Annals of Mathematical Sciences and Applications
Volume 1, Number 1, 195–215, 2016

A modified Pearson's χ^2 test with application to generalized linear mixed model diagnostics

Cecilia Dao and Jiming Jiang

We propose a modified version of Pearson's χ^2 test for goodness-of-fit that is applicable to generalized linear mixed models (GLMMs) diagnostics. The proposed test is based on cell frequencies, which is natural for many cases of GLMM. The procedure is simple and does not involve generalized inverse of a matrix, as was used in a previous study. Furthermore, the unknown parameters are estimated by solving a system of optimal estimating equations, which is computationally more efficient than the maximum likelihood estimators that were used in the previous study. Finally, the asymptotic null distribution of the proposed test is χ^2_{M-r-1}, where M is the number of cells and r is the number of unknown parameters that are estimated. A simulation study is carried out to demonstrate the asymptotic theory as well as finite-sample performance of the proposed test, including comparison with the previous method. An example of real-data application is considered.

Keywords and phrases: Asymptotic distribution, cell frequencies, chi-square, generalized linear mixed models, goodness-of-fit, model diagnostics.

1. Introduction

In classical mathematical statistics, one of the celebrated results is Pearson's χ^2 test for goodness-of-fit, or simply χ^2-test (Pearson 1900). The test statistic is given by

$$\chi^2 = \sum_{k=1}^{M} \frac{(O_k - E_k)^2}{E_k}, \tag{1}$$

where M is the number of cells, into which n observations are grouped, O_k and E_k are the observed and expected frequencies of the kth cell, $1 \le k \le M$, respectively. The expected frequency of the kth cell is given by $E_k = np_k$, where p_k is the known cell probability of the kth cell evaluated under the assumed model. The asymptotic theory associated with this test is simple: Under the null hypothesis of the assumed model, $\chi^2 \xrightarrow{\mathrm{d}} \chi^2_{M-1}$ as $n \to \infty$.

A good feature of Pearson's χ^2-test is that it can be used to test an arbitrary probability distribution, provided that the cell probabilities are completely known. However, the latter actually is a serious constraint, because in practice the cell probabilities often depend on certain unknown parameters of the probability distribution specified by the null hypothesis. For example, under the normal null hypothesis, the cell probabilities depend on the mean and variance of the normal distribution, which may be unknown. In such a case, one would, intuitively, replace the unknown parameters by their estimators, and thus obtain the estimated E_k, say $\hat{E}_k$, $1 \le k \le M$. The test statistic (1) then becomes

$$\hat{\chi}^2 = \sum_{k=1}^{M} \frac{(O_k - \hat{E}_k)^2}{\hat{E}_k}. \tag{2}$$

However, the test statistic, all of a sudden, may no longer have an asymptotic χ^2-distribution.

In a simple problem of assessing the goodness-of-fit to a Poisson or Multinomial distribution, it is known that the asymptotic null-distribution of (2) is χ^2_{M-p-1}, where p is the number of parameters estimated by the maximum likelihood method. This is the famous "subtract one degree of freedom for each parameter estimated" rule taught in many elementary statistics books (e.g., Rice 1995, pp. 242). However, the rule may not be generalizable to other probability distributions. For example, this rule does not even apply to testing normality with unknown mean and variance, as mentioned above. Note that here we are talking about MLE based on the original data, not the MLE based on cell frequencies. It is known that the rule applies in general to MLE based on cell frequencies. However, the latter are less efficient than the MLE based on the original data except for special cases where the two are the same, such as the above Poisson and Multinomial cases.

R. A. Fisher was the first to note that the asymptotic null-distribution of (2) is not necessarily χ^2 (Fisher 1922). He showed that if the unknown parameters are estimated by the so-called minimum chi-square method, the asymptotic null-distribution of (2) is still χ^2_{M-p-1}, but this conclusion may be false if other methods of estimation (including the ML) are used. Note that there is no contradiction of Fisher's result with the above results related to Poisson and Multinomial distributions, because the minimum chi-square estimators and the MLE are asymptotically equivalent when both are based on cell frequencies. A more thorough result was obtained by Chernoff and Lehmann (1954), who showed that when the MLE based on the original observations are used, the asymptotic null-distribution of (2) is not necessarily χ^2, but instead a "weighted" χ^2, where the weights are eigenvalues of certain nonnegative definite matrix. See Moore (1978) for a historical review of the χ^2-test.

Computational advances in modern era statistics have allowed consideration of much more complicated models than the Fisher-Pearson time that are more realistic and flexible. One class of such models is generalized linear mixed models, or GLMMs, which have become a popular and very useful class of statistical models. See, for example, Jiang (2007), McCulloch, Searle and Neuhaus (2008) for some wide-ranging accounts of GLMMs with theory and applications. Suppose that the observations come from m clusters and there are n_i observations, y_{ij}, $j = 1, \dots, n_i$, for the ith cluster. Suppose that there is a vector of random effects, α_i, associated with the ith cluster such that, given $\alpha_1, \dots, \alpha_m$, the observations y_{ij} are conditionally independent with conditional density

$$f(y_{ij}|\theta_{ij}, \phi_{ij}) = \exp\left\{\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{\phi_{ij}} + c(y_{ij}, \phi_{ij})\right\}, \tag{3}$$

where ϕ_{ij} is known up to a dispersion parameter, ϕ, which is known in some cases. Furthermore, the natural parameters θ_{ij}s satisfy

$$h(\theta_{ij}) = x'_{ij}\beta + z'_{ij}\alpha_i, \tag{4}$$

where h is a known function, and x_{ij}, z_{ij} are known vectors. A standard assumption for the random effects is that $\alpha_i, 1 \le i \le m$ are independent and distributed as $N(0, G)$, where the covariance matrix G depends on a vector, ψ, of variance components. Largely driven by practical interest, inference about GLMMs have received much attention since the early 1990s. See, for example, Jiang (2007) for a review of historical developments, and Torabi (2012), Jiang (2013) for some recent advances. On the other hand, model diagnostics, an important part of statistical modeling, is largely missing from the literature on GLMMs. For a special case of GLMMs, that is, linear mixed model (LMM), however, there has been significant advances. Lange and Ryan (1989) proposed to use the EBLUP of the random effects for informal checking of the normality of the random effects. Also see Calvin and Sedransk (1991). An important step in model diagnostics is goodness-of-fit test. This is often a first step in formal model checking (e.g., McCullagh and Nelder 1989, sec. 12.2) in that the result of goodness-of-fit test may lead to further investigation, if necessary, to identify the specific part(s) of the model assumptions that does(do) not hold. Jiang (2001) proposed a χ^2-type goodness-of-fit test for LMM diagnostics, whose asymptotic null distribution is a weighted χ^2, where the weights are eigenvalues of some non-negative definite matrix. More recently, Claeskens and Hart (2009) proposed an alternative approach to the χ^2 test for checking the normality assumption in LMM. The authors showed that, in some cases, the χ^2-test of Jiang (2001) is not sensitive to certain departures from the normality assumption; as a result, the test may have low power against certain alternatives. As a new approach, the authors considered a class of distributions that include the normal distribution as a reduced, special case. The test is based on the likelihood-ratio test that compares the "estimated distribution" and the null distribution (i.e., normal). A model selection procedure via the information criteria is used to determine the larger class of distributions for the LRT. In particular, the asymptotic null distribution is in the form of the distribution of $\sup_{r\geq 1}\{2Q_r/r(r+3)$, where $Q_r = \sum_{q=1}^{r}\chi^2_{q+1}$, and $\chi^2_2, \chi^2_3, \dots$ are independent such that χ^2_j has a χ^2 distribution with j degrees of freedom, $j \geq 2$. So far as GLMMs are concerned, the only existing literature in the form of goodness-of-fit test are Gu (2008) and Tang (2010), both being Ph. D. dissertations. Gu (2008) considered similar χ^2 tests to Jiang (2001) and applied them to mixed logistic models, a special case of GLMMs. She considered both minimum χ^2 estimator and method of simulated moments (MSM) estimator (Jiang 1998) of the model parameters, and derived asymptotic null distributions of the test statistics, which are weighted χ^2. Tang (2010) proposed a different χ^2-type goodness-of-fit test for GLMM diagnostics, which is not based on the cell frequencies. She proved that the asymptotic null distribution is χ^2. However, the test is based on the maximum likelihood estimator (MLE), which is known to be computationally difficult to obtain. Furthermore, the test statistic involves the Moore-Penrose generalized inverse (G-inverse) of a normalizing matrix, which does not have an analytic expression. The interpretation of such a G-inverse may not be straightforward for a practitioner.

For cases of discrete responses, such as in typical situations of GLMMs, χ^2 tests based on cell frequencies, such as Pearson's χ^2 test, are much more natural than for cases of continuous observations. For example, if the responses are binomial, the range of the responses is a set of nonnegative integers. Thus, a natural choice of the cells are exactly those integers. In contrast, if the responses are continuous, one has to divide the range of the responses into intervals in order to apply Pearson's χ^2 test, and there are infinitely many ways of choosing the number of cells as well as the cells, given the number of cells. In fact, the choice of the cells in the latter case is a difficult, unsolved problem. See, for example, Jiang (2001) for discussion on this issue and the reference therein. Also, a χ^2 asymptotic distribution is more desirable, in terms of simplicity, than the supremum of χ^2 of Claeskens and Hart (2009), or weighted χ^2 of Jiang (2001) and Gu (2008). Note that, in the latter cases, the weights are eigenvalues of some matrices, whose expressions are compicated, and involve unknown parameters. These parameters need to be estimated in order to obtain the critical values of the tests. Due to such a complication, Jiang (2001) suggests to use a Monte-Carlo method to compute the critical value; but, by doing so, the usefulness of the asymptotic result may be undermined. Furthermore, it would be beneficial if the test statistic does not involve a G-inverse. In particular, the latter has no analytic expression; thus, it is difficult to see how the G-inverse is affected by the parameters, and sample sizes. As is well known, there may be multiple factors contributing to the sample size under a mixed effects model. Finally, computation has been a big issue in GLMM, especially for likelihood-based inference (e.g., Jiang 2007, sec. 3.4; Torabi 2012). Any diagnostic technique for GLMM has to be computationally efficient. For example, a test that requires computation of the MLE, as in Tang (2010), would be computationally less efficient than one that only requires a consistent estimator that is computationally simpler, such as the MSM estimator used in Gu (2008); see Jiang (1998).

To summarize the points of the preceding paragraph, it is desirable to (i) develop a Pearson type χ^2 test, based on cell frequencies, that (ii) results in a χ^2 asymptotic null distribution, (iii) does not involve a G-inverse, and (iv) is computationally attractive. This development is carried out in the next two sections, with Section 2 focusing on construction of the test and idea of the derivation, and Section 3 providing rigorous statement of the result and the proof. In particular, the vector of unknown parameters, θ, is estimated by a generalized estimating equation (GEE) estimator, which is the solution to an optimal estimating equation, and the asymptotic null distribution is χ^2_{M-r-1}, where M is the number of cells and r is the dimension of θ. A simulation study is carried out in Section 4 to verify the asymptotic theory as well as to evaluate empirical performance of the proposed test. A real-data example and some discussion are presented in Section 5.

2. Construction and heuristic derivation

We first present a method of constructing a test that satisfies properties (i)–(iv) outlined at the end of the last section. We then discuss application of the method to GLMMs.

Let $Y_1, \ldots, Y_n$ be vectors of observations that are independent but not (necessarily) identically distributed. Let $C_1, \ldots, C_M$ denote the cells. In the original Pearson's χ^2 test, where the observations are i.i.d., and the cell probabilities, $p_k = \mathrm{P}(Y_i \in C_k), 1 \le k \le M$, are known, the asymptotic null distribution is χ^2_{M-1}. The "minus one degree of freedom" may be interpreted by the fact that the cell probabilities are subject to a sum constraint: $\sum_{k=1}^{M} p_k = 1$. A simply strategy to "free" the cell probabilities is to simply drop one of the cells, say, the last one. Therefore, we consider $O_i = [1_{(Y_i \in C_k)}]_{1 \le k \le M-1}$, and let $u_i(\theta) = \mathrm{E}_\theta(O_i) = [\mathrm{P}_\theta(Y_i \in C_k)]_{1 \le k \le M-1}$, where θ is the vector of parameters involved in the distribution of the observations, and E_θ and P_θ denote expectation and probability, respectively, when θ is the true parameter vector. Furthermore, let $b_i(\theta) = V_i^{-1/2}(\theta)\{O_i - u_i(\theta)\}$, where $V_i(\theta) = \mathrm{Var}_\theta(O_i)$, Var_θ denoting covariance matrix when θ is the true parameter vector, and $V_i^{-1/2}(\theta) = [\{V_i(\theta)\}^{-1}]^{1/2}$. Note that there is a simple expression for $\{V_i(\theta)\}^{-1}$. First, it is easy to show that

$$V_i(\theta) = P_i(\theta) - p_i(\theta)p_i'(\theta), \tag{5}$$

where $P_i(\theta) = \mathrm{diag}\{p_{ik}(\theta), 1 \le k \le M-1\}$, with $p_{ik}(\theta) = \mathrm{P}_\theta(Y_i \in C_k)$, $p_i(\theta) = [p_{ik}(\theta)]_{1 \le k \le M-1}$, and $p_i'(\theta) = \{p_i(\theta)\}'$. It follows, by a well-known formula of matrix inversion (e.g., Sen and Srivastava 1990, p. 275), that

$$\{V_i(\theta)\}^{-1} = \mathrm{diag}\left\{\frac{1}{p_{ik}(\theta)}, 1 \le k \le M-1\right\} + \frac{J_{M-1}}{1 - \sum_{k=1}^{M-1} p_{ik}(\theta)}, \tag{6}$$

assuming $\sum_{k=1}^{M-1} p_{ik}(\theta) < 1$, where J_a denotes the $a \times a$ matrix of 1's. Now consider

$$B(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_i(\theta). \tag{7}$$

If θ is the true parameter vector, then, by the central limit theorem (CLT), we have

$$B(\theta) \ \stackrel{\mathrm{d}}{\longrightarrow}\ N(0, I_{M-1}), \tag{8}$$

where I_a denotes the a-dimensional identity matrix. It follows that

$$|B(\theta)|^2 \ \stackrel{\mathrm{d}}{\longrightarrow}\ \chi^2_{M-1}, \tag{9}$$

as $n \to \infty$, where $|\cdot|$ denotes the Euclidean norm. In practice, θ is unknown; thus, we cannot use the result of (9) for testing. It is then customary to replace θ by an estimator, $\hat{\theta}$. Two immediate questions are: (I) what $\hat{\theta}$? and (II) how does the asymptotic distribution change when θ is replaced by $\hat{\theta}$?

Let us leave part of the answer to question (I) to a later stage. For now, all we require is that $\hat{\theta}$ be a solution to an estimating equation:

$$C(\theta) = 0, \tag{10}$$

where $C(\theta) = A(\theta)B(\theta)$ for some non-random matrix $A(\theta)$ that depends on θ, so that some standard asymptotic properties, used in the subsequent derivations, hold. Now let us focus on answering question (II) by showing that, with the construction of $B(\theta)$ given above and any $\hat{\theta}$ satisfying the above conditions, one always has

$$\hat{\chi}^2 \equiv |B(\hat{\theta})|^2 \xrightarrow{\mathrm{d}} \chi^2_{M-r-1}, \tag{11}$$

as $n \to \infty$, where $r = \dim(\theta)$. Hereafter in this section, θ denotes the true parameter vector; partial derivatives, such as $\partial B/\partial\theta'$, are understood as evaluated at the true θ. By Taylor series expansion, we have

$$\begin{aligned} B(\hat{\theta}) &\approx B(\theta) + \frac{\partial B}{\partial \theta'}(\hat{\theta} - \theta) \\ &\approx B(\theta) + \mathrm{E}_\theta\left(\frac{\partial B}{\partial \theta'}\right)(\hat{\theta} - \theta). \end{aligned} \tag{12}$$

Hereafter, each approximation, $\approx$, is in the sense that the omitted term is of lower order than the presented term, in the sense of convergence in probability (e.g., Jiang 2010, ch. 2). The arguments will be rigorized in the next section. Similarly, we have

$$\begin{aligned} 0 &= C(\hat{\theta}) \\ &\approx C(\theta) + \frac{\partial C}{\partial \theta'}(\hat{\theta} - \theta) \\ &\approx C(\theta) + \mathrm{E}_\theta\left(\frac{\partial C}{\partial \theta'}\right)(\hat{\theta} - \theta), \end{aligned}$$

using the definition of $\hat{\theta}$ for the first equation. It follows that

$$\hat{\theta} - \theta \;\approx\; -\left\{\mathrm{E}_\theta\left(\frac{\partial C}{\partial \theta'}\right)\right\}^{-1} C(\theta). \tag{13}$$

Combining (12), (13), we get

$$B(\hat{\theta}) \approx B(\theta) - \mathrm{E}_\theta\left(\frac{\partial B}{\partial \theta'}\right)\left\{\mathrm{E}_\theta\left(\frac{\partial C}{\partial \theta'}\right)\right\}^{-1} C(\theta). \tag{14}$$

Furthermore, note that $\mathrm{E}_\theta\{B(\theta)\} = 0$. Thus, we have

$$\mathrm{E}_\theta\left(\frac{\partial C}{\partial \theta'}\right) = A(\theta)\mathrm{E}_\theta\left(\frac{\partial B}{\partial \theta'}\right). \tag{15}$$

If we write $U = \mathrm{E}_\theta(\partial B/\partial\theta')$, $A = A(\theta)$, and $B = B(\theta)$, then, by (14) and (15), we have

$$B(\hat{\theta}) \approx \{I_{M-1} - U(AU)^{-1}A\}B. \tag{16}$$

Assume that, as $n \to \infty$, we have

$$\frac{1}{\sqrt{n}}U = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_\theta\left(\frac{\partial b_i}{\partial \theta'}\right) \longrightarrow Q, \tag{17}$$

$$\text{and} \quad \sqrt{n}A \longrightarrow R \tag{18}$$

for some limiting matrices Q and R. Then, we have $U(AU)^{-1}A \to Q(RQ)^{-1}R$. If we further assume that

$$U(AU)^{-1}A \text{ is symmetric,} \tag{19}$$

then, so is $Q(RQ)^{-1}R$. Thus, combined with (8) and (16), we have

$$B(\hat{\theta}) \xrightarrow{\mathrm{d}} N(0, P), \tag{20}$$

where $P = \{I_{M-1} - Q(RQ)^{-1}R\}^2 = I_{M-1} - Q(RQ)^{-1}R$, using the symmetry of the latter. It follows that P is idempotent; hence (e.g., Jiang 2007, p. 238), we have

$$|B(\hat{\theta})|^2 = B'(\hat{\theta})B(\hat{\theta}) \xrightarrow{\mathrm{d}} \chi^2_\nu,$$

where $\nu = \mathrm{tr}(P) = M - r - 1$. Therefore, (11) holds, giving answer to question (II).

It remains to find A that satisfies (18) and (19). Consider A having the following form: $A = n^{-1}U'W$, where W is a symmetric, non-random matrix to be determined. Then, by (13), we have $\hat{\theta} - \theta \approx -(U'WU)^{-1}U'WB$. It follows that

$$\text{(21)} \qquad \text{Var}_\theta(\hat{\theta}) \approx (U'WU)^{-1}U'W^2U(U'WU)^{-1} \geq (U'U)^{-1},$$

where for symmetric matrices A_1 and A_2, $A_1 \geq A_2$ iff $A_1 - A_2$ is nonnegative definite. See, for example, Lemma 5.1 of Jiang (2010) for the last inequality in (21). Furthermore, the equality on the right side of (21) holds when $W = I_{M-1}$. This shows that the optimal W, in the sense of minimizing the asymptotic covariance matrix (in the matrix order defined above), is $W = I_{M-1}$. Thus, the optimal A is

$$\text{(22)} \qquad A(\theta) = \frac{1}{n}U' = \frac{1}{n}\text{E}_\theta\left(\frac{\partial B'}{\partial \theta}\right) = -\frac{1}{n^{3/2}}\sum_{i=1}^{n}\frac{\partial u_i'}{\partial \theta}V_i^{-1/2}(\theta).$$

It is clear that (18) is expected to hold, and (19) obviously holds, for the optimal A. This determines A, hence completes our answer to question (I).

To apply the above general result to GLMMs, we obtain, as a first step, sufficient statistics at cluster levels. The idea is similar to Jiang (1998), which is straightforward when the link function is canonical, that is, when

$$\text{(23)} \qquad h(\theta_{ij}) = \theta_{ij}$$

in (4) for all i, j. Below we shall focus on this case. Let $p = \dim(\beta)$ and $d = \text{diag}(\alpha_i)$. Suppose that the covariance matrix G depends on a q-dimensional vector, ψ, of variance components, that is, $G = G(\psi)$. Let $G = DD'$ be the Cholesky decomposition of G, where $D = D(\psi)$. Then, α_i can be expressed as $\alpha_i = D(\psi)\xi_i$, where $\xi_i \sim N(0, I_d)$. Furthermore, suppose that ϕ_{ij} has the following special form (e.g., Jiang 2007, p. 191):

$$\text{(24)} \qquad \phi_{ij} = \frac{\phi}{w_{ij}},$$

where ϕ is an unknown dispersion parameter, and w_{ij} is a known weight, for every i, j. Then, it can be shown that the conditional density of $y_i = (y_{ij})_{1\leq j\leq n_i}$ given ξ_i (with respect to a σ-finite measure) can be expressed as

$$f(y_i|\xi_i) = \exp\left\{\left(\sum_{j=1}^{n_i} w_{ij}y_{ij}x_{ij}\right)'\left(\frac{\beta}{\phi}\right) + \left(\sum_{j=1}^{n_i} w_{ij}y_{ij}z_{ij}\right)'\left(\frac{D(\psi)}{\phi}\right)\xi_i\right.$$

$$-\sum_{j=1}^{n_i}\left(\frac{w_{ij}}{\phi}\right)b(x_{ij}'\beta+z_{ij}'D(\psi)\xi_i)+\sum_{j=1}^{n_i}c\left(y_{ij},\frac{\phi}{w_{ij}}\right)\Bigg\}.$$

Let $f(\cdot)$ denote the pdf of $N(0, I_d)$, and $\xi \sim N(0, I_d)$. It follows that

$$\begin{aligned}(25)\qquad f(y_i) &= \int f(y_i|\xi_i)f(\xi_i)d\xi_i \\ &= \exp\left\{\left(\sum_{j=1}^{n_i} w_{ij}y_{ij}x_{ij}\right)'\left(\frac{\beta}{\phi}\right)\right\} \\ &\quad \mathrm{E}\left[\exp\left\{\left(\sum_{j=1}^{n_i} w_{ij}y_{ij}z_{ij}\right)'\left(\frac{D(\psi)}{\phi}\right)\xi \right.\right. \\ &\quad \left.\left. -\sum_{j=1}^{n_i}\left(\frac{w_{ij}}{\phi}\right)b(x_{ij}'\beta+z_{ij}'D(\psi)\xi)\right\}\right] \\ &\quad \exp\left\{\sum_{j=1}^{n_i}c\left(y_{ij},\frac{\phi}{w_{ij}}\right)\right\},\end{aligned}$$

where the expectation is with respect to ξ. From (25), it is clear that a set of sufficient statistics for all of the unknown parameters involved, namely, β, ψ, and ϕ, are

$$(26)\qquad \sum_{j=1}^{n_i} w_{ij}y_{ij}x_{ij} \quad \text{and} \quad \sum_{j=1}^{n_i} w_{ij}y_{ij}z_{ij}.$$

Note that the first summation in (26) is a $p \times 1$ vector, while the second summation is a $d \times 1$ vector. So, in all, there are $p + d$ components of those vectors; however, some of the components may be redundant, or functions of the other components. After removing the redundants, and functions of others, the remaining components form a vector, denoted by Y_i, so that the sufficient statistics in (26) are functions of Y_i. The Y_i's will be used for goodness-of-fit test of the following hypothesis:

$$(27)\qquad H_0: \quad \text{The assumed GLMM holds}$$

versus the alternative that there is a violation of the model assumption. In many cases, the null hypothesis is more specific about one particular part of the GLMM, such as the normality of the random effects, assuming that other parts of the model hold; the alternative thus also changes correspondingly.

If the values of y_{ij}'s belong to a finite subset of R, such as in the Binomial situation, the possible values of $Y_1, \dots, Y_n$ is a finite subset $S \subset R^g$, where $g = \dim(Y_i)$, assuming that all of the Y_i's are of the same dimension. Let $C_1, \dots, C_M$ be the different (vector) values in S. These are the cells under the general set-up. If the values of y_{ij}'s are not bounded, such as in the Poisson case, let K be a positive number such that the probability that $\max_i |Y_i| > K$ is small. Let $C_1, \dots, C_{M-1}$ be the different (vector) values in $S \cap \{v \in R^g : |v| \leq K\}$, and $C_M = S \cap \{v \in R^g : |v| > K\}$, where S is the set of all possible values of the Y_i's. These are the cells under the general set-up. We illustrate with a simple example.

Example 1. Let y_{ij} be a binary outcome with $\text{logit}\{\text{P}(y_{ij} = 1|\alpha)\} = \mu + \alpha_i$, $1 \leq i \leq n$, $1 \leq j \leq M-1$, where $\alpha_1, \dots, \alpha_n$ are i.i.d. random effects with $\alpha_i \sim N(0, \sigma^2)$, $\alpha = (\alpha_i)_{1 \leq i \leq n}$, and μ, σ are unknown parameters with $\sigma \geq 0$. It is more convenient to use the following expression: $\alpha_i = \sigma \xi_i$, $1 \leq i \leq n$, where $\xi_1, \dots, \xi_n$ are i.i.d. $N(0,1)$ random variables. In this case, we have $x_{ij} = z_{ij} = w_{ij} = 1$, and $n_i = M-1, 1 \leq i \leq n$. Thus, both expressions in (26) are equal to $y_{i\cdot} = \sum_{j=1}^{M-1} y_{ij}$. It follows that the sufficient statistics are $Y_i = y_{i\cdot}, 1 \leq i \leq n$, which are i.i.d. The range of Y_i is $0, 1, \dots, M-1$; thus, we have $C_k = \{k-1\}, 1 \leq k \leq M$. Let $\theta = (\mu, \sigma)'$. It is easy to show that

$$p_{ik}(\theta) = \text{P}_\theta(Y_i \in C_k) = \binom{M-1}{k-1} e^{(k-1)\mu} \text{E}\left[\frac{\exp\{(k-1)\sigma\xi\}}{\{1 + \exp(\mu + \sigma\xi)\}^{M-1}}\right],$$

$1 \leq k \leq M-1$, where the expectation is with respect to $\xi \sim N(0,1)$. Note that, in this case, $p_{ik}(\theta)$ does not depend on i; and is no need to compute $p_{iM}(\theta)$. Also, we have

$$\begin{aligned} \frac{\partial p_{ik}(\theta)}{\partial \mu} &= \binom{M-1}{k-1} e^{(k-1)\mu} \text{E}\Bigg[\frac{\exp\{(k-1)\sigma\xi\}}{\{1 + \exp(\mu + \sigma\xi)\}^{M-1}} \Bigg\{(k-1) \\ &\quad - \frac{(M-1)\exp(\mu + \sigma\xi)}{1 + \exp(\mu + \sigma\xi)}\Bigg\}\Bigg], \\ \frac{\partial p_{ik}(\theta)}{\partial \sigma} &= \binom{M-1}{k-1} e^{(k-1)\mu} \text{E}\Bigg[\frac{\exp\{(k-1)\sigma\xi\}}{\{1 + \exp(\mu + \sigma\xi)\}^{M-1}} \Bigg\{(k-1) \\ &\quad - \frac{(M-1)\exp(\mu + \sigma\xi)}{1 + \exp(\mu + \sigma\xi)}\Bigg\}\xi\Bigg], \end{aligned}$$

$1 \leq k \leq M-1$. Again, there is no need to compute the derivatives for $k = M$.

In some cases, the range of Y_i may be different for different i. To avoid having zero cell probabilities, $p_{ik}(\theta)$, in (6), one strategy is to divide the data into (nonoverlapping) groups so that, within each group, the Y_i's have the same range. More specifically, let $Y_i, i \in I_l$ be the lth group whose corresponding cells are $C_{kl}, k = 1, \dots, M_l$ with $p_{ikl}(\theta) = \mathrm{P}_\theta(Y_i \in C_{kl}) > 0, i \in I_l, 1 \le k \le M_l$, $l = 1, \dots, L$. The method described above can be applied to each group of the data, $Y_i, i \in I_l$, leading to the χ^2 test statistic, $\hat{\chi}_l^2$, that has the asymptotic $\chi^2_{M_l - r - 1}$ distribution under the null hypothesis, $1 \le l \le L$. Then, because the groups are independent, the combined χ^2 statistic,

$$\hat{\chi}^2 = \sum_{l=1}^{L} \hat{\chi}_l^2, \tag{28}$$

has the asymptotic χ^2 distribution with $\sum_{l=1}^{L}(M_l - r - 1) = M. - L(r + 1)$ degrees of freedom, under the null hypothesis, where $M. = \sum_{l=1}^{L} M_l$. In conclusion, the goodness-of-fit test is carried out using (28) with the asymptotic $\chi^2_{M. - L(r+1)}$ null distribution.

3. Theorems and proofs

In this section, we rigorize the results of the previous section by providing regularity conditions, under which these results hold, and the rigorous proofs. Following the approach of the previous section, we first prove some results under the general setting in terms of $Y_i, 1 \le i \le n$; we then apply the general results to GLMMs.

3.1. Additional notation and regularity conditions

In this section, we use θ_0 to denote the true vector of parameters, and $\Theta \subset R^r$ the parameter space. Let $\|A\| = \{\lambda_{\max}(A'A)\}^{1/2}$ denote the spectral norm of matrix A, where $\lambda_{\max}$ denotes the largest eigenvalue, and $\|v\| = \sqrt{v'v}$ the Euclidean norm of vector v.

Note that the vectors and matrices $A(\theta), B(\theta)$, etc., introduced in Section 2, depend on the sample size, n. The dependence will be made explicit in this section through the corresponding notation $A_n(\theta), B_n(\theta)$, etc. In addition, write

$$c_{n,i}(\theta) = \frac{1}{\sqrt{n}} \mathrm{E}_\theta \left\{ \frac{\partial}{\partial \theta} B_n'(\theta) \right\} b_i(\theta),$$

where $b_i(\theta)$ is defined above (5), so that $C_n(\theta) = n^{-1} \sum_{i=1}^{n} c_{n,i}(\theta)$. Note that $c_{n,i}(\theta)$ is a r-dimensional vector. Let $\nabla C(\theta) = (\partial/\partial\theta')C(\theta)$. All of the general results below regarding $Y_1, \ldots, Y_n$ are under the limiting process $n \to \infty$.

We assume that the following regularity conditions are satisfied.

A1. Θ is open in R^r, and $c_{n,i}(\cdot)$ is continuously differentiable with respect to $\theta \in \Theta$.

A2. With probability tending to one, the matrix $\nabla C(\theta_0)$ is non-singular.

A3. The limit $\Sigma(\theta) = \lim_{n\to\infty} \mathrm{E}\{\nabla C(\theta)\}$ exists, and $\exists$ a constant $\delta > 0$ such that

$$\sup_{|\theta-\theta_0|<\delta} |\nabla C(\theta) - \Sigma(\theta)| \overset{\mathrm{P}}{\longrightarrow} 0.$$

A4. $M > r + 1$, and there exists a full-rank, $(M-1) \times r$ matrix Q such that

$$\frac{1}{n} \sum_{i=1}^{n} V_i^{-1/2}(\theta_0) \left. \frac{\partial u_i}{\partial \theta'} \right|_{\theta_0} \longrightarrow Q.$$

A5. There is a compact subset $\Theta_c \subset \Theta$ such that the $\max_{1\le i\le n} \|\cdot\|$ of $V_i(\theta)$ and its up to second order partial derivatives, and the $\max_{1\le i\le n} \|\cdot\|$ of first to third partial derivatives of $u_i(\theta)$ are bounded over $\theta \in \Theta_c$.

3.2. Existence, uniqueness, and consistency of $\hat{\theta}$

Before proving the main asymptotic result (11), we need to establish the existence of $\hat{\theta}$ as a solution to (10), and its consistency as an estimator of θ_0. The proof of the following result is very similar to that of Theorem 2 of Foutz (1977); and therefore omitted.

Theorem 1. *Under assumptions A1–A3, there exists a sequence of estimators, $\hat{\theta}_n$, such that $C(\hat{\theta}_n) = 0$ with probability tending to one, and $\hat{\theta}_n \overset{\mathrm{P}}{\to} \theta_0$. Furthermore, if $\check{\theta}_n$ also satisfies the above, one must have $\check{\theta}_n = \hat{\theta}_n$ with probability tending to one.*

Note. For the last part of Theorem 1, if $\hat{\theta}_n$ and $\check{\theta}_n$ are both consistent, of course one has $\hat{\theta}_n - \check{\theta}_n \overset{\mathrm{P}}{\to} 0$, but the conclusion, $\mathrm{P}(\check{\theta}_n = \hat{\theta}_n) \to 1$, is stronger.

3.3. Asymptotic null distribution of $B_n(\hat{\theta}_n)$

Theorem 2. *Under assumptions A1–A5, we have*

$$B_n(\hat{\theta}_n) \overset{\mathrm{d}}{\longrightarrow} N(0, P), \tag{29}$$

where $P = I_{M-1} - Q(Q'Q)^{-1}Q'$ is idempotent with rank $M - r - 1$.

Proof. Because $b_i(\theta_0), 1 \leq i \leq n$ are independent with mean vector 0 and covariance matrix I_{M-1}, by the central limit theorem (CLT; e.g., Jiang 2010, sec. 6.4), we have

$$B_n(\theta_0) \xrightarrow{\mathrm{d}} N(0, I_{M-1}). \tag{30}$$

Next, by Taylor series expansion, we have

$$\begin{aligned} 0 &= C_n(\hat{\theta}_n) \\ &= C_n(\theta_0) + \frac{\partial C_n}{\partial \theta'}(\hat{\theta}_n - \theta_0) \\ &+ \frac{1}{2}\left[(\hat{\theta}_n - \theta_0)' \left.\frac{\partial^2 C_{n,k}}{\partial\theta\partial\theta'}\right|_{\theta_{(k)}}\right]_{1\leq k\leq r} (\hat{\theta}_n - \theta_0), \end{aligned} \tag{31}$$

where, and hereafter, the derivatives $\partial C_n/\partial\theta'$, etc. without indicating at which θ are understood as evaluated at θ_0, $C_{n,k}$ denotes the kth component of C_n, and $\theta_{(k)}$ lies between θ_0 and $\hat{\theta}$, $1 \leq k \leq r$. By assumption *A5* and the law of large numbers (LLN; e.g., Jiang 2010, sec. 6.2), it can be shown that

$$\frac{\partial C_n}{\partial \theta'} = Q'Q + o_{\mathrm{P}}(1), \tag{32}$$

$$\left.\frac{\partial^2 C_{n,k}}{\partial\theta\partial\theta'}\right|_{\theta_{(k)}} = O_{\mathrm{P}}(1), \quad 1 \leq k \leq r, \tag{33}$$

where $o_{\mathrm{P}}(1)$ represents a term that converges to 0 in probability, and $O_{\mathrm{P}}(1)$ a term that is bounded in probability (e.g., Jiang 2010, sec. 3.4). By (31)–(33), and Theorem 1, we have $0 = C_n(\theta_0) + \{Q'Q + o_{\mathrm{P}}(1)\}(\hat{\theta}_n - \theta_0)$, or

$$\hat{\theta}_n - \theta_0 = -\{Q'Q + o_{\mathrm{P}}(1)\}^{-1} C_n(\theta_0). \tag{34}$$

On the other hand, again, by Taylor series expansion, we have

$$\begin{aligned} B_n(\hat{\theta}_n) &= B_n(\theta_0) + \frac{\partial B_n}{\partial \theta'}(\hat{\theta}_n - \theta_0) \\ &+ \frac{1}{2}\left[(\hat{\theta}_n - \theta_0)' \left.\frac{\partial^2 B_{n,k}}{\partial\theta\partial\theta'}\right|_{\theta^{(k)}}\right]_{1\leq k\leq r} (\hat{\theta}_n - \theta_0). \end{aligned} \tag{35}$$

where $B_{n,k}$ denotes the kth component of B_n, and $\theta^{(k)}$ lies between θ_0 and $\hat{\theta}_n$, $1 \leq k \leq r$. By (30), and assumption *A5*, it can be shown that

$C_n(\theta_0) = O_{\mathrm{P}}(n^{-1/2})$; hence, by (34), we have $\hat{\theta}_n - \theta_0 = O_{\mathrm{P}}(n^{-1/2})$. Thus, by assumption *A5*, it can be shown that the last term on the right side of (35) is $o_{\mathrm{P}}(1)$. Furthermore, by LLN, it can be shown that

$$\frac{1}{\sqrt{n}}\frac{\partial B_n}{\partial \theta'} = -Q + o_{\mathrm{P}}(1). \tag{36}$$

Also, by (22) and assumption *A4*, we have

$$\sqrt{n}A_n(\theta_0) \;\longrightarrow\; -Q'. \tag{37}$$

Therefore, combining (34)–(37), we have

$$\begin{aligned} B_n(\hat{\theta}_n) &= B_n(\theta_0) - \{-Q + o_{\mathrm{P}}(1)\} \\ &\quad \times \sqrt{n}\{Q'Q + o_{\mathrm{P}}(1)\}^{-1}\sqrt{n}A_n(\theta_0)\frac{1}{\sqrt{n}}B_n(\theta_0) + o_{\mathrm{P}}(1) \\ &= \{I_{M-1} - Q(Q'Q)^{-1}Q'\}B_n(\theta_0) + o_{\mathrm{P}}(1). \end{aligned} \tag{38}$$

The conclusion then follows by (30) and (38). The idempotentness of P is obvious. □

3.4. Asymptotic null distribution of $\hat{\chi}^2$ and application to GLMM

By Theorem 2 and the continuous mapping theorem (e.g., Jiang 2010, Theorem 2.12), and a property of multivariate normal distribution (e.g., Jiang 2007, p. 237), one immediately obtains the following.

Corollary 1. *Under assumptions A1–A5, (11) holds.*

To apply Corollary 1 to GLMMs, consider the construction of the $Y_i, 1 \le i \le n$ described in Section 2. We have the following results.

Corollary 2. *Under the GLMM assumption, suppose that A1–A5 hold.*

(I) If the Y_i's have the same range in the sense that $p_{ik}(\theta_0) > 0$ for every $1 \le k \le M$, the asymptotic distribution of $\hat{\chi}^2 = |B_n(\hat{\theta}_n)|^2$, under (27), is χ^2_{M-r-1}.

(II) If the Y_i's can be divided into L groups such that, within each group, the Y_i's have the same range in the sense above, the asymptotic distribution of the $\hat{\chi}^2$ given by (28), under (27), is $\chi^2_{M.-L(r+1)}$, where M. is defined below (28).

Example 1 (continued). It is easy to verify that assumptions *A1–A5* are satisfied. Note that such quantities as $V_i(\theta), u_i(\theta)$ do not depend on i in this case. Therefore, the limits such as the one in *A4* obviously exists, because it is a constant (i.e., not dependent on n).

4. Simulation study

We use a simulated example to verify the asymptotic theory and also study the finite-sample performance of the proposed test. The example is a special case of Example 1, with $n = 100$ or 200, and $M = 7$. The true values of the parameters are $\mu = 1$ and $\sigma = 2$.

Consider testing the normality of the random effects, α_i, assuming other parts of the GLMM assumptions hold. Then, the null hypothesis, (27), is equivalent to

$$H_0 : \quad \alpha_i \sim \text{Normal} \tag{39}$$

versus the alternative that the distribution of α_i is not normal. Two specific alternatives are considered. Under the first alternative, $H_{1,1}$, the distribution of α_i is a centralized exponential distribution, namely, the distribution of $\zeta - 2$, where $\zeta \sim$ Exponential(0.5). Under the second alternative, $H_{1,2}$, the distribution of α_i is a normal-mixture, namely, the mixture of $N(-3, 0.5)$ with weight 0.2, $N(2/3, 0.5)$ with weight 0.3, and $N(4/5, 0.5)$ with weight 0.5. All simulation results are based on $K = 1000$ simulation runs.

First, we compare the empirical and asymptotic null distributions of the test statistic under different sample sizes, n. According to Corollary 1, the asymptotic null distribution of the test is χ^2_4. The left figure of Figure 1 shows the histogram of the simulated test statistics, for $n = 100$, under H_0, with the pdf of χ^2_4 plotted on top. It appears that the histogram matches the theoretical (asymptotic) distribution quite well. The corresponding cdf's are plotted in the right figure, and there is hardly any visible difference between the two. Figure 2 shows the corresponding plots for $n = 200$. Here, the matches are even better, more visibly in the histogram-pdf comparison. Some numerical summaries, in terms of the 1st (Q_1), 2nd (Q_2), and 3rd (Q_3) quartiles, are presented in Table 1.

Next, we consider size and power of the test at the levels of significance 0.01, 0.05, and 0.10, respectively. The simulated sizes and powers are presented in Table 2. It appears that the simulated sizes are closer to the nominal levels for $n = 100$; on the other hand, the simulated powers are much higher for $n = 200$.

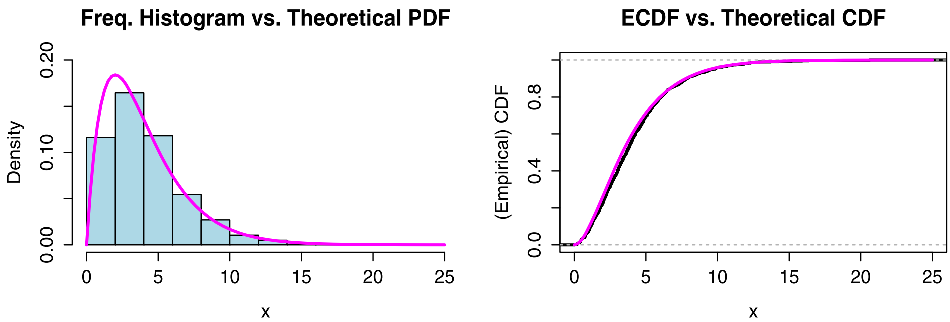

Figure 1: Theoretical vs Empirical Distributions: $n = 100$. Left: pdf's; Right: cdf's.

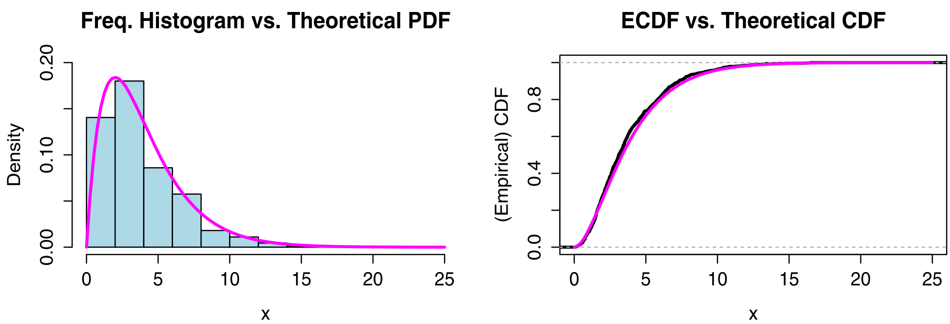

Figure 2: Theoretical vs Empirical Distributions: $n = 200$. Left: pdf's; Right: cdf's.

Table 1: Empirical and Theoretical Quartiles

Quartiles	$n = 100$	$n = 200$	χ^2_4
Q_1	2.094	1.848	1.923
Q_2	3.649	3.103	3.357
Q_3	5.493	5.203	5.385

Table 2: Simulated Size and Power

	Size		Power against $H_{1,1}$		Power against $H_{1,2}$	
Nominal Level	$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
0.01	0.010	0.007	0.148	0.408	0.329	0.766
0.05	0.051	0.043	0.345	0.650	0.577	0.912
0.10	0.104	0.080	0.483	0.768	0.704	0.952

Table 3: Litter Size and Number of Cases

Litter Size	6	7	8	9	10
# of Litters	116	253	368	310	144

As a comparison, we also studied performance of the χ^2 test proposed by Tang (2010) in same-data comparisons with our test. The simulated sizes and powers of Tang's test are both close to 0, under both sample sizes; thus, the detailed results are not presented. One possible reason is that the estimated covariance matrix, $\hat{\Sigma}_{\rm svd}$, in Tang's test, whose expression is complicated, is inaccurate, and there does not seem to be an easy fix.

5. A real-data example and discussion

Brooks et al. (1997) presented six datasets recording fetal mortality in mouse litters. As an application, we consider the HS2 dataset from Table 4 of their paper, which reports the number of dead implants in litters of mice from untreated experimental animals. Jiang and Zhang (2001) analyzed the data based on a GLMM assumption (also see Jiang 2007, sec. 4.4.1). Let $y_{ij}, i = 1, \ldots, m, j = 1, \ldots, n_i$ be binary responses such that $y_{ij} = 1$ if the jth implant in the ith litter is dead, and $y_{ij} = 0$ otherwise. Suppose that, given the litter-specific random effects, $\alpha_1, \ldots, \alpha_m$, the y_{ij}'s are conditionally independent such that

$$\text{logit}\{\text{pr}(y_{ij} = 1|\alpha)\} = \mu + \alpha_i, \quad 1 \leq j \leq n_i, \tag{40}$$

where μ is an unknown parameter. Furthermore, suppose that the α_i's are independent and distributed as $N(0, \sigma^2)$, where σ^2 is an unknown variance. Model (40) may be viewed as an extension of Example 1 in that the n_i's are different for different i's. Here we consider the cases with n_i ranging from 6 to 10. These are the cases so that, in each category, there are at least 100 litters; see Table 3. Altogether, these five categories account for about 90% of the totally 1328 litters that are involved.

Here we are interested in checking the overall GLMM assumption for the HS2 data. We use (28) as our test statistic. Here $L = 5$, and $M_l = 6 + l, l = 1, \ldots, L$, and $r = 2$. The test statistic has an asymptotic χ^2_{30} distribution. The value of the test statistic computed from the data is 1645.78. The p-value for the test is almost zero. Therefore, the null hypothesis that the proposed GLMM holds is rejected.

As noted earlier [see the discussion below (4)], goodness-of-fit test is often the first step for model diagnostics to check if there is an overall violation of the GLMM assumption. For this particular example, the proposed GLMM is very simple. Thus, the overall GLMM assumption consists of three parts: (i) the independence of the litters; (ii) binomial conditional distribution within each litter; and (iii) the normality assumption for the random effects. Thus, the rejection of the null hypothesis could mean that at least one of these three assumptions is violated.

Typically, model diagnostics does not stop when the goodness-of-fit test rejects the null hypothesis. In such a case, there will be follow-up steps after the goodness-of-fit test. While the follow-up steps are not the focus of the current paper, we would like to offer some thoughts for the current example. Assumption (i) has to do with the biological nature of the data, as well as how the data were collected. While one cannot be sure about the answer without further details, the assumption is not unreasonable if the litters are unrelated. If one can reasonably assume that (i) holds, rejection of the null hypothesis would suggest that either (ii) or (iii) are violated.

Between (ii) and (iii), the latter assumption is more critical. It is known that, under linear mixed effects models, the Gaussian REML estimators of the model parameters remain consistent even without the normality assumption (e.g., Richardson and Welsh 1994, Jiang 1996). However, this is not true under a non-linear GLMM, such as the current case (Jiang and Nguyen 2009). In other words, estimators of the model parameters, namely μ and σ^2, are not consistent if the distribution of the random effects is misspecified. On the other hand, assumption (ii) is less critical in the sense that, as long as the distribution of the random effects is correctly specified, the binomial assumption can be weakened to that only the first two moments, or even the first moment, of the conditional distribution is correctly specified (Jiang and Zhang 2001). Therefore, in a way, the focus shifts to replacing (iii) by a more reasonable distributional assumption. For example, such a distribution could be the broader class suggested by Claeskens and Hart (2009) (Section 1, second-to-last paragraph). One then carries out a similar goodness-of-fit test of a new hypothesis that the distribution of the random effects belong to the broader class. The latter null hypothesis is more likely to be accepted. The process can be repeated, if necessary.

Acknowledgements

Jiming Jiang research is partially supported by the NSF grants SES-1121794 and DMS-1510219, and the NIH grant R01-GM085205A1.

References

Brooks, S. P., Morgan, B. J. T., Ridout, M. S., and Pack, S. E. (1997), Finite mixture models for proportions, *Biometrics* 53, 1097–1115.

Calvin, J. A. and Sedransk, J. (1991), Bayesian and frequentist predictive inference for the patterns of care studies, *J. Amer. Statist. Assoc.* 86, 36–48.

Chernoff, H. and Lehmann, E. L. (1954), The use of maximum-likelihood estimates in χ^2 tests for goodness of fit, *Ann. Math. Statist.* 25, 579–586. MR0065109

Claeskens, G. and Hart, J. D. (2009), Goodness-of-fit tests in mixed models (with discussion), *TEST* 18, 213–239. MR2520330

Fisher, R. A. (1922), On the interpretation of chi-square from contingency tables, and the calculation of P, *J. Roy. Statist. Soc.* 85, 87–94.

Foutz, R. V. (1977), On the unique consistent solution to the likelihood equation, *J. Amer. Statist. Assoc.* 72, 147–148. MR0445686

Gu, Z. (2008), Model diagnostics for generalized linear mixed models, Ph. D. Dissertation, Dept. of Statist., Univ. of Calif., Davis, CA. MR2712300

Jiang, J. (1996), REML estimation: Asymptotic behavior and related topics, *The Annals of Statistics* 24, 255–286. MR1389890

Jiang, J. (1998), Consistent estimators in generalized linear mixed models, *J. Amer. Statist. Assoc.* 93, 720–729. MR1631373

Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, Springer, New York. MR2308058

Jiang, J. (2010), *Large Sample Techniques for Statistics*, Springer, New York. MR2675055

Jiang, J. (2013), The subset argument and consistency of MLE in GLMM: Answer to an open problem and beyond, *Ann. Statist.* 41, 177–195. MR3059414

Jiang, J. and Nguyen, T. (2009), Comments on: Goodness-of-fit tests in mixed models by G. Claeskens and J. D. Hart, *TEST* 18, 248–255. MR2520333

Jiang, J. and Zhang, W. (2001), Robust estimation in generalized linear mixed models, *Biometrika* 88, 753–765. MR1859407

Lange, N. and Ryan, L. (1989), Assessing normality in random effects models, *Ann. Statist.* 17, 624–642. MR0994255

McCulloch, C. E., Searle, S. R. and Neuhaus, J. M. (2008), *Generalized, Linear, and Mixed Models*, 2nd ed., Wiley, New York. MR2431553

Moore, D. S. (1978), Chi-square tests, in *Studies in Statistics* (R. V. Hogg, ed.), Mathematical Society of America, Providence, RI.

Pearson, K. (1900), On a criterion that a given system of deviations from the probable in the case of a corrected system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philos. Mag. 5th Series* 50, 157–175.

Rice, J. A. (1995), *Mathematical Statistics and Data Analysis*, 2nd ed., Duxbury Press, Belmont, CA.

Richardson, A. M. and Welsh, A. H. (1994), Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models, *Austral. J. Statist.* 36, 31–43. MR1309503

Tang, M. (2010), Goodness-of-fit tests for generalized linear mixed models, Ph. D. Dissertation, Dept. of Math., Univ. of Maryland, College Park, MD. MR2794873

Torabi, M. (2012), Likelihood inference in generalized linear mixed models with two components of dispersion using data cloning, *Comput. Statist. Data Anal.* 56, 4259–4265. MR2957869

Cecilia Dao
Department of Statistics
University of California, Davis
Davis, CA 95616
USA
E-mail address: cudao@ucdavis.edu

Jiming Jiang
Department of Statistics
University of California, Davis
Davis, CA 95616
USA
E-mail address: jimjiang@ucdavis.edu

Received November 1, 2015