Math 505 Notes

Chapter 6

Jesse Crawford

Department of Mathematics
Tarleton State University

Fall 2009

Outline

Definition

- Consider a statistical model with p.d.f. $f(x;\theta)$, for $\theta \in \Omega$.
- Then the *likelihood function* for the model is

$$\begin{array}{rcl} L : \Omega \times \mathbb{R}^n & \to & [0,1] \\ L(\theta; x) & = & \prod_{i=1}^n f(x_i;\theta). \end{array}$$

- θ is the unknown parameter
- $x = (x_1, \ldots, x_n)'$ can be thought of as the vector of observations from a random sample $X = (X_1, \ldots, X_n)'$.
- Let $\hat{\theta}$ be an estimator of θ such that the maximum value

$$\max\{L(\theta; X) \,|\, \theta \in \Omega\}$$

  is attained at $\hat{\theta}$ with probability one.
- Then $\hat{\theta}$ is called a *maximum likelihood estimator* for θ.

Definition

The *likelihood equation* for the model is

$$\frac{\partial}{\partial\theta} L(\theta; x) = 0.$$

Let θ_0 be the true value of the parameter.

Regularity Conditions

0. The pdfs are distinct, i.e., if $\theta \neq \theta'$, then $f(x_i; \theta) \neq f(x_i; \theta')$.
1. The pdfs have common support for all $\theta \in \Omega$.
2. The parameter θ_0 is an interior point of Ω.

Theorem

Under the regularity conditions (0) *and* (1)*,*

$$\lim_{n\to\infty} P_{\theta_0}[L(\theta_0, X) > L(\theta, X)] = 1, \text{ for all } \theta \neq \theta_0.$$

Theorem

- *Assume that the regularity conditions* (0) *through* (2) *are satisfied,*
- *and assume that* $f(x;\theta)$ *is differentiable wrt.* θ*.*
- *Suppose that the likelihood equation based on a sample of size* n *has a unique solution* $\hat{\theta}_n$*.*
- *Then* $\hat{\theta}_n$ *is a consistent estimator of* θ_0*.*

Outline

- Consider a statistical model with pdf $f(x;\theta)$, for $\theta \in \Omega$,
- where $\Omega \subseteq \mathbb{R}$ is an open interval.

Regularity Conditions

3. The pdf $f(x;\theta)$ is twice differentiable as a function of θ.
4. $$\frac{\partial^k}{\partial\theta^k}\int_{-\infty}^{\infty} f(x;\theta)\,dx = \int_{-\infty}^{\infty}\frac{\partial^k}{\partial\theta^k} f(x;\theta)\,dx \text{ for } k = 1, 2$$

- $\frac{\partial \log f(x;\theta)}{\partial\theta}$ is called the *score function* corresponding to the model.
- $E_\theta(\frac{\partial \log f(X;\theta)}{\partial\theta}) = 0$, for every $\theta \in \Omega$.
- Define the *Fisher information* (at θ) for the model to be

$$I(\theta) = \int_{-\infty}^{\infty}\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)dx = -\int_{-\infty}^{\infty}\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} f(x;\theta)dx$$

- Note that $I(\theta) = \mathrm{Var}_\theta(\frac{\partial \log f(X;\theta)}{\partial\theta})$.

Theorem (Rao-Cramér Lower Bound)

- *Let $X_1, \ldots, X_n$ be a sample from the aforementioned model, and*
- *assume the regularity conditions* (0) *through* (4) *are satisfied.*
- *Let $Y = u(X_1, \ldots, X_n)$ be a statistic with mean $E_\theta(Y) = k(\theta)$.*
- *Then*

$$Var(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)}.$$

- *In particular, if Y is an unbiased estimator of θ, then*

$$Var(Y) \geq \frac{1}{nI(\theta)}.$$

Definition

Let Y be an unbiased estimator whose variance attains the Rao-Cramér lower bound. Then Y is called an *efficient* estimator.

Regularity Conditions

5. The pdf $f(x;\theta)$ is three times differentiable wrt. θ. For every $\theta_0 \in \Omega$, there exists a constant c and a function $M(x)$ such that $E_{\theta_0}[M(X)] < \infty$, and

$$\left| \frac{\partial^3}{\partial \theta^3} \log f(x;\theta) \right| \leq M(x),$$

for all $\theta_0 - c < \theta < \theta_0 + c$ and all x in the support of X.

Theorem

- *Assume regularity conditions* (0) *through* (5) *are satisfied and*
- *let* $X_1, \ldots, X_n$ *be a sample from the pdf* $f(x;\theta_0)$.
- *Also, assume the Fisher information satisfies* $0 < I(\theta_0) < \infty$.
- *Then any consistent sequence of solutions to the mle equations satisfies*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_0)}\right).$$

Definition

- A sequence of random variables $\{X_n\}$ is *bounded in probability* if
- for every $\epsilon > 0$, there exists a $B_\epsilon > 0$ and $N_\epsilon \in \mathbb{N}$ such that

$$\text{if } n \geq N_\epsilon, \text{ then } P[|X_n| \leq B_\epsilon] \geq 1 - \epsilon.$$

- $$\hat{\theta}_n \approx N\left(\theta_0, \frac{1}{nI(\theta_0)}\right).$$
- Under the regularity conditions, MLEs are asymptotically normal and efficient.
- An approximate $1 - \alpha$ confidence interval for θ is

$$\hat{\theta}_n \pm z_{\alpha/2} \frac{1}{\sqrt{nI(\hat{\theta}_n)}}.$$

Outline

Definition

- Consider a statistical model given by the pdf $f(x;\theta)$, for $\theta \in \Omega \subseteq \mathbb{R}$.
- Let $\theta_0 \in \Omega$, and Consider the testing problem

$$\mathrm{H}_0 : \theta = \theta_0 \text{ vs. } \mathrm{H}_1 : \theta \neq \theta_0.$$

- Given a sample $X_1, \ldots, X_n$, the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(X_i;\theta).$$

- The *likelihood ratio statistic* is

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})}.$$

- $$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})}.$$
- $0 \leq \Lambda \leq 1$.
- Small values of Λ provide evidence against H_0.
- Likelihood ratio test:

$$\text{Reject } H_0 \text{ if } \Lambda \leq c,$$

- where c is chosen so that $\alpha = P_{\theta_0}[\Lambda \leq c]$.

Theorem

- *Assume the regularity conditions* (0) *through* (5) *hold.*
- *Under the null hypothesis,*

$$\chi_L^2 := -2\log\Lambda \xrightarrow{D} \chi^2(1).$$

- *We reject* H_0 *if* $\chi_L^2 \geq \chi_\alpha^2(1)$.
- *Alternatively, the statistic* χ_L^2 *can be replaced in the test above by the*
  - *Wald test statistic*

$$\chi_W^2 := \left[\sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0)\right]^2,$$

  - *or Rao's score test statistic*

$$\chi_R^2 := \left(\frac{l'(\theta_0)}{\sqrt{nI(\theta_0)}}\right)^2.$$