Section 8, 10/21/19

Theorem A (page 275):

Under appropriate smoothness conditions on the density f, the mle from an i.i.d. sample is consistent.

Proof

Recall that for an i.i.d. sample, the log likelihood of n observations is

$$l(\theta) = \sum_{i=1}^{n} log f(X_i|\theta)$$

Given a set of data, we wish to find the parameter set θ that maximizes our average likelihood

$$\frac{1}{n} l(\theta) = \frac{1}{n} \sum_{i=1}^{n} log f(X_i|\theta)$$

As n goes to infinity, by the law of large numbers

$$\begin{aligned} \frac{1}{n} l(\theta) &\to E\big[log f(X|\theta)\big] \\ &= \int log f(x|\theta)\, f(x|\theta_0)\, dx \end{aligned}$$

since the distribution of observations approaches the true distribution. In order to maximize this asymptotic distribution, we take the derivative and set it to zero.

$$\frac{\partial}{\partial\theta} \int log f(x|\theta)\, f(x|\theta_0)\, dx = \int \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)}\, f(x|\theta_0)\, dx = 0$$

If we set $\theta = \theta_0$, then

$$\int \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)}\, f(x|\theta_0)\, dx \bigg|_{\theta=\theta_0} = \int \frac{\partial}{\partial\theta} f(x|\theta)\, dx \bigg|_{\theta=\theta_0} = \frac{\partial}{\partial\theta} \int f(x|\theta)\, dx \bigg|_{\theta=\theta_0} = \frac{\partial}{\partial\theta}(1) = 0$$

on the basis of appropriate smoothness. So θ_0 is a stationary point, hopefully a maximum. A more involved argument rigorously shows that for such large n, the θ that maximizes the log likelihood also maximizes the expected log likelihood.

Theorem B

Under smoothness conditions on f, the probability distribution of $\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0)$ approaches a standard normal distribution.

Proof (sketch)

We begin by constructing a first order Taylor expansion approximation of the log likelihood derivative at 0 (maximized).

$$0 = l'(\hat{\theta}) \approx l'(\theta_0) + (\hat{\theta} - \theta_0)l''(\theta_0)$$

As n goes to infinity, by the law of large numbers

$$\begin{aligned} \frac{1}{n}l''(\theta_0) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\,log\,f(x_i|\theta_0) \\ &\to E\left[\frac{\partial^2}{\partial\theta^2}\,log\,f(X|\theta_0)\right] = -I(\theta_0) \end{aligned}$$

This latter equivalence following from Lemma A, from last week. Thus by rearranging terms to the other side we have

$$\begin{aligned} 0 &\approx l'(\theta_0) + (\hat{\theta} - \theta_0)I(\theta_0) \\ (\hat{\theta} - \theta_0) &\approx \frac{l'(\theta_0)}{nI(\theta_0)} \end{aligned}$$

We wish to examine the mean and variance of the distribution of our centered estimate $(\hat{\theta} - \theta_0)$. Note that the denominator is simple a constant. Thus

$$E\big[(\hat{\theta} - \theta_0)\big] = \frac{1}{nI(\theta_0)}E\big[l'(\theta_0)\big] = \frac{1}{nI(\theta_0)}\sum_{i=1}^{n}E\left[\frac{\partial}{\partial\theta}logf(X_i|\theta_0)\right] = 0$$

By last week's results. Taking variance,

$$Var\big[(\hat{\theta} - \theta_0)\big] = \frac{1}{n^2I^2(\theta_0)}Var\big[l''(\theta_0)\big] = \frac{1}{n^2I^2(\theta_0)}\sum_{i=1}^{n}E\left[\frac{\partial}{\partial\theta}logf(X_i|\theta_0)\right]^2 = \frac{1}{n^2I^2(\theta_0)}nI(\theta_0) = \frac{1}{nI(\theta_0)}$$

Since $l'(\theta)$ is the sum of i.i.d. random variables, we central limit theorem may be applied. Thus $\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \sim N(0,1)$ and for large data, we know the estimate of θ approaches the truth and we know our confidence in its estimate.