Foundations of non-asymptotic statistics

Uncultured Tramp

September 7, 2016

Contents

1 Basic set-up

Sample size $n \in \mathbb{N}$,
Sample $Y : n \to \texttt{RamdomVariable}(\mathbb{R}^d, \mathcal{B}\mathbb{R}^d)$,
True Distribution $\mathbb{P} : \texttt{Probability}(\Omega, \mathcal{F})$,
Model (0) : $\forall k \in n \,.\, Y_k \sim \mathbb{P}$,
Parameter Space $\Theta : \texttt{Set}$,
Parametric Family $P : \Theta \to \texttt{Probability}(\Omega, \mathcal{F})$,
Specification (1) : $\exists \theta \in \Theta : \mathbb{P} = P_\theta$,
True Parameter $\theta^* \in \Theta : \mathbb{P} = P_{\theta^*}$,

Kullbeck-Leiber Divergence

$$KL :: \texttt{Density}(\mathbb{R}^d, \mathcal{B}\mathbb{R}^d) \to \texttt{Density}(\mathbb{R}^d, \mathcal{B}\mathbb{R}^d) \to \mathbb{R}$$

$$KL(f, g) = \int_{\operatorname{supp} f} f \ln\left(\frac{f}{g}\right) d\lambda$$

$$\texttt{KLIsZero} :: KL(f, g) = 0 \iff f = g \quad \text{a . e . } [\lambda]$$

Assume that (1) is false. Model is misspecified.
Best parametric assumption $\theta^* = \arg\min_{\theta\in\Theta} KL(\mathbb{P}, P_\theta)$
Log-Likelihood

$$L :: \Theta \to \mathbb{R}$$

$$L(\theta) = \ln f_{\overline{P}_\theta}(Y) = \sum_{i=1}^{n} \ln f_{P_\theta}(Y_i)$$

Maximal likelihood estimation estimation $\tilde{\theta} = \arg\min_{\theta\in\Theta} KL(\hat{\mathbb{P}}, P_\theta) = \arg\max_{\theta\in\Theta} L(\theta)$
Distribution estimation

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(y_i, \infty)}(x)$$

$$\texttt{Regular} ::?(\Theta \to \texttt{Probability}(\Omega, \mathcal{F}))$$

$$P : \texttt{Regular} \iff \exists C : \texttt{Closed}(\mathbb{R}^d) : \forall \theta \in \Theta \,.\, \operatorname{supp} P_\theta = C$$

Linear model $Y = \Psi\theta + \epsilon$ where $\Psi : \texttt{Matrix}(n \times p, \mathbb{R}), \epsilon \sim \mathcal{N}(0, \sigma^2 I)$
In case of linear model $\tilde{\theta} = (\Psi\Psi^\top)^{-1}\Psi^\top Y = \Psi^\dagger Y$

2 Fisher Information

Fisher Information
If Θ is a smooth manifold and L is smooth

$$F : \prod \theta \in \Theta \ . \ \mathcal{L}(T_\theta \Theta, T_\theta \Theta; \mathbb{R})$$

$$F(\theta) = \mathbb{E}_{\mathbb{P}} \nabla^2 L|_\theta$$

In case $\Theta = \mathbb{R}^p$ $F(\theta) : \texttt{Mat}(p \times p, \mathbb{R})$.
Assume Euclidean case and that $F(\theta^*)$ is non singular.
Entropy? $D = \left(F(\theta^*)\right)^{\frac{1}{2}}$
Estimation error vectors $\xi = D^{-1} \nabla L(\theta^*)$
For linear model

$$L(\theta) = c - \frac{1}{2\sigma^2} \|Y - \Psi\theta\|^2$$

$$F(\theta) = \frac{1}{\sigma^2} \Psi^\top \Psi$$

$$\xi = \frac{1}{\sigma} \left(\Psi^\top \Psi\right)^{-\frac{1}{2}} \Psi(Y - \Psi\theta^*)$$

Asymptotic result

$$\lim_{n \to \infty} D(\tilde{\theta} - \theta^*) \to_d \xi \sim \mathcal{N}(0, I)$$

Non-asymptotic result

$$\mathbb{P}\left(\|D(\tilde{\theta} - \theta^*) - \xi\|) \leq \diamondsuit(x)\right) \geq 1 - \mathrm{e}^{-x}$$

where

$$\diamondsuit(x) \leq \sqrt{\frac{(p+x)^2}{n}}$$

Concentration sets

$$\Theta_\circ :: \mathbb{R}_{++} \to \quad ?\Theta$$

$$\Theta_\circ(r) = \{\theta \in \Theta : \|D(\theta - \theta^*)\| \leq r\}$$

Concentration result

$$\mathbb{P}(\tilde{\theta} \in \Theta(\diamondsuit(x))) \leq 1 - \mathrm{e}^{-x}$$

3 Wilkes Phenomenon

Asymptotic result
$\lim_{n\to\infty} 2(L(\tilde{\theta}) - L(\theta^*)) \to_d \|\xi\|^2 \sim \chi^2(p)$
Non-asymptotic
$\mathbb{P}(|2L(\tilde{\theta}) - 2L(\theta^*) - \lambda\xi\lambda^2| \leq \triangle(x)) \leq 1 - \mathrm{e}^{-x}$

4 Bernstein-von Mizes