Maximum spacing estimation

In statistics , maximum spacing estimation (MSE or MSP) , or maximum product of spacing estimation (MPS) , is a method for estimating the parameters of a univariate statistical model . The method requires maximization of the geometric mean of spacings in the data , which are the differences between the values of the cumulative distribution function at neighbouring data points .

The concept underlying the method is based on the probability integral transform , in that a set of independent random samples derived from any random variable should on average be uniformly distributed with respect to the cumulative distribution function of the random variable . The MPS method chooses the parameter values that make the observed data as uniform as possible , according to a specific quantitative measure of uniformity .

One of the most common methods for estimating the parameters of a distribution from data , the method of maximum likelihood (MLE) , can break down in various cases , such as involving certain mixtures of continuous distributions . In these cases the method of maximum spacing estimation may be successful .

Apart from its use in pure mathematics and statistics , the trial applications of the method have been reported using data from fields such as hydrology , econometrics , magnetic resonance imaging , and others .

History and usage

The MSE method was derived independently by Russel Cheng and Nik Amin at the University of Wales Institute of Science and Technology , and Bo Ranneby at the Swedish University of Agricultural Sciences . The authors explained that due to the probability integral transform at the true parameter , the ? spacing ? between each observation should be uniformly distributed . This would imply that the difference between the values of the cumulative distribution function at consecutive observations should be equal . This is the case that maximizes the geometric mean of such spacings , so solving for the parameters that maximize the geometric mean would achieve the ? best ? fit as defined this way . Ranneby (1984) justified the method by demonstrating that it is an estimator of the Kullback ? Leibler divergence , similar to maximum likelihood estimation , but with more robust properties for various classes of problems .

There are certain distributions , especially those with three or more parameters , whose likelihoods may become infinite along certain paths in the parameter space . Using maximum likelihood to estimate these parameters often breaks down , with one parameter tending to the specific value that causes the likelihood to be infinite , rendering the other parameters inconsistent . The method of maximum spacings , however , being dependent on the difference between points on the cumulative distribution function and not individual likelihood points , does not have this issue , and will return valid results over a much wider array of distributions .

The distributions that tend to have likelihood issues are often those used to model physical phenomena . Hall & al . (2004) seek to analyze flood alleviation methods , which requires accurate models of river flood effects . The distributions that better model these effects are all three @-@ parameter models , which suffer from the infinite likelihood issue described above , leading to Hall ? s investigation of the maximum spacing procedure . Wong & Li (2006) , when comparing the method to maximum likelihood , use various data sets ranging from a set on the oldest ages at death in Sweden between 1905 and 1958 to a set containing annual maximum wind speeds .

Definition

Given an iid random sample { x1 , ? , xn } of size n from a univariate distribution with the cumulative distribution function F (x ; ?0) , where ?0 ? ? is an unknown parameter to be estimated , let { x (1) , ? , x (n) } be the corresponding ordered sample , that is the result of sorting of all observations from smallest to largest . For convenience also denote x (0)

= ? ? and x (n + 1) =

+ ? .
Define the spacings as the ? gaps ? between the values of the distribution function at adjacent ordered points :
<formula>
Then the maximum spacing estimator of ?0 is defined as a value that maximizes the logarithm of the geometric mean of sample spacings :
<formula>
By the inequality of arithmetic and geometric means , function Sn (?) is bounded from above by ? ln (n + 1) , and thus the maximum has to exist at least in the supremum sense .
Note that some authors define the function Sn (?) somewhat differently . In particular , Ranneby (1984) multiplies each Di by a factor of (n + 1) , whereas Cheng & Stephens (1989) omit the 1 ? n + 1 factor in front of the sum and add the ? ? ? sign in order to turn the maximization into minimization . As these are constants with respect to ? , the modifications do not alter the location of the maximum of the function Sn .

= = Examples = =

This section presents two examples of calculating the maximum spacing estimator .

= = = Example 1 = = =