



Contribution ID: 439

Type: poster

Computationally efficient algorithms for the two-dimensional Kolmogorov-Smirnov test

Monday, September 3, 2007 8:00 AM (20 minutes)

Goodness-of-fit statistics measure the compatibility of random samples against some theoretical probability distribution function. The classical one-dimensional Kolmogorov-Smirnov test is a non-parametric statistic for comparing two empirical distributions, which defines the largest absolute difference between the two cumulative probability distribution functions as a measure of disagreement. Adapting this test to more than one dimension is a challenge because there are $2k-1$ independent ways of defining a cumulative probability distribution function when k dimensions are involved. However there are many applications in experimental physics where comparing two-dimensional data sets is important.

In this paper we discuss Peacock's version [3] of the Kolmogorov-Smirnov test for two-dimensional data sets which computes the differences between cumulative probability distribution functions in $4n^2$ quadrants and runs in $O(n^3)$, for a sample of size n . We also discuss Fasano and Franceschini's variation [2] of Peacock's test that runs in $O(n^2)$, Cooke's algorithm [1] for Peacock's test and ROOT's version of the two-dimensional Kolmogorov-Smirnov.

We establish a lower-bound limit on the work for computing Peacock's test in $W(n^2 \lg n)$, which contradicts Cooke's claim that it possible to perform this test with an $O(n \lg n)$ algorithm. We also establish a lower-bound limit of $W(n \lg n)$ for Fasano and Franceschini's test, and present an optimal sequential algorithm for it.

We finally discuss experimental results comparing two parallel algorithms implementing Peacock's test and an optimal algorithm for Fasano and Franceschini's test, and contrast these with the algorithm in ROOT which is based on the calculation of a mean of two one-dimensional Kolmogorov-Smirnov tests.

References

[1]

A. Cooke. The muac algorithm for solving 2d ks test.
<http://www.acooke.org/jara/muac/algorithm.html>.

[2]

G. Fasano and A. Franceschini. A multidimensional of the Kolmogorov-Smirnov test. Monthly Notices Royal Astronomy Society, 225:155-170, 1987.

[3]

J. A. Peacock. Two-dimensional goodness-of-fit testing in astronomy. Monthly Notices Royal Astronomy Society, 202:615-627, 1983.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS)

CMS

Primary authors: Dr REID, Ivan D. (School of Design and Engineering - Brunel University, UK); Prof. HOBSON, Peter R. (School of Design and Engineering - Brunel University, UK); Dr LOPES, Raul H. C. (School of Design

and Engineering - Brunel University, UK)

Presenter: Dr REID, Ivan D. (School of Design and Engineering - Brunel University, UK)

Session Classification: Poster 1

Track Classification: Software components, tools and databases