

Weighted average of scattered data: a Bayesian approach

Wednesday 12 June 2024 18:00 (2 hours)

A common problem for experimental scientists is computing averages of scattered data, i.e., with uncertainties much smaller than the distance between the different values. In such a case, a weighted average generally gives a very small uncertainty that does not reflect the bad agreement of the data. Different strategies are implemented in the literature, such as the application of a scale factor to the weighted average uncertainty proportional to a measure of the data scattering, like in the Particle Data Group (PDG) recommendations. Such methods are, however, based only on scientific intuitions without any strict statistical arguments.

We present here a rigorous and reproducible approach based on Bayesian statistics. The starting point, already known in the literature, is to consider each datum uncertainty σ_0 as a lower boundary estimation of the real unknown uncertainty σ . Once marginalised over σ values (using a Jeffreys prior), the expected statistical distribution is no longer a normal distribution but is characterised by smoothly decreasing wings. Like the derivation of the standard weighted average rule, our modified version is obtained by taking the maximum of the cumulative probability distribution and its second derivative value at the maximum. Unlike the standard weighted average, no analytical solution is available, and numerical methods have to be implemented. After a series of tests, the proposed method proved to be very reliable and robust for scattered data, but also with respect to the possible presence of outliers. In particular, it reproduces well the scale factor values reported by the PDG, but also suggests that the PDG method may be biased. When applied to the CODATA recommended values of the gravitational constant G our method proves a very good estimation of G for each CODATA edition, even for critical data sets (1996 edition) that included misleading measurements and where a scale factor of 37 was introduced in the final average value. The presented method will be extended in the future to correlated measurements, for applications to case such as the proton radius puzzle.

Authors: TRASSINELLI, Martino (CNRS, Sorbonne Université, Institut des NanoSciences de Paris); Ms MAX-TON, Marleen (Sorbonne Université, CNRS, Institut des NanoSciences de Paris)

Presenter: TRASSINELLI, Martino (CNRS, Sorbonne Université, Institut des NanoSciences de Paris)

Session Classification: Poster Session 2