A new approach to Bayesian consistency rates via Wasserstein dynamics

Emanuele Dolera, *Università di Pavia*

joint work with
Claudia Contardi, *Università di Pavia*
Stefano Favaro, *Università di Torino*
Matteo Giordano, *Università di Torino*
Edoardo Mainini, *Università di Genova*

This talk focuses on a topic in Mathematical Statistics known as *Bayesian consistency*. See [4, Ch. 6–10]. We present a new mathematical approach to obtain posterior contractions rates (PCRs), i.e. the speed at which the posterior distribution π_n concentrates on sequentially small neighborhoods of the true parameter, as the sample size goes to infinity.

We focus on statistical models presentable as a family $\{f_\theta\}_{\theta\in\Theta}$ of densities, with Θ included in some separable Banach space (possibly infinite-dimensional). In current literature, two main approaches have been developed to study rates of consistence (both Bayesian and freqentist): the former considers neighborhoods of the true density, say f_{θ_0}, in the space of densities (see, e.g., [5]); the latter takes account of neighborhoods of the true parameter, say θ_0, in the space Θ endowed with its natural norm-topology (see, e.g., [6]). We follow the latter approach by transferring the analysis to the space $\mathcal{P}_2(\Theta)$ of all probability measures (p.m.'s) on Θ with finite second moment, exploiting a well-known geodesic interpretation of the 2-Wasserstein $\mathcal{W}_2$ due to Benamou & Brenier. See [2] for our general theory. Actually, two main assumptions are at the core of our approach:

1) the existence of a Bayesian sufficient statistic;
2) a SLLN for such statistic.

Point 1) postulates the existence of a mapping $T_n = T_n(x_1,\dots,x_n)$, with values in some Banach space $\mathbb{B}$, for which π_n satisfies $\pi_n(\cdot|x_1,\dots,x_n) = \pi_n^*(\cdot|T_n)$ for some probability kernel $\pi_n^* : \mathcal{B}(\Theta)\times\mathbb{B}\to[0,1]$. Assumption 2) guarantees the existence of some $T_0\in\mathbb{B}$ such that $T_n(\xi_1,\dots,\xi_n)\to T_0$, when the ξ_i's are i.i.d. distributed according to f_{θ_0}. Thus, we connect the PCR with two other, more tractable, rates:

A) the speed at which $\pi_n^*(\cdot|T_0)\to\delta_{\theta_0}$ in $(\mathcal{P}_2(\Theta),\mathcal{W}_2)$;
B) the speed of $T_n(\xi_1,\dots,\xi_n)\to T_0$ in $L^1(\mathbb{B})$, i.e. $\mathbb{E}[\|T_n(\xi_1,\dots,\xi_n)-T_0\|]$, usually well-known.

To both quantify and exploit rates in A)-B), we develop some new mathematical tools which can be of independent interest. Concerning A), we investigate the *Laplace method* for approximating integrals in infinite-dimensional space. On the other hand, to bring the rates in B) into the game, we need a (local) Lipschitz-continuity of the map $\mathbb{B}\ni b\mapsto\pi_n^*(\cdot|b)\in\mathcal{P}_2(\Theta)$, previously investigated in [3]. Critical to this analysis is the behaviour of *weighted Poincaré-Wirtinger constants* in infinite-dimensional setting, with perturbed Gaussian and non-Gaussian weights, for which we present some new results.

The presentation of our new theory will be illustrated on some infinite-dimensional statistical models of interest such as: logistic-Gaussian model (see [2, Section 4.5]); white noise with Besov-Laplace priors (see [1]); Poisson-mixture models for Empirical Bayes (work in progress with CC and SF).

References

[1] DOLERA, E., FAVARO, S. AND GIORDANO, M. (2024). On strong posterior contraction rates for Besov-Laplace priors in the white noise model. *arXiv:2411.06981*

[2] DOLERA, E., FAVARO, S. AND MAININI, E. (2024). Strong posterior contraction rates via Wasserstein dynamics. *Probab. Theory Relat. Fields* **189**, 659–720.

[3] DOLERA, E. AND MAININI, E. (2023). Lipschitz continuity of probability kernels in the optimal transport framework. *Ann. Inst. H. Poincaré Probab. Statist.* **59**, 1778–1812.

[4] GHOSAL, S., AND VAN DER VAART, A.W. (2017). *Fundamentals of nonparametric Bayesian inference*. Cambridge University Press.

[5] GINÉ, E. AND NICKL, R. (2011). Rates of contraction for posterior distributions in L^r-metrics, $1 \leq r \leq \infty$. *The Annals of Statistics* **39**, 2883–2911.

[6] SRIPERUMBUDUR, B., FUKUMIZU, K., GRETTON, A, HYVÄRINEN, A. AND KUMAR, R. (2017). Density Estimation in Infinite Dimensional Exponential Families. *Journal of Machine Learning Research* **18**, 1–59.