Daniel YEKUTIELI

Department of Statistics and OR
School of Mathematical Sciences
Tel Aviv University

Research Statement

1 Overview

My research includes development, theoretical study, and implementation of False Discovery Rate controlling methodology. I currently participate in two NIH funded projects: Microarray analysis of morphine's behavioral effects – collaboration with Prof. Benjamini and a group of scientists from the University of Maryland, headed by Prof. Elmer; Synthetic lethality based assay for high through-put screening drug discovery – collaboration with Prof. Benjamini and Prof. Canaani from the Biochemistry Department at TAU. I collaborate with researchers from TAU on various projects: Microarray analysis with Prof. Chamovitz from the Department of Plant Sciences; discovery of over-represented Gene Ontology terms with Dr. Sharan from the School of Computer Science; statistical inference in Bioinformatic with Dr. Pupko from the Department of Cell Research and Immunology. I also collaborate with researchers from abroad: methodology for the analysis of dose response microarray experiments with Prof. Shkedy, from the Center for Statistics in Hasselt University, Belgium; behavioral phenotyping of exploratory rat behavioral for screening chemical compounds with Dr. Kafkafi and Prof. Elmer from the University of Maryland. The focus of my theoretical research is hierarchical FDR controlling procedures (supported by a grant from the Israel Science Foundation).

In Section 2 I provide a brief outline of some of the main research directions in FDR methodology, highlighting my contributions (underlined) and how they relate to the work of other researchers. In Section 3 I discuss hierarchical FDR testing. Section 4 is devoted to my planned research on Bayesian FDR.

2 FDR methodology

Benjamini and Hochberg introduced the FDR in 1995 as an alternative measure for type I error, in multiple testing, to the family wise error rate (FWE – the probability of making at least one type I error). Benjamini and Hochberg

(1995) argued that in many cases FWE control is not necessary, and that it is sufficient to control the proportion of false discoveries out of the total discoveries made. Benjamini and Hochberg (1995) also introduced a FDR controlling procedure (BH procedure), and proved that it controls the FDR for independently distributed test statistics.

Implementing the Westfall and Young (1993) resampling based multiple testing approach to control the FDR, Yekutieli and Benjamini (1999) introduced FDR adjusted p-values, and used them to scan the northern hemisphere for geographic regions whose atmospheric pressure is correlated to the precipitation in Israel; this methodology was later applied to microarray analysis in Reiner et al. (2003). However, the main significance of Yekutieli and Benjamini (1999) is that it turned control of the FDR from a multiple comparisons problem to an estimation problem much more amenable to statistical investigation: instead of comparing the sorted p-values to a series of critical values determined by the FDR level q (the BH procedure), estimate the FDR in a series of fixed rejection region tests. Storey (2002) and Storey (2003) discussed a Bayesian setting for the fixed rejection region FDR and introduced the positive FDR and the q-value with the following, very appealing, Bayesian interpretation: the conditional probability that a discovery is a false discovery given that its test statistic is in the rejection region. Efron et al. (2001) suggested empirical Bayes estimation of the FDR and even considered conditioning locally on the value of the test statistic, not just the rejection region. Genovese and Wasserman (2004) developed a framework in which the False Discovery Proportion, the number of false rejections divided by the number of rejections in a continuum of fixed rejection regions, is treated as a stochastic process.

Benjamini and Yekutieli (2001) generalized the work of Sarkar and Chang (1997) and Sarkar (1998) to derive an explicit expression for the FDR, proved that the BH procedure controls the FDR for positively dependent test statistics, and provided a general upper bound for the FDR of the BH procedure. Benjamini and Yekutieli (2005b) verified that the BH procedure can be validly applied to control the FDR in QTL mapping. Yekutieli (2006) presented a modification of the BH procedure for testing non positive dependent test statistics, and introduced a FDR controlling procedures for testing pairwise comparisons.

When some of the null hypotheses are not true, FDR controlling procedures are too conservative by a factor which is the proportion of the true null hypotheses among the the total number of hypotheses – m_0/m. In adaptive FDR procedures m_0/m is estimated and then used to derive a more powerful test-

ing procedure (Benjamini and Hochberg, 2000; Storey, Taylor and Siegmund, 2004; and others). Benjamini, Krieger and Yekutieli (2006) introduce an adaptive procedure in which the BH procedure is applied twice: once at level q to estimate m_0, and a second time at level $qm/\hat{m}_0$ to decide which hypotheses to reject; they also provide exact FDR computations and prove FDR control of adaptive procedure for independently distributed test statistics. Benjamini, Krieger and Yekutieli (2006) show in simulations that their procedure controls the FDR for dependent test statistics, and that the other, more powerful adaptive procedures, fail to control the FDR for dependent test statistics. Gittelman (2007) discusses the use of bootstrapping to approximate the sampling distribution of $\hat{m}_0$ and to construct a powerful testing procedure which controls the FDR for dependent test statistics.

Benjamini and Yekutieli (2005a) address the problem of providing statistical inference for selected parameters: e.g. consider a microarray experiment in which the goal is to discover differentially expressed genes, and then to construct a valid marginal confidence interval for each discovered gene. The solution suggested in Benjamini and Yekutieli (2005a) is control over the False Coverage statement Rate (FCR) – a generalization of the FDR, defined as the expected proportion of non-covering confidence intervals out of the total number of confidence intervals. For independently distributed parameter estimators Benjamini and Yekutieli (2005a) prove that the FCR of $1 - Rq/m$ marginal confidence intervals for a subset of R parameters chosen out of m parameters is less than or equal to q. Benjamini and Yekutieli (2005a) also show that the BH procedure is simply a particular example of selective inference, and provide a coherent framework for providing selective inference which includes the discovery of non-zero parameters, assigning a sign to each non-zero parameter, and constructing valid confidence intervals for each non-zero parameter.

In his 2001 NSF conference talks at Temple University, Yosef Hochberg identified the problem of Multiple Comparisons with Selective and Simultaneous Inference (he attributed this point of view to earlier work by Joseph Putter – reference unavailable). Benjamini and Yekutieli (2005a) argue that the problem of selective inference and the simultaneity problem are two distinct problems encountered when trying to provide inference for multiple parameters. The simultaneity problem is directly caused by multiplicity, it is the need to provide marginal inference that applies to all the parameters – e.g. marginal confidence intervals that cover all the parameters with probability 0.95, and the solution to this problem is FWE adjusted inference. Selective inference can occur when

considering a single parameter, however it is more common when initially considering multiple parameters. The construction of level q FCR adjusted confidence intervals usually implies that approximately $1-q$ of the confidence intervals cover their respective parameters; thus the probability that a FCR adjusted confidence interval covers a selected parameter is $\approx 1-q$.

3 Hierarchical FDR procedures

Existing FDR methodology can only be applied to test a single family of hypotheses. Yekutieli (2007) introduces a hierarchical testing approach that controls the FDR in more general settings – complex large-scale studies which involve considering multiple families of tested hypotheses. In this testing approach the tested hypotheses are arranged in a tree of disjoint subfamilies, and the subfamilies of hypotheses are hierarchically tested by the BH procedure. Yekutieli (2007) derives an approximation for the multiple family FDR for independently distributed test statistics: q – the level at which the BH procedure is applied, times the number of families tested plus the number of discoveries, divided by the number of discoveries plus 1; it is also shown that hierarchical testing inherently implies FDR control, thus the universal bound for the FDR in the new hierarchical testing approach is $2 \times 1.44 \times q$.

The BH procedure has been successfully applied to provide adaptive scalable thresholding (the threshold is determined by the proportion of discoveries in the data, not the number of tested hypotheses) which ensures reproducible results (approximately $1-q$ of the discoveries are true discoveries); the BH procedure has also been applied in model selection to produce models with small estimation error (Abramovich and Benjamini, 1996). Yekutieli (2007) explains that if the data analyzed has an hierarchical structure, the hierarchical FDR procedure passes over the families of tests with no discoveries and adaptively tests the families with the large proportion of discoveries. Yekutieli (2007) shows through simulations that hierarchical FDR procedures control the FDR and can yield substantially more discoveries than the BH procedure, and used in model selection it produces models with more terms and smaller MSE.

In general, I think the fact that the hierarchical approach yields more FDR-controlled discoveries points to a more important property – selective inference based on hierarchical testing has smaller estimation error. Recall that the main message in Benjamini and Yekutieli (2005a) is that selection biases and corrupts estimation; thus by reducing the dimension of the testing problem, and the

selection, the hierarchical testing produces better estimation. To illustrate this point consider the idealized example of families of m null hypotheses in a tree with L level: assume that there are two types of families – (a) families made of m true null hypotheses and (b) families in which π_1 of the hypotheses are false null hypotheses; the first level of the tree is a type (b) family; for Levels 1 through $L-1$, true null hypotheses are parent to type (a) families following level of the tree and false null hypotheses are parent to type (b) families on the following level of the tree. Thus the entire testing problem includes almost m^L hypotheses and the proportion of false null hypotheses is slightly more than π_1^L. On the other hand, the hierarchical approach essentially involves separately applying the BH procedure to the type (b) families – i.e. testing families of L hypotheses in which π_1 of the null hypotheses are false – logarithmically decreasing the dimension of the problem and the extent of selection.

The microarray analysis in Yekutieli et al. (2006) is the first application of the hierarchical FDR approach. The data includes microarrays measuring the expression levels of $27,000$ genes in two replicate assays taken from five areas in the brains of mice from ten inbred mice strains. A two-way ANOVA, with strain and brain region main effects, is fitted for each gene in order to identify genes with strain expression differences. The researchers are also interested in testing the interaction terms to locate areas in the brains and strains with abnormal expression levels. In the standard FDR approach the two questions are addressed separately: applied at level 0.05 to test the strain effect for the $27,000$ genes the BH procedure yielded 957 discoveries; however, there were no discoveries when the 0.05 BH procedure was applied to test the 1.35 million interaction terms. In the hierarchical approach the discovery of genes with significant strain effects is considered as the initial question for each gene, and localizing the effect to specific strains and brain regions are considered follow-up questions for genes with significant strain effects. Separately applying the 0.05 BH procedure in each of the 957 families of interactions yields 170 interaction discoveries.

The hierarchical FDR approach has also been used to select log-linear model in a Behavioral Genetic application: Kafkafi et al. (2007) present an algorithm for discovering behavior patterns that differentiate between mutant and wild-type rats. A filmed session of exploratory is coded into a series of nine behavioral relevant endpoints (Drai and Golani, 2001), which are discretized and summarized in a 9-way contingency table; the algorithm scans these immense, sparse, contingency tables for patterns with significant frequency differences between

mutant and wild-type rats, and the main challenge is to work at the highest resolution level and still avoid over-fitting. In the hierarchical approach this problem is overcome by using a hierarchical FDR procedure to fit a log-linear model to the mutant rat and wild-type rat contingency tables (Maltser, 2006): at the first stage the 0.05 BH procedure is applied to test the terms in the main effects model; at the second stage the 0.05 BH procedure is separately applied to test two-way interactions with each of the significant main effects from the first stage; in the third stage, the 0.05 BH procedure is separately applied to test three-way interactions with each of the significant two-way interaction; and so on. The tables are then scanned for the behavior patterns with the largest differences between the fitted values of the log-linear models.

Dr. Sharan and I are currently developing a hierarchical FDR procedure for the discovery of over-represented Gene-Ontologies. The Gene Ontology (GO) database (Ashburner et al., 2000) contains the hierarchy of current GO terms, and links between genes and the associated GOs that define their function. Scanning the database for GO terms which are over-represented in a list of differentially expressed genes from a microarray experiment (Beissbarth and Speed, 2004) is widely used for understanding biological processes. In our new approach the BH procedure is applied to test for over-representation in the high-level ontologies; and at later stages the BH is repeatedly applied to test for the conditional over-representation in families of terms corresponding to over-represented parent terms.

My current study of the hierarchical FDR approach is focused on three questions: (1) the results in Yekutieli (2007) were derived for independently distributed test statistics. In most applications the test statistics are dependent, thus it is very important to study the effect of dependency on the FDR of hierarchical FDR procedures. the expression for the hierarchical derived in Yekutieli (2007) is a multiplicative factor (shown to be < 1.44) times the aggregate FDR computed across families. The results in Benjamini and Yekutieli (2001) indicate that the FDR is controlled when the BH procedure is applied to test a single family of dependent test statistics, however my preliminary results suggest that dependency may affect the aggregation of the FDR and also the value of the multiplicative factor. (2) At the beginning of this section I speculated that estimation via hierarchical FDR may yield small estimation error. I aim to study the estimation error of selective inference based on hierarchical FDR testing and compare it with selective inference based on the BH procedure. Note that here we consider marginal estimation error for selected parameter, whereas in the

application of FDR for model selection (Abramovich et al., 2006) it is common to consider the estimation error for the m-dimensional parameter (computed in a l_p norm). (3) A related question is the construction of valid confidence intervals for hierarchical FDR discoveries. Note that the method suggested in Benjamini and Yekutieli (2005a) cannot be applied, because in the hierarchical FDR setting it is not clear how to define m – the number of tested parameters and R – the number of selected parameters. Furthermore, in the hierarchical approach we are often interested in confidence intervals for functions of several discoveries (e.g. the fitted values in the log-linear models are sums of model parameters estimated in the hierarchical log-linear model selection).

4 Bayesian adjusted inference for selected parameters

My goal in this research is to develop methodology for providing better selective inference. Control over the FCR is a frequentist mechanism for providing valid selective inference: it measures the strength of selection by the proportion of selected parameters – R/m, and adjusts the statistical inference accordingly. However it has two major limitations: (a) it attributes the same strength of selection to all the selected parameters – clearly there is no need to adjust for selection a parameter that is always selected; (b) the mode of adjustment is only determined by the type of statistical inference given (the same confidence interval but with a larger confidence level) and not by the selection criterion, whereas Benjamini and Yekutieli (2005a) show that adjustment is not necessary for some types of selection. Note that existing Bayesian FDR methodology (Storey, 2002; Storey, 2003; Efron et al., 2001) offers the same type of statistical inference as the BH procedure – thus, implicitly, it suffers from limitations (a) and (b). In my new research I will try to generalize Storey's Bayesian FDR approach, and I think that the key to the problem is to answer a question that has puzzled me for some time: is Bayesian adjustment for selected inference necessary?

References

[1] Abramovich F., Benjamini Y., (1996). "Adaptive thresholding of wavelet coefficients", *Computational Statistics and Data Analysis*; **22**, 351-361.

[2] Abramovich F., Benjamini Y., Donoho D.L., Johnstone I.M. (2006) "Adapting to unknown sparsity by controlling the False Discovery Rate", *The Annals of Statistics* **34**, 584-653.

[3] Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Davis, A., Dolinski, K., Dwight, S., Eppig, J., Harris, M., Hill, D., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J., Richardson, J., Ringwald, M., Rubin, G., Sherlock, G. (2000), "Gene Ontology: tool for the unification of biology" *Nature Genetics* **25** 2529

[4] Benjamini Y., Hochberg Y., (1995). "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing", *Journal of the Royal Statistical Society B*, **57**, 289-300.

[5] Benjamini, Y., Krieger, A.M., Yekutieli, D. (2006) "Adaptive Linear Step-up False Discovery Rate controlling procedures" *Biometrika* (3): 491-507.

[6] Benjamini Y., Yekutieli D., (2001) "The Control of the False Discovery Rate in Multiple Testing under Dependency" , *The Annals of Statistics*, **29**, 4, 1165-1188.

[7] Benjamini Y., Yekutieli D., (2005a) "False Discovery Rate-Adjusted Multiple Confidence Intervals for Selected Parameters" *Journal of the American Statistical Association*, **100**, 71.

[8] Benjamini Y., Yekutieli D., (2005b) "Quantitative Trait Loci Analysis using the False Discovery Rate", *Genetics*, **171**, 783-790.

[9] Beissbarth T., Speed T.P., (2004) "GOstat: find statistically overrepresented Gene Ontologies within a group of genes", *Bioinformatics* **20**, 1464-1465.

[10] Drai, D., Golani, I. (2001), "SEE: a tool for the visualization and analysis of rodent exploratory behavior", *Neurosci Biobehav Rev.*, **25** (5), 409-26.

[11] Efron, B., Tibshirani, R., Storey, J. D., Tusher, V. (2001) "Empirical Bayes Analysis of a Microarray Experiment", *Journal of the American Statistical Association*, **96**, 1151-1160.

[12] Genovese C., Wasserman L., (2004) "A Stochastic Process Approach to False Discovery Control", *The Annals of Statistics*, **32** (3) 1035-1061.

[13] Gittelman E., (2007) "Use of Bootstrapping to construct Adaptive Linear Step-up False Discovery Rate controlling procedures" M.Sc. Thesis in preparation , Department of Statistics and O.R., Tel Aviv University, Tel Aviv, Israel.

[14] Kafkafi N., Yekutieli D., Yarowsky P., Elmer G. I., (2007) "Pattern Array: a Novel Approach for Data Mining in Behavioral Tests", Submitted to *BMC Bioinformatics*.

[15] Maltser, A., (2006) "Log-Linear Model Selection Using Hierarchical False Discovery Rate controlling methodology" M.Sc. Thesis, Department of Statistics and O.R., Tel Aviv University, Tel Aviv, Israel.

[16] Reiner A., Yekutieli D., Benjamini Y., (2003) "Identifying differentially expressed genes using false discovery rate controlling procedures" *Bioinformatics*, **19**, 368-75.

[17] Sarkar, S. K., Chang, C. K., (1997) "The Simes method for Multiple Hypotheses Testing with Positively Dependent Test Statistics" *Journal of the American Statistical Association*, **92**, 1601-1608.

[18] Sarkar, S. K., (1998) "Some probability inequalities for ordered MTP_2 random variables: A proof of Simes' conjecture" *The Annals of Statistics*, **26**(2), 494-504.

[19] Storey J. D. (2002) "A direct approach to false discovery rates", *Journal of the Royal Statistical Society: Series B*, **64** 479-498.

[20] Storey J. D., (2003) "The positive false discovery rate: A Bayesian interpretation and the q-value", *The Annals of Statistics*, **31**, 2013-2035.

[21] Storey J. D., Taylor, J. E., and Siegmund D. (2004) "Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach", *Journal of the Royal Statistical Society: Series B*, **66**, 187-205(19).

[22] Yekutieli D., (2006) "False Discovery Rate control for non positively regression dependent test statistics", To appear in the special issue of the Journal of Statistical Planning and Infernce on Statistical Design of Medical Experiments III.

[23] Yekutieli D., (2007) "Hierarchical False Discovery Rate controlling methodology", under revision for Journal of the American Statistical Association.

[24] Yekutieli D., Benjamini Y. (1999), "Resampling based false discovery rate controlling procedure for dependent test statistics", *Journal of Statistical Planning and Inference*, **82** (1-2), 171-196.

[25] Yekutieli D., Reiner A., Elmer G. I., Kafkafi N., Letwin N. E., Lee N. H., Benjamini Y. (2006) "Approaches to multiplicity issues in complex research in microarray analysis", *Statistica Neerlandica*, **60** nr. 4, 414-437.