453

The
British
Psychological
Society

www.bpsjournals.co.uk

Beta-binomial ANOVA for multivariate randomized response data

Jean-Paul Fox*

Twente University, Enschede, The Netherlands

There is much empirical evidence that randomized response methods improve the cooperation of the respondents when asking sensitive questions. The traditional methods for analysing randomized response data are restricted to univariate data and only allow inferences at the group level due to the randomized response sampling design. Here, a novel beta-binomial model is proposed for analysing multivariate individual count data observed via a randomized response sampling design. This new model allows for the estimation of individual response probabilities (response rates) for multivariate randomized response data utilizing an empirical Bayes approach. A common beta prior specifies that individuals in a group are tied together and the beta prior parameters are allowed to be cluster-dependent. A Bayes factor is proposed to test for group differences in response rates. An analysis of a cheating study, where 10 items measure cheating or academic dishonesty, is used to illustrate application of the proposed model.

1. Introduction

When observing count data, it is often assumed that individual counts are generated from a binomial distribution. If, however, the counts exhibit extraneous variance, variance greater than expected under a binomial model, it is further assumed that the binomial probabilities vary between individuals according to a beta distribution. The marginal distribution of the counts is then beta-binomial. The beta-binomial model for psychological and educational testing was proposed by Lord (1965). The binomial probability function for describing a respondent's number-correct score is justified when each response is independent of the other, and when the respondent's response rate, the probability of a positive response, remains constant. In mental test theory where tests are usually measures of maximum performance, it is not to be expected that the items are of equal difficulty, which makes the binomial model unsatisfactory. However, items measuring an individual's interest, attitudes, or a specific type of

* Correspondence should be addressed to Jean-Paul Fox, Department of Research Methodology, Measurement and Data Analysis, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands (e-mail: fox@edte.utwente.nl).

behaviour (e.g. cheating or criminal behaviour) are often descriptive statements with which respondents agree or disagree. In personality assessment via questionnaires and self-report inventories which is focused on assessing an individual's interests, motives, or attitudes, it is likely that the individual's probability of a positive response remains constant. Domain responses relative to a respondent can be assumed as being more or less consistent. Further, it will be shown via a simulation study that the beta-binomial model is robust against violations of a constant individual response probability. The beta-binomial model has tractable mathematical properties and has proven to be a good descriptive model (Lord & Novick, 1968). Modifications of the beta-binomial model have been developed for analysing random guessing on multiple choice tests (Morrison & Brockway, 1979) and estimating domain scores (Lin & Hsiung, 1994), among others.

A particular problem is that respondents have a tendency to agree rather than disagree (acquiescence) and a tendency to give socially desirable answers (social desirability). Moreover, measuring incriminating or socially undesirable practices via direct questioning of respondents leads to some degree of evasiveness or non-cooperation. Obtaining valid and reliable information depends on the cooperation of the respondents, and the willingness of the respondents depends on the confidentiality of their responses. Warner (1965) developed a data collection procedure, the randomized response (RR) technique, in which a randomizing device is used to select a question from a group of questions and the respondent answers the selected question. The respondent is protected since the interviewer will not know which question is being answered. In this article, a related approach, a forced randomized response design, is used in which the randomized device determines whether the respondent is forced to say 'yes', say 'no', or answer the sensitive question. For example, in the study, described below, concerning cheating behaviour of students at a Dutch university, two dice were used. The respondents was asked to roll two dice and answer 'yes' if the sum of the outcomes was 2, 3, or 4, answer the sensitive question if the sum was between 4 and 11, and answer 'no' if the sum was 11 or 12. Again, the respondents were protected since the interviewer did not know the outcome of the dice.

In this paper, the traditional method (Warner, 1965; Greenberg, Abul-Ela, Simmons, & Horvitz, 1969) for analysing RR data is extended to handle multivariate RR data such that inferences are not limited to estimating population properties. Note that, up till now, there has been no straightforward method for analysing multivariate RR data that enables the computation of individual response estimates and corresponding variances without having to rely on large-scale survey data. A challenge in the analysis of RR data is that the true individual responses (that would have been observed via direct questioning) are masked due to the forced responses. However, individual response rates can be estimated when multiple RR observations are measured from each individual. First, the sum of randomized responses is modelled with a beta-binomial model. Second, a Bayes estimate of the individual's response rate and its variance is obtained by utilizing a probabilistic relationship between the randomized response and the response that would have been obtained via direct questioning. Different groups are modelled simultaneously in a common way, and it is shown how a Bayes factor can be used to test for group differences regarding the response rates, taking account of the RR sampling design. Below, for example, interest is focused on differences in cheating behaviour across faculties where a forced response sampling design is used.

Attempts have been made to modify item response theory (IRT) models for estimating an underlying construct given RR data. Fox (2005) developed a class of randomized IRT models within a Bayesian framework. Independently, Böckenholt and van der Heijden (2007) developed a comparable class of models within a frequentist framework. These types of models are complex, with many parameters, since all items and persons are parameterized. Stable parameter estimates are only to be expected for large data sets (e.g. Embretson & Reise, 2000, suggest the use of more than 500 respondents). Moreover, response patterns obtained via a randomized response sampling design contain less information about the underlying construct than response patterns obtained via direct questioning. Larger sample sizes are needed to obtain parameter estimates of underlying constructs with the same precision as those obtained via direct questioning data. There is relatively little information about the robustness of these IRT models for RR data, the computer algorithms for fitting them, and so far it is unknown how sensitive the models are to violation of the various assumptions. The main advantages of the proposed beta-binomial model are (1) its simplicity, (2) that stable parameter estimates can be obtained for small data sets, and (3) that no complex estimation methods are needed.

This paper is organized as follows. In section 2, the beta-binomial model is described for RR data and it is shown how RR data affect statistical inferences. Then, attention is focused on estimating the parameters of the model via parametric empirical Bayes. A closed-form expression is obtained for the Bayes risk of a Bayes estimator for an individual response rate. It is shown how to construct confidence intervals and to estimate probability statements with respect to a response rate. In section 5, a simulation study is given where (1) the robustness of the beta-binomial model is investigated, (2) the sensitivity of the proposed Bayes factors to hyperprior parameter values is shown, and (3) a risk comparison between the proposed Bayes estimator and a natural unbiased estimator is shown. An example is presented in which RR data from a cheating study in The Netherlands are used to illustrate the methodology. Finally, other extensions of the model are discussed.

2. The beta-binomial model

Here there are J groups, and participant i in group j has response probability or response rate p_{ij}. It is assumed that each person responds to $k = 1, \ldots, n_{ij}$ binary items. A random variable u_{ijk} is Bernoulli distributed with response probability p_{ij}. The random variable $u_{ij\cdot} = \sum_{k}^{n_{ij}} u_{ijk}$, the sum of independent Bernoulli trials, has binomial distribution with parameters n_{ij} and p_{ij}. This probability varies from respondent to respondent and has a beta distribution with group-specific parameters α_j and β_j. The beta-binomial hierarchy models the variation in individual responses via a binomial distribution, and models the variation between respondents' success probabilities via a beta distribution, that is,

$$u_{ij\cdot} | p_{ij} \sim \mathcal{BIN}(n_{ij}, p_{ij}),$$

$$p_{ij} \sim \mathcal{B}(\alpha_j, \beta_j).$$

This structure allows the conditional mean and variance of the individual success probability to vary across respondents and clusters. The posterior expectation and

variance of p_{ij} can be derived using Bayesian methodology;

$$\pi(p_{ij}|u_{ij.}, \alpha_j, \beta_j) = \frac{f(u_{ij.}|p_{ij})\pi(p_{ij}|\alpha_j, \beta_j)}{\int f(u_{ij.}|p_{ij})\pi(p_{ij}|\alpha_j, \beta_j)dp_{ij}}$$

$$= \frac{\Gamma(n_{ij} + \alpha_j + \beta_j)}{\Gamma(\alpha_j + u_{ij.})\Gamma(n_{ij} + \beta_j - u_{ij.})} p_{ij}^{u_{ij}+\alpha_j-1}(1 - p_{ij})^{n_{ij}-u_{ij.}+\beta_j-1},$$

which can be recognized as a beta distribution with parameter $u_{ij.} + \alpha_j$ and $n_{ij} - u_{ij.} + \beta_j$. The mean and variance of this beta distribution are

$$\mathcal{E}(p_{ij}|u_{ij.}, \alpha_j, \beta_j) = \frac{u_{ij.} + \alpha_j}{n_{ij} + \alpha_j + \beta_j},$$

$$\mathcal{V}(p_{ij}|u_{ij.}, \alpha_j, \beta_j) = \frac{(u_{ij.} + \alpha_j)(n_{ij} - u_{ij.} + \beta_j)}{(n_{ij} + \alpha_j + \beta_j + 1)(n_{ij} + \alpha_j + \beta_j)^2}.$$

The binary response data $\mathbf{u}$ are not observed, but RR data $\mathbf{y}$ are observed via a forced randomized response design. In this sampling design, a response u_{ijk} is given to a sensitive question k with probability φ_1, and a forced positive response is given with probability $(1 - \phi_1)\phi_2$. A probabilistic relationship can be specified that relates the observed randomized response y_{ijk} with the response u_{ijk}:

$$\pi(y_{ijk}|p_{ij}) = \phi_1\pi(u_{ijk}|p_{ij}) + (1 - \phi_1)\phi_2$$

$$= \phi_1 p_{ij} + (1 - \phi_1)\phi_2$$

$$= \Delta(p_{ij}),$$

where $\Delta(p_{ij})$ is a linear function with known parameters ϕ_1 and ϕ_2 and with inverse function $\Delta^-(p_{ij})$. It follows that for each respondent the sum of the randomized outcomes, $y_{ij.}$, of the n_{ij} independent Bernoulli trials has the binomial distribution,

$$y_{ij.}|p_{ij} \sim \mathcal{BIN}(n_{ij}, \Delta(p_{ij})), \tag{1}$$
$$\Delta(p_{ij}) \sim \mathcal{B}(\alpha_j, \beta_j),$$

using a beta prior distribution for the success probabilities $\Delta(p_{ij})$ with group-specific parameters α_j and β_j. The beta distribution describes the variation in the individual success probabilities of the binomial distribution within each cluster. It follows that

$$\mathcal{E}(p_{ij}|\alpha_j, \beta_j) = \Delta^-(\mu_j) = \Delta^-\left(\frac{\alpha_j}{\alpha_j + \beta_j}\right) = \phi_1^{-1}\frac{\alpha_j}{\alpha_j + \beta_j} + (1 - \phi_1^{-1})\phi_2,$$

$$\mathcal{V}(p_{ij}|\alpha_j, \beta_j) = \frac{1}{\phi_1^2}\left(\frac{\mu_j(1 - \mu_j)}{\alpha_j + \beta_j + 1}\right) = \mu_j(1 - \mu_j)\sigma_j/\phi_1^2$$

are the prior mean and variance of p_{ij}. The prior mean of p_{ij} is a weighted average of the prior mean, $\alpha_j/(\alpha_j + \beta_j)$, and the forced success probability ϕ_2. There are no

randomized responses if $\phi_1 = 0$ (the prior mean equals $\alpha_j/(\alpha_j + \beta_j)$), and there are only randomized responses if $\phi_1 = 0$ (the prior mean equals ϕ_2). In the present application, the beta distribution has zero density at 0 and 1. This is realized with $\alpha_j > 1$ and $\beta_j > 1$, and as a result $\sigma_j < 1/3$. Note that the prior variance will increase due to the RR sampling design, since $\phi_1 \in (0,1)$. The case of only binomial variation corresponds to $\sigma_j = 0$, where α_j and β_j have infinite values.

3. Empirical Bayes parameter estimation

The parameters of the two-stage model (1) are estimated using an empirical Bayes approach. In the empirical Bayes analyses, the parameters at the highest level of the hierarchy are estimated using the data. That is, there is no hyperprior and the data are used to provide information about the highest level in the hierarchy. In a parametric empirical Bayes approach (e.g. Casella, 1985; Morris, 1983), the parameters of the beta distribution are estimated using the marginal posterior distribution of the data, $p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\beta})$. An empirical Bayes estimator of the individual success probabilities is constructed by replacing these quantities by their estimates in the estimator. That is, $\varepsilon(p_{ij}|y_{ij.}, \hat{\alpha}_j, \hat{\beta}_j)$ is used to estimate $\varepsilon(p_{ij}|y_{ij.}, \alpha_j, \beta_j)$.

The marginal distribution of the prior parameters α_j, β_j is given by

$$\pi(y_{1j.}, \ldots, y_{Ij.})|\alpha_j, \beta_j) = \prod_i \int f(y_{ij.}|\Delta(p_{ij}))\pi(\Delta(p_{ij})|\alpha_j, \beta_j)dp_{ij}$$

$$= \prod_i \binom{n_{ij}}{y_{ij}} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)} \Delta(p_{ij})^{y_{ij.}+\alpha_j-1}(1 - \Delta(p_{ij}))^{n_{ij}-y_{ij.}+\beta_j-1}$$

$$= \prod_i \binom{n_{ij}}{y_{ij.}} \frac{\Gamma(\alpha_j + y_{ij.})\Gamma(n_{ij} + \beta_j - y_{ij.})\Gamma(\alpha_j + \beta_j)}{\Gamma(n_{ij} + \alpha_j + \beta_j)} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)} \tag{2}$$

and can be recognized as the beta-binomial distribution (Gelman, Carlin, Stern, & Rubin, 1995, p. 476). Note that, for each j, the marginal distributions of the $y_{ij.}$s, after integrating out the p_{ij}s, are identically distributed with parameters α_j and β_j if $n_{ij} = n_j$. These parameters can be estimated from equation (2).

There are two simple estimation methods for estimating each α_j and β_j. The method of moments, one of the oldest methods of finding point estimators (Casella & Berger, 2002, Chap. 7), provides closed-form expressions for the estimators. The first two sample moments are equated to the mean and variance of the beta-binomial distribution in equation (2) (Skellam, 1947). The moment estimators are found by solving the equations, that is,

$$\mathcal{E}(y_{ij.}|\alpha_j, \beta_j) = n_j \frac{\alpha_j}{\alpha_j + \beta_j} = \bar{y}_j,$$

$$\mathcal{V}(y_{ij.}|\alpha_j, \beta_j) = n_j \frac{\alpha_j\beta_j}{(\alpha_j + \beta_j)^2} \frac{(\alpha_j + \beta_j + n_j)}{(\alpha_j + \beta_j + 1)} = s_j^2,$$

where $\bar{y}_j$ and s_j^2 are the sample moments. It follows that the estimators are

$$\hat{\alpha}_j = \bar{y}_j \frac{(\bar{y}_j(n_j - \bar{y}_j) - s_j^2}{(n_j s_j^2 - \bar{y}_j(n_j - \bar{y}_j))},$$

$$\hat{\beta}_j = (n - \bar{y}_j) \frac{(\bar{y}_j(n_j - \bar{y}_j) - s_j^2)}{(n_j s_j^2 - \bar{y}_j(n_j - \bar{y}_j))}.$$

Maximizing the marginal likelihood for each j where the number of observations may vary across individuals yields the more efficient maximum-likelihood estimates. Ignoring constants involving only observations, the log likelihood, following from equation (2), is given by

$$l(\alpha_j, \beta_j | y_{ij\cdot}) = \sum_i \left[\sum_{k=0}^{y_{ij\cdot}-1} \log(\alpha_j + k) + \sum_{k=0}^{n_{ij}-y_{ij\cdot}-1} \log(\beta_j + k) - \sum_{k=0}^{n_{ij}-1} \log(\alpha_j + \beta_j + k) \right].$$
(3)

There are no closed-form expressions for the maximum-likelihood estimators. However, the equations of first-order derivatives equated to 0 can be solved iteratively using the Newton–Raphson method. Griffiths (1973) suggested estimating the parameters $\xi_j = \alpha_j/(\alpha_j + \beta_j)$ and $\omega_j = 1/(\alpha_j + \beta_j)$ since these parameters are more stable than α_j and β_j. The method-of-moments estimates and the maximum-likelihood estimates are, in most cases, nearly the same. However, on rare occasions, the method of moments gives poor results and, therefore, maximum-likelihood estimates are preferred (Wilcox, 1981). The method of moments has the advantage of yielding explicit answers, and they can also be used as starting values for obtaining maximum-likelihood estimates.

The posterior distribution of success probability, $\Delta(p_{ij})$ is $\mathcal{B}(y_{ij\cdot} + \alpha_j, \beta_j - y_{ij\cdot} + n_{ij})$. A natural estimator for the response rate, p_{ij}, is the mean of the posterior distribution. This gives the Bayes estimator, where the estimates of α_j and β_j are plugged in,

$$\mathcal{E}(p_{ij} | y_{ij\cdot}, \hat{\alpha}_j, \hat{\beta}_j) = \Delta^-(\mu_j) = \Delta^- \left[\frac{y_{ij\cdot} + \hat{\alpha}_j}{n_{ij} + \hat{\alpha}_j + \hat{\beta}_j} \right]$$

$$= \phi_1^{-1} \left[\left(\frac{n_{ij}}{n_{ij} + \hat{\alpha}_j + \hat{\beta}_j} \right) \frac{y_{ij\cdot}}{n_{ij}} + \left(\frac{\hat{\alpha}_j + \hat{\beta}_j}{n_{ij} + \hat{\alpha}_j + \hat{\beta}_j} \right) \frac{\hat{\alpha}_j}{\hat{\alpha}_j + \hat{\beta}_j} \right] + (1 - \phi_1^{-1})\phi_2, \quad (4)$$

with variance

$$\mathcal{V}(p_{ij} | y_{ij\cdot}, \hat{\alpha}_j, \hat{\beta}_j) = \frac{(y_{ij\cdot} + \hat{\alpha}_j)(n_{ij} - y_{ij\cdot} + \hat{\beta}_j)}{\phi_1^2(n_{ij} + \hat{\alpha}_j + \hat{\beta}_j + 1)(n_{ij} + \hat{\alpha}_j + \hat{\beta}_j)^2} \quad (5)$$

$$= \mu_j(1 - \mu_j)\sigma_j/\phi_1^2. \quad (6)$$

The Bayes estimate in (4) combines three kinds of information about p_{ij}. The prior distribution with mean $\hat{\alpha}_j/(\hat{\alpha}_j + \hat{\beta}_j)$ is combined with the sample mean $y_{ij\cdot}/n_{ij}$, where the weights are determined by n_{ij}, $\hat{\alpha}_j$ and $\hat{\beta}_j$. This weighted average is combined with the forced success probability ϕ_2 where the weight is defined by ϕ_1. As a result, the

Bayes estimate is a linear combination of the prior mean, sample mean, and the forced success probability.

The parametric empirical Bayes estimate fails to take account of the uncertainty about α_j and β_j. Therefore, the corresponding variance term in (6) is too small, and this term estimates only the variance $\mathcal{E}(\mathcal{V}(p_{ij}|y_{ij\cdot}, \hat{\alpha}_j, \hat{\beta}_j))$. Kass and Steffey (1989) proposed a first-order approximation of the term $\mathcal{V}(\mathcal{E}(p_{ij}|y_{ij\cdot}, \alpha_j, \beta_j))$, where $\mathcal{V}$ is the variance with respect to the posterior distribution of α_j, β_j. By writing $\mathcal{V}(p_{ij}|y_{ij\cdot}) = \mathcal{E}(\mathcal{V}(p_{ij}|y_{ij\cdot}, \alpha_j, \beta_j)) + \mathcal{V}(\mathcal{E}(p_{ij}|y_{ij\cdot}, \alpha_j, \beta_j))$ and applying first-order Taylor expansions, straightforward calculation then yields an approximation for the empirical Bayes variance under the model:

$$\hat{\mathcal{V}}(p_{ij}|y_{ij\cdot}) = \mathcal{E}(\mathcal{V}(p_{ij}|y_{ij\cdot}, \hat{\alpha}_j, \hat{\beta}_j)) + \hat{\mathcal{V}}(\mathcal{E}(p_{ij}|y_{ij\cdot}, \alpha_j, \beta_j))$$

$$= \mu_j(1 - \mu_j)\sigma_j/\phi_1^2 + \sum_{c,d}\hat{\sigma}_{c,d}\hat{\delta}_c\hat{\delta}_d/\phi_1^2, \tag{7}$$

where $\hat{\sigma}_{c,d}$ is the (c, d)th component of the negative Hessian of $l(\alpha_j, \beta_j|y_{ij\cdot})$, equation (3), and with $\hat{\delta}_c = (\partial/\partial\alpha_j)\mathcal{E}(p_{ij}|y_{ij\cdot}, \alpha_j, \beta_j)$ and $\hat{\delta}_d = (\partial/\partial\beta_j)\mathcal{E}(p_{ij}|y_{ij\cdot}, \alpha_j, \beta_j)$ evaluated at $\alpha_j = \hat{\alpha}_j$ and $\beta_j = \hat{\beta}_j$. Note that the accuracy of the approximation of the posterior distribution of α_j, β_j based on the normal distribution depends on the number of observations within each cluster, I_j, for $j = 1, \ldots, J$, rather than the number of observations, n_{ij}, per individual. When the number of observations per cluster becomes sufficiently large, with n_{ij} remaining small, the first term in (7) will suffice. The accuracy of the approximation can be improved by restricting the variance terms across clusters to be equal.

For a squared error loss function, a closed-form expression can be found of Bayes risk of the Bayes estimator $\delta(y) = \mathcal{E}(p_{ij}|y_{ij\cdot}, \hat{\alpha}_j, \hat{\beta}_j)$; see equation (4). The Bayes risk, defined as the expected posterior risk (the mean squared error) with respect to the marginal distribution of the data of estimator $\delta(y)$, can be written as

$$\mathcal{E}_y[\mathcal{E}_{p|y}(\delta(y) - p_{ij})^2] = \mathcal{E}_y[\mathcal{E}_{p|y}(\mathcal{E}(p_{ij}|y_{ij\cdot}, \hat{\alpha}_j, \hat{\beta}_j) - p_{ij})^2]$$

$$= \mathcal{E}_y[\mathcal{V}(p_{ij}|y_{ij\cdot}, \hat{\alpha}_j, \hat{\beta}_j)]$$

$$= \frac{\mathcal{E}_y[(y_{ij\cdot} + \hat{\alpha}_j)(n_{ij} - y_{ij\cdot} + \hat{\beta}_j)]}{\phi_1^2(n_{ij} + \hat{\alpha}_j + \hat{\beta}_j + 1)(n_{ij} + \hat{\alpha}_j + \hat{\beta}_j)^2} \tag{8}$$

$$= \hat{\alpha}_j\hat{\beta}_j/\phi_1^2(\hat{\alpha}_j + \hat{\beta}_j)(\hat{\alpha}_j + \hat{\beta}_j + 1)(\hat{\alpha}_j + \hat{\beta}_j + n_{ij}),$$

where the variance of the Bayes estimator is given in equation (5). Details of the computation of the expected value in the numerator of (8) can be found in Grosh (1972). The risk of the Bayes estimator will be compared by simulation with that of unbiased estimator $\Delta^-(y_{ij\cdot}/n_{ij})$.

4. Bayesian inference

The posterior distribution of $\Delta(p_{ij})$ can be used to construct a Bayesian credible interval for p_{ij}. The posterior distribution of $\Delta(p_{ij})$ is $\mathcal{B}(\tilde{\alpha}_j, \tilde{\beta}_j)$, where $\tilde{\alpha}_j = y_{ij\cdot} + \alpha_j$ and

$\tilde{\beta}_j = \beta_j - y_{ij\cdot} + n_{ij\cdot}$. A Bayesian $1 - 2v$ credible interval equals

$$1 - 2v = \pi(p_L \le \Delta(p_{ij}) \le p_U|y) \tag{9}$$

$$= \pi(\Delta^-(p_L) \le p_{ij} \le \Delta^-(p_U)|y),$$

where $p_L = 0$ if $\tilde{\alpha}_j \le 1$ and $p_U = 1$ if $\tilde{\beta}_j \le 1$. The computation of p_L and p_U requires the evaluation of incomplete beta functions and an algorithm for finding the roots in (9). This can be circumvented using the property $\Delta(p_{ij})/(1 - \Delta(p_{ij})) \sim \tilde{\alpha}_j/\tilde{\beta}_j F_{2\tilde{\alpha}_j, 2\tilde{\beta}_j}$, and a $1 - 2v$ credible interval for p_{ij} is

$$\Delta^-\left(\frac{1}{1 + (\tilde{\beta}_j + 1)(\tilde{\alpha}_j - 1)F_{2(\tilde{\beta}_j+1),\, 2(\tilde{\alpha}_j-1), v/2}}\right) \le p_{ij} < \Delta^-\left(\frac{(\tilde{\alpha}_j/\tilde{\beta}_j)F_{2\tilde{\alpha}_j, 2\tilde{\beta}_j, v/2}}{1 + (\tilde{\alpha}_j/\tilde{\beta}_j)F_{2\tilde{\alpha}_j, 2\tilde{\beta}_j, v/2}}\right),$$

where $F_{\alpha_j, \beta_j, v}$ is the upper v cut-off from an F-distribution with α_j and β_j degrees of freedom. The lower end-point is 0 if $y_{ij\cdot} = 0$ and the upper end-point is 1 if $y_{ij\cdot} = n_{ij\cdot}$. In the same way, the posterior probability that p_{ij} does not exceed some fixed value p_0 can be computed, that is,

$$\pi(p_{ij} \le p_0|y) = \pi(\Delta(p_{ij}) \le \Delta(p_0)|y)$$

$$= \pi\left(\Delta(p_{ij}) \le \frac{(\tilde{\alpha}_j/\tilde{\beta}_j)F}{1 + (\tilde{\alpha}_j/\tilde{\beta}_j)F}|y\right)$$

$$= \pi(F_{2\tilde{\alpha}_j, 2\tilde{\beta}_j} \le F|y),$$

where:

$$F = \frac{\tilde{\beta}_j}{\tilde{\alpha}_j}\frac{\Delta(p_0)}{1 - \Delta(p_0)}.$$

4.1. Homogeneity of proportions

There are J groups and each group has $I_j (j = 1, \ldots, J)$ respondents. Attention is focused on differences in latent response rates across groups. A reparameterization as suggested by Griffiths (1973), $\xi_j = \alpha_j/(\alpha_j + \beta_j)$ and $\omega_j = 1/(\alpha_j + \beta_j)$, is preferable. These parameters are more easily interpreted, with ξ_j, the mean success probability and ω_j a measure of variation in response probabilities in cluster j.

In this parameterization, the beta-binomial distribution becomes the binomial distribution when ω_j approaches 0 and this makes it possible to test for the extra-binomial variation. The Bayes factor can be used to test for the extra variability beyond binomial variance in each group j. The Bayes factor for testing H_0: $\omega_j = 0$ against H_1: $\omega_j > 0$ is written as

$$BF = \frac{P(y_j|H_0)}{P(y_j|H_1)} = \frac{\int \prod_{i|j} f(y_{ij\cdot}|\xi_j)\pi(\xi_j)d\xi_j}{\iint \prod_{i|j} f(y_{ij\cdot}|\xi_j, \omega_j)\pi(\xi_j, \omega_j)d\xi_j\, d\omega_j}. \tag{10}$$

Careful attention must be paid to the prior choices $\pi(\xi_j, \omega_j)$ and $\pi(\xi_j)$ since equation (10) is a test that a variance component lies on the boundary of its parameter space. Hsiao (1997) showed that the parameters ξ_j and ω_j are null orthogonal and that the parameters can be considered to be independent. A uniform prior distribution is

assumed for ξ_j and a half-normal unit information prior is assumed for ω_j centred at 0 with variance equal to the inverse of a unit information in group j evaluated at the null for the binomial case (Pauler, Wakefield, & Kass, 1999).

A Bayes factor can also be used to test the hypothesis $H_0 : \xi = \xi_1 = \ldots = \xi_j$ against the alternative that not all location parameters are the same. The Bayes factor for testing heterogeneity of population proportions between groups and allowing heterogeneity of variance between groups is written as

$$BF = \frac{\iint \prod_j f(y_{j \cdot}|\xi, \omega_j)\pi(\xi)\pi(\omega_j)d\xi \, d\omega_j}{\iint \prod_j f(y_{j \cdot}|\xi_j, \omega_j)\pi(\xi_j)\pi(\omega_j)d\xi_j \, d\omega_j}. \tag{11}$$

The same prior distributions can be used for ξ_j and ω_j terms, and a uniform prior is assumed for ξ. The Bayes factor in (11) is easily adjusted when assuming homogeneity of variance between groups, since it is possible that the groups may differ notably with respect to ξ_j but not to ω_j.

5. Simulation study

In a first simulation study, the robustness of the model was investigated. A second simulation study was performed in order to compare the risks of the proposed moment and maximum-likelihood estimators.

5.1. Robustness

The effects of a violation of the assumption of a constant response rate per individual across items were investigated. For each respondent, two response rates were simulated from a beta distribution, and each response rate was used to generate binomial distributed response data based on $n/2$ items. Within individuals, the response rates were allowed to vary in such a way that the approximately normally distributed differences had a mean of 0 with, under condition 1, a variance of 0.05 and, under condition 2, a variance of 0.10. In the so-called 'no-noise' condition, both data sets of $n/2$ items were analysed separately. In the other two conditions, the scores on both tests were summed to create one score based on n items. Subsequently, these scores for n items were analysed given the assumption of a constant individual response rate.

In Table 1, the results are given under the heading 'constant response rate'. Under each condition, the maximum-likelihood estimates of the beta parameters are reported under the heading ML. Furthermore, a mean squared error (MSE) of the estimated response rates was computed such that the estimated individual response rate(s) were compared with the true individual response rates. This means that in conditions 1 and 2 both individual response rates were estimated by an overall estimate based on the summed score. All estimates in Table 1 are averaged outcomes over 100 independent samples. Under the no-noise condition, the estimated beta parameters are close to the true values. The estimated beta parameters under conditions 1 and 2 are slightly biased due to the fact that the mean values of the generated beta-distributed response rates were not exactly beta distributed. The bias increases when the differences between individual response rates increase. It follows that a more informative prior leads to a lower MSE. The estimated MSEs in condition 1 are smaller than the estimated MSEs in the no-noise condition. Moreover, the estimated MSEs in condition 2 are also smaller

Table 1. Robustness of the beta-binomial model against different within-person response rates

(N,n)	(α, β)	No noise		Condition 1		Condition 2	
		ML	MSE_P	ML	MSE_P	ML	MSE_P
Constant response rate							
(100,10)	α = 2	2.028	0.047	2.064	0.032	2.292	0.036
	β = 2	2.034		2.071		2.284	
	α = 4	4.143	0.034	4.157	0.027	4.289	0.029
	β = 4	4.041		4.291		4.333	
(200,20)	α = 2	2.017	0.029	2.165	0.020	2.210	0.022
	β = 2	2.058		2.218		2.210	
	α = 4	4.042	0.024	4.479	0.017	4.640	0.021
	β = 4	4.071		4.307		4.380	
Beta prior							
(100,10)	α = 2	2.017	0.014	1.899	0.015	1.655	0.018
	β = 2	2.103		2.120		2.037	
	α = 4	4.062	0.012	3.980	0.014	3.585	0.016
	β = 4	4.081		4.472		4.300	
(200,20)	α = 2	2.014	0.008	1.776	0.009	1.622	0.013
	β = 2	2.026		2.010		2.026	
	α = 4	4.078	0.008	3.812	0.008	3.500	0.012
	β = 4	4.075		4.409		4.265	

than those in the no-noise condition. It can be concluded that the reduction in variance is larger than the increase in bias due to a violation of a constant response rate. The reduction in variance is caused by the fact that in conditions 1 and 2 more item information is available for estimating the response rates than in the no-noise condition. For the cases considered, the reduction in bias by taking account of different individual response rates across items has a smaller impact on the MSE than the reduction in variance by assuming a constant response rate. The differences between MSEs become smaller when increasing the number of items and persons, and only for large sample sizes does it become attractive to allow for different individual response rates.

The robustness to violations of the assumption of beta-distributed response rates was investigated. Therefore, two symmetric beta distributions for the response rates were specified, with $\alpha = \beta = 2$ and $\alpha = \beta = 4$. The generated response rates were contaminated with noise. Subsequently, binomial response data were generated given the noisy response rates. The noise was generated under two different conditions. Random noise was generated from a truncated normal distribution with a standard deviation of 0.2, denoted as condition 1, and 0.4, denoted as condition 2. In Table 1, it can be seen, under the heading 'Beta prior', that the estimated beta population parameters resemble the true values in the case of no noise. The estimated population parameters under the other two conditions are slightly biased. Although the simulated response rates do not follow the assumed beta prior distribution under conditions 1 and 2, even for small sample sizes both estimated beta prior parameters are close to the true values. It can be seen that the MSE of the estimated response rates given the estimated beta parameters is just slightly increasing when increasing the noise level. In conclusion, the model is robust against random disturbances in response rates since they hardly influence the estimates of beta parameters and response rates.

The robustness of the model was further examined by investigating its sensitivity to different item difficulties. Therefore, response data were generated under the Rasch model with item difficulties equal to 0 (no-noise condition), and item difficulties generated from a normal distribution with mean 0 and standard deviation 0.25 (condition 1) and standard deviation 0.50 (condition 2). A sum score was computed for each simulated item response vector, and response rates were estimated given the sum scores. Besides different item difficulties, a second model violation was introduced since the true simulated response rates were not beta but logistic distributed according to the Rasch model. The estimated response rates were compared with the true simulated item response probabilities at the item level by computing an MSE of the estimated response rates. In Table 2, the estimated MSEs based on 100 independent samples are given. The estimated MSEs in the no-noise condition in Table 2 are comparable to the estimated MSEs in the no-noise condition under the heading 'Beta prior' in Table 1. Thus, although the simulated response rates were not beta distributed, the estimated response rates are close to the true values. The estimated MSEs are quite small, and increasing the number of items and persons leads to a reduction of MSE values. It can be concluded that the estimated response rates are close to the true simulated values given the estimated MSEs. The beta-binomial model is quite robust against violations of varying item difficulties for small sample sizes.

Table 2. Robustness of the beta-binomial model against varying item difficulties

(N,n)	No noise $MSE_{\hat{p}}$	Condition 1 $MSE_{\hat{p}}$	Condition 2 $MSE_{\hat{p}}$
(100,10)	0.014	0.023	0.047
(200,20)	0.008	0.019	0.046

5.2. Risk comparison

Binomial data were generated for different values of the RR sampling design parameter φ_1 with a fixed forced success probability of $\phi_2 = .60$. Note that this parameter reflects the amount of noise in the simulated data due to forced randomized responses. For convenience, the binomial sample size selected was the same for each respondent, $n = n_{ij}$. A binomial sample size of 8 and 12 was selected and the number of respondents was set to 300. The beta prior distribution parameters were varied to allow for symmetric as well as for skewed prior distributions. A vague symmetric prior was specified with α and β equal to 1, and a more informative symmetric prior with α and β equal to 2. The symmetric priors both have a prior mean of 1/2 and a variance of 1/12 and 1/20, respectively. A skewed prior was specified with $\alpha = 2$ and $\beta = 3$, corresponding to a prior mean of 2/5 and a variance of 1/25.

In Table 3, the estimates are presented. The estimates given are averaged outcomes over 100 independent samples. The Bayes risks given moment estimates and maximum-likelihood estimates are denoted as $\delta_M(y)$ and $\delta_{ML}(y)$, respectively. The moment estimates of the beta parameters are quite comparable to the maximum-likelihood estimates for different values of the binomial sample size and parameter ϕ_1. The Bayes risk reduces when the proportion of forced responses decreases. This follows from the fact that there is less noise in the data when ϕ_1 increases. An extreme case is when there are no randomized forced responses, that is, when ϕ_1 equals 1. The Bayes risks for $\phi_1 = .70$ and $n = 12$ are comparable to the Bayes risks for $\phi_1 = .80$ and $n = 8$.

Table 3. Moment and maximum-likelihood estimates and corresponding Bayes risks for Bayes estimator given simulated RR data

| ϕ_1 | (α,β) | n = 8 | | | | n = 12 | | | |
| | | Estimator | | Bayes risk | | Estimator | | Bayes risk | |
		Moment	ML	$\delta_M(y)$	$\Delta_{ML}(y)$	Moment	ML	$\delta_M(y)$	$\delta_{ML}(y)$
.70	$\alpha = 1$	1.012	1.013	10.178	10.187	1.009	1.019	7.277	7.290
	$\beta = 1$	1.016	1.008			1.010	1.013		
	$\alpha = 2$	2.074	2.026	10.158	10.151	2.043	2.024	7.635	7.624
	$\beta = 2$	2.071	2.044			2.037	2.024		
	$\alpha = 2$	2.004	2.061	9.404	9.400	2.009	2.037	7.176	7.216
	$\beta = 3$	2.999	2.933			3.023	2.987		
.80	$\alpha = 1$	1.008	1.014	7.789	7.775	1.010	0.995	5.571	5.843
	$\beta = 1$	0.998	1.005			1.007	1.005		
	$\alpha = 2$	2.054	2.053	7.764	7.756	1.991	2.003	5.563	5.834
	$\beta = 2$	2.055	2.055			2.044	2.014		
	$\alpha = 2$	2.061	2.111	7.149	7.229	2.000	2.042	5.514	5.538
	$\beta = 3$	3.094	3.014			2.976	2.962		
1.00	$\alpha = 1$	0.993	0.998	4.983	4.981	1.009	1.000	3.564	3.736
	$\beta = 1$	0.987	0.994			1.007	0.997		
	$\alpha = 2$	2.069	2.009	4.961	4.954	2.033	2.049	3.571	3.735
	$\beta = 2$	2.078	2.022			2.020	2.038		
	$\alpha = 2$	2.038	2.091	4.590	4.539	2.036	2.072	3.571	3.534
	$\beta = 3$	3.053	2.979			3.040	3.018		

This means, in this case, that a comparable risk is found when the proportion of forced responses is increased from .20 to .30 and the number of items is also increased from 8 to 12. This is important since interest is focused on obtaining truthful answers and respondents are more willing to share sensitive answers when the probability that the randomizing device dictates a forced response is apparent. On the other hand, interest is also focused on obtaining an accurate and reliable estimate of the response rate which means a low risk of the corresponding empirical Bayes estimator.

This trade-off is further explored in Figure 1, where the risks are plotted for the empirical Bayes estimator and the unbiased estimator as a function of the probability ϕ_1 that the randomizing device dictates a truthful answer from the respondent, keeping the forced success probability constant $\phi_2 = .60$ for $N = 300$ and $n = 8$. The risk functions are given for a vague beta prior with α and β equal to 1, a symmetric more informative prior with α and β equal to 3, and a skewed more informative prior with $\alpha = 1$ and $\beta = 3$. The prior distributions correspond to a prior mean of 1/2, 1/2, and 1/4, and a variance of 1/12, 3/80, and 1/28, respectively. Several conclusions can be drawn by comparing the risk values. It follows that more informative priors lead to lower risk values. The empirical Bayes estimator outperforms the unbiased estimator with respect to a risk comparison. The risk functions are decreasing for increasing ϕ_1 values. The risk function of the empirical Bayes estimator corresponding to a more informative beta prior has a less steep slope for decreasing values of ϕ_1 in comparison to a risk function corresponding to a less informative risk function. This is not true for the unbiased estimator since the risk corresponding to the prior $\mathcal{B}(3, 3)$ is higher than the risk for the vague prior $\mathcal{B}(1, 1)$. Finally, the functions in Figure 1 can be used in practice since they
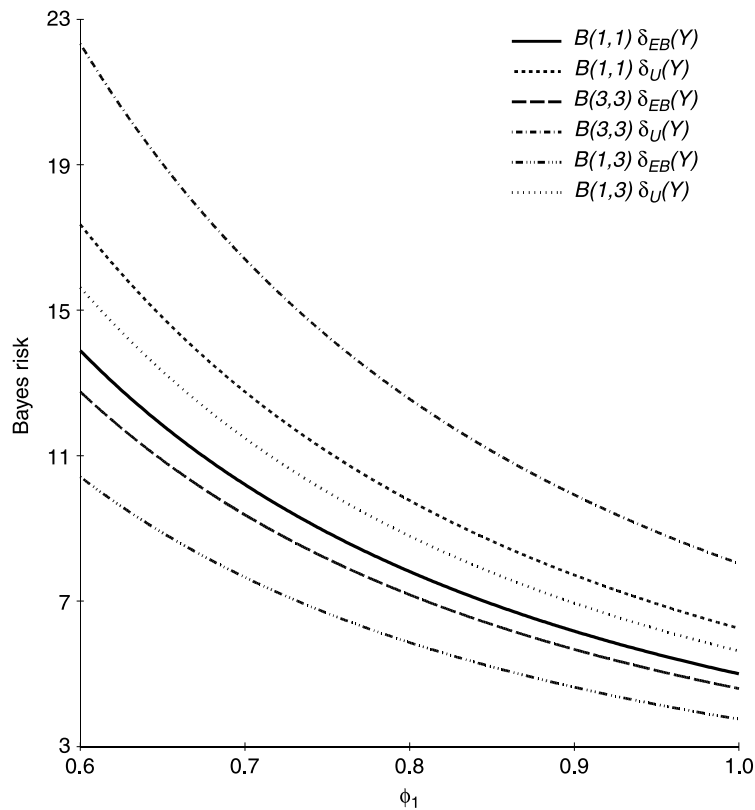
Figure 1. Bayes risk of empirical Bayes and unbiased estimator, $N = 300$, $n = 8$, and three different beta priors.

provide information about the risk in estimating the latent response rates for various values of ϕ_1 when population parameters are known.

6. An application using cheating data

Students at a university in The Netherlands were surveyed on the subject of to cheating in exams. Responses to questions were obtained via a forced randomized response technique, since it is known that most students are not eager to share information about frequency of and reasons for cheating in exams. Data were available from 349 students (229 male and 120 female) from one of the seven main disciplines at this university: Computer Science (CS), Educational Science and Technology (EST), Philosophy of Science (PS), Mechanical Engineering (ME), Public Administration and Technology (PAT), Science and Technology (ST), and Applied Communication Sciences (ACS). Within these seven disciplines, a stratified sample of students was drawn such that different studies were represented in proportion to their total number of students.

The students received an e-mail in which they were asked to participate in the survey. The forced alternative method was explained to increase the likelihood that students (1) participate in the study and (2) answer the questions truthfully. A web site

was developed containing 10 statements concerning cheating on exams and assignments (for the content of these items, see Appendix) and students were asked whether they agreed or disagreed with each statement. When a student visited the web site, an on-web dice server rolled two dice before a question could be answered. The result of both rolls determined whether the student should answer 'yes' (sum of the outcomes equalled 2, 3, or 4), 'no' (sum is 11 or 12), or answer the sensitive question truthfully. That is, the forced response technique was implemented with $\phi_1 = 3/4$ and $\phi_2 = 2/3$. Respondents were guaranteed confidentiality, and the questionnaires were filled in anonymously.

The posterior estimates of the mean response rate in the population and its variance equal .288 and .025, respectively, and these estimates indicate that student cheating is a serious problem. The estimated posterior distribution of the latent response rates, $\pi(p|y)$, is plotted in Figure 2. It can be seen that relatively high latent response rates of more than .5 are no exceptions. The estimated beta prior, $\pi(\mathbf{p}|\alpha, \beta)$, is shifted towards the right with parameters $\xi = .383$ and $\omega = .112$ in comparison with $\pi(\mathbf{p}|\mathbf{y})$, since it is the conjugated prior for the probabilities $\Delta(\mathbf{p})$. The corresponding beta prior parameters $\tilde{\alpha}$ and $\tilde{\beta}$ for the response rates $\mathbf{p}$ can be obtained from the equations

$$\xi = \Delta\left(\frac{\tilde{\alpha}}{\tilde{\alpha} + \tilde{\beta}}\right), \quad \frac{\xi(1 - \xi)}{\omega^{-1} + 1} = \frac{\phi_1^2 \tilde{\alpha}\tilde{\beta}}{(\tilde{\alpha} + \tilde{\beta})^2(\tilde{\alpha} + \tilde{\beta} + 1)}$$
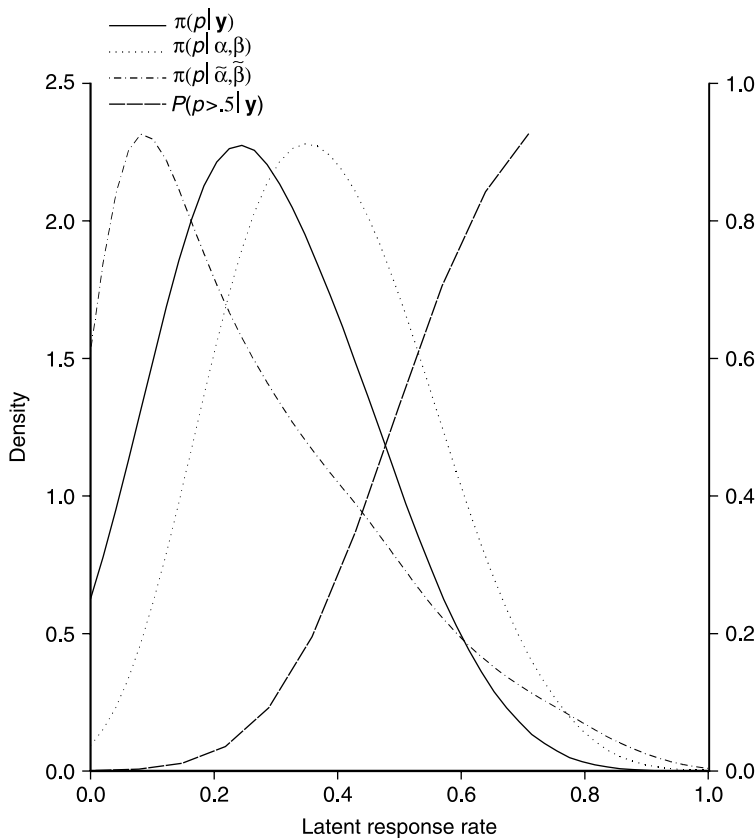

Figure 2. Posterior and prior distribution for the response rates.

since the transformed prior mean and variance of $\mathbf{p}$ are equal to the prior mean and variance of $\Delta(\mathbf{p})$. The corresponding estimated values of $\tilde{\xi}$ and $\tilde{\omega}$ are .288 and .259, respectively. Finally, the posterior probabilities that the latent response rates in the sample exceed .5 are plotted (corresponding to the right y-axis).

Via BF defined in (10), it was concluded that the response rates in the sample exhibit extraneous variance, that is, the null hypothesis $\omega = 0$ was rejected. Then, attention was focused on testing differences in mean response rates across gender and studies. In Figure 3, the reciprocal of estimated BFs is given for a uniform prior for ξ_j and a half-normal prior for ω_j, $\mathcal{N}(0, \sigma_\omega)$, where σ_ω ranged between 0 and 1. The plotted BFs correspond to the null hypothesis $\xi_j = \xi_{j'}$, and $\omega_j = \omega_{j'}$ for $j \neq j'$ against the alternative $\xi_j \neq \xi_{j'}$, $\omega_j \neq \omega_{j'}$. Values of BF^{-1} greater than 3 indicate substantial evidence against the null. In the case of grouping respondents by studies, the null hypothesis was rejected for all values of σ_ω between 0 and 1. In the case of grouping respondents by gender, the null was rejected when $\sigma_\omega > .170$. However, the null was rejected since the prior variance, defined as the inverse of the expected Fisher information, equalled .022. Note that increasing values of the normal variance, indicating more uncertainty about ω, result in values of BF that support the null hypothesis. It was concluded that separate beta-binomial models can be fitted for the different groups.

In Table 4, the parameter estimates of ξ_j and ω_j are given for the transformed response rates, $\Delta(p)$, and in brackets for the response rates, p, of the beta-binomial models. It can be seen that the males have a mean response rate lower than the females, meaning that females admit to cheating more than males do. Further, the response rates differ significantly across studies, and the largest difference was found between CS and ACS students.
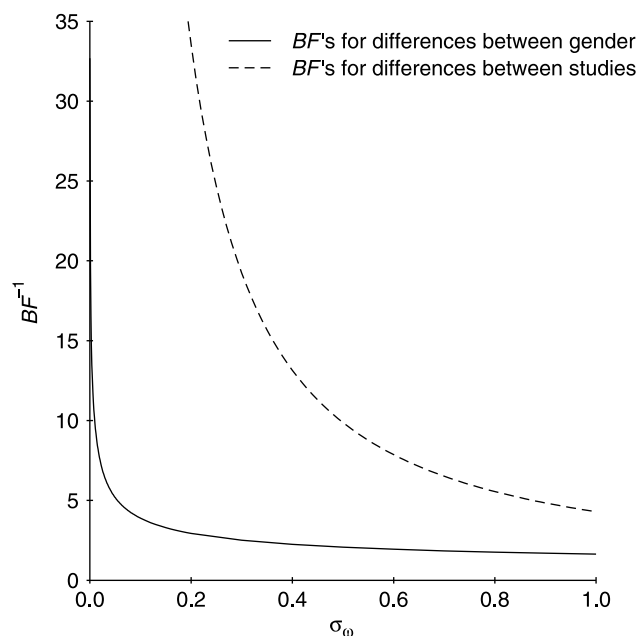

Figure 3. Bayes Factors for various prior variances of ω for testing differences in response rates between gender and studies.

Table 4. Posterior estimates of mean response rates and variation per group. Bayes factors for testing homogeneity in mean and variance between groups

Group	N	$\hat{\xi}$	$\hat{\omega}$	BF^{-1}
Gender	349	.383 (.288)	.112 (.259)	4.735
Male	229	.368 (.268)	.124 (.302)	
Female	120	.411 (.326)	.085 (.181)	
Study	349	.383 (.288)	.112 (.259)	179.087
CS	50	.299 (.176)	.119 (.374)	
PAT	53	.396 (.305)	.238 (.626)	
ACS	53	.420 (.337)	.068 (.140)	
ST	46	.405 (.317)	.142 (.325)	
EST	66	.411 (.325)	.091 (.195)	
ME	49	.369 (.269)	.058 (.130)	
PS	32	.371 (.272)	.016 (.034)	

Finally, the assumption of a constant response rate per individual was tested using a Bayes factor. The items were randomly grouped in two equal sets of five items, and the null hypothesis stated that the response rate to the first set of items, p_{ij1}, equals the response rate to the second set of items, p_{ij2}. Both response rates follow a beta distribution with parameters α_j and β_j. For each individual, a marginal likelihood was computed for the sum of responses and for the two sums of grouped item responses. In both cases, a log likelihood was defined based on equation (3) and the parameters were integrated out given a uniform prior for ξ_j and a half-normal prior for ω_j. In Figure 4, the reciprocal of the estimated BFs is plotted. An (inverse) Bayes factor value exceeding
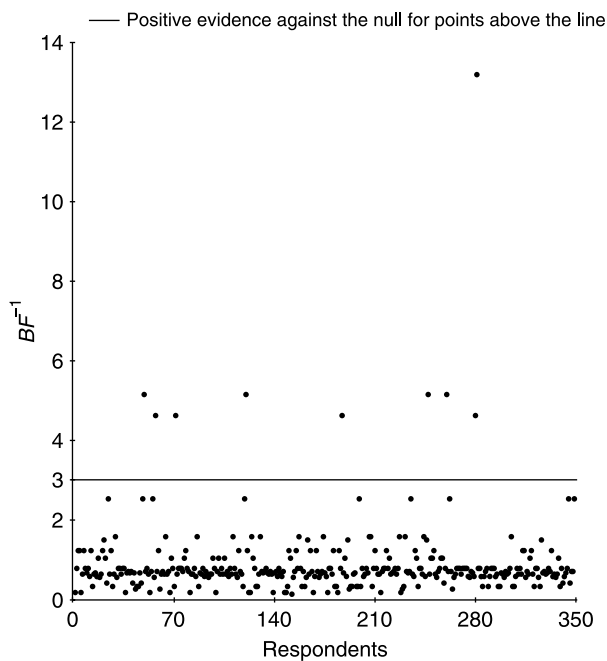

Figure 4. Bayes factors for testing non-constant individual response rates.

3 for the model with a constant response rate against the model with a non-constant response rate provides positive evidence in favour of a (non-)constant response rate. It can be seen that the null hypothesis is rejected for only nine respondents since each of the corresponding BF^{-1} value is greater than 3. The respondent with the maximum BF^{-1} value had a score of 0 for the first set of items and score of 5 for the second set of items, and this respondent can be marked as an outlier. It is concluded that the data do not support a non-constant response rate since the null was rejected for less than 2.6% of the respondents.

7. Discussion

In this paper, a beta-binomial model was proposed for analysing multivariate binary RR data. The model allows the computation of individual response rates, although the true individual responses are masked due to the RR sampling design. Moment estimates are easily obtained using the method of moments, and maximum-likelihood estimates can be obtained via the Newton–Raphson method. The empirical Bayes estimate of the individual response rate is a linear combination of the prior mean and sample mean and the forced success probability. As a result, the accuracy of the estimated response rates depends not only on the available prior knowledge, the binomial sample size, but also on properties of the randomizing device used in the sampling design.

An important problem is to compare proportions of a characteristic in several groups. A Bayes factors for testing homogeneity of proportions in the presence of over dispersion, given RR data, is presented. It is shown that the BF is sensitive to changes in the prior for parameter ω. The unit information prior is used but information for use in determining a prior for ω can be helpful.

The model can be extended in several ways. A generalization to multinomial data rather than binomial observations may be accomplished using the conjugated Dirichlet prior distribution. Explanatory variables can be incorporated by modelling the logit of response rates as a linear function of some covariates. This way, it is possible to model a grouping structure or to test for a group effect. Finally, the model can be extended to handle the entire class of related and unrelated or forced response sampling designs which are the two broad classes of RR designs. This can be difficult since the relationship between observed randomized responses and masked true responses is not necessarily linear as in the forced RR sampling design.

References

Böckenholt, U., & van der Heijden, P. G. M. (2007). Item randomized-response models for measuring noncompliance: Risk-return perceptions, social influences, and self-protective responses. *Psychometrika*, *72*, 245–262.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *American Statistician*, *39*, 83–87.

Casella, G., & Berger, R. L. (2002). *Statistical inference*. Pacific Grove, CA: Duxbury Thomson Learning.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fox, J.-P. (2005). Randomized item response theory models. *Journal of Educational and Behavioral Statistics*, *30*, 189–212.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman & Hall.

Greenberg, B. G., Abul-Ela, A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *American Statistician*, *64*, 520–539.

Griffiths, D. A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics*, *29*, 637–648.

Grosh, D. L. (1972). A Bayes sampling allocation scheme for stratified finite populations with hyperbinomial prior distributions. *Technometrics*, *14*, 599–612.

Hsiao, C. K. (1997). Approximate Bayes factors when a mode occurs on the boundary. *Journal of the American Statistical Association*, *92*, 652–663.

Kass, R. E., & Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association*, *84*, 717–726.

Lin, M.-H., & Hsiung, C. A. (1994). Empirical Bayes estimates of domain scores under binomial and hypergeometric distributions for test scores. *Psychometrika*, *59*, 331–359.

Lord, F. M. (1965). A strong true-score theory, with applications. *Psychometrika*, *30*, 239–270.

Lord, F. M., & Novick, R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Morris, D. G. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, *78*, 47–55.

Morrison, C. N., & Brockway, G. (1979). A modified beta binomial model with applications to multiple choice and taste tests. *Psychometrika*, *44*, 427–442.

Pauler, D. K., Wakefield, J. C., & Kass, R. E. (1999). Bayes factors and approximations for variance component models. *Journal of the American Statistical Association*, *94*, 1242–1253.

Skellam, J. G. (1947). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society*, *10*, 257–261.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*, 63–69.

Wilcox, R. R. (1981). A review of the beta-binomial model and its extensions. *Journal of Educational Statistics*, *6*, 3–32.

Appendix: Cheating questionnaire

During an exam or test (1–5):

(1)   Tried to confer with other students.
(2)   Allowed others to copy your work.
(3)   Used crib notes or cheat sheets.
(4)   Used unauthorized material such as books or notes.
(5)   Looked at another student's test paper with their knowledge.
(6)   Added information to authorized material.
(7)   Taken an exam illegally.
(8)   Lied to postpone a deadline.
(9)   Submitted coursework from others without their knowledge.
(10) Paraphrasing material from another source without acknowledging the author.