Topic 12

Overview of Estimation

Classical Statistics

Outline

Introduction

In the simplest possible terms, the goal of estimation theory is to answer the question:

What is that number?

Statistics has provided two distinct approaches this question - typically called

- classical or frequentist, and
- Bayesian.

Definition. A statistic is a function of the data that does not depend on any unknown parameter.

Exercise. Give a listing of statistics seen to this point.

Parameter Estimation

For parameter estimation, we consider $X = (X_1, \ldots, X_n)$, independent random variables chosen according to one of a family of probabilities P_θ where θ is element from the parameter space Θ. Based on our analysis, we choose an estimator $\hat{\theta}(X)$. If the data $\mathbf{x}$ takes on the values $x_1, x_2, \ldots, x_n$, then

$$\hat{\theta}(x_1, x_2, \ldots, x_n)$$

is called the estimate of θ. Thus we have three closely related objects.

1. θ - the parameter, an element of the parameter space, is a number or a vector.
2. $\hat{\theta}(x_1, x_2, \ldots, x_n)$ - the estimate, is a number or a vector obtained by evaluating the estimator on the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.
3. $\hat{\theta}(X_1, \ldots, X_n)$ - the estimator, is a random variable. We will analyze the distribution of this random variable to decide how well it performs in estimating θ.

Parameter Estimation

For Bernoulli trials $X = (X_1, \ldots, X_n)$, we have

1. p, a single parameter, the probability of success, with parameter space $[0, 1]$.
2. $\hat{p}(x_1, \ldots, x_n)$ is the sample proportion of successes in the data set.
3. $\hat{p}(X_1, \ldots, X_n)$, the sample mean of the random variables

$$\hat{p}(X_1, \ldots, X_n) = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}S_n$$

is an estimator of p. We can give the distribution of this estimator because S_n is a binomial random variable.

Parameter Estimation

Given pairs of observations $(\mathbf{x}, \mathbf{y}) = ((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$ that display a general linear pattern, we use ordinary least squares regression for

1. parameters - the slope β and intercept α of the regression line. So, the parameter space is $\mathbb{R}^2$, pairs of real numbers.
2. They are estimated using the statistics $\hat{\beta}$ and $\hat{\alpha}$ in the equations

$$\hat{\beta}(\mathbf{x}, \mathbf{y}) = \frac{cov(\mathbf{x}, \mathbf{y})}{var(\mathbf{x})}, \quad \bar{y} = \hat{\alpha}(\mathbf{x}, \mathbf{y}) + \hat{\beta}(\mathbf{x}, \mathbf{y})\bar{x}.$$

3. Later, when we consider statistical inference for linear regression, we will analyze the distribution of the estimators.

Classical Statistics

In classical statistics, the state of nature is assumed to be fixed, but unknown to us. Thus, one goal of estimation is to determine which of the P_θ is the source of the data. The estimate is a statistic

$$\hat{\theta} : \text{data} \to \Theta.$$

For estimation procedures, the classical approach to statistics is based on two fundamental questions:

- How do we determine estimators?
- How do we evaluate estimators?
  - Does this estimator in any way systematically under or over estimate the parameter?
  - Does it has large or small variance?
  - How does it compare to a notion of best possible estimator?
  - How easy is it to determine and to compute?
  - How does the procedure improve with increased sample size?

Densities and Likelihoods

Our analysis is based on the distribution of the random variables that underlie the data under any value θ. For each $\theta \in \Theta$, we have a density function

$$\mathbf{f}_X(\mathbf{x}|\theta).$$

For experimental designs based on a simple random sample, the observations $X_1, \ldots, X_n$, are drawn from a family of distributions each having density $f_X(x|\theta)$. For independent random variables, the joint density is the product of the marginal densities

$$\mathbf{f}_X(\mathbf{x}|\theta) = \prod_{k=1}^{n} f_X(x_k|\theta) = f_X(x_1|\theta) f_X(x_2|\theta) \cdots f_X(x_n|\theta).$$

In this circumstance, the data $\mathbf{x}$ are known and the parameter θ is unknown. Thus, we write the density function as

$$L(\theta|\mathbf{x}) = \mathbf{f}_X(\mathbf{x}|\theta)$$

and call L the likelihood function.

Densities and Likelihoods

- For Bernoulli trials with a known number of trials n but unknown success probability parameter p has joint density

$$\begin{aligned} \mathbf{f}_X(\mathbf{x}|p) &= p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2}\cdots p^{x_n}(1-p)^{1-x_n} = p^{\sum_{k=1}^n x_k}(1-p)^{\sum_{k=1}^n (1-x_k)} \\ &= p^{\sum_{k=1}^n x_k}(1-p)^{n-\sum_{k=1}^n x_k} = p^{n\bar{x}}(1-p)^{n(1-\bar{x})} \end{aligned}$$

- Normal random variables unknown mean μ and standard deviation σ has joint density

$$\begin{aligned} \mathbf{f}_X(\mathbf{x}|\mu,\sigma) &= \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2}\right)\cdots\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x_n-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{(\sigma\sqrt{2\pi})^n}\exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^n (x_k-\mu)^2\right) \end{aligned}$$

Exercise. Find the joint density of n independent $\Gamma(\alpha,\beta)$ random variables.