Estimators

Definition 1 *A sample is a sequence of realizations of the same random variable. If the random variable we want to estimate is X, then a sequence of realizations $X_1, X_2, ..., X_k$ is a sample.*

Remark 1 *We are generally interested in the distribution of the sample. If we assume that that the sample has a famous distribution, for example it is normally distributed, then we are trying to find the parameters that describe the distribution, namely the expected value mu and the variance σ^2.*

Definition 2 *An estimator is a function of a sample. Usual notation: T(X). The estimator is constructed to estimate an unknown parameters ϑ.*

Example 1 *Examples for statistics: minimum, maximum, mean. Another example is the range, which is the difference between the largest and smallest values, denoted by $X_n^* - X_1^*$. The last notation comes from the order statistics, i.e., given random variables $X_1, \ldots, X_n$, the order statistics $X_1^*, \ldots, X_n^*$ are also random variables, defined by sorting the values (realizations) in increasing order.*

Definition 3 (unbiasedness) *Suppose that we estimate a real parameter ϑ by the statistics $T(\underline{X})$. The estimate is called unbiased, if $E_\vartheta(T(\underline{X})) = \vartheta$ for all ϑ.*

Example 2 *estimating the probabilities using relative frequencies; estimating the expected value using sample mean; estimating the parameter of the Poisson distribution using sample mean*

If the expected value (m) is known, the unbiased estimator of variance is $s_n^2 = \frac{(x_1-m)^2+\ldots+(x_n-m)^2}{n}$. Contrary to the known expected value case, to get an unbiased estimator for variance, the formula needs a correction: $s_n^{*2} = \frac{(x_1-\overline{x})^2+\ldots+(x_n-\overline{x})^2}{n-1}$, i.e., in the denominator stands $n-1$, and $\overline{x}$ denotes the sample mean.

Comparison of estimations:

Definition 4 (minimum-variance unbiased estimator - MVUE) *The best amongst unbiased estimators is the minimum-variance unbiased estimator T, which has the smallest variance. In formula, for all T' unbiased statistics, $Var_\vartheta(T(\underline{X})) \leq Var_\vartheta(T'(\underline{X}))$ holds for all possible ϑ parameters. (Remark: we are looking for MVUE estimators amongst only unbiased statistics.)*

Definition 5 *A sequence $T_n(\underline{X})$ of estimators for parameter ϑ is* ***consistent****, if $E_\vartheta(T_n(\underline{X})) \underset{n\to\infty}{\longrightarrow} \vartheta$ (asymptotically unbiased), furthermore, $Var_\vartheta(T_n(\underline{X})) \underset{n\to\infty}{\longrightarrow} 0$. In words, if the number of sample elements increases, the estimation becomes more accurate.*

Example 3 *relative frequency to estimate probability; mean to estimate expected value; sample empirical variance (remember: s_n^{*2}, and it is unbiased) to estimate variance.*

A statistic $t = T(X)$ is sufficient for underlying parameter θ if and only if the conditional probability distribution of the data X, given the statistic $t = T(X)$, does not depend on the parameter θ.

Example 4 *As an example, the sample mean is sufficient for the mean of a normal distribution with known variance. Once the sample mean is known, no further information about the mean can be obtained from the sample itself. On the other hand, for an arbitrary distribution the median is not sufficient statistics for the mean: even if the median of the sample is known, knowing the sample itself would provide further information about the population mean. For example, if the observations that are less than the median are only slightly less, but observations exceeding the median exceed it by a large amount, then this would have a bearing on one's inference about the population mean.*

Estimation methods

Maximum likelihood method

Example:

- Consider a series of independent experiments, each successful with probability p, which is unknown.

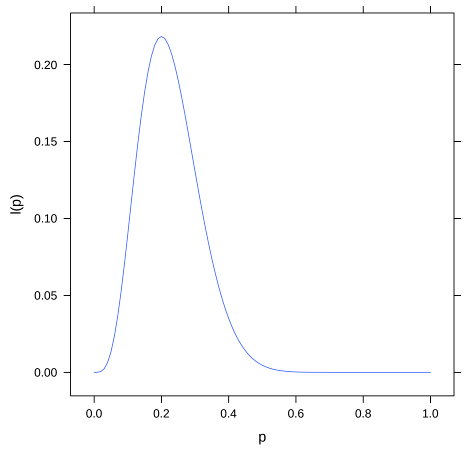

Figure 1. An example for likelihood function, when the sample size is $n = 20$, and the number of successful trials is $k = 4$.

- If n denotes the number of trials, our sample is an n-element Bernoulli sample: $X_1, \dots, X_n$.
- Assume that k trial were successful.
- $P\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n}{k} p^k (1-p)^{n-k}$ is called the likelihood function.
- The goal is to find p, and to do so, we will maximize the function in the parameter.

Definition 6 (Maximum Likelihood Method) *Let ϑ denote the vector of distribution parameters, and $\underline{x}$ the vector of sample. The* ***likelihood function*** *in discrete case is $L(\vartheta, \underline{x}) = \prod_{i=1}^{n} P_\vartheta(X_i = x_i)$, and in absolutely continuous case $L(\vartheta, \underline{x}) = \prod_{i=1}^{n} f_\vartheta(x_i)$. The* ***maximum likelihood estimator*** *of parameter ϑ is the value which maximizes the likelihood function. The method is the following.*

1. *Instead of maximizing the L function, we are looking for the maximum of the log-likelihood $l := \ln L$ (maximizing ϑ remains the same). Thus $l(\vartheta, \underline{x}) = \sum_{i=1}^{n} \ln f_\vartheta(x_i)$.*
2. *By differentiating the function in ϑ, the maximum is the solution of the following equation: $\frac{\partial}{\partial \vartheta} l(\vartheta, \underline{x}) = \sum_{i=1}^{n} \frac{\partial}{\partial \vartheta} \ln f_\vartheta(x_i) = 0$.*

Example 5 (maximum likelihood estimations)

ϑ	$\hat{\vartheta}$
probability	*relative frequency*
parameter of Poisson distribution	*sample mean* ($\overline{x}$)
parameter of Exponential distribution	*1/sample mean* ($\frac{1}{\overline{x}}$)
expected value of Normal distribution	*sample mean* ($\overline{x}$)