ON SOME LOWER BOUNDS IN TESTING OF HYPOTHESES AND INFORMATION THEORY

Marat V. Burnashev (IITP, Moscow)

1. Introduction

Many information theory problems =
= investigation of decision error prob. when testing large number M of hypotheses.

- $\mathbf{X}$ – observation space.
- $P_1(\mathbf{x}), \ldots, P_M(\mathbf{x})$ – probability measures (hypotheses) on $\mathbf{X}$
- Want to test $P_1, \ldots, P_M$ using one observation $\mathbf{x}$.
- Consider minimax problem statement (not important).

If $D_1, \ldots, D_M$ – decisions regions: $\mathbf{x} \in D_i \Rightarrow P_i$

$$P_{\mathrm{e}}(\{D_i\}) = \max_i P_i(D_i^c)$$

and

$$P_{\mathrm{e}} = P_{\mathrm{e}}(\{P_i\}) = \min_{\{D_i\}} P_{\mathrm{e}}(\{D_i\})$$

– minimal error prob. when testing $P_1, \ldots, P_M$.

Denote

$$D_{ij} = \{\mathbf{x} : P_i(\mathbf{x}) > P_j(\mathbf{x})\}\,.$$

If use ML decision (optimal for average error prob.)

$$\Rightarrow \qquad D_i = \bigcap_{j \neq i} D_{ij}, \qquad D_i^c = \bigcup_{j \neq i} D_{ji}$$

and

$$P_i(D_i^c) = P_i\left(\bigcup_{j \neq i} D_{ji}\right)$$

$\Rightarrow$ how to evaluate ?

Denote $P_{\rm e}(P_i, P_j)$ –
– minimal error prob. when testing only P_i, P_j.

Known for $P_{\rm e}(\{P_i\})$:
"*two closest hypotheses lower bound*"
and "*union upper bound*"

$$\max_{j\neq i} P_{\rm e}(P_i, P_j) \leq P_{\rm e}(\{P_i\}) \leq \max_i \sum_{j\neq i} P_{\rm e}(P_i, P_j) \quad (1)$$

- Lower bound: Chapman-Robbins (1953), Elias (1955), Cramer-Rao, Ziv-Zakai (1963), etc.
- Upper bound: ML decision with $P\left(\cup_i A_i\right) \leq \sum_i P(A_i)$ Elias (1955)
- When both sides of (1) are close to each other ?

$\Rightarrow$ If for each i main contribution to $P_i(D_i^c)$ is given by *few* closest to P_i neighbors.

Closest $\Rightarrow$ e.g. on variational distance $\|P_i - P_j\|_1$.

"Sphere–packing" lower bound

- very important in inform. theory (never used in math. stat.).

- If $|\mathbf{X}|$ – "volume" of $\mathbf{X}$, $|D_i|$ – "volume" of D_i

$$\Rightarrow \qquad \sum_i |D_i| \le |\mathbf{X}|.$$

- Want to have $P_i(D_i) \ge 1-\varepsilon$ for each i and given $\varepsilon > 0$.
- Let $D_1(\varepsilon), \ldots, D_M(\varepsilon)$ – "smallest" such sets (perhaps, intersecting).
- If $|D_1(\varepsilon)| = \ldots = |D_M(\varepsilon)|$

$$\Rightarrow \qquad |D_1(\varepsilon)| \le \frac{|\mathbf{X}|}{M}.$$

$\Rightarrow$ Gives lower bound (*sphere–packing*) on $P_e(\{P_i\})$.

- Why *sphere–packing* ?

$\Rightarrow$ In some interesting examples sets $\{D_i(\varepsilon)\}$ = spheres (balls).

- Corresponds to ideal packing of $\mathbf{X}$ by spheric. regions $\{D_i(\varepsilon)\}$.

2. Better lower bounds

▪ "Two closest hypotheses" and "sphere–packing" lower bounds:
– based on different ideas;
– use simplest characteristics of $\{P_i\}$;
– (how to combine them ?).

▪ Essentially need to lower bound $P(\cup_i A_i)$.

▪ "*Two closest hypotheses lower bound*" – based on

$$P(\cup_i A_i) \geq \max_i P(A_i)$$

▪ To get better lower bound
⇒ need *more information* on mutual properties of $\{P_i\}$ (what kind of information ?).

▪ One natural approach:

Assume: each P_i has $\geq M_1$ neighbors $\{P_j\}$ on "closest" distance.

⇒ Natural to expect that

$$P_{\mathrm{e}}(\{P_i\}) \gtrsim M_1 \max_{j \neq i} P_{\mathrm{e}}(P_i, P_j). \qquad (2)$$

Remark. Generally, (2) – not valid without restrictions (even from order point of view)

Example. Observe process $X(t)$ of form $(\varepsilon \to 0)$

$$dX(t) = s_i(t)dt + \varepsilon dW(t), \quad 0 \le t \le T\,,$$

$\{s_1, \ldots, s_M\} \in L_2(0,T)$ – orthonormal system, $(s_i, s_j) = \delta_{ij}$. Then

$$P_e(s_1, s_2) \sim e^{-\frac{1}{4\varepsilon^2}},$$

$$\Rightarrow \qquad M \max_{j \neq i} P_{\mathrm{e}}(P_i, P_j) \sim M e^{-\frac{1}{4\varepsilon^2}}.$$

$\Rightarrow$ If (2) – valid $\Rightarrow M \lesssim e^{\frac{1}{4\varepsilon^2}}$.

On other hand, we have

$$P_e\{s_1, \ldots, s_M\} \to 0 \quad \text{if} \quad M \lesssim e^{\frac{1}{2\varepsilon^2}}$$

- Possible justification of (2)
(second order Bonferroni inequality)

$$P\left(\cup_i A_i\right) \geq \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j).$$

- Used Burnashev (1980's)
⇒ sometimes – reasonable results;
⇒ sometimes – not (large M_1, "heavy intersections" $A_i \cap A_j$);
⇒ does not give "sphere–packing" lower bound.
- Lower bound below – "weakened" version of (2).

To simplify formulas $\Rightarrow$ consider particular case.
n – given integer (length, sample size).

$$\mathbf{U} = \mathbf{X} = \{0,1\}^n \text{ – all binary vectors of length } n\,,$$
$$\mathbf{U} \text{ – “transmitted signal” space,}$$
$$\mathbf{X} \text{ – observation space}$$

For $\mathbf{x}, \mathbf{u} \in \{0,1\}^n$ define (*Hamming* distance)

$d(\mathbf{x}, \mathbf{u}) = \{$ # of positions where $\mathbf{x}$ and $\mathbf{u}$ are different$\}$

$0 < p < 1/2$ – given parameter, $q = 1 - p$
(channel error prob.).

Each measure P_i is defined by n-vector $\mathbf{u}_i \in \mathbf{U}$.

By definition

$$P_i(\mathbf{x}) = P(\mathbf{x}|\mathbf{u}_i) = q^n \left(\frac{p}{q}\right)^{d(\mathbf{u}_i,\mathbf{x})}$$

(i.e. provided $\boldsymbol{u}_i$ – transmitted and $\boldsymbol{x}$ – received.).
In words, for given $\mathbf{u}_i$ all components of $\mathbf{x}$ – independent, and

$$P(x = u|u) = q\,, \qquad P(x \neq u|u) = p\,.$$

Collection $\{\mathbf{u}_1, \ldots, \mathbf{u}_M\}$ called (M, n)–code.

For integer t and each i define the (*ambiguity*) set:

$$\mathbf{Z}_t(i) = \left\{ \boldsymbol{x} : \begin{array}{c} d(\boldsymbol{x}, \boldsymbol{u}_i) = t \text{ and there exists} \\ \boldsymbol{u}_j \neq \boldsymbol{u}_i \text{ with } d(\boldsymbol{u}_j, \boldsymbol{x}) = t \end{array} \right\}.$$

S t a t e m e n t. *For decision error prob.* P_{e} *the lower bound holds*

$$P_{\mathrm{e}} \geq \frac{q^n}{2M} \sum_{t=0}^{n} \left(\frac{p}{q}\right)^t \sum_{i=1}^{M} |\mathbf{Z}_t(i)| \,. \tag{3}$$

Remark 1. "Sphere–packing" lower bound follows from (3) (in some interesting cases).

Remark 2. (3) gives exact exponent order of best P_{e} for any R:
Because for optimal (ML) decoding there are equalities (for exponents) in all their derivation steps.

Examples. Nonparametric stat.:
if *Fano* inequality was used $\Rightarrow$ can be replaced by (3).

1. M. V. Burnashev, “ Code spectrum and reliability function: binary symmetric channel”, *Probl. Inform. Transm.*, vol. 42, no. 4, pp. 3–22, 2006.

2. M. V. Burnashev, “ Code spectrum and reliability function: Gaussian channel”, *Probl. Inform. Transm.*, vol. 43, no. 2, pp. 3–24, 2007.