Nonparametric testing by convex optimization

Anatoli Juditsky*

joint research with Alexander Goldenshluger† and Arkadi Nemirovski†
*University J. Fourier, †University of Haifa, †ISyE, Georgia Tech, Atlanta

Gargantua, November 26, 2013
Array of 20 sensors on the uniform grid along the left and bottom edges of $[0, 1]^2$. 
“+” represent the points of the uniform $20 \times 20$-grid $\Gamma$, 
“•” are sensor positions, interposed with contour plot of the response of the 6th sensor.
Suppose that $m$ sensors are deployed on the domain $G \subseteq \mathbb{R}^d$. Given a grid $\Gamma = (\gamma_i)_{i=1,...,n} \subset G$.

An event at a node $\gamma_i \in \Gamma$ produces the signal $s = re[i] : \Gamma \to \mathbb{R}^n$ of known signature $e[i]$ with unknown real factor $r$.

The signal is contaminated by a nuisance (a background signal) $v \in \mathcal{V}$, where $\mathcal{V}$ is a known convex and compact set in $\mathbb{R}^n$.

Observation $\omega = [\omega_1;...;\omega_m]$ of the array of $m$ sensors is a linear transformation of the signal, contaminated with random noise:

$$\omega \sim P_\mu$$

– a random vector in $\mathbb{R}^m$ with the distribution parameterized by $\mu \in \mathbb{R}^m$, where

$$\mu = A(s + v),$$

and $A \in \mathbb{R}^{m \times n}$ is a known matrix of sensor responses.
Objective: testing the (null) hypothesis $H_0$ that no event happened against
the alternative $H_1$ that exactly one event took place.

We require that

- $A_e[i] \neq 0$ for all $i$
- under $H_1$, when an event occurs at a node $\gamma_i \in \Gamma$, we have $s = re[i]$ with $|r| \geq \rho_i$
  with some given $\rho_i > 0$.

Problem $(D_\rho)$: Given $\rho = [\rho_1; \ldots; \rho_n] > 0$, decide between

- hypothesis $H_0 : s = 0$
- alternative $H_1(\rho) : s = re[i]$ for some $i \in \{1, \ldots, n\}$ and $r$ with $|r| \geq \rho_i$.

The risk of the test is the maximal probability to reject $H_0$ when the hypothesis is true
or to accept $H_0$ when $H_1(\rho)$ is true.

Our goal is, given an $\epsilon \in (0, 1)$, construct a test with risk $\leq \epsilon$ for as wide as possible
(i.e., with as small $\rho$ as possible) alternative $H_1(\rho)$.  

A particular case: signal detection in convolution


We consider the model with observation

\[ \omega = A(s + v) + \sigma \xi, \]

where \( s, v \in \mathbb{R}^n \), and \( \xi \sim \mathcal{N}(0, I_m) \) with known \( \sigma > 0 \).

Let \( \mu = [\mu_1; \ldots; \mu_m] \) be the vector of \( m \) consecutive outputs of a discrete time linear dynamical system with a given impulse response \( \{g_k\}, k = 1, \ldots, T \), i.e. \( \mu \in \mathbb{R}^m \) is the convolution image of \( n \)-dimensional “signal” \( s \) (that is, \( n = m + T - 1 \)).

\( A \) is the Toeplitz \( m \times n \) matrix of the described linear mapping \( x \mapsto \mu \).

We want to detect the presence of the signal \( s = re[i], \) where \( e[i], i = 1, \ldots, n, \) are some given vectors in \( \mathbb{R}^n \).
Situation, formally

Given are

- “Observation space” $\Omega, P$
  - $\Omega$: Polish (complete separable metric) space
  - $P$: $\sigma$-finite $\sigma$-additive Borel measure on $\Omega$

- Family $\mathcal{P} = \{P_\mu(d\omega) = p_\mu(\omega)P(d\omega) : \mu \in \mathcal{M}\}$ of probability distributions on $\Omega$
  - $\mu$: distribution’s parameter running through “parameter space” $\mathcal{M} \subset \mathbb{R}^m$
  - $p_\mu$: density of distribution $P_\mu$ w.r.t. the reference measure $P$

- “Parameter spaces” – two nonempty convex compact subsets $M_0 \subset \mathcal{M}$ and $M_1 \subset \mathcal{M}$. 

Assumptions

We assume that

- $\mathcal{M} \subset \mathbb{R}^m$ is a convex set which coincides with its relative interior;
- distributions $P_\mu \in \mathcal{P}$ possess densities $p_\mu(\omega)$ w.r.t. the measure $P$ on the space $\Omega$. We assume that $p_\mu(\omega)$ is continuous in $\mu \in \mathcal{M}$ and is positive for all $\omega \in \Omega$;
- We are given a finite-dimensional linear space $\mathcal{F}$ of continuous functions on $\Omega$ containing constants such that $\ln(p_\mu(\cdot)/p_\nu(\cdot)) \in \mathcal{F}$ whenever $\mu, \nu \in \mathcal{M}$;
We assume that

- $\mathcal{M} \subset \mathbb{R}^m$ is a convex set which coincides with its relative interior;

- distributions $P_\mu \in \mathcal{P}$ possess densities $p_\mu(\omega)$ w.r.t. the measure $P$ on the space $\Omega$. We assume that $p_\mu(\omega)$ is continuous in $\mu \in \mathcal{M}$ and is positive for all $\omega \in \Omega$;

- We are given a finite-dimensional linear space $\mathcal{F}$ of continuous functions on $\Omega$ containing constants such that $\ln(p_\mu(\cdot)/p_\nu(\cdot)) \in \mathcal{F}$ whenever $\mu, \nu \in \mathcal{M}$;

- For every $\phi \in \mathcal{F}$, the function $F_\phi(\mu) = \ln \left( \int_\Omega \exp\{\phi(\omega)\} p_\mu(\omega)P(d\omega) \right)$ is well defined and concave in $\mu \in \mathcal{M}$.

We call the just described situation a good observation scheme.
... and goal

Given observation scheme \([\Omega, P]\) and family of distributions \(\{p_\mu(\cdot)\}_{\mu \in \mathcal{M}}\), “parameter spaces” \(M_0, M_1\), and random observation

\[ \omega \sim p_\mu(\cdot), \]

coming from some unknown \(\mu\), known to belong either to \(M_0\) (hypothesis \(H_0\)) or to \(M_1\) (hypothesis \(H_1\)), decide between \(H_0\) and \(H_1\).

Risk of the test: given a test (we interpret value 0 as accepting \(H_0\) and 1 as accepting \(H_1\)), we consider the quantities

\[ \epsilon_0 = \sup_{\mu \in M_0} \text{Prob}_{\omega \sim P_\mu} \{\text{test rejects } H_0\}, \]

\[ \epsilon_1 = \sup_{\mu \in M_1} \text{Prob}_{\omega \sim P_\mu} \{\text{test rejects } H_1\}, \]

We say that risk of the test is \(\leq \epsilon\), if both error probabilities are \(\leq \epsilon\).
Example: Gaussian case

Given an noisy observation

\[ \omega = \mu + \xi, \quad \xi \sim \mathcal{N}(0, I), \]

make conclusions about \( \mu \).

The observation scheme is

- \((\Omega, P)\): \( \mathbb{R}^m \) with Lebesgue measure
- \( p_\mu(\omega) = \mathcal{N}(\mu, I), \ \mu \in \mathcal{M} := \mathbb{R}^m \)
- \( \mathcal{F} = \{ \phi(\omega) = a^T \omega + b : a \in \mathbb{R}^m, \ b \in \mathbb{R} \} \), and

\[ \ln \left( \int_{\mathbb{R}^m} e^{a^T \omega + b} p_\mu(\omega) d\omega \right) = b + a^T \mu + \frac{a^T a}{2}, \]

is concave in \( \mu \)

Gaussian observation scheme is good!
Given \( m \) realizations of independent Poisson random variables

\[
\omega_i \sim \text{Poisson}(\mu_i)
\]

with parameters \( \mu_i \), make conclusions about \( \mu \).

The observation scheme is

- \( (\Omega, \mathcal{P}) : \mathbb{Z}_+^m \) with counting measure
- \( p_\mu(\omega) = \frac{\mu^\omega}{\omega!} e^{-\sum_i \mu_i}, \mu \in \mathcal{M} = \text{int} \mathbb{R}_+^m \)
- \( \mathcal{F} = \{ \phi(\omega) = a^T \omega + b : a \in \mathbb{R}^m, b \in \mathbb{R} \} \), and

\[
\ln \left( \sum_{\omega \in \mathbb{Z}_+^m} e^{a^T \omega + b} p_\mu(\omega) \right) = b + \sum_{i=1}^m \left[ e^{a_i} - 1 \right] \mu_i,
\]

is concave in \( \mu \)

Poisson observation scheme is good!
Example: discrete case

Given realization of random variable $\omega$ taking values $1, \ldots, m$ with probabilities $\mu_i$

$$\mu_i := \text{Prob}\{\omega = i\},$$

make conclusions about $\mu$.

The observation scheme is

- $(\Omega, P)$: \{1, ..., $m$\} with counting measure

- $p_\mu(\omega) = \mu_\omega$, $\mu \in \mathcal{M} = \left\{ \mu \in \mathbb{R}^m : \mu > 0, \sum_{\omega=1}^{m} \mu_\omega = 1 \right\}$

- $\mathcal{F} = \mathbb{R}(\Omega) = \mathbb{R}^m$, and

$$\ln \left( \sum_{\omega \in \Omega} e^{\phi(\omega)} p_\mu(\omega) \right) = \ln \left( \sum_{\omega=1}^{m} e^{\phi(\omega)} \mu_\omega \right),$$

is concave in $\mu$.

Discrete observation scheme is good!
Simple test

Simple (Cramer’s) test: a simple test is specified by a detector $\phi(\cdot) \in \mathcal{F}$; it accepts $H_0$, the observation being $\omega$, if $\phi(\omega) \geq 0$, and accepts $H_1$ otherwise.

We can easily bound the risk of a simple test $\phi$: for $\mu \in M_0$ we have

$$\text{Prob}_{\omega \sim P_\mu} (\phi(\omega) < 0) \leq E_{\omega \sim P_\mu} (e^{-\phi(\omega)}) = \int_\Omega e^{-\phi(\omega)} p_\mu(\omega) P(d\omega),$$

and for $\nu \in M_1$,

$$\text{Prob}_{\omega \sim P_\nu} (\phi(\omega) \geq 0) \leq E_{\omega \sim P_\nu} (e^{\phi(\omega)}) = \int_\Omega e^{\phi(\omega)} p_\nu(\omega) P(d\omega).$$

We associate with $\phi(\cdot) \in \mathcal{F}$, and $[\mu; \nu] \in M_0 \times M_1$ the aggregate

$$\Phi(\phi, [\mu; \nu]) = \ln \left( \int_\Omega e^{-\phi(\omega)} p_\mu(\omega) P(d\omega) \right) + \ln \left( \int_\Omega e^{\phi(\omega)} p_\nu(\omega) P(d\omega) \right)$$

Key observation: in a good observation scheme $\Phi(\phi, [\mu; \nu])$ is continuous on its domain, convex in $\phi(\cdot) \in \mathcal{F}$ and concave in $[\mu; \nu] \in M_0 \times M_1$. 
Main result

Theorem 1

(i) $\Phi(\phi, [\mu; \nu])$ possesses a saddle point $(\min \text{ in } \phi, \max \text{ in } [\mu; \nu]) (\phi_*(\cdot), [x_*; y_*])$ on $\mathcal{F} \times (M_0 \times M_1)$ with the saddle value

$$\min_{\phi \in \mathcal{F}} \max_{[\mu; \nu] \in M_0 \times M_1} \Phi(\phi, [\mu; \nu]) := 2 \ln(\varepsilon_*) .$$

The risk of the simple test associated with the detector $\phi_*$ on the composite hypotheses $H_{M_0}, H_{M_1}$ is $\leq \varepsilon_*$. 

Main result

Theorem 1

(i) \( \Phi(\phi, [\mu; \nu]) \) possesses a saddle point (\( \min in \phi \), \( \max in [\mu; \nu] \)) (\( \phi_*(\cdot), [x_*, y_*] \)) on \( \mathcal{F} \times (M_0 \times M_1) \) with the saddle value

\[
\min_{\phi \in \mathcal{F}} \max_{[\mu; \nu] \in M_0 \times M_1} \Phi(\phi, [\mu; \nu]) := 2 \ln(\varepsilon_*).
\]

The risk of the simple test associated with the detector \( \phi_* \) on the composite hypotheses \( H_{M_0}, H_{M_1} \) is \( \leq \varepsilon_* \).

(ii) The detector \( \phi_* \) is readily given by the \( [\mu; \nu] \)-component \( [\mu_*; \nu_*] \) of the associated saddle point of \( \Phi \), specifically,

\[
\phi_*(\cdot) = \frac{1}{2} \ln \left[ p_{\mu_*}(\cdot)/p_{\nu_*}(\cdot) \right].
\]
Main result

Theorem 1

(i) \( \Phi(\phi, [\mu; \nu]) \) possesses a saddle point \((\min \text{ in } \phi, \max \text{ in } [\mu; \nu]) \) \((\phi_*(\cdot), [x_*; y_*])\) on \( \mathcal{F} \times (M_0 \times M_1) \) with the saddle value

\[
\min_{\phi \in \mathcal{F}} \max_{[\mu; \nu] \in M_0 \times M_1} \Phi(\phi, [\mu; \nu]) := 2 \ln(\varepsilon_*).
\]

The risk of the simple test associated with the detector \( \phi_* \) on the composite hypotheses \( H_{M_0}, H_{M_1} \) is \( \leq \varepsilon_* \).

(ii) The detector \( \phi_* \) is readily given by the \([\mu; \nu]\)-component \([\mu_*; \nu_*]\) of the associated saddle point of \( \Phi \), specifically,

\[
\phi_*(\cdot) = \frac{1}{2} \ln \left[ p_{\mu_*}(\cdot)/p_{\nu_*}(\cdot) \right].
\]

(iii) Let \( \varepsilon \geq 0 \) be such that there exists a (whatever) test for deciding between two simple hypotheses

\[
(A) : \omega \sim p(\cdot) := p_{\mu_*}(\cdot), \quad (B) : \omega \sim q(\cdot) := p_{\nu_*}(\cdot)
\]

with the sum of error probabilities \( \leq 2\varepsilon \). Then \( \varepsilon_* \leq 2\sqrt{\varepsilon} \).
Example: Gaussian case

[Chencov, 70’s, Burnashev 1979, 1982, Ingster, Suslina, 2002,...]

Here \((\Omega, P)\) is \(\mathbb{R}^m\) with the Lebesgue measure, \(\mathcal{M} = \mathbb{R}^m\), \(p_\mu(\cdot)\) is the density of the Gaussian distribution \(\mathcal{N}(\mu, I)\), and \(\mathcal{F}\) is the space of all affine functions on \(\Omega = \mathbb{R}^m\).

Assuming that the nonempty convex compact sets \(M_0, M_1\) do not intersect, we get

\[
[\mu_*, \nu_*] \in \operatorname{Argmin}_{\mu \in M_0, \nu \in M_1} ||\mu - \nu||_2.
\]

and

\[
\phi_*(\omega) = \xi^T \omega - \alpha, \quad \text{where} \quad \xi = \frac{1}{2} [\mu_* - \nu_*], \quad \alpha = \frac{1}{2} \xi^T [\mu_* + \nu_*]
\]

The error probabilities of the associated simple test do not exceed

\[
1 - F_{\mathcal{N}}\left(||\mu_* - \nu_*||_2/2\right),
\]

where \(F_{\mathcal{N}}(\cdot)\) is the standard normal c.d.f.
Example: discrete case

[Birge 1982, 1983]

Let \((\Omega, P)\) be a finite set of cardinality \(m\) with counting measure \(P\), \(\mathcal{M} \subset \mathbb{R}^m\) is the relative interior of the standard simplex in \(\mathbb{R}^m\):

\[
\mathcal{M} = \{ \mu = \{ \mu_\omega : \omega \in \Omega \} : \mu > 0, \sum_\omega \mu_\omega = 1 \}
\]

with \(p_\mu(\omega) = \mu_\omega\), and \(\mathcal{F} = \mathbb{R}(\Omega)\) is the space of all real-valued functions on \(\Omega\).

Assuming that the sets \(M_0, M_1\) do not intersect, we get

\[
[\mu_*; \nu_*] \in \operatorname{Argmax}_{\mu \in M_0, \nu \in M_1} \sum_\omega \sqrt{\mu_\omega \nu_\omega},
\]

and

\[
\phi_*(\omega) = \ln \sqrt{[\mu_*]_\omega [\nu_*]_\omega}, \quad \varepsilon_* = \sum_{\omega \in \Omega} \sqrt{[\mu_*]_\omega [\nu_*]_\omega}.
\]
Example: Poisson case

Here $\Omega = \mathbb{Z}_+^m$ is the grid of nonnegative integer vectors in $\mathbb{R}^m$, $P$ is the counting measure on $\Omega$, $\mathcal{M} = \mathbb{R}^m_{++} := \{ \mu \in \mathbb{R}^m : \mu > 0 \}$, and

$$p_\mu(\omega) = \prod_{i=1}^m \left[ \frac{\mu_i^{\omega_i}}{\omega_i!} e^{-\mu_i} \right]$$

is the distribution of the random vector with independent Poisson entries $\omega_1, \ldots, \omega_m$. $\mathcal{F}$ is comprised of the restrictions onto $\mathbb{Z}_+^m$ of affine functions.

Assuming, same as above, that the sets $M_0$, $M_1$ do not intersect, we get

$$\begin{bmatrix} [\mu_*; \nu_*] & \in & \text{Argmin}_{\mu \in M_0, \nu \in M_1} \sum_{\ell=1}^m \left[ \sqrt{\mu_\ell} - \sqrt{\nu_\ell} \right]^2 \\ \text{Opt} & = & \frac{1}{2} \sum_{\ell=1}^m \left[ \sqrt{[\mu_*]_\ell} - \sqrt{[\nu_*]_\ell} \right]^2 \end{bmatrix},$$

and

$$\phi_*(\omega) = \sum_{\ell=1}^m \ln \left( \sqrt{[\mu_*]_\ell/[\nu_*]_\ell} \right) \omega_\ell - \frac{1}{2} \sum_{\ell=1}^m [\mu_* - \nu_*]_\ell$$

with $\varepsilon_* = \exp\{-\text{Opt}\}$. 
Illustration: PET

The collected data is the list of total numbers of coincidences registered in every bin (pair of detector cells) over a given time $T$. The goal is to infer about the density $\times$ of the tracer. After suitable discretization, we arrive at Poisson case

$$\omega = \{\omega_i \sim \text{Poisson}(\mu_i)\}_{i=1}^{m}, \quad \mu_i = \sum_{j=1}^{n} A_{ij} x_j$$

- $m$ bins and $n$ voxels (small cubes in which the field of view is split)
- $x_j$: average tracer’s density in voxel $j$
- $A_{ij}$: $T \times \begin{bmatrix} \text{probability for line of response originating in voxel } j \text{ to be registered in bin } i \end{bmatrix}$
We consider 2D PET with $m = 64$ detector cells and $40 \times 40$ field of view:

- $X \cup Y$: the set of tracer’s densities $x \in \mathbb{R}^{40 \times 40}$ satisfying some regularity assumptions and at average not exceeding 1
- $M_1 = AY$: $X$ is the set of densities with the average over the $3 \times 3$ red spot at least 1.1
- $M_0 = AX$: $Y$ is the set of densities with average over the red spot at most 1.
- The observation time is chosen to allow to decide on $H_0$ vs. $H_1$ with risk 0.01.
Results of 1024 simulations:

- Wrongly rejecting $H_0$ in 0% of cases
- Wrongly rejecting $H_1$ in 0.1% of cases

Top plot: $x_*$, middle plot: $y_*$, bottom plot: $x_* - y_*$
Case of repeated observations

Assume we are given a good observation scheme \(((\Omega, P), \{p_\mu(\cdot) : \mu \in \mathcal{M}\}, \mathcal{F})\), along with same as above \(M_0, M_1\).

We now observe a sample of \(K\) independent realizations

\[ \omega_k \sim p_\mu(\cdot), \quad k = 1, \ldots, K, \]

what corresponds to the observation scheme

- observation space \(\Omega^{(K)} = \{\omega^K = (\omega_1, \ldots, \omega_K) : \omega_k \in \Omega \ \forall k\}\) equipped with the measure \(P^{(K)} = P \times \ldots \times P\),
- family \(\left\{ p^{(K)}_\mu(\omega^K) = \prod_{k=1}^{K} p_\mu(\omega_k), \mu \in \mathcal{M} \right\}\) of densities of observations w.r.t. \(P^{(K)}\), and \(\mathcal{F}^{(K)} = \left\{ \phi^{(K)}(\omega^K) = \sum_{k=1}^{K} \phi(\omega_k), \ \phi \in \mathcal{F} \right\}\).

We want to decide between the hypotheses that the \((K\text{-element})\) observation \(\omega^K\) comes from a distribution \(p^{(K)}_\mu(\cdot)\) with \(\mu \in M_0\) (hypothesis \(H_0\)) or with \(\mu \in M_1\) (hypothesis \(H_1\)).
Detectors $\phi_*$, $\phi^{(K)}_*$ and risk bounds $\varepsilon_*$, $\varepsilon^{(K)}_*$ given by Theorem 1, as applied to the original and the K-repeated observation schemes are linked by the relations

$$\phi^{(K)}_*(\omega_1, \ldots, \omega_K) = \sum_{k=1}^{K} \phi_*(\omega_k), \quad \varepsilon^{(K)}_* = (\varepsilon_*)^K.$$  

As a result, the “near-optimality claim” Theorem 1.iii can be reformulated as follows:

**Corollary** Assume that for some integer $K^* \geq 1$ and some $\epsilon \in (0, 1/4)$, the hypotheses $H_0$, $H_1$ can be decided, by a whatever procedure utilising $K^*$ observations, with error probabilities $\leq \epsilon$. Then with

$$K^+ = \text{Ceil} \left( \frac{2 \ln(1/\epsilon)}{\ln(1/\epsilon) - 2 \ln(2)} K^* \right)$$  

observations, the simple test with the detector $\phi^{(K^+)}_*$ decides between $H_0$ and $H_1$ with risk $\leq \epsilon$. 
Assume that we are given

- convex compact sets $M_\ell$ in $\mathcal{M} \subset \mathbb{R}^m$, $1 \leq \ell \leq L$;
- a good observation scheme $((\Omega, P), \{p_\mu(\cdot), \mu \in \mathcal{M} \subset \mathbb{R}^m\}, \mathcal{F})$.

Given an observation $\omega \in \Omega$, our goal is to decide between the hypotheses $H_\ell$, $1 \leq \ell \leq L$, stating that the observation $\omega \sim p_\mu(\cdot)$ corresponds to $\mu \in M_\ell$. 
Pairwise testing

Consider all (unordered) pairs \( \{\ell, \ell'\} \) with \( \ell \neq \ell' \) and \( 1 \leq \ell, \ell' \leq L \), and associate with such a pair a simple test given by detector \( \phi_{\ell, \ell'}^* (\cdot) \), along with the upper bound \( \varepsilon_*[\ell, \ell'] \) on the risk of this test yielded by Theorem 1, as applied to \( M_0 = M_{\ell}, \ M_1 = M_{\ell'} \).

Let \( C \) be a collection of pairs \( \{\ell, \ell'\} \).

Testing procedure: given an observation \( \omega \), we “look” one by one at all pairs \( \{\ell, \ell'\} \in C \) and apply to our observation \( \omega \) the simple test, given by the detector \( \phi_{\ell, \ell'}^* (\cdot) \), to decide between the hypotheses \( H_\ell, H_{\ell'} \).

The outcome of the inference process is the list of these rejected hypotheses.

The (un)reliability of such an inference can be naturally upper-bounded by the quantity

\[
\varepsilon[C] := \max_{\ell \leq L} \sum_{\ell': \{\ell, \ell'\} \in C} \varepsilon_*[\ell, \ell'].
\]
Application to multisensor detection

The setting: We are given an observation \( \omega \sim P_\mu \) parameterized by the vector parameter \( \mu = A(s + v) \), where \( A \in \mathbb{R}^{m \times n} \) is a known matrix.

Useful signal \( s = \text{re}[i] \in \mathbb{R}^n \) is known up to its “position” \( i \in \{1, \ldots, n\} \) and the scalar factor \( r \), and \( v \) is the nuisance known to belong to a given set \( \mathcal{V} \subset \mathbb{R}^n \), which we assume to be convex and compact.

Objective: solve the testing problem \((\mathcal{D}_\rho)\), i.e., decide between \( H_0 : s = 0 \) and

\[
H_1(\rho = [\rho_1; \ldots; \rho_n]) = \{ s = \text{re}[i] \text{ for some } i \text{ and } r \text{ such that } |r| \geq \rho_i \}.
\]
Given a test $\phi(\cdot)$ and $\epsilon > 0$, we call a collection $\rho = [\rho_1; \ldots; \rho_n]$ of positive reals the $\epsilon$-rate profile of the test $\phi$ if

- whenever $s = 0$ and $v \in \mathcal{V}$, the probability for the test to reject $H_0$ is $\leq \epsilon$;
- whenever the signal $s$ underlying our observation is $re[i]$ for some $i$ and $r$ with $\rho_i \leq |r|$, and the nuisance $v \in \mathcal{V}$, the test rejects $H_0$ with probability $\geq 1 - \epsilon$.

Our goal is to design a test with the “best possible” $\epsilon$-rate profile:

**Definition.** Let $\kappa \geq 1$. A test $\phi$ with risk $\epsilon$ in the problem $(\mathcal{D}_\rho)$ is said to be $\kappa$–rate optimal, if there is no test with the risk $\epsilon$ in the problem $(\mathcal{D}_{\underline{\rho}})$ with $\underline{\rho} < \kappa^{-1}\rho$. 
Multisensor detection: Gaussian case

Let the distribution $P_\mu$ of $\omega$ be normal with the mean $\mu$, i.e. $\omega \sim N(\mu, \sigma^2 I)$ with known variance $\sigma^2 > 0$. For the sake of simplicity, assume also that the (convex and compact) nuisance set $\mathcal{V}$ is symmetric w.r.t. the origin.

- The null hypothesis is $H_0 : \mu \in A\mathcal{V} = \{\mu = Av, v \in \mathcal{V}\}$.
- The alternative $H_1(\rho)$ can be represented as the union, over $i = 1, \ldots, n$, of $2n$ hypotheses $H^{\pm,i}(\rho_i) : \mu \in \pm AX_i(\rho_i) = \{\mu = Ax, x \in \pm AX_i(\rho_i)\}$, where $X_i(\rho_i) = \{x \in \mathbb{R}^n : x = re[i] + v, v \in \mathcal{V}, \rho_i \leq r\}$. 
\[ \mu_k = A(\rho_k s[k] + v^k) \]

\[ A(\rho_k e[k] + \mathcal{V}) \]
Let $1 \leq i \leq n$ be fixed, and suppose we want to distinguish $H_0$ from $H_{i}^{+i}(\rho)$. The separation with risk $\epsilon$ is impossible unless
\[
\dist(AX_i(\rho)) \geq q_{\mathcal{N}}(\epsilon/2),
\]
meaning that
\[
\rho \geq \rho_{*,i}^G(\epsilon) = \max \{ r : \|Au - A(re[i] + v)\|_2 \leq 2\sigma q_{\mathcal{N}}(\epsilon/2), \; u, v \in \mathcal{V} \}.
\]
where $q_{\mathcal{N}}(s)$ is the $1 - s$–quantile of $\mathcal{N}(0, 1)$.

To ensure the “total risk” of separation of $H_0$ and $\bigcup_i H_{\pm,i}^{\pm}(\rho_i)$ to be $\leq \epsilon$, one can take
\[
\rho_i \geq \rho_i^G(\epsilon) = \max \{ r : \|Au - A(re[i] + v)\|_2 \leq 2\sigma q_{\mathcal{N}}(\epsilon/(4n)), \; u, v \in \mathcal{V} \}.
\]
Let $1 \leq i \leq n$ be fixed, and suppose we want to distinguish $H_0$ from $H_i^{+i}(\rho)$. The separation with risk $\epsilon$ is impossible unless

$$\text{dist}(A\mathcal{V}, AX_i(\rho)) \geq q_N(\epsilon/2),$$

meaning that

$$\rho \geq \rho_{*,i}^G(\epsilon) = \max_{\rho,r,u,v} \{ r : \|Au - A(re[i] + v)\|_2 \leq 2\sigma q_N(\epsilon/2), \ u, v \in \mathcal{V} \}.$$

where $q_N(s)$ is the $1 - s$–quantile of $\mathcal{N}(0,1)$.

We can be a bit smarter: when deciding between $H_0$ and each of $H^{\pm,i}(\rho_i)$ we can “skew” the test so that

- probability of wrongly rejecting $H_0$ is $\epsilon/4n$
- probability of wrongly rejecting $H^{\pm,i}(\rho_i)$ is $\epsilon/2$.

In this case, the risk $\epsilon$ is attained if

$$\rho_i \geq \rho_i^G(\epsilon) = \max_{\rho,r,u,v} \{ r : \|Au - A(re[i] + v)\|_2 \leq \sigma \left[ q_N \left( \frac{\epsilon}{4n} \right) + q_N \left( \frac{\epsilon}{2} \right) \right], \ u, v \in \mathcal{V} \}.$$
So, for $1 \leq i \leq n$ we set

$$
\rho^G_i(\epsilon) = \max_{\rho, r, u, v} \left\{ r : \|Au - A(\text{re}[i] + v)\|_2 \leq 2\sigma \left[ q_N \left( \frac{\epsilon}{4n} \right) + q_N \left( \frac{\epsilon}{2} \right) \right], \ u, v \in V \right\}.
$$

Let

$$
\phi_i,\pm(\omega) = \pm[Au^i - A(r^i e[i] + v^i)]^T \omega - \alpha_i,
$$

with

$$
\alpha_i = [Au^i - A(r^i e[i] + v^i)]^T \frac{q_N(\epsilon/4n)A(r^i e[i] + v^i) + q_N(\epsilon/2)Au^i}{q_N(\epsilon/4n) + q_N(\epsilon/2)},
$$

where $u^i, v^i, r^i$ are the $u, v, r$-components of an optimal solution to $(G^i_\epsilon)$ (of course, $r^i = \rho^G_i$).

Finally, set

$$
\rho^G[\epsilon] = [\rho^G_1(\epsilon); \ldots; \rho^G_n(\epsilon)],
$$

$$
\hat{\phi}_G(\omega) = \min_{1 \leq i \leq n} \phi_i,\pm(\omega).
$$
Consider the test (we refer to it as to $\hat{\phi}_G$) which

- accepts $H_0$ when $\hat{\phi}_G(\omega) \geq 0$ (i.e., with observation $\omega$, all simple tests with detectors $\phi_{i,\pm}$, $1 \leq i \leq n$, when deciding on $H_0$ vs. $H_{\pm,i}$, accept $H_0$),
- otherwise accepts $H_1(\rho)$.

**Proposition [Gaussian]**

(i) Whenever $\rho \geq \rho^G[\epsilon]$ the risk of the test $\hat{\phi}_G$ in the Gaussian case of problem $(\mathcal{D}_\rho)$ is $\leq \epsilon$.

(ii) When $\rho = \rho^G[\epsilon]$, the test is $\kappa_n$-rate optimal with

$$\kappa_n = \kappa_n(\epsilon) := \frac{q_{\mathcal{N}}(\frac{\epsilon}{4n}) + q_{\mathcal{N}}(\frac{\epsilon}{2})}{2q_{\mathcal{N}}(\frac{\epsilon}{2})}. $$

Note that $\kappa_n(\epsilon) \to 1$ as $\epsilon \to +0$. 
We consider here the “convolution model” with observation

\[ \omega = A(s + v) + \xi, \]

where \( s, v \in \mathbb{R}^n \), and \( \xi \sim \mathcal{N}(0, I_m) \), and \( A \) is the matrix of discrete convolution. We are to decide between the hypotheses

- \( H_0 : \mu \in A\mathcal{V} \) and
- \( H_1(\rho) = \bigcup_{1 \leq i \leq n} H^\pm, i(\rho_i) \), with the hypotheses \( H^\pm, i(\rho_i) \) as above.

\[ \mathcal{V}_L = \{ u \in \mathbb{R}^n : |u_i - 2u_{i-1} - u_{i-2}| \leq L, \ i = 3, \ldots, n \}, \]

where \( L \) is experiment’s parameter (\( L = 0.1 \) in the experiment below).
baseline and nominal $\rho$-profiles, $\epsilon = 0.1$

difference signal $s^i + v^i - u^i$, jump at $i = 100$

$\rho$-profiles ratio, $\epsilon = 0.1$

corresponding observation, $\epsilon = 0.1$
baseline and nominal $\rho$-profiles, $\varepsilon = 0.1$

difference signal $s^i + v^i - u^i$, jump at $i = 100$

$\rho$-profile ratio, $\varepsilon = 0.1$

corresponding observation and detector, $\varepsilon = 0.1$
Jump detection in convolution model: numerical lower bound

Question: can the $\log n$–factor be removed?

Answer (partial, theoretical): [Goldenshluger et al, 2008] *in certain (inverse) models the $\log n$–factor cannot be removed*

Answer (numerical): we can lower bound the performance of any test by the performance of the Bayesian test on the problem of testing of

- $H_0 : \mu = 0$, and
- $H_1(\rho)$ which is the union, over $i = 1, ..., n$, of $2n$ hypotheses

$$H^\pm_i(\rho_i) : \mu = \pm Ax^i := \pm A(\rho_i e[i] + \nu^i - u^i) \ [= \pm A(\rho_i e[i] + 2\nu^i)], \ \nu, u \in \mathcal{V}.$$
\[
\mu^k = A(\rho_k s[k] + v^k)
\]

\[
A(\rho_k e[k] + \mathcal{V})
\]
\[ \nu^k = A(\rho_k s[k] + \nu^k) - Au^k \]
Numerical lower bound in the periodic case

Sum $\varepsilon$ of error probabilities in testing $H_0$ versus $H_1(\rho)$ as a function of $\rho(=\rho_i)$, $n = 100$.

$-\log_{10}(\text{union upper bound})$

$-\log_{10}(\varepsilon)$ of the Bayesian test over uniform prior on $\nu^k$, $k = 1, \ldots, n$ (1e6 sim)

$-\log_{10}(\text{baseline error})$
Numerical lower bound in the periodic case

Sum $\varepsilon$ of error probabilities in testing $H_0$ versus $H_1(\rho)$ as a function of $\rho(=\rho_i)$, $n = 1000$.

- $-\log_{10}(\text{union upper bound})$
- $-\log_{10}(\varepsilon)$ of the Bayesian test over uniform prior on $\nu^k$, $k = 1, \ldots, n$ (1e6 sim)
- $-\log_{10}(\text{baseline error})$
Numerical example: event detection in sensor networks

Same as above, the available observation is

\[ \omega = A(s + v) + \xi, \]

where \( s, v \in \mathbb{R}^n \), and \( \xi \sim \mathcal{N}(0, I_m) \), \( A \) is the \( m \times n \) matrix of sensor responses. We are to decide between the hypotheses

- \( H_0 : \mu \in AV \) (observation is a result of a pure nuisance) and
- \( H_1(\rho) = \bigcup_{1 \leq i \leq n} H^{\pm,i}(\rho_i) \), with the hypothesis \( H^{\pm,i}(\rho_i) \) saying that an event at the node \( i \) produced a signal \( s = re[i], |r| \geq \rho_i \).

Setup: The signal signatures \( e[i], 1 \leq i \leq n \) are the standard basic orths in \( \mathbb{R}^n \), and the nuisance set \( V \) is defined as

\[ V_L = \{ u \in \mathbb{R}^n : |\mathcal{L}v| \leq L \}, \]

where \( \mathcal{L} \) is the discrete Laplace operator.

In the reported experiment \( m = 20, n = 20^2, L = 0.1 \).
response of the 6th sensor

signal $s + \nu$ of the event at $\gamma = (5, 20)$

$\rho$-profile, $\epsilon = 0.1$

corresponding detector, $\epsilon = 0.1$