International Journal of Computational Science and Mathematics.
ISSN 0974-3189 Volume 4, Number 2 (2012), pp. 69-76

http://www.irphouse.com

The Comparison and Scaling of Student Assessment Marks in Several Subjects using Optimization Method

Khizar H. Khan

alman bin Abdulaziz UniversityS,lA -Kharj
E-mail: khizarkhan@yahoo.com

Abstract

In many university examinations, particularly at final degree level, candidates are able to study particular options in order to bias their degree course in a particular direction. The result is a set of assessments in which not every candidate takes every paper. Yet for the purpose of ranking the candidates and classifying their degrees, a single overall mark must be assigned to each candidate. This paper discusses some of the methods for tackling the problem of producing such an average mark along with the difficulties that can result from their application.

It is assumed that each of *N* candidates takes q papers which are selected from a total number of *n* papers. The mark scored by candidate i on paper j is assumed to be m_{ij} and each paper is assumed to have equal weighting in the overall assessment. In order to harmonize standards across all papers, an adjustment parameter p_j (additive or multiplicative) for papers j is introduced along with an 'average' assessment a_i for candidate *i*. These parameters are calculated by minimizing a loss function which represents the disagreement between the actual marks and required properties of the assessment.

Software to compute the parameters, based on Fletcher-Reeves optimization method, has been prepared and applied to some actual examination marks.

Keywords: Fletcher-Reeves optimization Method; parameter computation; scaling

Introduction

In this paper we are interested in measuring the overall average of a student or, in a general sense, to check the ability of a student, so that a weak performance in one paper may be compensated by a strong performance in another paper.

Examinations usually consist of several components. These components can be papers, sections or questions of a paper on different subjects. We are interested in finding a fair and harmonious way of deriving overall marks from a set of components. Of course one important problem is to define how these component marks will be assigned to the candidates in the first place, but this will not be discussed. We shall assume that the component marks have been arrived at through a fair process and the only problem that will occupy us is that of combining them fairly and consistently.

One method that is often used is that of simply adding the component marks together to get the overall mark. This is based on the assumption that the components are all equally important and that the components are also treated equally. A common situation is for there to be some core courses with would be specialists in pure Mathematics, Applied Mathematics, Statistics or Computational Mathematics studying a number of special types. The result is a set of assessments in which not every candidate takes every paper. Yet for the purpose of ranking the candidates and classifying their degrees, a single overall 'average' mark must be assigned to each candidate. It is assumed that each of N candidates take q papers which are selected from a total numbers of n papers. The mark scored by candidate i on paper j is m_{ij} and each paper is assumed to have equal weighting in the overall assessment. Of course m_{ij} only exists for certain pairs. Even then it is only an imperfect measure of the ability of the candidate i in paper j. The papers will vary in their intrinsic difficulty, the examiners in their generosity. There is of course the added problem that ability is multivariate in its nature where as a numerical mark is one dimensional. The overall ability can be regarded as some function of the component marks in the individual topics.

In an attempt to harmonize standards across all papers, an adjustment parameter P_j (multiplicative or additive) for paper j is introduced, along with an 'average' measure a_i for candidate i. These parameters are calculated by minimizing a loss function which represents the disagreement between the scaled marks and the ability of the candidates. This idea was used by [6] in a some what different context. Murgatroyd ([7, 8]) also followed this philosophy although he mainly considered additive adjustments.

Comutation of the Paraneters

A simple form the loss function which treats all candidates and all papers on the same basis is the one proposed by [1, 2], viz.

$$S = \sum_i \sum_j (p_j m_{ij} - a_i)^2 \qquad (1)$$

We choose P_j and a_i to minimize this loss function. Unfortunately the solution to this problem is $p_j = 0$, and $a_i = 0$ for all i and j .Thus following [2, 3, 4] we introduced the constraints that the total of the marks remains the same. Thus we have

$$\sum_i \sum_j m_{ij} = \sum_i \sum_j p_j m_{ij} \tag{2}$$

Using (1) and (2) we can construct the Lagrangian

$$L = \sum_i \sum_j (p_j m_{ij} - a_i)^2 - 2\lambda \sum_i \sum_j m_{ij} (p_j - 1.0) \tag{3}$$

This leads to the equations

$$a_i = \sum_j m_{ij} p_j \,/\, n_i \tag{4}$$

and

$$p_j = \frac{\sum_i a_i m_{ij} + \lambda \sum_i m_{ij}}{\sum_i m_{ij}^2} \tag{5}$$

Where n_i is the number of papers taken by candidate *i*, (they don't all <u>have</u> to be the same). It is possible to set-up an iterative scheme to solve these equations.

It is perhaps worth commentary at this round that in an actual application we might have 100-200 candidates each taking say 8 options selected from 30-60 papers.

Thus with one variable (P_j) for each paper and another (a_i) for each candidate it would not be unusual to have a constrained optimization problem involving 140 variables. For this reason we have used Fletcher-Reeves optimization method to obtain the solution to the problem posed by [6, 7].

Unlike, variable metric methods such as Davidon-Fletcher-Powell or Broyden-Fletcher-Goldfarb-Shanno, which need for an approximating Hessian matrix, which gets updated at every iteration whereas the Fletcher-Reeves method (see [5]) uses and stores function values and gradient only. Thus it is suited to this problem with a large number of variables. It is an iterative method which searches along a set of mutually conjugate directions. To construct the next search direction we need the current gradient $\underline{g}$ and the last search direction (which is stored),

i.e.

$$d_{k+1} = -g_{k+1} + \alpha_k d_k \tag{6}$$

where

$$\alpha_k = \|g_{k+1}\|^2 / \|g_k\|^2 \tag{7}$$

and d_{k+1} is the current search direction. Since the constraint can be readily used to eliminate one of the variables (P_n say) we have

$$S = \sum_i \sum_j (p_j m_{ij} - a_i)^2$$

With

$$p_n = \frac{T}{T_n} - \sum_{j=1}^{n-1} p_j \frac{T_j}{T_n} \tag{8}$$

where $T_j = \sum_i m_{ij} \equiv$ Total marks for paper j, and $T = \sum_j \sum_i m_{ij} \equiv$ Total of all marks.

$$\frac{\partial S}{\partial a_i} = -2 \sum_j (p_j m_{ij} - a_i) \text{ for } i = 1,2,3,\cdots N. \tag{9}$$

$$\frac{\partial S}{\partial p_j} = \sum_i 2 m_{ij} (p_j m_{ij} - a_i) + \sum_i 2 m_{in} (p_n m_{in} - a_i)(-\frac{T_j}{T_n}) \tag{10}$$

for $j = 1,2,3,\cdots(n-1)$ give the gradient $\underset{\sim}{g}$.

With this formulation it was routine to calculate average assessment a_i, adjustment factor p_j and the results for the examination marks given in table -1 are provided in table 2 (a) and table 3(a).

Some Refinements

It is reasonable to suppose that each p_j should be near to one. To try to ensure that we could modify our loss function with a penalty component to become

$$S_1 = \sum_i \sum_j (p_j m_{ij} - a_i)^2 + c \sum_j (p_j - 1.0)^2 \tag{11}$$

Here $c = 0$ gives us (1) or (8) as before, whereas $c = \infty$ means $p_j = 1$ and we just calculate the straight forward average.

With Lagrangian (considering (11) and (2)) we get,

$$L_1 = \sum_i \sum_j (p_j m_{ij} - a_i)^2 + c \sum_j (p_j - 1.0)^2 + 2\lambda \sum_i \sum_j m_{ij} (p_j - 1.0) \tag{12}$$

This implies

$$a_i = \sum_j m_{ij} p_j / n_i$$

and

$$p_j = \frac{\sum_i a_i m_{ij} + \lambda \sum_i m_{ij} + c}{\sum_i m_{ij}^2 \;+\; c} \tag{13}$$

To have an effect c should be about m^2 where m is a typical mark. Values of around 2500-8000 seem reasonable although our calculations show that the final out come is not too sensitive to the actual values of c.

Biggins [1] show that if we include a fictitious candidate who score c on each paper so that $a = c\,\overline{p}$ (where $\overline{p}$ is the average value of p_j) for that candidate then (4) is unchanged but (5) becomes

$$p_j = \frac{\sum_i a_i m_{ij} + \lambda \sum_i m_{ij} + c^2\,\overline{p}}{\sum_i m_{ij}^2 + c^2} \tag{14}$$

which is equivalent to using the loss function

$$S_2 = \sum_i \sum_j (p_j m_{ij} - a_i)^2 + c^2 \sum_j (p_j - \overline{p})^2 \tag{15}$$

Some Numerical Results

The methods described have been applied to the following data.

Table 1: Sample data of marks for 20 candidates each taking 4 papers from seven and the paper-1 is compulsory.

Candidates	Paper-1	Paper-2	Paper-3	Paper-4	Paper-5	Paper-6	Paper-7
1	67	62	-	53	-	59	-
2	39	39	-	-	64	-	69
3	52	-	57	38	-	41	-
4	47	41	-	-	59	-	81
5	50	59	-	47	-	57	-
6	53	39	-	-	52	-	62
7	32	62	-	-	-	62	67
8	60	-	69	73	-	61	-
9	73	50	-	47	-	63	-
10	82	60		-	38	-	41
11	49	-	33	-	46	59	-

12	59	60	44	82	-	64	-
13	38	-	-	-	50	-	67
14	61	45	-	42	-	73	-
15	58	-	68	47	58	-	-
16	57	48	-	53	-	35	-
17	74	-	75	-	-	58	63
18	54	35	-	67	81	-	-
19	66	49	-	-	53	57	-
20	42	52	-	57	39	-	-

The outcome corresponding to a straight average with the loss function and other alternative methods is presented in tables 2 and 3.

Table 2: The overall marks a_i for the 20 candidates taking four papers out of seven and paper-1 compulsory. The columns (a)-(e) corresponds to the method described above.

Candidates	(a)	(b)	(c)	(d)	(e)
1	60.250	63.371	63.165	63.061	62.943
2	52.750	51.353	51.452	51.501	51.557
3	47.000	45.229	45.336	45.391	45.453
4	57.000	55.204	55.326	55.388	55.457
5	53.250	56.181	55.988	55.891	55.781
6	51.500	50.400	50.479	50.518	50.563
7	55.750	55.661	55.667	55.670	55.674
8	65.750	63.947	64.045	64.096	64.154
9	58.250	60.919	60.740	60.650	60.548
10	55.250	55.753	55.735	55.725	55.713
11	59.250	56.061	56.278	56.388	56.512
12	66.250	69.618	69.387	69.270	69.139
13	56.000	50.365	50.734	50.921	51.133
14	55.250	57.747	57.578	57.493	57.397
15	57.750	55.310	55.470	55.552	55.643
16	48.250	50.724	50.558	50.475	50.380
17	64.000	59.286	59.583	59.734	59.905
18	59.250	61.136	61.015	60.954	60.884
19	56.250	58.333	58.213	58.152	58.082
20	47.500	49.901	49.749	49.671	49.583

Table 3: The paper adjustment factors, P_j corresponding to the loss function (1) and the single fictitious candidate method (11) with different values of c.

Paper	(a)	(b)	(c)	(d)	(e)
1	1.000	1.003	1.003	1.003	1.002
2	1.000	1.140	1.132	1.128	1.123
3	1.000	0.828	0.840	0.846	0.853
4	1.000	1.040	1.037	1.035	1.033
5	1.000	0.997	0.998	0.998	0.999
6	1.000	1.025	1.023	1.022	1.021
7	1.000	0.841	0.850	0.855	0.861

Key:

The average raw marks.

The method using the loss function $S = \sum_i \sum_j (p_j m_{ij} - a_i)^2$

The modified (refined) method with c=1600

The modified method with c=4900

The modified method with c=6400

Conclusion

A comparative analysis shows that paper-2 is the most difficult paper or its examiner was tough enough or the paper was less popular. The paper-3 and paper-7 were although less popular but papers were easier and the examiners might be generous. We also found that if all the papers/subjects are compulsory, the changes in raw averages of students may be small and if majority of subjects are optional then the scaling process may produce large changes in the raw averages of the students. We reached to the conclusion that scaling methods give a fair indication of the abilities of the students/candidates.

Acknowledgement

The author (Khizar H. Khan) thankfully acknowledges the support provided by the Department of Mathematics, Salman bin Abdulaziz University, Al-Kharj and ministry of education, Saudi Arabia for providing the research facilities and the environment to work and publish in a reputed journal.

References

[1] Biggins, J.D. Loynes, R.M. and Walker, A.N. (1986). Combining Examination marks. Brit. J. Math. Statist. Psych. 39, 150-167.

[2] Biggins, J.D. and Yue, K.K. (1993). Scaling and combining examination marks. Brit. J. Math. Statist. Psych. 46, 153-179.

[3] Biggins, J.D. and Yue, K.K. (1995). A likelihood approach to scaling examination marks. Austral. J. Statist. 37, 273-282.

[4] Broyden, C.G. (1983). A mark-scaling algorithm. The Computer Journal, 26,109-112.

[5] Fletcher R , Reeves C M, (1964). "Function minimization by conjugate gradients," The Computer Journal, 7, 149-154.

[6] Healy, M.J.R. and Goldstein, H. (1976. An approach to the scaling of the categorized and attributes. Biomertrika 63, 219-229

[7] Murgatroyd, P.N (1975). An adjustment procedure for examination marks in optional subjects. International Journal of Mathematics Education in science technology, 6,435-444.

[8] Murgatroyd, P.N. (1979). Least squares estimation applied to a procedure for adjusting examination marks. International Journal of Mathematics Education in science technology. 10, 102-105.