

Weighted Estimation in Multilevel Ordinal Models to Allow for Informativeness of the Sampling Design

Stima Pesata nei Modelli Multilivello Ordinali con Disegno Campionario Informativo

Leonardo Grilli
Dipartimento di Statistica “G. Parenti”
Universita' degli Studi di Firenze
grilli@ds.unifi.it

Monica Pratesi¹
Dip. di Matem., Stat., Inform. e Appl.
Universita' degli Studi di Bergamo
pratesi@unibg.it

Riassunto: I modelli multilivello per variabili ordinali sono spesso utilizzati per l'analisi di dati raccolti con disegni campionari complessi a più stadi. Tuttavia, se il disegno è informativo, nel senso che le probabilità di inclusione dipendono dalla variabile di risposta anche condizionatamente alle variabili esplicative, allora gli usuali stimatori di massima verosimiglianza sono distorti. In questo lavoro proponiamo una procedura di stima pesata per i modelli multilivello ordinali che elimina la distorsione generata dall'informatività del disegno campionario: in particolare, i reciproci delle probabilità di inclusione ai vari stadi del campionamento sono usati per pesare il logaritmo della verosimiglianza. Un primo studio di simulazione, descritto di seguito, ha consentito di valutare l'efficacia della soluzione proposta nel caso particolare di una variabile di risposta dicotomica.

Keywords: Multilevel ordinal model, Multistage sampling, Pseudo Maximum Likelihood.

1. Weighted estimators for multilevel ordinal models

In a recent article Pfeiffermann et al. (1998) propose probability-weighting procedures of first and second level units that adjust for the effect of an informative design on the estimation of two-level models with continuous response variable. To ease the comparison with their results it is useful to write the ordinal model in terms of a latent linear model endowed with a set of thresholds. Suppose that for the i -th elementary unit ($i = 1, 2, \dots, N_j$) of the j -th cluster ($j = 1, 2, \dots, M$) the observed ordinal response variable Y_{ij} , with $k = 1, 2, \dots, K$ levels, is generated, through a set of thresholds, by a latent continuous variable $\tilde{Y}_{ij}$ following a variance component model $\tilde{Y}_{ij} = \beta' \mathbf{x}_{ij} + \omega u_j + \varepsilon_{ij}$, where $\mathbf{x}_{ij}$ is a covariate vector and β is the corresponding vector of slopes; the random variables ε_{ij} and u_j are the disturbances, respectively at the first (elementary) and second (cluster) level; and ω^2 is the second level variance component. For the disturbances we make the standard assumptions, i.e. a) the ε_{ij} 's are iid with zero mean and unknown variance σ^2 ; b) the u_j 's are Gaussian iid with zero mean and unit variance; c) the ε_{ij} 's and u_j 's are mutually independent.

The observed ordinal variable Y is linked to the latent one $\tilde{Y}$ through the following relationship: $\{Y_{ij} = k\}$ if and only if $\{\gamma_{k-1} < \tilde{Y}_{ij} \leq \gamma_k\}$, where the thresholds satisfy $-\infty = \gamma_0 \leq \gamma_1 \leq \dots \leq \gamma_{K-1} \leq \gamma_K = +\infty$. Note that, conditional on u_j ,

¹Financial support from MIUR research project “Statistical education and evaluation: instruments, methods and new technologies” is gratefully acknowledged.

$$P(\tilde{Y}_{ij} \leq \gamma_k | u_j) = P(\varepsilon_{ij} \leq \gamma_k - [\boldsymbol{\beta}' \mathbf{x}_{ij} + \omega u_j] | u_j) \quad (1)$$

$$= F\left(\frac{\gamma_k}{\sigma} - \left[\frac{1}{\sigma} \boldsymbol{\beta}' \mathbf{x}_{ij} + \frac{\omega}{\sigma} u_j\right]\right), \quad (2)$$

where $F(\cdot)$ is the distribution function of the standardized first level error term ε_{ij}/σ (note that $F(\cdot)$ is also the inverse of the link function of the ordinal model). All the model parameters are defined in terms of the unknown σ , so only the ratios of the model parameters to the standard deviation of the first level error term are identifiable; in the following we use the notation ψ_σ to indicate that the latent model parameter ψ is in σ units, i.e. $\psi_\sigma \equiv \psi/\sigma$.

Now let $\boldsymbol{\theta}$ denote the vector of all estimable parameters, which include $\boldsymbol{\beta}_\sigma, \omega_\sigma$ and $K-2$ thresholds $\{\gamma_\sigma, k: k=2, \dots, K-1\}$ ($\gamma_\sigma, 1$ is fixed to zero to insure identifiability). The conditional likelihood for subject i of cluster j , $L_{ij}(\boldsymbol{\theta} | u)$, is $\prod_{k=1}^K [P(Y_{ij} = k | u_j)]^{d_{ijk}}$, where d_{ijk} is the indicator function of the event $\{Y_{ij} = k\}$. Then, denoting with ϕ the standard Gaussian density function, the overall marginal log-likelihood is

$$\log L(\boldsymbol{\theta}) = \sum_{j=1}^M \log \int_{-\infty}^{+\infty} \left[\exp \left\{ \sum_{i=1}^{N_j} \log L_{ij}(\boldsymbol{\theta} | u) \right\} \right] \phi(u) du. \quad (3)$$

Now suppose that the whole population of M clusters (level 2 units) with N_j elementary units (level 1 units) per cluster is not observed; instead the following two-stage sampling scheme is used: a) first stage: m clusters are selected with inclusion probabilities π_j ($j = 1, \dots, M$); b) second stage: n_j elementary units are selected within the j -th selected cluster with probabilities $\pi_{i|j}$ ($i = 1, \dots, N_j$). The unconditional sample inclusion probabilities are then $\pi_{ij} = \pi_{i|j}\pi_j$.

When the sampling mechanism is informative, i.e. the π_j and/or the $\pi_{i|j}$ depend on the model disturbances and hence on the response variable, the maximum likelihood estimator of the parameters of the ordinal variance component model may be seriously biased. A standard solution to this problem is provided by the Pseudo Maximum Likelihood (PML) approach of Skinner (1989), which requires a design-consistent estimate of the population log-likelihood (3). Applying the Horvitz-Thompson principle, i.e. replacing each sum over the level 2 population units j by a sample sum weighted by $w_j \equiv 1/\pi_j$ and each sum over the level 1 units i by a sample sum weighted by $w_{i|j} \equiv 1/\pi_{i|j}$ we obtain:

$$\log \hat{L}(\boldsymbol{\theta}) = \sum_{j=1}^m w_j \log \int_{-\infty}^{+\infty} \left[\exp \left\{ \sum_{i=1}^{n_j} w_{i|j} \log L_{ij}(\boldsymbol{\theta} | u) \right\} \right] \phi(u) du. \quad (4)$$

Under mild conditions, the solution $\hat{\boldsymbol{\theta}}_{PML}$ derived from $\frac{\partial}{\partial \boldsymbol{\theta}} \log \hat{L}(\boldsymbol{\theta}) = \mathbf{0}$ is design-consistent for the finite-population maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ which, in turn, is model-consistent for the super-population parameter $\boldsymbol{\theta}$: therefore $\hat{\boldsymbol{\theta}}_{PML}$ is a consistent estimator of $\boldsymbol{\theta}$ with respect to the mixed design-model distribution (Skinner (1989)).

2. Simulation study

The simulation study was limited to the simplest case of the model defined in Section 1, with only two categories for the response variable (i.e. $K = 2$) and no explanatory

variables; moreover, for the first-level disturbances a Gaussian distribution was assumed, so the fitted model is in fact a random intercept probit binary model. The experiment has been designed following the example of Pfeiffermann et al. (1998). The values of the binary response variable Y_{ij} (1000 replicates for each sampling design) were generated using a two-stage scheme.

At the first stage, finite-population values Y_{ij} ($j = 1, \dots, M$; $i = 1, \dots, N_j$) were obtained by generating a value from the super-population latent model $\tilde{Y}_{ij} = \beta + u_j + v_{ij}$, with $u_j \sim N(0, \omega^2)$ and $v_{ij} \sim N(0, \sigma^2)$, and then putting $Y_{ij} = 0$ if $\tilde{Y}_{ij} \leq 0$ or $Y_{ij} = 1$ if $\tilde{Y}_{ij} > 0$. The latent model parameter values employed in the simulation are $\beta = 0$, $\omega^2 = 0.2$ and $\sigma^2 = 0.5$, so that the parameters estimable from the binary model are $\beta_\sigma \equiv \beta/\sigma = 0$ and $\omega_\sigma \equiv \omega/\sigma = 0.632$. The hierarchical structure of the population comprises $M = 300$ clusters, while the cluster sizes N_j were determined by $N_j = 75 \exp(\tilde{u}_j)$, with $\tilde{u}_j$ generated from $N(0, \omega^2)$, truncated below by -1.5ω and above by 1.5ω . As a result, in our population N_j lies in the range $[38, 147]$ with mean around 80.

At the second stage, given the finite-population values, we adopted one of the following sampling schemes: *a) Informative at both levels*: first, 35 clusters were selected with probability proportional to a 'measure of size' X_j , i.e. $\pi_j = mX_j / \sum_{j=1}^M X_j$; the measure X_j was determined in the same way as N_j but with $\tilde{u}_j$ replaced by u_j , the random effect at level 2. The elementary units in the j -th sampled cluster were then partitioned into two strata according to whether $v_{ij} > 0$ or $v_{ij} \leq 0$ and simple random samples of sizes $0.25n_j$ and $0.75n_j$ were selected from the respective strata. The sizes n_j were either fixed ($n_j = 38$, $n_j = 9$) or proportional to N_j ($n_j = 0.4N_j$, $n_j = 0.1N_j$). *b) Informative only at level 2*: the scheme is the same as the previous one, except that simple random sampling was employed for the selection of level 1 units within each sampled cluster. *c) Non-informative*: the scheme is the same as the previous one, except that the size measure X_j was set equal to N_j .

The maximization of the weighted log-likelihood (4) has been carried out by the NLMIXED procedure in SAS (SAS (1999)), using a dual quasi-Newton algorithm in conjunction with adaptive Gaussian quadrature. Though the NLMIXED procedure does not include an option for PML estimation, it is still possible to insert the weights in the likelihood, using different tricks for level 1 and level 2 weights. Note however that the conventional estimator of the covariance matrix of the estimators, which is used by SAS, is not appropriate for the PML method, so we used a bootstrap technique for finite populations (though the results are not reported here, we note that the bootstrap estimator performed very well in large samples, but some bias was still present in small samples).

Table 1 reports the results of the simulations for the fixed size designs with $n_j = 38$ and $n_j = 9$ (for the corresponding proportional allocation designs the results are similar, though they tend to be less favourable). For each sampling design the behavior of the point estimators of the intercept β_σ and the second level standard deviation ω_σ is summarized by the mean and standard deviation of their Monte Carlo sampling distribution. The point estimators under study are the standard maximum likelihood unweighted estimator and the weighted estimator obtained by maximizing expression (4) with scaled level 1 weights $w_{i|j}^{scaled} = w_{i|j}/\bar{w}_j$, where $\bar{w}_j = (\sum_{i=1}^{n_j} w_{i|j})/n_j$. Scaling is expected to reduce the small sample bias, discussed in Pfeiffermann et al. (1998); indeed the advantages of scaling are confirmed by our simulations (the results are not shown here).

First note that if the weights are introduced when there is no need to adjust for the effect of the design, we face a slight increase in the variability of the estimators, which is

Table 1: Simulation means and standard deviations (in parenthesis) of point estimators of the intercept (true value 0) and of the second level standard deviation (true value 0.632); number of replicates is 1000

Sampling design	β_σ		ω_σ	
	Unweighted estimator	Weighted estimator	Unweighted estimator	Weighted estimator
Inform. both levels $n_j = 38$	-0.120 (0.212)	0.015 (0.188)	0.671 (0.106)	0.604 (0.128)
Inform. both levels $n_j = 9$	-0.214 (0.204)	0.000 (0.185)	0.644 (0.145)	0.536 (0.222)
Inform. only level 2 $n_j = 38$	0.281 (0.169)	0.017 (0.169)	0.595 (0.100)	0.601 (0.111)
Inform. only level 2 $n_j = 9$	0.274 (0.187)	0.009 (0.196)	0.547 (0.121)	0.563 (0.133)
Non-inform. $n_j = 38$	0.000 (0.108)	0.001 (0.115)	0.611 (0.086)	0.617 (0.091)
Non-inform. $n_j = 9$	-0.007 (0.108)	-0.010 (0.117)	0.561 (0.105)	0.575 (0.111)

more pronounced in small samples.

When the sampling design is informative the unweighted estimator of β_σ becomes biased and less precise. The conclusions are the same for both types of informative design, though the bias has a different sign. The weighted based adjustment is effective in removing the bias in the estimation of β_σ , sometimes reducing also the sampling variance. As regards ω_σ , given the bias which is also present in the standard case of non-informative design, the behavior of the weighted estimator is satisfactory in nearly all situations, with the exception of the small samples when the design is informative at both levels.

In conclusion, the results of our simulation study on the random intercept probit binary model confirm the findings of Pfeiffermann et al. (1998) on the random intercept linear model: probability-weighted estimators are good for the intercept, while some relevant bias remains in the estimation of the variance component when the sample is small. A further simulation study is needed in order to assess the properties of the probability-weighted estimators when the response variable is truly ordinal, i.e. with three or more categories.

References

- Pfeiffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H. and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models, *Journal of the Royal Statistical Society B* **161**, 23–40.
- SAS (1999). *SAS/STAT User's Guide Version 8*, SAS Institute Inc., Cary.
- Skinner, C. J. (1989). Domain means, regression and multivariate analysis, in: *Analysis of complex surveys* (C.J. Skinner, D. Holt and T.M.F. Smith eds) pp. 59–87.