Statistics and Probability Letters 83 (2013) 2415–2421

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Existence of MLE and posteriors for a recognition-memory model

Xiaoyan Lin [a,*], Dongchu Sun [b,c], Paul L. Speckman [c], Jeffery N. Rouder [d]

[a] Department of Statistics, University of South Carolina, Columbia, SC 29208, United States
[b] School of Finance and Statistics, East China Normal University, Shanghai, 200241, China
[c] Department of Statistics, University of Missouri, Columbia, MO 65211, United States
[d] Department of Psychological Science, University of Missouri, Columbia, MO 65211, United States

ARTICLE INFO

Article history:
Received 31 May 2012
Received in revised form 26 June 2013
Accepted 26 June 2013
Available online 4 July 2013

Keywords:
Criterion
Fixed effect
Recognition memory
Sensitivity
Signal detection

ABSTRACT

Necessary and sufficient conditions are developed for the existence of the maximum likelihood estimate (MLE) for a recognition-memory model. The propriety of posteriors is shown for a class of bounded priors. Under a constant prior, an easy-to-implement Gibbs sampler is developed and illustrated via a real data set.

Published by Elsevier B.V.

1. Introduction

The theory of signal detection, first proposed by Tanner and Birdsall (1958), is a dominant measurement model in experimental psychology. The goal of the model is to decompose performance into an ability measure and a response bias. We describe its application here to *recognition memory* with the goal of estimating mnemonic ability. The recognition-memory task consists of a study and a test phase. At study, a participant is presented with a list of items and instructed to remember them for a later test. Then after a suitable delay, the participant is tested. At test, items are presented sequentially. Some of these items have been previously studied and are termed *old*, while others that have not been studied are termed *new*. The participant's task is to judge whether each test item is old or new.

The signal-detection model posits that each item has a mnemonic strength. In the simplest version of the model, the distribution of mnemonic strength is normal with fixed variance and mean depending on whether the item is new or old. Without loss of generality, the fixed variance is assumed 1, and the mnemonic strength is zero centered if the item is new while centered on parameter d with $d \geq 0$ if the item is studied. To reach a decision, the participant sets a *criterion* c on the mnemonic strength. If the strength is greater than c, then the participant judges the item as old; otherwise he or she judges the item as new. Parameter d serves as a measure of mnemonic ability and is referred to as *sensitivity*.

In the recognition-memory experiment, a *hit* event is an old-item response to a studied item; a *false* alarm event is an old-item response to a new item. The probability of a hit, denoted H, is the probability that mnemonic strength is larger than the criterion c for an old item: $H = \Phi(d - c)$, where $\Phi(\cdot)$ denotes the standard normal cdf; the probability of a false alarm, denoted F, is the probability that mnemonic strength is greater than c for a new item: $F = \Phi(-c)$.

* Corresponding author.
E-mail address: lin9@mailbox.sc.edu (X. Lin).

0167-7152/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.spl.2013.06.033

Suppose that there are I participants and each is tested on J items, with some subset of these J items previously studied. Let the response $y_{ij} = 1$ denote the case when participant i judges item j as old; otherwise $y_{ij} = 0$. Define $k_{ij} = 1$ if the ith participant studied the jth item before; otherwise $k_{ij} = 0$. Then, the joint probability mass function of $\boldsymbol{y} = (y_{11}, \ldots, y_{1J}, \ldots, y_{IJ})'$ is

$$[\boldsymbol{y} \mid \boldsymbol{H}, \boldsymbol{F}, \boldsymbol{k}] = \prod_{i=1}^{I} \prod_{j=1}^{J} \left[H_{ij}^{y_{ij}} (1 - H_{ij})^{1-y_{ij}} \right]^{k_{ij}} \left[F_{ij}^{y_{ij}} (1 - F_{ij})^{1-y_{ij}} \right]^{1-k_{ij}},$$

where H_{ij} and F_{ij} are the probabilities of hit and false alarm, respectively, for the ith participant tested at the jth item; $\boldsymbol{H}$ denotes the vector of all the H_{ij}, $\boldsymbol{F}$ denotes all the F_{ij}, and $\boldsymbol{k}$ denotes all the k_{ij}.

Signal detection is exceedingly popular because it is a convenient approach to separating ability and bias. Researchers generally use it for one of three reasons: (i) to understand how memory ability varies across individuals and subject variables such as pathology status, age, or IQ; (ii) to understand how memory ability varies across items and item variables such as item length or how frequently the item occurs in natural discourse; and (iii) to understand how memory ability varies across experimental manipulations (for instance, is memory retained even when attention is diverted by a concurrent task load). Unfortunately, although signal detection is useful for addressing a wealth of substantive questions, the conventional statistical analysis of the model is problematic. In practice, researchers have traditionally aggregated responses across items, participants, or both, to estimate the sensitivities and criteria. The aggregation method, however, requires very strict assumptions. When events are aggregated across items, it is implicitly assumed that each item has the same memorability and induces the same criterion. Likewise, when events are aggregated across participants, it is assumed that people have the same sensitivity and criterion. These implicit assumptions are usually too strict. When these assumptions do not hold, aggregation leads to an asymptotic bias of estimation (Rouder and Lu, 2005).

To estimate sensitivity and bias without recourse to aggregation, Rouder et al. (2007) expanded the hit and false alarm probabilities with participant and item effects. For the ith participant tested at the jth item, H_{ij} and F_{ij} are modeled as $H_{ij} = \Phi(\mu_h + \alpha_{hi} + \beta_{hj})$ and $F_{ij} = \Phi(\mu_f + \alpha_{fi} + \beta_{fj})$, respectively, where parameters μ_h and μ_f are overall effects; parameters α_{hi} and α_{fi} are participant effects for the ith participant; and parameters β_{hj} and β_{fj} are item effects for the jth item. Let $\boldsymbol{\alpha}_h = (\alpha_{h1}, \ldots, \alpha_{hI})'$, $\boldsymbol{\alpha}_f = (\alpha_{f1}, \ldots, \alpha_{fI})'$, $\boldsymbol{\beta}_h = (\beta_{h1}, \ldots, \beta_{hJ})'$, $\boldsymbol{\beta}_f = (\beta_{f1}, \ldots, \beta_{fJ})'$, and denote the joint vector of all parameters as $\boldsymbol{\lambda}$. The probability mass function of $\boldsymbol{y}$ is then expanded as

$$\begin{aligned}
[\boldsymbol{y} \mid \boldsymbol{\lambda}, \boldsymbol{k}] &= \prod_{i=1}^{I} \prod_{j=1}^{J} \left[\Big(\Phi(\mu_h + \alpha_{hi} + \beta_{hj}) \Big)^{y_{ij}} \Big(1 - \Phi(\mu_h + \alpha_{hi} + \beta_{hj}) \Big)^{1-y_{ij}} \right]^{k_{ij}} \\
&\quad \times \left[\Big(\Phi(\mu_f + \alpha_{fi} + \beta_{fj}) \Big)^{y_{ij}} \Big(1 - \Phi(\mu_f + \alpha_{fi} + \beta_{fj}) \Big)^{1-y_{ij}} \right]^{1-k_{ij}} \\
&\equiv L_1(\mu_h, \boldsymbol{\alpha}_h, \boldsymbol{\beta}_h \mid \boldsymbol{y}, \boldsymbol{k})\, L_2(\mu_f, \boldsymbol{\alpha}_f, \boldsymbol{\beta}_f \mid \boldsymbol{y}, \boldsymbol{k}),
\end{aligned} \tag{1}$$

where $L_1(\mu_h, \boldsymbol{\alpha}_h, \boldsymbol{\beta}_h \mid \boldsymbol{y}, \boldsymbol{k}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \Big(\Phi(\mu_h + \alpha_{hi} + \beta_{hj}) \Big)^{y_{ij}k_{ij}} \Big(1 - \Phi(\mu_h + \alpha_{hi} + \beta_{hj}) \Big)^{(1-y_{ij})k_{ij}}$; $L_2(\mu_f, \boldsymbol{\alpha}_f, \boldsymbol{\beta}_f \mid \boldsymbol{y}, \boldsymbol{k}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \Big(\Phi(\mu_f + \alpha_{fi} + \beta_{fj}) \Big)^{y_{ij}(1-k_{ij})} \Big(1 - \Phi(\mu_f + \alpha_{fi} + \beta_{fj}) \Big)^{(1-y_{ij})(1-k_{ij})}$. With this model, the participant-by-item sensitivity and criterion parameters are identified as $d_{ij} = \mu_h + \alpha_{hi} + \beta_{hj} - \mu_f - \alpha_{fi} - \beta_{fj}$ and $c_{ij} = -(\mu_f + \alpha_{fi} + \beta_{fj})$, respectively. The sensitivity and criterion parameters for the ith participant are $\mu_h + \alpha_{hi} - \mu_f - \alpha_{fi}$ and $-\mu_f - \alpha_{fi}$, respectively. The sensitivity and criterion parameters for the jth item are analogous expressions with β_{hj} and β_{fj}. Finally, the overall sensitivity and criterion parameters are $\mu_h - \mu_f$ and $-\mu_f$, respectively.

Rouder et al. (2007) treated the participant and item effects as random effects and proposed several hierarchical priors for the participant and item effects. However, the estimates are sensitive to the choice of hyperparameters in the inverse gamma or inverse Wishart hyper priors on the variance components. In general, when the variation of effects is of interest, random effect models are more suitable and provide shrinkage-style estimation. However, if the individual effects of participants and items are the focus to be estimated, then treating the effects as fixed effects is more natural. In this paper, we treat the individual participant and item effects as fixed effects and are interested in estimating them directly.

Under certain regularity conditions, the maximum likelihood estimate (MLE) is the same as the posterior mode under a constant prior. For binomial models, the existence of the MLE is equivalent to the existence of a proper posterior under the constant prior (Speckman et al., 2009). In this paper, we investigate the existence of the MLE of the participant and item effect parameters for the recognition-memory model. We propose a large class of noninformative or partial informative priors to conduct Bayesian analysis. The remainder of the paper is as follows. In Section 2, the nonidentifiability issue is discussed. In Section 3, necessary and sufficient conditions are provided for the existence of the MLE. Meanwhile, a class of priors that ensure the propriety of the joint posterior are given in Section 4. The constant prior is included as a special case. A Bayesian computation algorithm under the constant prior is presented in Section 5 and is illustrated with a real data set in Section 6. Some concluding remarks are in Section 7.

2. Nonidentifiability

Following Dawid (1979), if a probability density function (pdf) $f(y \mid \theta_1, \theta_2) = f(y \mid \theta_1)$, then the pdf is nonidentifiable, and θ_2 is the nonidentifiable parameter since y does not inform θ_2 at all.

By the above definition, both $L_1(\mu_h, \boldsymbol{\alpha}_h, \boldsymbol{\beta}_h \mid \boldsymbol{y}, \boldsymbol{k})$ and $L_2(\mu_f, \boldsymbol{\alpha}_f, \boldsymbol{\beta}_f \mid \boldsymbol{y}, \boldsymbol{k})$ in (1) have nonidentifiability problems. For example, if we transform parameters $\mu_h, \alpha_{hi},\ i = 1, \ldots, I$ and $\beta_{hj},\ j = 1, \ldots, J$ to $\mu_h, \eta_i = \mu_h + \alpha_{hi} + \beta_{hJ},\ i = 1, \ldots, I, \xi_j = \beta_{hj} - \beta_{hJ},\ j = 1, \ldots, J-1$, and β_{hJ}, then $L_1(\mu_h, \boldsymbol{\alpha}_h, \boldsymbol{\beta}_h \mid \boldsymbol{y}, \boldsymbol{k})$ is

$$\prod_{i=1}^{I}\prod_{j=1}^{J-1}\Big(\Phi(\eta_i + \xi_j)\Big)^{y_{ij}k_{ij}}\Big(1 - \Phi(\eta_i + \xi_j)\Big)^{(1-y_{ij})k_{ij}}\prod_{i=1}^{I}\Big(\Phi(\eta_i)\Big)^{y_{iJ}k_{iJ}}\Big(1 - \Phi(\eta_i)\Big)^{(1-y_{iJ})k_{iJ}},$$

which shows μ_h and β_{hJ} nonidentifiable. Therefore, the MLE of the whole set of parameters $\boldsymbol{\lambda}$ does not exist. To solve the nonidentifiability problem, constraints on parameters are introduced to L_1 and L_2. Conventionally, we let

$$\sum_{i=1}^{I}\alpha_{hi} = \sum_{i=1}^{I}\alpha_{fi} = \sum_{j=1}^{J}\beta_{hj} = \sum_{j=1}^{J}\beta_{fj} = 0. \tag{2}$$

One possible method is to let $\alpha_{hI} = -\sum_{i=1}^{I-1}\alpha_{hi}$, $\alpha_{fI} = -\sum_{i=1}^{I-1}\alpha_{fi}$, $\beta_{hJ} = -\sum_{j=1}^{J-1}\beta_{hj}$, and $\beta_{fJ} = -\sum_{j=1}^{J-1}\beta_{fj}$ and remove parameters $\alpha_{hI}, \alpha_{fI}, \beta_{hJ}$ and β_{fJ} from the model. Then the complete vector of identifiable parameters is $\boldsymbol{\lambda} = (\mu_h, \mu_f, \alpha_{h1}, \alpha_{f1}, \ldots, \alpha_{h(I-1)}, \alpha_{f(I-1)}, \beta_{h1}, \beta_{f1}, \ldots, \beta_{h(J-1)}, \beta_{f(J-1)})'$, which is $2I + 2J - 2$ dimensional. We next prove the existence of the MLE for the new defined $\boldsymbol{\lambda}$.

3. Existence of the MLE

A large number of papers (see e.g. Silvapulle, 1981; Albert and Anderson, 1984; Santner and Duffy, 1986) have investigated the existence of the MLE for binomial response models. The existence of the MLE is equivalent to the existence of a proper posterior under the constant prior for all parameters. Ghosh et al. (2000) investigated the posterior propriety for one-parameter item response models when using noninformative priors for participant and item effect parameters. Chen and Shao (2000) investigated the propriety of the posterior distribution for dichotomous response models with a general link when using the constant prior on the regression parameters and examined the relationship between the propriety of the posterior distribution and the existence of the MLE. Speckman et al. (2009) followed Albert and Anderson (1984) and developed necessary and sufficient conditions for the existence of the MLE and the propriety of posteriors for multinomial choice models when using the constant prior. The conditions in Chen and Shao (2000) are adopted to provide necessary and sufficient conditions for the existence of the MLE of the effect parameters for the recognition-memory model.

Without loss of generality, suppose I and J are even integers and the fractional balanced design is used. That is, each participant encounters $J/2$ old items and $J/2$ new items, and at the same time, each item is observed as an old item $I/2$ times and as a new item $I/2$ times. The following theorem provides necessary and sufficient conditions for the existence of the MLE of the effect parameters for the recognition-memory model.

Theorem 3.1. *The MLE of $\boldsymbol{\lambda}$ exists if and only if the following conditions hold.*

(i) $0 < \sum_{i=1}^{I} y_{ij}k_{ij} < \frac{I}{2}$, *for all j;*
(ii) $0 < \sum_{j=1}^{J} y_{ij}k_{ij} < \frac{J}{2}$, *for all i;*
(iii) $0 < \sum_{i=1}^{I} y_{ij}(1 - k_{ij}) < \frac{I}{2}$, *for all j;*
(iv) $0 < \sum_{j=1}^{J} y_{ij}(1 - k_{ij}) < \frac{J}{2}$, *for all i.*

The proof of Theorem 3.1 is sketched out as follows. Model (1) can be separated into two separate models. One is associated with L_1, and the other is associated with L_2. For submodel L_1, conditions (i) and (ii) ensure condition (b) of Chen and Shao's (2000) conditions hold. Therefore, by Theorem 3.1 in Chen and Shao (2000), the MLE of $(\mu_h, \boldsymbol{\alpha}_h, \boldsymbol{\beta}_h)$ exists. Likewise, conditions (iii) and (iv) ensure the existence of the MLE of $(\mu_f, \boldsymbol{\alpha}_f, \boldsymbol{\beta}_f)$. Chen and Shao's (2000) conditions and the connection to our conditions are presented in the Appendix.

The conditions of Theorem 3.1 are actually a more natural version of Chen and Shao's (2000) conditions in terms of our recognition-memory model and they are easy to check. Condition (i) states that each item when given in the studied condition evokes at least one correct (i.e. old) and at least one incorrect (i.e. new) response. Condition (iii) is the corresponding condition for each item when presented in the unstudied condition. Condition (ii) states that each participant has at least one correct and one incorrect response for the studied items, and condition (iv) is the corresponding condition for each participant facing unstudied items. The above theorem can be easily extended to an unbalanced design as long as there is at least one correct response and one incorrect response for each participant and for each item under both of the two study conditions.

4. Propriety of posteriors

For binomial response models, the existence of the MLE is equivalent to the existence of a proper posterior under the constant prior for all parameters (Chen and Shao, 2000; Speckman et al., 2009). Therefore, Theorem 3.1 also provides sufficient conditions for the existence of proper posterior distributions under a class of priors.

Theorem 4.1. *Assume the conditions in Theorem 3.1 hold. Under a bounded prior for* $\boldsymbol{\lambda}$*, the joint posterior of* $\boldsymbol{\lambda}$ *is proper.*

Theorem 4.1 provides a large class of useful priors for real data analysis. Proper priors and constant priors for all parameters are two extreme cases in this class. Proper priors usually reflect prior information of parameters while constant priors reflect little or no information about parameters. Instead of assigning all proper priors or all constant priors for all parameters, this class of priors allow us to use partial information of parameters. For example, we may assign proper informative priors for the participant effect parameters while using the constant noninformative prior for the item effect parameters, as it might be easier to access the prior information for participants than that of items.

It is worth mentioning that when the constant prior for all parameters is used, conditions in Theorem 3.1 are also necessary for the propriety of the joint posterior. Suppose condition (iii) is violated: for example, there is an item, say the j'th item, having all new responses when it performs not as previously studied (that is, $y_{ij'} = 0$ for all i with $k_{ij'=0}$). Let $\bar{\Phi}(\cdot)$ denote $1 - \Phi(\cdot)$. It is clear that

$$\int_{-\infty}^{+\infty} \prod_{i:k_{ij'}=0} \bar{\Phi}(\mu_f + \alpha_{fi} + \beta_{fj'})\, d\beta_{fj'} \geq \int_{-\infty}^{0} \prod_{i:k_{ij'}=0} \bar{\Phi}(\mu_f + \alpha_{fi})\, d\beta_{fj'} = +\infty$$

which leads to an improper joint posterior of $\boldsymbol{\lambda}$. When other conditions are violated, similar derivation can show the joint posterior improper.

5. Bayesian computation under the constant prior

In this section, Bayesian computation for the recognition-memory model (1) under the constant prior for all of the parameters is illustrated. Introduction of normal latent variables (Albert and Chib, 1993) can avoid the probit link function in the computation. For each y_{ij}, one unobserved latent variable w_{ij} is introduced: $w_{ij} \sim N(\mu_h + \alpha_{hi} + \beta_{hj}, 1)$ if $k_{ij} = 1$; and $w_{ij} \sim N(\mu_f + \alpha_{fi} + \beta_{fj}, 1)$ if $k_{ij} = 0$. Denote the set of all w_{ij} as $\boldsymbol{w}$. Therefore, the augmented density function of $\boldsymbol{y}$ and $\boldsymbol{w}$ is

$$[\boldsymbol{y}, \boldsymbol{w} \mid \boldsymbol{\lambda}, \boldsymbol{k}] = \prod_{i=1}^{I}\prod_{j=1}^{J}\left[1_{(w_{ij}>0,\, y_{ij}=1)} + 1_{(w_{ij}\leq 0,\, y_{ij}=0)}\right]^{k_{ij}}\left[1_{(w_{ij}>0,\, y_{ij}=1)} + 1_{(w_{ij}\leq 0,\, y_{ij}=0)}\right]^{1-k_{ij}} \times g_{H_{ij}}(w_{ij})^{k_{ij}} g_{F_{ij}}(w_{ij})^{1-k_{ij}}, \tag{3}$$

where $g_{H_{ij}}(w_{ij}) = \phi(w_{ij} - \mu_h - \alpha_{hi} - \beta_{hj})$ and $g_{F_{ij}}(w_{ij}) = \phi(w_{ij} - \mu_f - \alpha_{fi} - \beta_{fj})$ with $\phi(\cdot)$ a standard normal pdf. Using the conditional probability law, $[\boldsymbol{y}, \boldsymbol{w} \mid \boldsymbol{\lambda}, \boldsymbol{k}] = [\boldsymbol{y} \mid \boldsymbol{w}, \boldsymbol{\lambda}, \boldsymbol{k}][\boldsymbol{w} \mid \boldsymbol{\lambda}, \boldsymbol{k}] = [\boldsymbol{y} \mid \boldsymbol{w}, \boldsymbol{k}][\boldsymbol{w} \mid \boldsymbol{\lambda}, \boldsymbol{k}]$, where the second equality follows because $\boldsymbol{y}$ is conditionally independent of $\boldsymbol{\lambda}$ given $\boldsymbol{w}$. Combining this with the constant prior on $\boldsymbol{\lambda}$, we derive the full conditional posterior distributions. The conditional posterior distribution for the latent variable w_{ij} is a left or right truncated normal distribution, summarized in Proposition 5.1. The conditional posterior for $\boldsymbol{\lambda}$ is a multivariate normal distribution, summarized in Proposition 5.2.

Proposition 5.1. *For* $i = 1, \ldots, I, j = 1, \ldots, J$,

$$[w_{ij} \mid \boldsymbol{\lambda}; \boldsymbol{y}, \boldsymbol{k}] \propto \begin{cases} g_{H_{ij}}(w_{ij})1_{(w_{ij}>0)}, & \text{if } y_{ij} = 1,\ k_{ij} = 1, \\ g_{H_{ij}}(w_{ij})1_{(w_{ij}\leq 0)}, & \text{if } y_{ij} = 0,\ k_{ij} = 1, \\ g_{F_{ij}}(w_{ij})1_{(w_{ij}>0)}, & \text{if } y_{ij} = 1,\ k_{ij} = 0, \\ g_{F_{ij}}(w_{ij})1_{(w_{ij}\leq 0)}, & \text{if } y_{ij} = 0,\ k_{ij} = 0, \end{cases}$$

where $g_{H_{ij}}(w_{ij})$ *and* $g_{F_{ij}}(w_{ij})$ *are normal densities defined after (3).*

Proposition 5.2. *For* $\boldsymbol{\lambda}$*, the conditional posterior is*

$$(\boldsymbol{\lambda} \mid \boldsymbol{w}, \boldsymbol{k}) \sim N_{2I+2J-2}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{w}, (\boldsymbol{X}'\boldsymbol{X})^{-1}),$$

where $\boldsymbol{X}$ *is the* $IJ \times (2I + 2J - 2)$ *design matrix for the* w_{ij}*'s with respect to* $\boldsymbol{\lambda}$.

Having the full conditionals, Gibbs sampling is used to estimate the joint posterior distribution and marginal posterior distributions.

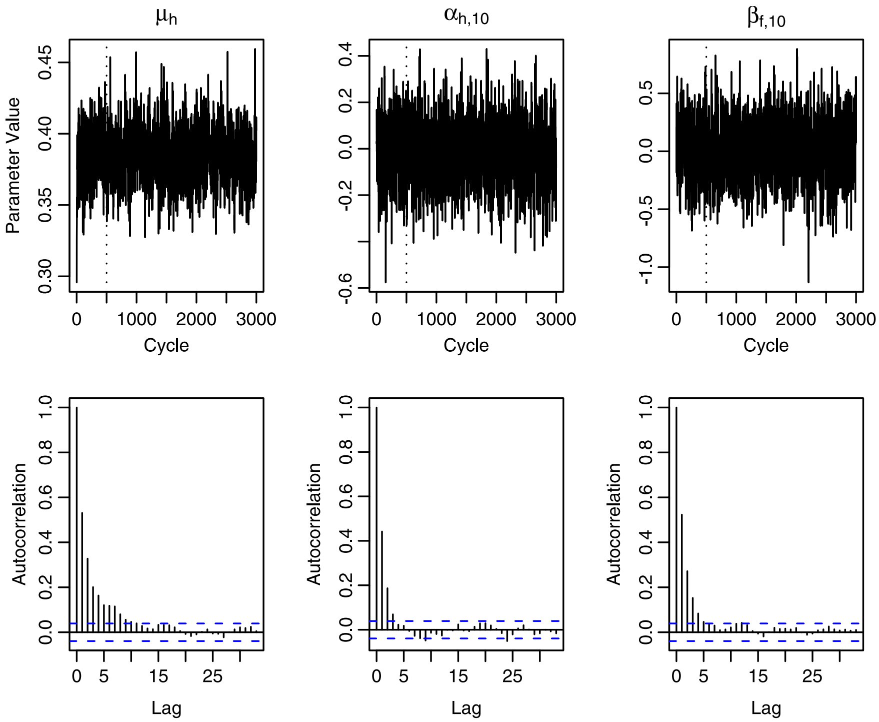

Fig. 1. MCMC chain convergence. The top row shows MCMC values for selected parameters. The vertical line denotes the end of the burn-in period. The bottom row shows autocorrelations for the same parameters after burn-in.

6. Analysis of a data set

The objective Bayesian method with the constant prior was used to analyze a recognition-memory data set. The data were collected at Rouder's Perception and Cognition Lab at the University of Missouri-Columbia. The main interest is to estimate the participants and items' sensitivities and criteria. There were 64 participants tested on 200 items. The design is fractional balanced. Before performing the analysis, we checked the conditions in Theorem 3.1. We noticed that all of the responses were new-item responses when the 77th item was not previously studied, which violates condition (iii). Including these data points in the analysis will lead to an improper joint posterior. For simplicity, here we deleted all data points associated with the 77th item to conduct the data analysis. In practice, if we are really interested in estimating the sensitivity of the 77th item, we may slightly modify the responses for the 77th item by, for example, randomly assigning old-item response for one or two data points. We ran 3000 Gibbs cycles with the first 500 serving as a burn-in. The convergence of the MCMC chain was good after 500 burn-in cycles (see Fig. 1). We used posterior means to estimate the effect parameters. The overall means μ_h and μ_f were estimated to be 0.3847 and −0.6989, respectively. The estimates of sensitivities and criteria for participants and items were calculated through the estimates of participant and item effects. Fig. 2 shows the estimates of the participant and item sensitivities and criteria and their 95% credible intervals. The 29th and 60th participants have negative estimates of sensitivity, reflecting the abnormal structure of the data: both of them gave more old responses for new items than for previously studied items. The 136th and 145th items have negative estimates of sensitivity, also reflecting the abnormal structure of the data: they got more old responses when they performed as new items than when they were previously studied.

7. Conclusion

In this paper, we investigate the existence of the MLE for the probit additive recognition-memory model. Under the certain parameter constraints, we develop necessary and sufficient conditions for the existence of the MLE. The likelihood conditions in Theorem 3.1 are easy to check, and usually easily satisfied when we have hundreds of participants and items. Under the same conditions, we provide a large class of priors that ensure the propriety of posteriors. Among the class of priors, the constant prior for all of the parameters is a special case. The Bayesian computation is illustrated when using the constant prior.

It is natural to consider that α_{hi} and α_{fi} are paired parameters since they represent the effects from the same participant. Likewise so are β_{hj} and β_{fj}. Ignoring the true parameter structure may lose partial power of the estimation. In practice, when there are hundreds of participants and items, it may be more suitable to model participant and item effects as random effects. Therefore, one possible Bayesian analysis to improve the estimation is to use bivariate normal priors for both paired

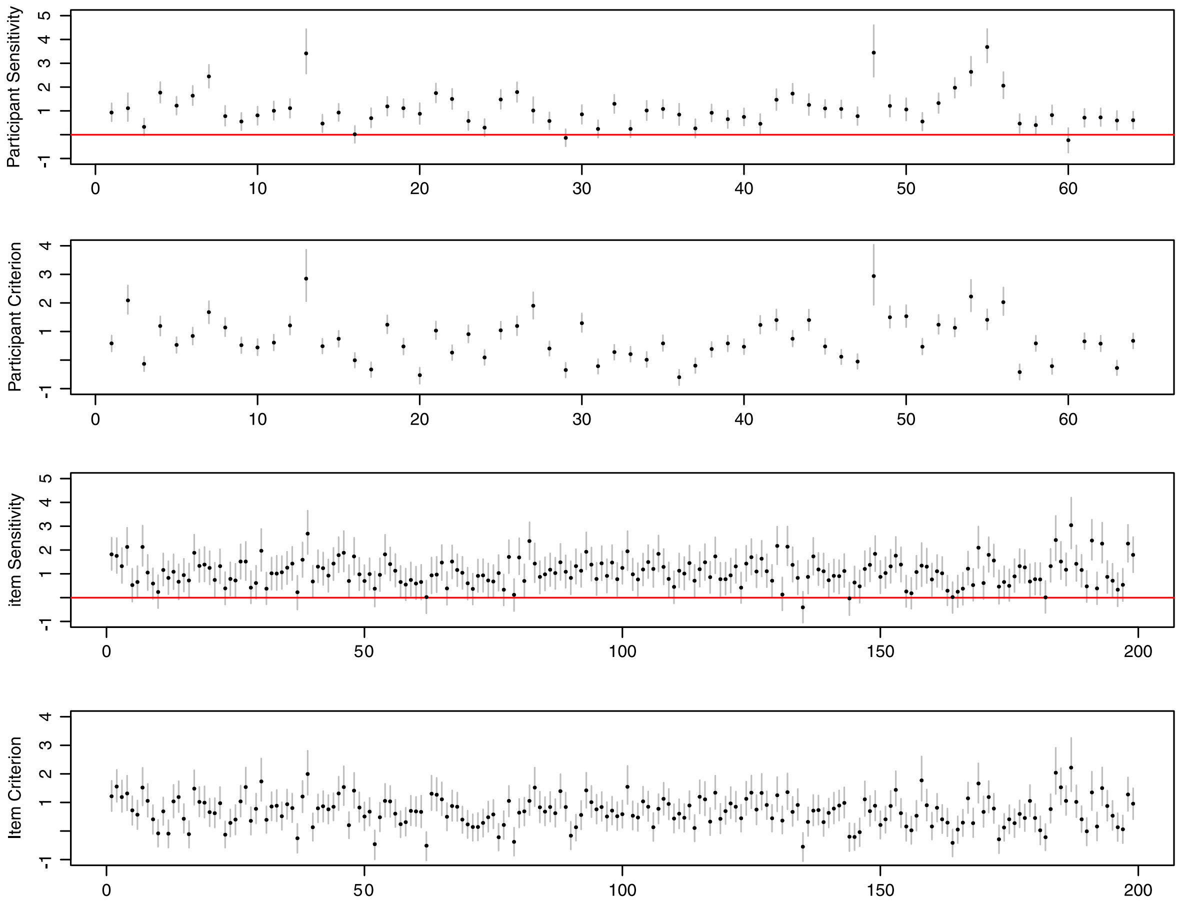

Fig. 2. Estimates of participant and item sensitivities and criteria and their 95% credible intervals.

participant effects and paired item effects and then apply some objective prior to the parameters in the bivariate normal priors. Some good candidates of objective priors for the bivariate normal may be found in Berger and Sun (2008).

Acknowledgments

The authors thank the associate editor and two referees for their helpful comments and suggestions. The project is supported by NSF grants SES-0720229, SES-1024080, and SES-1260806. Sun's research is also supported by the NSF grant DMS-1024080 and the NSF grant of China (11271136).

Appendix. Chen and Shao's (2000) conditions

Here we state the necessary and sufficient conditions for the existence of the MLE of the regression coefficient for a dichotomous probit or logit model. Assume a dichotomous response model, $P(y_i = 1) = G(x_i^T \beta)$, for $i = 1, \ldots, n$, where G is a standard normal or logistic cdf. Let X denote the design matrix whose ith row is x_i^T, and let Z denote the matrix whose ith row is z_i^T, where $z_i = x_i$ if $y_i = 0$ and $z_i = -x_i$ if $y_i = 1$.

Proposition 1 (*Chen and Shao, 2000*)**.** *The MLE of β exists if and only if*

(a) *the design matrix X has full column rank and*
(b) *there exists a positive vector $a = (a_1, \ldots, a_n)^T$ with strictly positive components such that $Z^T a = 0$.*

Condition (a) is satisfied for our recognition-memory model when the parameter constraints (2) are added. Condition (b) basically says that there exists some overlap between two spanned (with strictly positive coefficients) covariate spaces for the two response groups, one group of subjects with all responses 1 and the other group of subjects with all responses 0 (Silvapulle, 1981). It is clear that when there is a participant or item having all successes or failures, there is no positive a such that $Z^T a = 0$, and when there is at least one success and one failure for each participant and item, there is a positive

a such that $Z^T a = 0$. In general, assuming that X is full column rank, condition (b) can be straightforwardly checked with a simple linear program. For example, using the 'simplex' function from the 'boot' library in **R** (R Core Team, 2013). Details of the program can be found in Appendix A of Roy and Hobert (2007).

References

Albert, A., Anderson, J.A., 1984. On the existence of maximum likelihood estimates in logistic regression models. Biometrika 71, 1–10.
Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88, 669–679.
Berger, J.O., Sun, D., 2008. Objective priors for a bivariate normal model. Annals of Statistics 36, 963–982.
Chen, M.-H., Shao, Q.-M., 2000. Propriety of posterior distribution for dichotomous quantal response models with general link functions. Proceedings of the American Mathematical Society 129, 293–302.
Dawid, A., 1979. Conditional independence in statistical theory (with discussion). Journal of the Royal Statistical Society. Series B 41, 1–31.
Ghosh, M., Ghosh, A., Chen, M.-H., Agresti, A., 2000. Noninformative priors for one-parameter item response models. Journal of Statistical Planning and Inference 88, 99–115.
R Core Team, 2013. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available from http://www.R-project.org.
Rouder, J.N., Lu, J., 2005. An introduction to Bayesian hierarchical models with an application in the theory of signal detection. Psychonomic Bulletin and Review 12, 573–604.
Rouder, J.N., Lu, J., Sun, D., Speckman, P.L., Morey, R.D., Naveh-Benjamin, M., 2007. Signal detection models with random participant and item effects. Psychometrika 72, 621–642.
Roy, V., Hobert, J.P., 2007. Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. Journal of the Royal Statistical Society. Series B 69, 607–623.
Santner, T.J., Duffy, D.E., 1986. A note on A. Albert and J. A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. Biometrika 73, 755–758.
Silvapulle, M.J., 1981. On the existence of maximum likelihood estimators for the binomial response models. Journal of the Royal Statistical Society. Series B 43, 310–313.
Speckman, P.L., Lee, J., Sun, D., 2009. Existence of the mle and propriety of posteriors for a general multinomial choice models. Statistica Sinica 19, 731–748.
Tanner, J.W.P., Birdsall, T.G., 1958. Definition of d and n as psychophysical measures. Journal of the Acoustical Society of America 30, 922–928.