Journal of Modern Applied Statistical Methods

Volume 4 | Issue 2 Article 13

11-1-2005

A Bayesian Subset Analysis Of Sensory Evaluation Data

Balgobin Nandram
Worcester Polytechnic Institute, balnan@wpi.edu

Journal of Modern Applied Statistical Methods
November, 2005, Vol. 4, No. 2, 482-499

1538 – 9472/05/$95.00

A Bayesian Subset Analysis Of Sensory Evaluation Data

Balgobin Nandram
Department of Mathematical Sciences
Worcester Polytechnic Institute

In social sciences it is easy to carry out sensory experiments using say a *J*-point hedonic scale. One major problem with the *J*-point hedonic scale is that a conversion from the category scales to numeric scores might not be sensible because the panelists generally view increments on the hedonic scale as psychologically unequal. In the current problem several products are rated by a set of panelists on the *J*-point hedonic scale. One objective is to select the best subset of products and to assess the quality of the products by estimating the mean and standard deviation response for the selected products. A priori information about which subset is the best is incorporated, and a stochastic ordering is modified to select the best subset of the products. The method introduced in this article is sampling based, and it uses Monte Carlo integration with rejection sampling. The methodology is applied to select the best set of entrees in a military ration, and then to estimate the probability of at least a neutral response for the judged best entrees. A comparison is made with the method, which converts the category scales to numeric scores.

Key words: Bayes factor; composition method; stochastic ordering; rejection sampling.

Introduction

Consider the problem of selecting the best subset of a number of multinomial populations with ordinal categories. This can be accomplished by first converting the nominal data to numeric scores, and then a standard multiple comparison procedure can be performed on these scores. However, this procedure can go badly wrong when the conversion is made. It is, therefore, the purpose of this article to describe a straightforward method based on a stochastic ordering of the multinomial populations for selecting the best subset of populations and then to estimate parameters used to assess the quality of the best subset without conversion of the nominal data. A Bayesian approach is preferred because it is natural to incorporate a priori information about which subset is the best.

In sensory evaluation of food acceptability, judges are asked to rate several products on the 9-point scale with qualitative responses ranging from “dislike extremely” to “neither like nor dislike” to “like extremely” on an ordinal scale. Usually in the analysis these nominal values are converted to scores ranging from 1 to 9 where an attempt is made to associate “dislike extremely” with 1, “neither like nor dislike” with 5, “like extremely” with 9, and intermediate nominal values have graduated meanings. The use of scores has several disadvantages, which weaken the interpretation that can be placed on the analysis of sensory evaluation data.

First, the intervals between categories are psychologically unequal. Second, judges tend to avoid the use of extreme categories by grouping judgments into the center of the scale, and sometimes avoiding even “neither like nor dislike” response. Third, scale values have no numerical relationship. Thus, it is difficult to make conclusions concerning ratios of acceptability of the food products when

Balgobin Nandram is a Professor of Statistics, and a fellow of the American Statistical Association. His research interests are in survey methodology, Bayesian statistics, categorical data analysis, computational statistics and simulation, health, industrial and environmental statistics, and statistical education. Email him at balnan@WPI.EDU.

qualitative responses are converted to quantitative responses.

Newel (1982) applied the method of McCullagh (1980) to analyze sensory data and was able to overcome some of the advantages in using scores. This method for ordinal data treats the response categories as contiguous intervals on a continuous scale with unknown cutpoints $\theta_1, \ldots, \theta_{J-1}$, where for the *J*-point scale $J = 9$. Inherent in these models is the stochastic ordering with the use of scores unnecessary. Let π_{ij} denote the probability of the j^{th} response in the i^{th} population, and $\gamma_{ij} = \sum_{s=1}^{j} p_{is}$ be the cumulative probability of the i^{th} population. Then Newel (1982) entertained a model of the form

$$\log\{\gamma_{ij}/(1-\gamma_{ij})\} = (\theta_j - \beta_i)/\tau_i,\ i = 1, \ldots, I,$$
$$j = 1, \ldots,\ J-1,$$

where β_i and τ_i are relative measures of location and spread respectively of the i^{th} population. This model incorporates the location of the ratings and the consistency of the panelists' responses directly.

Such a model is usually fitted using nonlinear iteratively reweighted least squares; see, for example, Green (1985). While this is an attractive model, besides the cell probabilities, it introduces $2I + J$ new parameters. Moreover, while one can choose the best population as the one with the largest β_i, and perhaps the smallest τ_i, this modeling does not address the problem of selecting the best population directly, and in fact, it is difficult to assess the uncertainty in selecting the best population. Also as the analysis relies heavily on asymptotic theory, with sparse data this approach will provide poor estimates for the cutpoints θ_j, and hence the other parameters. A more appropriate method is associated with ranking and selection.

Recent Bayesian work on selection and ranking includes the approach of Morris and Christiansen (1996). They used a simple two-level Bayes empirical Bayes model to select the best mean. They generated samples from the product normal posterior distribution of the means, and obtained posterior probabilities that each of the means is the largest. Goldstein and Spiegelhalter (1996) described statistical issues in ranking institutions in the areas of health and education based on outcome data by using certain performance indicators. They obtained interval estimates of the ranks of these indicators for the different institutions, using both Bayesian and non-Bayesian methods. Similar to Morris and Christiansen (1996), Goldstein and Spiegelhalter (1996) did not incorporate uncertainty directly about the ranks of the performance indicators. Moreover, these authors did not consider the ranking of several multinomial populations nor did they consider sensory evaluation data. However, the sampling-based approach of these authors is closest in spirit to the work in this article.

In fact, Nandram (1997) obtained the best multinomial population (not best subset) among a set of populations, converting the nominal data on the hedonic scale to numeric scores. A number of independent nonidentical multinomial populations with the same ordinal categories are considered. This approach is different from that in the ranking and selection literature because it incorporates the prior belief about which population is the best by assigning a nonzero probability to the event that any population could be the best population (Nandram, 1997). The simple tree order (see Robertson, Wright and Dykstra, 1988) is used to obtain the most probable population under a variation of the stochastic ordering. Consider two discrete random variables, *P* and *Q*, which take the same values a_j (increasing in *j*) with probabilities p_j and q_j respectively, $j = 1, \ldots, J - 1$, where

$$\sum_{j=1}^{J} p_i = \sum_{j=1}^{J} q_i = 1.$$

then

$$P \overset{st}{\geq} Q$$

if, and only if,

$$\sum_{i=1}^{s} p_i \le \sum_{i=1}^{s} q_i, \quad s = 1, \ldots, J-1. \tag{1}$$

This is the situation for two multinomial populations which are stochastically ordered (*P* stochastically greater than *Q*) with the same ordered categories; see, for example, Sampson and Whitaker (1989). This stochastic ordering is modified to obtain a criterion which will be used to select the best population or best subset of populations without using the values a_j on the ordinal scale.

The Bayesian analysis is pertinent as there is useful information about which is the best product. In the non-Bayesian approach, it is difficult to express uncertainty about which population is the best. Moreover, as the non-Bayesian methods do not express uncertainty about the best population, estimation after selection becomes a delicate and tricky issue. In the Bayesian method the parameters can be estimated in a straightforward manner by mixing with appropriate weights (posterior probabilities); see Nandram (1997).

The objective is to select the best population (or subset) among a number of multinomial populations, whose cell counts arise from sensory evaluation, and to show how to estimate the parameters of the selected population. The method is sampling based, and it uses Monte Carlo integration which is accommodated by rejection sampling. A methodology is described, and it is shown how to compute efficiently the relevant quantities. Next, the sensory data obtained from the Natick food experiment is described and the methodology is applied to select the best entree. Finally, there are conclusions.

Methodology

The objective is to develop a method to judge the best multinomial population or the best subset of multinomial populations without converting the ordinal categories to numeric scores by modifying the stochastic ordering. Estimation is performed to make inference about the quality of product. In general, it is assumed that there are *I* multinomial populations, and the best subset of size $\ell < I$ subsets is to be selected. There are $T = I!/\ell!(I-\ell)!$ distinct subsets of size ℓ which are denoted by I_t, $t = 1 \ldots, T$. For example, with $I = 3$, $\ell = 2$, the set of all products is {1, 2, 3}, $T = 3$, and the subsets are I_1 = {1, 2}, I_2 = {1, 3} and I_3 = {2, 3}. The primary objective is to select the best subset among the I_t.

Model

I multinomial populations with *J* categories are considered. For the i^{th} population, the counts, denoted by $\underset{\sim}{n}_i = (n_{i1}, \ldots, n_{iJ})'$, $i = 1, \ldots, I$, are taken. In many applications it is reasonable to assume that the $\underset{\sim}{n}_i$ have independent multinomial distributions with probabilities $\underset{\sim}{p}_i = (p_{i1}, \ldots, p_{iJ})'$, $\sum_{j=1}^{J} p_{ij} = 1$.

Letting

$$\underset{\sim}{p} = (\underset{\sim}{p}'_1, \ldots, \underset{\sim}{p}'_I)',$$

the joint likelihood is

$$l(\underset{\sim}{p} \mid \underset{\sim}{n}) \propto \prod_{i=1}^{I} \prod_{j=1}^{J} p_{ij}^{n_{ij}}. \tag{2}$$

A priori, without any order restriction on the p_{ij}, we take independent Dirichlet distributions for the p_i,

$$\pi(\underset{\sim}{p}) = \prod_{i=1}^{I} \frac{\prod_{j=1}^{J} p_{ij}^{\alpha_{ij}-1}}{D(\underset{\sim}{\alpha}_i)}, \tag{3}$$

where the $\underset{\sim}{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{iJ})'$ and α_{ij} are fixed quantities to be specified. Note that in (3) $D(\underset{\sim}{a}) = \{\prod_{j=1}^{J} \Gamma(a_j)\}\{\Gamma(\sum_{j=1}^{J} a_j)\}^{-1}$ and $\Gamma(\cdot)$ is the gamma function. In (3), $\alpha_{ij} = ½$ is taken for three reasons. First, it is difficult to elicit information about α_{ij} even though they can be interpreted as cell counts in a prior sensory evaluation. Second, one does not want to model

similarity among the different products as it is believed that a priori some of them are better than others. Third, it simplifies the computation a lot if the α_{ij} are taken known, rather than if an assumption is made about their distributions a priori. Thus, to ensure the maximum heterogeneity (no preference) Jeffreys' reference prior is used (i.e., $\alpha_{ij} = ½$), a proper density in this application. In classical statistics, this is equivalent to adding a ½ to the cell counts; a recommendation usually made for sparse categorical tables. Rather, prior information will be inputted through the belief about which is the best product.

Criteria for Selection

One criterion that can be used is based on the random variable X_i representing values on the hedonic scale. That is, letting a_j denote the values on the ordinal scale,

$$Pr(X_i = a_j \Big| \underset{\sim}{p}_i) = p_{ij}, \quad j = 1,\ldots,J, \quad i = 1,\ldots,I$$

and the mean of X_i is denoted by $\mu_i = \sum_{j=1}^{J} a_j p_{ij}$.

First, to introduce the general criterion, suppose a single population is selected; let *b* denote the selected population. The best (selected) population is defined as the one for which

$$\mu_b \geq \max\{\mu_i, i = 1,\ldots,I\}. \qquad (4)$$

That is, the population with the largest mean is selected. Thus, the best population is defined by using the simple tree order; see Robertson, Wright and Dykstra (1988). Such an order restriction arises naturally in many situations. For example, if an investigator wishes to compare several treatments with a new one, the prior information that the new treatment mean is at least as large as the others might be entertained. Because of its simplicity, (4) is popular.

Nandram (1997) used criteria based on the mean, standard deviation and coefficient of variation of the X_i to obtain the best multinomial population (not best subset) among a set of populations. However, he used the scores on the hedonic scale to construct these criteria.

For subset selection, let I_b denote the set containing the ℓ best populations. (Note that I_b is a proper nonempty subset of the set of I products.) Then, based on the means, the (best) selected set of populations I_b is defined as the one for which

$$\min\{\mu_i; i \in I_b\} \geq \max\{\mu_i; i \notin I_b\}. \qquad (5)$$

Note that (4) is a special case of (5), and (5) can be viewed as an extension of the simple tree order.

Unfortunately, the method of subset selection based on the mean, uses the category scales. The a_j are almost always unknown and are usually taken as $a_j = j$, $j = 1,\ldots, J$. The thesis is that this is inaccurate, and an alternative solution based on a modification of the stochastic ordering is sought. However, the method of subset selection based on the mean will be used for comparison with the method which does not use the category scales.

A single criterion based on a version of the stochastic ordering is obtained, but first, an explanation for why the stochastic ordering cannot be used directly is provided. For simplicity, consider selecting the best population. Let $A_{is} = \{p : \sum_{j=1}^{s} p_{ij} \leq \max(\sum_{j=1}^{s} p_{tj}, t = 1,\ldots,I, t \neq i)\}$, $s = 1,\ldots,J-1$, and $S_i = \bigcap_{j=1}^{J-1} A_{ij}$. Then for each j the A_{ij} are mutually exclusive, $\sum_{i=1}^{I} P(A_{ij}) = 1$, and $P(S_i) \leq \min\{P(A_{ij}), j = 1,\ldots,J-1\}$. As the $P(A_{ij})$ are different for each *i*, for some choice of *s* and some *i*, $P(A_{is}) > \min\{P(A_{ij}), j = 1,\ldots,J-1\}$.

Then, $\sum_{i=1}^{I} P(S_i) < \sum_{i=1}^{I} P(A_{is}) = 1$. That is, while the S_i are mutually exclusive, they are not exhaustive. In fact, $P(S_i)$ is not the probability that the i^{th} population is the best; the $P(S_i)$ could be extremely small and $\sum_{i=1}^{I} P(S_i) << 1$. Thus, for each j $\{A_{ij}, I = 1,\ldots,I\}$ will be used as a partition to identify the best population or subset.

Letting

$$\Delta_{ik} = \sum_{j=k}^{J} p_{ij}, k = 2,\ldots,J,\ i = 1,\ldots,I, \qquad (6)$$

these Δ_{ik} are measures of the quality of the i^{th} product. Note that Δ_{ik} is the probability of getting at least response k on the ordinal scale (e.g., $\Delta_{i,\ \frac{J+1}{2}}$ is the probability of getting at least a neutral response). To express uncertainty about the best subset of populations, let B denote the random variable indicating the best population and κ denote exclusively the measure of quality which is used. Let $A_{tk} = \{p : \min\{\Delta_{ik}, i \in I_t\} \geq \max\ \{\Delta_{ik}, i \notin I_t\},\ \ t = 1,...,T, k = 2,...,J$, and $S_{t2} = A_{t2} = S_{tk} = A_{ts} - \bigcup_{j=2}^{s-1} A_{tj},\quad s = 3,\ldots,J$. Then, $\kappa = k$ if $\underset{\sim}{p} \in A_{tk},\ \ k = 2,\ldots,J$ is defined (However, note that κ is a nuisance parameter.). The criterion based on S_{bk} is defined as the modified stochastic ordering (MSO) criterion. Then,

$$Pr(B = b, \kappa = k) = \omega_{bk},\ b = 1,\ldots,T,$$
$$k = 2,\ldots,J, \sum_{b=1}^{I}\sum_{k=1}^{J} \omega_{bk} = 1, \qquad (7)$$

where the ω_{bk} are to be specified. Letting $\lambda_b = \sum_{k=1}^{J-1} \omega_{bk}$, a priori the best population is the b^{th} population for which $\lambda_b = \max\{\lambda_t, t = 1,\ldots,T\}$. The λ_b are to be updated using the data.

Incorporating prior information about which is the best entree through the ω_{bk} rather than the α_{ij} is preferred. It should be noted that it is conceptually simple and convenient to use the random variables B and κ to model uncertainty about which is the best entree. On the other hand, it is much more difficult to add information about which is the best entree through the α_{ij}. However, unless the α_{ij} are all equal, their specification will give latent information about which is the best entree, but this information is difficult to discern.

In addition, if there is a reluctance to specify the α_{ij}, then in the Bayesian paradigm they are random variables, and the problem of selection and estimation becomes extremely difficult, especially if one wants to incorporate uncertainty about which is the best population.

For the criterion given by (5) based on the mean, $k = 1$ will be taken and define $S_{b1} = \{p : \min\{\mu_i, i \notin I_b\} \geq \max\ (\mu_i, i \notin I_b\}, b = 1,\ldots,T$. The criterion based on S_{b1} will be called the mean response ordering (MRO) criterion.

Then the prior distribution on $\underset{\sim}{p}$ in (3) becomes

$$\pi\left(\underset{\sim}{p} \middle| B = b, \kappa\right) = \begin{cases} c_{bk}(\underset{\sim}{\alpha}) \prod_{i=1}^{I} \dfrac{\prod_{j=1}^{J} p_{ij}^{\alpha_{ij}-1}}{D(\underset{\sim}{\alpha_i})}, & \underset{\sim}{p} \in S_{bk}, \\ 0, & \textit{otherwise}, \end{cases} \qquad (8)$$

where

$$\underset{\sim}{\alpha} = (\underset{\sim}{\alpha}'_1, \ldots, \underset{\sim}{\alpha}'_I)'$$

and

$$c_{bk}(\underset{\sim}{\alpha})^{-1} = \int_{S_{bk}} \prod_{i=1}^{I} \frac{\prod_{j=1}^{J} p_{ij}^{\alpha_{ij}-1}}{D(\underset{\sim}{\alpha_i})} d\underset{\sim}{p},\ b = 1,\ldots,I,\ k = 2,\ldots,J.$$

Note that

$$c_{bk}(\underset{\sim}{\alpha})^{-1} = \Pr(\underset{\sim}{p} \in S_{bk}), b = 1,\ldots,T,\ k = 2,\ldots,J.$$

These quantities are to be updated by the data, and are to be used to update the ω_{bk} which, in turn, are to be used to judge the best product or set of products.

Bayesian Selection and Estimation

Now, it is shown how to use the data to judge the best subset, and then to make inference about the best set of populations.

Let

$$n'_{ij} = n_{ij} + \alpha_{ij},\ \underset{\sim}{n}'_i = (n'_{i1},\ n'_{i2},\ \ldots,\ n'_{iJ})'$$

and

$$\underset{\sim}{n}' = \{n'_{ij} : i = 1,\ldots,I;\ j = 1,\ldots,J)\}\ .$$

Using Bayes' theorem, the joint posterior distribution of $\underset{\sim}{p}$, B and κ is

$$\begin{aligned} &f(\underset{\sim}{p}, B=b, \kappa=k \mid \underset{\sim}{n}) \\ &= f(\underset{\sim}{p} \mid \underset{\sim}{n}, B=b, \kappa=k) P(B=b, \kappa=k \mid \underset{\sim}{n}) \end{aligned} \tag{9}$$

where

$$f(\underset{\sim}{p} \mid \underset{\sim}{n},\ B=b,\ \kappa=k)$$

and

$$P(B=b,\ \kappa=k \mid \underset{\sim}{n})$$

are to be described. First,

$$f\left(\underset{\sim}{p} \mid \underset{\sim}{n}, B=b,\ \kappa=k\right) = \begin{cases} c_{bk}(\underset{\sim}{\alpha}) \prod_{i=1}^{I} \dfrac{\prod_{j=1}^{J} p_{ij}^{n'_{ij}-1}}{D(\underset{\sim}{n}'_i)}, & \underset{\sim}{p} \in S_{bk} \\ 0, & \textit{otherwise,} \end{cases} \tag{10}$$

where

$$c_{bk}(\underset{\sim}{n}')^{-1} = \int_{S_{bk}} \prod_{i=1}^{I} \frac{\prod_{j=1}^{J} p_{ij}^{n'_{ij}-1}}{D(\underset{\sim}{n}'_i)}\ d\underset{\sim}{p},$$

b=1,…,T, k=2,…,J.

For convenience, letting $\bar{S}_{bk}$ be the complement of S_{bk},

The following is defined as,

$$\begin{aligned} &\bar{c}_{bk}(\underset{\sim}{n}')^{-1} \\ &= 1 - c_{bk}(\underset{\sim}{n}')^{-1} \int_{\bar{S}_{bk}} \prod_{i=1}^{I} \frac{\prod_{j=1}^{J} p_{ij}^{n'_{ij}-1}}{D(\underset{\sim}{n}'_i)}\ d\underset{\sim}{p}. \end{aligned}$$

Second, letting

$$r_{bk}(\underset{\sim}{n}') = c_{bk}(\underset{\sim}{\alpha}) c_{bk}(\underset{\sim}{n}')^{-1},$$

$$b = 1,\ldots,T,\ k = 2,\ldots,J,$$

$$\begin{aligned} &(B=b,\ \kappa=k \mid \underset{\sim}{n}) = \hat{\omega}_{bk} = \omega_{bk} r_{bk}(\underset{\sim}{n}') \\ &Pr\ \left\{ \sum_{t=1}^{T} \sum_{j=2}^{J} \omega_{tj} r_{tj}(\underset{\sim}{n}') \right\}^{-1}. \end{aligned} \tag{11}$$

Letting

$$\hat{\lambda}_b = \sum_{j=2}^{J} \hat{\omega}_{bj}, \tag{12}$$

in (11), a posteriori the best subset is the b^{th} subset for which $\hat{\lambda}_b = \max(\hat{\lambda}_t, t = 1,\ldots,T)$.

Consider testing H_o: b^{th} subset is the best versus h_1: b^{th} subset is not the best where $Pr(H_0) = \lambda_b = 1 - Pr(H_1)$. Then the Bayes factor, B_f, for testing H_0 versus H_1 is

$$B_f = \{\hat{\lambda}_b / (1-\hat{\lambda}_b)\} \{\lambda_b / (1-\lambda_b)\}^{-1}.$$

Letting

$$c_b^*(\underset{\sim}{\alpha})^{-1} = \sum_{j=2}^{J} c_{bj}(\underset{\sim}{\alpha})^{-1}$$

and

$$c_b^*(\underset{\sim}{n}')^{-1} = \sum_{j=2}^{J} c_{bj}(\underset{\sim}{n}')^{-1},$$

it follows easily from (11) and (12) that the Bayes factor is also given by

$$B_f = \{c_b^*(\underset{\sim}{\alpha})c_b^*(\underset{\sim}{n}')^{-1} - c_b^*(\underset{\sim}{n}')^{-1}\}\{1 - c_b^*(\underset{\sim}{n}')^{-1}\}^{-1}$$

$$\approx c_b^*(\underset{\sim}{\alpha})c_b^*(\underset{\sim}{n}')^{-1} \approx$$

$$(J-1)^{-1}\sum_{j=2}^{J} c_{bj}(\underset{\sim}{\alpha})c_{bj}(\underset{\sim}{n}')^{-1} = \bar{r}_b(\underset{\sim}{n}') \ . \qquad (13)$$

In (13) the first approximation follows because in many examples $c_b^*(\underset{\sim}{n}') >> 1$. This is true when there is a large number of subsets as in our application. Also in (13) the second approximation follows if the $c_{bk}(\underset{\sim}{\alpha})$ are approximately constant which is the case with a uniform prior on B and κ. Note that $\bar{r}_b(\underset{\sim}{n}')$ is the average of the $\bar{r}_{bk}(\underset{\sim}{n}')$ in (11). Thus, it is interesting to observe that one might interpret $\bar{r}_b(\underset{\sim}{n}')$ as the Bayes factor, which, in turn, can be interpreted as the odds for H_0 provided by the data. For a review of the literature on the Bayes factor and its interpretation see Kass and Raftery (1995).

Inference proceeds by first picking with uncertainty the best subset (i.e., the subset with the largest $\hat{\lambda}_b$). Whether the frequentist method or the Bayesian method is used, the statistician will be uncertain about which is the best subset of populations. However, in the Bayesian method, as presented here the statistician can incorporate uncertainty about the best population, and this is attractive because by (11) the uncertainty about the best population a posteriori can be quantified. In addition, a posteriori inference about the parameters of the judged best population is obtained by using the posterior distribution

$$\pi\left(\underset{\sim}{p}_b \middle| \underset{\sim}{n}\right) = \sum_{t=1}^{T} \hat{\lambda}_t \pi(\underset{\sim}{p}_b \middle| \underset{\sim}{n},\ B = t) \ . \qquad (14)$$

The elegance in the current approach is contained in (14), as the weakness in the classical approach, is that after the best population is obtained the methods usually proceed as though it is known with certainty which is the best population.

The expression in (14) can be simplified. For

$$\pi\left(\underset{\sim}{p}_b \middle| \underset{\sim}{n}\right) = \hat{\lambda}_b \pi(\underset{\sim}{p}_b \middle| \underset{\sim}{n},\ \underset{\sim}{p} \in S_b) + (1-\hat{\lambda}_b)\pi(\underset{\sim}{p}_b \middle| \underset{\sim}{n},\ \underset{\sim}{p} \in \bar{S}_b) \ , \qquad (15)$$

where

$$\pi(\underset{\sim}{p}_b \middle| \underset{\sim}{n}, \underset{\sim}{p} \in \bar{S}_b) = \begin{cases} c_b^*(\underset{\sim}{n}')\prod_{i=1}^{I} \dfrac{\prod_{j=1}^{J} p_{ij}^{n_{ij}'-1}}{D(\underset{\sim}{n}_i')}, & \underset{\sim}{p} \in \bar{S}_b \\ 0, & \text{otherwise}, \end{cases}$$

$S_b = \bigcup_{K=2}^{J} S_{bk}$, $\bar{S}_b$ is the component of S_b, and $\bar{c}_b^*(\underset{\sim}{n}')^{-1} = 1 - c_b^*(\underset{\sim}{n}')^{-1}$.

When the criterion based on the mean is used, the following is taken

$$\beta_i = \sum_{j=1}^{J} j p_{ij}$$

and

$$\tau_i = \left\{\sum_{j=1}^{J} p_{ij}(j-\beta_i)^2\right\}^{1/2}, \ i = 1, \ldots, I.$$

When the criterion based on the modified stochastic ordering is used, the following is taken

$$\ln\{\gamma_{ij}/(1-\gamma_{ij})\} = (\theta_j - \beta_i)/\tau_i, \ i = 1, \ldots, I, \ j = 1, \ldots, J-1,$$

where

$$\gamma_{ij} = \sum_{s=1}^{j} p_{is} \text{ and } \theta_1 < \theta_2 < \ldots < \theta_{J-1}$$

are the unknown cutpoints. A posteriori inference about β_i and τ_i can be obtained by using (15). Inference is made about the population means β_i and standard deviations τ_i, $i=1,\ldots,I$.

Computations

In this section, a description of how to compute $\hat{\lambda}_b$ in (12) and $\pi\left(p_b \middle| n\right)$ in (15) is provided.

First, consider $\hat{\lambda}_b$. Although it is more accurate to compute $\bar{r}_{bk}(n')$ directly rather than first computing $c_{bk}(\alpha)$ and $c_{bk}(n')$ separately, a simple method is proposed which first obtains $c_{bk}(\alpha)$ and $c_{bk}(n')$. How to obtain $c_{bk}(n')$, or $\bar{c}_{bk}(n')$ is described, for which the simple method suggested by Nandram, Sedransk and Smith (1997) is used. The problem of estimating $\bar{r}_{bk}(n')$ directly is a special case of the more general problem associated with estimating the ratio of two normalization constants; see, for example, Meng and Wong (1996) and Chen and Shao (1997) who used Markov chain Monte Carlo methods. (These refinements are unnecessary in this application.) Denoting the joint unrestricted posterior distribution of p by

$$f''\left(p \middle| n\right),$$

therefore,

$$f''(p|n)=\begin{cases}\prod_{i=1}^{I}\dfrac{\prod_{j=1}^{J}p_{ij}^{n'_{ij}-1}}{D(n'_i)} & 0\le p_{ij}\le 1,\ \sum_{j=1}^{J}p_{ij}=1\\ 0, & \textit{otherwise}.\end{cases} \tag{16}$$

N independent multivariate samples are selected from the unrestricted product Dirichlet distributions with parameters $n'_i, i=1,...,I$ in (16), and find the number $\frac{N_{n'_i}}{N}$ falling inside S_{bk}. (Note that $\bar{c}_{bk}(n')^{-1}$ is estimated by the proportion $1-T_{n'}$, falling outside S_{bk}.) The Monte Carlo sample size, *N*, is obtained by taking, for example,

$$Pr\left\{\left|c_{bk}(n')T_{n'}-1\right|<.01\right\}=0.95. \tag{17}$$

For the examples discussed, *N*=10,000 is taken. The computations for $c_{bk}(n')^{-1}$ or $\bar{c}_{bk}(n')^{-1}$ are performed for whichever requires smaller Monte Carlo sample size in (17). Estimates of the $c_{bk}(\alpha)$ are obtained in a similar manner. But note that with a uniform prior on *B* and κ, it is unnecessary to compute $c_{bk}(\alpha)$ since they are all equal. Otherwise, $r_{bk}(n')$ are obtained by monitoring the estimates of the ratios of $c_{bk}(\alpha)$ and $c_{bk}(n')$ for convergence. Again 10,000 iterates suffice.

Samples from the posterior distribution of p_b, $\pi\left(p_b \middle| n\right)$ in (15), can be obtained by using the composition method (Tanner 1993). First, draw a uniform random variate, $U \sim U(0,1)$. Then if $U \le \hat{\omega}_{bk}$, draw p_b from $\pi\left(p_b \middle| n, p\in S_{bk}\right)$; otherwise draw p_b from $\pi\left(p_b \middle| n, p\in \bar{S}_{bk}\right)$. Samples of p_b from $\pi\left(p_b \middle| n, p\in S_{bk}\right)$ can be obtained simply by drawing p_b from $f''\left(p \middle| n\right)$ and then if $p\in S_{bk}$, accept it. Similarly, samples from $\pi\left(p_b \middle| n, p\in \bar{S}_{bk}\right)$ are obtained by simply drawing p from $f''\left(p \middle| n\right)$ and then if $p\in \bar{S}_{bk}$, accept it. However, it is still possible to obtain

samples from $\pi''\left(p_b \middle| n\right)$ more efficiently.

It is not difficult to show that $\lambda_b c_b^*(\alpha) \neq 1$, then $\hat{\lambda}_b c_b^*(n') < 1$ if and only if $(1-\hat{\lambda}_b)\bar{c}_b^*(n') > 1$. Also, it is not difficult to show that if $\hat{\lambda}_b c_b^*(n') < 1$, then

$$\pi\left(p_b \middle| n\right) = \hat{\lambda}_b c_b^*(n')\, f''\left(p_b \middle| n\right) + (1-\hat{\lambda}_b c_b^*(n'))\pi\left(p_b \middle| n\right),\ p \in \bar{S}_b) \tag{18}$$

and if

$$(1-\hat{\lambda}_b)\bar{c}_b^*(n') < 1,$$

then

$$\pi\left(p_b \middle| n\right) = (1-\hat{\lambda}_b)\bar{c}_b^*(n')\, f''\left(p_b \middle| n\right) + (1-(1-\hat{\lambda}_b)\bar{c}_b^*(n'))\pi\left(p_b \middle| n\right),\ p \in S_b). \tag{19}$$

Note that $f''(p_n \mid n)$ is obtained by marginalization of the posterior distribution $f''(p \mid n)$, in (16). Related arguments are given by Bhattacharya and Nandram (1996). Note that the application $\hat{\lambda}_b$ could be very small and $c_b^*(n')^{-1}$ very close to 1, so that it is very likely that (18) is the choice.

Thus, samples from the posterior distribution $\pi\left(p_b \middle| n\right)$ can be obtained by using the composition method in either (15), (18) or (19). Notice that it is really simple to draw from $f''\left(p_b \middle| n\right)$. In practice, if $\hat{\lambda}_b c_b^*(n')$ is large but less than 1, draws can be made easily from (18), or if $(1-\hat{\lambda}_b)\bar{c}_b^*(n')$ is large but less than 1, draws can be made easily from (19). In the event that $\hat{\lambda}_b c_b^*(n')$ and $\bar{c}_b^*(n')^{-1}$ are small, or $(1-\hat{\lambda}_b)\bar{c}_b^*(n')$ and $c_b^*(n')^{-1}$ are small, one can draw efficiently from (15).

Posterior inference of any function of p_b (e.g., Δ_{bk}) can be obtained by using samples from $\pi''\left(p_b \middle| n\right)$ in a straightforward manner. Noting that p is first drawn from $\pi''\left(p_b \middle| n\right)$, and the components p_b are stripped off, one can take $\hat{p}_b^{(h)}$, $h=1,\ldots,M$ to be M vectors drawn from $\pi''\left(p_b \middle| n\right)$ and $\Delta_{bk}^{(h)} = \sum_{j=k}^{J} p_{bj}^{(h)}$, $h=1,\ldots,M$. Then $E\left(\Delta_{bk} \middle| n\right)$ is estimated by $\bar{\Delta}_{bk} = M^{-1}\sum_{h=1}^{M}\Delta_{bk}^{(h)}$ and $\mathrm{var}\left(\Delta_{bk} \middle| n\right)$ is estimated by $\bar{\Delta}_{bk} = (M-1)^{-1}\sum_{h=1}^{M}(\Delta_{bk}^{(h)} - \bar{\Delta}_{bk})^2$. Note that in these estimation procedures independent samples are used, not dependent samples as in Markov chain Monte Carlo methods.

To make inference about β_i and τ_i a random sample $p^{(1)},\ldots,p^{(M)}$ is first obtained from $\pi''\left(p_b \middle| n\right)$. Then using the criterion based on the mean, the following is computed

$$\beta_i^{(h)} = \sum_{j=1}^{J} j p_{ij}^{(h)} \quad \text{and} \quad \tau_i^{(h)} = \left\{\sum_{j=1}^{J} p_{ij}^{(h)} (j - \beta_i^{(h)})^2\right\}^{1/2},$$
$$i = 1,\ldots,I,\ h = 1,\ldots,M.$$

For the criterion based on the modified stochastic ordering, nonlinear least squares minimizing is used

$$\sum_{i=1}^{I}\sum_{j=1}^{J-1}\left\{\ln\{\gamma_{ij}^{(h)}/(1-\gamma_{ij}^{(h)})\} - (\theta_j^{(h)} - \beta_i^{(h)})/\tau_i^{(h)}\right\}^2$$

to obtain $\theta_j^{(h)}, \beta_i^{(h)}$ and $\tau_i^{(h)}$, *h=1,…,M;* see appendix A for the appropriate equations. (The iterative procedure converges quickly in less than 5 steps.) Then a posteriori we take

$$\hat{\beta}_i = M^{-1} \sum_{h=1}^{M} \beta_i^{(h)}$$

and

$$\hat{\tau}_i = M^{-1} \sum_{h=1}^{M} \tau_i^{(h)}$$

with corresponding standard deviation given by

$$\left\{ (M-1)^{-1} \sum_{h=1}^{M} \left(\beta_i^{(h)} - \hat{\beta}_i \right)^2 \right\}^{1/2}$$

and

$$\left\{ (M-1)^{-1} \sum_{h=1}^{M} \left(\tau_i^{(h)} - \hat{\tau}_i \right)^2 \right\}^{1/2} .$$

Analysis of the Military Data

In this section, the methodology is applied to the Natick Food Experiment. The Meal, Ready-To-Eat (MRE) has twelve meals (menus), each consisting of four to six food items. The system contains 39 distinct foods. Some of these items occur in more than one meal and are regarded as different items in different meals, so the total number of items studied is 52. These items can be classified into five principal types: entrees, pastries, vegetables, fruits and miscellaneous. Chen, Nandram and Ross (1996) analyzed these data to predict shelf lives of the entrees, and they classified the entrees according to whether their shelf lives are short, medium or long.

Meals were purchased through the military supply procedures of the armed-forces procurement system, and the taste testing was carried out at the Natick Laboratories (NLABS). On arrival at NLABS they were inspected for completeness, immediately tested at room temperature (21ºC) and stored at four different temperatures. Those stored at room temperature were withdrawn and tested at 12, 24 36, 48, 60 months' storage.

The meals were opened by test monitors, and each item served to a panel of 36 untrained subjects who judged its acceptability on a 9-point hedonic rating scale. At a session, each consumer evaluated all the items in one meal which consists of four to six items (including an entree) served one at a time in random order with a mouth-rinsing between items.

Each item in the entire meal, which consists of the entree and the other items, was rated on the 9-point hedonic scale by each panelist (Only one storage temperature was tested for that particular meal, and other temperatures for the same meal were judged mostly by other panelists.). The panelists were chosen from a pool of volunteers comprising both military and civilian staff at NLABS. At most, two meals were tested each day, one in the morning session and one in the afternoon. Care was taken so that no panelist was used twice in the same day. Thus, it is not unreasonable to entertain the assumption that the responses across meals and storage temperatures are uncorrelated.

The samples were coded alphabetically when presented to the test-subjects. The items were all served at room temperature as they came from the package, except for the dehydrated items, which were re-hydrated with water at 60ºC before serving. The tests took place in semi-isolated booths at NLABS under standard fluorescent lighting conditions. At any withdrawal period as many as 48 sessions (twelve menus at four temperatures) were required, which means that the tests went up to 5 weeks, and individual panelists were used about ten times during that period. Thus, it is natural to assume that the responses on each item in a meal follow a multinomial distribution, with different distributions for different entrees.

For each of the 23 combinations of time and temperature, there were sensory ratings for each of the 36 panelists, and so the data for each item consisted of 828 scores. The results were studied for 12 entrees: pork sausage (1), ham-chicken loaf (2), beef patty (3), barbecued beef (4), beef stew (5), frankfurters (6), turkey (7),

beef in gravy (8), chicken (9), meat balls (10), ham slices (11) and beef in sauce (12).

Our contact at NLABS suggested, of course with uncertainty, that among the best entrees are 5, 9 and 11. In fact, Chen, Nandram and Ross (1995) found that at room temperature the shelf lives of 5, 9 and 11 are very long (12, 8 and 14 years respectively) making these estimates less useful.

In Table 1 the responses of the 36 panelists for each entree are presented for the entrees withdrawn after 12 months' storage; the last two columns contain the average (avg) and standard deviation (std) of the 36 scores. Here, chicken (entree 9) has the largest average and the smallest standard deviation, and beef stew (entree 5) seems to be a good competitor.

Further, a Bonferroni multiple comparison procedure was performed using the ANOVA procedure of SAS on the raw data. Of course, this procedure assumes that the 36 scores are normally distributed. At 12 months' storage, the procedure indicated no significant differences between the means of the entrees, suggesting that there is no best entree at 12 months' storage. Thus, a procedure which is more sensitive than classical multiple comparison is needed.

Numerical Results

The data on the sensory evaluation of the twelve entrees withdrawn after twelve months' storage was used. Selection and estimation were studied in turn. The best subset of entrees with t entrees, t = 1,…, 4 were considered. First, a uniform prior on all subsets of size t was considered. That is, $\lambda_b = T^{-1}, b = 1,\ldots,T$ was taken. To make comparisons a much larger prior probability λ_b = .25 for a pre-assigned best subset and the remaining probability split equally among the (T - 1) subsets was also studied. To further assess difference between the criteria based on the mean response ordering (MRO) and the modified stochastic ordering (MSO) the observed data was perturbed by replacing each of the last two cell counts by the average of the observed cell counts for the last two cells for each entree.

In Table 2, the posterior probability $\hat{\lambda}_b$ and the Bayes factor *Bf* associated with the presumed best subsets which are {9} {5, 9} ,{ 5, 9, 11} , { 5, 7, 9, 11} by criterion, data and prior weight λ_b is presented. For the observed data when uniform prior weight is used, except for the best entree which is {9} when the MRO is used and {11} when the MSO is used, the determined subsets of size 2, 3 and 4 are the same, being exactly the presumed best subsets.

The best subsets with prior λ_b = .25 are the same as the presumed best subsets. The posterior probabilities increase as the number of subsets increase for both MRO and MSO, but much more rapidly for the MRO. For the perturbed data, there are substantial differences between the MRO and the MSO with the uniform prior. The posterior probability decreases with the number of subsets for the MRO and less rapidly for the MSO. But in both cases the Bayes factor increases rapidly with the number of subsets, more rapidly for the MRO.

Note that the best subsets of sizes 1, 2, 3, 4 with the MRO are {5}, {5, 9} {5, 9, 11}, {1, 5, 9, 11} respectively as compared with {11}, {9, 11} {5, 9, 10}, {5, 7, 9, 10}. The best subsets with the perturbed data and λ= .25 are the same as those for the observed data for both the MRO and the MSO. Thus, the two criteria can lead to different judged best subsets. However, if the prior probability on the best subset is substantial, the two criteria provide the same best subsets, the evidence with the MRO is slightly larger than with the MSO.

In Table 3, a sensitivity analysis to investigate misspecifications with the presumed best subsets is presented. A prior probability of λ_b = .25 is assigned to the possibly worst subsets {2}, {*2*, 4}, {*2*, 4, 6} and {*2*, 4, 6, 12} with a probability of .75 assigned equally to the remaining T – 1 subsets. Again, the observed and the perturbed data are considered. With the MSO the evidence for the presumed best subsets is very weak, and in fact, the best judged subsets are the ones expected. However, with the MSO the best subsets are the same as assigned for sizes 1, 2, 3 with very weak evidence, and for size 4 the best subset is {5, 7, 9, 10} rather than

{5, 7, 9, 11} as specified by the MRO (Note that the evidence is substantial.). Although the judged best subsets for the perturbed data and the observed data are the same, there are substantial differences between the MRO and the MSO for the perturbed data. The determined subsets are different at every size and interestingly the best subset of size 4 has associated with it fairly large Bayes factors (82.5 versus 29.2). Thus, it is important to specify the correct subset a priori especially if a large prior probability is placed on such a subset. Note that the determined subsets are different for the four scenarios.

Thus, the best subsets of any size are likely to be different for the two criteria, suggesting that it is risky to use the category scales when selecting the best subsets.

Next, consider estimation of the mean response β_i and the measure of variability τ_i for which the posterior mean and standard deviation are obtained. Letting δ denote either β_i or τ_i, we take $\text{AVG}_C = \hat{E}(\delta|\underset{\sim}{n})$ and $\text{STD}_C = \{\hat{\text{var}}(\delta|\underset{\sim}{n})\}^{1/2}$ under criterion based on C (MRO or MSO). Then, consider the ratio $R_{avg} = AVG_{mso}/AVG_{mro}$ and $R_{std} = STD_{mso}/STD_{mro}$.

In Table 4, results are presented for the observed data by prior weight for the modified stochastic ordering (MSO) for subsets of size 4. Columns 3 and 4, and 7 and 8, show there are minor differences between posterior means for β_i and τ_i respectively for $\lambda = T^{-1}$ and λ=.25. In addition, columns 5 and 9 show minor differences between the point estimates when the MRO and MSO are used. However, columns 6 and 10 show substantial differences between the MRO and MSO. R_{std} under the MSO is at least twice as large under the MRO for the β_i and at least one and a half times as large for the τ_i. Note also that there are differences for R_{std} between $\lambda = T^{-1}$ and λ=.25 (e.g., compare the values for entrees 7 and 10 in column 6). Thus, for estimation when little difference is expected between the posterior means with the MRO and MSO, there are substantial differences between the standard deviations.

In Table 5, ranges are considered for the ratios R_{avg} and R_{std} for subsets of sizes 1-4 $\lambda = T^{-1}$ and λ=.25 and for the observed data sets and the perturbed data sets for the β_i and the τ_i. The ranges for R_{avg} are very similar for both β_i and τ_i for all scenarios (i.e., the posterior means are very similar under MRO and MSO). The standard deviations are much larger under the MSO for β_i, but not so large for the τ_i, and there is a slight increase in the ranges of R_{std} from T^{-1} to λ=.25. In addition, as expected, note that there are virtually no differences in estimation for various sizes of the subsets.

Table 1: Panelists' responses for the military sensory evaluation Response Categories

Entree	1	2	3	4	5	6	7	8	9	avg	std
1	2	0	1	5	4	6	8	8	2	6.08	2.01
2	0	4	1	7	4	8	6	5	1	5.50	1.93
3	2	1	3	7	3	8	8	4	0	5.33	1.94
4	0	2	1	3	5	10	8	7	0	6.00	1.64
5	0	0	1	3	7	6	8	10	1	6.42	1.50
6	0	3	4	7	4	8	8	2	0	5.17	1.75
7	0	1	0	5	4	10	10	5	1	6.14	1.50
8	1	3	2	3	4	12	7	4	0	5.50	1.86
9	0	0	1	5	0	9	14	6	1	6.44	1.40
10	0	0	2	5	4	7	11	7	0	6.14	1.51
11	2	1	2	1	1	5	17	6	1	6.25	1.98
12	2	2	5	3	0	13	6	3	2	5.42	2.16

Note: Meals were withdrawn after twelve months' storage.

Table 2: Posterior probability, Bayes factor and the judged best subset (*b*) of entrees with a prior probability on the presumed best subset by data, criterion and prior weight

Observed Data						Perturbed Data					
MRO			MSO			MRO			MSO		
$\hat{\lambda}_b$	B_f	t_b	$\hat{\lambda}_b$	B_f	t_b	$\hat{\lambda}_b$	B_f	t_b	$\hat{\lambda}_b$	B_f	t_b
(a) $\lambda_b = T^{-1}$											
.36	5.1	9	.21	2.9	11	.32	5.2	5	.21	2.9	11
.72	22.7	5, 9	.34	4.7	5, 9	.20	16.6	5, 9	.10	7.5	9, 11
.85	50.2	5, 9, 11	.59	12.8	5, 9, 11	.13	31.3	5, 9, 11	.06	14.7	5, 9, 10
.88	64.4	5, 7, 9, 11	.69	20.1	5, 7, 9, 11	.11	62.0	1, 5, 9, 11	.04	22.0	5, 7, 9, 10
(b) $\lambda_b = .25$											
.63	5.1	9	.38	1.9	9	.59	4.3	9	.41	2.1	9
.88	22.7	5, 9	.61	4.7	5, 9	.85	16.6	5, 9	.62	4.9	5, 9
.94	50.2	5, 9, 11	.81	12.8	5, 9, 11	.91	31.3	5, 9, 11	.80	11.7	5, 9, 11
.96	64.4	5, 7, 9, 11	.87	20.1	5, 7, 9, 11	.94	44.9	5, 7, 9, 11	.88	21.7	5, 7, 9, 11

NOTE: The presumed best subsets are {9}, {5, 9}, {5, 9,11}, {5, 7, 9, 11}; a probability λ_b is assigned to each of these subsets and $(1-\lambda_b)(T-1)^{-1}$ is assigned to each of the remaining $(T-1)$ subsets; mean response ordering (MRO), modified stochastic ordering (MSO)

Table 3: Posterior probability, Bayes factor for the judged best subset (*b*) of entrees under misspecification of the presumed best subset by data, criterion and prior weight

Observed Data					Perturbed Data				
Preassigned		Determined			Preassigned		Determined		
$\hat{\lambda}_b$	B_f	$\hat{\lambda}_b$	B_f	t_b	$\hat{\lambda}_b$	B_f	$\hat{\lambda}_b$	B_f	t_b
(a) Mean Response Ordering (MRO)									
.24	0.3	.25	6.9	9	.10	0.3	.30	5.8	5
.02	0.0	.25	43.8	5, 9	.01	0.0	.20	21.8	5, 9
.00	0.0	.19	100.2	5, 9, 11	.00	0.0	.13	41.6	5, 9, 11
.00	0.0	.12	128.7	5, 7, 9, 11	.00	0.0	.11	82.5	1, 5, 9, 11
(b) Modified Stochastic Ordering (MSO)									
.39	0.6	.39	0.6	4	.20	0.7	.20	0.7	4
.25	0.3	.25	0.3	2, 4	.16	0.6	.16	0.6	2, 4
.37	0.6	.37	0.6	2, 4, 6	.19	0.7	.19	0.7	2, 4, 6
.05	0.1	.07	69.2	5, 7, 9, 10	.01	0.0	.04	29.2	5, 7, 9, 10

NOTE: The presumed worst subsets are {2}, {2, 4}, {2, 4, 6}, {2, 4, 6, 12}; a probability $\lambda_b = .25$ is assigned to each of these subsets and $(1-\lambda_b)(T-1)^{-1}$ is assigned to each of the remaining $(T-1)$ subsets.

Table 4: Posterior mean and standard deviation of μ and τ under MSO, and ratios of posterior means and standard deviations for all entrees based on the judged best four entrees using the observed data by prior weight

		μ				τ			
λ	Entree	AVG	STD	R_{avg}	R_{std}	AVG	STD	R_{avg}	R_{std}
T^{-1}	1	6.52	0.77	1.09	2.38	2.47	0.41	1.20	1.90
	2	5.76	0.64	1.06	2.05	1.83	0.35	0.93	2.07
	3	4.71	0.69	0.89	2.21	1.83	0.38	0.93	2.16
	4	5.82	0.68	0.99	2.46	1.49	0.31	0.86	1.60
	5	6.95	0.60	1.11	2.28	1.68	0.33	1.01	1.86
	6	4.91	0.65	0.95	2.27	1.48	0.30	0.82	2.04
	7	6.49	0.60	1.08	2.28	1.69	0.31	1.03	1.64
	8	4.94	0.68	0.91	2.25	1.74	0.35	0.91	1.86
	9	6.91	0.60	1.10	2.39	1.65	0.32	1.04	1.65
	10	6.11	0.67	1.02	2.56	1.42	0.29	0.85	1.76
	11	6.25	0.78	1.02	2.42	2.33	0.40	1.15	1.59
	12	5.37	0.76	1.00	2.22	2.51	0.41	1.16	2.21
.25	1	6.28	0.71	1.08	2.80	2.47	0.41	1.18	1.99
	2	5.69	0.64	1.05	2.18	1.83	0.35	0.93	2.07
	3	4.67	0.67	0.88	2.18	1.82	0.38	0.93	2.13
	4	5.65	0.65	0.98	2.85	1.51	0.32	0.85	1.73
	5	7.11	0.58	1.12	2.69	1.66	0.32	1.01	1.91
	6	4.89	0.65	0.95	2.29	1.48	0.31	0.82	2.08
	7	6.73	0.56	1.09	3.00	1.68	0.31	1.06	1.80
	8	4.85	0.66	0.89	2.28	1.74	0.36	0.91	1.93
	9	7.02	0.57	1.11	2.73	1.65	0.31	1.05	1.74
	10	5.90	0.64	1.01	3.09	1.43	0.31	0.84	1.82
	11	6.44	0.75	1.03	3.37	2.28	0.38	1.18	1.78
	12	5.28	0.74	0.99	2.25	2.51	0.41	1.16	2.22

Table 5: Ranges of ratios of posterior means and standard deviations of μ and τ based on the judged best subset of sizes 1– 4 by data and prior weight

	μ		τ	
λ	R_{avg}	R_{std}	R_{avg}	R_{std}
(a) Observed data				
T^{-1}	0.89-1.12	2.02-2.56	0.82-1.21	1.54-2.21
.25	0.88-1.13	2.05-3.37	0.82-1.21	1.61-2.22
(b) Perturbed data				
T^{-1}	0.92-1.13	1.78-2.23	0.83-1.21	1.68-2.26
.25	0.91-1.13	1.75-3.17	0.83-1.26	1.70-2.26

Conclusion

The method for how to obtain the best subset of a set of multinomial populations and how to estimate the parameters of any of the selected population has been shown. In addition, it has been shown that the judged best subset can be different under the modified stochastic ordering and the mean response ordering. The methodology applies generally to many sensory data problems when a nonparametric approach might be desirable and when there are small cell counts. For an alternative nonparametric Bayesian approach to estimate several similar multinomial populations see Quintana (1998). He used a Dirichlet process prior to obtain a more robust specification of exchangeability. The method to obtain the best subset of entrees that was outlined in this article is much simpler.

Specifically, five tasks were accomplished. First, a more formal framework for selection than Morris and Christiansen (1996) and Goldstein and Spiegelhalter (1996) has been obtained. The main feature of the estimation method is that it weighs the different subsets according to which one is believed to be best. As there is a joint posterior distribution of the best population and its parameters, estimation proceeds in a simple manner. Second, most non-Bayesian procedures in ranking and selection, use the normality assumption. A normal approximation was not used in this analysis; instead work was done directly with the multinomial assumption. Third, work was done with all the categories in the multinomial table (i.e., collapsing to remove sparseness has not been done). Fourth, this method is sampling based, facilitating a complete probabilistic analysis of the best subset of multinomial populations. Fifth, the method for how to estimate the average response score and standard deviation for each food without actually using the numeric scores has been shown.

With respect to the application discussed, future work will address more complicated issues associated with different storage temperatures, and the other items including the entrees in each meal. It will be useful to obtain the best subset at all temperatures for all rated items in each food. More generally, a number of items is usually rated in accordance with a number of different characteristics. Then, one might wish to find the best subset of items when all the characteristics are taken simultaneously.

References

Bhattacharya, B. & Nandram, B. (1996). Bayesian inference for multinomial populations under stochastic ordering. *Journal of Statistical Computation and Simulation*, *54*, 145- 163.

Chen, M-H., Nandram, B. & 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,* 377-392.

Chen, M-H. & Shao, Q-M. (1997). Estimating ratios of normalizing constants for densities with different dimensions. *Statistica Sinica, 7,* 607-630.

Goldstein, H. & Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, A 159,* 384-443.

Green, P.G. (1984). Iteratively reweighted least squares for the maximum likelihood estimation, and some robust and resistant alternatives (with discussion). *Journal of the Royal Statistical Society, B 46,* 149-192.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773-795.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of theRoyal Statistical Society, B 42,* (with discussions), 109-142.

Meng, X. L. & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica, 6,* 831-860.

Morris, C. N. & Christiansen, C. L. (1996). Hierarchical models for ranking and for identifying extremes, with applications. In *Bayesian Statistics V, eds.:* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Oxford University Press, 277-296.

Nandram, B. (1997). Bayesian inference for the best ordinal multinomial population. In *Case Studies in Bayesian Statistics, eds.: C. Gatsonis, J. S. Hodges, R. E. Kass and N. D. Singpurwalla, Volume III, Springer-Verlag,* 399-418.

Nandram, B., Sedransk, J., & Smith, S. J. (1997). Order restricted bayesian estimation of the age composition of a population of Atlantic cod. *Journal of the American Statistical Association, 92,* 33-40.

Newell, G. J. (1982). Use of linear logistic model for the analysis of sensory evaluation data. *Journal of Food Science, 47,* 818-820.

Quintana, F. A. (1998). Nonparametric Bayesian analysis for assessing homogeneity in $k \times \ell$ contingency tables with fixed right margin totals. *Journal of the American Statistical Association, 93,* 1140-1149.

Robertson, T., Wright, F. T. & Dykstra, R. L. (1988). *Order restricted statistical inference.* John Wiley & Sons, Inc., New York.

Sampson, A. R. & Whitaker, L. R. (1989). Estimation of multivariate distributions under stochastic ordering. *Journal of the American Statistical Association, 84,* 541-548.

Tanner, M. A. (1993). *Tools for statistical inference: methods for exploration of posterior distributions and likelihood functions.* Springer-Verlag, New York.

Appendix A

For the iterative nonlinear least squares, one would take

$$\Delta_{ij} = \ln\{\gamma_{ij}/(1-\gamma_{ij})\} = (\theta_j - \beta_i)/\tau_i$$

where

$$\theta_1 < \theta_2 < \ldots < \theta_{J-1}, \gamma_{ij} = \sum_{p=1}^{j} p_{ip},\ i = 1, 2, \ldots, I,\ j = 1, 2, \ldots, J-1.$$

Let

$$\bar{\theta} = (J-1)^{-1}\sum_{j=1}^{J-1}\theta_j,\ \bar{\Delta}_i = (J-1)^{-1}\sum_{j=1}^{J-1}\Delta_{ij},$$

$$\omega_{ij} = \left\{\sum_{j=1}^{J-1}(\theta_j - \beta_i)/\Delta_{ij}\right\}^{-1}\{(\theta_j - \beta_i)/\Delta_{ij}\},$$

$$i = 1, 2, \ldots, I,\ j = 1, 2, \ldots, J-1.$$

Then, the normal equations, obtained by minimizing

$$\sum_{i=1}^{I}\sum_{j=1}^{J-1}\{\Delta_{ij} - (\theta_j - \beta_i)/\tau_i\}^2$$

over θ_j, β_i, and τ_i, are

$$\theta_j = \left(\sum_{i=1}^{I} \tau_i^{-2} \right)^{-1} \sum_{i=1}^{I} \tau_i^{-2} (\tau_i \Delta_{ij} + \beta_i) \quad j = 1, 2, \ldots, J-1, \tag{A.1}$$

$$\beta_i = \bar{\theta} - \tau_i \bar{\Delta}_i \qquad i = 1, 2, \ldots, I, \tag{A.2}$$

$$\tau_i = \sum_{j=1}^{J-1} \omega_{ij} (\theta_j - \beta_i) \Delta_{ij}^{-1}. \tag{A.3}$$

Letting

$$\hat{p}_{ij} = n_{ij} / n_i., \; \Delta_{ij}^{*} = \ln \left\{ \frac{(\hat{p}_{ij} + 1/2n_i.)}{(1 - \hat{p}_{ij} + 1/2n_i.)} \right\}$$

with

$$n_i. = \sum_{j=1}^{J} n_{ij,} \text{ for } i=1, 2,\ldots, I, j=1, 2,\ldots, J\text{-}1,$$

$$\beta_i = \sum_{j=1}^{J} j \hat{p}_{ij}, \quad \tau_i = \left\{ \sum_{j=1}^{J} \hat{p}_{ij} (j - \beta_i)^2 \right\}^{1/2},$$

$$\theta_j = \left(\sum_{i=1}^{I} \tau_i^{-2} \right)^{-1} \sum_{j=1}^{I} \tau_i^{-2} (\tau_i \Delta_{ij}^{*} + \beta_i).$$

and starting values are obtained by taking

Starting with a random sample $\underset{\sim}{p}^{(1)}, \underset{\sim}{p}^{(2)}, \ldots, \underset{\sim}{p}^{(M)}$, taking $\Delta_{ij}^{(h)} = \ln\{\gamma_{ij}^{(h)} / (1 - \gamma_{ij}^{(h)})\}$ and solving the normal equations (A.1), (A.2), (A.3), samples $\theta_j^{(h)}$, $\beta_i^{(h)}$, and $\tau_i^{(h)}$, *h=1, 2,…, M* are obtained from their empirical posterior distributions.