Convex optimization for nonparametric estimation and test

A. Juditsky[1]

joint work with A. Nemirovski[2] and A. Goldenshluger[3]

[1]LJK, Université J. Fourier de Grenoble
[2]ISyE, GeorgiaTech, Atlanta
[3]University of Haifa

Zvenigorod
June, 20 2012

Motivating example: particle detector

Continuous time model:

$$dY_t = \left[\int_{-\infty}^{t} K(t-s)(\mu(s)+w(s))ds\right] dt + \epsilon dW_t.$$

with

- *arrival signal* – Dirac function $\mu_s = \mu\delta(s-\theta)$;
- *drift of the measurement device* w;
- *measurement noise* W, modelled by a standard Brownian motion;
- $K(s)$ impulse response of the detector.

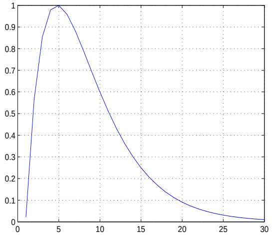
Example: particle detector

When sampled with period Δ:

$$Y_n = \sum_{k \le n} K_{n-k}[\mu(k\Delta) + \Delta w(k\Delta)] + \epsilon\sqrt{\Delta}\xi_n,$$

where

- Y_n is the observation at the instant $t = \Delta n$,
- $\xi_n \sim N(0,1)$ are i.i.d Gaussian disturbances,
- K_k is the sampled response



General model

Let in the model

$$Y = A_0\mu + A_1 w + \sigma\xi,$$

$\mu \in \mathbf{R}^m$ be (unknown) *useful signal* and $w \in \mathbf{R}^k$ be (unknown) *nuisance signal*.

It is known *a priori* that $\mu \in S$ and $w \in N$ (convex and compact).

On may be interested in

- estimating the component μ_i of μ;
- testing if, for instance, $\mu_i = 0$ against $\mu_i > 0$;
- etc..

If we set $x = [\mu; w]$, $X = S \times N$ and $A = [A_0, A_1]$, we come to the model

$$Y = Ax + \sigma\xi.$$

White noise model

We consider the observation

$$Y = Ax + \sigma\xi$$

with

- *observable* $Y \in \mathbf{R}^M$;
- *unknown parameter* $x \in \mathbf{R}^N$; it is known *a priori* that $x \in X$, X being convex and compact;
- *sensing matrix* $A \in \mathbf{R}^{M \times N}$;
- *observation noise* $\xi \sim N(0, I_M)$, $\sigma > 0$ is known.

Our interest may be

- estimate the value $g(x)$ at x of a given linear functional $g(z) = g^T z$;
- test the hypothesis

$$H_0 : \ x \in X_0 \ \text{ against } \ H_1 : x \in X_1,$$

where X_0 and X_1 are disjoint subsets of X.

Noisy distribution model

Let ζ be a random variable on $\{1, ..., N\}$, $\zeta \sim x$ (i.e.$P(\zeta = k) = x_k$, $k = 1, ..., N$),

$$x \in \mathcal{P}_N = \{x \in \mathbf{R}^N : \; x \geq 0, \sum_i x_i = 1\}.$$

We suppose that it is known *a priori* that $x \in X \subset \mathcal{P}$ (convex and closed).

Let $Y \in \{1, ..., M\}$ be a discrete random variable, $Y \sim p$:

$$P(Y = k) = p_k, \;\; p_k > 0, \;\; k = 1, ..., M,$$

where $p = Ax$.

Our objective is, given n i.i.d. realisations $Y_1, ..., Y_n$ of Y,

- estimate the value at x of a given linear functional $g^T z$;
- test the hypothesis

$$H_0 : \; x \in X_0 \;\text{ against }\; H_1 : x \in X_1,$$

where X_0 and X_1 are disjoint subsets of X.

Example: convolution model

Suppose we observe realizations of the random variable $\zeta = \eta + \xi$ of two independent r.v..

The distribution F_ξ of ξ (noise) is supposed to be known. It is also known *a priori* that the distribution F_η of η belongs to a certain *regularity class*.

- Let us cut the domain of η into N segments Δ_i (say, $N-2$ equal segments of length Δ and 2 "complements" in $\mathbf{R}$);
- cut the observation domain into M segments δ_j;
- compute values A_{ij}, $i = 1, ..., N$, $j = 1, ..., M$:

$$A_{ij} = \Delta_j^{-1} \int_{u \in \delta_i} \int_{v \in \Delta_j} f_\xi(u - v) du dv.$$

Then the distribution of $Y = \sum_{k=1}^M k 1_{\xi+\eta}$ may be approximate with noisy distribution model.

Poisson (PET) model

We observe a vector Y of M independent Poisson r.v., $Y = [Y_1; ...; Y_M]$, distributed with parameter $\lambda = [\lambda_1; ...; \lambda_M] > 0$. Assume that

$$\lambda = Ax, \quad \text{with} \quad A = [A_{ij} \geq 0, \; i = 1, ..., M, \; j = 1, ..., n],$$

where the signal of interest $x \in X \subset \mathbf{R}^n_+$, convex and compact.
Our objective is, given the observation Y to infer the value of the linear form $g^T x$ at x.

PET problem: we have N independent particle sources of intensity $x = [x_1; ...; x_N]$ and M detectors. We observe the vector vector $Y = [Y_1; ...; Y_M]$ of numbers of particles registered in each detector.

In this case A_{ij} represents the probability for a particle from the source j to be registered in the detector i.

General problem description

We observe a sample $Y = [Y_1; ...; Y_M]$ of M independent r.v.. We assume that the distribution of Y_i are the same up to (a parameter) $\lambda \in \mathbf{R}^M$.

The unknown λ *affinely* parameterized by an n-dimensional "signal" x, $x \in X$ (a known convex compact set in $\mathbf{R}^n$):

$$\lambda = A(x) = [A_1(x); ...; A_M(x)],$$

where $A(\cdot)$ is a given affine mapping.

Our objective is

- estimate a given linear form $g^T z$ of $z \in \mathbf{R}^n$ at x;
- or, given X_i, $i = 0, 1$, disjoin subsets of X, test l'hypothesis $x \in X_0$ vs $x \in X_1$.

Testing simple hypothesis

Assume that an observation of a r.v. $Y \sim P$ is available, $P \in \{P_0, P_1\}$. We want to distinguish between two hypotheses:

$$H_0 : \; P = P_0 \;\text{ and }\; H_1 : \; P = P_1.$$

We call *test* any function $\phi(X) \to \{0, 1\}$.
We say that

$$\phi(X) = \begin{cases} 1, & \text{test rejects } H_0, \\ 0, & \text{test does not reject } H_0. \end{cases}$$

We associate with ϕ its *rejection region* R such that $\phi(x) = 1_{x \in R}$.

We measure the risk of the test with the quantity

$$\mathrm{Risk}(\phi) = P_0(\phi(X) = 1) + P_1(\phi(X) = 0).$$

Neyman-Pearson lemma

Suppose that P_0 and P_1 possess the densities p_0 and p_1 with respect to some measure μ.
The test ϕ^* such that

$$\phi^*(x) = 1_{p_1(x) \geq p_0(x)}$$

is referred to as *likelihood ratio test* or Neyman-Pearson test.

Lemma [Neyman-Pearson] *For any test T*

$$\mathrm{Risk}(\phi) \geq \mathrm{Risk}(\phi^*) = \int [p_0(x) \wedge p_1(x)]\mu(dx) = 1 - \frac{1}{2}\|p_0 - p_1\|_1.$$

In other words, the likelihood ratio test minimizes the risk among all tests.

Proof:

Note that for any ϕ valued in $\{0,1\}$,

$$P_0(\phi(X)=1)=E_0\phi(X)=\int\phi(x)p_0(x)\mu(dx),\;\; P_1(\phi=0)=\int(1-\phi(x))p_1(x)\mu(dx),$$

and we have

$$\begin{aligned}\mathrm{Risk}(\phi) &= P_0(\phi(X)=1)+P_1(\phi(X)=0)\\ &= \int\phi(x)p_0(x)\mu(dx)+\int(1-\phi(x))p_1(x)\mu(dx)\\ &\geq \int[p_0(x)\wedge p_1(x)]\mu(dx).\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathrm{Risk}(\phi^*) &= P_0(\phi^*(X)=1)+P_1(\phi^*(X)=1)\\ &= \int 1_{p_0(x)\leq p_1(x)}\,p_0(x)\mu(dx)+\int 1_{p_1(x)<p_0(x)}\,p_1(x)\mu(dx)\\ &= \int[p_0(x)\wedge p_1(x)]\mu(dx).\end{aligned}$$

Application into Gaussian setting

Consider the problem of testing between the hypotheses

$$H_0 : \; x = x_0 \;\; \text{and} \;\; H_1 : \; x = x_1$$

given an observation $Z \sim \mathcal{N}(x, \sigma^2 I)$, $x \in \mathbf{R}^n$.

We have

$$f_x(z) = (2\pi)^{-n/2}\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}(z-x)^T(z-x)\right),$$

and the likelihood ratio writes

$$\frac{f_{x_1}(z)}{f_{x_0}(z)} = \exp\left(\frac{1}{2\sigma^2}(2z - [x_1 + x_0])^T(x_1 - x_0)\right).$$

Thus the N.-P. test $\quad R^* = \left\{z : \; (x_1 - x_0)^T(z - \frac{x_1+x_0}{2}) \geq 0\right\},$
and its risk

$$\begin{aligned} \mathrm{Risk}(R^*) &= P_{x_0}(X \in R^*) + P_{x_1}(X \notin R^*) = 2P\left(N(0,\sigma^2) \geq \tfrac{1}{2}\|x_1 - x_0\|_2\right) \\ &= 2P\left(N(0,1) \geq \frac{\|x_1 - x_0\|_2}{2\sigma}\right) := 2\mathrm{erf}\left(\frac{\|x_1 - x_0\|_2}{2\sigma}\right). \end{aligned}$$

Testing convex hypotheses

We assume the white noise model: $Y = Ax + \sigma\xi$, where $Y \in \mathbf{R}^M$ is observable and $x \in X \subset \mathbf{R}^N$, convex and compact.

Our first objective is to test

$$H_0 : \; x \in X_0 \; \text{ against } \; H_1 : x \in X_1$$

where $X_i, i = 0, 1$ are convex compact subsets of X.

Let ϕ be a test. We measure the performance of ϕ with the maximal risk over $X_0 \cup X_1$:

$$\mathrm{Risk}(\phi) = \sup_{x \in X_0} P_x(\phi(Y) = 1) + \sup_{x \in X_1} P_x(\phi(Y) = 0).$$

Testing convex hypotheses

Consider the following construction: let

$$[x_0, x_1] \in \mathrm{Argmin}\{\|Az_0 - Az_1\|_2,\ z_0 \in X_0,\ z_1 \in X_1\}. \qquad (1)$$

Let $y_0 = Ax_0$ and $y_1 = Ax_1$. Then

$$\|y_1 - y_0\|_2 = \mathrm{dist}(AX_0, AX_1).$$

We set $\phi = \frac{y_1 - y_0}{\|y_1 - y_0\|_2}$, and let

$$\phi^* = 1_{\phi^T(y - \frac{y_1 + y_0}{2}) \geq 0}$$

(the likelihood ratio test for $H_0 :\ y = y_0$ against $H_1 : y = y_1$).

Lemma *For any test ϕ, the maximal risk satisfies*

$$\mathrm{Risk}(\phi) \geq \mathrm{Risk}(\phi^*) = 2\mathrm{erf}\left(\frac{\|y_1 - y_0\|_2}{2\sigma}\right) = 2\mathrm{erf}\left(\frac{\mathrm{dist}(AX_0, AX_1)}{2\sigma}\right).$$

Application to particle detector

We have M observations

$$Y_t = \sigma\xi_t + \sum_{\tau=t-p+1}^{t} (\mu_\tau + w_\tau \Delta t) K_{t-\tau},\ t = 0, 1, ..., M-1,$$

where $w_i = w(i\Delta)$ and $K_i = K(i\Delta)$.
We are to apply the test R^* in the white noise model $Y = Ax + \sigma\xi$ with $x = [\mu; w]$, $A = [A_0, A_0]$, where A_0 is a Toeplitz matrix with $A_{ij} = K_{i-j}$.

We assume that

- the nuisance w satisfies $\|w''\|_\infty < L$, $\|w\|_\infty < L$. It amounts to the convex compact polyhedral set

$$N = \left\{ w : \left| \frac{1}{\Delta^2}[w_i - 2w_{i+1} + w_{i+2}] \right| \leq L,\ \ |w_i| \leq L \right\};$$

- the useful signal μ satisfies: $\mu \in S = \{0 \leq \mu < \bar{\mu},\ \sum_i \mu_i \leq 1\}$.

Application to particle detector

Suppose that $i \in \{0, ..., M-1\}$ is fixed. We set

- $S_i(0) = \{\mu \in S : \ \mu_i = 0\}$, $X_i(0) = S_i(0) \times N$,
- $S_i(\rho) = \{\mu \in S : \ \mu_i \geq \rho > 0\}$, $X_i(\rho) = S_i(\rho) \times N$.

Given $0 \leq i \leq M-1$ we may want to decide between

$$H_0 : \ x \in X_i(0) \ \text{ and } \ H_1 : \ x \in X_i(\rho).$$

Note that solving (1) in this case amounts to solving a quadratic problem.

Application to particle detector

Let us consider a slightly different problem:

- Let us fix the confidence level $0 < \epsilon < 1$. We may be interested to find the rate of the test – the smallest $\rho = \rho_i$ such that the maximal risk of distinguishing between $X_i(0)$ and $X_i(\rho)$ is $\leq \epsilon$.

In order to solve the latter problem we have to find

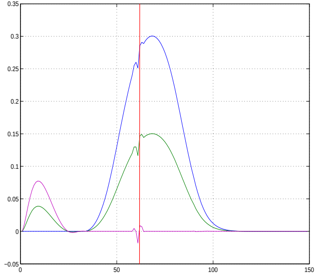
$$\begin{aligned} \rho^* &= \max\{\rho : \ dist(X_i(\rho), X_i(0)) \leq 2\sigma \mathrm{erfinv}(\epsilon/2)\} \\ &= \max_{\rho, x_0, x_1} \{\rho : \ \|Ax_0 - Ax_1\|_2 \leq 2\sigma \mathrm{erfinv}(\epsilon/2), \ x_0 \in X_i(0), x_1 \in X_i(\rho)\} \quad (2) \end{aligned}$$

(here erfinv stands for the inverse error function:
$\mathrm{erf}(\mathrm{erfinv}(p)) = (2\pi)^{-1/2} \int_{\mathrm{erfinv}(p)}^{\infty} e^{-\frac{t^2}{2}} dt = p$).

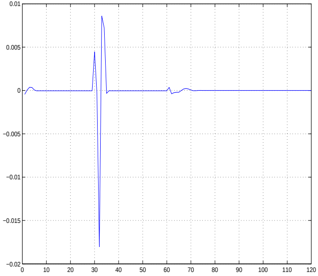
Observe that (2) is a convex conic problem.

Numerical illustration: particle detector

Detecting μ_{31}: $\sigma = 0.02$, $\epsilon = 0.05$.



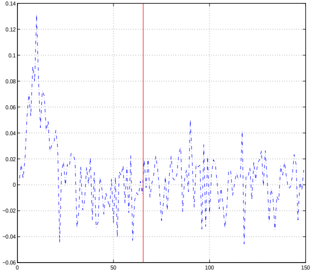
Response $A\bar{\mu}_1$, $A\bar{y} = A(\bar{\mu}_1 + \bar{w}_1)$
and $A(\bar{x} - \bar{y}) = A(\bar{\mu}_0 + \bar{w}_0) - A(\bar{\mu}_1 + \bar{w}_1)$



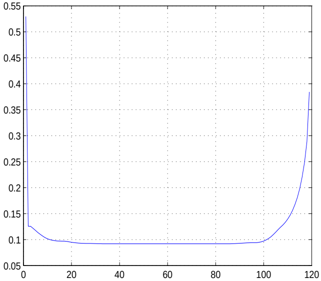
optimal detector ϕ

Numerical illustration: particle detector

Detecting μ: $\sigma = 0.02$, $\epsilon = 0.05$.



"Worst observation" for $\mu_{31} = 0.1$
(available from $t = 30$)



Detection threshold ρ^*
for μ_t as a function of t

Application: "jump" detection

Let us consider two hypotheses abou the signal $\mu \in \mathbf{R}^N$:

H_0 : $\mu = 0$,

H_1 : $\mu \in \mathcal{F}_N$. where $\mathcal{F} = \bigcup_{i=1}^N \{\mu_i \geq \rho_i^+ \text{ or } \mu_i \geq \rho_i^-\}$.

We are to distinguish between H_0 and H_1 given the observation

$$Y = A_0\mu + w + \sigma\xi,$$

where $A_0 \in \mathbf{R}^{M\times N}$, $w \in N \subset \mathbf{R}^M$, convex and compact, and $\xi \sim \mathcal{N}(0, \sigma^2 I)$, $\sigma > 0$ is known.

Let us fix $0 < \epsilon < 1$. We are interested in the test which attains the risk ϵ for the smallest values ρ_i^+, ρ_i^-, $i = 1, ..., N$.

Application: "jump" detection

Let us consider the following test procedure:

- For every $1 \leq i \leq N$ we test the hypothesis

$$H_0 : \ \mu = 0 \ \text{ against } \ H_i^+ : \ \mu_i \geq \rho_i^+, \ \mu_j = 0, \ j \neq i;$$
$$H_0 : \ \mu = 0 \ \text{ against } \ H_i^- : \ \mu_i \leq \rho_i^-, \ \mu_j = 0, \ j \neq i.$$

Clearly, if the risk of test of each hypothesis is $\leq \frac{\epsilon}{2N}$, then the risk of test of H_0 against $\cup_i H_i^{\pm}$ is $\leq \epsilon$.

The optimal ρ_i^+ for testing H_0 vs H_i^+ with risk $\frac{\epsilon}{2N}$ may be computed as follows:

$$\rho_i^+ = \max \left\{ \rho : \ \|\mu A_0 e_i + w - v\|_2 \leq 2\sigma \mathrm{erfinv}\left(\frac{\epsilon}{4N}\right), \ w, v \in N, \ 0 \leq \mu \leq \rho \right\} \quad (3)$$

(here e_i the i-th canonical orth of $\mathbf{R}^N$).

Application: "jump" detection

Let $[\bar{w}_i; \bar{v}_i]$ be an optimal solution to (3). Let

$$\bar{u}_i = \rho_i A_0 e_i + \bar{w}_i, \quad \phi_i = \frac{\bar{u}_i - \bar{v}_i}{\|\bar{u}_i - \bar{v}_i\|_2}.$$

Now consider the test φ_i^+ of H_0 against H_i^+:

$$\varphi_i^+(Y) = 1_{\phi_i^T(Y - \frac{\bar{u}+\bar{v}}{2}) \geq 0}.$$

Proposition *Let φ_* be the testing procedure based on multiple tests $\varphi_i^\pm$, $i = 1, ..., N$. Then for any test φ in the problem of testing H_0 vs H_1 it holds*

$$\rho_i^\pm(\varphi) \geq \frac{\mathrm{erfinv}\left(\frac{\epsilon}{2}\right)}{\mathrm{erfinv}\left(\frac{\epsilon}{4N}\right)} \rho_i^\pm(\varphi_*)$$

(here $\rho_i^\pm(\cdot)$ are the corresponding test thresholds).

Testing convex hypotheses

Let $\mathcal{P}_M = \{x \in \mathbf{R}^M : \; x \geq 0, \sum_i x_i = 1\}$.
Assume that we are given n independent realizations $Y = [Y_1; ...; Y_n]$ of a random variable $Y \sim x \in \mathcal{P}_M$.
Suppose that X_0 and X_1 are two closed convex subsets of $\mathcal{P}_M$. Given Y we want to distinguish between the hypotheses

$$H_0 : \; x \in X_0, \quad \text{and} \quad x \in X_1.$$

The risk of the decision rule $\phi(Y) \in \{0, 1\}$ is measured with the maximal risk:

$$\mathrm{Risk}(\phi) = \sup_{x \in X_0} P_x(\phi(Y) = 1) + \sup_{x \in X_1} P_x(\phi(Y) = 0).$$

Total variation test

Theorem *Denote for $x \in X_0$ and $y \in X_1$*

$$h(x,y) = \sum_i [x \wedge y] = 1 - \tfrac{1}{2}\|x-y\|_1.$$

Suppose that there exist probabilities

$$[\bar{x};\bar{y}] \in \mathrm{Argmax}\{h(x,y):\ x \in X_0,\ y \in X_1\}.$$

Then for any test $\varphi(Y)$ which decides between H_0 and H_1 given an observation Y,

$$\mathrm{Risk}(\varphi) \geq \mathrm{Risk}(\varphi^*) = h(\bar{x},\bar{y}),$$

where φ^* is the "likelihood ratio test"

$$\varphi^*(Y) = \phi_Y,\ \ \phi = [\phi_1, ..., \phi_M],\ \phi_i = 1_{\bar{x}_i \leq \bar{y}_i}.$$

Proof:

Note that $\phi \in \partial_x\{-h(\bar{x},\bar{y})\}$, so, by the optimality conditions,

$$\phi^T(x-\bar{x}) \le 0 \;\; \forall x \in X_0.$$

For the same reasons,

$$(1-\phi)^T(y-\bar{y}) \le 0 \;\; \forall y \in X_1.$$

Thus

$$\begin{aligned}\mathrm{Risk}(T^*) &= \max_{x\in X_0} P_x(\phi_Y=1) + \max_{y\in X_1} P_y(\phi_Y=0) = \max_{x\in X_0} \phi^T x + \max_{y\in X_1}(1-\phi)^T y \\ &\le \phi^T \bar{x} + (1-\phi)^T \bar{y} = h(\bar{x},\bar{y}).\end{aligned}$$

On the other hand, for any other decision rule $\varphi(Y)=\psi_Y$, $\psi_i \in \{0,1\}$, $i=1,...,M$,

$$\mathrm{Risk}(\varphi) \ge P_{\bar{x}}(\varphi=1) + P_{\bar{y}}(\varphi=0) = \psi^T\bar{x} + (1-\psi)^T\bar{y} \ge \sum_1^M [\bar{x}_i \wedge \bar{y}_i] = h(\bar{x},\bar{y}).$$

□

Bernstein Test

Assume we have n i.i.d. observations $Y = [Y_1, ..., Y_n]$.

The test we intend to use is as follows:

Test $\varphi_{\phi,c}$:

- choose a "weight vector" $\phi \in \mathbf{R}^N$ and a threshold $c \in \mathbf{R}$;
- set

$$\varphi(Y) = 1_{\sum_i \phi_{Y_i} \geq c}.$$

To choose the test parameters (ϕ, c) we rely on the Bernstein approximation of the risk of the test.

Bernstein Test

Let $y \in X_1$, the probability to accept H_0 does not exceed

$$E_y\left(\exp\{\sum_{k=1}^{n} \phi_{Y_k} - c\}\right) = e^{-c}\left(\sum_{i=1}^{N} y_i e^{\phi_i}\right)^n.$$

And we conclude that if $\epsilon \in (0,1)$, then the condition

$$n \max_{y \in X_1} \ln\left(\sum_{i=1}^{N} y_i e^{\phi_i}\right) - c \leq \ln(\epsilon)$$

is sufficient for the y-probability to accept H_0 to be $\leq \epsilon$. We write it

$$n \max_{y \in X_1} \ln\left(\sum_{i=1}^{N} y_i e^{\phi_i}\right) - c + \ln(1/\epsilon) \leq 0.$$

By similar argument we come to the condition

$$n \max_{x \in X_0} \ln\left(\sum_{i=1}^{N} x_i e^{-\phi_i}\right) - c + \ln(1/\epsilon) \leq 0$$

which guarantees $\max_{x \in X_0} P_x(\varphi(Y) = 1) \leq \epsilon$.

Bernstein's Test

We come to
Lemma *Assume that* $\phi \in \mathbf{R}^N$ *is such that*

$$F(\phi; x, y) := n \max_{x \in X_0} \ln \left(\sum_{i=1}^{N} x_i e^{-\phi_i} \right) + n \max_{y \in X_0} \ln \left(\sum_{i=1}^{N} y_i e^{\phi_i} \right) + 2 \ln(1/\epsilon) \leq 0.$$

When setting

$$c = \frac{1}{2} \left[n \max_{y \in X_0} \ln \left(\sum_{i=1}^{N} y_i e^{-\phi_i} \right) - n \max_{x \in X_0} \ln \left(\sum_{i=1}^{N} x_i e^{-\phi_i} \right) \right]$$

we ensure that $\mathrm{Risk}(\varphi) \leq 2\epsilon$.

Bernstein's Test

Observe that $F(\phi; x, y)$ is

- continuous and concave in $(x, y) \in X_0, X_1$,
- convex in $\phi \in \mathbf{R}^N$.

We have

$$\inf_{\phi} \max_{x \in X_0, y \in X_1} F(\phi; x, y) = \max_{x \in X_0, y \in X_1} \inf_{\phi} F(\phi; x, y).$$

It is immediately seen that

$$\inf_{\phi} F(\phi; x, y) = 2n \ln \left(\sum_i \sqrt{x_i y_i} \right) + 2 \ln(1/\epsilon).$$

When $x_i, y_i > 0$, the minimizer of F is given by

$$\phi_i = \frac{1}{2} \ln(y_i / x_i), \quad i = 1, ..., N.$$

When $x_i, y_i \geq \kappa > 0$, we have $|\phi_i| \leq -\frac{1}{4} \ln \kappa$.

Bernstein's Test

Proposition *Let* $\epsilon \in (0,1)$.
Suppose that all vectors in $X_0 \cup X_1$ *are strictly positive, so that*

$$\kappa \equiv \min_{x \in X_0, y \in X_1, i} \min[x_i, y_i] > 0,$$

and that

$$\max_{x \in X_0, y \in X_1} \left[n \ln \left(\sum_i \sqrt{x_i y_i} \right) \right] + \ln(1/\epsilon) \leq 0. \qquad (4)$$

Then
(i) *there is a test* $\varphi^*_{\phi,c}$ *with* $\|\phi\|_\infty \leq \frac{1}{4} \ln(1/\kappa)$ *with the risk* $\mathrm{Risk}(\varphi^*) \leq 2\epsilon$.

(ii) *Assume that* (4) *is not satisfied. Then one can find* $\bar{x} \in X_0$ *and* $\bar{y} \in X_1$ *such that for every test* φ *one has*

$$\mathrm{Risk}(\varphi) \geq P_{\bar{x}^n}\{\varphi(Y) = 1\} + P_{\bar{y}^n}\{\varphi(Y) = 0\} \geq \epsilon_+ \equiv \epsilon^2/2.$$

Proof:

We only need to prove (ii). Note that under the premise of (ii) there exist $\bar{x} \in X_0$ and $\bar{y} \in X_1$ such that

$$n \ln \left(\sum_i \sqrt{\bar{x}_i \bar{y}_i} \right) + \ln(1/\epsilon) \geq 0. \quad (5)$$

Assume that there exists a test φ such that

$$P_{\bar{y}}\{\varphi(Y) = 0\} + P_{\bar{x}}\{\varphi(Y) = 1\} < \epsilon_+.$$

Then by Neyman-Pearson lemma, the distributions $\bar{x}^n$ and $\bar{y}^n$ satisfy

$$h(\bar{x}^n_{i^n}, \bar{y}^n_{i^n}) = \sum_{i^n} [\bar{x}^n_{i^n} \wedge \bar{y}^n_{i^n}] \leq \epsilon_+.$$

If we denote $A = \{i^n : \bar{x}^n_{i^n} < \bar{y}^n_{i^n}\}$ and $A = \{i^n : \bar{y}^n_{i^n} \leq \bar{x}^n_{i^n}\}$

$$\left(\sum_i \sqrt{\bar{x}_i \bar{y}_i} \right)^n = \sum_{i^n} \sqrt{\bar{x}^n_{i^n} \bar{y}^n_{i^n}} = \sum_A \sqrt{\bar{x}^n_{i^n} \bar{y}^n_{i^n}} + \sum_B \sqrt{\bar{x}^n_{i^n} \bar{y}^n_{i^n}}$$

$$\leq \left(\sum_A \bar{x}^n_{i^n} \right)^{1/2} \left(\sum_A \bar{y}^n_{i^n} \right)^{1/2} + \left(\sum_B \bar{x}^n_{i^n} \right)^{1/2} \left(\sum_B \bar{y}^n_{i^n} \right)^{1/2} < \sqrt{2\epsilon_+} = \epsilon,$$

what is forbidden by (5).

Numerical illustration: convolution problem

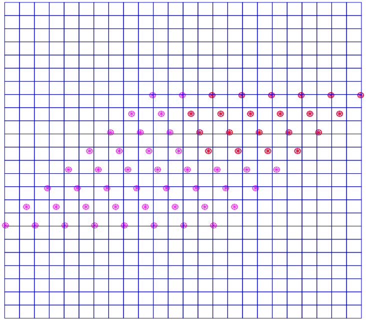
Suppose we observe the sum $\eta+\xi$ of two independent 2D random vectors, with $\xi \sim N(0, I_2)$ and η taking values on the 8×8-grid G.
Denote x the distribution of η ($x \in X$, the space of all probability distributions on G).
Let G_+ be the "non-negative part" of G belonging to the non-negative quadrant. We want to distinguish between

$$H_0 : x \in X_0 = \{x \in X : \sum_{g \in G_+} x_g \leq 0.4\}$$

and

$$H_1 : x \in X_1 = \{x \in X : \sum_{g \in G_+} x_g \geq 0.6\}.$$

Numerical illustration: convolution problem



For $N = 1000$ we have $\mathrm{Risk}(\varphi^*_{\phi,c}) \approx 0.00094$.

In an experiment with 20 000 samples with 10 000 randomly generated distributions in X_0 and the remaining 10 000 in X_1, the empirical probability to reject the hypothesis H_i when it is true was 10^{-3} for $i = 0$ and 0 for $i = 1$.

Estimating linear functionals

Our next objective is

- given an observation Y estimate the value at x of a given linear functional $g^T z$.

Definition of the estimation risk:

- given $\epsilon \in (0,1)$, we define the risk of an estimator $\widehat{g}(Y)$ of $g^T x$ on X as the width of maximal confidence interval of level $1-\epsilon$:

$$\mathrm{Risk}(\widehat{g};\epsilon) = \inf\left[\delta : \sup_{x\in X} P_x\left\{|\widehat{g}(Y) - g^T x| > \delta\right\} < \epsilon\right],$$

We are interested in the *minmax risk*

$$\mathrm{Risk}_*(\epsilon) = \inf_{\widehat{g}(\cdot)} \mathrm{Risk}(\widehat{g};\epsilon).$$

Linear functional estimation: white noise model

Let us consider *white noise model*

$$Y = Ax + \sigma\xi,$$

where

- $Y \in \mathbf{R}^M$ is observable;
- $x \in \mathbf{R}^N$ is the unknown parameter, $x \in X$, X being convex and compact;
- $A \in \mathbf{R}^{M\times N}$ is a given sensing matrix;
- $\xi \sim N(0, I_M)$ observation noise, $\sigma > 0$ is known.

We say that an *estimator* $\widehat{g}_{c,\phi}$ *is affine* if it is of the form

$$\widehat{g}_{c,\phi}(Y) = \phi^T Y + c, \quad \phi \in \mathbf{R}^M, \ c \in \mathbf{R}^r.$$

We would like to compare the minmax $\mathrm{Risk}_*(\epsilon)$ to the *maximal on X risk of the best affine estimator*:

$$\mathrm{RiskA}(\epsilon) = \inf_{c,\phi} \mathrm{Risk}(\widehat{g}_{c,\phi}; \epsilon).$$

White noise model : Donoho theorem

Thorem [Donoho, 1995] *In the white noise model*

$$\mathrm{RiskA}(\epsilon) \leq C\ \mathrm{Risk}_*(\epsilon)$$

avec $C \leq 1.25$ *pour* $\epsilon \leq 0.01$.

We prove an "extended" version of the result as follows:
Proposition *Let* $g(x, \xi) = g^T x + h^T \xi$. *There is an affine estimator* $\widehat{g}(Y)$ *of* $g(x, \xi)$ *such that*

$$\mathrm{Risk}(\widehat{g}; \epsilon) \leq \frac{\mathrm{erfinv}(\epsilon/2)}{\mathrm{erfinv}(\epsilon)} \mathrm{Risk}_*(\epsilon).$$

Proof of the proposition:

Note that $\forall u \in X$,

$$\begin{aligned} \widehat{g}(Y) - g(u,\xi) &= \phi^T Y + c - (g^T u + h^T u) = \phi^T Au - g^T u + (\sigma\phi - h)^T \xi \\ &\leq (A^T\phi - g)u + c + \mathrm{erfinv}(\epsilon/2)\|h - \sigma\phi\|_2 \end{aligned}$$

with probability $\epsilon/2$.
Thus, for $d = \max_{u\in X}(A^T\phi - g)u + c + \mathrm{erfinv}(\epsilon/2)\|h - \sigma\phi\|_2$,

$$\sup_{u\in X} P_x(\widehat{g}(Y) - g(u,\xi) \geq d) \leq \epsilon/2.$$

For the same reasons,

$$\sup_{v\in X} P_x(\widehat{g}(Y) - g(v,\xi) \leq -d') \leq \epsilon/2.$$

where $d' = \max_{v\in X}(g - A^T\phi)v - c + \mathrm{erfinv}(\epsilon/2)\|h - \sigma\phi\|_2$

Proof of the proposition:

Let us put

$$c = \frac{1}{2}\left[\max_{u\in X}(g - A^T\phi)x - \max_{v\in X}(A^T\phi - g)v\right].$$

With this choice of c we have $d = d'$ and

$$d = \frac{1}{2}\max_{u,v\in X}\left\{\Phi(\phi; u, v) := [g(v-u) + \phi^T A(u-v)] + 2\mathrm{erfinv}(\epsilon/2)\|h - \sigma\phi\|_2\right\}.$$

is the half-length of the confidence interval of $\widehat{g}(Y)$.
Observe that $\Phi(\phi; u, v)$ is convex in ϕ and concave in u, v, so that

$$\min_{\phi}\max_{u,v\in X}\Phi(\phi; u, v) = \max_{u,v\in X}\min_{\phi}\Phi(\phi; u, v)$$

The minimum in ϕ of the right-hand side is $-\infty$ unless $\|A(u-v)\|_2 \le 2\mathrm{erfinv}(\epsilon/2)\sigma$, and in the latter case we have

$$\max_{u,v\in X}\min_{\phi}\Phi(\phi; u, v) = \max_{u,v\in X} g^T(v-u) + h^T A(u-v).$$

Proof of the proposition:

Let now $\widehat{g}_*(Y) = \phi_*^T Y + c_*$ where ϕ_* is the corresponding minimizer. We conclude that $\mathrm{Risk}(\widehat{g}_*(Y), \epsilon) \leq \Phi_*$, where

$$\Phi_* = \max\left\{\frac{1}{2}(g - A^T h)^T (v - u) :\ u, v \in X,\ \|A(v-u)\|_2 \leq 2\mathrm{erfinv}(\epsilon/2)\sigma\right\}. \quad (6)$$

We need to show that

$$\Phi_* \leq \frac{\mathrm{erfinv}(\epsilon/2)}{\mathrm{erfinv}(\epsilon)} \mathrm{Risk}_*(\epsilon) = \psi(\epsilon)\,\mathrm{Risk}_*(\epsilon).$$

Let us assume, on the contrary, that $\rho = \mathrm{Risk}_*(\epsilon) < \psi(\epsilon)^{-1}\Phi_*$.

Let $[\bar{u}; \bar{v}]$ be an optimal solution to (6). We have $2\rho < (g - A^T h)^T (\bar{v} - \bar{u})$. Let now $[\widehat{u}; \widehat{v}]$ be a "contraction" of $[\bar{u}; \bar{v}]$ such that

$$2\rho = (g - A^T h)^T (\widehat{v} - \widehat{u}).$$

Then

$$\|A(\widehat{v} - \widehat{u})\|_2 = \underbrace{\frac{\rho}{\Phi_*}}_{\psi^{-1}} \|A(\bar{v} - \bar{u})\|_2 < 2\mathrm{erfinv}(\epsilon)\sigma.$$

Proof of the proposition:

Now let $H_0 : \; x = \widehat{u}$ and $H_1 : \; x = \widehat{v}$.

Let $\bar{g}(Y)$ be the minmax estimator of $g(u, \xi)$. We consider the following decision rule to decide between H_0 and H_1:

$$\bar{\varphi}(Y) = 1_{\bar{g}(Y) - h^T y - \widehat{c} \geq 0}, \;\; \widehat{c} = \frac{1}{2}(g - A^T h)^T (\widehat{v} + \widehat{u}).$$

The probability that $\bar{\varphi}(Y)$ rejects H_i, $i = 0, 1$ when it is true is $\leq \epsilon$.
On the other hand,

$$\|A\widehat{v} - A\widehat{u}\|_2 < 2\text{erfinv}(\epsilon)\sigma.$$

Now the Neyman-Pearson lemma states that there is no test for H_0 vs H_1 with the risk 2ϵ, and we arrive at a contradiction. □

Computational issues

In order to find $\widehat{g}$ one has to solve a conic convex problem (6):

$$\max\left\{\tfrac{1}{2}(g-A^Th)^T(v-u):\ u,v\in X,\ f(u,v):=\|A(v-u)\|_2\leq 2\mathrm{erfinv}(\epsilon/2)\sigma\right\}.$$

It is a "well structured" problem and can be solved very efficiently if X is *computationally tractable*.

An optimal solution $(\bar{u},\bar{v})$ to this problem can be augmented by Lagrange multiplier $\nu\geq 0$ such that the vectors

$$e_u=\frac{\partial}{\partial u}\Big|_{(u,v)=(\bar{u},\bar{v})}\left[\tfrac{1}{2}(g-A^Th)^Tu+\nu f(u,v)\right],$$
$$e_v=\frac{\partial}{\partial v}\Big|_{(u,v)=(\bar{u},\bar{v})}\left[-\tfrac{1}{2}(g-A^Th)^Tv+\nu f(u,v)\right]$$

belong to normal cones of X at the points $\bar{u}$, $\bar{v}$:

$$\forall(u,v\in X):e_u^T(u-\bar{u})\geq 0,\ e_v^T(v-\bar{v})\geq 0. \qquad (7)$$

Computational issues

There are two possible cases:

- $\nu = 0$: in this case the constraint $\|A(v-u)\|_2 \leq 2\mathrm{erfinv}(\epsilon/2)$ is not active, and

$$[\bar{u};\bar{v}] \in \mathrm{Argmax}\left\{(g - A^T h)^T (v-u) : \; u, v \in X\right\}.$$

  We have $\phi_* = 0$ and $c_* = (g - A^T h)^T(\bar{u} + \bar{v})$;

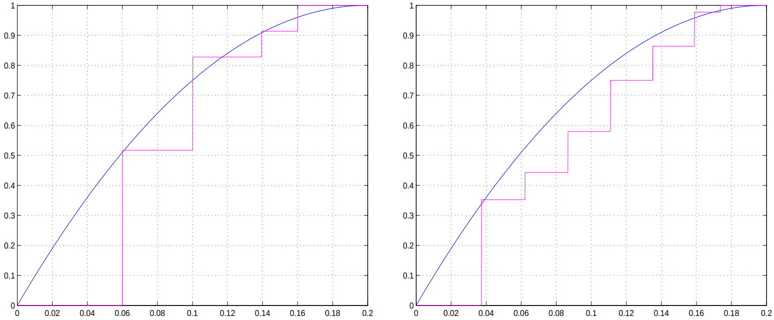
- $\nu > 0$: in this case we set

$$\phi_* = \frac{\nu}{\sigma \mathrm{erfinv}(\epsilon/2)} A(\bar{u} - \bar{v}) - \frac{2\nu}{\sigma} h, \quad c_* = \tfrac{1}{2}(g - A^T h)^T (\bar{u} + \bar{v}).$$

  One verifies straightforwardly that $\Phi(\phi_*; u, v)$ attains its maximum over $u, v \in X$ at the point $(\bar{u}, \bar{v})$, and the maximal value is exactly $2\Phi_*$.

Numerical illustration: particle detector

Estimation of μ: $\epsilon = 0.05$. C.d.f. of μ_i (blue) and its "plug-in" estimation (magenta):



$\sigma = 0.005$ $\sigma = 0.001$

Noisy distribution model

We assume that n i.i.d. realisations $Y_1, ..., Y_n$ of r.v. $Y \in \{1, ..., M\}$ are available. We have

$$P(Y = k) = p_k, \;\; p_k > 0, \;\; k = 1, ..., M,$$

where $p = Ax$, $x \in X$ (we suppose that X is compact and convex).

Our objective is to estimate estimate the value at x of a given linear functional $g^T z$.

We consider an affine (in empirical distribution) estimator

$$\widehat{g}_{c,\phi}(Y) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{M} \phi_{Y_i} + c = \sum_{k=1}^{M} \phi_k \widehat{p}_k + c = \phi^T \widehat{p} + c.$$

where $\widehat{p}_k = n^{-1} \sum_{i=1}^{n} 1_{Y_i = k}$, $k = 1, ..., M$.

To bound the risk of $\widehat{g}(Y)$ we use the "Bernstein device".

Risk of affine estimator

We have $\forall a \in \mathbf{R},\ s > 0$,

$$P_x(\widehat{g}(Y) - g^T x \geq a) \leq E_x \exp\left\{t^{-1}(\widehat{g}(Y) - g^T x) - t^{-1}a\right\}.$$

Thus, for

$$a(t) \geq t \ln\left[E_x \exp\left\{t^{-1}(\widehat{g}(Y) - g^T x)\right\}\right] + t\ln(2/\epsilon),$$

we have $P_x(\widehat{g}(Y) - g^T x \geq a(t)) \leq \epsilon/2$.
In particular, $\forall t > 0,\ x \in X,\ P_x(\widehat{g}(Y) - g^T x \geq \bar{a}(t))$ for

$$\begin{aligned}\bar{a}(t) &= \sup_{x \in X} t \ln\left[E_x \exp\left\{t^{-1}(\widehat{g}(Y) - g^T x)\right\}\right] + t\ln(2/\epsilon) \\ &= \sup_{x \in X} nt \ln\left[E_x e^{(tn)^{-1}\phi_Y}\right] - g^T x + c + t\ln(2/\epsilon) \\ &= \max_{x \in X} nt \ln\left[\sum_i [Ax]_i e^{(tn)^{-1}\phi_i}\right] - g^T x + c + t\ln(2/\epsilon).\end{aligned}$$

Similarly, $\forall t > 0,\ y \in X$ we have $P_y(\widehat{g}(Y) - g^T y \geq -\bar{b}(s))$ if

$$\bar{b}(s) = \max_{y \in X} nt \ln\left[\sum_i [Ay]_i e^{-(tn)^{-1}\phi_i}\right] + g^T y - c + t\ln(2/\epsilon).$$

Risk of affine estimator

If we set $c = \frac{1}{2}(\bar{a}(t) - \bar{b}(t))$, we obtain $\forall \phi \in \mathbf{R}^M,\ t > 0$:

$$\text{Risk}(\widehat{g}(Y);\epsilon) \le \frac{1}{2} \inf_{s,t>0} (\bar{a}(s) + \bar{b}(s))$$

Let us denote

$$\Phi(\phi, t; x, y) = nt \ln \left[\sum_i [Ax]_i e^{(tn)^{-1}\phi_i} \right] + nt \ln \left[\sum_i [Ay]_i e^{-(tn)^{-1}\phi_i} \right] + g^T(y-x) + 2t\ln(2/\epsilon)$$

Then

$$\text{Risk}(\widehat{g}(Y);\epsilon) \le \min_{t>0} \max_{x,y\in X} \frac{1}{2}\Phi(\phi, t; x, y)$$

We note that $\Phi(\phi, t; x, y)$ is convex in ϕ and t and concave in $x, y \in X$. Thus

$$\min_{\phi,t>0} \max_{x,y\in X} \Phi(\phi, t; x, y) = \max_{x,y\in X} \min_{\phi,t>0} \Phi(\phi, t; x, y)$$

Risk of affine estimator

The minimization in ϕ is immediate: the minimizer ϕ_* is given with,

$$[\phi_*]_i = \frac{tn}{2}\frac{[Ay]_i}{[Ax]_i} + \alpha,$$

and

$$\min_{\phi} \Phi(\phi, t; x, y) = 2t^{-1}\left(n \ln\left[\sum_{i=1}^{M}\sqrt{[Ax]_i[Ay]_i}\right] + 2\ln(2/\epsilon)\right) + g^T(y-x).$$

The minimum in t of the right-hand side is $-\infty$ if

$$f(x,y;\epsilon) := n\ln\left[\sum_{i=1}^{M}\sqrt{[Ax]_i[Ay]_i}\right] + \ln(2/\epsilon) < 0,$$

and is equal to $g^T(y-x)$ when $f(x,y;\epsilon) \geq 0$.
Let $\widehat{g}_*(Y) = \widehat{g}_{\phi_*,c_*}(Y)$. We come to the bound for the risk of $\widehat{g}_*(Y)$:

$$\mathrm{Risk}(\widehat{g}_*(Y);\epsilon) \leq \left\{\tfrac{1}{2}g^T(y-x) : \; f(x,y;\epsilon) \geq 0\right\}.$$

Risk of affine estimator

We come to the following result:

Theorem *Let $\epsilon < 1/4$. Then the risk of the affine estimator $\widehat{g}_*(\cdot)$ satisfies the relation*

$$\mathrm{Risk}(\widehat{g}_*(Y);\epsilon) \le \vartheta(\epsilon)\,\mathrm{Risk}_*(\epsilon), \quad \vartheta(\epsilon) = \frac{2\ln\left(\frac{2}{\epsilon}\right)}{\ln\left(\frac{1}{4\epsilon}\right)}$$

(note that $\vartheta(\epsilon) \to 2$ as $\epsilon \to +0$.

Proof: skipped?

Proof:

Let us denote

$$\Phi_*(\ln(2/\epsilon)) = \frac{1}{2} \min_{\phi, t>0} \max_{x,y \in X} \Phi(\phi, t; x, y).$$

Let us prove that

$$\mathrm{Risk}_*(\delta^2/4) \geq \Phi_*(\ln(1/\delta)).$$

To this end, suppose that $\mathrm{Risk}_*(\delta^2/4) < \Phi_*(\ln(1/\delta))$, so that there is $\delta' < \delta^2/4$ and $R' < \Phi_*(\ln(1/\delta))$, and an estimate $\widehat{g}$ such that

$$P_x(|\widehat{g}(Y) - g^T x| \geq R') < \delta', \quad \forall x \in X.$$

Now let $\bar{x}$ and $\bar{y}$ be the maximizers of $\Phi(\cdot)$ for $\epsilon = \delta$.

Proof:

Suppose that we are to decide between $H_0 : x = \bar{x}$ and $H_0 : x = \bar{y}$, and to this end we use the rule

$$\varphi(Y) = 1_{\widehat{g}(Y) > \frac{1}{2} g^T(\bar{x}+\bar{y})}.$$

Observe that since $R' < \Phi_*(\ln(2/\delta))$, $\mathrm{Risk}(\varphi) < 2\delta'$.

As we have already seen, the latter means that $h([A\bar{x}]^n, [A\bar{y}]^n) \leq 2\delta'$, and

$$\left(\sum_i \sqrt{[A\bar{x}]_i [A\bar{y}]_i}\right)^n \leq \sqrt{4\delta'} < \delta,$$

what is forbidden by the constraint

$$n \ln\left[\sum_i \sqrt{[A\bar{x}]_i [A\bar{y}]_i}\right] + \ln(1/\delta) \geq 0.$$

We conclude that for $\delta = 2\sqrt{\epsilon}$,

$$\mathrm{Risk}_*(\epsilon) \geq \mathrm{Risk}_*(\delta^2/4) \geq \Phi_*(\ln(1/\delta)) = \Phi_*\left(\frac{1}{2}\ln\left(\frac{1}{4\epsilon}\right)\right).$$

Proof:

Now we are done. Indeed, since

$$\Phi_*(\ln(2/\epsilon)) = \min_{t>0} \; \ldots \; + 2t^{-1}\ln(2/\epsilon),$$

we conclude that $\Phi_*(\ln(2/\epsilon))$ is a concave function of $\gamma = \ln(2/\epsilon)$.

Since $\Phi_*(0) = 0$, we have for $\ln(2/\delta) \leq \ln(2/\epsilon)$:

$$\frac{\Phi_*(\ln(2/\delta))}{\ln(2/\delta)} \geq \frac{\Phi_*(\ln(2/\epsilon))}{\ln(2/\epsilon)}.$$

In particular, for $\delta = 2\sqrt{\epsilon}$ we have

$$\Phi_*\left(\frac{1}{2}\ln\left(\frac{1}{4\epsilon}\right)\right) \geq \frac{\ln\frac{1}{4\epsilon}}{2\ln(2/\epsilon)}\Phi_*(\ln(2/\epsilon)).$$

□

Computational issues

In order to compute the estimator $\widehat{g}_*$ we have to

- Find a solution $[\bar{x}, \bar{y}]$ to the problem

$$\max\left\{\frac{1}{2}g^T(y-x):\; x,y\in X,\; n\ln\left[\sum_{i=1}^{M}\sqrt{[Ax]_i[Ay]_i}\right]+\ln(2/\epsilon)\geq 0\right\}.$$

pause The above problem is equivalent to the conic quadratic problem

$$\max_{x,y\in X}\left\{\frac{1}{2}g^T(y-x):\; x,y\in X,\;\sum_{i=1}^{M}\sqrt{[Ax]_i[Ay]_i}\geq(2/\epsilon)^{1/n}\right\}. \qquad (8)$$

- Let $[\bar{x};\bar{y}]$ be an optimal solution to (8). We set

$$[\phi_*]_i=\nu\left(\sum_{i=1}^{M}\sqrt{[A\bar{x}]_i[A\bar{y}]_i}\right)\ln\sqrt{\frac{[A\bar{y}]_i}{[A\bar{x}]_i}},\;\; c_*=\frac{1}{2}g^T(\bar{x}+\bar{y}),$$

where ν is the Lagrange multiplier of the conic constraint.

Example: case of direct observations

We are in the case $Ax = x$. To compute the minmax affine estimator $\widehat{g}_*$, we need the maximizers $\bar{x}$ and $\bar{y}$ of

$$\max\left\{\frac{1}{2}g^T(y-x):\ x,y\in X,\ \sum_{k=1}^N \sqrt{x_k y_k}\geq(\epsilon/2)^{1/n}\right\},\qquad (9)$$

then set

$$[\phi_*]_k=\nu\left(\sum_{i=1}^N\sqrt{x_i y_i}\right)\ln\sqrt{\frac{\bar{y}_k}{\bar{x}_k}},\quad c_*=\frac{1}{2}g^T(\bar{y}+\bar{x}).$$

and, finally,

$$\widehat{g}(Y)=\phi_*^T\widehat{p}+c_*,\ \ \widehat{p}_k=\frac{1}{n}\sum_{i=1}^n 1_{Y_i=k},\ k=1,...,N.$$

Recovering parameters of Bernoulli distribution

Suppose that each component Y^j of the random vector $Y \in \mathbf{R}^N$ is distributed according Bernoulli law with parameter x_j, $j = 1, ..., N$. We know that $x \in X$, where $X \in \mathbf{R}^N$ is a convex subset of the $[0,1]^N$ N-dimensional cube.

Our objective is to estimate the linear functional $g^T z$ at x, given n i.i.d. realizations $Y_1, ..., Y_n$ of Y.

Let us consider the optimization problem

$$\max\left\{\frac{1}{2}g^T(y-x): \ x, y \in X, \ -n\sum_{j=1}^{N} \ln\left[\sqrt{x_j y_j} + \sqrt{(1-x_j)(1-y_j)}\right] \le \ln\left(\frac{2}{\epsilon}\right)\right\}. \tag{10}$$

Recovering parameters of Bernoulli distribution

Let $[\bar{x};\bar{y}]$ be an optimal solution to (10). We set

$$c_* = \frac{1}{2} g^T(\bar{x}+\bar{y}), \quad \text{and} \quad \bar{\phi}_{j0} = n\nu \ln\sqrt{\frac{1-\bar{y}_j}{1-\bar{x}_j}}, \quad \bar{\phi}_{j1} = n\nu \ln\sqrt{\frac{\bar{y}_j}{\bar{x}_j}}, \; j=1,...,N,$$

where ν is the Lagrange multiplier of the conic constraint.

Finally, we put $\widehat{p}_j = \frac{1}{n}\sum_{i=1}^n 1_{Y_i^j=1}$, and

$$\widehat{g}_*(Y) \quad = \quad \sum_{j=1}^N \left(\bar{\phi}_{j0}(1-\widehat{p}_j) + \bar{\phi}_{j1}\widehat{p}_j\right) + c_*.$$

Theorem *Let $\epsilon \in (0, 1/4]$. The risk of the estimator $\widehat{g}_*$ satisfies:*

$$\mathrm{Risk}(\widehat{g}_*;\epsilon) \le \frac{2\ln\left(\frac{2}{\epsilon}\right)}{\ln\left(\frac{1}{4\epsilon}\right)} \mathrm{Risk}_*(\epsilon).$$

Recovering a parameter of Bernoulli distribution

ϵ	n	c	ϕ	upper risk bound	lower risk bound	ratio of bounds	$\vartheta(\epsilon)$
0.05	10	2.91e-1	4.18e-2	3.61e-1	2.49e-1	1.45	4.58
0.05	100	4.13e-2	9.17e-3	1.33e-1	8.19e-2	1.63	4.58
0.05	1000	4.29e-3	9.91e-4	4.29e-3	2.60e-3	1.65	4.58
0.01	10	3.58e-1	2.83e-2	4.04e-1	3.29e-1	1.23	3.29
0.01	100	5.83e-2	8.84e-2	1.59e-1	1.15e-1	1.38	3.29
0.01	1000	6.15e-3	9.88e-4	5.13e-2	3.67e-3	1.40	3.29
0.001	10	4.19e-1	1.61e-2	4.42e-1	3.98e-1	1.11	2.75
0.001	100	8.15e-2	8.37e-3	1.88e-1	1.51e-1	1.24	2.75
0.001	1000	8.79e-3	9.82e-4	6.14e-3	4.88e-3	1.26	2.75

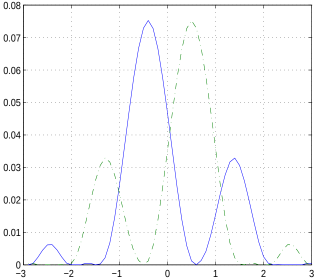
Numerical experiment: convolution model

Suppose that r.v. $\eta \sim F_\eta$, $\operatorname{supp}\eta \subseteq [-3,3]$, $\xi \sim F_\xi$, F_ξ is known. Our objective is, given $n = 100$ i.i.d observations Y_i of a r.v. $\eta + \xi$, recover $F_\eta(t)$ for a given $t \in [-3,3]$.

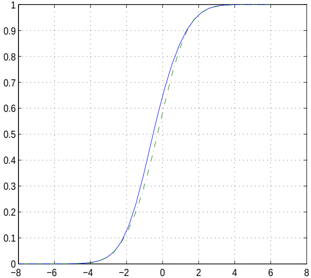
- We cut $[-3,3]$ into $N = 60$ equal segments Δ_i of length Δ, with the "sampled signal" $x_i = P(\eta \in \Delta_i)$. On suppose
$$\frac{x_i - 2x_{i-1} + x_{i-2}}{\Delta^2} \leq 1$$
(the density f_η of η is twice differentiable with $|f''_\eta(t)| \leq 1$).
- We cut the domaine of Y into $M = 42$ pieces δ_i (40 equal segment between $]Y_{\min}, Y_{\max}[$ and two "complements" in $\mathbf{R}$. The distribution of $\eta + \xi$ is approximated with $p = Ax$, where
$$A_{ij} = \Delta_i^{-1} \int_{v \in \delta_i} \int_{u \in \Delta_j} f_\xi(v-u)dv\,du.$$
- We compute the empirical probabilities
$$\widehat{p}_k = \frac{1}{n} \sum_{i=1}^{n} 1_{Y_i \in \delta_k}.$$

Numerical experiment: convolution model

Estimation of c.d.f. F_η, $n = 100$ observations, $\epsilon = 0.05$.



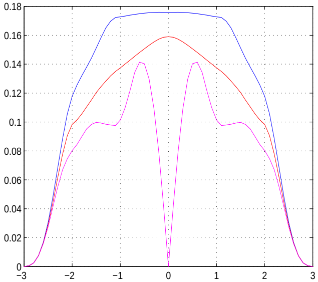
Optimal $\bar{x}$ et $\bar{y}$ signals
for estimation of $F_\eta(0)$
normal nuisance ξ



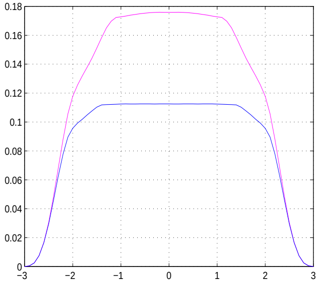
Corresponding
c.d.f. of $A\bar{x}$ and of $A\bar{y}$

Numerical experiment: convolution model

Estimation of F_η, $n = 100$ observations, $\epsilon = 0.05$.



Estimation risk
blue: minmax affine estimator
red: constrained to $E\eta = 0$
magenta: under symmetry constraint



Estimation risk
magenta: normale noise ξ
blue: Laplace noise ξ