BAYES' ANSWER TO PERFECTION

Ben Wright

University of Chicago

(7/20/86)

ABSTRACT

When a measurement model is used to transform scores into measures, perfect scores produce infinite estimates. If the prior belief that it was reasonable to administer the test is acceptable, then Bayes' estimation can be applied to Rasch measures to provide finite estimates for perfect scores which are simple functions of the measure and error of a next-to-perfect score. General equations for Bayes' estimation with Rasch measures are provided.

KEYWORDS: Bayes, estimation, item response theory, measurement, perfect scores, Rasch.

BAYES' ANSWER TO PERFECTION

Ben Wright

University of Chicago

(7/20/86)

We know the perfection of answering every test item correctly is conditional. How much a perfect score is worth depends on the difficulty of the items. But, when we count correct answers in order to calibrate items and measure persons, perfection makes trouble.

When all items used to measure a person are exceeded, the mathematical answer to the worth of a perfect score is "infinity". We cannot determine from a perfect performance alone how much more that person might know. The sensible answer to the worth of a perfect performance is "give harder items". But this answer is unsatisfactory, when it is inconvenient to give a harder test and we must, nevertheless, record finite measures for "perfect" performers.

What to do? What finite measure shall we use to represent a perfect performance? What standard error of measurement shall qualify this measure? If we seek objectivity (Rasch 1961 1968, Roskam and Jansen 1984, Fischer 1985, Wright 1985, Douglas and Wright 1986) or fundamental measurement (Luce and Tukey 1964, Brogden 1977, Perline, Wright and Wainer 1979) or just simplicity and use a Rasch model to govern measuring, then a satisfying and very simple Bayesian answer is available. This answer is obtained by adding to the observed performance a prior belief as to how much the person probably knows, along with an indication of our confidence in this belief.

EXPRESSING PRIOR BELIEF AS AN IMAGINARY PERFORMANCE

We can make our prior belief and our confidence in it explicit by imagining the distribution of all values we think possible for the person's measure and letting:

> M be the mean of this "distribution" and
> S be its standard deviation.

Then M can be our best guess as to how much this person knows and S can indicate the amount of our uncertainty.

The way M and S can represent prior belief is easy to grasp but inconvenient to apply. The translation of M and S into an imaginary test performance, however, would expedite application because then we could append this prior "performance" to the observed performance for a posterior synthesis of prior belief and current experience. We can achieve this translation by thinking of M and S as the result of a just completed test on which this person's estimated measure was M with standard error S .

An easy way to do that is to imagine that the performance which resulted in M and S took place on a test which was perfectly targeted on this person with all items of difficulty M. For such a test to produce a measure M with error S, the number of items on the test would have to be $N = 4/(S*S)$ and the person's test score would have to be $N/2$. This means we can express our prior belief concerning how much this person knows as a performance of:

$N/2$	correct answers, on a test which is
$N = 4/(S*S)$	items in length, with each item of
M	difficulty.

To integrate this belief with an observed performance, we append this imaginary "performance" which represents our prior belief to the performance observed and calculate a single measure and error for this person from this concatenation. Even when the observed performance is perfect the result is not a perfect score because our prior belief equips the person with $N/2$ incorrect answers.

We can also express our prior belief as M and N directly, rather than as M and S. This is equivalent to saying that we believe the person to be near M and that our belief is worth the amount of information provided by a test of N well-targeted items.

ESTIMATING A MEASURE FROM A PERFORMANCE AND A BELIEF

A simple way to estimate a measure B from a performance of R correct answers on a test of L items of known difficulties Di for i = 1 to L is to deduce the equation for B (as a function of R, L and the Di's) that would result if the distribution of item difficulties were normal (Wright 1977, Wright and Stone 1979 Chapter 2 or 7).

If H = (sum Di)/L is the item difficulty mean

and W = (sum Di*Di)/L - H*H is the item difficulty variance,

then the equation for B based on the observed performance is:

$$B = H + \mathrm{root}(1 + W/2.9) * \log(R/(L - R)) \tag{1}$$

with standard error

$$E = \mathrm{root}((1 + W/2.9)L/R(L - R)) \ . \tag{2}$$

When R = L , as is the case with a perfect score, then L - R = 0 so that the estimates of B and E become infinite.

When we append an imaginary prior "performance" to an observed performance we have:

$R + N/2$ correct answers, on a test which is

$L + N$ items in length with difficulty mean

$G = (L*H + N*M)/(L + N)$ and difficulty variance

$V = (L*(W + H*H) + N*M*M)/(L + N) - G*G$

This makes the equation for the person's composite posterior measure:

$$B' = G + \mathrm{root}(1 + V/2.9) * \log((R + N/2)/(L - R + N/2)) \tag{3}$$

with standard error

$$E' = \mathrm{root}((1 + V/2.9)(L + N)/(R + N/2)(L - R + N/2)) . \tag{4}$$

ESTIMATING A MEASURE FROM A PERFECT SCORE

These equations for B' and E' enable us to deal with the situation where a person has obtained the perfect score, $R = L$. Now the equations for B' and E' become:

$$B' = G + \mathrm{root}(1 + V/2.9) * \log(2L/N + 1) \tag{5}$$

$$E' = \mathrm{root}((1 + V/2.9)(L + N)/(L + N/2)(N/2)) . \tag{6}$$

DECIDING WHAT TO BELIEVE AND HOW MUCH

What values shall we choose for M and N to introduce a prior belief into these equations? If the act of giving this test to this person can be construed as evidence of a belief that a performance on this test would be useful for measuring this person, then it must follow that the test was aimed at the person. But, if the test was aimed at the person, then $M = H$.

If we further decide that our belief that the test was aimed at the person was worth as much as the information in one test item response, then $N = 1$.

A FINITE MEASURE FOR A PERFECT SCORE

Entering these values in the equations for B' and E' produces:

$$B' = H + \text{root}(1 + W*L/2.9(L + 1)) * \log(2L + 1) \tag{7}$$

$$E' = \text{root}((1 + W*L/2.9(L + 1))(2L + 2)/(2L + 1)) \tag{8}$$

as the measure and error for a perfect score.

Now the equations for B and E for an observed next-to-perfect performance of $R = L - 1$ are

$$B = H + \text{root}(1 + W/2.9) * \log(L - 1) \tag{9}$$

$$E = \text{root}((1 + W/2.9)L/(L - 1)) \,. \tag{10}$$

This pair of equations provide a finite measure and error for the next-to-perfect performance. When we combine them with the Bayesian equations (7) and (8) for B' and E' , which are based on one item response worth of conviction that it was reasonable to give this test to this person, we get the approximations:

$$B' = B + \text{root}(1 + W/2.9) * \log(2 + 3/(L - 1)) \tag{11}$$

$$\begin{aligned} E' &= E * \text{root}(2(L - 1/L)/(L + 1/2)) \\ &= 1.4E * \text{root}((L - 1/L)/(L + 1/2)) \,. \end{aligned} \tag{12}$$

But

a. the value of root(1 + W/2.9) varies from 1.1, when W = 0.6, to 1.2, when W = 1.3 (a range encompassing the item dispersion of most tests),

b. the value of $\log(2 + 3/(L - 1))$ varies from .75, when $L = 25$, to .70, when $L = 217$, and

c. the value of $\mathrm{root}(L - 1/L)/(L + 1/2)$ varies from .99, when $L = 25$, to 1.00, when L is infinite.

As a result we can represent the relations between B' and B and E' and E quite well with the approximations:

$$B' = B + (1.1)(0.7) = B + .8 \tag{13}$$

$$E' = 1.4E \tag{14}$$

These approximations hold regardless of the number of items or the shape of the test, providing we have used at least 25 items and the range of item difficulty does not exceed 4 logits. Adjustments for shorter tests are easy to calculate but not different enough to bother with. For $L = 20$ the addition to B is .85 and the coefficient for E is 1.395 .

IN CASE OF A STRONGER BELIEF

Should our belief that administering this particular test to this perfectly scoring person was reasonable be worth the amount of information in two item responses, then equations (5) and (6) lead to the approximations:

$$B' = B + .1 \tag{15}$$

$$E' = E . \tag{16}$$

With two item responses worth of belief in test relevance, we might be tempted to use the B and E for a next-to-perfect score of $R = L - 1$ as reasonable estimates for those who earned the perfect score of $R = L$. But that would equate those who did perfectly with those who got one wrong. If we took this course of action, then we would probably want to apply equations (3) and (4) to all performances so that $R = L$ received a higher measure than $R = L - 1$.

REFERENCES

Brogden, H.E. (1977). The Rasch model, the law of comparative judgement and additive conjoint measurement. PSYCHOMETRIKA, 37, 29-51.

Douglas, G.A. and Wright, B.D. (1986) THE TWO CATEGORY MODEL FOR OBJECTIVE MEASUREMENT. Memorandum No. 34, MESA Psychometric Laboratory, Department of Education, University of Chicago.

Fischer, G. (1985) Some consequences of specific objectivity for the measurement of change. In Edw. Roskam, MEASUREMENT AND PERSONALITY ASSESSMENT (pp.39-56). Amsterdam: Elsevier.

Luce, R.D. and Tukey, J.W. (1964) Simultaneous conjoint measurement: A new type of fundamental measurement. JOURNAL OF MATHEMATICAL PSYCHOLOGY, 1, 1-27.

Perline, R., Wright, B.D. and Wainer, H. (1979) The Rasch model as additive conjoint measurement. APPLIED PSYCHOLOGICAL MEASUREMENT, 3, 237-256.

Rasch, G. (1961) On general laws and the meaning of measurement in psychology. PROCEEDINGS OF THE FOURTH BERKELEY SYMPOSIIUM ON MATHEMATICAL STATISTICS AND PROBABILITY (pp. 321-333). Berkeley: University of California Press.

Rasch, G. (1968) A mathematical theory of objectivity and its consequences for model construction. In REPORT FROM THE EUROPEAN MEETING ON STATISTICS, ECONOMETRICS AND MANAGEMENT SCIENCES, Amsterdam.

Roskam, E.E. and Jansen, P.G.W. (1984) A new derivation of the Rasch model. In E. Degreef and J. van Buggenhaut, TRENDS IN MATHEMATICAL PSYCHOLOGY (pp.293-307). Amsterdam: Elsevier.

Wright, B.D. (1977) Solving measurement problems with the Rasch model. JOURNAL OF EDUCATIONAL MEASUREMENT, 14, 97-116.

Wright, B.D. (1985) Additivity in psychological measurement. In Edw. Roskam, MEASUREMENT AND PERSONALITY ASSESSMENT (pp. 39-56). Amsterdam: Elsevier.

Wright, B.D. and Stone, M.H. (1979) BEST TEST DESIGN, Chicago: MESA Press.