Contributed Discussion

Marta Catalano* and Sirio Legramanti†

We congratulate the authors on this paper, in which they propose a distance-based Bayesian exponentially-tilted empirical likelihood (D-BETEL) approach. As a distance, they define a new variant of the Wasserstein metric, termed ANDREW, which is tailored to elliptical mixture models.

The proposed D-BETEL may be viewed as a likelihood-free Bayesian procedure, in the same broad spirit as Approximate Bayesian Computation (ABC; Marin et al., 2012), especially if we consider recent, discrepancy-based, ABC variants; see, e.g., Park et al. (2016), Jiang (2018), Bernton et al. (2019), Nguyen et al. (2020), Frazier (2020), Fujisawa et al. (2021), Legramanti et al. (2025). Notably, D-BETEL departs from traditional moment-based BETEL in a way that parallels how discrepancy-based ABC has moved away from summary statistics, often also related to moments. Both D-BETEL and ABC replace the likelihood with a term involving a discrepancy between an empirical distribution supported on the observed data and a parametric family F_θ. There are, however, important differences. In ABC, the discrepancy is typically computed between the empirical distributions (or some summary statistics) of the observed data and of synthetic data generated from F_θ. In contrast, in D-BETEL the discrepancy is evaluated between F_θ and a *weighted* discrete distribution supported on the observed data. More precisely, D-BETEL constructs a pseudo-likelihood by finding the maximum-entropy weighted empirical distribution supported on the observed sample that lies within an ε-neighborhood of F_θ under the chosen discrepancy. This entropy-based reweighting scheme provides a direct robustness mechanism, in which observations that are difficult to reconcile with F_θ are adaptively downweighted.

The connection between D-BETEL and ABC suggests two natural directions for further investigation. First, D-BETEL could borrow some ideas from the abovementioned recent advancements in ABC, which may be useful when simulation from F_θ is possible but the corresponding density, or the distance from F_θ to a discrete measure, is not available in a tractable form. In such settings, one could consider a simulation-based version of D-BETEL in which F_θ is replaced by a Monte Carlo empirical approximation. Naturally, this would require careful choice of the empirical-to-empirical discrepancy, since this discrepancy would be evaluated repeatedly inside the entropy-constrained optimization over weights. In this simulation-based D-BETEL variant, integral probability semimetrics (IPS), and in particular the maximum mean discrepancy (MMD; Muandet et al., 2017), appear to be natural candidates. Such discrepancies have proved effective within ABC (Legramanti et al., 2025), are directly computable between empirical measures, and may offer computational and robustness advantages in settings where a model-specific transport discrepancy such as ANDREW is not available. Moreover, MMD has been shown to yield estimators with robustness properties under outlier

*Department of AI, Data and Decision Sciences, Luiss University, Italy, mcatalano@luiss.it
†Department of Economics, University of Bergamo, Italy, sirio.legramanti@unibg.it

contamination (Alquier and Gerber, 2024). Combining such a robust discrepancy with the entropy-based reweighting mechanism of D-BETEL could provide an additional source of robustness, which would then arise both from the choice of discrepancy and the adaptive downweighting of outliers. It would be interesting to understand whether MMD-based or, more generally, IPS-based versions of D-BETEL can deliver posterior robustness properties comparable to those obtained with ANDREW.

Second, D-BETEL may provide useful ideas for ABC. In fact, computational scalability is a central challenge in ABC, which typically requires tens of thousands of discrepancy evaluations, and the ANDREW discrepancy is explicitly designed to balance multivariate geometry, tail sensitivity, and computational efficiency. Namely, in ANDREW, a tractable restricted transport component is enriched by a marginal quantile augmentation, which adds coordinate-wise tail information while preserving efficient computation. Since this added tail information is marginal, a natural extension would be to investigate whether joint tail dependence could also be incorporated, for example through copula- or projection-based terms. Moreover, it would be of interest to assess whether ANDREW (or related augmented and restricted transport discrepancies) can be successfully employed within ABC, and whether they yield posterior concentration properties similar to those proved by Bernton et al. (2019) and Legramanti et al. (2025) for the Wasserstein distance and generic integral probability semimetrics. This would clarify whether the computational advantages that make ANDREW effective within D-BETEL can also benefit ABC and, more broadly, likelihood-free inference.

References

Alquier, P. and Gerber, M. (2024). Universal robust regression via maximum mean discrepancy. *Biometrika* 111, 71–92. doi:10.1093/biomet/asad031. 1004

Bernton, E., Jacob, P. E., Gerber, M., and Robert, C. P. (2019). Approximate Bayesian computation with the Wasserstein distance. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* 81, 235–269. doi:10.1111/rssb.12312. 1003, 1004

Frazier, D. T. (2020). Robust and efficient approximate Bayesian computation: A minimum distance approach. arXiv:2006.14126. MR4147553. 1003

Fujisawa, M., Teshima, T., Sato, I., and Sugiyama, M. (2021). γ-ABC: Outlier-robust approximate Bayesian computation based on a robust divergence estimator. In *International Conference on Artificial Intelligence and Statistics*, 1783–1791. PMLR. 1003

Jiang, B. (2018). Approximate Bayesian computation with Kullback-Leibler divergence as data discrepancy. In *International Conference on Artificial Intelligence and Statistics*, 1711–1721. PMLR. doi:10.1201/b22400. 1003

Legramanti, S., Durante, D., and Alquier, P. (2025). Concentration of discrepancy-based approximate Bayesian computation via Rademacher complexity. *The Annals of Statistics* 53, 37–60. doi:10.1214/24-aos2453. 1003, 1004

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing* 22, 1167–1180. doi:10.1007/s11222-011-9288-2. 1003

Muandet, K., Fukumizu, K., Sriperumbudur, B., and Schölkopf, B. (2017). Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning* 10, 1–141. doi:10.1561/2200000060. 1003

Nguyen, H. D., Arbel, J., Lü, H., and Forbes, F. (2020). Approximate Bayesian computation via the energy statistic. *IEEE Access* 8, 131683–131698. doi:10.1109/ACCESS.2020.3009878. 1003

Park, M., Jitkrittum, W., and Sejdinovic, D. (2016). K2-ABC: Approximate Bayesian computation with kernel embeddings. In *Artificial Intelligence and Statistics*, 398–407. PMLR. 1003