On Consistent Hypotheses Testing

Mikhail Ermakov
St.Petersburg
E-mail: erm2512@gmail.com

History

is rather distinctive.

Stein (1954); Hodges (1954); Hoefding and Wolfowitz (1956), Le Cam and Schwartz (1960).

special original setups:
Shepp, Cover, Zeitouni, Dembo and Peres

Modern semiparametric and nonparametric problems:
Bickel and Rosenblatt, Donoho,De Vroye and Lugoshi

Motivation

The problems of consistent estimation and consistent classification were rather well studied.

In hypothesis testing the situation is more complicated. The results have a disordered character.

The paper goal is to present a systematic viewpoint to this problem and to establish new results.

Abstract

In the talk we discuss the links between different types of consistency:
point-wise consistency and usual consistency, uniform consistency and strong consistency (discernibility).

On the base of these results the sufficient conditions and necessary conditions for existence of consistent tests (in above mentioned senses) are studied for the problems of hypotheses testing
on a probability measure of i.i.d.r.v.'s,
in semiparametric problems,
on a mean measure of Poisson process,
on solutions of linear ill-posed problem in Hilbert space if the noise is Gaussian,
on deconvolution problem
and in the problem of signal detection in Gaussian white noise.

The main interest represents the necessary conditions.

Basic technique

Let we have random observation X with pm P defined on probability space $(\Omega, \mathfrak{F})$ and let we want to test the hypothesis

$$H_0 : P \in \Theta_0$$

versus alternative

$$H_1 : P \in \Theta_1$$

We define the test $K(X), 0 \leq K(X) \leq 1$ such that our decision is

$$H_0 : P \in \Theta_0 \quad \text{with probability} \quad 1 - K(X)$$

and

$$H_1 : P \in \Theta_1 \quad \text{with probability} \quad K(X)$$

Probability errors equal

$$\alpha_P(K) = E_P[K], \qquad P \in \Theta_0$$

and

$$\beta_Q(K) = E_Q[1 - K], \qquad Q \in \Theta_1$$

We can always define the test $K(X) \equiv \alpha$ and get

$$\alpha_P(K) + \beta_Q(K) = 1 \quad \text{for all} \quad P \in \Theta_0, Q \in \Theta_1$$

Thus it is of interest to search for the test K such that

$$\alpha_P(K) + \beta_Q(K) < 1 - \delta, \quad \delta > 0 \quad \text{for all} \quad P \in \Theta_0, Q \in \Theta_1$$

or, other words,

$$\int K dP + \int (1 - K) dQ < 1 - \delta \tag{1}$$

In this case we say that the hypotheses and alternatives are weakly distinguishable.

For any $\epsilon > 0$ we can approximate the function K by simple function

$$K_0(x) = \sum_{i=1}^{k} c_i 1_{A_i}(x), \quad x \in \Omega$$

such that

$$|K(x) - K_0(x)| < \epsilon, \quad x \in \Omega.$$

Here $\{A_1, \ldots, A_k\}$ is a partition of Ω.
Therefore, by (1), we get

$$\sum_{i=1}^{k} c_i(P(A_i) - Q(A_i)) \le 2\epsilon - \delta \qquad (2)$$

Proposition

The hypothesis H_0 and alternative H_1 are weakly distinguishable if there is a partition $A_1, \dots, A_k$ of Ω such that the sets

$$V_0 = \{v = (v_1, \dots, v_k) : v_1 = P(A_1), \dots, v_k = P(A_k), P \in \Theta_0\} \subset R^k$$

and

$$V_1 = \{v = (v_1, \dots, v_k) : v_1 = Q(A_1), \dots, v_k = Q(A_k), Q \in \Theta_0\} \subset R^k$$

have disjoint closures.

Thus the problem of distinguishability becomes a finite parametric problem.
Le Cam (1973) has implemented similar reasoning implicitly for the proof of exponential decay of type I and type II error probabilities.

Exponential decay of type I and Type II error probabilities

The proposition reduce the problem of hypotheses testing on probability measure of i.i.d.r.v.'s to the problem of hypothesis testing on multinomial distribution. Hence we get exponential decay of type I and type II error probabilities (Schwartz (1965) and Le Cam (1973)).

There is a sequence of tests K_n and constant n_0 such that

$$\alpha(K_n) \le \exp\{-cn\} \quad \text{and} \quad \beta(K_n) \le \exp\{-cn\} \tag{3}$$

for all $n > n_0$.

Proposition allows also to establish the links between different types of consistency.

Setup

Let we be given a sequence of statistical experiments $\mathfrak{E}_n = (\Omega_n, \mathfrak{B}_n, \mathfrak{P}_n)$ where $(\Omega_n, \mathfrak{B}_n)$ is sample space with σ-fields of Borel sets $\mathfrak{B}_n$ and let $\mathfrak{P}_n = \{P_{\theta,n}, \theta \in \Theta\}$ be a sequence of probability measures.

One needs to test a hypothesis $H_0 : \theta \in \Theta_0 \subset \Theta$ versus alternative $H_1 : \theta \in \Theta_1 \subset \Theta$.

For any tests K_n denote $\alpha_\theta(K_n), \theta \in \Theta_0$, and $\beta_\theta(K_n), \theta \in \Theta_1$, their type I and type II error probabilities respectively.

Denote

$$\alpha(K_n) = \sup_{\theta \in \Theta_0} \alpha_\theta(K_n) \quad \text{and} \quad \beta(K_n) = \sup_{\theta \in \Theta_1} \beta_\theta(K_n).$$

Point-wise Consistency and Consistency

Tests K_ϵ are point-wise consistent (see Lehmann and Romano, van der Vaart), if

$$\limsup_{n\to\infty} \alpha(K_n, \theta_0) = 0, \quad \text{and} \quad \limsup_{n\to\infty} \beta(K_n, \theta_1) = 0$$

for all $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$.
Tests K_n, are consistent (Lehmann and Romano, van der Vaart), if

$$\limsup_{\epsilon\to 0} \alpha(K_n) = 0, \quad \text{and} \quad \limsup_{\epsilon\to 0} \beta(K_n, \theta_1) = 0$$

for all $\theta_1 \in \Theta_1$.

Uniform Consistency

Tests $K_n, \alpha(K_n) < \alpha$, are uniformly consistent if

$$\lim_{n\to\infty} \alpha(K_n) = 0 \quad \text{and} \lim_{n\to\infty} \beta(K_n) = 0.$$

for all $0 < \alpha < 1$.
Hypothesis H_0 and alternative H_1 are called distinguishable (Hoefding and Wolfowitz (1958)) if there is uniformly consistent tests.

In i.i.d.r.v.'s case, implementing Proposition, we get that weak distinguishability implies distinguishability (Hoefding and Wolfowitz (1958)).

Strong Consistency

Tests K_n are called discernible (Devroye, Lugosi (2002) and Dembo, Peres (1994)) or strong consistent (van der Vaart) if

$$P(K_n = 1 \quad \text{for only finitely many} \quad n) = 1 \quad \text{for all} \quad P \in \Theta_0 \quad (4)$$

and

$$P(K_n = 0 \quad \text{for only finitely many} \quad n) = 1 \quad \text{for all} \quad P \in \Theta_1. \quad (5)$$

If there are discernible tests then hypotheses and alternatives are called discernible

Link of consistency and uniform consistency

Theorem *There are consistent tests iff there are nested subsets* $\Theta_{1i} \subseteq \Theta_{1,i+1}, 1 \le i \le \infty$ *such that*

$$\Theta_1 = \cup_{i=1}^{\infty} \Theta_{1i}$$

and the set Θ_0 *of hypotheses and the set* Θ_{1i} *of alternatives are distinguishable for each* i.

Proof. Let K_i be consistent sequence of tests. Let $0 < \alpha, \beta < 1$ be such that $\alpha + \beta < 1$. For each i define the subsets $\Theta'_{1i} = \{P : \beta(K_i, P) \le \beta, \alpha(K) < \alpha, P \in \Theta_1\}$. The sets Θ_0 and Θ'_{1i} are weakly distinguishable and therefore they are distinguishable. It is easy to show that the sets Θ_0 and $\Theta_{1i} = \cup_{j=1}^{i} \Theta'_{1j}$ are also distinguishable.

The link of point-wise consistency and uniform consistency

There are point-wise consistent tests iff there are nested subsets $\Theta_{0i} \subseteq \Theta_{0,i+1}$ *and* $\Theta_{1i} \subseteq \Theta_{1,i+1}, 1 \le i \le \infty$ *such that*

$$\Theta_0 = \cup_{i=1}^{\infty} \Theta_{0i} \quad \textit{and} \quad \Theta_1 = \cup_{i=1}^{\infty} \Theta_{1i},$$

the sets Θ_{0i} *of hypotheses and* Θ_{1i} *of alternatives are distinguishable for each* i.

The link of strong consistency and point-wise consistency

There are strong consistent tests iff there are a point-wise consistent tests.

Uniform strong consistency

We say that tests K_n are uniformly strong consistent if

$$\lim_{k\to\infty} \sup_{P\in\Theta_0} P(K_n = 1 \text{ for all } n > k) = 0 \tag{6}$$

and

$$\lim_{k\to\infty} \sup_{P\in\Theta_1} P(K_n = 0 \text{ for all } n > k) = 0.$$

Theorem. *If hypothesis H_0 and alternative H_1 are distinguishable, then there are uniformly strong consistent tests.*

If there are consistent tests, then there are strong consistent tests satisfying (6).

Hypothesis testing on a probability measure of independent sample

Let $X_1, \ldots, X_n$ be i.i.d.r.v.'s on a probability space $(\Omega, \mathfrak{B}, P)$ where $\mathfrak{B}$ is σ-field of Borel sets on Hausdorff topological space Ω. Denote Λ the set of all probability measures on $(\Omega, \mathfrak{B})$.

Define the τ_Φ-topology as the weak topology generated by all measurable functions.

Le Cam and Schwartz Theorem (1960). *Let $A \subset \Lambda$ and $B \subset \Lambda$ are relatively compact in τ_Φ-topology. Then the set A of hypotheses and the set B of alternatives are distinguishable iff "there is a finite family $\{f_j, j = 1, \dots, m\}$ of measurable bounded functions on" Ω "such that*

$$\sup \left| \int f_j \, dP - \int f_j dQ \right| < 1 \qquad (7)$$

implies that either both P and Q are elements of A or both are elements of B".

τ-topology

The coarsest topology in Λ providing the continuous mapping

$$P \to P(A), \quad P \in \Lambda$$

for all measurable sets A is called the τ-topology or the topology of set-wise convergence on all Borel sets.
For any set $A \subset \Lambda$ denote $\mathfrak{cl}_\tau(A)$ the closure of A in τ-topology.

Compacts in τ-topology

If set Ψ is relatively compact in τ-topology, then, by Theorem 2.6 in Ganssler (1971), the set Ψ is equicontinuous and there exists probability measure ν such that $P << \nu$ for all $P \in \Psi$.

This implies that for any $\delta > 0$ there exists $\epsilon > 0$ such that, if $\nu(B) < \epsilon, B \in \mathfrak{B}$, then $P(B) < \delta$ for all $P \in \Psi$.

The set of densities $\mathfrak{F} = \left\{ f : f = \frac{dP}{d\nu}, P \in \Psi \right\}$ is uniformly integrable.

Simple version of Le Cam and Schwartz Theorem (1960)

.

Let Θ_0 *and* Θ_1 *be relatively compact in* τ *- topology. Then the hypothesis* H_0 *and alternative* H_1 *are distinguishable iff*

$$\mathfrak{cl}_\tau(\Theta_0) \cap \mathfrak{cl}_\tau(\Theta_1) = \emptyset.$$

Remark. If Ω is a metric space and set $\Theta \subset \Lambda$ is relatively compact in τ-topology then weak and τ-topologies coincide in Θ.

The main problem in the proof of Theorem is that the partition in Proposition may exist only for $\Omega^k, k > 1$.
The proof of Theorem is based on the statement that the map $P \to P \otimes P, P \in \Theta$ is continuous in τ- topology if Θ is relatively compact. Hence, if there is a partition for Ω^2 then such a partition exists also for Ω.

Example

Let ν is Lebesgue measure in $(0,1)$ and let we consider the problem of hypothesis testing on a density f of probability measure P. Let $H_0 : f(x) = 1$, $x \in (0,1)$ and $\Theta_1 = \{f_1, f_2, \dots\}$ with $f_i(x) = 1 + \sin(2\pi i x)$, $x \in (0,1)$; $i = 1, 2, \dots$.

For any measurable set $B \in \mathfrak{B}$ we have

$$\lim_{i \to \infty} \int_B f_i(x)\,dx = \int_B dx.$$

Therefore H_0 and H_1 are indistinguishable

Hypothesis testing on a value of functional

The sets of alternatives described Le Cam –Schwartz Theorem are very poor for nonparametric hypothesis testing. At the same time all traditional nonparametric problems admits semiparametric interpretation as the problems of hypothesis testing on a value of functional $T : \Lambda \to R^1$

$$H_0 : \ T(P) \in \Theta_0$$

versus

$$H_1 : \ T(P) \in \Theta_1$$

Semiparametric setup

If there exists compact set $K \subset \Lambda$ in the τ-topology such that $T(K) = [a, b], 0 \in a, b$ and T is continuous we can consider the problem of hypothesis testing in the following form

$$H_0 : \ P \in T^{(-1)}(\Theta_0) \cap K$$

versus

$$H_1 : \ P \in T^{(-1)}(\Theta_1) \cap K$$

After that we can implement Le Cam - Schwartz Theorem.

Example. Kolmogorov tests.

Let $\Omega = (0, 1)$ and let $T(P) = \max_{x \in (0,1)} |F(x) - x|$ where $F(x)$ is distribution function of probability measure $P \in \Lambda$. Define probability measures $P_u = P_0 + uG, 0 < u < 1$ where P_0 is Lebesgue measure and signed measure G has the density $dG/dP_0(x) = -1$ if $x \in (0, 1/2)$ and $dG/dP_0(x) = -1$ if $x \in (1/2, 1)$.

Uniform distances

Hoefding and Wolfowitz (1958) proposed classification of distances on consistent and uniformly consistent.
Let ρ be a distance on the set Λ of all probability measures. The distance ρ is consistent in Θ, if for each $\epsilon > 0$ and $P \in \Theta$, there holds

$$\lim_{n\to\infty} P(\rho(\hat{P}_n, P) > \epsilon) = 0. \tag{8}$$

The distance ρ is uniformly consistent in Θ if the convergence in (8) is uniform for P with $P \in \Theta$.

Let the map $T : \Lambda \to R^d$ be uniformly continuous for uniformly consistent distance ρ in Λ.
The problem is to test the hypothesis $T(P) \in \Theta_0 \subset R^d, P \in \Lambda$ versus alternative $H_1 : T(P) \in \Theta_1 \subset R^d, P \in \Lambda$.
For any set $A \subset \Psi$ denote $\mathfrak{cl}(A)$ the closure of A in R^d.

Theorem *i. Let Θ_0 be bounded. Then the hypothesis H_0 and alternative H_1 are distinguishable if $\mathfrak{cl}(\Theta_0) \cap \mathfrak{cl}(\Theta_1) = \emptyset$.*
ii. The condition $\mathfrak{cl}(\Theta_0) \cap \mathfrak{cl}(\Theta_1) = \emptyset$ is necessary if T satisfies the following type of differentiability assumption.

For each $P \in \Theta$ there are non-singular matrix D, signed measures $G_1, \dots, G_d$ and $\delta > 0$ such that $P + \sum_{i=1}^{d} u_i G_i \in \Lambda$ for all $\vec{u} = (u_1, \dots, u_d) : |\vec{u}| < \delta, \delta > 0$ and

$$\left| T\left(P + \sum_{i=1}^{d} u_i G_i\right) - T(P) - D\vec{u} \right| = o(|\vec{u}|) \qquad (9)$$

as $|\vec{u}| \to 0$.

Hypothesis testing on a mean measure of Poisson random process

Let we be given n independent realizations $\kappa_1, \ldots, \kappa_n$ of Poisson random process with mean measure P defined on Borel sets $\mathfrak{B}$ of Hausdorff space Ω. The problem is to test a hypothesis $H_0 : P \in \Theta_0 \subset \Theta$ versus $H_1 : P \in \Theta_1 \subset \Theta$ where Θ is the set of all measures $P, P(\Omega) < \infty$.

The all obtained results on existence of different types of consistent tests in i.i.d.r.v.'s model can be extended on this setup. Only one Theorem is provided below.

Theorem of Distinguishability

Let Θ_0 and Θ_0 be relatively compact in τ - topology. Then the hypothesis H_0 and alternative H_1 are distinguishable iff

$$\mathfrak{cl}_\tau(\Theta_0) \cap \mathfrak{cl}_\tau(\Theta_1) = \emptyset$$

.

Signal detection in L_2

Suppose we observe a realization of stochastic process $Y_\epsilon(t), t \in (0,1)$, defined by the stochastic differential equation

$$dY_\epsilon(t) = S(t)dt + \epsilon dw(t), \quad \epsilon > 0$$

where $S \in L_2(0,1)$ is unknown signal and $dw(t)$ is Gaussian white noise.
We wish to test the hypothesis $H_0 : S \in \Theta_0 \subset L_2(0,1)$ versus alternative $H_1 : S \in \Theta_1 \subset L_2(0,1)$.
The results are provided in terms of the *weak topology in* $L_2(0,1)$.

Theorem

i. Let Θ_0 and Θ_1 be bounded sets in L_2. Then H_0 and H_1 are distinguishable iff Θ_0 and Θ_1 have disjoint closures.

ii. Let Θ_0 be bounded set in L_2. Then there are consistent tests iff the sets Θ_0 and Θ_1 are contained in disjoint closed set and F_σ- set respectively.

iii. There are point-wise consistent tests iff the sets Θ_0 and Θ_1 are contained in disjoint F_σ-sets.

Hypothesis testing on a solution of ill-posed problem

In Hilbert space H we wish to test a hypothesis on a vector $\theta \in \Theta \subset H$ from the observed Gaussian random vector

$$Y = A\theta + \epsilon\xi.$$

Hereafter $A : H \to H$ is known operator and ξ is Gaussian random vector having known covariance operator $R : H \to H$ and $E\xi = 0$. For any operator $U : H \to H$ denote $\mathfrak{R}(U)$ the rangespace of U. Suppose that the nullspaces of A and R equal zero and $\mathfrak{R}(A) \subset \mathfrak{R}(R)$.

Theorem *Let the operator $R^{-1/2}A$ be bounded. Then the statements i.-iii. of previous Theorem hold for the weak topology in H.*

Hypothesis testing on a solution of deconvolution problem

Let we observe i.i.d.r.v.'s $Z_1,\dots,Z_n$ having density $h(z), z \in R^1$ with respect to Lebesgue measure. It is known that $Z_i = X_i + Y_i, 1 \le i \le n$ where $X_1,\dots,X_n$ and $Y_1,\dots,Y_n$ are i.i.d.r.v.'s with densities $f(x), x \in R^1$ and $g(y), y \in R^1$ respectively. The density g is known.

Let P be the probability measure of f. The problem is to test the hypothesis $H_0 : P \in \Theta_0$ versus the alternative $H_1 : P \in \Theta_1$ where $\Theta_0, \Theta_1 \subset \Lambda$.

Suppose $g \in L_2(R^1)$.

Denote

$$\hat{g}(\omega) = \int_{-\infty}^{\infty} \exp\{i\omega x\} g(y)\, dy, \quad \omega \in R^1.$$

Define the sets $\Psi_i = \{f : f = dP/dx, P \in \Theta_i\}$ with $i = 0, 1$.

Theorem

Suppose the sets Θ_0 *and* Θ_1 *are tight and the sets* Ψ_0 *and* Ψ_1 *are bounded in* $L_2(R^1)$. *Let*

$$\mathrm{essinf}_{\omega \in (-a,a)} |\hat{g}(\omega)| \neq 0$$

for all $a > 0$.

Then the statement i. of Theorem holds for the weak topology in $L_2(R^1)$.

THANK YOU FOR YOUR ATTENTION