Quasi-Bayes methods for categorical data
under informative censoring
Abstract

Bayesian methods are discussed for finite sampling when some of the observations are censored (i.e., suffer missing distinctions between categories). Such problems have been researched over the years, as they are important in many application areas. However, most previous work assumed such strong restrictions as non-informative censoring, truthful reporting, etc. Here, we attempt to remove these restrictions. Computational difficulties of Bayesian approaches are first discussed. An approximate computational quasi-Bayes procedure is then given. Simulation studies show that this procedure has desirable convergence properties. In addition, it is shown that the procedure gives the same results as previous Bayesian methods when the familiar restrictions are imposed.

Key Words: Bayesian inference; Generalized Dirichlet distributions; Informative censoring; Quasi-Bayes sequential methods.
§ 1 INTRODUCTION

Bayesian treatments of categorical sampling with censored, or partially-classified, data have been given by Karson and Wrobleski (1970), Antelman (1972), Kaufman and King (1973), Albert and Gupta (1983), Gunel (1984), Smith and Gunel (1984), Smith, Choi, and Gunel (1985), Kadane (1985), and Gibbons and Greenberg (1989). These all deal with $2 \times 2$ contingency tables with information missing regarding row or column variables. Dickey, Jiang, and Kadane (1987) extended consideration to the general multinomial. But all of these studies were restricted to noninformatively censored categorical data. (For treatments from the frequentist viewpoint, see e.g., Hartley (1958), Chen and Fienberg (1974, 1976), Dempster, Laird, and Rubin (1977), Little and Rubin (1987).)

In practice, it is very likely that the pattern of censoring has information regarding parameters of interest, and/or the reported data may not even match the true categories. For example, suppose that each personal income falls into one of five categories. An individual, whose income is in the highest category, may report his/her income as being in the second highest category. This is an example of nontruthful-reporting. In addition, to discourage a refusal to respond, an individual may be allowed to report a set union of two or more categories. However, then, an individual who is truly in the highest category, may be more likely to report himself/herself as being in the top two categories than another individual who is truly in the second highest category. This is an example of informatively censored reporting.

while Walker (1996) used MAP (maximum a posteriori) to make inference.

In this paper, we assume that observations are obtained sequentially. We will propose “quasi-Bayes” methods, to analyze informatively censored and nontruthfully reported data, analogous to methods by Makov and Smith (1977), Smith and Makov (1978) and Titterington, Smith and Makov (1985, Chapter 6). Our quasi-Bayes methods give approximate posterior distributions and approximate posterior means for a more general prior distribution family than that considered by Paulino and Pereira (1995). We further show that our quasi-Bayes posterior means would be the same as those using Paulino and Pereira under their prior distribution family. Finally, from our extensive simulations, the quasi-Bayes estimate based on forward data order and that based on backward data order are seen to converge to the same value when sample size is large.

§ 2 MULTIPLE BERNOULLI SAMPLING PROCESS

In a sequence of n (n prespecified) multiple-Bernoulli trials having $I$ categories, let $Y_1, Y_2, \ldots, Y_n$ denote the first, second, \ldots, $n$-th trial variable. With $\theta_i$ as the probability of a trial outcome in the $i$-th category for $i = 1, 2, \ldots, I$, write

\[
\Pr(Y_k = i) = \theta_i, \quad \text{for } k = 1, \ldots, n.
\]

Then $\theta_+ = 1$, where $\theta_+ = \sum_{i=1}^{I} \theta_i$.

The Dirichlet distributions are a conjugate prior family for samples from such a multiple-Bernoulli distribution. The random vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)$ is said to have the Dirichlet distribution $D(\boldsymbol{b})$, denoted by $\boldsymbol{\theta} \sim D(\boldsymbol{b})$, with parameter vector $\boldsymbol{b} = (b_1, \ldots, b_I)$, each $b_i > 0$, if $\boldsymbol{\theta}$ has the density in any $I - 1$ of its coordinates,

\[
f(\boldsymbol{\theta}; \boldsymbol{b}) = B(\boldsymbol{b})^{-1} \prod_{i=1}^{I} \theta_i^{b_i-1}, \quad (2.1)
\]
where $B(b) = \left[\prod_{i=1}^{I} \Gamma(b_i)\right] / \Gamma(b_+)$, for all $\theta$ in the probability simplex $\{ \theta : \text{each } \theta_i > 0, \theta_+ = 1 \}$. (Throughout this paper, a variable or a parameter with “+” in a subscript represents the sum of this variable or parameter with all possible such subscript values; for example, $\theta_+ = \sum_{i=1}^{I} \theta_i$, and $b_{i+} = \sum_{j=1}^{J} b_{ij}$, when the possible values of $j$, for this $i$, are 1, \ldots, $J$.)

The corresponding prior moment, for a prior distribution, $\theta \sim D(b)$, is $g(c; b) = E_{\theta|b} \prod_{i=1}^{I} \theta_i^{c_i} = B(b + c)/B(b)$. The predictive distribution is then the Dirichlet-multiple-Bernoulli with mass

$$\Pr(Y = y) = g(x; b),$$

where $x = (x_1, \ldots, x_I)$ is the vector of frequency counts of the outcomes vector $Y = y$ in each of $I$ categories. That is, $x_i$ denotes the number of $y_i$'s equal to $i$, and $\sum_{i=1}^{I} x_i = n$.

First, consider the situation when all the data is fully and truthfully categorized. Based on the first trial outcome $Y_1 = y_1$, the posterior distribution, starting from the conjugate prior (2.1), is again a Dirichlet distribution, with updated parameters,

$$\theta | Y_1 = y_1 \sim D(b + \delta_{y_1}),$$

where $\delta_{y_1}$ is an $I$-coordinate vector with value 1 for its $y_1$-th coordinate, and 0 otherwise. The posterior Dirichlet density is then $f(\theta; b + \delta_{y_1})$ and the posterior moment is $g(c; b + \delta_{y_1})$. Before receiving outcome $y_2$, we regard (2.2) as the prior distribution. The posterior distribution, after $Y_2 = y_2$, is then

$$\theta | Y_1 = y_1, Y_2 = y_2 \sim D(b + \delta_{y_1} + \delta_{y_2}).$$

This process continues until we have received all $n$ outcomes $y_1, y_2, \ldots, y_n$. The eventual posterior distribution is

$$\theta | y_1, y_2, \ldots, y_n \sim D(b + \delta_{y_1} + \delta_{y_2} + \cdots + \delta_{y_n}) \sim D(b + x),$$
where \( \mathbf{x} = (x_1, \ldots, x_I) \), and \( x_i, i = 1, \ldots, I \), is the number of \( y_j \)'s equal to \( i \). The corresponding posterior density is \( f(\mathbf{\theta} ; \mathbf{b} + \mathbf{x}) \) and the posterior moment is \( g(\mathbf{c} ; \mathbf{b} + \mathbf{x}) \). Hence, there are closed forms for the posterior moments.

In practice, it is very likely that some of the outcomes may not be reported completely and truthfully. We shall use \( R_k \) to denote the report of the \( k \)-th respondent (or subject), where \( k = 1,2, \ldots, n \). Here, the value of variable \( R_k \) is a category set, a non-empty subset of \( \{ 1,2, \ldots, I \} \). For example, suppose the first respondent, who is truly from the second category, reports as being in either the third or the fourth category. Then, from this example, \( R_1 = \{3,4\} \), a non-truthful report, since the true category of the first subject is not in the reported category set \( R_1 = \{3,4\} \). Assume that there are only \( J \) (\( J \leq 2^I - 1 \)) different possible reported category sets. We shall use \( 1, \ldots, J \) to index these category sets. Further, let \( \lambda_{i,r} \) and \( \lambda_{ij} \), respectively, with and without a comma, denote the conditional probabilities that a respondent, who is truly from the \( i \)-th category, reports in the category set \( r \), or the \( j \)-th category set. Here, for each \( i \), \( \sum_r \lambda_{i,r} = \sum_{j=1}^J \lambda_{ij} = 1 \). Let \( \Lambda \) be the probabilities \( \lambda_{ij} \) arranged in matrix form. Then, \( \Lambda \) is an \( I \times J \) conditional probability matrix. In \( n \) trials, the independent marginal probabilities of receiving reports \( R_1 = r_1, R_2 = r_2, \ldots, R_n = r_n \), are each of the form,

\[
\Pr(R_k = r_k \mid \mathbf{\theta}, \Lambda) = \sum_{i=1}^I \lambda_{i,r_k} \cdot \theta_i, \tag{2.5}
\]

for \( k = 1, \ldots, n \). In this paper, we treat \( \Lambda \) as an unknown parameter matrix.

Dickey, Jiang, and Kadane (1987) developed Bayesian inferences with three assumptions concerning the censoring process \( \lambda_{i,r} \):

(i) Reporting is truthful:

\[ \lambda_{i,r} = 0 \text{ where } i \notin r. \]
(ii) Every report $R_k = r$ is differentially noninformative among the categories within $r$;
\[ \lambda_{i,r} = \lambda_{i',r}, \text{ where both } i \in r \text{ and } i' \in r. \]

(iii) Prior independence is assumed between the parameter arrays $\theta$ and $\Lambda$.

In this paper, we shall drop assumptions (i) and (ii) and assume the $I+1$ independent Dirichlet
prior distributions, $\theta \sim D(\alpha)$ and $\lambda_{i\alpha} \sim D(b_{i\alpha})$, where $i = 1, 2, \ldots, I$. (In this paper, a variable
or a parameter with “*” notation in a subscript stands for a vector of this variable or parameter
with all the possible subscript values as its components.) Then, the joint prior probability density
function of $\theta$ and $\Lambda$ is proportional to
\[ \left( \prod_{i=1}^{I} \theta_i^{a_i-1} \right) \cdot \left[ \prod_{i=1}^{I} \left( \prod_{j=1}^{J} \lambda_{ij}^{b_{ij}-1} \right) \right]. \tag{2.6} \]

As an example, consider the structure with all subsets included, $\lambda_{i\alpha} = (\lambda_{i\{1\}}, \lambda_{i\{2\}}, \ldots, \lambda_{i\{I\}},
\lambda_{i\{1,2\}}, \ldots, \lambda_{i\{1,2,\ldots,J\}} = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{iI+1}, \ldots, \lambda_{iJ})$, where $J = 2^I - 1$. This can be considered
the most general case, in the following sense. As usual, we understand that a zero Dirichlet
parameter coordinate, $a_i = 0$ or $b_{ij} = 0$, corresponds to a singular prior distribution in which
$\theta_i = 0$ or $\lambda_{ij} = 0$, with probability one, respectively. Then if the experimental data agrees with
such presumed singularities, they would be preserved in the posterior distribution. If the data
disagrees, a partially noninformative prior density would apply.

Consider now, the inference following a report only of the first trial (first respondent). The
posterior probability density function corresponding to (2.5) for $k = 1$ (first trial) and the joint
prior (2.6) is then proportional to
\[ \left( \prod_{i=1}^{I} \theta_i^{a_i-1} \right) \cdot \left[ \prod_{i=1}^{I} \left( \prod_{j=1}^{J} \lambda_{ij}^{b_{ij}-1} \right) \right] \cdot \left( \sum_{i=1}^{I} \lambda_{i,r_1} \cdot \theta_i \right) \]
\[ = \sum_{m=1}^{I} \left\{ \left( \prod_{i=1}^{I} \theta_i^{a_i+\delta^m_{i}-1} \right) \cdot \left[ \prod_{i=1}^{I} \left( \prod_{j=1}^{J} \lambda_{ij}^{b_{ij}+\delta^m_{ij}-1} \right) \right] \right\}, \tag{2.7} \]
where $\delta_i^m$ is 1 for $i = m$, and is 0 otherwise, and $\delta_{ij}^{mr_1}$ is 1 for $i = m$ and $j = r_1$, and is 0 otherwise.

This posterior density can be expressed as

$$
\sum_{m=1}^{I} \frac{A_m}{A_+} \left\{ \left( \prod_{i=1}^{I} \theta_i^{\delta_i^m + \delta_i^m - 1} \right) \cdot \left[ \prod_{j=1}^{J} \left( \prod_{i=1}^{I} \lambda_{ij}^{b_{ij} + \delta_{ij}^{mr_1} - 1} \right) \right] \right\} / A_m, \tag{2.8}
$$

where $A_m = B(a + \delta^m) \cdot \prod_{i=1}^{I} B(b_{is} + \delta_{is}^{mr_1})$, $a = (a_1, a_2, \ldots, a_I)$, $\delta^m = (\delta_1^m, \delta_2^m, \ldots, \delta_I^m)$, $b_{is} = (b_{i1}, b_{i2}, \ldots, b_{ij})$, $\delta_{is}^{mr_1} = (\delta_{i1}^{mr_1}, \delta_{i2}^{mr_1}, \ldots, \delta_{ij}^{mr_1})$, and $A_+ = \sum_{m=1}^{I} A_m$.

Here the posterior density (2.8) is a weighted average of products of Dirichlet densities. The posterior moment is the weighted average of $I$ ratios of $A$'s, each of the form $A_m'/A_m$, where $A_m'$ is, similar to $A_m$, the product of $B$'s. For example, the posterior mean of $\theta_1$ is $\sum_{m=1}^{I} (A_m/A_+) \cdot \left( A_m^{(1)} / A_m \right)$, where $A_m^{(1)} = B(a^{(1)} + \delta^m) \cdot \prod_{i=1}^{I} B(b_{is} + \delta_{is}^{mr_1})$, and $a^{(1)} = (a_1 + 1, a_2, \ldots, a_I)$.

If we then receive the further report $R_2 = r_2$, the updated posterior probability density function is then proportional to the product of (2.8) and (2.5) for $k = 2$. This would give the new posterior p.d.f. as a mixture of $I^2$ products of Dirichlet densities. The posterior moment is now the weighted average of $I^2$ ratios of $A$'s. As the number of reports received increases, the number of ratios of $A$'s increases drastically. This would make the computation of posterior moments unfeasible. In the next section, we shall give an approximate posterior distribution, which is aimed at achieving both computational simplicity and convergence.

§ 3 QUASI-BAYES METHODS

First, for convenience, we denote the joint prior probability density function (2.6) by

$$
p_0(\theta, \Lambda) = Q(a) \cdot \prod_{i=1}^{I} Q(b_{is}), \tag{3.1}
$$

where $Q(a) \equiv f(\theta \mid a)$ denotes the probability density function of the Dirichlet distribution with parameter $a$. The posterior density (2.8), after receipt of the first report $R_1 = r_1$, then has the
form,
\[
p_1(\theta, \Lambda \mid R_1 = r_1) = \sum_{m=1}^{I} (A_m / A_+) \cdot Q(a + \delta^m) \cdot \prod_{i=1}^{I} Q(b_{is} + \delta_{is}^{m,r_1}). \tag{3.2}
\]

If we were further informed the true category of the first subject, then the posterior density would become
\[
Q(a + c^{(1)}) \cdot \prod_{i=1}^{I} Q(b_{is} + c^{(1)}_i \cdot \delta_{is}^{r_1}), \tag{3.3}
\]
where \( c^{(1)} = (c_1^{(1)}, c_2^{(1)}, \ldots, c_I^{(1)}) \) and each \( c_i^{(1)} \) is 1 if the true category of the first subject is \( i \), and is 0 otherwise. However, we are not further informed the true category of the first subject. Instead, we base our decision upon the quasi-datum \( d_i^{(1)} = E \left[ C_i^{(1)} \mid R_1 = r_1 \right] = \Pr \left( C_i^{(1)} = 1 \mid r_1 \right) \), the expected value of \( C_i^{(1)} \) posterior to the datum \( r_1 \). Whether we use \( c_i^{(1)} \) or \( C_i^{(1)} \) depends on whether the true category of the first subject, \( i \), is known or not. Now we have
\[
d_i^{(1)} = \Pr \left( C_i^{(1)} = 1 \mid R_1 = r_1 \right) = \frac{\Pr \left( R_1 = r_1 \mid C_i^{(1)} = 1 \right) \cdot \Pr \left( C_i^{(1)} = 1 \right)}{\sum_{m=1}^{I} \Pr \left( R_1 = r_1 \mid C_m^{(1)} = 1 \right) \cdot \Pr \left( C_m^{(1)} = 1 \right)}.
\]
Let
\[
\hat{d}_i^{(1)} = \left[ \hat{\lambda}_{i,r_1}^{(0)} \cdot \hat{\theta}_i^{(0)} \right] / \sum_{m=1}^{I} \left[ \hat{\lambda}_{m,r_1}^{(0)} \cdot \hat{\theta}_m^{(0)} \right],
\]
where \( \hat{\theta}_i^{(0)} \) and \( \hat{\lambda}_{i,r_1}^{(0)} \) are prior means of \( \theta_i \) and \( \lambda_{i,r_1} \), respectively. We use \( \hat{d}_i^{(1)} \) to estimate \( d_i^{(1)} \). This provides the key to our proposed procedure. We approximate (3.2) by the density (3.3),
\[
\hat{p}_1(\theta, \Lambda \mid R_1 = r_1) = Q(a + d^{(1)}) \cdot \prod_{i=1}^{I} Q(b_{is} + \hat{d}_i^{(1)} \cdot \delta_{is}^{r_1})
\]
\[
= Q(a^{(1)}) \cdot \prod_{i=1}^{I} Q(b_{is}^{(1)}), \tag{3.4}
\]
where $\mathbf{a}^{(1)}$ and $\mathbf{b}_{is}^{(1)}$ are the updated parameter vector at the first step and $\mathbf{a}^{(1)} = \left( \tilde{a}_1^{(1)}, \tilde{a}_2^{(1)}, \ldots, \tilde{a}_I^{(1)} \right)$. That is, $\mathbf{a}^{(1)} = \mathbf{a} + \tilde{\mathbf{a}}^{(1)}$ and $\mathbf{b}_{is}^{(1)} = \mathbf{b}_{is} + \tilde{\mathbf{b}}_{is}^{(1)} \cdot \mathbf{b}_{is}^{(1)}$, for each $i$. This approximate posterior distribution is within the prior distribution family. Subsequent updating proceeds in the identical manner. For the $n$-th step, after receiving the $n$-th report $R_n = r_n$, our approximate posterior distribution of $\mathbf{\theta}$ and $\Lambda$ is

$$
\hat{p}_n(\mathbf{\theta}, \Lambda \mid R_1 = r_1, \ldots, R_{n-1} = r_{n-1}, R_n = r_n) = Q \left( \mathbf{a}^{(n-1)} + \tilde{\mathbf{a}}^{(n)} \right) \prod_{i=1}^{I} Q \left( \mathbf{b}_{is}^{(n-1)} + \tilde{\mathbf{b}}_{is}^{(n)} \cdot \mathbf{b}_{is}^{(n)} \right),
$$

(3.5)

where $\mathbf{a}^{(n-1)}$ and $\mathbf{b}_{is}^{(n-1)}$ are the updated parameter vector at the $(n-1)$st step. That is, $\mathbf{a}^{(n-1)} = \mathbf{a}^{(n-2)} + \tilde{\mathbf{a}}^{(n-1)}$ and, for each $i$, $\mathbf{b}_{is}^{(n-1)} = \mathbf{b}_{is}^{(n-2)} + \tilde{\mathbf{b}}_{is}^{(n-1)} \cdot \mathbf{b}_{is}^{(n-1)}$. The approximate general posterior moment of $\theta_i$'s and $\lambda_{ij}$'s is

$$
\hat{E} \left[ \left( \prod_{i=1}^{I} \theta_i^{q_i} \right) \cdot \left( \prod_{i=1}^{I} \prod_{j=1}^{J} \lambda_{ij}^{q_{ij}} \right) \right] = \hat{E} \left( \prod_{i=1}^{I} \theta_i^{q_i} \right) \cdot \left\{ \prod_{i=1}^{I} \left[ \hat{E} \prod_{j=1}^{J} \lambda_{ij}^{q_{ij}} \right] \right\} = \left[ B \left( \mathbf{a}^{(n)} + \mathbf{a'} \right) / B \left( \mathbf{a}^{(n)} \right) \right] \cdot \prod_{i=1}^{I} \left[ B \left( \mathbf{b}_{is}^{(n)} + \mathbf{b}_{is}^{(n)} \right) / B \left( \mathbf{b}_{is}^{(n)} \right) \right].
$$

(3.6)

Hence, for example, the posterior means of $\theta_i$ and $\lambda_{ij}$, based on $r_1, r_2, \ldots, r_n$, can be approximated, for each $i = 1, 2, \ldots, I$, by

$$
\hat{\theta}_i^{(n)} = \left[ a_i^{(n)} / a_i^{(n)} \right] \equiv \left[ a_i^{(n-1)} + \tilde{a}_i^{(n)} \right] / \left[ a_i^{(n-1)} + \tilde{a}_i^{(n)} \right] + \tilde{a}_i^{(n)} / \left[ a_i^{(n-1)} + \tilde{a}_i^{(n)} \right]
$$

and

$$
\hat{\lambda}_{ij}^{(n)} = \left[ b_{ij}^{(n)} / b_{ij}^{(n)} \right] \equiv \left[ b_{ij}^{(n-1)} + \tilde{b}_{ij}^{(n)} \right] / \left[ b_{ij}^{(n-1)} + \tilde{b}_{ij}^{(n)} \right], \text{ for } j = r_n,
$$

$$
\hat{\lambda}_{ij}^{(n)} = \left[ b_{ij}^{(n)} / b_{ij}^{(n)} \right] \equiv \left[ b_{ij}^{(n-1)} + \tilde{b}_{ij}^{(n)} \right] / \left[ b_{ij}^{(n-1)} + \tilde{b}_{ij}^{(n)} \right], \text{ for } j \neq r_n.
$$
where \( b_{i+}^{(n)} = \hat{b}_{i+}^{(n-1)} + \hat{\Pr}(C_i^{(n)} = 1 \mid R_n = r_n) = b_{i1} + b_{i2} + \cdots + b_{iJ} + \hat{\Pr}(C_i^{(1)} = 1 \mid R_1 = r_1) + \hat{\Pr}(C_i^{(2)} = 1 \mid R_2 = r_2) + \cdots + \hat{\Pr}(C_i^{(n-1)} = 1 \mid R_n = r_n) \), \( a_+^{(0)} = \sum_{i=1}^{I} a_i \) and \( \hat{\Pr}(C_i^{(n)} = 1 \mid R_n = r_n) \).

\[ r_n = \left[ \hat{\lambda}_{i,r_n}^{(n-1)} \cdot \hat{\theta}_i^{(n-1)} \right] \bigg/ \left\{ \sum_{m=1}^{J} \left[ \hat{\lambda}_{m,r_n}^{(n-1)} \cdot \hat{\theta}_m^{(n-1)} \right] \right\}. \]

\section{Properties of the Quasi-Bayes Methods}

In this section, we assume that there are \( J \) different possible types (or category sets) and the data are denoted by \( n = (n_1, \ldots, n_J) \), where \( n_j \) is the count of reports with the \( j \)-th category set. We further assume that we have prior independent Dirichlet distributions

\[ \theta \sim D(a) \text{ and } \lambda_{i*} \sim D(b_{i*}), \quad \forall i = 1, \ldots, I, \quad (4.1) \]

where \( \theta = (\theta_1, \ldots, \theta_I) \), \( a = (a_1, \ldots, a_I) \), \( \lambda_{i*} = (\lambda_{i1}, \ldots, \lambda_{iJ}) \), and \( b_{i*} = (b_{i1}, \ldots, b_{iJ}) \). Several cases with additional assumptions are considered here.

\textbf{Case A.} Each \( a_i \) is the sum of all entries of the \( i \)-th row of parameter matrix \([b_{ij}]\). That is,

\[ a_i = b_{i+} = \sum_{j=1}^{J} b_{ij}, \quad \forall i = 1, \ldots, I. \quad (4.1A) \]

\textbf{Case B.} All \( a_i \)'s are equivalent and all \( b_{i*} \)'s are equivalent. That is,

\[ a_1 = \cdots = a_I, \text{ and } b_{1*} = \cdots = b_{I*}. \quad (4.1B) \]

\textbf{Case C.} There exist positive real numbers \( e_1, \ldots, e_J \) such that

\[ b_{ij} = a_i \cdot e_j, \quad \forall i = 1, \ldots, I, \text{ and } j = 1, \ldots, J. \quad (4.1C) \]

We shall show that the quasi-Bayes estimates take the exact values of the actual posterior means under the restrictions on the Dirichlet parameters in Case A, and take fixed approximate posterior means under the other restrictions on the Dirichlet parameters, Cases B, C.
For convenience, define $w_{ij}$ as the probability that the true category of a subject is $i$ and this subject reports in the $j$-th category set (or the $j$-th type). That is, $w_{ij} = \theta_i \cdot \lambda_{ij}$. Hence,

$$\theta_i = \sum_{j=1}^{J} w_{ij}, \text{ for all } i = 1, \ldots, I \text{ and } \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ij} = 1.$$  

Furthermore, $[w_{ij}] \sim D([\tilde{b}_{ij}])$ if $\mathbf{w} \sim D(\tilde{b})$, where $\mathbf{w}$ is the row vector of all entries of matrix $[w_{ij}]$ arranged from the first row of matrix $[w_{ij}]$ to the last row, as in $\tilde{b}$. Hence $\mathbf{w} = (w_{11}, \ldots, w_{1J}, w_{21}, \ldots, w_{2J}, \ldots, w_{I1}, \ldots, w_{IJ})$ if $[w_{ij}]$ is an $I \times J$ matrix. Theorem 4.3 of Dickey, Jiang, and Kadane (1987) says that the independent Dirichlet distributions (4.1) and (4.1A) are equivalent to

$$[w_{ij}] \sim D([b_{ij}]), \quad (4.2)$$

where $[w_{ij}]$ is an $I \times J$ matrix with $i,j$-th entry being $w_{ij}$. In addition, it can be shown that the posterior mean of the $i$-th categorical probability is

$$E(\theta_i \mid \mathbf{n}) = \left[ b_{i+} + \sum_{j=1}^{J} n_j (b_{ij} / b_{++}) \right] / (b_{++} + n_+), \quad (4.3)$$

where $n_+$ is the sum of all $n_j$’s, as is $b_{++}$. Note that Paulino and Pereira (1995) (see their equation (9)) give a similar posterior mean when all reports are further assumed truthful (i.e., the reported category set must contain the true category.)

Next, we show that the posterior means of $\mathbf{\theta}$ with the quasi-Bayes method are the same as those with the fully Bayesian method when the prior distribution is assumed to be (4.1) with (4.1A). We give, first, the approximate posterior distribution using the quasi-Bayes method.

**Theorem 1** Assume the prior distribution (4.1) and (4.1A), and censored data $\mathbf{n} = (n_1, \ldots, n_J)$ and let the sum of all entries of $\mathbf{n}$ be denoted by $n_+ (n = n_+)$. Then the approximate posterior distribution of $(\mathbf{\theta}, \Lambda)$ with the quasi-Bayes method is

$$D(\alpha^{(n)}) \cdot D([b_{ij}]^{(n)}),$$

11
where \([b_{ij}]^{(n)}\) is an \(I \times J\) matrix with \((i, j)\)th element \(b_{ij}^{(n)} = b_{ij}(1 + n_j/b_{+j})\), for all \(i = 1, 2, \ldots, I\), and \(j = 1, \ldots, J\), and the \(i\)-th entry of \(a^{(n)}\) is

\[
a_i^{(n)} = \sum_{j=1}^{J} b_{ij}^{(n)} = b_{i+} + \sum_{j=1}^{J} n_j \cdot \frac{b_{ij}}{b_{+j}}.
\]

The following lemma is an immediate consequence of Theorem 1 and equation (4.3).

**Lemma 2** For any informatively censored data with prior distribution (4.1) and (4.1A), the posterior means under quasi-Bayes are identical to the fully Bayesian posterior means.

To have some ideas about the similarities and differences of the quasi-Bayes methods and Bayes methods, we shall give the posterior distribution and the approximate posterior distribution under the prior distribution (4.1) and (4.1A). First, the approximate posterior distribution in Theorem 1 can be reexpressed as, independently,

\[
(w_{+1}, \ldots, w_{+J})|\mathbf{n} \sim D(b_{+1} + n_1, \ldots, b_{+J} + n_J), \quad \text{and} \quad (4.4A)
\]

\[
\frac{1}{w_{+j}}(w_{1j}, \ldots, w_{IJ})|\mathbf{n} \sim D \left[ \left(1 + \frac{n_j}{b_{+j}}\right)(b_{1j}, \ldots, b_{IJ}) \right], \quad \forall j = 1, \ldots, J. \quad (4.4B)
\]

In addition, from Theorem 1 of Jiang, Kadane, and Dickey (1992), the posterior distribution under the same prior distribution (4.1) and (4.1A) can also be expressed as, independently,

\[
(w_{+1}, \ldots, w_{+J})|\mathbf{n} \sim D(b_{+1} + n_1, \ldots, b_{+J} + n_J), \quad \text{and} \quad (4.5A)
\]

\[
\frac{1}{w_{+j}}(w_{1j}, \ldots, w_{IJ})|\mathbf{n} \sim D(b_{1j}, \ldots, b_{IJ}), \quad \forall j = 1, \ldots, J. \quad (4.5B)
\]

From the above expressions, it can be seen that (4.4A) and (4.5A) are equivalent, while the parameter vectors in (4.4B) and (4.5B) are proportional. Note that the posterior distribution under the more general prior distribution is a generalized Dirichlet distribution, (see Dickey (1983), Dickey,
Jiang, and Kadane (1987)), while the approximate posterior distribution is still a Dirichlet distribution.

In the next two lemmas, we show that the approximate posterior means with the different prior assumptions (4.1B) or (4.1C) take fixed values regardless of the censored data.

**Lemma 3** For any informatively censored data with prior distribution (4.1) and (4.1B), the approximate posterior mean of the $i$-th categorical probability $\theta_i$ using quasi-Bayes is $1/I$, for any $i$.

**Lemma 4** For any informatively censored data with prior distribution (4.1) and (4.1C), the approximate posterior mean of the $i$-th categorical probability $\theta_i$ using quasi-Bayes is $a_i/a_+$, for any $i$.

Lemmas 2, 3 and 4 show that the approximate posterior means with quasi-Bayes take the same values as those from the fully Bayesian under some restrictions on the parameters $\mathbf{a}$ and $[b_{ij}]$ and take fixed values, independent of censored data outcomes, under other restrictions. The proofs of Theorem 1, Lemma 3 and Lemma 4 are given in the appendix.

From the proof of Lemma 3, it can be seen that the property of the updated parameters, $a_1^{(k)} = \cdots = a_r^{(k)}$ and $b_{1s}^{(k)} = \cdots = b_{rs}^{(k)}$ after receiving $k$ censored data is always the same as that of the prior (4.1B). This is because the prior parameter matrix $[b_{ij}]$ of $[\lambda_{ij}]$ in (4.1B) has the property that each column vector of $[b_{ij}]$ is proportional to the prior parameter vector $\mathbf{a}$ of $\mathbf{\theta}$. Hence, the quasi-Bayes posterior means of $\mathbf{\theta}$ would always be the same as the quasi-Bayes prior means of $\mathbf{\theta}$ and would not depend on the data. The results and reasons are similar for Lemma 4. Of course, such identifiability issues can be handled well in prior-posterior analysis, because the posterior variance of an unidentifiable parameter will not be diminished relative to the prior variance.
§ 5 SIMULATION STUDIES

We performed simulation studies to investigate additional properties of the quasi-Bayes method. In § 5.1, we study the convergence properties of the method. As quasi-Bayes is a sequential method, we study, in § 5.2, when the estimation results would almost never depend on the order of data received.

§ 5.1 Convergence Properties

First, we consider specific cases, with figures, to see the behavior of the posterior means as the sample size increases. We then make extensive simulations to check if this behavior is still true for other potential cases.

Assume there are three categories and seven possible reported category sets, i.e., $I = 3$, $J = 7$. We generate reported data from a population with the vector of true categorical probabilities $\theta = (0.5, 0.3, 0.2)$ and the conditional probability matrix

$$
\Lambda = \begin{pmatrix}
0.40 & 0.05 & 0.05 & 0.15 & 0.15 & 0.05 & 0.15 \\
0.05 & 0.50 & 0.05 & 0.10 & 0.05 & 0.10 & 0.15 \\
0.05 & 0.05 & 0.40 & 0.05 & 0.10 & 0.10 & 0.25
\end{pmatrix},
$$

where $\Lambda$ is interpreted as in § 2. Hence, given that a subject is truly from the $i$-th category, the $i$-th row of $\Lambda$ consists of the conditional probability vector that this subject reports that it belongs in each of the 7 possible sets of categories, {$1$}, {$2$}, {$3$}, {$1,2$}, {$1,3$}, {$2,3$}, or {$1,2,3$}, which are, according to the order, also called category sets 1, 2, …, 7, respectively. For example, category set 4 is {$1,2$}. We generated data with probability $w_{+,j} = \theta_1 \cdot \lambda_{1j} + \theta_2 \cdot \lambda_{2j} + \theta_3 \cdot \lambda_{3j}$ that each subject reports as being in the $j$-th set of categories, where $j = 1, 2, \ldots, 7$. Therefore the generating probability vector is $w_{+,+} = (0.225, 0.185, 0.12, 0.115, 0.11, 0.075, 0.07)$. Note that $\sum_{j=1}^{7} w_{+,j} = 1.$
There are many other pairs of $\theta$, $\Lambda$ that would have this same generating distribution. This is the problem of identifiability, which is not discussed here. (For such discussion see Paulino and Pereira (1995) and references therein). We study the estimates with various prior parameters $\alpha$ and $[b_{ij}]$ for $\theta$ and $\Lambda$. They include some extreme values, such as $\alpha_1 = (5, 5, 5)$ or $\alpha_2 = (10, 10, 10)$, and

$$
[b_{ij}] = \begin{pmatrix}
    100 & 200 & 300 & 400 & 500 & 600 & 700 \\
    3 & 4 & 5 & 6 & 7 & 1 & 2 \\
    5 & 6 & 7 & 1 & 2 & 3 & 4
\end{pmatrix}
$$

Figures 1 and 2 show our approximate posterior estimates, based on $\alpha_1$ and $[b_{ij}]$, and $\alpha_2$ and $[b_{ij}]$, respectively, as these prior parameters. These simulation studies suggest that the approximate posterior means of $\theta = (\theta_1, \theta_2, \theta_3)$ and the marginal probabilities $w_{+*} = (w_{+1}, w_{+2}, w_{+3}, w_{+4}, w_{+5}, w_{+6}, w_{+7})$ converge, as the number of observations (subjects) increases, as can be seen in Figures 1–2. Note that the posterior means of $\theta$ may converge to different values for different prior parameters $\alpha$ and $[b_{ij}]$. This is expected by the problem of identifiability. For the purpose of the comparison, the estimated marginal probabilities $w_{+j} = \sum_{i=1}^{3} E[\theta_i \mid Y] \cdot E[\lambda_{ij} \mid Y]$ are also given in the figures.

We next confirmed the suggestion that the posterior means of $\theta$, based on the quasi-Bayes method, converge. Our simulations generated different values of $\theta$, $\Lambda$, $\alpha$, $[b_{ij}]$ of (4.1), in addition to generating the sample data. Details follow:

**Step A.** All values of $\theta$, $\Lambda$, $\alpha$, $[b_{ij}]$ of (4.1) are selected randomly.

**Step B.** A report (or reported datum) is randomly selected and posterior means based on this report and all previous reports (if any) are computed. Relative differences of all $\theta_i$'s between two consecutive posterior means are also computed (e.g., $|\theta_1^{(n+1)} - \hat{\theta}_1^{(n)}|/\hat{\theta}_1^{(n)}$, where $\hat{\theta}_1^{(n)}$ is the posterior mean of $\theta_1$ from the first $n$ reports). This step B is continued until the relative difference of all $\theta_i$'s is no more than 1%.
Steps A and B were repeated more than 100 times and we found, in each case, that it took less than 55 observations (less than 20 observations for most cases) for the relative differences to become less than 1%. This evidenced that the approximate posterior means of the quasi-Bayes method converged with less than 55 observations.

§ 5.2 Sensitivity to the Order of the Data

In this section, we study the relative difference of the posterior means of $\theta^{(n)}_i$s if the data of size $n$ are received in different order. Random samples of size 2000 are generated. The posterior means of $\theta^{(n)}_i$s, posterior to the first $n$ observations, where $n = 1, \ldots, 2000$, using the quasi-Bayes method, both according to the original forward data order and the backward order, are computed. The relative differences of these posterior means are also computed (i.e., $\left| \frac{\hat{\theta}^{(n)}_{ib} - \hat{\theta}^{(n)}_{if}}{\hat{\theta}^{(n)}_{if}} \right|$, for all $i = 1, \ldots, I$, where $\hat{\theta}^{(n)}_{if}$ and $\hat{\theta}^{(n)}_{ib}$ are the posterior means of $\theta_i$ corresponding to original forward data order and backward data order, respectively). We then calculated the maximum of these $I$ relative differences, denoted by $\text{max}_{\text{err}}(n)$. That is, $\text{max}_{\text{err}}(n) = \max_{1 \leq i \leq I} \left| \frac{\hat{\theta}^{(n)}_{ib} - \hat{\theta}^{(n)}_{if}}{\hat{\theta}^{(n)}_{if}} \right|$. The above process was repeated many dozen times. All of our simulation studies showed that the trend of $\text{max}_{\text{err}}(n)$ increases for the first few values of $n$ and then decreases for the remaining values of $n$. Note that $\text{max}_{\text{err}}(1) = 0$. In addition, most of our simulation studies needed sample sizes $n$ no more than 100 for $\text{max}_{\text{err}}(n)$ to decrease to less than 5%.

§ 6 CONCLUSIONS

By simulation studies, the posterior estimates in the quasi-Bayes method appear to converge. Because of the identifiability problem, the posterior means of the categorical probability vector $\theta$ may converge to different values for different prior parameters. If estimation accuracy is a concern,
it is suggested that one can calculate and compare the approximate mean based on forward data order and that based on backward data order. If the difference is not small enough, one needs to gather more data before estimating.

Our methods generalize Dickey, Jiang, and Kadane (1987)’s fully Bayesian methods by removing two of their three conditions. In addition, the computer programs for the quasi-Bayes methods are easy to implement and CPU times are relatively small. Hence, by using quasi-Bayes, we are able to approximate new fully Bayesian inferences with informatively censored categorical and/or non-truthfully reported data quickly. These quasi-Bayes methods also generalize those given by Paulino and Pereira (1995), which assume additional restrictions on the prior family. It can be seen from the results in § 4 that our methods and Paulino and Pereira (1995)’s methods have the same posterior means under their more restrictive prior family.

Finally, we give some suggestions for future studies. Although priors (4.1B) or (4.1C) rarely occur in practice, it would be interesting to show that the posterior means are the same as the quasi-Bayes means under the assumptions of Lemma 3 or Lemma 4 in § 4. It would also be interesting to compare this quasi-Bayes approach with that given by Jiang, Kadane, and Dickey (1992) under the cases that the latter is workable. In addition, this quasi-Bayes approach may also be used as a new or an alternative computation method in other problems, e.g., Bayesian smoothing problems which are discussed by Dickey and Jiang (1998).
APPENDIX: PROOFS

Proof of Theorem 1

Assume data are received in the order of \( n_1 \) 1's, \( n_2 \) 2's, \( \ldots \), \( n_J \) \( J \)'s. For convenience, we shall denote the prior distribution parameter vector \( \mathbf{a} \) and matrix \([b_{ij}]\) by \( \mathbf{a}^{(0)} \) and \([b_{ij}]^{(0)}\), respectively, and the posterior version after receiving \( n \) reports by \( \mathbf{a}^{(n)} \) and \([b_{ij}]^{(n)}\), respectively. Hence, the prior parameters are as follows:

\[
\mathbf{a}^{(0)} = \begin{pmatrix}
    a_1 \\
a_2 \\
    \vdots \\
a_I \\
\end{pmatrix} \quad \text{and} \quad [b_{ij}]^{(0)} = \begin{pmatrix}
    b_{11} & b_{12} & \cdots & b_{1J} \\
b_{21} & b_{22} & \cdots & b_{2J} \\
    \vdots & \vdots & \cdots & \vdots \\
b_{I1} & b_{I2} & \cdots & b_{IJ} \\
\end{pmatrix}.
\]

First, by equation (3.4), we can see that \( a_i^{(1)} = a_i + d_i^{(1)} \) and \( b_{ij}^{(1)} = \sum_{j=1}^{J} b_{ij}^{(0)} = \sum_{j=1}^{J} b_{ij} + d_i^{(1)} = b_{i+} + d_i^{(1)} \) and hence \( a_i^{(1)} = b_{i+} \), if \( a_i = b_{i+} \). Similarly, by (3.5), we have \( a_i^{(k)} = b_{i+}^{(k)} \), if \( a_i^{(k-1)} = b_{i+}^{(k-1)} \), where \( k = 2, 3, \ldots, n \). By assumption that \( a_i = b_{i+} \), for all \( i \), and the mathematical induction, we have

\[
a_i^{(k)} = b_{i+}^{(k)}, \quad \text{for all } k = 1, 2, \ldots, n \text{ and for all } i = 1, 2, \ldots, I. \quad (A.1)
\]

Now, from § 3 and Theorem 4.3 of Dickey, Jiang and Kadane (1987), we have

\[
\hat{d}_i^{(k)} = \frac{\hat{\lambda}_{i,r_k}^{(k-1)} \hat{\theta}_i^{(k-1)}}{\sum_{m=1}^{I} \hat{\lambda}_{m,r_k}^{(k-1)} \hat{\theta}_m^{(k-1)}} = \frac{\sum_{j=1}^{J} \hat{b}_{ij}^{(k-1)} \hat{\lambda}_j^{(k-1)} \hat{\theta}_i^{(k-1)}}{\sum_{j=1}^{J} \sum_{i=1}^{I} \hat{a}_i^{(k-1)}} = \frac{\sum_{m=1}^{I} \hat{b}_{im}^{(k-1)} \hat{\lambda}_m^{(k-1)} \hat{\theta}_i^{(k-1)}}{\sum_{m=1}^{I} \sum_{i=1}^{I} \hat{a}_i^{(k-1)}}.
\]
where $k = 1, 2, \ldots, n$ and $i = 1, 2, \ldots, J$. Since $r_1 = 1$ (i.e., first category set), $r_2 = 1, \ldots, r_{n_1} = 1,$ $r_{n_1+1} = 2$, $r_{n_1+2} = 2, \ldots, r_{n_1+n_2} = 2, \ldots, r_n = J$ (i.e., $J^{th}$ category set), then $d_i^{(1)} = b_{i1}/b_{1+1}$, $d_i^{(2)} = b_{i1}^{(1)}/b_{1+1}^{(1)} = [b_{i1}(1 + 1/b_{1+1})]/(b_{1+1} + 1) = b_{i1}/b_{1+1}$, $d_i^{(n_1)} = b_{i1}/b_{1+1}$, $d_i^{(n_1+1)} = b_{i2}/b_{2+2}$, $d_i^{(n_1+2)} = b_{i2}^{(1)}/b_{2+2}^{(1)} = [b_{i2}(1 + 1/b_{2+2})]/(b_{2+2} + 1) = b_{i2}/b_{2+2}$, $d_i^{(n_1+n_2)} = b_{i2}/b_{2+2}$, $\ldots$, $d_i^{(n)} = b_{iJ}/b_{J+J}$. Hence, the posterior distribution parameter matrix and parameter vector, after observing $r_1 = 1, r_2 = 1, \ldots, r_{n_1} = 1, r_{n_1+1} = 2, r_{n_1+2} = 2, \ldots, r_{n_1+n_2} = 2, \ldots, r_n = J$, can be expressed as

$$
[b_{ij}]^{(n)} = \begin{pmatrix}
\begin{array}{cccc}
 b_{11} \left(1 + \frac{n_1}{b_{1+1}}\right) & b_{12} \left(1 + \frac{n_2}{b_{2+2}}\right) & & b_{1J} \left(1 + \frac{n_J}{b_{J+J}}\right) \\
 b_{21} \left(1 + \frac{n_1}{b_{1+1}}\right) & b_{22} \left(1 + \frac{n_2}{b_{2+2}}\right) & & b_{2J} \left(1 + \frac{n_J}{b_{J+J}}\right) \\
 & \vdots & \ddots & \vdots \\
 b_{i1} \left(1 + \frac{n_1}{b_{1+1}}\right) & b_{i2} \left(1 + \frac{n_2}{b_{2+2}}\right) & & b_{iJ} \left(1 + \frac{n_J}{b_{J+J}}\right) \\
 & \vdots & \ddots & \vdots \\
 b_{I1} \left(1 + \frac{n_1}{b_{1+1}}\right) & b_{I2} \left(1 + \frac{n_2}{b_{2+2}}\right) & & b_{IJ} \left(1 + \frac{n_J}{b_{J+J}}\right)
\end{array}
\end{pmatrix}
$$

and

$$
a^{(n)} = \begin{pmatrix}
a_1^{(n)} \\
a_2^{(n)} \\
\vdots \\
a_i^{(n)} \\
\vdots \\
a_J^{(n)}
\end{pmatrix},
$$

where $a_i^{(n)} = \sum_{j=1}^{J} b_{ij}^{(n)} = b_{i+} + n_1 b_{i1}/b_{1+1} + n_2 b_{i2}/b_{2+2} + \cdots + n_J b_{ij}/b_{j+J}$. This completes the proof for a special order of the data. It is apparent from the proof that the posterior distribution based on quasi-Bayes is the same if we change the order of the data.
Proof of Lemma 3

The approximate posterior distribution of $\theta$ and $\Lambda$ is, after receiving reports $r_1, r_2, \ldots, r_k,$

$$D \left( a^{(k)} \right) \cdot \prod_{i=1}^{I} Q \left( b_{is}^{(k)} \right),$$

(A.2)

where $a^{(k)} = a^{(k-1)} + d^{(k)}$, $b_{is}^{(k)} = b_{is}^{(k-1)} + d_{is}^{(k)}$, and $k = 1, 2, \ldots, n$. In addition, $d_{is}^{(k)} = \left( \sum_{m=1}^{I} \hat{\theta}_{m,r_k}^{(k-1)} \right)^2 \left( \sum_{m=1}^{I} \hat{\theta}_{m,r_k}^{(k-1)} \right)$, for all $i = 1, 2, \ldots, I$. By the assumption (4.1B), it can be shown that $d_{i}^{(1)} = 1/I$. Hence, we have $a_{1}^{(1)} = a_{2}^{(1)} = \cdots = a_{I}^{(1)}$ and $b_{1s}^{(1)} = b_{2s}^{(1)} = \cdots = b_{Is}^{(1)}$.

If $a_{1}^{(k-1)} = a_{2}^{(k-1)} = \cdots = a_{I}^{(k-1)}$ and $b_{1s}^{(k-1)} = b_{2s}^{(k-1)} = \cdots = b_{Is}^{(k-1)}$, by the same arguments above, it can be shown that $a_{1}^{(k)} = a_{2}^{(k)} = \cdots = a_{I}^{(k)}$ and $b_{1s}^{(k)} = b_{2s}^{(k)} = \cdots = b_{Is}^{(k)}$. Hence, by the mathematical induction, we have $a_{1}^{(n)} = a_{2}^{(n)} = \cdots = a_{I}^{(n)}$ and $b_{1s}^{(n)} = b_{2s}^{(n)} = \cdots = b_{Is}^{(n)}$. Therefore, the approximate posterior mean of $\theta_i$ is $1/I$.

Proof of Lemma 4

By (A.2) and the assumption (4.1C), it can be shown that $\tilde{d}_{i}^{(1)} = a_i / a_+$. Then, we have $a_{i}^{(1)} = a_i [(a_+/a_+)/a_+]$, $b_{ij}^{(1)} = a_i \cdot e_j$ if $j \neq r_1$ and $b_{ij}^{(1)} = a_i (e_j + 1/a_+)$ if $j = r_1$. Therefore, $b_{ij}^{(1)} = a_i^{(1)} \cdot e_j^{(1)}$, where $e_{j}^{(1)}$ is $[a_+/a_+ + 1] e_j$ if $j \neq r_1$ and $e_{j}^{(1)} = [a_+/a_+ + 1] (e_j + 1/a_+)$ if $j = r_1$. Continue this argument or by the mathematical induction, the lemma can then easily be shown.

REFERENCES


Figure 1: Estimated probabilities, when $a = a_1 = (5, 5, 5)$.

prior means: $\hat{\theta} = (0.3333, 0.3333, 0.3333)^T$

$\hat{w}_{+\ast} = \hat{\theta}' \cdot \hat{\Lambda} = (0.1071, 0.1429, 0.1786, 0.1310, 0.1667, 0.1190, 0.1548)$

posterior means: $\hat{\theta} = (0.2471, 0.3316, 0.4213)^T$

$\hat{w}_{+\ast} = \hat{\theta}' \cdot \hat{\Lambda} = (0.2067, 0.1722, 0.1541, 0.1049, 0.1325, 0.0775, 0.1521)$

Figure 2: Estimated probabilities, when $a = a_2 = (10, 10, 10)$.

posterior means: $\hat{\theta} = (0.2630, 0.3331, 0.4039)^T$

$\hat{w}_{+\ast} = \hat{\theta}' \cdot \hat{\Lambda} = (0.2059, 0.1716, 0.1537, 0.1051, 0.1329, 0.0790, 0.1518)$