Contributed Discussion

Stefano Rizzelli*

Preliminaries I congratulate the authors on this work, which stimulates reflection on the comparison between Bayesian and frequentist interval estimation. As they mention in Section 4, in many practical cases, the parameter which is assigned a prior by mixing a point mass and an absolutely continuous distribution, say ξ, is not unique and contributes together with others, say θ, to statistical modelling of the data. Typically, therein the inferential goal is estimating a functional of (ξ, θ), or predicting future observations through the predictive distribution, which incorporates the uncertainty as to whether or not a reduced model should be preferred to a more complex one through model averaging. Hereafter, the focus is on the former problem and possible discrepancies between a frequentist and a Bayesian approach are outlined (in informal terms).

Some heuristics Let's consider the prototypical setup of a sequence of observables $Y = (Y_1, \dots, Y_n)$ modelled through a parametric family of probability densities $\mathcal{M} = \{p^{(n)}_{\xi,\theta},\ \xi \in \Xi, \theta \in \Theta\}$, with $\Xi \subset \mathbb{R}$ and $\Theta \subset \mathbb{R}^k$. Assume that parameter (ξ, θ) is given a product prior distribution $\Pi = (w\delta_0 + (1-w)\Pi_\Xi) \otimes \Pi_\Theta$, where Π_Ξ and Π_Θ are absolutely continuous: this corresponds to giving mass w to submodel $\mathcal{M}_0 = \{p^{(n)}_{0,\theta},\ \theta \in \Theta\}$ and $1 - w$ to complete model $\mathcal{M}_1 = \{p^{(n)}_{\xi,\theta},\ \xi \in \Xi_0, \theta \in \Theta\}$ a priori, with $\Xi_0 = \Xi \setminus \{0\}$. Let (ξ^*, θ^*) be the true parameters: if $\xi^* = 0$, under regularity conditions the posterior probability of submodel $\mathcal{M}_0$ satisfies

$$\pi(\mathcal{M}_0|Y) = \frac{w \int_\Theta p^{(n)}_{0,\theta}(Y)\Pi_\Theta(\mathrm{d}\theta)}{w \int_\Theta p^{(n)}_{0,\theta}(Y)\Pi_\Theta(\mathrm{d}\theta) + (1-w) \int_{\Xi_0 \times \Theta} p^{(n)}_{\xi,\theta}(Y)\Pi_\Xi \otimes \Pi_\Theta(\mathrm{d}\xi, \mathrm{d}\theta)} = 1 + o_{\mathbb{P}}(1)$$

and, by Bernstein-von Mises theorem (van der Vaart, 1998, Ch. 10), the posterior of θ given Y and $\xi = 0$ merges in total variation with the Gaussian sequence $\mathcal{N}(\hat\theta_{0,n}, n^{-1}I_0^{-1})$, where

$$\hat\theta_{0,n} \in \operatorname{argmax}\{p^{(n)}_{0,\theta}(Y) :\ \theta \in \Theta\}, \quad I_0 = -n^{-1}\nabla^2_\theta \log p^{(n)}_{0,\theta}(Y)\big|_{\theta=\theta^*} + o_{\mathbb{P}}(1).$$

As a result, if $\phi \equiv \phi(\xi, \theta)$ is a real-valued continuously differentiable functional, its posterior distribution is asymptotically Gaussian with variance proportional to $J_0^T I_0^{-1} J_0$, where $J_0 = \nabla_\theta \phi(0, \theta)|_{\theta=\theta^*}$. Precisely, with $\mathcal{B}$ the Borel σ-field of $\mathbb{R}$, as $n \to \infty$

$$\sup_{B \in \mathcal{B}} \Big| \Pi\big(\sqrt{n}(\phi - \phi(0, \hat\theta_{0,n})) \in B|Y\big) - \mathcal{N}\big(B; 0, J_0^T I_0^{-1} J_0\big) \Big| = o_{\mathbb{P}}(1)$$

(e.g., Dombry et al., 2023, Section 2.3). Although the posterior of ϕ may give rise to credible intervals of whatever level of credibility using posterior quantiles, these may have

*Department of Statistical Science, University of Padova, stefano.rizzelli@unipd.it

substantially different amplitude than confidence intervals obtained through the maximum likelihood estimator $\phi(\hat{\xi}_n, \hat{\theta}_n)$ calculated on the full model, since $\sqrt{n}(\phi(\hat{\xi}_n, \hat{\theta}_n) - \phi(\xi^*, \theta^*)) \rightsquigarrow \mathcal{N}(0, J^T I^{-1} J)$, where now $J = \nabla_{\xi,\theta}\phi(\xi,\theta)|_{(\xi,\theta)=(\xi^*\theta^*)}$ and

$$(\hat{\xi}_n, \hat{\theta}_n) \in \operatorname{argmax}\big\{p^{(n)}_{\xi,\theta}(Y) : \xi \in \Xi, \theta \in \Theta\big\},$$
$$I = -n^{-1}\nabla^2_{\xi,\theta} \log p^{(n)}_{\xi,\theta}(Y)\big|_{(\xi,\theta)=(\xi^*,\theta^*)} + o_{\mathbb{P}}(1).$$

Instead, since $\sqrt{n}(\phi(0, \hat{\theta}_{0,n}) - \phi(\xi^*, \theta^*)) \rightsquigarrow \mathcal{N}(0, J_0^T I_0^{-1} J_0)$, a frequentist who correctly uses the reduced model relying on $\phi(0, \hat{\theta}_{0,n})$ ends up with an interval whose width is comparable to the Bayesian's: the two produce similar inferences, upon consistent model selection. These considerations concern a first-order asymptotic behaviour; a higher-order asymptotic analysis may give better indications for moderate sample sizes. Moreover, the priors of ξ and θ may not be assigned independently (as assumed for simplicity).

An interesting example Assume a Generalised Extreme Value (GEV) density is fitted to a sample of i.i.d. (univariate) periodic maxima, i.e.

$$p^{(n)}_{\xi,\theta}(Y) = \prod_{i=1}^{n} \frac{1}{\theta_2}\left(1 + \xi\frac{Y_i - \theta_1}{\theta_2}\right)_+^{-1/\xi-1}$$

with parameter spaces $\Xi = \mathbb{R}$ and $\Theta = \mathbb{R} \times (0, \infty)$. In this model, widely used for the analysis of extremes, positive and negative values of ξ yield heavy-tailed and short-tailed distributions, respectively, while $\xi = 0$ yields the Gumbel subclass with exponentially decaying tails. The nature of the observed phenomenon may suggest assigning a priori a certain mass w to the event $\xi = 0$ (Stephenson and Tawn, 2004). Interest is typically in the quantile $\phi(\xi, \theta) = \theta_1 + \theta_2((-\ln(1-p))^{-\xi} - 1)/\xi$, with $p \in (0, 1)$. When Y_i's are yearly maxima, it is commonly referred to as the *return level* and interpreted as the value of the observed phenomenon which is expected to occur or be exceeded every $1/p$ years. Due to the irregularity of the GEV model, the heuristics presented above do not apply directly to return level estimation, but one can conjecture that similar conclusions are in order. A further complication is that the GEV model is usually misspecified for a sample of block maxima, see Padoan and Rizzelli (2024, Sections 3.2 and 5) for results with smooth priors. Extending their comparison of frequentist and Bayesian interval estimation to priors with point masses could be an interesting exercise.

References

Dombry, C., Padoan, S. A., and Rizzelli, S. (2023). "Asymptotic theory for Bayesian inference and prediction: from the ordinary to a conditional Peaks-Over-Threshold method." arXiv:2310.06720. 973

Padoan, S. A. and Rizzelli, S. (2024). "Empirical Bayes inference for the block maxima method." *Bernoulli*, 30: 2154–2184. MR4746603. doi: https://doi.org/10.3150/23-bej1668. 974

Stephenson, A. and Tawn, J. (2004). "Bayesian inference for extremes: accounting for the three extremal types." *Extremes*, 7: 291–307. MR2212389. doi: https://doi.org/10.1007/s10687-004-3479-6. 974

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press. MR1652247. doi: https://doi.org/10.1017/CBO9780511802256. 973