

SDT におけるベイズ的データ分析

Bayesian Data Analysis in SDT¹⁾

OKAMOTO Yasuharu

岡 本 安 晴

[Abstract] Bayesian data analyses in equal- or unequal-variance Gaussian signal detection theory (SDT) are proposed. Lee (2008) proposed a Bayesian analysis in SDT, but his method is useful for a simple equal-variance SDT with two response categories, yes and no, and uses a Markov chain Monte Carlo (MCMC) method, which requires some experience by a user. Before Lee's proposal, Okamoto (2005) had proposed a simple grid method of Bayesian analysis in the simple SDT, and Okamoto (2006) had proposed an MCMC method for an unequal-variance SDT model with more than two response categories. The grid method by Okamoto (2005) does not require a special experience and a user can obtain an appropriate result of Bayesian analysis by adjusting values displayed on the form window. The MCMC method by Okamoto (2006) can be applied to data on SDT, which may have unequal-variances. Evidence of unequal variance is presented in Green and Swets (1966). Both of Okamoto's methods, which were presented in Japanese, are explained here in English.

1. Introduction

Maximum likelihood estimation (MLE) is a prominent method in statistical analysis, but in some cases has some disadvantages (Whichmann and Hill, 2001). Theoretically, MLE is based on the assumption that data X_1, \dots, X_N are independent and identically distributed (iid) random variables. The properties of MLE are asymptotic and hold as N becomes sufficiently large (Hogg, McKean and Craig, 2005). In some applications, these assumptions are violated when the sample size may not be large enough or the iid assumption may not hold. For example, in signal detection theory (SDT), the probability of a "yes" response to a signal stimulus is not the same as to a noise stimulus. As an alternative method to MLE, a Bayesian data analysis is recommended (e.g., Wichmann and Hill, 2001; Raudenbush and Bryk, 2002). Okamoto (2005, 2006) reported successful results of a Bayesian data analysis in SDT by simulation.

Lee (2008) presented a Bayes SDT MATLAB-based package, which uses WinBUGS. Lee's Bayes SDT is restricted to equal-variance Gaussian SDT with two response categories, yes and no. However, the unequal-variance case arises quite naturally (Wickens, 2002). For example, a Poisson distribution suggests that random variables with larger means (i.e., in case of a signal stimulus) may have larger variances. In contrast, in a word recognition experiment, noise stimuli are new words, so they can make their impressions more variable than the impressions of signal stimuli, which were presented recently and have the variability of their distributions reduced (Wickens, 2002). Evidence of unequal variance is presented in

Green and Swets (1966). An algorithm for unequal-variance SDT is needed.

Lee's Bayes SDT software package uses a Markov chain Monte Carlo (MCMC) simulation, a stochastic simulation to estimate a probability distribution. But, in case of a small number of parameters of the distributions, more straightforward estimation is possible. A discrete approximation to the distribution is simple and can be made accurate enough by using a fine grid to be comparable to the result by MCMC.

Before Lee's proposal of Bayesian data analysis in SDT (Lee, 2008), where the method used (MCMC) is unnecessarily complex and the model considered (an equal-variance model for two responses) is restrictive, I had proposed Bayesian methods. In these methods, a simple grid method is used for the simple model (Okamoto, 2005), and an MCMC algorithm is applied to data from an SDT experiment, which uses more than three response categories and allows unequal variances (Okamoto, 2006). These papers were written in Japanese, so an explanation of the methods used is presented here in English.

2. A Grid Method for Equal-Variance SDT

Table 1. Number of Responses to the Stimuli.

Stimulus	Response	
	"yes"	"no"
Signal	N_h	N_m
Noise	N_{fa}	N_{cr}

In the case of data shown in Table 1, the likelihood function is given by

$$L(\mu_s, k \mid \text{data}) = P(\text{yes} \mid \text{signal})^{N_h} \cdot P(\text{no} \mid \text{signal})^{N_m} \\ \cdot P(\text{yes} \mid \text{noise})^{N_{fa}} \cdot P(\text{no} \mid \text{noise})^{N_{cr}}$$

where

$$P(\text{yes} \mid \text{signal}) = 1 - \Phi(k - \mu_s),$$

$$P(\text{no} \mid \text{signal}) = \Phi(k - \mu_s),$$

$$P(\text{yes} \mid \text{noise}) = 1 - \Phi(k),$$

$$P(\text{no} \mid \text{noise}) = \Phi(k),$$

and $\Phi(z)$ is the standard Gaussian cumulative distribution.

From the likelihood function, we have a posterior distribution

$$P(\mu_s, k) \propto p_0(\mu_s, k) \cdot L(\mu_s, k \mid \text{data})$$

where $p_0(\mu_s, k)$ is a prior distribution. As a prior distribution, a Jeffreys prior is sometimes recommended as a non-informative one, but it was pointed out that a Jeffreys prior is not non-informative (Okamoto, 2011). As another type of non-informative prior distribution, truncated uniform distributions are used in this paper. When we employ a truncated uniform prior, which is not zero on a sufficiently large region, the posterior distribution is proportional to the likelihood function. In this case, we have

$$P(\mu_s, k \mid \text{data}) = C_1 L(\mu_s, k \mid \text{data}), \tag{2.1}$$

where C_1 is some constant to make the right hand side of Eq. (2.1) a probability distribution.

To perform a discrete approximation to Eq. (2.1), probability distribution (2.1) is discretized onto a suitably fine grid of points (μ_{si}, k_j) , where μ_{si} and k_j are given as follows

$$\begin{aligned}\mu_{si} &= \mu_{\min} + (\mu_{\max} - \mu_{\min}) \cdot i / N_g, \quad i = 0, \dots, N_g \\ k_j &= k_{\min} + (k_{\max} - k_{\min}) \cdot j / N_g, \quad j = 0, \dots, N_g\end{aligned}\quad (2.2)$$

where N_g is the number of grid elements. Minimum values, μ_{\min} and k_{\min} , and maximum values, μ_{\max} and k_{\max} , are determined to make the discrete approximation cover the region where posterior probabilities are practically not zero. Adjustment of these minimum and maximum values can be done when using the program. The program in Visual C++ is uploaded at the author's website:

<http://y-okamoto-psy1949.la.coocan.jp/booksetc/bayesiansdt/>

Applying the grid by Eq. (2.2) to posterior distribution (2.1), we get the following discrete approximation

$$P(\mu_{si}, k_j | \text{data}) = C_2 L(\mu_{si}, k_j | \text{data}), \quad (2.3)$$

where C_2 is chosen so that the sum of $P(\mu_{si}, k_j | \text{data})$ over μ_{si} and k_j equals 1.

Marginal distributions of μ_{si} and k_j can be derived from Eq. (2.3) as follows

$$\begin{aligned}P(\mu_{si} | \text{data}) &= \sum_j P(\mu_{si}, k_j | \text{data}) \\ P(k_j | \text{data}) &= \sum_i P(\mu_{si}, k_j | \text{data})\end{aligned}$$

For each of the marginal distributions, point and interval estimations can be derived as the median and the 95% highest probability density (HPD) interval. The 95% HPD interval can be estimated by the Chen-Shao HPD estimation algorithm (Chen, Shao and Ibrahim, 2000).

In the case of data, $N_h = 75$, $N_m = 25$, $N_{fa} = 25$, and $N_{cr} = 75$, the author's program produces the results shown in Figures 1 and 2, where the number of grid elements $N_g = 1001$, and the curves are drawn as sequential line graphs.

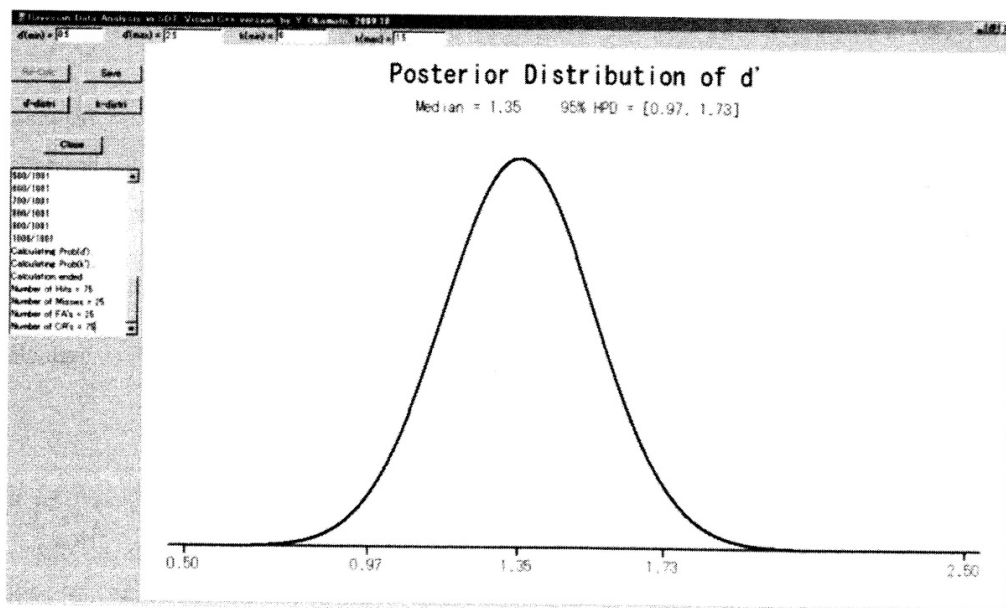


Figure 1. Posterior distribution and the 95% HPD interval of d' . The number of grid elements $N_g = 1001$.

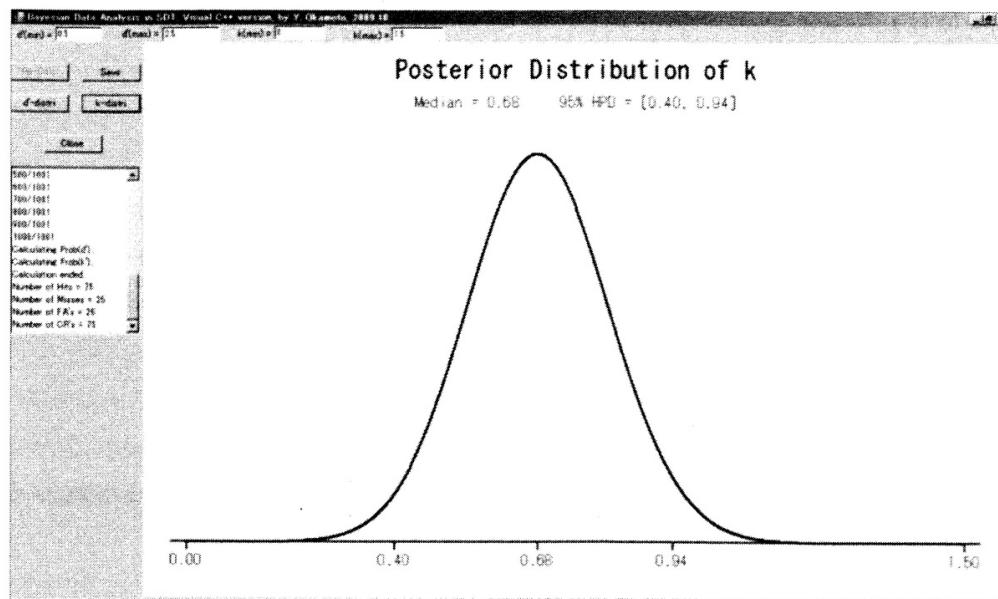


Figure 2. Posterior distribution and the 95% HPD interval of k . The number of grid elements $N_g = 1001$.

3. An MCMC Method for Unequal-Variance SDT

A model for unequal-variance SDT has more than two parameters, so more than two independent data points are required to estimate the parameters of an unequal-variance SDT model. The rating experiment, providing more than two independent data points, is the one where an observer rates his/her confidence about the response “yes” or “no.” If an observer is required to rate his/her confidence by assigning one of three categories: “sure,” “uncertain,” and “guessing,” then a response can be categorized as one of the six ordered categories: “no-sure,” “no-uncertain,” “no-guessing,” “yes-guessing,” “yes-uncertain,” and “yes-sure.” In general, suppose that an observer is required to rate his/her observations according to ordered categories: C_1, C_2, \dots, C_{N_c} . The observer’s response is category C_i when the sensory effect X belongs to interval (k_{i-1}, k_i) , where $-\infty = k_0 < k_1 < \dots < k_{N_c} = +\infty$. Hence, the likelihood function is given as follows:

$$L(\mu_s, \sigma_s, k_1, \dots, k_{N_c-1} | \text{data}) = \prod_{i=1}^{N_c} \{P(C_i | \text{signal})^{N_{si}} \cdot P(C_i | \text{noise})^{N_{ni}}\},$$

where N_{si} is the number of responses of category C_i to the signal and N_{ni} to the noise. Probabilities of response categories are given by the following equations:

$$P(C_i | \text{signal}) = \begin{cases} \Phi\left(\frac{k_1 - \mu_s}{\sigma_s}\right), & i = 1 \\ \Phi\left(\frac{k_i - \mu_s}{\sigma_s}\right) - \Phi\left(\frac{k_{i-1} - \mu_s}{\sigma_s}\right), & 1 < i < N_C \\ 1 - \Phi\left(\frac{k_{N_C-1} - \mu_s}{\sigma_s}\right), & i = N_C \end{cases}$$

$$P(C_i | \text{noise}) = \begin{cases} \Phi(k_i), & i = 1 \\ \Phi(k_i) - \Phi(k_{i-1}), & 1 < i < N_C \\ 1 - \Phi(k_{N_C-1}), & i = N_C \end{cases}$$

where

$$\sigma_s > 0 \text{ and } k_1 < k_2 < \dots < k_{N_C-1}. \quad (3.1)$$

The posterior distribution is given as follows

$$P(\mu_s, \sigma_s, \dots, k_{N_C-1} | \text{data}) = C_3 \cdot \prod_{i=1}^{N_C} \{P(C_i | \text{signal})^{N_{si}} \cdot P(C_i | \text{noise})^{N_{ni}}\}, \quad (3.2)$$

$$\times P_0(\mu_s, \sigma_s, k_1, \dots, k_{N_C-1}),$$

where $P_0(\mu_s, \sigma_s, k_1, \dots, k_{N_C-1})$ is a prior distribution, that represents constraint (3.1) on the parameters and is assumed to be a truncated non-zero uniform distribution on a sufficiently large region. If parameters k_1 to k_{N_C-1} do not satisfy condition (3.1) or σ_s is not positive, then $P_0(\mu_s, \sigma_s, k_1, \dots, k_{N_C-1}) = 0$. A uniform prior for σ is suggested by Carlin and Louis (2009, p.46).

The probability distribution (3.2) can be estimated by a stochastic simulation called Markov chain Monte Carlo (MCMC). There are many algorithms proposed for MCMC. Considering the complexity of the model (3.2) and condition (3.1), the author employs the componentwise Metropolis-Hastings algorithm (Ntzoufras, 2009), which in this case is as follows.

1. Select the initial values $\mu_s^{(0)}, \sigma_s^{(0)}, k_1^{(0)}, \dots, k_{N_C-1}^{(0)}$, and set $t \leftarrow 0$, where the arrow \leftarrow means that the right hand value is assigned to the left hand variable.

For $t = 0, \dots, T$, repeat from step 2 to step 14:

2. Select $y \sim N(\mu_s^{(t)}, c_\mu^2)$, where $N(\mu_s^{(t)}, c_\mu^2)$ represents a normal distribution with mean $\mu_s^{(t)}$ and variance c_μ^2 .
3. Calculate

$$\alpha = \min \left\{ 1, \frac{P(y, \sigma_s^{(t)}, k_1^{(t)}, \dots, k_{N_C-1}^{(t)} | \text{data})}{P(\mu_s^{(t)}, \sigma_s^{(t)}, k_1^{(t)}, \dots, k_{N_C-1}^{(t)} | \text{data})} \right\}$$

4. Set $\mu_s^{(t+1)} \leftarrow y$ with probability α and $\mu_s^{(t+1)} \leftarrow \mu_s^{(t)}$ with probability $1 - \alpha$.
5. Select $y \sim N(\sigma_s^{(t)}, c_\mu^2)$.
6. Calculate

$$\alpha = \min \left\{ 1, \frac{P(\mu_s^{(t+1)}, y, k_1^{(t)}, \dots, k_{N_c-1}^{(t)} | \text{data})}{P(\mu_s^{(t+1)}, \sigma_s^{(t)}, k_1^{(t)}, \dots, k_{N_c-1}^{(t)} | \text{data})} \right\}$$

7. Set $\sigma_s^{(t+1)} \leftarrow y$ with probability α and $\sigma_s^{(t+1)} \leftarrow \sigma_s^{(t)}$ with probability $1 - \alpha$.
8. Set $h \leftarrow 1$.
9. Select $y \sim N(k_h^{(t)}, c_{kh}^2)$.
10. Set

$$\alpha = \min \left\{ 1, \frac{P(\mu_s^{(t+1)}, \sigma_s^{(t+1)}, k_1^{(t+1)}, \dots, k_{h-1}^{(t+1)}, y, k_{h+1}^{(t)}, \dots, k_{N_c-1}^{(t)} | \text{data})}{P(\mu_s^{(t+1)}, \sigma_s^{(t+1)}, k_1^{(t+1)}, \dots, k_{h-1}^{(t+1)}, k_h^{(t)}, \dots, k_{N_c-1}^{(t)} | \text{data})} \right\}$$

11. Set $k_h^{(t+1)} \leftarrow y$ with probability α and $k_h^{(t+1)} \leftarrow k_h^{(t)}$ with probability $1 - \alpha$.
12. Set $h \leftarrow h + 1$.
13. If $h \leq N_c - 1$, jump back to step 9.
14. Set $t \leftarrow t + 1$.

From the sampled parameter values, $\mu_s^{(t)}, \sigma_s^{(t)}, k_1^{(t)}, \dots, k_{N_c-1}^{(t)}$, estimations of medians and 95% HPD intervals of the marginal posterior distributions can be derived. The Chen-Shao HPD estimation algorithm (Chen et al., 2000) can be used to estimate the 95% HPD intervals.

Table 2. Number of responses to the stimuli (the rating experiment).

Stimulus	Response					
	Yes			No		
	Sure	Uncertain	Guessing	Guessing	Uncertain	Sure
Signal	62	45	24	26	22	21
Noise	11	15	29	58	41	46

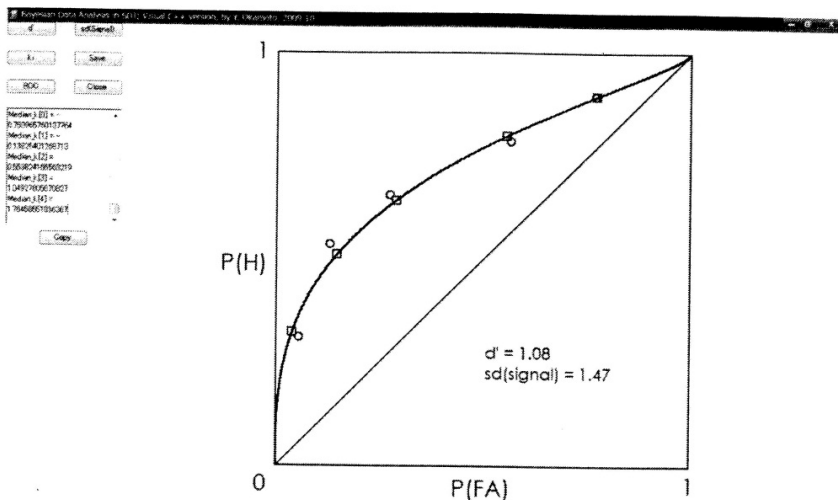


Figure 3. ROC curve estimated. Small circles represent the data points and small squares corresponding points on the ROC curve.

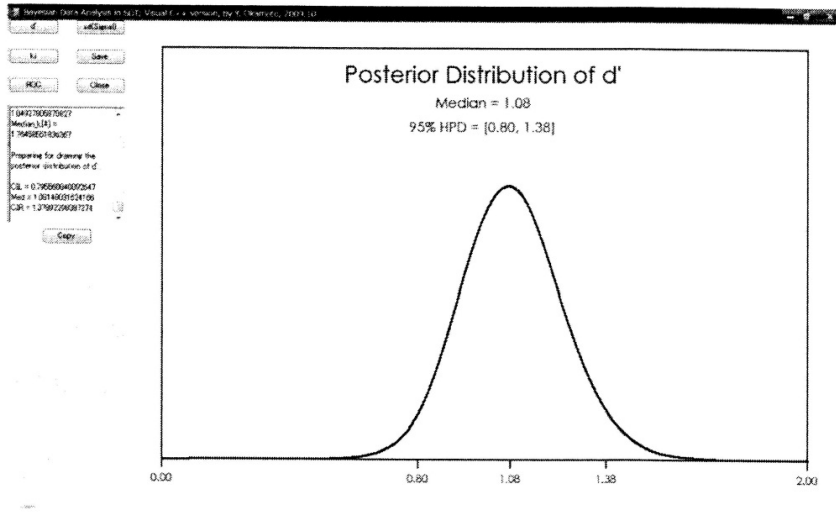


Figure 4. Posterior distribution of d' .

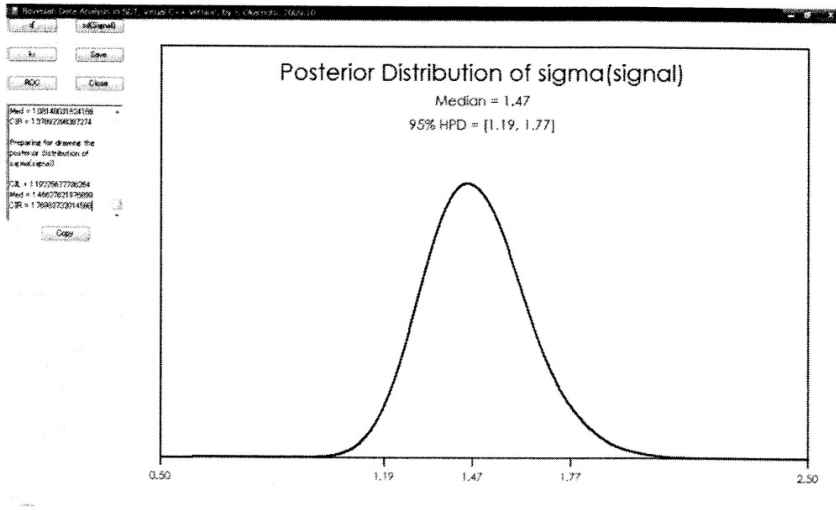


Figure 5. Posterior distribution of σ s.

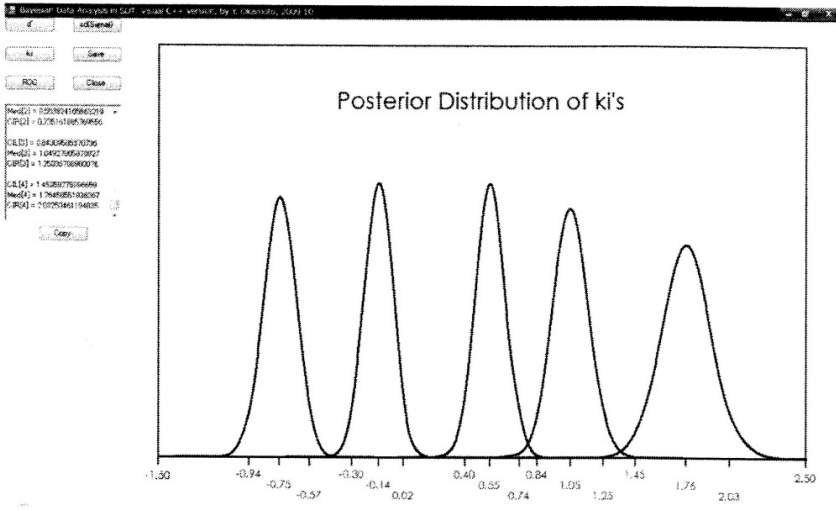


Figure 6. Posterior distributions of criterion parameters k_i 's.

The program at the author's website produces the results shown in Figures 3, 4, 5, and 6 for the data of Table 2. Figure 3 shows the ROC curve with small circles and small squares, that represent the data $((\sum_{i=1}^r N_{ni}) / (\sum_{i=1}^{N_c} N_{ni}), (\sum_{i=1}^r N_{si}) / (\sum_{i=1}^{N_c} N_{si}))$ and the estimated points $(\Phi(k_r), \Phi((k_r - \mu_s) / \sigma_s))$; $r = 1, \dots, N_c - 1$, respectively. The curves of the posterior distributions (Figures 4, 5, and 6) are estimated by a ker-

nel method with a Gaussian kernel (Wand and Jones, 1995).

4. Summary

Theoretical and practical problems in applying maximum likelihood methods to the cases where the sample size is not large enough or the iid assumption is not satisfied were pointed out. Data from an SDT experiment do not satisfy the iid assumption and may not be large enough. In these cases, Bayesian approaches can be considered as alternative ones. Basic results of Bayesian data analysis are posterior distributions. The two algorithms of Bayesian data analysis for equal- and unequal-variance SDT models estimate posterior distributions by a grid method and by MCMC, respectively. Examples of the proposed analyses are successfully presented.

Footnote

This research was supported in part by Grants-in-Aid for Scientific Research, Ministry of Education, Culture, Sports, Science and Technology, 20530668.

References

- Carlin, B. P. and Louis, T. A. (2009). *Bayesian Methods for Data Analysis*, third edition. Boca Raton, Flor.: CRC Press.
- Chen, M.-H, Shao, Q.-M. and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.
- Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: John Wiley & Sons, Inc.
- Hogg, R. V., McKean, J. W. and Craig, A. T. (2005). *Introduction to Mathematical Statistics*, sixth edition. Upper Saddle River, NJ: Pearson Prentice Hall.
- Lee, M. D. (2008). Bayes SDT: Software for Bayesian inference with signal detection theory. *Behavior Research Methods*, 40, 450-456.
- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: John Wiley & Sons, Inc.
- Okamoto, Y. (2005). Bayesian data analysis in SDT. Paper presented at the annual meeting of the Behaviormetric Society of Japan, Nagaoka (in Japanese). Retrieved from <http://ci.nii.ac.jp/naid/110006373126>
- Okamoto, Y. (2006). Bayesian data analysis in SDT: Analysis of rating data by MCMC. Paper presented at the annual meeting of the Behaviormetric Society of Japan, Kamio (in Japanese). Retrieved from <http://ci.nii.ac.jp/naid/110006329629>
- Okamoto, Y. (2011). Prior Information in a Jeffreys Prior: Distinction between Invariance and Non-informativeness. *Faculty of Integrated Arts and Social Sciences Journal, Japan Women's University*. 21, 33-36.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models*, second edition. Thousand Oaks: Sage Publications.
- Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Boca Raton: Chapman & Hall/CRC.
- Wichmann, F. A. and Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63, 1293-1313.
- Wickens, T. D. (2002). *Elementary Signal Detection Theory*. Madison Avenue, NY: Oxford University Press.