On a law of ordinal error

David Andrich and Pender Pedler
Graduate School of Education, The University of Western Australia, Crawley WA 6009, Australia
Email: david.andrich@uwa.edu.au

Abstract. When no systematic factor disturbs replicated measurements of the same entity with the same instrument, the observed or inferred distribution is expected to satisfy the Gaussian law of measurement error. A characteristic of this distribution, which ensures it is unimodal with a smooth transition between adjacent probabilities, is that it is strictly log-concave. Many assessments in the social sciences begin by analogy to measurements in that an idealised linear continuum is partitioned by successive thresholds into contiguous, ordered categories. However, the distances between successive thresholds, generally finite, are not equal and assessments remain ordinal. The paper establishes that if the thresholds in the probabilistic Rasch measurement model used to transform ordinal assessments into measurements are in their natural order, then the distribution is also strictly log-concave. Therefore it is proposed that the Rasch model with ordered thresholds be referred to as the law of ordinal error. Accordingly, by analogy to the expectation that the distribution of replicated measurements satisfy the law of measurement error, it is proposed that the observed or inferred distribution of replicated ordinal assessments be expected to satisfy the proposed law of ordinal error.

1. Introduction

The prototypic measurement, that of mapping the amount of a property on a real line divided by successive thresholds from a conventional or natural origin into equal intervals called a unit, is understood by elementary school children. However, the function of measurement in science, which is the basis of the many remarkable advances in the physical sciences, is the understanding of variables and their mathematical formalization into scientific laws [1]. Although measured variables are conceived of as continuous, every measurement is discrete in the unit of measurement. Moreover, these measurements have errors and the role of errors continue to exercise scientists at both the conceptual and instrumental or empirical levels [2].

The long history of the recognition, acceptance, formalization and exploitation of random errors of measurement has been studied in Eisenhart [3].

“Laws of error,” i.e., probability distributions assumed to describe the distribution of the errors arising in repeated measurement of a fixed quantity by the same procedure under constant conditions, …culminating in the quadratic exponential law of Gauss,

\[ f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \]

upon which Gauss based his first formulation of the method of
least squares, which became almost universally regarded in the nineteenth century as “the law of error”. [3: p.1].

When no measuring instrument of the kind used in the physical sciences is available, measurement in the social sciences generally begins with assessments in a set of ordered categories. A hypothesized linear continuum, by analogy to a unit of measurement, is partitioned by successive thresholds into ordered categories. However, unlike in measurement, the number of categories is finite and small and the successive thresholds are not expected to be equidistant. In advanced analyses these ordinal assessments are modelled by probabilistic models. Although such modelling has a substantial history [4], there is no law of ordinal error analogous to the Gaussian law of measurement error. This paper proposes such a law based on the Rasch measurement model (RMM). The function of the proposed law for ordinal assessments is the same as the function of the law of error in measurement – to disclose factors that generate errors which are systematic rather than random.

2. The Rasch measurement model

There is not the space in this paper to detail the case that when data conform to the RMM, its transformation of the ordinal assessments provides measurements which in principle are of the kind obtained in the natural sciences. The case is made in Rasch [5], [6], and developed further in Andrich [7], Fisher & Stenner [8], Wright [9] and Stone & Stenner [10]. The model was derived from the theory that measurement arises from invariant comparisons – that within a specified empirical frame of reference, the comparisons of entities are independent of the instruments of measurement, and that the comparisons of instruments are independent of the entities [6]. Without loss of generality, the ordinal assessments can be assigned successive integers beginning with 0, giving ordinal counts \( x = 0, 1, 2, \ldots, m \).

Then the RMM used to model such ordinal counts is expressed in the form

\[
P_{ni} = \Pr(X_{ni} = x; \beta_n, (\delta_i)) = \exp\left(\sum_{i=0}^{m} \delta_i - \sum_{i=1}^{x} \delta_i - \sum_{i=x}^{m} \delta_i \right) + x \beta_n / \gamma_m^{-1},
\]

where (i) \( X_{ni} \) is the random variable of the ordinal counts of assessed entity \( n \) with instrument \( i \) which take the values \( x = 0, 1, 2, \ldots, m \) for the \( m+1 \) successive categories, (ii) \( \beta_n \) is the measure of person \( n \), (iii) \( \delta_{i0}, \delta_{i1}, \ldots, \delta_{ix}, \ldots, \delta_{im} \) is a vector of \( m_i \) thresholds where \( \delta_{lx} \) is the point on the continuum where the probabilities in the two adjacent categories \( x-1 \) and \( x, x > 0 \) are equal, with \( \delta_{i0} = 0 \) for convenience, and (iv) where \( \gamma_m = \sum_{\delta_{i0}}^{m} \exp(\sum_{\delta_{i0}}^{m} \delta_i) \) is a normalizing factor which ensures Equation (1) is a probability distribution. In the rest of the paper we focus on the responses to a single instrument and drop the subscript \( i \) from \( m_i \). Subscripts \( n, i \) are retained to emphasise that the replications of concern are of the assessment of a single person with a single instrument. Incidentally, in the social sciences, a set of ordered categories is referred to as an item.

3. The functions of the laws of measurement and ordinal errors

The Gaussian law of measurement error (LME) is derived theoretically and not to fit any observed distribution of measurements. Like hypothesized variables of measurement, the Gaussian distribution is continuous. However, because all measurements are discrete in the unit of measurement, we immediately consider the discretised Gaussian distribution in the unit of measurement. It is discretised by treating the ordinate at each possible measurement as its probability and to ensure a probability distribution, the ordinates in the range of possible measurements are normalised by their sum. Panel A of Figure 1 shows such a discretised distribution. The RU ratio referred to in Figure 1 is considered later in the paper as is the distribution of Panel B.
The LME is unimodal with a smooth transition between the adjacent probabilities. These properties are generally taken for granted. For the purpose of this paper in which these features are to be generalised to ordinal assessments, they need to be made explicit. To convey smoothness and unimodality, such a distribution will be referred to as randomly unimodal.

The functions of the LME arise from its comparison with a real or inferred empirical distribution of replicated measurements. If the empirical distribution is deemed to deviate significantly from the LME then two implications follow. First, the mean of the measurements cannot be taken as an estimate of the measure of the entity. Second, no less importantly, an unknown factor(s) have interfered with the measurements. That interference can lead to an empirical and theoretical search which can further the understanding of the variable and its measurement.

Like the Gaussian distribution, the RMM was derived theoretically. Its present form was derived with no data analyses in three successive papers [5], [11], [12]. The present paper shows that if the thresholds in the RMM are ordered, then the distribution satisfies the definition of random unimodality and therefore can be referred to as the law of ordinal error (LOE). By analogy to the functions of the LME, if it is deemed that real or replicated ordinal assessments deviate significantly from the LOE, then a scalar cannot be taken as an estimate of the measure of an entity and some unknown factor(s) has interfered with the assessments. By analogy to the functions of the LME, a search for a factor(s) can lead to further understanding of the variable and its assessment.

4. The definition of random unimodality: Strict log-concavity

The random unimodality of the continuous Gaussian distribution arises from its property of strict log-concavity [13]. Keilson and Gerber [14] generalise strict log-concavity to discrete distributions showing, for example, that the Poisson and binomial distributions are strictly log-concave. A discrete distribution \( P_i \) of counts \( m \) is strictly log-concave if

\[
\frac{P_{i+1}}{P_{i}} > \frac{P_{i}}{P_{i-1}}, \quad 0 < x < m.
\]

We refer to the ratio on the left of Equation (2) as the RU ratio. The term log-concave arises from Equation (2) where \( \ln P_i \) is greater than the mean of \( \ln P_{i-1} \) and \( \ln P_{i+1} \).

4.1. Strict log-concavity and the law of measurement error

Figure 1 shows two distributions. The distribution in Panel A, with mean 15.0, satisfies the LME with measurements in the range of 0 to 30. Its RU ratios are not only greater than 1 (1.041), \( 0 < x < 30 \), but constant. It can be shown readily that for the LME, the

\[
RU \ ratio: \frac{P_{i}^{2}}{(P_{i-1}) P_{i+1}} = \exp[(-\sigma^2)^{-1}] \quad 0 < x < m.
\]

The constant ratio of Equation (3) shows that the LME is uniformly smooth. The mean of 15.0 estimates the entity’s measure and there is no evidence of a systematic factor(s) interfering with the measurements. In contrast the distribution in Panel B, also with mean 15.0, is not randomly unimodal.
with five measurements with RU ratios less than 1 – the distribution is also bimodal. Its mean of 15.0 does not estimate the entity’s measure and some unknown factor(s) has interfered with the measurements. This extreme non-randomly unimodal distribution, unlikely in measurement, provides a basis for identifying non-randomly unimodal distributions of ordinal assessments.

4.2. Strict log-concavity and the law of ordinal error
It can be shown readily that the RU ratio in the RMM of Equation (1) is given by
$$RU\ ratio:\left(\frac{P_{x_i}}{P_{x_{i+1}}}\right)^{1/2} = \exp(\delta_{i+x} - \delta_i), \quad 0 < x < m. \quad (4)$$
Therefore if $\delta_{i+x} > \delta_i$, in which case the thresholds are in their natural order, then independently of the measure $\beta$ the distribution is strictly log-concave and randomly unimodal. It justifies the proposal that the RMM with ordered thresholds characterises the law of ordinal error.

5. Replications and analyses of ordinal assessments with the RMM
Many assessments in the social sciences, such as those of performance and attitude, involve multiple instruments (items) which assess the same variable for a well-defined class of persons. Multiple instruments provide a kind of replication and are used, by analogy to replicated measurements in the natural sciences, to improve the precision and validity of the assessment.

5.1. Estimating the parameters of an instrument
One distinctive property of the RMM is that the vector of each instrument’s parameters can be estimated independently of the parameters $\beta_n$ of any sample of $n = 1, 2, ..., N$ persons. The estimation exploits the property that the sum score $r_n = \sum_{i=1}^{m} x_{ni}$ of each person across the $I$ instruments is sufficient in the sense that, if the profile of responses fits the model, then there is no further information about $\beta_n$ in the vector of assessments across the instruments. For example, for two instruments $(i, j)$, the conditional estimation equation takes the form
$$Pr(x, y) = \exp((-\delta_{0i} - \delta_{1i} - \ldots - \delta_{ni}) + (-\delta_{0j} - \delta_{1j} - \ldots - \delta_{nj}))^{\gamma_{nj}} \quad (5)$$
where $\gamma_{nj} = \sum_{i=1}^{m-1} \exp((-\delta_{0i} - \delta_{1i} - \ldots - \delta_{ni}) + (-\delta_{0j} - \delta_{1j} - \ldots - \delta_{nj}))$ is the sum of the exponent in Equation (5) over all meaningful scores of $r$. Because there is only one possible response pair for each of $r_n = 0$, $r_n = 2m$, these conditional probabilities are uninformative. The distinctive property of Equation (5) is that it contains only the vectors of thresholds $(\delta_i, (\delta_j)$ of $i, j$ and does not contain $\beta_n$. Equation (5) shows the first distinctive property of the RMM. Even though each response pair $(x_{ni}, y_{nj})$ is governed respectively by a different parameter, $\lambda_{ni} = f((\delta_i, \beta_n), \lambda_{nj} = f((\delta_j, \beta_n))$, the distribution of conditional responses is a function of the same set of instrument parameters $(\delta_i, \delta_j)$; therefore across a sample of persons it reflects genuine replications. Second, when these replications are used to obtain the estimates $(\hat{\delta}_i, \hat{\delta}_j)$, no assumptions need be made regarding the distribution $\beta$ of the persons in the sample. Third, the instrument parameters characterize the assessment features of the instrument for any potential measure $\beta$. These distinctive properties are relevant when concerned with distributions of replications of assessments of a single person with a single instrument. Equation (5) readily generalises to the total score of many instruments and variable numbers of thresholds among them. Therefore, given estimates $(\hat{\delta}_i)$ of instrument $i$, and assessments with one or more instruments, each estimate $\hat{\beta}_n$ can be obtained individually and is taken as a measurement with an arbitrary conventional unit and an arbitrary conventional origin which can be transformed linearly for convenience.

5.2. Inference of replicated ordinal assessments
Substituting the estimates ($\hat{\delta}_i$) into the RMM of Equation (1) gives

$$P_{nix} = \Pr(X_{ni} = x; \beta, (\hat{\delta}_i)) = \exp((-\hat{\delta}_{i0} - \hat{\delta}_{i1} - \hat{\delta}_{i2} - ... - \hat{\delta}_{in}) + x\beta_n)\gamma_n^{x+1}.$$  \hspace{1cm} (6)

for any $\beta$. Equation (6) is the distribution of inferred replications of the assessment of person $n$ with instrument $i$. Given the parameter values ($\hat{\delta}_i$), it is the distribution of the assessment of any person with any $\beta$. It can be interpreted as if the person was assessed independently with the instrument on multiple occasions even though each person is assessed only once by each instrument. Clearly the response of a person to the instrument provides an observation in this distribution.

A second distinctive property of the RMM is that the threshold estimates may not be in their natural order even if the assessments fit the RMM statistically. That is, the inferred distribution of replicated assessments is not forced by the RMM to be randomly unimodal - random unimodality is an empirical property of the assessments as disclosed by the RMM, and not a property of the RMM. However, it is expected that the threshold estimates of any instrument will be in their natural order, that is, that the distribution for any measure will be deemed to satisfy the LOE. This is analogous to the expectation that replicated measurements, which are not forced to be Gaussian, will satisfy the LME.

### 5.3. Examples of the analysis of ordinal assessments

Two examples of ordinal assessments analysed with the RMM are shown below in Figure 2. Panel A shows a distribution from R. A. Fisher [15: p.290] which involved 12 actual replicated serological readings for five cells classified into five ordered levels of reaction. The threshold estimates are $\hat{\delta} = (-3.543, -2.095, 2.357, 3.282)$, though not equidistant, they are in their natural order - $\delta_{(x+1)} > \delta_{xx}$, $x = 1, 2, 3$. The distribution shown in Panel A, which includes the RU ratios, is for a cell measurement of $\beta = 0.00$. Clearly for this measurement, and from Equation (5) for any other cell location, the RU Ratio is greater than 1 and the distribution is randomly unimodal. If the profile of responses across five readings of a cell fits the RMM, a calculated value of $\hat{\beta}$ can be taken to estimate $\beta$ for the cell. In addition, there is no evidence of a systematic factor interfering with the assessments.

Panel B shows a distribution from Embretson and Reise [16] concerning a 12-item (instrument) neuroticism questionnaire where each instrument involves ratings in five ordered categories. The distributions are from an analysis of responses of persons to their Item 3, highlighted by the authors. The vector of threshold estimates is $\hat{\delta} = (-1.800, 0.167, -1.168, 1.117)$. In this example the thresholds are not in their natural order; specifically, $\hat{\delta}_{13} < \hat{\delta}_{12}$. Therefore for $x = 2$, the RU Ratio is $0.263 < 1$ and the distribution $P_{nix}, x = 0, 1, 2, 3, 4$ for any $\beta$ is not randomly unimodal. The distribution in Panel B is specifically for $\beta = -0.50$ and in this case the distribution is even bimodal. In addition to Item 3, all the instruments in the questionnaire exhibit similarly non-naturally ordered thresholds. Therefore, even if the profile of responses of a person across instruments of the questionnaire fits the RMM, a calculated value of $\hat{\beta}$ cannot be taken to estimate $\beta$. In addition, there is evidence that there is a systematic factor(s) interfering with the responses and that the empirical ordering of the categories, specifically the middle category, is not the intended ordering. This factor(s) needs to be studied theoretically and empirically. Understanding and correcting the interference which renders the middle category to malfunction would add to the understanding of the assessment of neuroticism.
Figure 2. Panel A: R. A. Fisher distribution $P_{nix}$ for $\beta = 0.00$. Panel B: Embretson and Reise distribution for $\beta = -0.50$. The RU Ratio, $x = 1, 2, 3$ is also shown.

6. Conclusion

Surprisingly, many ordinal assessments when analysed with the RMM exhibit distributions of the kind shown in Panel B of Figure 2. Perhaps this situation arises from the lack of formalisation, up to this point, of a LOE. With its formalisation, the properties of errors of ordinal assessments used in the social sciences to construct measurements can be studied from the same perspective as errors of measurements in the physical sciences.

References

[1] Kuhn T S 1961 The function of measurement in modern physical science Isis 52 161-93
[2] JCGM 2008 Evaluation of measurement data - guide to the uncertainty in measurement (GUM)
[3] Eisenhart C 1983 Law of error I: development of the concept Encyclopedia of statistical sciences vol 4, ed S Kotz and N L Johnson (Toronto: Wiley) pp 530-47
[4] Clogg C C and Shihadeh E S 1994 Statistical models for ordinal variables (Thousand Oaks, California: Sage Publications)
[5] Rasch G 1961 On general laws and the meaning of measurement in psychology Proc. Fourth Berkeley Symp. on Math. Statis. and Prob. vol 4, ed J Neyman (Berkeley CA: University of California) pp 321-34
[6] Rasch G 1977 On specific objectivity: an attempt at formalising the request for generality and validity of scientific statements Danish Yearbook of Philosophy 14 58-94
[7] Andrich D 2017 Advances in social measurement: a Rasch measurement theory Perceived health and adaptation in chronic disease: stakes and future challenge ed F Guillemin, A Leplège, S Briançon, E Spitz and J Coste (Taylor and Francis) In Press
[8] Fisher W P Jr and Stenner A J 2016 Theory-based metrological traceability in education: a reading measurement network Measurement 92 489-96
[9] Wright B D 1977 Solving measurement problems with the Rasch model J. Educ Meas. 14(2) 97-116
[10] Stone M H and Stenner A J 2014 Comparison is key J. Appl. Meas. 15(1) 26-39
[11] Andersen E B 1977 Sufficient statistics and latent trait models Psychometrika 42 69-81
[12] Andrich D 1978 A rating formulation for ordered response categories Psychometrika 43(4) 561-74
[13] Ibragimov I A 1956 On the composition of unimodal distributions Theor. Probab. Appl. 1(2) 255-60
[14] Keilson J and Gerber H 1971 Some results for discrete unimodality J. Am. Stat. Assoc. 66 386-9
[15] Fisher R A 1958 Statistical methods for research workers 13th edition (New York: Hafner)
[16] Embretson S E and Reise P R 2000 Item Response Theory for Psychologists (Mahwah: Lawrence Erlbaum)