



WIKIPEDIA
The Free Encyclopedia

[Main page](#)

[Contents](#)

[Featured content](#)

[Current events](#)

[Random article](#)

[Donate to Wikipedia](#)

[Wikipedia store](#)

Interaction

[Help](#)

[About Wikipedia](#)

[Community portal](#)

[Recent changes](#)

[Contact page](#)

Tools

[What links here](#)

[Related changes](#)

[Upload file](#)

[Special pages](#)

[Permanent link](#)

[Page information](#)

[Wikidata item](#)

[Cite this page](#)

Print/export

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

Languages

[العربية](#)

[Español](#)

[Euskara](#)

[فارسی](#)

[한국어](#)

[Italiano](#)

[עברית](#)

[Magyar](#)

[Norsk bokmål](#)

[Polski](#)

[Português](#)

[Українська](#)

[中文](#)

[Edit links](#)

Article [Talk](#)

[Read](#)

[Edit](#)

[View history](#)



Estimation theory

From Wikipedia, the free encyclopedia

For other uses, see [Estimation \(disambiguation\)](#).

"Parameter estimation" redirects here. It is not to be confused with [Point estimation](#) or [Interval estimation](#).

Estimation theory is a branch of [statistics](#) that deals with estimating the values of parameters based on measured/empirical data that has a random component. The parameters describe an underlying physical setting in such a way that their value affects the distribution of the measured data. An [estimator](#) attempts to approximate the unknown parameters using the measurements.

For example, it is desired to estimate the proportion of a population of voters who will vote for a particular candidate. That proportion is the parameter sought; the estimate is based on a small random sample of voters.

Or, for example, in [radar](#) the goal is to estimate the range of objects (airplanes, boats, etc.) by analyzing the two-way transit timing of received echoes of transmitted pulses. Since the reflected pulses are unavoidably embedded in electrical noise, their measured values are randomly distributed, so that the transit time must be estimated.

In estimation theory, two approaches are generally considered. ^[1]

- The probabilistic approach (described in this article) assumes that the measured data is random with [probability distribution](#) dependent on the parameters of interest
- The [set-membership approach](#) assumes that the measured data vector belongs to a set which depends on the parameter vector.

For example, in electrical communication theory, the measurements which contain information regarding the parameters of interest are often associated with a [noisy signal](#). Without randomness, or noise, the problem would be [deterministic](#) and estimation would not be needed.

Contents [hide]

- Basics
- Estimators
- Examples
 - Unknown constant in additive white Gaussian noise
 - Maximum likelihood
 - Cramér–Rao lower bound
 - Maximum of a uniform distribution
- Applications
- See also
- Notes
- References
- References

Basics ^[edit]

To build a model, several statistical "ingredients" need to be known. These are needed to ensure the estimator has some mathematical tractability.

The first is a set of [statistical samples](#) taken from a [random vector](#) (RV) of size *N*. Put into a [vector](#),

$$\mathbf{x} = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}.$$

Secondly, there are the corresponding *M* parameters

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{bmatrix},$$

which need to be established with their continuous [probability density function](#) (pdf) or its discrete counterpart, the [probability mass function](#) (pmf)

$$p(\mathbf{x}|\boldsymbol{\theta}).$$

It is also possible for the parameters themselves to have a probability distribution (e.g., [Bayesian statistics](#)). It is then necessary to define the [Bayesian probability](#)

$$\pi(\boldsymbol{\theta}).$$

After the model is formed, the goal is to estimate the parameters, commonly denoted $\hat{\boldsymbol{\theta}}$, where the "hat" indicates the estimate.

One common estimator is the [minimum mean squared error](#) estimator, which utilizes the error between the estimated parameters and the actual value of the parameters

$$\mathbf{e} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$$

as the basis for optimality. This error term is then squared and minimized for the MMSE estimator.

Estimators [\[edit\]](#)

Main article: [Estimator](#)

Commonly used estimators and estimation methods, and topics related to them:

- [Maximum likelihood](#) estimators
- [Bayes estimators](#)
- [Method of moments](#) estimators
- [Cramér–Rao bound](#)
- [Minimum mean squared error](#) (MMSE), also known as Bayes least squared error (BLSE)
- [Maximum a posteriori](#) (MAP)
- [Minimum variance unbiased estimator](#) (MVUE)
- [nonlinear system identification](#)
- [Best linear unbiased estimator](#) (BLUE)
- Unbiased estimators — see [estimator bias](#).
- [Particle filter](#)
- [Markov chain Monte Carlo](#) (MCMC)
- [Kalman filter](#), and its various derivatives
- [Wiener filter](#)

Examples [\[edit\]](#)

Unknown constant in additive white Gaussian noise [\[edit\]](#)

Consider a received [discrete signal](#), $x[n]$, of N [independent samples](#) that consists of an unknown constant A with [additive white Gaussian noise](#) (AWGN) $w[n]$ with known [variance](#) σ^2 (i.e., $\mathcal{N}(0, \sigma^2)$). Since the variance is known then the only unknown parameter is A .

The model for the signal is then

$$x[n] = A + w[n] \quad n = 0, 1, \dots, N - 1$$

Two possible (of many) estimators are:

- $\hat{A}_1 = x[0]$
- $\hat{A}_2 = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$ which is the [sample mean](#)

Both of these estimators have a [mean](#) of A , which can be shown through taking the [expected value](#) of each estimator

$$\mathbf{E} \left[\hat{A}_1 \right] = \mathbf{E} [x[0]] = A$$

and

$$\mathbb{E}[\hat{A}_2] = \mathbb{E}\left[\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right] = \frac{1}{N} \left[\sum_{n=0}^{N-1} \mathbb{E}[x[n]] \right] = \frac{1}{N} [NA] = A$$

At this point, these two estimators would appear to perform the same. However, the difference between them becomes apparent when comparing the variances.

$$\text{var}(\hat{A}_1) = \text{var}(x[0]) = \sigma^2$$

and

$$\text{var}(\hat{A}_2) = \text{var}\left(\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right) \stackrel{\text{independence}}{=} \frac{1}{N^2} \left[\sum_{n=0}^{N-1} \text{var}(x[n]) \right] = \frac{1}{N^2} [N\sigma^2] = \frac{\sigma^2}{N}$$

It would seem that the sample mean is a better estimator since its variance is lower for every $N > 1$.

Maximum likelihood [\[edit\]](#)

Main article: [Maximum likelihood](#)

Continuing the example using the [maximum likelihood](#) estimator, the [probability density function](#) (pdf) of the noise for one sample $w[n]$ is

$$p(w[n]) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} w[n]^2\right)$$

and the probability of $x[n]$ becomes ($x[n]$ can be thought of a $\mathcal{N}(A, \sigma^2)$)

$$p(x[n]; A) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (x[n] - A)^2\right)$$

By independence, the probability of \mathbf{x} becomes

$$p(\mathbf{x}; A) = \prod_{n=0}^{N-1} p(x[n]; A) = \frac{1}{(\sigma\sqrt{2\pi})^N} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2\right)$$

Taking the [natural logarithm](#) of the pdf

$$\ln p(\mathbf{x}; A) = -N \ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2$$

and the maximum likelihood estimator is

$$\hat{A} = \arg \max \ln p(\mathbf{x}; A)$$

Taking the first [derivative](#) of the log-likelihood function

$$\frac{\partial}{\partial A} \ln p(\mathbf{x}; A) = \frac{1}{\sigma^2} \left[\sum_{n=0}^{N-1} (x[n] - A) \right] = \frac{1}{\sigma^2} \left[\sum_{n=0}^{N-1} x[n] - NA \right]$$

and setting it to zero

$$0 = \frac{1}{\sigma^2} \left[\sum_{n=0}^{N-1} x[n] - NA \right] = \sum_{n=0}^{N-1} x[n] - NA$$

This results in the maximum likelihood estimator

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

which is simply the sample mean. From this example, it was found that the sample mean is the maximum likelihood estimator for N samples of a fixed, unknown parameter corrupted by AWGN.

Cramér–Rao lower bound [\[edit\]](#)

For more details on this topic, see [Cramér–Rao bound](#).

To find the [Cramér–Rao lower bound](#) (CRLB) of the sample mean estimator, it is first necessary to find the [Fisher information](#) number

$$\mathcal{I}(A) = \mathbb{E} \left(\left[\frac{\partial}{\partial A} \ln p(\mathbf{x}; A) \right]^2 \right) = -\mathbb{E} \left[\frac{\partial^2}{\partial A^2} \ln p(\mathbf{x}; A) \right]$$

and copying from above

$$\frac{\partial}{\partial A} \ln p(\mathbf{x}; A) = \frac{1}{\sigma^2} \left[\sum_{n=0}^{N-1} x[n] - NA \right]$$

Taking the second derivative

$$\frac{\partial^2}{\partial A^2} \ln p(\mathbf{x}; A) = \frac{1}{\sigma^2} (-N) = \frac{-N}{\sigma^2}$$

and finding the negative expected value is trivial since it is now a deterministic constant

$$-\mathbb{E} \left[\frac{\partial^2}{\partial A^2} \ln p(\mathbf{x}; A) \right] = \frac{N}{\sigma^2}$$

Finally, putting the Fisher information into

$$\text{var}(\hat{A}) \geq \frac{1}{\mathcal{I}}$$

results in

$$\text{var}(\hat{A}) \geq \frac{\sigma^2}{N}$$

Comparing this to the variance of the sample mean (determined previously) shows that the sample mean is *equal to* the Cramér–Rao lower bound for all values of N and A . In other words, the sample mean is the (necessarily unique) [efficient estimator](#), and thus also the [minimum variance unbiased estimator](#) (MVUE), in addition to being the [maximum likelihood](#) estimator.

Maximum of a uniform distribution [\[edit\]](#)

Main article: [German tank problem](#)

One of the simplest non-trivial examples of estimation is the estimation of the maximum of a uniform distribution. It is used as a hands-on classroom exercise and to illustrate basic principles of estimation theory. Further, in the case of estimation based on a single sample, it demonstrates philosophical issues and possible misunderstandings in the use of [maximum likelihood](#) estimators and [likelihood functions](#).

Given a [discrete uniform distribution](#) $1, 2, \dots, N$ with unknown maximum, the [UMVU](#) estimator for the maximum is given by

$$\frac{k+1}{k}m - 1 = m + \frac{m}{k} - 1$$

where m is the [sample maximum](#) and k is the [sample size](#), sampling without replacement.^{[\[2\]](#)[\[3\]](#)} This problem is commonly known as the [German tank problem](#), due to application of maximum estimation to estimates of German tank production during [World War II](#).

The formula may be understood intuitively as:

"The sample maximum plus the average gap between observations in the sample",

the gap being added to compensate for the negative bias of the sample maximum as an estimator for the population maximum.^{[\[note 1\]](#)}

This has a variance of^{[\[2\]](#)}

$$\frac{1}{k} \frac{(N-k)(N+1)}{(k+2)} \approx \frac{N^2}{k^2} \text{ for small samples } k \ll N$$

so a standard deviation of approximately N/k , the (population) average size of a gap between samples; compare $\frac{m}{k}$ above. This can be seen as a very simple case of [maximum spacing estimation](#).

The sample maximum is the [maximum likelihood](#) estimator for the population maximum, but, as discussed above, it is biased.

Applications [\[edit\]](#)

Numerous fields require the use of estimation theory. Some of these fields include (but are by no means limited to):

- Interpretation of scientific [experiments](#)
- Signal processing
- Clinical trials
- Opinion polls
- Quality control
- Telecommunications
- Project management
- Software engineering
- Control theory (in particular [Adaptive control](#))
- Network intrusion detection system
- Orbit determination

Measured data are likely to be subject to [noise](#) or uncertainty and it is through statistical [probability](#) that [optimal](#) solutions are sought to extract as much [information](#) from the data as possible.

See also [\[edit\]](#)

- [Category:Estimation theory](#)
- [Category:Estimation for specific distributions](#)
- [Best linear unbiased estimator \(BLUE\)](#)
- [Chebyshev center](#)
- [Completeness \(statistics\)](#)
- [Cramér–Rao bound](#)
- [Detection theory](#)
- [Efficiency \(statistics\)](#)
- [Estimator, Estimator bias](#)
- [Expectation-maximization algorithm \(EM algorithm\)](#)
- [Fermi problem](#)
- [Grey box model](#)
- [Information theory](#)
- [Kalman filter](#)
- [Least-squares spectral analysis](#)
- [Markov chain Monte Carlo \(MCMC\)](#)
- [Matched filter](#)
- [Maximum a posteriori \(MAP\)](#)
- [Maximum likelihood](#)
- [Maximum entropy spectral estimation](#)
- [Method of moments, generalized method of moments](#)
- [Minimum mean squared error \(MMSE\)](#)
- [Minimum variance unbiased estimator \(MVUE\)](#)
- [Nonlinear system identification](#)
- [Nuisance parameter](#)
- [Parametric equation](#)
- [Particle filter](#)
- [Rao–Blackwell theorem](#)
- [Spectral density, Spectral density estimation](#)
- [Statistical signal processing](#)
- [Sufficiency \(statistics\)](#)
- [Wiener filter](#)

Notes [\[edit\]](#)

- [^] The sample maximum is never more than the population maximum, but can be less, hence it is a [biased estimator](#): it will tend to *underestimate* the population maximum.

References [\[edit\]](#)

- [^] Walter, E.; Pronzato, L. (1997). *Identification of Parametric Models from Experimental Data*. London, UK: Springer-Verlag.
- [^] ^a ^b Johnson, Roger (1994), "Estimating the Size of a Population", *Teaching Statistics* [↗](#) **16** (2 (Summer)): 50,

doi:10.1111/j.1467-9639.1994.tb00688.x

3. ^ Johnson, Roger (2006), "Estimating the Size of a Population" (PDF), *Getting the Best from Teaching Statistics*

References [edit]

- *Theory of Point Estimation* by E.L. Lehmann and G. Casella. (ISBN 0387985026)
- *Systems Cost Engineering* by Dale Shermon. (ISBN 978-0-566-08861-2)
- *Mathematical Statistics and Data Analysis* by John Rice. (ISBN 0-534-209343)
- *Fundamentals of Statistical Signal Processing: Estimation Theory* by Steven M. Kay (ISBN 0-13-345711-7)
- *An Introduction to Signal Detection and Estimation* by H. Vincent Poor (ISBN 0-387-94173-8)
- *Detection, Estimation, and Modulation Theory, Part 1* by Harry L. Van Trees (ISBN 0-471-09517-6; website)
- *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches* by Dan Simon website
- Ali H. Sayed, *Adaptive Filters*, Wiley, NJ, 2008, ISBN 978-0-470-25388-5.
- Ali H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, NJ, 2003, ISBN 0-471-46126-1.
- Thomas Kailath, Ali H. Sayed, and Babak Hassibi, *Linear Estimation*, Prentice-Hall, NJ, 2000, ISBN 978-0-13-022464-4.
- Babak Hassibi, Ali H. Sayed, and Thomas Kailath, *Indefinite Quadratic Estimation and Control: A Unified Approach to H2 and Hoo Theories*, Society for Industrial & Applied Mathematics (SIAM), PA, 1999, ISBN 978-0-89871-411-1.
- V.G.Voinov, M.S.Nikulin, "Unbiased estimators and their applications. Vol.1: Univariate case", Kluwer Academic Publishers, 1993, ISBN 0-7923-2382-3.
- V.G.Voinov, M.S.Nikulin, "Unbiased estimators and their applications. Vol.2: Multivariate case", Kluwer Academic Publishers, 1996, ISBN 0-7923-3939-8.

v · t · e	Digital signal processing		[hide]
Theory	Detection theory · Discrete signal · Estimation theory · Nyquist–Shannon sampling theorem		
Sub-fields	Audio signal processing · Digital image processing · Speech processing · Statistical signal processing		
Techniques	Advanced Z-transform · Bilinear transform · Discrete Fourier transform (DFT) ·		
	Discrete-time Fourier transform (DTFT) · Impulse invariance · Matched Z-transform method · Z-transform · Zak transform		
Sampling	Aliasing · Anti-aliasing filter · Downsampling · Nyquist rate / frequency · Oversampling · Quantization · Sampling rate · Undersampling · Upsampling		
Categories: Estimation theory Statistical inference Signal processing			

This page was last modified on 11 April 2015, at 06:35.

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

Privacy policy | About Wikipedia | Disclaimers | Contact Wikipedia | Developers | Mobile view

