SSC Annual Meeting, June 2001
Proceedings of the Survey Methods Section, pp. 131-136

EXISTENCE AND CONSISTENCY OF GEE ESTIMATORS - APPLICATION TO DESIGN-BASED INFERENCE

Ioana Şchiopu-Kratina[1]

ABSTRACT

Conditions for existence of roots of weighted generalized estimation equations are given. The design-consistency of the estimators thus obtained is studied.

KEY WORDS: Consistency of estimator; Design-based inference; GEE.

RÉSUMÉ

Nous présentons des conditions d'existence des racines d'équations généralisées pondérées. Nous étudions aussi la convergence de ces estimateurs dans la probabilité induite par le plan de sondage.

MOTS CLÉS : Convergence des estimateurs; GEE; inférence statistique avec des données d'enquêtes complexes.

1. INTRODUCTION

This article presents general results on existence and consistency of estimators obtained as roots of estimating equations (REE). We prove weak consistency of the estimators thus obtained since only convergence in probability is appropriate in the context of design-based inference. However, our main results, Theorems 1-2, actually hold a.s. in the appropriate context. Assumption 2 presented here is considerably weaker than Assumption 2 in Yuan & Jennrich (1998) and consequently our Theorem 2 is stronger than their Theorem 2. We illustrate the improvement in Example 3. In addition, this article contains more general results than those presented in Şchiopu-Kratina (2001). Firstly, the existence theorem has been extended to cover the multivariate case using the proof of the classical inverse function theorem (see also Yuan and Jennrich (1998)). Secondly, the design-based inference is self-contained. Superpopulation parameters are not needed as fixed points in obtaining asymptotic results and no model-based assumption is used. Proofs of all statements contained in this article are available and will be presented elsewhere.

The original purpose of this research was to provide a theoretical foundation for design-based inference with data from longitudinal sample surveys. A starting point was the seminal paper of Liang & Zeger (1986). Their article contains asymptotic results but no proofs of these results are given. Results on design - consistency, asymptotic normality and consistency of the jackknife estimator of the asymptotic variance are given in Şchiopu-Kratina (2001) for a specific design. The existence of a root of the generalized estimating equation (GEE) is proved in the univariate case only ($p = 1$). In the case of estimating equations (EE) that are not gradients there seems to be no trivial extension from the univariate to the multivariate case. The proof of the existence of REE for GEE was first given by Yuan and Jennrich (1998) (see also Chen et al (1999)). It is this result that we generalize and apply to GEE associated with marginal models as in Liang and Zeger (1986) in the context of design - based inference. For the design-based inference, we followed the approach of Binder (1983) in the context of marginal model as in Rao (1998).

In longitudinal data, observations on the same subject are dependent, and this dependence is different from the clustering effect due to the sampling selection. Liang

[1]Ioana Şchiopu-Kratina, Statistics Canada, R.H. Coats Building, 16th floor, Ottawa, Ontario, Canada, K1A 0T6, ioana.schiopu-kratina@statcan.ca.

and Zeger (1986) introduced GEE's, which require only specification of the marginal model mean and variance for each individual. Correlation across time for the same individual is assumed to exist, but it is not specifically modelled. In the special situation when the observations across time are assumed independent for each individual (the working independence assumption), GEE becomes the Independence Estimating Equation (IEE).

2. GEE AND MARGINAL MODELS

We describe briefly the set-up in Liang and Zeger (1986). Consider N individuals observed on d_i occasions (i = 1, ... N). The univariate responses y_{it} and the p - covariates x_{it} are recorded, t = 1, ... d_i, i = 1,... N. We assume that $d_i = d$, i = 1, ... N. Typically, d is small for marginal models. Otherwise, time series techniques may be more appropriate. Liang and Zeger (1986) consider probability densities of the following type: $\rho(y_{it}) = \exp[\{y_{it}\,\theta_{it} - a(\theta_{it}) + b(y_{it})\}\,\varphi]$ with $\theta_{it} = h(\eta_{it})$, $\eta_{it} = x_{it}^T\beta$, where a, b and h are known (differentiable) functions, $\{\theta_{it}\}$, φ are parameters , x_{it}^T is an $1 \times p$ matrix of covariates and β is a $p \times 1$ vector of main parameters, for all t, i≥ 1. Here T stands for transposition of matrices. Note that for random variables with such densities we have:

(1) $\quad E_m[y_{it}] = \mu_{it} = a'(\theta_{it})$, for all t, i≥ 1:

Let $\mu(\eta) = a'(\eta)$, for any η in a space of parameters Θ. The function g is a link function if $g \circ \mu(\theta_{it}) = x_{it}^T\,\beta$, for all t, i≥ 1. If $g = \mu^{-1}$, then g is called the canonical or natural link, the function *h* above can be taken to be the identity and the parametric form of the model is the natural one. With binary response, the logit link function $g(\mu) = \log\{\mu/(1-\mu)\}$ is the natural link associated with the logistic regression model. EE's are formed that mimic log likelihood equations associated with exponential distributions (e.g. normal, binomial, logistic, Poisson). These are quasi - likelihood equations if the original distributions belong to the normal family upon further restrictions, e.g. knowledge on the dispersion parameter φ (Shao 1999, p. 242). The idea is to produce estimators for β which are REE by making only few assumptions on the distribution of the observed data, and then study the properties of these estimators.

When GEE's are used, it is assumed that correlation of observations y_{it} across time for the same individual is the same for all individuals, and is represented by a matrix $R(\alpha)$, with α a "nuisance parameter". More precisely, let $U_i(\beta, \alpha, \varphi) = D_i^T V_i^{-1} S_i$, $V_i = 1/\varphi\,[A_i^{1/2} R(\alpha) A_i^{1/2}]$, $D_i = A_i \Delta_i X_i$, $S_i = Y_i - a'(\theta_i)$, A_i = diag $a''(\theta_{it})$ in R^d and $\Delta_i = \mathrm{diag}\,[d\,\theta_{it} / d\,\eta_{it}]$, which could be taken to be the identity matrix I_d, for all i ≥ 1. Notice that the covariates are contained in D_i and that A_i as well as S_i (through a') contain the main parameter β, i ≥ 1. The GEE, or equation (7) of Liang and Zeger (1986), is:

(2) $$\sum_{i=1}^{N} U_i(\beta, \tilde{\alpha}(\beta), \varphi(s, \beta)) = 0$$

Equation (2) above is called a pseudo-likelihood equation in Shao (1999), p. 315. Note that it consists of p scalar equations. In equation (2) $\tilde{\alpha}$ and $\varphi(s, \beta)$ are estimates of nuisance parameters that are obtained from the sample and generally contain β. When the solution to (2) exists and is unique , i.e. when β is defined implicitly by (2), it is denoted by $\hat{\beta}_G$ in Liang and Zeger (1986). Note that this approach is different from the one presented in Section 5 of Rao (1998). It is important to note that (2) contains only β as unknown parameter and that, due to the estimation of the nuisance parameters, the left hand side of (2) is, in general, a nonlinear function of the sample observations.

3. GEE AND DESIGN-BASED INFERENCE

The conditions for model based inference are often not met by the data collected according to the survey design, even if the finite population is large. Design - based inference, introduced by Binder (1983), offers a solution, as it allows for the use of modelling techniques in the context of survey randomization. When the observations across time are assumed independent for each individual (the working independence assumption), equation (2) becomes IEE. In this case $R(\alpha) = I_d$ and there is no need to estimate nuisance parameters in (2). This is the situation discussed, in a design randomization context, by Binder (1983). In the context of IEE and survey randomization $\hat{\beta}_G$ becomes the "census" parameter defined in Binder (1983).

Definition 1 In the case of the GEE (2), the census parameter β_N is defined as the solution (when it exists and is unambiguously defined) of equation (3) below:

(3) $$\psi_N(\beta) = \sum_{k=1}^{N} U_k(\beta, \alpha_N(\beta), \varphi_N(\beta)) = 0$$

■

The example below illustrates the calculation of $\beta_N = \hat{\beta}_G$ from an IEE. Notice that the presence of the time dimension is accounted for by the increase in the number of data points (from N to 2×N in this case).

Example 1 (the linear model, p = 1). Assume that the individual observations are independent, identically distributed (i.i.d.) and that they follow a normal distribution. Take $\varphi = 1$ and d = 2 occasions. We have $R(\alpha) = I_2$ (case IEE). Assume that x_{it}, β are scalars, i,t ≥ 1 and that *h* is the identity.

$$\rho(y_{it}) = \exp - \frac{(y_{it} - \theta_{it})^2}{2} \Rightarrow a(\theta_{it}) = \frac{\theta_{it}^2}{2},\quad b(y_{it}) = -\frac{y_{it}^2}{2}$$

$$E[y_{it}] = \theta_{it} = \frac{da}{d\theta_{it}};\quad \frac{d^2a}{d\theta_{it}^2} = 1,\ \theta_{it} = x_{it}\beta,\ i,t \geq 1.$$

For i, t ≥ 1:

$$\frac{d\log\rho(y_{it})}{d\beta} = y_{it}x_{it} - x_{it}^2\beta$$

Now $a'(\theta_{it}) = \theta_{it} = x_{it}\beta$, i, t ≥ 1 and (2) is:

$$(4)\qquad \sum_{i=1}^{N}\sum_{t=1}^{2} x_{it}y_{it} - \sum_{i=1}^{N}\sum_{t=1}^{2} x_{it}^2\beta = 0 \Rightarrow$$

$$\beta_N = \hat{\beta}_G = \frac{\sum_{i=1}^{N}\sum_{t=1}^{2} x_{it}y_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{2} x_{it}^2}$$

■

We will define next a sample - based estimator $\hat{\beta}_N$, which will serve to make design based inference on the census parameter β_N. In conjunction with the GEE (2), we define, for $\beta \in \Theta$:

$$(5)\quad \hat{\psi}_N(\beta) = \psi_N(s,\beta) = \sum_{k\in s} w_k U_k(\beta, \hat{\alpha}_N(\beta), \varphi_N(s,\beta))$$

In (5) $\hat{\alpha}_N(\beta)$ and $\varphi_N(s,\beta)$ are sample based estimators of the census parameters α_N and φ_N, respectively, with w_k the design weights divided by N. Notice that in case of with - replacement sampling, s is an ordered sample, i.e. the same p.s.u.'s may appear several times in the sample *s* (Särndal et al 1992, p.72)

Definition 2 The REE estimator $\hat{\beta}_N$ of the census parameter β_N in (4) is defined as a solution to $\psi_N(s,\beta) = 0$, with $\psi_N(s,\beta)$ as in (5) above. ■

Example 2 (the linear model, p = 1). Consider the simpler situation of an IEE presented in Example 1. The census parameter is β_N in (4). A design based estimator $\hat{\beta}_N$ is a solution to $\hat{\psi}_N(\beta) = \psi_N(s,\beta) = 0$, where:

$$\psi_N(s,\beta) = \sum_{k\in s} w_k \sum_{t=1}^{2} x_{kt}(y_{kt} - x_{kt}\beta)$$

This estimator can be found explicitly as the EE above has the unique solution:

$$(6)\qquad \hat{\beta}_N = \frac{\sum_{k\in s}\sum_{t=1}^{2} w_k x_{kt} y_{kt}}{\sum_{k\in s}\sum_{t=1}^{2} w_k x_{kt}^2}$$

Note that in (6) the normalized weights can be replaced by the original design weights. ■

4. THE MAIN RESULTS

In what follows we consider quite general EE indexed by a parameter $\theta \in \mathbb{R}^p$.

Assumption 1. $[\hat{\psi}_N(\theta_N)] \to 0$ in p_N , where θ_N is a nonrandom vector such that $\psi_N(\theta_N) = 0$ and p_N is the design probability. ■

Comment 1. Assumption 1 is closely related to design consistency. It holds if $\hat{\psi}_N(\theta) - \psi_N(\theta) \to 0$ in p_N and uniformly in θ. ■

Examples 1 - 2 We take all covariates equal to 1 in (4) and (6). The census parameter is $\beta_N = \theta_N = \bar{Y}_N$ and Assumption 1 becomes: $N^{-1}\sum_{i\in s} w_i y_i - \bar{Y}_N \to 0$ in p_N , which is design-consistency of the sample mean. ■

Assumption 2: Let $\hat{D}_N(\theta)$ be the pxp matrix of partial derivatives of $N\hat{\psi}_N(\theta)$, which are assumed to be continuous in a neighbourhood of θ_N , N = 1,.... For large N, we assume that $A_N = \hat{D}_N(\theta_N)$ is nonsingular. As in Lemma 2 of Yuan & Jennrich (1998), or the proof of the inverse function theorem in Rudin (1964 , p.193), we define L_N such that $2L_N\|A_N^{-1}\|_o = 1$, where the subscript *o* stands for the operator norm of the matrix (see Rudin (1964), p. 185). By the continuity assumption:

(i) $\exists\, r_N \ni \|\theta - \theta_N\| \leq r_N \Rightarrow \|\hat{D}_N(\theta) - A_N\|_e \leq L_N$, for each large N. Here the subscript *e* stands for the Euclidian norm of the matrix. We assume that,

$\forall\, \eta > 0$, $\exists\, l_\eta > 0$, $n_0(\eta, \delta) = n_0$, n_0 a sample size, $N_0 > n_0$ such that, for all n > n_0, N > max{ n , N_0 }, we have :

(ii) $\qquad p_N(r_N \times L_N / N > l_\eta) > 1 - \eta$ ■

Example 2 Recall from (5) $\hat{\psi}_N(\beta) = \sum_{i \in s} w_i u_i(\beta)$, where $u_i(\beta) = x_i(y_i - x_i\beta)$, i = 1,...N.

Then $\hat{D}_N(\beta) = -\sum_{i \in s} w_i x_i^2 = A_N$. We have:

$L_N = \frac{|A_N|}{2} = 2^{-1} \sum_{i \in s} w_i x_i^2$. Because $\hat{D}_N$ does not depend on the parameter, we may take $r_N = N$, $L_N \to \infty$ in p_N so $r_N \times L_N / N \to \infty$ in p_N and (ii) holds. ■

Comment 2 Note that, on the r.h.s. of '$\Rightarrow$' in (i) we can divide by N both terms of the inequality. ■

Let us denote by $B_r(\theta_0)$ the ball $\{ \theta : \| \theta - \theta_0 \| < r \}$. Let $B_N = B_{r_N}(\theta_N)$, N =1,....

Theorem 1 (existence of REE, p > 1). Assume that Assumptions 1 and 2 hold. Then there exist estimators $\hat{\theta}_N$ such that: $\forall\, \eta > 0$, $\exists N_0 > n_0, n_0 = n_0(\eta)$ a sample size, such that for all n $> n_0$, N > max$\{ n, N_0 \}$, we have :

$$p_N(\hat{\theta}_N \in B_N, \ \hat{\psi}_N(\hat{\theta}_N) = 0) \geq 1 - \eta$$

■

Theorem 2 (existence and consistency of REE). Assumption 1 holds. We replace (ii) in Assumption 2 by two conditions, which together imply (ii): the functions $\hat{D}_N / N$ are equi(in N)continuous functions of θ at θ_N and, with the same notation as in (ii) and with $\bar{L}_N = L_N / N$, we have the weaker condition:

(iii) $\qquad p_N(\bar{L}_N > l_\eta) > 1 - \eta$

There exist then estimators $\hat{\theta}_N$ such that, $\forall\, \eta, \delta > 0$, $\exists N_0 > n_0, n_0 = n_0(\eta, \delta)$ a sample size and for all n $> n_0$, N > max$\{ n, N_0 \}$, we have :

$$p_N(\,|\,\hat{\theta}_N - \theta_N\,| \leq \delta, \ \hat{\psi}_N(\hat{\theta}_N) = 0) \geq 1 - \eta$$

This result follows from Theorem 1 above. Theorem 1 relies on the proof of the inverse function theorem (see Rudin (1964) p. 193 or Lemma 2 of Yuan & Jennrich (1998)). ■

Comment 3 We note that the functions $\hat{\psi}_N$ are invertible on B_N, N = 1,.... ■

5. APPLICATIONS

The first application of Theorem 2 is to the linear model. Assumption 2 of Yuan & Jennrich (1998) would require that $\lim \sum_{i \in s} w_i x_i^2 / N$ exist and be > 0 in p_N as $n(s)$, $N \to \infty$, with n(s) the sample size. This is stronger than (ii) in our Assumption 2, which reduces in Example 2 to:

(7) $\qquad \sum_{i \in s} w_i x_i^2 \to \infty$ in p_N as $n(s), N \to \infty$

Examples 1 - 2 If, in addition to (7), we have $\sum_{i \in s} w_i x_i y_i - \sum_{i=1}^{N} x_i y_i \to 0$ and $\sum_{i \in s} w_i x_i^2 - \sum_{i=1}^{N} x_i^2 \to 0$ as $n(s), N \to \infty$ in p_N, where n(s) is the sample size, then $\hat{\beta}_N - \beta_N \to 0$ in p_N. Note that, under (7), Assumption 1 becomes $\hat{\beta}_N - \beta_N \to 0$ in p_N. ■

The next example is the example discussed in section 5 of Yuan & Jennrich (1998). An application of our Theorem 2 weakens considerably the conditions on the sequence of covariates. Indeed, for Assumption 2 of Yuan & Jennrich to hold, the empirical distributions F_N of $\{ x_i, i = 1, \ldots N \}$ must converge uniformly to a distribution F. We only require that the sequence of covariates be bounded from above and also bounded away from 0.

Example 3 (GLM) The components of the EE are: $u_i(\theta) = x_i(y_i - \mu_i(\theta))$, $\mu_i(\theta) = e^{-x_i\theta}$, $x_i \geq 0, \theta > 0$, i = 1 , . . . Assume that limsup $x_i \leq M$ and liminf $x_i = l > 0$. Then estimators which are REE's exist and are consistent. ■

In order to do statistical inference for GEE, we must find a sample based estimator of $V_k = V_k(\alpha, \beta)$ (see Rao (1998)), and replace it in $U_k(\alpha, \beta, \varphi) = D_k^T V_k^{-1} S_k$, k = 1,... N. This corresponds to the case when $R(\alpha)$ is completely unspecified in Example 5 of Liang and Zeger (1986). In this instance there is no need to estimate the overdispersion parameter φ. To estimate $V_k(\beta) = A_k^{1/2} C_N(\alpha, \beta) A_k^{1/2}$, k ≥ 1 for fixed values of the parameters, we estimate the common correlation structure across time, denoted here $C_N(\alpha,\beta)$, by $\sum_{k\in s} w_k A_k^{-1/2}(\beta) S_k(\beta) S_k^T(\beta) A_k^{-1/2}(\beta)$. The entries of this matrix are:

$$\hat{c}_N^{ij}(\beta) = \sum_{k\in s} w_k [a''(\eta_{ki}(\beta)) a''(\eta_{kj}(\beta))]^{-1/2} s_{ki}(\beta) s_{kj}(\beta)$$

where $s_{ki}(\beta) = y_{ki} - \mu_{ki}(\beta)$, k = 1,... M, i , j = 1,...d, $\beta \in \Theta$. Let $\hat{g}_N^{ij}(\beta)$, i, j, = 1, ...d, $\beta \in \Theta$, be the entries of $\hat{C}_N^{-1}(\beta)$, which is assumed to exist. Then $\hat{V}_k^{-1}(\beta)$ has entries $\hat{g}_N^{ij}(\beta)[a''(\eta_{ki}(\beta)) a''(\eta_{kj}(\beta))]^{-1/2}$, i, j = 1,...d. We substitute in GEE (5): $U_k(\beta, \hat{V}_k(\beta)) = D_k^T \hat{V}_k^{-1}(\beta) S_k$, $S_k = Y_k - a'(\theta_k)$, for any $k \geq 1$ and obtain:

$$\psi_N(s,\beta) = \sum_{i,j=1,..d} \hat{g}_N^{ij}(\beta)\ \psi_N^{ij}(s,\beta),$$

$$\psi_N^{ij}(s,\beta) = \sum_{k\in s} w_k [\frac{a''(\eta_{ki}(\beta))}{a''(\eta_{kj}(\beta))}]^{1/2} x_{ki} s_{kj}(\beta) \quad i, j = 1, \dots d.$$

Therefore, the GEE in (5) can be written as a finite sum of terms with each of these terms equal to a product of two estimators. Furthermore, each $\psi_N^{ij}(s,\beta)$, i, j = 1,...d, is a sum of random variables for which the conditions for consistency in Theorem 2 can easily be applied.

Theorem 3 We assume that, for large N and all β around β_N, the matrices $\hat{C}_N(\beta)$ are nonsingular. We assume further that the hypotheses of Theorem 2 hold for $\hat{\psi}_N^{ij}$, i, j = 1,... d. The following conditions are assumed to hold in probability and for all i, j = 1,...d. :

(iv) $\frac{\partial \hat{g}_N^{ij}}{\partial \beta_k}(\beta)$ are equi and uniformly (in β) bounded around β_N.

(v) $\hat{g}_N^{ij}(\beta_N)$ are equibounded from above.

(vi) $\frac{\partial \hat{\psi}_N^{ij}}{\partial \beta_k}(\beta)$ are equibounded from above.

Then the conclusions of Theorem 2 hold , i.e. we have existence and consistency of the estimators obtained as REE in (5). ■

Comment 4 The assumptions of Theorem 3 are not difficult to verify. Since all functions involved are averages, it suffices to verify these assumptions on the components. For example, conditions on $\hat{g}_N^{ij}(\beta)$ (if well defined) can be expressed as conditions on $\hat{c}_N^{ij}(\beta)$, which reduce to conditions on the functions a'' and $s_{ki}(\beta)$, j = 1,..d, k = 1, ... As a minimum requirement in the nonlinear case, one must assume equiboundedness of the covariates (see (2.3) and Remark 1 of Chen et al (1999)). ■

REFERENCES

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, 279-292.

Binder , D.A. and Patak, Z. (1994). Use of estimating functions for estimation from complex surveys. *JASA*, **89**, 1035- 1043.

Chen, K., Hu, I., Ying, Z. (1999). Strong consistency of maximum qusi-likelihood estimators in generalized linear models with fixed and addaptive designs. *The Annals of Statistics*, **27**, 1155-1163.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics*, **13**, 342 - 368.

Fuller W.A. (1975). Regression analysis for sample surveys. *Sankhya* (Series C), **37**, 117-132.

Hájek, J. (1960). Limiting distributions in simple random sampling from finite populations. *Publ. Math. Inst. Hungarian Acad. Sci.*, **5**, 361 - 374.

Krewski, D. and Rao, J.N.K. (1981). Inference from stratified samples: Properties of linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*, **9**, 1010-1019.

Liang, K-Y and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

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

Rao, J.N.K. (1998). Marginal models for repeated observations: inference with survey data. *Proceedings of the ASA*, 1998.

Rudin,W. (1964). *Principles of Mathematical Analysis* (2nd. ed.). New York: Mc Graw-Hill.

Rubin-Bleuer, S. (1998). Sabbatical Research Report, Statistics Canada.

Rubin-Bleuer, S. (2000). Some issues in the analysis of complex survey data. Statistics Canada technical report, BSMD - 20 - 001 E.

Särndal, C-E, Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling.* New York: Springer-Verlag.

Şchiopu-Kratina, I. (2001). Marginal models for longitudinal data analysis using complex survey data. Technical Report Series of the Laboratory for Research in Statistics and Probability. No. 351 - May 2001.

Serfling, R.J. (1980). *Approximation Tems of Mathematical Statistics*. New York: Wiley.

Shao, J. (1999). *Mathematical Statistics*. New York: Springer Verlag.

Wu, C.F.J. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics* **9**, 501-513.

Yuan, K-H & Jennrich, R.I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis,* **65**, 245 - 260.

Yung, W. (1996). Contributions to poststratification in stratified multi-stage samples. Ph. D. thesis, Carleton University, Ottawa, Canada