9.07 Introduction to Statistics for Brain and Cognitive Sciences
Emery N. Brown

Lecture 9 Addendum 1: The Observed Fisher Information and Fisher Information for the Binomial Model

Recall that $k = \sum_{i=1}^{n} x_i$ is the number of correct responses in the binomial model. The log likelihood function for this model is

$$\log L(p) \propto k \log p (n-k) \log(1-p) . \tag{A.1}$$

The score function is

$$\frac{\partial \log L(p)}{\partial p} = \frac{k}{p} - \frac{n-k}{(1-p)} \tag{A.2}$$

and the second derivative

$$\frac{\partial^2 \log L(p)}{\partial p^2} = -[\frac{k}{p^2} + \frac{n-k}{(1-p)^2}] . \tag{A.3}$$

We defined the **observed Fisher information** as $\left. \frac{-\partial^2 \log L(p)}{\partial p^2} \right|_{\hat{p}_{ML}}$ If we compute the ML estimate as $\hat{p}_{ML} = \frac{k}{n}$, then we get

$$\begin{aligned} I(\hat{p}_{ML}) &= -[\frac{\partial^2 \log L(p)}{\partial p^2}] = \left. [\frac{k}{p^2} + \frac{n-k}{(1-p)^2}] \right|_{p=\hat{p}_{ML}} \\ &= \frac{n}{n}[\frac{k}{\hat{p}_{ML}{}^2} + \frac{n-k}{(1-\hat{p}_{ML})^2}] \\ &= n[\frac{\hat{p}_{ML}}{\hat{p}_{ML}{}^2} + \frac{(1-\hat{p}_{ML})}{(1-\hat{p}_{ML})^2}] \\ &= n[\frac{1}{\hat{p}_{ML}} + \frac{1}{(1-\hat{p}_{ML})}] \\ &= n[\hat{p}_{ML}(1-\hat{p}_{ML})]^{-1} \end{aligned} \tag{A.4}$$

Recall that $E(k) = np$. Therefore from Eqs. 9.38 and 9.39, the **Fisher information** for this model can be computed two ways

$$\begin{aligned} I(p) &= -E[\frac{\partial^2 \log L(p)}{\partial p^2}] = E[\frac{k}{p^2} + \frac{n-k}{(1-p)^2}] \\ &= [\frac{np}{p^2} + \frac{n-np}{(1-p)^2}] \\ &= n[\frac{1}{p} + \frac{1}{(1-p)^2}] \\ &= n[p(1-p)]^{-1} \end{aligned} \tag{A.5}$$

The second way the Fisher information can also be computed is as the expectation of the square of the score function. To see this we take

$$\begin{aligned} I(p) = E[\frac{\partial \log L(p)}{\partial p}]^2 &= E[\frac{k}{p} - \frac{n-k}{(1-p)}]^2 \\ &= E[\frac{k - kp - np + kp}{p(1-p)}]^2 \\ &= E[\frac{(k-np)^2}{p(1-p)^2}] \\ &= \frac{Var(k)}{[p(1-p)]^2} = \frac{np(1-p)}{[p(1-p)]^2} = \frac{n}{p(1-p)} \end{aligned} \tag{A.6}$$

To estimate the Fisher information we evaluate Eqs. A.5 and A.6 at $p = \hat{p}_{ML}$. Plugging $p = \hat{p}_{ML}$ into either Eq. A.5 or A.6 gives the same estimate Eq. A.4. This is in general not the case because the expectations in Eqs. A.5 and A.6 can be difficult to compute. To summarize, the observed Fisher information is

$$I(\hat{p}_{ML}) = -[\frac{\partial^2 \log L(p)}{\partial p^2}]\Bigg|_{\hat{p}_{ML}} = \frac{n}{\hat{p}_{ML}(1-\hat{p}_{ML})}. \tag{A.7}$$

and it is the estimate of the Fisher information which can be computed two ways as

$$-E[\frac{\partial^2 \log f(x \mid p)}{\partial p^2}] = E[\frac{\log f(x \mid p)}{\partial p}]^2 = \frac{n}{p(1-p)}. \tag{A.8}$$

Remember $L(p) = f(x \mid p)$. The general results (Eq. 9.38 and 9.39) are

$$\begin{aligned} I(\theta) &= -E[\frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2}] \\ &= -\int (\frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2}) f(x \mid \theta) dx \\ &= E[(\frac{\partial \log f(x \mid \theta)}{\partial \theta})^2] \\ &= \int (\frac{\partial \log f(x \mid \theta)}{\partial \theta})^2 f(x \mid \theta) dx. \end{aligned} \tag{A.9}$$

The estimate of the Fisher information is the observed Fisher information defined as

$$I(\hat{\theta}_{ML}) = -[\frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2}]\Bigg|_{\theta = \hat{\theta}_{ML}}. \tag{A.10}$$