BvM for Mixtures

Stefan Franssen — *Oxford* — stefan.franssen@stats.ox.ac.uk
Joint work with Jeanne Nguyen and Aad van der Vaart

BvM for semiparametric mixtures

Dirichlet process mixtures

Mixture distributions are densities of the form

$$p_{\theta,F} := \int p_\theta(x|\, z)\, \mathrm{d}\, F(z) \tag{1}$$

where $p_\theta(x|\, z)$ is a given kernel, which has two parameters: θ, z. z can be seen as a marginal variable that we marginalize over given some distribution F. We will endow F with a Dirichlet process prior and θ with a parameter prior.

Bernstein-von Mises

Theorem 1. *If $\Pi_n(\theta \in \Theta_n, F \in \mathcal{F}_n|\, X_1, \ldots, X_n) \to 1$, in P_0^n-probability, and (2)-(3) hold, then*

$$\sqrt{n}(\theta - \hat{\theta}_n)|X^{(n)} \rightsquigarrow N(0, \tilde{I}_0^{-1}).$$

The likelihood condition

For given $(\theta, F) \in \Theta \times \mathcal{F}$, assume that there exists a map $t \mapsto F_t(\theta, F)$ from a given neighbourhood of $0 \in \mathbb{R}^d$ to $\mathcal{F}$ such that, for given measurable subsets $\Theta_n \subset \Theta$ and $\mathcal{F}_n \subset \mathcal{F}$,

$$\begin{aligned}\ell_n\Big(\theta + \frac{t}{\sqrt{n}}, F_{t/\sqrt{n}}(\theta, F)\Big) &- \ell_n(\theta, F) \\ &= t^T \mathbb{G}_n \tilde{\ell}_0 - t^T\big(\tilde{I}_0 + R_{n,1}(\theta, F)\big)\sqrt{n}(\theta - \theta_0) \\ &\quad - \frac{1}{2} t^T \tilde{I}_0 t + R_{n,2}(\theta, F),\end{aligned}$$

for a matrix-valued process $R_{n,1}$ and and scalar process $R_{n,2}$ such that

$$\sup_{\theta \in \Theta_n, F \in \mathcal{F}_n} \big\|R_{n,1}(\theta, F)\big\| + |R_{n,2}(\theta, F)| \xrightarrow{P_0^n} 0. \tag{2}$$

The change of measure condition

$$\frac{\int_{\Theta_n \times \mathcal{F}_n} e^{\ell_n\left(\theta - t/\sqrt{n}, F_{-t/\sqrt{n}}(\theta, F)\right)}\, d\Pi(\theta, F)}{\int_{\Theta_n \times \mathcal{F}_n} e^{\ell_n(\theta, F)}\, d\Pi(\theta, F)} \xrightarrow{P_0^n} 1. \tag{3}$$

Consistency by compactness

Lemma 2. *Assume the assumptions as in [1] and suppose that (θ_0, F_0) belongs to the Kullback-Leibler support of Π. Then the posterior distribution is consistent at (θ_0, F_0).*

Consistency by Glivenko-Centelli class

Lemma 3. *Suppose that $\{\log(p_{\theta,F})\}$ is Glivenko-Cantelli, (θ_0, F_0) is identifiable and that (θ_0, F_0) belongs to the Kullback-Leibler support of Π. Then the posterior distribution is consistent at (θ_0, F_0).*

Verifying the conditions for the Bernstein-von Mises

Broadly speaking, we can verify Condition (2) by using entropy conditions and Condition (3) by showing that the posterior for (θ, F) contracts at a rate faster than $n^{-1/4}$.

An example: Frailty models

We will now sketch how to verify the rates for frailty models. In a frailty model, the kernel is given by

$$p_{\theta_0}(x, y|\, z) = z^2 \theta_0 e^{-z(x+\theta y)}.$$

Restricting to compacts

Lemma 4. *Let F_0 be a probability distribution on $[0, \infty$ such that there exists a $\gamma > 0$ such that $\int z^\gamma + z^{-\gamma}\, d\, F_0(z) < \infty$. Then for every $0 < \epsilon < \frac{1}{4}$ there exists $\underline{m}_\epsilon, \overline{m}_\epsilon$ and a probability distribution F^* supported on $[\underline{m}_\epsilon, \overline{m}_\epsilon]$ such that*

$$\|p_{\theta_0,F_0} - p_{\theta_0,F^*}\|_1 < 4\epsilon.$$

Finite approximation

Lemma 5. *Let F be a finite measure on $[m, M]$ with $0 < m < M < \infty$. For every $0 < \epsilon < \frac{1}{2}$ there exists a discrete measure F_ϵ with $|F| = |F_\epsilon|$, with fewer than $c \log(\frac{M}{m}) \log(\frac{1}{\epsilon})$ support points, for some universal constant $C > 0$, such that*

$$\|p_{\theta,F} - p_{\theta,F_0}\| < |F|\epsilon.$$

Moreover, in every interval $I_i = [e^i m, e^{i+1} m]$ there are atoms unless $F(I_i) = 0$.

Bounding for general approximations

Lemma 6. *Let $(0, \infty) = \cup_{j=0}^\infty A_j$ be a partition and $F_N = \sum_{j=1}^N w_j \delta_{z_j}$ a probability measure with $Z_j \in A_j$ for $j = 1, \ldots, N$. Then*

$$\|p_{\theta_0,F} - p_{\theta_0,F_N}\|_1 \le 4 \max_{1 \le j \le N} \frac{diam A_j}{\min(a\colon a \in A_j)} + \sum_{j=1}^N |F(A_j) - w_j| + F(A_0).$$

Defining $\Theta_\epsilon \times \mathcal{F}_\epsilon$

Define

$$\Theta_\epsilon \times \mathcal{F}_\epsilon := \{\theta, F\colon \|\theta - \theta_0\| < \epsilon\theta_0, \sum_{i=1}^{N_\epsilon} |F(A_i) - F_0(A_i)| < \epsilon, \min_i F(A_i) > \frac{\epsilon^2}{2}\}$$

Bounded likelihood ratio

Lemma 7. *Let F_0 be a fixed probability distribution. Suppose that $\int z^{2\delta}\, d\, F_0(z) < \infty$ for some $\delta > 0$. Let $0 < \epsilon < \frac{1}{3}$. Then for all $(\theta, F) \in \Theta_\epsilon \times \mathcal{F}_\epsilon$*

$$P_{\theta_0,F_0}\left(\frac{p_{\theta_0,F_0}}{p_{\theta,F}}\right)^\delta \lesssim \frac{1}{\epsilon^\delta \underline{m}_\epsilon^{2\delta}}.$$

Prior mass bound

Lemma 8. *Let F_0 be a probability distribution on $[0, \infty)$ and $\theta_0 \in \mathbb{R}_{\ge 0}$. Suppose that there exists a $\gamma > 0$ such that $\int (z^{-\gamma} + z^\gamma)\, d\, F_0(z) < \infty$. For suitable center measures G, and priors π*

$$\Pi((\theta, F)\colon V_2 + KL(P_{\theta_0,F_0}; P_{\theta,F}) < \epsilon) \ge e^{-c \log(\epsilon)^3}.$$

UNIVERSITY OF OXFORD
DEPARTMENT OF STATISTICS

References

1. Kiefer, J. & Wolfowitz, J. Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. **27,** 887–906. ISSN: 0003-4851 (1956).
2. Franssen, S. E. M. P., Nguyen, J. & van der Vaart, A. W. *Bernstein-von Mises for Semiparametric Mixtures*