# A BAYESIAN TREATMENT OF NONRESPONSE WHEN SAMPLING FROM A DICHOTOMOUS POPULATION

James H. Albert, Bowling Green State University

ABSTRACT. Suppose a sample survey is taken to learn about the proportion of a population in favor of a particular statement. A significant proportion of the sample does not respond to the survey and a priori the experimenter feels that the groups of respondents and nonrespondents possess different attitudes towards the statement. A prior distribution is developed which can reflect vague prior beliefs about the differences in the attitudes of respondents and nonrespondents. This distribution is used to develop an interval estimate of the population proportion.

Key words and phrases: Cross-product ratio, Dirichlet mixture, credible interval.

1. INTRODUCTION. Suppose a mail survey is sent to learn about the attitudes of a population toward a particular subject. Let $n$ denote the total number of surveys sent out and let $n_r$ ($n_s$) denote the number of people who respond (don't respond). The $n_r$ respondents are classified dichotomously into the two groups "favorable" and "unfavorable"; let $n_{fr}$ ($n_{ur}$) denote the number observed in the favorable (unfavorable) group. The observed counts and corresponding probability model are shown below.

| | Resp. | Nonresp. | | | | |
|---|---|---|---|---|---|---|
| Fav. | $n_{fr}$ | | | $p_{fr}$ | $p_{fs}$ | $p_f$ |
| Unfav. | $n_{ur}$ | | | $p_{ur}$ | $p_{us}$ | $p_u$ |
| | $n_r$ | $n_s$ | $n$ | $p_r$ | $p_s$ | |

(Note that two cells in the count table are empty, since these counts are unobservable.) If the population is assumed infinite, the probability of observing the triple $\underset{\sim}{x} = (n_{fr}, n_{ur}, n_s)$ is proportional to $p_{fr}^{n_{fr}} p_{ur}^{n_{ur}} (p_{fs}+p_{us})^{n_s}$. One general problem of interest is to estimate $p_f$, the probability that a given member of the population is favorable.

From a classical point of view, there is no data available to estimate $p_{fs}$ (the probability an individual does not respond and is favorable), and therefore the probability $p_f = p_{fr} + p_{fs}$ is not estimable. The usual practice is to use the statistic $n_{fr}/n_r$, the proportion of respondents in favor, as an estimate of $p_f$. This estimator is unbiased for $p_{fr}/p_r$, the conditional probability that a respondent is favorable. However, if the bias $p_{fr}/p_r - p_f$ is large, $n_{fr}/n_r$ will be an unsuitable estimate of the probability of interest $p_f$. (Birnbaum and Sirken (1950) and Hansen and Hurwitz (1946) have studied the errors incurred from using the estimate $n_{fr}/n_r$.)

This classical problem of estimability can be avoided by means of the Bayesian method. A user states his prior beliefs about the vector $\underset{\sim}{p} = (p_{fr}, p_{fs}, p_{ur}, p_{us})$ in terms of a prior distribution placed on $\underset{\sim}{p}$ and a posterior distribution for $\underset{\sim}{p}$ is computed which combines the prior beliefs of the user with the observed sample counts. The posterior distribution is then used to provide point and interval estimates for $p_f$. This Bayesian method of estimating $p_f$, of course, depends strongly on the form of the prior distribution chosen. The most convenient family of prior distributions on $\underset{\sim}{p}$ is the Dirichlet, with density given by

$$(1.1)\quad \pi_D(\underset{\sim}{p}|K,\underset{\sim}{\eta}) \propto p_{fr}^{K\eta_{fr}-1}\, p_{fs}^{K\eta_{fs}-1}\, p_{ur}^{K\eta_{ur}-1}\, p_{us}^{K\eta_{us}-1},$$

where $K > 0$, $\eta_{ij} > 0$ for all $i,j$, $\eta_{fr} + \eta_{fs} + \eta_{ur} + \eta_{us} = 1$ and $\underset{\sim}{\eta} = (\eta_{fr}, \eta_{fs}, \eta_{ur}, \eta_{us})$. To use this family of prior distributions, a user specifies $\underset{\sim}{\eta}$, which represents a guess at the vector of probabilities $\underset{\sim}{p}$, and $K$, which reflects the precision of the guess $\underset{\sim}{\eta}$. Kaufman and King (1973) use the Dirichlet family of priors to develop a posterior mean estimate of $p_f$ and, in addition, develop solutions to various two-stage sampling problems.

The Bayesian method of estimating $p_f$ using the Dirichlet family of priors is attractive because of its computational simplicity. However, before this Bayesian procedure is recommended in practice, we should investigate whether typical prior beliefs about $\underset{\sim}{p}$ can be modeled by the Dirichlet family. One common prior belief that the user may possess is that the group of respondents are not representative of the entire population. To state in terms of conditional probabilities, it may be believed that $p_{f\cdot r} = p_{fr}/p_r$, the probability a respondent is favorable, is significantly different than $p_{f\cdot s} = p_{fs}/p_s$, the probability a nonrespondent is favorable. Alternatively, the user may believe that $p_{r\cdot f} = p_{fr}/p_f$, the probability a favorable subject responds is much smaller or larger than $p_{r\cdot u} = p_{ur}/p_u$, the probability an unfavorable subject responds. In either case, the user is making a statement a priori about the association structure in the $2\times 2$ table formed by the classes response/nonresponse and favorable/unfavorable.

One implication of the Dirichlet prior is that the conditional probabilities $p_{f\cdot r}$ and $p_{f\cdot s}$ (or $p_{r\cdot f}$ and $p_{r\cdot u}$) are independent. Thus this convenient class of priors will be unsuitable for reflecting prior beliefs about the similarity or dissimilarity of these conditional probabilities. In other words, the Dirichlet family is not a rich enough family to incorporate certain prior beliefs about the bias due to the non-availability of all the subjects.

In the estimation of cell probabilities from a 2×2 table, Albert and Gupta (1982) introduced a class of priors, a mixture of Dirichlet distributions, which is designed to reflect prior beliefs about the association structure of the table. This class of priors accepts the input of two parameters $\alpha_0$ and K. The parameter $\alpha_0$ is a guess at the cross product ratio $\alpha = (p_{fr}p_{us})/(p_{fs}p_{ur})$, reflecting one's prior belief about the cross-classification structure. To understand how one guesses at $\alpha$ in this situation, note that the measure can be rewritten as

$$(1.2) \qquad \alpha = \frac{p_{f\cdot r}/(1-p_{f\cdot r})}{p_{f\cdot s}/(1-p_{f\cdot s})}$$

or

$$(1.3) \qquad \alpha = \frac{p_{r\cdot f}/(1-p_{r\cdot f})}{p_{s\cdot f}/(1-p_{s\cdot f})} .$$

In either expression, $\alpha$ is written as a ratio of odds of conditional probabilities. To illustrate the use of (1.2), one interpretation of the guess $\alpha_0 = 2$ is that the odds of a respondent favoring is believed twice as large as the odds of a nonrespondent favoring. The parameter K reflects the precision of the guess at $\alpha$, or equivalently the sureness of one's prior belief about the association in the table.

Using the prior of Albert and Gupta (1982), section 2 gives the posterior distribution for $p_f$ and uses this distribution to find posterior moments for $p_f$. Since this prior is a mixture of Dirichlets, these results are built on results in the Dirichlet prior case given in Kaufman and King (1973). These posterior moments are then used in the development of an approximate $(1-\gamma)100$ percent credible interval for $p_f$. In section 3, we conclude our discussion by illustrating the computation of the posterior moments in examples. Our main concern is to investigate the effect of the choice of $\alpha_0$ and K on the estimate and the precision of the estimate of $p_f$.

2. PRIOR TO POSTERIOR ANALYSIS. Albert and Gupta (1982) introduced the following two-stage prior distribution to reflect prior beliefs about association in a 2×2 table:

Stage I: The vector $\underset{\sim}{p}$ possess the Dirichlet density (1.1), where the components in $\underset{\sim}{\eta}$ have row margins $\eta_f$, $1-\eta_f$, column margins $\eta_r$, $1-\eta_r$, and cross product ratio $\alpha_0$. That is, the prior means satisfy the configuration

$$(2.1) \qquad \begin{array}{|c|c|c} \hline \eta_{fr}(\eta_f,\eta_r) & \eta_f-\eta_{fr}(\eta_f,\eta_r) & \eta_f \\ \hline \eta_r-\eta_{fr}(\eta_f,\eta_r) & 1-\eta_r-\eta_f+\eta_{fr}(\eta_f,\eta_r) & 1-\eta_f \\ \hline \multicolumn{1}{c}{\eta_r} & \multicolumn{1}{c}{1-\eta_r} & \end{array}$$

where $\alpha_0 = [\eta_{fr}(\cdot,\cdot)(1-\eta_r-\eta_f+\eta_{fr}(\cdot,\cdot))]/[(\eta_f-\eta_{fr}(\cdot,\cdot))(\eta_r-\eta_{fr}(\cdot,\cdot))]$. From (2.1), it can be shown that

$$\eta_{fr}(\eta_f,\eta_r) = d-\operatorname{sgn}(\alpha_0-1)[d^2-\alpha_0(\alpha_0-1)^{-1}\eta_f\eta_r]^{\frac{1}{2}} \qquad \alpha_0 \neq 1$$
$$= \eta_f\eta_r \qquad \alpha_0 = 1,$$

where $d = [(\alpha_0-1)^{-1}+\eta_f+\eta_r]/2$.

Stage II: The parameters $\eta_f$ and $\eta_r$ represent guesses at the probability of favoring $p_f$ and the probability of responding $p_r$, respectively. We assume that the user is unable to make guesses at either probability, and therefore, $(\eta_f,\eta_r)$ is given a noninformative uniform distribution on the unit square.

The resulting prior density for $\underset{\sim}{p}$ is given by

$$(2.2) \qquad \pi_1(\underset{\sim}{p}) = \int_0^1\int_0^1 \pi_D(\underset{\sim}{p}|K,\underset{\sim}{\eta}^*)d\eta_f d\eta_r,$$

where $\underset{\sim}{\eta}^* = (\eta^*_{fr},\eta^*_{fs},\eta^*_{ur},\eta^*_{us})$ is the vector of prior means with configuration (2.1). To use the prior (2.2), the user need only specify two parameters $\alpha_0$ and K. As mentioned in section 1, the parameter $\alpha_0$ is a guess at the cross-product ratio of the table. The prior parameter K represents the precision of this guess. Typically, vague prior information will exist about the association structure in the table and a small positive value will be chosen for K.

If $\underset{\sim}{p}$ is given the Dirichlet $(\underset{\sim}{\eta},K)$ prior, then Kaufman and King (1973) show that the posterior density of $p_f$ can be represented by

$$(2.3) \qquad \pi_2(p_f|\underset{\sim}{x},\underset{\sim}{\eta}) = C\sum_{j=0}^{n_s} f_{Bb}(j|K\eta_{fs},K\eta_s,n_s) \cdot f_\beta(p_f|K\eta_f+n_{fr}+j,K\eta_u+n_{ur}+n_s-j),$$

where $f_\beta(\cdot|a,b)$ is the beta density with parameters a,b, $f_{Bb}(\cdot|a,b,c)$ is the beta-binomial mass function as defined in Raiffa and Schlaifer (1961), p. 218, C is a proportionality constant, and $\eta_s,\eta_f$ and $\eta_u$ are the prior means of $p_s$, $p_f$ and $p_u$ respectively. Using this representation, one can show that the posterior mean and variance of $p_f$ are given by

$$(2.4) \qquad E(p_f|\underset{\sim}{x},\underset{\sim}{\eta}) = \frac{n_{fr}+K\eta_f+n_s\eta_{fs}/\eta_s}{n+K}$$

and

$$(2.5) \qquad \mathrm{Var}(p_f|\underset{\sim}{x},\underset{\sim}{\eta}) = \frac{1}{n+K+1}E(p_f|\underset{\sim}{x},\underset{\sim}{\eta})(1-E(p_f|\underset{\sim}{x},\underset{\sim}{\eta})) + \frac{n_s}{(n+K)(n+K+1)}\frac{\eta_{fs}\eta_{us}}{\eta_s^2}\frac{(n_s+K\eta_s)}{(1+K\eta_s)} .$$

Note that (2.3) is the posterior distribution of $p_f$ conditional on a value of the prior mean $\underset{\sim}{\eta}$. If $\underset{\sim}{p}$ is given the two-stage prior (2.2), then the posterior distribution of $p_f$ is given by

$$(2.6)\quad \pi_4(p_f|\underset{\sim}{x}) = \int_0^1 \int_0^1 \pi_2(p_f|\underset{\sim}{x},\underset{\sim}{\eta}^*) \cdot \pi_3(\eta_f,\eta_r|\underset{\sim}{x})d\eta_f d\eta_r,$$

where

$$(2.7)\quad \pi_3(\eta_f,\eta_r|\underset{\sim}{x}) \propto \frac{\Gamma(K\eta_{fr}^*+n_{fr})\Gamma(K\eta_{ur}^*+n_{ur})\Gamma(K(1-\eta_r)+n_s)}{\Gamma(K\eta_{fr}^*)\quad\Gamma(K\eta_{ur}^*)\quad\Gamma(K(1-\eta_r))}.$$

Using conditioning arguments, we can use the representation (2.6) together with (2.4) and (2.5) to find posterior moments of $p_f$. To illustrate, the posterior mean of $p_f$ is given by

$$(2.8)\quad E(p_f|\underset{\sim}{x}) = E[E(p_f|\underset{\sim}{x},\eta_f,\eta_r)] = E\left[\frac{n_{fr}+K\eta_f+n_s\eta_{fs}^*/(1-\eta_r)}{n+K} \Big| \underset{\sim}{x}\right] = \frac{n_{fr}+KE(\eta_f|\underset{\sim}{x})+n_sE(\eta_{fs}^*/(1-\eta_r)|\underset{\sim}{x})}{n+K}.$$

Using similar techniques, the posterior variance of $p_f$ can be computed and is given by

$$(2.9)\quad \mathrm{Var}(p_f|\underset{\sim}{x}) = (n+K+1)^{-1}E(p_f|\underset{\sim}{x})(1-E(p_f|\underset{\sim}{x})) + (n+K)^{-1}(n+K+1)^{-1} \cdot [n_sE(\eta_{f\cdot s}^*(1-\eta_{f\cdot s}^*)\frac{n_s+K(1-\eta_r)}{1+K(1-\eta_r)}) + \mathrm{Var}(K\eta_f+n_s\eta_{f\cdot s}^*)],$$

where $\eta_{f\cdot s}^* = \eta_{fs}^*/(1-\eta_r)$.

The posterior mean (2.8) can be used as a point estimate of the probability $p_f$. A credible interval for $p_f$ can be easily developed by assuming that the central portion of the distribution is approximately normally distributed. With this assumption, a $(1-\gamma)100$ per cent credible interval is given by

$$(2.10)\quad (E(p_f|\underset{\sim}{x})-z_{\gamma/2}[\mathrm{Var}(p_f|\underset{\sim}{x})]^{\frac{1}{2}},\ E(p_f|\underset{\sim}{x})+z_{\gamma/2}[\mathrm{Var}(p_f|\underset{\sim}{x})]^{\frac{1}{2}}),$$

where $z_{\gamma/2}$ is the upper $\gamma/2$ percentage point of a standard normal distribution.

3. EXAMPLE. In this section, we will conduct a preliminary investigation of the behavior of the posterior distribution of $p_f$ (2.6). In particular, we will investigate the effect of one's choice of $\alpha_0$ and $K$ on the posterior mean and variance of $p_f$.

To begin, a few comments are necessary about the computation of the posterior quantities (2.8) and (2.9). These expressions all involve expectations using the posterior density of $(\eta_f,\eta_r)$ (2.7), which is not expressible in closed form. Thus it is necessary to compute expectations of the form

$$(3.1)\quad E[g(\eta_f,\eta_r)|\underset{\sim}{x}] = \frac{\int_0^1\int_0^1 g(\eta_f,\eta_r)\pi_3(\eta_f,\eta_r)d\eta_f d\eta_r}{\int_0^1\int_0^1 \pi_3(\eta_f,\eta_r)d\eta_f d\eta_r}$$

where $g$ is an arbitrary function of $\eta_f$ and $\eta_r$. One efficient way of computing the integrals in (3.1) uses the notion of importance sampling. First if $\alpha_0 = 1$ and the parameter $K$ approaches infinity, then it can be shown that

$$(3.2)\quad \lim_{K\to\infty} \pi_3(\eta_f,\eta_r|\underset{\sim}{x}) = \pi_L(\eta_f,\eta_r|\underset{\sim}{x}) = f_\beta(\eta_f|n_{fr}+1,n_{ur}+1)f_\beta(\eta_r|n_r+1,n_s+1),$$

a product of two beta densities. The limiting distribution (3.2) can serve as a rough approximation to $\pi_3(\eta_f,\eta_r)$ for values of $\alpha_0$ near one and moderate values of $K$. Next rewrite the expectation (3.1) as

$$(3.3)\quad E[g(\eta_f,\eta_r)|\underset{\sim}{x}] = \frac{\int_0^1\int_0^1 g(\eta_f,\eta_r)\left[\frac{\pi_3(\eta_f,\eta_r)}{\pi_L(\eta_f,\eta_r)}\right]\pi_L(\eta_f,\eta_r)d\eta_f d\eta_r}{\int_0^1\int_0^1 \left[\frac{\pi_3(\eta_f,\eta_r)}{\pi_L(\eta_f,\eta_r)}\right]\pi_L(\eta_f,\eta_r)d\eta_f d\eta_r}.$$

Finally, to approximate the integrals in (3.3) using simulation, $n_0$ values of $(\eta_f,\eta_r)$ are randomly generated from the beta densities in (3.2). Call the randomly generated values $(e_{fi},e_{ri})$, $i = 1,\ldots,n_0$. Then (3.3) is approximated by

$$(3.4)\quad \frac{\sum_{i=1}^{n_0} g(e_{fi},e_{ri})\pi_3(e_{fi},e_{ri})/\pi_L(e_{fi},e_{ri})}{\sum_{i=1}^{n_0} \pi_3(e_{fi},e_{ri})/\pi_L(e_{fi},e_{ri})}$$

In Table I, the posterior mean (2.8) and posterior variance (2.9) have been computed for the table of counts $\underset{\sim}{x} = (70,48,78)$ and for different values of the prior parameters $\alpha_0$ and $K$. In each case, the approximation (3.4) was used with $n_0 = 10000$ iterations. A few general observations can be made about these posterior estimates. First, note that the posterior mean of $p_f$ (2.8) can be expressed as a weighted mean of three terms:

$$(3.5)\quad E(p_f|\underset{\sim}{x}) = \frac{n_r}{n+K}\frac{n_{fr}}{n_r} + \frac{K}{n+K}E(\eta_f|\underset{\sim}{x}) + \frac{n_s}{n+K}E(\eta^*_{f\cdot s}|\underset{\sim}{x}).$$

The parameter $\eta^*_{f\cdot s}$ represents the user's prior guess at $p_{f\cdot s}$, the probability a nonrespondent is in favor, and the posterior expectation $E(\eta^*_{f\cdot s}|\underset{\sim}{x})$ appears to roughly satisfy the equality

$$\frac{(n_{fr}/n_f)/(1-n_{fr}/n_f)}{E(\eta^*_{f\cdot s}|\underset{\sim}{x})/(1-E(\eta^*_{f\cdot s}|x))} = \alpha_0.$$

The posterior mean (3.5) appears to use the $n_s$ nonrespondent counts by first partitioning them into the (fs), (us) cells so that the cross-product ratio of the 2×2 table of counts is $\alpha_0$, and then pooling the counts in the favorable cells to estimate $p_f$. A second general comment is that the posterior variance of $p_f$ appears to be a decreasing function of K and $|\alpha_0 - 1|$. These posterior variance values can be contrasted in size with the variance of the "usual" estimate $n_{fr}/n_r$, which is equal to $(n_{fr}/n_r)(1-n_{fr}/n_r)/n_r = .00205$. Typically, the user will possess vague prior beliefs about the differences between respondents and nonrespondents and thus will choose a small value of K. As indicated in Table I, this choice of K will result in a much larger posterior variance than the classical variance .00205. Thus, in this brief example, the posterior distribution appears to reflect both the belief in association (inputted through $\alpha_0$) and the strength of the belief in the association (inputted through K).

TABLE I

Computed Values of Posterior Means and Posterior Variances. $\underset{\sim}{x} = (70,48,78)$

| | $\alpha_0$ .06 | .25 | 1 | 4 | 16 |
|---|---|---|---|---|---|
| K = 4 | | | | | |
| $E(p_f\|\underset{\sim}{x})$ | .702 | .648 | .592 | .520 | .467 |
| $Var(p_f\|\underset{\sim}{x})$ | .0078 | .0125 | .0167 | .0145 | .0159 |
| K = 25 | | | | | |
| $E(p_f\|\underset{\sim}{x})$ | .716 | .665 | .592 | .514 | .420 |
| $Var(p_f\|\underset{\sim}{x})$ | .0021 | .0041 | .0053 | .0047 | .0042 |
| K = 100 | | | | | |
| $E(p_f\|\underset{\sim}{x})$ | .713 | .682 | .591 | .500 | .414 |
| $Var(p_f\|\underset{\sim}{x})$ | .0010 | .0018 | .0028 | .0022 | .0012 |

REFERENCES


1. Albert, J. H., and A. K. Gupta (1982). Bayesian estimation methods for 2×2 contingency tables using mixtures of Dirichlet distributions. Department of Mathematics and Statistics, Bowling Green State University, Technical Report.
2. Birnbaum, Z. W., and M. G. Sirken (1950). Bias due to nonavailability in sample surveys. *J. Amer. Statist. Assoc.*, 45, 98-111.
3. Hansen, M. H., and W. N. Hurwitz (1946). The problem of non-response in sample surveys. *J. Amer. Statist. Assoc.*, 41, 517-529.
4. Kaufman, G. M., and B. King (1973). A Bayesian analysis of nonresponse in dichotomous processes. *J. Amer. Statist. Assoc.*, 68, 670-678.
5. Raiffa, H., and R. O. Schlaifer (1961). *Applied Statistical Decision Theory*, Boston: Harvard Business School.