STAT 450

Lecture 18

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

- Define the Fisher information matrix.
- Illustrate standard MLE theory on $N(\mu, \sigma^2)$ example.

Today's notes

Estimating Equations

An equation of the form

$$U(X_1, \ldots, X_n; \theta) = 0$$

is called an **estimating equation**; we find an estimate of θ by solving the equation.

Examples:

1. The likelihood equations:

$$\frac{\partial \ell(\theta)}{\partial \theta} = U(X_1, \ldots, X_n; \theta) = 0$$

2. The normal equations in a linear regression model:

$$X^t X \beta = X^t Y$$

   are of the form

$$X^t(Y - X\beta) \equiv U(X, Y, \beta) = 0$$

3. The method of moments (later in this course)
4. Quasi-likelihood estimation as in STAT 402.

Definition: : An estimating equation is **unbiased** if

$$\mathrm{E}_\theta \left[U(X_1, \ldots, X_n; \theta) \right] = 0$$

It is a fact that in regular models the likelihood equations are unbiased.

It is a fact that unbiased estimating equations often lead to consistent estimates:

Theorem: Assume "regularity conditions". If

a) $\hat\theta$ is a root of $U(X_1, \ldots, X_n; \theta) = 0$ and

b) $\mathrm{E}_\theta \left[U(X_1, \ldots, X_n; \theta) \right] = 0$

then

1. $\hat\theta$ is consistent.

2. $\hat\theta - \theta \approx MVN(0,\Sigma)$ where

$$\Sigma = A^{-1}BA^{-1}$$

with

$$B = \mathrm{Var}_\theta(U(X_1,\ldots,X_n;\theta))$$

and

$$A = -\mathrm{E}\left[\frac{\partial}{\partial\theta}U(X_1,\ldots,X_n;\theta)\right]$$

Example: $N(\mu,\sigma^2)$ sample: Suppose $X_1,\ldots,X_n$ are iid $N(\mu,\sigma^2)$. The score function is

$$U(\mu,\sigma) = \left[\begin{array}{c} \frac{\sum(X_i-\mu)}{\sigma^2} \\ \frac{\sum(X_i-\mu)^2}{\sigma^3} - \frac{n}{\sigma} \end{array}\right]$$

I want to illustrate the fact that the likelihood equations are unbiased and to hint at how the first assertion is proved. To do so, I need notation for the true value of μ and σ. I use μ_0 and σ_0 for these values and then let μ and σ without the subscripts be the arguments of the random function U. Since $\mathrm{E}(X_i)=\mu_0$ and

$$\begin{aligned}
\mathrm{E}[(X_i-\mu)^2] &= \mathrm{E}[(X_i-\mu_0+\mu_0-\mu)^2] \\
&= \mathrm{E}[(X_i-\mu_0)^2 + 2(X_i-\mu_0)(\mu_0-\mu) + (\mu_0-\mu)^2] \\
&= \sigma_0^2 + (\mu_0-\mu)^2
\end{aligned}$$

we can compute

$$\mathrm{E}_{\mu_0,\sigma_0}[U(X,1\ldots,X_n;\mu,\sigma)] = \left[\begin{array}{c} \frac{n(\mu_0-\mu)}{\sigma^2} \\ \frac{n(sigma_0^2+(\mu_0-\mu)^2)}{\sigma^2} - \frac{n}{\sigma^2} \end{array}\right]$$

I want you to see that if $\mu=\mu_0$ and $\sigma=\sigma_0$ then this expectation is 0 **and** vice-versa.

Now imagine $\epsilon > 0$ is tiny. If n is large then (probably) $|\bar X - \mu_0| < \epsilon$. This would guarantee that

$$\frac{\sum(X_i-(\mu_0+\epsilon))}{\sigma^2} < 0$$

and

$$\frac{\sum(X_i-(\mu_0-\epsilon))}{\sigma^2} > 0$$

Since the first component of U is a continuous function of μ and has opposite signs at $\mu_0 \pm \epsilon$ the root of the equation must be within ϵ of the true answer. This is roughly we prove **consistency**.

Now we will work out the other pieces of the theorem. I will compute B, A and Σ in this example. In order to do so I will use a variety of rules for working with variances and covariances. Remember that if X is a random vector with $\mathrm{E}(X)=\mu$ then

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)(X-\mu)^t] = \mathrm{E}[XX^t] - \mu\mu^t$$

The ijth entry in this matrix is

$$\text{Cov}(X_i, X_j) = \text{E}[(X_i - \mu_i)(X_j - \mu_j)] = \text{E}[X_i X_j] - \mu_i \mu_j$$

We need the following properties of Var and Cov. In what follows as and bs are constants while V_is and W_is are random variables, not independent unless I say so.

1. $\text{Var}(aV + b) = a^2 \text{Var}(V)$.
2. If $W_1, \ldots, W_n$ are independent then
$$\text{Var}(\sum a_i W_i) = \sum a_i^2 \text{Var}(W_i)$$
3. If V and W are independent the $\text{Cov}(V, W) = 0$.
4. $\text{Cov}(\sum a_i W_i, V) = \sum a_i \text{Cov}(W_i, V)$
5. $\text{Cov}(V, \sum a_i W_i) = \sum a_i \text{Cov}(V, W_i)$
6. $\text{Cov}(V, W) =$
$textCov(W, V)$
7. $\text{Cov}(V, V) = \text{Var}(V)$
8. $\text{Cov}(W, a) = 0$

I now intend to compute $\text{Var}_{\theta_0}(U(\theta_0))$. To save typing effort I drop the subscripts on θ_0 but I remind you that the calculation nowonly works then the argument of U is the true value of the parameter θ.

We have

$$B = \begin{bmatrix} \text{Var}(\sum(X_i - \mu)/\sigma^2) & \text{Cov}(\sum(X_i - \mu)/\sigma^2, \sum(X_i - \mu)^2/\sigma^3 - n/\sigma) \\ \text{Cov}(\sum(X_i - \mu)/\sigma^2, \sum(X_i - \mu)^2/\sigma^3 - n/\sigma) & \text{Var}(\sum(X_i - \mu)^2/\sigma^3 - n/\sigma) \end{bmatrix}$$

To compute the individual entries it is simpler to define $Z_i = (X_i - \mu)/\sigma$ and remember that $Z_i \sim N(0, 1)$. Thus $\text{E}(Z) = 0$, $\text{E}(Z^2) = 1$, $\text{E}(Z^3) = 0$ and $\text{E}(Z^4) = 3$. Then

$$\text{Var}(\sum(X_i - \mu)/\sigma^2) = \text{Var}(\sum Z_i)/\sigma^2 = n/\sigma^2$$

$$\text{Var}(\sum(X_i - \mu)^2/\sigma^3 - n/\sigma) = \text{Var}(\sum Z_i^2)/\sigma^2$$

SInce $\text{Var}(Z_i^2) = \text{E}(Z_i^4) - \left[\text{E}(Z_i^2)\right]^2 = 3 - 1 = 2$ we find

$$\text{Var}(\sum(X_i - \mu)^2/\sigma^3 - n/\sigma) = 2n/\sigma^2$$

Finally

$$\text{Cov}(\sum(X_i - \mu)/\sigma^2, \sum(X_i - \mu)^2/\sigma^3 - n/\sigma) = \text{Cov}(\sum Z_i, \sum Z_j^2)/\sigma^2 = \sum_{ij} \text{Cov}(Z_i, Z_i^2)/\sigma^2$$

For $i \neq j$, $\mathrm{Cov}(Z_i, Z_j^2) =$ by independence. For $i = j$

$$\mathrm{Cov}(Z_i, Z_i^2) = \mathrm{E}(Z_i^3) - \mathrm{E}(Z_i)\mathrm{E}(Z_i^2) = 0$$

Assemble these to discover

$$B = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

Next: compute $A = -\mathrm{E}(\partial U / \partial \theta)$. First

$$\begin{aligned} \frac{\partial U(\mu, \sigma)}{\partial \theta} &= \begin{bmatrix} \frac{\partial}{\partial \mu} \frac{\sum(X_i - \mu)}{\sigma^2} & \frac{\partial}{\partial \sigma} \frac{\sum(X_i - \mu)}{\sigma^2} \\ \frac{\partial}{\partial \mu} \left(\frac{\sum(X_i - \mu)^2}{\sigma^3} - \frac{n}{\sigma} \right) & \frac{\partial}{\partial \sigma} \left(\frac{\sum(X_i - \mu)^2}{\sigma^3} - \frac{n}{\sigma} \right) \end{bmatrix} \\ &= \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{2\sum(X_i - \mu)}{\sigma^3} \\ -\frac{2\sum(X_i - \mu)}{\sigma^3} & -\frac{3\sum(X_i - \mu)2}{\sigma^4} + \frac{n}{\sigma^2} \end{bmatrix} \end{aligned}$$

Take expected values to get

$$A = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix} = B$$