Bayesian multidimensional nominal response model for observer study of radiologists

Nishio, Mizuho ; Kobayashi, Daigo ; Matsuo, Hidetoshi ; Urase, Yasuyo ; Nishioka, Eiko ; Murakami, Takamichi

(Citation)
Japanese Journal of Radiology,41(4):449-455

(Issue Date)
2022-12-05

(Resource Type)
journal article

(Version)
Accepted Manuscript

Title

Bayesian Multidimensional Nominal Response Model for Observer Study of Radiologists

Article Type

Original Manuscript

Abstract

Purpose: This study proposes a Bayesian multidimensional nominal response model (MD-NRM) to statistically analyze the nominal response of multiclass classifications.

Materials and methods: First, for MD-NRM, we extended the conventional nominal response model to achieve stable convergence of the Bayesian nominal response model and utilized multidimensional ability parameters. We then applied MD-NRM to a 3-class classification problem, where radiologists visually evaluated chest X-ray images and selected their diagnosis from one of the three classes. The classification problem consisted of 150 cases, and each of the six radiologists selected their diagnosis based on a visual evaluation of the images. Consequently, 900 (= 150×6) nominal responses were obtained. In MD-NRM, we assumed that the responses were determined by the softmax function, the ability of radiologists, and the difficulty of images. In addition, we assumed that the multidimensional ability of one radiologist were represented by a 3×3 matrix. The latent parameters of the MD-NRM (ability parameters of radiologists and difficulty parameters of images) were estimated from the 900 responses. To implement Bayesian MD-NRM and estimate the latent parameters, a probabilistic programming language (Stan, version 2.21.0) was used.

Results: For all parameters, the Rhat values were less than 1.10. This indicates that the latent parameters of the MD-NRM converged successfully.

Conclusion: The results show that it is possible to estimate the latent parameters (ability and difficulty parameters) of the MD-NRM using Stan. Our code for the implementation of the MD-NRM is available as open source.

Short Abstract

To statistically analyze the nominal response of multiclass classifications, this study proposes a Bayesian multidimensional nominal response model (MD-NRM). With MD-NRM, it is possible to statistically analyze the nominal response of multiclass classifications obtained by radiologists.

Keywords

item response theory; nominal response model; probabilistic programming language; chest X-ray

1. Introduction

Item response theory (IRT), also known as latent trait theory, is a statistical model/paradigm used for analyzing tests [1–3]. The IRT can be used to evaluate test items and test takers quantitatively and to perform quality assurance of tests. In addition, IRT can be used to equate the test results. Statistical analysis of the IRT is performed using the results of binary classifications (e.g., correct and incorrect answers to test items).

IRT can be implemented using probabilistic programming languages (e.g., WinBUGS, JAGS, and Stan) [4–6]. Extension of IRT is also possible using probabilistic programming languages. For example, the graded response model and nominal response model (NRM) as multiclass extensions of IRT were implemented in Stan [7]. In previous studies, the Bayesian IRT and NRM models were implemented in Stan, and statistical analyses were performed using them [7,8].

In medical diagnosis, the results of various diagnostic procedures are frequently defined as binary classifications: the presence or absence of a disease, the presence or absence of distant metastasis of cancer, the prediction results of cancer death, etc. However, not all medical diagnoses can be expressed as binary classifications, and multiclass classifications are sometimes necessary. For example, the T, N, and M factors of the TNM classification system are expressed as multiclass classifications for many cancer types [9–11]. The statistical analysis of these multiclass classification results requires multiclass analysis (such as NRM).

IRT can be applied to medical diagnosis, by considering (i) the patient as the test item, (ii) the doctor as the test taker, and (iii) the results of the binary classification obtained through medical diagnosis as test results. For example, a previous study used Bayesian IRT to statistically evaluate improvements in radiologists' diagnostic performance [8]. However, to the best of our knowledge, Bayesian NRM has not been applied to the medical diagnosis of multiclass classifications.

The purpose of this study was to use the Bayesian NRM for multiclass medical diagnosis. Although a Stan implementation of the Bayesian NRM has already been reported [7], it does not work stably [12]. Therefore, for the Bayesian NRM, we extended the conventional NRM in this study. There are two major differences between conventional and our extended NRM. First, while conventional NRM is frequently implemented by extending the 2PL-IRT, our NRM is implemented by extending the 1PL-IRT. We speculate that, because the existing Stan implementation of NRM extends the 2PL-IRT, the existing implementation produced unstable results. The second change in our NRM is to evaluate the ability of test takers using multidimensional parameters rather than a single parameter. Since this study used the results of multiclass classifications, it is natural for test-takers' abilities to be evaluated based on multidimensional parameters rather than a single parameter. Our Stan implementation of the Bayesian NRM proposed in this study is disclosed as open source through GitHub (https://github.com/jurader/MDNRM).

2. Materials and methods

2.1. Binary and multiclass classification

Generally, the results of binary classification are summarized by a 2×2 confusion matrix between the ground truth and prediction. This 2×2 confusion matrix is composed of true positive, true negative, false positive, and false negative. Figure 1 shows the 2×2 confusion matrix. In a previous study [8], Bayesian IRT was applied by considering true positive and true negative as "correct" and false positive and false negative as "incorrect".

In general, the results of multiclass classification are summarized by an N×N confusion matrix. As an example, a 3×3 confusion matrix is shown in Figure 2. Here, it is possible to define the diagonal values of the N×N confusion matrix as "correct" and the other values as "incorrect" and apply Bayesian IRT to these values. However, this method does not allow class-by-class evaluations. To solve this problem, conventional NRM was extended in this study.

2.2. IRT and NRM

IRT is a statistical model for analyzing the results of binary classifications. There are several types of IRT models. Here, 1PL-IRT and 2PL-IRT are shown [1]. In addition, conventional NRM, as a multiclass extension of 2PL-IRT, is also described [7]. Finally, an extended NRM is proposed. Hereafter, “case” and “radiologist” are used instead of “test item” and “test taker” of IRT, respectively.

2.2.1. 1PL-IRT

In 1PL-IRT, one parameter (β_i) is used to represent case i, and another parameter (θ_j) is used to represent radiologist j. The following equations represent the 1PL-IRT.

$$\Pr(r_{ij} = 1) = \frac{1}{1 + \exp\left(-z_{ij}\right)}$$

$$z_{ij} = \theta_j - \beta_i$$

Here,

- r_{ij} is the response (prediction) of radiologist j to case i,
- $r_{ij} = 1$ means that the response r_{ij} is correct,
- $\Pr(r_{ij} = 1)$ represents the probability that the response of radiologist j to case i is correct,
- β_i is the difficulty parameter of case i,
- θ_j is the ability parameter of the radiologist j.

The equation of 1PL-IRT indicates that for converting logit (z_{ij}) to probability, 1PL-IRT uses the sigmoid function, which is used for logistic regression. In the 1PL-IRT, $\theta_j\ and\ \beta_i$ are estimated based on the prediction results of the radiologists.

2.2.2. 2PL-IRT

In the 2PL-IRT, two parameters (α_i and β_i) are used to represent case *i*. The following equations represent 2PL-IRT.

$$\Pr(r_{ij} = 1) = \frac{1}{1 + \exp\,(-z_{ij}\,)}$$

$$z_{ij} \;=\; \alpha_i(\theta_j - \beta_i)$$

Here,

- α_i and β_i are the discrimination and difficulty parameters of case *i*.

In 2PL-IRT, θ_j, α_i, and β_i are estimated based on the prediction results of radiologists as in 1PL-IRT.

2.2.3. Conventional NRM

Conventional NRM can be viewed as an extension of 2PL-IRT for multiclass classifications. The following equation represents the conventional NRM.

$$\Pr(r_{ij} = k) = \frac{\exp(\alpha_{ik}\theta_j + \beta_{ik})}{\sum_{h=1}^{c} \exp(\alpha_{ih}\theta_j + \beta_{ih})}$$

Here,

- $\Pr(r_{ij} = k)$ represents the probability that the response of radiologist j to case i is class k,
- The number of classes is c,
- α_{ik} and β_{ik} are the two parameters of case i on class k,
- θ_j is the ability parameter of radiologist j.

The equation of conventional NRM can be rewritten as follows:

$$\Pr(r_{ij} = k) = \frac{\exp(z_{ijk})}{\sum_{h=1}^{c} \exp(z_{ijh})}$$

$$z_{ijk} = \alpha_{ik}\theta_j + \beta_{ik}$$

This means that the NRM uses the softmax function to convert logit (z_{ijk}) to probability. Based on the results of the multiclass classification, $\alpha_{ik}, \beta_{ik}, and\ \theta_j$ are estimated. Although a previous study provided the Stan code for conventional NRM (Bayesian NRM) [7], its calculation results were unstable [12].

2.2.4. Extended NRM (1PL-NRM and multidimensional NRM (MD-NRM))

To stabilize the results of the Bayesian NRM, we extended the conventional NRM. First, the discrimination parameter (α) is removed from conventional NRM. In other words, our extended NRM is based on 1PL-IRT. This deletion stabilizes the calculation results of our Bayesian NRM. Hereafter, this extension of NRM is referred to as 1PL-NRM. The following equations represent 1PL-NRM.

$$\Pr(r_{ij} = k) = \frac{\exp(z_{ijk})}{\sum_{h=1}^{c} \exp(z_{ijh})}$$

$$z_{ijk} = \theta_j - \beta_{ik}$$

In this equation, $\theta_j - \beta_{ik}$ is used instead of $\theta_j + \beta_{ik}$, to strengthen the meaning of the difficulty parameter of β_{ik}.

In both conventional NRM and 1PL-NRM, the probability of $r_{ij} = k$ (the response of radiologist *j* to case *i* is class *k*) depends on only one parameter of the radiologist (θ_j). This assumption is unnatural, and the equations of both conventional NRM and 1PL-NRM indicate that the difficulty parameter of the case has a greater influence on which class a radiologist chooses for the diagnosis than their ability parameter. Because there are *c* classes in multiclass classification, the probability should be dependent on several parameters of the radiologists. To address this problem, we extended 1PL-NRM.

Herein, we propose a multidimensional NRM (MD-NRM) based on 1PL-NRM, which is represented by the following equations:

$$\Pr\left(r_{ij} = t \mid ground\ truth\ of\ case\ i = s\right) = \frac{\exp\left(z_{ijst}\right)}{\sum_{h=1}^{c} \exp\left(z_{ijsh}\right)}$$

$$z_{ijst} = \theta_{jst} - \beta_{is}$$

- The number of classes is *c*,
- *s* is the ground truth of case *i*,
- $r_{ij} = t$ means that the response of radiologist *j* to case *i* is *t*,
- θ_{jst} is the ability parameter of radiologist *j* in class *t* when the ground truth of the case is *s* ($\theta_{j..}$ can be represented by a matrix (*c*×*c*)),
- β_{is} is the difficulty parameter of case *i* on class *s*.

This extended NRM (MD-NRM) means that the multidimensional ability parameters of a radiologist are represented by a matrix, which corresponds to the confusion matrix of multiclass classification.

2.2.5. Relationship between IRT and NRM

The 2PL-IRT can be derived from the conventional NRM as a special case [13]. Here, we use only two classes (*k* or *k'*) for the NRM. Additionally, we consider the conditional probability for a response in class *k* given that the response is one of classes *k* or *k'*. Generally, the conditional probability is given by the following equation:

$$\Pr(r = k \mid k, k') = \frac{\Pr(r = k)}{\Pr(r = k) + \Pr(r = k')}$$

Here, the conventional NRM is used for Pr(*k*) or Pr(*k'*). The conditional probability is represented by the following equation (case *i* and radiologist *j* are omitted for brevity):

$$\Pr(k \mid k, k') = \frac{1}{1 + \exp\left(-\alpha_k^c \theta + {\beta'}_k^c\right)}$$

$${\beta'}_k^c = \beta_{k'} - \beta_k$$

$$\alpha_k^c = \alpha_k - \alpha_{k'}$$

Further, this equation can be converted to the following equations.

$$\Pr(k \mid k, k') = \frac{1}{1 + \exp\left(-\alpha_k^c(\theta - \beta_k^c)\right)}$$

$$\beta_k^c = \frac{{\beta'}_k^c}{\alpha_k^c}$$

These equations represent the 2PL-IRT. This derivation (conventional NRM to 2PL-IRT) can be applied to our 1PL-NRM; In our 1PL-NRM, the conditional probability is represented by the 1PL-IRT. This is one of the major reasons for using the extended NRM.

2.3. Experiments

Our institutional review board approved this retrospective study and waived the requirement for informed consent.

It is assumed that MD-NRM can be applied to data from medical diagnoses of multiclass classification. We applied MD-NRM to the classification results obtained by radiologists in a previous study [14]. In the previous study, there were three classes of medical diagnosis: novel coronavirus pneumonia (COVID), non-novel-coronavirus pneumonia (PNEUMONIA), and normal (NORMAL). Therefore, the classification results were summarized as a 3×3 confusion matrix. In total, 150 cases (50 COVID, 50 PNEUMONIA, and 50 NORMAL) were reported in the previous study. From the three classes, six radiologists determined their diagnoses based on visual evaluation of chest X-ray images. Therefore, 900 (150×6) nominal responses were obtained. The Supplementary material shows the ground truth of the case and the classification results of the six radiologists.

In this study, MD-NRM was applied to multiclass classification results to analyze the ability of six radiologists for the three classes. For all ability and difficulty parameters, a normal distribution with mean = 0 and standard deviation = 2 was used as the prior distribution. The following parameters were used for sampling in Stan: chains=8, iter=8000, warmup=4000, thin=1, adapt_delta = 0.9, and max_treedepth = 15. The convergence check of the MD-NRM was performed by evaluating the Rhat values of all parameters [7,8,15]. Since we focused on the ability parameters, the estimation results of the difficulty parameters were omitted in this study. We used the following software to implement MD-NRM: R, version 4.1.1; Stan, version 2.21.0; rstan, version 2.21.2; shinystan, version 2.5.0; and tidybayes, version 3.0.1.

3. Results

Figure 3 shows the convergence check of the Stan results for all the parameters. The Rhat values for all parameters (difficulty and ability parameters) were less than 1.10, which means that the MD-NRM converged successfully [7,8,15]. Thus, 1PL-NRM is useful for stabilizing Bayesian NRM.

Table 1 shows the estimation results of the MD-NRM for the ability parameters of Radiologist 1. In the MD-NRM, radiologist ability is represented in a 3×3 matrix, so one radiologist has nine ability parameters. In Table 1, if θ_{111}, θ_{122}, and θ_{133} are high, the diagnostic performance of radiologist 1 is high for NORMAL, PNEUMONIA, and COVID.

Figure 4 shows a 3 × 3 ability matrix created from the median values of the ability parameters of radiologist 1 in Table 1. In Figure 4, if the diagonal values of the ability matrix are high, the diagnostic ability of radiologist 1 is also high. In contrast, if the non-diagonal values of the ability matrix are low, the diagnostic ability is high.

Figure 5 (A)–(C) show the summary of θ_{i11} (i = 1, 2, …, 6), θ_{i22} (i = 1, 2, …, 6), and θ_{i33} (i = 1, 2,

…, 6) for radiologists 1–6 for NORMAL, PNEUMONIA, and COVID, respectively. The figures in the Supplementary Materials show all the ability parameters of radiologists 1–6.

Our Stan and R codes of the MD-NRM are shown in the Supplementary materials. In addition, these codes were disclosed as open source through GitHub (https://github.com/jurader/MDNRM).

4. Discussion

Using the MD-NRM, stable convergence in the Bayesian NRM calculation was achieved with our Stan code. This is a significant improvement because stable convergence cannot be obtained with the conventional NRM Stan code [7]. In addition, compared to the conventional NRM, MD-NRM makes it possible to determine the multidimensional ability parameters of a doctor. Furthermore, this extension of multidimensional parameters allows MD-NRM to provide a more detailed evaluation of a doctor's ability rather than a single parameter.

In IRT and NRM, parameters (ability and difficulty) are evaluated in the latent space [1,8]. Because of this characteristic, IRT is known as the latent trait theory. Logistic regression is frequently used to analyze the results of medical diagnoses [16]. In logistic regression, the effect of covariates on the latent space is statistically evaluated. Therefore, IRT, NRM, and logistic regression share similar functionalities in the evaluation of the parameters.

In receiver operating characteristic (ROC) analysis, it is common to assign a multilevel score for each case to statistically analyze the results of binary classifications [17,18]. For example, a radiologist's ability to differentiate between benign and malignant tumors is frequently analyzed by assigning a score to each case on a 5-point scale [19]. However, there is no commonly used method of ROC analysis for multiclass classification results. Instead, for the analysis of multiclass classifications, it is relatively frequent to perform multiple ROC analyses of binary classifications in a one-vs-rest fashion [20]. On the other hand, MD-NRM can analyze the results of multiclass classifications by a single analysis, which is a major difference between MD-NRM and ROC analysis. In addition, MD-NRM does not require the multilevel scoring (please see the data of 150 cases in the Supplementary material), which is another major difference from ROC analysis.

As in a previous study [8], it is possible to introduce covariates into MD-NRM. This extension makes it possible to quantitatively evaluate the effect of covariates using MD-NRM; for example, the improvement in diagnostic performance with various experiences of doctors or with different diagnostic modalities/examinations.

This study had several limitations. First, the data from 150 patients were used. It has not been evaluated whether stable convergence will be obtained in MD-NRM with smaller data. Second, although an extension of the MD-NRM (MD-NRM with covariates) is proposed in this paper, no actual experiments have been conducted. We expect that the extension of MD-NRM will be investigated in the future based on our open-source code. Third, because MD-NRM has more latent parameters than 1PL-

NRM, it will be difficult to obtain stable convergence in MD-NRM, compared with 1PL-NRM.

In conclusion, the MD-NRM was proposed as a statistical analysis method for multiclass classification. The MD-NRM achieved successful convergence of parameter estimation in the Bayesian NRM. In addition, using the MD-NRM, a more detailed evaluation of the doctor's ability was possible by using multidimensional ability parameters.

References

1. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. Clin Ther. Clin Ther; 2014;36:648–62.

2. Embretson SE, Reise SP. Item response theory for psychologists. Item Response Theory for Psychologists. Taylor and Francis; 2000.

3. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. Med Care. 2000;38:II28-42.

4. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. Journal of Statistical Software. 2017;76:1–32.

5. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing. 2000;10:325–37.

6. Depaoli S, Clifton JP, Cobb PR. Just Another Gibbs Sampler (JAGS). Journal of Educational and Behavioral Statistics. 2016;41:628–49.

7. Luo Y, Jiao H. Using the Stan Program for Bayesian Item Response Theory: Educational and Psychological Measurement. SAGE PublicationsSage CA: Los Angeles, CA; 2017;78:384–408.

8. Nishio M, Akasaka T, Sakamoto R, Togashi K. Bayesian Statistical Model of Item Response

Theory in Observer Studies of Radiologists. Academic Radiology. Elsevier; 2020;27:e45–54.

9. Lim W, Ridge CA, Nicholson AG, Mirsadraee S. The 8th lung cancer TNM classification and clinical staging system: review of the changes and clinical implications. Quantitative Imaging in Medicine and Surgery. AME Publications; 2018;8:709–18.

10. Kalli S, Semine A, Cohen S, Naber SP, Makim SS, Bahl M. American joint committee on cancer's staging system for breast cancer, eighth edition: What the radiologist needs to know. Radiographics. Radiological Society of North America Inc.; 2018;38:1921–33.

11. Allen PJ, Kuk D, Castillo CF del, Basturk O, Wolfgang CL, Cameron JL, et al. Multi-institutional Validation Study of the American Joint Commission on Cancer (8th Edition) Changes for T and N Staging in Patients With Pancreatic Adenocarcinoma. Ann Surg. Ann Surg; 2017;265:185–91.

12. Non convergence issue on polytomous IRT model - Modeling - The Stan Forums [Internet]. [cited 2022 Jul 12]. Available from: https://discourse.mc-stan.org/t/non-convergence-issue-on-polytomous-irt-model/12576

13. Thissen D, Cai L, Bock RD. The Nominal Categories Item Response Model. Handbook of Polytomous Item Response Theory Models. Routledge Handbooks Online; 2010.

14. Nishio M, Kobayashi D, Nishioka E, Matsuo H, Urase Y, Onoue K, et al. Deep learning model for the automatic classification of COVID-19 pneumonia, non-COVID-19 pneumonia, and the

healthy: a multi-center retrospective study. Sci Rep. Sci Rep; 2022;12:8214.

15. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Bayesian Data Analysis. Chapman and Hall/CRC; 2013.

16. Kleinbaum DG, Klein M. Logistic Regression. New York, NY: Springer; 2010.

17. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. Radiology. Radiology; 2003;229:3–8.

18. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry. Oxford Academic; 1993;39:561–77.

19. Park SH, Goo JM, Jo CH. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. Korean Journal of Radiology. Korean Society of Radiology; 2004;5:18.

20. Receiver Operating Characteristic (ROC) — scikit-learn 1.1.1 documentation [Internet]. [cited 2022 Jul 12]. Available from: https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html#plot-roc-curves-for-the-multiclass-problem

Tables

Table 1

Result of MD-NRM for ability parameters of radiologist 1.

Note: 2.50%, 25%, 50%, 75%, and 97.5% are percentile values of the estimated parameters. For example, the pair of -0.137 at 2.50% and 4.665 at 97.5% represents a 95% credible interval of θ_{111}.

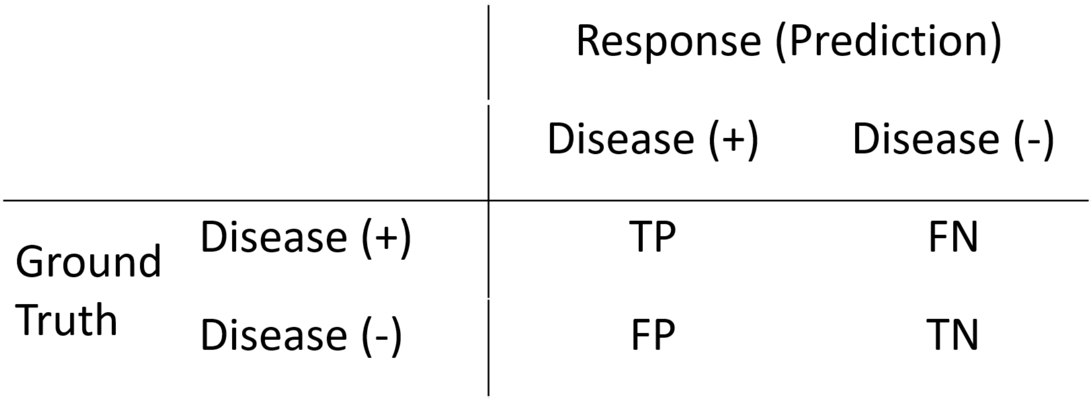
Parameters	2.50%	25%	50% (median)	75%	97.50%
θ_{111}	-0.137	1.46	2.288	3.117	4.665
θ_{112}	-2.835	-1.242	-0.403	0.452	2.035
θ_{113}	-4.44	-2.763	-1.896	-1.009	0.655
θ_{121}	-4.659	-2.987	-2.137	-1.262	0.372
θ_{122}	-0.796	0.762	1.589	2.418	3.969
θ_{123}	-1.845	-0.268	0.556	1.387	2.959
θ_{131}	-3.652	-2.077	-1.249	-0.422	1.155
θ_{132}	-2.572	-1.049	-0.233	0.593	2.122
θ_{133}	-0.831	0.682	1.491	2.297	3.824

Figure and figure legends

Figure 1

2×2 confusion matrix between ground truth and prediction.

Abbreviations: TP, true positive; TN, true negative; FP, false positive; FN, false negative.

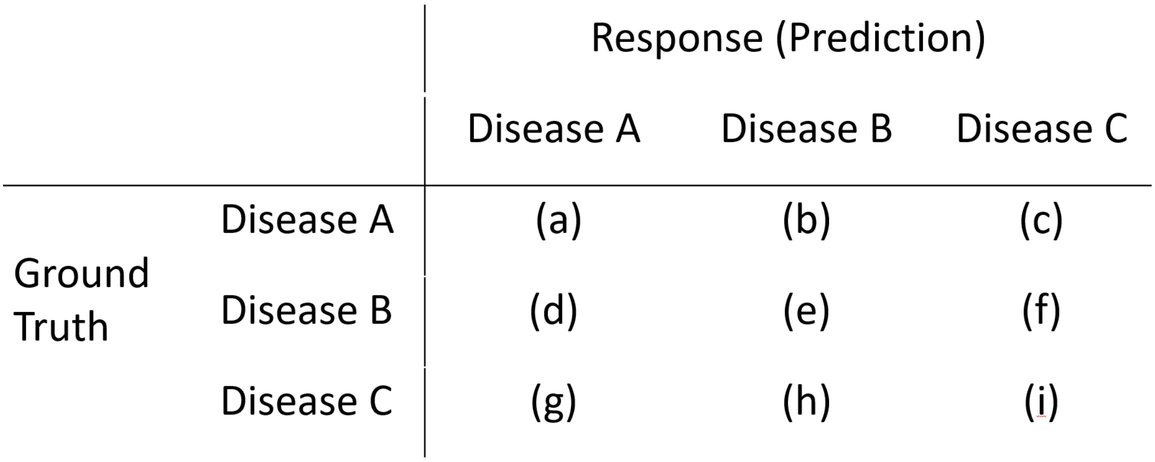


		Response (Prediction)	
		Disease (+)	Disease (-)
Ground Truth	Disease (+)	TP	FN
	Disease (-)	FP	TN

Figure 2

3×3 confusion matrix between ground truth and prediction for 3-class classifications.

Note: (a)-(i) represent the frequencies of cells.



		Response (Prediction)		
		Disease A	Disease B	Disease C
Ground Truth	Disease A	(a)	(b)	(c)
	Disease B	(d)	(e)	(f)
	Disease C	(g)	(h)	(i)

Figure 3

Evaluation of convergence in MD-NRM.

Note: For all parameters, their Rhat values are less than 1.10.

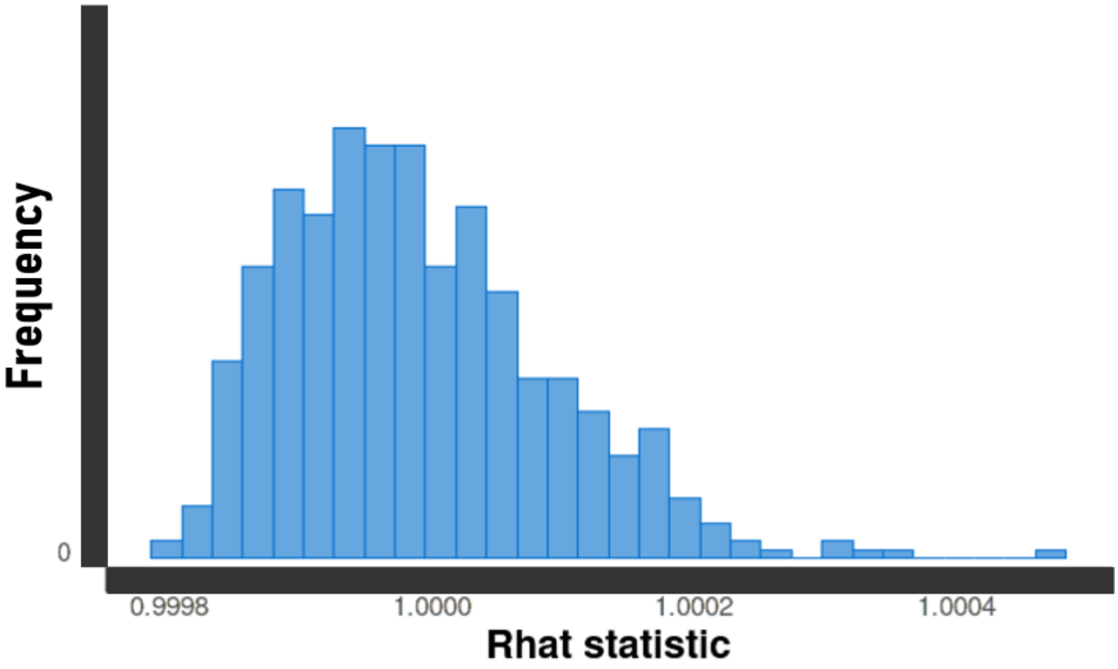
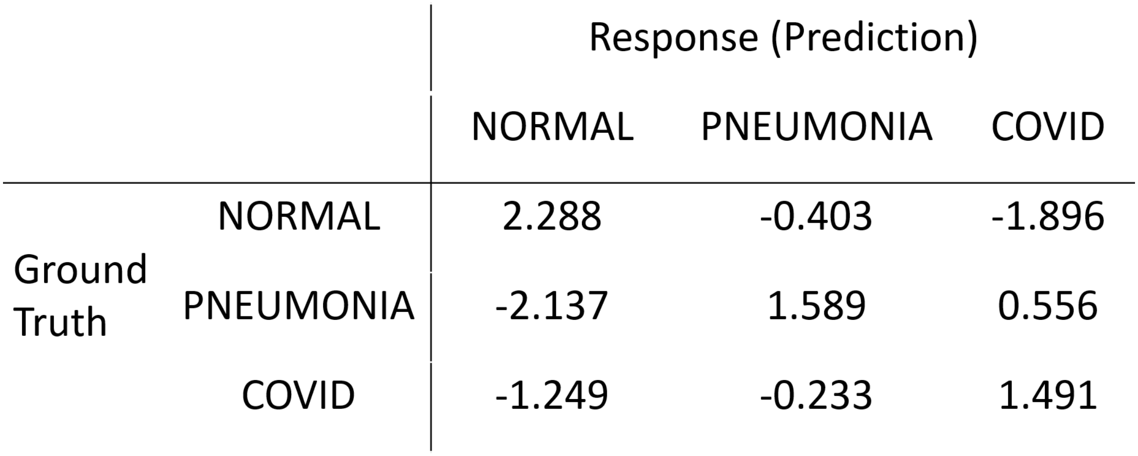


Figure 4

Ability matrix of radiologist 1.

Note: This matrix was created from the median values of the ability parameters of radiologist 1 in Table 1.



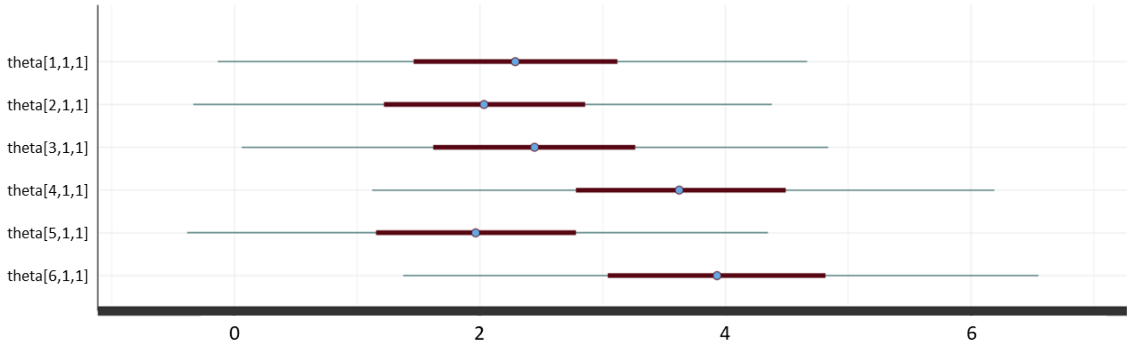
		Response (Prediction)		
		NORMAL	PNEUMONIA	COVID
Ground Truth	NORMAL	2.288	-0.403	-1.896
	PNEUMONIA	-2.137	1.589	0.556
	COVID	-1.249	-0.233	1.491

Figure 5

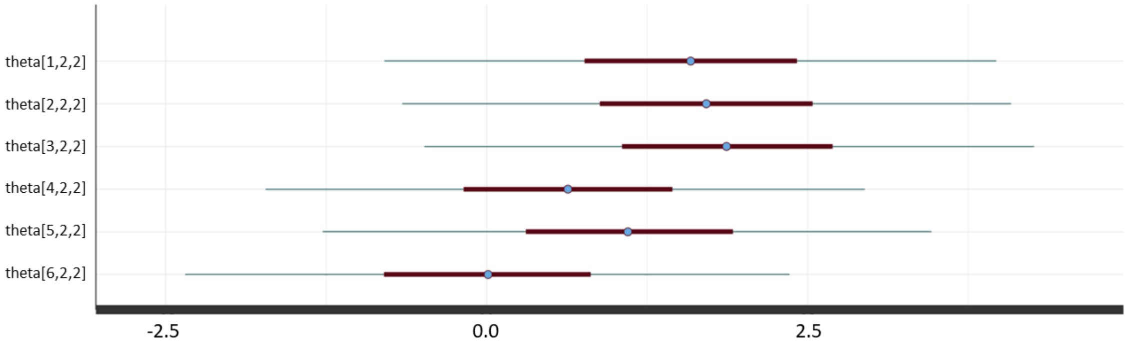
Summary of ability parameters of six radiologist. (A) θ_{i11} (i = 1, 2, …, 6) for NORMAL, (B) θ_{i22} (i = 1, 2, …, 6) for PNEUMONIA, (C) θ_{i33} (i = 1, 2, …, 6) for COVID.

Note: In (A)–(C), "theta[i,j,k]” represents θ_{ijk}. For example, “theta[1,1,1]” represents θ_{111}. Circles, thick bars, and thin bars represent the median values, 50% credible interval (interquartile range), and 95% credible interval, respectively.

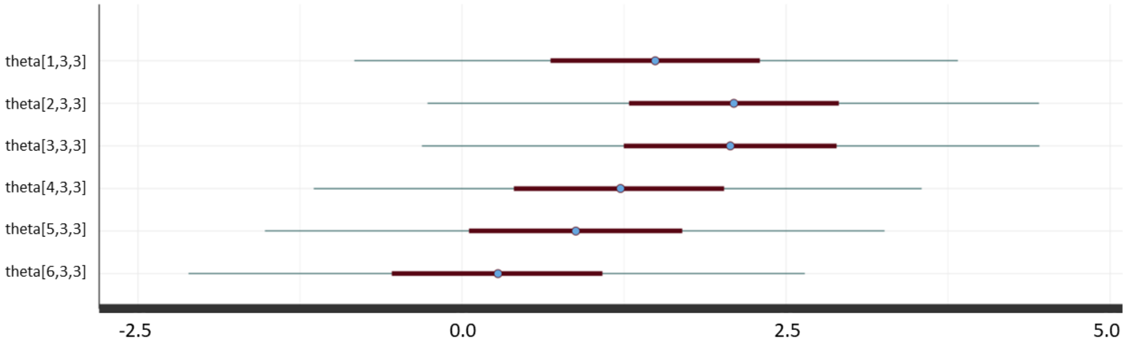
(A)



(B)


theta[1,2,2]
theta[2,2,2]
theta[3,2,2]
theta[4,2,2]
theta[5,2,2]
theta[6,2,2]
-2.5
0.0
2.5

(C)


theta[1,3,3]
theta[2,3,3]
theta[3,3,3]
theta[4,3,3]
theta[5,3,3]
theta[6,3,3]
-2.5
0.0
2.5
5.0