, 400, 800, 1600, and 3200. For every scenario, the simulation was repeated 100 times.

5.2 Method

Six variations of the proposed estimation method and one existing estimation method were examined in the experiment. The bivariate fixed kernel (FK) density estimation method was used as the existing method, and the product method and shuffle method were used as the proposed methods. Three methods for bandwidth selection were applied to each estimation method: likelihood cross-validation, least-square cross-validation, and weighted least-square cross-validation.

The errors of the methods are calculated as MISE (mean integrated squared error).

5.3 Results and Discussion

The resulting data are shown in Tables 2, 3 and 4. First, we compared our six proposed estimation methods. At a sample size of n = 50, the product methods were better than the shuffle methods at a sample size of n = 3200, and the shuffled least-square cross-validation and shuffled weighted least-square cross-validation methods were worse than other methods. Therefore, the shuffle methods performed worse than did the product methods. One possible explanation is that the randomization of the shuffle method increased the variance of the estimation methods. In summary, the product likelihood cross-validation method provided stable estimates under a wide range of conditions.

Second, we compared the existing method and the product likelihood method at different target densities. As shown in Table 4, the MISE of the proposed estimation method displayed no clear dependence on the number of modes of the target densities. Table 4 illustrates the performance of the product likelihood method relative to that of the existing method under varying average mutual information of the two random variables of the target densities.

Table 2. The MISE of One Existing Method and Our Six Proposed Methods.

N = 50			**Product Method**			**Shuffle Method**		
m	**σ**	**FK**	**LiKCV**	**LSCV**	**WLSCV**	**LiKCV**	**LSCV**	**WLSCV**
1	1	0.009156	0.006182	0.006182	0.007017	0.009318	0.009318	0.009196
2	1	0.005926	0.007133	0.007506	0.018301	0.006055	0.008499	0.02111
3	1	0.005504	0.009235	0.009235	0.009118	0.010653	0.010653	0.010495
4	1	0.004227	0.005914	0.005909	0.005872	0.006781	0.006842	0.006777
3	1.25	0.003367	0.004335	0.004335	0.004298	0.005186	0.005186	0.00507
3	1.5	0.002332	0.002419	0.002419	0.002427	0.002989	0.002989	0.002896
3	1.75	0.001722	0.001526	0.001526	0.001597	0.002118	0.002118	0.002064
3	2	0.001287	0.001067	0.001067	0.00116	0.001422	0.001422	0.001419
3	2.25	0.001177	0.00087	0.00087	0.000991	0.001279	0.001279	0.001271
3	2.5	0.001002	0.000724	0.000724	0.00079	0.001001	0.001001	0.000991
3	2.75	0.000903	0.000635	0.000635	0.00072	0.000902	0.000902	0.000898
3	3	0.000677	0.000484	0.000484	0.000536	0.000711	0.000711	0.000697

The proposed method estimates more efficiently than do conventional estimators when the average mutual information of the target density is low. The MISE of the proposed method is lower than that of the conventional methods when the average mutual information of the target density is lower than approximately 0.15 bits and the sample size is 50, as well as when the average mutual information of the target density is lower than approximately 0.08 bit and the sample size is 3200. When the average mutual information and sample size are small, the relative efficiency of the proposed method is high. One possible explanation for the effect of average mutual information is that the larger bandwidth for $g(x, y)$ decreases the variance when the average mutual information of $g(x, y)$ is low.

In contrast to the previous method, the proposed method does not exhibit nonnegativity. Moreover, the proposed method does not integrate to one. This weakness should be addressed in future research.

Table 3. The MISE of One Existing Method and Our Six Proposed Methods.

N = 3200			Product Method			Shuffle Method		
m	**σ**	**FK**	**LiKCV**	**LSCV**	**WLSCV**	**LiKCV**	**LSCV**	**WLSCV**
1	1	0.000599	0.000254	0.000254	0.000564	0.000614	0.000614	0.000614
2	1	0.000457	0.000876	0.000459	0.001799	0.000515	0.000567	0.001814
3	1	0.000943	0.001201	0.001452	0.004352	0.000971	0.001502	0.004453
4	1	0.000351	0.000551	0.000803	0.002784	0.000367	0.000892	0.003036
3	1.25	0.000578	0.000698	0.000972	0.001902	0.000664	0.00105	0.001973
3	1.5	0.000371	0.000478	0.000635	0.000908	0.000513	0.000703	0.000952
3	1.75	0.000258	0.000361	0.000441	0.000486	0.000404	0.000493	0.000513
3	2	0.000187	0.000282	0.000302	0.000282	0.000333	0.000357	0.000293
3	2.25	0.00014	0.000182	0.000184	0.00018	0.000224	0.000227	0.000193
3	2.5	0.000108	0.000119	0.000119	0.000125	0.000156	0.000156	0.000135
3	2.75	0.000089	0.000083	0.000083	0.000097	0.000114	0.000114	0.000105
3	3	0.000072	0.00006	0.00006	0.000077	0.000086	0.000086	0.00008

Table 4. Relative MISE for Product Method using LikCV to Fixed Kernel Method.

		Average Mutual	**Sample Size**						
m	σ	**Information**	**50**	**100**	**200**	**400**	**800**	**1600**	**3200**
1	1.00	0.00	**0.68**	**0.65**	**0.61**	**0.54**	**0.50**	**0.47**	**0.42**
3	3.00	0.04	**0.72**	**0.70**	**0.69**	**0.71**	**0.72**	**0.77**	**0.84**
3	2.75	0.05	**0.70**	**0.71**	**0.70**	**0.74**	**0.79**	**0.84**	**0.93**
3	2.50	0.06	**0.72**	**0.76**	**0.77**	**0.79**	**0.87**	**0.95**	1.10
3	2.25	0.08	**0.74**	**0.74**	**0.85**	**0.91**	1.04	1.19	1.30
3	2.00	0.11	**0.83**	**0.85**	**0.99**	1.09	1.30	1.49	1.50
3	1.75	0.16	**0.89**	1.01	1.17	1.43	1.69	1.80	1.40
3	1.50	0.25	1.04	1.26	1.56	1.84	2.12	1.61	1.29
3	1.25	0.42	1.29	1.66	2.03	2.31	1.79	1.38	1.21
4	1.00	0.63	1.40	1.80	2.30	1.63	1.40	1.50	1.57
3	1.00	0.77	1.68	2.12	2.51	1.74	1.39	1.34	1.27
2	1.00	0.84	1.20	1.60	1.52	1.73	1.83	1.82	1.91

6. Conclusion

We proposed a method for estimating joint density from estimates of pointwise mutual information. We evaluated the proposed estimator through Monte Carlo simulation. Our experimental results showed that the proposed method was superior in efficiency than the previous method when the average mutual information of target density was low. The proposed method is thus useful for the analysis of noisy, weak dependent bivariate data.

References

Bouma G. (2009) “Normalized (pointwise) mutual information in collocation extraction” *Proceedings of the Biennial GSCL Conference*, 31-40.

Clark, A. B. and A. B. Lawson (2004) “An evaluation of non-parametric relative risk estimators for disease maps” *Computational Statistics & Data Analysis*, **47**(1), 63-78.

Davies, T. M., Hazelton M. L. and J. C. Marshall (2011) “Sparr: analyzing spatial relative risk using fixed and adaptive kernel density estimation in R” *Journal of Statistical Software*, **39**(i01).

Dionísio A, Menezes R and D. A. Mendes (2004) “Mutual Information: a measure of dependency for nonlinear time series” *Physica A*. 344, 326–329.

Fernando, W., Ganesalingam S. and M. L. Hazelton (2014) “A comparison of estimators of the geographical relative risk function” *Journal of Statistical Computation and Simulation*, **84**(7), 1-15.

Hazelton, M. L. and T. M. Davies (2009) “Inference based on kernel estimates of the relative risk function in geographical epidemiology” *Biometrical Journal*, **51**(1), 98-109.

Kelsall J. E. and P. J. Diggle (1995) “Kernel estimation of relative risk” *Bernoulli*, 1 (1-2), 3-16.

Kelsall J. E. and P. J. Diggle (1998) “Spatial variation in risk of disease: a nonparametric binary regression approach” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **47**(4), 559-573.

Moon Y. I., Rajagopalan B. and U. Lall (1995) “Estimation of mutual information using kernel density estimators” *Physical Review E*, **52**(3), 2318-2321.

Sheather, S. J. and M. C. Jones (1991) “A reliable data-based bandwidth selection method for kernel density estimation” *Journal of the Royal Statistical Society. Series B (Methodological)*, 683-690.

Silverman B. W. (1978) “Density ratios, empirical likelihood and cot death” *Applied Statistics*, **27**, 26–33.

Takada T. (2012) “Mining local and tail dependence structures based on pointwise mutual information” *Data Mining and Knowledge Discovery*, **24**(1), 78-102.

Tjøstheim D and K. O. Hufthammer (2013) “Local Gaussian correlation: A new measure of dependence” *Journal of Econometrics* **172**(1), 33–48.