Amortized Mixture and Multilevel Models

Paul Bürkner

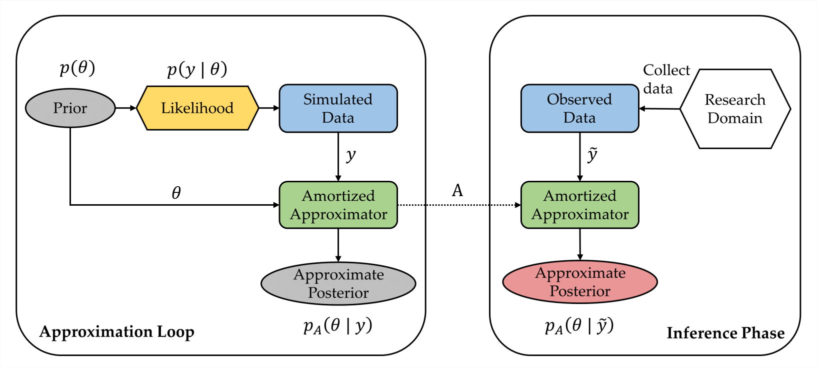
Team

Stefan T. Radev

Daniel Habermann

Šimon Kucharský

Amortized Bayesian Inference (ABI)



Neural Posterior Estimation

Example: Maximum likelihood loss for normalizing flow q_ψ coupled with summary architecture h_ξ:

$$\widehat{\psi}, \hat{\xi} = \underset{\psi,\xi}{\operatorname{argmin}} \, \mathbb{E}_{(\theta,y)\sim p(\theta,y)} \left[-\log q_\psi(\theta \mid h_\xi(y)) \right]$$

In practice approximated via S samples $(\theta^{(s)}, y^{(s)}) \sim p(\theta, y)$:

$$\widehat{\psi}, \hat{\xi} = \underset{\psi,\xi}{\operatorname{argmin}} \, \frac{1}{S} \sum_{s=1}^{S} \left[-\log q_\psi(\theta^{(s)} \mid h_\xi(y^{(s)})) \right]$$

Factorized Bayesian Models

Example: Suppose we have data y and two sets of parameters θ and τ with the following forward model:

$$p(y, \theta, \tau) = p(\tau)\, p(\theta \mid \tau)\, p(y \mid \theta)$$

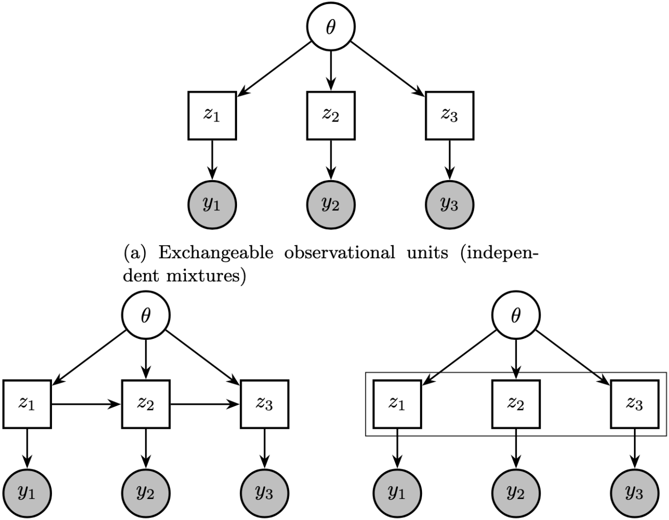
Then we can obtain simple factorization of the posterior:

$$p(\theta, \tau \mid y) = p(\theta \mid y)\, p(\tau \mid \theta)$$

Mixture Models

Kucharský S. & Bürkner P. C. (in review). Amortized Bayesian Mixture Models. *ArXiv preprint*. https://arxiv.org/abs/2501.10229

Mixture Models as Graphical Models



(a) Exchangeable observational units (independent mixtures)

(b) Hidden Markov model

(c) Joint prior over mixture indicators

Factorization for Mixture Models

Forward model (independent observational units):

$$\theta \sim p(\theta)$$
$$z_i \sim p(z \mid \theta)$$
$$y_i \sim p_{z_i}(y \mid \theta)$$

for $i = 1, \ldots, N$ observational units

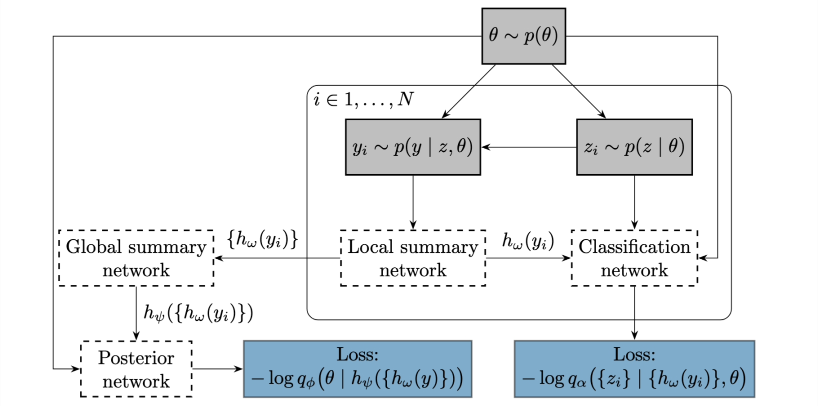
Inverse factorization (independent units):

$$p(\theta, z \mid \{y_{ij}\}) = p(\theta \mid \{y_{ij}\}) \prod_{i=1}^{I} p(z_i \mid y_i, \theta)$$

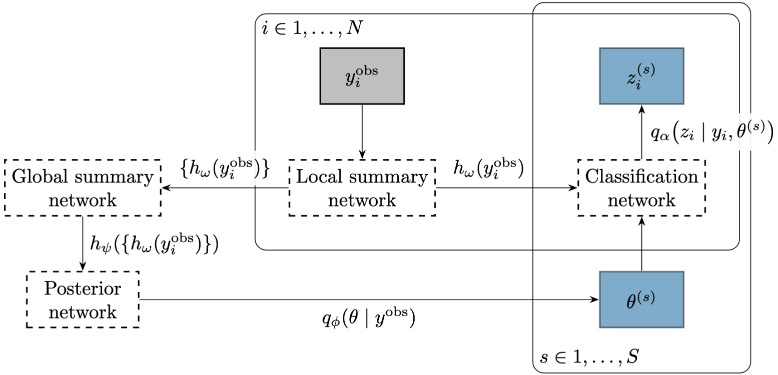
Inverse factorization (dependent units):

$$p(\theta, z \mid \{y_i\}) = p(\theta \mid \{y_i\}) \, p(z \mid \{y_i\}, \theta)$$

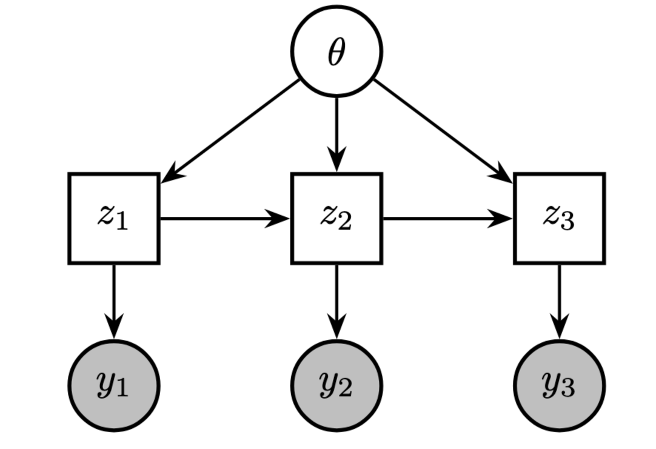
Neural Archictecutres for Mixture Models: Forward



Neural Archictecutres for Mixture Models: Inverse



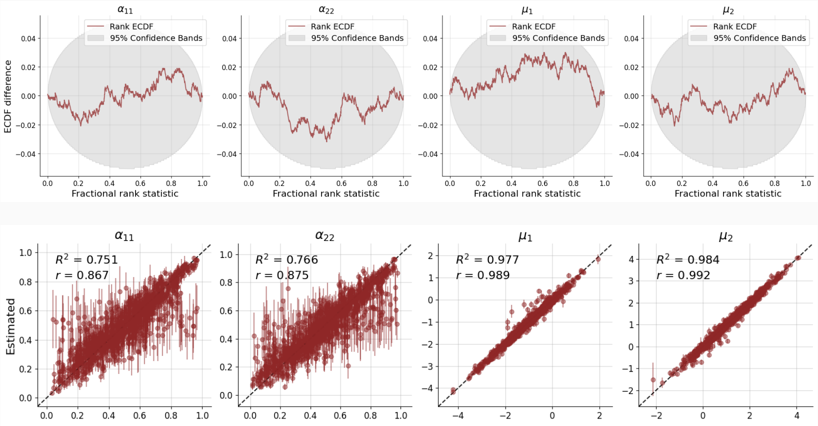
Gaussian Hidden Markov Models



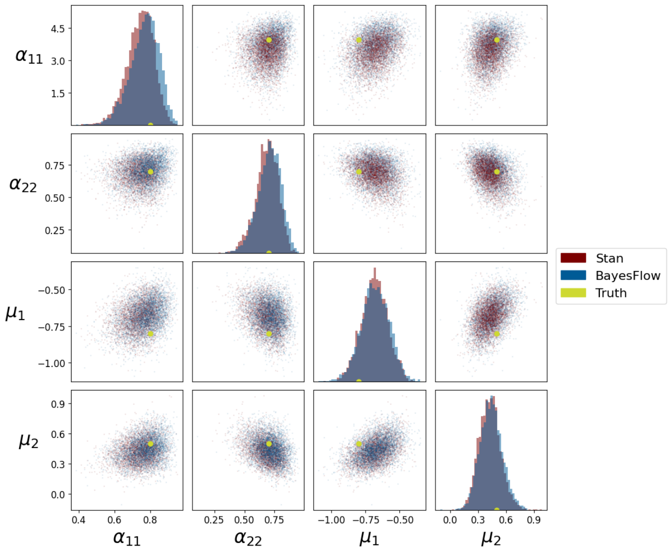
Example: Gaussian Hidden Markov Model with Two States

$$
\begin{aligned}
p_i &\sim \mathsf{Uniform}(2, 5) \quad i = 1, \dots, N \\
\alpha_k &\sim \mathsf{Dirichlet}(2, 2) \quad k = 1, 2 \\
z_1 &\sim \mathsf{Categorical}(0.5, 0.5) \\
z_i &\sim \mathsf{Categorical}(\alpha_{z_{i-1}}) \quad i = 2, \dots, N \\
(\mu_1, \mu_2) &\sim \mathsf{Normal}\Big((-1.5, 1.5), \mathbb{I}\Big)_{\mu_1 < \mu_2} \\
y_{ij} &\sim \mathsf{Normal}(\mu_{z_i}, 1) \quad i = 1, \dots, N;\ j = 1, \dots, p_i
\end{aligned}
$$

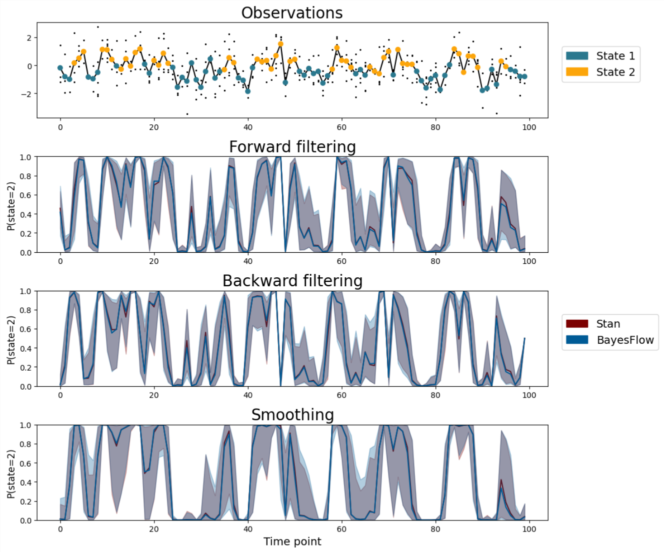
Results: Calibration and Recovery



Results: Joint Posterior



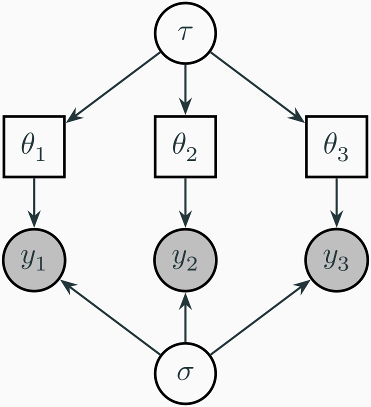
Results: Classification



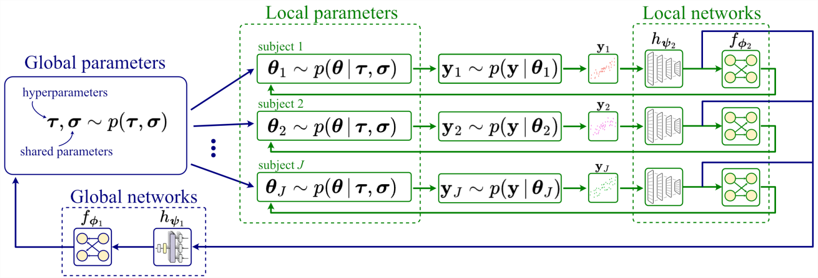
Multilevel Models

Habermann D., Schmitt M., Kühmichel L., Bulling A., Radev S. T., & Bürkner P. C. (in review). Amortized Bayesian Multilevel Models. *ArXiv preprint*. https://arxiv.org/abs/2408.13230

Factorization for Simple Multilevel Models



Neural Archictecture for Two-Level Models



Example: European Air Passenger Traffic

Likelihood of the difference in air passenger traffic for country j between time $t+1$ and t:

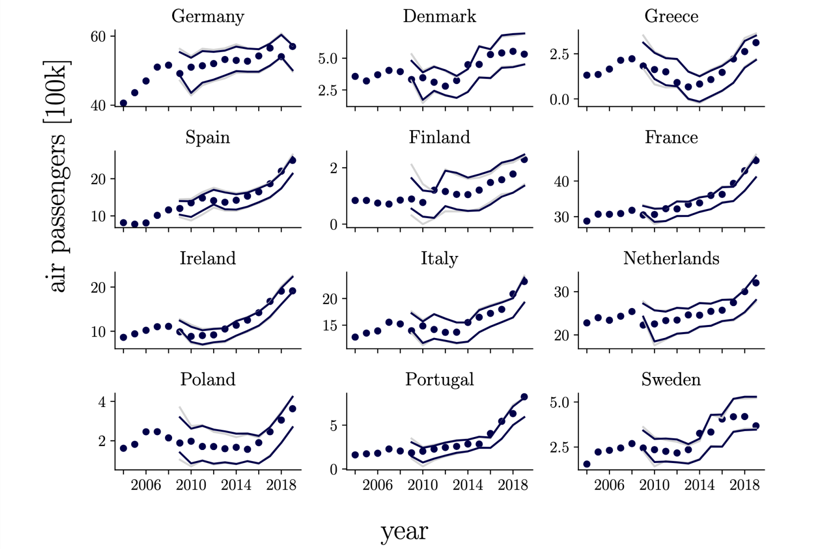
$$y_{j,t+1} \sim \mathrm{Normal}(\alpha_j + y_{j,t}\beta_j + u_{j,t}\gamma_j + w_{j,t}\delta_j, \sigma_j)$$

Local priors:

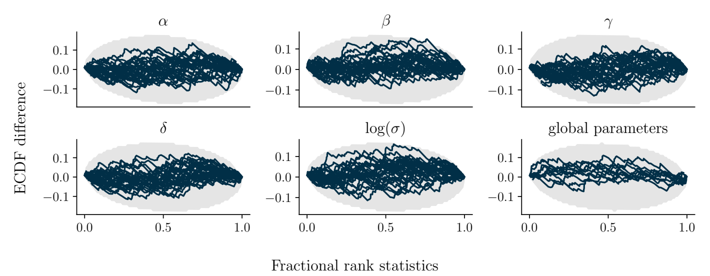
$$\begin{aligned}
\alpha_j &\sim \textsf{Normal}(\mu_\alpha, \sigma_\alpha) \\
\beta_j &\sim \textsf{Normal}(\mu_\beta, \sigma_\beta) \\
\gamma_j &\sim \textsf{Normal}(\mu_\gamma, \sigma_\gamma) \\
\delta_j &\sim \textsf{Normal}(\mu_\delta, \sigma_\delta) \\
\log(\sigma_j) &\sim \textsf{Normal}(\mu_\sigma, \sigma_\sigma)
\end{aligned}$$

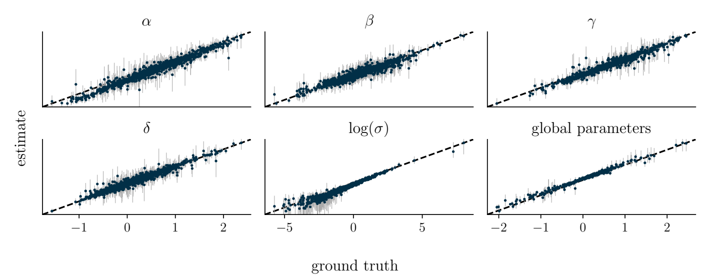
Global priors not shown for simplicity

Results: Posterior Predictions

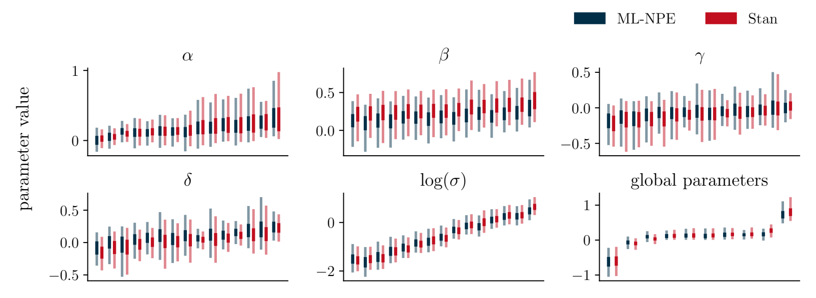


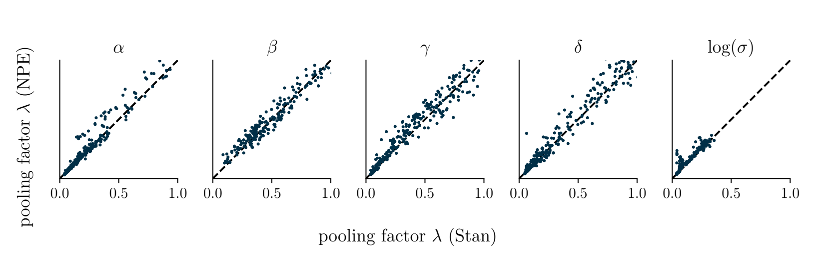
Results: Calibration and Recovery





Results: Comparison with Stan





Summary

Multilevel and mixture models follow simple probabilistic factorizations

Utilizing these factorizations enables amortized Bayesian inference across varying number of observational units and observations within units

Much of the potential of neural ABI is yet to be realized