Topic 15

Maximum Likelihood Estimation

Introduction and Procedure

Outline

Introduction

Procedure

- Bernoulli Trials

Introduction

We begin with observations $X = (X_1, \ldots, X_n)$ of random variables chosen according to one of a family of probabilities P_θ indexed by the parameter space, Θ. In addition,

$$\mathbf{f}(\mathbf{x}|\theta), \quad \mathbf{x} = (x_1, \ldots, x_n)$$

will be used to denote the joint density function when θ is the true state of nature.

Definition. The likelihood function is the density function regarded as a function of θ.

$$\mathbf{L}(\theta|\mathbf{x}) = \mathbf{f}(\mathbf{x}|\theta), \ \theta \in \Theta.$$

The maximum likelihood estimate (MLE),

$$\hat{\theta}(\mathbf{x}) = \arg\max_{\theta \in \Theta} \mathbf{L}(\theta|\mathbf{x}).$$

Thus, we are presuming that a *unique* global maximum exists.

Introduction

This class of estimators has an important property.

- If $\hat{\theta}(\mathbf{x})$ is a maximum likelihood estimate for θ, then $g(\hat{\theta}(\mathbf{x}))$ is a maximum likelihood estimate for $g(\theta)$.
  - If $\hat{\theta}$ is the maximum likelihood estimate for the variance, then $\sqrt{\hat{\theta}}$ is the maximum likelihood estimator for the standard deviation.

For independent observations, the likelihood

$$\mathbf{L}(\theta|\mathbf{x}) = f(x_1|\theta)f(x_2|\theta)\cdots f(x_n|\theta).$$

is the product of density functions. Using the properties of the logarithm of a product,

$$\ln \mathbf{L}(\theta|\mathbf{x}) = \ln f(x_1|\theta) + \ln f(x_2|\theta) + \cdots + \ln f(x_n|\theta).$$

Finding zeroes of the score function, $\partial \ln \mathbf{L}(\theta|\mathbf{x})/\partial\theta$, the derivative of the logarithm of the likelihood, will be easier.

Bernoulli Trials

If the experiment consists of n Bernoulli trials with success probability p, then

$$\mathbf{L}(p|\mathbf{x}) = p^{x_1}(1-p)^{(1-x_1)} \cdots p^{x_n}(1-p)^{(1-x_n)} = p^{(x_1+\cdots+x_n)}(1-p)^{n-(x_1+\cdots+x_n)}.$$

$$\ln \mathbf{L}(p|\mathbf{x}) = \ln p(\sum_{i=1}^{n} x_i) + \ln(1-p)(n - \sum_{i=1}^{n} x_i) = n(\bar{x}\ln p + (1-\bar{x})\ln(1-p)).$$

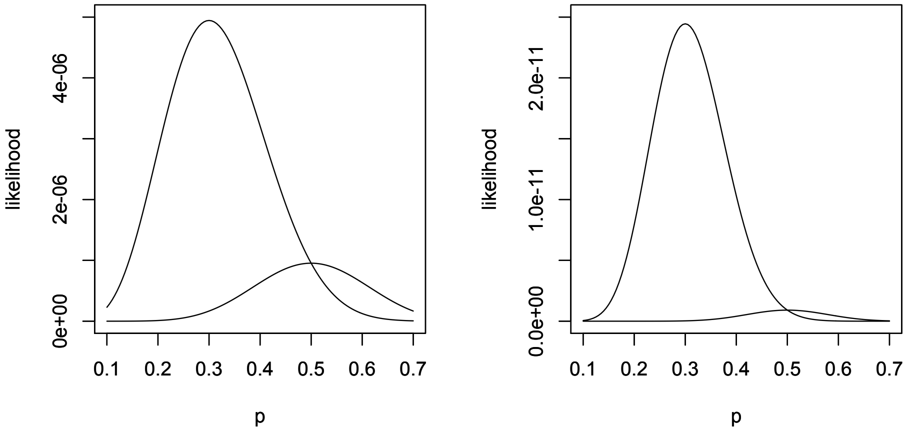
$$\frac{\partial}{\partial p}\ln \mathbf{L}(p|\mathbf{x}) = n\left(\frac{\bar{x}}{p} - \frac{1-\bar{x}}{1-p}\right) = n\frac{\bar{x}-p}{p(1-p)}$$

This equals zero when $p = \bar{x}$.

Exercise. Check that this is a maximum.

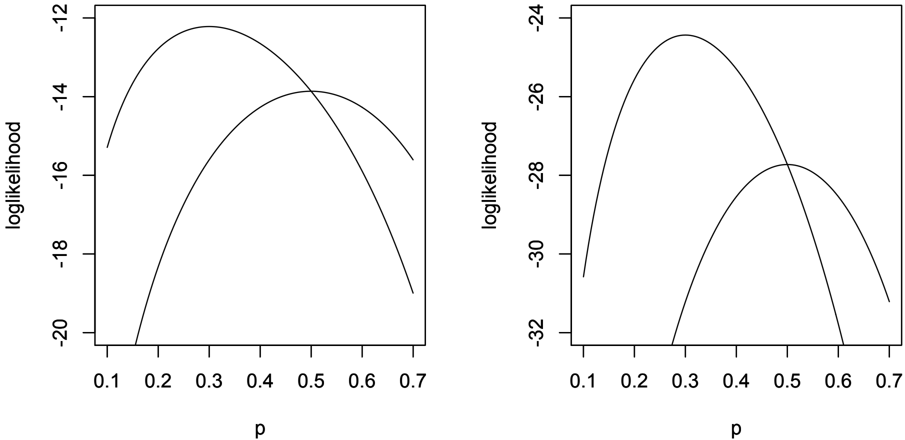
In this case, the maximum likelihood estimator is also unbiased.

Bernoulli Trials



Graph of $\mathbf{L}(p|\mathbf{x})$with (left) 6 and 10 successes in 20 trials and (right) 12 and 20 successes in 40 trials.

Bernoulli Trials



Graph of $\ln \mathbf{L}(p|\mathbf{x})$ with (left) 6 and 10 successes in 20 trials and (right) 12 and 20 successes in 40 trials.

Bernoulli Trials

Notice

- Both $\mathbf{L}(p|\mathbf{x})$ and $\ln \mathbf{L}(p|\mathbf{x})$ have their maximum at $p = \bar{x}$.
- The maxima when $\bar{x} = 0.3$ is greater than the corresponding maxima when $\bar{x} = 0.5$. However, for the case $n = 20$ there is a factor of

$$\binom{20}{10} \Big/ \binom{20}{6} = \frac{143}{30}$$

  that produce 10 successes than produce 6.
- The maxima are more peaked with larger values of n.
  - We will soon learn that the variance in the estimator is closely tied to the curvature of the log likelihood function at the maximum likelihood estimate.