Conditional Logistic Regression With Longitudinal Follow-up and Individual-Level Random Coefficients: A Stable and Efficient Two-Step Estimation Method

Radu V. CRAIU, Thierry DUCHESNE, Daniel FORTIN,
and Sophie BAILLARGEON

The analysis of data generated by animal habitat selection studies, by family studies of genetic diseases, or by longitudinal follow-up of households often involves fitting a mixed conditional logistic regression model to longitudinal data composed of clusters of matched case-control strata. The estimation of model parameters by maximum likelihood is especially difficult when the number of cases per stratum is greater than one. In this case, the denominator of each cluster contribution to the conditional likelihood involves a complex integral in high dimension, which leads to convergence problems in the numerical maximization. In this article we show how these computational complexities can be bypassed using a global two-step analysis for nonlinear mixed effects models. The first step estimates the cluster-specific parameters and can be achieved with standard statistical methods and software based on maximum likelihood for independent data. The second step uses the EM-algorithm in conjunction with conditional restricted maximum likelihood to estimate the population parameters. We use simulations to demonstrate that the method works well when the analysis is based on a large number of strata per cluster, as in many ecological studies. We apply the proposed two-step approach to evaluate habitat selection by pairs of bison roaming freely in their natural environment. This article has supplementary material online.

Key Words: CREML; EM-algorithm; Habitat selection; Mixed effects; Mixed multinomial logit; One-step estimator; REML; Two-step analysis.

Radu V. Craiu is Associate Professor, Department of Statistics, University of Toronto, Toronto, Ontario M5S 3G3, Canada (E-mail: *craiu@utstat.toronto.edu*). Thierry Duchesne is Professor (E-mail: *Thierry.Duchesne@mat.ulaval.ca*) and Sophie Baillargeon is Research Associate (E-mail: *Sophie.Baillargeon@mat.ulaval.ca*), Departement de mathematiques et de statistique, Universite Laval, Quebec, Quebec G1V 0A6, Canada. Daniel Fortin is Associate Professor and co-holder of the NSERC-Universite Laval Industrial Research Chair in Silviculture and Wildlife, Centre d'etude de la foret and Departement de biologie, Universite Laval, Quebec, Quebec G1V 0A6, Canada (E-mail: *Daniel.Fortin@bio.ulaval.ca*).

Journal of Computational and Graphical Statistics, Volume 20, Number 3, Pages 767–784
DOI: 10.1198/jcgs.2011.09189

1. INTRODUCTION

In many ecological, health, or social science studies, measures are collected on (groups of) individuals who are followed over time. In some instances, the experimental units do not generate a single measurement at each observation time, but rather a stratum of measurements. In the following, *a cluster* will be the collection of all strata generated by an individual or group of correlated individuals. For instance, in the study which motivated this work, each pair of bison (i.e., two individuals roaming together) wearing global positioning system (GPS) collars generates a data stratum composed of $m = 2$ visited locations (with covariates) and $n - m$ random locations (with covariates), and this several times per day for many weeks. Similar data may occur in a medical study where a number of practices are followed up over time and, for each practice at each measurement time, a group of m cases and $n - m$ controls are sampled.

To take within-cluster correlation and between-cluster heterogeneity into account, regression models with cluster-level random effects are often fitted to such data. Though methods to fit generalized linear mixed models to standard longitudinal data have been thoroughly investigated (see Tuerlinckx et al. 2006; Dean and Nielsen 2007, for reviews), the same is not true when we consider data that are both clustered and stratified. There are nonetheless two special cases for which a treatment can be found in the literature. The first special case is when there is a single stratum per cluster as in, for example, family studies of gene-environment interactions. Pfeiffer, Gail, and Pee (2001) analyzed the data from such a study where families with $m = 2$ or more cases of a given rare disease (nasopharyngeal carcinoma) are sampled. They considered maximum likelihood inference for the mixed conditional logistic regression model with family-level random effects. The optimization of the marginal (i.e., with random effects integrated out) likelihood is solved numerically using classic Monte Carlo integration coupled with a hybrid Newton's method/grid search algorithm. The second special case is when there is a single case in each stratum, that is, $m = 1$. This yields the mixed multinomial logit model. This model has been studied in the econometric literature (see Train 2003; Train and Weeks 2005, for reviews). It is usually fitted by numerical maximization of the marginal likelihood that relies on Monte Carlo integration coupled with a Newton-type method (Bhat 2001; Sándor and Train 2004; Hesse, Train, and Polak 2006).

Neither of the methods used in the two special cases above can be used for the type of data produced by the bison study. First, the method must be able to handle the situation where $m > 1$, which is not the case in the mixed multinomial logit model. Second, we have many strata per cluster (several hundred strata in some clusters) which is much more than can be found in family genetic studies where, usually, there is a single stratum per cluster. As will be seen later, this renders numerical evaluation of the mixed conditional logistic regression likelihood function especially difficult and unstable. Indeed, trying to apply a combination of numerical integration and numerical optimization to such a likelihood function is quite problematic and leads to slow execution and lack of convergence. Moreover, we will see in Section 2 that having a very large number of strata per cluster complicates the likelihood function to the point that this type of numerical approach fails to converge when applied to the bison data. One possibility to circumvent such numerical issues could be to use the EM-algorithm to maximize the likelihood, but its complex form due to the conditioning (see Section 2) makes the E- and M-steps as difficult to implement as the numerical integration and maximization discussed above.

One viable option to fit mixed models with large clusters is to use previously developed two-step approaches proposed for the (unconditional) mixed logistic regression model by Korn and Whittemore (1979) and Stiratelli, Laird, and Ware (1984). But this yields highly biased estimators when applied to the case of conditional logistic regression, even though the large number of strata in each cluster justifies the normal approximation for the cluster-level maximum likelihood estimates.

The approach that proved to be most efficient and stable to obtain valid results in our case is the two-step method based on restricted maximum likelihood (REML) estimation, as suggested by Chervoneva, Iglewicz, and Hyslop (2006). In this approach, ordinary (fixed effects) regression models are fitted separately to each cluster in the first step. Then the cluster-level estimates are combined in the second step using REML. However, unlike Chervoneva, Iglewicz, and Hyslop (2006) who used the Newton–Raphson approach of Lindstrom and Bates (1988), in this article we follow Meng and van Dyk (1998) and derive an EM-algorithm to implement the REML estimation in the second step. From a practical point of view, the resulting methodology leads to fast and stable computation of consistent estimators of the parameters of a mixed conditional logistic regression model when the number of strata per cluster is large. More importantly, it is the only method, to our knowledge, that can be used to handle computation for the conditional logistic regression model under such conditions. As an added benefit, we note that this approach can also be used to fit any mixed model where the cluster-level estimates are approximately normal and for which a large number of observations per cluster makes the use of maximum likelihood methods or two-step methods with Newton-type maximization at the second step unstable.

The article is organized as follows. In Section 2 we describe the model and our proposed EM-REML method for inference on the model parameters. Finite sample performance of the estimators, including a comparison with maximum likelihood, is done in Section 3. We apply the method to real data from a habitat selection study of pairs of female bison traveling together in Prince Albert National Park (Canada) in Section 4. We conclude with a discussion in Section 5.

2. MODEL AND METHODS

The data consist of observations gathered together in K clusters, each cluster consisting of a number of strata, each strata containing m cases and $n - m$ controls. For instance, in the ecological study described in Section 4, each cluster consists of all observations relating to the same pair of animals and each stratum consists of the data on the characteristics of $m = 2$ locations the two animals have visited and $n - m = 10$ randomly sampled locations these two animals could have chosen instead.

To illustrate the computational difficulties associated with the model, let us present a simple example of the sampling mechanism. Consider a single pair of animals that is followed for 4 time steps and that measurements are taken at the 4 pairs of locations visited at these time steps. At each time step, in addition to the 2 visited locations (cases), a random selection of 3 additional sites (controls) that could have been chosen by the animals instead of the visited ones are measured. In this case, there are 4 strata (one per time step) and each stratum contains the data from the 2 cases and 3 controls. The cluster consists of all observations in all 4 strata. The measurements for stratum $s \in \{1, 2, 3, 4\}$ thus contain

(i) the vector of binary responses $\tilde{y}_s = (y_{s1}, \dots, y_{s5})$ where $y_{sj} = 1$ if the location j is one of the two visited locations and is zero otherwise,

(ii) the covariate values $\tilde{x}_s = \{x_{s1}, \dots, x_{s5}\}$ where x_{sj} are the characteristics of the location j.

Naturally, the responses in the same cluster are not independent since they are produced by the same pair of animals. In a prospective study where the stratum sums would not be fixed before sampling, such longitudinal binary data could be modeled using a mixed logistic regression model, that is, a logistic regression model where unobserved cluster-level random effects are added in the linear predictor to induce within-cluster correlation and between-cluster heterogeneity. Conditional on the cluster-specific random effects **b** whose distribution in the population has c.d.f. $F(\mathbf{b}|\theta)$, the observations are assumed independent. So the cluster's contribution to the (observed data) likelihood is

$$\int \prod_{s=1}^{4} \prod_{i=1}^{5} \mathrm{P}(Y_{si} = y_{si} | x_{si}, \mathbf{b}) \, dF(\mathbf{b}|\theta). \tag{2.1}$$

Throughout the article we use the logit link, that is,

$$\mathrm{P}(Y_{si} = y_{si} | x_{si}, \mathbf{b}) = \frac{\exp\{y_{si}(\beta^\top x_{si} + \mathbf{b}^\top z_{si})\}}{1 + \exp(\beta^\top x_{si} + \mathbf{b}^\top z_{si})}, \tag{2.2}$$

where the z_{si} are the subset of the covariates for which we assume random coefficients.

However, the standard mixed effects model defined by (2.1) and (2.2) does not account for the fact that the number of cases per stratum is fixed by sampling design. We must therefore consider a *retrospective* conditional logistic model that conditions on the stratum sums. The contribution of the cluster to the likelihood becomes

$$\begin{aligned} L(\beta, \theta | \tilde{x}, \tilde{y}) &= \mathrm{P}\left(\tilde{Y}_s = \tilde{y}_s, 1 \le s \le 4 \,\middle|\, \tilde{X}_s = \tilde{x}_s, \sum_{i=1}^{5} Y_{si} = 2, 1 \le s \le 4 \right) \\ &= \frac{\mathrm{P}(\tilde{Y}_s = \tilde{y}_s, 1 \le s \le 4 | \tilde{x}_s, 1 \le s \le 4)}{\mathrm{P}(\sum_i Y_{si} = 2, 1 \le s \le 4 | \tilde{x}_s, 1 \le s \le 4)} \\ &= \frac{\int \prod_{s=1}^{4} \prod_{i=1}^{5} \mathrm{P}(\tilde{Y}_{si} = \tilde{y}_{si} | \tilde{X}_s = \tilde{x}_s, \mathbf{b}) \, dF(\mathbf{b}|\theta)}{\int \prod_{s=1}^{4} \mathrm{P}(\sum_i Y_{si} = 2 | \tilde{x}_s, \mathbf{b}) \, dF(\mathbf{b}|\theta)}. \end{aligned} \tag{2.3}$$

The main difficulty comes from the denominator in (2.3) that involves a product of sums of $\binom{5}{2} = 10$ terms corresponding to all the unordered vectors of length 5 containing 3 zeros and 2 ones.

2.1 NOTATION, MODEL, AND LIKELIHOOD

Although the above description is illustrative for the dependence structure of the data, the dimensionality of the real application is much higher and thus, the computational complexity is seriously amplified. The methods that we propose in this article are appropriate for situations where the number of strata is large in every cluster; in the application studied in Section 4, this number varies between 31 and 349, with most clusters containing more than 100 strata.

Let us now consider a general situation with more than one cluster, then assume that in the cth cluster we observe S^c strata where, in the sth stratum, we have n_s^c observations each made of a binary response y and a vector of covariates $\mathbf{x} = (x_1, \ldots, x_p)^\top$. By study design, the sum of the y's in the sth stratum of the cth cluster is fixed to m_s^c. Note that in many applications, all strata have the same size (i.e., $n_s^c = n$ for all s, c) and the sum of the y's in all clusters is fixed to m (i.e., $m_s^c = m$ for all s, c).

More precisely, let Y_{si}^c denote the ith response in the sth stratum of the cth cluster and let $\mathbf{x}_{si}^c$ be the corresponding vector of covariates. The covariate vector for the whole stratum is denoted $\mathbf{x}_s^c$. Let $\mathbf{b}_c$ be a vector of cluster-level random effects. The contribution of the cth cluster to the likelihood is

$$\begin{aligned} L_c(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \mathrm{P}\left[Y_{si}^c = y_{si}^c, \forall s, i \,\middle|\, \mathbf{x}_{si}^c, \sum_i Y_{si}^c = m_s^c, \forall s, i. \right] \\ &= \frac{\int \prod_s \prod_i \mathrm{P}[Y_{si}^c = y_{si}^c | \mathbf{b}_c, \mathbf{x}_{si}^c]\, dF(\mathbf{b_c}; \boldsymbol{\theta})}{\int \prod_s \mathrm{P}[\sum_i Y_{si}^c = m_s^c | \mathbf{b}_c, \mathbf{x}_s^c]\, dF(\mathbf{b_c}; \boldsymbol{\theta})}. \end{aligned} \tag{2.4}$$

Combining (2.2) and (2.4) yields the likelihood function

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{c=1}^{K} \frac{\exp(\sum_{si} y_{si}^c \boldsymbol{\beta}^\top \mathbf{x}_{si}^c) \int \exp(\sum_{si} y_{si}^c \mathbf{b}^\top \mathbf{z}_{si}^c) d^c(\boldsymbol{\beta}, \mathbf{b})\, dF(\mathbf{b}; \boldsymbol{\theta})}{\int d^c(\boldsymbol{\beta}, \mathbf{b}) \prod_s \sum_{\ell \in \mathcal{L}_s^c} \exp\{\sum_i v_{\ell si}^c (\boldsymbol{\beta}^\top \mathbf{x}_{si}^c + \mathbf{b}^\top \mathbf{z}_{si}^c)\}\, dF(\mathbf{b}; \boldsymbol{\theta})}, \tag{2.5}$$

where $d^c(\boldsymbol{\beta}, \mathbf{b}) = \prod_s \prod_i \{1 + \exp(\boldsymbol{\beta}^\top \mathbf{x}_{si}^c + \mathbf{b}^\top \mathbf{z}_{si}^c)\}^{-1}$ and $\sum_{\ell \in \mathcal{L}_s^c}$ denotes the sum over all vectors $\mathbf{v}_{\ell s}^c = (v_{\ell s1}^c, v_{\ell s2}^c, \ldots, v_{\ell s n_s^c}^c)^\top$ such that $v_{\ell si}^c \in \{0, 1\}$ and $\sum_i v_{\ell si}^c = m_s^c$.

The "direct maximization" of (2.5) using, for instance, the method proposed by Pfeiffer, Gail, and Pee (2001), leads to a heavy computational load and numerical instability. Moreover, in cases where the number of observations in a cluster is large (e.g., in our case study, we have well over a hundred strata of size 12 in most clusters), both the numerator and denominator in (2.5) involve sums of products of thousands of probabilities which quickly becomes indistinguishable from zero in double precision. To overcome this problem, we have to implement the numerical evaluation of the likelihood with all figures coded in normalized scientific notation. The probabilities are stored using two distinct values: a double for the significand and an integer for the exponent. This allows working with much smaller numbers compared to using a unique double precision value. For example, the number 6.25×10^{-355} is too small to be stored as one double precision number. However, one can store its significand 6.25 as a double precision number and its exponent -355 as an integer number. The numerical algorithm that we used relied on iterating quasi-Monte Carlo integration and the variable metric quasi-Newton algorithm (the BFGS option of the function `optim` in R; see, for instance, Broyden, Dennis, and Moré 1973) for finding the maximum likelihood estimates.

Though this type of approach works well for generalized linear mixed models with clusters that are not too big or with the models used in genetic studies or in econometrics, with our specific model we will see in Section 3 that it only achieves moderate success when the number of strata in a cluster is relatively small. As a matter of fact, we were unable to get this approach to converge when we applied it to the bison data considered in Section 4. One should also note that the denominator in (2.5) makes it very difficult to implement the usual EM- or Monte Carlo EM-algorithms one can sometimes use for generalized mixed effects models (McCulloch 2003; Caffo, Jank, and Jones 2005, McCulloch, Searle, and Neuhaus 2008). We thus have to find a different approach that is numerically efficient and stable to make inferences about $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

2.2 TWO-STEP ESTIMATION

For the remainder of this article, we make the assumption that $F(\mathbf{b};\boldsymbol{\theta})$ is the multivariate normal distribution with mean $\mathbf{0}$ and variance–covariance matrix $\boldsymbol{\Sigma}$. Various structural forms can be assumed for $\boldsymbol{\Sigma}$ depending on the subject-matter knowledge, with more structured forms (e.g., diagonal) sometimes entailing faster convergence of the iterative algorithm proposed below. Assuming a joint normal distribution for unobserved random effects in mixed models is usually the norm in practice (Tuerlinckx et al. 2006). Several authors have investigated the robustness of the inferences to departures from this normality assumption. In the case of mixed models for binary responses, Agresti, Caffo, and Ohman-Strickland (2004) have found that inferences tend to be unbiased under general departures from normality and only found a drop in efficiency when the random effects' true distribution is discrete with large variance. As a matter of fact, for this type of model Litière, Alonso, and Molenberghs (2007) formally showed (see their theorem 1 and corollary 1) that a significant finding (i.e., β significantly different from zero) is a reliable result even under a misspecified distribution of the random effects. We make this normality assumption because (i) it allows us to use two-step estimation methods instead of direct maximum likelihood and (ii) it yields closed-form E- and M-steps in the EM-algorithm that we propose for the second step of the two-step estimation procedure.

Two-step estimation methods are among the first approaches used to make inferences on the parameters of mixed regression models for binary responses (Korn and Whittemore 1979; Stiratelli, Laird, and Ware 1984). Because in these models the observations are assumed independent given the value of the cluster-level random effects, cluster-level inferences on the model parameters can be done using ordinary (fixed-effects) regression models. Hence in a first step, classical regression models are fitted by maximum likelihood separately to each of the K clusters and cluster-level estimates of the regression coefficients are obtained, along with an estimate of their variance matrix. If the number of observations in each cluster is large, then likelihood theory implies that each of these K regression coefficient estimates is approximately normal. When combined with the normality of the random effects, we then have that estimating the population-level coefficients amounts to estimating the mean of a multivariate normal distribution based on K independent observations (the cluster-level estimates from the first step) with different variances (the variance estimates from the first step). The procedures designed by Davidian and Giltinan (1993) and Chervoneva, Iglewicz, and Hyslop (2006) consist in using REML estimation for linear mixed models in the second step to estimate the population parameters on the basis of the cluster-level estimates obtained in the first step. We show in Section 2.3 that under conditional logistic regression and our sampling scheme, inferences will be valid and efficient.

Mathematically, given p covariates we generally assume that the model has q random coefficients and $p - q$ fixed coefficients. The linear predictor for observation (s, i) in the cth cluster is thus of the form $(\boldsymbol{\beta} + \mathbf{b}_c)^\top \mathbf{x}_{si}^c$. If $p > q$, $\mathbf{b}_c$ is extended to dimension p by adding $p - q$ zeros in positions corresponding to nonrandom coefficients, with corresponding elements in $\boldsymbol{\Sigma}$ also set to zero. For instance, if we have that $p = 3$ and that the coefficients in front of x_1 and x_3 are random, then $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ and $\mathbf{b}_c = (b_{c1}, 0, b_{c3})^\top$. Let $\{\widehat{\boldsymbol{\beta}}_c,\ c = 1, \ldots, K\}$ denote the K cluster-specific estimators of $\boldsymbol{\beta}$, obtained by fitting the model that assumes only fixed effects separately to each cluster and let $\{\hat{\mathbf{R}}_c,\ c = 1, \ldots, K\}$ be their respective variance estimates. In the case of the conditional logistic regression model, procedures that fit the Cox model using the exact method to handle ties (Gail, Lubin, and Rubinstein 1981) can be used to obtain the $\hat{\boldsymbol{\beta}}_c$ and $\hat{\mathbf{R}}_c$.

These first-step estimates must then be “combined” to obtain population-level estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. In our initial trials we followed Korn and Whittemore (1979) and Stiratelli, Laird, and Ware (1984) and used maximum likelihood to estimate $\boldsymbol{\theta}$ in the second step. This was unsatisfactory because the variance parameters in $\boldsymbol{\theta}$ were badly underestimated, which led to biased estimates of $\boldsymbol{\beta}$. Following a suggestion made by Stiratelli, Laird, and Ware (1984) and implemented elsewhere by Lindstrom and Bates (1990), Davidian and Giltinan (1993), and Chervoneva, Iglewicz, and Hyslop (2006), we derived a REML approach to estimate $\boldsymbol{\theta}$ in the second step. Unlike these authors, however, we implemented REML estimation using the EM-algorithm instead of the Newton–Raphson method. As discussed by Meng and van Dyk (1998), using the EM-algorithm has certain advantages over Newton-type methods in the context of mixed effects models: it requires less monitoring and it does not produce negative variance estimates. Moreover, in the precise context of mixed conditional logistic regression, the retrospective sampling design implies that the data might not contain much information about some parameters in $\boldsymbol{\theta}$ and hence numerical robustness and stability of the algorithm is preferable to numerical efficiency (for more detail, see Pfeiffer, Gail, and Pee (2001), who actually performed a grid search instead of Newton–Raphson for maximization over some elements of $\boldsymbol{\theta}$). Note that in our case the price to pay in terms of numerical efficiency was not significant: in our simulated and real analyses, the proposed two-step EM-REML method is fast and stable and never took more than a few seconds to run. Nevertheless, the algorithm proposed can still be further sped-up if need be by adapting the general algorithm and strategies for REML estimation in linear mixed-effects models presented by Meng and van Dyk (1998) to the situation in which the noise variance–covariance matrix is known.

We now illustrate the EM-REML algorithm with a simple example. The general method is detailed in the Appendix (available as supplemental file from the *JCGS* website). Suppose that we have three pairs of animals, so that $c = 1, 2, 3$ and that we would like to fit a model with one random coefficient $(\beta_1 + b_{c1})$ for covariate x_1 and one fixed coefficient $(\beta_2 + 0)$ for covariate x_2. Thus here $K = 3$, $p = 2$, $q = 1$, $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} \theta_1^2 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, $\text{var}(b_{c1}) = \theta_1^2$, and $\mathbf{D} = \mathbf{I}_3 \otimes \boldsymbol{\Sigma}$, where $\mathbf{I}_3$ denotes the identity matrix and $\otimes$ denotes the Kronecker product. The two-step method proceeds in this way:

Step 1—Cluster-level estimation: For each $c \in \{1, 2, 3\}$: obtain $\widehat{\boldsymbol{\beta}}_c = (\widehat{\beta}_{c1}, \widehat{\beta}_{c2})^\top$, the maximum likelihood estimator of $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$, and its variance–covariance estimate $\widehat{\mathbf{R}}_c$ using only the data from the cth pair of animals and under the assumption that all strata in cluster c are independent. It should be noted that by separately fitting the model to each cluster we perform conditional inference so $\hat{\beta}_{c1}$ actually estimates $\beta_1 + b_{c1}$. Put $\widehat{\boldsymbol{\beta}}_{1st} = (\widehat{\boldsymbol{\beta}}_1^\top, \widehat{\boldsymbol{\beta}}_2^\top, \widehat{\boldsymbol{\beta}}_3^\top)^\top$.

Step 2—EM-REML estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$:

1. In line with the notation used in the Appendix, here we set $\tilde{\boldsymbol{\Sigma}} = \theta_1^2$, $\mathbf{U} = (\hat{\beta}_{11}, \hat{\beta}_{21}, \hat{\beta}_{31})^\top$, $\tilde{\mathbf{R}} = \text{diag}(\tilde{\mathbf{R}}_1, \tilde{\mathbf{R}}_2, \tilde{\mathbf{R}}_3)$ where $\tilde{\mathbf{R}}_c = \widehat{\text{var}}(\hat{\beta}_{c1})$, and $\tilde{\mathbf{D}} = \theta_1^2 \mathbf{I}_3$.
2. Compute $\mathbf{W}_1 = (1\ \ 1\ \ 1)^\top$ and $\mathbf{M} = \mathbf{I}_3 - \mathbf{W}_1 \mathbf{W}_1^\top / 3$.
   Put $\mathbf{S}^{\tilde{\mathbf{D}}} = \{\mathbf{M}(\mathbf{M}^\top \tilde{\mathbf{R}} \mathbf{M})^{-1} \mathbf{M}^\top + \tilde{\mathbf{D}}^{-1}\}^{-1}$ and $\boldsymbol{\mu}^{\tilde{\mathbf{D}}} = \mathbf{S}^{\tilde{\mathbf{D}}} \mathbf{M}(\mathbf{M}^\top \tilde{\mathbf{R}} \mathbf{M})^{-1} \mathbf{M}^\top \mathbf{U}$.
3. Set an initial value for θ_1^2.
4. Iterate the E- and M-steps of the EM-algorithm described below until convergence:
   E-step: Compute
   $$Q(\tilde{\mathbf{D}} | \tilde{\mathbf{D}}^*) = -\tfrac{3}{2} \ln(\theta_1^2) - \tfrac{1}{2} \text{tr}\big\{\tilde{\mathbf{D}}^{-1}\big(\mathbf{S}^{\tilde{\mathbf{D}}^*} + \boldsymbol{\mu}^{\tilde{\mathbf{D}}^*} \boldsymbol{\mu}^{\tilde{\mathbf{D}}^* \top}\big)\big\}.$$
   M-step: Get
   $$\theta_1^2 = \frac{1}{3} \sum_{c=1}^{3} \big(\mathbf{S}_{cc}^{\tilde{\mathbf{D}}^*} + \boldsymbol{\mu}_c^{\tilde{\mathbf{D}}^*} \boldsymbol{\mu}_c^{\tilde{\mathbf{D}}^* \top}\big),$$
   where $\mathbf{S}_{cc}^{\tilde{\mathbf{D}}^*}$ is the cth element on the diagonal of $\mathbf{S}^{\tilde{\mathbf{D}}^*}$ and $\boldsymbol{\mu}_c^{\tilde{\mathbf{D}}^*}$ is the cth entry in the vector $\boldsymbol{\mu}^{\tilde{\mathbf{D}}^*}$.
5. Compute $\mathbf{Q} = (1\ \ 1\ \ 1) \otimes \mathbf{I}_2$ and $\hat{\mathbf{V}} = \hat{\mathbf{D}} + \hat{\mathbf{R}}$. The estimator $\hat{\boldsymbol{\beta}}$ and its variance are respectively given by
   $$\hat{\boldsymbol{\beta}} = (\mathbf{Q} \hat{\mathbf{V}}^{-1} \mathbf{Q}^\top)^{-1} \mathbf{Q} \hat{\mathbf{V}}^{-1} \widehat{\boldsymbol{\beta}}_{1st}$$
   and
   $$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{Q} \hat{\mathbf{V}}^{-1} \mathbf{Q}^\top)^{-1}. \tag{2.6}$$

2.3 PROPERTIES OF THE ESTIMATORS

The estimation strategy used in the article falls within the general scope of two-step conditional restricted maximum likelihood (CREML) introduced by Chervoneva, Iglewicz, and Hyslop (2006). The assumptions made by CREML are that the first-step estimates, $\{\widehat{\boldsymbol{\beta}}_c\}_{1 \le c \le K}$, are consistent and asymptotically normal and the variance estimates $\widehat{\mathbf{R}}_c$ are consistent. Maximum likelihood estimators for conditional logistic regression do fulfill these requirements as the number of strata tends to infinity (Arbogast and Lin 2004). Because these estimators still fulfill these requirements if data are *missing completely at random* (MCAR) or *missing at random* (MAR) (Ibrahim and Molenberghs 2009), the method proposed here yields valid inferences under these two missing-data schemes.

It should be noted that the estimates obtained with the above EM-REML algorithm are different from standard REML estimates since we assume the noise covariance matrix known. Under mild regularity conditions the CREML estimators of $\mathbf{D}$ and $\boldsymbol{\beta}$, denoted $\widehat{\mathbf{D}}_{CREML}$ and $\widehat{\boldsymbol{\beta}}_{CREML}$ respectively, are consistent estimators with rate of convergence $O_P[\max\{K^{-1/2}, (\min_{1\le c\le K} n_s^c S^c)^{-1/2}\}]$ (Chervoneva, Iglewicz, and Hyslop 2006). In addition, if $\min_c n_s^c S^c/K \to \infty$, then $\widehat{\boldsymbol{\beta}}_{CREML}$ is asymptotically efficient with

$$\sqrt{K}(\widehat{\boldsymbol{\beta}}_{CREML} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{D}). \tag{2.7}$$

In studies where the number of strata is much larger than the number of clusters, the asymptotic result (2.7) provides reasonable approximations for standard errors and confidence intervals. But in finite samples $\widehat{\mathbf{D}}_{CREML}$ generally underestimates the variance of $\widehat{\boldsymbol{\beta}}_{CREML}$. Our simulations in the next section show that the variance estimate given in (2.6) provides an accurate estimate of the variance of $\widehat{\boldsymbol{\beta}}_{CREML}$.

2.4 Hybrid Estimator

As a compromise between full maximum likelihood estimation and the proposed two-step EM-REML estimator, we can consider a *hybrid* estimate, in which the two-step estimate is used as the starting value of the quasi-Newton algorithm which is run for *only one* iteration. The latter is actually the traditional one-step maximum likelihood estimator (Simpson, Ruppert, and Carroll 1992, for instance) but we prefer to call it "hybrid" here because "one-step" might cause confusion with "two-step." It is hoped that the hybrid estimator has better efficiency than the two-step estimator, but without the numerical instability of the full maximum likelihood estimator. We investigate this question empirically in the simulation study section.

3. SIMULATION STUDY

The purpose of the simulations is: (1) to compare the performance of the proposed two-step estimator with that of the hybrid estimator and the MLE obtained using a numerical quasi-Newton method, and (2) to study the behavior of the two-step estimator when the competing methods do not converge.

Within each step of the algorithm used for computing the MLE we approximate the value of the log-likelihood using Monte Carlo integration. For low-dimensional integrals, the estimators based on quasi-Monte Carlo (QMC) samples have been known to often be more efficient than those using Monte Carlo samples (Owen 2003). In our approximation we use randomized Korobov quasi-random sequences (Lemieux 2009) to approximate the integrals involved in (2.5) (see also Bhat 2001; Jank 2005, for other uses of QMC in similar contexts). Although we have used large quasi-Monte Carlo samples ($M = 4000$) to estimate the likelihood function, in a few replications (about 2–5%) the numerical quasi-Newton does not stabilize and we discard the output produced. During pilot runs it has been found that the number of clusters has a small effect on the estimators' performance so we have fixed $K = 30$ throughout the simulation. The number of strata is $S = 30$ or 60 and the covariance matrix for the random effects, $\boldsymbol{\Sigma} \in \mathbf{R}^{2\times 2}$, is defined such that $\boldsymbol{\Sigma}_{ii} = s$ and $\boldsymbol{\Sigma}_{ij} = \rho s$ whenever $i \neq j$. We considered values for $s \in \{0.5, 1.25\}$ and $\rho \in \{0, 0.6\}$. In Figures 1 and 2 we report the results when we assume, correctly, $\rho = 0$ (diagonal $\boldsymbol{\Sigma}$) and $p = q = 2$. In addition, while we use $n = 5$, we vary the number of cases in each stratum as $m = 1, 2,$ or 3. Also fixed throughout are the fixed effects $\beta_1 = 0.75$ and $\beta_2 = 1.25$, and the distribution of the two covariates, x_1 and x_2, which we chose as independent $N(0, 0.5)$. The two-step EM-REML method was entirely coded in `R`. For the maximum likelihood and hybrid estimates, we used `R`'s implementation of the quasi-Newton method (i.e., the `optim()` function), but the quasi-Monte Carlo evaluation of the likelihood function (2.5) was done in `C`. The computer code for the two-step EM-REML method is available in package `TwoStepCLogit` on the Comprehensive R Archive Network (CRAN).

The results shown in Figures 1 and 2 suggest that the two-step estimators mimic the performance of the MLEs when the latter can be computed, that is, when the quasi-Newton converges. The two-step estimates are very accurate when the number of strata is large, and even more so when $m > 1$. The hybrid estimate is usually numerically between the two-step and maximum likelihood estimates. Numerically, it is also much more stable than the latter. One bonus of the two-step strategy is that asymptotic estimates for the variances of the estimators are easily obtainable via (2.6), while the Hessian of (2.5) would need to be estimated and inverted to get standard errors for the MLEs.

Additional simulation results obtained when using an unstructured model for $\boldsymbol{\Sigma}$ are reported in Table 1 when $s \in \{0.2, 0.5\}$, $\beta_1 = 0.75$, $\beta_2 = 1.25$, $K = 30$, $S = 60$, $m = 2$, and $n = 12$.

The results suggest that:

(i) the estimates for $\boldsymbol{\beta}$ remain on target even when the random effects are correlated;

(ii) the number of parameters (or covariates included in the model) affects more the estimators' bias than their variance;

(iii) the precision of the estimates is affected by the variance of the random effects;

(iv) additional simulations reported in the supplemental material suggest that the estimates for $\boldsymbol{\beta}$ and the diagonal elements of $\boldsymbol{\Sigma}$ are fairly robust to misspecification of the matrix $\boldsymbol{\Sigma}$.

In all simulations performed, the Monte Carlo mean of the standard errors based on the variance estimator (2.6) for the two-step method closely agrees with the observed Monte Carlo standard deviations. It is striking that in all cases the Monte Carlo standard deviations for the hybrid and maximum likelihood estimates are actually *larger* than the two-step, which contradicts theoretical expectations. This loss in efficiency of the hybrid and MLE is a result of the numerical integration error that is added to the statistical estimation error when the Monte Carlo integrations are performed within the quasi-Newton algorithm.

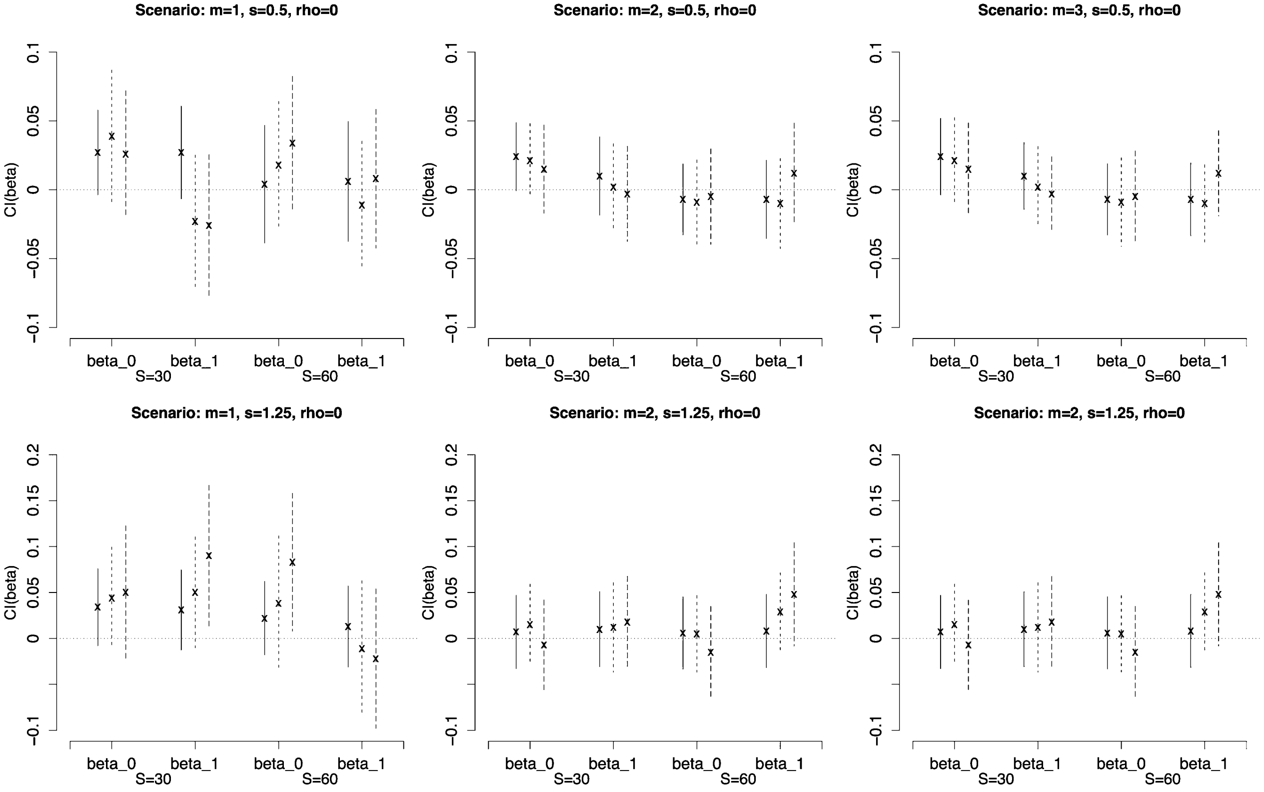

Figure 1. Comparison of the two-step (solid line), hybrid (short-dash line), and ML (long-dash line) estimators for β_0 and β_1 when $\mathbf{\Sigma}$ is assumed diagonal. Each segment has as bounds $\text{CI}(\beta) = \hat{\beta}_{MC} - \beta \pm 1.96\sqrt{\widehat{\text{var}}_{MC}(\hat{\beta})}$ where $\hat{\beta}_{MC}$ and $\widehat{\text{var}}_{MC}(\hat{\beta})$ are, respectively, the Monte Carlo average estimate and Monte Carlo variance computed from 1000 replicates.

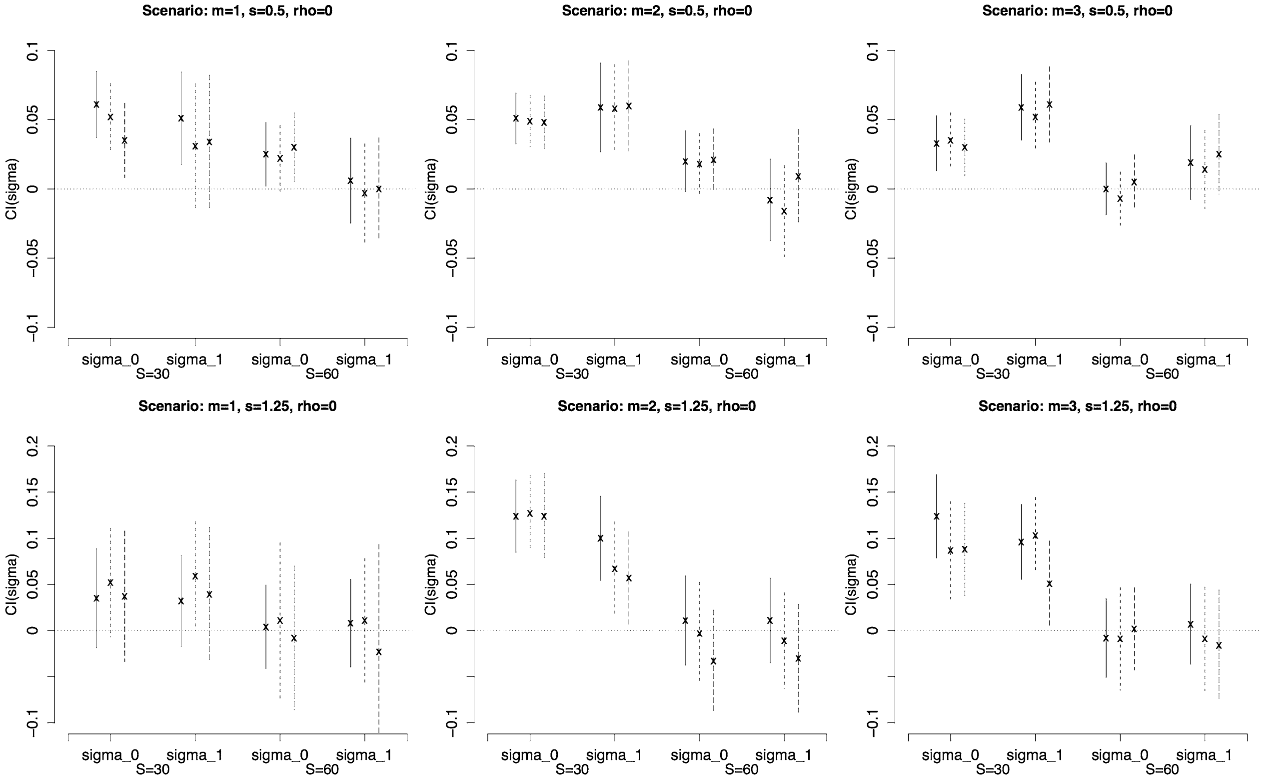

Figure 2. Comparison of the two-step (solid line), hybrid (short-dash line), and ML (long-dash line) estimators for $\sigma_0 = \mathbf{\Sigma}_{11}$ and $\sigma_1 = \mathbf{\Sigma}_{22}$ when $\mathbf{\Sigma}$ is assumed diagonal. Each segment has as bounds $\mathrm{CI}(\sigma) = \hat{\sigma}_{MC} - \sigma \pm 1.96\sqrt{\widehat{\mathrm{var}}_{MC}(\hat{\sigma})}$ where $\hat{\sigma}_{MC}$ and $\widehat{\mathrm{var}}_{MC}(\hat{\sigma})$ are, respectively, the Monte Carlo average estimate and Monte Carlo variance computed from 1000 replicates.

Table 1. Simulation results when no structure is assumed for $\mathbf{\Sigma}$. Throughout $\mathbf{\Sigma}_{11} = \mathbf{\Sigma}_{22} = s$, $\mathbf{\Sigma}_{12} = \rho s$, $\beta_1 = 0.75$, and $\beta_2 = 1.25$. True values of the parameters p, q, s, ρ are reported in the column "Scenario." Each cell entry shows the Monte Carlo average estimate and the Monte Carlo standard error (between brackets) for the two-step estimator.

Scenario	$\beta_1 = 0.75$ $\beta_2 = 1.25$	$\mathbf{\Sigma}_{11} = s$ $\mathbf{\Sigma}_{22} = s$	$\mathbf{\Sigma}_{12} = \rho s$
$(p = q = 2, \rho = 0, s = 0.2)$	0.744 (0.093) 1.234 (0.095)	0.195 (0.063) 0.198 (0.066)	0.008 (0.046)
$(p = q = 2, \rho = 0.6, s = 0.2)$	0.742 (0.092) 1.238 (0.094)	0.196 (0.065) 0.197 (0.067)	0.128 (0.053)
$(p = q = 2, \rho = 0, s = 0.5)$	0.747 (0.132) 1.239 (0.133)	0.482 (0.148) 0.481 (0.144)	0.010 (0.100)
$(p = q = 2, \rho = 0.6, s = 0.5)$	0.748 (0.132) 1.240 (0.133)	0.484 (0.145) 0.483 (0.144)	0.303 (0.116)
$(p = q = 8, \rho = 0, s = 0.2)$	0.779 (0.099) 1.300 (0.101)	0.220 (0.082) 0.219 (0.084)	0.011 (0.058)
$(p = q = 8, \rho = 0.6, s = 0.2)$	0.786 (0.101) 1.310 (0.105)	0.236 (0.082) 0.246 (0.084)	0.153 (0.074)
$(p = q = 8, \rho = 0, s = 0.5)$	0.784 (0.146) 1.304 (0.155)	0.554 (0.175) 0.547 (0.176)	0.011 (0.129)
$(p = q = 8, \rho = 0.6, s = 0.5)$	0.794 (0.155) 1.318 (0.162)	0.588 (0.196) 0.606 (0.225)	0.385 (0.166)

An additional important element to consider in the comparison is the time required for obtaining each of these estimates. Timewise, on average, the hybrid estimate and the maximum likelihood are, respectively, between 600–1800 and 1500–3500 times more expensive than the two-step calculation. Moreover, once the number of parameters in the model is increased (e.g., a larger number of covariates are included as in the application example described in the next section), then the ML method cannot be completed at all due to numerical errors.

4. APPLICATION: HABITAT SELECTION BY PAIRS OF BISON

Gregarious animals generally move as social units so it is reasonable to assume that group members do not select their habitat independently. On the other hand, as they move across the landscape, group members occupy a certain area that can be more or less extensive, depending in part on group size and how individuals react to habitat heterogeneity. Individuals of the same group may thus experience different habitat characteristics at any one time, depending on the extent of the area occupied by the group and the spatial autocorrelation in these characteristics. To gain a general understanding of habitat selection by gregarious animals, multiple individuals can be followed within each group, and the analysis can be conducted while accounting for the dependence between these individuals.

The objective of this analysis was to evaluate habitat selection by groups of free-ranging bison. For each observed group, we followed simultaneously two individuals (dyad) equipped with GPS radio-collars, and evaluated the habitat selection of the bison dyads with mixed-effects conditional logistic regression.

Bison were followed in Prince Albert National Park (53°44′N, 106°40′W), Saskatchewan (Canada). Within the park, the bison range is composed of approximately 85% forests, 10% meadows, and 5% lakes and rivers (see Fortin et al. 2003). From 2005 to 2007, 16 bison were followed for up to 2 years each, between 15 November and 15 April (except in 2005 when radio-tracking started on 10 March). GPS collars took a location every 3 hours.

We considered that two radio-collared bison belonged to the same group when they were <100 m apart for at least five consecutive locations. Because group members can move with a certain delay, we also assumed that individuals from a given pair had remained part of the same group even if they had been separated by >100 m for up to two locations before reuniting.

In our analyses, we considered specific pairs of bison as our experimental units (clusters). A given individual could be part of more than one dyad, but data from different dyads involving a particular individual were never collected at the same time. It was therefore reasonable to assume independence among dyads.

Habitat selection was studied based on resource selection functions (RSFs), an analytical approach commonly used in ecological studies (Fortin et al. 2009). RSFs compare landscape attributes at animal locations and at random locations (see Manly et al. 2002), and can be estimated based on a case-control design as in the work of Craiu, Duchesne, and Fortin (2008). A stratum is generally composed of a visited location ($Y = 1$) paired with a set of random locations ($Y = 0$). In our case, however, the interest was to evaluate how bison dyads selected their habitat. A stratum was therefore composed of two GPS locations (one for each individual) gathered at the same time, together with 10 random locations (five drawn within 700 m of each of the two focal bison). The 700-m distance included $>90\%$ of all distances moved by bison within the 3-hr relocation interval.

We fit the mixed conditional logistic regression model where all regression coefficients are assumed to be independent zero-mean normal random variables, with the variance of the jth coefficient unknown and denoted by θ_j. Covariates include a dummy variable indicating whether bison were observed inside or outside a meadow, the proportion of meadow in a circular plot (700 m in radius) centered at the locations, and an above-ground vegetation biomass index measured (in $\mathrm{kg/m^2}$) only at locations within meadows (otherwise the index is set to zero).

We found that bison dyads selected meadows with an intermediate level of vegetation biomass (Figure 3). Indeed, with either a backward or forward selection approach the best relationship was quadratic, regardless of the sampling scheme considered (Table 2). Ecological theory provides an explanation for this preference. Energy maximization principles predict that grazers should select vegetation of intermediate biomass, a prediction known as the forage maturation hypothesis (Fryxell 1991; Hebblewhite, Merrill, and McDermid 2008). As plants mature and the overall vegetation biomass increases, individual plants become less digestible for large herbivores. On the other hand, herbivores can consume vegetation more rapidly under high than low vegetation biomass. The combined effect is

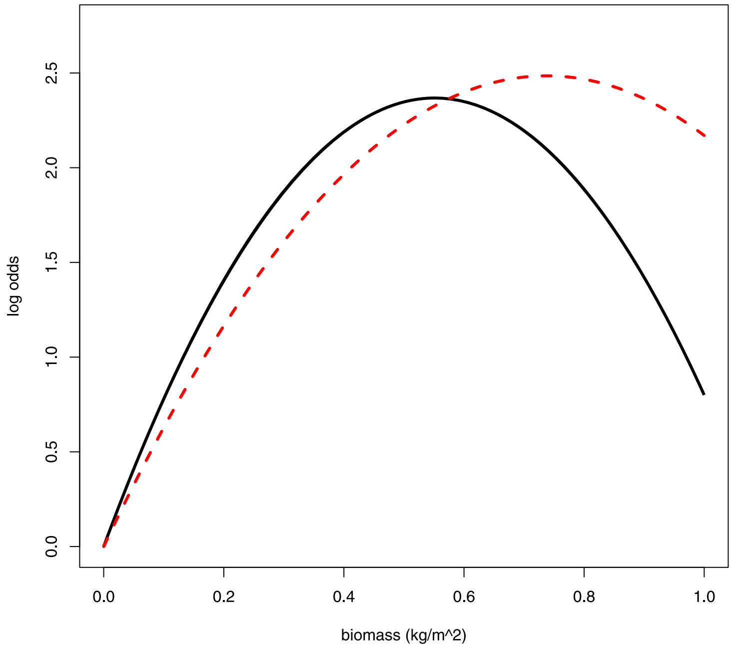

Figure 3. Log-odds of selection for a given value of biomass versus biomass = 0. The solid line is estimated using all data and the dashed line is estimated using only data from dyads with at least 30 strata. The online version of this figure is in color.

Table 2. Bison habitat selection: estimates of the regression coefficients with standard errors (between brackets), estimates of the variance components ($\hat{\theta}_1$ and $\hat{\theta}_2$), number of dyads (K), and total number of strata ($\sum_c S^c$).

Sampling scheme	Biomass ($\hat{\beta}_1$)	Biomass^2 ($\hat{\beta}_2$)	$\hat{\theta}_1$	$\hat{\theta}_2$	K	$\sum_c S^c$
All dyads	8.59 (1.13)	−7.79 (2.07)	13.8	48.4	20	1410
Dyads with $S > 30$	6.74 (1.07)	−4.57 (1.96)	7.36	26.2	13	1261

such that the intake rate of digestible energy is often maximized at intermediate biomass. The pattern of habitat selection that we detected for bison dyads is therefore consistent with expectations from the forage maturation hypothesis. Because biomass was assigned a value of 0 outside meadows, this general selection for plant biomass also implies that bison selected meadows over other areas. We could not explain additional variation by considering the proportion of meadow within a 700-m radius.

5. DISCUSSION

In this article we proposed a two-step method to fit the conditional logistic regression model with mixed effects in situations where case-control strata are collected longitudinally and the number of cases per stratum may be more than one. From the theoretical investigations of Chervoneva, Iglewicz, and Hyslop (2006) and the simulation study, we see that the method proposed yields estimators that are consistent and efficient when the number of strata per cluster is large enough to make cluster-level estimators approximately normal.

The main advantages of the method proposed over maximum likelihood are its ease of implementation, its computational efficiency, and its numerical stability. The simulation-based comparison conveys that: (i) the two-step procedure is numerically more stable than the numerical Monte Carlo quasi-Newton method, even in relatively small models, (ii) the loss in inferential efficiency is negligible when the number of strata is large (>30); moreover, the MLE variance is inflated due to the Monte Carlo variability introduced by numerical integration within each quasi-Newton iteration, (iii) the hybrid procedure is stable and may be a reasonable compromise (from a numerical standpoint) when the number of strata is less than 30, (iv) the reduction in computation times obtained using the two-step procedure is of order 10^{-2}.

In our illustration on habitat selection by pairs of bison, likelihood evaluation had to be reprogrammed in scientific notation and a single evaluation of the log-likelihood using Korobov quasi-Monte Carlo integration with 4000 points took well over 15 minutes in C on a Pentium 4, 3.8 GHz CPU. Moreover, we were unable to get the quasi-Newton iterations to converge even when the estimators obtained with the two-step method were fed in as starting values. In comparison, the entire two-step estimation process was performed in a few seconds.

Some potential generalizations of the method could be explored. For instance, in the prospective models, the observations from a same cluster are assumed independent conditionally on the random effects. Perhaps this could be relaxed by generalizing the multi-period multinomial probit model (Geweke, Keane, and Runkle 1997). Another avenue of interest is to account more explicitly for the spatial sampling process in the inference, as recently suggested by McCullagh (2008). Indeed, he showed that, in cases in which the configuration of covariates is random, the conditional distribution of the response on the sampling units may be different than what is inferred from the marginal model (2.4). We are currently investigating whether the computational approach proposed here is general enough to tackle the type of challenges encountered in the wider class of random-effects models that incorporate sampling design into their specification.

SUPPLEMENTARY MATERIALS

Appendix: Portable document format file containing the notation for the model in its full generality and additional simulation results. (appendix.pdf)

Simulation Code: An example of the code that we used for one batch of simulations from Section 3. (simulation.R)

ML Estimation: C code to compute the maximum likelihood estimate and that is called by the R code in `simulation.R`. (full_ml2.c)

File Description: A text file with a short description and instructions for use of all files listed above. (README.txt)

ACKNOWLEDGMENTS

We thank the editor, an associate editor, and two referees for a set of thorough reviews that have greatly improved the article. We are grateful to Xulin Guo for her work with the GIS, and to Dan Frandsen, Lloyd O'Brodovich, Marie-Eve Fortin, Sabrina Courant, and Karine Dancose for their help with the acquisition of bison data. This research has been supported by the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, Recherche Québec, and Parks Canada Ecological Integrity Innovation Fund.

[Received October 2009. Revised October 2010.]

REFERENCES

Agresti, A., Caffo, B., and Ohman-Strickland, P. (2004), "Examples in Which Misspecification of a Random Effects Distribution Reduces Efficiency, and Possible Remedies," *Computational Statistics & Data Analysis*, 47, 639–653. [772]

Arbogast, P. G., and Lin, D. Y. (2004), "Goodness-of-Fit Methods for Matched Case-Control Studies," *The Canadian Journal of Statistics*, 32, 373–386. [775]

Bhat, C. (2001), "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," *Transportation Research Part B: Methodological*, 35, 677–693. [768,775]

Broyden, C. G., Dennis, J. E., and Moré, J. J. (1973), "On the Local and Superlinear Convergence of Quasi-Newton Methods," *IMA Journal of Applied Mathematics*, 12, 223–245. [772]

Caffo, B., Jank, W., and Jones, G. (2005), "Ascent-Based Monte Carlo Expectation-Maximization," *Journal of the Royal Statistical Society, Ser. B*, 67, 235–251. [772]

Chervoneva, I., Iglewicz, B., and Hyslop, T. (2006), "A General Approach for Two-Stage Analysis of Multilevel Clustered Non-Gaussian Data," *Biometrics*, 62, 752–759. [769,773-775,781]

Craiu, R. V., Duchesne, T., and Fortin, D. (2008), "Inference Methods for the Conditional Logistic Regression Model With Longitudinal Data," *Biometrical Journal*, 50, 97–109. [780]

Davidian, M., and Giltinan, D. M. (1993), "Some Simple Methods for Estimating Intraindividual Variability in Nonlinear Mixed Effects Models," *Biometrics*, 49, 59–73. [773]

Dean, C., and Nielsen, J. (2007), "Generalized Linear Mixed Models: A Review and Some Extensions," *Lifetime Data Analysis*, 13, 497–512. [768]

Fortin, D., Fortin, M.-E., Beyer, H. L., Duchesne, T., Sourant, S., and Dancose, K. (2009), "Group Size Mediated Habitat Selection and Group Fusion–Fission Dynamics of Bison Under Predation Risk," *Ecology*, 90, 2480–2490. [780]

Fortin, D., Fryxell, J., O'Brodovich, L., and Frandsen, D. (2003), "Foraging Ecology of Bison at the Landscape and Plant Community Levels: The Applicability of Energy Maximization Principles," *Oecologia*, 134, 219–227. [780]

Fryxell, J. (1991), "Forage Quality and Aggregation by Large Herbivores," *American Naturalist*, 138, 478–498. [780]

Gail, M. H., Lubin, J. H., and Rubinstein, L. V. (1981), "Likelihood Calculations for Matched Case-Control Studies and Survival Studies With Tied Death Times," *Biometrika*, 68, 703–707. [773]

Geweke, J. F., Keane, M. P., and Runkle, D. E. (1997), "Statistical Inference in the Multinomial Multiperiod Probit Model," *Journal of Econometrics*, 80, 125–165. [782]

Hebblewhite, M., Merrill, E., and McDermid, G. (2008), "A Multi-Scale Test of the Forage Maturation Hypothesis in a Partially Migratory Ungulate Population," *Ecological Monographs*, 78, 141–166. [780]

Hesse, S., Train, K. E., and Polak, J. (2006), "On the Use of a Modified Latin Hypercube Sampling (MLHS) Method in the Estimation of Mixed Logit Model for Vehicle Choice," *Transportation Research Part B: Methodological*, 40, 147–163. [768]

Ibrahim, J. G., and Molenberghs, G. (2009), “Missing Data Methods in Longitudinal Studies: A Review,” *Test*, 18, 1–43. [775]

Jank, W. (2005), “Quasi-Monte Carlo Sampling to Improve the Efficiency of Monte Carlo EM,” *Computational Statistics & Data Analysis*, 48, 685–701. [775]

Korn, E. L., and Whittemore, A. S. (1979), “Methods for Analyzing Panel Studies of Acute Health Effects of Air Pollution,” *Biometrics*, 35, 795–802. [769,772,773]

Lemieux, C. (2009), *Monte Carlo and Quasi-Monte Carlo Sampling*. *Springer Series in Statistics*, New York: Springer. [775]

Lindstrom, M. L., and Bates, D. M. (1988), “Newton–Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measure Data,” *Journal of the American Statistical Association*, 83, 1014–1022. [769]

——— (1990), “Nonlinear Mixed Effects Models for Repeated Measures Data,” *Biometrics*, 46, 673–687. [773]

Litière, S., Alonso, A., and Molenberghs, G. (2007), “Type I and Type II Error Under Random-Effects Misspecification in Generalized Linear Mixed Models,” *Biometrics*, 63, 1038–1044. [772]

Manly, B. F. J., McDonald, L. L., Thomas, D. L., McDonald, T. L., and Erickson, W. P. (2002), *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*, Dordrecht: Kluwer Academic. [780]

McCullagh, P. (2008), “Sampling Bias and Logistic Models” (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 70, 643–677. [782]

McCulloch, C. E. (2003), *Generalized Linear Mixed Models*, Vol. 7, Beachwood, OH: Institute of Mathematical Statistics. [772]

McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008), *Generalized, Linear, and Mixed Models*, 2nd ed., Hoboken, NJ: Wiley. Available at *http://www.loc.gov/catdir/enhancements/fy0827/2008002724-d.html*.

Meng, X.-L., and van Dyk, D. (1998), “Fast EM-Type Implementations for Mixed Effects Models,” *Journal of the Royal Statistical Society, Ser. B*, 60, 559–578. [769,773]

Owen, A. (2003), “Quasi-Monte Carlo Sampling,” in *Monte Carlo Ray Tracing: Siggraph 2003 Course 44*, ed. H. W. Jensen, SIGGRAPH, New York: ACM, pp. 69–88. [775]

Pfeiffer, R. M., Gail, M. H., and Pee, D. (2001), “Inference for Covariates That Accounts for Ascertainment and Random Genetic Effects in Family Studies,” *Biometrika*, 88, 933–948. [768,771,773]

Sándor, Z., and Train, K. E. (2004), “Quasi-Random Simulation of Discrete Choice Models,” *Transportation Research Part B: Methodological*, 38, 313–327. [768]

Simpson, D., Ruppert, D., and Carroll, R. J. (1992), “On One-Step GM Estimates and Stability of Inferences in Linear Regression,” *Journal of the American Statistical Association*, 87, 439–450. [775]

Stiratelli, R., Laird, N., and Ware, J. H. (1984), “Random-Effects Models for Serial Observations With Binary Response,” *Biometrics*, 40, 961–971. [769,772,773]

Train, K. (2003), *Discrete Choice Methods With Simulation*, New York: Cambridge University Press. Available at *http://www.loc.gov/catdir/samples/cam033/2002071479.html*. [768]

Train, K., and Weeks, M. (2005), “Discrete Choice Models in Preference Space and Willlingness-to-Pay Space,” in *Applications of Simulation Methods in Environmental and Resource Economics*, eds. R. Scarpa and A. Alberini, Dordrecht: Springer. [768]

Tuerlinckx, F., Rijmen, F., Verbeke, G., and De Boeck, P. (2006), “Statistical Inference in Generalized Linear Mixed Models: A Review,” *British Journal of Mathematical & Statistical Psychology*, 59, 225–255. [768, 772]