STAT 450 Lecture 16

Reading for Today's Lecture:
Goals of Today's Lecture:

- Do MLE examples.
- Discuss parametriztion invariance of MLE.
- Introduce large sample theory for MLE.

Today's notes

Maximum Likelihood Estimation

To find an MLE we maximize L. This is a typical function maximization problem which we approach by setting the gradient of L equal to 0 and then checking to see that the root is a maximum, not a minimum or saddle point.

The Binomial Distribution

If X has a Binomial(n, θ) distribution then

$$\begin{aligned}
L(\theta) &= \binom{n}{X} \theta^X (1-\theta)^{n-X} \\
\ell(\theta) &= \log \binom{n}{X} + X \log(\theta) + (n-X)\log(1-\theta) \\
U(\theta) &= \frac{X}{\theta} - \frac{n-X}{1-\theta}
\end{aligned}$$

The function L is 0 at $\theta = 0$ and at $\theta = 1$ unless $X = 0$ or $X = n$ so for $1 \le X \le n$ the MLE must be found by setting $U = 0$ and getting

$$\hat\theta = \frac{X}{n}$$

For $X = n$ the log-likelihood has derivative

$$U(\theta) = \frac{n}{\theta} > 0$$

for all θ so that the likelihood is an increasing function of θ which is maximized at $\hat\theta = 1 = X/n$. Similarly when $X = 0$ the maximum is at $\hat\theta = 0 = X/n$.

If $X_1, \ldots, X_n$ are independent then the log likelihood is of the form

$$\ell(\theta) = \sum \log(f(X_i, \theta))$$

The score function is

$$U(\theta) = \sum \frac{\partial \log f}{\partial\theta}(X_i, \theta)$$

The mle $\hat\theta$ maximizes ℓ. If the maximum occurs at the interior of the parameter space and the log likelihood continuously differentiable then $\hat\theta$ solves the likelihood equations

$$U(\theta) = 0$$

Examples

N(μ, σ^2)

There is a unique root of the likelihood equations. It is a global maximum.

[Remark: Suppose we had called $\tau = \sigma^2$ the parameter. The score function would still have two components with the first component being the same as before but now the second component is

$$\frac{\partial}{\partial \tau}\ell = \frac{\sum(X_i - \mu)^2}{2\tau^2} - \frac{n}{2\tau}$$

Setting the new likelihood equations equal to 0 still gives

$$\hat{\tau} = \hat{\sigma}^2$$

This is a general **invariance** (or **equivariance**) principal. If $\phi = g(\theta)$ is some reparametrization of a model (a one to one relabelling of the parameter values) then $\hat{\phi} = g(\hat{\theta})$. We will see that this does not apply to other estimators.]

Cauchy: location θ

There is at least 1 root of the likelihood equations but often several more. One of the roots is a global maximum, others, if they exist may be local minima or maxima.

Binomial(n, θ)

If $X = 0$ or $X = n$ there is no root of the likelihood equations; in this case the likelihood is monotone. For other values of X there is a unique root, a global maximum. The global maximum is at $\hat{\theta} = X/n$ even if $X = 0$ or n.

The 2 parameter exponential

The density is

$$f(x; \alpha, \beta) = \frac{1}{\beta} e^{-(x-\alpha)/\beta} 1(x > \alpha)$$

The resulting log-likelihood is $-\infty$ for $\alpha > \min\{X_1, \ldots, X_m n\}$ and otherwise is

$$\ell(\alpha, \beta) = -n \log(\beta) - \sum(X_i - \alpha)/\beta$$

As a function of α this is increasing till α reaches

$$\hat{\alpha} = X_{(1)} = \min\{X_1, \ldots, X_m n\}$$

which gives the mle of α. Now plug in this value $\hat{\alpha}$ for α and get the so-called profile likelihood for β:

$$\ell_{\text{profile}}(\beta) = -n \log(\beta) - \sum(X_i - X_{(1)})/\beta$$

Take the β derivative and set it equal to 0 to get

$$\hat{\beta} = \sum(X_i - X_{(1)})/n$$

Notice that the mle $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ does *not* solve the likelihood equations; we had to look at the edge of the possible parameter space. The parameter α is called a *support* or *truncation* parameter. ML methods behave oddly in problems with such parameters.

Three parameter Weibull

The density in question is

$$f(x;\alpha,\beta,\gamma) = \frac{1}{\beta}\left(\frac{x-\alpha}{\beta}\right)^{\gamma-1}\exp[-\{(x-\alpha)/\beta\}^\gamma]1(x>\alpha)$$

There are 3 derivatives to take to solve the likelihood equations. Setting the β derivative equal to 0 gives the equation

$$\hat\beta(\alpha,\gamma) = \left[\sum (X_i-\alpha)^\gamma/n\right]^{1/\gamma}$$

where we use the notation $\hat\beta(\alpha,\gamma)$ to indicate that the mle of β could be found by finding the mles of the other two parameters and then plugging in to the formula above. It is not possible to find explicitly the remaining two parameters; numerical methods are needed. However, you can see that putting $\gamma < 1$ and letting $\alpha \to X_{(1)}$ will make the log likelihood go to ∞. The mle is not uniquely defined, then, since any $\gamma < 1$ and any β will do.

If the true value of γ is more than 1 then the probability that there is a root of the likelihood equations is high; in this case there must be two more roots: a local maximum and a saddle point! For a true value of $\gamma > 0$ the theory we detail below applies to the local maximum and not to the global maximum of the likelihood equations.

Large Sample Theory

In the next few lectures we will be working toward explaining and "proving" the following theorem:

Theorem: Under "suitable regularity conditions"

1. The MLE is consistent.

2. The MLE is asymptotically normal.

The meaning of the first assertion is that in a precise mathematical sense

$$\hat\theta \to \theta_0$$

as the sample size n goes to infinity. In this course we will simply provide some heuristics which help indicate why this theorem ought to be true.

The second assertion means that

$$n^{1/2}(\hat\theta-\theta_0) \Rightarrow N(0,\sigma^2)$$

for a certain σ^2; we will later show how to compute σ, how to estimate σ and how to use this to get confidence intervals and hypothesis tests.