Contraction rates for generalized posteriors based on f-divergences: a diffusion process approach

Ioar Casado Telletxea, Enric Alberola Boloix

Basque Center for Applied Mathematics (BCAM)

icasado@bcamath.org

Resumen. Recent advances in Bayesian statistics have led to methodological innovations that extend the scope of classical Bayesian inference. In particular, several alternatives and generalizations of the standard Bayesian posterior have been introduced to address its limitations. These approaches are commonly referred to under the broad term "generalized posteriors".
In its most general form [1], a generalized posterior is the solution to the following variational problem:

$$q^* = \arg\min_{q \in \mathcal{Q}} \left\{ \mathbb{E}_q \left[\sum_{i=1}^{n} \ell(\theta, x_i) \right] + D(q \| \pi) \right\}, \tag{1}$$

where $\mathcal{Q} \subseteq \mathcal{P}(\mathbb{R}^d)$, $\{x_i\}_{i=1}^n$ are observations in an Euclidean space $\mathcal{X}$, $\ell : \mathbb{R}^d \times \mathcal{X} \to \mathbb{R}_+$ is a loss function, $\pi \in \mathcal{P}(\mathbb{R}^d)$ is a prior, and $D : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}_+$ is a divergence measure. Observe that choosing $\mathcal{Q} = \mathcal{P}(\mathbb{R}^d)$, $\ell(\theta, x) = -\log p(x|\theta)$ for some likelihood $p(x|\theta)$, and $D = D_{\mathrm{KL}}$ (the Kullback-Leibler divergence) recovers the standard Bayesian posterior.
Beyond the most basic cases, theoretical analysis of generalized posteriors arising from (1) is still lacking, especially when $D \neq D_{\mathrm{KL}}$. In this work we study the finite-sample behavior of generalized posteriors defined via f-divergences [2], a broad class of divergence measures widely used in statistics. Our main technical contribution is to extend the stochastic differential equation (SDE) framework of [3] beyond the standard Bayesian posterior. Using those tools, we obtain non-asymptotic posterior contraction rates for f-divergence posteriors by bounding the moments of their associated SDEs. Our results yield nearly optimal rates and clarify how different divergence choices affect posterior concentration. Finally, we illustrate the general framework with concrete examples.

Palabras clave: generalized Bayesian inference; posterior contraction rates; f-divergences; diffusion processes.

References

[1] J. Knoblauch, J. Jewson and T. Damoulas (2022). An optimization-centric view on Bayes' rule: Reviewing and generalizing variational inference. *Journal of Machine Learning Research*, 23(132):1–109.

[2] I. Csiszár (1967). On information-type measure of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318.

[3] W. Mou, N. Ho, M. Wainwright, P. Bartlett and M. Jordan (2024). A diffusion process perspective on posterior contraction rates for parameters. *SIAM Journal on Mathematics of Data Science*, 6(2):553–577.

Indicar la preferencia (subrayar la opción elegida): póster o <u>charla</u>.

Indicar la preferencia (subrayar la opción elegida): <u>Lunes/Martes</u> o Jueves/Viernes.