Estimation, Section 2.4

Grethe Hystad

September 16, 2012

Estimation

- In estimation theory one wants to answer the question: **what is the number?**
- The number could be temperature, length, mean lifetime, the slope and intercept of a line, mean income, the probability that small classes improve test scores, the probability for heritage a particular gene, cholesterol level in a certain population, mean weight of a product produced.
- One performs a well-designed experiment to estimate the number(s).

Statistics

Definition

A statistics is a function of the observed data that is independent of the unknown parameter.

- If the population is large, it is often impossible to measure the mean or standard deviation of the entire population. Instead we measure the sample statistics.

Examples of sample statistics are:

- sample mean: $\bar{x}$
- sample variance: s^2
- sample standard deviation: s
- sample median, sample quartiles, percentiles.

Statistics

- We consider random variables in which the p.m.f. is known but the distribution depends on an unknown parameter, say θ, that we want to estimate.
- Define the parameter space: $\Omega = \{\theta\}$ as the set of all possible parameters θ.
- Sometimes, one wants to select one member of the family $\{f(x, \theta) : \theta \in \Omega\}$ as the most likely p.m.f. of the random variable. Then one needs an estimate of the parameter θ.
- Example: The p.m.f for the Binomial distribution, $f(x; p) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, 2, ...$, where the parameter space is $\Omega = \{p : 0 < p < 1\}$.

Statistics

- Take a random sample $X_1, ..., X_n$ from a distribution (population) to get information about the unknown parameter θ. That is we perform the same experiment n times independently from each other in which each X_i has the same distribution. Such a random sample, $X_1, ..., X_n$, is said to be **independent and identically distributed (iid)**. We use the observations $X_1 = x_1$,...., $X_n = x_n$ to estimate the value of θ.

Estimator and estimate

- The **parameter** θ describes a property of the statistical population that we want information about.
- The **statistics**, $\mu(X_1, ..., X_n)$, which is a function of the random sample, $X_1, ..., X_n$ is called an **estimator** of θ if $\mu(X_1, ..., X_n)$ is usually close to θ. The estimator $\mu(X_1, ..., X_n)$ is a random variable.
- $\mu(x_1, ...x_n)$ is called the **estimate** of θ, where $x_1, ..., x_n$ are the observed values of $X_1, ..., X_n$.
- The estimate is called a **point estimate** since the estimate is a single number.

Likelihood function

- Since the random variables, $X_1, ..., X_n$, are independent, we have

  $$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta).$$

- $f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$ is called the joint p.m.f. of $X_1, ..., X_n$.

Definition

Define the likelihood function,

$$L(\theta) = L(\theta; x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i;\theta) \text{ for } \theta \in \Omega,$$

where θ is unknown and $x_1, ..., x_n$ are known.

Likelihood function

Example

- Consider n independent Bernoulli trials, $X_1, ..., X_n$, with an unknown success probability, p, with observed values $x_1, ..., x_n$. We want to estimate p.

Likelihood function

Example

- Consider n independent Bernoulli trials, $X_1, ..., X_n$, with an unknown success probability, p, with observed values $x_1, ..., x_n$. We want to estimate p.
- Each X_i has p.m.f, $f(x_i) = p^{x_i}(1-p)^{1-x_i}$.

Likelihood function

Example

- Consider n independent Bernoulli trials, $X_1, ..., X_n$, with an unknown success probability, p, with observed values $x_1, ..., x_n$. We want to estimate p.
- Each X_i has p.m.f, $f(x_i) = p^{x_i}(1-p)^{1-x_i}$.
- The joint p.m.f. is

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \\ &= p^{n\bar{x}}(1-p)^{n(1-\bar{x})} \end{aligned}$$

Likelihood function, example continue

Example

- Want to find the value of p that maximizes the likelihood function $L(p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$.

Likelihood function, example continue

Example

- Want to find the value of p that maximizes the likelihood function $L(p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$.
- It is easier to find the value of p that maximize $\ln(L(p))$.

Likelihood function, example continue

Example

- Want to find the value of p that maximizes the likelihood function $L(p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$.
- It is easier to find the value of p that maximize $\ln(L(p))$.
- We have $\ln(L(p)) = n\bar{x}\ln(p) + n(1-\bar{x})\ln(1-p)$.

Likelihood function, example continue

Example

- Want to find the value of p that maximizes the likelihood function $L(p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$.
- It is easier to find the value of p that maximize $\ln(L(p))$.
- We have $\ln(L(p)) = n\bar{x}\ln(p) + n(1-\bar{x})\ln(1-p)$.
- Taking the derivatives, $\frac{d}{dp}(\ln(L(p))) = n\bar{x}\frac{1}{p} - n(1-\bar{x})\frac{1}{1-p} = 0$, gives $p = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Likelihood function, example continue

Example

- Want to find the value of p that maximizes the likelihood function $L(p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$.
- It is easier to find the value of p that maximize $\ln(L(p))$.
- We have $\ln(L(p)) = n\bar{x}\ln(p) + n(1-\bar{x})\ln(1-p)$.
- Taking the derivatives,
  $\frac{d}{dp}(\ln(L(p))) = n\bar{x}\frac{1}{p} - n(1-\bar{x})\frac{1}{1-p} = 0$, gives
  $p = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
- The statistics $\hat{p} = \hat{p}(X_1, .., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ is called the maximum likelihood estimator of p.

Likelihood function, example continue

Example

- Want to find the value of p that maximizes the likelihood function $L(p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$.
- It is easier to find the value of p that maximize $\ln(L(p))$.
- We have $\ln(L(p)) = n\bar{x}\ln(p) + n(1-\bar{x})\ln(1-p)$.
- Taking the derivatives,
  $\frac{d}{dp}(\ln(L(p))) = n\bar{x}\frac{1}{p} - n(1-\bar{x})\frac{1}{1-p} = 0$, gives
  $p = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
- The statistics $\hat{p} = \hat{p}(X_1, .., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ is called the maximum likelihood estimator of p.
- $\hat{p}(x_1, ..., x_n)$ is the sample proportion of successes out of n trials.

Maximum Likelihood estimator

Example

- Suppose $X_1, ..., X_5 \in b(1, p)$ with $X_1 = 1$, $X_2 = 1$, $X_3 = 0$, $X_4 = 0$, $X_5 = 1$.
- Then $\hat{p}(1, 1, 0, 0, 1) = \bar{x} = \frac{1}{5}(1 + 1 + 0 + 0 + 1) = \frac{3}{5}$ is an estimate of p.

Maximum Likelihood estimator

Definition

Given the likelihood function $L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$, for $\theta \in \Omega$. The maximum likelihood estimator (MLE) of θ is

$$\hat{\theta} = \hat{\theta}(X_1, ..., X_n) = \arg\max_{\theta}(L(\theta))$$

which maximizes $L(\theta)$ in Ω.
The observed value of this statistics, $\hat{\theta}(x_1, ..., x_n)$, is called a maximum likelihood estimate.

- The maximum likelihood estimator, $\hat{\theta}$, is a choice of estimator for the parameter, θ, that is most likely to have produced the sample values $x_1, ..., x_n$.

Maximum likelihood estimator

The Maximum Likelihood estimators have the following property:

- If $\hat{\theta}$ is a maximum likelihood estimator of θ, then $h(\hat{\theta})$ is a maximum likelihood estimator of $h(\theta)$. For example, if $\hat{\theta}$ is a maximum likelihood estimator of the variance, θ, then $\sqrt{\hat{\theta}}$ is a maximum likelihood estimator of the standard deviation, $\sqrt{\theta}$.
- The unbiased estimators does not have this property.
- We will consider more MLE properties in chapter 3.5.