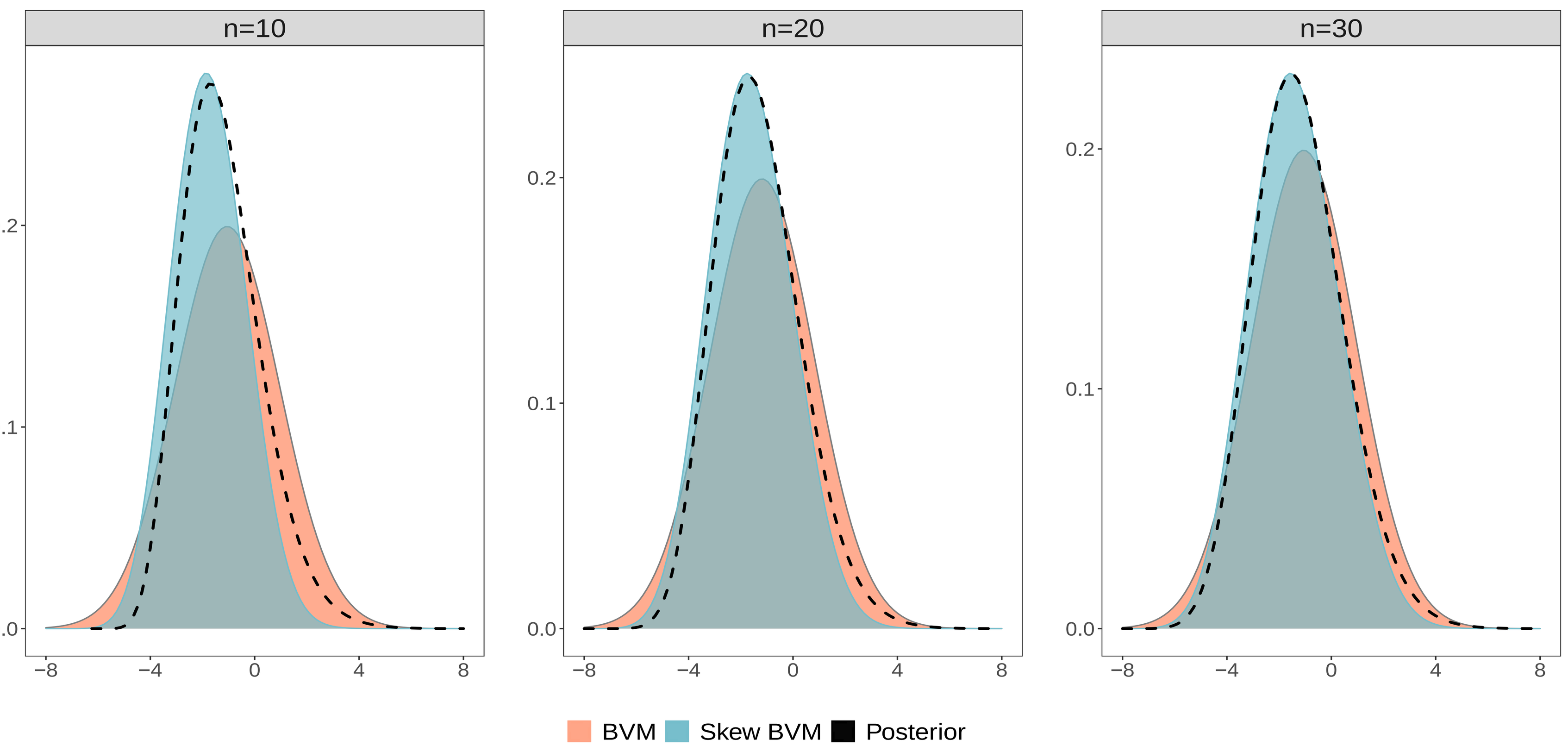
Accurate skewed asymptotic approximations of posterior distributions

Francesco Pozza, Daniele Durante & Botond Szabo

Statistical methods and models for complex data — Poster Session: 21 September 2022

1. Bernstein-Von Mises Theorem

Under regularity conditions, the total variation distance between the posterior distribution and a particular Gaussian distribution converges to 0 in probability.



The limiting behaviour may require a large sample size before becoming visible. By adopting as limiting law a skewed generalization of the Gaussian distribution, we show that it is possible to obtain substantially more accurate results

2. Notation

- $\{X_i\}_{i=1}^n$ sequence of independent random variables with probability measure $P_{\theta_0}^{(n)}$
- $\left\{P_\theta^{(n)}, \theta \in \Theta \in \mathbb{R}^p\right\}$
- $\ell(\theta)$: log-likelihood
- $\pi(\theta)$ and $\pi(\theta \mid X^{(n)})$: prior and posterior
- First three log-likelihood derivatives at θ_0 : $\ell^{(1)} = \left[\ell_r^{(1)}\right]$, $\ell^{(2)} = \left[\ell_{rs}^{(2)}\right]$, $\ell^{(3)} = \left[\ell_{rst}^{(3)}\right]$ for $r, s, t = 1, \ldots, p$
- Einstein's summation convention adopted

3. Skew-Symmetric distributions

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

$$p(\theta) = 2p_0(\theta)G\{w(\theta)\}$$

Where

- $p_0(\cdot)$: pdf symmetric about 0
- $w(\cdot)$: odd function from $\mathbb{R}^p \to \mathbb{R}$
- $G\{\cdot\}$: continuous univariate cdf with $G\{-\theta\} = 1 - G\{\theta\}$

Inclusion of location and scale parameters:

$$Y = \mu + \Sigma\theta$$

Special case: $p_0(\cdot)$ is a multivariate normal pdf, $G\{\cdot\}$ is the standard normal cdf and $w(\cdot)$ is an odd polynomial function

Simulation:

1. Simulate from $\theta^* \sim P_0$
2. Simulate a Bernoulli random variable Z with probability $G\{w(\theta^*)\}$
3. $\theta = \theta^*(2Z - 1)$ is skew-symmetric with density $2p_0(\theta)G\{w(\theta)\}$

4. Main result: skew Bernstein-Von Mises

For $h = \sqrt{n}(\theta - \theta_0)$ consider as limiting distribution

$$p_{\delta,V,n}(h) = 2\phi_p\left(h \mid \delta, V\right)\Phi\left(\alpha(h)\right)$$

where

- $\phi_p(\cdot \mid \delta, V)$: is a p-variate Gaussian density with mean δ and covariance matrix V. Moreover, $\Phi(\cdot)$ is the standard normal cdf
- $\delta_s = (-\ell^{(2)}/n)_{st}^{-1} w_t$, where $w_t = \{(\ell^{(1)} + \log \pi^{(1)})/\sqrt{n}\}_t$
- $V = [v_{st}] = \left[-\ell_{st}^{(2)}/n - (\delta_l \ell_{stl}^{(3)})/n^{3/2}\right]$
- $\alpha(h) = \frac{\sqrt{2\pi}}{12\sqrt{n}} \frac{\ell_{stl}^{(3)}}{n} \{(h-\delta)_s(h-\delta)_t(h-\delta)_l + 3(h-\delta)_s\delta_t\delta_l\}$

Theorem (Skew Bernstein-von Mises)

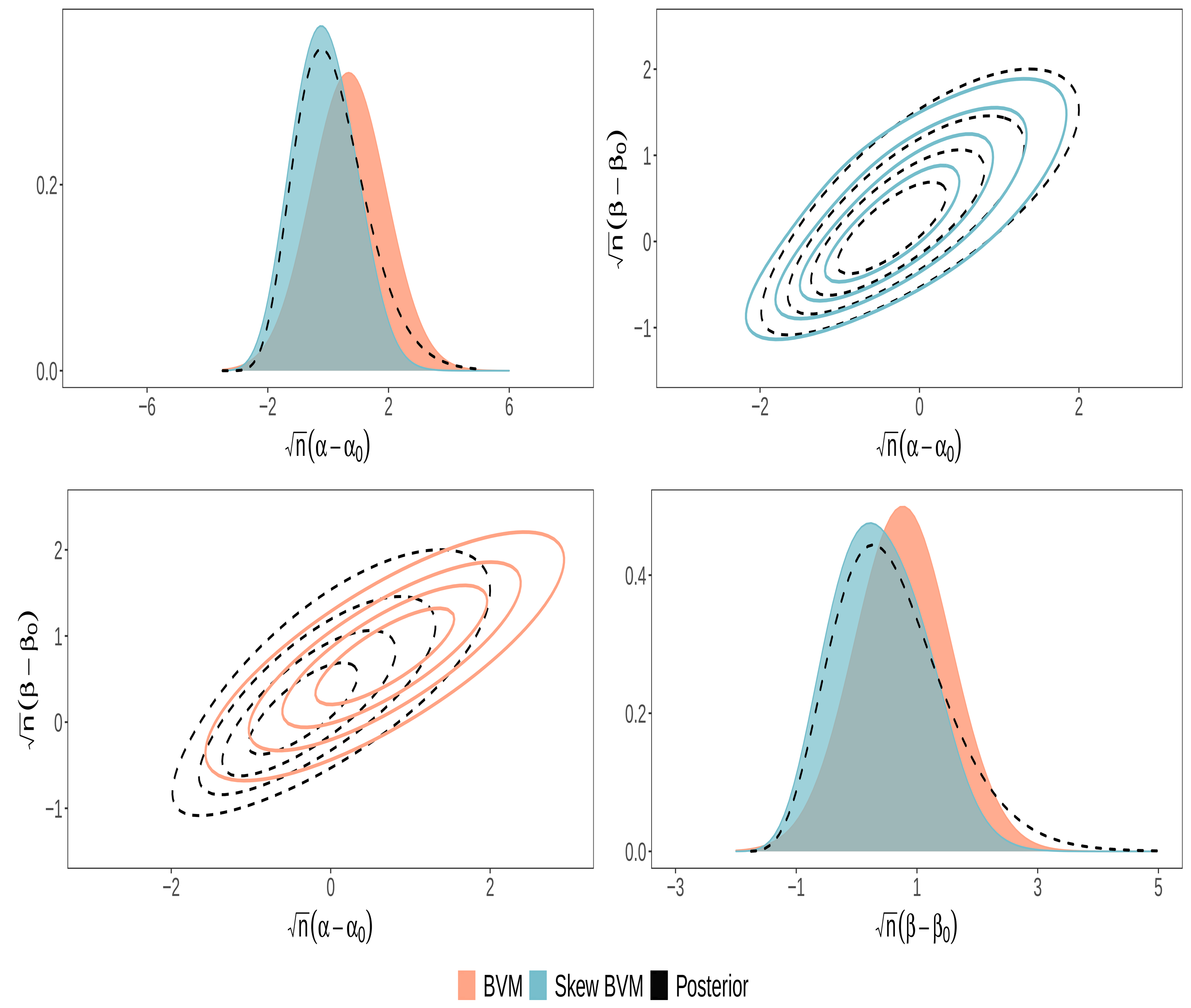
Under regularity conditions on the log-likelihood ratio and its derivatives, if the map $\theta \to P_\theta^n$ is one-to-one, θ_0 is an inner point of Θ and the prior measure is absolutely continuous with bounded and positive density in a neighbourhood of θ_0, then

$$|| \Pi(\cdot \mid X^{(n)}) - P_{se}(\cdot) ||_{TV} = O_p(\{\log n\}^{p/2+3}/n)$$

where $|| \cdot ||_{TV}$ is the total variation norm and $P_{se}(S) = \int_S p_{se}(h)dh$ for $S \subset \mathbb{R}^p$

5. Skew Bernstein-Von Mises (Gamma)

Simulation with $n = 15$, $X_i \overset{iid}{\sim} Ga(\alpha, \beta)$, $\alpha \sim Exp(2)$ and $\beta \sim Exp(2)$



6. Skew modal approximation

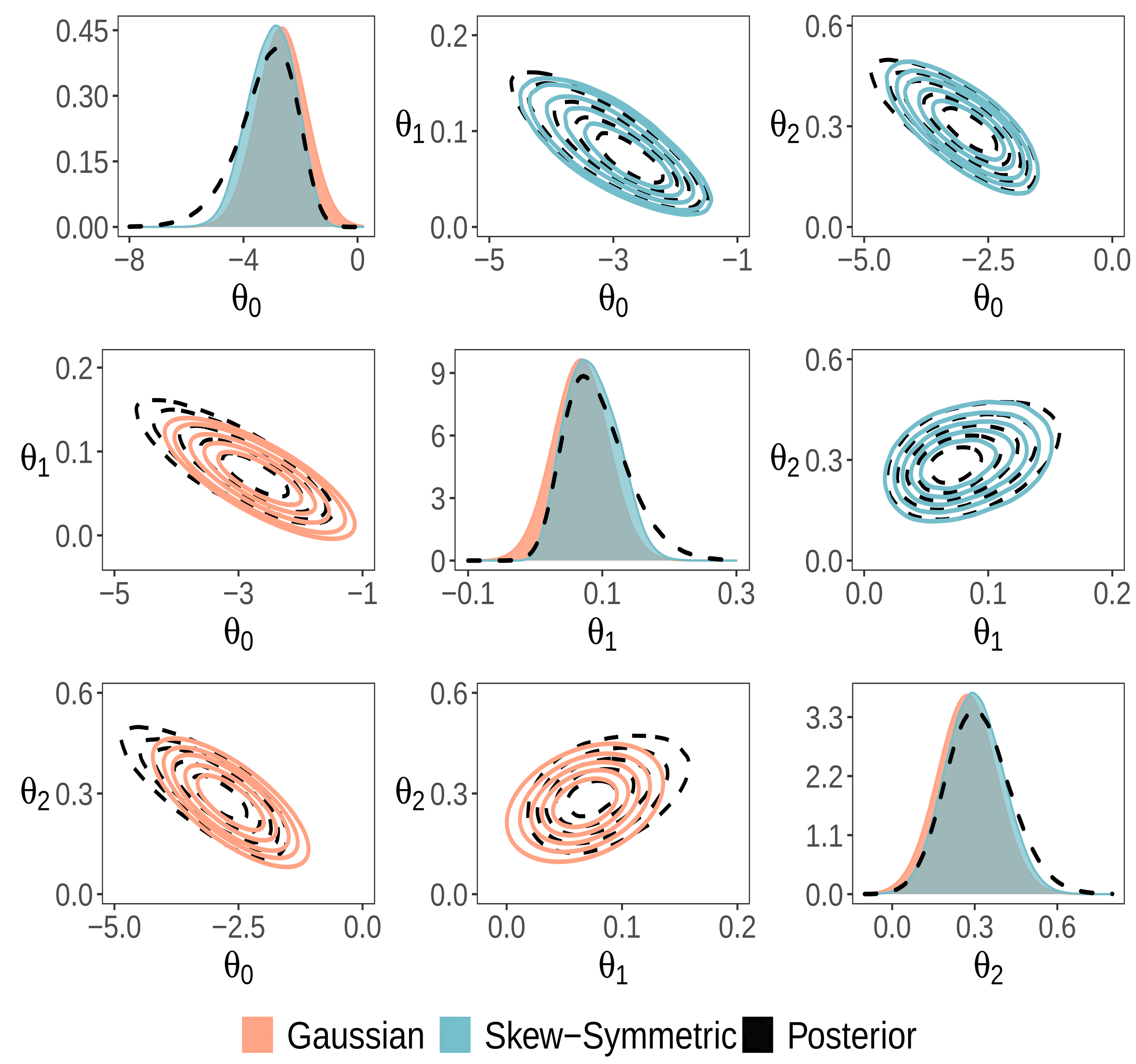
Let $\tilde{\ell}$ be the log-posterior distribution and $\tilde{\theta}$ its maximum. Under some additional regularity conditions it holds that

$$\|\pi(\theta \mid X^{(n)}) - 2\phi_p\left(\theta \mid \tilde{\theta}, \Omega\right)\Phi\left(\alpha(\theta)\right)\|_{TV} = O_p\left(\{\log n\}^{p/2+3}/n\right)$$

where $\Omega = [\omega_{st}] = \left[-\tilde{\ell}_{st}^{(2)}\right]$ and $\alpha(\theta) = \frac{\sqrt{2\pi}}{12}\tilde{\ell}_{stl}^{(3)}(\theta - \tilde{\theta})_s(\theta - \tilde{\theta})_t(\theta - \tilde{\theta})_l$

7. Real data example (Probit model)

Cushings datset: Prediction of carcinoma in $n = 27$ patients with Cushing's syndrome. Explanatory variables: Z_1 = Tetrahydrocortisone and Z_2 = Pregnanetriol.
Model: $X_i \sim Be(\pi_i)$ where $\pi_i = \Phi(\theta_0 + \theta_1 Z_{i1} + \theta_2 Z_{i2})$. The prior distribution is $\theta_j \overset{iid}{\sim} N(0, 25)$, $j = 0, 1, 2$



Reference

Pozza, F., Durante, D., Szabo, B. (2022+). Accurate skewed asymptotic approximations of posterior distributions. *Forthcoming paper*

Azzalini, A. (2013). *The skew-normal and related families* Cambridge University Press, Cambridge.

Van der Vaart, A. W. (2000). *Asymptotic statistics* Cambridge University Press, Cambridge.

Contact information

Francesco Pozza, Ph.D. student
Università degli Studi di Padova
francesco.pozza.2@phd.unipd.it
https://francesco16p.github.io/