Testing significance of mixing and demixing coefficients in ICA

Shohei Shimizu[1 2], Aapo Hyvärinen[2], Yutaka Kano[1], Patrik O. Hoyer[2] and Antti J. Kerminen[2]

[1] Graduate School of Engineering Science, Osaka University, Japan
[2] Helsinki Institute for Information Technology, Basic Research Unit, Finland
`http://www.cs.helsinki.fi/hiit_bru/index_neuro.html`

Abstract. Independent component analysis (ICA) has been extensively studied since it was originated in the field of signal processing. However, almost all the researches have focused on estimation and paid little attention to testing. In this paper, we discuss testing significance of mixing and demixing coefficients in ICA. We propose test statistics to examine significance of these coefficients statistically. A simulation experiment implies the good performance of our testing procedure. A real example in psychometrics, which is a new application area of ICA, is also presented.

1 Introduction

Independent component analysis (ICA) [1] is a multivariate analysis technique that aims at recovering linearly mixed unobserved multidimensional independent signals from the mixed observable variables. Let $\boldsymbol{x}$ be an m-dimensional observed vector. The ICA model for $\boldsymbol{x}$ is written as

$$\boldsymbol{x} = \mathbf{A}\boldsymbol{s}, \tag{1}$$

where $\mathbf{A}$ is called a mixing matrix and $\boldsymbol{s}$ is an n-dimensional vector of independent components with zero mean and unit variance. Typically, the number of observed variables m is assumed to equal that of latent variables n, that is, $m = n$, which we assume in the following. The main goal of ICA is to estimate the mixing matrix $\mathbf{A}$ or the demixing matrix $\mathbf{B}^T = \mathbf{A}^{-1}$. (Some authors use $\mathbf{B} = \mathbf{A}^{-1}$ *without* the transpose [1].)

The ICA has been extensively studied since identification conditions for the model were provided in [2]. However, almost all the researches have focused on estimation [3–5], e.g., consistency, stability, robustness and asymptotic variance [6–8], and have not paid very much attention to testing. In this paper, we discuss testing of significance of mixing and demixing coefficients a_{ij} and b_{ij}. Such a test of significance is an important process in psychometrics for example [9].

The paper is structured as follows. First, in Section 2, we briefly review asymptotic variance of ICA and provide asymptotic covariance matrices of mixing and demixing coefficients estimated by ICA based on non-gaussianity maximization with constraints of orthogonality, e.g., FastICA [5]. In Section 3, we derive test statistics to evaluate the magnitude of significance of these coefficients using the asymptotic variances. We also consider multiple comparison procedures since we usually test significance of more than one coefficient. In Sections 4 and 5, we conduct a simulation study and provide a real data example to study how the test statistics work empirically. We conclude the paper in Section 6.

2 Asymptotic variance of ICA

Several authors studied asymptotic variance of ICA [7, 8, 10, 11], where the theory of estimating functions [12] was often used. Let us consider a semiparametric model $p(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a r-dimensional parameter vector of interest. Note that the density function $p(\boldsymbol{x}|\boldsymbol{\theta})$ is unknown. Let us denote by $\boldsymbol{\theta}_0$ the true parameter vector of interest. A r-dimensional vector-valued function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})$ is called an estimating function when it satisfies the following conditions for any $p(\boldsymbol{x}|\boldsymbol{\theta}_0)$:

$$E[\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}_0)] = \mathbf{0} \tag{2}$$

$$|\det \mathbf{Q}| \neq 0, \qquad \text{where } \mathbf{Q} = E\left[\frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] \tag{3}$$

$$E[\|\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}_0)\|^2] < \infty, \tag{4}$$

where the expectation E is taken over $\boldsymbol{x}$ with respect to $p(\boldsymbol{x}|\boldsymbol{\theta}_0)$.

Let $\boldsymbol{x}(1), \cdots, \boldsymbol{x}(N)$ be a random sample from $p(\boldsymbol{x}|\boldsymbol{\theta}_0)$. Then an estimator $\hat{\boldsymbol{\theta}}$ is obtained by solving the estimating equation:

$$\sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}(i), \boldsymbol{\theta}) = \mathbf{0}. \tag{5}$$

Under some regularity conditions including identification conditions for $\boldsymbol{\theta}$, the estimator $\hat{\boldsymbol{\theta}}$ is consistent when N goes to infinity and asymptotically distributes according to the gaussian distribution $N(\boldsymbol{\theta}_0, \mathbf{G})$, and

$$\mathbf{G} = \frac{1}{N} \mathbf{Q}^{-1} E[\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}_0) \boldsymbol{f}^T(\boldsymbol{x}, \boldsymbol{\theta}_0)] \mathbf{Q}^{-T}. \tag{6}$$

In [7], an estimating function for (quasi-) maximum likelihood estimation was derived. In [13], an estimating function for JADE [4] was provided, and an estimating function for ICA based on non-gaussianity maximization with orthogonality (uncorrelatedness) constraints including FastICA [5] was also introduced.

In this paper, we restrict ourselves to testing mixing and demixing coefficients estimated by FastICA. In FastICA, we first center the data to make its mean zero and whiten the data by computing a matrix $\mathbf{V}$ such that the covariance matrix of $\boldsymbol{z} = \mathbf{V}\boldsymbol{x}$ is the identity matrix. After that, we find an orthogonal matrix $\mathbf{W}$ so that components of $\mathbf{W}^T \boldsymbol{z} = \mathbf{W}^T \mathbf{V} \boldsymbol{x}$ have maximum non-gaussianity. Then we obtain estimates of $\mathbf{A}$ and $\mathbf{B}$ by $\mathbf{A} = \mathbf{V}^{-1}\mathbf{W}$ and $\mathbf{B} = \mathbf{V}^T \mathbf{W}$.

Let us consider the following function:

$$\mathbf{F}(\boldsymbol{x}, \mathbf{W}) = \boldsymbol{y}\boldsymbol{y}^T - \mathbf{I} + \boldsymbol{y}\boldsymbol{g}^T(\boldsymbol{y}) - \boldsymbol{g}(\boldsymbol{y})\boldsymbol{y}^T, \tag{7}$$

where $\boldsymbol{y} = \mathbf{B}^T\boldsymbol{x} = \mathbf{W}^T\mathbf{V}\boldsymbol{x} = \mathbf{W}^T\boldsymbol{z}$ and $g(u)$ is the nonlinearity. The estimating function for FastICA is obtained as $\boldsymbol{f} = \text{vec}(\mathbf{F})$ taking $\boldsymbol{\theta} = \text{vec}(\mathbf{W})$ [13], where $\text{vec}(\cdot)$ denotes the vectorization operator which creates a column vector from a matrix by stacking its columns.

According to the estimating function theory, we obtain the asymptotic covariance matrix of $\text{vec}(\mathbf{W})$ by (6) (see the Appendix for the complete formula). Here we assume that the variance in the estimate of $\mathbf{V}$ is negligible with respect to the variance in $\mathbf{W}$, which is validated empirically in the simulation below. Then we obtain the asymptotic covariance matrix of $\text{vec}(\mathbf{A})$ and $\text{vec}(\mathbf{B})$ as follows:

$$\begin{aligned}
\text{acov}\{\text{vec}(\mathbf{A})\} &= \text{acov}\{\text{vec}(\mathbf{V}^{-1}\mathbf{W})\} \\
&= (\mathbf{I} \otimes \mathbf{V}^{-1})\text{acov}\{\text{vec}(\mathbf{W})\}(\mathbf{I} \otimes \mathbf{V}^{-1})^T && (8) \\
\text{acov}\{\text{vec}(\mathbf{B})\} &= \text{acov}\{\text{vec}(\mathbf{V}^T\mathbf{W})\} \\
&= (\mathbf{I} \otimes \mathbf{V}^T)\text{acov}\{\text{vec}(\mathbf{W})\}(\mathbf{I} \otimes \mathbf{V}^T)^T, && (9)
\end{aligned}$$

where $\otimes$ denotes the Kronecker product. Given an $m \times n$ matrix $\mathbf{T}$ and $p \times q$ matrix $\mathbf{U}$, the Kronecker product of $\mathbf{T}$ and $\mathbf{U}$ is the following $mp \times nq$ matrix

$$\mathbf{T} \otimes \mathbf{U} := \begin{bmatrix} t_{11}\mathbf{U} & \cdots & t_{1n}\mathbf{U} \\ \vdots & \ddots & \vdots \\ t_{m1}\mathbf{U} & \cdots & t_{mn}\mathbf{U} \end{bmatrix}. \tag{10}$$

Matlab codes to compute $\text{acov}\{\text{vec}(\mathbf{A})\}$ and $\text{acov}\{\text{vec}(\mathbf{B})\}$ are available online at the webpage: `http://chobi.sigmath.es.osaka-u.ac.jp/~shimizu/acov/`

3 Testing significance of mixing and demixing coefficients

3.1 Wald statistics

In this paper, we would like to test if mixing or demixing coefficients are zero or not. Such tests are related to the fundamental question typically posed in empirical sciences: Does the independent component s_j have a statistically significant effect on the observed variable x_i? Here, the null and alternative hypotheses H_0 and H_1 are as follows:

$$H_0: \; a_{ij} = 0 \qquad \text{versus} \qquad H_1: \; a_{ij} \neq 0 \tag{11}$$

or

$$H_0: \; b_{ij} = 0 \qquad \text{versus} \qquad H_1: \; b_{ij} \neq 0. \tag{12}$$

One can use the following Wald statistics

$$\frac{\hat{a}_{ij}^2}{\text{avar}(\hat{a}_{ij})} \text{ or } \frac{\hat{b}_{ij}^2}{\text{avar}(\hat{b}_{ij})} \tag{13}$$

to test significance of a_{ij} and b_{ij}, where $\mathrm{avar}(\hat{a}_{ij})$ and $\mathrm{avar}(\hat{b}_{ij})$ denote the asymptotic variances of $\hat{a}_{ij}$ and $\hat{b}_{ij}$ computed by (8) and (9). The Wald statistics can be used to test the null hypothesis H_0. Under H_0, the Wald statistic asymptotically approximates to a chi-square variate with one degree of freedom [9]. Then we can obtain the probability of having a value of the Wald statistic larger than or equal to the empirical one computed from data. We reject H_0 if the probability is smaller than a significance level, and otherwise we accept H_0. Acceptance of H_0 implies that the assumption $a_{ij} = 0$ (or $b_{ij} = 0$) fits data. Rejection of H_0 suggests that the assumption is in error so that H_1 holds [9].

3.2 Multiple comparison

Usually, mixing and demixing matrices have more than one element. In many cases, we need to perform more than one testing simultaneously to find out if all or all of a set of the coefficients are significantly large in an absolute value sense. Although a given significance level may be appropriate for each individual testing, it is not for the set of all the testing. We are bound to have a lot of spurious significance if we just repeat testing without any corrections. Suppose we repeat testing 1,000 times at significance level 5%. Assume all the null hypotheses are true. Nevertheless, we can always expect that approx. 50 null models are rejected. However, we should not reject the null models. We have to control the probability of having at least one spurious false positive. In such a case, we should employ multiple comparison procedures. A simple and basic method is the Bonferroni correction, where we simply divide a significance level by the number of testing to obtain the significance level for individual testing. See [14] for details. We employ the Bonferroni correction in the simulation and real data analysis below.

4 Simulation

We conducted simulations in an attempt to confirm the theoretical results above. We employed FastICA, where the hyperbolic tangent function was taken as the nonlinearity and the symmetric orthogonalization was applied.

The simulation consisted of 10,000 replications. We employed the following mixing matrix that was lower triangular:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.65 & 0.7 & 1 \end{bmatrix}, \tag{14}$$

and then the demixing matrix $\mathbf{B}^T = \mathbf{A}^{-1}$ was

$$\mathbf{B}^T = \begin{bmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0 \\ -0.3 & -0.7 & 1 \end{bmatrix}. \tag{15}$$

In each replication, we generated independent sources and created observed signals following the ICA model (1). First, we created three independent components s_i with the sample size $N = 300, 500, 1000$ where their components were independently distributed according to the Laplace distribution. The independent components were normalized to have zero means and unit variances.

The FastICA was then applied to the data, and estimates of mixing and demixing coefficients a_{ij} and b_{ij} were obtained. Then we computed Wald statistics for these coefficients and tested the null hypotheses of the coefficients being zero as described above. The significance level was set at 5%. We computed how many null hypotheses on the coefficients in these matrices were rejected to know if the chi-square approximation worked for finite sample sizes. We also counted the numbers of cases where at least one of null hypotheses on the coefficients with zero values (here, elements in strictly upper triangular parts of $\mathbf{A}$ and $\mathbf{B}^T$) was rejected to know if the Bonferroni correction was effective.

The results are shown in Tables 1 and 2. First in Table 1, we shall examine the empirical significance levels (number of rejections) for a_{12}, a_{13} and a_{23}, and b_{12}, b_{13} and b_{23} that had zero values. Overall, we would say that the numbers of rejections of the null models were very close to the theoretically expected number 500. Second, Table 1 allows us to examine the statistical power of the test for the other coefficients that had non-zero values. The power of 0.99 (9,900 rejections) was achieved for all the conditions other than when testing b_{31} with $N = 300$. Thus, Table 1 showed that the Wald statistics were well approximated by the chi-square distribution, and the power of test was quite good.

Table 1. Numbers of rejected null hypotheses with significance level 5% (10,000 replications)

	a_{11}	a_{21}	a_{31}	a_{22}	a_{32}	a_{33}	a_{12}	a_{13}	a_{23}
N =									
300	9,999	9,914	9,931	9,995	9,946	9,984	467	499	478
500	10,000	9,997	9,995	10,000	9,997	10,000	506	473	475
1,000	10,000	10,000	9,996	10,000	10,000	10,000	488	468	477

	b_{11}	b_{21}	b_{31}	b_{22}	b_{32}	b_{33}	b_{12}	b_{13}	b_{23}
N =									
300	9,994	9,903	9,159	9,990	9,962	9,999	406	452	422
500	10,000	9,994	9,893	10,000	9,998	10,000	442	454	464
1,000	10,000	10,000	9,999	10,000	10,000	10,000	428	456	507

Note: N is sample size in estimation.

Next in Table 2, we examine the numbers of cases where at least one of null hypotheses on the coefficients with zero values to study the performance of the Bonferroni correction for multiple comparison discussed in Section 3.2. Overall, we would say that the numbers of rejections with the Bonferroni correction were rather close to the theoretically expected number 500 for all the

Table 2. Numbers of cases where at least one of null hypotheses on coefficients with zero values was rejected (10,000 replications)

	Testing a_{ij}			Testing b_{ij}		
	$N=$ 300	500	1,000	$N=$ 300	500	1,000
Bonferroni correction	497	471	447	383	408	411
No corrections	1,307	1,344	1,326	1,139	1,222	1,249

Note: N is sample size in estimation.

conditions, though the null models were rejected a bit less often than the theoretically expected number 500 when testing b_{ij}. On the other hand, the numbers of rejections with no Bonferroni correction were much larger than the theoretically expected number for all the conditions. Thus, Table 2 showed that the Bonferroni correction was effective and should be applied to real data analyses.

5 Example with real data

Questionnaire data about criminal psychology were analyzed as an example. The survey was conducted with university students in Japan [15]. The sample size was 222. Observed variables were standardized to have zero mean and unit variance. The labels of the variables x_1 and x_2 are "x_1: Sum of scores of items that ask subjective evaluation on frequency of your criminal opportunities when you went to high school" and "x_2: Sum of scores of items that ask subjective evaluation on frequency of your criminal behavior when you went to high school". A Kolmogorov-Smirnov test showed that all variables could not be assumed to come from the gaussian distribution (significance level 1%). Thus, ICA should be applicable to this kind of non-gaussian data.

We employed FastICA with the nonlinearity $g(u) = \tanh(u)$ and the symmetric orthogonalization. We set the significance level at 5% and used the Bonferroni method for multiple comparison. The estimated $\mathbf{A}$ by FastICA was

$$\begin{bmatrix} 0.93 & 0.36 \\ 0.77 & 0.64 \end{bmatrix}, \tag{16}$$

where a_{11}, a_{21} and a_{22} were significant, and a_{12} was not significant. See Table 3 for the Wald statistics. Thus, the matrix $\mathbf{A}$ could be seen to be lower triangular.

Table 3. Estimates, Wald statistics and p values

	a_{11}	a_{21}	a_{12}	a_{22}
Estimates	0.93	0.77	0.36	0.64
Wald statistics	48.70	17.57	1.69	9.24
p values	0.00	0.00	0.19	0.00

The fact that $\mathbf{A}$ is lower triangular allows us to interpret the results in terms of a causal ordering of the variables [16]. The result implied the causal order, $x_1 \to x_2$, that is, criminal opportunities at high schools → criminal behaviors at high schools. The link between the lower triangularity of $\mathbf{A}$ and the causal order can be seen as follows. For the lower triangular mixing matrix, x_1 is essentially equal to s_1, up to a multiplicative constant, a_{11}. On the other hand, x_2 is a function of s_1 and s_2, $a_{21}s_1 + a_{22}s_2$. Thus, x_2 is a function of x_1 and a new independent variable, s_2, that is, $(a_{21}/a_{11})x_1 + a_{22}s_2$. This indicates that x_1 may cause x_2, but x_2 cannot cause x_1. See [16] for details.

In fact, the order $x_1 \to x_2$ was reasonable to the criminal psychology theory. According to a criminal psychology theory [17], the frequency of criminal opportunities (x_1) is a typical environmental cause of the frequency of criminal behaviors (x_2) [15]. Therefore, the possible causal order from background knowledge was $x_1 \to x_2$. Thus, the causal order founded by our method would be reasonable to the background knowledge.

6 Conclusion

In this paper, we proposed Wald statistics to test significance of mixing and demixing coefficients in ICA. We conducted a small simulation experiment, which implied that our testing procedure worked well even for finite sample sizes, although more simulation studies are needed to study to what extent the result can be generalized. We also provided a real data example in psychometrics that would be a promising new area that ICA applies.

Appendix: A complete formula of acov{vec(W)}

The formula of acov{vec($\mathbf{W}$)} for FastICA is written as

$$\text{acov}\{\text{vec}(\mathbf{W})\} = \frac{1}{N}\mathbf{Q}^{-1}E[\text{vec}\{\mathbf{F}(\boldsymbol{x},\mathbf{W})\}\text{vec}\{\mathbf{F}(\boldsymbol{x},\mathbf{W})\}^T]\mathbf{Q}^{-T}. \tag{17}$$

Denote by F_{pq} the (p,q)-element of $\mathbf{F}$ and by $\mathbf{F}_q$ the q-th column of $\mathbf{F}$, respectively. We shall provide $E(F_{pq}F_{rs})$ to compute $E\{\text{vec}(\mathbf{F})\text{vec}(\mathbf{F})^T\}$. Denote by i, j, k, l four different subscripts. Then we have

$$\begin{aligned}
E(F_{ii}F_{ii}) &= E(s_i^4) + 1,\ E(F_{ii}F_{jj}) = 2,\ E(F_{ki}F_{ij}) = -E\{g(s_k)\}E\{g(s_j)\}\\
E(F_{ki}F_{li}) &= E\{g(s_k)\}E\{g(s_l)\},\ E(F_{ki}F_{kj}) = E\{g(s_i)\}E\{g(s_j)\}\\
E(F_{ii}F_{li}) &= -E(s_i^3)E\{g(s_l)\},\ E(F_{ki}F_{ii}) = -E(s_i^3)E\{g(s_k)\}\\
E(F_{ii}F_{ij}) &= E(s_i^3)E\{g(s_j)\},\ E(F_{ji}F_{jj}) = E(s_j^3)E\{g(s_i)\}\\
E(F_{ii}F_{lj}) &= 0,\ E(F_{ki}F_{jj}) = 0,\ E(F_{ki}F_{lj}) = 0\\
E(F_{ji}F_{ij}) &= 1 + 2E\{s_ig(s_i)\}E\{s_jg(s_j)\} - E\{g(s_i)^2\} - E\{g(s_j)^2\}\\
E(F_{ji}F_{lj}) &= 1 + E\{s_ig(s_i)\} + E\{s_jg(s_j)\}\\
&\quad + E\{s_ig(s_i)\}E\{s_jg(s_j)\} - E\{g(s_i)\}E\{g(s_l)\}
\end{aligned}$$

$$E(F_{ki}F_{ki}) = 1 + 2E\{s_i g(s_i)\} - 2E\{s_k g(s_k)\} + E\{g(s_i)^2\} + E\{g(s_k)^2\} - 2E\{s_i g(s_i)\}E\{s_k g(s_k)\}.$$

We also give $E\left\{(\partial \mathbf{F}_i)/(\partial \boldsymbol{w}_j^T)\right\}$ to compute $\mathbf{Q} = E\left[\{\partial \mathrm{vec}(\mathbf{F})\}/\{\partial \mathrm{vec}(\mathbf{W})^T\}\right]$:

$$E\left[\frac{\partial \mathbf{F}_i}{\partial \boldsymbol{w}_i^T}\right] = \begin{cases} 2\boldsymbol{w}_i^T & (i\text{−th row}) \\ \left[1 - E\{s_k g(s_k)\} + E\{g'(s_i)\}\right] \boldsymbol{w}_k^T & (k\text{−th row},\ k \neq i) \end{cases}$$

$$E\left[\frac{\partial \mathbf{F}_i}{\partial \boldsymbol{w}_j^T}\right] = \begin{cases} [1 - E\{g'(s_j)\} + E\{s_i g(s_i)]\boldsymbol{w}_i^T & (j\text{−th row},\ j \neq i) \\ \mathbf{0}^T & (k\text{−th row},\ k \neq j) \end{cases}.$$

References

1. Hyvärinen, A., Karhunen, J., Oja, E.: Independent component analysis. Wiley, New York (2001)
2. Comon, P.: Independent component analysis. a new concept? Signal Processing **36** (1994) 62–83
3. Amari, S.: Natural gradient learning works efficiently in learning. Neural Computation **10** (1998) 251–276
4. Cardoso, J.F., Souloumiac, A.: Blind beamforming for non Gaussian signals. IEE Proceedings-F **140** (1993) 362–370
5. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Networks **10** (1999) 626–634
6. Amari, S., Chen, T.P., Cichocki, A.: Stability analysis of learning algorithms for blind source separation. Neural Networks **10** (1997) 1345–1351
7. Pham, D.T., Garrat, P.: Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. Signal Processing **45** (1997) 1457–1482
8. Hyvärinen, A.: One-unit contrast functions for independent component analysis: A statistical analysis. In: Neural Networks for Signal Processing VII (Proceedings of IEEE Workshop on Neural Networks for Signal Processing). (1997) 388–397
9. Bollen, K.A.: Structural Equations with Latent Variables. Wiley, New York (1989)
10. Cardoso, J.F., Laheld, B.H.: Equivariant adaptive source separation. IEEE Transactions on Signal Processing **44** (1996) 3017–3030
11. Tichavský, P., Koldovský, Z., Oja, E.: Peformance analysis of the FastICA algorithm and Cramèr-Rao bounds for linear independent component analysis. IEEE Transactions on Signal Processing (2005) In press.
12. Godambe, V.P.: Estimating functions. Oxford University Press, New York (1991)
13. Kawanabe, M., Müller, K.R.: Estimating functions for blind separation when sources have variance dependencies. Journal of Machine Learning Research **6** (2005) 453–482
14. Hochberg, Y., Tamhane, A.C.: Multiple comparison procedures. John Wiley & Sons, New York (1987)
15. Murakami, N.: Research on causes of criminal and deviant behavior. Bachelor thesis, Osaka University, School of Human Sciences (2000) (In Japanese).
16. Shimizu, S., Hyvärinen, A., Kano, Y., Hoyer, P.O.: Discovery of non-gaussian linear causal models using ICA. In: Proc. the 21st Conference on Uncertainty in Artificial Intelligence (UAI-2005). (2005) 526–533
17. Gottfredson, M.R., Hirschi, T.: A general theory of crime. Stanford University Press, Stanford, CA (1990)