Markov Chain Monte Carlo Exact Inference for Binomial and Multinomial Logistic Regression Models

Jonathan J. Forster,[1] John W. McDonald,[2] and Peter W. F. Smith[2]

Abstract

We develop Metropolis-Hastings algorithms for exact conditional inference, including goodness-of-fit tests, confidence intervals and residual analysis, for binomial and multinomial logistic regression models. We present examples where the exact results, obtained by enumeration, are available for comparison. We also present examples where Monte Carlo methods provide the only feasible approach for exact inference.

Keywords: Conditional inference; Exact confidence interval; Exact test; Goodness-of-fit test; Logistic regression; Markov chain Monte Carlo; Metropolis-Hastings algorithm; Residual analysis.

1 Introduction

Asymptotic inference for logistic regression models, particularly goodness of fit testing, is often unreliable for sparse data. In this case, it is preferable to base inference on an exact conditional distribution, rather than on a large-sample approximation such as the chi-squared distribution. Exact conditional inference is based on the distribution of sufficient statistics for the parameters of interest given sufficient statistics for the nuisance parameters. This distribution does not depend on the nuisance parameters. See Agresti (1992) and accompanying discussion for a review of exact inference for log-linear and logistic regression models.

For inference concerning parameters of a logistic regression model, calculation of the exact conditional distribution is conceptually simple. However, enumerating this distribution may be infeasible. Monte Carlo sampling from the exact conditional distribution enables estimation of quantities of interest, such as tail areas and hence exact

---

[1] Department of Mathematics, University of Southampton, Southampton, SO17 1BJ, U.K.

[2] Department of Social Statistics, University of Southampton, Southampton, SO17 1BJ, U.K.

conditional p-values. In general, direct generation from the exact conditional distribution is infeasible.

Markov chain Monte Carlo (MCMC) methods can be used to generate samples from complex or high-dimensional multivariate distributions, typically only known up to a constant of proportionality. Markov chain Monte Carlo methods for exact inference involve the construction of a Markov chain, which is straightforward to generate from, and whose equilibrium distribution is the exact conditional distribution of interest. Ideally the Markov chain should be irreducible, i.e., it must be possible to reach any point in the sample space from any other by transitions of the chain.

Diaconis and Sturmfels (1998) described a general theory for constructing MCMC algorithms for sampling from conditional distributions. Their approach, based on Gröbner bases, can in principle be applied to any discrete regression problem. However, our experience is that in many situations computing the required basis is impracticable. In this paper, we develop Metropolis-Hastings algorithms for exact inference for binomial and multinomial logistic regression models. The proposed algorithms are intuitive and easy to construct. However, the resulting Markov chain is not necessarily irreducible. When the Markov chain is not irreducible, the resulting inferences may still be interpreted as exact conditional inferences, where conditioning is also on being in a particular reduced component of the sample space. For example, comparing the resulting MCMC p-value with a fixed significance level still results in a test of exact size. This is essentially the MCMC approach developed by Besag and Clifford (1989). In this paper we present examples where the results of our algorithms are consistent with exact conditional inference obtained by enumeration.

2 Binomial Logistic Regression Models

Let $y_i \sim \text{binomial}(m_i, \pi_i)$ and $\text{logit}(\pi_i) = \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\gamma}$, $i = 1, \ldots, n$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a vector of regression parameters corresponding to the vector of p explanatory variables $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^T$ is a vector of regression parameters corresponding to the vector of q explanatory variables $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iq})^T$. The conditional distribution of the vector of responses $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, given $\boldsymbol{X}^T\boldsymbol{y}$, the vector of sufficient statistics for $\boldsymbol{\beta}$, is given by

$$f(\boldsymbol{y}|\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{X}^T\boldsymbol{y}_{obs}; \gamma) \propto \exp(\gamma^T \boldsymbol{Z}^T \boldsymbol{y}) \prod_{i=1}^{n} \binom{m_i}{y_i}, \tag{1}$$

where $\boldsymbol{X} = (x_{ij})$ is a $n \times p$ matrix, $\boldsymbol{y}_{obs}$ denotes the observed vector of responses and the right hand side of (1) is subject to the conditioning constraints. Exact conditional inference concerning γ is based on the marginal distribution of $\boldsymbol{Z}^T\boldsymbol{y}$, the vector of sufficient statistics for γ, where $\boldsymbol{Z} = (z_{ij})$ is a $n \times q$ matrix.

Consider an exact conditional test of the hypothesis $\mathrm{H}_{\boldsymbol{\gamma}}$: $\gamma = \gamma^0$. When $q = 1$, the test is based on the marginal distribution of the univariate statistic $\boldsymbol{Z}^T\boldsymbol{y}$ where $\gamma = \gamma^0$. When $q > 1$, the test is usually based on a one-dimensional function of the multivariate statistic $\boldsymbol{Z}^T\boldsymbol{y}$, such as a likelihood ratio statistic. A test of goodness of fit of the model $\text{logit}(\pi_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}$, $i = 1, \ldots, n$, is a special case of a test of $\mathrm{H}_{\boldsymbol{\gamma}}$, where the columns of $\boldsymbol{X}$ and $\boldsymbol{Z}$ span $\mathrm{I\!R}^n$ and $\gamma^0 = \boldsymbol{0}$. In this case, the exact conditional distribution is equivalent to (1) with $\gamma = \boldsymbol{0}$. When $q = 1$, an exact confidence interval for γ can be constructed by inverting a test of $\mathrm{H}_{\boldsymbol{\gamma}}$; see for example Agresti (1992).

For a binomial logistic regression example, Bedrick and Hill (1990) enumerated all possible responses consistent with the observed values of the sufficient statistics for the regression parameters in order to perform exact inference based on (1). However, enumeration of the support of (1), that is all $\boldsymbol{y}$ satisfying the conditioning constraints $\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{X}^T\boldsymbol{y}_{obs}$, is only feasible for moderate-sized datasets and therefore, as discussed by Bedrick and Hill (1991), an efficient algorithm to generate from the exact conditional distribution is required.

Monte Carlo exact inference is based on a sample generated from (1). For example, when $q = 1$ an exact p-value for H_γ is estimated using a tail area of the empirical distribution of $\boldsymbol{Z}^T\boldsymbol{y}$. The lower (upper) end point of an exact $(1 - 2\alpha)$ confidence interval for γ can be estimated by finding the value γ^0 such that the observed value of $\boldsymbol{Z}^T\boldsymbol{y}$ is the upper (lower) α quantile of the empirical distribution. Note that given a sample from (1) for $\gamma = \gamma^*$, the exact p-value under H_γ: $\gamma = \gamma^0$ can be estimated by weighting the sample by $\exp\{(\gamma^0 - \gamma^*)\boldsymbol{Z}^T\boldsymbol{y}\}$. Hence, in principle, a grid search for both end points of a confidence interval may be based on a single Monte Carlo sample. A natural choice is $\gamma^* = 0$, if a Monte Carlo test of $\gamma = 0$ has already been performed. Alternatively, $\gamma^* = \hat{\gamma}$, the maximum likelihood estimate, is a value which is supported by the observed data.

Direct Monte Carlo generation is typically infeasible as, in general, this also requires enumeration of the support of (1). Mehta, Patel and Senchaudhuri (2000) proposed a network-based direct Monte Carlo sampling approach for binomial logistic regression. This approach efficiently represents the support of (1) by a network which must be constructed and stored. Unfortunately, memory requirements often exceed availability even for data sets of modest size. For a binomial logistic regression model with a single covariate, Forster, McDonald and Smith (1996) used Gibbs sampling to perform exact inference. However, the resulting Markov chain may be reducible. For a similar example, Diaconis and Sturmfels (1998) used their Gröbner basis approach to obtain an irreducible Markov chain. For more complex examples neither of these approaches may be practical: the Gibbs sampler tends to become highly reducible and finding the Gröbner basis is not feasible.

3 Metropolis-Hastings Sampling

Metropolis-Hastings sampling (Hastings, 1970) is a MCMC method for sampling from arbitrary, possibly unnormalized, multivariate distributions. The procedure is:

1. Given current value $\boldsymbol{y}$, generate a new value $\boldsymbol{y}'$ from some probability distribution $q(\boldsymbol{y}, \boldsymbol{y}')$.

2. Accept $\boldsymbol{y}'$ as the next realization of the chain with probability

$$a(\boldsymbol{y}, \boldsymbol{y}') = \min\left\{\frac{f(\boldsymbol{y}')q(\boldsymbol{y}', \boldsymbol{y})}{f(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{y}')}, 1\right\} \tag{2}$$

   otherwise, retain $\boldsymbol{y}$.

Provided q is chosen appropriately, then f is the equilibrium distribution for this chain. See Smith and Roberts (1993) and references therein for details.

For Monte Carlo exact conditional inference for logistic regression models, the required distributions are margins of (1). Therefore, if we construct a Metropolis-Hastings algorithm whose equilibrium distribution is (1), samples from the required marginal distributions may be extracted. Given a current vector of responses $\boldsymbol{y}$, we generate a new vector $\boldsymbol{y}'$ so as to maintain the sufficient statistics for the nuisance parameters. Our proposed Metropolis-Hastings algorithm involves setting $\boldsymbol{y}' = \boldsymbol{y} + d\boldsymbol{v}$, where $0 \leq y_i' \leq m_i$ for all i, d is an integer and $\boldsymbol{v}$ is a vector of integers such that $\boldsymbol{X}^T\boldsymbol{v} = \boldsymbol{0}$, thus maintaining the sufficient statistics for the nuisance parameters. Diaconis, Graham and Sturmfels (1996) proposed a Markov chain of this form, and presented theory, based on primitive partition identities, for ensuring irreducibility. Takken (1999) noted that there is no easy way to generate these primitive partition identities, so a practical method for constructing an irreducible Markov chain for a general logistic regression model does not yet exist.

We propose algorithms which are intuitive and easy to construct, although not necessarily irreducible. Initially we enumerate all possible $\boldsymbol{v}$ subject to the constraint $\sum_{i=1}^n |v_i| \leq r$ for a given r, chosen so that the enumeration is feasible. We allow only those $\boldsymbol{v}$ where the v_i are coprime. Usually, the vector of ones is in the column space of $\boldsymbol{X}$, as a constant term is included in the model, so $\sum_{i=1}^n v_i$ must equal zero and therefore, $\sum_{i=1}^n |v_i|$ must be even. Typically, we choose r to be 4, 6 or 8.

Our Metropolis-Hastings algorithm involves selecting one of the possible $\boldsymbol{v}$ with equal probability and then generating d using

$$q(d|\boldsymbol{v}) \propto \exp\{\gamma^T \boldsymbol{Z}^T(\boldsymbol{y} + d\boldsymbol{v})\} \prod_{i=1}^{n} \binom{m_i}{y_i + dv_i}, \tag{3}$$

where $0 \leq y_i + dv_i \leq m_i$ for all i. This expression reduces to a product of at most $r + 1$ terms, as terms where $v_i = 0$ are absorbed into the constant of proportionality. Furthermore, for sparse data the support of (3) is small. Hence, obtaining the normalizing constant for $q(d|\boldsymbol{v})$ by enumeration is straightforward.

The proposal distribution $q(\boldsymbol{y}, \boldsymbol{y}')$ is given by $q(d|\boldsymbol{v})q(\boldsymbol{v})$, which is proportional to $q(d|\boldsymbol{v})$, since $q(\boldsymbol{v})$ is uniform. By comparing (1) and (3) we see that $q(d|\boldsymbol{v}) \propto f(\boldsymbol{y}')$, so $q(\boldsymbol{y}, \boldsymbol{y}') \propto f(\boldsymbol{y}')$ for every $\boldsymbol{y}$, $\boldsymbol{y}'$. As transitions from $\boldsymbol{y}'$ to $\boldsymbol{y}$ can be made only by selecting $-(d\boldsymbol{v})$, we have $q(\boldsymbol{y}', \boldsymbol{y}) \propto f(\boldsymbol{y})$ with the same constant of proportionality. Hence, $a(\boldsymbol{y}, \boldsymbol{y}') = 1$ and there is no rejection step.

4 Binomial Examples

We present two examples: one where the exact results, obtained by enumeration, are available for comparison and one where MCMC provides the only feasible method for

Table 1: Dose-response Data

Log-dose	m_i	y_i
0.301	19	19
0.000	20	18
−0.301	19	19
−0.602	21	14
−0.903	19	15
−1.208	20	4
−1.509	16	0
−1.807	19	0
−2.108	40	0
−2.710	81	2

exact inference. For these examples, and those in Section 6, each estimated exact conditional p-value is based on a sample of one million realizations. The Monte Carlo error of estimated p-values is summarised by approximate 99% confidence intervals, calculated using the method of batch means (Geyer, 1992).

4.1 Dose-response Experiment

Table 1, taken from Bedrick and Hill (1990), presents data from a dose-response experiment on the tumorigenicity of benzepyrene in mice. Doses were approximately equally spaced on a log-dose scale, except that no observations were made at log(dose) ≈ -2.4. We test the goodness of fit of the logistic regression model with log(dose) as an explanatory variable using both L^2, the likelihood ratio statistic, and X^2, the Pearson chi-squared statistic. We use $r = 8$ in the Metropolis-Hastings algorithm. The observed values of L^2 and X^2 are displayed in Table 2 along with the asymptotic and estimated exact conditional p-values. Also displayed are the exact conditional p-values obtained by enumeration (Michaelides, 1997, pp. 88–89). The MCMC estimates are in close agreement with the exact p-values. This is despite the fact that this chain is reducible, as it can be shown that no chain with $r < 14$ will be irreducible. This is an example of a reducible chain which still provides a reasonable inference.

Bedrick and Hill (1990) used an exact outlier test to determine whether lack of fit is due to a single extreme observation, say the ith observation. One of their tests uses

Table 2: Test Statistics and P-values for Table 1

	Observed value	df	Asymptotic p-value	Estimated exact p-value	Exact p-value
L^2	26.68	8	0.0008	0.0064	0.0064
X^2	32.10	8	0.0001	0.0116	0.0132

the marginal conditional probability $w_i = f(y_i|\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{X}^T\boldsymbol{y}_{obs})$ as a test statistic with corresponding p-value $p(w_i) = Pr(w_i \leq w_{i\,obs}|\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{X}^T\boldsymbol{y}_{obs})$. In order to perform simultaneous inference, they proposed using the test statistic $T = \min\{p(w_1), \ldots, p(w_n)\}$. They obtained an observed value for T of 0.04, corresponding to the observation at log(dose)$= -0.602$, and enumerated the exact p-value as 0.13. We use our Markov chain to estimate first $w_1, \ldots, w_{10}$, second $p(w_1), \ldots, p(w_{10})$, third T, and finally the exact p-value. Our MCMC estimate of the exact p-value is 0.11, which is in close agreement with the enumerated value.

4.2 Greying of the Hair and Mortality

Lasker and Kaplan (1974) studied the relationship between hair greyness in 1948 (none, slight, moderate, general) and natural death between 1948 and 1969, controlling for sex and age. Their data, on a cohort of 469 adult Mexicans, are presented in Table 3. Note that y_i is the number of deaths for each of the 65 observed covariate patterns where the number of observations, m_i, is positive. We consider two logistic regression models for mortality: SEX+AGE and SEX+AGE+GREY, where AGE (1 to 12) and GREY (1 to 4) are covariates with equally-spaced scores. We used $r = 4$ in the Metropolis-Hastings algorithms, which resulted in 50 495 and 9 697 different $\boldsymbol{v}$ for the tests of goodness of fit of the SEX+AGE and SEX+AGE+GREY models respectively. For both models, the observed goodness-of-fit statistics L^2 and X^2 are presented in Table 4 along with the asymptotic and MCMC estimated exact p-values. In all cases, the estimated exact p-values suggest a better fit than do the corresponding asymptotic p-values. Note that the SEX+AGE+GREY model is rejected at the 5% level using the asymptotic p-value for L^2, but not using the estimated exact p-value.

In order to compare the two models, we perform a test of H_γ, as described in Section 2,

Table 3: Greying of the Hair and Mortality

	Sex	Male								Female							
	Greyness	None		Slight		Moderate		General		None		Slight		Moderate		General	
		y_i	m_i	y_i	m_i	y_i	m_i	y_i	m_i	y_i	m_i	y_i	m_i	y_i	m_i	y_i	m_i
	17–24	1	46	0	1					2	34						
	25–29	1	29							0	21	0	1				
	30–34	3	23	0	3					1	13						
	35–39	4	33	3	7					0	23	0	5	0	1		
	40–44	2	12	3	12	0	2			0	11	0	2	1	1	1	1
Age	45–49	1	12	5	15	3	7	0	2	0	8	4	7			0	3
	50–54	1	1	3	16	0	1	5	8	0	3	1	7	0	2	1	4
	55–59	1	2	5	6	1	4	3	9	1	2	0	6	1	4	3	7
	60–64	0	3	1	4	3	6	3	6	1	1	0	2			1	1
	65–69					2	3	3	5			0	1	0	2	0	1
	70–74			1	1	2	2	3	4			1	1			2	2
	75+					2	2	3	3			1	1	1	1		

Table 4: Goodness-of-fit Test statistics and P-values for Table 3

Model	Observed statistic	df	Asymptotic p-value	Estimated exact p-value*
SEX+AGE	$L^2 = 87.80$	62	0.0172	0.0487 ± 0.0059
	$X^2 = 85.81$	62	0.0244	0.0518 ± 0.0054
SEX+AGE+GREY	$L^2 = 84.01$	61	0.0270	0.0959 ± 0.0091
	$X^2 = 77.05$	61	0.0806	0.0973 ± 0.0089

* with approximate 99% confidence interval

where γ is the greyness score parameter. Therefore, we extract from the Markov chain for the SEX+AGE model a sample of $\boldsymbol{Z}^T\boldsymbol{y}$, the sufficient statistic for γ. Here $\boldsymbol{Z}$ is a 65×1 matrix containing the values of the greyness scores. The exact p-value is estimated by ranking the observed value of the sufficient statistic, $\boldsymbol{Z}^T\boldsymbol{y}_{obs} = 235$, among this sample. Against the one-sided alternative that hair greyness is deleterious, the estimated exact conditional p-value is 0.0314 ± 0.0068.

Lack of fit of a model may be investigated by residual analysis. The deviance residual divided by its standard error has an asymptotic standard normal distribution (McCullagh and Nelder, 1989). We examine the standardized deviance residuals for the SEX+AGE model in order to investigate whether the lack of fit is due to a small

Table 5: Standardized Deviance Residuals and P-values for Table 3

		Covariate values			Deviance	Asymptotic	Estimated	Support
y_i	m_i	Sex	Age	Greyness	residual	p-value	exact p-value	points
0	3	Male	60–64	None	−2.229	0.0258	0.0878	4
3	7	Male	35–39	Slight	2.008	0.0446	0.0484	6
4	7	Female	45–49	Slight	2.703	0.0068	0.0075	7
1	1	Female	40–44	Moderate	2.161	0.0306	0.0978	2
1	1	Female	40–44	General	2.161	0.0306	0.1000	2

number of extreme cases. Residuals greater than two in absolute value are presented in Table 5, along with their asymptotic and estimated exact p-values. The estimated exact conditional p-values are calculated using the empirical distribution of each residual, extracted from the MCMC sample used to estimate the p-value for the goodness-of-fit test. Also presented in Table 5 are the number of support points in the empirical distribution of each residual. The estimated exact p-values for the 65 residuals do not indicate that the lack of fit is due to a small number of extreme cases. Note that the asymptotic p-values are closer to the estimated exact p-values for the empirical distributions with the larger numbers of support points.

The estimated exact 95% confidence intervals for the greyness score parameter are $(-0.015, 0.613)$ and $(-0.010, 0.600)$ using $\gamma^* = 0$ and $\gamma^* = \hat{\gamma} = 0.295$ respectively. These show relatively little sensitivity to the choice of γ^* and are similar to the asymptotic interval $(-0.001, 0.592)$.

5 Multinomial Logistic Regression Models

We now consider a polytomous response with categories $0, \ldots, K$. The ith observation may be represented by the $K+1$ counts $(y_{i0}, y_{i1}, \ldots, y_{iK})$, $i = 1, \ldots, n$, with the total count $m_i = \sum_{k=0}^{K} y_{ik}$, assumed fixed. Let $\boldsymbol{Y} = (y_{ik})$ denote a $n \times K$ matrix of responses, where k runs from 1, and denote the K columns of $\boldsymbol{Y}$ by $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K$ with $\boldsymbol{y}_0 = \boldsymbol{m} - \sum_{k=1}^{K} \boldsymbol{y}_k$ where $\boldsymbol{m} = (m_1, \ldots, m_n)^T$. Let $(y_{i0}, y_{i1}, \ldots, y_{iK})$ be multinomially distributed with probabilities $(\pi_{i0}, \ldots, \pi_{iK})$, $i = 1, \ldots, n$. Then a baseline-category multinomial logistic regression model, with baseline category 0, is

$$\log\left(\frac{\pi_{ik}}{\pi_{i0}}\right) = \boldsymbol{x}_i^T \boldsymbol{\beta}_k + \boldsymbol{z}_i^T \gamma_k \qquad i = 1, \ldots, n, \quad k = 1, \ldots, K, \tag{4}$$

where $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp})^T$ is a vector of regression parameters corresponding to $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$ and $\boldsymbol{\gamma}_k = (\gamma_{k1}, \ldots, \gamma_{kq})^T$ is a vector of regression parameters corresponding to $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iq})^T$. We denote the $p \times K$ matrix $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$ by $\boldsymbol{B}$ and the $q \times K$ matrix $(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_K)$ by $\boldsymbol{\Gamma}$.

When the response is ordinal, it may be more appropriate to express this model as the equivalent adjacent-category model

$$\log\left(\frac{\pi_{ik}}{\pi_{i\,k-1}}\right) = \boldsymbol{x}_i^T\boldsymbol{\beta}'_k + \boldsymbol{z}_i^T\boldsymbol{\gamma}'_k \qquad i = 1, \ldots, n, \quad k = 1, \ldots, K, \tag{5}$$

where $\boldsymbol{B}' = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_K) = \boldsymbol{B}\boldsymbol{S}^{-T}$, $\boldsymbol{\Gamma}' = (\boldsymbol{\gamma}'_1, \ldots, \boldsymbol{\gamma}'_K) = \boldsymbol{\Gamma}\boldsymbol{S}^{-T}$ and $\boldsymbol{S}$ is a lower triangular matrix of ones. The matrix of sufficient statistics for $\boldsymbol{B}'$ is $\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{S}$. See Hirji (1992) for details. As $\boldsymbol{S}$ is invertible, $\boldsymbol{X}^T\boldsymbol{Y}$ is also sufficient here.

For model (4), the conditional distribution of $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K)$, given $\boldsymbol{X}^T\boldsymbol{Y}$, the matrix of sufficient statistics for $\boldsymbol{B}$ (or $\boldsymbol{B}'$), is

$$f(\boldsymbol{Y}|\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{Y}_{obs}; \boldsymbol{\Gamma}) \propto \exp\left\{\mathrm{tr}\,(\boldsymbol{\Gamma}^T\boldsymbol{Z}^T\boldsymbol{Y})\right\} \prod_{i=1}^{n} \binom{m_i}{y_{i0} \;\cdots\; y_{iK}}, \tag{6}$$

where $\boldsymbol{Y}_{obs}$ denotes the observed matrix of responses and the right hand side is subject to the conditioning constraints. Here, exact conditional inference is based on the marginal distribution of a univariate function of $\boldsymbol{Z}^T\boldsymbol{Y}$, the matrix of sufficient statistics for $\boldsymbol{\Gamma}$. For model (5), exact conditional inference is based on the distribution of a univariate function of $\boldsymbol{Z}^T\boldsymbol{Y}\boldsymbol{S}$, the matrix of sufficient statistics for $\boldsymbol{\Gamma}'$. This is a marginal distribution of (6) where $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}'\boldsymbol{S}^T$. For a test of goodness of fit, the exact conditional distribution is equivalent to (6) where $\boldsymbol{\Gamma} = \boldsymbol{0}$.

Our Metropolis-Hastings algorithm for binomial logistic regression models can be extended to the multinomial logistic regression models above. The binomial is a special case of the multinomial where $K = 1$, $\boldsymbol{y}_1 = \boldsymbol{y}$ and $\boldsymbol{y}_0 = \boldsymbol{m} - \boldsymbol{y}$, and each proposed step of our binomial algorithm may be thought of as addition of $d\boldsymbol{v}$ to $\boldsymbol{y}_1$ together with subtraction of $d\boldsymbol{v}$ from $\boldsymbol{y}_0$. In the multinomial case, where there are $K + 1$ vectors of outcomes, $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_K$, a proposal is obtained by selecting at random a $\boldsymbol{v}$ such that $\boldsymbol{X}^T\boldsymbol{v} = 0$, and an integer vector $\boldsymbol{w} = (w_0, w_1, \ldots, w_K)^T$ of length $K + 1$ such that $\boldsymbol{1}_{K+1}^T\boldsymbol{w} = 0$, where $\boldsymbol{1}$ is a vector of ones of the given dimension. Then $\boldsymbol{Y}' = \boldsymbol{Y} + d\boldsymbol{v}\boldsymbol{w}_{\backslash 0}^T$ and $\boldsymbol{y}'_0 = \boldsymbol{y}_0 + dw_0\boldsymbol{v}$, where $\boldsymbol{w}_{\backslash 0} = (w_1, \ldots, w_K)^T$ and d is generated using

$$q(d|\boldsymbol{v}) \propto \exp\left(\mathrm{tr}\,\{\boldsymbol{\Gamma}^T\boldsymbol{Z}^T(\boldsymbol{Y} + d\boldsymbol{v}\boldsymbol{w}_{\backslash 0}^T)\}\right) \prod_{i=1}^{n}\prod_{k=0}^{K} \frac{1}{(y_{ik} + dv_iw_k)!}, \tag{7}$$

where $y_{ik} + dv_i w_k \geq 0$ for all i and k.

For computational convenience, the set of possible $\boldsymbol{w}$ is restricted to those for which $\sum_{k=0}^{K} |w_k| = 2$ and (7) reduces to a product of at most $2r+1$ terms, as terms where $v_i w_k = 0$ are absorbed into the constant of proportionality. The procedure is then equivalent to selecting at random $k_1, k_2 \in \{0, 1, \ldots, K\}$, $k_1 \neq k_2$, adding $d\boldsymbol{v}$ to $\boldsymbol{y}_{k_1}$ and subtracting $d\boldsymbol{v}$ from $\boldsymbol{y}_{k_2}$. Therefore, this is a simple extension of the binomial algorithm. As potential $\boldsymbol{v}$ only depend on $\boldsymbol{X}$, and not on K, enumeration of all possible $\boldsymbol{v}$ subject to the constraint $\sum_{i=1}^{n} |v_i| \leq r$ is as described in Section 3. As with (3), obtaining the normalizing constant for $q(d|\boldsymbol{v})$ by enumeration is straightforward and $a(\boldsymbol{Y}, \boldsymbol{Y}') = 1$.

Hirji (1992) considered the baseline-category multinomial logistic regression model

$$\log\left(\frac{\pi_{ik}}{\pi_{i0}}\right) = \theta_k + \boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{\gamma} \qquad i = 1, \ldots, n, \quad k = 1, \ldots, K, \tag{8}$$

and the adjacent-category model

$$\log\left(\frac{\pi_{ik}}{\pi_{i\,k-1}}\right) = \theta_k + \boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{\gamma} \qquad i = 1, \ldots, n, \quad k = 1, \ldots, K, \tag{9}$$

which are more parsimonious than (4) and (5) as the regression parameters do not depend on the category. The sufficient statistics for $(\theta_1, \ldots, \theta_K)$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $\boldsymbol{1}_n^T \boldsymbol{Y}$, $\boldsymbol{X}^T \boldsymbol{Y} \boldsymbol{1}_K$ and $\boldsymbol{Z}^T \boldsymbol{Y} \boldsymbol{1}_K$ for the baseline-category model and $\boldsymbol{1}_n^T \boldsymbol{Y} \boldsymbol{S}$, $\boldsymbol{X}^T \boldsymbol{Y} \boldsymbol{S} \boldsymbol{1}_K$ and $\boldsymbol{Z}^T \boldsymbol{Y} \boldsymbol{S} \boldsymbol{1}_K$ for the adjacent-category model. Note that models (8) and (9) are not equivalent and there is no one-to-one correspondence between the sufficient statistics for the respective regression parameters. However, the sufficient statistics for the respective category effects $(\theta_1, \ldots, \theta_K)$ are equivalent, and conditioning on their observed values fixes $\boldsymbol{1}^T \boldsymbol{y}_k$, the total count in each category $k = 0, 1, \ldots, K$.

The exact conditional distribution of the sufficient statistics for γ is proportional to the right hand side of (6), with the exponential term replaced by $\exp(\gamma^T \boldsymbol{Z}^T \boldsymbol{Y} \boldsymbol{1}_K)$ for model (8) and by $\exp(\gamma^T \boldsymbol{Z}^T \boldsymbol{Y} \boldsymbol{S} \boldsymbol{1}_K)$ for model (9). Hence, the only difference between the conditional distributions for tests of $\gamma = \boldsymbol{0}$ for models (8) and (9) and of $\boldsymbol{\Gamma} = \boldsymbol{0}$ for models (4) and (5) is the support imposed by the conditioning constraints. Again, when $q = 1$, a confidence interval for γ can be constructed by weighting a single Monte Carlo sample from the exact conditional distribution where $\gamma = \gamma^*$.

Our Metropolis-Hastings algorithm described above for models (4) and (5) needs modification for models (8) and (9). If the algorithm is based on a set of $\boldsymbol{v}$ such that

$\mathbf{1}_n^T \boldsymbol{v} = 0$, $\boldsymbol{X}^T\boldsymbol{v} = \mathbf{0}$ and $\sum_{i=1}^n |v_i| \leq r$ for a given r, then for any proposed $\boldsymbol{Y}' = \boldsymbol{Y} + d\boldsymbol{v}\boldsymbol{w}_{\backslash 0}^T$, $\mathbf{1}_n^T\boldsymbol{Y}' = \mathbf{1}_n^T\boldsymbol{Y}_{obs}$ and $\boldsymbol{X}^T\boldsymbol{Y}' = \boldsymbol{X}^T\boldsymbol{Y}_{obs}$. Hence, the constraints on the sufficient statistics for the category effects and for the regression parameters for either models (8) or (9) are maintained. However, the constraint $\boldsymbol{X}^T\boldsymbol{Y}' = \boldsymbol{X}^T\boldsymbol{Y}_{obs}$ is over-restrictive, since we only require $\boldsymbol{X}^T\boldsymbol{Y}'\mathbf{1}_K = \boldsymbol{X}^T\boldsymbol{Y}_{obs}\mathbf{1}_K$ for the baseline-category model and $\boldsymbol{X}^T\boldsymbol{Y}'\boldsymbol{S}\,\mathbf{1}_K = \boldsymbol{X}^T\boldsymbol{Y}_{obs}\boldsymbol{S}\,\mathbf{1}_K$ for the adjacent-category model. Therefore, additional proposals are required which allow $\boldsymbol{X}^T\boldsymbol{Y}$ to vary, while maintaining either $\boldsymbol{X}^T\boldsymbol{Y}\mathbf{1}_K$ or $\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{S}\,\mathbf{1}_K$.

We consider an extra set of integer vectors $\boldsymbol{u}$ where $\mathbf{1}_n^T\boldsymbol{u} = 0$ but $\boldsymbol{X}^T\boldsymbol{u} \neq \mathbf{0}$. For computational convenience, we only consider vectors $\boldsymbol{u}$ for which $\sum_{i=1}^n |u_i| = 2$. In our modified algorithm, a $\boldsymbol{v}$ or a $\boldsymbol{u}$ is selected at random from the combined set of all feasible $\boldsymbol{v}$ and $\boldsymbol{u}$. If a $\boldsymbol{v}$ is selected, then the algorithm proceeds as before. If a $\boldsymbol{u}$ is selected, then a $\boldsymbol{w}$ is again selected at random and the proposal is given by $\boldsymbol{Y}' = \boldsymbol{Y} + d\boldsymbol{u}\boldsymbol{w}_{\backslash 0}^T$ and $\boldsymbol{y}_0' = \boldsymbol{y}_0 + dw_0\boldsymbol{u}$, with d generated using (7). Clearly $\boldsymbol{w}^T\mathbf{1}_{K+1} = 0$, in order to maintain $\boldsymbol{y}_0' + \boldsymbol{Y}'\mathbf{1}_K = \boldsymbol{m}$. For the baseline-category model, $\boldsymbol{w}$ is selected such that $\boldsymbol{w}_{\backslash 0}^T\mathbf{1}_K = 0$ and $\sum_{k=1}^K |w_i| = 2$. This proposal maintains $\boldsymbol{Y}\mathbf{1}_K$ and hence $\boldsymbol{X}^T\boldsymbol{Y}'\mathbf{1}_K = \boldsymbol{X}^T\boldsymbol{Y}_{obs}\mathbf{1}_K$. For the adjacent-category model, $\boldsymbol{w}$ is selected such that $\boldsymbol{w}_{\backslash 0}^T\boldsymbol{S}\,\mathbf{1}_K = 0$ and $\sum_{k=0}^K |w_i| = 4$. This proposal maintains $\boldsymbol{Y}\boldsymbol{S}\mathbf{1}_K$ and hence $\boldsymbol{X}^T\boldsymbol{Y}'\boldsymbol{S}\mathbf{1}_K = \boldsymbol{X}^T\boldsymbol{Y}_{obs}\boldsymbol{S}\mathbf{1}_K$.

6 Multinomial Examples

We present three examples: two where the exact results, obtained by enumeration, are available for comparison and one where MCMC provides the only feasible method for exact inference.

6.1 Rhabdomyosarcoma Data

Hirji (1992) considered data on tumor type (embryonal, alveolar, pleomorphic), age (0: $\leq$ 15 years; 1: $>$ 15 years) and sex (0: male; 1: female) for patients with rhabdomyosarcoma. He considered the baseline-category model (8) with tumor type as the multinomial response, embryonal as baseline tumor type, and age as a covariate. He calculated the exact conditional distribution of the sufficient statistic, $\boldsymbol{Z}^T\boldsymbol{Y}\mathbf{1}_2$, for the sex parameter γ, where $\boldsymbol{Z}$ is a 4×1 matrix containing the values coding sex for each of the four covariate patterns. We use $r = 4$ in our Metropolis-Hastings algorithm. The distribution function of the exact distribution, calculated by Hirji (1992), and our MCMC estimate are presented in Figure 1a. The two distributions are so close that it is hard to discern any difference between them. The estimated exact 95% confidence intervals for γ are $(-1.92, 2.22)$ and $(-1.91, 2.21)$ using $\gamma^* = 0$ and $\gamma^* = \hat{\gamma} = 0.116$, respectively. These show relatively little sensitivity to the choice of γ^* and are very close to the true interval $(-1.94, 2.25)$, calculated using the exact conditional distribution presented by Hirji (1992). For comparison, note that the asymptotic confidence interval is $(-1.65, 1.89)$.

6.2 Neuroleptic Treatment Data

Hirji (1992) also presented data on patients undergoing neuroleptic treatment. The variables considered were final Clinical Global Impression (CGI) scale (3 ordinal categories), initial CGI (scored $0, 1, 2, 3$), sex (0: female; 1: male) and treatment period (0: two weeks; 1: three weeks). He considered the adjacent-category model (9) with final CGI as the ordinal response and initial CGI and sex as covariates. He calculated the exact conditional distribution of the sufficient statistic, $\boldsymbol{Z}^T\boldsymbol{Y}\boldsymbol{S}\mathbf{1}_2$, for the treatment period parameter γ, where $\boldsymbol{Z}$ is a 12×1 matrix containing the values coding treatment period for each of the 12 covariate patterns. We use $r = 4$ in our Metropolis-Hastings algorithm. The distribution function of the exact distribution, calculated by Hirji (1992), and our MCMC estimate are presented in Figure 1b. The two distributions are again so close that it is hard to discern any difference between them. The estimated exact 95% confidence intervals for γ are $(-1.40, 1.73)$ and $(-1.47, 1.74)$ using $\gamma^* = 0$ and $\gamma^* = \hat{\gamma} = 0.137$, respectively. Again, these show relatively little sensitivity to the choice of γ^* and are very close to the true interval $(-1.42, 1.74)$, calculated using the exact conditional distribution presented by Hirji (1992). Here the asymptotic confidence interval is $(-1.26, 1.53)$.

6.3 Pregnancy Outcome in Consanguineous Marriages

Schull (1958) studied pregnancy outcome in three districts of Shizuoka City, Japan, according to the degree of consanguinity between the parents. The outcomes of 6 358

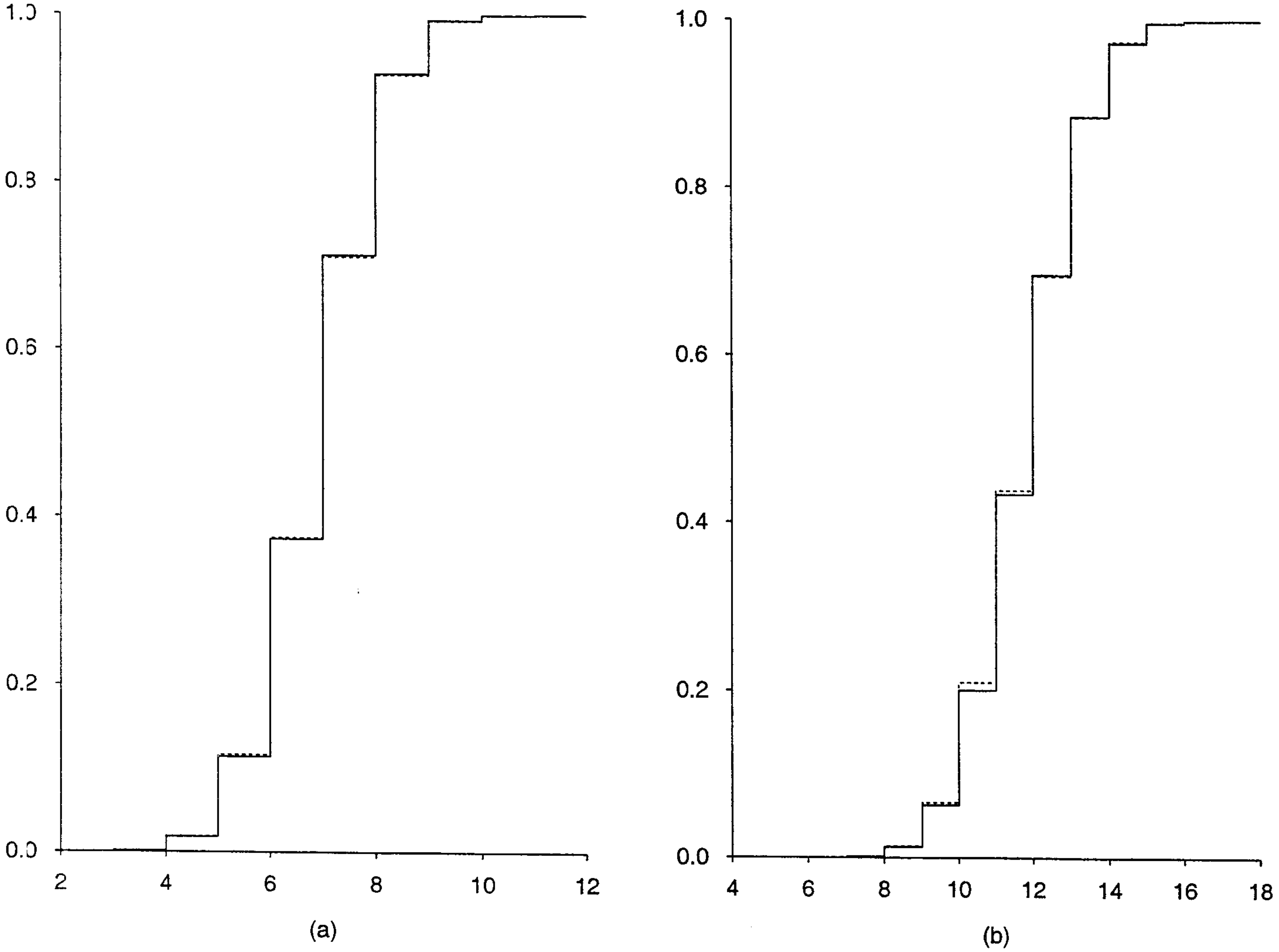

Figure 1: Exact (solid line) and estimated (dashed line) distribution functions for (a) rhabdomyosarcoma data and (b) neuroleptic treatment data.

Table 6: Pregnancy Outcome in Consanguineous Marriages

				Deaths		
		Abortion	Stillbirth	$\leq$ 12 months	13–60 months	Survived
Residence	Consanguinity	y_4	y_3	y_2	y_1	y_0
Rural district	No relation	27	15	57	25	834
	2nd cousins	1	1	13	6	139
	$1\frac{1}{2}$ cousins	3	2	7	2	51
	1st cousins	12	2	18	11	250
Intermediate district	No relation	67	20	128	76	2379
	2nd cousins	11	1	25	10	291
	$1\frac{1}{2}$ cousins	11	4	14	12	196
	1st cousins	23	6	40	27	558
Urban district	No relation	7	5	21	14	496
	2nd cousins	4	0	1	2	63
	$1\frac{1}{2}$ cousins	3	0	5	2	100
	1st cousins	7	1	15	11	226

pregnancies are presented in Table 6. Scores $0, 1, 2$ and 4 are assigned to the no relation to first cousins groups respectively, as the corresponding probabilities that any single locus is homozygous by descent from a common ancestor are $0/64, 1/64, 2/64$ and $4/64$.

We consider multinomial regression models (8) and (9) for pregnancy outcome, with district as a categorical explanatory variable. Note that, with the category labelling in Table 6, the baseline category for model (8) is 'Survived'. In order to test, for each model, that the consanguinity score parameter γ is zero, we extract from the respective Markov chain a sample of $\boldsymbol{Z}^T\boldsymbol{Y}\mathbf{1}_4$ or $\boldsymbol{Z}^T\boldsymbol{Y}\boldsymbol{S}\,\mathbf{1}_4$. Here $\boldsymbol{Z}$ is a 12×1 matrix containing the values of the consanguinity scores for each observation. The observed values of the sufficient statistic for the consanguinity score parameter are $\boldsymbol{Z}^T\boldsymbol{Y}_{obs}\mathbf{1}_4 = 897$ and $\boldsymbol{Z}^T\boldsymbol{Y}_{obs}\boldsymbol{S}\,\mathbf{1}_4 = 2\,034$, and the MCMC estimated exact p-values for the test against the one-sided alternative that consanguinity is deleterious are 0.0004 and 0.0010 respectively. Therefore, there is strong evidence that consanguinity has a significant effect on pregnancy outcome.

For the baseline category model, the estimated exact 95% confidence intervals for γ are $(0.033, 0.196)$ and $(0.039, 0.131)$ using $\gamma^* = 0$ and $\gamma^* = \hat{\gamma} = 0.085$, respectively. Here, there is a substantial discrepancy between the upper end points of the two intervals. The interval calculated using $\gamma^* = 0$ is unreliable because of the large difference between

Table 7: Goodness-of-fit Test Statistics and P-values for Table 6

	Observed statistic	df	Asymptotic p-value	Estimated exact p-value*
Model (8)	$L^2 = 40.00$	41	0.5150	0.8200 ± 0.0037
	$X^2 = 39.83$	41	0.5226	0.7478 ± 0.0063
Model (9)	$L^2 = 42.27$	41	0.4159	0.5293 ± 0.0170
	$X^2 = 43.11$	41	0.3811	0.3849 ± 0.0201
Model (4)	$L^2 = 32.06$	32	0.4638	0.5813 ± 0.0114
	$X^2 = 32.18$	32	0.4576	0.4633 ± 0.0128

* with approximate 99% confidence interval

zero and the upper end point. The interval calculated using $\gamma^* = \hat{\gamma}$ is close to the asymptotic interval $(0.039, 0.131)$. For the adjacent category model, the estimated exact 95% confidence intervals for γ are $(0.0108, 0.0674)$ and $(0.0109, 0.0468)$ using $\gamma^* = 0$ and $\gamma^* = \hat{\gamma} = 0.0294$, respectively. Again, there is a substantial discrepancy between the upper end points of the two intervals, due to the large difference between zero and the upper end point, and again the interval calculated using $\gamma^* = \hat{\gamma}$ is close to the asymptotic interval $(0.0115, 0.0472)$.

We can also perform an exact test of the goodness of fit of models (8) and (9), including both district and consanguinity score as explanatory variables. For both models, and for the more general model (4), the observed values of L^2 and X^2 are displayed in Table 7 along with the asymptotic and estimated exact p-values. By any of these criteria, all three models fit the data well, and choice between them will depend on which is considered more substantively plausible.

7 Discussion

The Metropolis-Hastings algorithms developed have proved efficient for exact inference for binomial and multinomial logistic regression models. In the examples where enumeration was feasible, our MCMC results are in close agreement with the results obtained by enumeration. One advantage of estimating, rather than enumerating, an exact distribution is that the required computational effort is much less dependent on the sample size. As expected, the discrepancy between asymptotic and estimated exact p-values and confidence intervals is largest when the data are sparse or unbalanced. Sparse data also commonly arise in matched case-control studies. Bedrick and Hill (1996) discuss exact methods of model checking for matched case-control studies, and our MCMC methods can be applied when enumeration is infeasible.

The exact confidence intervals estimated in this paper were obtained from a single Monte Carlo sample. When the sample is based on the maximum likelihood estimate, the observed value of the sufficient statistic for γ, the parameter of interest, typically is close to the median of the empirical distribution. Hence, when the empirical probabilities are reweighted so that the observed value of the sufficient statistic is either the upper or lower α quantile, approximately half the observations are in the tail. Therefore, we expect that the upper and lower end points of the interval are equally well estimated. Conversely, if the observed value of the sufficient statistic is an extreme quantile of the empirical distribution, there will be very few observations in one of the tails, and hence it is likely that at least one of the end points will be poorly estimated. Therefore, if a Monte Carlo test rejects the hypothesis $\gamma = 0$, the Monte Carlo sample is likely to be unreliable for estimating an exact confidence interval.

In certain examples, the Markov chain may not be irreducible. If a sufficiently large value of r is chosen, then clearly any $\boldsymbol{Y}$ satisfying the conditioning constraints may be obtained from any other in one transition of the Markov chain, and the chain is necessarily irreducible. For example, for a binomial logistic regression model with a single covariate, $r = 2\boldsymbol{y}_{obs}^T\mathbf{1}_n$ will permit all possible transitions. In practice, the chain will be irreducible for smaller values of r. For example, for binomial logistic regression with a single equally-spaced covariate we have proved that $r = 4$ is sufficient. Takken (1999) also proved this and several related results. The results of Diaconis, Graham and Sturmfels (1996) can be used to give sufficient conditions on r for irreducibility of MCMC algorithms for any logistic regression model. For example, for the dose-response data presented by Bedrick and Hill (1990), $r = 6\,022$ is sufficient although, as $\boldsymbol{y}_{obs}^T\mathbf{1}_n = 91$, it is clear that $r = 182$ allows transition between any two points in the sample space of the exact conditional distribution. However, using such high values of r is impractical, as enumeration of all possible $\boldsymbol{v}$ is typically infeasible.

For continuous covariates the required conditional distribution is often degenerate, as only the observed data satisfy the conditioning constraints. However, when covariate values are integer or evenly spaced, the exact conditional distribution is not usually degenerate. For the dose-response data presented by Bedrick and Hill (1990), log-dose may be truncated to one decimal place, resulting in equally-spaced values, except that there are no observations at log(dose) $= -2.4$. Then the conditional distribution for exact inference changes and the number of possible $\boldsymbol{v}$ increases from 268 to 1 924. The Markov chain used in Section 4.1 with $r = 8$ is now irreducible. The estimated exact p-value for the outlier test is 0.0256, and there is much stronger evidence of an outlier, with the most extreme observation at log(dose)$= -2.710$.

Acknowledgements

This research was supported by Economic and Social Research Council award R000236626, Methods of Exact Inference for Sparse Data.

References

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

Bedrick, E. J. and Hill, J. R. (1990). Outlier tests for logistic regression: A conditional approach. *Biometrika*, **77**, 815–827.

Bedrick, E. J. and Hill, J. R. (1991). Model checking for logistic regression: A conditional approach. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface,* ed. E. M. Keramidas, 208–214. Fairfax Station: Interface Foundation.

Bedrick, E. J. and Hill, J. R. (1996). Assessing the fit of the logistic regression model to individual matched sets of case-control data. *Biometrics*, **52**, 1–9.

Besag, J. and Clifford, P. (1989). Generalized Monte Carlo significance tests. *Biometrika*, **76**, 633–642.

Diaconis, P., Graham, R. L., and Sturmfels, B. (1996). Primitive partition identities. *Combinatorics. Paul Erdös is Eighty*, **2**, 173–192.

Diaconis, P., and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, **26**, 363–397.

Forster, J. J., McDonald, J. W., and Smith, P. W. F. (1996). Monte Carlo exact conditional tests for log-linear and logistic models. *Journal of the Royal Statistical Society Series B*, **58**, 445–453.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statistical Science*, **7**, 473–511.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

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

Lasker, G. W. and Kaplan, B. (1974). Graying of the hair and mortality. *Social Biology*, **21**, 290–295.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). London: Chapman and Hall.

Mehta, C. R., Patel, N. R. and Senchaudhuri, P. (2000). Efficient Monte Carlo methods for conditional logistic regression. *Journal of the American Statistical Association*, **95**, 99–108.

Michaelides, D. T. (1997). *Exact tests via complete enumeration: a distributed computing approach.* Unpublished Ph.D. thesis, Department of Social Statistics, University of Southampton, UK.

Schull, W. J. (1958). Empirical risks in consanguineous marriages: sex ratio, malformation and viability. *American Journal of Human Genetics*, **30**, 294–212.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society Series B*, **55**, 3–23.

Takken, A. R. (1999). *Monte Carlo Goodness-of-fit Tests for Discrete Data.* Unpublished Ph.D. dissertation, Department of Statistics, Stanford University.