557: MATHEMATICAL STATISTICS II
THE FISHER INFORMATION

A random sample of size n is available from a distribution with pdf $f_0(x)$. We seek to use a parametric "approximating" model $f_X(x;\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}^k$, defined in the family $\mathcal{F}_\theta$. We define the "true" value of θ, denoted θ_0, using a minimum Kullback-Leibler divergence criterion:

$$\theta_0 = \arg\min_\theta KL(f_0, f_X(.;\theta)) = \arg\min_\theta \left\{ \mathbb{E}_{f_0}\left[\log\left(\frac{f_0(X)}{f_X(X;\theta)} \right) \right] \right\}$$

which, after expansion, equates to

$$\theta_0 = \arg\max_\theta \{ \mathbb{E}_{f_0}[\log f_X(X;\theta)] \} = \arg\max_\theta \left\{ \int \log f_X(x;\theta) f_0(x)\, dx \right\}.$$

We seek to solve this problem using calculus: we solve

$$\frac{\partial}{\partial\theta}\{ \mathbb{E}_{f_0}[\log f_X(X;\theta)] \} = 0_k$$

which, under regularity conditions that allow the interchange of integration and differentiation, may be re-written

$$\mathbb{E}_{f_0}[U(X;\theta)] \equiv \mathbb{E}_{f_0}\left[\frac{\partial}{\partial\theta}\{\log f_X(X;\theta)\} \right] = 0_k.$$

Thus at the solution $\theta = \theta_0$, we have by construction that

$$\mathbb{E}_{f_0}\left[\frac{\partial}{\partial\theta}\{\log f_X(X;\theta)\}_{\theta=\theta_0} \right] = 0_k.$$

To verify that the solution corresponds to a maximum, we inspect the second derivative matrix if it is available: again, passing the derivatives through the integral, we obtain the $(k \times k)$ Hessian matrix

$$\mathbb{E}_{f_0}\left[\frac{\partial^2}{\partial\theta\partial\theta^\top}\{\log f_X(X;\theta)\} \right] = \mathbb{E}_{f_0}[\Psi(X;\theta)]$$

say, which we require to be **negative definite**; alternatively, we require that

$$\mathcal{I}_{f_0}(\theta) = \mathbb{E}_{f_0}[-\Psi(X;\theta)]$$

be **positive definite**.

Correct specification: If $f_0 \in \mathcal{F}_\theta$, that is, the true f_0 is a member of the parametric family, then there is, under the assumption of identifiability, a unique θ_0 such that $f_0(x) \equiv f_X(x;\theta_0)$. In this case, we write

$$\mathbb{E}_X\left[\frac{\partial}{\partial\theta}\{\log f_X(X;\theta)\}_{\theta=\theta_0}; \theta_0 \right] = 0_k.$$

and

$$\mathcal{I}_{\theta_0}(\theta) = \mathbb{E}_X[-\Psi(X;\theta);\theta_0]$$

The quantity $\mathcal{I}_{\theta_0}(\theta)$ is termed the **Fisher Information**, which is most commonly computed at $\theta = \theta_0$.

An alternative representation of the second partial derivative matrix is available under correct specification: we have that

$$\frac{\partial^2}{\partial\theta\partial\theta^\top}\{\log f_X(x;\theta)\} = \frac{\partial}{\partial\theta}\left\{ \frac{\partial}{\partial\theta^\top}\{\log f_X(x;\theta)\} \right\} = \frac{\partial}{\partial\theta}\left\{ \frac{\dfrac{\partial f_X(x;\theta)}{\partial\theta^\top}}{f_X(x;\theta)} \right\} = \frac{\partial}{\partial\theta}\left\{ \frac{\{\dot{f}_X(x;\theta)\}^\top}{f_X(x;\theta)} \right\}$$

say, where

$$\{\dot{f}_X(x;\theta)\}^\top = \frac{\partial f_X(x;\theta)}{\partial \theta^\top} = \left\{\frac{\partial f_X(x;\theta)}{\partial \theta}\right\}^\top.$$

Writing $\ddot{f}_X(x;\theta)$ for the $(k \times k)$ second partial derivative (Hessian) matrix

$$\ddot{f}_X(x;\theta) = \frac{\partial^2 f_X(x;\theta)}{\partial \theta \partial \theta^\top}$$

we have that

$$\frac{\partial}{\partial \theta}\left\{\frac{\{\dot{f}_X(x;\theta)\}^\top}{f_X(x;\theta)}\right\} = \frac{\ddot{f}_X(x;\theta)}{f_X(x;\theta)} - \left\{\frac{\dot{f}_X(x;\theta)}{f_X(x;\theta)}\right\}\left\{\frac{\dot{f}_X(x;\theta)}{f_X(x;\theta)}\right\}^\top$$

But as before

$$\left\{\frac{\dot{f}_X(x;\theta)}{f_X(x;\theta)}\right\} = \frac{\partial}{\partial \theta}\{\log f_X(x;\theta)\} = U(x;\theta)$$

so therefore

$$\Psi(x;\theta) = \frac{\ddot{f}_X(x;\theta)}{f_X(x;\theta)} - U(x;\theta)\{U(x;\theta)\}^\top$$

Moving to the random variable version, and taking expectations with respect to the data generating distribution f_0 yields the identity

$$\mathbb{E}_{f_0}[\Psi(X;\theta)] = \mathbb{E}_{f_0}\left[\frac{\ddot{f}_X(X;\theta)}{f_X(X;\theta)}\right] - \mathbb{E}_{f_0}\left[U(X;\theta)\{U(X;\theta)\}^\top\right]$$

Now for the first term on the right hand side, we have

$$\mathbb{E}_{f_0}\left[\frac{\ddot{f}_X(X;\theta)}{f_X(X;\theta)}\right] = \int \frac{\ddot{f}_X(x;\theta)}{f_X(x;\theta)} f_0(x)dx$$

Under correct specification, $f_0(x) \equiv f_X(x;\theta_0)$, and so in this case we have that

$$\mathbb{E}_X\left[\frac{\ddot{f}_X(X;\theta)}{f_X(X;\theta)};\theta_0\right] = \int \frac{\ddot{f}_X(x;\theta)}{f_X(x;\theta)} f_X(x;\theta_0)dx$$

and when evaluating at $\theta = \theta_0$ we have that the integral is

$$\int \frac{\ddot{f}_X(x;\theta_0)}{f_X(x;\theta_0)} f_X(x;\theta_0)dx = \int \ddot{f}_X(x;\theta_0)dx = \mathbf{0}_{k\times k}$$

as under regularity conditions

$$\int \ddot{f}_X(x;\theta_0)dx = \int \frac{\partial^2}{\partial\theta\partial\theta^\top}\{f_X(x;\theta)\}_{\theta=\theta_0} dx = \frac{\partial^2}{\partial\theta\partial\theta^\top}\left\{\int f_X(x;\theta)dx\right\}_{\theta=\theta_0} = \mathbf{0}_{k\times k}$$

Therefore, in this case

$$\mathcal{I}_{\theta_0}(\theta_0) = \mathbb{E}_X[-\Psi(X;\theta_0);\theta_0] = \mathbb{E}_X\left[U(X;\theta_0)\{U(X;\theta_0)\}^\top;\theta_0\right].$$

Also,

$$\mathbb{E}_X\left[U(X;\theta_0)\{U(X;\theta_0)\}^\top;\theta_0\right] = \mathrm{Var}_X[U(X;\theta_0);\theta_0]$$

– the $(k \times k)$ covariance matrix of the random vector $U(X;\theta_0)$ – as $\mathbb{E}_X[U(X;\theta_0);\theta_0] = \mathbf{0}_k$.