Some Observations on Novel Statistical Issues in Analysis of High Dimensional Problems of Inference about Genes

Ritabrata Dutta* **and Jayanta K. Ghosh**
Department of Statistical Science, Purdue University, West Lafayette, IN 47907, USA

SUMMARY

Genomic studies and data have revolutionized genetic studies. Two kinds of studies, namely, microarrays for gene expression and SNP's for possible association with various diseases, have become very popular. We survey briefly both of these areas, highlighting some recent theoretical and methodological work in Statistics.

Keywords : High dimensional inference, Microarrays, SNP's, Benjamini-Hochberg test, mBIC, Lasso.

1. INTRODUCTION

This review is a tribute to Professor Prem Narain for his illustrous career as one of India's leading geneticists and Director of IASRI. In view of Professor Prem Narain's interest in Genetics and important current trends in Statistics and our own work, we focus on inference relating to Microarrays and SNP's, both of which have attracted a lot of interest of geneticists in recent years.

The microarray may be modeled as consisting of m independent normal mixtures, the model for the expression of the i-th gene being,

$$X_i \sim (1-p)N(0, \sigma^2) + pN(\mu_i, \sigma^2) \qquad (1)$$

where p is the probability that the gene was expressed (corresponding to the i-th alternative hypothesis) and $(1-p)$ is the probability that the gene is not expressed (corresponding to the i-th null hypothesis H_{0i}). We remind the reader that expression is essentialy measured by the amount of mRNA produced in the cell with that gene. Typically p is small while m is large and, to be detectable, the magnitude of the expressed genes, often referred to as signals, has to be rather large. Further simplication occurs if we assume under H_{1i}, μ_i itself is also normal $N(0, \tau^2)$ where τ^2 is typically rather large and a measure of the magnitude of μ_i. Such models are regarded as Parametric Empirical Bayes (PEB) and have become quite popular, see for example Bogdan *et al*. (2011) for references to papers based on PEB models. Nonparametric Empirical Bayes models have also been introduced by Efron (2008) but identifiability issues need to be settled. A reasonably complete theory of inference based on PEB and the Benjamini Hochberg multiple test is now available. We will survey this briefly in Section 2.

We now turn to SNP's. We quote from (Thain *et al*. 2004) single nucleotide polymorphisms are "single DNA base alterations between human individuals", which "are being analysed" as part of association studies between genes (or markers) and diseases.

The statistical problem for identifying significant SNP's from among a huge number, easily a few thousands, is like choosing a few markers in Quantitative Trait Loci (QTL) studies from among many, see for example, Bogdan *et al*. (2004). This is a regression problem involving variable selection. If the design matrix were orthogonal, so that the least squares

*Corresponding author : Ritabrata Dutta
E-mail address : rdutta@purdue.edu

estimates based on the full model are basically independent of model (*i.e.*, of all models that include the particular variable under study), the methods that have been successful for microarrays can be used. Without orthogonality, the high dimensional regression problem is much more difficult. Theoretical study has just begun. We discuss this a bit below.

We illustrate the difficulties by explaining why the theory of optimality of the Benjamini-Hochberg rule will not apply at all. Without independence the notion of the Bayes oracle of Bogdan *et al.* (2011) is not available. Even more fundamentally, without independence, the Benjamini-Hochberg multiple test isn't easy to define, nor is the theorem of Benjamini and Hochberg (1995) applicable. On the other hand in new work on multiple regression, Bickel *et al.* (2009) study optimality of popular procedures like Lasso or its relatively recent competitor, the Dantzig selector of Candes and Tao (2007), by studying suitable oracle attaining properties in the sparse case. Oracles are lower bounds to measure of risk of a decision rule.

In the case of SNP's the present study owes a lot to Frommlet *et al.* (2010, to appear) for several basic ideas. Like them we study the following variable selection procedures for a simulated example, namely mBIC of Bogdan *et al.* (2004, 2008) and Lasso due to Tibshirani (1996).

However, we also study Lasso with a different penalty that is suggested in Bickel *et al.* (2009) as suitable in the context of their Oracle, *i.e.*, a lower bound that Lasso attains in the sparse case.

Finally we evaluate each procedure by its predictive performance as given by the ratio of Residual Sum of Squares and Total Sum of Squares from the ANOVA table and the accuracy of estimating the number of SNP's in the data. Associating significant β's and true non-zero β's which represent significant SNP's is much more tricky and requires some form of bootstrap sampling and clustering of covariates. It appears one can associate significant SNP's only with such clusters of covariates, which are our proxy clusters of markers in real experiments. Our evalution of Lasso and mBIC is quite different from Frommlet *et al.* (2010). We are indebted to Frommlet *et al.* (2010) for all these insights about SNP's and in the way we generate simulated data. We strongly advise interested readers to read their paper carefully. Two additional remarks explaining possible extensions are collected together in Section 4. They were in response to several constructive suggestions and comments of the editor and referee.

2. INFERENCE ABOUT MICROARRAYS

We discuss and survey very briefly the rich, current literature on this topic from the point of view of three paradigms, classical statistics, PEB and Full Bayes.

The most important multiple test that is often applied to microarrays is the BH test based on the mixture model. X_i's are independent $N(\mu_i, \sigma^2)$. $\mu_i = 0$ under H_{0i} and $\mu_i \neq 0$ under H_{1i}, $i = 1..., m$. The error variance σ^2 may be estimated well and hence assumed known if there are replicates for each i. Often the normality assumption is due to normalizing transformation of a t-statistics, in which case also the variance is assumed known approximately. Some basic references are given in Bogdan *et al.* (2011).

Let p_i be the P-value for the test that rejects H_{oi} if $|X_i|$ is large, i.e.

$$p_i = P_{H_{oi}} \{|X_i| > \text{observed value of } |X_i|\} \quad (2)$$

Order p_i's as $p_{(1)} < p_{(2)} < ... < p_{(m)}$. Suppose we wish to control False Discovery Rate (FDR), *i.e.* $E_\mu\left(\frac{V}{R} I_{R>0}\right)$, where $\mu = (\mu_1, ..., \mu_m)$, V=#H_{0i}'s wrongly rejected (*i.e.* true but rejected), R = total number of H_{0i}'s rejected.

Fix a small α, we want

$$\max_{\mu} E_\mu\left(\frac{V}{R} I_{R>0}\right) \leq \alpha \quad (3)$$

Benjamini and Hochberg (1995) proved that their multiple test described below satisfies the above condition. This is a remarkable theorem. A relatively simple proof is given in Ewens and Grant (2005, pp 460-462).

We now describe the BH test. Order P-values as $p_{(1)} < ... < p_{(m)}$. Let i_0 be the last i such that

$$p_{(i)} \leq \frac{i}{m} \alpha. \quad (4)$$

(*i.e.* after i_0, the inequality is violated $\forall\, i = i_0 + 1, ..., m$).

Think of (4) as indicating $p_{(i)}$ is significantly small. The $p_{(i_0)}$ is the biggest significant value.

The BH test rejects all the null hypotheses corresponding to $p_{(1)}, \ldots, p_{(i_0)}$.

This is the famous Benjamini-Hochberg (BH) multiple test which attains the classical requirement. It is known that the test was proposed earlier by others (Seeger 1968 and Simes 1986) but the test was thoroughly discussed and made popular by Benjamini and Hochberg. It was they who proved the beautiful result.

This is a classical multiple test and valid under the high dimensional classical model. We next turn to the PEB model in which μ_i's are themselves iid with distribution

$$(1 - p)\delta_0 + pN(0, \tau^2) \qquad (5)$$

where δ_0 is the probability distribution which has total mass at zero (indicating H_{0i} is true). The second component gives the conditional distribution of μ_i under H_{1i}, namely $N(0, \tau^2)$.

The original high-dimensional problem has now reduced to a two-dimensional problem with p and τ^2 as unknown parameters. In the PEB approach p and τ^2 are estimated from the full data $(X_1, \ldots, X_m)$, and $\hat{p}$ is treated as the data based prior probability of $H_{(0i)}$. One can use the Bayes test. For details see Bogdan *et al.* (2008). Since $p \to 0$ as $m \to \infty$, estimating p well isn't easy, see the discussion of this in Bogdan *et al.* (2008) and Scott and Berger (2010). In the full Bayes approach, p and τ^2 are not estimated from data but given a hierarchical prior distribution. This is discussed in Scott and Berger (2006, 2010).

Bogdan *et al.* (2011) provide a theoretical justification of the BH test in the PEB model. It is shown in Bogdan *et al.* (2011) that the BH test behaves like the PEB test and asymptotically does as well as the lower bound to misclassification probability provided by the full Bayes test with known p and τ^2. It is also shown there unless τ^2 is sufficiently large, as in Assumption A in Bogdan *et al.* (2011), even the best test is quite poor. In high dimensional analysis good inference is possible only under two strong assumptions, sparsity of signal, *i.e.* small p, and sufficiently large signal, *i.e.* large τ^2, and these two have to be related as in the Assumption A of Bogdan *et al.* (2011).

To sum up there is a well developed, useful theory in each paradigm for inference on microarray.

3. INFERENCE ABOUT SNPs

3.1 Model assumptions

The SNP based Genome Wise Association Study (GWAS), can be easily seen as a multiple linear regression problem with variable selection as the key issue. Ideally each SNP corresponds to a covariate, so identifying SNP's is equivalent to identifying significant regression coefficients.

Actually, the problem is much more delicate because of correlation between covariates. We discuss the more realistic version towards the end of this paper. Let us treat the quantitative trait of n observations as the response varible $y_i : i \in \{1, \ldots, n\}$ and the corresponding genotype of person i and SNP j as $x_{ij} \in \{-1, 0, 1\} : 1 \in \{1, \ldots, n\}, j \in \{1, \ldots, p\}$. Now if the subset j^* of the p SNPs, having $k << p$ SNPs $1 \leq j_1^* \leq \ldots \leq j_k^* \leq p$, are causal for the trait y, then we can assume the additive true model as

$$M_j^* : y_i = \sum_{l=1}^{k} b_{j_l^*} x_{ij_l^*} + \varepsilon, \text{ when } \varepsilon \sim N(0, \sigma^2 I_k) \qquad (6)$$

But we can also create 2^p similar models using all p SNPs. A model not having any SNP will be denoted as M_0 or the null model and all other models will be denoted by M_j, where j is an ordered subset of elements of the set $\{1, \ldots, p\}$. Following the standard conventions, we write $q = q_j$ for the number of SNPs in a model. We define for each model M_j the matrix $\mathbf{X}^j$ containing the genotype of the SNPs in the model. Then the model becomes

$$M_j : y = \mathbf{X}^j \beta_j + \varepsilon^j \qquad (7)$$

Now as in multiple regression problem with variable selection, we want to choose the best model containing all the causal variables here. So we will discuss some variable selection methods in the next subsection under sparsity assumption, as our $k << p$.

3.2 Lasso, Lars and Stepwise selection

For variable selection in linear regression problems different skrinkage estimators like ridge regression are very common in use. Following the idea of shrinkage as in ridge regression, the Lasso method introduced by

Tibshirani (1996), tries to minimize the least square error of the regression with an upperbound on the L_1 norm of the parameter vector. So the estimate is defined by

$$\hat{\beta}^{lasso} = \text{argmin}_{\beta} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \quad (8)$$

$$\text{subject to } \sum_{j=1}^{p} |\beta_j| \le t$$

Here the upper bound t for the L_1-penalty of the parameter-vector controls the amount of shrinkage. Lasso chooses subsets of variables depending on the tuning-parameter t. When t is very small, then almost all the parameters are zero, similarly for large enough t value, the parameter estimate $\hat{\beta}$ is same as the least square estimate. The shrinkage constraint makes the solution nonlinear in y_i, needing a quadratic programming algorithm to compute the estimate. Efron *et al*. (2004) have shown that Lasso is closely related to another novel shrinkage estimation scheme Lars, introduced by them. In a general setup, they show that the subset selected by Lars, Lasso and Stepwise Selection are similar. Here we will concentrate on the most popular of them, namely, Lasso, its solution path obtained from Lars algorithm for fast variable selection. This has become the standard method for Lasso. The tuning parameter t is chosen by minimizing the cross-validation error.

Suppose we have a statistical inference problem, *e.g*., testing or estimation. An oracle depending on unkown parameters, is a lower bound for the risk or loss function of all decision functions we consider, and which is asymptotically attained by our chosen rule. If one can construct such an oracle for a particular decision rule, it immediately proves the asymptotic optimality of the chosen decision function. Such oracles were first proposed for AIC, Shibata (1984), Li (1987) and Shao (1997). An early oracle (not stated as such) is the Cramer-Rao inequality, which is an oracle for the mle. The results in Bogdan *et al*. (2008, 2011) are based on a Bayes oracle for all multiple tests.

Recently Bickel *et al*. (2009) have derived an Oracle property for Lasso under sparsity assumption. The Lasso constraint $\sum|\beta_j| \le t$ is equivalent to the addition of a penalty term $r\sum|\beta_j|$ to the residual sum of squares [Murray *et al*. (1981)]. While an explicit mathematical relation between t and r isn't available, the basic idea of convex optimization makes it easy to move from the one to the other. We will use both versions of Lasso. This can be written as following,

$$\hat{\beta}^{lasso} = \text{argmin}_{\beta} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + r\sum_{j=1}^{p} |\beta_j| \quad (9)$$

According to Bickel *et al*. (2009), when the errors ε_i are independent $N(0, \sigma^2)$ random variables with $\sigma^2 > 0$, all the diagonal elements of the matrix $\mathbf{X}'\mathbf{X}/n$ be equal to 1, then under some additional conditions on the Gram matrix, with $r = A\sigma\sqrt{\frac{\log(p)}{n}}$ and $A > 2\sqrt{2}$, with probability $1 - M^{1-\frac{A^2}{8}}$, we have

$$|\hat{\beta}^{lasso} - \beta_0|_1 \le \frac{16A}{c(s)}\sigma s\sqrt{\frac{\log(p)}{n}} \quad (10)$$

$$|\mathbf{X}(\hat{\beta}^{lasso} - \beta_0)|_2^2 \le \frac{16A^2}{c(s)}\sigma^2 s\log(p) \quad (11)$$

$$\#\{\hat{\beta}_j^{lasso} \ne 0\} \le \frac{64}{c_1(s)}s \quad (12)$$

when s is the no. of non-zero components in β_0, and $c(s)$, $c_1(s)$ are constants depending on s and the Gram matrix.

3.3 Modified Bayesian Information Criterion

For linear regression under assumption of normal error term $\varepsilon \sim N(0, \sigma^2)$ the likelihood function of each model M_j is given by

$$L_j(y|\beta_j, \sigma) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left(-\frac{(y - X^j\beta_j)'(y - X^j\beta_j)}{2\sigma^2} \right) \quad (13)$$

The maximum likelihood estimator of β_j is same as the least square estimate $\hat{\beta}_j$. So we know for fixed σ using BIC is then equivalent to minimizing a standardized residual sum of squares RSS_j/σ^2 under model M_j with penalty $q_j \log(n)$ for model dimension q_j.

$$\frac{RSS_j}{\sigma^2} + q_j \log(n) \quad (14)$$

But it is known (Bogdan *et al*. 2004) that under sparsity, BIC chooses too many regressors. As a remedy,

Bogdan *et al.* (2004) introduced a modification of BIC, as

$$mBIC : -2\log L_j(\hat{\beta}_j) + q_j(\log(n) - 2\log(w)) \quad (15)$$

where, w can be interpreted as a probability of a particular covariate being relevant. When such prior information isn't available there is also a default choice of w. In recent unpublished work, following a similar idea, Frommlet *et al.* (2010) introduced a different criterion, namely

$$mBIC2 : -\log L_j(\hat{\beta}_j) + q_j(\log(n) - 2\log(w)) - 2\log(q_j!) \quad (16)$$

which is suitable for multiple regression and works similarly to the Benjamini-Hochberg correction for multiple testing. We study the performance of this new criterion with Lasso for a simulation based GWAS study. In a GWAS study, the number of parameters is too big, so to apply *m*BIC2, they used a pre-screening scheme, which picks a smaller but more relevant subset of parameters to explore further. They conducted 1-variable β-significance test for each variables and then created a subset consisting only of the variables having a p-value below a pre-specified threshold. They chose the threshold by using their (assumed) rough prior knowledge about the expected number of significant variables. Following Frommlet *et al.* (2010), we also take the threshold here as 1.5. After the first stage of screening, we choose the final model minimizing *m*BIC2 criterion with a forward selection procedure. This prior information is used only for *m*BIC2 but not for Lasso.

3.4 Simulation Study

Like Frommlet *et al.* (2010) we generated a dataset to mimic a SNP dataset in real life, but still having control over the parameters. We assumed the sample size $n = 100$, and the total number of SNP's under study to be $p = 30{,}000$. Each covariate has been set to be a $\{-1, 0, 1\}$ valued r.v. as in the case of SNP, with the minor allele frequency always lying below 0.5. For our study the minor allele was taken as "1". We also checked the covariance between any two covariates in most of the cases to be in between –.1 to .1, signifying weak but not negligible covariance between the covariates. Our data set is somewhat smaller than the simulated data set of Frommlet *et al.* (2010). We denote the covariate matrix as $\mathbf{X}$, which is an $n \times p$ matrix. Forty causal SNP's have been selected randomly from the set of all "p" SNPs. Then we have chosen a vector β_0 having non-zero coefficient value lying between .5-1 for those 40 SNPs and we simulate the response as following

$$y = \beta_0'\mathbf{X} + \varepsilon, \text{ when } \varepsilon \sim N(0, \sigma^2\mathbf{I}) \quad (17)$$

Three variable selection procudres, namely Lasso with Cross validation, Lasso with Oracle-penalty and *m*BIC2, have been compared in this setup. The Lasso with Oracle property is new, not considered in Frommlet *et al.* (2010). Since detection of non-zero β's is a much more tricky task, we compare their predictive performance by the ratio of Residual Sum of Squares (RSS) and Total Sum of Squares (TSS) in Table 1, when we define RSS and TSS as following

$$RSS = \sum (Y_i - X'\hat{\beta})^2 \quad (18)$$
$$TSS = \sum (Y_i - \bar{Y})^2.$$

Also in the definition of *m*BIC2, we have used the prior knowledge w about the sparsity. So to compare it with Lasso, where we don't use any prior information, we consider *m*BIC2 for 3 different values of $w = \frac{30}{30000}, \frac{60}{30000}, \frac{120}{30000}$. *m*BIC2 does best but also chooses more variables depending on the information of the sparsity. It chooses 15, 56 and 99 variables correspondingly for $w = \frac{30}{30000}, \frac{60}{30000}, \frac{120}{30000}$. But both the versions of Lasso choose much smaller number

Table 1. $\frac{RSS}{TSS}$ for different methods

	Lasso-CV	Lasso-Oracle	*m*BIC2$\left(w = \frac{60}{30000}\right)$	*m*BIC2$\left(w = \frac{30}{30000}\right)$	*m*BIC2$\left(w = \frac{120}{30000}\right)$
$\frac{RSS}{TSS}$	.227	.092	.02	.018	.033

of variables, 23 and 31 in Lasso-CV and Lasso-Oracle correspondingly, *i.e.*, they were much more parsimonious. Lasso-Oracle comes nearest to estimating the correct number of SNP's. Unfortunately our simulation would need to be strengthened with Bootstrap before we can identify clusters of co-variates as causal variables, as mentioned in Introduction.

To check if the estimate of RSS/TSS will increase substantially under cross-validation, we simulated another data set and calculated RSS/TSS with the same estimates obtained earlier. We get this time the value in a ±.01 interval of the value found earlier. So our earlier conclusions do not change substantially, as we expected since the total data size is much bigger than the number of unknown parameters.

4. POSSIBLE EXTENSIONS

R1. The editor and referee have suggested we consider different locations for the experiment. Suppose the experiment is done at k locations, *i.e.*, with different Y_i's coming from one of these locations. We introduce k indicator covariates I_j, $j = 1, \ldots, k$. For the i-th sampling unit the j-th covariate I_j is equal to 1 if the unit is from the j-th location and zero otherwise. This will not change our linear model but we now have $p+k$ covariates, with the regression coefficients for the k location-covariates taking on the role of k intercepts. The formal equation (6) for the model remains same. We may still apply Lasso or modify it by targetting shrinkage for the non-locational regression coefficients only. These two versions will lead to the following versions of Lasso.

(a) Usual Lasso

(b) Modied Lasso

We do not know which one will be better. Many simulations would be needed to get a clear picture.

R2. Another suggestion from the editor and the referee is that we take note of the recent work on mixed directional (*md*FDR). Based on the pioneering work of Peddada *et al.* (2003) and Benjamini *et al.* (1995), there is now an extended notion of FDR, called *md*FDR and its control. In Guo *et al.* (2010) a new multiple test controlling *md*FDR is proposed, extending earlier work. We decribe this briefly since it is very useful when the microarray provides a time series for each gene. The time-series may show the response to increasing doses over time. Guo *et al.* (2010) present some simulation comparing their method with both the BH multiple test and the multiple test of Benjamini *et al.* (1995). We will return to this topic in our concluding remarks.

In this case each component test is about l parameters and in case of rejection, we wish to get the sign of each $\theta_{j1}, \ldots, \theta_{jl}$ right. A correct rejection of the j-th null along with a wrong sign is treated as false discovery. The corresponding random number of false discoveries is denoted by S and $md\text{FDR} \equiv E\left(\frac{S}{R \vee 1}\right)$. The total FDR $\equiv \left(\frac{V+S}{R \vee 1}\right)$ where V = # usual false discoveries and S is as explained earlier and R = the total number of rejections of nulls. The basic papers are Benjamini and Yekutieli (2005) and Guo *et al.* (2010).

5. CONCLUDING DISCUSSION

Even under sparsity there may be approximate colinearity between covariates, as in our simulated example. In this case the correct model may not be identifiable. This will often be the case for studies involving SNP's. Such problems require a thorough simulation and theoretical study. It appears that the forthcoming lecture by Bickel and Ritov at the Joint Statistical Meeting of ASA this year will provide major insights and new modications of Lasso to cope with these problems.

Other heuristic options would be to use cross-validation to choose one of several good estimators, or use clusters of covariates like Frommlet *et al.* (2010). It is evident from the abstract of an invited talk at the JSM (Florida, Summer 2011) by Bickel and Ritov that a major new modification of Lasso for the sparse, colinear case will be offered. A second problem is to

explore how large the signals should be for a given level of sparsity in the used regression setting. As mentioned before this problem is thoroughly discussed for micro-arrays in Bogdan *et al.* (2011). In the regression setting one can prove the necessity of such scaling by considering the case of orthogonal covariates (with each covariate normalized to have l_2-norm equal to one). It appears a sufficient condition can also be proved when there is no colinearity, the conditions for Theorem 7.2 of Bickel *et al.* (2009) hold and $\sqrt{\frac{\log(M)}{n}} \to 0$.

We would also like to suggest some future work on controlling FDR and *md*FDR, as invited by the editor and referee. In Bogdan *et al.* (2007) it is shown via simulations that many multiple tests, not just BH multiple test, attain a Bayes oracle approximately. They include Parametric Bayes and Parametric Empirical Bayes multiple tests as well as multiple test based on Bayesian Nonparametrics. Further studies of these other tests in the context of *md*FDR should be useful.

Finally, we note that one of us, Jyotishka Datta and Ghosh, have successfully used State Space models and multiple testing to identify gene pathways when a time series for each gene is provided by a microarray. This part is Jyotishka Datta's ongoing doctoral work.

ACKNOWLEDGEMENT

We thank Professor Malgortaza Bogdan for sharing (Frommlet *et al.* 2010) with us and for several illuminating discussions. We also thank Debapriya Sengupta of ISI for suggesting that the Cramer-Rao inequality is probably the first oracle.

REFERENCES

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate : A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc.,* **B57**, 289-300.

Benjamini, Y. and Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Amer. Statist. Assoc.*, **100**, 71-93.

Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.*, **37**, 1705-1732.

Bogdan, M., Ghosh, J.K. and Doerge, R.W. (2004). Modifying the Schwarz Bayesian Information Criterion to locate multiple interacting quantitive trait loci. *Genetics*, **167**, 989-999.

Bogdan, M., Ghosh, J.K. and Tokdar, S.T. (2007). A comparison of the Benjamini-Hochberg procedure with some Bayesian rule for multiple testing. In: Fetschript for P. K. Sen. IMS Lecture Notes-Monograph Series (N. Balakrishnan, Edsel A. Pena and Mervyn J. Silvapulle eds.), Beachwood, Ohio, USA: IMS, 2008, 211-230.

Bogdan, M., Biecek, P., Cheng, R., Frommlet, F., Ghosh, J.K. and Doerge., R.W. (2008). Extending the modified Bayesian Information Criterion (mBIC) to dense markers and multiple interval mapping. *Biometrics*, **64(4)**, 1162-1169.

Bogdan, M., Chakrabarti, A., Frommlet, F. and Ghosh, J.K., (2011). Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. To appear in *Ann. Statist.*

Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.*, **35(6)**, 2313-2351.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression *Ann. Statist.*, **32(2)**, 407-499.

Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.*, **23**, 122.

Ewens, W.J. and Grant, G.R. (2005). *Statistical Methods in Bioinformatics*. Springer.

Frommlet, F., Ruhaltinger, F., Twarog, P. and Bogdan, M. (2010). A model selection approach to genome wide association studies. (submitted), http://arxiv.org/abs/1010.0124

Guo, W., Sarkar, S.K. and Peddada, S.D. (2010). Controlling false discoveries in multidimensional directional decisions, with applications to gene expression data on ordered categories. *Biometrics*, **66(2)**, 485-492.

Li, K.C. (1987). Asymptotic optimality for Cp, CL, cross-validation and generalized crossvalidation: Discrete index set. *Ann. Statist.*, **15**, 958-975.

Murray, W., Gill, P.E. and Wright, M.H. (1981). *Practical Optimization*. Academic Press, London.

Peddada, S., Lobenhoer, E., Li, L., Afshari, C., Weinberg, C. and Umbach, D. (2003). Gene selection and clustering for time-course and dose response microarray experiments using order-restricted inference. *Bioinformatics*, **19**, 834-841.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica*, **7**, 221-264.

Scott, J.G. and Berger, J.O. (2006). An exploration of aspects of Bayesian multiple testing. *J. Statist. Plann. Inf.*, **136(7)**, 2144-2162.

Scott, J.G. and Berger, J.O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Statist.*, **38(5)**, 2587-2619.

Seeger, P. (1968). A note on a method for the analysis of significance en masse. *Technometrics*, **10**, 586-593.

Shibata, R. (1984). Approximate efficiency of a selection procedure for the number of regression variables. *Biometrika*, **71**, 43-49.

Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**, 751-754.

Thain, M., Hickman, M., Abercrombie, M., Hickman, C.J. J., Johnson, N.I. and Turvey, R. (2004). *The Penguin Dictionary of Biology*. Penguin.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc.*, **B58(1)**, 267-288.