Biometrika (2010), **xx**, x, *pp.* 1–9

Printed in Great Britain

Parameter Redundancy with Covariates

By D. J. Cole and B. J. T. Morgan

School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7NF, England.

d.j.cole@kent.ac.uk b.j.t.morgan@kent.ac.uk

Abstract

We show how to determine the parameter redundancy status of a model with covariates from that of the same model without covariates, thereby simplifying calculation considerably. A matrix decomposition is necessary to ensure that symbolic computation computer programs return correct results. The paper is illustrated by mark-recovery and latent-class models, with associated Maple code.

Keywords: Computer algebra; identifiability; latent class models; Maple; mark-recovery models; symbolic computation.

1. Introduction

A model is parameter redundant if it can be reparameterised in terms of a smaller number of parameters than the size of its defining parameter set, so that using classical inference it would not be possible to estimate all the original parameters. One approach to removing parameter redundancy is to include covariates in a model, by setting parameters to be appropriate functions of covariates, as explained below. However this has not been formally evaluated. In this paper we show how to determine whether a model with covariates is parameter redundant using information from the equivalent model without covariates. We also show how to establish what combinations of parameters may be estimated and whether the result applies to the whole of the parameter space or just a region.

A method for detecting parameter redundancy was developed by Catchpole & Morgan (1997) for members of the exponential family, where the observations have expectation μ and the unknown parameters are θ. This involves forming the derivative matrix $D = \partial\mu/\partial\theta$ and calculating its symbolic rank, which determines how many parameters are estimable. The symbolic rank is the rank of a matrix that has symbolic rather than numeric entries. It is also possible to replace μ by the natural parameters (Bekker et al., 1994), due the one-to-one transformation between μ and the natural parameters. Early work was by Rothenberg (1971) and see for example Evans & Chappell (2000) and Little et al. (2009) for further developments. For a parameter redundant model it is further possible to find which parameter combinations and/or individual parameters are estimable by solving a set of Lagrange partial differential equations (Catchpole et al., 1998). Extension theorems allow conclusions to be drawn for models of arbitrary dimension; see Catchpole & Morgan (1997) and Catchpole & Morgan (2001). The symbolic computation involved can be performed by a symbolic computation computer program (Catchpole et al., 2002), and Maple code for the examples in this paper can be downloaded from `www.kent.ac.uk/ims/personal/djc24/covariates.htm`.

A model that is not parameter redundant is termed full rank. We define an essentially full rank model as one that is full rank for all θ and all permitted values of any covariates that are present in the model, whereas a conditionally full rank model is full rank for some but not all θ and permitted values of the covariates. A modified PLUR decomposition of D (Corless & Jeffrey, 1997) can be used to distinguish between essentially and conditionally full rank models. We set $D = PLUR$, where P is a permutation matrix, L is a lower triangular matrix with ones on the diagonal, U is an upper triangular matrix and R is a matrix in reduced echelon form. A model is conditionally full rank if $\det(U) \neq 0$ for any θ. Otherwise the model is essentially full rank. This result follows from Theorem 2 of Corless & Jeffrey (1997).

Adding covariates increases the structural complexity of D, so that symbolic computation programs may lack the memory to calculate the matrix symbolic rank. In addition, the work of Catchpole & Morgan (1997) was for models expressed in terms of rational functions only, and inclusion of covariates requires a check to be made of the symbolic rank of U if a symbolic computation computer package is used.

2. Parameter redundancy of models with covariates

Consider first a model without covariates, which has p parameters. Suppose the symbolic rank of D is q. A q-dimensional set of estimable parameters, β, may be found as the solution to the partial differential equations $\sum_{r=1}^{p} \alpha_{r,j} \partial f / \partial \theta_r = 0$, where $\alpha_{i,j}$ is the ith element of the vector solution of $\alpha_j^T D = 0$ and where $j = 1, ..., p - q$, with $q < p$ (Catchpole et al., 1998). The derivative matrix $D_\beta = \partial \mu / \partial \beta$ has full row rank; if $q = p$ then $\beta = \theta$.

We add covariates to a model by setting elements of θ to be differentiable functions of combinations of covariates, resulting in p_c parameters θ_c. For instance we might set

$\theta_i = \Psi_i(\alpha^T x_i)$ where x_i are covariates, for certain i. We assume that $\partial\theta/\partial\theta_c$ is full rank, which will normally be the case; in the above illustration this follows if $d\Psi_i(z)/dz \neq 0$ for all i and (x_{ji}) is full rank. This is true in the examples below, when logistic functions are used.

Let $\tilde{D} = \partial\beta/\partial\theta_c$. Rather than finding the symbolic rank of the derivative matrix $D_c = \partial\mu/\partial\theta_c$, we may use the result below.

THEOREM 1. *The ranks of the derivative matrices D_c and $\tilde{D}$ are both equal to min(p_c, q).*

For proof see the Appendix.

Remark: If the covariate model is parameter redundant, a set of partial differential equations can be derived from $\tilde{D}$ in the same way as Catchpole et al. (1998), to find the set of estimable parameters. If the model is full rank, whether it is essential or conditionally full rank may be determined from the modified PLUR decomposition of $\tilde{D}$.

3. TWO EXAMPLES

3·1. *Example 1 Conditional analyses of ring recovery data*

In ring recovery data the total number of birds ringed in each year may be unknown or unreliable. In such a case a model can only be fitted by conditioning on the number of birds recovered from each cohort. The most commonly used model for such a conditional analysis assumes that there is a constant probability of recovery, λ. However there is evidence that the reporting probability of wild birds in Britain in recent years has been decreasing over time (Baillie & Green, 1987). If we try to account for time-variation in the recovery probability directly, then the resulting model is parameter redundant, as shown below. However introducing a logistic regression of λ on time can result in a full rank model.

It is assumed that the probability of surviving the first year of life is ϕ_1, the probability of surviving other years is ϕ_a and the recovery probability in year j, λ_j, is dependent on time. Thus the probability of a bird being ringed in year i and recovered in year j conditional on being found dead is

$$Q_{ij} = \begin{cases} (1-\phi_1)\lambda_j/F & i = j \\ \phi_1\phi_a^{j-i-1}(1-\phi_a)\lambda_j/F & i < j \end{cases} \tag{1}$$

with $F = (1-\phi_1)\lambda_i + \sum_{k=i+1}^{J} \phi_1\phi_a^{k-i-1}(1-\phi_a)\lambda_k$ for $i = 1,\ldots,I$, $j = 1,\ldots,J$, and constants I and J. Catchpole & Morgan (1997) show that we may consider the derivative matrix of Q, rather than that of μ. We just analyse models for which $I = J \geq 3$. If $I = J = 3$, the derivative matrix has rank 3, and as there are 5 parameters this model is parameter redundant. From solving the appropriate Lagrange equations we find that the estimable parameter combinations are $(\phi_a - \phi_1)/\{\phi_a(1-\phi_1)\}$, $\phi_a\lambda_2/\lambda_1$ and $\phi_a^2\lambda_3/\lambda_1$. Using an extension theorem we can show that an estimable set of parameters for this model in general is $\beta = \left[(\phi_a - \phi_1)/\{\phi_a(1-\phi_1)\}, \phi_a\lambda_2/\lambda_1, \ldots, \phi_a^{J-1}\lambda_J/\lambda_1\right]$. If we add covariates to the reporting probabilities, then from Theorem 1 as long as $p_c < J$, the number of terms in β, the resulting covariate model will be full rank. For example, suppose $\lambda_j = 1/\{1+\exp(\alpha+\beta j)\}$. This model has 4 parameters $\theta_c = (\phi_1, \phi_a, \alpha, \beta)$. If $I = J = 3$ this model with covariates is parameter redundant, however if $I = J \geq 4$ the model with covariates is full rank. If $J = 4$, the PLUR decomposition of the derivative matrix $\tilde{D}$, see Maple code, results in $\mathrm{Det}(\tilde{U}) = 0$ at $\phi_a = 1$ or $\beta = 0$, where $\tilde{U}$ is the upper triangular matrix formed from a PLUR decomposition of $\tilde{D}$. An annual survival probability of unity is unrealistic, and the case $\beta = 0$ results in constant reporting probability, when the model is parameter redundant. Therefore this model with covariates is conditionally full rank for $I = J \geq 4$.

Remark: This example can also be examined directly by Maple, without using Theorem 1. However, in the 3×3 case Maple incorrectly gives the rank of D_c as 4, whereas in fact the correct rank is 3. This is due to the failure of Maple to simplify exponential algebraic terms completely in this case. A modified PLUR decomposition of D_c reveals that $\mathrm{Det}(U) = 0$ everywhere, which is incorrect from the definition of a PLUR decomposition, as shown in Theorem 1 of Corless & Jeffrey (1997); such an error can arise when a symbolic computation package is used with non-rational terms. The diagnostic check of $\mathrm{Det}(U)$ should always be performed in such a case. If $\mathrm{Det}(U) = 0$ everywhere then simplification by hand is necessary.

3·2. *Example 2 Adding Individual Covariates to Latent Class Models*

Forcina (2008) considers three latent class models with individual covariates, and concludes that the models are almost certainly full rank, using $20,000$ numerical evaluations of derivative matrices. Here we use symbolic computation for the first example of Forcina (2008); treatment of the further two examples of that paper can be found in the Maple code. The first example consists of two binary response variables Y_1 and Y_2, which are conditionally independent given a binary latent variable Z. Let $\theta_{j,k} = \mathrm{pr}(Y_j = 1 \mid Z = k)$ and $p = \mathrm{pr}(Z = 1)$. It may be verified using the Maple code that the rank of the appropriate derivative matrix is 3 and that the set of estimable parameters is $\beta = \{p(\theta_{1,0} - \theta_{1,1}) - \theta_{1,0}, p(1-p)(\theta_{2,1} - \theta_{2,0})(\theta_{1,1} - \theta_{1,0}), p(\theta_{2,1} - \theta_{2,0}) + \theta_{2,0}\}$.

The model is generalised for n individuals, so that for individual i, $\mathrm{pr}(Y_{j,i} = 1 \mid Z_i = k) = \theta_{j,k,i}$ and $\mathrm{pr}(Z_i = 1) = p_i$. In this case the elements of μ are proportional to the joint probability functions of $Y_{1,i}$ and $Y_{2,i}$. Using an extension theorem it can be shown that in this case the derivative matrix will have rank $3n$ and estimable parameters $\{p_i(\theta_{1,0,i} - \theta_{1,1,i}) - \theta_{1,0,i}, p_i(1-p_i)(\theta_{2,1,i} - \theta_{2,0,i})(\theta_{1,1,i} - \theta_{1,0,i}), p_i(\theta_{2,1,i} - \theta_{2,0,i}) + \theta_{2,0,i}\}$, for $i = 1, \ldots, n$. Then by Theorem 1, if individual covariates are added, as long as there are fewer than $3n$ parameters in the model with covariates it is full rank. This is true for the example in Forcina (2008), where $n = 5$ and $p_c = 8$; no numerical investigation is necessary.

4. Discussion

We have shown that the parameter-redundancy status of a model with covariates may be determined from the equivalent model without covariates. This is valuable, because adding covariates to a model can make derivative matrix calculations much more complex. Symbolic algebra programs such as Maple can fail to simplify terms involving exponentials correctly, such as when a logistic link function is used, as in both examples. If such a program is used when a model contains non-rational terms then a PLUR decomposition is necessary to check that the correct results are obtained. An advantage of Theorem 1 is that it may avoid the need for symbolic computation of indeterminate functions.

In this paper we have considered the parameter redundancy status of models irrespective of the extent of data, and in the two examples missing data could change conclusions. Missing data are easily dealt with, effectively by appropriately redefining μ, as explained in Catchpole & Morgan (2001).

Acknowledgements: We thank Daniel Brown for discussion of the first example, Remi Choquet for discussion of the limitations of symbolic computation packages, and Ted

Catchpole, an Associate Editor and two referees for their helpful comments on a first draft of the paper. The work of DJC was supported by the EPSRC grant supporting the National Centre for Statistical Ecology.

5. Appendix: Proof of Theorem 1

The model without covariates has rank q and estimable parameter combinations β. Hence $D_\beta = \partial\mu/\partial\beta$ has full row rank q. By the chain rule the derivative matrix of the model with covariates is $D_c = \partial\mu/\partial\theta_c = \partial\beta/\partial\theta_c \times \partial\mu/\partial\beta$, giving

$$\text{rank}(D_c) = \text{rank}\left(\frac{\partial\beta}{\partial\theta_c}\frac{\partial\mu}{\partial\beta}\right) = \text{rank}\left(\frac{\partial\beta}{\partial\theta_c}\right) \equiv \text{rank}(\tilde{D}),$$

which is due to the result of Horn & Johnson (1985) that $\text{rank}(AB) = \text{rank}(A)$, if B has full row rank. Also by the chain rule,

$$\tilde{D} = \frac{\partial\beta}{\partial\theta_c} = \frac{\partial\theta}{\partial\theta_c}\frac{\partial\beta}{\partial\theta}.$$

Now $\partial\theta/\partial\theta_c$ has full rank p_c, and by the chain rule $\partial\mu/\partial\theta = \partial\beta/\partial\theta \times \partial\mu/\partial\beta$. Thus using the above result from Horn & Johnson (1985)

$$\text{rank}\left(\frac{\partial\beta}{\partial\theta}\right) = \text{rank}\left(\frac{\partial\beta}{\partial\theta}\frac{\partial\mu}{\partial\beta}\right) = \text{rank}\left(\frac{\partial\mu}{\partial\theta}\right) = q,$$

so that $\partial\beta/\partial\theta$ has full column rank. The rank of a matrix can be calculated using Gaussian elimination, which continues until the matrix is in reduced echelon form. The rank of $\tilde{D}$ is then equal to the number of ones on the main diagonal of the matrix in reduced echelon form. This is equivalent to writing $\tilde{D} = PLUR$, as in Section 1. We show in the Maple code that as $\partial\theta/\partial\theta_c$ and $\partial\beta/\partial\theta$ are full rank we have $\det(U) \neq 0$. Then, by Corless & Jeffrey (1997), as $\det(U) \neq 0$, $\text{rank}(\tilde{D}) = \text{rank}(R)$. The matrix R has a complete unit diagonal, so that its rank depends on its dimensions, and $\text{rank}(\tilde{D}) = \min(p_c, q)$.

References

Baillie, S. R. & Green, R. E. (1987). The importance of variation in recovery rates when estimating survival rates from ringing recoveries. *Acta Ornithol* **23**, 41–60.

Bekker, P. A., Merchens, A. & Wansbeek, T. J. (1994). *Identification, Equivalent Models and Computer Algebra.* Boston: Academic Press.

Catchpole, E. A. & Morgan, B. J. T. (1997). Detecting parameter redundancy. *Biometrika* **84**, 187–196.

Catchpole, E. A. & Morgan, B. J. T. (2001). Deficiency of parameter-redundant models. *Biometrika* **88**, 593–598.

Catchpole, E. A., Morgan, B. J. T. & Freeman, S. N. (1998). Estimation in parameter redundant models. *Biometrika* **85**, 462–468.

Catchpole, E. A., Morgan, B. J. T. & Viallefont, A. (2002). Solving problems in parameter redundancy using computer algebra. *J. Appl. Statist.* **29**, 625–636.

Corless, R. M. & Jeffrey, D.J. (1997). Detecting parameter redundancy. *Sigsam Bulletin* **31**, 20–28.

Evans, N. D. & Chappell, M. J. (2000). Extensions to a procedure for generating locally identifable reparameterisations of unidentifable systems *Mathematical Biosciences* **168**, 137–159.

Forcina, A. (2008). Identifiability of extended latent class models with individual covariates. *Comput. Stat. Data. An.* **52**, 5263–5268.

Horn, R. A. & Johnson, C. R. (1985). *Matrix Analysis.*. Cambridge: Cambridge University Press.

Little, M. P., Heidenreich, W. F. & Li, G. (2009). Parameter identifiability and redundancy in a general class of stochastic carcinogenesis models. *PLoS ONE* **4**, e8520.

Rothenberg, T. J. (1971). Identification in parametric models *Econometrica* **39**, 577–591.