

Department of Mathematical and Statistical Methods¹, Department of Genetics and Animal Breeding²,
August Cieszkowski Agricultural University of Poznań, Poland

KRZYSZTOF MOLIŃSKI¹, MACIEJ SZYDŁOWSKI², TOMASZ SZWACZKOWSKI²,
ANITA DOBEK¹, and EWA SKOTARCZAK¹

An algorithm for genetic variance estimation of reproductive traits under a threshold model

Summary

Some binary traits are determined by a large number of loci. The standard approach, in this case, is to model an unobservable variable (liability) with a fixed threshold. We present a method for the estimation of the genetic additive variance component under a threshold animal model with fixed effects included. The method can be applied to data with repeated measurement per animal. The unknown parameters of the model have been estimated by Gibbs sampling. The numerical properties of the method are investigated on simulated data for a large real pedigree of breeding stock of laying hens. The algorithm shows good mixing properties, producing consistent estimates from many distinct runs. The application of the method is exemplified on fertility data recorded for the same pedigree.

Key Words: binary data, Gibbs sampling, threshold model, simulation study

Zusammenfassung

Titel der Arbeit: **Ein Algorithmus für die Schätzung genetischer Varianzen von Reproduktionsmerkmalen mittels eines Schwellenwertmodells**

Viele binäre Merkmale werden von vielen Genen determiniert. Die Standardmethode ist in solchen Fällen die Modellierung der nichtbeobachtbaren Variablen durch einen festen Schwellenwert. Es wird eine Methode für die Schätzung der additiv-genetischen Varianzkomponente mittels eines Schwellenwert Tiermodells unter Einbeziehung von festen Effekten vorgestellt. Diese Methode kann für Daten aus wiederholten Messungen am Einzeltier angewendet werden. Die unbekannten Modellparameter werden mittels Gibbs-Sampling geschätzt. Die numerischen Eigenschaften der Methode wurden untersucht an simulierten Daten für ein umfangreiches Pedigree aus dem Zuchtstamm von Legehennen. Der Algorithmus zeigte gute 'Mischungs' Eigenschaften und lieferte konsistente Schätzungen für die Parameter bei mehrfach wiederholten Simulationsläufen.

Schlüsselwörter: Binäre Merkmale, Gibbs-Sampling, Schwellenwertmodell, Simulationsuntersuchungen

Introduction

Some reproductive traits (e.g. fertility) are determined by many loci but are binary scored and, in consequence, usually troublesome in genetic analysis. The standard approach is to assume an unobservable continuous scale with a fixed threshold. The threshold model has become a preferred method in genetic evaluations of reproductive binary traits. The methodology (for sire model) has been described by GIANOLA and FOULLEY (1983) and developed by others authors (e.g. MISZTAL et al., 1989; MATOS et al., 1997). SORENSSEN et al. (1995) implemented Gibbs sampler for binary data where all conditional posterior distributions have a known form, so the sampler is relatively easy to apply.

In poultry, reproductive eggs are binary scored (eg. hatched or not) and may be analysed in this form. Alternatively, the set of binary observations can be summarised

as a proportion variable recorded for each dam (SEWALEM et al., 1998). Both approaches have been implemented and may lead to various estimates of genetic parameters (SZWACZKOWSKI et al., 2002). In general, the inclusion of binary data instead of summarised ‘percentage’ data increases the number of observations.

In this paper we present Gibbs sampling algorithm for genetic evaluation of a binary trait under the threshold model. Simulation study on the efficiency of sampling scheme is reported. The method is exemplified with genetic evaluation of fertility in laying hens.

Method

Model. Let us assume that for each of the observed individuals we have a number of repeated measurements of a binomial random variable describing this individual. The binominal variable is an expression of an underlying, unobservable, continuous random variable. As in all threshold models we assume that there exists an unknown threshold for the continuous variable which partitions the real line into two intervals corresponding to categories of response (MATOS et al., 1997; SORENSEN et al., 1995). In our case without loss of generality the threshold can be taken as 0.

The unobserved variable can be modelled in the following way

$$w_{ij} = \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a} + e_{ij} \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n_i$$

where $\boldsymbol{\beta}$ is a p -vector of fixed effects, \mathbf{a} is a q -vector of random additive genetic effects, \mathbf{x}_i and \mathbf{z}_i are vectors choosing the appropriate fixed and random effects and e_{ij} is a random error with 0 expectation and variance equal to 1. The random additive genetic effects follow a normal distribution $N(\mathbf{0}, \mathbf{A}\sigma_a^2)$, where \mathbf{A} is the additive relationship matrix. The expression of the unobservable variable w_{ij} is the observation y_{ij} which takes value 1 or 0 (e.g. fertilized or unfertilized).

Parameter estimation. The estimation of model parameters is realized by the Gibbs sampling procedure. The consecutive parameters are sampled from the following distributions:

$$w_{ij} \mid \text{all other parameters, and } y_{ij} = 1 \sim N(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a}, 1) \cdot \delta(w_{ij} > 0)$$

where $\delta(\cdot)$ is an indicator taking value 1 if the condition is satisfied and 0 otherwise;

$$w_{ij} \mid \text{all other parameters, and } y_{ij} = 0 \sim N(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a}, 1) \cdot \delta(w_{ij} \leq 0)$$

$$\beta_k \mid \text{all other parameters} \sim N\left(\frac{1}{r_k} \sum w_{ij}^{(\beta_k)}, \frac{1}{r_k}\right)$$

where the summation goes over all w_{ij} connected with the k -th level of $\boldsymbol{\beta}$, r_k being the number of observations;

$$\sigma_a^2 \mid \text{all other parameters} \sim \mathbf{a}' \mathbf{A}^{-1} \mathbf{a} / \chi_q^2.$$

The conditional posterior density of additive genetic effect a_{i^*} is normal with expected value E and variance V of the following form

- when i^* is a dam:

$$E = c_1 J_p \frac{a_{S_i^*} + a_{D_i^*}}{2} + c_3 \sum_{p(\text{progeny})} (2a_p - a_{(i^*)}) + c_4 w_{i^*}^a;$$

- when i^* is a sire:

$$E = c_1 J_p \frac{a_{S_{i^*}} + a_{D_{i^*}}}{2} + c_2 \sum_{p(final)} (2w_p^a - a_{(i^*)}) + c_3 \sum_{p(nonfinal)} (2a_p - a_{(i^*)}) + c_4 w_{i^*}^a;$$

- when i^* is a final progeny:

$$E = c_1 J_p \frac{a_{S_{i^*}} + a_{D_{i^*}}}{2} + c_4 w_{i^*}^a.$$

$$V = \frac{1}{D} 2\sigma_a^2 (\sigma_a^2 + 2),$$

where:

$$w_{i^*}^a = \frac{1}{n_{i^*}} \sum_j w_{i^*j} - \mathbf{x}_{i^*}' \boldsymbol{\beta},$$

$a_{(i^*)}$ is the additive genetic effect of spouse of i -th individual,

$$D = [2(J_p + 1) + J_s n_p + 2J_D \sigma_a^2] (\sigma_a^2 + 2)$$

and

$$c_1 = \frac{2}{D} (J_p + 1) (\sigma_a^2 + 2), \quad c_2 = \frac{1}{D} J_s \sigma_a^2,$$

$$c_3 = \frac{1}{D} J_s (\sigma_a^2 + 2), \quad c_4 = \frac{2}{D} J_D \sigma_a^2 (\sigma_a^2 + 2)$$

with n_p being a total number of progeny. The parameters $J_s = 1$ or 0 as the individual has no offspring, $J_p = 0$ or 1 as the individual is, or is not a founder and $J_D = 1$ or 0 as the individual is or is not observed.

Mixing properties. Due to a correlation among variables a sampling process may show slow mixing, i.e. move slowly in the parameter space. This may limit the practical application of the method. There is no way in which convergence can be proved beyond any doubt. However, by monitoring many distinct variables we get much power to detect a lack of convergence, if any. The standard procedure is to compare estimates from different chains starting from distinct initial values or/and seeds for a random number generator. To evaluate the mixing properties we monitored the estimates of \mathbf{a} - the vector of random additive effects and compared the two estimates from two independent runs. The random values for genetic additive variance were monitored to visualize the speed with which the process converges to the true values.

The simulation study was performed on data generated for a real pedigree structure. The pedigree was nine-generation breeding stock of laying hens, consisting of 396 founders and 5980 non-founders. The pedigree included large full-sibs nested within the sire group. The average inbreeding coefficient was 0.09. Three data sets were generated under a pure genetic additive model assuming a heritability $h^2=0$, $1/3$ and $5/6$, respectively (Table). The estimates were taken as means from 10000 random values after discarding the first 1000 samples. The samples were collected every 20-th iteration.

Results

The algorithm shows good mixing properties. The random values of genetic additive variance converged rapidly to true values from distinct extreme starting values ($\sigma_a^2=0.001$ and $\sigma_a^2=5$). The process was stable through out all iterations (Fig. 1). For each of the three data sets the two independent estimates of vector **a** were very consistent (Fig. 2). This strongly suggests a very good mixing performance of the method, at least for the pedigree structure used here. The heritability estimates were quite close to the true values used for data simulation (Table). This was also the case for null genetic contribution in the first data set.

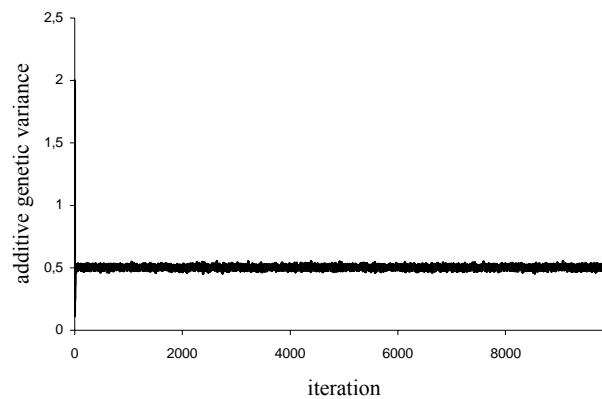


Fig. 1: The random values of genetic additive variance – example of simulation study. The true value was 0.5 (Zufallswerte der additiv-genetischen Varianz am Beispiel der Simulation bei dem Realwert von 0,5)

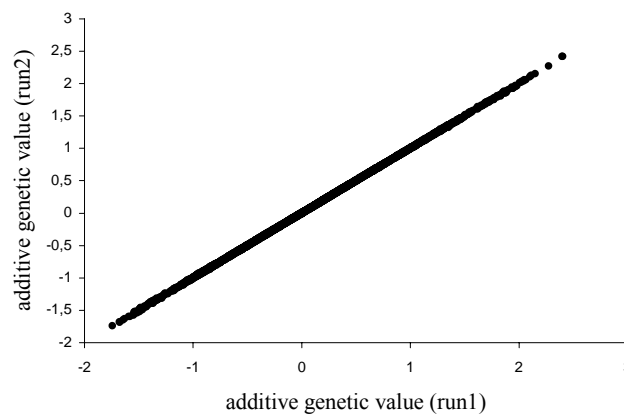


Fig. 2: Plot of two estimates of individual genetic additive values. Position of each point is determined by two estimates based on two independent runs of the algorithm (Diagramm von zwei Schätzungen der individuellen additiv-genetischen Varianz. Die Lage jedes Punktes wird durch zwei Schätzungen auf der Basis von zwei unabhängigen Programmabläufen bestimmt)

Table

The true and estimated values of genetic additive variance (σ_a^2) for three simulated data sets (Real- und Schätzwerte der additiv-genetischen Varianzen (σ_a^2) für drei simulierte Dateien)

| True value Realwert | Estimated value Schätzwert |
|------------------------|-------------------------------|
| 0.0 | 0.001 |
| 0.5 | 0.502±0.0195 |
| 5.0 | 4.990±0.057 |

Numerical example

We used the method to analyse fertility data of laying hens. The pedigree was the same as used for simulation study. The fertility data were collected from 1989 through 1997. Eggs were examined by candling on 8th day of incubation. The trait was measured on 30 eggs per dam, hence 30 binary observations were available for each of the 5936 hens. The mean percentage of fertilized eggs per dam was $79 \pm 30\%$. The following model was used for liability variable

$$y_{ijkl} = \mu + g_i + h_j + a_{ijk} + e_{ijkl},$$

where g_i is the fixed effect of i -th year (generation), h_j is the fixed effect of j -th hatch period, a_{ijk} is the random genetic additive effect of ijk -th dam, and e_{ijkl} is random residual effect connected with $ijkl$ -th egg.

The estimated density for genetic variance is given in Figure 3. The mean and mode of distribution were 0.776 ± 0.021 and 0.785, respectively. This gives a relatively high heritability estimate of 0.437 ± 0.021 . The genetic trend for fertility was obtained by averaging the individuals' genetic effects within each generation (Fig. 4).

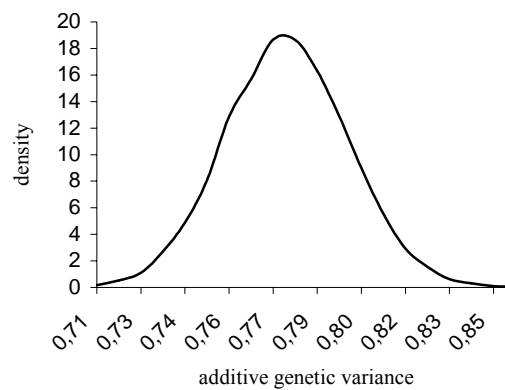


Fig. 3: Probability density estimate of genetic additive variance - case study of fertility in laying hens (Wahrscheinlichkeitsdichte für die Schätzung der additiv-genetischen Varianz – Fruchtbarkeitsanalyse der Legehennen)

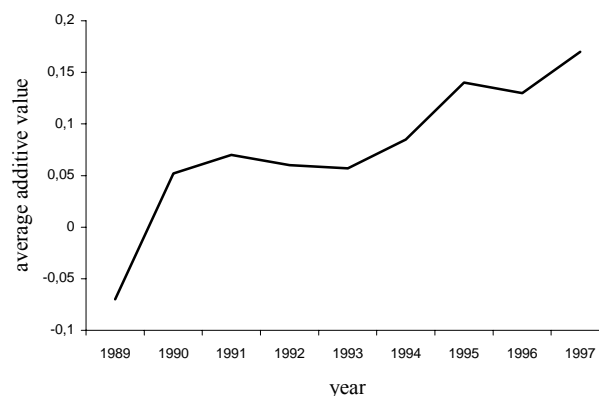


Fig. 4: Estimate of genetic trend of fertility in laying hens (Schätzung des genetischen Trends der Fruchtbarkeit von Legehennen)

Practical implications

The threshold models were developed a few decades ago (see e.g. DEMPSTER and LERNER, 1950), but their application was limited, mainly because of the difficulty to

integrate over random effects analytically. Over the past decade several approaches to analysing threshold characters have been developed. In this study we describe methods of estimating variance component as well as fixed and random genetic effects based of simple implementation of the Monte Carlo approach. The method is presented in application to poultry data, but it can be applied to any binary threshold data of other livestock and poultry species.

Although the Monte Carlo procedures may provide excellent solutions to complex problems, their efficiency totally depends on particular applications. For genetic problems there are a number of samplers to study additive genetic effects. In general, simple samplers suffer mixing difficulties due to strong dependencies on pedigree structure. More advanced samplers use a blocking strategy and sample the entire vector of additive effects. This, however, makes the sampler complex and more expensive. Our sampler used limited blocking and proved to be efficient for the pedigree studied in this paper. Although we have encountered no problem for other pedigree structures, the efficiency of the algorithm needs to be verified for each particular pedigree under study.

We estimated a relatively high heritability of fertility. Many authors (e.g. MATOS et al., 1997) reported higher heritability for the threshold than the linear model. This corresponds with our earlier studies carried out by SZWACZKOWSKI et al. (2002). The difference can be partly explained by non-normality of the trait when analysed as percentage data. Although several transformation functions have been proposed, they rarely satisfy model assumptions. This suggests that the data should be rather analysed in their original binary scale, when possible.

Higher heritability from the threshold models suggests that a selection index based on the threshold model may give higher accuracy of selection. The algorithm presented in this study allows for unitrait genetic analysis. In practice, however, multitrait evaluation is required. The algorithm to sampling of additive genetic additive values for an arbitrary number of continuous traits has been presented by SZWACZKOWSKI et al. (2001). The method presented here can be relatively easily extended to mixed threshold-linear problems.

Acknowledgements

This study was supported by the State Committee for Scientific Research, grant No. 5 P06D02019.

References

- DEMPSTER, E.R.; LERNER, I.M.:
Heritability of threshold characters. *Genetics* **35** (1950), 212-236
- DOBEK, A.; MOLIŃSKI, K.; SZYDŁOWSKI, M.; SZWACZKOWSKI, T.:
Analysis of single gene multitrait effects in livestock by the use of Gibbs sampling. *J. Appl. Genet.* **41** (2000) 4, 275-283
- GIANOLA, D.; FOULLEY, J.L.:
Sire evaluation for ordered categorical data with a threshold model. *Genet. Sel. Evol.* **15** (1983), 201-224
- HARTMANN, C.:
Selection for yolk production in laying hens. *Acta Universitatis Agriculturae Sueciae, Agraria* 298, Swedish University of Agricultural Sciences, Uppsala, Sweden, 2001
- MATOS, C.A.P.; THOMAS, D.L.; GIANOLA, D.; PEREZ-ENCISO, M.; YOUNG, L.D.:
Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear models. I. Estimation

- of genetic parameters. *J. Anim. Sci.* **75** (1997), 76- 87
- MISZTAL, I.; GIANOLA, D.; FOULLEY, J.L.:
Computing aspects of a nonlinear method of sire evaluation for categorical data. *J. Dairy Sci.* **72** (1989), 1557-1568
- SEWALEM, A.; JOHANSSON, K.; CARLGREN, A.B.; WILHELMSON, M.; LILLPERS, K.:
Are reproductive traits impaired by selection for egg production in hens? *J. Anim. Breed Genet.* **115** (1998), 281-297
- SORENSEN, D.A.; ANDERSEN, S.; GIANOLA, D.; KORSGAARD, I.:
Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* **27** (1995), 229-249
- SZWACZKOWSKI, T.; MOLINSKI, K.; SZYDŁOWSKI, M.; SKOTARCZAK, E.; PIOTROWSKI, P.; DOBEK, A.:
Comparison of heritability estimates of hatchability layer traits obtained with linear and threshold model. 7-th World Congress on Genetics Applied to Livestock Production (accepted for publication). (2002)
- SZWACZKOWSKI, T.; SZYDŁOWSKI, M.; MOLINSKI, K.; DOBEK, A.:
Multivariate analysis of mixed inheritance model of performance traits in layers using Gibbs sampling. *J. Anim. Breed. Genet.* **118** (2001), 205-211

Received: 2002-05-15

Accepted: 2002-12-03

Authors' addresses

Prof. Dr. KRZYSZTOF MOLINSKI
Department of Mathematical and Statistical Methods
August Cieszkowski Agricultural University of Poznań
Wojska Polskiego 28 str.
PL 60-637 Poznań
Poland

Dr. MACIEJ SZYDŁOWSKI
Department of Genetics and Animal Breeding
August Cieszkowski Agricultural University of Poznań
Wołyńska 33 str.
PL 60-637 Poznań
Poland

Prof. Dr. TOMASZ SZWACZKOWSKI
Department of Genetics and Animal Breeding
August Cieszkowski Agricultural University of Poznań
Wołyńska 33 str.
PL 60-637 Poznań
Poland

Prof. Dr. ANITA DOBEK
Department of Mathematical and Statistical Methods
August Cieszkowski Agricultural University of Poznań
Wojska Polskiego 28 str.
PL 60-637 Poznań
Poland

Ms. EWA SKOTARCZAK
Department of Mathematical and Statistical Methods
August Cieszkowski Agricultural University of Poznań
Wojska Polskiego 28 str.
PL 60-637 Poznań
Poland