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 $\Phi(0) = 0$.

2 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 1 (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

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

□

3 Th classical (local) Fano method

References