Bayesian Analysis (2026) **TBA**, Number TBA, pp. 1

Discussion of "Robust Probabilistic Inference via a Constrained Transport Metric"

Gourab Mukherjee*

Robust empirical Bayes prediction using D-BETEL

The authors are to be congratulated for a timely contribution to robust statistical inference. In Chakraborty et al. [1], they propose D-BETEL, a distributionally robust Bayesian exponentially tilted empirical likelihood method that replaces exact parametric specification with a constrained transport neighborhood around a working model. The resulting procedure is both computationally tractable and conceptually appealing.

We consider empirical Bayes (EB) prediction in generalized linear mixed models (GLMMs) under acute data scarcity, where only a small subset of units contributes repeated observations and most units are observed at most once. Such regimes arise in modern prediction problems and are especially challenging because limited replication, severe imbalance, and possible covariate shift make the latent-effects distribution difficult to estimate reliably from the data [4]. In these settings, shrinkage is essential for stable prediction, since unit-specific estimates would otherwise be too noisy. Empirical Bayes is therefore attractive because it borrows strength across units while adapting to the limited data available for each one. We study whether the robustness of D-BETEL can be leveraged for EB prediction under data scarcity, imbalance, and covariate shift, and whether its performance is close to that of widely used EB methods based on nonparametric maximum likelihood estimation (NPMLE) [2] and g-modeling [3].

In Section 3.2 of Chakraborty et al. [1], the authors consider a Poisson GLM example. Here we study a related problem, but with random effects whose distribution is unknown and must be estimated. For $i = 1, \ldots, n$, consider $Y_{ik} = \text{Poisson}(x_{ik}^\top \beta + b_i)$, $k = 1, \ldots, K_i$, and $\tilde{Y}_{i\tilde{k}} = \text{Poisson}(\tilde{x}_{i\tilde{k}}^\top \beta + b_i)$, $\tilde{k} = 1, \ldots, \tilde{K}_i$. The random effects b_i are i.i.d. from g, β is a common fixed effect, and both are assumed to be invariant across training and test samples. The goal is to predict $\{\tilde{Y}_{i\tilde{k}}\}$ based on the observed $\{Y_{ik}\}$. NPMLE estimates g by marginal likelihood over a discrete support, while g-modeling fits g within a low-dimensional parametric family on a fixed grid [2, 3]. D-BETEL instead starts from a working family of priors and replaces exact fit by a transport-relaxed entropy criterion [1]. On the common support grid used here, the difference is therefore between exact mixture fitting and transport-relaxed fitting. Table 1 summarizes the results from three simulation experiments comparing the performance of g-modeling, NPMLE, D-BETEL, and the unshrunken estimator. For Experiments 1 and 2, we report estimation error in the parameters. In Experiment 1, we consider Gaussian random effects and, across Cases A–E, examine different data-scarce imbalanced designs. The ratio of observations to parameters is 1, 2, 1.5, 1.2, and 1.4, respectively. The first two designs

---

*Department of Data Sciences and Operations, University of Southern California, gourab@usc.edu

are balanced, whereas Cases C, D, and E exhibit low, moderate, and high imbalance, respectively. Experiments 2A–2C all have ratio equal to 1 and consider, respectively, a two-component Gaussian mixture, a heavy-tailed t distribution, and a contaminated prior with 10% large positive outliers as the random-effects distribution. Experiments 3A–3D consider the settings of Experiment 2 together with Gaussian random effects in Case A. They evaluate prediction under covariate shift, where future covariates follow a different distribution from the training covariates, and prediction accuracy is measured using log-scale mean error.

Table 1: Mean (SD) loss across 25 replicates. Experiments 1–2 report mean absolute estimation error. Experiment 3 reports the log mean absolute prediction error under covariate shift.

Experiment	ORACLE	g-modeling	NPMLE	D-BETEL	Unshrunken
1A	0.7782 (0.0881)	0.9525 (0.1783)	1.0195 (0.2636)	1.0966 (0.1244)	2.1301 (0.2001)
1B	0.7399 (0.0865)	0.8192 (0.1233)	0.8279 (0.1652)	1.0734 (0.1122)	1.6000 (0.1804)
1C	0.7127 (0.0803)	0.8248 (0.0990)	0.9487 (0.2070)	1.1201 (0.0854)	1.9337 (0.1614)
1D	0.7550 (0.0572)	0.9048 (0.1478)	1.0145 (0.3020)	1.1252 (0.1014)	2.0772 (0.1628)
1E	0.7346 (0.0808)	0.8856 (0.1561)	1.0080 (0.3826)	1.0803 (0.1089)	2.0961 (0.1600)
2A	1.6199 (0.2294)	1.6003 (0.2572)	1.6279 (0.2844)	1.8131 (0.2394)	2.3792 (0.3159)
2B	5.8146 (8.9803)	5.9295 (9.0564)	6.0208 (8.9841)	6.2165 (9.0033)	6.5493 (8.9969)
2C	1.6418 (0.3968)	1.4505 (0.2834)	1.5721 (0.2895)	1.9656 (0.2028)	2.3048 (0.2338)
3A	0.0086 (0.0016)	0.0625 (0.0051)	0.0622 (0.0049)	0.0622 (0.0050)	—
3B	0.0095 (0.0024)	0.0616 (0.0049)	0.0612 (0.0050)	0.0611 (0.0050)	—
3C	0.8331 (0.5800)	0.8809 (0.5783)	0.8806 (0.5778)	0.8806 (0.5776)	—
3D	0.2186 (0.1026)	0.2678 (0.1009)	0.2676 (0.1008)	0.2676 (0.1008)	—

All code for these results is available at `https://github.com/gmukherjee/dbetel-eb`. Across all regimes, D-BETEL yields performance that is close to that of state-of-the-art EB methods and is substantially better than the unshrunken MLE. In D-BETEL, the transport radius ϵ controls the degree of shrinkage; in our table we used $\epsilon = 0.35$ with a Wasserstein metric. Improved tuning may further enhance performance. In addition, using the full subject-level response vector in D-BETEL leads to a multivariate Wasserstein transport problem that is computationally expensive. The results reported here are instead based on a reduced model using summary statistics, which is computationally simpler but less informative; this suggests there is further room for improvement.

References

[1] Chakraborty, A., Bhattacharya, A. and Pati, D. (2025). Robust Probabilistic Inference via a Constrained Transport Metric. *Bayesian Analysis*, Advance Publication, 1–33. 1

[2] Jiang, W. and Zhang, C.-H. (2009). General Maximum Likelihood Empirical Bayes Estimation of Normal Means. *The Annals of Statistics*, 37(4), 1647–1684. 1

[3] Narasimhan, B. and Efron, B. (2020). deconvolveR: A G-Modeling Program for Deconvolution and Empirical Bayes Estimation. *Journal of Statistical Software*, 94(11). 1

[4] Sarkar, A., Mukherjee, G., Yano, K. (2026). Empirical Bayes Predictive Density Estimation under Covariate Shift in Large Imbalanced Linear Mixed Models. *arXiv:2603.27984*. 1