Invariance under reparametrization.

Moulinath Banerjee

University of Michigan

January 17, 2003

What words may have failed to convey, I hope this will!!

We have a regular parametric model

$$\mathcal{P} \equiv \{P_\theta : \theta \in \Theta\}$$

where Θ is an open subset of k–dimensional Euclidean space. Now, $\theta \mapsto P_\theta$ is *a* parametrization of the model $\mathcal{P}$, but by no means is it the *only* parametrization. A parametrization is just some association of the probability distributions in $\mathcal{P}$ with some subset of a finite–dimensional space. Adequate smoothness assumptions in the parameter θ allow us to do meaningful mathematical computations.

Consider a finite–dimensional functional defined on $\mathcal{P}$. Thus we consider, $\nu : \mathcal{P} \to \mathbb{R}^m$ (with $m \leq k$). Because of the one–one association between θ and P_θ we can write $\nu(P_\theta)$ as a function of θ; thus,

$$\nu(P_\theta) \equiv q(\theta)\,.$$

Consider a point $P \in \mathcal{P}$, which is P_θ for some θ in Θ. We define the information bound and the efficient influence function for the estimation of ν at the point $P \equiv P_\theta$ as follows:

$$I^{-1}\,(P \;\mid\; \nu, \mathcal{P}) = \dot{q}^T(\theta)\, I^{-1}(\theta)\, \dot{q}(\theta)$$

and

$$\tilde{l}(\cdot, P \;\mid\; \nu, \mathcal{P}) = \dot{q}^T(\theta)\, I^{-1}(\theta)\, \dot{l}(\cdot, \theta)\,.$$

Here $\dot{q}(\theta)_{k\times m}$ is the (Frechet) derivative of q at the point θ, and

$$\dot{l}(x,\theta) = \frac{\partial}{\partial\,\theta}\log\, f(x,\theta)$$

is the usual score function for θ and

$$I(\theta) = E_\theta\left[\dot{l}(X,\theta)\,\dot{l}^T(X,\theta)\right]$$

is the dispersion matrix of the score function – the information matrix. We will show that the information bound for ν at the point P and the efficient influence function *do not depend* on the parametrization. They only depend on the underlying probability measure P.

To this end, consider a different parametrization of $\mathcal{P} = \{Q_\gamma : \gamma \in \Gamma\}$, where $Q_\gamma = P_{\psi(\gamma)}$. Here, Γ is an open subset of $\mathbb{R}^k$ and ψ is a bijection from Γ to Θ that is continuously differentiable with non–singular derivative $\dot{\psi}(\gamma)_{k\times k}$. If we now consider a fixed probability measure P in $\mathcal{P}$ with $P \equiv P_\theta$, where $\theta = \psi(\gamma)$, then

$$P = Q_\gamma = P_{\psi(\gamma)}\,.$$

Also, note that the functional ν can be written as a function of γ, since,

$$\nu(Q_\gamma) = \nu(P_{\psi(\gamma)}) = q(\psi(\gamma)) \equiv r(\gamma)\,.$$

The densities corresponding to the family $\mathcal{P}$ can be written as $\{g(x,\gamma) : \gamma \in \Gamma\}$ where

$$g(x,\gamma) \equiv \frac{d\,Q_\gamma}{d\,\mu} = \frac{d\,P_{\psi(\gamma)}}{d\,\mu} = f(x,\psi(\gamma)) \equiv f(x,\theta)\,.$$

The score function using the parametrization γ is,

$$\dot{l}(x,\gamma) = \frac{\partial}{\partial\,\gamma}\,\log\,g(x,\gamma) \equiv \frac{\partial}{\partial\,\gamma}\,\log\,f(x,\psi(\gamma))\,.$$

Using the chain rule it follows immediately that,

$$\dot{l}(x,\gamma)_{k\times 1} \equiv \dot{\psi}(\gamma)^T_{k\times k}\,\dot{l}(x,\psi(\gamma))_{k\times 1} \equiv \dot{\psi}(\gamma)^T\,\dot{l}(x,\theta)\,. \quad (1)$$

Consequently, the information matrix at parameter value γ is

$$\tilde{I}(\gamma) = \mathrm{Disp}(\dot{l}(X,\gamma)) = \mathrm{Disp}(\dot{\psi}(\gamma)^T\,\dot{l}(X,\theta)) = \dot{\psi}(\gamma)^T\,I(\theta)\,\dot{\psi}(\gamma)\,. \quad (2)$$

Now, in terms of the parametrization using Γ, the information bound and the influence function for the estimation of ν at the point P is,

$$I^{-1}\,(P \mid \nu,\mathcal{P}) = \dot{r}^T(\gamma)\,\tilde{I}^{-1}(\gamma)\,\dot{r}(\gamma)$$

and

$$\tilde{l}(\cdot,P \mid \nu,\mathcal{P}) = \dot{r}^T(\gamma)\,\tilde{I}^{-1}(\gamma)\,\dot{l}(\cdot,\gamma)\,.$$

Now, using the chain rule once again,

$$\dot{r}(\gamma) = \frac{\partial}{\partial\,\gamma}\,q(\psi(\gamma)) = \dot{\psi}(\gamma)^T\,\dot{q}(\psi(\gamma)) \equiv \dot{\psi}(\gamma)^T\,\dot{q}(\theta)\,. \quad (3)$$

Now, on using (2) and (3)

$$\dot{r}^T(\gamma)\,\tilde{I}^{-1}(\gamma)\,\dot{r}(\gamma) = \dot{q}(\theta)^T\,\dot{\psi}(\gamma)\,(\dot{\psi}(\gamma))^{-1}\,I(\theta)^{-1}\,(\dot{\psi}(\gamma)^T)^{-1}\,\dot{\psi}(\gamma)^T\,\dot{q}(\theta) = \dot{q}^T(\theta)\,I^{-1}(\theta)\,\dot{q}(\theta)\,.$$

This shows that the information bound for the estimation of ν under probability measure $P \equiv P_\theta \equiv Q_\gamma$ is indeed invariant under the parametrization. To show the invariance of the efficient influence function, use (1), (2) and (3) to get that,

$$\dot{r}^T(\gamma)\,\tilde{I}^{-1}(\gamma)\,\dot{l}(x,\gamma) = \dot{q}(\theta)^T\,\dot{\psi}(\gamma)\,(\dot{\psi}(\gamma))^{-1}\,I(\theta)^{-1}\,(\dot{\psi}(\gamma)^T)^{-1}\,\dot{\psi}(\gamma)^T\,\dot{l}(x,\theta) = \dot{q}(\theta)^T\,I(\theta)^{-1}\,\dot{l}(x,\theta)\,.$$

This establishes the claim.