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

Lecture 9 Addendum 2: Properties of the Maximum Likelihood Estimates

Remark 9.4. If $\hat{\theta}_{ML}$ is the maximum likelihood estimate of $\theta,$ then $\hat{\theta}_{ML} \xrightarrow{p} \theta$ (in the sense of the Weak Law of Large Numbers) as n goes to infinity. We can show this for elementary ML estimates such as the sample mean by the $LLN.$ For a more general result see Theorem A, pp. 275-276 of Rice. This property is called **consistency**.

Remark 9.5. If $\hat{\theta}_{ML}$ is the maximum likelihood estimate of $\theta,$ then the large sample (asymptotic) distribution of $\hat{\theta}_{ML}$ is

$$\hat{\theta}_{ML} \approx N(\theta, I(\theta)^{-1})$$

(see Theorem B for the 2-dimensional case, Rice, pp. 277-279). In general, we can construct 95% confidence intervals for θ as

$$\hat{\theta}_{ML,i} \pm 2I(\hat{\theta}_{ML})_{ii}^{-\frac{1}{2}}$$

where $\hat{\theta}_{ML} = (\hat{\theta}_{ML,1}, \ldots, \hat{\theta}_{ML,d})$ and $I(\hat{\theta}_{ML})_{ii}$ is the i^{th} diagonal element of $I(\theta)$ evaluated at $\hat{\theta}_{ML}$ and $\hat{\theta}_{ML,i}$ is i^{th} element of $\hat{\theta}_{ML}$ for $i = 1, \ldots, d.$

Remark 9.6. If $\hat{\theta}_{ML}$ is the maximum likelihood estimate of $\theta,$ then the maximum likelihood estimate of $h(\theta)$ is $h(\hat{\theta}_{ML})$ and

$$h(\hat{\theta}_{ML}) \approx N(h(\theta), h'(\theta)^2 I(\theta)^{-1})$$

and the approximate 95% confidence interval is

$$h(\hat{\theta}_{ML}) \pm 2h'(\hat{\theta}_{ML}) I(\hat{\theta}_{ML})^{-\frac{1}{2}}.$$

Proof (Heuristic): Taking the first two terms of the Taylor series of $h(\hat{\theta}_{ML})$ about $h(\theta),$ which we can do because of the consistency of $\hat{\theta}_{ML}$ (**Remark 9.4**), we have that

$$h(\hat{\theta}_{ML}) \approx h(\theta) + h'(\theta)(\hat{\theta}_{ML} - \theta).$$

Thus,

$$E(h(\hat{\theta}_{ML})) \approx h(\theta) + h'(\theta)E(\hat{\theta}_{ML} - \theta) = h(\theta) + 0 = h(\theta)$$

by consistency of $\hat{\theta}_{ML}$. Also,

$$\mathrm{Var}(h(\hat{\theta}_{ML})) \approx E[(h(\hat{\theta}_{ML}) - h(\theta)]^2 = E[h'(\theta)(\hat{\theta}_{ML} - \theta)]^2 = h'(\theta)^2 \mathrm{Var}(\hat{\theta}_{ML}) = h'(\theta)^2 I(\theta)^{-1}.$$

Example 9.1 (continued). The second derivatives of the log likelihood are computed as

$$\frac{\partial L(\alpha,\beta)}{\partial \beta} = \frac{n\alpha}{\beta} - \sum_{i=1}^{n} x_i$$

$$\frac{\partial L(\alpha,\beta)}{\partial \alpha} = -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + n\log\beta + \sum_{k=1}^{n} \log(x_k)$$

$$\frac{\partial^2 L(\alpha,\beta)}{\partial \alpha^2} = -n\left[\frac{\Gamma(\alpha)\Gamma''(\alpha) - \Gamma'(\alpha)^2}{\Gamma(\alpha)^2}\right]$$

$$\frac{\partial^2 L(\alpha,\beta)}{\partial \beta^2} = \frac{-n\alpha}{\beta^2}$$

$$\frac{\partial^2 L(\alpha,\beta)}{\partial \alpha \partial \beta} = \frac{\partial^2 L(\alpha,\beta)}{\partial \beta \partial \alpha} = \frac{n}{\beta}$$

Hence, the Fisher Information is

$$I(\alpha,\beta) = -E\begin{bmatrix} \dfrac{\partial^2 L(\alpha,\beta)}{\partial \alpha^2} & \dfrac{\partial^2 L(\alpha,\beta)}{\partial \alpha,\beta} \\ \dfrac{\partial^2 L(\alpha,\beta)}{\partial \beta \partial \alpha} & \dfrac{\partial^2 L(\alpha,\beta)}{\partial \beta^2} \end{bmatrix} = n\begin{bmatrix} \dfrac{\Gamma(\alpha)\Gamma''(\alpha) - \Gamma'(\alpha)^2}{\Gamma(\alpha)^2} & -\beta^{-1} \\ -\beta^{-1} & \alpha\beta^{-2} \end{bmatrix}.$$

The observed Fisher Information is

$$I(\alpha,\beta)\Big|_{\hat{\alpha}_{ML},\hat{\beta}_{ML}} = n\begin{bmatrix} \dfrac{\Gamma(\hat{\alpha}_{ML})\Gamma''(\hat{\alpha}_{ML})-\Gamma'(\alpha_{ML})^2}{\Gamma(\hat{\alpha}_{ML})^2} & -\hat{\beta}_{ML}^{-1} \\ -\hat{\beta}_{ML}^{-1} & \hat{\alpha}_{ML}\hat{\beta}_{ML}^{-2} \end{bmatrix}.$$

Hence, we have 95% confidence intervals

$$\hat{\alpha}_{ML} \pm 2I(\hat{\alpha}_{ML},\hat{\beta}_{ML})^{-\frac{1}{2}}$$

$$\hat{\beta}_{ML} \pm 2I(\hat{\alpha}_{ML},\hat{\beta}_{ML})^{-\frac{1}{2}}.$$

What is an approximate correlation coefficient between $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$?

Example 2.1 (continued). The 95% confidence interval for p is

$$\hat{p}_{ML} \pm 2[I(\hat{p}_{ML})]^{-\frac{1}{2}} = \hat{p}_{ML} \pm 2[\frac{\hat{p}_{ML}(1-\hat{p}_{ML})}{n}]^{\frac{1}{2}}.$$

Example 3.3 (continued). For a Poisson model, the ML estimate of λ is $\hat{\lambda}_{ML} = \bar{x}$. The Fisher Information is

$$I(\lambda) = -\frac{n}{\lambda}$$

and an approximate 95% confidence interval is

$$\hat{\lambda}_{ML} \pm 2(\frac{\hat{\lambda}_{ML}}{n})^{\frac{1}{2}}.$$

What is an approximate 95% confidence interval for $\Pr(X=0)$?

$$\Pr(x=0)=\frac{\lambda^0 e^{-\lambda}}{0!}=e^{-\lambda}$$

By the **Invariance Property** of the maximum likelihood estimate (**Lecture 9**, p. 8), the ML estimate is

$$e^{-\hat{\lambda}_{ML}}.$$

Hence, $h(\lambda)=e^{-\lambda}$ and $h'(\lambda)=-e^{-\lambda}$ and

$$\mathrm{Var}(e^{-\hat{\lambda}_{ML}})\approx(-e^{-\lambda})^2\cdot\frac{\lambda}{n}=\frac{\lambda}{n}e^{-2\lambda}$$

and the approximate 95% confidence interval is

$$e^{-\hat{\lambda}_{ML}}\pm 2e^{-\hat{\lambda}_{ML}}\left(\frac{\hat{\lambda}_{ML}}{n}\right)^{\frac{1}{2}}$$