Projected partial likelihood and its application to longitudinal data

SUSAN MURPHY AND BING LI

Department of Statistics, Pennsylvania State University,

326 Classroom Building, University Park, PA 16802, U.S.A.

SUMMARY

An estimating equation, which we call the projected partial score, is introduced for longitudinal data analysis. The estimating equation is obtained by projecting the partial likelihood score function onto the vector space spanned by a class of "conditionally linear" estimating equations. We demonstrate that removing certain terms from the projection of the full likelihood score does not alter important inferential properties of the estimating equation, and doing so is advantageous in handling missing data and time-varying covariates. Within a prequential frame of reference it is shown that the estimating equation is optimal among the largest collection of estimating equations determined by the conditional moments. Furthermore, the method possesses similar properties to generalized estimating equations; in particular, the correct conditional variance specification is necessary for efficiency but not for asymptotic consistency and distribution theory.

AMS 1991 subject classifications: 62J12, 62M09, 62P10

Key words: Partial likelihood; Quasi-likelihood; Generalized estimating equations; life history data; longitudinal data

1. Introduction

Longitudinal or life history data typically involve the observation of subjects over a period of time, thereby resulting in dependent observations on the same subject. Much research has been focused on modelling the marginal mean of the response as a function of time and covariates and accommodating the within-subject dependence via specification of a covariance matrix. Alternatively one can use a conditional model which is based on a model for the conditional mean of the response as a function of past responses, and of present and past covariates.

Both approaches can be carried out under a fully parametric assumption, using maximum likelihood. Recently, particularly for the marginal model, much progress has been made in relaxing the parametric assumption to a semiparametric one, making the inference less sensitive to distributional assumptions. Liang & Zeger (1986) and Zeger & Liang (1986) propose the use of a generalized estimating equation, which requires only correct specification of the marginal mean for asymptotic consistency and normality of the estimator. This is attractive since, due to the complexity of biological and epidemiological responses, researchers are often reluctant to assume a parametric distribution for the responses.

One difficulty associated with the marginal model is drop-out, i.e. measurements for subjects may terminate prematurely; for example, a subject may leave the study before it is completed. Liang & Zeger (1986) note that when the drop-out depends on the previous observations, the use of a generalized estimating equation will, in general, result in biased estimators. The term, random drop-out, is used (Diggle & Kenward, 1994) to describe drop-out depending on the past observations. Another difficulty (Zeger, Liang, & Albert, 1988) is that generally only a linear parameterization of the marginal mean will allow statements concerning the effects of changes in the covariate value of the average individual on the average individual's response.

In contrast, the conditional model, combined with the partial likelihood method, has advantage. The first is our ability to delete conditional likelihood factors from the full likelihood so as to focus on processes of interest. This is because a partial likelihood is the product of any subset of conditional likelihoods from the full likelihood, so long as the conditioning events are nested as time progresses (Wong, 1986). In our setting, the marginal or conditional mean response is of primary interest; the drop-out, and covariates are, at best, of secondary interest. By leaving out those factors associated with the distribution of the covariate or of drop-out given the past, we avoid modelling them; yet we still obtain unbiased estimating functions, which behave similarly to derivatives of a full likelihood.

A second advantage is conceptual. To view a longitudinal data set progressively in time also helps us to understand phenomena special to the way the data are collected, that are not understood as transparently through a marginal model. For example, from the point of view of marginal modelling, it is tempting to think that in order to allow the drop-out to depend on the past, it is necessary to specify the full likelihood: see, for example, Diggle & Kenward (1994). However, from a conditional perspective, we realize that the essential property of random drop-out is that it does not alter the conditional distribution of the response given the observed past. Additionally, given a desired parametrization of the marginal mean, one can choose among various conditional mean parameterizations leading to that marginal mean by distributional assumptions. We illustrate this in Section 3.

Since we consider situations in which one is reluctant to describe the distribution fully, it is natural to replace the sequence of parametrized conditional likelihoods, conditioned on nested events, by estimating functions based on the corresponding sequence of conditional moments. In this paper we provide a method for constructing such estimating functions. When applied to longitudinal data, this handles covariates and random drop-out satisfactorily, as does partial likelihood; it avoids distributional assumptions, as does generalized estimating equations. It is obtained by projecting the partial score function onto a collection of Hilbert spaces with inner product specified by conditional moments, conditioned on nested events. We also demonstrate, within a prequential frame of reference (Dawid 1984, 1991), that the estimating function is optimal among the largest collection of estimating functions that can be described by the postulated conditional moments.

Just as quasi-likelihood can be viewed as a generalization of least squares, projected partial likelihood can be viewed as a generalization of least squares for the partial likelihood as employed by Scheike (1994). Our projection of the partial likelihood produces a martingale estimating function (Godambe, 1985). The projection of the full likelihood score, as a martingale, onto a set of Hilbert spaces associated with conditional moments, was studied in Godambe & Heyde (1987), & McLeish & Small (1988, page 99). However, as we shall see in Section 2, removing certain terms from the partial score prior to projection does not change the essential inferential properties, and as illustrated via an example in Section 3, can help us in analyzing longitudinal data.

2. The projected partial likelihood

Suppose that we observe n individuals at time points $j = 1, 2, ...$, and obtain observations of the form

$$Y = (Z_0, Z_1, X_1, Z_2, X_2, Z_3, X_3, ...).$$

The random vector X_j represents the responses from the subjects at time j, and Z_j is the associated covariate matrix. Since there may be drop-out, we may have fewer than

n observations at each time point. In other words, X_j is a vector of dimension at most n, and Z_j is a matrix of dimension $\dim(X_j) \times p$. On each subject only a finite number of observations are recorded; however in general one may not be able to specify a priori the maximum number of observations for a subject. See section 3 for an example. Denote $(Z_0, Z_1, X_1, ..., X_{j-1}, Z_j)$ by U_{j-1}. We shall assume that, conditionally on U_{j-1}, the subjects are independent at time j. In other words, $p(X_j|U_{j-1})$ may be factorized as $\prod_i p(X_{ij}|U_{j-1})$, where X_{ij} stands for the component of X_j corresponding to individual i. We are interested in understanding the effect of the past, and the present covariate, on the response X_j; this is modelled through a regression parameter β. We assume that the conditional expectation, $E_\beta(X_j|U_{j-1})$ depends only on the parameter β.

We consider a partial likelihood obtained by removing from the full likelihood those factors associated with the covariate process $\{Z_i\}$, and those associated with drop-out. This results in

$$\prod_{j \geq 1} p_\beta(X_j|U_{j-1}) = \prod_{j \geq 1} \prod_i p_\beta(X_{ij}|U_{j-1}), \tag{1}$$

where, at each j, only those individuals who remain in the study contribute to $\prod_i p_\beta(X_{ij}|U_{j-1})$. Because of the nested structure of $\{U_j\}$, removing drop-out factors in this way will not cause bias in the resulting partial score function, as long as the subjects' drop-out depends only on the past. This is in contrast to a generalized estimating equation, which is biased under random drop-out. We write the partial score function corresponding to (1) as

$$s(\beta; Y) = \sum_{j \geq 1} (\partial/\partial\beta) \log p_\beta(X_j|U_{j-1}) \equiv \sum_{j \geq 1} s_j(\beta).$$

Now suppose that we are willing to parametrize only the first two conditional moments $\mu_j(\beta) = E_\beta(X_j|U_{j-1})$, and $V_j(\beta) = \mathrm{cov}_\beta(X_j|U_{j-1})$. For fixed j, let $\mathcal{G}_j$ be the class of "conditionally linear" estimating functions of the form $a^T X_j + b$, a and b being functions of the past observations U_{j-1}, and define, for members of $\mathcal{G}_j$, an inner product by

$$\langle g_j, h_j \rangle_j = E_\beta\{g_j h_j | U_{j-1}\}, \quad g_j, h_j \in \mathcal{G}_j. \tag{2}$$

The member of $\mathcal{G}_j$ which is nearest to $s_j(\beta)$ in terms of the inner product (2) is the quasi-score function based on the conditional moments

$$g_j^*(\beta) \equiv \dot{\mu}_j^T(\beta) V_j^{-1}(\beta)\{X_j - \mu_j(\beta)\},$$

where $\dot{\mu}_j(\beta)$ is the gradient matrix of the conditional mean vector $\mu_j(\beta)$: see Wedderburn (1974), McCullagh (1983) & Small & McLeish (1994, page 73). We define the projected partial score function $g^*(\beta, Y)$ to be the estimating function

$$\sum_{j \geq 1} g_j^*(\beta).$$

This construction might at first glance seem heuristic and arbitrary: we simply sum the projections of $\{s_j\}$, which are onto different linear spaces $\{\mathcal{G}_j\}$ associated with different inner products $\{\langle \cdot, \cdot \rangle_j\}$. However, this is entirely natural within a prequential frame of reference. In Dawid (1984, 1991), it is argued that if we are to assess the adequacy of the joint distribution P of a sequence of random variables $\{X_j\}$ based on the realization $\{x_j\}$, all that is relevant is the realization $\{x_j\}$, together with the associated sequence of conditional distributions $\{P_j\}$, $P_j(\cdot)$ being the conditional distribution $P(\cdot | X_1 = x_1, ..., X_{j-1} = x_{j-1})$. Thus we can choose any joint distribution Q whose conditional distributions $Q(\cdot | X_1 = x_1, ..., X_{j-1} = x_{j-1})$ coincide with P_j. A particularly convenient choice of Q is that under which $\{X_j\}$ are independent with marginal distributions $\{P_j\}$.

To specialize to the present context, let $P_{\beta,j}(\cdot) = P_\beta(\cdot | u_{j-1})$, and let Q_β be the joint distribution of $\{X_j\}$ under which the X_j's are independent with marginal distribution $P_{\beta,j}$. We abbreviate $E(\cdot | P_\beta)$ to E_P, or simply E, and $E(\cdot | Q_\beta)$ to E_Q. Notice that we do not assume the knowledge of either P_β or Q_β beyond the moments $\{\mu_j(\beta), V_j(\beta)\}$. Under the distribution Q_β, the maximal space of estimating functions whose first two moments are determined by $\{\mu_j(\beta), V_j(\beta)\}$ is the tensor product space $\otimes_{j\geq 1}\mathcal{G}_j$: see Proposition 2.1 of McLeish & Small (1992). Since, for each element $g \in \otimes_{j\geq 1}\mathcal{G}_j$,

$$E_Q\{(s-g^*)\times g\} = E_Q\left\{\sum_{j\geq 1}(s_j - g_j^*) \times \prod_{j\geq 1} g_j\right\} = \sum_{j\geq 1} E_Q\{(s_j - g_j^*)\times g_j\} = 0,$$

the member g^* of $\otimes_{j\geq 1}\mathcal{G}_j$ is the projection of s onto $\otimes_{j\geq 1}\mathcal{G}_j$ under the distribution Q_β. In other words, if we accept $\{Q_\beta : \beta \in B\}$ as the frame of reference, then g^* is the optimal estimating function (Godambe, 1960) in the largest possible collection of estimating functions whose first two moments are determined by $\{\mu_j(\beta), V_j(\beta)\}$.

That g^* is an optimal estimating function can also be justified under the original model $\{P_\beta : \beta \in B\}$ using the unconditional moments, but this time, among a smaller collection of estimating functions. Let $\mathcal{G}$ be the class of estimating functions with each member defined by $g(\beta; Y) = \sum_j g_j(\beta)$, $g_j(\beta) \in \mathcal{G}_j$. If g is in $\mathcal{G}$, then

$$\begin{aligned} E_P\{(g^*-s)\times g\} &= E_P[\sum_{j\geq 1}\{g_j^*(\beta) - s_j(\beta)\} \times \sum_{l\geq 1} g_l(\beta)] \\ &= \sum_{j\geq 1}\sum_{l\geq 1} E_P[\{g_j^*(\beta) - s_j(\beta)\}\times g_l(\beta)] \equiv \sum_{j\geq 1}\sum_{l\geq 1} A_{jl}. \end{aligned}$$

If $j \neq l$, then A_{jl} vanishes because $E_P[\{g_j^*(\beta) - s_j(\beta)\} \times g_l(\beta) | U_{k-1}] = 0$, where $k = \min\{j, l\}$. If $j = l$, then $E_P[\{g_j^*(\beta) - s_j(\beta)\} \times g_j(\beta) | U_{j-1}] = 0$, because $g_j^*(\beta)$ is by definition the projection of $s_j(\beta)$ in terms of the conditional inner product (2) on to the space $\mathcal{G}_j$ that $g_j(\beta)$ lives in. Hence $s - g^*$ is orthogonal to all members in $\mathcal{G}$; that is, g^* is the projection of s on to $\mathcal{G}$ in terms of the unconditional inner product $E_P(gh)$.

The following proposition summarizes a number of properties of g^* that are similar to those of a likelihood score function. They are either immediate consequences of the preceding discussions or may easily be verified using techniques of the theory of estimating functions

(Small & McLeish, 1994). For simplicity, we assume $\beta \in R^1$, though similar results hold for vector-valued β.

PROPOSITION

(i) For the information identity,

$$E_P\{(g^*)^2\} = -E_P(\partial g^*/\partial\beta), \quad \textit{and} \quad E_Q\{(g^*)^2\} = -E_Q(\partial g^*/\partial\beta).$$

(ii) Optimality in the sense of Godambe holds in that for each g in $\mathcal{G}$ and h in $\otimes_{j\geq 1}\mathcal{G}_j$,

$$\frac{E_P^2(\partial g/\partial\beta)}{E_P(g^2)} \leq \frac{E_P^2(\partial g^*/\partial\beta)}{E_P(g^{*2})}, \quad \textit{and} \quad \frac{E_Q^2(\partial h/\partial\beta)}{E_Q(h^2)} \leq \frac{E_Q^2(\partial g^*/\partial\beta)}{E_Q(g^{*2})}.$$

(iii) The sequence of increments $\{g_j^*\}$ of g^* is a martingale difference sequence.

When one can not, a priori, specify a maximal number of observations per subject, a rigorous justification of the optimality properties described above would utilize the fact that the score functions and estimating functions can be written as integrals with respect to a marked point process. We leave this topic for further research.

In some applications, the covariance V_i may depend on an additional parameter α, say of dimension s. Suppose that, for each fixed β, there is an estimator, $\hat{\alpha}(\beta)$ as in the theory of generalized estimating equations (Liang & Zeger, 1986). The projected partial score, with α replaced by its estimator, is

$$g^*(\beta; Y) \equiv \sum_{j\geq 1} \dot{\mu}_j^T(\beta) V_j^{-1}\{\beta, \hat{\alpha}(\beta)\}\{X_j - \mu_j(\beta)\}. \tag{3}$$

Note that the estimating function, evaluated at the true value of β and any finite value of α, will have expectation zero, even if the form of the conditional variance is misspecified. As noted by Zeger & Liang (1986), this implies that under general conditions, misspecification of the conditional variance will not preclude consistent estimation of β. That is, if $\hat{\alpha}(\beta)$ converges in probability to some constant $\tilde{\alpha}$, then, under regularity conditions, the equation

$$g^*(\beta; Y) = 0 \tag{4}$$

has a consistent sequence of solutions. Furthermore, if $\sqrt{n}\{\hat{\alpha}(\beta) - \tilde{\alpha}\}$ is bounded in probability, $\hat{\alpha}(\beta)$ is differentiable and $\{\partial\hat{\alpha}(\beta)/\partial\beta : n = 1, 2, ...\}$ is bounded in probability, then $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically multivariate Gaussian with zero mean, and variance given by the limit in probability of

$$\left(n^{-1}\sum_{j\geq 1}\dot{\mu}_j^T V_j^{-1}\dot{\mu}_j\right)^{-1}\left[n^{-1}\sum_{j\geq 1}\dot{\mu}_j^T V_j^{-1}\mathrm{cov}\{X_j|U_{j-1}\}V_j^{-1}\dot{\mu}_j\right]\left(n^{-1}\sum_{j\geq 1}\dot{\mu}_j^T V_j^{-1}\dot{\mu}_j\right)^{-1}. \tag{5}$$

Approximately correct specification of the variance results in a lower asymptotic variance of the estimators than an incorrect specification. Note that the asymptotics given above are for the number of subjects increasing without bound and do not necessarily require that the numbers of observations per subject increase. A sketch of the proof is given in the Appendix.

To consistently estimate the variance replace the diagonal elements of $\mathrm{cov}(X_j|U_{j-1})$ by the diagonal elements of $\{X_j - \mu_j(\hat{\beta})\}\{X_j - \mu_j(\hat{\beta})\}^T$, and replace α and β by their estimators. Denote the lth diagonal element of $V_j(\beta, \alpha)$ by $V_{jll}(\beta, \alpha)$. Ordinarily the estimators of α for a given β are moment estimators. For example, one could estimate α by solving

$$\sum_{j\geq 1}\sum_{l\geq 1}\frac{\partial}{\partial\alpha}V_{jll}(\beta, \alpha)[\{X_{jl} - \mu_{jl}(\beta)\}^2 - V_{jll}(\beta, \alpha)] = 0. \tag{6}$$

In the above, it is assumed that both the conditional mean and conditional variance have a finite dimensional parametrization. If the parametrization is infinite dimensional, then the above theorem is not valid. In these cases, $\hat{\alpha}$ may not be $\sqrt{n}$-consistent and care must be exercised. An example of an infinite-dimensional parameterization is given in Huffer & McKeague (1991).

3. Example

This work was initially motivated by the analysis of menstrual data which presents many challenges to the researcher. We focus on a few of the challenges below. First, one usually observes women over a fixed length of time, e.g. one year, and records the lengths of the menstrual cycles. This means that women who have a short mean cycle-length contribute more cycles than women with long mean cycle-length. Since the resulting data set will be overly weighed toward short cycles, data of this type are called length biased. In addition, the last cycle is only partially observed. Two additional characteristics of menstrual data are the presence of time-varying covariates, and a right-skewed cycle-length distribution. There is not, as yet, a biological model justify a parametric distribution for cycle length.

In order to address the length bias, note that if a woman is followed for C days then a cycle is observed only if the sum of the prior cycle lengths is less than C. This observation, combined with a parameterization of the conditional mean, allows us to construct an unbiased estimating function even though the observed cycles are length biased. For simplicity of presentation we will not address issues raised by only a partial observation of the last cycle: see Murphy, Bentley & O'Hanesian (1994), for a thorough discussion.

The data considered here are for 46 women from a longitudinal study by Bentley, Harrigan & Ellison, of the menstrual patterns of Lese women of the Ituri Forest, Zaire. The study follows women through eight months containing three seasons: a prehunger season, a hunger season, and a posthunger season. It is thought that, as women lose weight, their cycles will lengthen.

If equal numbers of cycles had been collected on each of the women, a marginal linear model for the mean and a equi-correlation matrix for the covariance matrix would have been the working model of choice (Harlow & Matanoski, 1991). Since these data are length-biased, a generalized estimating equation based on a marginal model will yield biased estimators of the regression coefficients (Murphy, Bentley & O'Hanesian, 1994). Alternatively we consider a parametrization of the conditional mean of a cycle length given the previous cycles yielding the same parametrization of the marginal mean as above. We parametrize the mean, μ_{ji}, of the ith woman's jth cycle, X_{ji}, given the previous cycles, by

$$\mu_{ji} = Z_{ji}\beta + \frac{\rho}{\rho(j-1)+1-\rho}\left(\sum_{l=1}^{j-1} X_{li} - \sum_{l=1}^{j-1} Z_{li}\beta\right)$$

where Z_{ji} is the associated covariate. In the mixed effects model, ρ is the intra-woman correlation coefficient. In a model assuming only the above form of the conditional mean, ρ functions as an intra-woman correlation coefficient, in that if ρ is zero then the covariance between $X_{ij} - Z_{ij}\beta$ and $X_{il} - Z_{il}\beta$ is zero for $j \neq l$ and the past is no longer helpful in estimating the mean of the next cycle length.

We also use the working model to specify the conditional variance as,

$$V_{ji} = \sigma^2\left(1 + \frac{\rho}{\rho(j-1)+1-\rho}\right).$$

The estimating function is given by,

$$\sum_{j\geq 1}\sum_{i=1}^{n} \dot{\mu}_{ji} V_{ji}^{-1}(X_{ji} - \mu_{ji})\delta_{ji},$$

where δ_{ji} is 1 if $\sum_{l=1}^{j-1} X_{li} < C_i$, 0 otherwise. The presence of δ_{ji} allows for the drop-out of subjects, and, since δ_{ji} is a function of the past, the estimating function is unbiased, assuming the conditional mean is correctly specified.

The additional complication that the last cycle length is not fully observed requires an adjustment to the above method and is described in Murphy, Bentley & O'Hanesian (1994), where a more complete analysis of this data set can be found.

Table 1: Model Results

cycles=176	women=46	$\hat{\sigma}^2$ =17.3
Covariate	*Estimate*	*Stderr*
Intercept	1.17	0.43
locale = far	−1.23	0.98
BMI*(locale=near)	−0.44	0.20
BMI*(locale=far)	0.27	0.65
ρ	0.10	0.08

This analysis was conducted using the SAS procedure IML. The response, cycle length, is centred at 28 days and the body mass index is centred at 21 kilograms/centimeters^2. From Table 1, we see that there is evidence that, as the body mass of women in the nearer locale decreases, their cycle lengths increase. The lack of an effect for women in the further locale may be due to the women being wealthier and not being followed as closely as the women in the nearer locale.

4. Discussion

Even when the inferential questions are best answered by a marginal model, a researcher confronted with random dropout can consider a conditional model which leads to the desired marginal model. An additional important advantage is that estimation via our variant of generalized estimating equations allows one to produce consistent estimators under only conditional moment assumptions, much as generalized estimating equations allow consistent estimation under marginal moment assumptions. When the conditional mean is linear in previous responses, a likelihood analysis is not necessary for consistent estimation of the parameters in the marginal model. Instead an analysis via a projected partial likelihood can be made. This analysis requires only the correct parametrization of the conditional mean.

A full theoretical treatment of the ideas in this paper would utilize the fact that the projected partial likelihood can be expressed as a martingale which is an integral with respect to a marked point process. As in Greenwood & Wefelmeyer (1991), the model we consider is a partially specified marked point process model, since we do not specify the distribution of the covariate or drop-out. However in their comparison of Godambe's finite sample optimality with asymptotic optimality, they take the length of observation time to increase without bound. Our work is based on the number of subjects increasing without bound. Finally the prequential frame of reference is a natural viewpoint from which to consider finite sample optimality of the projected partial likelihood. Perhaps this frame of reference would be useful in constructing an optimality criterion for martingale estimating equations. Careful study is needed in this regard.

5. Appendix

Asymptotic Distribution Theory

Use a Taylor series on the equation $0 = g^*\{\hat\beta, \hat\alpha(\hat\beta)\}$, the assumption that $(\partial/\partial\beta)\hat\alpha(\beta) = O_p(1)$, and the assumption that there exists a finite α_0 for which $\sqrt{n}(\hat\alpha(\beta) - \alpha_0) = O_p(1)$, to obtain

$$0 = g^*\{\beta, \hat\alpha(\beta)\} + \frac{\partial g^*\{\beta, \hat\alpha(\beta)\}}{\partial\beta}(\hat\beta - \beta) + O_p(n|\hat\beta - \beta|^2).$$

Rearranging terms results in

$$0 = g^*(\beta, \alpha_0) + \frac{\partial}{\partial\beta} g^*(\beta, \alpha_0)(\hat\beta - \beta) + O_p(n|\hat\beta - \beta|^2) +$$

$$[g^*\{\beta, \alpha_0\} - g^*\{\beta, \hat\alpha(\beta)\}] +$$

$$\Big[\{\frac{\partial}{\partial\beta}g^*(\beta,\alpha)|_{\alpha=\hat{\alpha}} - \frac{\partial}{\partial\beta}g^*(\beta,\alpha_0)\} + \frac{\partial}{\partial\alpha}g^*(\beta,\alpha)|_{\alpha=\alpha_0}\frac{\partial}{\partial\beta}\hat{\alpha}(\beta)\Big](\hat{\beta}-\beta).$$

The fourth term is $o_p(n|\hat{\alpha}-\alpha_0|)$, the last term is $O_p(n|\hat{\alpha}-\alpha_0| \times |\hat{\beta}-\beta|) + o_p(n|\hat{\beta}-\beta|)$. Therefore

$$\sqrt{n}(\hat{\beta}-\beta) = \Big\{-1/n\frac{\partial}{\partial\beta}g^*(\beta,\alpha_0) + o_p(1)\Big\}^{-1} n^{-1/2}g^*(\beta,\alpha_0).$$

All that is left is to verify that $n^{-1/2}g^*(\beta,\alpha_0)$ has an asymptotic normal distribution, and that $-n^{-1}\frac{\partial}{\partial\beta}g^*(\beta,\alpha_0)$ converges in probability to a positive definite matrix. If g^* is the sum of independent observations, then asymptotic normality can be verified with the help of the Lindeberg-Feller central limit theorem and assumptions on the convergence of appropriate averages. However if the drop-out or appointment scheme induces dependence among subjects then it is useful to consider the data, Y, as a marked point process. Corollary 2 and the example in the unpublished manuscript by Murphy (1993) give central limit theorems for a marked point process.

6. References

Dawid, A.P. (1991). Fisherian inference in likelihood and prequential frames of reference. *J. R. Statist. Soc.* B **53**, 79-109.

Dawid, A.P. (1984). Statistical theory: the prequential approach (with discussion). *J. R. Statist. Soc.* A, **147**, 278-292.

Diggle, P.J., K. & Kenward, M.G. (1994). Informative Drop-out in Longitudinal Data Analysis. *Appl. Statist.* **43**, 49-93.

Godambe, V.P. (1960). An Optimum Property of Regular Maximum Likelihood Estimation. *Ann. Math. Statist.* **31**, 1208-1211.

Godambe, V.P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika.* **72** 419-428.

Godambe, V.P. & Heyde, C.C. (1987). Quasi-likelihood and optimal estimation. *International Statistical Review.* **55** 231-244.

Greenwood, P.E., & W. Wefelmeyer (1991). On optimal estimating functions for partially specified counting process models. In *Estimating Functions,* Ed. V.P. Godambe, pp. 147-160. Oxford: Claredon Press.

Harlow, S.D., & Matanoski, M.G. (1991). The Association between Weight, Physical Activity, and Stress and Variation in the Length of the Menstrual Cycle. *Am. J. Epidemiol.* **133**, 38-49.

Huffer, F.W., & McKeague, I.W. (1991). Weighted Least Squares Estimation for Aalen's Additive Risk Model. *J. Am. Statist. Assoc.* **86**, 114-129.

Liang, K.Y. & Zeger, S.L. (1986). Longitudinal Data Analysis using Generalized Linear Models. *Biometrika* **73**, 13-22.

McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist.* **11**, 59-67.

McLeish, D.L. & Small, C.G. (1988). *The Theory and Applications of Statistical Inference Functions,* Lecture Notes in Statistics 44. New York: Springer.

McLeish, D.L. & Small, C.G. (1992). A projected likelihood function for semiparametric models, *Biometrika* **79**, 93-102.

Murphy, S.A., Bentley, G.R. & O'Hanesian, M.A. (1994). An Analysis for Menstrual Data with Time-Varying Covariates. *to appear in Stat. Medicine.*

Scheike, T.H. (1994). Parametric Regression for Longitudinal Data with Counting Process Measurement Times. *Scand. J. Statist.* **21**, 245-264.

Small, C.G. & McLeish, D.L. (1994). *Hilbert Space Methods in Probability and Statistical*

Inference. New York: John Wiley & Sons.

Wedderburn, R.W.M. (1974). Quasi-likelihood, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439-447.

Wong, W.H. (1986). Theory of Partial Likelihood. *Ann. Statist.* **14**, 88-123.

Zeger, S.L. & Liang, K.Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics* **42**. 121-130.

Zeger, S.L., Liang, K.Y., & Albert, P.S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics* **44**. 1040-1060.

Acknowledgements

Both authors acknowledge the perspicacious comments by the editor, A. P. Dawid, and two anonymous referees. This research has been supported by the Division of Mathematical Sciences at the National Science Foundation.