Bayesian score calibration for approximate models

Prof Chris Drovandi

Co-authors: Joshua Bon, David Warne, David Nott

5 February 2024

Bayesian Statistics

Let $y \sim P_0$ denote the observed data.

The model P_0 is approximated using a parametric family of models $\{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^{d_\theta}\}$.

Place prior distribution on $\theta \sim \Pi$ with density $\pi(\theta)$.

We wish to estimate θ via the posterior distribution with density:

$$\pi(\theta \mid y) \propto p(y \mid \theta)\pi(\theta),$$

where $\pi(\theta)$ is the prior density and $p(y \mid \theta)$ is the likelihood.

For complex models, $p(y \mid \theta)$ may be intractable.

Approximate Models to Rescue?

Likelihood-free approaches often require a large number of model simulations.

Often we can formulate a simpler version of the model with the same parameterisation. Some examples include:

- Deterministic versions of stochastic models.
- Whittle likelihood for time series data
- Composite likelihoods
- Approximate inference algorithms

It is tempting to proceed with inference using the approximate model, $\hat{\pi}(\theta \mid y) \propto \hat{p}(y \mid \theta)\pi(\theta)$ for an approximate likelihood $\hat{p}(y \mid \theta)$.

But could lead to bias and poor uncertainty quantification.

Calibrate Approximate Posterior

We propose to calibrate/adjust samples from the approximate posterior to improve the approximation, in some way.

Our proposed approach:

- Requires only the ability to generate a small number of true model simulations $\sim \mathcal{O}(10^2)$.
- Is numerically stable.
- Finite-sample based.

Scoring Rules

We use a scoring rule $S(U, V)$ to define a discrepancy between the calibrated approximate posterior U and true posterior $V = \Pi(\cdot \mid y)$.

Consider the optimisation problem

$$U^\star = \max_{U \in \mathcal{P}} \mathbb{E}_{\theta \sim \Pi(\cdot \mid y)} \left[S(U, \theta) \right].$$

If $S(\cdot, \cdot)$ is strictly proper and $\mathcal{P}$ is rich enough we recover $U^\star(\cdot) = \Pi(\cdot \mid y)$.

Hassle-Free Implementation

Since the optimisation on previous slide is intractable, we consider the alternative optimisation problem:

$$U^\star \equiv \underset{U \in \mathcal{K}}{\arg\max}\, \mathbb{E}_{\theta \sim \bar{\Pi}} \mathbb{E}_{\tilde{y} \sim P(\cdot \mid \theta)} \left[S(U(\cdot \mid \tilde{y}), \theta) \right],$$

where $\bar{\Pi} = \bar{\Pi}(\theta|y)$ is the approximate posterior.

We have a more general formulation with theoretical justification in the paper.

We can approximate above expectation using m Monte Carlo samples $(\theta_i, \tilde{y}^i)_{i=1}^m \sim P(\tilde{y} \mid \theta)\bar{\Pi}(\theta|y)$.

Energy Score

Here we choose the energy score:

$$S(U,\theta) = \frac{1}{2}\mathbb{E}_{u,u'\sim U}\|u - u'\|^{\beta} - \mathbb{E}_{u\sim U}\|u - \theta\|^{\beta},$$

which can be approximated by Monte Carlo as:

$$S^{N}(U,\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}\|u_i - u_i'\|^{\beta} - \|u_i - \theta\|^{\beta}\right), \quad u_i, u_i' \sim U.$$

Moment-Correcting Transformation

For this work we use: $\mathcal{K} = \{f_{\#}\hat{\Pi}(\cdot \mid \cdot) : f \in \mathcal{F}\}$

For $f_{\#}$ we use moment-correcting transformation:

$$f(x) = L[x - \hat{\mu}(y)] + \hat{\mu}(y) + b$$

where:

- Mean $\hat{\mu}(y) = \mathbb{E}(\theta), \quad \theta \sim \hat{\Pi}(\cdot \mid y)$ for $y \in \mathsf{Y}$.
- L is a lower triangular matrix with positive elements on diagonal.

The mean and covariance of the adjusted approximate posterior, $f_{\#}\hat{\Pi}(\cdot \mid y)$, is $\mu(y) = \hat{\mu}(y) + b$ and $\Sigma(y) = L\hat{\Sigma}(y)L^{\top}$, respectively.

Summarising the Hassle-Free Method

1. Transform the parameter vector θ to $\tilde{\theta}$ so that each parameter is unconstrained.
2. Obtain samples $\{\tilde{\theta}^{(i)}\}_{i=1}^{N}$ from the approximate posterior $\hat{\pi}(\tilde{\theta}|y)$.
3. Inflate the approximate posterior standard deviation of each component of $\tilde{\theta}$ by a factor of s. This now re-defines $\{\tilde{\theta}^{(i)}\}_{i=1}^{N}$.
4. Take a random subset of m samples from $\{\tilde{\theta}^{(i)}\}_{i=1}^{N}$, denoted $\{\tilde{\theta}^{(j)}\}_{j=1}^{m}$ (these are the calibration parameter values).
5. Generate calibration datasets $y^{(j)}$ from $p(y|\tilde{\theta}^{(j)})$ for $j = 1, \ldots, m$.
6. Obtain a sample from the approximate posterior $\hat{\pi}(\tilde{\theta}|y^{(j)})$ for $j = 1, \ldots, m$. For notational simplicity, denote the sample from the jth approximate posterior as $\hat{\pi}^{(j)}$. Store the results from steps 1-6.
7. Solve the optimisation problem $g^{\star} \equiv \arg\min_{f \in \mathcal{F}} \frac{1}{m} \sum_{j=1}^{m} S(f(\hat{\pi}^{(j)}), \tilde{\theta}^{(j)})$.
8. Apply the transform $g^{\star}$ to $\{\tilde{\theta}^{(i)}\}_{i=1}^{N}$ to obtain samples from the adjusted approximate posterior, and back-transform to the original space of θ.

Example 1: OU Process (limiting approximation)

$$\mathrm{d}X_t = \gamma(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t$$

Observe final observation at time T ($n = 100$):

$$X_T \sim \mathcal{N}\left(\mu + (x_0 - \mu)e^{-\gamma T}, \frac{D}{\gamma}(1 - e^{-2\gamma T})\right)$$

where $D = \frac{\sigma^2}{2}$. Fix $\gamma = 2$, $T = 1$, $x_0 = 10$

Infer μ and D with approximate likelihood based on

$$X_\infty \sim \mathcal{N}\left(\mu, \frac{D}{\gamma}\right)$$

Example 1: OU Process (limiting approximation)

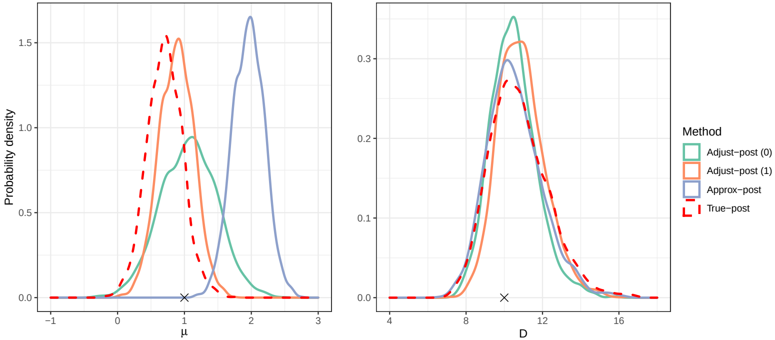

Figure 1: Univariate densities estimates of approximations to the OU Process model posterior distribution for a single dataset.

Example 2: Bivariate OU Process (VI approximation)

$$\mathrm{d}X_t = \gamma(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t$$

$$\mathrm{d}Y_t = \gamma(\mu - Y_t)\mathrm{d}t + \sigma \mathrm{d}W_t$$

$$Z_t = \rho X_t + (1-\rho) Y_t$$

Model (X_t, Z_t) with setup as in the univariate case, $(x_0, z_0) = (5, 5)$.

Mean-field variational approximation.

Example 2: Bivariate OU Process (VI approximation)

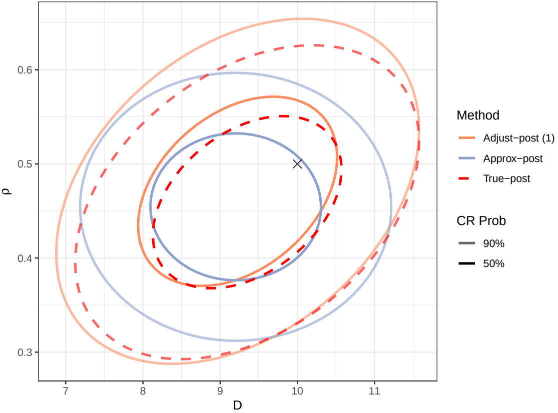

Figure 2: Bivariate OU Process model posterior distribution for a single dataset.

Future Work

- Extend to more complicated transforms
- Apply to crude likelihood-free algorithms
- Possibly different scoring rules
- Extend to model misspecification?

Paper reference:

Bon et al (2023). Bayesian score calibration for approximate models.
https://arxiv.org/abs/2211.05357.