Paper AS14

Penalized Likelihood Logistic Regression for Sparse Data Using Data Priors

Martin Dörr, Clinipace, Marburg, Germany

ABSTRACT

Penalization of the likelihood is a probate means to stabilize the estimate when fitting logistical regression models with sparse data. The default Firth method, available in PROC LOGISTIC, shows good results in bias and MSE properties for the parameters. However, if one is interested in several marginalized endpoints, a switch to PROC NLMIXED is necessary. Here the Firth method cannot be implemented. A suitable alternative are logF(1,1) data priors. This presentation will introduce a logistic regression on sparse data with supporting data priors which demonstrate the custom PROC NLMIXED code for modeling.

KEYWORDS

logistic regression, sparse data, rare events, data priors, PROC NLMIXED

INTRODUCTION

If a logistic regression model has to be fit and the underlying data consists of sparse data, rare events or covariables show a high degree of collinearity, fit results will drift to extreme estimates with a large variability. Frequentists solve this issue by adding a penalization term *A(β)* to the log-likelihood.

$$\log L^{\text{pen}}(\beta) = \log L(\beta) + A(\beta)$$

These constraint terms describe a model parameter statistic which exerts more penalization on the basic log-likelihood the greater the included parameters become. In consequence, the coefficients are reduced (shrunk) towards zero, thus showing reasonable estimates (Cole et al., 2014).

Bayesians do not need to think of any extensions of their models. Their parameter priors *p(ϑ)* already extend the sampling distribution (likelihood) *p(X | ϑ)*, which in combination result into the posterior distribution *p(ϑ | X)*.

$$p(\theta \mid X) \propto p(X \mid \theta)\, p(\theta)$$

It is shown that some priors and penalization terms are corresponding to each other. Table 1 depicts such pairs.

The Firth method is commonly accepted for its good results in bias and MSE properties in binary models (Firth, 1993). Its big advantage is the easy use in a four-field table setting with (quasi-) complete separation: one has just to add 0.5 to each cell before fitting the desired odds ratio or relative risk estimate. When working with SAS®, the Firth

Method	Penalization term $A(\beta)$	Prior $p(\vartheta)$
Ridge	$-\lambda \sum \beta^2$	Normal
LASSO	$-\lambda \sum \lvert\beta\rvert$	Double exponential
Firth	$\frac{1}{2} \log \det(I(\beta))$ where I(β) is the Fisher information matrix.	Jeffreys'
Greenland	$\sum \left(\frac{\beta}{2} - \log\left(1 + \exp \beta\right) \right)$	logF(1,1)

Table 1 Selected penalization terms and corresponding Bayesian priors.

method is available within PROC LOGISTIC as option in the MODEL statement for more elaborate logistic regression models.
When several endpoints (odds ratio, relative risk and absolute risk difference) have to be marginalized at-the-means out of the same model, a switch to PROC NLMIXED becomes necessary (Dörr, 2018). Implementing a partial (record-wise) penalization term for the Firth method is not feasible with basic SAS code. A suitable alternative are logF(1,1) data priors, which extend the original data in an easy manner. For each variable, excluding the intercept, two data prior records (pseudo observations) are added, having a 1 for that variable and a 0 for all other variables. One of these two records is a responder, the other not. Finally, a new variable for partial likelihood contributions is added, having a weight of 1 for original and 0.5 for data prior records (Greenland et al., 2015).

EXAMPLE DATA

An arbitrary dataset of twelve observations is used, see Table 2. It consists of three dichotomous variables, treatment arm (trt), site and sex, and of the continuous variable baseline measurement (base). The z-transformed parallel basez of base will be used in the models. To force a necessity for a penalized likelihood, a rare event situation with no response in one of the treatment arms is set up.

trt	site	sex	base	basez	response
1	1	0	45	-0.63	1
2	1	0	56	0.69	0
1	1	0	51	0.09	1
2	1	1	59	1.05	0
1	1	0	58	0.93	0
2	1	1	39	-1.35	0
1	2	0	48	-0.27	0
2	2	1	47	-0.39	0
1	2	1	36	-1.71	1
2	2	0	64	1.65	0
1	2	0	46	-0.51	0
2	2	1	54	0.45	0

Table 2 Example data.

intercept	trt2	site2	sex2	basez	response
1	0	0	0	-0.63	1
1	1	0	0	0.69	0
1	0	0	0	0.09	1
1	1	0	1	1.05	0
1	0	0	0	0.93	0
1	1	0	1	-1.35	0
1	0	1	0	-0.27	0
1	1	1	1	-0.39	0
1	0	1	1	-1.71	1
1	1	1	0	1.65	0
1	0	1	0	-0.51	0
1	1	1	1	0.45	0

Table 3 Design matrix using reference coding.

intercept	trt2	site2	sex2	basez	response	pweight
1	0	0	0	-0.63	1	*1*
1	1	0	0	0.69	0	*1*
1	0	0	0	0.09	1	*1*
1	1	0	1	1.05	0	*1*
1	0	0	0	0.93	0	*1*
1	1	0	1	-1.35	0	*1*
1	0	1	0	-0.27	0	*1*
1	1	1	1	-0.39	0	*1*
1	0	1	1	-1.71	1	*1*
1	1	1	0	1.65	0	*1*
1	0	1	0	-0.51	0	*1*
1	1	1	1	0.45	0	*1*
0	*1*	*0*	*0*	*0*	*0*	*0.5*
0	*1*	*0*	*0*	*0*	*1*	*0.5*
0	*0*	*1*	*0*	*0*	*0*	*0.5*
0	*0*	*1*	*0*	*0*	*1*	*0.5*
0	*0*	*0*	*1*	*0*	*0*	*0.5*
0	*0*	*0*	*1*	*0*	*1*	*0.5*
0	*0*	*0*	*0*	*1*	*0*	*0.5*
0	*0*	*0*	*0*	*1*	*1*	*0.5*

Table 4 Design matrix with data priors and partial likelihood weights.

As PROC NLMIXED does not have a CLASS statement, a design matrix has to be derived from the dataset manually. Table 3 shows the result with reference coding, in which the first level was used as reference.
Following the instructions from Greenland (2015), the design matrix shown in Table 3 is extended with a new variable pweight for the partial likelihood contributions and with specially designed data prior records. Table 4 illustrates these addendums.

PROC NLMIXED CODE

The syntax of PROC NLMIXED is different compared to e.g. PROC LOGISTIC. Not only the variables of the dataset need to be specified but also the related parameter to be fitted; in the code examples they are composed of the variable name and prefix “b_”. The user can introduce interim variables to keep the code readable, as done in the

```
proc nlmixed
  data   = example_design
  tech   = newrap
  df     = %eval(12 - 5)
;
  eta = b_intercept * intercept
      + b_trt2      * trt2
      + b_site2     * site2
      + b_sex2      * sex2
      + b_agez      * agez
  ;
  prob = 1 / (1 + exp(-1 * eta));

  model response ~ binary(prob);
run; quit;
```

SAS Code 1 Default logistic regression model in PROC NLMIXED.

illustrating code with eta and prob. The model statement finally describes that a binary log-likelihood function is used for the response variable. SAS Code 1 presents the syntax for an ordinary logistic regression model.
To work with partial likelihood weights, the user has to specify the complete log-likelihood manually. SAS Code 2 shows the update of the probability mapping, how the partial log-likelihood is linked with its weight stored in the dataset, and that the model statement works now with a general log-likelihood function.

EXAMPLE RESULTS

Four models have been fitted. Two of them use just the original data and are calculated in PROC LOGISITC and PROC NLMIXED, the third model uses the Firth method on the original data in PROC LOGISTIC, and the last one uses the data prior extension in PROC NLMIXED.

```
proc nlmixed
  data   = example_design_pen
  tech   = newrap
  df     = %eval(12 - 5)
;
  eta = b_intercept * intercept
      + b_trt2      * trt2
      + b_site2     * site2
      + b_sex2      * sex2
      + b_agez      * agez
  ;
  if response eq 1 then prob = 1 / (1 + exp(-1 * eta));
                   else prob = 1 / (1 + exp(     eta));

  prob_adj = (prob>0 and prob<=1)*prob + (prob<=0)*1e-8 + (prob>1);

  wloglike = pweight * log(prob_adj);

  model response ~ general(wloglike);
run; quit;
```

SAS Code 2 Logistic regression model in PROC NLMIXED allowing partial likelihood weights.

		Ordinary fit				Penalized fit			
Parameters		PROC LOGISTIC		PROC NLMIXED		PROC LOGISTIC with Firth method		PROC NLMIXED with logF(1,1) data priors	
Intercept		57.28	(106.6)	14.53	(430.4)	5.62	(4.6)	-0.08	(1.1)
trt	/ b_trt2	-30.20	(68.86)	-56.03	(1013.7)	-3.15	(2.9)	-2.34	(2.0)
site	/ b_site2	-20.38	(41.7)	-41.26	(855.0)	-2.00	(1.9)	-1.34	(1.5)
sex	/ b_sex2	-1.59	(67.9)	-10.72	(1175.9)	1.32	(2.9)	-0.17	(1.6)
agez	/ b_agez	-13.51	(29.0)	-29.26	(813.7)	-0.65	(1.1)	-1.34	(1.1)

Table 5 Model fit results. Estimates and standard errors are presented.

Both procedures show the classical symptoms of quasi-complete separation of data points in the ordinary model fit: the parameter estimates are extreme with a tremendous standard error. On the contrary, the penalized fits provide fair estimates and standard errors.

CONCLUSION

logF(1,1) data priors are a probate means to stabilize the fit of a logistic regression when the dataset consists of sparse data and PROC NLMIXED is necessary for deriving marginalized endpoints.

REFERENCES

(1) Greenland S, Mansournia M A: Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. Statist. Med., 34: 3133–3143. doi: 10.1002/sim.6537. 2015.

RECOMMENDED READING

(1) Cole S R, Chu H, Greenland S: Maximum Likelihood, Profile Likelihood, and Penalized Likelihood: A Primer. Am J Epidemiol., 179(2): 252–260. doi: 10.1093/aje/kwt245. 2014.

(2) Firth D: Bias reduction of maximum likelihood estimates. Biometrika 80:27-38, 1993.

(3) Dörr M: A single study that solves multiple endpoint preferences - statistical solution for binary endpoint models. PhUSE SDE, Brussels, Sep 21st 2018. www.phusewiki.org/docs/2018%20SDE/brussels/presentations/11.%20Martin%20D%c3%b6rr%20-%20A%20single%20study%20that%20solves%20multiple%20endpoint%20preferences.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Email: mdoerr@clinipace.com