Electronic Journal of Statistics
Vol. 9 (2015) 1720–1722
ISSN: 1935-7524
DOI: 10.1214/14-EJS980

Discussion of "Hypotheses testing by convex optimization"*

Axel Munk

Felix Bernstein Institute for Mathematical Statistics in the Bioscience
University of Göttingen, Germany
e-mail: munk@math.uni-goettingen.de

and

Frank Werner

Statistical Inverse Problems in Biophysics Group
Max Planck Institute for Biophysical Chemistry
Göttingen, Germany
e-mail: Frank.Werner@mpibpc.mpg.de

Received December 2014.

We congratulate the authors on a stimulating paper on a very intuitive and general approach to construct hypotheses tests. Restricting the considered class of tests to simple ones determined by a detector function ϕ, it seems most natural to minimize over ϕ and maximize over the pair $(x, y) \in X \times Y$ where X is the hypothesis and Y the alternative. Particularly, this guarantees that the corresponding test minimizes the worst case error which might happen under the given pair of hypothesis and alternative. It is evident that the resulting test is just the likelihood-ratio test for the worst-case hypothesis x^* against the worst-case alternative y^* with corresponding risk ε_*.

The used approach naturally leads to some restrictions yielding solvability of the resulting saddle point problem, including that X and Y need to be compact and convex. The authors propose an aggregation scheme to overcome this restriction, which is of independent interest from our point of view. Even though it is limited to X and Y being convex hulls of finitely many convex and compact subsets, the construction might be very helpful in many cases.

In the following we will comment on the impact of the proposed methodology to the problem of change point detection. Suppose m observations of the form

$$Y_i = \mu_m\left(\frac{i}{m}\right) + \sigma\varepsilon_i, \qquad 1 \leq i \leq m \tag{1}$$

are given, where $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and μ is a bump function of the form $\mu_m(x) = \Delta_m 1_{I_m}$ with a subinterval $I_m \subset [0,1]$. We want to test the hypothesis $\mu = 0$

*Main article 10.1214/15-EJS1054.

against the alternative $\mu \neq 0$ and ask for the detection boundary in terms of $\Delta_m, |I_m|$ and σ as $m \to \infty$. This problem can be handled with Goldenshluger, Juditsky and Nemirovski's approach by the construction used in Section 4.2.3. W.l.o.g. we will assume $n := |I_m|^{-1} \in \mathbb{N}$, set $\mathcal{V} = \{0\}$, $e[i] = e_i$ (the ith unit vector in $\mathbb{R}^n$) and define the matrix $A \in \mathbb{R}^{m \times n}$ by

$$A = \begin{pmatrix} \mathbf{1}_{m|I_m|} & 0 & & 0 \\ 0 & \mathbf{1}_{m|I_m|} & & \vdots \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & & \mathbf{1}_{m|I_m|} \end{pmatrix}.$$

Here $\mathbf{1}_{m|I_m|} = (1, 1, \dots, 1)^T \in \mathbb{R}^{m|I_m|}$. Note that m is the number of observations and $n = n(m)$ is related to the complexity of the alternative, more precisely it equals the number of possible non-overlapping positions for a bump of width $|I_m|$ in $[0, 1]$.

Now the ϵ-rate profile of the test is given by $\rho_i^G(\epsilon)$ in $(G_\epsilon^{i,\chi})$, i.e.

$$\begin{aligned} \rho_i^G(\epsilon) &= \max_{\rho,r} \left\{\rho : \quad |r| \, \|Ae[i]\|_2 \leq \sigma \left[\operatorname{ErfInv}\left(\frac{\epsilon}{4n}\right) + \operatorname{ErfInv}\left(\frac{\epsilon}{2}\right)\right], \quad r \geq \rho \right\} \\ &= \frac{\sigma \left[\operatorname{ErfInv}\left(\frac{\epsilon}{4n}\right) + \operatorname{ErfInv}\left(\frac{\epsilon}{2}\right)\right]}{\sqrt{m\,|I_m|}}. \end{aligned}$$

Here $\operatorname{ErfInv}(\operatorname{Erf}(t)) = t$ for all $t \in \mathbb{R}$ with $\operatorname{Erf}(t) = (2\pi)^{-1/2} \int_t^\infty \exp(-s^2/2)\,\mathrm{d}s$, which is in fact a multiple of the conjugate error function. Asymptotically it holds

$$\operatorname{Erf}(t) = \frac{\exp\left(-\frac{t^2}{2}\right)}{\sqrt{2\pi}t}\,(1 + o(1)), \qquad t \to \infty$$

and thus

$$\operatorname{ErfInv}(s) = \sqrt{2}\sqrt{-\ln(s)}\,(1 + o(1)), \qquad s \searrow 0.$$

The test gained by convex optimization is thus able to decide between $\mu = 0$ and $\mu \neq 0$ in the specified model if

$$\sqrt{m\,|I_m|}\Delta_m \geq \sqrt{2}\sigma\sqrt{-\ln(|I_m|)}, \qquad m \to \infty,$$

and is hence asymptotically optimal (see e.g. [2]).

If we consider the problem of testing $\mu = 0$ against k signals, the complexity of the alternative needs to equal $\binom{|I_m|^{-1}}{k}$. Thus $n \approx -k(\ln(|I_m|) + \ln(k))$, and thus the test gained by convex optimization will be able to decide between $\mu = 0$ and $\mu \neq 0$ in this situation if

$$\sqrt{m\,|I_m|}\Delta_m \geq \sqrt{2k}\sigma\sqrt{-\ln(|I_m|) - \ln(k)}, \qquad m \to \infty.$$

Note that $\sqrt{2k}$ is not optimal anymore, see again Sect. 2.5 in [2] (we also refer to [1] in a related context) where we obtained the constants 4 for a bounded and 12 for an unbounded number of change points. For the proposed testing procedure an unbounded number of signals in the alternative would (as shown above) lead to a diverging factor $k_m \to \infty$ instead of k.

At this point we want to emphasize that the above discussion shows the question of optimality to be answered in a conservative fashion in Proposition 4.2. In the discussed situation, no test achieving the κ_n-improved ϵ-rate can exist, but from Proposition 4.2 this cannot be concluded. It remains unclear if the same is true in many other interesting situations.

A similar behavior can be observed in case of indirect measurements. Suppose (1) is replaced by

$$Y_i = (h * \mu_m)\left(\frac{i}{m}\right) + \sigma\varepsilon_i, \qquad 1 \leq i \leq m$$

with a smooth and 1-periodic kernel h, and $*$ denotes periodic convolution. Again we want to test $\mu = 0$ against $\mu = \Delta_m 1_{I_m}$ with a subinterval $I_m \subset [0,1]$. This can be handled similarly as above by replacing A with $H \circ A$ where H is the $m \times m$-matrix describing the discretized convolution with h. It can readily be seen that

$$\begin{aligned}\|(H \circ A)\, e\,[i]\|_2 &= \sqrt{m}\, \|h * 1_{I_m}\|_{L^2([0,1])}\, (1 + o\,(1)) \\ &= \sqrt{m}\, |I_m|\, \|h\|_{L^2([0,1])}\, (1 + o\,(1))\,, \qquad m \to \infty.\end{aligned}$$

Thus the test gained by convex optimization is able to decide between $\mu = 0$ and $\mu \neq 0$ in this indirect model if

$$\sqrt{m}\, |I_m|\, \|h\|_{L^2([0,1])}\, \Delta_m \geq \sqrt{2}\sigma\sqrt{-\ln\left(|I_m|\right)}, \qquad m \to \infty.$$

Proposition 4.2 only ensures this rate to be optimal up to a $\sqrt{-\ln(|I_m|)}$-factor. In view of the case of direct observations this rate is even though likely to be optimal with optimal constant.

References

[1] Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36** 1758–1785. MR2435455

[2] Frick, K., Munk, A. and Sieling, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. With discussion and a rejoinder by the authors. MR3210728