Classical and Bayesian inference

AMS 132

March 14, 2018

Definition and Properties of Fisher Information

- The Fisher information is one property of a distribution that can be used to measure how much information one is likely to obtain from a random variable or a random sample.
- Also, the Fisher information allows to lower bound the variance of an estimator.

Definition and Properties of Fisher Information

Definition (Fisher Information in a Random Variable)

Let X be a random variable whose distribution depends on a parameter θ and let the p.f. or p.d.f. of X be $f(x \mid \theta)$. Assume that

- θ takes values in an open interval of the real line,
- the set of x such that $f(x \mid \theta) > 0$ is the same for all θ,
- $\lambda(x \mid \theta) = \log f(x \mid \theta)$ is twice differentiable as a function of θ.

Then, the Fisher information $I(\theta)$ in the random variable X is defined as

$$I(\theta) = E_\theta \left\{ \left[\frac{d}{d\theta} \lambda(X \mid \theta) \right]^2 \right\}.$$

- If, additionally, the two derivatives of $f(x \mid \theta)$ with respect to θ can be calculated by reversing the order of integration and differentiation, then the Fisher information also equals

$$I(\theta) = -E_\theta \left[\frac{d^2}{d\theta^2} \lambda(X \mid \theta) \right] = Var_\theta \left[\frac{d}{d\theta} \lambda(X \mid \theta) \right].$$

Definition and Properties of Fisher Information

Example

a) Suppose that X has the Bernoulli distribution with parameter p. Determine the Fisher information $I(p)$ in X.

b) Suppose that X has the normal distribution with unknown mean μ and known variance σ^2. Determine the Fisher information $I(\mu)$ in X.

Definition and Properties of Fisher Information

- There is a simple relation between the Fisher information $I_n(\theta)$ in the entire sample and the Fisher information $I(\theta)$ in a single observation X_i.

Theorem (The Fisher Information in a Random Sample)

Under the conditions of the definition of $I(\theta)$,

$$I_n(\theta) = nI(\theta).$$

- So, the Fisher information in a random sample of n observations is simply n times the Fisher information in a single observation.

Definition and Properties of Fisher Information

Example (Customer arrivals)

A store owner is interested in learning about customer arrivals. She models arrivals during an hour as a Poisson process with unknown rate λ. She thinks of two different possible sampling plans to obtain information about customer arrivals:

a) One plan is to choose a fixed number, n, of customers and to see how long, X, it takes until n customers arrive.

b) The other plan is to observe for a fixed length of time, t , and count how many customers, Y , arrive during time t.

Discuss which sampling plan provides more information.

Efficient estimators

- We will show how to use the Fisher information to determine a lower bound for the variance of an unbiased estimator of the parameter θ in a given problem.

Corollary (Cramér-Rao Lower bound of the Variance of an Unbiased Estimator)

Assume the conditions of the definition of $I(\theta)$. Let T be an unbiased estimator of θ. Then

$$Var_\theta(T) \geq \frac{1}{nI(\theta)}.$$

Efficient estimators

- An estimator whose variance equals the Cramér-Rao lower bound makes the most efficient use of the data in some sense.

Definition (Efficient Estimator)

It is said that an unbiased estimator, T, is an efficient estimator of θ if, for every value of $\theta \in \Omega$,

$$Var_\theta(T) = \frac{1}{nI(\theta)}.$$

- In some problems, efficient estimators do not exist.

Efficient estimators

Example

a) Suppose that $X_1, \ldots, X_n$ form a random sample from the Bernoulli distribution with parameter p. Show that $\overline{X}_n$ is an efficient estimator for p.

b) Suppose that $X_1, \ldots, X_n$ form a random sample from the normal distribution with unknown mean μ and known variance σ^2. Show that $\overline{X}_n$ is an efficient estimator for μ.

- The results for finding a lower bound for an unbiased estimator or an efficient estimator can be extended for any estimator that has mean $m(\theta)$. In this case $Var_\theta(T) \geq \frac{[m'(\theta)]^2}{nI(\theta)}$, and T is efficient when $Var_\theta(T) = \frac{[m'(\theta)]^2}{nI(\theta)}$.

Properties of Maximum Likelihood Estimators for Large Samples

Theorem (Asymptotic Distribution of the M.L.E.)

Assume that $X_1, \ldots, X_n$ form a random sample from a distribution that has p.f. of p.d.f. $f(x \mid \theta)$. Assume the conditions of the definition of $I(\theta)$ are satisfied. Let $\widehat{\theta}_n$ denote the M.L.E. of θ. Then, if n is large, the distribution of $\widehat{\theta}_n$ is approximately normal with mean θ and variance $\frac{1}{nI(\theta)}$, this is,

$$[nI(\theta)]^{1/2}(\widehat{\theta}_n - \theta)$$

has a standard normal distribution.

Properties of Maximum Likelihood Estimators for Large Samples

Example

Suppose that $X_1, \ldots, X_n$ form a random sample from the Bernoulli distribution with parameter p.

a) Find the asymptotic distribution of the M.L.E. of p.
b) Use the above results for finding an approximate 95% confidence interval for p.