Skew-Symmetric Approximations of Posterior Distributions

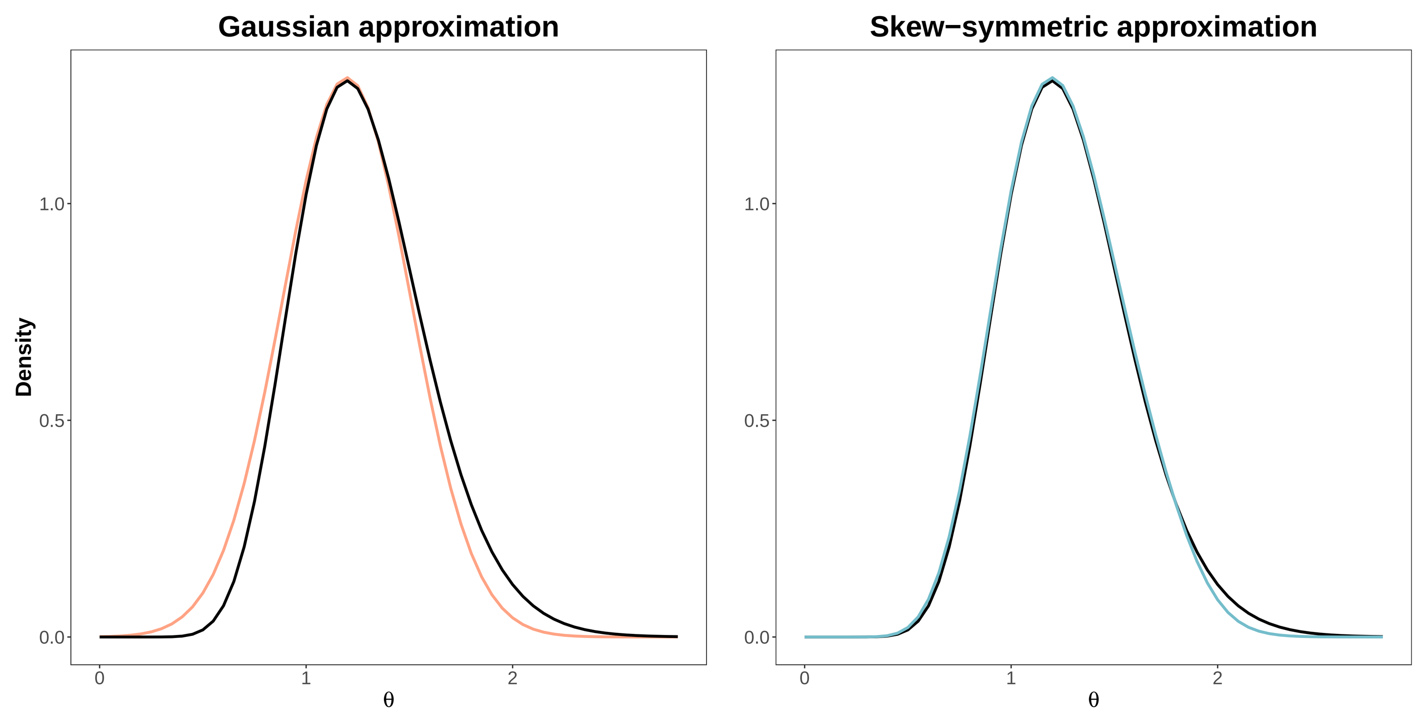
Francesco Pozza

Department of Statistical Sciences, University of Padova, Padova.

[1] **email**: francesco.pozza.2@phd.unipd.it **web**: https://francesco16p.github.io

1. Introduction

In Bayesian statistics, common deterministic approximations of posterior distributions are typically Gaussian. This choice facilitates optimization and inference, but it may compromise the quality of the overall approximation. Indeed, even in simple parametric models the posterior distribution can be asymmetric.



Recent research has moved toward more flexible classes of approximating densities incorporating skewness. However, current solutions are model specific and/or lack of general supporting theory.

2. Notation

- $\{X_i\}_{i=1}^n$ sequence of random variables with probability measure P_0^n
- $\{P_\theta^n, \theta \in \Theta \in \mathbb{R}^p\}$
- $L(\theta)$ likelihood and $\ell(\theta)$ log-likelihood
- $\pi(\theta)$ prior and $\pi_n(\theta) = \pi(\theta)L(\theta)/m(X^n)$ posterior
- $\hat{\theta}$: posterior mode
- First three log-likelihood derivatives at $\hat{\theta}$ $\ell_{\hat{\theta}}^{(1)} = \left[\ell_{\hat{\theta},r}^{(1)}\right]$, $\ell_{\hat{\theta}}^{(2)} = \left[\ell_{\hat{\theta},rs}^{(2)}\right]$, $\ell_{\hat{\theta}}^{(3)} = \left[\ell_{\hat{\theta},rst}^{(3)}\right]$ for $r, s, t = 1, \ldots, p$
- Einstein's summation convention adopted

3. Skew-symmetric random variables

Definition: *A random variable θ is skew-symmetric if its probability density function takes the form*

$$p(\theta) = 2p_\xi(\theta)G(w(\theta - \xi))$$

Where

- $p_\xi(\cdot)$: probability density function symmetric about $\xi \in \mathbb{R}^d$
- $w(\cdot)$: odd function from $\mathbb{R}^d \to \mathbb{R}$
- $G(\cdot)$: continuous univariate cumulative distribution function satisfying $G(-\theta) = 1 - G(\theta)$

Special case: $p_\xi(\cdot) = \phi_d(\cdot; \xi, \Sigma)$ is a multivariate normal with mean ξ and covariance matrix Σ, $G(\cdot) = \Phi(\cdot)$ is the standard normal cumulative distribution function and $w(\cdot)$ is an odd polynomial function

Simulation from skew-symmetric random variables:

1. Simulate from the symmetric distribution $\theta^* \sim P_0$ where $\xi = 0$
2. Simulate a Bernoulli random variable Z with probability $G(w(\theta^*))$
3. $\theta = \theta^*(2Z - 1) + \xi$ is skew-symmetric with density $2p_\xi(\theta)G(w(\theta - \xi))$

4. Skew modal approximation

We propose to approximate the posterior density $\pi_n(\theta)$ with the asymmetric density

$$\hat{p}^n_{\text{SKS}}(\theta) = 2\phi_d(\theta; \hat{\theta}, J_{\hat{\theta}}^{-1})\Phi(w(\theta - \hat{\theta})) \tag{1}$$

where

- $J_{\hat{\theta}} = -\ell_{\hat{\theta}}^{(2)}$
- $\phi_d(\cdot; \hat{\theta}, J_{\theta}^{-1})$: is a d-variate Gaussian density with mean $\hat{\theta}$ and covariance matrix $J_{\hat{\theta}}^{-1}$
- $w(\theta) = (\sqrt{2\pi}/12)\ell^{(3)}_{\hat{\theta},stl}(\theta - \hat{\theta})_s(\theta - \hat{\theta})_t(\theta - \hat{\theta})_l$

Theorem 1. *Under appropriate regularity conditions*

$$\frac{1}{2}\int \left|\pi_n(\theta) - \hat{p}^n_{\text{SKS}}(\theta)\right| d\theta = O_p(\{\log n\}^{c_1}/n)$$

for some $c_1 > 0$

Remark: Under similar assumptions the classical Gaussian approximation $\phi_d(\theta; \hat{\theta}, J_{\hat{\theta}}^{-1})$ has asymptotic error of order $O_p(\{\log n\}^{c_2}/\sqrt{n})$, for some $c_2 > 0$.

5. Skew modal marginal approximation

In general, the marginal densities of approximation (1) are not available in closed form. If the interest is on a subset $\theta_{\mathcal{C}}$ of θ, it is possible to approximate $\pi_{\mathcal{C}}(\theta_{\mathcal{C}}) = \int \pi_n(\theta) d\theta_{\mathcal{C}^c}$ with the alternative skew-modal marginal approximation

$$\hat{p}^n_{\text{SKS},\mathcal{C}}(\theta_{\mathcal{C}}) = 2\phi_d(\theta_{\mathcal{C}}; \hat{\theta}_{\mathcal{C}}, J^{-1}_{\hat{\theta},\mathcal{CC}})\Phi\Big(w_{\mathcal{C}}(\theta_{\mathcal{C}})\Big) \tag{2}$$

where $J^{-1}_{\hat{\theta},\mathcal{CC}}$ is the sub-matrix of $J^{-1}_{\hat{\theta}}$ in which only the entries associated to the elements of $\theta_{\mathcal{C}}$ are maintained, while

$$w_{\mathcal{C}}(\theta_{\mathcal{C}}) = \frac{\sqrt{2\pi}}{12}\{a_{1,s}(\theta_{\mathcal{C}} - \hat{\theta}_{\mathcal{C}})_s + a_{2,stl}(\theta_{\mathcal{C}} - \hat{\theta}_{\mathcal{C}})_s(\theta_{\mathcal{C}} - \hat{\theta}_{\mathcal{C}})_t(\theta_{\mathcal{C}} - \hat{\theta}_{\mathcal{C}})_l\}$$

where $a_1 \in \mathbb{R}^d$ and $a_2 \in \mathbb{R}^{d \times d \times d}$ depend both on $\ell_{\hat{\theta}}^{(2)}$ and $\ell_{\hat{\theta}}^{(3)}$

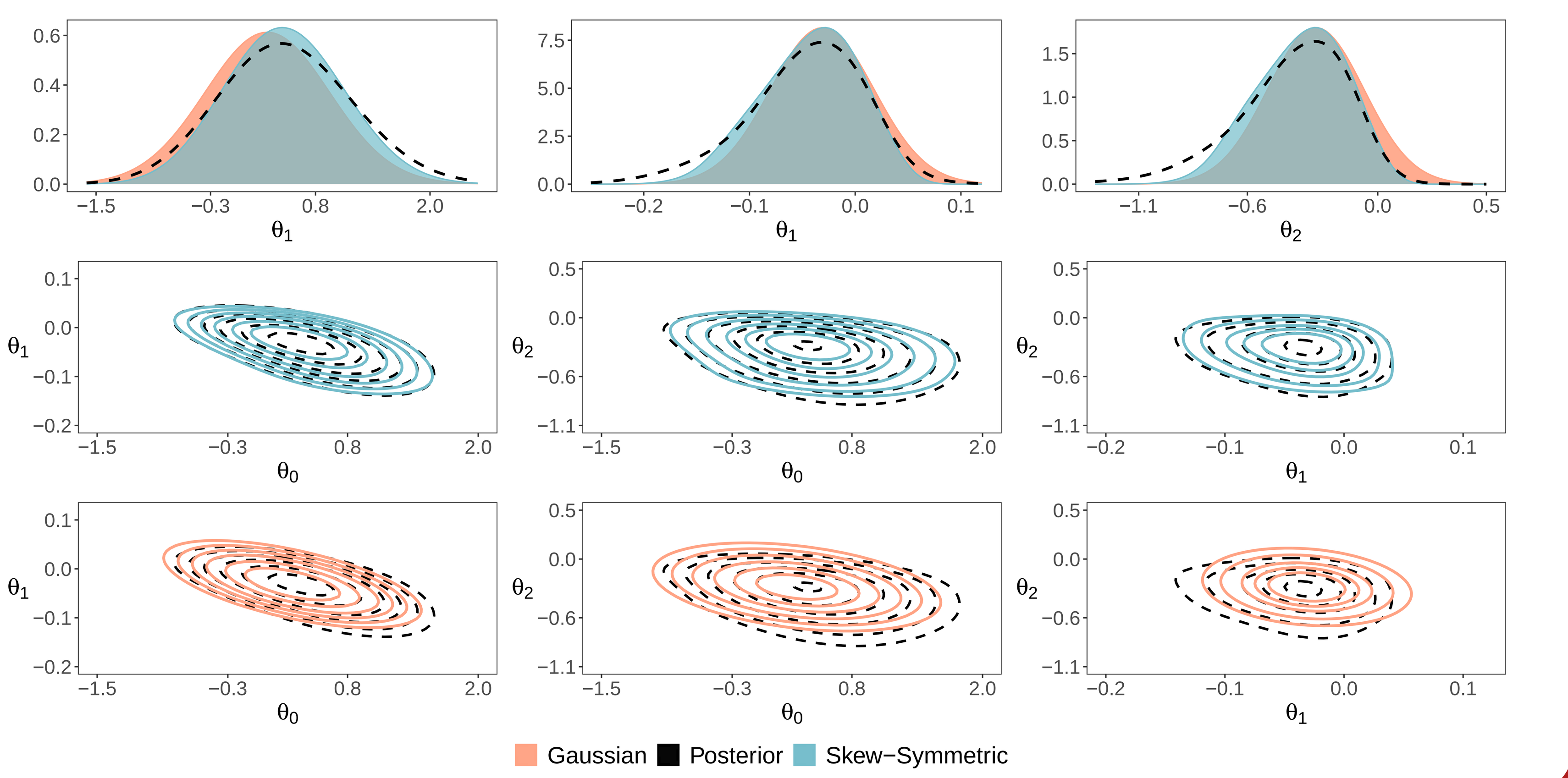
Theorem 2. *Under appropriate regularity conditions*

$$\frac{1}{2}\int \left|\pi_{\mathcal{C}}(\theta_{\mathcal{C}}) - \hat{p}^n_{\text{SKS},\mathcal{C}}(\theta_{\mathcal{C}})\right| d\theta = O_p(\{\log n\}^{c_3}/n)$$

for some $c_3 > 0$

6. Binary regression model (Logistic)

We compare the performance of the proposed skew-symmetric approximation on a logistic regression with $n = 27$ and $d = 3$. For each coefficient, we assume weakly informative N(0,25) prior.



8. References

Durante, D., Pozza, F., & Szabo, B. (2023). Skewed bernstein-von mises theorem and skew-modal approximations. *arXiv preprint arXiv:2301.03038*.