Frontiers of Statistics: Contraction theory for posterior distributions **Spring 2019**

Lecture 1: March 28

Lecturer: Anirban Bhattacharya & Debdeep Pati *Scribes: Satwik Acharyya*

Note: *LaTeX template courtesy of UC Berkeley EECS dept & CMU's convex optimization course taught by Ryan Tibshirani.*

Theorem 1.1. *Fix $\alpha \in (0,1)$. Then for any $\epsilon \in (0,1)$,*

$$\int \left\{ \frac{1}{n} D_{\alpha}^{(n)}(\theta, \theta_0) \right\} \Pi_{n,\alpha}(d\theta | X^{(n)}) \leq \frac{\alpha}{n(1-\alpha)} \int r_n(\theta, \theta_0) \rho(d\theta) + \frac{1}{n(1-\alpha)} D(\rho, \Pi_n) + \frac{1}{n(1-\alpha)} \log(1/\epsilon)$$

$\forall$ probability measure $\rho \ll \Pi$, with $\mathbb{P}_{\theta_0}^{(n)}$ probability at least $(1-\epsilon)$.

Let $T_1 = \int r_n(\theta, \theta_0)\rho(d\theta)$ and $T_2 = D(\rho \| \Pi_n)$.

Comments

1. There is a trade off going on in between T_1 and T_2. We need ρ to place more and more mass near $\theta^\star$ to make T_1 small. We have to set $\rho = \Pi_n$ to make T_2 small. Our choice of ρ [ρ : the prior Π_n restricted to $B_n(\theta^\star, \epsilon; \theta_0)$] provides an optimal trade off.
2. Let $\psi(\rho) = T_1 + T_2$. $\psi(\rho)$ is minimized (over all $\rho \ll \Pi_n$) when $\rho = \Pi_{n,\alpha}$ and $\psi(\Pi_{n,\alpha}) = -\log$ marginal likelihood under the fractional posterior.
3. The eventual bound (after you choose ρ as mentioned) only depends on the "prior concentration" $(\Pi_n[B_n(\theta^\star, \epsilon; \theta_0)])$. You do not need the metric entropy and testing conditions to obtain this risk bounds. Also unified analysis of well-specified and miss-specified models can be obtained. This in particular allows to work with heavy-tailed priors.