Estimation*

Ewa Paszek

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 2.0†

Abstract

This course is a short series of lectures on Introductory Statistics. Topics covered are listed in the Table of Contents. The notes were prepared by Ewa Paszek and Marek Kimmel. The development of this course has been supported by NSF 0203396 grant.

1 ESTIMATION

Once a model is specified with its parameters and data have been collected, one is in a position to evaluate the model's goodness of fit, that is, how well the model fits the observed pattern of data. Finding parameter values of a model that best fits the data — **a procedure called parameter estimation, which assesses goodness of fit**.

There are two generally accepted methods of parameter estimation. They are **least squares estimation (LSE)** and **maximum likelihood estimation (MLE)**. The former is well known as linear regression, the sum of squares error, and the root means squared deviation is tied to the method. On the other hand, MLE is not widely recognized among modelers in psychology, though it is, by far, the most commonly used method of parameter estimation in the statistics community. LSE might be useful for obtaining a descriptive measure for the purpose of summarizing observed data, but MLE is more suitable for statistical inference such as model comparison. LSE has no basis for constructing confidence intervals or testing hypotheses whereas both are naturally built into MLE.

1.1 Properties of Estimators

UNBIASED AND BIASED ESTIMATORS

Let consider random variables for which the functional form of the p.d.f. is know, but the distribution depends on an unknown parameter θ , that may have any value in a set θ , which is called the **parameter space**. In estimation the random sample from the distribution is taken to elicit some information about the unknown parameter θ. The experiment is repeated n independent times, the sample $X_1, X_2, ..., X_n$ is observed and one try to guess the value of θ using the observations $x_1, x_2, ...x_n$.

The function of $X_1, X_2, ..., X_n$ used to guess θ**is called an estimator of** θ . We want it to be such that the computed estimate $u(x_1, x_2, ...x_n)$ is usually close to θ. Let $Y = u(x_1, x_2, ...x_n)$ be an estimator of θ. If Y to be a good estimator of θ , a very desirable property is that **it means be equal to** θ , namely $E(Y) = \theta$.

Definition 1:

If $E[u(x_1, x_2, ..., x_n)] = \theta$ is called **an unbiased estimator of** θ. Otherwise, it is said to be **biased**.

*Version 1.2: Oct 8, 2007 3:23 pm -0500

†http://creativecommons.org/licenses/by/2.0/

It is required not only that an estimator has expectation equal to θ, but also the variance of the estimator should be as small as possible. If there are two unbiased estimators of θ, it could be probably possible to choose the one with the smaller variance. In general, with a random sample $X_1, X_2, ..., X_n$ of a fixed sample size **n**, a statistician might like to find the estimator $Y = u(X_1, X_2, ..., X_n)$ of an unknown parameter θ which minimizes the mean (expected) value of the square error (difference) $Y - \theta$ that is, minimizes

$$E\left[(Y-\theta)^2\right] = E\{[u(X_1, X_2, ..., X_n) - \theta]^2\}.$$

The statistic **Y** that minimizes $E\left[(Y-\theta)^2\right]$ is the one with minimum mean square error. If we restrict our attention to unbiased estimators only, then

$$Var(Y) = E\left[(Y-\theta)^2\right],$$

and the unbiased statistics **Y** that minimizes this expression is said to be **the unbiased minimum variance estimator of** θ .

1.2 Method of Moments

One of the oldest procedures for estimating parameters is **the method of moments**. Another method for finding an estimator of an unknown parameter is called **the method of maximum likelihood**. In general, in the method of moments, if there are **k** parameters that have to be estimated, the first **k** sample moments are set equal to the first **k** population moments that are given in terms of the unknown parameters.

Example 1
Let the distribution of **X** be $N(\mu, \sigma^2)$. Then $E(X) = \mu$ and $E(X^2) = \sigma^2 + \mu^2$. Given a random sample of size **n**, the first two moments are given by

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

We set $m_1 = E(X)$ and $m_2 = E(X^2)$ and solve for μ and σ^2,

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \mu$$

and

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \sigma^2 + \mu^2.$$

The first equation yields $\overline{x}$ as the estimate of μ . Replacing μ^2 with $\overline{x}^2$ in the second equation and solving for σ^2 ,
we obtain

$$\frac{1}{n}\sum_{i=1}^{n} x_i - \overline{x}^2 = v$$

for the solution of σ^2 .

Thus the method of moment estimators for μ and σ^2 are $\tilde{\mu} = \overline{X}$ and $\tilde{\sigma}^2 = V$. Of course, $\tilde{\mu} = \overline{X}$ is unbiased whereas $\tilde{\sigma}^2 = V$. is biased.

1.3

At this stage arises the question, which of two different estimators $\hat{\theta}$ and $\tilde{\theta}$, for a parameter θ one should use. Most statistician select he one that has the smallest mean square error, for example,

$$E\left[\left(\hat{\theta}-\theta\right)^2\right] < E\left[\left(\tilde{\theta}-\theta\right)^2\right],$$

then $\hat{\theta}$ seems to be preferred. This means that if $E\left(\hat{\theta}\right) = E\left(\tilde{\theta}\right) = \theta$, then one would select the one with the smallest variance.

Next, other questions should be considered. Namely, given an estimate for a parameter, how accurate is the estimate? How confident one is about the closeness of the estimate to the unknown parameter?

NOTE: CONFIDENCE INTERVALS I[1] and CONFIDENCE INTERVALS II[2]

[1] "CONFIDENCE INTERVALS I": Section CONFIDENCE INTERVALS I <http://cnx.org/content/m13494/latest/#sec_1>

[2] "CONFIDENCE INTERVALS II": Section CONFIDENCE INTERVALS II <http://cnx.org/content/m13496/latest/#sec_1>