Adversarial Inference is Efficient

By Tetsuya Kaji, Elena Manresa, and Guillaume A. Pouliot*

Online Appendix

We denote the empirical measure corresponding to X_i by $\mathbb{P}_n$, to X_i^θ by $\mathbb{P}_m^\theta$, and to Z by $\tilde{\mathbb{P}}_m$. Recall $p_\theta : \mathbb{R}^d \to \mathbb{R}$ is a positive density function over $X_i \in \mathbb{R}^d$. Let $p = p_{\theta_0}$. The population D_θ^* is then $p/(p+p_\theta)$. The infeasible objective function for θ is $\mathbb{P}_n \log \frac{p}{p+p_\theta} + \tilde{\mathbb{P}}_m \log \frac{p_\theta}{p+p_\theta} \circ G_\theta$, where $G_\theta(Z) = G(\theta, Z)$ as defined in the text. Assume p_θ and G_θ are twice continuously differentiable in both arguments, where θ is a $p \times 1$ column vector and the value of G_θ is a $d \times 1$ column vector. We denote $\dot{p}_\theta = dp_\theta/d\theta$ (a $p \times 1$ column vector), $p'_\theta = dp_\theta/dx$ (a $d \times 1$ column vector), $\dot{p}'_\theta = dp_\theta/(d\theta dx)$ (a $p \times d$ matrix), $\ddot{p}_\theta = dp_\theta/(d\theta d\theta)$ (a $p \times p$ matrix), $p''_\theta = dp_\theta/(dxdx)$ (a $d \times d$ matrix), and $\dot{G}_\theta = dG_\theta/d\theta$ (a $d \times p$ matrix. The first-order derivative with respect to θ is

$$\mathbb{P}_n\biggl(-\frac{\dot{p}_\theta}{p+p_\theta}\biggr) + \tilde{\mathbb{P}}_m\biggl(\biggl[\frac{\dot{p}_\theta}{p_\theta} - \frac{\dot{p}_\theta}{p+p_\theta}\biggr] \circ G_\theta + \frac{p+p_\theta}{p_\theta} \circ G_\theta \cdot \dot{G}_\theta^\intercal \biggl[\frac{p_\theta}{p+p_\theta}\biggr]' \circ G_\theta\biggr).$$

At $\theta = \theta_0$ (so $p = p_\theta$), this equals

$$-\mathbb{P}_n\biggl(\frac{\dot{p}}{2p}\biggr) + \tilde{\mathbb{P}}_m\biggl(\frac{\dot{p}}{2p} \circ G\biggr) = -\frac{1}{2}(\mathbb{P}_n - \mathbb{P}_m)\biggl(\frac{\dot{p}}{p}\biggr).$$

Since $\bigl[\frac{p_\theta}{p+p_\theta}\bigr]' = 0$ at $\theta = \theta_0$, the second-order derivative at $\theta = \theta_0$ for population can be evaluated as

$$\begin{aligned} &P_0\biggl(-\frac{\ddot{p}_\theta}{p+p_\theta} + \frac{\dot{p}_\theta \dot{p}_\theta^\intercal}{(p+p_\theta)^2}\biggr)\bigg|_{\theta_0} + \tilde{P}_0\biggl(\biggl[\frac{\ddot{p}_\theta}{p_\theta} - \frac{\dot{p}_\theta \dot{p}_\theta^\intercal}{p_\theta^2} - \frac{\ddot{p}_\theta}{p+p_\theta} + \frac{\dot{p}_\theta \dot{p}_\theta^\intercal}{(p+p_\theta)^2}\biggr] \circ G_\theta \\ &+ \biggl(\biggl[\frac{\dot{p}_\theta}{p_\theta} - \frac{\dot{p}_\theta}{p+p_\theta}\biggr]' \circ G_\theta\biggr)\dot{G}_\theta + \frac{p+p_\theta}{p_\theta} \circ G_\theta \cdot \dot{G}_\theta^\intercal \biggl[\frac{\dot{p}_\theta^\intercal}{p+p_\theta} - \frac{p_\theta \dot{p}_\theta^\intercal}{(p+p_\theta)^2}\biggr]' \circ G_\theta\biggr)\bigg|_{\theta_0}. \end{aligned}$$

Note that at $\theta = \theta_0$,

$$\begin{aligned} &\biggl(\biggl[\frac{\dot{p}_\theta}{p_\theta} - \frac{\dot{p}_\theta}{p+p_\theta}\biggr]' \circ G_\theta\biggr)\dot{G}_\theta + \frac{p+p_\theta}{p_\theta} \circ G_\theta \cdot \dot{G}_\theta^\intercal \biggl[\frac{\dot{p}_\theta^\intercal}{p+p_\theta} - \frac{p_\theta \dot{p}_\theta^\intercal}{(p+p_\theta)^2}\biggr]' \circ G_\theta \\ &= \biggl(\biggl[\frac{\dot{p}}{2p}\biggr]' \circ G_\theta\biggr)\dot{G}_\theta + \dot{G}_\theta^\intercal \biggl(\biggl[\frac{\dot{p}}{2p}\biggr]' \circ G_\theta\biggr)^\intercal. \end{aligned}$$

Moreover,

$$\tilde{P}_0\biggl(\biggl[\frac{\dot{p}}{p}\biggr]' \circ G_\theta \cdot \dot{G}_\theta\biggr)\bigg|_{\theta_0} = \tilde{P}_0\biggl(\biggl[\frac{\dot{p}}{p} \circ G_\theta\biggr]^{\boldsymbol{\cdot}}_{\theta_0}\biggr) = \biggl[P_\theta\biggl(\frac{\dot{p}}{p}\biggr)\biggr]^{\boldsymbol{\cdot}}_{\theta_0} = \int \biggl(\frac{\dot{p}}{p}\biggr)\dot{p}_\theta^\intercal\bigg|_{\theta_0} = P_0\biggl(\frac{\dot{p}\dot{p}^\intercal}{p^2}\biggr)$$

and the same applies to the other term.

* Kaji: Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA, tkaji@chicagobooth.edu. Manresa: Department of Economics, New York University, 19 West 4th Street, 6 Floor New York, NY, 10012 USA, elena.manresa@nyu.edu. Pouliot: Harris School of Public Policy, University of Chicago, 1155 E 60th St, Chicago, IL 60637, USA, guillaumepouliot@uchicago.edu.

Hence, the second derivative is equal to half the information matrix

$$-\frac{1}{2}P_0\left(\frac{\dot{p}\dot{p}^\intercal}{p^2}\right)+P_0\left(\frac{\dot{p}\dot{p}^\intercal}{p^2}\right)=\frac{1}{2}P_0\left(\frac{\dot{p}\dot{p}^\intercal}{p^2}\right),$$

which shows efficiency of $\hat{\theta}_{n,m}$. Note that, if $n/m \to 0$, the first-order derivative is half the score, so twice the objective function yields the unscaled score and information.