Fano Method

Rui Wang

Thursday 6th July, 2017

Abstract

123

1 Basic framework

Throughout, we let $\mathcal{P}$ denote a class of distributions on a sample space $\mathcal{X}$, and let $\theta : \mathcal{P} \to \Theta$ denote a function defined on $\mathcal{P}$, that is, a mapping $P \mapsto \theta(P)$. The goal is to estimate the parameter $\theta(P)$ based on observations X_i drawn from the distribution P.

To evaluate the quality of an estimator $\hat{\theta}$, we let $\rho : \Theta \times \Theta \to \mathbb{R}_+$ denote a semimetric on the space Θ, which we use to measure the error of an estimator for the parameter θ, and let $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-decreasing function with $\Phi(0) = 0$.

From estimation to testing Given an index set $\mathcal{V}$ of finite cardinality, consider a family of distributions $\{P_v\}_{v\in\mathcal{V}}$ contained within $\mathcal{P}$. This family induces a collection of parameters $\{\theta(P_v)\}_{v\in\mathcal{V}}$; we call the family a 2δ-packing in the ρ-semimetric if

$$\rho(\theta(\theta(P_v), \theta(P_{v'}))) \geq 2\delta \quad \text{for all } v \neq v'.$$

We use this family to define the canonical hypothesis testing problem:

- first, nature chooses V according to the uniform distribution over $\mathcal{V}$;
- second, conditioned on the choice $V = v$, the random sample $X = X_1^n = (X_1, \ldots, X_n)$ is drawn from the n-fold product distribution P_v^n.

Proposition 1. *The minimax error has lower bound*

$$\mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho) \geq \Phi(\delta) \inf_{\Psi} \mathbb{P}(\Psi(X_1, \ldots, X_n) \neq V).$$

2 Metric entropy and packing numbers

- Covering number: $N(\delta, \Theta, \rho)$
- Metric entropy: $\log N(\delta, \Theta, \rho)$
- Packing number: $M(\delta, \Theta, \rho)$

Lemma 1. $M(2\delta, \Theta, \rho) \leq N(\delta, \Theta, \rho) \leq M(\delta, \Theta, \rho)$

Lemma 2 (Gilbert-Vershamov bound)**.** *Let $d \geq 1$. There is a subset $\mathcal{V}$ of the d-dimensional hypercube $\mathcal{H}_d = \{-1, 1\}^d$ of size $|\mathcal{V}| \geq \exp(d/8)$ such that the ℓ_1-distance*

$$\|v - v'\|_1 = 2\sum_{j=1}^{d} \mathbf{1}\{v_j \neq v'_j\} \geq \frac{d}{2}$$

for all $v \neq v'$ with $v, v' \in \mathcal{V}$.

Lemma 3. *Let $\|\cdot\|$ be any norm in $\mathbb{R}^d$. Let $\mathbb{B}$ denote the unit $\|\cdot\|$-ball in $\mathbb{R}^d$. Then*

$$\left(\frac{1}{\delta}\right)^d \leq N(\delta, \mathbb{B}, \|\cdot\|) \leq \left(1 + \frac{2}{\delta}\right)^d.$$

3 Fano inequality

Let V be a random variable taking values in a finite set $\mathcal{V}$, and assume that we observe a random variable X, and then must estimate or guess the true value of $\hat{V}$. That is, we have the Markov chain

$$V \to X \to \hat{V}.$$

Let the function $h_2(p) = -p\log p - (1-p)\log(1-p)$ denote the binary entropy.

Proposition 2 (Fano inequality)**.** *For any Markov chain $V \to X \to \hat{V}$, we have*

$$h_2(\Pr(\hat{V} \neq V)) + \Pr(\hat{V} \neq V)\log(|\mathcal{V}| - 1) \geq H(V|\hat{V}).$$

Proof. Let E be the indicator for the event that $\hat{X} \neq X$, that is, $E = 1$ if $\hat{V} \neq V$ and is 0 otherwise. Then we have

$$
\begin{aligned}
&H(V|\hat{V}) = H(V, E|\hat{V}) = H(V|E, \hat{V}) + H(E|\hat{V}) \\
=& \Pr(E = 0) \underbrace{H(V|E = 0, \hat{V})}_{0} + \Pr(E = 1)H(V|E = 1, \hat{V}) + H(E|\hat{V}) \\
\leq& \Pr(E = 1) \log(|\mathcal{V}| - 1) + H(E)
\end{aligned}
$$

□

Remark 1. During the proof, X is not needed.

Corollary 1. *Assume V is uniform on $\mathcal{V}$, then*

$$\Pr(\hat{V} \neq V) \geq 1 - \frac{I(V; X) + \log 2}{\log(|\mathcal{V}|)}.$$

Proof. Note that $h_2(\Pr(\hat{V} \neq V)) \leq \log 2$ and

$$H(V|\hat{V}) = H(V) - I(V; \hat{V}) \geq H(V) - I(V; X) = \log(|\mathcal{V}|) - I(V; X).$$

□

4 The classical (local) Fano method

Proposition 3. *Let $\{\theta(P_v)\}_{v \in V}$ be a 2δ-packing in the ρ-semimetric. Assume that V is uniform on the set $\mathcal{V}$, and conditional on $V = v$, we draw a sample $X \sim P_v$. Then the minimax risk has lower bound*

$$\mathfrak{M}(\theta(\mathcal{P}); \Phi \circ \rho) \geq \Phi(\delta) \left(1 - \frac{I(V; X) + \log 2}{\log |\mathcal{V}|}\right).$$

References