Computer Methods and Programs in Biomedicine 71 (2003) 181–187

Computer Methods and Programs in Biomedicine

www.elsevier.com/locate/cmpb

Fixing the nonconvergence bug in logistic regression with SPLUS and SAS

Georg Heinze *, Meinhard Ploner

Department of Medical Computer Sciences, University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria

Received 5 June 2001; received in revised form 7 May 2002; accepted 21 May 2002

Abstract

When analyzing clinical data with binary outcomes, the parameter estimates and consequently the odds ratio estimates of a logistic model sometimes do not converge to finite values. This phenomenon is due to special conditions in a data set and known as 'separation'. Statistical software packages for logistic regression using the maximum likelihood method cannot appropriately deal with this problem. A new procedure to solve the problem has been proposed by Heinze and Schemper (Stat. Med. 21 (2002) pp. 2409–3419). It has been shown that unlike the standard maximum likelihood method, this method always leads to finite parameter estimates. We developed a SAS macro and an SPLUS library to make this method available from within one of these widely used statistical software packages. Our programs are also capable of performing interval estimation based on profile penalized log likelihood (PPL) and of plotting the PPL function as was suggested by Heinze and Schemper (Stat. Med. 21 (2002) pp. 2409–3419).

Keywords: Monotone likelihood; Nonexistence of parameter estimates; Penalized likelihood; Separation

1. Introduction

For analyzing clinical studies with binary outcomes, the logistic regression model [1,2] is often used. The straightforward interpretation of the estimated parameters as log odds ratios favored its popularity in medical research, and the capability of allowing models with more than one covariate enables estimation of odds ratios that are adjusted for other covariates [2]. Parameter estimation is usually based on maximization of the (log) likelihood function (maximum likelihood method) via an iteratively weighted least-squares algorithm [3]. However, it is also known that there are certain situations particularly occurring in samples with a high number of parameters relative to sample size where finite maximum likelihood parameter estimates do not exist. In those cases the likelihood converges to a finite value while at least one parameter estimate diverges to $\pm\infty$ [4]. This phenomenon is due to special conditions in a

* Corresponding author. Tel.: +43-1-40400-6684; fax: +43-1-40400-6687
E-mail address: georg.heinze@akh-wien.ac.at (G. Heinze).

0169-2607/02/$ - see front matter
PII: S0169-2607(02)00088-3

data set and known as 'separation'. The simplest example of separation arises in the analysis of a binary outcome and a single binary covariate if the resulting 2×2 table has one zero cell count. Generally, the probability of occurrence of separation is too high to be negligible [5,6].

Some statistical software packages for logistic regression may warn the user in case of separation that the parameter estimates do not have converged [7]. Others simply base convergence of the model fitting algorithm on the deviance or the log likelihood and will not detect separation [8]. In both cases, the resulting odds ratio estimates are based on the last iteration carried out. For covariates causing separation the resulting estimates are completely arbitrary and thus extremely inaccurate [5] and misleading. Although exact logistic regression [9] which has been implemented in the new SAS version 8.1 [10] can provide finite and accurate estimates in some situations, it cannot generally be used as a tool to cope with separation [5].

In this paper, we present SPLUS [8] and SAS [11] programs to solve the separation problem. In our programs parameter estimation is based on the penalized maximum likelihood approach originally developed by Firth [12] and suggested for the logistic regression model by Heinze and Schemper [5]. This approach provides an ideal solution to the problem of separation. It has been shown that parameter estimates from this approach are always finite and have lower small sample bias than maximum likelihood estimates. Because of asymmetric shapes of the profile penalized likelihood in case of separation, Heinze and Schemper [5] recommend the construction of confidence intervals based on profile penalized likelihood instead of using the simpler Wald method. Our programs can be used to perform both ways of interval estimation and to compare them graphically by plotting the profile penalized log likelihood (PPL) function as has been suggested [5].

In Section 2 we describe the algorithms used by the SPLUS library logistf and by the SAS macro FL. Section 3 gives an overview of the application of our programs. Finally, in Section 4 we compare results obtained from the penalized maximum likelihood approach with those from standard analysis by means of a worked example and describe the availability of the program.

2. Computing estimates and confidence limits

A logistic regression model is given by $Pr(y_i = 1|\boldsymbol{x}_i) = \pi_i = 1/\{1 + \exp(-\boldsymbol{x}_i\beta)\}$ where $(y_i, \boldsymbol{x}_i)$, $y_i \in \{0, 1\}$, $i = 1, \ldots, n$, denotes a sample of n observations of the outcome variable y and the $1 \times k$ covariate vector $\boldsymbol{x}$. Usually, $\boldsymbol{x}_{i1} = 1$ denotes the constant. Maximum likelihood estimates of the regression parameters β_r, $r = 1, \ldots, k$, are usually obtained by solving the k score equations $\partial \log L/\partial\beta_r \equiv U(\beta_r) \equiv \Sigma_{i=1}^n (y_i - \pi_i)\boldsymbol{x}_{ir} = 0$, $r = 1, \ldots, k$, where L is the likelihood function. However, in small samples these estimates may be seriously biased. In order to remove the bias of order $O(n^{-1})$ from the parameter estimates Firth [12] suggested to maximize the profile penalized log likelihood $\log L(\beta)^* = \log L(\beta) + 1/2 \log|I(\beta)|$. We recommend the original contribution of the Firth [12] to the interested and mathematically inclined reader. An additional advantage of the resulting penalized maximum likelihood or Firth-type estimates is their finiteness in case of separation [5].

In our programs we based estimation of the Firth-type logistic regression (FL-type) parameter estimates on a Newton–Raphson algorithm. Parameter estimates at iteration $s+1$ are thus obtained by applying $\hat{\beta}^{(s+1)} = \hat{\beta}^{(s)} + I(\hat{\beta}^{(s)})^{-1} U(\hat{\beta}^{(s)})^*$, where $U(\hat{\beta}_r)^* = \Sigma_{i=1}^n [y_i - \hat{\pi}_i + h_i(1/2 - \hat{\pi}_i)]\boldsymbol{x}_{ir}$ denotes the modified score function, and the h_i's are the ith diagonal elements of the 'hat' matrix $\mathrm{H} = \hat{W}^{1/2}X(X^T\hat{W}X)^{-1}X^T\hat{W}^{1/2}$ with $\hat{W} = \mathrm{diag}\{\hat{\pi}_i(1 - \hat{\pi}_i)\}$. If the profile penalized log likelihood evaluated at $\hat{\beta}^{(s+1)}$ is less than that evaluated at $\hat{\beta}^{(s)}$ then $\hat{\beta}^{(s+1)}$ is recomputed by halving the step width of that iteration. Furthermore the maximum step width is restricted to a prespecified value. By this scheme numerical problems during estimation should be avoided. $U(\hat{\beta}_r)$, H and $\hat{W}$ are updated at each iteration. Convergence is declared if $\Sigma_{r=1}^k |\hat{\beta}_r^{(s+1)} - \hat{\beta}_r^{(s)}| < \varepsilon$ with ε denoting any user-suppliable value of the convergence criterion. Unless any offset values have been specified by the user, starting values $\hat{\beta}_r^{(0)}$ of parameters β_r,

$r = 2, \ldots, k$ are taken as zero. The starting value of the intercept parameter is set to $\hat{\beta}_1^{(0)} = \log(\bar{p}/1 - \bar{p}) - \eta$ where $\bar{p} = n^{-1}\Sigma_{i=1}^{n} y_i$, $\bar{\eta} = n^{-1}\Sigma_{i=1}^{n} \eta_i$, $\eta_i = \Sigma_{r-2}^{k} \boldsymbol{x}_{ir}\hat{\beta}_r^{(0)}$. This setting proved satisfactory in the analysis of various data sets.

Standard errors $\hat{\sigma}_r$, $r = 1, \ldots, k$; of the regression parameters are estimated by taking the square roots of the diagonal elements of $I(\hat{\beta})^{-1} = (X'\hat{W}X)^{-1}$ $A(1-\alpha) \times 100\%$ confidence interval for parameter β_r is then defined as $[\hat{\beta}_r + \Phi_{\alpha/2}\hat{\sigma}_r,\ \hat{\beta}_r + \Phi_{1-\alpha/2}\hat{\sigma}_r]$, with Φ_α denoting the α-quantile of the standard normal distribution function. The adequacy of Wald confidence intervals for parameter estimates should be verified by plotting the profile penalized likelihood function. While a symmetric shape of this function allows use of Wald intervals, an asymmetric shape demands profile penalized likelihood intervals [5]. These intervals are based on the inversion of a penalized likelihood ratio test. For testing the hypothesis of $\gamma = \gamma_0$, let the likelihood ratio statistic: $\mathrm{LR} = 2[\log L(\hat{\gamma}, \hat{\delta})^* - \log L(\gamma_0, \hat{\delta}_{\gamma_0})^*]$, where $(\hat{\gamma}, \hat{\delta})$ is the joint penalized maximum likelihood estimate of $\beta = (\gamma, \delta)$, and $\hat{\delta}_{\gamma_0}$ is the penalized maximum likelihood estimate of δ when $\gamma = \gamma_0$. The profile penalized likelihood confidence interval are those values of γ_0 for which LR does not exceed the $(1-\alpha)$100th percentile of the χ_1^2-distribution. Computation of profile penalized likelihood confidence intervals for parameters follows the algorithm of Venzon and Moolgavkar [13] which, for penalized maximum likelihood estimation, has been described by Heinze and Ploner [14]. In our programs, the profile penalized likelihood method is the default for obtaining interval estimates. However, computation of profile penalized likelihood confidence intervals may be time consuming since an iterative algorithm has to be repeated for the lower and for the upper confidence limits of each of the k parameters, resulting in $2k$ maximizations. Since in some situations, e.g. in bootstrap applications, the user may not be interested in interval estimation, one can request computation of confidence intervals and P-values to be based on the Wald method instead of the profile penalized likelihood method.

3. Program description

Both the SPLUS library logistf and the SAS macro FL apply the penalized maximum likelihood approach outlined above. Several simple options allow the user to specify the logistic regression model for which parameter estimates should be obtained. While the complete User's Guide to those two programs can be found in Technical Reports [15,6] here we can only give a brief summary of the parameters that can be set by the user.

3.1. SPLUS library logistf

The SPLUS library logistf includes the functions `logistf`, `print.logistf`, `summary.logistf`, `logistftest`, `print.logistftest`, and `logistfplot`. The main function logistf follows the structure of standard SPLUS functions such as `lm` or `glm`, requiring a data frame and a formula for the model specification. The associated generic functions `summary.logistf` and `print.logistf` are used to show detailed and brief results of `logistf`, respectively. The most important arguments of the function `logistf` are the following:

- formula: an SPLUS formula object with the binary response variable (responses and non-responses coded as 1 or T, and 0 or F, respectively) on the left of the operator and the model terms on the right. The model terms may include contrasts, interactions, nested effects, splines and all other SPLUS features (e.g. `formula = response ~ X1 + X2 + X3 + X4`);
- data: an SPLUS dataframe containing the relevant variables;
- pl: specifies if confidence intervals should be based on the PPL method (`pl = T`, the default) or on the Wald method (`pl = F`);
- alpha: the significance level (1 minus the confidence level, 0.05 as default).

The arguments maxit, maxstep, maxhs, epsilon, and beta0 control the Newton–Raphson algorithm (maximum number of iterations, maximum step width per iteration, maximum number of step-halvings per iteration, convergence criterion and starting values for the parameter estimates, respectively).

The function `logistftest` performs a penalized likelihood ratio test for a set of variables (e.g. a set of $k-1$ dummy variables that are related to a k-level factor). In addition to the arguments of the function `logistf` the test function includes:

- test: specifies the factors which will be tested (e.g. `test = ~X1 + X2`);
- values: null hypothesis values in a vector, default = 0 (e.g. `values = c(2, 0)` will test the null hypothesis H_0: $\beta_{X1} = 2$, $\beta_{X2} = 0$).

To plot the PPL function the function `logistfplot` can be used. It allows the specification of the number of and distances between the points plotted.

3.2. SAS macro FL

The most important macro options of FL are the following:

- DATA: SAS data set which contains the survival data.
- Y: response variable, coded in 0 (nonresponse) and 1 (response).
- VARLIST: list of covariates to be included in the model.
- ODDS: specifies if in addition to parameter estimates odds ratio estimates and associated confidence intervals should be given in the output.
- PL: specifies if confidence intervals should be based on the profile PPL (the default) or on the Wald method.
- PROFILE: requests plots of the profile PPL for the covariates specified in this option.
- TEST: requests a penalized likelihood ratio test for a set of variables (e.g. a set of $k-1$ dummy variables that are related to a k-level factor).

Further macro options control the Newton–Raphson algorithm used for estimation of the FL-type parameter estimates (maximum number of iterations, maximum step length per iteration, maximum number of step-halvings per iteration, convergence criterion for the parameter estimates) and the names of SAS data sets containing the output of the macro. In particular, these data sets comprise a table of parameter estimates, standard errors, confidence intervals and P-values (OUTTAB), a summary of the number of responses and nonresponses, the number of iterations needed and the P-value for the global penalized likelihood ratio test (OUTMOD), the estimate and its variance–covariance matrix (OUTEST), and a data set containing the predicted response probabilities $\hat{\pi}_i$ and the diagonal elements of the hat matrix h_i for each observation (OUT). Part of the printed output for a typical macro call can be seen from Fig. 1. It is possible to request separate analyzes for observations in groups defined by a grouping variable (option BY) and to supress all printed output (option PRINT). Thus, FL can also be efficiently applied in simulation studies or bootstrap applications, which makes it compatible to virtually all SAS procedures.

4. Example and availability

Use of logistf and FL is exemplified by means of the analysis of an epidemiological data set [16] that can be downloaded from the WWW location http://www.cytel.com/examples/sex.dat. Purpose of this case-control study was to evaluate the effects of condom use, lubricated condom use, spermicide use, oral contraceptivce use, diaphragm use and age on risk of acquiring first urinary tract infection. The data set, 'sex' contains data on these variables (CONDOM, LUBRI, SPERM, ORAL, DIAPHRAG, AGE) for 130 infected cases (CASE = 1) and 109 healthy controls (CASE = 0). If a logistic regression model is fit to this data set using PROC LOGISTIC of SAS/STAT [11], separation is detected and a message is printed warning the user that the 'maximum likelihood estimate may not exist'. Fig. 2 shows the (shortened) output of PROC LOGISTIC. In particular, the odds ratio estimate for DIAPHRAG is given as '>999.999' with a 95% Wald confidence interval of [<0.001, >999.999]. There is no warning for nonconvergence of the parameter estimates from SPLUS [8] function glm, resulting

(Part of the output omitted)

Model fitting information

Iterations	Penalized log likelihood	Number of responses	Number of nonresponses	Number of observations
8	-132.539	130	109	239

FL estimates, profile penalized likelihood confidence limits and penalized likelihood ratio tests

Variable	Parameter estimate	Standard Error	Lower 95% c.l.	Upper 95% c.l.	Pr > Chi-Square
INTERCEP	-0.98573	0.60032	-2.17886	0.16260	0.0928
AGE	1.10598	0.42366	0.30743	1.97379	0.0061
ORAL	-0.06882	0.44379	-0.94143	0.78920	0.8752
CONDOM	2.26887	0.54842	1.27302	3.43543	<.0001
LUBRI	-2.11141	0.54308	-3.26086	-1.11773	<.0001
SPERM	-0.78832	0.41737	-1.60809	0.01519	0.0545
DIAPHRAG	3.09601	1.67501	0.77457	8.03029	0.0050

FL odds ratio estimates, profile penalized likelihood confidence limits and penalized likelihood ratio tests

Variable	Odds ratio	Lower 95% c.l.	Upper 95% c.l.	Pr > Chi-Square
AGE	3.0222	1.35992	7.20	0.0061
ORAL	0.9335	0.39007	2.20	0.8752
CONDOM	9.6685	3.57163	31.04	<.0001
LUBRI	0.1211	0.03836	0.33	<.0001
SPERM	0.4546	0.20027	1.02	0.0545
DIAPHRAG	22.1095	2.16965	3072.64	0.0050

Fig. 1. Analysis of epidemiological data set: output of SAS macro FL.

in parameter estimates that are based on that iteration at which the deviance converged. The odds ratio point and interval estimates for DIAPHRAG obtained by SPLUS are thus depending on the convergence criterion specified, yielding an effect of DIAPHRAG which is marginally significant on the 6% level if the convergence criterion is 0.001 (the default) and a totally insignificant yet larger effect if it is set to 0.00001 (see Table 1).

Estimation via SPLUS function logistf or SAS macro FL, however, arrives at plausible odds ratio point and interval estimates for DIAPHRAG of 22.1 and [2.17, 3072.64], respectively, meaning that the odds of contracting an infection increase by a factor of 22.1 with diaphragm use, and that this finding is statistically significant at the 1%-level (see Fig. 1 for the SAS output, SPLUS output is similar). The corresponding SPLUS and SAS calls are simply logistf(formula = case ~ age + oral +

```
                    (Part of the output omitted)

                      Model Convergence Status

           Quasi-complete separation of data points detected.

WARNING: The maximum likelihood estimate may not exist.

WARNING: The LOGISTIC procedure continues in spite of the above warning.
         Results shown are based on the last maximum likelihood iteration.
         Validity of the model fit is questionable.


                       The LOGISTIC Procedure
WARNING: The validity of the model fit is questionable.

               Analysis of Maximum Likelihood Estimates

                                     Standard
      Parameter    DF    Estimate       Error    Chi-Square    Pr > ChiSq

      Intercept     1     -1.0360      0.6106        2.8791        0.0897
      age           1      1.1644      0.4345        7.1810        0.0074
      oral          1     -0.0736      0.4497        0.0268        0.8700
      condom        1      2.4059      0.5695       17.8454        <.0001
      lubri         1     -2.2462      0.5643       15.8427        <.0001
      sperm         1     -0.8201      0.4216        3.7843        0.0517
      diaphrag      1     14.4663       424.1        0.0012        0.9728


                        Odds Ratio Estimates

                          Point          95% Wald
            Effect     Estimate      Confidence Limits

            age           3.204      1.367       7.508
            oral          0.929      0.385       2.243
            condom       11.089      3.632      33.858
            lubri         0.106      0.035       0.320
            sperm         0.440      0.193       1.006
            diaphrag   >999.999     <0.001    >999.999
```

Fig. 2. Analysis of epidemiological data set: output of PROC LOGISTIC.

Table 1
Dependence of odds ratio estimate for variable DIAPHRAG on deviance convergence criterion in SPLUS function glm

Criterion	Estimate	95%-Confidence interval		P-value
0.001	220.02	0.8068	6.5×10^5	0.0594
0.0001	1726.25	0.0004	6.8×10^9	0.3358
0.00001	34 773.80	0.0000	1.1×10^{34}	0.7628

condom + lubri + sperm + diaphrag, data = sex) and %fl(data = sex, y = case, varlist = age oral condom lubri sperm diaphrag, odds = l), respectively.

The complete User's Guide for the SPLUS library logistf and the SAS macro FL, along with installation notes can be found in Technical Reports 2/2001 [15] and 10/1999 [6] which are available along with the programs at the WWW location http://www.akh-wien.ac.at/imc/biometrie/fl.

References

[1] D.R. Cox, Analysis of Binary Data, Methuen, London, 1970, pp. 26–105.
[2] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Wiley, New York, NY, 1989, pp. 25–37.
[3] P. McCullagh, J.A. Nelder, Generalized Linear Models, 2nd ed., Chapman & Hall, London, 1989, pp. 40–43.
[4] A. Albert, J.A. Anderson, On the existence of maximum likelihood estimates in logistic regression models, Biometrika 71 (1984) 1–10.
[5] G. Heinze, M. Schemper, A solution for the problem of separation in logistic regression, to appear in Statistics in Medicine 21 (2002) 2409–2419.
[6] G. Heinze, Technical Report 10: the Application of Firth's Procedure to Cox and Logistic Regression, Department of Medical Computer Sciences, Section of Clinical Biometrics, Vienna University, Vienna, 1999.
[7] SAS/STAT User's Guide, Version 8, SAS Institute Inc., Cary, NC, 1999.
[8] SPLUS 4.0, MathSoft Inc., Cambridge, MA, 1997.
[9] K.F. Hirji, A.A. Tsiatis, C.R. Mehta, Median unbiased estimation for binary data, The American Statistician 43 (1989) 7–11.
[10] R.E. Derr, Performing exact logistic regression with the SAS system, SAS Paper P254-25, (2000) 1–10.
[11] SAS Language Reference, Version 8, SAS Institute Inc., Cary, NC, 1999.
[12] D. Firth, Bias reduction of maximum likelihood estimates, Biometrika 80 (1993) 27–38.
[13] D.J. Venvon, S.H. Moolgavkar, A method for computing profile-likelihood based confidence intervals, Applied Statistics 37 (1988) 87–94.
[14] G. Heinze, M. Ploner, SAS and SPLUS programs to perform Cox regression without convergence problems, Computer Methods and Programs in Biomedicine 67 (2002) 217–223.
[15] M. Ploner, Technical Report 2: An S-PLUS Library to Perform Logistic Regression without Convergence Problems, Department of Medical Computer Sciences, Section of Clinical Biometrics, Vienna University, Vienna, 2001.
[16] B. Foxman, I. Marsh, B. Gillespie, N. Rubin, J.S. Koopman, S. Spear, Condom use and first-time urinary tract infection, Epidemiology 8 (1997) 637–641.