Electronic Journal of Statistics
Vol. 9 (2015) 1744–1748
ISSN: 1935-7524
DOI: 10.1214/15-EJS1054REJ

Rejoinder of "Hypothesis testing by convex optimization"[*]

Alexander Goldenshluger[†]

Department of Statistics, University of Haifa, 31905 Haifa, Israel
e-mail: goldensh@stat.haifa.ac.il

Anatoli Juditsky[‡]

LJK, Université Grenoble Alpes, B.P. 53, 38041 Grenoble Cedex 9, France
e-mail: anatoli.juditsky@imag.fr

and

Arkadi Nemirovski[§]

Georgia Institute of Technology, Atlanta, Georgia 30332, USA,
e-mail: nemirovs@isye.gatech.edu

MSC 2010 subject classifications: Primary 62G10, 62C20; secondary 62M02, 62M10, 90C25, 65K10.
Keywords and phrases: Hypothesis testing, nonparametric testing, composite hypothesis testing, statistical applications of convex optimization.

First of all, we would like to thank all the discussants for their interesting, thought–provoking comments and thorough investigation. We also thank the editors for the opportunity to comment briefly on a few issues raised in the discussions.

The comments of the discussants underline importance of the topic discussed in our paper, namely, that of application of convex optimization methodology to statistical inference problems. Of special interest is the diversity of perspectives, which include theoretical and practical issues. Before addressing comments of the discussants, we would like to restate as simply as possible the main point of this paper.

Our approach. In the nutshell, our approach is as follows. Let $\mathcal{P}$, $\mathcal{Q}$ be two families of probability distributions on an observation space Ω. Given observation $\omega \sim P \in \mathcal{P} \cup \mathcal{Q}$, our goal is to decide on the hypotheses H_1, H_2 stating

[*] Main article 10.1214/15-EJS1054.

[†] Research of this author was supported by grants BSF 2010466, and ISF 104/11.

[‡] This author was supported by the CNRS-Mastodons project Titan and the LabEx PERSYVAL-Lab (ANR-11-LABX-0025).

[§] Research of this author was supported by NSF grants CMMI-1232623, CMMI-1262063, CCF-1415498, CCF-1523768.

that $P \in \mathcal{P}$ and $P \in \mathcal{Q}$, respectively. Consider the optimization problem

$$\begin{aligned}\epsilon_* &= \min_{\phi(\cdot),\epsilon}\left\{\epsilon : \sup_{P\in\mathcal{P}}\int_\Omega \exp\{-\phi(\omega)\}P(d\omega) \le \epsilon,\ \sup_{P\in\mathcal{Q}}\int_\Omega \exp\{\phi(\omega)\}P(d\omega) \le \epsilon\right\}\\ &= \min_{\phi(\cdot),\epsilon}\left\{\epsilon : \sup_{P\in\bar{\mathcal{P}}}\int_\Omega \exp\{-\phi(\omega)\}P(d\omega) \le \epsilon,\ \sup_{P\in\bar{\mathcal{Q}}}\int_\Omega \exp\{\phi(\omega)\}P(d\omega) \le \epsilon\right\},\end{aligned} \tag{1}$$

where $\bar{\mathcal{P}}$ and $\bar{\mathcal{Q}}$ are convex hulls of $\mathcal{P}$ and $\mathcal{Q}$.

Then

- it is immediately seen that a feasible solution $(\phi(\cdot), \epsilon)$ to the problem induces a test deciding on H_1, H_2 with risk $\le \epsilon$; given observation ω, this test accepts H_1 when $\phi(\omega) \ge 0$, and accepts H_2 otherwise;
- the best risk ϵ_* achievable with this approach is not too far from the "ideal" risk: if "in the nature" there exists a (perhaps, randomized) test deciding on H_1, H_2 with risk $\delta < 1/2$, then (1) admits a feasible solution $(\bar{\phi}(\cdot), \bar{\epsilon})$ with $\bar{\epsilon} \le 2\sqrt{\delta(1-\delta)} < 1$. Moreover, for every $K = 1, 2, ...,$ the risk of the test which, given an i.i.d. sample $\omega_t \sim P \in \mathcal{P} \cup \mathcal{Q}$, $1 \le t \le K$, accepts H_1 when $\sum_{t=1}^K \bar{\phi}(\omega_t) \ge 0$, and accepts H_2 otherwise, does not exceed $\bar{\epsilon}^K$. As a result, if the ideal risk δ_* of deciding on H_1, H_2 via a single observation is, say, ≤ 0.016, and repeated observations are allowed, then the accuracy of the test given by a (near)-optimal solution to (1) and based on the sample size $K = 3$ is as good as the one of the ideal single–observation test.

The bottom line is that an optimal, or nearly so, solution to (1) induces a test with attractive near-optimality properties. Although this fact is well known for more than thirty years (it can be traced back to [1] and [3], cf. Lucien Birgé's discussion), it is of limited "practical value" by itself. Unfortunately, even though (1) is a convex program, it is typically *computationally intractable*: it is infinite-dimensional, and the constraints are, in general, difficult to compute. Furthermore, in this respect, (1) is not different from the infinite-dimensional convex optimization problem which is responsible for the ideal (randomized) test, that is, the problem

$$\delta_* = \min_{\psi(\cdot),\epsilon}\left\{\epsilon : \sup_{P\in\bar{\mathcal{P}}}\int_\Omega (1-\psi(\omega))P(d\omega) \le \epsilon,\ \sup_{P\in\bar{\mathcal{Q}}}\int_\Omega \psi(\omega)P(d\omega) \le \epsilon,\ 0 \le \psi(\cdot) \le 1\right\}. \tag{2}$$

The principal observation underlying all other developments in our paper is that in several special cases (1) becomes computationally tractable, namely, in the cases of (stationary K-repeated) Gaussian, Poisson, and Discrete observation schemes with "convex" hypotheses H_1, H_2. The convexity here means that $\mathcal{P}$, $\mathcal{Q}$ are generated by *convex* sets in the spaces of parameters of the corresponding distributions.[1]

[1] In contrast, the only known to us cases when the problem (2) responsible for *exactly*

One of our principal objectives in this work was to demonstrate that already this (restricted) statistical framework, "augmented" with some matrix calculus, encompasses several classical applications. In particular, our attention was attracted to inverse problems, where the proposed approach leads to a "universal" problem treatment – in order to build a near-optimal testing procedure one should solve a certain optimization problem (which admits a numerically efficient solution). For instance, in Gaussian o.s. $\omega = Ax+\xi$, where $\xi \sim \mathcal{N}(0,I)$, the tests are "tuned" precisely for the problem matrix A; in the case of indirect observation $\omega \sim P_\mu,\ \mu = Ax$ in Discrete o.s., the exact structure of A (which may describe noisy observations, censored observations, their composition, etc) is taken into account "automatically", and so on. We also believe that this approach to testing is also important for "real" problems: in many cases its direct application leads to testing procedures with reasonable (that is, nearly the best possible) "practical performance". Last but not least, it can be easily implemented and tested using widely available optimization tools, such as CVX [2].

Discussion. We would like to thank Lucien Birgé who more than thirty years ago made fundamental contributions to development of the approach presented in our paper. He is certainly the best person to put this work into the right historical perspective, and he has done this with sparkle in his thought–provoking discussion. In his work Lucien Birgé studied various consequences of the aforementioned theoretical result. He also pioneered different weighting schemes for test aggregations, in particular those arising in the problem of nonparametric estimation through multiple testing.

An important point which is recurrent in several discussions (Alekh Agarwal; Fabienne Comte, Céline Duval and Valentine Genon–Catalot; Axel Munk and Frank Werner, and Philippe Rigollet) is related to principal limitations of the proposed approach. In our opinion, extending the testing framework to the case where the requirements of good o.s. are not satisfied is of high interest. The following questions appear essential in this regard:

1. can the proposed construction be extended to encompass a larger variety of distribution families?
2. is it possible to extend the discussed framework beyond testing of unions of "not very large" number of convex hypotheses?

The question which is subsidiary to question 1) above is that of existence of good o.s. other than Gaussian, Poisson and Discrete one, described in section

optimal test is tractable are the cases of 1) Gaussian o.s. with convex hypotheses H_1, H_2, where the exactly optimal test, "by chance," is given by an optimal solution to (1), and 2) Discrete o.s. $\emptyset$ with convex hypotheses H_1, H_2. The latter case is of very limited consequences, since usually one is interested in the case of repeated observations, that is, in the case of K-repeated o.s. $\emptyset^K$ as defined in section 2.4.1 of the paper. On the other hand, problem (2) associated with $\emptyset^K$, $K \geq 2$, is usually intractable, the difficulties stemming from the necessity to describe in a computation-friendly manner the convex hulls of *direct powers* of probability distributions from $\mathcal{P}$ and $\mathcal{Q}$, not speaking about exponential growth with K of the cardinality of observation space associated with $\emptyset^K$ (and thus – of the design dimension of (2)).

2.3 of the paper. The simple answer is "we do not know", but, this being said, the main result (Theorem 2.1) allows for some extensions. For instance, we currently work on extending Discrete o.s. to the case where $\mathcal{P}$ and $\mathcal{Q}$ are families of continuous distributions P, given by bounds on a finite number of linear functionals of P (i.e., on expectations of some vector-valued functions $\Psi(\omega)$). In the latter case the optimal solution ϕ_* to the optimization problem in (1) admits an alternative description in terms of the problem dual to (1), and under favorable circumstances it can be computed efficiently. An alternative approach uses quadratic approximation of the functional

$$\int_{\Omega} \exp\{-\phi(\omega)\} P(d\omega)$$

and under analogous conditions allows for constructing efficiently computable tests (a similar approach is described in Alekh Agarwal's discussion).

Some directions to follow in order to answer question 2) above are given in discussions by Munk and Werner, and Rigollet. Axel Munk and Frank Werner consider the problem of testing multiple change–points in the nonparametric regression model. Note that direct application of our approach to this problem is problematic for at least two reasons. First of all, the complexity of the testing becomes prohibitive as the number of hypotheses to test grows exponentially with the number of signal jumps. Furthermore, the pairwise detectors corresponding to testing of different signals are strongly correlated. Meanwhile, the calculus of tests, given in section 3 of the paper does not take this possibility into account, and the resulting detection boundaries become suboptimal. Philippe Rigollet discusses testing a sparsity pattern of the normal mean vector. The basic idea here is to use convex relaxations in order to construct efficiently computable tests in the case where the hypotheses are not associated with compact convex sets. Usually, analysis of accuracy of a statistical procedure obtained in this way is a difficult task, and one cannot expect here any "universal results", as those described in our paper. For instance, quality of the resulting decision rules depends heavily on the chosen relaxation of the non-convex constraints. However, in some specific cases such approximations lead to near-optimal statistical procedures. In particular, this is the case where the non-convex component of the problem is due to sparsity constraint on the signals, which can be approximated by a union of "not very large" number of convex sets.

Arnak Dalalyan in his discussion presents interesting connections between our approach to testing problems and a widely adopted approach to classification in Learnig Theory. His revealing comments shed some additional light on construction of tests used in our paper. We have also implemented some changes in the paper following comments of Fabienne Comte, Céline Duval and Valentine Genon–Catalot.

In conclusion we again express our deep thanks to all the discussants and editors.

References

[1] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952. MR0057518

[2] M. Grant and S. Boyd. *The CVX Users' Guide. Release 2.1*, 2014. http://web.cvxr.com/cvx/doc/CVX.pdf.

[3] C. Kraft. Some conditions for consistency and uniform consistency of statistical procedures. *Univ. of California Publ. Statist.*, 2:493–507, 1955. MR0073896