Mixture data-dependent priors

Leonardo Egidi*, Francesco Pauli**, Nicola Torelli**
*Dipartimento di Scienze Statistiche, University of Padova (egidi@stat.unipd.it)
**Dipartimento di Scienze Economiche, Aziendali, Matematiche e Statistiche 'Bruno de Finetti' (DEAMS), University of Trieste

Why data in the prior?

In many cases it has been proposed to use a prior **dependent on the data**: in the Adjusted Data-Dependent Bayesian paradigm [1]; as an approximation of a hierarchical model [2]; for obtaining a proper posterior in finite mixture models [3].
We have a further motivation. When the sample size n is *small*, an overly informative prior could dominate the inference. Conversely, a noninformative prior neglects relevant information. Hence, *we need something in between*.

The proposal

We propose the **mixture data-dependent** (MDD) class of priors, consisting of mixtures of a baseline (flat) prior π_b, and an informative prior π; they are weighted in such a way to prefer π_b if an additional set of $\varkappa$ values generated under π appears to be far from the data at hand, as measured by ψ_{n*}, an indicator of surprise of the data with respect to the prior, where $n^* = n + \varkappa$. Then, the MDD prior is:

$$\varphi(\theta) = \psi_{n^*}\pi_b(\theta) + (1 - \psi_{n^*})\pi(\theta)$$

Let $\boldsymbol{y}_n = (y_1, \dots, y_n)$ be a sample from a given sampling distribution $f(\boldsymbol{y}_n|\theta)$, with $\theta \in \mathbb{R}$.

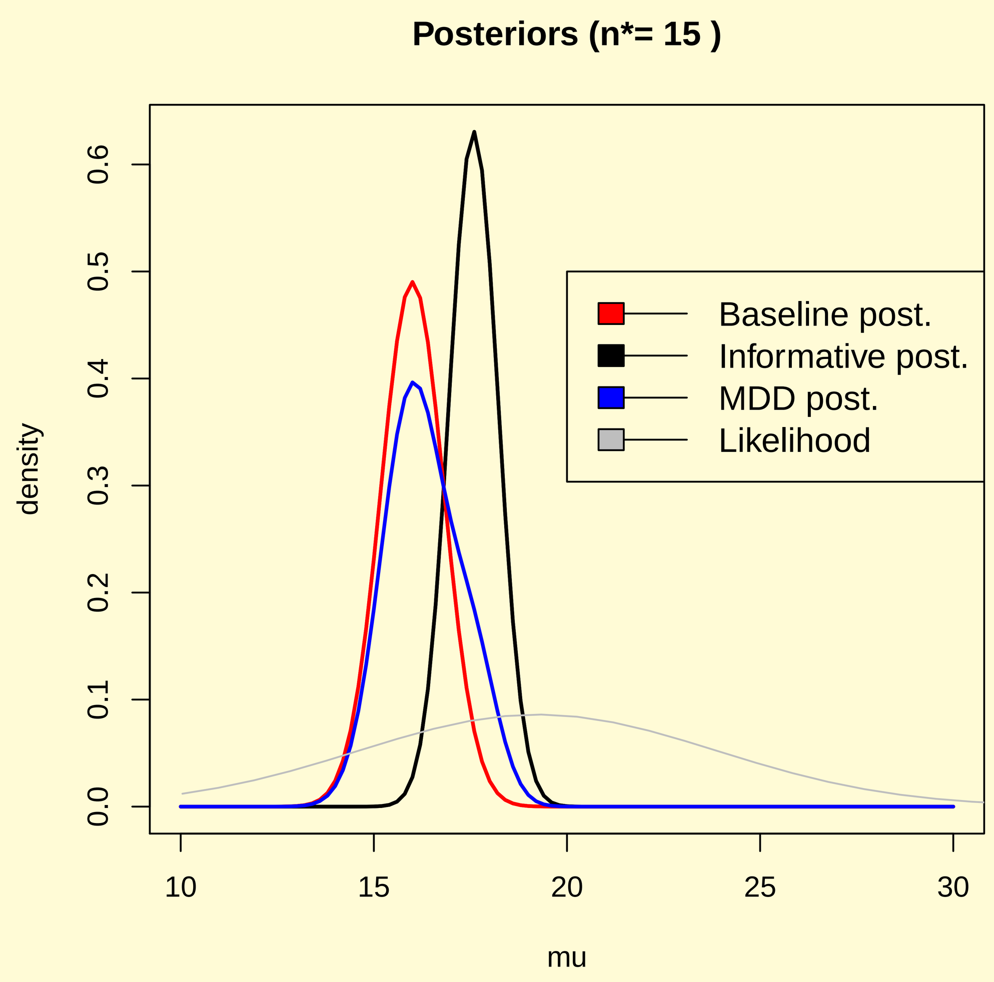
Resampling-algorithms:

→ generate $\theta^* \sim \pi(\theta)$, fix a tolerance ε.

→ sequentially generate new values from $f(\boldsymbol{y}_n|\theta^*)$ (MDD-Res1) or from $f(\boldsymbol{y}_n|\theta_0)$ (MDD-Res2), until the Hellinger distance between the baseline posterior and the informative posterior is greater than the fixed tolerance ε.

→ save the new sample size n^* and ψ_{n^*}, that is the observed value of the Hellinger distance between our data at hand and those generated through one of the two algorithms.

Theoretical results



We proved some **theoretical** results: **(1)**: the information provided by the MDD prior φ, measured by the effective sample size (ESS), is never greater than the information provided by π. **(2)** if the Fisher information of the model we are using doesn't depend on θ, then the observed Hellinger distance ψ_{n^*} is distribution-constant and the MDD prior reduces to choosing a genuine prior.

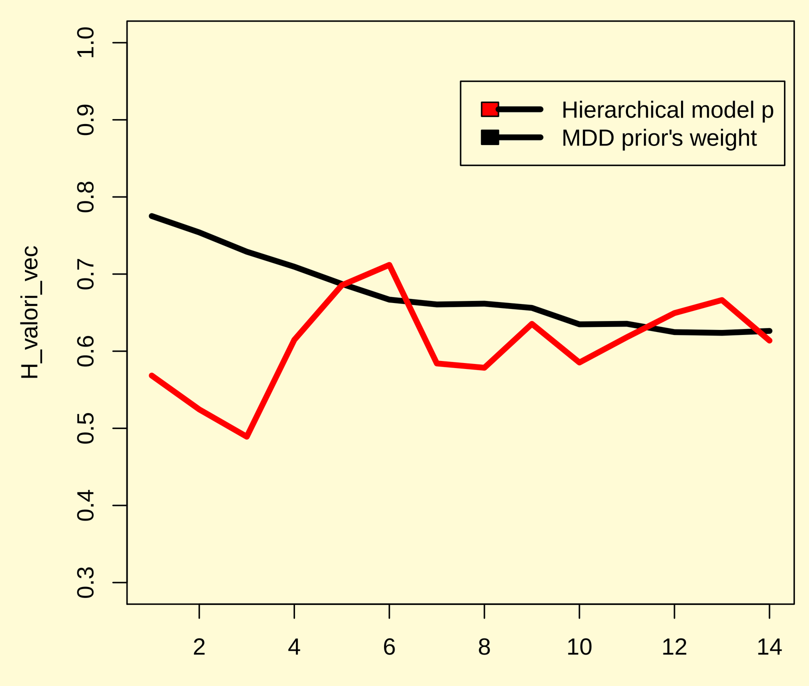
Hierarchical approximation

Following the motivations in [2], let consider the following hierarchical model:

$$y_{ij} \sim \mathcal{N}(\mu, \sigma^2_{j[i]}),\ i = 1 \dots n$$

$$\sigma^2_j = \begin{cases} \tau^2 \text{ with } p \\ c\tau^2 \text{ with } 1-p \end{cases},\ j = 1,2$$

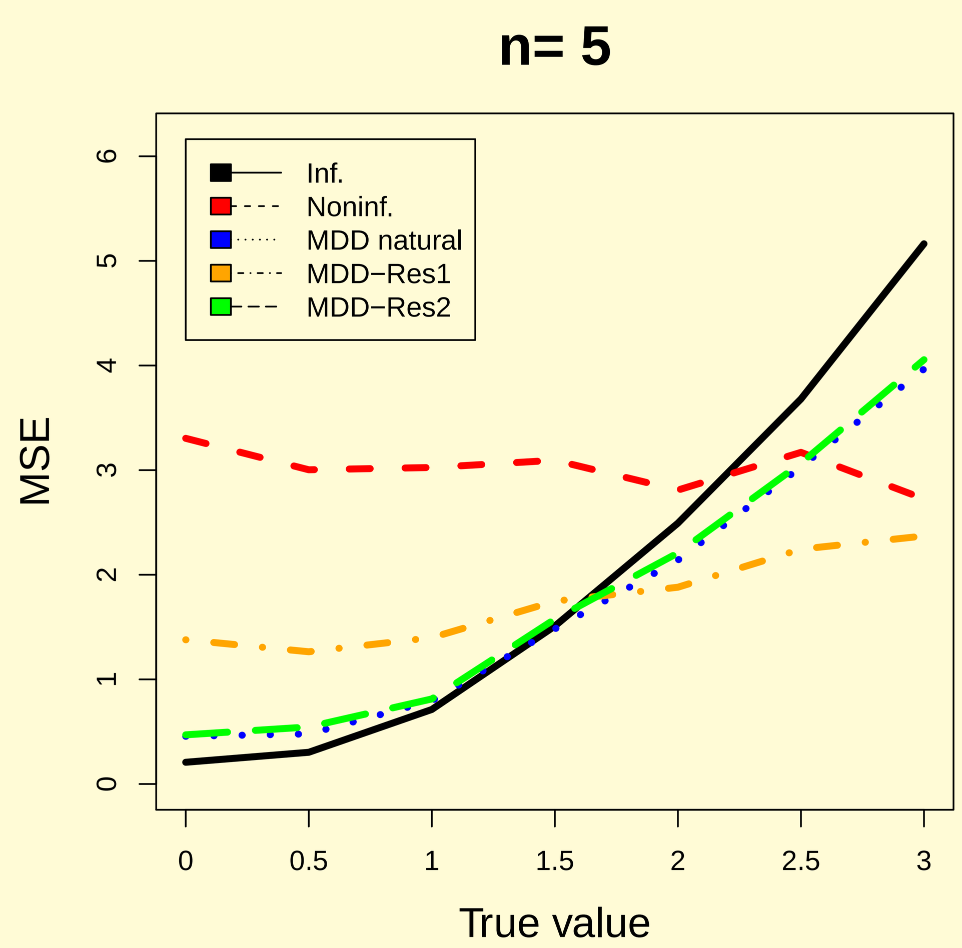
where the variance may assume two different values with probabilities p and $1-p$. The MDD class of prior is a natural approximation of the model above: a-priori, no value is assigned to the parameter p. Instead, it is estimated from the data through ψ_{n^*}.

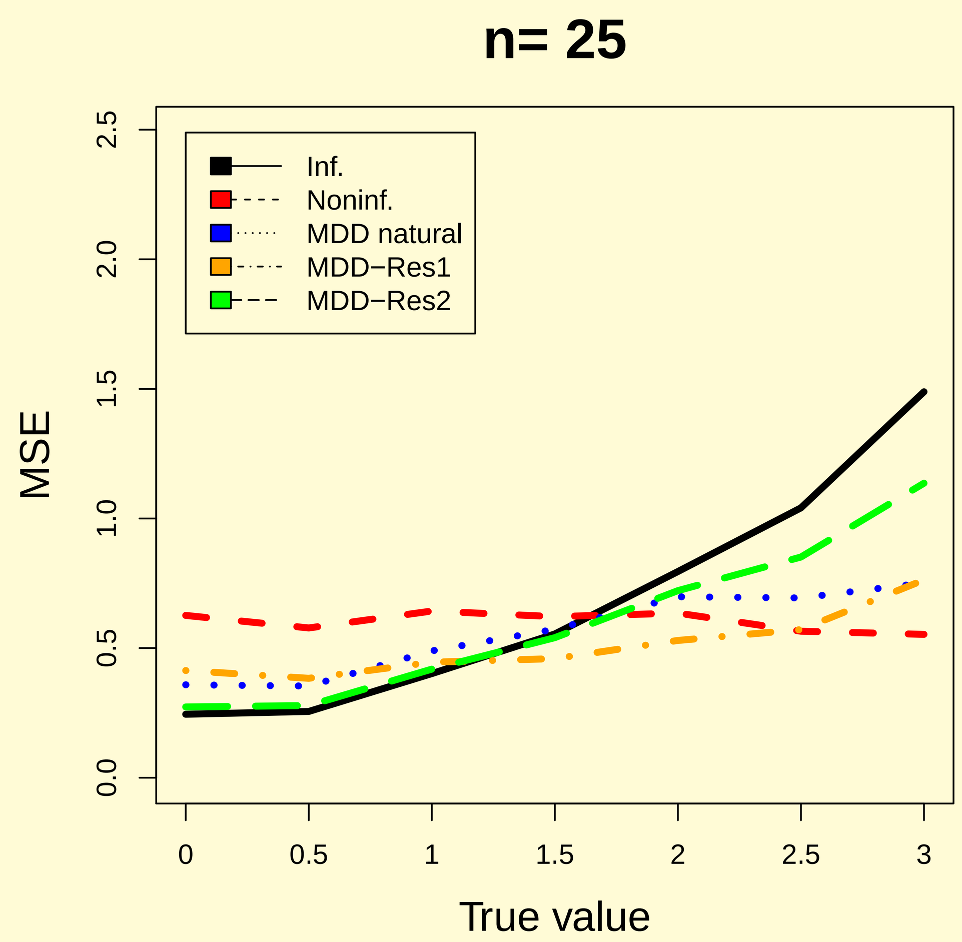


First evidences from simulation studies

We simulate $\boldsymbol{y}_n \sim \mathcal{N}(\theta_0, 15)$, and we specify $\pi_b(\theta) = \mathcal{N}(0, 1000)$, $\pi(\theta) = \mathcal{N}(0, 1)$.

- When $n = 5$, as the true parameter value increases (and thus the prior distribution is far from it), the point estimate (we take here the posterior median) obtained by using MDD-Res1 yields the lowest mean squared errors;
- When $n = 25$, the MSEs are more shrunk each other, but the pattern is similar. They appear to be lower than those obtained under $n = 5$;
- The MSEs for the estimates obtained with MDD-Res2, explicitly designed for tackling the model misspecification, in both cases exhibit a similar pattern to the one of informative prior;
- Note that the natural MDD is obtained without resampling, but just computing the distance between the informative prior and posterior.





Concluding remarks

→ Definition of a new class of data-dependent priors, that weights a non-informative and an informative prior.

→ Likely to not dominate the inference when the sample size is small.

→ Evidence from simulation studies shows good results in case of prior and/or model misspecification.

References

[1] William Francis Darnieder. *Bayesian methods for data-dependent priors.* PhD thesis, The Ohio State University, 2011.

[2] Andrew Gelman. Data-dependent prior as an approximation to hierarchical model. *Statistical modeling, causal inference, and social science blog*, 2016.

[3] Larry Wasserman. Asymptotic inference for mixture models by using data-dependent priors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):159–180, 2000.