Score Tests

Consider a statistical problem involving an unknown parameter $\theta \in \mathbb{R}^p$. Denote the log-likelihood $\ell(\theta)$, its gradient (vector of first derivatives) by $U(\theta)$, and its Hessian (matrix of second derivatives) by $J(\theta)$. Let θ^* denote the true value of θ. In the development below, all expectations and variances are computed under the distribution resulting from θ^*. Recall that the Fisher information matrix is defined by $\mathcal{I}(\theta) = -E[J(\theta)]$. In typical settings, we have the following results:

$$E[U(\theta^*)] = 0, \tag{1}$$

$$\mathrm{Cov}(U(\theta^*)] = \mathcal{I}(\theta^*), \tag{2}$$

$$U(\theta^*) \dot{\sim} N(0, \mathcal{I}(\theta^*)), \tag{3}$$

$$\mathcal{I}(\theta^*)^{-1} J(\theta^*) \xrightarrow{P} I, \tag{4}$$

$$U(\theta^*) \dot{\sim} N(0, J(\theta^*)). \tag{5}$$

Suppose we want to test the hypothesis $H_0 : \theta = \theta_0$ for some special value θ_0. Define

$$S(\theta) = U(\theta)^T J(\theta)^{-1} U(\theta). \tag{6}$$

From equation (5) above we get that

$$S(\theta_0) \dot{\sim} \chi^2_p \quad \text{under } H_0. \tag{7}$$

We can thus use $S(\theta_0)$ as a test statistic for testing H_0. This test is called the "score test."

We can extend the development to testing a subset of the elements of θ while estimating the others. Let us partition θ as $\theta = (\phi, \psi)$, where $\phi \in \mathbb{R}^q$ is the vector of parameters of interest and $\psi \in \mathbb{R}^{p-q}$ is the vector of remaining parameters (so-called "nuisance parameters"). Partition U correspondingly as $U(\phi, \psi) = [U_\phi(\phi, \psi) \;\; U_\psi(\phi, \psi)]$, partition $\mathcal{I}$ as

$$\mathcal{I}(\phi, \psi) = \begin{bmatrix} \mathcal{I}_{\phi\phi}(\phi, \psi) & \mathcal{I}_{\phi\psi}(\phi, \psi) \\ \mathcal{I}_{\psi\phi}(\phi, \psi) & \mathcal{I}_{\psi\psi}(\phi, \psi) \end{bmatrix},$$

and partition J similarly. Note that $\mathcal{I}_{\psi\phi} = \mathcal{I}_{\phi\psi}^T$ and similarly for J. Denote the true values of ϕ and ψ by ϕ^* and ψ^*.

Denote by $\hat{\psi}(\phi)$ the MLE of ψ with ϕ held fixed at the specified value. We have that $\hat{\psi}(\phi)$ is the solution to the equation

$$U_\psi(\phi, \hat{\psi}(\phi)) = 0. \tag{8}$$

By the same kind of standard Taylor expansion argument as is used to develop the asymptotic properties of the MLE in general, we have

$$\hat{\psi}(\phi^*) - \psi^* \doteq -J_{\psi\psi}(\phi^*, \psi^*)^{-1} U_\psi(\phi^*, \psi^*). \tag{9}$$

Now, suppose we want to test $H_0 : \phi = \phi_0$. We can build a test statistic based on $U_\phi(\phi_0, \hat{\psi}(\phi_0))$. We need to work out the asymptotic distribution of this quantity under H_0. Assume for now that H_0 holds. We then have

$$\begin{aligned} U_\phi(\phi_0, \hat{\psi}(\phi_0)) &\doteq U_\phi(\phi_0, \psi^*) + J_{\phi\psi}(\psi_0, \psi^*)(\hat{\psi}(\phi_0) - \psi^*) \\ &\doteq U_\phi(\phi_0, \psi^*) - \mathcal{I}_{\phi\psi}(\phi_0, \psi^*)\mathcal{I}_{\psi\psi}(\phi_0, \psi^*)^{-1} U_\psi(\phi_0, \psi^*), \end{aligned} \tag{10}$$

where the first step is based on a Taylor expansion and the second step is based on (4) and (9). Using the results (1)-(3), we find that $U_\phi(\phi_0, \hat{\psi}(\phi_0))$ is asymptotically mean-zero multivariate normal with covariance matrix given as follows (where U_ϕ, U_ψ, and the various blocks of the matrix $\mathcal{I}$ are all evaluated at (ϕ_0, ψ^*), but we omit these arguments for brevity):

$$\begin{aligned} \Omega &= \mathrm{Cov}(U_\phi, U_\phi) - \mathrm{Cov}(U_\phi, U_\psi)\mathcal{I}_{\psi\psi}^{-1}\mathcal{I}_{\phi\psi}^T - \mathcal{I}_{\phi\psi}\mathcal{I}_{\psi\psi}^{-1}\mathrm{Cov}(U_\psi, U_\phi) + \mathcal{I}_{\phi\psi}\mathcal{I}_{\psi\psi}^{-1}\mathrm{Cov}(U_\psi, U_\psi)\mathcal{I}_{\psi\psi}^{-1}\mathcal{I}_{\phi\psi}^T \\ &= \mathcal{I}_{\phi\phi} - \mathcal{I}_{\phi\psi}\,\mathcal{I}_{\psi\psi}^{-1}\,\mathcal{I}_{\phi\psi}^T - \mathcal{I}_{\phi\psi}\,\mathcal{I}_{\psi\psi}^{-1}\,\mathcal{I}_{\psi\phi} + \mathcal{I}_{\phi\psi}\,\mathcal{I}_{\psi\psi}^{-1}\,\mathcal{I}_{\phi\psi}^T\,\mathcal{I}_{\phi\psi}\,\mathcal{I}_{\psi\psi}^{-1}\,\mathcal{I}_{\psi\psi}\,\mathcal{I}_{\psi\psi}^{-1}\,\mathcal{I}_{\phi\psi}^T \\ &= \mathcal{I}_{\phi\phi} - \mathcal{I}_{\phi\psi}\,\mathcal{I}_{\psi\psi}^{-1}\,\mathcal{I}_{\psi\phi}. \end{aligned} \tag{11}$$

We can form an estimate $\hat{\Omega}(\phi_0)$ of Ω by replacing $\mathcal{I}(\phi_0, \psi^*)$ with $J(\phi_0, \hat{\psi}(\phi_0))$ throughout. We then get the score statistic

$$S(\phi_0) = U_\phi(\phi_0, \hat{\psi}(\phi_0))^T \hat{\Omega}(\phi_0)^{-1} U_\phi(\phi_0, \hat{\psi}(\phi_0)), \tag{12}$$

for which we have

$$S(\phi_0) \mathrel{\dot\sim} \chi_q^2 \quad \text{under } H_0. \tag{13}$$

Remark: By the formula for the inverse of a partitioned matrix, the matrix $\hat{\Omega}(\phi_0)^{-1}$ is equal to the $\phi\phi$ block of $J(\phi_0, \hat{\psi}(\phi_0))^{-1}$.