Supplementary material for

Toward computerized efficient estimation in infinite-dimensional models

Verification of the KL divergence and Hellinger distance as appropriate divergences

Condition (B1) is well-known to hold for both $\mathbb{D}_{KL}$ and $\mathbb{D}_H$ (see, e.g., Pardo, 2005).

Condition (B2) holds trivially for $\mathbb{D}_H$ with $K_0(m_0, m_1) = K_1(m_0, m_1) = 1$ irrespective of m_0 and m_1. The lower bound of (B2) is also known to hold for $\mathbb{D}_{KL}$ with $K_0(m_0, m_1) = 1$ since

$$\begin{aligned}\mathbb{D}_{KL}(P_1, P_2) \ &\geq \ -\int 2\left[\sqrt{\frac{dP_1}{dP_2}(u)} - 1\right] dP_2(u)\\ &= \ 2 - 2\int \sqrt{\frac{dP_1}{dP_2}(u)}dP_2(u) \ = \ \int \left[1 - \sqrt{\frac{dP_1}{dP_2}(u)}\right]^2 dP_2(u)\end{aligned}$$

in view of the fact that $\log(x) \leq 2(\sqrt{x} - 1)$ for each $x > 0$. A Taylor expansion of the function $w \mapsto \log(1+w)$ around $w = 0$ suggests further study of the function f defined as

$$f(w) := \frac{2\log(1 + w) - 2w + w^2}{2w^3}$$

for $w \neq 0$ and $f(0) = 1/3$, which then allows us to write $2\log(1+w) = 2w - w^2 + 2w^3 f(w)$. It is not difficult to show that f is decreasing and non-negative on $(-1, +\infty)$, and so, over any interval of the form $[a, +\infty)$ for $a > -1$, f is bounded in $[0, f(a)]$ with $0 < f(a) < +\infty$. We will set $w = \sqrt{\frac{dP_1}{dP_2}(u)} - 1$ in the above. With simple manipulations, we can show that

$$\int \left\{2\log\left[1 + \sqrt{\frac{dP_1}{dP_2}(u)} - 1\right] - 2\left[\sqrt{\frac{dP_1}{dP_2}(u)} - 1\right] + \left[\sqrt{\frac{dP_1}{dP_2}(u)} - 1\right]^2\right\} dP_2(u)$$

simplifies to $2\mathbb{D}_H(P_1, P_2) - \mathbb{D}_{KL}(P_1, P_2)$. This gives that

$$2\mathbb{D}_H(P_1, P_2) - \mathbb{D}_{KL}(P_1, P_2) \ = \ 2\int \left[\sqrt{\frac{dP_1}{dP_2}(u)} - 1\right]^3 f\left(\sqrt{\frac{dP_1}{dP_2}(u)} - 1\right) dP_2(u) \ . \qquad (1)$$

Of course, we can always consider that $0 < m_0 < 1 < m_1 < +\infty$ without loss of generality. In view of (1), defining $M_0 := \sqrt{m_0} - 1$, $M_1 := \sqrt{m_1} - 1$ and $M := \max(-M_0, M_1)$, we find that

$$2\mathbb{D}_H(P_1, P_2) - \mathbb{D}_{KL}(P_1, P_2) \;\geq\; -2f(M_0) \int \left| \sqrt{\frac{dP_1}{dP_2}(u)} - 1 \right|^3 dP_2(u) \;\geq\; -2Mf(M_0)\mathbb{D}_H(P_1, P_2) \;.$$

As such, the upper bound of (B2) holds for $\mathbb{D}_{KL}$ with $K_1(m_0, m_1) := 2 + 2Mf(M_0)$. We have thus shown that condition (B2) holds for both $\mathbb{D}_H$ and $\mathbb{D}_{KL}$.

Condition (B3) can be established readily for $\mathbb{D}_{KL}$ – the argument was sketched in the body of the paper. In fact, as was argued before, if $\mathbb{D} = \mathbb{D}_{KL}$, then $\int h(u) dP_2(u) = 0$ for each $h \in T_{\mathcal{M}}(P_2^*)$, which is a stronger property that implies condition (B3). Establishing that this condition holds for $\mathbb{D}_H$ requires more work. For arbitrary bounded h in the interior of $T_{\mathcal{M}}(P_2^*)$, we define the parametric path $P_{2,\eta}^*$ via $dP_{2,\eta}^* = (1 + \epsilon h) dP_2^*$ with index η taking values in a neighborhood of zero. We can compute that

$$\frac{d}{d\eta}\mathbb{D}_{KL}(P_{2,\eta}^*, P_2)\bigg|_{\eta=0} = \int h(u) dP_2(u) \quad \text{and} \quad \frac{d}{d\eta}\mathbb{D}_H(P_{2,\eta}^*, P_2)\bigg|_{\eta=0} = 0 \;,$$

where the latter equality follows from the fact that P_2^* is a global minimizer of the Hellinger distance over $\mathcal{M}(P_2)$. We now focus on the right-hand side of (1). Writing

$$\mathbb{E}_h(P_2^*, P_2) := 2\,\frac{d}{d\eta} \int \left[\sqrt{\frac{dP_{2,\eta}^*}{dP_2}(u)} - 1 \right]^3 f\left(\sqrt{\frac{dP_{2,\eta}^*}{dP_2}(u)} - 1 \right) dP_2(u) \Bigg|_{\eta=0} \;,$$

by the product rule for differentiation, we find that $\mathbb{E}_h(P_2^*, P_2) = \mathbb{E}_{1,h}(P_2^*, P_2) + \mathbb{E}_{2,h}(P_2^*, P_2)$, where

$$\begin{aligned}
\mathbb{E}_{1,h}(P_2^*, P_2) \;&:=\; 3 \int \left[\sqrt{\frac{dP_2^*}{dP_2}(u)} - 1 \right]^2 f\left(\sqrt{\frac{dP_2^*}{dP_2}(u)} - 1 \right) \sqrt{\frac{dP_2^*}{dP_2}(u)} h(u) dP_2(u) \\
\mathbb{E}_{2,h}(P_2^*, P_2) \;&:=\; \int \left[\sqrt{\frac{dP_2^*}{dP_2}(u)} - 1 \right]^3 f'\left(\sqrt{\frac{dP_2^*}{dP_2}(u)} - 1 \right) \sqrt{\frac{dP_2^*}{dP_2}(u)} h(u) dP_2(u) \;.
\end{aligned}$$

By direct calculation, we find that

$$f'(w) = \frac{1}{w} - \frac{1}{1+w} - \frac{3}{2w^2} + \frac{3}{w^3} - \frac{3\log(1+w)}{w^4}$$

for $w \neq 0$ and $f'(0) = -1/4$. Since f is decreasing on $(-1, +\infty)$, f' is negative there. We can also show that f' is non-decreasing on $(-1, +\infty)$. As such, over any interval of the form $[a, +\infty)$ for $a > -1$, f' is bounded in $[f'(a), 0]$ with $-\infty < f'(a) < 0$. Using these facts about f and f', we then obtain that

$$
\begin{aligned}
|\mathbb{E}_{1,h}(P_2^*, P_2)| &\leq 3\sqrt{m_1} f(M_0) \|h\|_{\infty, \mathfrak{X}(P_2)} \left\| \frac{dP_2^*}{dP_2} - 1 \right\|_{2,P_2}^2 \\
|\mathbb{E}_{2,h}(P_2^*, P_2)| &\leq \sqrt{m_1} M |f'(M_0)| \|h\|_{\infty, \mathfrak{X}(P_2)} \left\| \frac{dP_2^*}{dP_2} - 1 \right\|_{2,P_2}^2 .
\end{aligned}
$$

Setting $C(m_0, m_1) := \sqrt{m_1}\,[M|f'(M_0)| + 3f(M_0)]$, this then implies that

$$
\left| \int h(u) dP_2(u) \right| = |E_h(P_2^*, P_2)| \leq C(m_0, m_1) \|h\|_{\infty, \mathfrak{X}(P_2)} \left\| \frac{dP_2^*}{dP_2} - 1 \right\|_{2,P_2}^2 .
$$

This inequality allows us to conclude that condition (B3) holds for $\mathbb{D}_H$ with $B(\delta) := C(m_0, m_1) K \delta$, where K is the uniform bound considered on h. Clearly, $B(\delta)$ tends to zero as δ tends to zero.

Direct verification of the validity of the proposed representation of the EIF in examples

Below, we directly verify that the key representation proposed in this paper indeed yields the EIF in the particular problems studied numerically in Section 5.

Example 1. We denote by μ_0 the mean of P. We define $\mathbb{U}_\lambda(\epsilon, \xi) := \int \left[\frac{u}{1-\xi(u-\mu_0)} - \mu_0 \right] dP_{\epsilon,\lambda}(u)$, and take $\xi_0(\epsilon, \lambda)$ to be a solution in ξ of $\mathbb{U}_\lambda(\epsilon, \xi) = 0$. We note that $\xi_0(0, \lambda) = 0$. We have that

$$
\begin{aligned}
\left. \frac{\partial}{\partial \epsilon} \mathbb{U}_\lambda(\epsilon, \xi) \right|_{\epsilon=0, \xi=0} &= \int u dH_{x,\lambda}(u) - \int u dP(u) = \int u dH_{x,\lambda}(u) - \mu_0 \\
\left. \frac{\partial}{\partial \xi} \mathbb{U}_\lambda(\epsilon, \xi) \right|_{\epsilon=0, \xi=0} &= \int (u - \mu_0)^2 dP(u) .
\end{aligned}
$$

It is easy to verify that $\mathbb{U}_\lambda$ is continuously differentiable around $(0, 0)$, and since $\left. \frac{\partial}{\partial \xi} \mathbb{U}_\lambda(\epsilon, \xi) \right|_{\epsilon=0, \xi=0} \neq 0$, the Implicit Function Theorem applies. In particular, it states that there exists an open set $B \subset \mathbb{R}$ containing 0 such that there is a unique continuously differentiable function η_λ such that $\mathbb{U}_\lambda(\epsilon, \eta_\lambda(\epsilon)) = 0$ for each $\epsilon \in B$, and additionally that

$$
\left. \frac{\partial}{\partial \epsilon} \eta_\lambda(\epsilon) \right|_{\epsilon=0} = - \left[\left. \frac{\partial}{\partial \xi} \mathbb{U}_\lambda(\epsilon, \xi) \right|_{\epsilon=0, \xi=0} \right]^{-1} \left. \frac{\partial}{\partial \epsilon} \mathbb{U}_\lambda(\epsilon, \xi) \right|_{\epsilon=0, \xi=0} = - \frac{\int (u - \mu_0) dH_{x,\lambda}(u)}{\int (u - \mu_0)^2 dP(u)} .
$$

Now, we note that the density $p_{\epsilon,\lambda}^*$ of $P_{\epsilon,\lambda}^*$ is given pointwise as $p_{\epsilon,\lambda}^*(u) = p_{\epsilon,\lambda}(u)/[1 - \xi_0(\epsilon, \lambda)u]$, and thus, we find that $\Psi(P^*_{\epsilon,\lambda}) = \int \{p_{\epsilon,\lambda}(u)/[1-\xi_0(\epsilon,\lambda)(u-\mu_0)]\}^2 du$. We can then compute

$$\begin{aligned}
\frac{d}{d\epsilon}\Psi(P^*_{\epsilon,\lambda})\Big|_{\epsilon=0} &= 2\int p(u)dH_{x,\lambda}(u) - 2\Psi(P) + 2\,\frac{\partial}{\partial\epsilon}\xi_0(\epsilon,\lambda)\Big|_{\epsilon=0}\int (u-\mu_0)p(u)dP(u) \\
&= 2\int p(u)dH_{x,\lambda}(u) - 2\Psi(P) - 2\int (u-\mu_0)dH_{x,\lambda}(u)\frac{\int (u-\mu_0)p(u)dP(u)}{\int (u-\mu_0)^2 dP(u)} \\
&= 2\left[\int p(u)dH_{x,\lambda}(u) - \Psi(P)\right] - \int (u-\mu_0)dH_{x,\lambda}(u)\frac{\int (u-\mu_0)\phi_{\mathrm{NP},P}(u)dP(u)}{\int (u-\mu_0)^2 dP(u)},
\end{aligned}$$

from which we see that

$$\lim_{\lambda\to 0}\frac{d}{d\epsilon}\Psi(P^*_{\epsilon,\lambda})\Big|_{\epsilon=0} = 2\left[p(x)-\Psi(P)\right] - (x-\mu_0)\frac{\int (u-\mu_0)\phi_{\mathrm{NP},P}(u)dP(u)}{\int (u-\mu_0)^2 dP(u)} = \phi_P(x)\ .$$

Example 2. We define $\mathbb{U}_\lambda(\epsilon,v) := \frac{\partial}{\partial v}\mathbb{L}_{\epsilon,\lambda}(v)$ with $\mathbb{L}_{\epsilon,\lambda}(v) := \int \log\left[\frac{p_{\epsilon,\lambda}(u)+p_{\epsilon,\lambda}(2v-u)}{2}\right] dP_{\epsilon,\lambda}(u)$, and we consider taking $\mu^*_{\epsilon,\lambda}$ to be a solution in v of the equation $\mathbb{U}_\lambda(\epsilon,v) = 0$ around $v=\mu$. In fact, it is this equivalent formulation in terms of an equation that we use in the implementation of Example 2 in the numerical results provided. For simplicity, suppose that $\mu = 0$ and that ν is the Lebesgue measure on $\mathbb{R}$. It is not difficult to show that the mapping $\mathbb{U}_\lambda$ is continuously differentiable around $(0,0)$, and also that

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\mathbb{U}_\lambda(\epsilon,v)\Big|_{\epsilon=0,v=0} &= -\frac{1}{2}\int\left[\frac{\dot p(u)}{p(u)}\right]dH_{x,\lambda}(u) \\
\frac{\partial}{\partial v}\mathbb{U}_\lambda(\epsilon,v)\Big|_{\epsilon=0,v=0} &= -\frac{1}{2}\int\left[\frac{\dot p(u)}{p(u)}\right]^2 dP(u) = -\frac{1}{2}\cdot I(f)
\end{aligned}$$

using basic calculus techniques and algebraic manipulations. Since $\frac{\partial}{\partial v}\mathbb{U}_\lambda(\epsilon,v)\big|_{\epsilon=0,v=0} \neq 0$, the Implicit Function Theorem indicates that there exists an open set $B \subset \mathbb{R}$ containing 0 such that there is a unique continuously differentiable function η_λ such that $\mathbb{U}_\lambda(\epsilon,\eta_\lambda(\epsilon)) = 0$ for each $\epsilon \in B$, and additionally that

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\eta_\lambda(\epsilon)\Big|_{\epsilon=0} &= -\left[\frac{\partial}{\partial v}\mathbb{U}_\lambda(\epsilon,v)\Big|_{\epsilon=0,v=0}\right]^{-1}\frac{\partial}{\partial\epsilon}\mathbb{U}_\lambda(\epsilon,v)\Big|_{\epsilon=0,v=0} \\
&= -\frac{1}{I(f)}\int\left[\frac{\dot p(u)}{p(u)}\right]dH_{x,\lambda}(u)\ .
\end{aligned}$$

By the uniqueness of η_λ, it must be that $\mu^*_{\epsilon,\lambda} = \eta_\lambda(\epsilon)$ for ϵ small enough, and so, we find that

$$\lim_{\lambda\to 0}\frac{d}{d\epsilon}\Psi(P^*_{\epsilon,\lambda})\Big|_{\epsilon=0} = \lim_{\lambda\to 0}\frac{d}{d\epsilon}\eta_\lambda(\epsilon)\Big|_{\epsilon=0} = -\frac{1}{I(f)}\frac{\dot f(x-\mu)}{f(x-\mu)} = \phi_P(x)\ .$$

Similar arguments can be used to show this result using the Hellinger distance instead.

Example 3. Writing $E_{\epsilon,\lambda}$ to denote expectation under $P_{\epsilon,\lambda}$, we define

$$\mathbb{U}_\lambda(\epsilon,\beta) := E_{\epsilon,\lambda}\left(ZY\right) - E_{\epsilon,\lambda}\left[\frac{E_{\epsilon,\lambda}\left(Y \mid W\right) E_{\epsilon,\lambda}\left(Ze^{\beta W} \mid W\right)}{E_{\epsilon,\lambda}\left(e^{\beta Z} \mid W\right)}\right].$$

Using the law of total expectation, we have that $E\left(YZ\right) = E\left[e^{g_0(W)}E\left(Ze^{\beta_0 Z} \mid W\right)\right]$ and $E\left(Y \mid W\right) = e^{g_0(W)}E\left(e^{\beta_0 Z} \mid W\right)$. Through rather tedious calculations, it is possible to show that

$$\frac{\partial}{\partial\epsilon}\mathbb{U}_\lambda(\epsilon,\beta)\bigg|_{\epsilon=0,\beta=\beta_0} = y\left[z - \int a_P(u)H_{w,\lambda}(du)\right] - e^{\beta_0 z}\int e^{g_0(u)}H_{w,\lambda}(du)\left[z - \frac{\int e^{g_0(u)}a_P(u)H_{w,\lambda}(du)}{\int e^{g_0(u)}H_{w,\lambda}(du)}\right]$$

and furthermore, that

$$\begin{aligned}\frac{\partial}{\partial\beta}\mathbb{U}_\lambda(\epsilon,\beta)\bigg|_{\epsilon=0,v=0} &= -E\left(Z^2Y\right) + E\left[\frac{E^2\left(ZY \mid W\right)}{E\left(Y \mid W\right)}\right] = -E\left(Z^2Y\right) + E\left[a_P^2(W)E\left(Y \mid W\right)\right] \\ &= -E\left(Z^2Y\right) + E\left[Ya_P^2(W)\right] .\end{aligned}$$

We can use the fact that $[Z - a_P(W)][Y - E\left(Y \mid Z, W\right)]$ has mean zero given (Z, W) to write

$$\begin{aligned} var\left\{[Z - a_P(W)][Y - E\left(Y \mid Z,W\right)]\right\} &= E\left(var\left\{[Z - a_P(W)][Y - E\left(Y \mid Z,W\right)] \mid Z, W\right\}\right) \\ &= E\left\{[Z - a_P(W)]^2 var\left(Y \mid Z,W\right)\right\} \\ &= E\left\{[Z - a_P(W)]^2 E\left(Y \mid Z,W\right)\right\} = E\left\{Y[Z - a_P(W)]^2\right\} .\end{aligned}$$

Using that $E\left[YZa_P(W)\right] = E\left[E(YZ \mid W)a_P(W)\right] = E\left[Ya_P^2(W)\right]$, we thus see that

$$\frac{\partial}{\partial\beta}\mathbb{U}_\lambda(\epsilon,\beta)\bigg|_{\epsilon=0,v=0} = -var\left\{[Z - a_P(W)][Y - E\left(Y \mid Z,W\right)]\right\}.$$

It can be shown that the mapping $\mathbb{U}_\lambda$ is continuously differentiable around $(0,\beta_0)$ and $\frac{\partial}{\partial\beta}\mathbb{U}_\lambda(\epsilon,\beta)\Big|_{\epsilon=0,v=0} \neq 0$. Hence, the Implicit Function Theorem indicates that there exists an open set $B \subset \mathbb{R}$ containing 0 such that there is a unique continuously differentiable function η_λ such that $\mathbb{U}_\lambda(\epsilon, \eta_\lambda(\epsilon)) = 0$ for each $\epsilon \in B$, and additionally that

$$\begin{aligned}\frac{\partial}{\partial\epsilon}\eta_\lambda(\epsilon)\bigg|_{\epsilon=0} &= -\left[\frac{\partial}{\partial\beta}\mathbb{U}_\lambda(\epsilon,\beta)\bigg|_{\epsilon=0,\beta=\beta_0}\right]^{-1}\frac{\partial}{\partial\epsilon}\mathbb{U}_\lambda(\epsilon,\beta)\bigg|_{\epsilon=0,\beta=\beta_0} \\ &= \frac{y\left[z - \int a_P(u)H_{w,\lambda}(du)\right] - e^{\beta_0 z}\int e^{g_0(u)}H_{w,\lambda}(du)\left[z - \frac{\int e^{g_0(u)}a_P(u)H_{w,\lambda}(du)}{\int e^{g_0(u)}H_{w,\lambda}(du)}\right]}{var\left\{[Z - a_P(W)][Y - E\left(Y \mid Z,W\right)]\right\}} .\end{aligned}$$

By the uniqueness of η_λ, it must be that $\mu^*_{\epsilon,\lambda} = \eta_\lambda(\epsilon)$ for ϵ small enough, and so, we find that

$$\lim_{\lambda\to 0} \frac{d}{d\epsilon}\Psi(P^*_{\epsilon,\lambda})\Big|_{\epsilon=0} \;=\; \lim_{\lambda\to 0} \frac{d}{d\epsilon}\eta_\lambda(\epsilon)\Big|_{\epsilon=0} \;=\; \frac{[z-a_P(w)][y-E\,(Y\mid Z,W)]}{var\,\{[Z-a_P(W)][Y-E\,(Y\mid Z,W)]\}} \;=\; \phi_P(x)\ .$$

Example 4. In this example, there is a closed form expression for the projected distribution $P^*_{\epsilon,\lambda}$. Specifically, the projection of $p_{L_j,1,\epsilon,\lambda}$ onto $\mathcal{M}$ has explicit form

$$\begin{aligned} p^*_{L_j,1,\epsilon,\lambda}(\ell'_j\mid\ell'_{j-1}) \;&=\; \frac{\int_{\ell'_0}\int_{\ell'_1}\cdots\int_{\ell'_{j-2}} p_{\overline{L}_j,\overline{A}_{j-1},\epsilon,\lambda}(\overline{\ell}'_j,1_j)d\overline{\ell}'_j}{\int_{\ell'_0}\int_{\ell'_1}\cdots\int_{\ell'_{j-2}} p_{\overline{L}_{j-1},\overline{A}_{j-1},\epsilon,\lambda}(\overline{\ell}'_{j-1},1_j)d\overline{\ell}'_{j-1}} \\ &=\; \frac{(1-\epsilon)\int_{\ell'_0}\int_{\ell'_1}\cdots\int_{\ell'_{j-2}} p_{\overline{L}_j,\overline{A}_{j-1}}(\overline{\ell}'_j,1_j)d\overline{\ell}'_j + \epsilon a_0a_1\ldots a_{j-1}K_\lambda(\ell_j-\ell'_j)K_\lambda(\ell_{j-1}-\ell'_{j-1})}{(1-\epsilon)\int_{\ell'_0}\int_{\ell'_1}\cdots\int_{\ell'_{j-2}} p_{\overline{L}_{j-1},\overline{A}_{j-1}}(\overline{\ell}'_{j-1},1_j)d\overline{\ell}'_{j-1} + \epsilon a_0a_1\ldots a_{j-1}K_\lambda(\ell_{j-1}-\ell'_{j-1})} \end{aligned}$$

for $j=1,2,\ldots,K+1$, and furthermore, it is easy to see that $p^*_{L_0,\epsilon,\lambda} = p_{L_0,\epsilon,\lambda}$. We can compute

$$\frac{d}{d\epsilon}\Psi(P^*_{\epsilon,\lambda})\Big|_{\epsilon=0} \;=\; \frac{\partial}{\partial\epsilon}\int_{\ell'_0}\int_{\ell'_1}\cdots\int_{\ell'_{K+1}} \ell'_{K+1}\left[\prod_{j=1}^{K+1} p^*_{L_j,1,\epsilon,\lambda}(\ell'_j\mid\ell'_{j-1})\right] p_{L_0,\epsilon,\lambda}(\ell'_0)d\overline{\ell}'_{K+1}\Big|_{\epsilon=0} \;=\; \sum_{r=0}^{K+1}\mathbb{A}_{r,\lambda}\ ,$$

where we write, for $r=1,2,\ldots,K+1$,

$$\begin{aligned} \mathbb{A}_{r,\lambda} \;&=\; \int_{\ell'_0}\int_{\ell'_1}\cdots\int_{\ell'_{K+1}} \ell'_{K+1}\left[\prod_{j\neq r} p_{L_j,1}(\ell'_j\mid\ell'_{j-1})\right]\frac{\partial}{\partial\epsilon}p^*_{L_r,1,\epsilon,\lambda}(\ell'_r\mid\ell'_{r-1})\Big|_{\epsilon=0} p_{L_0}(\ell'_0)d\overline{\ell}'_{K+1}\ , \\ \mathbb{A}_{0,\lambda} \;&=\; \int_{\ell'_0}\int_{\ell'_1}\cdots\int_{\ell'_{K+1}} \ell'_{K+1}\left[\prod_{j=1}^{K+1} p_{L_j,1}(\ell'_j\mid\ell'_{j-1})\right]\left[K_\lambda(\ell_0-\ell'_0)-p_{L_0}(\ell'_0)\right]d\overline{\ell}'_{K+1}\ . \end{aligned}$$

We can calculate that

$$\frac{\partial}{\partial\epsilon}p^*_{L_j,1,\epsilon,\lambda}(\ell'_j\mid\ell'_{j-1})\Big|_{\epsilon=0} \;=\; \frac{a_0a_1\ldots a_{j-1}}{p_{L_{j-1},\overline{A}_{j-1}}(\ell'_{j-1},1_j)}K_\lambda(\ell_{j-1}-\ell'_{j-1})\left[K_\lambda(\ell_j-\ell'_j)-p_{L_j,1}(\ell'_j\mid\ell'_{j-1})\right].$$

We first investigate how $\mathbb{A}_{r,\lambda}$, $r\in\{1,2,\ldots,K+1\}$, can be simplified in view of this result. It is not difficult to see that $\mathbb{A}_{r,\lambda} = a_0a_1\ldots a_{r-1}\tilde{A}_{r,\lambda}$, where we define $\tilde{A}_{r,\lambda}$ to be

$$\int_{\ell'_0}\int_{\ell'_1}\cdots\int_{\ell'_{r-1}}\left[\int_{\ell'_r} m_{r,P}(\ell'_r)K_\lambda(\ell_r-\ell'_r)d\ell'_r - m_{r-1,P}(\ell'_{j-1})\right]K_\lambda(\ell_{r-1}-\ell'_{r-1})\frac{\prod_{j=0}^{r-1}p_{L_j,1}(\ell'_j\mid\ell'_{j-1})}{p_{L_{r-1},\overline{A}_{r-1}}(\ell'_{r-1},1_r)}d\overline{\ell}'_{r-1}\ .$$

We note that, as λ tends to zero, $\tilde{\mathbb{A}}_{r,\lambda}$ tends to

$$\frac{m_{r,P}(\ell_r) - m_{r-1,P}(\ell_{r-1})}{p_{L_{r-1},\overline{A}_{r-1}}(\ell_{r-1}, 1_r)} \int_{\ell_0'} \int_{\ell_1'} \cdots \int_{\ell_{r-2}'} \left[p_{L_{r-1},1}(\ell_{r-1} \mid \ell_{r-2}') \prod_{j=0}^{r-2} p_{L_j,1}(\ell_j' \mid \ell_{j-1}') \right] d\overline{\ell}_{r-2}'$$

$$= \frac{[m_{r,P}(\ell_r) - m_{r-1,P}(\ell_{r-1})] \int_{\ell_0'} \int_{\ell_1'} \cdots \int_{\ell_{r-2}'} p_{L_r,1}(\ell_r \mid \ell_{r-1}) p_{L_{r-1},1}(\ell_{r-1} \mid \ell_{r-2}') \prod_{j=0}^{r-2} p_{L_j,1}(\ell_j' \mid \ell_{j-1}') d\overline{\ell}_{r-2}'}{p_{L_r,L_{r-1},\overline{A}_{r-1}}(\ell_r, \ell_{r-1}, 1_r)}$$

$$= \frac{[m_{r,P}(\ell_r) - m_{r-1,P}(\ell_{r-1})] \int_{\ell_0'} \int_{\ell_1'} \cdots \int_{\ell_{r-2}'} p_{L_r,L_{r-1},\overline{L}_{r-2},\overline{A}_{r-1}}(\ell_r, \ell_{r-1}, \overline{\ell}_{r-2}', 1_r) d\overline{\ell}_{r-2}'}{p_{A_{r-1}|L_{r-1},\overline{L}_{r-2},\overline{A}_{r-2}}(1 \mid \ell_{r-1}, \ell_{r-2}', 1_{r-1}) \left[\prod_{j=0}^{r-2} p_{A_j|\overline{L}_j,\overline{A}_{j-1}}(1 \mid \ell_j', 1_j)\right] p_{L_r,L_{r-1},\overline{A}_{r-1}}(\ell_r, \ell_{r-1}, 1_r)}$$

$$= [m_{r,P}(\ell_r) - m_{r-1,P}(\ell_{r-1})] \frac{p_{\overline{L}_{r-2}|L_r,L_{r-1},\overline{A}_{r-1}}(\overline{\ell}_{r-2}' \mid \ell_r, \ell_{r-1}, 1_r)}{p_{A_{r-1}|L_{r-1},\overline{L}_{r-2},\overline{A}_{r-2}}(1 \mid \ell_{r-1}, \ell_{r-2}', 1_{r-1}) \left[\prod_{j=0}^{r-2} p_{A_j|\overline{L}_j,\overline{A}_{j-1}}(1 \mid \ell_j', 1_j)\right]}$$

$$= [m_{r,P}(\ell_r) - m_{r-1,P}(\ell_{r-1})] \, E_P \left[\frac{1}{\prod_{j=0}^{r-1} p_{A_j|\bar{L}_j,\overline{A}_{j-1}}(1 \mid \overline{L}_j, \overline{A}_{j-1})} \;\middle|\; L_r = \ell_r, L_{r-1} = \ell_{r-1}, \overline{A}_{r-1} = 1_r \right],$$

and so, we find that $\mathbb{A}_{r,\lambda}$ tends to $\phi_{r,P}(x)$. We can also verify that

$$\mathbb{A}_{0,\lambda} = \int_{\ell_0'} m_{0,P}(\ell_0') \left[K_\lambda(\ell_0 - \ell_0') - p_{L_0}(\ell_0')\right] d\ell_0' = \int_{\ell_0'} m_{0,P}(\ell_0') K_\lambda(\ell_0 - \ell_0') d\ell_0' - \Psi(P)$$

and so, we readily see that $\mathbb{A}_{0,\lambda}$ tends to $\phi_{0,P}(x)$ as λ tends to zero. We thus conclude that

$$\left.\frac{d}{d\epsilon} \Psi(P_{\epsilon,\lambda}^*)\right|_{\epsilon=0} = \sum_{r=0}^{K+1} \mathbb{A}_{r,\lambda} \longrightarrow \sum_{r=0}^{K+1} \phi_{r,P}(x) = \phi_P(x)$$

as λ tends to zero, which confirms the validity of our proposed representation in this problem.

Example in which the use of an *inappropriate* divergence leads to an invalid result

To illustrate the importance of choosing an appropriate divergence, e.g., a divergence satisfying conditions (B1)–(B3), we explicitly calculate the limit provided in (2.3) in a simple parametric problem in which an inappropriate divergence has been chosen, and we show that it does not coincide with the EIF.

We take $\mathcal{M} := \{P_\theta : \theta \in \Theta\}$ for some compact subset $\Theta \subset \mathbb{R}$, where all elements of $\mathcal{M}$ are dominated by a common measure ν, and consider the divergence $\mathbb{D}$ defined as $(P_1, P_2) \mapsto \int [p_1(u) - p_2(u)]^2 \, d\nu(u)$, where p_1 and p_2 denotes, respectively, the density functions of P_1 and P_2 relative to ν. As the $L_2(\nu)$ distance between two densities, this particular divergence is very natural. Nevertheless, it fails to be appropriate for our purposes, as we show below.

Taking $P_0 = P_{\theta_0}$, fixing $x \in \mathbb{R}$ and defining $\Psi(P_\theta) = \theta$, we focus on obtaining the value of the EIF of Ψ relative to $\mathcal{M}$ evaluated at distribution P_0 and observation value x. It is well known that this value is given by $\phi_{P_0}(x) := -I(\theta_0)^{-1}s_0(x)$, where $s_0 := s_{\theta_0}$, $s_\theta(x) := \frac{\partial}{\partial\theta}\log p_\theta(x)$ and $I(\theta_0) := -\int \frac{\partial}{\partial\theta}s_\theta(x)\big|_{\theta=\theta_0} dP_0(x)$. We define the projection

$$\theta^*_{\epsilon,\lambda} := \operatorname*{argmin}_{\theta\in\Theta} \int \left[p_\theta(u) - p_{\epsilon,\lambda}(u)\right]^2 d\nu(u)$$

of $p_{\epsilon,\lambda} = (1-\epsilon)p_0 + \epsilon h_{x,\lambda}$, where p_0 and $h_{x,\lambda}$ are, respectively, the density functions of P_0 and $H_{x,\lambda}$ relative to ν, onto the parametric model $\mathcal{M}$. As such, $\theta^*_{\epsilon,\lambda}$ is the solution in θ of the equation $\mathbb{U}_\lambda(\theta,\epsilon) = 0$, where

$$\mathbb{U}_\lambda(\theta,\epsilon) := \int \left[p_\theta(u) - p_{\epsilon,\lambda}(u)\right] \dot{p}_\theta(u) d\nu(u)$$

and we write $\dot{p}_\theta(x) := \frac{\partial}{\partial\theta}p_\theta(x)$. Writing $\dot{p}_0 := \dot{p}_{\theta_0}$, we can calculate that

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\mathbb{U}_\lambda(\epsilon,\theta)\bigg|_{\epsilon=0,\theta=\theta_0} &= \int \left[p_0(u) - h_{x,\lambda}(u)\right] s_0(u) dP_0(u) \;=\; \int \dot{p}_0(u)dP_0(u) - \int \dot{p}_0(u)dH_{x,\lambda}(u) \\
\frac{\partial}{\partial\theta}\mathbb{U}_\lambda(\epsilon,\theta)\bigg|_{\epsilon=0,\theta=\theta_0} &= \int \left[\dot{p}_0(u)\right]^2 d\nu(u)\ .
\end{aligned}$$

Using the Implicit Function Theorem as before, we have that

$$\begin{aligned}
\frac{d}{d\epsilon}\theta^*_{\epsilon,\lambda}\bigg|_{\epsilon=0} &= -\left[\frac{\partial}{\partial\theta}\mathbb{U}_\lambda(\epsilon,\theta)\bigg|_{\epsilon=0,\theta=\theta_0}\right]^{-1} \frac{\partial}{\partial\epsilon}\mathbb{U}_\lambda(\epsilon,\theta)\bigg|_{\epsilon=0,\theta=\theta_0} \\
&= \frac{\int \dot{p}_0(u)dH_{x,\lambda}(u) - \int \dot{p}_0(u)dP_0(u)}{\int \left[\dot{p}_0(u)\right]^2 d\nu(u)} \;\longrightarrow\; \frac{\dot{p}_0(x) - \int \dot{p}_0(u)dP_0(u)}{\int \left[\dot{p}_0(u)\right]^2 d\nu(u)}
\end{aligned}$$

as λ tends to zero. In general, this limit does not coincide with the true value $\phi_P(x)$ exhibited above, and thus, we find that representation (2.3) fails.

Defining $\dot{s}_\theta(x) := \frac{\partial}{\partial\theta}s_\theta(x)$ and setting $\dot{s}_0 := \dot{s}_{\theta_0}$, we note that

$$\begin{aligned}
\int s_{\theta^*_{\epsilon,\lambda}}(u)dP_{\epsilon,\lambda}(u) &= \int s_{\theta^*_{\epsilon,\lambda}}(u)dP_{\epsilon,\lambda}(u) - \int s_0(u)dP_0(u) \\
&= \int s_0(u)(P_{\epsilon,\lambda} - P)(du) + \int \left[s_{\theta^*_{\epsilon,\lambda}}(u) - s_0(u)\right] dP(u) + \int \left[s_{\theta^*_{\epsilon,\lambda}}(u) - s_0(u)\right](P_{\epsilon,\lambda} - P)(du) \\
&= \epsilon \int s_0(u)dH_{x,\lambda}(u) + (\theta^*_{\epsilon,\lambda} - \theta_0)\int \dot{s}_0(u)dP(u) + o(\epsilon)\ ,
\end{aligned}$$

from which it follows that

$$\frac{\int \phi^*_{\epsilon,\lambda}(u)dP_{\epsilon,\lambda}(u)}{\epsilon} \;=\; -I(\theta^*_{\epsilon,\lambda})^{-1}\frac{1}{\epsilon}\int s_{\theta^*_{\epsilon,\lambda}}(u)dP_{\epsilon,\lambda}(u)$$

$$\to \ -I(\theta_0)^{-1}\left\{\int s_0(u)dH_{x,\lambda}(u) + \int \dot{s}_0(u)dP_0(u)\left[\frac{\int \dot{p}_0(u)dH_{x,\lambda}(u) - \int \dot{p}_0(u)dP_0(u)}{\int \left[\dot{p}_0(u)\right]^2 d\nu(u)}\right]\right\}$$

as ϵ tends to zero. Clearly, the latter quantity is not generally zero, which implies that condition (A1) fails to hold. This observations explains why representation (2.3) does not hold here, and suggests that the divergence used in this example is inappropriate and should not generally be used.

References

Leandro Pardo. *Statistical inference based on divergence measures.* CRC press, 2005.