

Calibration of assessors

R.S.MacKay

January 29, 2009

It frequently happens that a set of objects has to be evaluated by a panel of assessors but each assessor evaluates only a subset of the objects. There may be too many objects or the range of expertise of an assessor might not cover the range of objects.

The aim of this miniproject is to test and develop a method to achieve robust calibration of assessors and hence robust scores for the objects.

The basic version of the method is to ask each assessor a to provide for each object o they evaluate, a score s_{ao} and a confidence c_{ao} (i.e. indication of the reciprocal of how much they'd be willing to move their score up or down), and fit a model to the full set of results of the form

$$s_{ao} = v_o + b_a + \frac{\eta_{ao}}{c_{ao}},$$

with η_{ao} independent random variables of zero mean and unit variance. Then the resulting v_o is interpreted as the true value of the object o and b_a as the bias of assessor a . The obvious degeneracy that one could add the same constant k to all v_o and subtract k from all b_a can be removed by fixing $\sum_a b_a = 0$, for example. Least-squares fitting reduces to a simple linear algebra problem, which has a unique solution if the above degeneracy is removed and the graph linking assessors to objects they have evaluated is connected.

The project requires the student to:

- search the literature rapidly to find out what may already have been done of this nature (which could of course lead to drastic revision of the rest of this plan!),
- write a computer programme to do the fitting and to compute associated measures of robustness of the outcome,
- get hold of one or more databases to test out the method (this may not be easy because the data is confidential, so you will have to persuade the giver to put in the effort of anonymising it first, but Colin Sparrow offered undergraduate mathematics examination results for example),
- analyse the limitations and defects of the basic method, and ideally to propose and test refinements.

A draft paper is available from me (I might put it on my website if I receive a lot of requests, but it is in progress so I update it quite frequently).