PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 32, Fasc. 2 (2012), pp. 47–55

WEAK CONSISTENCY OF MODIFIED VERSIONS OF BAYESIAN INFORMATION CRITERION IN A SPARSE LINEAR REGRESSION*

BY

PIOTR SZULC (WROCŁAW)

Abstract. We consider the regression model in the situation when the number of available regressors p_n is much bigger than the sample size n and the number of nonzero coefficients p_{0n} is small (the sparse regression). To choose the regression model, we need to identify the nonzero coefficients. However, in this situation the classical model selection criteria for the choice of predictors like, e.g., the Bayesian Information Criterion (BIC) overestimate the number of regressors. To address this problem, several modifications of BIC have been recently proposed. In this paper we prove weak consistency of some of these modifications under the assumption that both n and p_n as well as p_{0n} go to infinity.

2000 AMS Mathematics Subject Classification: Primary: 62J05; Secondary: 92D20.

Key words and phrases: Sparse linear regression, mBIC, mBIC2, consistency.

1. INTRODUCTION

If we want to choose the regression model in the case when we have a lot of exogenous variables, we should at first identify the nonzero coefficients. We can use model selection criteria for the choice of predictors but the classical ones, e.g. Akaike Information Criterion (AIC, Akaike 1974) or Bayesian Information Criterion (BIC, Schwarz 1978), were derived based on the assumption that the sample size n goes to infinity, while the total number of available regressors p_n remains constant. In our case p_n is much bigger than n and those classical criteria are inappropriate (they overestimate the number of regressors [9]). Specifically, Bogdan et al. [8] showed that if $p_n/\sqrt{n} \to c \in (0, \infty]$, then the expected number of false positives (false regressors) detected by BIC may go to infinity. Since AIC selects more regressors than BIC, it will also overestimate the number of predictors.

* This research was funded by the Budget of Ministry of Science and Higher Education, the Republic of Poland, as a research project N N201 414139.

In the situation when p_n is larger than n, least squares estimators for regression coefficients are not unique and regression models are not identifiable. We need some prior knowledge, e.g. concerning the number of nonzero coefficients p_{0n}. In many applications we assume the sparsity, i.e. that p_{0n}/p_n is very small. The appropriate asymptotic assumption under which we examine theoretical properties of model selection criteria (consistency, the optimality) is $p_{0n}/p_n \to 0$. This assumption was used to construct modifications of BIC: mBIC [7], mBIC2 ([14], [15], [18]) and EBIC [11]. In 2011, Chen and Luo [13] proved weak consistency of EBIC. In this paper we use techniques from [13] and in a similar way we prove weak consistency of mBIC and mBIC2.

2. MODIFIED VERSIONS OF BIC

We consider the following linear model:

$$y_i = \sum_{j=1}^{p_n} \beta_{nj} x_{ij} + \epsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where ϵ_i's are independent variables with normal distribution $\mathcal{N}(0, \sigma^2)$. Equivalently, we can write

$$\mathbf{y}_n = \mathbf{X}_n \boldsymbol{\beta}_n + \boldsymbol{\epsilon}_n, \tag{2.2}$$

where $\mathbf{y}_n = (y_1, \ldots, y_n)^T$, $\mathbf{X}_n = (x_{ij})_{i=1,\ldots,n, j=1,\ldots,p_n}$, $\boldsymbol{\beta}_n = (\beta_{n1}, \ldots, \beta_{np_n})^T$, $\boldsymbol{\epsilon}_n = (\epsilon_1, \ldots, \epsilon_n)^T$. We denote by s a subset of $\{1, \ldots, p_n\}$, by $M(s)$ the set of explanatory variables with indices in s, and by $v(s)$ the number of elements in s. Finally, let $s_{0n} = \big\{j : \beta_{nj} \neq 0, j \in \{1, \ldots, p_n\}\big\}$ and $p_{0n} = v(s_{0n})$.

Our goal is the identification of important predictors. One of the most popular model selection criteria is the Bayesian Information Criterion (BIC, Schwarz 1978) which suggests choosing the model minimizing the following formula:

$$BIC(s) = n \ln \big(RSS(s)\big) + v(s) \ln n, \tag{2.3}$$

where $RSS(s)$ is the residual sum of squares.

BIC was derived in the Bayesian context and it is used to approximate the logarithm of the posterior probability of the given model. The posterior probability of the model $M(s)$ is proportional to $m(s)\pi_{M(s)}$, where $m(s)$ is the integrated likelihood of the data given the model $M(s)$ and $\pi_{M(s)}$ is the prior probability of $M(s)$. BIC neglects $\pi_{M(s)}$ which correspond to assigning the same prior probability to all models. As a result, the prior on the number of the nonzero coefficients p_{0n} is $\mathcal{B}(p_n, \frac{1}{2})$, where $\mathcal{B}$ is a binomial distribution. This distribution is concentrated almost entirely on $[p_n/2 - 2\sqrt{p_n}, p_n/2 + 2\sqrt{p_n}]$, which does not agree with the assumption that p_{0n} is small and leads to the overestimation of the number of regressors.

We can modify BIC using an informative prior distribution on p_{0n}. In 2004, Bogdan et al. [7] proposed the modification of BIC, called *mBIC*. In this criterion the prior distribution on p_{0n} is $\mathcal{B}\big(p_n, (Ep_{0n})/p_n\big)$, where Ep_{0n} is the expected value of p_{0n}. The resulting formula for mBIC is

$$mBIC(s) = n\ln\big(RSS(s)\big) + v(s)\ln n + 2v(s)\ln\frac{p_n}{c}, \tag{2.4}$$

where $c = Ep_{0n}$. If we do not know Ep_{0n}, we can use $c = 4$ (to control the overall type I error at the level below 10%). Good properties of mBIC were documented based on simulation studies and real data analysis in many papers, e.g. [2], [3], [8]. In 2011, Frommlet et al. [14] showed consistency and the asymptotic optimality (ABOS) of mBIC under sparsity and when the design matrix $\mathbf{X}_n$ is orthogonal. Specifically, they showed that mBIC is ABOS if $Ep_{0n} = \text{const}$ (i.e. it does not increase when $n \to \infty$).

In Bogdan et al. [8] and Frommlet et al. [14] it is shown that the additional penalty in mBIC is closely related to the Bonferroni correction for multiple testing. While the Bonferroni correction has been shown to have some asymptotic optimality properties under very sparse designs, Abramovich et al. [1] and Bogdan et al. [6] prove that it is substantially worse than the popular Benjamini–Hochberg procedure [4] for multiple testing, which is asymptotically optimal in a much wider range of sparsity parameters. Exploiting these good properties of the B–H procedure, several new model selection criteria for multiple regression have been proposed (see, e.g., [1], [16]). In this paper we will analyze one of these criteria, mBIC2, proposed in Frommlet et al. ([14], [15]). The formula for mBIC2 is

$$mBIC2(s) = n\ln\big(RSS(s)\big) + v(s)\ln n + 2v(s)\ln p_n - 2\ln\big(v(s)!\big). \tag{2.5}$$

In 2011, Frommlet et al. [14] showed that mBIC2 is ABOS when the design matrix $\mathbf{X}_n$ is orthogonal and when $Ep_{0n} = \text{const}$ or $Ep_{0n} \to \infty$ so that $Ep_{0n}/p_n \to 0$. Simulation studies confirm these good properties and show that mBIC2 usually performs better than mBIC.

In 2008, Chen and Chen [11] proposed another modification of BIC, called *EBIC*. Let S_j be the set of all combinations of j indices in $\{1,\ldots,p_n\}$ and let $\tau(S_j)$ be the size of S_j,

$$\tau(S_j) = \binom{p_n}{j}.$$

We assume that the probability of choosing the model s is

$$P(s) = \binom{p_n}{j}^{-\gamma}$$

if the model s has the size j. This assumption gives the formula for EBIC family:

$$EBIC_\gamma(s) = n\ln\big(RSS(s)\big) + v(s)\ln n + 2\gamma\ln\binom{p_n}{v(s)}, \tag{2.6}$$

where $\gamma \geqslant 0$. Simulation studies and genetic data analysis showed good properties of this criterion ([17], [19]). In 2008, Chen and Chen [11] showed that EBIC is consistent for $p_{0n} = \text{const}$ and when the maximum size of searched models is limited. In 2011, Chen and Luo [13] extended this result and proved consistency of EBIC for $p_{0n} \to \infty$.

In this paper we use techniques from [13] to confirm the good properties of mBIC and mBIC2 under the nonorthogonal design by proving their consistency.

3. WEAK CONSISTENCY OF MBIC AND MBIC2

To present the theorems about consistency, we need to introduce some additional notation. We denote by $X_n(s)$ the matrix composed of columns of X_n with indices in s. Let $H_n(s)$ be the matrix of the orthogonal projection on the space spanned by columns of $X_n(s)$, $H_n(s) = X_n(s)[X_n(s)^T X_n(s)]^{-1} X_n(s)^T$. Let $\Delta_n(s) = \mu_n^T [I_n - H_n(s)]\mu_n$, where $\mu_n = E\mathbf{y}_n = \mathbf{X}_n(s_{0n})\boldsymbol{\beta}_n(s_{0n})$.

In our case, when the number of available regressors is much bigger than the sample size, almost every column in the experimental matrix can be represented as a linear combination of others. Therefore, some models with small number of predictors can be represented by more than one combination of available regressors. To prevent this situation, we need a condition guaranteeing the identification of "small models." Such conditions are presented e.g. in [5] and [10] and depend on the experimental matrix. Here we use the following identifiability condition from [13], which assumes the identification of the true model and depends on the vector of expected values μ_n.

IDENTIFIABILITY CONDITION:

$$\lim_{n\to\infty} \min \left\{ \frac{\Delta_n(s)}{p_{0n} \ln p_n} : s_{0n} \not\subset s, v(s) \leqslant k_n \right\} = \infty, \tag{3.1}$$

where $k_n = kp_{0n}$ for some fixed $k > 1$.

According to [13], the identifiability condition (3.1) is implied by

$$\sqrt{\frac{n}{p_{0n} \ln p_n}} \min\{|\beta_{nj}| : j \in s_{0n}\} \to \infty \tag{3.2}$$

and the sparse Riesz condition:

$$\begin{aligned} 0 < c_{\min} &\leqslant \min \left\{ \lambda_{\min} \left(\frac{1}{n} X_n(s)^T X_n(s) \right) : v(s) \leqslant k_n \right\} \\ &\leqslant \min \left\{ \lambda_{\max} \left(\frac{1}{n} X_n(s)^T X_n(s) \right) : v(s) \leqslant k_n \right\} \leqslant c_{\max} < \infty, \end{aligned} \tag{3.3}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and the largest eigenvalues, respectively.

The first of the theorems says that the probability of finding the "true model" using mBIC goes to one when n goes to infinity. Notice that we have to limit the number of searching models to kp_{0n}.

THEOREM 3.1. *Assume the model* (2.1), *the identifiability condition* (3.1) *and that* $p_{0n} \ln p_n = o(\sqrt{n})$. *Then*

$$P\big(\min_{\substack{s:v(s)\leqslant k_n\\ s\neq s_{0n}}} mBIC(s) > mBIC(s_{0n})\big) \to 1.$$

Proof. Let s be any submodel. We want to show that with probability converging to one the difference $mBIC(s) - mBIC(s_{0n})$ is greater than zero if n is big enough. Similarly as in Chen and Luo [13], we can write

$$mBIC(s) - mBIC(s_{0n}) = T_1 + T_2, \tag{3.4}$$

where

$$T_1 = n \ln \frac{y_n^T\big(I_n - H_n(s)\big)y_n}{y_n^T\big(I_n - H_n(s_{0n})\big)y_n}, \tag{3.5}$$

$$T_2 = \big(v(s) - p_{0n}\big)\ln n + 2\big(v(s) - p_{0n}\big)\ln p_n. \tag{3.6}$$

Assume without loss of generality that $\sigma^2 = 1$ and consider $s_{0n} \not\subset s$. Chen and Luo [13] showed that

$$T_1 = n \ln\left(1 + \frac{\Delta_n(s)}{n}\big(1 + o_p(1)\big)\right). \tag{3.7}$$

We can write

$$\begin{aligned}
&(3.8)\quad mBIC(s) - mBIC(s_{0n})\\
&= n \ln\left(1 + \frac{\Delta_n(s)}{n}\big(1 + o_p(1)\big)\right) + \big(v(s) - p_{0n}\big)\ln n + 2\big(v(s) - p_{0n}\big)\ln p_n\\
&\geqslant \left[n \ln\left(1 + \frac{Cp_{0n}\ln p_n}{n}\big(1 + o_p(1)\big)\right) - p_{0n}\ln n - 2p_{0n}\ln p_n\right]\\
&\geqslant \left[n \ln\left(1 + \frac{Cp_{0n}\ln p_n}{n}\big(1 + o_p(1)\big)\right) - 3p_{0n}\ln p_n\right]\\
&= \left[Cp_{0n}\ln p_n \ln\left(1 + \frac{1}{n/(Cp_{0n}\ln p_n)}\big(1 + o_p(1)\big)\right)^{n/(Cp_{0n}\ln p_n)} - 3p_{0n}\ln p_n\right]
\end{aligned}$$

because $\Delta_n(s) > Cp_{0n}\ln p_n$ for any large $C > 0$ if n is large enough, by the consistency condition. The expression

$$\left(1 + \frac{1}{n/(Cp_{0n}\ln p_n)}\big(1 + o_p(1)\big)\right)^{n/(Cp_{0n}\ln p_n)} \tag{3.9}$$

goes to e because $p_{0n} \ln p_n = o(\sqrt{n})$, so when C is large enough, the difference above is greater than zero for all s with $v(s) \leqslant k_n$.

Chen and Luo [13] showed that when $s_{0n} \subset s$, we have

$$T_1 \geqslant -2j\big(\ln p_n + \ln(j \ln p_n)\big)\big(1 + o_p(1)\big), \tag{3.10}$$

where $j = v(s) - p_{0n}$ and $o_p(1)$ goes to zero faster than

$$\frac{c}{2j\big(\ln p_n + \ln(j \ln p_n)\big)} q_n^{k_n - p_{0n}}$$

with $c > 0$ and $q_n \to 0$. Consequently,

$$\begin{aligned} (3.11)\quad & mBIC(s) - mBIC(s_{0n}) \\ & \geqslant -2j\big(\ln p_n + \ln(j \ln p_n)\big)\big(1 + o_p(1)\big) + j \ln n + 2j \ln p_n \\ & = -2j \ln p_n o_p(1) + 2j \ln(\sqrt{n}) - 2j \ln(j \ln p_n)\big(1 + o_p(1)\big). \end{aligned}$$

We have $-2j \ln p_n o_p(1) \to 0$ and under the assumption that $p_{0n} \ln n = o(\sqrt{n})$ it follows that $\ln\big(\sqrt{n}/(j \ln p_n)\big) \to \infty$, so the difference above is greater than zero uniformly for all s with $v(s) \leqslant k_n$ with probability converging to one. ■

The second theorem says that the probability of finding the "true model" using mBIC2 goes to one when n goes to infinity.

THEOREM 3.2. *Assume the model* (2.1), *the identifiability condition* (3.1) *and that* $p_{0n}^2 \ln p_n = o(\sqrt{n})$. *Then*

$$P\Big(\min_{\substack{s: v(s) \leqslant k_n \\ s \neq s_{0n}}} mBIC2(s) > mBIC2(s_{0n})\Big) \to 1.$$

Proof. Consider $s_{0n} \not\subset s$. We can estimate $n \ln\big(RSS(s)\big) - n \ln\big(RSS(s_{0n})\big)$ in the same way as before, so we have

$$\begin{aligned} (3.12)\quad & mBIC2(s) - mBIC2(s_{0n}) \\ & = n \ln\left(1 + \frac{\Delta_n(s)}{n}\big(1 + o_p(1)\big)\right) + \big(v(s) - p_{0n}\big) \ln n \\ & \quad + 2\big(v(s) - p_{0n}\big) \ln p_n + 2 \ln(p_{0n}!) - 2 \ln\big(v(s)!\big) \\ & \geqslant n \ln\left(1 + \frac{C p_{0n} \ln p_n}{n}\big(1 + o_p(1)\big)\right) - p_{0n} \ln n - 2p_{0n} \ln p_n - 2kp_{0n} \ln(kp_{0n}) \end{aligned}$$

because $-2 \ln\big(v(s)!\big) \geqslant -2 \ln\big((kp_{0n})!\big) \geqslant -2kp_{0n} \ln(kp_{0n})$. The last inequality is the result of $(kp_{0n})! \leqslant (kp_{0n})^{kp_{0n}}$. Next, we can write

$$\begin{aligned} (3.13)\quad & -p_{0n} \ln n - 2p_{0n} \ln p_n - 2kp_{0n} \ln(kp_{0n}) \\ & \qquad \geqslant -3p_{0n} \ln p_n - 2kp_{0n} \ln p_n \geqslant -(2k + 3)p_{0n} \ln p_n, \end{aligned}$$

so in the same way as before we show that $mBIC2(s) - mBIC2(s_{0n})$ is greater than zero (we only need larger C).

Now consider $s_{0n} \subset s$. We can estimate $n\ln\big(RSS(s)\big) - n\ln\big(RSS(s_{0n})\big)$ in the same way as before, so we have

$$\begin{aligned}
(3.14)\quad & mBIC2(s) - mBIC2(s_{0n}) \\
& \geqslant -2j\Big(\ln p_n + \ln(j\ln p_n)\big(1+o_p(1)\big) + j\ln n\Big) \\
& \quad + 2j\ln p_n + 2\ln(p_{0n}!) - 2\ln\big(v(s)!\big) \\
& = -2j\ln p_n o_p(1) + 2j\ln\left(\frac{\sqrt{n}}{j\ln p_n}\right)\big(1+o_p(1)\big) + 2\ln(p_{0n}!) - 2\ln\big(v(s)!\big).
\end{aligned}$$

Since $\ln n! \geqslant n\ln n - n$ and $\ln n! \leqslant n\ln n$, we obtain

$$\begin{aligned}
(3.15)\quad \ln(p_{0n}!) - \ln\big(v(s)!\big) & \geqslant p_{0n}\ln p_{0n} - p_{0n} - (j+p_{0n})\ln(j+p_{0n}) \\
& \geqslant p_{0n}\ln p_{0n} - p_{0n}\ln e - p_{0n}\ln(kp_{0n}) - j\ln(kp_{0n}) \\
& = p_{0n}\ln\left(\frac{p_{0n}}{ekp_{0n}}\right) - j\ln(kp_{0n}) \\
& \geqslant -j\ln(ek) - j\ln(kp_{0n}) = -j\ln(ek^2p_{0n}).
\end{aligned}$$

Thus,

$$\begin{aligned}
(3.16)\quad & mBIC2(s) - mBIC2(s_{0n}) \\
& \geqslant -2j\ln p_n o_p(1) + 2j\ln\left(\frac{\sqrt{n}}{kp_{0n}\ln p_n}\right)\big(1+o_p(1)\big) - 2j\ln(ek^2p_{0n}).
\end{aligned}$$

We have $-2j\ln p_n o_p(1) \to 0$ and under the assumption that $p_{0n}^2\ln n = o(\sqrt{n})$ we get

$$\ln\left(\frac{\sqrt{n}}{ek^3p_{0n}^2\ln p_n}\right) \to \infty,$$

so the difference above is greater than zero uniformly for all s with $v(s) \leqslant k_n$ with probability converging to one. ■

The assumptions for consistency of EBIC are weaker, $p_{0n}\ln p_n$ can be $o(n)$ (see [14]). Assumptions for consistency of mBIC2 are stronger than mBIC.

4. DISCUSSION

We proved weak consistency of mBIC and mBIC2 in the sparse regression problem under some assumptions on the design matrix (the identifiability condition) and limitations of the total number of the available regressors and the nonzero coefficients. These theorems formally explain the good properties of mBIC and mBIC2 when identifying significant regressors in large data bases, e.g. in the problem of locating QTLs. However, consistency does not mean that a criterion is good at prediction, e.g. BIC is consistent but AIC is better if we want to predict response variables. In a further research we plan to analyze properties of mBIC and mBIC2 in terms of prediction and construct the criteria which will be optimal with respect to prediction under sparsity.

Frommlet et al. [14] showed that mBIC and mBIC2 are asymptotically optimal in the Bayesian context (ABOS) when the design matrix $\mathbf{X}_n$ is orthogonal. We expect that the current study will be useful in proving that these criteria are ABOS under nonorthogonal design.

Acknowledgments. I thank Z. Chen and Z. Luo for sharing the early version of their unpublished manuscript and M. Bogdan and F. Frommlet for many helpful discussions.

REFERENCES

[1] F. Abramovich, Y. Benjamini, D. L. Donoho and I. M. Johnstone, *Adapting to unknown sparsity by controlling the false discovery rate*, Ann. Statist. 34 (2006), pp. 584–653.

[2] A. Baierl, M. Bogdan, F. Frommlet and F. Futschik, *On locating multiple interacting quantitative trait loci in intercross designs*, Genetics 173 (2006), pp. 1693–1703.

[3] A. Baierl, F. Futschik, M. Bogdan and P. Biecek, *Locating multiple interacting quantitative trait loci using robust model selection*, Comput. Statist. Data Anal. 51 (2007), pp. 6423–6434.

[4] Y. Benjamini, Y. Hochberg and A. B. Tsybakov, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, J. Roy. Statist. Soc. Ser. B 57 (1) (1995), pp. 289–300.

[5] P. J. Bickel, Y. Ritov and A. B. Tsybakov, *Simultaneous analysis of LASSO and Dantzig selector*, Ann. Statist. 37 (2009), pp. 1705–1732.

[6] M. Bogdan, A. Chakrabarti, J. K. Ghosh and F. Frommlet, *Asymptotic Bayes-optimality under sparsity of some multiple testing procedures*, Ann. Statist. 39 (2011), pp. 1551–1579.

[7] M. Bogdan, J. K. Ghosh and R. W. Doerge, *Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci*, Genetics 167 (2004), pp. 989–999.

[8] M. Bogdan, J. K. Ghosh, M. Żak-Szatkowska, *Selecting explanatory variables with the modified version of Bayesian Information Criterion*, Qual. Reliab. Eng. Int. 24 (2008), pp. 627–641.

[9] K. W. Broman and T. P. Speed, *A model selection approach for the identification of quantitative trait loci in experimental crosses*, J. Roy. Statist. Soc. Ser. B 64 (2002), pp. 641–656.

[10] E. Candes and T. Tao, *The Dantzig selector: statistical estimation when p is much larger than n*, Ann. Statist. 35 (2007), pp. 2313–2351.

[11] J. Chen and Z. Chen, *Extended Bayesian information criterion for model selection with large model space*, Biometrika 94 (2008), pp. 759–771.

[12] J. Chen and Z. Chen, *Extended BIC for small n-large-P sparse GLM* (2010) (submitted, available at www.stat.nus.edu.sg/~stachenz/ChenChen.pdf).

[13] Z. Chen and Z. Luo, *Extended BIC for linear regression models with diverging number of parameters and high or ultra-high feature spaces* (2011) (technical raport available at arxiv.org/abs/1107.2502v1).

[14] F. Frommlet, M. Bogdan and A. Chakrabarti, *Asymptotic Bayes optimality under sparsity for general priors under the alternative* (2011) (technical raport available at arxiv.org/abs/1005.4753v2).

[15] F. Frommlet, F. Ruhaltinger, P. Twaróg and M. Bogdan, *A model selection approach to genome wide association studies*, Comput. Statist. Data Anal. (2011) (doi:10.1016/j.csda.2011.05.005).

[16] E. I. George and D. P. Foster, *Calibration and empirical Bayes variable selection*, Biometrika 87 (2000), pp. 731–747.

[17] W. Li and Z. Chen, *Multiple interval mapping for quantitative trait loci with a spike in the trait distribution*, Genetics 182 (2) (2009), pp. 337–342.

[18] M. Żak-Szatkowska and M. Bogdan, *Modified versions of Bayesian Information Criterion for sparse Generalized Linear Models*, Comput. Statist. Data Anal. 55 (11) (2011), pp. 2908–2924.

[19] J. Zhao and Z. Chen, *A two-stage penalized logistic regression approach to case-control genome-wide association studies* (2010) (submitted, available at www.stat.nus.edu.sg/~stachenz/MS091221PR.pdf).

Institute of Mathematics and Computer Science
Wrocław University of Technology
ul. Janiszewskiego 14a
50-372 Wrocław, Poland
E-mail: piotr.a.szulc@pwr.wroc.pl

Received on 11.10.2011;
revised version on 1.2.2012