Skew–symmetric representations of posterior densities

Skew 2026, January 7–9, Padova

09.01.2026

Daniele Durante

daniele.durante@unibocconi.it

In this talk

Skew–symmetric densities have been originally proposed to increase the flexibility of symmetric ones via the inclusion of a suitably designed skewing factor that preserves the tractability of the unperturbed counterparts.

In this talk

Skew–symmetric densities have been originally proposed to increase the flexibility of symmetric ones via the inclusion of a suitably designed skewing factor that preserves the tractability of the unperturbed counterparts.

In this talk:

1. Any generic posterior distribution admits a skew-symmetric representation decomposing its density as the product of a suitably symmetrized version of the target posterior and a tractable closed–form skewing factor.

Skew–symmetric densities have been originally proposed to increase the flexibility of symmetric ones via the inclusion of a suitably designed skewing factor that preserves the tractability of the unperturbed counterparts.

In this talk:

1. Any generic posterior distribution admits a skew-symmetric representation decomposing its density as the product of a suitably symmetrized version of the target posterior and a tractable closed–form skewing factor.

2. This allows the design of a general strategy to perturb any given symmetric approximation of a generic posterior density (Laplace, VB and EP) for obtaining an improved, yet tractable, skew–symmetric counterpart.

In this talk

Skew–symmetric densities have been originally proposed to increase the flexibility of symmetric ones via the inclusion of a suitably designed skewing factor that preserves the tractability of the unperturbed counterparts.

In this talk:

1. Any generic posterior distribution admits a skew-symmetric representation decomposing its density as the product of a suitably symmetrized version of the target posterior and a tractable closed–form skewing factor.

2. This allows the design of a general strategy to perturb any given symmetric approximation of a generic posterior density (Laplace, VB and EP) for obtaining an improved, yet tractable, skew–symmetric counterpart.

Reference: Pozza, F., Durante, D. and Szabo, B. [2025+]. "Skew–symmetric approximations of posterior distributions". *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, In Press.

Skew–symmetric densities

Definition 1. Let $\bar{q}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ denote a density for $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$, which is symmetric at $\boldsymbol{\theta}^*$, i.e. $\bar{q}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = \bar{q}_{\boldsymbol{\theta}^*}(2\boldsymbol{\theta}^* - \boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$, and define with $w_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ a skewing factor such that $w_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}) \in [0, 1]$ and $w_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 1 - w_{\boldsymbol{\theta}^*}(2\boldsymbol{\theta}^* - \boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \Theta$. Then

$$q_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 2\bar{q}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})w_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}), \tag{1}$$

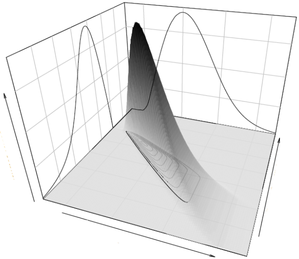
is a skew-symmetric density.

Skew–symmetric densities

Definition 1. Let $\bar{q}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ denote a density for $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$, which is symmetric at $\boldsymbol{\theta}^*$, i.e. $\bar{q}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = \bar{q}_{\boldsymbol{\theta}^*}(2\boldsymbol{\theta}^* - \boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$, and define with $w_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ a skewing factor such that $w_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}) \in [0, 1]$ and $w_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 1 - w_{\boldsymbol{\theta}^*}(2\boldsymbol{\theta}^* - \boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \Theta$. Then

$$q_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 2\bar{q}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}) w_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}), \tag{1}$$

is a skew-symmetric density.



Relies on general skewness–inducing mechanism that leverages a suitably designed perturbation factor $w_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ to redistribute the density between each point $\boldsymbol{\theta}$ and its polar opposite $2\boldsymbol{\theta}^* - \boldsymbol{\theta}$.

Encompasses popular examples, such as skew–elliptical densities, which admit, as special cases, skew–normal, skew–t and skew–Cauchy.

Two important results

Proposition 1(a) [Tractability]. Let $\boldsymbol{\theta}$ be a random variable having density $q_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ as in (1), and denote with $\mathbb{1}[\cdot]$ the indicator function, then

$$\boldsymbol{\theta} \stackrel{d}{=} \mathbb{1}[\mathrm{U} \leq w_{\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}})]\bar{\boldsymbol{\theta}} + (1 - \mathbb{1}[\mathrm{U} \leq w_{\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}})])(2\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}),$$

where $\bar{\boldsymbol{\theta}}$ has symmetric density $\bar{q}_{\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}})$, and U is uniformly distributed in $[0, 1]$.

Two important results

Proposition 1(a) [Tractability]. Let $\boldsymbol{\theta}$ be a random variable having density $q_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ as in (1), and denote with $\mathbb{1}[\cdot]$ the indicator function, then

$$\boldsymbol{\theta} \overset{d}{=} \mathbb{1}[\mathrm{U} \leq w_{\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}})]\bar{\boldsymbol{\theta}} + (1 - \mathbb{1}[\mathrm{U} \leq w_{\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}})])(2\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}),$$

where $\bar{\boldsymbol{\theta}}$ has symmetric density $\bar{q}_{\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}})$, and U is uniformly distributed in $[0,1]$.

Proposition 1(b) [Generality]. Consider the generic density $\pi_n(\boldsymbol{\theta})$ for the variable $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$, and let $\boldsymbol{\theta}^* \in \Theta$ be a symmetry point. Then $\pi_n(\boldsymbol{\theta})$ can be equivalently re–expressed in skew–symmetric form as

$$\pi_n(\boldsymbol{\theta}) = \frac{\pi_n(\boldsymbol{\theta}) + \pi_n(2\boldsymbol{\theta}^* - \boldsymbol{\theta})}{2} 2 \frac{\pi_n(\boldsymbol{\theta})}{\pi_n(\boldsymbol{\theta}) + \pi_n(2\boldsymbol{\theta}^* - \boldsymbol{\theta})} = 2\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}), \tag{2}$$

where $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is symmetric around $\boldsymbol{\theta}^* \in \Theta$, while $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) \in [0,1]$ and $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 1 - w_{n,\boldsymbol{\theta}^*}(2\boldsymbol{\theta}^* - \boldsymbol{\theta})$.

In Bayesian inference

In Bayesian inference the focus is on $\pi_n(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})L(\boldsymbol{\theta}; \mathbf{y}_{1:n})/c(\mathbf{y}_{1:n})$, where $\pi(\boldsymbol{\theta})$ is the prior, $L(\boldsymbol{\theta}; \mathbf{y}_{1:n})$ the likelihood, and $c(\mathbf{y}_{1:n})$ the normalizing constant.

In Bayesian inference

In Bayesian inference the focus is on $\pi_n(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})L(\boldsymbol{\theta}; \mathbf{y}_{1:n})/c(\mathbf{y}_{1:n})$, where $\pi(\boldsymbol{\theta})$ is the prior, $L(\boldsymbol{\theta}; \mathbf{y}_{1:n})$ the likelihood, and $c(\mathbf{y}_{1:n})$ the normalizing constant.

Proposition 2. Consider the generic posterior density $\pi_n(\boldsymbol{\theta})$ and denote with

$$\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = \frac{\pi_n(\boldsymbol{\theta}) + \pi_n(2\boldsymbol{\theta}^* - \boldsymbol{\theta})}{2}, \tag{3}$$

the symmetrized form of this posterior density about a point $\boldsymbol{\theta}^* \in \Theta$. Let $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = \pi_n(\boldsymbol{\theta})/[\pi_n(\boldsymbol{\theta}) + \pi_n(2\boldsymbol{\theta}^* - \boldsymbol{\theta})]$, which implies

$$w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta}; \mathbf{y}_{1:n})/\cancel{c(\mathbf{y}_{1:n})}}{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta}; \mathbf{y}_{1:n})/\cancel{c(\mathbf{y}_{1:n})} + \pi(2\boldsymbol{\theta}^* - \boldsymbol{\theta})L(2\boldsymbol{\theta}^* - \boldsymbol{\theta}; \mathbf{y}_{1:n})/\cancel{c(\mathbf{y}_{1:n})}}. \tag{4}$$

Then, $\pi_n(\boldsymbol{\theta})$ can be equivalently re–expressed in skew–symmetric form as

$$\pi_n(\boldsymbol{\theta}) = 2\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})[\pi_n(\boldsymbol{\theta})/2\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})] = 2\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}), \tag{5}$$

for any $\boldsymbol{\theta}^* \in \Theta$ and n, with $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) \in [0, 1]$, $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 1 - w_{n,\boldsymbol{\theta}^*}(2\boldsymbol{\theta}^* - \boldsymbol{\theta})$.

Some comments

- <u>**Positive aspect:**</u> The skewing factor $w_{n,\theta^*}(\theta)$ is available in closed form.

Some comments

- **Positive aspect:** The skewing factor $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is available in closed form.
- **Negative aspect:** The symmetrized posterior $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is intractable.

- **Positive aspect:** The skewing factor $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is available in closed form.
- **Negative aspect:** The symmetrized posterior $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is intractable.

  $\longrightarrow$ **Solution:** Replace $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ with a tractable symmetric density $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$, such as the Gaussians resulting from state–of–the–art deterministic approximations of intractable posterior densities from e.g., Laplace method, VB, EP, and others.

Some comments

- **Positive aspect:** The skewing factor $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is available in closed form.
- **Negative aspect:** The symmetrized posterior $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is intractable.

  $\longrightarrow$ **Solution:** Replace $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ with a tractable symmetric density $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$, such as the Gaussians resulting from state–of–the–art deterministic approximations of intractable posterior densities from e.g., Laplace method, VB, EP, and others. *Is this a good idea?*

Some comments

- **Positive aspect:** The skewing factor $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is available in closed form.
- **Negative aspect:** The symmetrized posterior $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is intractable.

  $\longrightarrow$ **Solution:** Replace $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ with a tractable symmetric density $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$, such as the Gaussians resulting from state–of–the–art deterministic approximations of intractable posterior densities from e.g., Laplace method, VB, EP, and others. *Is this a good idea?*

Lemma 1. Let $\pi_n(\boldsymbol{\theta})$ be a generic posterior density for $\boldsymbol{\theta} \in \Theta$, and denote with $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ an already–derived approximation of $\pi_n(\boldsymbol{\theta})$ which is symmetric about the point $\boldsymbol{\theta}^* \in \Theta$. Moreover, consider the symmetrized posterior density $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ about $\boldsymbol{\theta}^*$ defined as in (3). Then, for any $\boldsymbol{\theta}^* \in \Theta$ and n, it holds

$$\mathcal{D}[\bar{\pi}_{n,\boldsymbol{\theta}^*} \mid\mid \bar{q}_{n,\boldsymbol{\theta}^*}] \leq \mathcal{D}[\pi_n \mid\mid \bar{q}_{n,\boldsymbol{\theta}^*}],$$

where $\mathcal{D}$ is either the TV distance, KL, reverse–KL or a generic α–divergence.

Skew–symmetric approximations

Lemma 1 justifies replacing $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ with $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ in the skew–symmetric representation of $\pi_n(\boldsymbol{\theta})$ in (5). This implies that $2\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) \approx \pi_n(\boldsymbol{\theta})$. Thus, setting $q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 2\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ gives the improved skew–symmetric approximation of $\pi_n(\boldsymbol{\theta})$ in **Definition 2**.

Skew–symmetric approximations

Lemma 1 justifies replacing $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ with $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ in the skew–symmetric representation of $\pi_n(\boldsymbol{\theta})$ in (5). This implies that $2\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) \approx \pi_n(\boldsymbol{\theta})$. Thus, setting $q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 2\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ gives the improved skew–symmetric approximation of $\pi_n(\boldsymbol{\theta})$ in **Definition 2**.

Definition 2. Consider the posterior density $\pi_n(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta$, and let $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ denote an already–derived approximation of $\pi_n(\boldsymbol{\theta})$ which is symmetric about $\boldsymbol{\theta}^* \in \Theta$. Moreover, let $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) \in [0,1]$ be the skewing factor in (4). Then, the skew–symmetric approximation of $\pi_n(\boldsymbol{\theta})$ is

$$\begin{aligned} q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) &= 2\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) \\ &= 2\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})\frac{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta};\mathbf{y}_{1:n})}{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta};\mathbf{y}_{1:n})+\pi(2\boldsymbol{\theta}^*-\boldsymbol{\theta})L(2\boldsymbol{\theta}^*-\boldsymbol{\theta};\mathbf{y}_{1:n})}, \end{aligned} \tag{6}$$

for every symmetry point $\boldsymbol{\theta}^* \in \Theta$ and sample size n, where $\pi(\boldsymbol{\theta})$ and $L(\boldsymbol{\theta};\mathbf{y}_{1:n})$ are, respectively, the prior and likelihood inducing the posterior $\pi_n(\boldsymbol{\theta})$.

In practice

1. `Obtain` $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ `via e.g., Laplace, VB, EP, ... targeting` $\pi_n(\boldsymbol{\theta})$

for $s = 1, \ldots, N_{\text{SAMPL}}$ **do**

2a. `Sample` $\bar{\boldsymbol{\theta}}^{(s)}$ `from the distribution with density` $\bar{q}_{n,\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}})$

2b. `Sample` $\mathtt{U}^{(s)} \sim \mathtt{Unif}[0,1]$

2c. `If` $\mathtt{U}^{(s)} \leq w_{n,\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}}^{(s)})$ `set` $\boldsymbol{\theta}^{(s)} = \bar{\boldsymbol{\theta}}^{(s)}$`, otherwise` $\boldsymbol{\theta}^{(s)} = 2\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}^{(s)}$

end for

3. `Compute any functional of interest of the approximate posterior` $q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ `via Monte Carlo leveraging the above samples`

In practice

1. `Obtain` $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ `via e.g., Laplace, VB, EP, ... targeting` $\pi_n(\boldsymbol{\theta})$

for $s = 1, \dots, N_{\text{SAMPL}}$ **do**

2a. `Sample` $\bar{\boldsymbol{\theta}}^{(s)}$ `from the distribution with density` $\bar{q}_{n,\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}})$

2b. `Sample` $\texttt{u}^{(s)} \sim \texttt{Unif}[0, 1]$

2c. `If` $\texttt{u}^{(s)} \leq w_{n,\boldsymbol{\theta}^*}(\bar{\boldsymbol{\theta}}^{(s)})$ `set` $\boldsymbol{\theta}^{(s)} = \bar{\boldsymbol{\theta}}^{(s)}$, `otherwise` $\boldsymbol{\theta}^{(s)} = 2\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}^{(s)}$

end for

3. `Compute any functional of interest of the approximate posterior` $q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ `via Monte Carlo leveraging the above samples`

Remark: The above algorithm only requires simulation from the symmetric approximation $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ (e.g., Gaussian) and computation of the skewing factor $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$, which is analytically–available and it does not depend on intractable quantities (requires prior and likelihood evaluation).

Finite sample guarantees

Theorem 1. Consider the generic posterior density $\pi_n(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta$, and let $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ correspond to an already–derived approximation for $\pi_n(\boldsymbol{\theta})$ which is symmetric about $\boldsymbol{\theta}^* \in \Theta$. Moreover, let $q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 2\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$, where $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is defined as in (4). Then, for any $\boldsymbol{\theta}^* \in \Theta$ and n, it holds

$$\mathcal{D}[\pi_n \mid\mid q_{n,\boldsymbol{\theta}^*}] = \mathcal{D}[\bar{\pi}_{n,\boldsymbol{\theta}^*} \mid\mid \bar{q}_{n,\boldsymbol{\theta}^*}], \tag{7}$$

where $\mathcal{D}$ is either the TV distance, KL, reverse–KL or a generic α–divergence, while $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ corresponds to the symmetrized posterior density defined in (3). In view of Lemma 1, the result in (7) implies also

$$\mathcal{D}[\pi_n \mid\mid q_{n,\boldsymbol{\theta}^*}] \leq \mathcal{D}[\pi_n \mid\mid \bar{q}_{n,\boldsymbol{\theta}^*}], \tag{8}$$

for any $\boldsymbol{\theta}^* \in \Theta$ and sample size n.

Finite sample guarantees

Theorem 1. Consider the generic posterior density $\pi_n(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta$, and let $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ correspond to an already–derived approximation for $\pi_n(\boldsymbol{\theta})$ which is symmetric about $\boldsymbol{\theta}^* \in \Theta$. Moreover, let $q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 2\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$, where $w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ is defined as in (4). Then, for any $\boldsymbol{\theta}^* \in \Theta$ and n, it holds

$$\mathcal{D}[\pi_n \mid\mid q_{n,\boldsymbol{\theta}^*}] = \mathcal{D}[\bar{\pi}_{n,\boldsymbol{\theta}^*} \mid\mid \bar{q}_{n,\boldsymbol{\theta}^*}], \tag{7}$$

where $\mathcal{D}$ is either the TV distance, KL, reverse–KL or a generic α–divergence, while $\bar{\pi}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ corresponds to the symmetrized posterior density defined in (3). In view of Lemma 1, the result in (7) implies also

$$\mathcal{D}[\pi_n \mid\mid q_{n,\boldsymbol{\theta}^*}] \leq \mathcal{D}[\pi_n \mid\mid \bar{q}_{n,\boldsymbol{\theta}^*}], \tag{8}$$

for any $\boldsymbol{\theta}^* \in \Theta$ and sample size n.

Remark: It is also possible to prove the optimality of $q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ within the approximating family $\mathcal{Q} = \{\tilde{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) : \tilde{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta}) = 2\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})\tilde{w}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})\}$

Proof

Let us focus on the TV distance $\mathcal{D}_{\text{TV}}(\cdot||\cdot)$ and assume, without loss of generality, that $\boldsymbol{\theta}^* = 0$. Moreover, consider the partition $\Theta = \Theta_+ \cup \Theta_-$, where $\boldsymbol{\theta} \in \Theta_+$ implies $-\boldsymbol{\theta} \in \Theta_-$. Leveraging (5)–(6), along with $w_n(-\boldsymbol{\theta}) = 1 - w_n(\boldsymbol{\theta})$, and the symmetry of $\bar{\pi}_n(\boldsymbol{\theta})$ and $\bar{q}_n(\boldsymbol{\theta})$ (omit $\boldsymbol{\theta}^*$ for notational convenience):

$$\mathcal{D}_{\text{TV}}[\pi_n \mid\mid q_n] = \frac{1}{2} \int \left| 2\bar{\pi}_n(\boldsymbol{\theta}) w_n(\boldsymbol{\theta}) - 2\bar{q}_n(\boldsymbol{\theta}) w_n(\boldsymbol{\theta}) \right| d\boldsymbol{\theta}$$

Proof

Let us focus on the TV distance $\mathcal{D}_{\mathrm{TV}}(\cdot||\cdot)$ and assume, without loss of generality, that $\boldsymbol{\theta}^* = 0$. Moreover, consider the partition $\Theta = \Theta_+ \cup \Theta_-$, where $\boldsymbol{\theta} \in \Theta_+$ implies $-\boldsymbol{\theta} \in \Theta_-$. Leveraging (5)–(6), along with $w_n(-\boldsymbol{\theta}) = 1 - w_n(\boldsymbol{\theta})$, and the symmetry of $\bar{\pi}_n(\boldsymbol{\theta})$ and $\bar{q}_n(\boldsymbol{\theta})$ (omit $\boldsymbol{\theta}^*$ for notational convenience):

$$\mathcal{D}_{\mathrm{TV}}[\pi_n \mid\mid q_n] = \frac{1}{2} \int \Big| 2\bar{\pi}_n(\boldsymbol{\theta}) w_n(\boldsymbol{\theta}) - 2\bar{q}_n(\boldsymbol{\theta}) w_n(\boldsymbol{\theta}) \Big| d\boldsymbol{\theta}$$

$$= \frac{1}{2} \int_{\Theta_+} 2w_n(\boldsymbol{\theta}) \Big| \bar{\pi}_n(\boldsymbol{\theta}) - \bar{q}_n(\boldsymbol{\theta}) \Big| + \frac{1}{2} \int_{\Theta_+} 2w_n(-\boldsymbol{\theta}) \Big| \bar{\pi}_n(-\boldsymbol{\theta}) - \bar{q}_n(-\boldsymbol{\theta}) \Big| d\boldsymbol{\theta}$$

Proof

Let us focus on the TV distance $\mathcal{D}_{\mathrm{TV}}(\cdot||\cdot)$ and assume, without loss of generality, that $\boldsymbol{\theta}^* = 0$. Moreover, consider the partition $\Theta = \Theta_+ \cup \Theta_-$, where $\boldsymbol{\theta} \in \Theta_+$ implies $-\boldsymbol{\theta} \in \Theta_-$. Leveraging (5)–(6), along with $w_n(-\boldsymbol{\theta}) = 1 - w_n(\boldsymbol{\theta})$, and the symmetry of $\bar{\pi}_n(\boldsymbol{\theta})$ and $\bar{q}_n(\boldsymbol{\theta})$ (omit $\boldsymbol{\theta}^*$ for notational convenience):

$$\mathcal{D}_{\mathrm{TV}}[\pi_n \mid\mid q_n] = \frac{1}{2} \int \left|2\bar{\pi}_n(\boldsymbol{\theta}) w_n(\boldsymbol{\theta}) - 2\bar{q}_n(\boldsymbol{\theta}) w_n(\boldsymbol{\theta})\right| d\boldsymbol{\theta}$$

$$= \frac{1}{2} \int_{\Theta_+} 2w_n(\boldsymbol{\theta}) \left|\bar{\pi}_n(\boldsymbol{\theta}) - \bar{q}_n(\boldsymbol{\theta})\right| + \frac{1}{2} \int_{\Theta_+} 2w_n(-\boldsymbol{\theta}) \left|\bar{\pi}_n(-\boldsymbol{\theta}) - \bar{q}_n(-\boldsymbol{\theta})\right| d\boldsymbol{\theta}$$

$$= \frac{1}{2} \int_{\Theta_+} [2w_n(\boldsymbol{\theta}) \left|\bar{\pi}_n(\boldsymbol{\theta}) - \bar{q}_n(\boldsymbol{\theta})\right| + 2(1 - w_n(\boldsymbol{\theta})) \left|\bar{\pi}_n(\boldsymbol{\theta}) - \bar{q}_n(\boldsymbol{\theta})\right|] d\boldsymbol{\theta}$$

Proof

Let us focus on the TV distance $\mathcal{D}_{\mathrm{TV}}(\cdot||\cdot)$ and assume, without loss of generality, that $\boldsymbol{\theta}^* = 0$. Moreover, consider the partition $\Theta = \Theta_+ \cup \Theta_-$, where $\boldsymbol{\theta} \in \Theta_+$ implies $-\boldsymbol{\theta} \in \Theta_-$. Leveraging (5)–(6), along with $w_n(-\boldsymbol{\theta}) = 1 - w_n(\boldsymbol{\theta})$, and the symmetry of $\bar{\pi}_n(\boldsymbol{\theta})$ and $\bar{q}_n(\boldsymbol{\theta})$ (omit $\boldsymbol{\theta}^*$ for notational convenience):

$$\mathcal{D}_{\mathrm{TV}}[\pi_n \mid\mid q_n] = \frac{1}{2}\int \left|2\bar{\pi}_n(\boldsymbol{\theta})w_n(\boldsymbol{\theta}) - 2\bar{q}_n(\boldsymbol{\theta})w_n(\boldsymbol{\theta})\right| d\boldsymbol{\theta}$$

$$= \frac{1}{2}\int_{\Theta_+} 2w_n(\boldsymbol{\theta})\left|\bar{\pi}_n(\boldsymbol{\theta}) - \bar{q}_n(\boldsymbol{\theta})\right| + \frac{1}{2}\int_{\Theta_+} 2w_n(-\boldsymbol{\theta})\left|\bar{\pi}_n(-\boldsymbol{\theta}) - \bar{q}_n(-\boldsymbol{\theta})\right| d\boldsymbol{\theta}$$

$$= \frac{1}{2}\int_{\Theta_+} [2w_n(\boldsymbol{\theta})\left|\bar{\pi}_n(\boldsymbol{\theta}) - \bar{q}_n(\boldsymbol{\theta})\right| + 2(1 - w_n(\boldsymbol{\theta}))\left|\bar{\pi}_n(\boldsymbol{\theta}) - \bar{q}_n(\boldsymbol{\theta})\right|] d\boldsymbol{\theta}$$

$$= \int_{\Theta_+} \left|\bar{\pi}_n(\boldsymbol{\theta}) - \bar{q}_n(\boldsymbol{\theta})\right| d\boldsymbol{\theta} = \frac{1}{2}\int \left|\bar{\pi}_n(\boldsymbol{\theta}) - \bar{q}_n(\boldsymbol{\theta})\right| d\boldsymbol{\theta} = \mathcal{D}_{\mathrm{TV}}[\bar{\pi}_n \mid\mid \bar{q}_n],$$

Asymptotic guarantees

Theorem 1 ensures systematic accuracy gains of $q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ over $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ in approximating the target posterior, but does not quantify such gains.

Asymptotic guarantees

Theorem 1 ensures systematic accuracy gains of $q_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ over $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ in approximating the target posterior, but does not quantify such gains.

Theorem 2. Let $q_{n,\boldsymbol{\theta}_{\mathrm{MAP}}}(\boldsymbol{\theta}) = \phi_d(\boldsymbol{\theta}; \boldsymbol{\theta}_{\mathrm{MAP}}, \mathbf{J}^{-1}_{\boldsymbol{\theta}_{\mathrm{MAP}}}) w_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$. Then, under suitable assumptions, it holds

$$\mathcal{D}[\pi_n \mid\mid q_{n,\boldsymbol{\theta}_{\mathrm{MAP}}}] = O_{P_0^n}\left(M_n^{c_1} d^3/n\right), \tag{9}$$

where $M_n = \sqrt{c_0 \log n}$, c_0 and c_1 are fixed positive constants not depending on n and d, while $\mathcal{D}$ denotes either the TV distance, KL, reverse–KL or a generic α–divergence.

- $q_{n,\boldsymbol{\theta}_{\mathrm{MAP}}}(\boldsymbol{\theta})$ improves by an $\sqrt{n}$ factor the rate of $\phi_d(\boldsymbol{\theta}; \boldsymbol{\theta}_{\mathrm{MAP}}, \mathbf{J}^{-1}_{\boldsymbol{\theta}_{\mathrm{MAP}}})$
- rates rely of bounds that vanish also for growing d, as long as $d = o(n^{1/3})$
- can be improved by skewing higher–order symmetric approximations

Simulation study

◦ **Setting:** Simulate i.i.d. data $\mathbf{y}_{1:n}$, for growing sample size n from $n = 15$ to $n = 145$ with step size 10, from a Poisson distribution having rate $\exp(\theta_0) = 1$.

◦ **Focus:** For each n, assess the accuracy in approximating the posterior for θ, under a Student-t prior with one degree of freedom ($\rightarrow$ quantify **rates**).

◦ **Setting:** Simulate i.i.d. data $\mathbf{y}_{1:n}$, for growing sample size n from $n = 15$ to $n = 145$ with step size 10, from a Poisson distribution having rate $\exp(\theta_0) = 1$.

◦ **Focus:** For each n, assess the accuracy in approximating the posterior for θ, under a Student-t prior with one degree of freedom ($\rightarrow$ quantify **rates**).

Table: For routinely employed divergences, slope of the fitted linear regression between $\log \mathcal{D}(\pi_n \,||\, q_n)$ and $\log n$, where q_n is either $\bar{q}_{n,\theta^*}$ (i.e. Gaussian) or q_{n,θ^*} (i.e. Skew–symmetric), and n is defined on a grid from 15 to 145 with step size 10. Results are averaged across 50 replicated experiments (standard error within brackets).

	TV	KL	reverse-KL
Laplace (Gaussian)	−0.48 (0.01)	−0.93 (0.02)	−0.97 (0.02)
Laplace (Skew-symmetric)	**−1.04** (0.02)	**−1.80** (0.08)	**−3.11** (0.26)
black–box VB (Gaussian)	−0.48 (0.01)	−0.95 (0.02)	−0.98 (0.02)
black–box VB (Skew–symmetric)	**−1.05** (0.02)	**−1.73** (0.12)	**−3.18** (0.29)
EP (Gaussian)	−0.47 (0.01)	−0.93 (0.02)	−0.99 (0.02)
EP (Skew–symmetric)	**−0.99** (0.02)	**−1.76** (0.11)	**−3.47** (0.26)

Real–data application

◦ **Setting:** Hierarchical semiparametric logistic regression applied to data from a demographic study investigating factors underlying the use of contraceptives in a sample of $n = 30{,}524$ women in India ($d = 62$).

Real–data application

◦ **Setting:** Hierarchical semiparametric logistic regression applied to data from a demographic study investigating factors underlying the use of contraceptives in a sample of $n = 30{,}524$ women in India ($d = 62$).

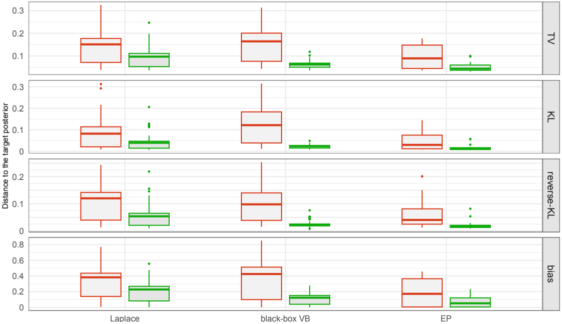


Figure: First three panels: boxplots of $\mathcal{D}(\pi_{j,n} \mid\mid q_{j,n})$, $j = 1, \ldots, 62$, where $q_{j,n}$ is the jth marginal of either $\bar{q}_{n,\boldsymbol{\theta}^*}$ (Gaussian) or $q_{n,\boldsymbol{\theta}^*}$ (Skew–symmetric). Fourth panel: boxplot of the absolute differences between approximated and actual posterior means.

Conclusion

Main message: Generic posterior densities can be re–expressed as members of the skew–symmetric family, with tractable skewing factor. This result has direct impact in the design of a general and theoretically-supported strategy to perturb any given symmetric approximation of a generic posterior density for obtaining an improved, yet similarly tractable, skew–symmetric counterpart.

Conclusion

Main message: Generic posterior densities can be re–expressed as members of the skew–symmetric family, with tractable skewing factor. This result has direct impact in the design of a general and theoretically-supported strategy to perturb any given symmetric approximation of a generic posterior density for obtaining an improved, yet similarly tractable, skew–symmetric counterpart.

- Explore general notions of symmetry, yielding more flexible perturbations mechanisms (this requires extending the skew–symmetric family).
- Optimize also with respect to $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$ (this would require extracting a symmetric component from $\pi_n(\boldsymbol{\theta})$, which is both close to $\pi_n(\boldsymbol{\theta})$ and can be accurately approximated by a tractable symmetric density $\bar{q}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\theta})$).
- Study the impact of the skew–symmetric approximation when applied either directly or as a building block in other popular strategies (ALA, INLA, stochastic VB, delta–method VB, importance sampling, ...).