C-R lower bound, Fisher Information

May 5, 2014

Some identities related to the log likelihood functions

.

Since we always have, (for a family of density functions $f(x,\theta)$)

$$\int f(x,\theta)dx = 1$$

(true for all θ) therefore taking derivative on both side wrt θ and assume we can taking the derivative inside the integral (this requires some conditions), we have

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

multiply and devide inside the integral by $f(x,\theta)$, we can rewrite the equality as

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

which is

$$E\left[\frac{\partial}{\partial \theta} \log f(X,\theta)\right] = 0 \tag{1}$$

Taking one more derivative (and exchange the integral and derivative again) we have

$$\int \left(\frac{\partial^2}{\partial \theta^2} \log f(x,\theta) \cdot f(x,\theta) + [\frac{\partial}{\partial \theta} \log f(x,\theta)] \frac{\partial}{\partial \theta} f(x,\theta)\right) dx = 0$$

multiple and devide inside the second integral by $f(x,\theta)$, that is

$$E\frac{\partial^2}{\partial\theta^2}\log f(X,\theta) + E\left[\frac{\partial}{\partial\theta}\log f(X,\theta)\right]^2 = 0 \qquad (2)$$

We shall call these Bartlett identities one and two. More derivative will result more identities.

Cramer-Rao Lower Bound

Now supppose we have an unbiased estimator of θ, i.e.

$$E\hat{\theta}(X) = \theta$$

(here x can be a vector) that is

$$\int \hat{\theta}(x) f(x,\theta)dx = \theta$$

Taking derivative wrt θ on both sides (and assuming we can taking the derivative insde the integral)

$$\int \hat{\theta}(x)\frac{\partial}{\partial\theta} f(x,\theta)dx = 1$$

(by definition, the estimator $\hat{\theta}$ cannot depend on the parameter θ, it is a function of data x.) Multiply and devide by $f(x,\theta)$;

$$\int \hat{\theta}(x)\frac{\partial \log f(x,\theta)}{\partial\theta} f(x,\theta)dx = 1$$

Write it as expectation:

$$E\hat{\theta}(X)\frac{\partial \log f(X,\theta)}{\partial\theta} = 1$$

We know

$$E\frac{\partial \log f(X,\theta)}{\partial\theta} = 0$$

(first Bartlett identity) therefore the above is also

$$cov(\hat{\theta}(X))(\frac{\partial \log f(X,\theta)}{\partial\theta})) = 1$$

Recall $cov(XY) = E(X - EX)(Y - EY) = E(X)(Y - EY)$.

Using the inequality $[cov(XY)]^2 \leq VarX \cdot VarY$ (or Cauchy-Schwarz inequality) we have

$$Var\hat{\theta}(X) \times Var\frac{\partial \log f(X,\theta)}{\partial \theta} \geq 1$$

Another way to write it is

$$Var\hat{\theta}(X) \geq \frac{1}{Var\left(\frac{\partial \log f(X,\theta)}{\partial \theta}\right)} \tag{3}$$

This is called The Cramer-Rao lower bound.

Please note (1) this lower bound is for ALL the unbiased estimators; and (2) the lower bound is the same as the approximate variance of the MLE.

In other words, the MLE is (at least) approximately the best estimator. (well, MLE may have a small, finite sample bias, sometimes).

Using the identity, we can rewrite the C-R lower bound as

$$Var\hat{\theta}(x) \geq \frac{1}{-E\left(\frac{\partial^2 \log f(X,\theta)}{\partial \theta^2}\right)} = \frac{1}{E\left(\frac{\partial \log f(X,\theta)}{\partial \theta}\right)^2}$$

Please note the condition of interchange the derivative and integration will exclude the uniform $[0, \theta]$ distribution as $f(x, \theta)$.

Fisher Information

One fundamental quantity in statistical inference is *Fisher Information.* We will define Fisher information, two kinds.

As with MLEs, we concentrate on the log likelihood.

Definition: The expected Fisher Information about a parameter θ is the expectation of the minus second derivative of the log likelihood: $E\{-\frac{\partial^2}{(\partial\theta)^2} \log lik(\theta)\}$.

Definition: The observed Fisher Information about a parameter θ is the minus second derivative of the log likelihood, and then with the parameter replaced by the MLE. $-\frac{\partial^2}{(\partial\theta)^2} \log lik(\theta)|_{\theta=\theta(MLE)}$.

Example 1

Let $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ with μ unknown and σ^2 known. The loglikelihood is

$$-(n/2)\ln(2\pi\sigma^2) - (1/2\sigma^2)\sum(y_i - \mu)^2$$

Taking derivatives twice with respect to μ,

$$\frac{\partial^2 \log L(\mu)}{\partial\mu^2} = \frac{\partial}{\partial\mu}(1/2\sigma^2)\sum 2(y_i - \mu) = -n/\sigma^2$$

Here, both the expected and the observed Fisher information are equal to n/σ^2. Because there is no random quantity to take expectation of, and there is no unknown parameter, μ, to be replaced of.

Example 2

Let $Y_1, \ldots, Y_n \sim Poisson(\lambda)$. The loglikelihood is

$$-n\lambda + \sum y_i \ln\lambda + \ln(1/\prod_i(y_i!))$$

Taking derivatives twice with respect to λ we find

$$\frac{\partial^2}{\partial\lambda^2}\log L(\lambda) = \frac{\partial}{\partial\lambda}\left(-n + \sum y_i/\lambda\right) = -\sum y_i/\lambda^2$$

The $\hat\lambda_{mle} = \bar{Y}$. The observed Fisher information here is $\sum y_i/\lambda^2$, with λ replaced by $\hat\lambda_{mle} = \bar{Y}$. So it simplify to $n^2/\sum y_i$.

The expected Fisher information is

$$E\sum y_i/\lambda^2 = n\lambda/\lambda^2 = n/\lambda = I_{fisher}(\lambda)$$

The observed Fisher information is something you can calculate its value given the data. It depends on the data, but not explicitly involve parameter. It usually is random.

The expected Fisher information is the theoretical value, often involving unknown parameters. But it is non-random (do not depend on the observed sample), as in the Poisson example.