BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 55, Number 1, January 2018, Pages 97–102
http://dx.doi.org/10.1090/bull/1602
Article electronically published on October 23, 2017

SELECTED MATHEMATICAL REVIEWS

related to the paper in the previous section by
SUSAN HOLMES

MR1263027 (95d:01025) 01A75; 62J15

Tukey, John W.

The collected works of John W. Tukey. Vol. VIII. (English)

With a preface by William S. Cleveland. With a biography by Frederick Mosteller.
Chapman & Hall, *New York*, 1994, lxii+475+i10 pp., US$62.95,
ISBN 0-412-05121-4

The volume contains writings by Tukey on the problem of making simultaneous inferences about a set of parameter values from a single experiment. Most of the book consists of an unpublished monograph, dated 1953. In addition to other work from this period, there are also three publications from 1965 and one from 1983. Topics addressed include: possible definitions of the error rate and reasons for preferring the one that is now generally accepted, relating to a statement about a complete batch of values or family of comparisons; designing experiments to have prescribed accuracy; budgeting of the error rate; short-cut procedures based on ranges; procedures in which the decision about a pair of means is influenced by the configuration of other means. General comments are also made on the theory and practice of statistics, and issues raised include: the rôle of distributional assumptions in theoretical and applied statistics; the tension between tightness of statement and sensitivity to nonnormality; situations in which problems would be more relevantly addressed in terms of “confidence” rather than “significance”. Extensive tables are included.

{Volumes I–VII have been reviewed [MR 86h:62134; MR 87b:62125; MR 88d:62003a; MR 88d:62003b; MR 89b:01080; MR 92a:01086; MR 93a:01068].}

A. D. Gordon

From MathSciNet, October 2017

MR1869245 (2002i:62135) 62J15; 47N30, 62H20

Benjamini, Yoav; Yekutieli, Daniel

The control of the false discovery rate in multiple testing under dependency.

The Annals of Statistics **29** (2001), *no.* 4, 1165–1188.

The common approach to multiple testing problems is to control the familywise error rate (FWER). Pointing out a few faults in this approach, the first author and Y. Hochberg [J. Roy. Statist. Soc. Ser. B **57** (1995), no. 1, 289–300; MR1325392] suggested that the expected proportion of falsely rejected hypotheses, named the false discovery rate (FDR), may be the appropriate error rate to control in many applied problems. They gave a simple FDR controlling procedure for independent statistics and showed it to be more powerful than comparable procedures that control FWER. In the paper under review, the authors show that this same procedure also controls the FDR when the test statistics have positive regression dependency on each of the test statistics corresponding to the true null hypotheses. This situation covers many problems of practical interest such as the comparisons of several treatments with a single control, multivariate normal test statistics with positive correlation matrix and multivariate t. The authors also show that, without posing special difficulties, issues such as discrete test statistics, composite null hypotheses, general step-up procedures and general dependency can be addressed.

S. Panchapakesan
From MathSciNet, October 2017

MR2065195 (2005e:62066) 62G10; 62G20, 62G32

Donoho, David; Jin, Jiashun

Higher criticism for detecting sparse heterogeneous mixtures.

The Annals of Statistics **32** (2004), *no.* 3, 962–994.

Consider n independent tests of unrelated hypotheses with test statistics $X_1, \dots,$ X_n. Under the null hypothesis H_0, the X_i are i.i.d. with a certain distribution D_0, while under the alternative hypothesis H_1, the X_i are i.i.d. with a distribution $(1-\epsilon)D_0 + \epsilon D_1$ for $1 \leq i \leq n$ i.e. under H_1 a small fraction ϵ, $(0 < \epsilon < 1)$ of the data comes from another distribution D_1. This multiple comparison problem for testing the global intersection null hypothesis against a specific element in its complement has been called higher criticism, or second-level significance testing. Tukey suggested for this kind of problem a standardized statistic (z-score). Here, this approach is generalized by maximizing the z-score over a range of significance levels $0 < \alpha \leq \alpha_0$. It turns out that there exists a detection boundary which denotes a demarcation which describes how big the nonzero effect must be to be detectable as a function of the rarity of nonzero effects. The new approach is compared to other multiple comparison procedures (studentized range, maximum statistic, Bonferroni, false-discovery-rate controlling methods, etc.).

Joachim Krauth
From MathSciNet, October 2017

MR2329475 (2008j:62060) 62H12; 62C10, 62C12, 62C20, 62F12, 62G20

Donoho, David; Jin, Jiashun

Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data.

The Annals of Statistics **34** (2006), *no.* 6, 2980–3018.

The authors consider a vector of measurements whose coordinates are independently exponentially distributed with individual means. The vector of means is thought to be sparse, with most coordinates equal to one, but with a small fraction significantly larger than one. Thus, most coordinates can be considered as simply ‘noise’, but a small fraction contain ‘signal’. Such situations arise in several areas of application.

One example is failure times of many comparable independent systems where a small fraction of the systems have significantly higher expected lifetimes than the typical system. However, the size of the proportion of these systems with long lifetimes, and which they are, are unknown.

Another example is testing of many independent statistical hypotheses, each yielding its separate p-value, and that the vast majority of these tests correspond to cases where the null hypothesis is true, i.e. the mean value is equal to one, in the case considered by the authors, while a few are false, i.e., the mean value is greater than one, but it is unknown which these are.

The present paper is heavily based on a recent paper by F. P. Abramovich et al. [Ann. Statist. **34** (2006), no. 2, 584–653; MR2281879], in which similar problems were addressed for vectors of normally distributed variables. In that paper a methodological structure to treat problems of the actual kind was developed, and the relevant concepts and methodologies were discussed and explained in detail and compared with other available methods, in all cases with reference to normally distributed variables. The aim of the present paper is to adapt this methodology to exponentially distributed variables.

A basic concept is sparsity, which was defined in several ways by Abramovich et al. The present authors utilize a definition in which sparsity is characterized by two parameters.

As their main technique the authors use control of the false discovery rate (FDR), a principle for design of simultaneous testing procedures developed by Y. Benjamini and Y. Hochberg [J. Roy. Statist. Soc. Ser. B **57** (1995), no. 1, 289–300; MR1325392]. In a setting where one is testing many hypotheses, the principle imposes control on the expected ratio of the number of erroneously rejected hypotheses to the total number of rejected hypotheses by demanding that this expected ratio be less than or equal to a given value q. They construct an FDR-thresholding estimator of the mean vector. Those components for which the observed value exceeds a data-dependent threshold depending on the FDR control parameter q keep their values; otherwise, the value of the component is put equal to one.

It is shown that this FDR-thresholding estimator is asymptotically minimax if q is less than 0.5. To reach this result the authors utilize an adaptive minimaxity technique developed by Abramovich et al.

A couple of important points are illustrated graphically.

Jon Stene
From MathSciNet, October 2017

MR2752616 62J15; 62P10

Hu, James X.; Zhao, Hongyu; Zhou, Harrison H.

False discovery rate control with groups.

Journal of the American Statistical Association **105** (2010), *no.* 491, 1215–1227.

The article contributes to the theory of the false discovery rate (FDR) in multiple statistical hypothesis testing. It deals with a method to enhance the power of the classical linear step-up (LSU) test [cf. Y. Benjamini and Y. Hochberg, J. Roy. Statist. Soc. Ser. B **57** (1995), no. 1, 289–300; MR1325392] for FDR control under positive dependency in the case that the hypotheses to be tested can be divided into disjoint groups before testing starts. This is especially relevant for applications in genetics, where genes can typically be grouped according to external information, e.g., from ontology databases.

The authors introduce a weighting scheme that assigns group-specific weights to marginal p-values. These weights depend on the proportion of true null hypotheses in the corresponding group and are first motivated in an oracle setting with known proportions by Bayesian arguments. In the oracle case, the authors' weighting scheme allows one to carry out the LSU test with weighted p-values at the relaxed level $\alpha/(1-\pi_0)$, where π_0 is the overall proportion of true nulls, while still strongly controlling the FDR at level α. Under mild assumptions, this can lead to an improved power (in terms of the expected number of rejections) of the resulting multiple test compared with the LSU procedure.

In a second theoretical part, the authors are concerned with the problem of pre-estimating unknown group-specific proportions of true nulls and briefly review several approaches towards that from the literature. Two conservative procedures (on average overestimating those proportions) are then picked and utilized as a preceding step of the authors' weighting procedure. It turns out that for these conservative estimation approaches the weighting procedure with pre-estimation still controls the FDR level asymptotically.

By means of computer simulations and by evaluating real-life data sets from a breast cancer application, the authors finally conclude the superiority of their method over the classical LSU test if there are marked differences in proportions of true null hypotheses among groups.

Thorsten Dickhaus
From MathSciNet, October 2017

MR3469131 62A01

Häggström, Olle

Why the empirical sciences need statistics so desperately.

European Congress of Mathematics, 347–360, *European Mathematical Society*, 2013.

In this well-informed and thought-provoking appeal, the author makes the case that the empirical sciences need the involvement of formally trained statisticians to correct the cognitive biases that plague scientists and other mortals. Scientists desperately need statistics to save them from over-interpreting their data, from overconfidence in conclusions, from using statistical methods without understanding their assumptions, from focusing on statistical significance at the expense of estimated effect sizes, and from other practices resulting from ignorance of the statistical literature. Such practices lead to faulty scientific conclusions, unnecessarily inhibiting scientific progress.

Statistics graduates should intervene not only through education but also through collaboration. That will be much more effective when scientists come to realize many of their "self-taught" authorities on statistics cannot rescue them from their desperate plight.

The author especially laments that so many scientists still interpret a p value as a posterior probability that the null hypothesis is true given the data. He nevertheless approves of Fisher's argument that since a low p value indicates either that an improbable event occurred or that the null hypothesis is false, the p value quantifies evidence against the null hypothesis (pp. 355–356). That measure of evidence conveniently does not require considering prior distributions, the choice of which is called "the Achilles heel of Bayesian statistics" (p. 357).

Since a null hypothesis would be rejected given a sufficiently low p value or given a sufficiently low posterior probability of the null hypothesis, could it be that continued advocacy for Fisher's argument fuels widespread misinterpretation of the p value? That would help explain the failure of sustained educational efforts described by the American Statistical Association: "Statisticians and others have been sounding the alarm about these matters for decades, to little avail" [R. L. Wasserstein and N. A. Lazar, Amer. Statist. **70** (2016), no. 2, 129–133; MR3511040; see also D. R. Cox, *Principles of statistical inference*, Cambridge Univ. Press, Cambridge, 2006 (Section 3.7); MR2278763]. For in Bayesian thinking, to reject a null hypothesis without regard to its posterior probability is to do so without regard for whether one doubts it enough to bet against it. Such a hypothesis rejection then would mean something other than betting against it, much as acceptance can differ from belief [L. J. Cohen, *An essay on belief and acceptance*, Clarendon Press, Oxford, 1992]. In any case, widespread disagreement among mainstream statisticians on the value of Fisher's argument [see R. M. Royall, *Statistical evidence*, Monogr. Statist. Appl. Probab., 71, Chapman & Hall, London, 1997 (Section 3.3); MR1629481] may indicate that cognitive biases afflict more than the empirical sciences.

David R. Bickel
From MathSciNet, October 2017

MR3454203 62G10; 62J05, 62J15

Grazier G'Sell, Max; Wager, Stefan; Chouldechova, Alexandra; Tibshirani, Robert

Sequential selection procedures and false discovery rate control.

Journal of the Royal Statistical Society. Series B. Statistical Methodology **78** (2016), *no.* 2, 423–444.

New procedures for multiple hypothesis testing that guarantee the given level α for the false discovery rate (FDR) are proposed. This is motivated by the sequential model selection problems for linear regression models with a large number of predictors. The problem of providing FDR control in regression models has been well studied before [see, for example, Y. Wu, D. D. Boos and L. A. Stefanski, J. Amer. Statist. Assoc. **102** (2007), no. 477, 235–243; MR2345541; Y. Benjamini and Y. Gavrilov, Ann. Appl. Stat. **3** (2009), no. 1, 179–198; MR2668704; N. Meinshausen and P. Bühlmann, J. R. Stat. Soc. Ser. B Stat. Methodol. **72** (2010), no. 4, 417–473; MR2758523; D. Lin, D. P. Foster and L. H. Ungar, J. Amer. Statist. Assoc. **106** (2011), no. 493, 232–247; MR2816717]. Many other examples can be found in the references of the last article just cited.

In the paper under review, the authors thoroughly study parallels between the procedures of multiple testing with FDR control (in the spirit of Benjamini and Y. Hochberg [J. Roy. Statist. Soc. Ser. B **57** (1995), no. 1, 289–300; MR1325392]) and the procedures of the variable selection using a path-based regression algorithm of a type of forward stepwise regression [see R. R. Hocking, Biometrics **32** (1976), no. 1, 1–49; MR0398008] and least-angle regression [B. Efron et al., Ann. Statist. **32** (2004), no. 2, 407–499; MR2060166]. In such problems a statistician deals with a sequence of p-values: $p_1, \dots, p_m \in [0,1]$, corresponding to some hypotheses $H_1, \dots, H_m$. If all the hypotheses are true, these p-values form a sequence of i.i.d. random variables with the uniform distribution on $[0,1]$. The statistical problem consists of a construction of a selection procedure of a set of hypotheses $H_{i_1}, \dots, H_{i_k}$ which are declared to be false, but the mean value of relative proportion of the true hypotheses in this set (FDR) should not exceed the given level α. The typical procedure of solving the problem of multiple testing is in the increasing reordering of p-values $p_{(1)} \le \cdots \le p_{(m)}$ and defining number $\widehat{k}$ such that all hypotheses

$H_{(1)}, \ldots, H_{(\widehat{k})}$ corresponding to the reordered p-values are declared to be false and the FDR is controlled at the given level α.

Two procedures of the multiple testing are suggested: ForwardStop with

$$\widehat{k}_F = \max\left\{k \in \{1, \ldots, m\} : \ -\frac{1}{k}\sum_{i=1}^{k} \log(1 - p_{(i)}) \leq \alpha\right\}$$

and StrongStop with

$$\widehat{k}_S = \max\left\{k \in \{1, \ldots, m\} : \ \exp\left\{\sum_{j=k}^{m} \frac{\log p_{(j)}}{j}\right\} \leq \frac{\alpha k}{m}\right\}.$$

It is proved that both of these procedures control FDR at the given level α and StrongStop also controls FWER (the probability that a procedure makes even a single false discovery). A more specialized version of StrongStop is developed which takes advantage of special properties of some of the proposed decision rules. An extensive investigation of the power properties of the suggested tests by the results of statistical modeling is provided and applications to real data are also given.

I. N. Volodin

From MathSciNet, October 2017