Math 505 Notes
Chapter 6

Jesse Crawford

Department of Mathematics
Tarleton State University

Fall 2009
Outline

1. Section 6.1: Maximum Likelihood Estimation
2. Section 6.2: Rao-Cramér Lower Bound and Efficiency
3. Section 6.3: Maximum Likelihood Tests
Definition

- Consider a statistical model with p.d.f. \( f(x; \theta) \), for \( \theta \in \Omega \).
- Then the *likelihood function* for the model is
  \[
  L : \Omega \times \mathbb{R}^n \rightarrow [0, 1] \\
  \quad L(\theta; x) = \prod_{i=1}^{n} f(x_i; \theta).
  \]
- \( \theta \) 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 \( \theta \) such that the maximum value
  \[
  \max\{L(\theta; X) \mid \theta \in \Omega\}
  \]
  is attained at \( \hat{\theta} \) with probability one.
- Then \( \hat{\theta} \) is called a *maximum likelihood estimator* for \( \theta \).
Definition
The *likelihood equation* for the model is

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

Let \( \theta_0 \) 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 \( \theta_0 \) is an interior point of \( \Omega \).

Theorem

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

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

- Assume that the regularity conditions (0) through (2) are satisfied,
- and assume that \( f(x; \theta) \) is differentiable wrt. \( \theta \).
- 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 \( \theta_0 \).
Outline

1. Section 6.1: Maximum Likelihood Estimation

2. Section 6.2: Rao-Cramér Lower Bound and Efficiency

3. Section 6.3: Maximum Likelihood Tests
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 $\theta$.

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.

\[
E_\theta\left( \frac{\partial \log f(X; \theta)}{\partial \theta} \right) = 0, \text{ for every } \theta \in \Omega.
\]

Define the *Fisher information* (at $\theta$) 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 $I(\theta) = \text{Var}_\theta\left( \frac{\partial \log f(X; \theta)}{\partial \theta} \right)$. 
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

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

- In particular, if $Y$ is an unbiased estimator of $\theta$, then

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

Definition

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

The pdf $f(x; \theta)$ is three times differentiable wrt. $\theta$. 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

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

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

\[\hat{\theta}_n \approx N \left( \theta_0, \frac{1}{nl(\theta_0)} \right).\]

Under the regularity conditions, MLEs are asymptotically normal and efficient.

An approximate \( 1 - \alpha \) confidence interval for \( \theta \) is

\[
\hat{\theta}_n \pm z_{\alpha/2} \frac{1}{\sqrt{nl(\hat{\theta}_n)}}.
\]
Outline

1. Section 6.1: Maximum Likelihood Estimation
2. Section 6.2: Rao-Cramér Lower Bound and Efficiency
3. Section 6.3: Maximum Likelihood Tests
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
  
  $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$.

- 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 \Lambda provide evidence against H_0.

Likelihood ratio test:

Reject H_0 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,

\[
\chi^2_{L} := -2 \log D \xrightarrow{D} \chi^2(1).
\]

- We reject \( H_0 \) if \( \chi^2_{L} \geq \chi^2_{\alpha}(1) \).
- Alternatively, the statistic \( \chi^2_{L} \) can be replaced in the test above by the
  - Wald test statistic

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

  - or Rao's score test statistic

\[
\chi^2_{R} := \left( \frac{l'(\theta_0)}{\sqrt{nI(\theta_0)}} \right)^2.
\]