KM 13(1984)
pag 120-130

APPLICATIONS OF A BAYESIAN POISSON MODEL FOR MISREADINGS*

Margo G.H. Jansen**
University of Groningen

Abstract

Apart from the widely known model for binary scored items, Rasch has developed several other models for the analysis of achievement test data. The model we consider here is the so-called multiplicative Poisson model for misreadings. This model assumes that the total number of misreadings on a text for a certain individual is approximately Poisson distributed with an intensity parameter which depends on the ratio of two other parameters, one partaining to the ability of the individual and one to the difficulty of the text.

A Bayesian version of this model was developed by Owen (1969). In this paper we adopt a slightly different formulation in the tradition of the general approach by Lindley (Lindley, 1970; Lindley and Smith, 1972). The method is illustrated by two examples, one based on empirical and the other on artificial data.

1. Introduction.

Apart from the so-called one-parameter logistic model for binary scored items, Rasch has developed several other latent trait models for the analysis of achievement test data. Among these are the models for oral reading speed and for misreadings in a text. These models are described in the monograph "Probabilistic models for some intelligence and attainment tests" (Rasch, 1960) and summarized by Lord and Novick (1968).

*Paper read at the 14th European Mathematical Psychology Group Meeting: Brussels, 1983.

**Dr. G.G.H. Jansen, Institute of Education, Westerhaven 16, Groningen.

Compared to the logistic model these models have attracted much less attention. As a result these models are much less developed and practical applications apart from those of Rasch himself are not known to the author. This seems unfortunate because tests that would lend themselves to analyses based on Poisson process models are widely used in educational practice. Bayesian versions of the Poisson model for misreadings were already developed in 1969 by Owen. Related models were developed by Van der Ven (1969).

2. The Poisson model for misreadings.

In contrast to the one-parameter logistic model, the Poisson model for misreadings is a latent trait model for tests rather than items. Rasch however begins his development of the latter with the assumption that a text to be read can be considered as a succession of mutually independent Bernouilli trials, corresponding with the words in the text, where the chances of an error are relatively small. A second, but extremely important assumption is that the probability of a misreading is a product of two factors, one pertaining to the ability of the pupil and one pertaining to the difficulty of the word. Then, given that the error probabilities are indeed small and the number of words in the text (= number of items) is large, the distribution of the total number of errors in the text (= test score) may be approximated by a Poisson distribution with intensity parameter equal to the sum of the Bernouilli parameters. The intensity parameter has the same multiplicative structure as the parameters on item-level.

Rasch assumes in his derivations that the error probabilities per item are constant given a certain test. However, because of the kinds of data we would like to describe, it seems more useful to base the derivation of the Poisson model on the equally standard but more general case in which the Bernouilli parameters are variable (Lord & Novick, 1968, p. 485 ff).

Looking at the examples given by Rasch, it is also clear that it is the model at test-level for which we have expirical evidence. So in fact, we could take the Poisson model for the total test scores as our starting point, with the item-level model as used by Rasch as one among other possible underlying structures.

The model on test-level is completed by the assumption that we have two or more tests for which a Poisson model holds.

Formally, consider a sample of N persons taking k tests, where the expected number of errors y_{ij} made by person i on test j is given by

$$P\ (Y_{ij}=y_{ij}|\lambda_{ij}) = \exp(-\lambda_{ij})\ \lambda_{ij}^{\ y_{ij}}/y_{ij}! \tag{1}$$

We suppose that the intensity parameter λ_{ij} is a ratio of two other parameters, the difficulty δ_j of the test and the ability ξ_i of the person

$$\lambda_{ij} = \delta_j\ /\ \xi_i\quad . \tag{2}$$

The expected number of errors will be higher if the test is more difficult or the person less able and lower if the test is less difficult and/or the person more able.

Obviously neither the δ nor the ξ are uniquely determined. This state of affairs can be remedied by imposing an arbitrary restriction on the set of parameters. Rasch proposes taking the most difficult test as a point of reference by choosing a value of one for the corresponding difficulty parameter. Notice that, because all Poisson parameters have to be nonnegative, setting one difficulty to be positive implies that all difficulty parameters δ and all ability parameters ξ are also nonnegative. In the Bayesian frame work identifiability problems are solved differently, by choosing suitable prior distributions which tell us how to identify the parameters, as we will show in the next section.

From a theoretical point of view, one very desirable characteristic of the Poisson model, which it shares with the "ordinary" Rasch model, is the separability of the parameters. Test parameters can be estimated independently of the person parameters and vice versa. For a discussion of what in this context is also known as "specific objectivity", see Rasch (1960), Wright & Stone (1979) and Jansen (1983).

Maximum Likelyhood estimates for the relative difficulty parameters are given by equation (3).

$$\hat{\rho}_j = \frac{y_{+j}}{y_{++}}\quad , \tag{3}$$

where $\rho_j = \delta_j\ /\ \delta_+\quad ,\quad \delta_+ = \Sigma_j \delta_j\quad ,$

$$y_{+j} = \Sigma_i y_{ij} \quad \text{and} \quad y_{++} = \Sigma_{i,j}\ y_{ij}$$

for $j = 1, \ldots, k$ and $i = 1, \ldots, N$.

Estimates for the difficulties δ_j are obtained by rescaling the ρ_j's. Estimates for the ability parameters can be obtained by solving equation (4) for the ξ's.

$$\xi_s = \frac{\delta_+}{s} \quad , \qquad s = 1,2,3, \ldots \quad . \tag{4}$$

Where s is the total score summed over all k tests. Note that we do not have a finite ability estimate for persons with a total score equal to zero.

3. Bayesian version of the multiplicative Poisson model.

A full Bayesian model for the multiplicative Poisson model has been developed by Owen (1969). He took the formulation in Eqns 1 and 2 as a starting point but adopted another definition of the parameters. The person parameter he used is inversily proportional to its counterpart in Rasch's formulation. This change was brought about to provide a complete symmetry in the interpretation of the parameter values of both test and person parameters. Prior information was specified by choosing mutually independent gamma distributions for all test and person parameters and several posterior densities, including conditional and marginal posterior densities for the individual parameters were derived. A somewhat related approach was used by Van der Ven (1969), in order to find suitable strong true-score models for the analysis of time-limit tests.

The formulation which we will use here is based on the method developed by Lindley for Bayesian estimation in the linear model, which uses the exchangeability theorem in order to specify prior knowledge concerning the parameters of the model. This approach has been succesfully used by various other authors in such different contexts as multiple regression (Lindley, 1970), the one-parameter logistic model (Swaminathan & Gifford, 1982; Jansen, 1981, Jansen & Lewis, 1983), multinomial data and contingency tables (Leonard, 1971; Lindley, 1964).

We adopt the following transformation of the original ability and difficulty parameters.

$$\theta_i = -\ln\xi_i \quad , \quad b_j = -\ln\delta_j \quad \text{and} \quad \ln\lambda_{ij} = \theta_i - b_j \tag{5}$$

Rewriting equation (1) in terms of the new transformed parameters gives

$$P\,(Y_{ij}=y_{ij}\mid\theta_i\ ,\ b_j) =$$
$$\exp(-\exp(\theta_i-b_j))\quad \exp(\theta_i-b_j)^{y_{ij}}/y_{ij}! \tag{6}$$

The likelihood function is given by

$$L\,(\underset{\sim}{\theta},\underset{\sim}{b}\mid\underset{\sim}{Y}) =$$
$$\exp\{\sum_{i,j} y_{ij}\,(\theta_i-b_j) - \sum_{i,j}\exp(\theta_i-b_j)\}/\prod_{i,j}(y_{ij}!). \tag{7}$$

Treating the $\theta_1, \ldots, \theta_N$ and $b_1, \ldots, b_k$, respectively as exchangeable, we assume the following mutually independent marginal prior distributions

$$\theta_i \sim N(\mu_\theta,\phi_\theta)\quad ,\quad i=1,\ \ldots,\ N\quad , \tag{8}$$

and

$$b_j \sim N(0,\phi_b)\quad ,\quad j=1,\ \ldots,\ k\quad , \tag{9}$$

where ϕ_θ, ϕ_b and are supposed to be known constants. By choosing the prior mean of the difficulty parameters equal to zero, we incorporate the restriction necessary for identifiability as part of the model.

A more convenient and adaptable formulation is based on the following. Let

$$\theta_i^* = \theta_i - \mu_\theta\quad . \tag{10}$$

Then the prior distributions can be formulated as follows

$$\theta_i^* \sim N(0,\phi_\theta) \tag{11}$$

and

$$b_j \sim N(0,\phi_b)\quad . \tag{12}$$

The prior variances are assumed to be known, as in the previous formulation. The prior μ_θ can also be assumed known or be retained as an explicit parameter. In the latter case we assume that μ_θ is distributed uniformly over the real line. Then, the posterior distribution of $(\underset{\sim}{\theta}^*,\underset{\sim}{b},\mu_\theta)$ is, by combining the likelihood in (7) and the priors, obtained as

$$p(\underset{\sim}{\theta}^*,\underset{\sim}{b},\mu_\theta|\underset{\sim}{y},\phi_\theta,\phi_b) \propto \exp\{\sum_{i,j} y_{ij}(\theta_i^*+\mu_\theta-b_j) - \sum_{i,j} \exp(\theta_i^*+\mu_\theta-b_j)\} *$$

$$\phi_\theta^{-\frac{1}{2}n} \exp(-\tfrac{1}{2}\sum_i \frac{\theta_i^{*2}}{\phi_\theta})\ \phi_b^{-\frac{1}{2}k} \exp(-\tfrac{1}{2}\sum_j \frac{b_j^2}{\phi_b}) \quad . \tag{13}$$

To simplify the notation we will in the following derivations refer to μ_θ as μ.

Estimates for the parameters can be obtained by taking the logarithm of the posterior density (13), by differentiating with respect to μ, θ_i and b_j, setting their derivatives equal to zero and solving the resulting equations:

$$\sum_{i,j} y_{ij} - \sum_{i,j} \exp\,(\theta_i^* + \mu - b_j) = 0 \quad , \tag{14}$$

$$-\sum_j y_{ij} + \sum_j \exp\,(\theta_i^* + \mu - b_j) + \theta_i^* / \phi_\theta = 0 \quad , \tag{15}$$

$$-\sum_i y_{ij} + \sum_i \exp\,(\theta_i^* + \mu - b_j) - b_j / \phi_b = 0 \quad , i = 1, .. , N \quad j = 1, .. , K. \tag{16}$$

As is obvious from equations (14) through (16) an iterative procedure is required to solve them. A computer program, which performs the necessary computations, using a simplified Newton-Raphson procedure has been written by the author*.

4. Examples.

To better understand the possibilities for working with the approach described in the previous sections, some concrete illustrations will be given. The first example is based on an artificial data set. This data set was obtained by generating the responses of 100 individuals, who were randomly sampled from a standard normal distribution, on three tests according to the Poisson model. The difficulties of the tests were set to

* available at request

resp. $b_1 = -0.5$, $b_2 = 0.0$ and $b_3 = 0.5$. For tests consisting of 50 items this correspondends to expected error percentages of 3, 2 and 1 percent for a "standard" individual with an ability parameter value of $\theta = 0$. This data set was analysed and both Bayesian and maximum likelihood estimates for the test and person parameters were obtained. For the Bayesian analysis the prior variances were both set as one. Maximum likelihood estimates for the relative difficulties were obtained by using formula (3). From these we derived b - estimates by taking logarithms changing signs and rescaling to a mean of zero.

The results are shown in table 1. The Bayesian estimates show the typical regression effect that also characterizes Bayesian parameter estimates obtained in other situations, as for example for the better known one - parameter logistic model (Jansen & Lewis, 1983). The regression effect is more marked for the person than for the test parameters, due to the asymmetry in the sample information (N much larger than k). Another illustration of the regression effect is given by figure 1.

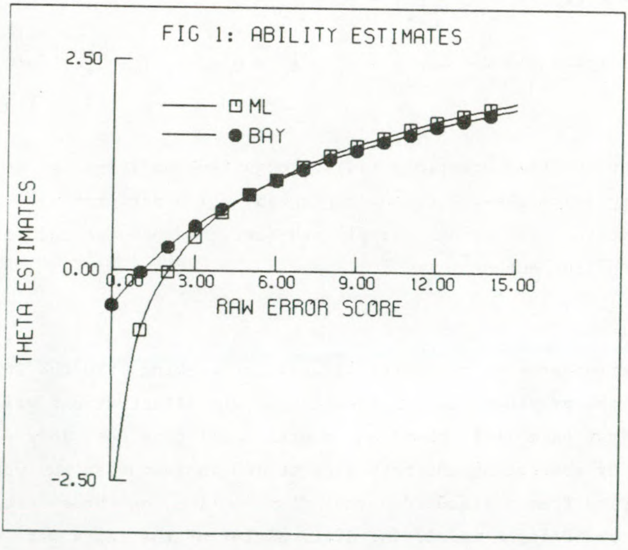

Table 1. Maximum Likelihood (ML) and Bayesian estimates (BAY) for the Person and Test Parameters for the Artificial Data Set. (N = 100, k = 3 and Number of Errors = 0 to 15).

Raw Score	Theta Estimates*		Test Nr.	Beta estimates	
	ML	BAY		ML	BAY
0	-----	-1.16	1	-0.507	-0.503
1	-1.19	-0.72	2	-0.028	-0.027
2	-0.49	-0.38	3	0.535	0.530
3	-0.09	-0.10			
4	0.20	0.13			
5	0.42	0.33			
6	0.61	0.50			
7	0.76	0.64			
8	0.89	0.77			
9	1.01	0.89			
10	1.12	1.00			
11	1.21	1.09			
12	1.30	1.18			
13	1.38	1.27			
14	1.45	1.34			
15	1.52	1.41			

* $\theta = \theta^{*} + \mu$

Another interesting feature of the Bayesian method is that we can obtain finite ability estimates for persons with a perfect score (raw error score of zero). This is important since, with easy tests as are required by the model, a relatively large proportion of the sample consists of just these persons.

For the second sample we used an empirically obtained data set. This data set consisted of the response of 716 pupils on a spelling test that was administered twice under different format conditions. The first was the "regular" administration where the test leader dictated the words that had to be written down by the pupils. Under the second format condition the pupils had to guess the words from pictures and then try to spell them correctly. The research questions involved the importance of auditive clues in spelling. The total test length was 30 words. The test responses were analysed using the same prior specifications as in the first example.

Maximum likelihood and Bayesian estimates for the person and test parameters can be found in table 2. The results are comparable to those reported in table 1. The Bayesian estimates of the test parameters are somewhat regressed to a common mean of zero compared to the maximum likelihood estimates, though the differences are practically negligeable. The regression effect for the person parameter estimates is more pronounced, especially for the raw error scores in the range from 0 to 3.

Table 2. Maximum Likelihood and Bayesian Estimates for the Test and Person Parameters of the Data Obtained by Administering the Same Spelling Test under Two Different Format Conditions. (N = 716 , k = 2).

Raw Score	Theta Estimates*		Test Nr.	Beta estimates	
	ML	BAY		ML	BAY
0	-----	-0.41			
1	-0.70	-0.02	1	-0.116	-0.115
2	-0.01	0.28	2	0.116	0.115
3	0.40	0.53			
4	0.69	0.73			
5	0.91	0.91			
6	1.09	1.07			
7	1.25	1.21			
8	1.38	1.33			
9	1.50	1.44			
10	1.60	1.54			
11	1.70	1.63			
12	1.79	1.72			
13	1.87	1.80			
14	1.94	1.87			
15	2.01	1.94			

* $\theta = \theta^{*} + \mu$

5. Discussion.

The results presented in the previous sections show that Bayesian estimation procedures are potentially useful for estimating the parameters of the multiplicative Poisson model. A from a more practical point of view attractive feature of the model is, that we are able to provide finite ability estimates for individuals having a perfect score. In other contexts it has been shown that Bayesian estimates are in general more accurate than ML estimates, especially in the case where the number of observations available is relatively small (Swaminathan & Gifford, 1982). More research however is needed to substantiate this claim.

A drawback of the approach used here is, that we have to be rather precise in the specification of our prior beliefs. It still has to be investigated how robust the procedure is against differences in prior specifications. Is is also possible to extend the model to the case were the prior variances are not assumed to be known exactly but where the prior information about these variances is also specified in the form of prior distributions. This socalled hierarchical two-stage Bayesian model is used by Jansen & Lewis (1983) and Swaminathan & Gifford (1982) for the one-parameter logistic model. Despite its advantages, it also tends to produce anomalous results in some situations, for which the reasons are not, at this moment. fully understood.

References

Andersen, E.B., Discrete Statistical Models with Social Science Applications. North-Holland, Amsterdam, 1980.

Jansen, G.G.H., A Bayesian Latent Trait Model for Binary Test Items. Paper presented at the 12th European Mathematical Psychologic Group Meeting, Birmingham, 1981.

Jansen, G.G.H. & Lewis, Ch., Problems in Estimating Abilities for the Rasch Model. Paper presented at the 1983 European Psychometric Society Meeting.

Jansen, P.G.W., Rasch Analysis of Attitudinal Data. Unpublished Doct. Diss., Nijmegen, 1983.

Leonard, T., Bayesian Methods for Multinomial Data. ACT Technical Bulletin no. 4. The American Testing Program, Iowa City, 1971.

Lindley, D.V., A Bayesian Solution for Some Educational Prediction Problems; III. Research Bulletin RB-70-33. ETS, Princeton, 1970.

Lindley, D.V., The Bayesian Analysis of Contingency Tables. Ann. of Math. Statist, 1964, 35, 1622-43.

Lindley, D.V. & Smith, A.F.M., Bayesian Estimates for the Linear Model. Journal of the Royal Statist. Soc.(Series B), 1972, 34, 1-41.

Lord, F.M. & Novick, M.R., Statistical Theories for Mental Test Scores. Addison-Wesley, Reading, Mass., 1968.

Rasch, G., Probabilistic Models for some Intelligence and Attainment Tests. Copenhagen, 1960. (Reprint: The University of Chicago Press, Chicago, 1980).

Owen, R.J., A Bayesian Analysis of Rasch's Multiplicative Poisson Model for Misreadings. (Res. Bull. 69-64) E.T.S., Princeton, 1969.

Ven, A.H.G.S. van der, The Binomial Error Model Applied to Time-Limit Tests. Unpublished doctoral dissertation, Nijmegen, 1969.

Wright, B.D. & Stone, M.H., Best Test Design. MESA Press, Chicago, 1979.

Ontvangen: 5-10-1983