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 Species Sampling process prior and θ with a parameter prior.

Bernstein-von Mises

Theorem 1. *If $\Pi_n(\theta \in \Theta_n, F \in \mathcal{F}_n|\, X_1, \dots, X_n) \to 1$, 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

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

LAN expansion

Condition (2) can be split in a random and a deterministic part:

$$\mathbb{G}_n\Big[\sqrt{n}\log\frac{p_{\theta+n^{-1/2}t, F_{n^{-1/2}t}(\theta,F)}}{p_{\theta,F}} - t^T\tilde{\ell}_0\Big] = o_P(1), \tag{4}$$

$$nP_0\log\frac{p_{\theta+n^{-1/2}t, F_{n^{-1/2}t}(\theta,F)}}{p_{\theta,F}} = -t^T\big(\tilde{I}_0 + o_P(1)\big)\sqrt{n}(\theta - \theta_0) - \tfrac{1}{2}t^T\tilde{I}_0 t + o_P(1). \tag{5}$$

Lemma 2. *Suppose that the map $t \mapsto \ell(t; \theta, F)(x) := \log p_{\theta+n^{-1/2}t, F_{n^{-1/2}t}(\theta,F)}(x)$ is twice continuously differentiable in a neighbourhood of zero, for every $(\theta, F) \in \Theta_n \times \mathcal{F}_n$ and $x \in \mathcal{X}$.*

- *If the classes of functions $\{\dot{\ell}(t/\sqrt{n}; \theta, F) : \|t\| < 1, (\theta, F) \in \Theta_n \times \mathcal{F}_n\}$ are contained in a given P_0-Donsker class and $P_0\|\dot{\ell}(t_n/\sqrt{n}; \theta_n, F_n) - \tilde{\ell}_0\|^2 \to 0$, for every $\|t_n\| < 1$ and $(\theta_n, F_n) \in \Theta_n \times \mathcal{F}_n$, then (4) is valid.*
- *If $\|P_0\ddot{\ell}(t_n/\sqrt{n}; \theta_n, F_n) - \tilde{I}_0\| \to 0$, for every $\|t_n\| \le 1$ and $(\theta_n, F_n) \in \Theta_n \times \mathcal{F}_n$, then (5) is satisfied if also*

$$\sup_{\theta\in\Theta_n, F\in\mathcal{F}_n} \frac{\|P_0\dot{\ell}(0; \theta, F) + \tilde{I}_0(\theta - \theta_0)\|}{\|\theta - \theta_0\| + n^{-1/2}} \xrightarrow{P_0^n} 0.$$

Consistency

Consistency by compactness

Lemma 3. *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 4. *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).*

Change of Measure condition

Lemma 5. *Let Π be a a species sampling process with center measure G, such that G has density g. Let π be be a probability distribution on Θ with density h. Suppose that for all $t \in T$, with T a open neighbourhood of 0 Assumptions 6 to 8 hold. Then*

$$\frac{\iint_{\Theta_n, \mathcal{F}_n} e^{\ell_n\left(\theta + \frac{t}{\sqrt{n}}, F_{\frac{t}{\sqrt{n}}}(\theta, F)\right)}\, d\,\pi(\theta)\, d\,\Pi(F)}{\iint_{\Theta_n, \mathcal{F}_n} e^{\ell_n(\theta, F)}\, d\,\pi(\theta)\, d\,\Pi(F)} \rightsquigarrow 1.$$

Assumption 6. *If F_t is the map as in Equation (3), we assume that $p_{\theta+t, F_t}(X) = \int p_{\theta+t}(x|\, \phi_{t,\theta}(z))\, d\,F(z)$*

Assumption 7. *For h the density of the prior π and G the center measure of the species sampling process with density g, we assume that there exists a open set U of θ and constants $C_t, C_t' > 0$ such that, for all $\theta \in U$ and all z*

$$\left\|\frac{g(\phi^{-1}_{\frac{t}{\sqrt{n}}, \theta - \frac{t}{\sqrt{n}}}(z))\, \left\|\det \dot{\phi}^{-1}_{\frac{t}{\sqrt{n}}, \theta - \frac{t}{\sqrt{n}}}(z)\right\|}{g(z)} - 1\right\| \le \frac{C_t}{\sqrt{n}}; \tag{6}$$

$$\left\|\frac{h(\theta + \frac{t}{\sqrt{n}})}{h(\theta)} - 1\right\| \le \frac{C_T'}{\sqrt{n}}. \tag{7}$$

Assumption 8. *We assume that for all t in a neighbourhood of 0, and for U a neighbourhood of θ_0,*

$$\pi \otimes \Pi_{\frac{t}{\sqrt{n}}, \theta}\left(\theta + \frac{t}{\sqrt{n}} \notin U, K_n \ge k_n|\, X^{(n)}\right) = o(1),$$

for some sequence $k_n = o(\sqrt{n})$.

Assumption 9. *Assume that for all t the following limit converges to 1.*

$$\frac{\iint_{\Theta_n, \mathcal{F}_n} e^{\ell_n\left(\theta + \frac{t}{\sqrt{n}}, F_{\frac{t}{\sqrt{n}}}(\theta, F)\right)}\, d\,\pi(\theta)\, d\,\Pi(F)}{\iint e^{\ell_n\left(\theta + \frac{t}{\sqrt{n}}, F_{\frac{t}{\sqrt{n}}}(\theta, F)\right)}\, d\,\pi(\theta)\, d\,\Pi(F)} \rightsquigarrow 1.$$

Examples

We give two explicit examples of models were all the assumptions can be verified:

Frailty model

In a frailty model, the kernel is given by

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

Errors-in-Variables model

The errors in variables model is defined by a kernel which is given by

$$p_\theta(x, y|\, z) = \phi_\sigma(x - z)\phi_\tau(y - \alpha - \beta z),$$

where ϕ_s is the density of the mean zero normal distribution with variance s^2.

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*