

Estimator

In statistics, an **estimator** is a rule for calculating an estimate of a given quantity based on observed data: thus the rule (the estimator), the quantity of interest (the estimand) and its result (the estimate) are distinguished.^[1]

There are point and interval estimators. The point estimators yield single-valued results, although this includes the possibility of single vector-valued results and results that can be expressed as a single function. This is in contrast to an interval estimator, where the result would be a range of plausible values (or vectors or functions).

Estimation theory is concerned with the properties of estimators; that is, with defining properties that can be used to compare different estimators (different rules for creating estimates) for the same quantity, based on the same data. Such properties can be used to determine the best rules to use under given circumstances. However, in robust statistics, statistical theory goes on to consider the balance between having good properties, if tightly defined assumptions hold, and having less good properties that hold under wider conditions.

Contents

Background

Definition

Quantified properties

- Error
- Mean squared error
- Sampling deviation
- Variance
- Bias
- Relationships among the quantities

Behavioral properties

- Consistency
- Asymptotic normality
- Efficiency
- Robustness

See also

Notes

References

External links

Background

An "estimator" or "point estimate" is a statistic (that is, a function of the data) that is used to infer the value of an unknown parameter in a statistical model. The parameter being estimated is sometimes called the *estimand*. It can be either finite-dimensional (in parametric and semi-parametric models), or infinite-dimensional (semi-parametric and non-parametric models).^[2] If the parameter is denoted

θ then the estimator is traditionally written by adding a circumflex over the symbol: $\hat{\theta}$. Being a function of the data, the estimator is itself a random variable; a particular realization of this random variable is called the "estimate". Sometimes the words "estimator" and "estimate" are used interchangeably.

The definition places virtually no restrictions on which functions of the data can be called the "estimators". The attractiveness of different estimators can be judged by looking at their properties, such as unbiasedness, mean square error, consistency, asymptotic distribution, etc. The construction and comparison of estimators are the subjects of the estimation theory. In the context of decision theory, an estimator is a type of decision rule, and its performance may be evaluated through the use of loss functions.

When the word "estimator" is used without a qualifier, it usually refers to point estimation. The estimate in this case is a single point in the parameter space. There also exists another type of estimator: interval estimators, where the estimates are subsets of the parameter space.

The problem of density estimation arises in two applications. Firstly, in estimating the probability density functions of random variables and secondly in estimating the spectral density function of a time series. In these problems the estimates are functions that can be thought of as point estimates in an infinite dimensional space, and there are corresponding interval estimation problems.

Definition

Suppose a fixed *parameter* θ needs to be estimated. Then an "estimator" is a function that maps the sample space to a set of *sample estimates*. An estimator of θ is usually denoted by the symbol $\hat{\theta}$. It is often convenient to express the theory using the algebra of random variables: thus if X is used to denote a random variable corresponding to the observed data, the estimator (itself treated as a random variable) is symbolised as a function of that random variable, $\hat{\theta}(X)$. The estimate for a particular observed data value x (i.e. for $X = x$) is then $\hat{\theta}(x)$, which is a fixed value. Often an abbreviated notation is used in which $\hat{\theta}$ is interpreted directly as a random variable, but this can cause confusion.

Quantified properties

The following definitions and attributes are relevant.^[3]

Error

For a given sample x , the "error" of the estimator $\hat{\theta}$ is defined as

$$e(x) = \hat{\theta}(x) - \theta,$$

where θ is the parameter being estimated. The error, e , depends not only on the estimator (the estimation formula or procedure), but also on the sample.

Mean squared error

The mean squared error of $\hat{\theta}$ is defined as the expected value (probability-weighted average, over all samples) of the squared errors; that is,

$$\text{MSE}(\hat{\theta}) = \mathbf{E}[(\hat{\theta}(X) - \theta)^2].$$

It is used to indicate how far, on average, the collection of estimates are from the single parameter being estimated. Consider the following analogy. Suppose the parameter is the bull's-eye of a target, the estimator is the process of shooting arrows at the target, and the individual arrows are estimates (samples). Then high MSE means the average distance of the arrows from the bull's-eye is high, and low MSE means the average distance from the bull's-eye is low. The arrows may or may not be clustered. For example, even if all arrows hit the same point, yet grossly miss the target, the MSE is still relatively large. However, if the MSE is relatively low then the arrows are likely more highly clustered (than highly dispersed) around the target.

Sampling deviation

For a given sample x , the *sampling deviation* of the estimator $\hat{\theta}$ is defined as

$$d(x) = \hat{\theta}(x) - \mathbf{E}(\hat{\theta}(X)) = \hat{\theta}(x) - \mathbf{E}(\hat{\theta}),$$

where $\mathbf{E}(\hat{\theta}(X))$ is the expected value of the estimator. The sampling deviation, d , depends not only on the estimator, but also on the sample.

Variance

The *variance* of $\hat{\theta}$ is simply the expected value of the squared sampling deviations; that is, $\text{var}(\hat{\theta}) = \mathbf{E}[(\hat{\theta} - \mathbf{E}[\hat{\theta}])^2]$. It is used to indicate how far, on average, the collection of estimates are from the *expected value* of the estimates. (Note the difference between MSE and variance.) If the parameter is the bull's-eye of a target, and the arrows are estimates, then a relatively high variance means the arrows are dispersed, and a relatively low variance means the arrows are clustered. Even if the variance is low, the cluster of arrows may still be far off-target, and even if the variance is high, the diffuse collection of arrows may still be unbiased. Finally, even if all arrows grossly miss the target, if they nevertheless all hit the same point, the variance is zero.

Bias

The *bias* of $\hat{\theta}$ is defined as $B(\hat{\theta}) = \mathbf{E}(\hat{\theta}) - \theta$. It is the distance between the average of the collection of estimates, and the single parameter being estimated. The bias of $\hat{\theta}$ is a function of the true value of θ so saying that the bias of $\hat{\theta}$ is b means that for every θ the bias of $\hat{\theta}$ is b .

The bias also is the expected value of the error, since $\mathbf{E}(\hat{\theta}) - \theta = \mathbf{E}(\hat{\theta} - \theta)$. If the parameter is the bull's-eye of a target, and the arrows are estimates, then a relatively high absolute value for the bias means the average position of the arrows is off-target, and a relatively low absolute bias means the average position of the arrows is on target. They may be dispersed, or may be clustered. The relationship between bias and variance is analogous to the relationship between accuracy and precision.

The estimator $\hat{\theta}$ is an *unbiased estimator* of θ if and only if $B(\hat{\theta}) = 0$. Bias is a property of the estimator, not of the estimate. Often, people refer to a "biased estimate" or an "unbiased estimate," but they really are talking about an "estimate from a biased estimator," or an "estimate from an unbiased estimator." Also, people often confuse the "error" of a single estimate with the "bias" of an

estimator. That the error for one estimate is large, does not mean the estimator is biased. In fact, even if all estimates have astronomical absolute values for their errors, if the expected value of the error is zero, the estimator is unbiased. Also, an estimator's being biased does not preclude the error of an estimate from being zero in a particular instance. The ideal situation is to have an unbiased estimator with low variance, and also try to limit the number of samples where the error is extreme (that is, have few outliers). Yet unbiasedness is not essential. Often, if just a little bias is permitted, then an estimator can be found with lower MSE and/or fewer outlier sample estimates.

An alternative to the version of "unbiased" above, is "median-unbiased", where the median of the distribution of estimates agrees with the true value; thus, in the long run half the estimates will be too low and half too high. While this applies immediately only to scalar-valued estimators, it can be extended to any measure of central tendency of a distribution: see median-unbiased estimators.

Relationships among the quantities

- The MSE, variance, and bias, are related: $\text{MSE}(\hat{\theta}) = \text{var}(\hat{\theta}) + (B(\hat{\theta}))^2$, i.e. mean squared error = variance + square of bias. In particular, for an unbiased estimator, the variance equals the MSE.
- The standard deviation of an estimator $\hat{\theta}$ of θ (the square root of the variance), or an estimate of the standard deviation of an estimator $\hat{\theta}$ of θ , is called the standard error of $\hat{\theta}$.

Behavioral properties

Consistency

A consistent sequence of estimators is a sequence of estimators that converge in probability to the quantity being estimated as the index (usually the sample size) grows without bound. In other words, increasing the sample size increases the probability of the estimator being close to the population parameter.

Mathematically, a sequence of estimators $\{t_n; n \geq 0\}$ is a consistent estimator for parameter θ if and only if, for all $\epsilon > 0$, no matter how small, we have

$$\lim_{n \rightarrow \infty} \Pr \{|t_n - \theta| < \epsilon\} = 1.$$

The consistency defined above may be called weak consistency. The sequence is *strongly consistent*, if it converges almost surely to the true value.

An estimator that converges to a *multiple* of a parameter can be made into a consistent estimator by multiplying the estimator by a scale factor, namely the true value divided by the asymptotic value of the estimator. This occurs frequently in estimation of scale parameters by measures of statistical dispersion.

Asymptotic normality

An asymptotically normal estimator is a consistent estimator whose distribution around the true parameter θ approaches a normal distribution with standard deviation shrinking in proportion to $1/\sqrt{n}$ as the sample size n grows. Using \rightarrow to denote convergence in distribution, t_n is asymptotically normal if

$$\sqrt{n}(t_n - \theta) \xrightarrow{D} N(0, V),$$

for some V .

In this formulation V/n can be called the *asymptotic variance* of the estimator. However, some authors also call V the *asymptotic variance*. Note that convergence will not necessarily have occurred for any finite " n ", therefore this value is only an approximation to the true variance of the estimator, while in the limit the asymptotic variance (V/n) is simply zero. To be more specific, the distribution of the estimator t_n converges weakly to a dirac delta function centered at θ .

The central limit theorem implies asymptotic normality of the sample mean \bar{X} as an estimator of the true mean. More generally, maximum likelihood estimators are asymptotically normal under fairly weak regularity conditions — see the asymptotics section of the maximum likelihood article. However, not all estimators are asymptotically normal; the simplest examples are found when the true value of a parameter lies on the boundary of the allowable parameter region.

Efficiency

Two naturally desirable properties of estimators are for them to be unbiased and have minimal mean squared error (MSE). These cannot in general both be satisfied simultaneously: a biased estimator may have lower mean squared error (MSE) than any unbiased estimator; see estimator bias.

Among unbiased estimators, there often exists one with the lowest variance, called the minimum variance unbiased estimator (MVUE). In some cases an unbiased efficient estimator exists, which, in addition to having the lowest variance among unbiased estimators, satisfies the Cramér–Rao bound, which is an absolute lower bound on variance for statistics of a variable.

Concerning such "best unbiased estimators", see also Cramér–Rao bound, Gauss–Markov theorem, Lehmann–Scheffé theorem, Rao–Blackwell theorem.

Robustness

See also

- Best linear unbiased estimator (BLUE)
- Invariant estimator
- Kalman filter
- Markov chain Monte Carlo (MCMC)
- Maximum a posteriori (MAP)
- Method of moments, generalized method of moments
- Minimum mean squared error (MMSE)
- Particle filter
- Pitman closeness criterion
- Sensitivity and specificity
- Shrinkage estimator
- Signal Processing
- Testimator
- Wiener filter
- Well-behaved statistic