Statistics and Its Interface Volume 11 (2018) 369–384

Bayesian analysis of a ROC curve for categorical data using a skew-binormal model

Balgobin Nandram and Thelge Buddika Peiris*

In a taste-testing experiment, foods are withdrawn from storage at various times and a panel of tasters is asked to rate the foods on a nine-point hedonic scale. We provide a statistical procedure that can assess the difference between fresh foods and foods withdrawn a few months later. Thus, we have two sets of ordinal data, one for the fresh foods and the other for the stored foods that are withdrawn later. A natural and popular way to compare two withdrawals is to use the receiver operating characteristic (ROC) curve and the area under the curve (AUC). It is a standard practice to use a binormal model to obtain the ROC curve and the AUC, and Bayesian methods have been used. One drawback of the binormal model is that it has non-identifiable parameters. First, we look more carefully at non-identifiability, and we use robust measures to obtain a more non-parametric analysis of the Bayesian binormal model. Second, in a more innovative approach we extend the robust binormal model to a skew-binormal model. Like recent approaches to ROC curve analysis, we also incorporate a stochastic ordering. We use the Gibbs sampler to fit both models in order to estimate the ROC curves and the AUCs. Using both models, these AUCs demonstrate that there is not much practical difference between fresh foods and those withdrawn later, but there are some differences in inference between the two models. We have also shown, using marginal likelihoods, that the skew-binormal model may be better. A small simulation study shows that the skew-binormal model provides improved precision over the binormal model with similar AUCs but somewhat different ROC curves.

Keywords and phrases: Area under the curve, Griddy Gibbs sampler, Marginal likelihood, Non-indentifiability, Robust measures, Sensory data, Stochastic order.

1. INTRODUCTION

Receiver operating characteristic (ROC) analysis can be used to assess performance in any two-group classification task. The ROC curve is a plot of the true positive rate (the rate of actually positive cases correctly classified as positive) and false positive rate (the rate of actually negative cases incorrectly classified as positive). Most ROC studies involving human judgments have employed an ordinal scale that contains a fixed number of discrete response categories. Smoothed ROC curves can be easily obtained by fitting a parametric binormal model to two-sample data, cases and non-cases. Several frequentist and Bayesian methods for fitting "binormal" ROC curves to such data and estimating the area under the curve (AUC) have been discussed. Our application is on assessing food deterioration in an experiment.

Most of the frequentist methods are used when the populations for two groups (actually positive and actually negative) are known. Metz, Herman, and Shen (1998), Metz (1989), Dorfman and Alf (1969) and Grey and Morgan (1972) introduced different techniques to obtain maximum likelihood estimate of the ROC curve. Tosteson, Pinto, Holsinger, and Lamb (1994) showed limitations of some of the above techniques and they illustrated how ROC curve regression analysis facilitates diagnostic test assessment. Faraggi and Reiser (2002) discussed and compared some nonparametric estimation methods of area under the curve with the binormal assumption.

Bayesian statistical inferences of ROC curves using continuously distributed data have been used quite successfully in a large number of areas. O'Malley and Zou (2006) analyzed hierarchical structured outcome data from a diagnostic test that requires cluster-specific monotone transformations. Here they have developed a hierarchical Bayesian multivariate transformation method based on data with covariates. Inácio, Jara, Hanson, and De Carvalho (2013) used a Bayesian nonparametric approach to study a covariate-dependent ROC curve with a dependent Dirichlet process for clustered data. In their work they evaluated the influence of age on the performance of blood glucose to accurately diagnose individuals with diabetes. Here they used a nonparametric approach that entails a modeling framework, which requires specifying a prior distribution over all probability measures. As they pointed out, this does not mean an absence of parameters in the model, on the contrary it has an infinite number of parameters. But the objectives in our work are different from theirs because we use standard robust measures.

There are many studies on Bayesian statistical inference of ROC curves using ordinal data too. Peng and Hall (1996) introduced a Bayesian approach to generalized ordinal regression models for rating data using a probit link func-

*Corresponding author.

tion. The approach that they have used admits latent binormal distributions for diseased and non-diseased populations, even though other parametric distributions could be considered. Johnson and Johnson (1996) addressed a situation frequently observed in radiology, in which several radiologists rate, on an ordinal scale, multiple exams collected from the same individual. These authors proposed a hierarchical Bayesian latent variable model for this multi-rater correlated ordinal data. Further Hellmich et al. (1998) and Ishwaran and Gatsonis (2000) also used regression models with ordinal data for ROC curve analysis.

Like many of these discussions, we assume that the decision variable axis is partitioned into K categories with $K-1$ cut-points, $\underset{\sim}{\theta} = \{\theta_1, \theta_2, \theta_3, \cdots, \theta_{K-1}\}$ and values beginning from the "actually negative" end of the axis. We denote the cumulative distribution functions of the decision variable x for respectively the first group (actually negative) and the second group (actually positive) by $F_1(x)$ and $F_2(x)$. Let μ_1 and σ_1 denote respectively the mean and standard deviation of the population for the first group and μ_2 and σ_2 denote respectively the mean and standard deviation of the population for the the second group. With $\theta_0 \equiv -\infty$ and $\theta_K \equiv \infty$, the probability of a response in category $i, i = 1, 2, \cdots, K-1$, is then,

$$p_{ji} = F_j\left(\frac{\theta_i - \mu_j}{\sigma_j}\right) - F_j\left(\frac{\theta_{i-1} - \mu_j}{\sigma_j}\right), \qquad j = 1, 2.$$

The ROC curve represents the plot $(1 - F_1(\theta), 1 - F_2(\theta))$ for all θ. Let $S_1(\theta) = 1 - F_1(\theta)$ and $S_2(\theta) = 1 - F_2(\theta)$. Then the ROC curve is given

$$ROC(u) = S_2\{S_1^{-1}(u)\} = 1 - F_2\{F_1^{-1}(1-u)\}$$

for $u \in [0, 1]$. In addition to the ROC curve, an important parameter of interest is the corresponding $AUC = \int_0^1 ROC(u)du$. AUC has the interpretation of being the probability that a randomly selected individual has a value for the decision variable that is greater than that for a randomly selected individual from the second group (Bamber, 1975).

In general, subjects in Group 1 are more likely to test positively than the subjects in Group 2 which implies that $ROC(u) > u$ and further $AUC > 0.5$. But $ROC(u) > u$ (i.e., above the 45° straight line) if and only if F_1 is stochastically larger than F_2 (i.e., $F_1(t) \geq F_2(t)$ for all t). Here we use the technique introduced in Gelfand and Kuo (1991) and developed in Gelfand and Kottas (2001) to have stochastically ordered distributions. Theoretical work on the symmetry of the concave ROC curve has been investigated by others (e.g., Bhattacharya and Hughes 2011, 2015 and Hughes and Bhattarcharya 2013). However, we do not follow these directions and we simply add the stochastic order to enforce $ROC(u) > u$ with robust measures.

Here our goal is to estimate the ROC curve and AUC using a Bayesian approach assuming the two groups are known to the analyst. In our example of food preservation, we assume that the panelists are accurate and coherent in assigning scores, and we wish to determine whether the food deteriorates after several withdrawals. Jhonson and Jhonson (1996) discussed a similar problem but since data are from two known groups, our approach is slightly different than theirs. We also discuss how to robustify our method using the skew-normal data model. Our new model, the skew-binormal model, has two skew-normal densities (Azzalini, 2014). Our skew-binormal model takes care of a degree of skewness beyond the binormal model, and therefore, more robust to non-normality.

The plan of the rest of the paper is as follows. In Section 2, we describe a general model (i.e., without making any specific distribution assumptions) and a special case, the binormal model with some degree of robustness. In Section 3, we describe the skew-binormal model (both cases and non-cases). In Section 4, we describe an application on a food experiment and use it to compare the binormal model and the skew-binormal model via the ROC curves and AUCs. We also perform a simulation study to assess any difference between the binormal model and skew-binormal model. Section 5 has concluding remarks.

2. GENERAL MODEL AND BINORMAL MODEL

In Section 2.1, we describe the general model and, in Section 2.2, we describe the binormal model, which has a degree of robustification.

Suppose that data, collected on a discrete K-category scale, are obtained from n_1 non-cases and n_2 cases, and the true states of the non-cases and the cases are known. If the responses are mutually independent, a standard assumption is that the categorical frequencies within each true state follow a multinomial distribution. The observed data are $\underset{\sim}{n}_1 = \{n_{11}, n_{12}, n_{12}, \cdots, n_{1K} : \sum n_{1i} = n_1\}$ for the non-cases and $\underset{\sim}{n}_2 = \{n_{21}, n_{22}, n_{22}, \cdots, n_{2K} : \sum n_{2i} = n_2\}$ for the cases. Then, the joint probability mass function of $\underset{\sim}{n}_1$ and $\underset{\sim}{n}_2$ is

$$\begin{aligned} p(\underset{\sim}{n}_1, \underset{\sim}{n}_2|\theta) = {} & \frac{n_1!}{n_{11}!.n_{12}!.\cdots.n_{1K}!} \times \frac{n_2!}{n_{21}!.n_{22}!.\cdots.n_{2K}!} \\ & \times \left[F_1\left(\frac{\theta_1 - \mu_1}{\sigma_1}\right)\right]^{n_{11}} \left[F_1\left(\frac{\theta_2 - \mu_1}{\sigma_1}\right)\right. \\ & \left. - F_1\left(\frac{\theta_1 - \mu_1}{\sigma_1}\right)\right]^{n_{12}} \cdots \left[1 - F_1\left(\frac{\theta_{K-1} - \mu_1}{\sigma_1}\right)\right]^{n_{1K}} \\ & \times \left[F_2\left(\frac{\theta_1 - \mu_2}{\sigma_2}\right)\right]^{n_{21}} \left[F_2\left(\frac{\theta_2 - \mu_2}{\sigma_2}\right)\right. \\ & \left. - F_2\left(\frac{\theta_1 - \mu_2}{\sigma_2}\right)\right]^{n_{22}} \cdots \left[1 - F_2\left(\frac{\theta_{K-1} - \mu_2}{\sigma_2}\right)\right]^{n_{2K}}. \end{aligned} \tag{1}$$

It is worth noting that there are $K+3$ parameters in this K-category data model and they are $\theta_1, \ldots, \theta_{K-1}, (\mu_s, \sigma_s^2), s = 1, 2$. These parameters are not identifiable, and without further information they create a serious problem for computation and data analysis. Therefore, without additional information, one possibility is to reduce some parameters.

2.1 General model

Here we show how to reduce the number of parameters in this model. When the first and the second groups are normally distributed, Metz, Herman, and Shen (1998) showed that

$$F_1(\theta) = \Phi(\theta) \text{ and } F_2(\theta) = \Phi(b\theta - a),$$

where $a = \frac{|\mu_2 - \mu_1|}{\sigma_2}$, $b = \frac{\sigma_1}{\sigma_2}$ and $\Phi(\cdot)$ represent the standard normal cumulative distribution function. Using a similar argument and because

$$\frac{d[\theta - c - (\mu - c)]}{d\sigma} = \frac{\theta - \mu}{\sigma}$$
$$\text{for } d > 0 \text{ and } -\infty < c < \infty,$$

without lost of generality, we can take $\mu_1 = 0$ and $\sigma_1 = 1$. Under these specifications there are now $K + 1$ parameters, and μ_2 and σ_2 are still not identifiable. So here we assume the mean and the standard deviation for the second group are functions of $\underset{\sim}{\theta}$. That is, $\mu = \mu(\underset{\sim}{\theta}) > 0$ and $\sigma = \sigma(\underset{\sim}{\theta})$. In addition, for robustification we take

$$\mu(\underset{\sim}{\theta}) = \begin{cases} \text{med}(\underset{\sim}{\theta}), & \text{if med}(\underset{\sim}{\theta}) \geq \underset{\sim}{0}, \\ \text{smallest positive } \theta_i, & \text{if med}(\underset{\sim}{\theta}) < 0, \end{cases}$$

and

$$\sigma(\underset{\sim}{\theta}) = \text{med}|\underset{\sim}{\theta} - \mu(\underset{\sim}{\theta})|,$$

where $\text{med}(\underset{\sim}{\theta}) = \text{median of } \{\theta_1, \theta_2, \cdots, \theta_{K-1}\}$ and $\text{med}|\underset{\sim}{\theta} - \mu(\underset{\sim}{\theta})| = \text{median of } \{|\theta_1 - \mu(\underset{\sim}{\theta})|, |\theta_2 - \mu(\underset{\sim}{\theta})|, |\cdots, |\theta_{K-1} - \mu(\underset{\sim}{\theta})|\}$.

To enforce the stochastic ordering, we define two distribution functions, say G and F, such that

$$F_1(\theta) = G(\theta) \text{ and } F_2\left(\frac{\theta - \mu(\underset{\sim}{\theta})}{\sigma(\underset{\sim}{\theta})}\right) = G(\theta)F\left(\frac{\theta - \mu(\underset{\sim}{\theta})}{\sigma(\underset{\sim}{\theta})}\right).$$

This construction is motivated by the works on stochastic ordering of Gelfand and Kuo (1991), developed in Gelfand and Kottas (2001) and used by Kottas and Gelfand (2001) for a semi-parametric Bayesian approach and Kottas, Branco, and Gelfand (2002) and Hanson, Kottas, and Branscum (2008) for a non-parametric Bayesian approach.

Further, rather than a normal distribution, we assume the more heavy tailed and actually more convenient standard logistic density function (location 0 and scale 1) for $\underset{\sim}{\theta}$

$$\theta_1, \theta_2, \cdots, \theta_{K-1} \sim^{iid} \log(0, 1)$$
$$\text{such that } \theta_1 < \theta_2 < \cdots < \theta_{K-1},$$

where $\log(0, 1)$ denotes the logistic distribution with location 0 and scale 1 (i.e., $f(x) = e^x/(1+e^x)^2, -\infty < x < \infty$). That is, $\theta_1, \ldots, \theta_{G-1}$ are order statistics from a standard logistic distribution. Then, the joint prior distribution for $\underset{\sim}{\theta}$ is

$$f(\underset{\sim}{\theta}) = (K-1)! \prod_{i=1}^{K-1} \frac{e^{\theta_i}}{(1+e^{\theta_i})^2}, \quad \theta_1 < \theta_2 < \cdots < \theta_{K-1}.$$

Therefore, using Bayes' theorem, the joint posterior density of $\underset{\sim}{\theta}$ is

$$\pi(\underset{\sim}{\theta} \mid \underset{\sim}{n}_1, \underset{\sim}{n}_2) \propto \prod_{i=1}^{K-1} \Bigg\{ [G(\theta_i) - G(\theta_{i-1})]^{n_{1i}} \times \Bigg[[G(\theta_i) F\left(\frac{\theta_i - \mu(\underset{\sim}{\theta})}{\sigma(\underset{\sim}{\theta})}\right) - G(\theta_{i-1}) F\left(\frac{\theta_{i-1} - \mu(\underset{\sim}{\theta})}{\sigma(\underset{\sim}{\theta})}\right) \Bigg]^{n_{2i}} \frac{e^{\theta_i}}{(1+e^{\theta_i})^2} \Bigg\},$$

where $-\infty = \theta_0 < \theta_1 < \theta_2 < \cdots < \theta_{K-1} < \theta_K = \infty$.

It is convenient to transform the parameters $\theta_i, i = 1, \ldots, K-1$, to parameters in $(0, 1)$. This can potentially remove some computational difficulties when we use the griddy sampler for such awkward posterior densities. So we define

$$\tau_i = \frac{e^{\theta_i}}{1 + e^{\theta_i}}, \quad i = 1, 2, \cdots, K-1,$$

and then note that

$$\theta_i = \text{logit}\,(\tau_i) = \log\left(\frac{\tau_i}{1 - \tau_i}\right), \quad i = 1, 2, \cdots, K-1.$$

It is convenient to work with $\tau_i, i = 1, \ldots, K-1$.

Then, without any information about $\tau_i, i = 1, \ldots, K-1$, we have

$$\tau_1, \tau_2 < \cdots, \tau_{K-1} \sim^{iid} U(0, 1),$$
$$\text{such that } 0 < \tau_1 < \tau_2 < \cdots < \tau_{K-1} < 1.$$

That is, $\tau_1, \ldots, \tau_{K-1}$ are order statistics from a standard uniform distribution. Thus, the joint distribution function of $\underset{\sim}{\tau}$ is

$$f(\underset{\sim}{\tau}) = (K-1)!, \quad \tau_1 < \tau_2 < \cdots < \tau_{K-1}.$$

Therefore, using Bayes' theorem, the joint posterior distribution function of $\underset{\sim}{\tau}$ is

(2)

$$\pi(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2) \propto \prod_{i=1}^{K-1} [G(\text{logit}\,(\tau_i)) - G(\text{logit}\,(\tau_{i-1}))]^{n_{1i}} \times \Bigg[G\left(\text{logit}\,(\tau_i)\right) F\left(\frac{\text{logit}\,(\tau_i) - \mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right) - G\left(\text{logit}\,(\tau_{i-1})\right) F\left(\frac{\text{logit}\,(\tau_{i-1}) - \mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right) \Bigg]^{n_{2i}},$$

where $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{K-1} < \tau_K = 1$.

Finally, we use the griddy Gibbs sampler (Ritter and Tanner 1992) to draw $\tau_i, i = 1, \ldots, K-1$, from their conditional posterior densities,

$$\tau_i|\underset{\sim}{\tau}_{(i)}, \underset{\sim}{n}_1, \underset{\sim}{n}_2, i = 1, 2, \cdots, K-1,$$

where $\underset{\sim}{\tau}_{(i)}$ is the vector of all parameters excluding the i^{th} one. The τ_i are drawn using a grid method; see Nandram and Yin (2016) for an efficient grid method. Of course, this Gibbs sampler will depend on actual forms of G and F.

2.2 Binormal model

In this section, we specify the parametric cumulative distribution functions. Assuming that G and F are normal cumulative distribution distributions, the corresponding posterior function of $\underset{\sim}{\tau}$ in (2) is

$$\pi(\underset{\sim}{\tau}|\underset{\sim}{n}_1,\underset{\sim}{n}_2)\propto\prod_{i=1}^{K-1}\left[\Phi(\text{logit}\,(\tau_i))-\Phi(\text{logit}\,(\tau_{i-1}))\right]^{n_{1i}}\times\left[\Phi\left(\text{logit}\,(\tau_i)\right)\Phi\left(\frac{\text{logit}\,(\tau_i)-\mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)-\Phi\left(\text{logit}\,(\tau_{i-1})\right)\Phi\left(\frac{\text{logit}\,(\tau_{i-1})-\mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)\right]^{n_{2i}},$$

where $0=\tau_0<\tau_1<\tau_2<\cdots<\tau_{K-1}<\tau_K=1$ and $\Phi(.)$ is the cumulative distribution function of standard normal distribution.

We use the griddy Gibbs sampler to draw $\tau_i, i=1,\ldots,K-1$. The conditional posterior density of τ_i is

$$\pi(\tau_i|\underset{\sim}{\tau}_{(i)},\underset{\sim}{n}_1,\underset{\sim}{n}_2)\propto\left[\Phi(\text{logit}\,(\tau_i))-\Phi(\text{logit}\,(\tau_{i-1}))\right]^{n_{1i}}\times\left[\Phi\left(\text{logit}\,(\tau_i)\right)\Phi\left(\frac{\text{logit}\,(\tau_i)-\mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)-\Phi\left(\text{logit}\,(\tau_{i-1})\right)\Phi\left(\frac{\text{logit}\,(\tau_{i-1})-\mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)\right]^{n_{2i}},$$

where $0=\tau_0<\tau_1<\tau_2<\cdots<\tau_{G-1}<\tau_K=1$ and $\underset{\sim}{\tau}_{(i)}$ is the vector of values except the i^{th} component. The Gibbs sampler is run by drawing the τ_i, each in turn, from its conditional posterior density.

In addition, we use $\mu(\underset{\sim}{\theta})=\text{med}(\underset{\sim}{\theta})$ if $\text{med}(\underset{\sim}{\theta})$ is positive or $\mu(\underset{\sim}{\theta})=$ the first non-negative order statistic larger than $\text{med}(\underset{\sim}{\theta})$ if $\text{med}(\underset{\sim}{\theta})$ is negative, and $\sigma(\underset{\sim}{\theta})=\text{med}|\underset{\sim}{\theta}-\mu(\underset{\sim}{\theta})|$. These can be written as functions of the τ_i via the logit transformation.

We obtained starting values for the griddy Gibbs sampler as follows. First, we define

$$\hat{\theta}_{1i}=\Phi^{-1}\left[\frac{\sum_{j=1}^{i}n_{1i}}{\sum_{j=1}^{K-1}n_{1i}}\right],\qquad\hat{\theta}_{2i}=\Phi^{-1}\left[\frac{\sum_{j=1}^{i}n_{2i}}{\sum_{j=1}^{K-1}n_{2i}}\right],$$

and

$$\hat{\sigma}_1^2=\frac{\sum_{i=1}^{K-1}\left(\hat{\theta}_{1i}-\hat{\bar{\theta}}_1\right)^2}{K-2},\qquad\hat{\sigma}_2^2=\frac{\sum_{i=1}^{K-1}\left(\hat{\theta}_{2i}-\hat{\bar{\theta}}_1\right)^2}{K-2}.$$

Then, we take

$$\mu(\hat{\underset{\sim}{\theta}})=\frac{\sum_{i=1}^{K-1}\left(\hat{\theta}_{1i}-\hat{\sigma}_1|\hat{\theta}_2|\right)}{K-1},\qquad\sigma(\hat{\underset{\sim}{\theta}})=\sqrt{\hat{\sigma}_1^2\;\hat{\sigma}_2^2},$$

and

$$\theta_i=\hat{\theta}_{1i},\qquad\tau_i=\frac{e^{\hat{\theta}_i}}{1+e^{\hat{\theta}_i}},\quad i=1,2,3,\cdots,K-1.$$

3. SKEW-BINORMAL MODEL

We use the skew-normal density for both "non-cases" and "cases". The standard skew-normal density with location 0, scale 1 and skewness parameter λ is given by,

$$f(z|\lambda)=2\phi(z)\Phi(\lambda z),\qquad-\infty<z<\infty,$$

where $\phi(z)$ is the standard normal density and $\Phi(z)$ is the cumulative distribution function; see Azzalini (2014) for a book-length review of the skew-normal density and related densities. Note that λ is not the coefficient of skewness, γ_1, but λ and γ_1 have the same sign. For example, if $\lambda<0$, the skew-normal density is negatively skewed, and if $\lambda>0$, it is positively skewed.

It is noted by Azzalini (2014)

(3)
$$E(Z)=\mu_Z=\sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}},\quad\text{Var}(Z)=\sigma_Z^2=1-\frac{1}{\pi}\frac{\lambda^2}{1+\lambda^2},$$

and the coefficient of skewness is given by

$$\gamma_1=\left(\frac{4-\pi}{2}\right)R^3,\tag{4}$$

where $R=\frac{\mu_Z}{\sigma_Z}$. It is also true that

$$\lambda=\frac{R}{\sqrt{\frac{2}{\pi}-\left(1-\frac{2}{\pi}\right)R^2}}.$$

Note that R can also be written as

$$R=\left\{\frac{2\gamma_1}{4-\pi}\right\}^{\frac{1}{3}},$$

which we take as

$$R=\text{sign}(\gamma_1)\left\{\frac{2|\gamma_1|}{4-\pi}\right\}^{\frac{1}{3}}.$$

We use a method of Brys, Hubert and Struyf (2012) to obtain a robust measure of γ_1. Let k_1 of $\theta_1,\theta_2,\cdots,\theta_{K-1}$ be such that $\theta_i\le\text{med}(\underset{\sim}{\theta}),\;\;i=1,2,\cdots,K-1$ and $k_2=(K-1-k_1)$ of $\theta_1,\theta_2,\cdots,\theta_{K-1}$ be such that $\theta_i>\text{med}(\underset{\sim}{\theta}),\;\;i=1,2,\cdots,K-1$. We define $q_{i_0}=\theta_i-\text{med}(\underset{\sim}{\theta})$ for the θ_i values with $\theta_i>\text{med}(\underset{\sim}{\theta})$ and $q_{i_1}=\theta_i-\text{med}(\underset{\sim}{\theta})$ for the θ_i values with $\theta_i>\text{med}(\underset{\sim}{\theta})$. Then we define the fractions,

$$h_{i_0,i_1}=\frac{q_{i_1}+q_{i_0}}{q_{i_1}-q_{i_0}}$$
for all $i_0=1,2,\cdots,k_1$ and $i_1=1,2,\cdots,k_2$.

Finally

$$\gamma_1=\text{med}\{h_{i_0,i_1},\;\;i_0=1,2,\cdots,k_1,\;\;i_1=1,2,\cdots,k_2\}.$$

One drawback of the skew-normal density is that it models moderate skewness. In fact, $|\gamma_1| \leq 0.9953$ (see Azzalini 2014). It is also true that $|\lambda| \leq 10$ and γ_1 is mostly constant outside this range. We will express γ_1 in terms of the skewness of $\theta_1, \theta_2, \cdots, \theta_{K-1}$. Then we will make R a function of $\theta_1, \theta_2, \cdots, \theta_{K-1}$. So that λ will be a function of $\theta_1, \theta_2, \cdots, \theta_{K-1}$.

In our application, the continuous latent variable for the "non-cases" will have a skew-normal density with location 0 scale 1 and skewness parameter $\lambda_1(\underset{\sim}{\theta})$. The "cases" will have a skew-normal density with location $\mu(\underset{\sim}{\theta})$ scale $\sigma(\underset{\sim}{\theta})$ and skewness parameter $\lambda_2(\underset{\sim}{\theta})$. We model $\lambda_1(\underset{\sim}{\theta})$ such that

$$\frac{1}{c}\lambda_2(\underset{\sim}{\theta}) < \lambda_1(\underset{\sim}{\theta}) < c\lambda_2(\underset{\sim}{\theta}), \quad c \geq 1. \tag{5}$$

Here we note an interesting symmetric relation between $\lambda_1(\underset{\sim}{\theta})$ and $\lambda_2(\underset{\sim}{\theta})$ (i.e.,

$$\frac{1}{c}\lambda_1(\underset{\sim}{\theta}) < \lambda_2(\underset{\sim}{\theta}) < c\lambda_1(\underset{\sim}{\theta})).$$

In addition, we assume that

$$\lambda_1(\underset{\sim}{\theta}) = \nu + \lambda_2(\underset{\sim}{\theta}),$$

which implies that

$$\frac{1}{c}\lambda_2(\underset{\sim}{\theta}) < \nu + \lambda_2(\underset{\sim}{\theta}) < c\lambda_2(\underset{\sim}{\theta})$$

and

$$\left(\frac{1}{c} - 1\right)\lambda_2(\underset{\sim}{\theta}) < \nu < (c-1)\,\lambda_2(\underset{\sim}{\theta}).$$

We use $\mu(\underset{\sim}{\theta})$ and $\sigma(\underset{\sim}{\theta})$ in a manner similar to the binormal model.

For our application, we denote the distribution of the latent variable for "non-cases" as $G_{\lambda_2(\underset{\sim}{\theta})+\nu}(z)$ and for the "cases" as $F_{\lambda_2(\underset{\sim}{\theta})}\left(\frac{z-\mu(\underset{\sim}{\theta})}{\sigma(\underset{\sim}{\theta})}\right)$. Then the likelihood function is

$$\begin{aligned}\pi(\underset{\sim}{\theta}, \nu \mid \underset{\sim}{n}_1, \underset{\sim}{n}_2) \propto \prod_{i=1}^{K-1} &\left[G_{\lambda_2(\underset{\sim}{\theta})+\nu}(\theta_i) - G_{\lambda_2(\underset{\sim}{\theta})+\nu}(\theta_{i-1})\right]^{n_{1i}} \\ &\times \left[\left[G_{\lambda_2(\underset{\sim}{\theta})+\nu}(\theta_i) F_{\lambda_2(\underset{\sim}{\theta})}\left(\frac{\theta_i - \mu(\underset{\sim}{\theta})}{\sigma(\underset{\sim}{\theta})}\right)\right.\right. \\ &\left.\left. - G_{\lambda_2(\underset{\sim}{\theta})+\nu}(\theta_{i-1}) F_{\lambda_2(\underset{\sim}{\theta})}\left(\frac{\theta_{i-1} - \mu(\underset{\sim}{\theta})}{\sigma(\underset{\sim}{\theta})}\right)\right]^{n_{2i}},\end{aligned}$$

where $-\infty = \theta_0 < \theta_1 < \theta_2 < \cdots < \theta_{K-1} < \theta_K = \infty$.

We have the same prior distribution of $\theta_1, \theta_2, \cdots, \theta_{K-1}$ as in the binormal model

$$f(\underset{\sim}{\theta}) = (K-1)! \prod_{i=1}^{K-1} \frac{e^{\theta_i}}{(1+e^{\theta_i})^2}, \quad \theta_1 < \theta_2 < \cdots < \theta_{K-1}.$$

In addition, since

$$\left(\frac{1}{c} - 1\right)\lambda_2(\underset{\sim}{\theta}) < \nu < (c-1)\,\lambda_2(\underset{\sim}{\theta}),$$

we use a uniform prior for ν. Specifically,

$$\nu \mid \theta \sim \text{uniform}\left[(\frac{1}{c} - 1)\lambda_2(\underset{\sim}{\theta}), \quad (c-1)\lambda_2(\underset{\sim}{\theta})\right] \quad \text{if } \lambda_2(\underset{\sim}{\theta}) \geq 0$$

or

$$\nu \mid \theta \sim \text{uniform}\left[(c-1)\lambda_2(\underset{\sim}{\theta}), \quad (\frac{1}{c} - 1)\lambda_2(\underset{\sim}{\theta})\right] \quad \text{if } \lambda_2(\underset{\sim}{\theta}) < 0.$$

We actually choose $c = 2$ in our application. A referee has pointed out that inference may be sensitive to the choice of c; see our sensitivity analysis in Section 4.

Here, also we use the transformation from θ to τ as in binormal case. So we define

$$\tau_i = \frac{e^{\theta_i}}{1+e^{\theta_i}}, \quad \text{and then} \quad \theta_i = \text{logit}\,(\tau_i) = \log\left(\frac{\tau_i}{1-\tau_i}\right), \quad i = 1, 2, \cdots, K-1.$$

Then the joint prior distribution function of τ is

$$f(\underset{\sim}{\tau}) = (K-1)!, \quad \tau_1 < \tau_2 < \cdots < \tau_{K-1}.$$

Therefore, the joint posterior distribution function of $\underset{\sim}{\tau}$ and ν is

$$\begin{aligned}&\pi(\underset{\sim}{\tau}, \nu | \underset{\sim}{n}_1, \underset{\sim}{n}_2) \\ &\quad \propto \prod_{i=1}^{K-1} \left[G_{\lambda_2(\underset{\sim}{\tau})+\nu}(\text{logit}\,(\tau_i)) - G_{\lambda_2(\underset{\sim}{\tau})+\nu}(\text{logit}\,(\tau_{i-1}))\right]^{n_{1i}} \\ &\qquad \times \left[G_{\lambda_2(\underset{\sim}{\tau})+\nu}\left(\text{logit}\,(\tau_i)\right) F_{\lambda_2(\underset{\sim}{\tau})}\left(\frac{\text{logit}\,(\tau_i) - \mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)\right. \\ &\qquad - G_{\lambda_2(\underset{\sim}{\tau})+\nu}\left(\text{logit}\,(\tau_{i-1})\right) \\ &\qquad \left. \times F_{\lambda_2(\underset{\sim}{\tau})}\left(\frac{\text{logit}\,(\tau_{i-1}) - \mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)\right]^{n_{2i}},\end{aligned}$$

where $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{K-1} < \tau_K = 1$ with the appropriate range on ν and $G_{\lambda_2(\underset{\sim}{\tau})+\nu}(.)$ is the corresponding function of $G_{\lambda_2(\underset{\sim}{\theta})+\nu}(.)$ for transformed parameter $\underset{\sim}{\tau}$ (i.e., when θ_i is replaced by $\text{logit}\,(\tau_i)$, $i = 1, 2, \cdots, K-1$).

The griddy Gibbs sampler (Ritter and Tanner 1992) is used to draw $\tau_i, i = 1, 2, \cdots, K-1$, and ν. The conditional posterior density of τ_i is

$$\begin{aligned}&\pi(\underset{\sim}{\tau}_i | \underset{\sim}{\tau}_{(i)}, \nu, \underset{\sim}{n}_1, \underset{\sim}{n}_2) \\ &\quad \propto \prod_{i=1}^{K-1} \left[G_{\lambda_2(\underset{\sim}{\tau})+\nu}(\text{logit}\,(\tau_i)) - G_{\lambda_2(\underset{\sim}{\tau})+\nu}(\text{logit}\,(\tau_{i-1}))\right]^{n_{1i}} \\ &\qquad \times \left[G_{\lambda_2(\underset{\sim}{\tau})+\nu}\left(\text{logit}\,(\tau_i)\right) F_{\lambda_2(\underset{\sim}{\tau})}\left(\frac{\text{logit}\,(\tau_i) - \mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)\right. \\ &\qquad - G_{\lambda_2(\underset{\sim}{\tau})+\nu}\left(\text{logit}\,(\tau_{i-1})\right) \\ &\qquad \left. \times F_{\lambda_2(\underset{\sim}{\tau})}\left(\frac{\text{logit}\,(\tau_{i-1}) - \mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)\right]^{n_{2i}},\end{aligned} \tag{6}$$

where $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{K-1} < \tau_K = 1$ and $\underset{\sim}{\tau}_{(i)}$ is the vector of values except the i^{th} component.

Once $\theta_1, \theta_2, \cdots, \theta_{K-1}$ have been drawn, we use $\mu(\underset{\sim}{\theta}) = \text{med}(\underset{\sim}{\theta})$ if $\text{med}(\underset{\sim}{\theta})$ is positive or $\mu(\underset{\sim}{\theta}) =$ the first non negative order statistic larger than $\text{med}(\underset{\sim}{\theta})$ if $\text{med}(\underset{\sim}{\theta})$ is negative, and $\sigma(\underset{\sim}{\theta}) = \text{med}|\underset{\sim}{\theta} - \mu(\underset{\sim}{\theta})|$ as we have used in binormal case. Here, $\lambda_2(\underset{\sim}{\tau})$ is obtained using the procedure outlined at the beginning of this section starting at equation (3) coupled with the robust measure of skewness (Brys, Hubert and Sturyf 2012).

Finally, the conditional density of ν is

(7)

$$\begin{aligned}
&\pi(\nu|\underset{\sim}{\tau}, \nu, \underset{\sim}{n}_1, \underset{\sim}{n}_2) \\
&\quad \propto \prod_{i=1}^{K-1} \left[G_{\lambda_2(\underset{\sim}{\tau})+\nu}(\text{logit}\,(\tau_i)) - G_{\lambda_2(\underset{\sim}{\tau})+\nu}(\text{logit}\,(\tau_{i-1}))\right]^{n_{1i}} \\
&\qquad \times \left[G_{\lambda_2(\underset{\sim}{\tau})+\nu}\left(\text{logit}\,(\tau_i)\right) F_{\lambda_2(\underset{\sim}{\tau})}\left(\frac{\text{logit}\,(\tau_i) - \mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)\right. \\
&\qquad - G_{\lambda_2(\underset{\sim}{\tau})+\nu}\left(\text{logit}\,(\tau_{i-1})\right) \\
&\qquad \left. \times F_{\lambda_2(\underset{\sim}{\tau})}\left(\frac{\text{logit}\,(\tau_{i-1}) - \mu(\underset{\sim}{\tau})}{\sigma(\underset{\sim}{\tau})}\right)\right]^{n_{2i}}.
\end{aligned}$$

with the appropriate range. Again, the grid method is used to draw ν in (7). Therefore, to execute the Gibbs sampler, we draw $\tau_1, \ldots, \tau_{K-1}$ from (6), each in turn, and then ν from (7) and the entire procedure is repeated until convergence; see Section 4.

Here also we use the starting values for $\mu(\underset{\sim}{\theta})$, $\sigma(\underset{\sim}{\theta})$ and τ_i for the griddy Gibbs sampler as for the binormal case. We also use $\nu = 0$ and $\lambda_1(\underset{\sim}{\theta}) = \lambda_2(\underset{\sim}{\theta})$ as starting values.

4. APPLICATIONS TO THE MEAL, READY-TO-EAT

In this section, we present numerical results based on the Natick Food Experiment (NFE). In Section 4.1, we describe the NFE. In Section 4.2, we first show how to use marginal likelihoods to choose between the binormal model and skew-binormal model for the foods. Then, in Section 4.3, we compare the ROC curves and the AUCs obtained from these two models. We also discuss sensitivity to choice of c in (5). In Section 4.4, we perform a small simulation study.

4.1 A brief description of the MREs

The data we study arise from the Natick Food Experiment (NFE), in which at each storage time and temperature, sensory data on a nine-point hedonic scale are obtained for the Meal, Ready-to-Eat (MRE) from 36 panelists. The MRE has twelve meals (menus), each meal consisting of four to six food items. On arrival at Natick Laboratories (NLABS), the meals were inspected for completeness and stored at four different temperatures, 4^0C, 21^0C, 30^0C, and 38^0C, then withdrawn and tested at 6, 12, 18, 24, 30, 36, 48, and 60 months. Description of the NFE is given by Chen, Nandram and Ross (1996) and the references therein. More recent discussions, but very limited because of confidentiality, is given in US ARMY RDECOM (2014).

The meals were opened by test monitors, and each item served to a panel of 36 untrained subjects who judged its acceptability on a nine-point hedonic rating scale where 1 = dislike extremely, 9 = like extremely, and intermediate integer scores have graduated meanings. For a detailed analysis of these data, see Chen, Nandram and Ross (1996). Other interpretations of these data are given by Nandram (2005, 1998, 2012). Our analysis in this paper is much different as we want to assess deterioration over two withdrawals rather than the prediction of shelf-lives that are too large. Such analyses have not been done with these data. We study the entrees kept at 21 degrees centigrade withdrawn after 12, 18, 24, 30, 36, 48 and 60 months.

For the nine-point hedonic scale, many of the cells are empty because there are only thirty six panelists. Therefore, our methods become unstable, so we have to collapse the table into a five-point hedonic scale. Specifically, we add cells 1 and 2, 3 and 4 to form respectively new cells 1 and 2, and we add 6 and 7 and 8 and 9 to form respectively new cells 4 and 5; the old cell 5 is the new cell 3. There are still one or two empty cells and some foods and some withdrawals, but our methods are much more stable on a 5-point scale rather than a 9-point scale.

For each food we study two withdrawal times, fresh foods and foods withdrawn later. Thus, for each withdrawal time, there is a five-cell categorical table of 36 panelists. Our task is to see if fresh foods are different from foods withdrawn later. To this end, we apply our methodology to these data to obtain the ROC curves and the AUCs to assess any deterioration. We are making a reasonable assumption that the panelists are very accurate in scoring.

As stated above, this is an older dataset that has twelve meals (menus), but nowadays the MRE has twenty-four meals. The procedure that is used nowadays is very similar to the one we described in this section. The new datasets are confidential and are not available to the public. However, our methodology is still useful to NLABS and we will make it accessible to them.

4.2 Comparison of the binormal model and the skew-binormal model

We use marginal likelihoods to compare the binormal model and the skew-binormal model. The integration needed is done using Monte Carlo methods.

For the binormal model, the marginal likelihood is

$$P_{BN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2) = \int\int\cdots\int_{\tau_1<\tau_2<\cdots<\tau_{K-1}} P(\underset{\sim}{n}_1, \underset{\sim}{n}_2|\underset{\sim}{\tau})\pi(\underset{\sim}{\tau})d\underset{\sim}{\tau},$$

and for the skew-binormal model, the marginal likelihood is

$$P_{SBN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2)$$
$$= \int\int\cdots\int_{\tau_1<\tau_2<\cdots<\tau_{K-1}} P(\underset{\sim}{n}_1, \underset{\sim}{n}_2|\underset{\sim}{\tau}, \nu)\pi(\underset{\sim}{\tau}, \nu)d\underset{\sim}{\tau}d\nu.$$

If $P_{BN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2) > P_{SBN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2)$, the binormal model is better than the skew-binormal model and vice versa.

We use the method of Nandram and Kim (2002) (with a slight modification) to compute $P_{BN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2)$ and $P_{SBN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2)$. Specifically,

$$P_{BN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2)$$
$$= \frac{\int\int\cdots\int_{\tau_1<\tau_2<\cdots<\tau_{K-1}} P(\underset{\sim}{n}_1, \underset{\sim}{n}_2|\underset{\sim}{\tau})\frac{\pi(\underset{\sim}{\tau})}{\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)}\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)d\underset{\sim}{\tau}}{\int\int\cdots\int_{\tau_1<\tau_2<\cdots<\tau_{K-1}} \frac{\pi(\underset{\sim}{\tau})}{\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)}\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)d\underset{\sim}{\tau}},$$

where $\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$ is an approximation to $\pi(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$ and it is used to draw sample from $\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$, of course, such that $\tau_1 < \tau_2 < \cdots < \tau_{K-1}$. Suppose we can have a sample $\tau^{(h)}, h = 1, 2, \cdots, M$ (e.g. M=1000) from $\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$. Then we approximate $P_{BN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2)$ by

$$\widehat{P_{BN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2)} = \sum_{h=1}^{M} w_h P(\underset{\sim}{n}_1, \underset{\sim}{n}_2|\underset{\sim}{\tau}^{(h)}),$$
$$w_h = \frac{\pi(\underset{\sim}{\tau}^{(h)})}{\pi_a(\underset{\sim}{\tau}^{(h)}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)}, \quad h = 1, 2, \cdots, M.$$

Similarly, for $P_{SBN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2)$ we calculate

$$\widehat{P_{SBN}(\underset{\sim}{n}_1, \underset{\sim}{n}_2)} = \sum_{h=1}^{M} w_h P(\underset{\sim}{n}_1, \underset{\sim}{n}_2|\underset{\sim}{\tau}^{(h)}, \nu^{(h)}),$$
$$w_h = \frac{\pi(\underset{\sim}{\tau}^{(h)}, \nu^{(h)})}{\pi_a(\underset{\sim}{\tau}^{(h)}, \nu^{(h)}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)}, \quad h = 1, 2, \cdots, M.$$

Here $(\underset{\sim}{\tau}^{(h)}, \nu^{(h)}), \quad h = 1, 2, \cdots, M$, is a sample from $\pi_a(\underset{\sim}{\tau}, \nu|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$.

We have samples from $\pi(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$ and $\pi(\underset{\sim}{\tau}, \nu|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$ that we can use to construct samples from $\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$ and $\pi_a(\underset{\sim}{\tau}, \nu|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$. Here

$$\pi_a(\underset{\sim}{\tau}, \nu|\underset{\sim}{n}_1, \underset{\sim}{n}_2) = \pi_a(\nu|\underset{\sim}{\tau}, \underset{\sim}{n}_1, \underset{\sim}{n}_2)\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2),$$

where

$$\nu \sim U\left(\left(\frac{1}{c} - 1\right)\lambda_2(\underset{\sim}{\theta}), (c-1)\,\lambda_2(\underset{\sim}{\theta})\right).$$

In Appendix A we show how to obtain the proposal density, $\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$ and in Appendix B we show how to obtain the standard errors of the log-marginal likelihood and the log Bayes factor. We obtain two proposal densities which are both based on the beta distribution.

In Figure 1, we plot the marginal log-likelihood of the skew-binormal model versus the binormal model for all foods and all seven withdrawal times. We see that most of the points lie above the 45° straight line for both proposal densities (55/84 for the top panel and 58/84 for the bottom panel). The one based on the uniform proposal density is much simpler than that based on the general beta distribution; see Appendix A.

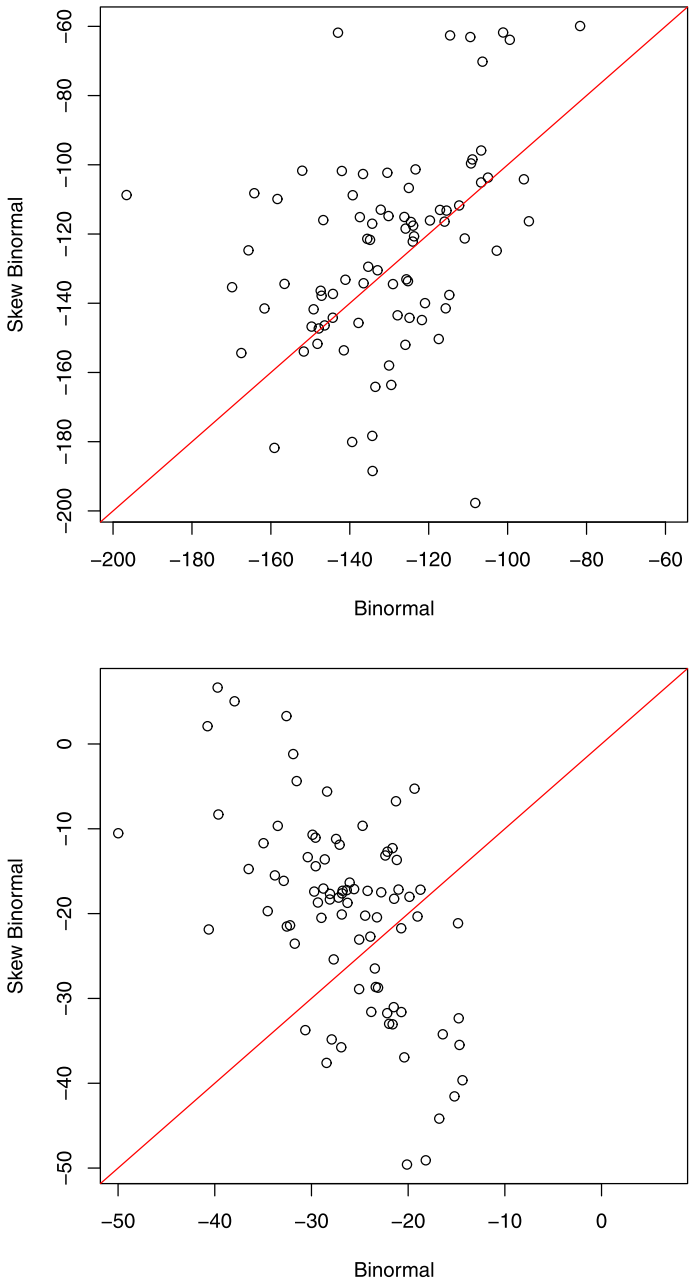

Figure 1. Comparison of the log-marginal likelihoods under the binormal model and the skew-binormal model using beta prior (top) and uniform prior (bottom).

For greater detail in Table 1, we present these comparisons for two withdrawal times (12 months and 60 months). First, note that for all foods the standard errors are reasonably small with the standard errors under the binormal model smaller than those under the skew-binormal model. This is due to the fact that it is more difficult to compute

Table 1. Log marginal likelihoods (LML) and their standard errors (SE) under the binormal model (BN) and skew binormal model (SBM), log Bayes factors (LBF) and their standard errors for fresh food versus food withdrawn at 12 months (top) and for fresh food versus food withdrawn at 60 months (bottom)

Food	LML(BN)	SE(BN)	LML(SBN)	SE(SBN)	LBF	SE(LBF)
1	-29.581	1.444	-11.046	2.962	18.536	2.586
2	-30.383	1.176	-13.342	2.248	17.041	1.916
3	-27.452	1.371	-11.201	2.657	16.251	2.276
4	-27.090	1.235	-11.876	2.916	15.213	2.641
5	-21.261	1.353	-6.753	3.021	14.508	2.702
6	-16.785	1.670	-44.174	2.426	-27.389	1.760
7	-20.721	1.251	-21.728	2.420	-1.007	2.072
8	-25.076	1.145	-28.903	2.015	-3.827	1.658
9	-23.910	1.595	-22.713	3.033	1.197	2.579
10	-14.841	1.363	-21.135	2.314	-6.294	1.869
11	-40.747	1.598	2.093	3.869	42.841	3.523
12	-26.922	1.266	-35.761	2.048	-8.839	1.610

Food	LML(BN)	SE(BN)	LML(SBN)	SE(SBN)	LBF	SE(LBF)
1	-34.957	1.444	-11.697	2.952	23.260	2.575
2	-24.201	1.396	-17.319	2.333	6.882	1.870
3	-49.985	1.242	-10.524	2.543	39.461	2.220
4	-28.972	1.328	-20.498	2.974	8.474	2.662
5	-28.386	1.445	-5.611	3.095	22.774	2.737
6	-15.207	1.547	-41.548	2.491	-26.341	1.952
7	-24.448	1.289	-20.240	2.311	4.208	1.918
8	-14.380	1.310	-39.649	1.950	-25.269	1.444
9	-32.550	1.484	-21.516	3.014	11.034	2.624
10	-32.863	1.247	-16.136	2.329	16.727	1.968
11	-31.531	1.631	-4.373	3.823	27.158	3.458
12	-28.443	1.218	-37.599	2.063	-9.156	1.665

NOTE: Formulas for SEs are shown in Appendix B.

the marginal likelihood under the skew-binormal model. For 12 months' withdrawal, as judged by the Bayes factor, we notice that there are six foods (1, 2, 3, 4, 5, 11) for which the skew-binormal model is better, three foods (6, 10, 12) in which the binormal model is better, and three foods (7, 8, 9) in which it is difficult to tell the better model when the standard errors are considered. For 60 months' withdrawal, again as judged by the Bayes factor, we notice that there are nine foods (1, 2, 3, 4, 5, 7, 9, 10, 11) in which the skew-binormal model is better than the binormal model, and three foods (6, 8, 12) in which the binormal model is better than the skew-binormal. Based on the marginal likelihoods, there are no foods with inconclusive decisions.

4.3 Numerical results

We now discuss the key results for the MREs. We present the posterior mean (PM), posterior standard deviation (PSD) and 95% credible interval for each of the AUCs. We note that in all foods and both models the numerical standard errors (NSE) are small.

We first give a summary of the performance of the two Gibbs samplers (the binormal model and the skew-binormal model). We have run both samplers in exactly the same way. That is, we drew 11,000 iterates, used 1,000 as a "burn in" and took every 10^{th} iterate thereafter. When we fit both models, we looked at the auto-correlations among the iterates, used the Geweke test of stationarity with trace plots and the effective sample size. We have done this for all 84 Gibbs samplers (12 foods at 7 withdrawals) for each model. The auto-correlations were always negligible, the Geweke test almost always gave acceptable p-values (larger than .05) and effective sample sizes were mostly 1,000 except for a few Gibbs samplers under the skew-binormal model where the effective sample sizes reach 500. So for the skew-binormal model, we have increased the number of runs to 15,000, used 5,000 as a "burn-in" and choose every 10^{th} iterate. This brought the effective sizes to nearly 1,000. The entire procedure is set up to run automatically with no intervention from the user; we can run fewer iterations but would need individual attention to run the 84 Gibbs samplers for each model. In the end, the computer time to run one Gibbs sampler for the binormal model on our Linux Computational Node with 2.70GHz and 8 CPU Cores is about three seconds whereas the skew-binormal model took about three minutes. Much of the time taken by the Gibbs sampler un-

Table 2. Posterior mean (PM), standard deviation (PSD), numerical standard error (NSE) and 95% credible interval (CI) of the area under the curve using the binormal model (left) and the skew-binormal model (right) for fresh food versus food withdrawn at 12 months

Food	PM	PSD	NSE	95% CI
1	.650	.018	.00058	(.631, .670)
2	.652	.014	.00043	(.633, .678)
3	.649	.015	.00041	(.630, .667)
4	.659	.028	.00107	(.632, .682)
5	.668	.028	.00095	(.645, .688)
6	.690	.031	.00108	(.639, .753)
7	.660	.024	.00075	(.637, .683)
8	.655	.016	.00052	(.631, .687)
9	.650	.026	.00084	(.634, .663)
10	.658	.024	.00078	(.640, .678)
11	.730	.083	.01071	(.626, .833)
12	.647	.013	.00036	(.630, .670)

Food	PM	PSD	NSE	95% CI
1	.724	.022	.00069	(.685, .769)
2	.676	.041	.00196	(.640, .783)
3	.777	.027	.00080	(.729, .828)
4	.754	.025	.00073	(.703, .798)
5	.729	.025	.00073	(.682, .776)
6	.727	.032	.00264	(.666, .788)
7	.760	.023	.00072	(.714, .805)
8	.679	.025	.00094	(.642, .723)
9	.731	.020	.00066	(.693, .771)
10	.778	.027	.00085	(.729, .827)
11	.684	.018	.00054	(.650, .721)
12	.679	.028	.00081	(.636, .714)

NOTE: NSEs are obtained by the batch-means method.

Table 3. Posterior mean (PM), standard deviation (PSD), numerical standard error (NSE) and 95% credible interval (CI) of the area under the curve using the binormal model (left) and the skew-binormal model (right) for fresh food versus food withdrawn at 60 months

Food	PM	PSD	NSE	95% CI
1	.645	.014	.00044	(.628, .665)
2	.664	.016	.00057	(.637, .700)
3	.640	.009	.00026	(.627, .657)
4	.654	.028	.00100	(.628, .675)
5	.661	.033	.00107	(.637, .679)
6	.675	.028	.00087	(.633, .728)
7	.657	.019	.00049	(.635, .679)
8	.666	.019	.00070	(.640, .693)
9	.647	.020	.00057	(.629, .661)
10	.647	.014	.00043	(.630, .665)
11	.739	.077	.00544	(.639, .848)
12	.649	.016	.00054	(.628, .681)

Food	PM	PSD	NSE	95% CI
1	.728	.021	.00064	(.687, .768)
2	.705	.062	.00554	(.644, .842)
3	.776	.028	.00087	(.722, .824)
4	.754	.025	.00069	(.712, .805)
5	.728	.026	.00077	(.679, .776)
6	.728	.034	.00291	(.664, .791)
7	.762	.023	.00072	(.717, .807)
8	.679	.026	.00114	(.634, .729)
9	.733	.020	.00068	(.696, .772)
10	.765	.044	.00438	(.654, .833)
11	.684	.018	.00055	(.649, .719)
12	.682	.028	.00095	(.639, .719)

NOTE: NSEs are obtained by the batch-means method.

der the skew-binormal model comes from computing cdfs and inverse cdfs of the skew-binormal densities; the standard normal cdf and its inverse take respectively less time, of course.

In Table 2, we present posterior inference about the AUCs for foods withdrawn after 12 months. Under the binormal model, the AUCs are all around .650 with food 11 having an AUC of .730. Also, the PSDs are small, thereby making the 95% credible intervals (CI) narrow. On the other hand, under the skew-binormal model, the PMs are roughly around .750, with some of them a bit smaller. The PMs of the AUCs under the skew-binormal model are larger than those under the binormal model. There are some foods (e.g., 1, 3, 4, 7, 9, 10) for which the 95% credible intervals under the binormal model are completely to the left of those under the skew-binormal model and, in fact, for four of these foods the skew-binormal model fits better. In Figure 2, we present the corresponding ROC curves for both the binormal model and the skew-binormal model. We notice that for some foods (1, 3, 4, 7, 9, 10), the ROC curves under the skew-normal model are completely above those of the binormal model; other ROC curves overlap. For all foods, the 95% point-wise bands all being above the 45° straight line. Therefore, we can say that the panelists were able to distinguish between fresh foods and foods withdrawn 12 months later, albeit not very strongly. It is important to enforce the stochastic ordering; otherwise the lower band of the ROC curves near to the origin can go below the 45° (not appealing for ROC curves). However, we note that the ROC curves under the skew-binormal model are much smoother with more curvature than under the binormal model.

In Table 3 we present posterior inference about the AUCs for foods withdrawn after 60 months. In Figure 3, we present the corresponding ROC curves with the 95% pointwise bands all being above the 45° straight line. We notice that for three foods (1, 3, 7) the ROC curves under the skew-binormal model are almost always above those under the binormal model. Overall, the conclusions are similar to foods withdrawn after 12 months.

If we assume that the panelists can discriminate reasonably well among the different withdrawals, we can make the following conclusion. The 95% credible intervals are all to

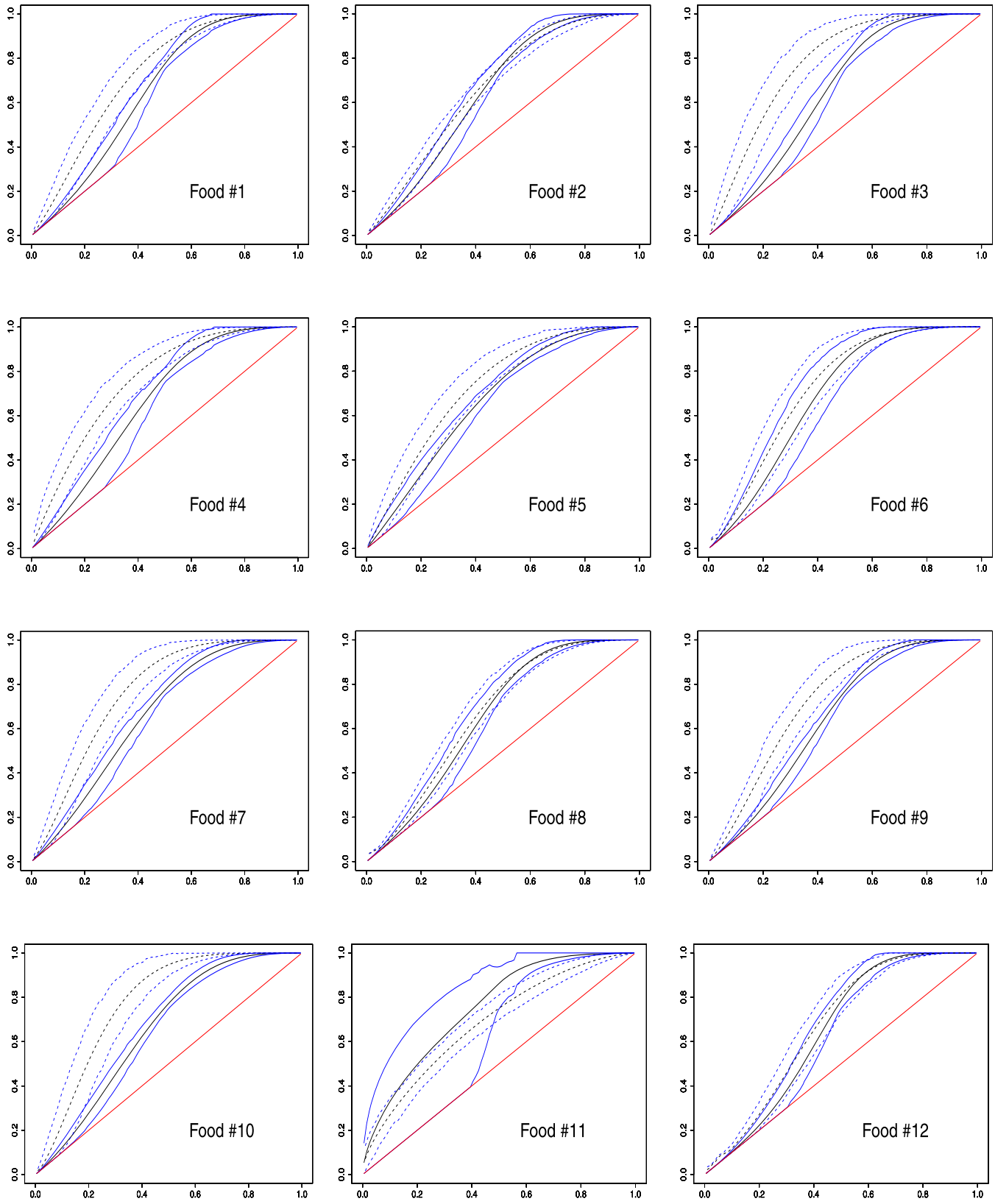

Figure 2. ROC curves and 95% pointwise credible intervals under the binormal model (solid) and the skew-binormal model (dotted) for fresh foods versus foods withdrawn at 12 months. Axes: X- false positive rate, Y- true positive rate.

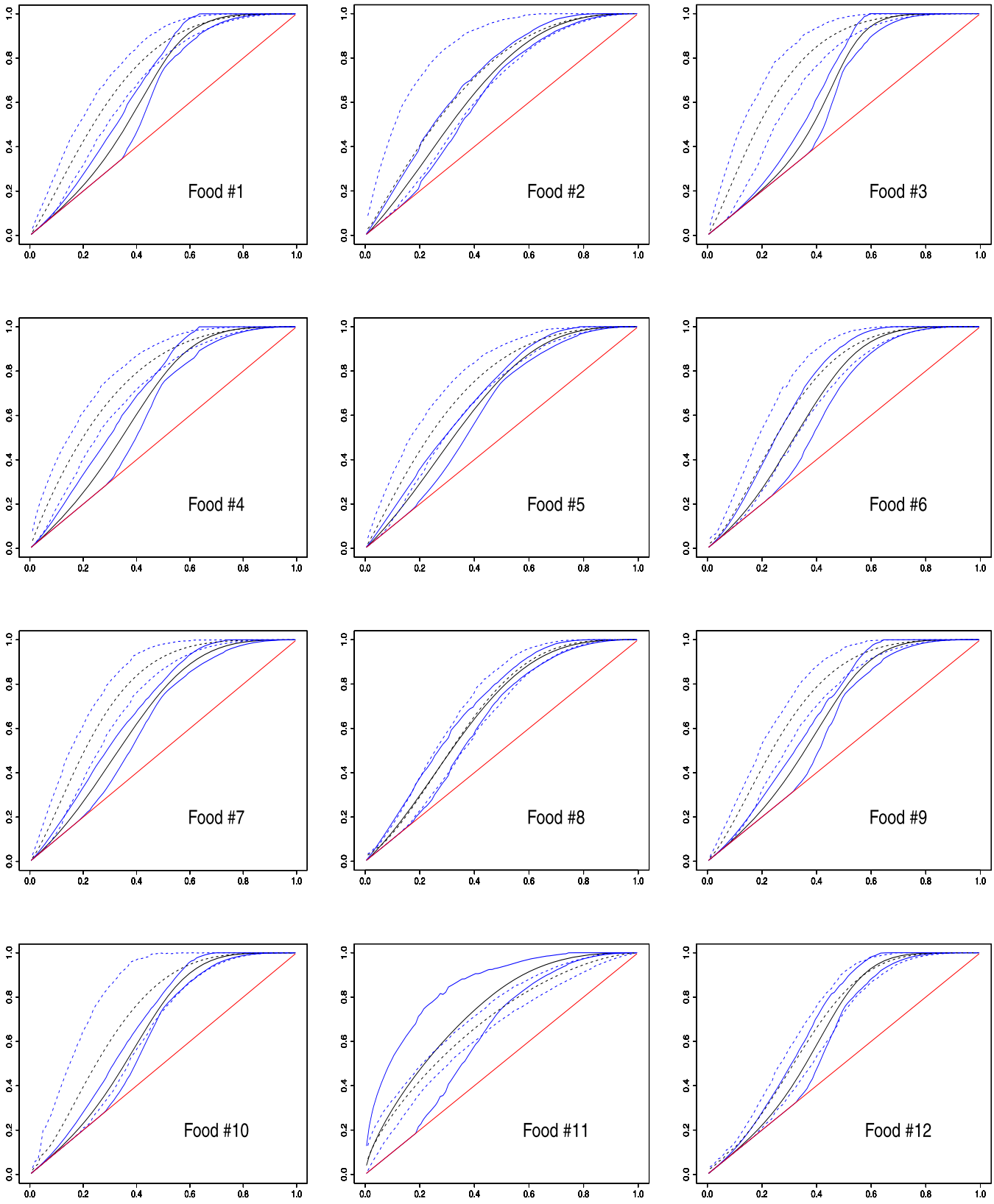

Figure 3. ROC curves and 95% pointwise credible intervals under the binormal model (solid) and the skew-binormal model (dotted) for fresh foods versus foods withdrawn at 60 months. Axes: X- false positive rate, Y- true positive rate.

Table 4. Sensitivity analysis under the skew-binormal model: Posterior mean (PM) and posterior standard deviation (PSD) of the area under the curve by c for fresh food versus food withdrawn at $W = 12$ months and $W = 60$ months

		c=1.5		c=2		c=3		c=5		c=10		c=15	
Food	W	PM	PSD	PM	PSD	PM	PSD	PM	PSD	PM	PSD	PM	PSD
1	12	0.720	0.023	0.724	0.022	0.726	0.025	0.727	0.026	0.724	0.026	0.725	0.026
	60	0.723	0.023	0.728	0.021	0.734	0.024	0.737	0.026	0.734	0.026	0.734	0.027
2	12	0.675	0.036	0.676	0.041	0.663	0.039	0.660	0.040	0.654	0.040	0.654	0.039
	60	0.689	0.049	0.705	0.062	0.725	0.075	0.750	0.076	0.768	0.070	0.766	0.072
3	12	0.775	0.030	0.777	0.027	0.783	0.025	0.788	0.025	0.785	0.025	0.785	0.026
	60	0.774	0.030	0.776	0.028	0.778	0.028	0.779	0.028	0.776	0.028	0.775	0.031
4	12	0.752	0.028	0.754	0.025	0.757	0.025	0.759	0.025	0.758	0.024	0.758	0.024
	60	0.752	0.028	0.755	0.025	0.762	0.021	0.769	0.021	0.770	0.021	0.771	0.021
5	12	0.727	0.026	0.729	0.025	0.727	0.028	0.727	0.028	0.726	0.028	0.727	0.028
	60	0.726	0.026	0.728	0.026	0.726	0.029	0.726	0.029	0.725	0.029	0.726	0.029
6	12	0.733	0.032	0.727	0.032	0.731	0.032	0.730	0.032	0.730	0.033	0.730	0.032
	60	0.733	0.033	0.728	0.034	0.731	0.033	0.731	0.033	0.730	0.034	0.732	0.034
7	12	0.755	0.026	0.760	0.023	0.769	0.024	0.773	0.026	0.772	0.026	0.771	0.026
	60	0.756	0.026	0.762	0.023	0.775	0.022	0.782	0.024	0.782	0.025	0.782	0.025
8	12	0.679	0.021	0.679	0.025	0.667	0.023	0.661	0.024	0.658	0.021	0.657	0.025
	60	0.679	0.022	0.679	0.026	0.665	0.022	0.659	0.024	0.656	0.026	0.655	0.027
9	12	0.724	0.021	0.731	0.020	0.746	0.024	0.754	0.028	0.757	0.031	0.761	0.033
	60	0.725	0.021	0.733	0.020	0.750	0.023	0.759	0.027	0.764	0.031	0.769	0.032
10	12	0.775	0.029	0.778	0.026	0.782	0.028	0.778	0.038	0.771	0.046	0.770	0.047
	60	0.771	0.033	0.765	0.044	0.682	0.059	0.673	0.060	0.654	0.050	0.641	0.032
11	12	0.682	0.019	0.684	0.018	0.690	0.020	0.694	0.022	0.699	0.023	0.701	0.024
	60	0.682	0.019	0.684	0.018	0.689	0.020	0.693	0.022	0.697	0.023	0.699	0.024
12	12	0.679	0.025	0.679	0.028	0.669	0.025	0.666	0.026	0.663	0.024	0.677	0.054
	60	0.681	0.025	0.682	0.028	0.676	0.028	0.764	0.081	0.672	0.030	0.671	0.030

the right of .50 and so statistically there is a deterioration of the foods under either model. However, in practice because the ROC curves are not too far above the 45° line, we can conclude that the foods actually deteriorate but not substantially over the five-year period.

We have performed a sensitivity analysis to see how posterior inference changes with c that we specified in our analysis as $c = 2$. With this in mind, in Table 4 we have compared posterior mean (PM) and posterior standard deviation (PSD) of the AUCs for the 12 foods and two withdrawal times, $W = 12$, $W = 60$, at $c = 1.5, 2.0, 3.0, 5.0, 10.0, 15.0$. For each food, there are negligible changes in the PMs and PSDs as c changes. This is true at both $W = 12$ and $W = 60$. There are some foods in which there is an increase of PM with c, the variation in the PSDs is small. For example, for Food 7, PM increases from .755 to .773 at $W = 12$ and .756 to .782 at $W = 60$. But these increases are minor. We conclude that around $c = 2$, inference about the AUCs is hardly sensitive as c changes.

4.4 Simulation study

We have performed a simulation study to assess the difference between the binormal model and the skew-binormal model. As in our data analysis, we consider the five-point hedonic scale. We choose the probabilities that a panelist responded in the j^{th} cell of five-cell contingency table as $p_j, j = 1, \ldots 5$. We have selected two scenarios for these probabilities $p_1 = (.50, .25, .15, .05, .05)$ for the non-cases and $p_2 = (.05, .20, .50, .20, .05)$ for the cases for a small degree of separation and $p_1 = (.50, .25, .15, .05, , 05)$ for the non-cases and $p_2 = (.05, .05, .150, .25, .50)$ for the cases for a medium degree of separation. We have chosen two sets of panel sample sizes, 36 and 72. Thus, we have a 2^2 factorial design with separation (2 levels) and sample size (2 levels).

We generated 200 runs of the cell counts according to separation probabilities and sample sizes from multinomial distributions. We noticed one or two cells are zeros, so we enforced a restriction of positive cell counts in the simulation to avoid difficulties in fitting the models. Except for parallel computing, we fit the binormal model and the skew-binormal model in the same manner as we did for the food data. The MCMC diagnostics for all 800 runs show acceptable performance for convergence. In Figure 4 and Table 5, we averaged the posterior means of the ROC curves, and the posterior means and posterior standard deviations of the AUCs over the 200 runs at each of the four design points by model.

In Table 5 we present the results of the AUCs by design point and model. For all design points under the skew-binormal model, the posterior means are nearly the same.

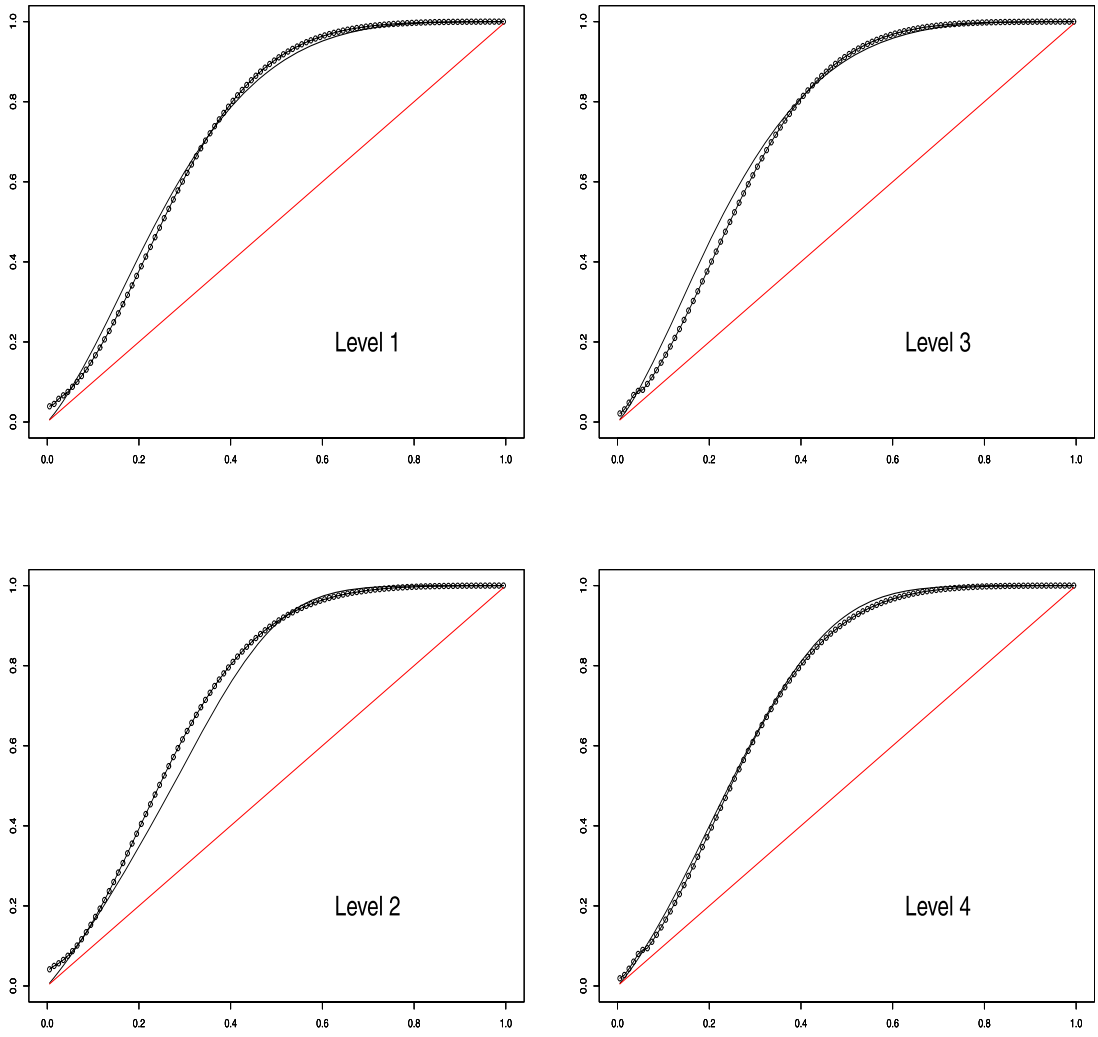

Figure 4. ROC curves for the simulated data for the binormal model (solid) and the skew-binormal model (dotted). Axes: X-false positive rate, Y-true positive rate. The design points are (a) Level-1: small separation, 36 panelists; (b) Level-2: medium separation, 36 panelists; (c) Level-3: small separation, 72 panelists; and (d) Level-4: medium separation, 72 panelists.

Table 5. Simulation study: Comparison of the binormal model and the skew-binormal model using the AUC averaged over the 200 runs by separation and sample size

			PM	PSD
Binormal	36	1	$0.733_{[0.0022]}$	$0.034_{[0.0003]}$
		2	$0.720_{[0.0027]}$	$0.041_{[0.0007]}$
	72	1	$0.748_{[0.0018]}$	$0.025_{[0.0001]}$
		2	$0.741_{[0.0023]}$	$0.034_{[0.0005]}$
Skew-binormal	36	1	$0.731_{[0.0020]}$	$0.032_{[0.0002]}$
		2	$0.735_{[0.0019]}$	$0.033_{[0.0002]}$
	72	2	$0.736_{[0.0016]}$	$0.025_{[0.0001]}$
		2	$0.734_{[0.0016]}$	$0.025_{[0.0001]}$

NOTE: The notation $a_{[b]}$ means that a is the average and b the standard error over the 200 runs.

There are some differences under the binormal model. For 36 panelists the summaries are a bit different at medium separation (compare PM= .733 and PSD= .034 with PM= .720 and PSD= .041). We notice a significant drop in the PSDs from 36 panelists to 72 panelists regardless of separation and model. Surprisingly, for 72 panelists at medium separation PSD = .034 under the binormal model and PSD = .025 under the skew-binormal model.

In Figure 4 we present the average posterior mean ROC curves under the binormal model (solid) and the skew-binormal (dotted) by design point. [Because of space, we put all four figures on the same page, but here we describe the figures from the original single pages.] For 36 panelists and at small separation the two curves cross each other near .4, thereby explaining why the AUCs are similar. At medium separation, there is a significant difference in the ROC curves with the one for the skew-binormal model sig-

nificantly above the one for the binormal model mostly in the interval (.15, .45), again explaining why the AUC is lower at this design point. For 72 panelists and at the first separation the two curves cross each other near .4 with the ROC curve much lower for the skew-binormal below .4, thereby again explaining why the AUCs are similar. For 72 panelists and at the second separation the two curves are very similar but still crossing each other, again explaining why the AUCs are similar.

In the simulations we have noticed some differences between the two models at the four design points, but it is difficult to pin down any useful pattern. We have attempted greater separation, but the skew-binormal model becomes difficult to fit because there are one or two cells with zero counts.

5. CONCLUDING REMARKS

We have developed a robust Bayesian analysis to study count data obtained from a nine-point hedonic scale at two time points. We have robustified a standard binormal model by modeling skewness. Our robust procedure is not as complicated as in Bayesian non-parametric statistics (e.g. Dirichlet Process). In fact, we have used robust summaries to get an improvement over a standard binormal model. We were able to get around the difficulties created by a set of non-identifiable parameters in a clever reparametrization that naturally speeds up the Gibbs sampler (Nandram and Chen 1996).

We note that the novelty in our paper is threefold. First, we have made the binormal model more robust by dealing with the non-identifiablity in a more non-parametric manner still incorporating the stochastic ordering. Second, we have introduced the skew-binormal model into the ROC curve literature and we have included the stochastic ordering as well to extend the binormal model for ROC curve analysis. Third, the computation using the griddy Gibbs sampler is also innovative even with the difficulties involved in computing the cdf's and inverse cdf's for the skew-normal densities.

Our substantive conclusion is the foods actually deteriorate over time. But the deterioration may be slow because over a five-year period the ROC curves are not too far above the 45° line. The AUCs are statistically much larger than 0.50, but they are far from say .95. Shelf-live estimation for these foods are difficult just because the food deterioration is slow (e.g., see Chen, Nandram and Ross 1996).

We have performed simulation study to compare the binormal model and the skew-binormal model. While the AUCs are very similar across the four design points (two separations versus two sample sizes), the skew-binormal model can give improved precision. The similarity in the AUCs is partly due to the crossing of the ROC curves, thereby leading to nearly equal AUCs. It is possible to have much more improved ROC curves under the skew-binormal model by increasing the number of panelists, but this is time-consuming and costly to NLABS.

One important extension of our model can be done. It may be possible to provide a combined analysis of all twelve foods in the food system. Because of the sparseness of the data from each food (responses of 36 panelists on a 9-point hedonic scale), it will be sensible to combine the data from all foods as in small area estimation. Another team of researchers is working on this project but not on the robustness side (using the binormal model). Also, this research will benefit from using a nonparametric Bayesian approach (e.g., a Dirichlet process mixture model; see Inácio, Jara, Hanson, and De Carvalho 2013).

APPENDIX A. CONSTRUCTION OF $\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$

We show how to construct the proposal density, $\pi_a(\underset{\sim}{\tau}|\underset{\sim}{n}_1, \underset{\sim}{n}_2)$.

Let $X_i \sim^{iid} \text{Gamma}(\alpha_i, 1)$, $i = 1, 2, \cdots, K$, then the joint density of $X_1, X_2, \cdots, X_K$ is

$$f(x_1, x_2, \cdots, x_K) = \prod_{i=1}^{K} \left\{ \frac{1}{\Gamma(\alpha_i)} x_i^{\alpha_i - 1} e^{-x_1} \right\}.$$

Let

$$Y_j = \frac{\sum_{i=1}^{j} X_i}{\sum_{i=1}^{K} X_i}, \quad i = 1, 2, \cdots, K, \quad \text{and} \quad T = \sum_{i=1}^{K} X_i.$$

Then

$$\sum_{i=1}^{j} X_i = TY_j \quad \text{and} \quad X_j = T(Y_j - Y_{j-1}), \quad j = 1, 2, \cdots, K,$$

and the jacobian is t^{K-1} when $\underset{\sim}{X}$ is transformed to $\underset{\sim}{Y}$. So the joint density function of $Y_1, Y_2, \cdots, Y_K$ and T is

(A.1)

$$\begin{aligned} &f(y_1, y_2, \cdots, y_K, t) \\ &\quad = t^{K-1} \frac{(ty_1)^{\alpha_1 - 1} e^{-ty_1}}{\Gamma(\alpha_1)} \times \frac{(t(y_2 - y_1))^{\alpha_2 - 1} e^{-t(y_2 - y_1)}}{\Gamma\alpha_2} \times \cdots \\ &\qquad \times \frac{(t(1 - y_{K-1}))^{\alpha_K - 1} e^{-t(1 - y_{K-1})}}{\Gamma\alpha_K} \\ &\quad = \frac{t^{\sum_{i=1}^{K} \alpha_i - 1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} e^{-t} \\ &\qquad \times y_1^{\alpha_1 - 1}(y_2 - y_1)^{\alpha_2 - 1} \cdots (1 - y_{K-1})^{\alpha_1 - 1}. \end{aligned}$$

Therefore, the joint density function of $Y_1, Y_2, \cdots, Y_K$ is

(A.2)

$$\begin{aligned} &f(y_1, y_2, \cdots, y_K) \\ &\quad = \frac{t^{\sum_{i=1}^{K} \alpha_i - 1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \, y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1} \cdots y_{K-1}^{\alpha_{K-1} - 1} \end{aligned}$$

$$\times (1-\frac{y_1}{y_2})^{\alpha_2-1}(1-\frac{y_2}{y_3})^{\alpha_3-1}\cdots(1-\frac{y_{K-1}}{y_{K-2}})^{\alpha_{K-2}-1},$$
$$0<y_1<y_2<\cdots<y_{K-1}<1.$$

So that $\tau_i, i=1,2,\cdots,K-1$, can be drawn from

$$\tau_i \sim Beta\left(\sum_{j=1}^{i}\alpha_j,\ \sum_{j=i+1}^{K}\alpha_j\right),$$

where $\alpha_K=\frac{\sum_{j=1}^{K-1}\alpha_j}{K}$. Note that if we take $\alpha_i=1,\quad i=1,2,\cdots,K$,

$$\tau_i \sim Beta(i,K-i),\qquad \tau_1<\tau_2<\cdots<\tau_{K-1}.$$

That is, $\tau_1,\tau_2,\cdots\tau_{K-1}$, are the order statistics from the $U(0,1)$ distribution.

We can obtain a possibly better proposal density as follows. We now use the iterates $\tau_i^{(h)},\quad i=1,2,\cdots,K,\quad h=1,2,\cdots,M$ is from Gibbs sampling. Take

$$\tau_1^{(h)} \sim Beta(\alpha_1,\ \sum_{i=2}^{K}\alpha_i),$$
$$\tau_2^{(h)} \sim Beta(\alpha_1+\alpha_2,\ \sum_{i=3}^{K}\alpha_i),\ldots,\tau_K^{(h)}\sim Beta(\sum_{i=1}^{K-1}\alpha_i,\ \alpha_K).$$

Denote the means and variances of the τ_i are

$$\hat{\mu}_i=\frac{1}{M}\sum_{h=1}^{M}\tau_i^{(h)}\qquad \hat{\sigma}_i^2=\frac{1}{M-1}\sum_{h=1}^{M}(\tau_i^{(h)}-\hat{\mu}_i)^2.$$

Then, equating moments we get

$$\frac{\alpha_1}{\tau}=\hat{\mu}_1,\quad \frac{\alpha_1-\alpha_2}{\tau}=\hat{\mu}_2,\cdots$$
$$\frac{\sum_{i=1}^{K}\alpha_i^{(h)}}{\tau}=\hat{\mu}_K,\quad \hat{\tau}=\frac{\sum_{i=1}^{K}\hat{\mu}_1(1-\hat{\mu}_i)}{\sum_{i=1}^{K}\hat{\sigma}_i^2}.$$

Finally,

$$\alpha_1=\hat{\tau}\hat{\mu}_1,\quad \alpha_2=\hat{\tau}(\hat{\mu}_2-\hat{\mu}_1),\ldots,\alpha_K=\hat{\tau}(\hat{\mu}_K-\hat{\mu}_{K-1}).$$

So we have two proposal densities, one in which the α_i are all unity and the other is based on the iterates from the Gibbs sampler.

APPENDIX B. MONTE CARLO STANDARD ERROR OF THE MARGINAL LIKELIHOODS

We obtain a general approximation of the Monte Carlo error, also called numerical standard error, of a typical log-marginal likelihood, which includes the weights. Let $T=\sum_{j=1}^{n}w_je^{X_j}$, where the X_j and the w_j are respectively independent and identically distributed. Let $\bar{x}$ and S_x and $\bar{w}$ and S_w denote the respective sample means and standard deviations. The lemma below gives the standard deviation of $\log(T)$ in the special case in which the variation in the weights is small. This is different from the batch-means method that cannot be used here.

Lemma. Assume the variation in the w_j is small, then approximation of estimated standard error of $\log\left\{\sum_{j=1}^{n}w_je^{X_j}\right\}$ is $\frac{S_X}{\sqrt{n}}$.

Proof. Let $T=\sum_{j=1}^{n}w_je^{X_j}$ then using the first order Taylor series approximation, we can show that

$$Var(log(T))=\frac{Var(T)}{(E(T))^2}.$$

Suppose w_j and $X_j,\quad j=1,2,\cdots,n$, are independent and identically distributed with means μ_w and μ_X and variances σ_w^2 and σ_X^2 respectively. Then, using first order Taylor's series expansion

$$E(T)\simeq\sum_{j=1}^{n}\mu_we^{\mu_X}=n\mu_we^{\mu_X}$$

and

$$\begin{aligned}Var(T)&=Var(\sum_{j=1}^{n}w_je^{X_j})\\&=\sum_{j=1}^{n}\left[E(w_j^2e^{2X_j})-(E(w_je^{X_j})^2\right]\\&=\sum_{j=1}^{n}\left[E(w_j^2)E(e^{2X_j})-(E(w_j(E(e^{X_j})^2\right]\\&\simeq\sum_{j=1}^{n}\left[(\mu_w^2+\sigma_w^2)(e^{2\mu_X}+\sigma_X^2e^{2\mu_x})-\mu_w^2e^{2\mu_x}\right]\\&=n\left[\mu_w^2+\sigma_w^2+\frac{\sigma_w^2}{\sigma_X^2}\right]e^{2\mu_x}\sigma_X^2.\end{aligned}$$

It follows that

$$Var(log(T))\simeq\frac{n\left[\mu_w^2+\sigma_w^2+\frac{\sigma_w^2}{\sigma_X^2}\right]e^{2\mu_x}\sigma_X^2}{(n\mu_we^{\mu_X})^2},$$

and the estimated variance is given by

$$\begin{aligned}\widehat{Var}(log(T))&\simeq\frac{n\left[\bar{X}^2+S_w^2+\frac{S_w^2}{S_X^2}\right]e^{2\bar{X}}S_X^2}{(n\bar{w}e^{\bar{X}})^2}\\&=\frac{S_X^2}{n}\left[1+\frac{S_w^2}{\bar{w}^2}\left[1+\frac{1}{S_X^2}\right]\right].\end{aligned}$$

When the coefficient of variation of the w_j is small, $Var(log(T))\simeq\frac{S_X^2}{n}$. So estimated standard error of $\log\left\{\sum_{j=1}^{n}w_je^{X_j}\right\}$ is $\frac{S_X}{\sqrt{n}}$.

ACKNOWLEDGMENTS

The authors are grateful to the two reviewers for their careful reading of the manuscript and their suggestions. Also this work was supported by a grant from the Simons Foundation (353953, Balgobin Nandram).

Received 14 July 2016

REFERENCES

Azzalini, A. (2014), *The Skew-Normal and Related Families,* Cambridge University Press, Cambridge: New York. MR3468021

Bamber, D. (1975), "The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph," *Journal of Mathematical Psychology*, 12(4), 387–415. MR0384214

Bhattacharya, B. and Hughes, G. (2015), "On Shape Properties of Receiver Operator Characteristic Curve," *Statistics and Probability Letters*, 103, 73–79. MR3350864

Bhattacharya, B. and Hughes, G. (2011), "Symmetry of Receiver Operating Characteristic Curves and Kullbach Leibler Divergences Between the signal and Noise Population," *Journal of Mathematical Psychology*, 55, 365–367. MR2838424

Brys, G., Hubert, M., and Struyf, A. (2012), "A Robust Measure of Skewness," *Journal of Computational and Graphical Statistics*, 13(4), 996–1017. MR2109062

Chen, M-H., Nandram, B. and Ross E. W. (1996), "Bayesian Prediction of the Shelf Life of a Military Ration With Sensory Data," *Journal of Agricultural, Biological, and Environmental Statistics*, 1(3), 377–392. MR1807755

Dofman, D. D. and Alf E. Jr. (1969), "Maximum Likelihood Estimation of Parameters of Signal-Detection Theory and Determination of Confidence Intervals-Rating-Method Data," *Journal of Mathematical Psychology*, 6, 487–496.

Faraggi, D. and Reiser, B. (2002), "Estimation of the Area Under the ROC Curve," *Statistics in Medicine*, 21(20), 3093–3106.

Gelfand, A. E. and Kottas, A. (2001), "Nonparametric Bayesian Modeling for Stochastic Order," *Annals of the Institute of Statistical Mathematics*, 53(4), 865–876. MR1880817

Gelfand, A. E. and Kuo, L. (1991), "Nonparametric Bayesian Bioassay Including Ordered Polychotomous Response," *Biometrika*, 78(3), 657–666. MR1130934

Grey, D. R. and Morgan, B. J. T. (1972), "Some Aspects of ROC Curve Fitting: Normal and Logistic Models," *Journal of Mathematical Psychology*, 9, 128–139.

Hanson, T. E., Kottas, A. and Branscum, A. J. (2008), "Modelling Stochastic Order in the Analysis of Receiver Operating Characteristic Data: Bayesian Non-parametric Approaches," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(2), 207–225. MR2420437

Hellmich, M., Abrams, K. R., Jones, D. R. and Lambert, P. C. (1998), "A Bayesian Approach to a General Regression Model for ROC Curves," *Medical Decision Making*, 18, 436–443.

Hughes, G. and Bhattacharya, B. (2013), "Symmetry Properties of the Bi-Normal and Bi-Gamma Receiver Operator Characteristic Curve are Described by Kullbach-Leibler Divergences," *Entropy*, 15, 1342–1356. MR3048289

Inácio, V., Jara, A., Hanson, T. E. and De Carvalho, M. (2013), "Bayesian Nonparametric ROC Regression Modeling," *Bayesian Analysis*, 8(3), 623–646. MR3102228

Ishwaran, H. and Gatsonis, C. A. (2000), "A General Class of Hierarchical Ordinal Regression Models with Applications to Correlated ROC Analysis," *Canadian Journal of Statistics*, 28(4), 731–750. MR1821431

Johnson, T. D. and Johnson, V. E. (1996), "A Bayesian Hierarchical Approach to Multirater Correlated ROC Analysis," *Statistics in Medicine*, 25(11), 1858–1871. MR2227433

Kottas, A., Branco, M. D. and Gelfand, A. E. (2002), "A Nonparametric Bayesian Modeling Approach for Cytogenetic Dosimetry," *Biometrics*, 58(3), 593–600. MR1925555

Kottas, A. and Gelfand, A. E. (2001), "Modeling variability order: A Semiparametric Bayesian Approach," *Methodology and Computing in Applied Probability*, 3(4), 427–442. MR1912186

Metz, C. E. (1989), "Some Practical Issues of Experimental Design and Data Analysis in Radiological ROC Studies," *Investigative radiology*, 24, 234–245.

Metz, C. E., Herman, B. A. and Shen J. H. (1998), "Maximum Likelihood Estimation of Receiver Operating Characteristic (ROC) Curves from Continuously-Distributed Data," *Statistics in Medicine*, 17(9), 1033–1053.

Nandram, B. (2005), "A Bayesian Subset Analysis of Sensory Evaluation Data," *The Journal of Modern Applied Statistical Methods*, 4(2), 482–499.

Nandram, B. (2012), "Bayesian Inference for the Best Ordinal Multinominal Population in a Taste Test," *Case Studies in Bayesian Statistics*, 121, 399–418.

Nandram, B. (1998), "A Bayesian Analysis of the Three-Stage Hierarchical Multinomial Model," *Journal of Statistical Computation and Simulation*, 61(1–2), 97–126. MR1702669

Nandram, B. and Chen, M.-H. (1996), "Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence," *Journal of Statistical Computation and Simulation*, 54, 129–144. MR1700909

Nandram, B. and Kim, H. (2002), "Marginal Likelihood for a Class of Bayesian Generalized Linear Models," *Journal of Statistical Computation and Simulation*, 72(4), 319–340. MR1923106

Nandram, B. and Yin, J. (2016), "A Nonparametric Bayesian Prediction Interval for a Finite Population Mean," *Journal of Statistical Computation and Simulation*, 86, 1–17. MR3534572

O'Malley, A. J. and Zou, K. H. (2006), "Bayesian Multivariate Hierarchical Transformation Models for ROC Analysis," *Statistics in Medicine*, 25(3), 459–479. MR2225177

Peng, F. and Hall, W. J. (1996), "Bayesian Analysis of ROC Curves Using Markov-chain Monte Carlo Methods," *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 16(4), 404–411.

Ritter, C. and Tanner, M. A. (1992), "Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler," *Journal of the American Statistical Association*, 87(419), 861–868.

Tosteson, M. T., Pinto, L. H., Holsinger, L. J., and Lamb, R. A. (1994), "Reconstitution of the Influenza Virus M2 Ion Channel in Lipid Bilayers," *The Journal of Membrane Biology*, 142(1), 117–126.

US Army Rdecom (2014), *US Army Natick Soldier Research, Developing & Engineering Center Testing Facilities and Equipment*, 3^{rd} Edition, May 2014, pg. 1–46.

Balgobin Nandram
Department of Mathematical Sciences
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
USA
E-mail address: balnan@wpi.edu

Thelge Buddika Peiris
Department of Mathematical Sciences
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
USA
E-mail address: tbpeiris@wpi.edu