UCLA
Department of Statistics Papers

Title
Exact Analysis for Binomial Data with Covariates

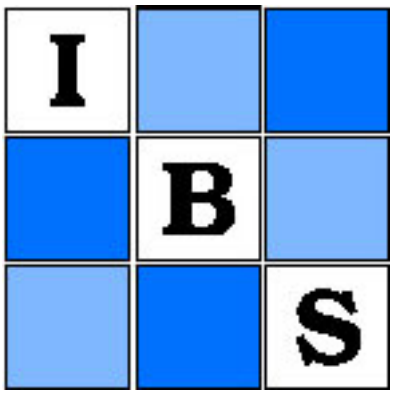
Permalink
https://escholarship.org/uc/item/8rr572db

Author
Hirji, Karim F.

Publication Date
1994-07-04

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California



Exact Analysis for Paired Binary Data
Author(s): Karim F. Hirji
Source: *Biometrics*, Vol. 50, No. 4 (Dec., 1994), pp. 964-974
Published by: International Biometric Society
Stable URL: http://www.jstor.org/stable/2533436
Accessed: 26/05/2011 14:15

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=ibs.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

International Biometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Biometrics*.

http://www.jstor.org

BIOMETRICS 50, 964–974
December 1994

Exact Analysis for Paired Binary Data

Karim F. Hirji

Department of Biomathematics, UCLA School of Medicine,
Los Angeles, California 90024-1766, U.S.A.

SUMMARY

This paper provides an efficient algorithm to generate exact distributions for the bivariate logistic model with common and sub-unit-specific covariates. The algorithm can be used to analyze correlated paired binary response data from studies lacking a large sample size. Possible applications include a clinical trial with two distinct binary outcomes, a binary outcome cross-over or two-time point cohort study, and a pair-matched prospective study with binary outcome. Analysis of data from an ophthalmologic study is provided to illustrate the method. Extension to three or more correlated binary responses is also outlined.

1. Introduction

Paired data arise in a variety of circumstances. In an ophthalmologic investigation, a reaction may be elicited in both eyes (Rosner, 1984). Pharmacologic investigations frequently adopt a two-period cross-over design (Jones and Kenward, 1989). Other examples include a two-time-point cohort study, a prospective matched pairs or identical twins study, and a size-two cluster sample study. In this paper we consider such paired studies in which, moreover, both responses are binary. In general, these studies aim to explicate the relationship between response and one or more of a set of factors. It is not prudent to analyze data from such a study by assuming that responses within each pair are independent (Glynn and Rosner, 1992). Further, in some studies, the sample size or sparse structure of the data may contraindicate use of large-sample approximations. This paper provides a method for performing tests and constructing confidence regions for small-sample correlated paired binary response data with covariates. Using a bivariate logistic model, we derive the joint distribution of the sufficient statistics for the model parameters, and indicate how exact inference may be performed using appropriate conditional distributions. We also give an efficient algorithm for computing such distributions, and use the algorithm to analyze a relevant data set.

In Section 2, we first consider a bivariate logistic model without explanatory factors. This is subsequently extended to a model with two types of covariates, namely, common and sub-unit covariates. In Section 3, we give a recursive algorithm to generate exact distributions for such a model. A grouped-data version of this algorithm together with other techniques to enhance computational efficiency are described in the Appendix. Section 4 gives an example of the exact analysis for paired binary data, Section 5 discusses computational issues, and we conclude by indicating extensions to settings with more than two binary responses.

2. The Bivariate Logistic Model

Consider first a model with no explanatory factors or covariates. Suppose there are n independent pairs of binary responses, with the ith pair denoted by (Y_{1i}, Y_{2i}). Assume a bivariate logistic model with

$$\Pr(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}) = \frac{\exp(\alpha_1 y_{1i} + \alpha_2 y_{2i} + \alpha_{12} y_{1i} y_{2i})}{1 + \exp(\alpha_1) + \exp(\alpha_2) + \exp(\alpha_1 + \alpha_2 + \alpha_{12})}, \tag{1}$$

where $y_{1i}, y_{2i} \in \{0, 1\}$. Models equivalent to (1) were given by Cox and Snell (1989), and Jones and Kenward (1989). In model (1), conditioning upon $Y_{2i} = y_{2i}$ produces the usual univariate logistic model. Further, in this model, $\alpha_{12} = 0$ signifies independence within pairs, whereupon the model becomes a product of two univariate binary logistic models.

Key words: Bivariate logistic model; Paired binary data; Recursive algorithm.

For model (1), one can obtain exact unconditional and conditional distributions explicitly. These distributions may be used to construct exact tests for hypotheses of interest, or obtain exact confidence regions for parameters of interest. For example, one may want to test the hypothesis $\alpha_1 = \alpha_2$, postulating the identity of the marginal distributions $\Pr(Y_{1i} = y) = \Pr(Y_{2i} = y)$ for all y, and all $i = 1, \ldots, n$. This may be done by testing $\delta = 0$ in the reparametrization $\alpha_1 = \alpha$ and $\alpha_2 = \alpha + \delta$, and using the conditional distribution of the sufficient statistic for δ.

Now consider a situation with covariates. Two types of covariates need to be considered here. For example, in an ophthalmologic study in which each eye is being tested for visual acuity, suppose that age and presence or absence of cataract are the pertinent covariates. The former covariate is common for both eyes whereas the latter is not. Another instance is a twin study considering risk factors for a disease. Parental medical history is a common covariate whereas occupation of the subject is not. In general, assume that each pair of responses (Y_{1i}, Y_{2i}) has a common p-dimensional covariate vector $\mathbf{x}_i$, and a pair of sub-unit-specific q-dimensional covariate vectors $(\mathbf{z}_{1i}, \mathbf{z}_{2i})$. In some settings, such covariates are labelled baseline and time-dependent covariates, respectively. Let $\Delta_{1i} = \alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}_i + \boldsymbol{\gamma}_1'\mathbf{z}_{1i}$, $\Delta_{2i} = \alpha_2 + \boldsymbol{\beta}_2'\mathbf{x}_i + \boldsymbol{\gamma}_2'\mathbf{z}_{2i}$, and $\Delta_{12i} = \alpha_{12} + \boldsymbol{\beta}_{12}'\mathbf{x}_i + \boldsymbol{\gamma}_{12}'(\mathbf{z}_{1i} \odot \mathbf{z}_{2i})$, where $\odot$ denotes elementwise multiplication for two vectors. Then consider the model

$$\Pr(Y_{1i} = y_{1i}, Y_{2i} = y_{2i} | \mathbf{x}_i, \mathbf{z}_{1i}, \mathbf{z}_{2i}) = \frac{\exp(\Delta_{1i}y_{1i} + \Delta_{2i}y_{2i} + \Delta_{12i}y_{1i}y_{2i})}{1 + \exp(\Delta_{1i}) + \exp(\Delta_{2i}) + \exp(\Delta_{1i} + \Delta_{2i} + \Delta_{12i})}. \tag{2}$$

Model (2) is a generalization of a model for paired data given by Cox and Snell (1989). A similar model is also given by Liang, Zeger, and Qaqish (1992). The conditions $\alpha_{12} = 0$, $\boldsymbol{\beta}_{12} = \mathbf{0}$, and $\boldsymbol{\gamma}_{12} = \mathbf{0}$ in model (2) correspond to a situation of within-pairs independence. The parameters in $\boldsymbol{\beta}_{12}$ and $\boldsymbol{\gamma}_{12}$ reflect dependence between paired responses mediated through covariates. Model (2) can be extended easily to include terms representing interactions between covariates; we have not included such terms here in order to avoid further notational complexities. If we assume that $\alpha_1 = \alpha_2$, and if there is only a single binary common covariate, then an exact test for $\beta_1 = \beta_2$ corresponds to the familiar McNemar's test for paired data.

Let $\Sigma_{i=1}^n$ be denoted by Σ_i. Further let the sufficient statistics for $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_{12})'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_{12}')'$, and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \boldsymbol{\gamma}_{12}')'$ be denoted by $\mathbf{R} = (R_1, R_2, R_{12})'$, $\mathbf{S} = (\mathbf{S}_1', \mathbf{S}_2', \mathbf{S}_{12}')'$, and $\mathbf{T} = (\mathbf{T}_1', \mathbf{T}_2', \mathbf{T}_{12}')'$, respectively. Then from the joint probability of all n pairs, we can show that

$$R_1 = \sum_i Y_{1i}, \quad R_2 = \sum_i Y_{2i}, \quad R_{12} = \sum_i Y_{1i}Y_{2i},$$

$$\mathbf{S}_1 = \sum_i Y_{1i}\mathbf{x}_i, \quad \mathbf{S}_2 = \sum_i Y_{2i}\mathbf{x}_i, \quad \mathbf{S}_{12} = \sum_i Y_{1i}Y_{2i}\mathbf{x}_i,$$

$$\mathbf{T}_1 = \sum_i Y_{1i}\mathbf{z}_{1i}, \quad \mathbf{T}_2 = \sum_i Y_{2i}\mathbf{z}_{2i}, \quad \mathbf{T}_{12} = \sum_i Y_{1i}Y_{2i}(\mathbf{z}_{1i} \odot \mathbf{z}_{2i}).$$

Further, the distribution of $(\mathbf{R}, \mathbf{S}, \mathbf{T})$, $\Pr(\mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s}, \mathbf{T} = \mathbf{t})$, is

$$\frac{c_n(\mathbf{r}, \mathbf{s}, \mathbf{t})\exp(\boldsymbol{\alpha}'\mathbf{r} + \boldsymbol{\beta}'\mathbf{s} + \boldsymbol{\gamma}'\mathbf{t})}{\Sigma_{(\mathbf{u},\mathbf{v},\mathbf{w}) \in \Omega} c_n(\mathbf{u}, \mathbf{v}, \mathbf{w})\exp(\boldsymbol{\alpha}'\mathbf{u} + \boldsymbol{\beta}'\mathbf{v} + \boldsymbol{\gamma}'\mathbf{w})}, \tag{3}$$

where $c_n(\mathbf{r}, \mathbf{s}, \mathbf{t})$ is the number of paired zero–one binary sequences $\{(y_{1i}, y_{2i}): i = 1, \ldots, n\}$ for which $\Sigma_i y_{1i} = r_1$, $\Sigma_i y_{2i} = r_2$, $\Sigma_i y_{1i}y_{2i} = r_{12}$, $\Sigma_i y_{1i}\mathbf{x}_i = \mathbf{s}_1$, $\Sigma_i y_{2i}\mathbf{x}_i = \mathbf{s}_2$, $\Sigma_i y_{1i}y_{2i}\mathbf{x}_i = \mathbf{s}_{12}$, $\Sigma_i y_{1i}\mathbf{z}_{1i} = \mathbf{t}_1$, $\Sigma_i y_{2i}\mathbf{z}_{2i} = \mathbf{t}_2$, and $\Sigma_i y_{1i}y_{2i}(\mathbf{z}_{1i} \odot \mathbf{z}_{2i}) = \mathbf{t}_{12}$, and Ω is the set of all possible realizations of the random vector $(\mathbf{R}, \mathbf{S}, \mathbf{T})$.

Distribution (3) is of the same form as the exact distribution for a unitary sequence of independent binary responses, indicating that exact analysis for the parameters of model (2) can be carried out along the lines suggested by Cox and Snell (1989). For instance, suppose the parameters in $\boldsymbol{\gamma}$ are of primary interest. The other parameters can then be factored out by using the conditional distribution, $\Pr(\mathbf{T} = \mathbf{t} | \mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s})$, which is

$$\frac{c_n(\mathbf{r}, \mathbf{s}, \mathbf{t})\exp(\boldsymbol{\gamma}'\mathbf{t})}{\Sigma_{\mathbf{w} \in \Omega_*} c_n(\mathbf{r}, \mathbf{s}, \mathbf{w})\exp(\boldsymbol{\gamma}'\mathbf{w})}, \tag{4}$$

where Ω_* is the conditional section of the subset of Ω with $\mathbf{R} = \mathbf{r}$ and $\mathbf{S} = \mathbf{s}$. Before providing examples of analysis based on such distributions, we consider the issue of computing these exact distributions.

3. The Computational Algorithm

Without covariates, both conditional and unconditional exact distributions can be represented explicitly. When the model has covariates, that may not be possible. Another strategy to compute the exact distribution then would be through exhaustive enumeration of all sequences of n binary pairs. This would generate a distribution of the form (3), from which the required conditional distribution could be extracted. For n pairs, the computational effort would be proportional to 2^{2n}, making the approach not an appealing one for, say, $n \geq 10$. For the case of unitary response, whether binary or polytomous, it has been shown that recursive derivation of the exact distribution is more efficient than total enumeration of all sequences, especially if the covariates are discrete (Hirji, Mehta, and Patel, 1987; Hirji, 1992). For the problem of paired binary sequences, a recursive approach can be developed as follows. Let us define

$$\mathbb{W}_i(y_1, y_2) = \begin{bmatrix} y_1 & y_2 & y_1y_2 \\ y_1\mathbf{x}_i & y_2\mathbf{x}_i & y_1y_2\mathbf{x}_i \\ y_1\mathbf{z}_{1i} & y_2\mathbf{z}_{2i} & y_1y_2(\mathbf{z}_{1i} \odot \mathbf{z}_{2i}) \end{bmatrix}, \quad \text{and} \quad \mathbb{T} = \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \\ \mathbf{t} \end{bmatrix},$$

and write $c_i(\mathbf{r}, \mathbf{s}, \mathbf{t})$ as $c_i(\mathbb{T})$. Then using arguments of the type given in the two papers cited above, it can be shown that the recursions for generating the exact distribution (3) are given by

$$c_i(\mathbb{T}) = \sum_{y_1,y_2\in\{0,1\}} c_{i-1}(\mathbb{T} - \mathbb{W}_i(y_1, y_2)),$$

$$\Omega_i = \Omega_{i-1} \oplus \bigcup_{y_1,y_2\in\{0,1\}} \{\mathbb{W}_i(y_1, y_2)\}.$$

These recursions can be used to generate distributions of the form (3) with both common and paired covariates in the model. The efficiency of these recursions can be demonstrated in a manner similar to that given in Hirji et al. (1987). The basic idea underlying such a demonstration is equivalent to arguing that, say, when multiplying three polynomials of degree n, it generally requires fewer multiplications if one initially multiplies two of them, and then multiplies the result with the third instead of multiplying all three polynomials simultaneously. The latter, which can be readily programmed on a computer using a "DO loop," would use $2(1 + n)^3$ multiplicative operations, whereas the former would need $(n + 1)(3n + 2)$ such operations, the reduction being due to an aggregation of terms that occurs with the multiplication of first two polynomials. Computing unconditional exact distributions for the situation under study is equivalent to recursively multiplying the polynomial-type product

$$\prod_{i=1}^{n} \{1 + \exp(\Delta_{1i}) + \exp(\Delta_{2i}) + \exp(\Delta_{1i} + \Delta_{2i} + \Delta_{12i})\}.$$

It can be shown that the above recursions also induce aggregations of terms, thus improving computational efficiency.

The efficiency of the above recursions can be further improved by using the data in a grouped format. Appendix A gives the grouped data versions of these recursions. Also note that exact analysis of some parameters in model (2) requires use of conditional distributions of the sufficient statistics for these parameters. Hirji et al. (1987) showed that an efficient method of obtaining conditional distributions is to supplement the recursions for computing unconditional distributions with methods of recursively trimming these distributions so that eventually only the required conditional slice of the joint distribution is generated. In Appendix B, we give some methods for recursively trimming unconditional distributions for paired binary data.

4. An Application

Berson et al. (1991) present, among other things, ocular findings from 17 unrelated patients with autosomal dominant retinitis pigmentosa and a specific genetic defect. Table 2 of this paper has data on two binary visual findings, namely, lens opacity and presence of pigment. For all these patients, the findings in both eyes were identical. Thus per-eye analysis is not an issue here. For the ith patient, let $Y_{1i} = 1$ if lens had posterior subcapsular opacity, and $Y_{1i} = 0$ if it was clear; $Y_{2i} = 1$ if all four quadrants had pigments, and $Y_{2i} = 0$, otherwise. The distribution of these findings by age group is shown in Table 1.

An issue of interest for such data would be to jointly relate both findings to age. For the ith patient, let x_i be the age group variable (0 if age $\leq$ 30; 1 if 31 $\leq$ age $\leq$ 40; 2 if 41 $\leq$ age $\leq$ 50; 3 if 51 $\leq$ age). We use model (2) with $\Delta_{1i} = \alpha_1 + \beta_1 x_i$, $\Delta_{2i} = \alpha_2 + \beta_2 x_i$, and $\Delta_{12i} = \alpha_{12} + \beta_{12} x_i$, and test

Table 1
Eye data

No. of (y_1, y_2) pairs				Age
(0, 0)	(0, 1)	(1, 0)	(1, 1)	(years)
2	0	1	1	≤30
0	4	0	3	31–40
0	3	1	1	41–50
0	1	0	0	≥51

Source: Berson et al. (1991).

whether $\beta_1 = \beta_2 = \beta_{12} = 0$. This is a model with a single common covariate where the interaction effect between responses mediated through the covariate is also included. The sample size is small, and Table 1 shows that most of the cell counts are zero or nearly zero. Under such circumstances, one would hesitate to perform the traditional large-sample analysis. Indeed it turns out that such an analysis for these data is fraught with difficulties. The Newton–Raphson algorithm, applied to the likelihood function based on (2), is unable to converge, with output from successive iterations indicating that maximum likelihood estimates for model parameters are not finite. This precludes performance of the asymptotic Wald test, which is perhaps the most commonly used large-sample test. A likelihood ratio test performed by using the upper bound of the unrestricted log-likelihood gave the value of the likelihood ratio statistic to be 11.67. Referring to a chi-square distribution with 3 degrees of freedom, the asymptotic test based on this statistic yields a P-value of approximately .01.

Alternatively, we can compute the exact distribution of the sufficient statistics. The realized values of these statistics are: $R_1 = 7$, $R_2 = 13$, $R_{12} = 5$, $T_1 = 7$, $T_2 = 18$, $T_{12} = 5$, where $\mathbf{T} = (T_1, T_2, T_{12})$ is the vector of sufficient statistics for $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_{12})$. For exact inference, the trivariate conditional distribution $\Pr(\mathbf{T} = \mathbf{t}|\mathbf{R} = \mathbf{r})$ is needed. This is of the form

$$\frac{c_n(\mathbf{r}, \mathbf{t})\exp(\boldsymbol{\beta}'\mathbf{t})}{\Sigma_{\mathbf{u}\in\Omega_*} c_n(\mathbf{r}, \mathbf{u})\exp(\boldsymbol{\beta}'\mathbf{u})}.$$

We compute this distribution using grouped data recursions given in Appendix A, supplemented with the trimming criteria given in Appendix B. It has 309 sample points, and a listing of it is available from the author.

The score statistic (Rao, 1973) based on this conditional distribution can be used to construct an exact test for H: $\boldsymbol{\beta} = \mathbf{0}$. Let $\mathbb{D}$ denote the vector of the first derivatives, with respect to the parameters, of the logarithm of the above probability, and let $\mathbb{V}$ be the matrix of the second derivatives. It can be shown that the ith element of $\mathbb{D}$ is $t_i - E[T_i]$, and the (i, j)th element of $\mathbb{V}$ is $E[T_iT_j] - E[T_i]E[T_j]$, where $E[\cdot]$ denotes the conditional expectation operator. The score statistic is given by $\partial(\mathbf{t}) = \mathbb{D}'\mathbb{V}^{-1}\mathbb{D}$. If $\mathbf{t}^*$ is the observed value of $\mathbf{T}$, then an exact α $(0 < \alpha < 1)$ level test based on this statistic obtains upon computing a conservative P-value, $P_c = \Pr_0\{\mathbf{t}: \partial(\mathbf{t}) \geq \partial(\mathbf{t}^*)\}$ and rejecting if $P_c \leq \alpha$. Conservative bias of such a test can be reduced by using a mid-P-value, $P_m = P_c - \frac{1}{2}\Pr_0\{\mathbf{t}: \partial(\mathbf{t}) = \partial(\mathbf{t}^*)\}$, and rejecting if $P_m \leq \alpha$. For the data in question, we obtained $P_c = .1169$ and $P_m = .1161$. The observed value of the conditional score statistic is 5.90. With reference to a chi-square distribution with 3 degrees of freedom, this gives an approximate P-value of .1166.

For these data, the traditional approximate unconditional analysis either fails or yields results that are suspect. On the other hand, both exact and approximate conditional analyses yield similar results. These conditional analyses are made possible by computation of the exact conditional distribution.

In principle, it is also possible to compute an exact $(1 - \alpha)100\%$ three-dimensional confidence region for $(\beta_1, \beta_2, \beta_{12})$ as $\{(\beta_1, \beta_2, \beta_{12}): P_c(\boldsymbol{\beta}) > \alpha\}$, and a mid-$P(1 - \alpha)100\%$ confidence region as $\{(\beta_1, \beta_2, \beta_{12}): P_m(\boldsymbol{\beta}) > \alpha\}$. Once the exact conditional distribution has been obtained, the confidence regions can be constructed by computing values of the score statistics along a three-dimensional parameter grid. In general, this is quite a time-consuming exercise. Further, visualization and presentation of such a three-dimensional region is not a straightforward task. Therefore, instead we construct a bivariate confidence region for (β_1, β_2) in a simpler model with $\beta_{12} = 0$. First we compute the bivariate exact conditional distribution of (T_1, T_2). This is followed by a grid search construction to identify the set $\{(\beta_1, \beta_2): P_m(\beta_1, \beta_2) > .05\}$. This is the 95% score-statistic-based mid-P confidence region for (β_1, β_2), and is shown in Figure 1. Much of this region is over the quadrant where both

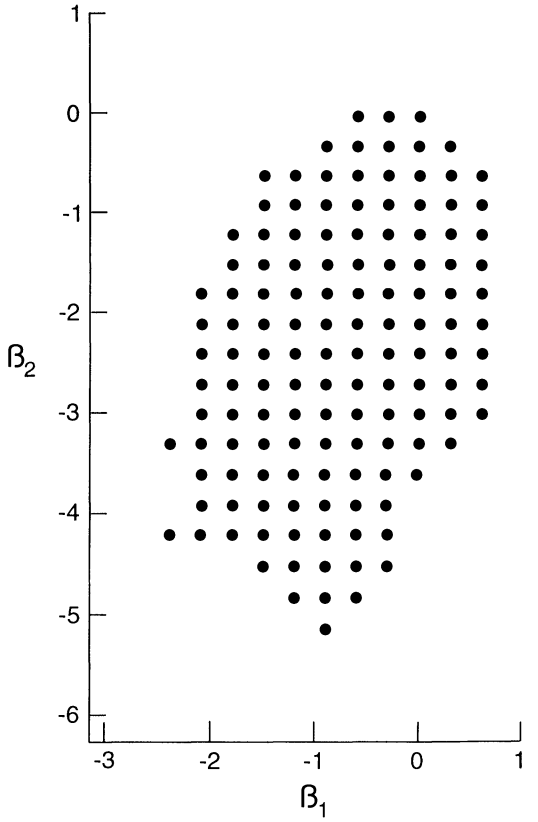


Figure 1. Mid-P confidence region for (β_1, β_2).

parameter values are negative; however, as expected from the mid-P-value, it also includes the null value.

5. Computational Capabilities of the Algorithm

In this section, we deal with some practical aspects of computer implementation of the algorithm given in Section 3. For this purpose, all computations were done on a PC-Brand™ 486/33 IBM compatible microcomputer with 4 Mb of RAM. The algorithm was programmed in FORTRAN-77 with double-precision accuracy. Elements of multidimensional arrays were stored using a hash function as described in Hirji et al. (1987). The program was compiled with the F77L-EM/32 Lahey FORTRAN compiler version 5.01. This compiler can embed a DOS-extender module in the executable code, enabling it to access all the RAM memory available on the machine.

With this code, computation of the trivariate and bivariate exact conditional distributions for the eye data in Section 4 took 1.8 and 1.6 seconds of CPU time, respectively. Computation of P-values from these distributions took less than one-tenth of a second. However, using a grid search method with 1,600 grid points, the bivariate confidence region depicted in Figure 1 needed a further 19.6 seconds to be generated.

In this particular example, age was grouped at four levels. If instead, age is not grouped, computation of exact distributions is a more demanding exercise. With age given in years, the bivariate distribution $\Pr((T_1, T_2)|(R_1, R_2, R_{12}))$, which has 7,058 sample points, took 139 seconds. Computation of the trivariate distribution $\Pr((T_1, T_2, T_{12})|(R_1, R_2, R_{12}))$ required more than 4 Mb of memory, and thus was not done. In general, if a covariate for which an exact distribution is needed takes on many distinct values, then the task of computing such a distribution is quite onerous. The number of records in the intermediate stages of the recursion may be very large, requiring a large amount of memory for storage and efficient retrieval. In general, computing exact distributions for binary covariates is more feasible than for non-binary covariates.

Features of the data that may impact computational time include sample size, the degree of cell-count imbalance within the data, number and types of covariates, and the type of model being fitted. We illustrate some of these issues with a help of three hypothetical datasets given in Table 2.

First consider Datasets I and II from this table. These datasets have 40 and 80 paired observa-

Table 2
Illustrative datasets

Dataset I

No. of (y_1, y_2) pairs					
(0, 0)	(0, 1)	(1, 0)	(1, 1)	x_1	x_2
0	1	1	9	0	0
1	1	2	8	0	1
1	2	1	5	1	0
1	3	0	4	1	1

Dataset II

No. of (y_1, y_2) pairs					
(0, 0)	(0, 1)	(1, 0)	(1, 1)	x_1	x_2
40	1	1	0	0	0
11	1	0	0	0	1
10	2	1	1	1	0
1	3	0	3	1	1

Dataset III

No. of (y_1, y_2) pairs						
(0, 0)	(0, 1)	(1, 0)	(1, 1)	v_1	v_2	v_3
0	1	8	8	0	0	0
1	1	4	6	0	0	1
1	0	1	0	0	1	0
1	0	0	1	0	1	1
0	1	0	1	1	0	0
0	1	0	1	1	0	1
1	0	0	0	1	1	0
0	1	0	1	1	1	1

tions, respectively. Suppose the variables x_1 and x_2 represent two values of a paired binary covariate. Then define two models for these data as: MODEL A: $\Delta_{1i} = \alpha_1 + \beta_1 x_{1i}$, $\Delta_{2i} = \alpha_2 + \beta_2 x_{2i}$, and $\Delta_{12i} = \alpha_{12}$; MODEL B: $\Delta_{1i} = \alpha_1 + \beta_1 x_{1i}$, $\Delta_{2i} = \alpha_2 + \beta_2 x_{2i}$, and $\Delta_{12i} = \alpha_{12} + \beta_{12} x_{1i} x_{2i}$. Alternatively, we may assume that variables x_1 and x_2 represent two common covariates. Then we consider two other models for these data: MODEL C: $\Delta_{1i} = \alpha_1 + \beta_{11} x_{1i} + \beta_{12} x_{2i}$, $\Delta_{2i} = \alpha_2 + \beta_{21} x_{1i} + \beta_{22} x_{2i}$, and $\Delta_{12i} = \alpha_{12}$; MODEL D: $\Delta_{1i} = \alpha_1 + \beta_{11} x_{1i} + \beta_{12} x_{2i}$, $\Delta_{2i} = \alpha_2 + \beta_{21} x_{1i} + \beta_{22} x_{2i}$, and $\Delta_{12i} = \alpha_{12} + \beta_{121} x_{1i} + \beta_{122} x_{2i}$. In Table 3, we show the size and computing times required for the

Table 3
Distribution size and computing times for Dataset I *and Dataset* II

		Dataset I		Dataset II	
Model	Conditional distribution	Sample points	CPU time	Sample points	CPU time
		1 Paired Covariate			
A	(T_1, T_2)	88	7.7	84	4.5
B	(T_1, T_2, T_{12})	634	8.4	240	4.8
		2 Common Covariates			
C	(T_{11}, T_{12})	76	6.0	64	3.0
C	(T_{21}, T_{22})	76	2.0	56	1.2
D	$(T_{11}, T_{12}, T_{112})$	160	7.4	120	3.0
D	$(T_{11}, T_{12}, T_{112})$	160	1.8	108	1.2

Note: CPU times are in seconds.

relevant conditional distributions for these models for Datasets I and II. Note that although Dataset II has almost twice as many pairs as Dataset I, the pattern of paired responses within it is more imbalanced, resulting in smaller final distributions and shorter computational times for Dataset II.

Now consider Dataset III, which has three binary variables, v_1, v_2, and v_3, and 40 paired

responses. Suppose one of these variables represents a common covariate x, and the other two, paired values of a sub-unit-specific covariate (z_1, z_2). Then consider the following two models: MODEL E: $\Delta_{1i} = \alpha_1 + \beta_1 x_i + \gamma_1 z_{1i}$, $\Delta_{2i} = \alpha_2 + \beta_2 x_i + \gamma_2 z_{2i}$, and $\Delta_{12i} = \alpha_{12}$; MODEL F: $\Delta_{1i} = \alpha_1 + \beta_1 x_i + \gamma_1 z_{1i}$, $\Delta_{2i} = \alpha_2 + \beta_2 x_i + \gamma_2 z_{2i}$, and $\Delta_{12i} = \alpha_{12} + \beta_{12} x_i + \gamma_{12} z_{1i} z_{2i}$. Let the sufficient statistics for $(\beta_1, \beta_2, \beta_{12})$ and $(\gamma_1, \gamma_2, \gamma_{12})$ be (T_1, T_2, T_{12}) and (S_1, S_2, S_{12}), respectively. With the data as given, we compute conditional distributions of these sufficient statistics for both models for each of the following three scenarios: (i) $x = v_1, z_1 = v_2, z_2 = v_3$, (ii) $x = v_2, z_1 = v_1, z_2 = v_3$, (iii) $x = v_3, z_1 = v_1, z_2 = v_2$. Computing times for each of these problems are shown in Table 4.

Table 4
Distribution size and computing times for Dataset III

Model	Conditional distribution	Sample points	CPU time
	(i) $x = v_1, z_1 = v_2, z_2 = v_3$		
E	(T_1, T_2)	55	8.7
E	(S_1, S_2)	128	7.9
F	(T_1, T_2, T_{12})	102	9.2
F	(S_1, S_2, S_{12})	409	7.9
	(ii) $x = v_2, z_1 = v_1, z_2 = v_3$		
E	(T_1, T_2)	55	8.8
E	(S_1, S_2)	136	40.0
F	(T_1, T_2, T_{12})	102	6.9
F	(S_1, S_2, S_{12})	448	37.5
	(iii) $x = v_3, z_1 = v_1, z_2 = v_2$		
E	(T_1, T_2)	131	29.3
E	(S_1, S_2)	64	24.8
F	(T_1, T_2, T_{12})	381	38.0
F	(S_1, S_2, S_{12})	164	37.4

Note: CPU times are in seconds.

From this table, we see that computing times for bivariate conditional distributions in MODEL E range from 7.9 to 40.0 seconds, and those for trivariate distributions in MODEL F range from 6.9 to 38.0 seconds. The number of sample points in bivariate distributions is generally less than half of that in corresponding trivariate distributions. Although the larger number of points in the final distribution in the trivariate distributions imposes an additional computational burden, this is offset by the additional trimming engendered by the third component of the vector of sufficient statistic being conditioned on. Of course, models with covariate-mediated interaction terms need more RAM memory for storage of intermediate stage arrays.

The examples given above indicate the range of computing times needed for such data. The data in these examples are quite imbalanced, with total number of pairs ranging from 40 to 80. All these computations were done within 2 Mb RAM memory. If data are more balanced, both computing times and memory requirements will be higher.

For microcomputer applications, we recommend usage of the algorithm given in this paper for imbalanced data with up to 80 paired responses and two binary covariates. For other applications, use of a fast processor in a workstation or a mainframe environment would be needed. However, further research on ways of improving efficiency of the algorithm is in progress. These include techniques for data sorting and development of more efficient trimming criteria, especially for non-binary covariates, which hopefully will enable somewhat larger sample sizes and balanced data to be tackled with greater ease.

6. Extensions and Conclusion

The model considered in this paper can be generalized to the case of three or more correlated binary responses. To simplify the presentation, we illustrate the case of three binary responses with common covariates. Extension to four or more responses, and inclusion of sub-unit-specific covariates into the model follows along similar lines. Let $\Delta_{1i} = \alpha_1 + \boldsymbol{\beta}_1' \mathbf{x}_i$, $\Delta_{2i} = \alpha_2 + \boldsymbol{\beta}_2' \mathbf{x}_i$, $\Delta_{3i} = \alpha_3 + \boldsymbol{\beta}_3' \mathbf{x}_i$, $\Delta_{12i} = \alpha_{12} + \boldsymbol{\beta}_{12}' \mathbf{x}_i$, $\Delta_{13i} = \alpha_{13} + \boldsymbol{\beta}_{13}' \mathbf{x}_i$, $\Delta_{23i} = \alpha_{23} + \boldsymbol{\beta}_{23}' \mathbf{x}_i$, and $\Delta_{123i} = \alpha_{123} + \boldsymbol{\beta}_{123}' \mathbf{x}_i$. Further let $\Lambda(y_{1i}, y_{2i}, y_{3i}) = \Delta_{1i} y_{1i} + \Delta_{2i} y_{2i} + \Delta_{3i} y_{3i} + \Delta_{12i} y_{1i} y_{2i} + \Delta_{13i} y_{1i} y_{3i} + \Delta_{23i} y_{2i} y_{3i} + \Delta_{123i} y_{1i} y_{2i} y_{3i}$. Then consider the model

$$\Pr(Y_{1i}=y_{1i},\ Y_{2i}=y_{2i},\ Y_{3i}=y_{3i}|\mathbf{x}_i)=\frac{\exp(\Lambda(y_{1i},\ y_{2i},\ y_{3i}))}{\Sigma_{y_{1i},y_{2i},y_{3i}\in\{0,1\}}\exp(\Lambda(y_{1i},\ y_{2i},\ y_{3i}))}.$$

It can be shown that the sufficient statistics for $(\alpha_1, \alpha_2, \alpha_3, \alpha_{12}, \alpha_{13}, \alpha_{23}, \alpha_{123})$ are $R_1 = \Sigma_i Y_{1i}$, $R_2 = \Sigma_i Y_{2i}$, $R_3 = \Sigma_i Y_{3i}$, $R_{12} = \Sigma_i Y_{1i}Y_{2i}$, $R_{13} = \Sigma_i Y_{1i}Y_{3i}$, $R_{23} = \Sigma_i Y_{2i}Y_{3i}$, $R_{123} = \Sigma_i Y_{1i}Y_{2i}Y_{3i}$, respectively, and those for $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3', \boldsymbol{\beta}_{12}', \boldsymbol{\beta}_{13}', \boldsymbol{\beta}_{23}', \boldsymbol{\beta}_{123}')$ are $\mathbf{S}_1 = \Sigma_i Y_{1i}\mathbf{x}_i$, $\mathbf{S}_2 = \Sigma_i Y_{2i}\mathbf{x}_i$, $\mathbf{S}_3 = \Sigma_i Y_{3i}\mathbf{x}_i$, $\mathbf{S}_{12} = \Sigma_i Y_{1i}Y_{2i}\mathbf{x}_i$, $\mathbf{S}_{13} = \Sigma_i Y_{1i}Y_{3i}\mathbf{x}_i$, $\mathbf{S}_{23} = \Sigma_i Y_{2i}Y_{3i}\mathbf{x}_i$, $\mathbf{S}_{123} = \Sigma_i Y_{1i}Y_{2i}Y_{3i}\mathbf{x}_i$, respectively. The joint distribution of these sufficient statistics can be shown to have a form similar to that in (3), with appropriate conditional distributions similar to that in (4). Further by defining

$$\mathbb{W}_i(y_1, y_2, y_3)=\begin{bmatrix} y_1 & y_2 & y_3 & y_1y_2 & y_1y_3 & y_2y_3 & y_1y_2y_3 \\ y_1\mathbf{x}_i & y_2\mathbf{x}_i & y_3\mathbf{x}_i & y_1y_2\mathbf{x}_i & y_1y_3\mathbf{x}_i & y_2y_3\mathbf{x}_i & y_1y_2y_3\mathbf{x}_i \end{bmatrix},$$

$$\mathbb{T}=\begin{bmatrix} R_1 & R_2 & R_3 & R_{12} & R_{13} & R_{23} & R_{123} \\ S_1 & S_2 & S_3 & S_{12} & S_{13} & S_{23} & S_{123} \end{bmatrix},$$

we can develop recursive algorithms for generating such distributions as extensions of those given earlier for paired binary data. Parametric parsimony in such models can be effected by imposing a structure of dependence between observations. One method would be to consider Markov models (Bonney, 1987). For example, in the model with three binary responses, a one-step Markov model obtains by setting $\Delta_{13i} = \Delta_{123i} = 0$. The distribution of sufficient statistics for this model, defined as above, can then be generated by using an appropriate algorithm.

In principle, therefore, the model considered in this paper can be extended to more general situations, and can be used for exact analysis of, say, longitudinal or clustered binary response problems. In practice, some difficulties may ensue. One obstacle is computational complexity. The models would in general have a large number of parameters, and a fast processor with a large random access memory would be needed to generate the required conditional distributions. Sometimes the dimensionality of the problem may render it computationally intractable even with a theoretically efficient algorithm. One may then need to resort to techniques such as importance sampling (Mehta, Patel, and Senchaudhuri, 1988) to obtain reasonable approximations to exact *P*-values.

The problem of computational intractability may arise even in paired response data if the number of covariates exceeds 3. In that case it may be feasible to fit only simpler models. These include models without covariate-mediated interaction terms, and models in which the covariate effect parameter is identical for both responses. In the context of the model given in Section 2, the former implies $\boldsymbol{\beta}_{12} = \boldsymbol{\gamma}_{12} = \mathbf{0}$, and the latter, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ and $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$. Another problem associated with the method given in this paper is conceptual in nature. This problem is the possibility of overconditioning. In some instances, conditional inference has been shown to have optimal properties in an unconditional setting (Lehmann, 1986; Hirji, Tan, and Elashoff, 1991). But when one is conditioning on the sufficient statistics for a large number of parameters, conditional distributions have a greater tendency to be unduly coarse or even degenerate. In this case, conditional inference would entail extensive loss of information, manifested by, among other things, excessive conservatism (Agresti, 1992). This problem needs detailed investigation, and calls for exploration of the possibility, both theoretical as well as practical, for unconditional exact inference from joint distributions of the form (4). Selective parametric reduction through use of simpler models can also help avoid overconditioning. If there is a particular covariate of interest, then overconditioning can be reduced by keeping the covariate-mediated interaction terms for this covariate in the model, and letting $\beta_{j12} = \gamma_{j12} = 0$ for all other covariates.

As efficient algorithms for performing exact conditional and unconditional inference become available for many different problems, one is inevitably faced with the issue of comparison of such methods with those developed under a more traditional large-sample-theory-based rubric. Some such investigations have already been done but many more are needed. The algorithm given in this paper will hopefully provide a further incentive to conduct such studies. Improvement of the algorithm given here, especially through use of better infeasibility criteria, and use of modern computer science techniques also needs to be considered. Extension of this algorithm to non-binary discrete responses is, of course, now on the agenda.

In conclusion, in this paper, we provide results of an initial effort for exact analysis of dependent categorical data with covariates with the hope that it will give an impetus to further research in the field. This algorithm can be used to compute exact distributions for small-sample paired binary response data within a reasonable time, especially if there is a high degree of sparseness or

imbalance in the data. It is with such data that one would hesitate to use large-sample approximations. A FORTRAN program implementing the method described in this paper is available from the author free of charge.

ACKNOWLEDGEMENT

This research was supported by USHHS Grant No. CA 16042.

RÉSUMÉ

Le présent article propose un algorithme permettant de calculer des distributions exactes pour des modèles logistiques bivariés, modèles dans lesquels on trouve aussi bien des covariables portant simultanément sur l'ensemble des deux réponses binaires étudiées que des covariables portant sur chacune de ces réponses prises séparément. Cet algorithme peut s'utiliser pour l'analyse de données binaires appariées, corrélées entre elles, dans des études où l'on ne dispose pas d'un effectif très important. Parmi les applications possibles, on peut citer les essais cliniques où il y a deux réponses binaires distinctes, les cross-overs comportant une seule réponse binaire, les études en cohorte où l'on s'intéresse à des résultats mesurés à deux temps différents et les études cas–témoin avec réponse binaire. La méthode est illustrée par une analyse de données issues d'un essai en ophtalmologie. Enfin, on aborde brièvement une généralisation de notre méthode au cas où le nombre de réponses binaires corrélées entre elles est supérieur ou égal à trois.

REFERENCES

Agresti, A. (1992). A survey of exact inference for contingency tables (with Dscussion). *Statistical Science* **7,** 131–177.

Berson, E. L., Rosner, B., Sandberg, M. A., and Dryja, T. P. (1991). Ocular findings in patients with autosomal dominant retinitis pigmentosa and a rhodopsin gene defect (Pro-23-His). *Archives of Ophthalmology* **109,** 92–101.

Bonney, G. E. (1987). Logistic regression for dependent binary observations. *Biometrics* **43,** 951–973.

Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data*, 2nd edition. London: Chapman and Hall.

Glynn, R. J. and Rosner, B. (1992). Accounting for correlation between fellow eyes in regression analysis. *Archives of Ophthalmology* **110,** 381–387.

Hirji, K. F. (1992). Computing exact distributions for polytomous response data. *Journal of the American Statistical Association* **87,** 487–492.

Hirji, K. F., Mehta, C. R., and Patel, N. R. (1987). Computing distributions for exact logistic regression. *Journal of the American Statistical Association* **82,** 1110–1117.

Hirji, K. F., Tan, S-J., and Elashoff, R. M. (1991). A quasi-exact test for comparing two binomial proportions. *Statistics in Medicine* **10,** 1137–1153.

Jones, B. and Kenward, M. G. (1989). *Design and Analysis of Cross-Over Trials*. London: Chapman and Hall.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd edition. New York: Wiley.

Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data (with Discussion). *Journal of the Royal Statistical Society, Series B* **54,** 3–40.

Mehta, C. R., Patel, N. R., and Senchaudhuri, P. (1988). Importance sampling for estimating exact probabilities in exact permutational inference. *Journal of the American Statistical Association* **83,** 999–1005.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. New York: Wiley.

Rosner, B. (1984). Multivariate methods in ophthalmology with application to other paired-data situations. *Biometrics* **40,** 1025–1035.

Received June 1992; *revised April and July* 1993; *accepted August* 1993.

APPENDIX A

Recursions for Grouped Data

Let m_i be the number of covariate vectors $(\mathbf{x}_i, \mathbf{z}_{1i}, \mathbf{z}_{2i})$, and let $\mathbb{S}_i$ be $\{j: \mathbf{x}_j = \mathbf{x}_i, \mathbf{z}_{1j} = \mathbf{z}_{1i}, \text{ and } \mathbf{z}_{2j} = \mathbf{z}_{2i}\}$. Further, let m be the number of distinct covariate vectors. Define grouped random variables N_{ji}

$(j = 1, 2, 3; i = 1, \ldots, m\}$ by $N_{1i} = \Sigma_{j \in \mathbb{S}_i} Y_{1j}(1 - Y_{2j})$, $N_{2i} = \Sigma_{j \in \mathbb{S}_i} Y_{2j}(1 - Y_{1j})$, and $N_{3i} = \Sigma_{j \in \mathbb{S}_i} Y_{1j} Y_{2j}$. In terms of these variables, if we let Σ_i denote $\Sigma_{i=1}^m$, then in model (2) the sufficient statistics for $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ are respectively given by

$$R_1 = \sum_i (N_{1i} + N_{3i}), \quad R_2 = \sum_i (N_{2i} + N_{3i}), \quad R_3 = \sum_i N_{3i},$$

$$\mathbf{S}_1 = \sum_i (N_{1i} + N_{3i})\mathbf{x}_i, \quad \mathbf{S}_2 = \sum_i (N_{2i} + N_{3i})\mathbf{x}_i, \quad \mathbf{S}_3 = \sum_i N_{3i}\mathbf{x}_i,$$

$$\mathbf{T}_1 = \sum_i (N_{1i} + N_{3i})\mathbf{z}_{1i}, \quad \mathbf{T}_2 = \sum_i (N_{2i} + N_{3i})\mathbf{z}_{2i}, \quad \mathbf{T}_3 = \sum_i N_{3i}(\mathbf{z}_{1i} \odot \mathbf{z}_{2i}),$$

with $0 \leq N_{ji} \leq m_i$ $(j = 1, 2, 3)$ and $N_{1i} + N_{2i} + N_{3i} \leq m_i$. Further, the combinatorial coefficient $c_n(\mathbf{r}, \mathbf{s}, \mathbf{t})$ is given by

$$c_n(\mathbf{r}, \mathbf{s}, \mathbf{t}) = \sum \tau(m_i, n_{1i}, n_{2i}, n_{3i})$$

where $\tau(m_i, n_{1i}, n_{2i}, n_{3i}) = m_i!/\{n_{1i}!n_{2i}!n_{3i}![m_i - (n_{1i} + n_{2i} + n_{3i})]!\}$, and further, where the summation is over all sequences $\{(n_{1i}, n_{2i}, n_{3i})$: $0 \leq n_{ji} \leq m_i$, $(j = 1, 2, 3)$, and $n_{1i} + n_{2i} + n_{3i} \leq m_i\}$ such that $\Sigma_i (n_{1i} + n_{3i}) = r_1$, $\Sigma_i (n_{2i} + n_{3i}) = r_2$, $\Sigma_i n_{3i} = r_3$, $\Sigma_i (n_{1i} + n_{3i})\mathbf{x}_i = \mathbf{s}_1$, $\Sigma_i (n_{2i} + n_{3i})\mathbf{x}_i = \mathbf{s}_2$, $\Sigma_i n_{3i}\mathbf{x}_i = \mathbf{s}_3$, $\Sigma_i (n_{1i} + n_{3i})\mathbf{z}_{1i} = \mathbf{t}_1$, $\Sigma_i (n_{2i} + n_{3i})\mathbf{z}_{2i} = \mathbf{t}_2$, and $\Sigma_i n_{3i}(\mathbf{z}_{1i} \odot \mathbf{z}_{2i}) = \mathbf{t}_3$. Let $\mathbb{W}_i(n_1, n_2, n_3)$ be equal to

$$\mathbb{W}_i(n_1, n_2, n_3) = \begin{bmatrix} n_1 + n_3 & n_2 + n_3 & n_3 \\ (n_1 + n_3)\mathbf{x}_i & (n_2 + n_3)\mathbf{x}_i & n_3\mathbf{x}_i \\ (n_1 + n_3)\mathbf{z}_{1i} & (n_2 + n_3)\mathbf{z}_{2i} & n_3(\mathbf{z}_{1i} \odot \mathbf{z}_{2i}) \end{bmatrix}.$$

Then using the arguments of the type given in unpublished work by Hirji (UCLA Statistics Series #28, 1989), it can be shown that the recursions for generating the exact distributions of the form (3) are given by

$$c_i(\mathbb{T}) = \sum_{0 \leq n_1 + n_2 + n_3 \leq m_i} \tau(m_i, n_1, n_2, n_3) c_{i-1}(\mathbb{T} - \mathbb{W}_i(n_1, n_2, n_3)),$$

$$\Omega_i = \Omega_{i-1} \oplus \bigcup_{0 \leq n_1 + n_2 + n_3 \leq m_i} \{\mathbb{W}_i(n_1, n_2, n_3)\}.$$

Appendix B

Computing Conditional Distributions

We describe some simple but quite effective methods for trimming recursively generated unconditional distributions. Suppose we are at the ith step of the recursions for grouped data given in Appendix A. Further assume that we need to compute a distribution of some sufficient statistic conditional on values of the sufficient statistics for the constant term parameters, i.e., $(R_1, R_2, R_3) = (r_1^*, r_2^*, r_3^*)$. Then to each element of $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ of Ω_i, we apply the following test: If for any $j = 1, 2, 3$, $r_j + \Sigma_{k>i} n_k < r_j^*$ or $r_j > r_j^*$, then the vector $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ is not included in Ω_i. If we are conditioning on some common covariate term generically denoted as $(S_1, S_2, S_3) = (s_1^*, s_2^*, s_3^*)$, then for any element of $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ of Ω_i, if for any $j = 1, 2, 3$, $s_j + \Sigma_{k>i} x_k n_k < s_j^*$ or $s_j > s_j^*$, then the vector $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ is not included in Ω_i. If we are conditioning on a paired covariate statistic generically denoted by $(T_1, T_2, T_3) = (t_1^*, t_2^*, t_3^*)$, then for any element of $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ of Ω_i, if either

$$\text{(i)} \quad t_1 + \sum_{k>i} x_{1k} n_k < t_1^* \quad \text{or} \quad t_1 > t_1^*,$$

$$\text{or (ii)} \quad t_2 + \sum_{k>i} x_{2k} n_k < t_2^* \quad \text{or} \quad t_2 > t_2^*,$$

$$\text{or (iii)} \quad t_3 + \sum_{k>i} x_{1k} x_{2k} n_k < t_3^* \quad \text{or} \quad t_3 > t_3^*,$$

then the vector $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ is not included in Ω_i.

Trimming criteria sharper than the ones given above can be derived upon consideration of values

of two or more of the sufficient statistics being conditioned on. For example, assume that the vectors $(R_1, R_2, R_3) = (r_1^*, r_2^*, r_3^*)$, and $(T_1, T_2, T_3) = (t_1^*, t_2^*, t_3^*)$ are both in the conditional set. For $j = 1$, 2, let

$$\mu_{ji} = \sum_{k>i} x_{jk} n_k \quad \text{and} \quad \nu_{ji} = \sum_{k>i} (1 - x_{jk}) n_k.$$

Also let $\mu_{3i} = \Sigma_{k>i} x_{1k} x_{2k} n_k$, $\nu_{3i} = \Sigma_{k>i} (1 - x_{1k} x_{2k}) n_k$. Then we replace conditions (i), (ii), and (iii) given above for deleting the vector under consideration from the feasible set by deleting if, for any $j = 1, 2, 3$, either $t_j + \min(r_j^* - r_j, \mu_{ji}) < t_j^*$, or $t_j + \max(0, r_j^* - r_j - \nu_{ji}) > t_j^*$.