On the equivalence of prospective and retrospective likelihood methods in case-control studies

BY ANA-MARIA STAICU

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.
staicu@stat.ncsu.edu

SUMMARY
We present new approaches to analyzing case-control studies using prospective likelihood methods. In the classical framework, we extend the equality of the profile likelihoods to the Barndorff-Nielsen modified profile likelihoods for prospective and retrospective models. This enables simple and accurate approximate conditional inference for stratified case-control studies of moderate stratum size. In the Bayesian framework, we provide sufficient conditions on priors for the prospective model parameters to yield a prospective marginal posterior density equal to its retrospective counterpart. Our results extend the prospective-retrospective equivalence in the Bayesian paradigm with a more general class of priors than has previously been investigated.

Some key words: Approximate conditional inference; Case-control study; Equivalent Bayesian analysis; Modified profile likelihood; Prospective likelihood; Retrospective likelihood.

1. INTRODUCTION
Case-control studies are important in the analysis of effects of covariates \( X_i \) on a binary disease outcome \( D_i \). They involve a retrospective model for the distribution of covariates given disease status and can accommodate various degrees of stratification by matching on important stratification variables; see Breslow & Day (1980) and Breslow (1996) for background and references. The standard method for analysis of case-control studies is the logistic regression corresponding to the prospective model for the probability of disease given covariates, which ignores the retrospective design. The validity of this approach relies on the classic result of Cornfield (1956), that the prospective and retrospective odds ratios are equivalent.

From a classical point of view, the prospective and retrospective models yield identical maximum likelihood estimates and identical profile likelihood functions for the odds ratios, provided that the underlying distribution of the covariates is unrestricted (Andersen 1970; Prentice & Pyke, 1979). Moreover, Prentice & Pyke (1979) showed that the large sample inference agrees for the two models. Nevertheless, this inference can be seriously biased when the study involves many small- or moderate-sized strata. One alternative that reduces the bias is conditional analysis (Andersen 1970; Breslow, 1982), which eliminates the nuisance parameters by conditioning. The prospective and retrospective conditional analyses for the odds ratio are equivalent (Breslow & Day, 1980) and the conditional logistic regression can be successfully applied to case-control studies of small stratum sizes, but it can be cumbersome to perform when there are many strata of larger size. From a Bayesian viewpoint, the equivalence between prospective and retrospective analyses, studied in terms of the marginal posterior density of the log odds ratio parameter, was established in recent work of Seaman & Richardson (2004) by assuming specific priors for the nuisance parameters involved. Still, the problem of studying the Bayesian equivalence, when other prior distributions are assumed, attracts considerable scientific interest (Mukherjee et al., 2005).
The goal of this note is two-fold. First, we discuss a new approach to analyzing case-control studies via simple approximate methods to the conditional likelihood that are reliable and readily available for moderate stratum sizes, in stratified designs. These accurate approximations are based on a modified profile likelihood (Barndorff-Nielsen & Cox, 1979); see also Davison (1988) and Pierce & Peters (1992) who discussed approximate conditional inference for exponential family models. We show that the Barndorff-Nielsen modified profile likelihood is identical for prospective and retrospective models; for prospective models, the modified profile likelihood can be computed straightforwardly using standard output (Davison, 1988). Our finding is closely related to the result of Wang & Carroll (1999), who show that the maximum likelihood estimator of the odds ratio has the same marginal saddlepoint approximation (Daniels, 1954) to order \(O(n^{-1})\), irrespective of the sampling scheme, where \(n\) is the total sample size.

Second, we provide an explicit answer to the problem: given a retrospective model and a prior for its parameters, find prior distributions for the prospective model parameters that yield identical marginal posterior densities for the log odds ratio parameter as the retrospective analysis. In Bayesian analysis of retrospective models, a Dirichlet distribution for the nuisance parameters is typically assumed (Seaman & Richardson, 2001, 2004). Seaman & Richardson (2004) showed that the resulting marginal posterior density of the log odds ratio parameter is identical to that obtained from prospective likelihood with a uniform prior distribution for the stratum parameter. We extend the equivalence between the prospective and retrospective marginal posterior densities to other prior distributions than had previously been investigated. This approach provides access to a simpler model, which is readily implementable in freely available software such as WinBUGS (Lunn et al., 2000). The findings shown here represent an important motivation to increase the scope for prior sensitivity analysis in case-control studies.

2. APPROXIMATE CONDITIONAL INFEERENCE IN RETROSPECTIVE MODELS

Consider a source population divided into strata determined by a stratification variable \(Z_i\) with levels \(z_1, \ldots, z_K\). Let \(D_i\) be a binary indicator of a disease, \(X_i\) be a vector of exposure covariates that may be related to the disease incidence and \(Z_i\) be the stratification variable for the \(i\)th subject. Under the stratified case-control sampling scheme, the design involves direct sampling of persons with disease, cases, from \(\Pr(X_i \mid D_i = 1, Z_i = z_k)\), and of persons without disease, controls, from \(\Pr(X_i \mid D_i = 0, Z_i = z_k)\) within the same stratum determined by \(z_k\).

Our approach is based on modeling the covariate distribution in the control population; see also Seaman & Richardson (2001, 2004). Denote by \(X = \{x_1, \ldots, x_J\}\), the set of the unique values of the covariates vector \(X\) that are observed in the entire case-control sample, and assume that the maximum likelihood estimator of the stratum-specific covariates distribution in the control population allows positive mass only within this set. For a generic stratum \(k\), such an estimator can be parameterized with respect to the probability masses \(\delta_{k1}, \ldots, \delta_{kJ}\) that it assigns to the values \(x_1, \ldots, x_J\). For convenience, we instead use the one-to-one reparameterization \(\delta_{k1}, \ldots, \delta_{kJ}\), where \(\delta_{k1} = 0\) and the probability of the value \(x_j\) in the control group is equal to \(\exp(\delta_{kj}) / \sum_{l=1}^J \exp(\delta_{kl})\). Under the retrospective sampling scheme, we assume that

\[
\log \frac{\Pr(X_i = x_j \mid D_i = d, Z_i = z_k)}{\Pr(X_i = x_j \mid D_i = d, Z_i = z_k)} = dx_j^T \beta + \delta_{kj} \quad (d = 0, 1),
\]

where \(\beta\) is the log odds ratio parameter of interest which is the same to all strata, while \(\delta_k = (\delta_{k2}, \ldots, \delta_{kJ})^T\) are stratum-dependent nuisance parameters. We assume that the observations are independent across strata and that for each stratum \(k\) determined by \(Z_i = z_k\), there are \(n_{k+1}\) independent observations, comprised of \(n_{kd}\) outcomes of type \(X_i = x_j\) given disease status \(D_i = d\), and \(d = 0, 1\). The likelihood for case-control data is \(L^{\text{ret}}(\beta, \delta) = \prod_{k=1}^K L^{\text{ret}, k}(\beta, \delta_k)\), where

\[
L^{\text{ret}, k}(\beta, \delta_k) = \prod_{d=0}^1 \prod_{j=1}^J \left\{ \frac{\exp \left( dx_j^T \beta + \delta_{kj} \right)}{\sum_{l=1}^J \exp \left( dx_j^T \beta + \delta_{kl} \right)} \right\}^{n_{kd}},
\]

(1)
and involves the nuisance parameters \( \delta_s \). Conditioning on the sufficient statistic associated with the nuisance parameters \( \delta_s \) yields a distribution that does not depend on these parameters. The resulting conditional distribution agrees with the conditional logistic regression corresponding to a prospective analysis of the data, and many statistical packages provide algorithms for computing it. However, the calculations become more intensive as the number and size of strata increase (Breslow & Day, 1980). An alternative approach is to use simpler approximations, such as double saddlepoint approximation (Daniels, 1954). The corresponding likelihood function is equal to an adjusted version of the profile likelihood of the model (Barndorff-Nielsen & Cox, 1979), also known as the modified profile likelihood and discussed in Barndorff-Nielsen (1983), Davison (1988) and others.

For retrospective models, the modified profile likelihood has a simple expression: 
\[
L_{\text{ret}}^{\text{MP}}(\beta) = L_{\text{ret}}(\beta) \times |j_{\beta}(\beta, \hat{\delta}_\beta)|^{1/2},
\]
where the subscript MP refers to modified profile, \( L_{\text{ret}}^{\text{MP}}(\beta) = L_{\text{ret}}(\beta, \hat{\delta}_\beta) \) is the profile likelihood, \( \hat{\delta}_\beta \) is the value of \( \delta \) that maximizes the retrospective likelihood function \( L_{\text{ret}}(\beta, \delta) \) and \( j_{\beta}(\beta, \delta) = -\hat{\delta}_\beta^2 \log L_{\text{ret}}(\beta, \delta)/\partial \delta \partial \delta \) is the submatrix corresponding to the parameter \( \delta \) of the Hessian matrix related to the retrospective likelihood. For a rich model with many covariates, the retrospective model (1) involves a large number of nuisance parameters and calculating the modified profile likelihood can be very difficult. The following result provides an equivalent representation of \( L_{\text{ret}}^{\text{MP}}(\beta) \) that simplifies its calculation.

**Theorem 1.** Define \( \alpha_k = \log [\text{pr}(D = 1 \mid X = x_1, Z = z_k)/\text{pr}(D = 0 \mid X = x_1, Z = z_1)] \) and let \( L_{\text{pro}}^{\text{MP}}(\beta, \alpha) \) be the prospective logistic likelihood that ignores the retrospective design, where \( \alpha = (\alpha_1, \ldots, \alpha_K)^T \). The retrospective modified profile likelihood, \( L_{\text{ret}}^{\text{MP}}(\beta) \), is equal to the prospective modified profile likelihood, \( L_{\text{pro}}^{\text{MP}}(\beta) \), where
\[
L_{\text{pro}}^{\text{MP}}(\beta) = L_{\text{pro}}(\beta, \hat{\alpha}_\beta) \times |j_{\alpha\alpha}(\beta, \hat{\alpha}_\beta)|^{1/2},
\]
and \( j_{\alpha\alpha}(\beta, \alpha) = -\hat{\alpha}_\beta^2 \log L_{\text{pro}}(\beta, \alpha)/\partial \alpha \partial \alpha \). Here, \( \hat{\alpha}_\beta = (\hat{\alpha}_{\beta,1}, \ldots, \hat{\alpha}_{\beta,K})^T \) is defined as the solution of the equations
\[
n_{k1+} = \sum_{j=1}^J n_{k+j} \exp(\alpha_k + x_j^T \beta)/(1 + \exp(\alpha_k + x_j^T \beta)) \quad (k = 1, \ldots, K),
\]
where \( n_{k1+} = \sum_{j=1}^J n_{k+j} \) and \( n_{k+j} = n_{k0j} + n_{k1j} \).

For general models under an arbitrary parameterization, say \( L(\beta, \eta) \), where \( \beta \) is the parameter of interest and \( \eta \) a nuisance parameter, the modified profile likelihood has the complex formula \( L_{\text{MP}}(\beta) = |\hat{\eta}_\beta / \partial \hat{\eta}_\beta|^{-1} |j_{\eta\eta}(\beta, \hat{\eta}_\beta)|^{1/2} L(\beta, \hat{\eta}_\beta) \), using notation similar to that introduced earlier (Barndorff-Nielsen, 1983). However, because both the prospective and retrospective models are formulated as exponential families in the canonical parameter representation, the modified profile likelihood can be simplified to the expressions given above, where \( \eta \) is the vector of \( K \) for prospective models or the vector of \( 3s \) for retrospective models. This simplification is due to the fact that, for full exponential family models in the canonical parameterization, we have \( |\hat{\eta}_\beta / \partial \hat{\eta}_\beta| \propto |j_{\eta\eta}(\beta, \hat{\eta}_\beta)|^{-1} \), where the proportionality refers to a function of data.

The proof of Theorem 1 is sketched in the Appendix. Theorem 1 extends the equivalence of prospective-retrospective profile likelihoods (Birch, 1963) to modified profile likelihoods. This provides theoretical grounds for employing the standard approximate conditional inference for logistic regressions (Davison, 1988) to analyze case-control studies. To illustrate numerically the advantage of the modified profile-based inference over the profile-based inference, we considered a small-sample simulation study. We generated 1000 datasets from a retrospective design with \( X = \{0, 1, 2\} \) and a number of \( K \) strata, each stratum containing two cases and five controls. We found that the coverage probabilities for the nominal 95% confidence interval were 84-3% with the profile loglikelihood and 96-2% with the modified profile loglikelihood in the case \( K = 10 \), and 54-5% and 95-2%, respectively, when \( K = 50 \).

For many large strata, approximate conditional inference provides excellent results. More importantly, for many moderate strata, approximate conditional inference is still very good. In the latter settings, it is fair to state that large sample approximation assumptions are not satisfied and inference may be unreliable. Sartori (2003) shows that the modified profile likelihood inference allows a number of strata \( K = o(n^{1/4}) \), where \( n \) is the total sample size; the analogous condition for the profile likelihood is \( K = o(n^{1/2}) \). Here, the limits in \( o(\cdot) \) are taken as \( nK \to \infty \). From this viewpoint, Theorem 1 addresses the equivalence between the prospective and retrospective analyses for stratified data with many modest strata. However,
Consider, for example, the partition of the log odds ratio parameters modified profile likelihood methods, using the former only for very small strata. An attractive overall approach would be to combine the exact conditional and the modified profile likelihood methods, without any additional cost. Special purpose packages in R are available, for example, cond by Brazzale in a 2000 dissertation from the Ecole Polytechnique Fédérale de Lausanne; while general purpose programs may be also adapted (Davison, 1988). These findings are limited to discrete covariates, and extending them to covariates that are not discrete remains an open problem.

3. Bayesian Analysis Equivalence

3.1. Beyond the Dirichlet priors

We now turn to analyzing case-control studies in the Bayesian framework. This section extends the equivalence of the retrospective and prospective marginal posterior distribution of the log odds ratio established in Seaman & Richardson (2004) for a specific class of priors, to a wider class of priors. For simplicity of presentation, we consider unstratified designs and use throughout the parameterization outlined in the Appendix.

Theorem 2. Let \( L^{\text{ret}}(\beta, \delta^*) \) be the retrospective likelihood (1) and assume a priori the density \( \pi^{\beta,\delta}(\beta, \delta^*) \). Then \( p^{\text{ret}}(\beta \mid y) \) is equal to \( p^{\text{pro}}(\beta \mid y) \propto \pi^{\beta,\alpha}(\beta, \alpha) \times L^{\text{pro}}(\beta, \alpha) \), if

\[
\pi^{\beta,\alpha}(\beta, \alpha) \propto \prod_{j=1}^{J} \left(1 + \exp(\alpha + \beta^T x_j)\right)^{-1} \times K(\beta, \alpha),
\]

and \( K(\beta, \alpha) = \int \prod_{j=1}^{J} \lambda_j^{\ln \delta_j^*} \exp(-\lambda_j) \times p^{(1)}(\beta, \alpha, \lambda) \, d\lambda \). exists is finite and does not depend on data. Here, \( p^{(1)}(\beta, \alpha, \lambda) = g(\mu \sum_{j=1}^{J} \delta_j^* \exp(\alpha + x_j^T \beta)) \times h(\mu) \times \pi^{\beta,\delta}(\beta, \delta^*) \times \left\{ \sum_{j=1}^{J} \delta_j^* \exp(\alpha + x_j^T \beta) \right\} \times \mu^{-1} \) after substituting \( \mu = \sum_{j=1}^{J} \lambda_j \{1 + \exp(\alpha + x_j^T \beta)\}^{-1} = \lambda_j / [\mu \{1 + \exp(\alpha + x_j^T \beta)\}] \) for \( j = 1, \ldots, J \). \( \pi^{\eta}(\cdot) \) and \( h(\cdot) \) are positive functions such that \( \int \pi^{\eta}(\eta) \exp(-\eta \delta^* \sum_{j=1}^{J} x_j^T \beta) \, d\eta < \infty \), and \( \int \pi^{\mu}(\mu) \exp(-\mu \lambda^x) \, d\mu < \infty \) for all positive integers \( x \).

The induced prior density, \( \pi^{\beta,\alpha}(\beta, \alpha) \), implied by Theorem 2 may be improper and one should check that the joint posterior \( \pi(\beta, \alpha \mid y) \) is proper. The result readily extends to stratified studies: it suffices to have separate \( \alpha \) parameters for each stratum and for these to be a priori independent. The main consequence of Theorem 2 is that the marginal posterior density of \( \beta \) can be obtained by fitting a prospective model with an induced prior on the prospective model parameters; this is readily implementable in freely available software such as WinBUGS.

The sufficient condition (2) requires calculation of \( K(\beta, \alpha) \), which may seem cumbersome; often the functions \( g(\cdot) \) and \( h(\cdot) \) can be conveniently chosen to simplify the integration. In the following examples this calculation reduces to a product of \( J \) one-dimensional easy integrations.

Example 1. Consider a retrospective model parameterized in terms of \( \beta \) and \( \delta^* \) and assume the joint prior \( \pi^{\beta,\delta^*}(\beta, \delta^*) \propto p(\beta) \prod_{j=1}^{J} \delta_j^* a_j^{-1} \), for some \( a_j \geq 0 \). For \( g(\eta) = \eta^{-1} \) and \( h(\mu) = \mu \sum_{j=1}^{J} a_j^{-1} \), which satisfy Theorem 2, the expression (2) gives the induced prior \( \pi^{\beta,\alpha}(\beta, \alpha) \propto p(\beta) \prod_{j=1}^{J} \{1 + \exp(\alpha + x_j^T \beta)\}^{-a_j} \). The retrospective marginal posterior density of \( \beta \) with \( \pi^{\beta,\delta^*}(\beta, \delta^*) \) is equal to the prospective marginal posterior density of \( \beta \) with prior \( \pi^{\beta,\alpha}(\beta, \alpha) \). For \( a_j \neq 0 \) for some \( j \), this is an example of retrospective
model parameters \((\beta, \delta^*)\) that are assumed a priori independent, while the prospective model parameters \((\alpha, \beta)\) are taken to be dependent.

**Example 2.** Consider the prior \(\pi^{(\beta, \delta^*)} \propto p(\beta) (\sum_{j=1}^J \delta_j \exp(x_j^T \beta))^{-a} \prod_{j=1}^J \delta_j^{a_j - 1}\) for the retrospective model parameters, where \(a, a_j\) are real numbers \((j = 1, \ldots, J)\) and \(\sum_{j=1}^J a_j > a\). This is a case where the retrospective model parameters \((\beta, \delta^*)\) are assumed a priori to be dependent, and hence it is not covered by Seaman & Richardson’s (2004) class of priors. Take \(g(\eta) = \exp(-\eta)\eta^{a-1}\) and \(h(\mu) = \exp(-\mu)\mu^{-a+\sum_{j=1}^J a_j - 1}\) two functions that satisfy Theorem 2. The retrospective marginal posterior distribution of \(\beta\) is equal to the prospective counterpart with the prior for the prospective model parameters equal to,

\[
\pi^{(\beta, \alpha)}(\beta, \alpha) \propto p(\beta) \exp(a\alpha) \prod_{j=1}^J \{1 + \exp(\alpha + x_j^T \beta)\}^{-a_j}.
\]

In this example, both priors, the joint prior for the retrospective model parameters and the joint prior for the prospective model parameters, are derived from a joint prior for \((\beta, \alpha, \delta^*)\); see the 2007 PhD thesis from the University of Toronto by Staicu. One can show that the prospective prior \(\pi^{(\beta, \alpha)}(\beta, \alpha)\) leads to a proper posterior.

Theorem 2 seems to lead to covariate-dependent priors. Priors with such characteristics are not unknown in the literature; for example Zellner’s \(g\)-priors for Bayesian variable selection (Zellner, 1986) are covariate dependent. It is noteworthy that Theorem 2 is applicable to case-control studies with discrete exposure variables, for which the set of possible values is finite and known. However, this is not a limitation: as one referee pointed out, epidemiologists almost always make their variables discrete, either in terms of tertiles, quartiles or quintiles (Violantia et al., 2008; Tellez-Plaza et al., 2008). The results open many new practical problems on finding which prior to be used, in the case one has multiple choices. Addressing all these problems exceeds the scope of the current note and will be addressed elsewhere.

### 3.2. Lung cancer illustration

We consider an illustrative application inspired by the French case-control study investigating the effect of genotype and smoking on the risk of lung cancer (Seaman & Richardson, 2001; 2004). The study involved two risk factors: the CYP2D6 genotype, a six-level factor, and tobacco smoking coded as low, medium or high tertile of consumption (Seaman & Richardson, 2001).

A prospective model obtained by ignoring the retrospective design is fitted and assuming two choices of priors for the prospective model parameters. The two prior distributions provide identical prospective and retrospective marginal posterior densities for the log odds ratio of disease. The first prior is that of Seaman & Richardson (2004), \(\pi^{(\beta, \alpha)}(\beta, \alpha) \sim p(\beta)\), and corresponds to the Dirichlet prior for the retrospective model parameters. The second prior is (2), for \(a = -5, a_j = 5\ (j = 1, \ldots, 18)\), and corresponds to the retrospective model parameters prior of Example 2. The choices of \(a\) and \(a_j\) ensure that the marginal posterior is proper; other choices may be possible. Both priors use the same marginal prior for \(\beta\), \(p(\beta) \propto 1\). Table 1 summarizes the results obtained using WinBUGS v14 (Lunn et al., 2000). In order to ensure that Monte Carlo standard errors associated with all log odds ratio estimates were less than 0.005, it was necessary to generate a chain of length 50 000 for Seaman & Richardson’s (2004) prior and 100 000 iterations for the prior (2).

The choice of the retrospective model parameters prior affects the inferential results about the log odds ratio of disease: the point estimates and standard deviations of the parameter of interest are expected to be sensitive to the prior selected. Our approach allows the researcher or practitioner to assume the most appropriate prior for the retrospective model parameters and still be able to simplify the analysis by prospective modelling, without any loss of accuracy.
It is sufficient to show the main result for unstratified designs.

Consider \( \{N_{dj} : d = 0, 1; j = 1, \ldots, J \} \), where \( N_{dj} \sim \text{Po}(\gamma_{dj}) \) are independent, \( \log \gamma_{dj} = \mu + d\alpha + \delta_j + dx_j^1\beta \) with \( \delta_1 = 0 \) and \( \{x_1, \ldots, x_J\} \) is the set of covariates in the case-control setting. Let \( \beta \) be the parameter of interest and let \( \alpha, \mu \) and \( \delta = (\delta_1, \delta_2, \ldots, \delta_J)^T \) be the nuisance parameters. Assume that the observed data are identical with \( n_{dj} \) from the case-control scenario, for \( d = 0, 1 \) and \( j = 1, \ldots, J \).

The main idea is to show that the modified profile likelihood for the prospective and the retrospective models is equal to the modified profile likelihood of the multinomial Poisson transformation model introduced by Birch (1963); see also Lang (1996) and Seaman & Richardson (2004). It is sufficient to show the main result for unstratified designs.

Appendix

**Proof of Theorem 1.** The key ingredient is the use of the Poisson likelihood, which corresponds to the multinomial Poisson transformation model introduced by Birch (1963); see also Lang (1996) and Seaman & Richardson (2004). It is sufficient to show the main result for unstratified designs.

Consider \( \{N_{dj} : d = 0, 1; j = 1, \ldots, J \} \), where \( N_{dj} \sim \text{Po}(\gamma_{dj}) \) are independent, \( \log \gamma_{dj} = \mu + d\alpha + \delta_j + dx_j^1\beta \) with \( \delta_1 = 0 \) and \( \{x_1, \ldots, x_J\} \) is the set of covariates in the case-control setting. Let \( \beta \) be the parameter of interest and let \( \alpha, \mu \) and \( \delta = (\delta_1, \delta_2, \ldots, \delta_J)^T \) be the nuisance parameters. Assume that the observed data are identical with \( n_{dj} \) from the case-control scenario, for \( d = 0, 1 \) and \( j = 1, \ldots, J \).

The main idea is to show that the modified profile likelihood for the prospective and the retrospective models is equal to the modified profile likelihood of the Poisson model, \( L_{\text{MP}}^p(\beta) \). To show the equality \( L_{\text{MP}}^p(\beta) = L_{\text{MP}}^r(\beta) \), we take the appropriate orthogonal reparameterization \( (\beta^*, \alpha^*, \lambda^*) \), where \( \lambda = (\lambda_1, \ldots, \lambda_J)^T \) and \( \lambda_j = \mu + \delta_j + \log[1 + \exp(\alpha + x_j^1\beta)] \) for \( j = 1, \ldots, J \). The Poisson likelihood, corresponding to this reparameterization is the product of two distinct components, one which is the prospective likelihood function of \( (\beta, \alpha) \) and one which depends solely on \( \lambda \). For this reason, the parameter components \( \lambda \) and \( (\beta, \alpha) \) are typically called likelihood orthogonal or likelihood independent (Lindsey, 1996).

The proof follows now immediately.

**Proof of Theorem 2.** Denote by \( p^{\text{ret}}(\beta | y) \) the marginal posterior density for \( \beta \) given the data \( y \), corresponding to the retrospective likelihood, \( L_{\text{ret}}^r(\beta, \delta^*) \) and the joint prior density \( \pi^{\delta, \lambda^*}(\beta, \delta^*) \) and let \( g(\cdot) \) and \( h(\cdot) \) be two positive functions satisfying the conditions required by the theorem. By assuming that \( \int \int L^r(\eta, \mu | y)\pi^{\delta, \lambda^*}(\mu) d\eta d\mu < \infty \), where \( L^r(\eta, \mu) = \eta^{n_+} \exp(-\eta) \mu^{n_0} \exp(-\mu) \) and \( n_{0+} = \sum_{j=1}^{J} n_{0j}, n_{1+} = \sum_{j=1}^{J} n_{1j} \) and then using Fubini’s theorem, we obtain

\[
p^{\text{ret}}(\beta | y) \propto \int \int L^r(\beta, \delta^*)L^2(\eta, \mu) \pi^{\delta, \lambda^*}(\beta, \delta^*)g(\eta)h(\mu) d\delta^* d\eta d\mu.
\]

(A1)

It is noteworthy that \( L_{\text{ret}}^r(\beta, \delta^*)L^2(\eta, \mu) \) coincides with the likelihood function of the multinomial Poisson transformation model \( L_{\text{MP}}^r(\beta, \alpha, \mu, \delta^*) \) discussed above, with Poisson mean \( \gamma_{dj} \), where \( \log \gamma_{dj} = \log \mu + d\alpha + \log \delta_j + dx_j^1\beta \) such that \( \sum_{j=1}^{J} \delta_j = 1 \) and reparameterization \( (\beta, \eta, \mu, \delta^*) \), where \( \eta = \mu \sum_{j=1}^{J} \delta_j^* \exp(\alpha + x_j^1\beta) \). Let \( p^r(\beta, \alpha, \lambda) \) be defined as in Theorem 2.

**ACKNOWLEDGEMENT**

I would like to thank N. Reid for very helpful comments and feedback on previous drafts and C. Crainiceanu and R. Samworth for very useful suggestions. I am grateful for valuable comments provided by the editor, associate editor and two anonymous reviewers. This work was supported by the Natural Sciences and Engineering Research Council of Canada.
Under the reparameterization \((\beta, \alpha, \delta^*, \mu)\), one can show that the right hand side of (A1) equals
\[
p_{\text{ret}}(\beta \mid y) \propto \int L^\text{pro}(\beta, \alpha) \prod_{j=1}^{J} \left\{ 1 + \alpha \exp(\beta^T z_j) \right\}^{-1} \left\{ \int L^1(\lambda) \times p^{(1)}(\beta, \alpha, \lambda) \, d\lambda \right\} \, d\alpha.
\]
The proof concludes by choosing the prior density for \((\beta, \alpha)\) to be
\[
\pi_{\beta, \alpha}(\beta, \alpha) \propto \prod_{j=1}^{J} \left\{ 1 + \exp(\alpha + x_j^T \beta) \right\}^{-1} \int L^1(\lambda) \times p^{(1)}(\beta, \alpha, \lambda) \, d\lambda,
\]
provided the integral is finite and independent of \(y\). \(\square\)

REFERENCES


[Received April 2009. Revised June 2010]