Statistica Sinica **11**(2001), 199-211

EXACT GOODNESS-OF-FIT TEST FOR BINARY LOGISTIC MODEL

Man-Lai Tang

The Chinese University of Hong Kong

Abstract: Logistic regression is a widely applied tool for the analysis of binary response variables. Several test statistics have been proposed for the purpose of assessing the goodness of fit of the logistic regression model. Unfortunately, analysis based on these test statistics requires a moderately large sample size so that the chi-square approximation can be applied. When the sample size is small or the data structure is sparse, the asymptotic approximation becomes unreliable. In this article, an exact conditional goodness-of-fit test for the logistic regression model with grouped binomial response data is proposed. Two efficient algorithms are presented for carrying out the exact conditional goodness-of-fit test in small sample studies. Two data sets from an animal carcinogenesis experiment and a study on self-esteem are analyzed to demonstrate the methods.

Key words and phrases: Exact inference, goodness-of-fit test, grouped binomial response data, recursive algorithm.

1. Introduction

The logistic regression model is a commonly used technique for relating a binary response variable Z $\in \{0, 1\}$ (e.g., tumour incidence - present or absent) to a set of covariates $(X_1, \ldots, X_k)$ (e.g., risk factors) according to

$$\log\left\{\frac{p_{\boldsymbol{X}}}{1-p_{\boldsymbol{X}}}\right\} = \boldsymbol{\beta}'\boldsymbol{X},$$

where $p_{\boldsymbol{X}}$ represents the conditional probability of response given the vector $\boldsymbol{X}'$ $= (X_0, \ldots, X_k)$, $X_0 = 1$, and $\boldsymbol{\beta}' = (\beta_0, \ldots, \beta_k)$ represents the regression coefficients. Two approaches are available for statistical inference for the regression coefficients of the logistic regression model, the asymptotic and the exact approach. The asymptotic approach has been very popular and widely used due to its simplicity in computation (see, Hosmer and Lemeshow (1989)). However, it has been felt that this approach is unreliable for small, sparse, or skewed data (see, Mehta and Patel (1995)). Under these situations, statistical inference based on the exact approach is recommended. Since the exact approach requires the determination of the permutational distributions of appropriate sufficient statistics, it has been believed to be computationally impractical. Thanks to the

development of some fast algorithms for deriving these permutational distributions (e.g., Tritchler (1984), Hirji, Mehta and Patel (1987), and Hirji (1992)), statistical inference based on the exact approach now becomes computationally feasible, at least for a reasonable sample size. In fact, some of the algorithms for exact logistic regression analysis have been incorporated into commercial statistical packages, such as LogXact-4 for Windows (1999) and StatXact-4 for Windows (1999), enhancing their availability to the applied statistician.

Due to the popularity of the exact approach, more and more statisticians are now tempted to apply it without checking the adequacy of fit of the assumed logistic regression models. There has been a scattering of literature (e.g., Tsiatis (1980) and Hosmer and Lemeshow (1980)) on assessing the goodness of fit for binary logistic regression models. The methods proposed by these authors, however, require partitioning the space of covariates or subjects into regions or groups. A goodness-of-fit statistic is then calculated as a quadratic form of the observed minus expected counts in these regions or groups, and is shown to have a chi-square distribution, or a distribution which can be well approximated by a chi-square, when the sample size is large (see, e.g., Hosmer and Lemeshow (1980) and Lipsitz, Fitzmaurice and Molenberghs (1996)). Nevertheless, the choice of the number of partitioned regions or groups is quite subjective. Moreover, when the sample size is small and the data structure is sparse, the accuracy of the asymptotic approximation is questionable. Forster, McDonald and Smith (1996) presented a Gibbs sampling approach for estimating the exact P-value of the goodness-of-fit likelihood ratio statistic. However, as stated in their paper, how best to implement the Gibbs sampler is still an area of much current research and vigorous debate. More importantly, we note that the application of the Gibbs sampler for obtaining the exact P-value is sometimes unnecessary when the data are organized in a grouped fashion and the sample size is not large.

Mehta and Patel (1995) provided an excellent review and discussion on the exact conditional approach, an alternative to the maximum likelihood method, for making inferences about the parameters of the logistic regression model. Various applications of such an exact conditional approach were given to biomedical data sets. Unfortunately, a potential but important application of the existing exact conditional approach to the problem of assessing the goodness of fit of the logistic regression model was not discussed in their paper. The purpose of this paper is to propose an exact conditional goodness-of-fit test for the logistic regression model with grouped binomial response data.

In Section 2, we describe the logistic regression model for grouped binomial response data. A conditional goodness-of-fit test for the model will be discussed. In Section 3, an algorithm based on nested DO loops will be presented for calculating the exact P-value of the conditional goodness-of-fit test. An example

from a study of self-esteem is used for illustration. In Section 4, another efficient recursive algorithm will be proposed. We demonstrate the algorithm with an example from an animal carcinogenesis experiment. A brief discussion is given at the end of the article.

2. Exact Goodness-of-Fit Test

Suppose that in a study with a binary response variable, individuals can be classified according to common experimental conditions or covariate values. Let G be the number of groups, n_g be the number of subjects in group g in which observations possess a common p-dimensional covariate vector $(x_{g1}, \ldots, x_{gk})'$, Y_g be the corresponding random number of subjects who express a response (e.g., develop a tumor or not) in the gth group, and p_g be the probability that a subject exhibits a response in the gth group, where $Y_g \in \{0, \ldots, n_g\}$ and $g = 1, \ldots, G$. The logistic regression model that relates the probability of response p_g to the covariate vector $\boldsymbol{x}_g$ is given by

$$\log\left\{\frac{p_g}{1-p_g}\right\} = \boldsymbol{\beta}'\boldsymbol{x}_g, \tag{1}$$

where $\boldsymbol{x}_g = (x_{g0}, \ldots, x_{gk})'$ and $\boldsymbol{\beta}' = (\beta_0, \ldots, \beta_k)$ with $x_{g0} = 1$ for $g = 1, \ldots, G$. That is, we consider the logistic regression model with intercept. Let $q = k + 1$. Following Forster et al. (1996), to test the hypothesis concerning a model with $q < G$ covariates (i.e., treating the intercept as a covariate), we consider the following saturated model

$$\log\left\{\frac{p_g}{1-p_g}\right\} = \boldsymbol{\beta}'\boldsymbol{x}_g + \delta_g, \tag{2}$$

where $g = 1, \ldots, G$ with $\delta_1 = \cdots = \delta_q = 0$. It can be shown that the vectors of sufficient statistics for the regression parameters, i.e., $\boldsymbol{\beta}$, and $(\delta_{q+1}, \ldots, \delta_G)'$ are respectively given by

$$\boldsymbol{T} = \sum_{g=1}^{G} Y_g \boldsymbol{x}_g, \text{ and } (Y_{q+1}, \ldots, Y_G)'. \tag{3}$$

McCullagh (1986) argued that the appropriate distribution for testing goodness-of-fit is conditional on the vector of the sufficient statistics for the regression parameters. In our framework, a goodness-of-fit test for the model at (1) corresponds to the exact conditional test of $H_0 : \delta_{q+1} = \cdots = \delta_G = 0$. Let $\boldsymbol{y}^* = (y_1^*, \ldots, y_G^*)'$ be the observed vector $\boldsymbol{Y}$, $\boldsymbol{t}^* = \sum_{g=1}^{G} y_g^* \boldsymbol{x}_g$ be the corresponding observed vector $\boldsymbol{T}$, $\boldsymbol{y}_{q+1}^{+} = (y_{q+1}, \ldots, y_G)'$, and $\boldsymbol{y}_{q+1}^{*+} = (y_{q+1}^*, \ldots, y_G^*)'$. Hence,

the required exact conditional distribution under the null hypothesis H_0 is given by

$$Pr_{H_0}[\boldsymbol{Y}_{q+1}^{+} = \boldsymbol{y}_{q+1}^{+} | \boldsymbol{T} = \boldsymbol{t}^*] = \frac{\prod_{g=1}^{G} b(n_g, y_g)}{\sum_{\boldsymbol{v} \in \Omega(\boldsymbol{t}^*)} \prod_{g=1}^{G} b(n_g, v_g)}, \tag{4}$$

where $b(m,k)$ denotes the binomial coefficient $m!/[k!(m-k)!]$, $\Omega(\boldsymbol{t}^*) = \{\boldsymbol{v} = (v_1, \ldots, v_G)' : \sum_{g=1}^{G} v_g \boldsymbol{x}_g = \boldsymbol{t}^*,\ v_g = 0, \ldots, n_g, \text{ for } g = 1, \ldots, G\}$ and $(y_1, \ldots, y_q)'$ is the (unique) solution of the equations: $\sum_{g=1}^{G} Y_g \boldsymbol{x}_g = \boldsymbol{t}^*$, and $Y_g = y_g$, for $g = q+1, \ldots, G$. That is, we treat the regression parameters as nuisance parameters and factor them out by conditioning on their respective sufficient statistics. It should be noted that under the null hypothesis H_0, the exact conditional distribution (4) is free of any parameters. Following Hirji et al. (1987) (see also, Mehta and Patel (1995)), to test the null hypothesis H_0, the exact P-value based on the conditional probabilities test can be calculated by summing the probabilities of all configurations $\{y_1, \ldots, y_q\}$ with a probability not greater than that of the observation configuration, $\boldsymbol{y}^*$. In other words, the exact P-value of the conditional goodness-of-fit test can be computed as

$$P_{\boldsymbol{y}^*} = 1 - \sum_{\boldsymbol{y} \in \Omega(\boldsymbol{t}^*)} Pr_{H_0}[\boldsymbol{y}_{q+1}^{+} | \boldsymbol{T} = \boldsymbol{t}^*] I_{\{Pr_{H_0}[\boldsymbol{y}_{q+1}^{+} | \boldsymbol{T} = \boldsymbol{t}^*] > Pr_{H_0}[\boldsymbol{y}_{q+1}^{*+} | \boldsymbol{T} = \boldsymbol{t}^*]\}}, \tag{5}$$

where I denotes an appropriate indicator function.

According to the previous formulation, it is clear that the calculation of the exact P-value, $P_{\boldsymbol{y}^*}$, involves the identification of any vector $\boldsymbol{y}$ such that $\sum_{g=1}^{G} y_g \boldsymbol{x}_g = \boldsymbol{t}^*$ with $Y_g \in \{0, \ldots, n_g\}$ for $g = 1, \ldots, G$. Once the set $\Omega(\boldsymbol{t}^*)$ is constructed, the P-value, $P_{\boldsymbol{y}^*}$, can be computed according to (4) and (5). However, computing $\Omega(\boldsymbol{t}^*)$ according to the method proposed by Hirji et al. (1987) can be quite laborious. For example, we note that the statistical package LogXact-4 (1999) provides an option to report the value of $P_{\boldsymbol{y}^*}$. Unfortunately for the two examples we consider in this paper, LogXact failed to produce the value of $P_{\boldsymbol{y}^*}$ after more than twelve hours of computations and gave an "insufficient memory" message. All these analyses were running on a Dell Optiplex GX1 computer with 32Mb of RAM and a processor running at 400 MHz. We observe that the "insufficient memory" problem is mainly due to the fact that the existing package needs to generate and store all the elements in $\Omega(\boldsymbol{t}^*)$. In the following section, we will present an efficient algorithm for "identifying", but not "storing", elements in $\Omega(\boldsymbol{t}^*)$.

3. An Algorithm Based on Nested DO Loops

A trivial approach to identifying those elements in the set $\Omega(\boldsymbol{t}^*)$ is based on the method of exhaustive enumeration. That is, for every possible sequence

$(y_1,\ldots, y_G)$ with $y_g \in \{0,\ldots,n_g\}$ for $g = 1,\ldots,G$, we check if $\sum_{g=1}^{G} y_g \boldsymbol{x}_g = \boldsymbol{t}^*$. It should be noted that if the data are manipulated in an ungrouped format, i.e., each individual is treated as a separate entity, the total number of possible outcome sequences is 2^N, where $N = \sum_{g=1}^{G} n_g$. On the other hand, if individuals with common covariate values are processed in grouped format, then the total number of possible sequences is reduced to $\prod_{g=1}^{G}(1+n_g)$. Therefore, if the number $\prod_{g=1}^{G}(1+n_g)$ is not too large, the exhaustive enumeration approach is computationally practical. However, in the simplest situation in which the covariates under study are all binary, the number of possible groups is 2^p. In this situation, exhaustive enumeration becomes inefficient for $p \geq 4$. The situation is worse when some of the covariates are nonbinary.

We consider the following algorithm based on nested DO loops:

Step 1. Generate the sequence $\boldsymbol{y} = (y_1,\ldots,y_G)$ by using the following backward nested loops

$$L_g \leq y_g \leq U_g, \text{ for } g = 1,\ldots,G, \tag{6}$$

where $L_g = \max\{0, L_{gj}, j = 1,\ldots,q\}$; $U_g = \min\{n_g, U_{gj}, j = 1,\ldots,q\}$; $L_{gj} = \min\{0, I_{x_{gj}}[\frac{1}{x_{gj}}\{t_j^* - \sum_{g'=1}^{g-1} x_{g'j} n_{g'} - \sum_{g'=g+1}^{G} x_{g'j} y_{g'}\}]\}$, $j = 1,\ldots,q$; $U_{gj} = \max\{0, I_{x_{gj}}[\frac{1}{x_{gj}}\{t_j^* - \sum_{g'=g+1}^{G} x_{g'j} y_{g'}\}]\}, j = 1,\ldots,q$; $I_{x_{gj}} = 1$ if $x_{gj} \neq 0$; $= 0$ otherwise; and $[D]$ = the largest integer less than or equal to D;

Step 2. In the innermost loop, i.e., $g = 1$, check if $\sum_{j=1}^{G} y_j \boldsymbol{x}_j = \boldsymbol{t}^*$.

Step 2a. If "No", skip to the next $\boldsymbol{y}$ sequence;

Step 2b. If "Yes", then

(1) Accumulate $\prod_{j=1}^{G} b(n_j, y_j)$ to P_t; and

(2) If $\sum_{j=1}^{G} \ln[b(n_j, y_j)] > \sum_{j=1}^{G} \ln[b(n_j, y_j^*)]$, then accumulate $\prod_{j=1}^{G} b(n_j, y_j)$ to P_{gt};

Step 3. If the nested DO loops are exhausted, the exact P-value is computed as $P_{\boldsymbol{y}^*} = 1 - P_{gt}/P_t$.

The above algorithm is basically the same as the one proposed by Bedrick and Hill (1992). In their case, the entire reference set (i.e., $\Omega(\boldsymbol{t}^*)$) is stored. In our case, we are only interested in generating the right-tailed probability neccessary for the exact P-value calculation and storing the entire reference set is unneccessary.

By noting that any linear transformation on the covariates does not change the exact P-value of the conditional goodness-of-fit test, we assume without loss of generality that all covariates take nonnegative integer values. It is easy to see that if all the covariates under consideration are binary, then those sequences $\{y_1,\ldots,y_G\}$ generated by the nested DO loops (6) will automatically satisfy $\sum_{g=1}^{G} y_g \boldsymbol{x}_g = \boldsymbol{t}^*$. Finally, we would like to point out that the Nested DO-LOOPS

algorithms proposed by O'Flaherty and MacKenzie (1982) or Garcia-Perex (1995) can be readily adopted in our framework.

3.1. Example: a study on self-esteem

Demo and Parker (1987) studied the effect of academic achievement on self-esteem among black and white college students. The data from this study are given in Table 1. By performing a series of separate primary analyses, they showed that (i) there was no significant difference between the self-esteem levels of black and white college students; (ii) the effect of gender on self-esteem was significant; and (iii) there was no association between academic achievement and overall self-esteem. Let $g = 1, \ldots, 8$ label the eight groups listed in Table 1. We can reanalyze the data by considering the following logistic model relating various factors to the probability of expressing low self-esteem

$$\log\left\{\frac{p_g}{1-p_g}\right\} = \beta_0 + \beta_1 Gender_g + \beta_2 GPA_g + \beta_3 Race_g. \tag{7}$$

Table 1. Self-esteem study data.

Gender	Cumulative GPA	Race	No. of Students with Low Self-esteem	Total no. of Students
0	0	0	26	48
0	0	1	17	43
0	1	0	10	27
0	1	1	9	24
1	0	0	17	20
1	0	1	23	47
1	1	0	32	54
1	1	1	22	35

Note: Gender, 0 = male, 1 = female; Cumulative GPA, 0 = low, = high; Race, 0 = white, 1 = black.
Source: Demo and Parker (1987).

We implemented the exhaustive enumeration approach on a SUN SPARC20B station and obtained the exact P-value of the conditional goodness-of-fit test, equal to 0.1371, in 52 seconds, while the algorithm based on the nested DO loops took only 9 seconds to get the same result. Therefore, a conditional goodness-of-fit test shows that the fit of the logistic regression model (7) is adequate. To test the significance of each individual effect, we performed a two-sided hypothesis testing of H_0: $\beta_i = 0$ vs. H_1: $\beta_i \neq 0$, for $i = 1$, 2, and 3, using the exact approach. In this case, we adopt "the twice the smaller tail" method to compute the two-sided exact P-value (see, e.g., Tang, Hirji and Vollset (1995)). The corresponding

exact P-value for H_0: $\beta_i = 0$ vs. H_1: $\beta_i \neq 0$ is determined by

$$2\min(Pr[T_i \geq t_i^* | T_j = t_j^*, \text{ for all } j \neq i; \beta_i = 0], \\ Pr[T_i \leq t_i^* | T_j = t_j^*, \text{ for all } j \neq i; \beta_i = 0]).$$

The P-values for different hypotheses are reported in Table 2. The exact analysis indicates that there is a significant gender effect on self-esteem and the effect of cumulative GPA on self-esteem is not significant. However, a non-significant race effect on self-esteem is reported. In this case, exact analysis of these data leads to the same conclusions as those of Demo and Parker (1987).

Table 2. Exact P values* for self-esteem data.

Variable	$H_0 : \beta_i = 0$ vs. $H_1 : \beta_i \neq 0$ Exact P Value*
Gender (β_1)	0.0027
GPA (β_2)	0.3786
Race (β_3)	0.0685

*: Exact P value is calculated according to the "Twice the Smaller Tail" Method.

4. An Algorithm Based on Recursion

In practice, we do not need to generate the entire exact conditional distribution under the null hypothesis to compute the exact P-value of the conditional goodness-of-fit test. From equation (5), the main components in determining the exact P-value, $P_{\boldsymbol{y}^*}$, include (i) the calculation of the normalizing constant $\sum_{\boldsymbol{v} \in \Omega(\boldsymbol{t}^*)} \prod_{g=1}^{G} b(n_g, v_g)$; and (ii) the summation $\sum \prod_{g=1}^{G} b(n_g, v_g)$ for any sequence $\boldsymbol{v}$ in $\Omega(\boldsymbol{t}^*)$ which satisfies $\sum_{g=1}^{G} \ln[b(n_g, v_g)] > \sum_{g=1}^{G} \ln[b(n_g, y_g^*)]$. The first task can be done by implementing a modified procedure based on the recursive algorithm proposed by Hirji et al. (1987), the second task can be accomplished by using the same recursive procedure together with a simple trimming criterion. We discuss the implementation as follows.

4.1. Computation of the normalizing constant

In the formulation at (1), we include the intercept term. Then T_1 is always equal to $\sum_{g=1}^{G} Y_g$, and $t_1^* = \sum_{g=1}^{G} y_g^*$. Consider the joint distribution of the vector of sufficient statistics for the regression parameters in $\boldsymbol{\beta}$, i.e., $\boldsymbol{T}$. The joint distribution of $\boldsymbol{T}$ is given by

$$Pr[\boldsymbol{T} = \boldsymbol{t}] = \frac{c_G(\boldsymbol{t}) \exp(\boldsymbol{\beta}' \boldsymbol{t})}{\sum_{\boldsymbol{u} \in \Omega_G} c_G(\boldsymbol{u}) \exp(\boldsymbol{\beta}' \boldsymbol{u})}, \tag{8}$$

where $c_G(\boldsymbol{t})$ is the number of ways of selecting y_g ($\in \{0, \ldots, n_g\}$) subjects from group $g (g = 1, \ldots, G)$ such that the sequence (y_1,..., y_G) satisfies $\sum_{g=1}^{G} y_g \boldsymbol{x}_g = \boldsymbol{t}$, and where Ω_G is the set of all possible values of $\boldsymbol{T}$. It is easy to see that $c_G(\boldsymbol{t}^*)$ is our desired normalizing constant, $\sum_{\boldsymbol{v} \in \Omega(\boldsymbol{t}^*)} \prod_{g=1}^{G} b(n_g, v_g)$.

Let Ω_g be the set of the feasible values of $\boldsymbol{T}$ obtained by using the first g groups only ($g = 1, \ldots, G$) and, for each $\boldsymbol{t} \in \Omega_g$, let $c_g(\boldsymbol{t})$ be the resultant combinatorial coefficient. Hirji et al. (1987) presented an efficient recursive algorithm that generates the joint distribution of the sufficient statistics, $\boldsymbol{T}$, when $n_g = 1$ for $g = 1, \ldots, G$. In this case, the recursions for generating Ω_G and $\{c_G(\boldsymbol{t}) : \boldsymbol{t} \in \Omega_G\}$ are given by $\Omega_g = \Omega_{g-1} \cup \{\Omega_{g-1} \oplus \boldsymbol{x}_g\}$; and $c_g(\boldsymbol{t}) = c_{g-1}(\boldsymbol{t}) + c_{g-1}(\boldsymbol{t} - \boldsymbol{x}_g)$, for $\boldsymbol{t} \in \Omega_g$, where $\oplus$ symbolizes the addition of a vector to every member of a set. By mathematical induction, we can show that for the general case in which $n_g \geq 1$ $(g = 1, \ldots, G)$, the corresponding recursions can be summarized as follows:

$$\Omega_g = \bigcup_{y_g=0}^{n_g} \{\Omega_{g-1} \oplus y_g \boldsymbol{x}_g\}; \text{ and} \tag{9}$$

$$c_g(\boldsymbol{t}) = \sum_{y_g=0}^{n_g} b(n_g, y_g) c_{g-1}(\boldsymbol{t} - y_g \boldsymbol{x}_g), \text{ for } \boldsymbol{t} \in \Omega_g. \tag{10}$$

Obviously, if only the normalizing constant is needed, generation of the entire set $\{c_G(\boldsymbol{t}) : \boldsymbol{t} \in \Omega_G\}$ is not necessary. Following the arguments in Hirji et al. (1987), we can calculate the desired normalizing constant $c_G(\boldsymbol{t}^*)$ by introducing a set of appropriate infeasibility criteria.

For $g = 0, \ldots, G-1$, and $j = 2, \ldots, q$, let $\phi_{gj}(z)$ and $\psi_{gj}(z)$ be the minimum and the maximum, respectively, of $\sum_{i=g+1}^{G} Y_i x_{ij}$ subject to $\sum_{i=g+1}^{G} Y_i = \mathrm{z}$ and $Y_i \in \{0, \ldots, n_i\}$. To obtain $\phi_{gj}(z)$, we first arrange the subsequence $x_{g+1,j}$, $x_{g+2,j}$,..., x_{Gj} in a nondecreasing order. Let the ordered (nondecreasing) subsequence be $x_{(g+1,j)} \leq x_{(g+2,j)} \leq \cdots \leq x_{(Gj)}$; and let $n_{(g+1,j)}$, $n_{(g+2,j)}$,..., $n_{(Gj)}$ be the corresponding sample sizes. Let $M_i = \min(n_{(i)}, z_i)$ and $z_{i+1} = z_i$ - M_i, for $i = g+1, g+2, \ldots, G$, where $z_{g+1} = z$. Then $\phi_{gj}(z)$ is $\sum_{i=g+1}^{G} M_i x_{(ij)}$ and $\psi_{gj}(z)$ can be computed in a similar manner, after arranging the subsequence $x_{g+1,j}$, $x_{g+2,j}$,..., x_{Gj} in a nonincreasing order. Let $\phi_{g1}(z) = 0$ and $\psi_{g1}(z) = \sum_{i=g+1}^{G} n_i$, for $g = 0, \ldots, G-1$ and for any z. Now, suppose that we are at the gth stage of the recursions given by (9) and (10). Let $\boldsymbol{t} \in \Omega_{g-1}$. For any given $y_g \in \{0, \ldots, n_g\}$, and any $j = 1, \ldots, q$, if either $t_j + y_g x_{gj} + \phi_{gj}(t_1^* - t_1 - y_g) > t_j^*$ or $t_j + y_g x_{gj} + \psi_{gj}(t_1^* - t_1 - y_g) < t_j^*$, then given the sub-sequence $(y_1, \ldots, y_{g-1})$ that resulted in the vector $\boldsymbol{t}$, and y_g, there cannot exist any sub-sequence $(y_{g+1}, \ldots, y_G)$ for which $\sum_{j=1}^{G} Y_j \boldsymbol{x}_j = \boldsymbol{t}^*$. Thus, we can delete the vector $\boldsymbol{t} + y_g \boldsymbol{x}_g$ from further consideration. Therefore, by implementing the recursions (9) and (10) together

with checking those aforementioned infeasibility criteria, the desired normalizing constant $c_G(\boldsymbol{t}^*)$ will be produced in the final stage of the recursions, i.e., when $g = G$.

Here we compute the minimums and maximums subject only to the constraint based on the sufficient statistic of the intercept parameter. However, "stricter" minimums and maximums can be obtained by imposing more constraints based on other sufficient statistics of the regression parameters. The rationale for using stricter bounds can be found in Hirji (1992) and we omit the details here. However, solutions may not be available and computations are lengthy.

4.2. Summation $\sum \prod_{j=1}^{G} b(n_j, v_j)$ for appropriate sequences

For $g \leq G-1$ and $z \leq t_1^*$, let $\lambda_g(z) = \max \sum_{j=g+1}^{G} \ln[b(n_j, y_j)]$, with the maximization done over all $(y_{g+1},\ldots, y_G)$ satisfying $y_{g+1} + \cdots + y_G = z$, $0 \leq y_j \leq n_j$; $y_j \in \{0,\ldots,n_j\}$ with $j = g+1,\ldots,G$. Let $\lambda_G(\mathrm{z}) = 0$ for any z. $\lambda_g(.)$ can be readily computed by backward induction (see, Mehta and Patel (1980)). To compute the summation $\sum \prod_{j=1}^{G} b(n_j, v_j)$ for those sequences $\boldsymbol{v}$'s in $\Omega(\boldsymbol{t}^*)$ which satisfy $\sum_{j=1}^{G} \ln[b(n_j, v_j)] > \sum_{j=1}^{G} \ln[b(n_j, y_j^*)]$, we introduce an additional trimming criterion.

Suppose we are at the gth stage of the recursions given by (9) and (10). Let $\boldsymbol{t} \in \Omega_{g-1}$. For any given $y_g \in \{0,\ldots,n_g\}$, and any $j = 1,\ldots,q$, if $\sum_{j=1}^{g} \ln[b(n_j, y_j)] + \lambda_g(t_1^* - \sum_{j=1}^{g} y_j) \leq \sum_{j=1}^{G} \ln[b(n_j, y_j^*)]$, then given the sub-sequence $(y_1,\ldots,y_{g-1})$ that resulted in the vector $\boldsymbol{t}$, and y_g, there cannot exist any sub-sequence $(y_{g+1}, \ldots, y_G)$ for which $\sum_{j=1}^{G} \ln[b(n_j, y_j)] > \sum_{j=1}^{G} \ln[b(n_j, y_j^*)]$. Thus, we can delete the vector $\boldsymbol{t} + y_g \boldsymbol{x}_g$ from further consideration. By implementing the recursions (9) and (10) together with the trimming criterion above, and the infeasibility criteria described in Section 4.1, the desired summation $\sum \prod_{j=1}^{G} b(n_j, v_j)$ for those sequences $\boldsymbol{v}$'s in $\Omega(\boldsymbol{t}^*)$ which satisfy $\sum_{j=1}^{G} \ln[b(n_j, v_j)] > \sum_{j=1}^{G} \ln[b(n_j, y_j^*)]$ will be produced in the final stage of the recursions, i.e., when $g = G$.

Once again, one may argue that a "stricter" maximum for $\sum_{j=g+1}^{G} \ln[b(n_j, y_j)]$ can be adopted if we restrict the maximization to sub-sequences $(y_{g+1},\ldots, y_G)$ satisfying $\sum_{j=g+1}^{G} y_j \boldsymbol{x}_j = \boldsymbol{t}^* - \sum_{j=g}^{G} y_j \boldsymbol{x}_j$. In this case, we may need to use general integer programming methods (Hadley (1964)). However, the more constraints the greater the computational effort required in the maximization problem. This is an important factor to be considered before using stricter bounds.

4.3. Example: study of tolazamide in an animal carcinogensis experiment

We consider the example discussed in Tarone and Gart (1980), a National Cancer Institute animal carcinogensis experiment in which the drug tolazamide

was administered to male and female mice and rats. Animals assigned to control, low dose, and high dose groups, respectively, were fed the drug at levels 0.0, 0.5, and 1.0 percent of their diet. Proportions of animals in each of the species/sex strata with leukemia or lymphoma are reported in Table 3. To illustrate the efficiency of our algorithms, we fit the following to the data.

$$\log\left\{\frac{p_g}{1-p_g}\right\} = \beta_0 + \beta_1 Gender_g + \beta_2 Species_g + \beta_3 Dose_g. \tag{11}$$

In this case, the sufficient statistics for β_0, β_1, β_2, and β_3 are respectively given by $T_1 = \sum_{g=1}^{12} Y_i$; $T_2 = \sum_{g=7}^{12} Y_i$; $T_3 = Y_4 + Y_5 + Y_6 + Y_{10} + Y_{11} + Y_{12}$; and $T_4 = Y_2 + 2Y_3 + Y_5 + 2Y_6 + Y_8 + 2Y_9 + Y_{11} + 2Y_{12}$.

Table 3. Carcinogenesis bioassay of tolazamide.

Gender	Species	Dosage	No. of Animals Developed Disease	Total no. of Animals
0	0	0	4	14
0	0	1	5	35
0	0	2	1	34
0	1	0	2	15
0	1	1	1	35
0	1	2	4	35
1	0	0	6	15
1	0	1	2	33
1	0	2	4	34
1	1	0	4	15
1	1	1	3	33
1	1	2	2	35

Note: Gender, 0 = male, 1 = female; Species, 0 = mice, 1 = rat; Dosage, 0 = control, 1 = low, 2 = high.
Source: Tarone and Gart (1980).

The corresponding observed values of the sufficient statistics are given by 38, 21, 16, and 33. There are 3,672,542 sequences which satisfy $\sum_{j=1}^{G} y_j \boldsymbol{x}_j = \boldsymbol{t}^*$. Among these sequences, only 88,257 of them satisfy the inequality $\sum_{j=1}^{G} \ln[b(n_j, y_j)] > \sum_{j=1}^{G} \ln[b(n_j, y_j^*)]$. To check if the proposed logistic model fits the data, we run the exhaustive enumeration procedure, the algorithm based on nested DO loops, and the recursive algorithm. Exhaustive enumeration took 2,980 seconds to produce $P_{\boldsymbol{y}^*} = 0.1965$. The algorithm based on nested DO loops took 80 seconds to obtain the same result, while the recursive algorithm took only 30 seconds.

Now that the underlying logistic model (11) fits the data, further analyses and conclusions based on this model are warranted.

5. Discussion

We have considered an exact conditional test for assessing the goodness-of-fit of the logistic regression model. Our test is different from the previous goodness-of-fit tests proposed by Hosmer and Lemeshow (1980) and Lipsitz, Fitzmaurice and Molenberghs (1996) in two respects. First, we do not require the partitioning of the space of covariates or subjects into regions or groups. Second, we consider exact analysis rather than asymptotic analysis. The advantage of our method is its reliability in small samples. Moreover, our proposed algorithms can be readily modified to give an exact P-value for other goodness-of-fit statistics, such as the Pearson chi-square or the deviance statistic.

The existing statistical package LogXact-4 (1999) provides another option for producing the desired exact P-value. However, in the two examples we considered, it failed to produce the results due to insufficient computer memory and it took a long time to report failure. We present two efficient algorithms for calculating the exact P-value of the conditional goodness-of-fit test. They are the algorithm based on nested DO loops, and the recursive algorithm. The former algorithm is the one proposed by Bedrick and Hill (1992). But, we do not store the entire reference set. The advantage of using this method is that it is ideally suited for a nested-Do-loop program, say, in FORTRAN. More importantly, this method does not require additional memory to store or retrieve intermediate records. Theoretically, it can be used for problems of any size and any number of covariates. From our experience, the recursive algorithm is more efficient than the algorithm based on nested DO loops when the sample size and the number of covariates are large. However, the recursive algorithm may require extensive memory to manipulate the hashing table used for implementing the recursions (9) and (10). For details in implementation of the hashing table, consult Hirji et al. (1987).

In closing, we would like to discuss various factors that would affect the efficiency and applicability of the proposed algorithms. (i) The efficiency of the proposed algorithms depend heavily on data labeling. According to our experience, processing the algorithms with sample sizes already sorted in descending order usually accelerates the computation. This finding agrees with those of Bedrick and Hill (1992), Hirji and Vollset (1994), and Hirji et al. (1996). (ii) Type of covariates is another critical factor in efficiency. As noted in Section 3, if all the covariates under consideration are binary, then any sequence $\{y_1, \ldots, y_G\}$ generated by the nested DO loops (6) will automatically satisfy $\sum_{j=1}^{G} y_j \boldsymbol{x}_j = \boldsymbol{t}^*$ and computational time is saved. For categorical covariates, efficiency decreases

as the number of categories increases. We have found our algorithms computationally practical for categorical covariates with 4 to 5 categories. (iii) For covariates that are continuous, each observation in the reference set may possess the same conditional probability. As a result, the conditional probabilities test (and even the deviance test) may provide no information about lack-of-fit. In this case, we suggest the conversion of the measurements to binary or ordinal types. (v) According to our experience, the proposed algorithms are applicable to problems with 4 to 6 covariates and total sample size up to 300 observations. Applicability could be considerably extended to data that are extremely unbalanced or highly grouped. Also, we note that problems in which the observed response rate is close to 0 or 1 are less time consuming.

Acknowledgement

The work described in this paper was fully supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region (Project No. CUHK4170/99M). I would like to express my sincere thanks to Lauren Nelson for her kind assistance in running the two examples in LogXact, to the Editor and to two referees for several helpful comments and useful suggestions which led to an improved manuscript.

References

Bedrick, E. J. and Hill, J. R. (1992). Discussion on "A survey of exact inference for contingency tables" by Agresti, A. *Statist. Sci.* **7**, 153-157.

Demo, D. H. and Parker, K. D. (1987). Academic achievement and self-esteem among black and white college students. *J. Soc. Psych.* **127**, 345-355.

Garcia-Perex, M. A. (1995). Algorithm AS 320: decomposing an integer N into all sets of J positive integer summands by simulation of dynamically varying nested DO loops. *Appl. Statist.* **46**, 522-533.

Forster J. J., McDonald, J. W. and Smith, P. W. F. (1996). Monte Carlo exact conditional tests for log-linear and logistic models. *J. Roy. Statist. Soc. Ser. B* **58**, 445-453.

Hadley, G. (1964). *Nonlinear and Dynamic Programming.* Addison-Wesley, Reading, MA.

Hirji, K. F. (1992). Exact distributions for polytomous data. *J. Amer. Statist. Assoc.* **87**, 487-492.

Hirji, K. F., Mehta, C. R. and Patel, N. R. (1987). Computing distributions for exact logistic regression. *J. Amer. Statist. Assoc.* **82**, 1110-1117.

Hirji, K. F. and Vollset, S. E. (1994). Computing exact distributions for several 2×2 tables. *Appl. Statist.* **45**, 270-274.

Hirji, K. F., Vollset, S. E., Reis, I. M. and Afifi, A. A. (1996). Exact tests for interaction in several 2×2 tables. *J. Comput. Graph. Statist.* **5**, 209-224.

Hosmer, D. W. and Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Comm. Statist. Theory Method* **10**, 1043-1069.

Hosmer, D. W. and Lemeshow, S. (1989). Applied logistic regressions. Wiley, New York.

Lipsitz, S. R., Fitzmaurice, G. M. and Molenberghs, G. (1996). Goodness-of-fit tests for ordinal response regression models. *J. Roy. Statist. Soc. Ser. B* **58**, 175-190.

LogXact-4 for Windows (1999). *Software for Exact Logistic Regression*. Cytel Software Corporation, Cambridge, MA.

McCullagh, P. (1986). The conditional distribution of goodness-of-fit statistics for discrete data. *J. Amer. Statist. Assoc.* **81**, 104-107.

Mehta, C. R. and Patel, N. R. (1980). A network algorithm for exact treatment of the 2×K contingency table. *Commun. Statist. Simulation Comput.* **B9**, 649-664.

Mehta, C. R. and Patel, N. R. (1995). Exact logistic regression: theory and examples. *Statist. Medicine* **14**, 2143-2160.

O'Flaherty, M. and MacKenzie, G. (1982). Algorithm AS 172: Direct simulation of nested Fortran DO-LOOPS. *Appl. Statist.* **31**, 71-74.

StatXact-4 for Windows (1999). *Software for Exact Nonparametric Inference*. Cytel Software Corporation, Cambridge, MA.

Tang, M. L., Hirji, K. F. and Vollset, S. E. (1995). Exact power computation for dose-response studies. *Statist. Medicine* **14**, 2261-2272.

Tarone, R. E. and Gart, J. J. (1980). On the robustness of combined tests for trends in proportions. *J. Amer. Statist. Assoc.* **75**, 110-116.

Tritchler, D. (1984). An algorithm for exact logistic regression. *J. Amer. Statist. Assoc.* **79**, 709-711.

Tsiatis, A. A. (1980). A note on a goodness-of-fit test for the logistic regression model. *Biometrika* **67**, 250-251.

Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong.

E-mail: tang@sparc20b.sta.cuhk.edu.hk

(Received March 1999; accepted June 2000)