

A marginal quasi-likelihood approach to the analysis of Poisson variables with generalized linear mixed models

JL Foulley*, S Im²

¹ INRA, Institut National de la Recherche Agronomique, Station de Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas Cedex;

² INRA, Station de Biométrie et d'Intelligence Artificielle, 31326 Castanet Tolosan Cedex, France

(Received 19 June 1992; accepted 16 November 1992)

Summary – This paper extends to Poisson variables the approach of Gilmour, Anderson and Rae (1985) for estimating fixed effects by maximum quasi-likelihood in the analysis of threshold discrete data with a generalized linear mixed model.

discrete variable / Poisson distribution / generalized linear mixed model / quasi-likelihood

Résumé – Une approche de quasi-vraisemblance pour l'analyse de variables de Poisson en modèle linéaire mixte généralisé. Cet article généralise à des variables de Poisson l'approche de Gilmour, Anderson et Rae (1985) destinée à l'estimation par maximum de quasi-vraisemblance des effets fixés lors de l'analyse de variables discrètes à seuils sous un modèle mixte.

variables discrètes / distribution de Poisson / modèle linéaire mixte généralisé

INTRODUCTION

As shown by Ducrocq (1990), there has been recently some interest in non linear statistical procedures of genetic evaluation. Examples of such modelling procedures involve: 1) the threshold liability model for categorical data (Gianola and Foulley, 1983; Harville and Mee, 1984) and for ranking data in competitions (Tavernier,

* Correspondence and reprints

1991); 2) Cox's proportional hazard model for survival data (Ducrocq *et al*, 1988); and 3) a Poisson model for reproductive traits (Foulley *et al*^a, 1987). In FGI, estimation of fixed (β) and random (u) effects involved in the model is based on the mode of the joint posterior distribution of those parameters. As discussed by Foulley and Manfredi (1991), this procedure is likely to have some drawbacks regarding estimation of fixed effects due to the lack of integration of random effects. A popular alternative is the quasi-likelihood approach (Mc Cullagh and Nelder, 1989) for generalized linear models (GLM) which only requires the specification of the mean and variance of the distribution of data. This procedure has been used extensively by Gilmour *et al*^b (1985) in genetic evaluation for threshold traits. In particular, GAR derived a very appealing algorithm for computing estimates of fixed effects which resembles the so-called mixed model equations of Henderson (1984). The purpose of this note is to show how the GAR procedure can be extended to Poisson variables.

THEORY

The same model as in FGI is postulated. Let Y_k be the random variable (with realized value $y_k = 0, 1, 2, \dots$) pertaining to the k th observation ($k = 1, 2, \dots, K$). Given λ_k , the Y_k 's have independent Poisson distributions with parameter λ_k , *ie*:

$$\Pr(Y_k = y_k) = \lambda_k^{y_k} \exp(-\lambda_k)/y_k!. \quad [1]$$

As the canonical link for the Poisson distribution is the logarithm (Mc Cullagh and Nelder, 1989), λ_k is modelled as:

$$\ln(\lambda_k) = \mathbf{x}'_k \beta + \mathbf{z}'_k u, \quad [2]$$

where β and u are $(p \times 1)$ and $(q \times 1)$ vectors of fixed and random effects respectively, and \mathbf{x}'_k and \mathbf{z}'_k are the corresponding row incidence vectors, the parameterization in β being assumed to be of full rank.

Notice that [2] is an extension to mixed models of the structure of "linear predictors" originally restricted to fixed effects in the GLM theory: see *eg* Breslow and Clayton (1992) and Zeger *et al* (1988) for more detail about the so-called general linear mixed models (GLMM). Moreover, it will be assumed as in other studies (Hinde, 1982; Im, 1982; Foulley *et al*, 1987) that u has a multivariate normal distribution $N(0, G)$ with mean zero and variance covariance matrix G , thus resulting in what is called the Poisson-lognormal distribution (Reid, 1981; Aitchinson and Ho, 1989).

FGI showed that the mixed model structure in [2] can cope with most modelling situations arising in animal breeding such as, *eg*, sire, sire and maternal grand sire, and animal models on one hand, and direct and maternal effects on the other hand. A simple example of that is the classical animal model $\ln(\lambda_{ij}) = \mathbf{x}'_{ij} \beta + a_i + p_i$, for the j th performance of the i th female (*eg* ovulation rate of an ewe) as a function of

^a From now on referred to as FGI (Foulley, Gianola and Im); ^b from now on referred to as GAR (Gilmour, Anderson and Rae).

the usual fixed effects (*eg* herd \times year, parity), the additive genetic value a_i and a permanent environmental component p_i for female i .

According to the GLM theory, the quasi-likelihood estimating equations for $\boldsymbol{\beta}$ are obtained by differentiating the log quasi-likelihood function $Q(\boldsymbol{\mu}; \mathbf{y}, \mathbf{G})$ (\mathbf{G} being assumed known), with respect to $\boldsymbol{\beta}$, and equating the corresponding quasi-score function to zero, *viz*:

$$\frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\mu}; \mathbf{y}, \mathbf{G}) = \mathbf{D}'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}, \quad [3]$$

where

$$\boldsymbol{\mu}_{(K \times 1)} = \{\mu_k = E(Y_k) = E_u(\lambda_k)\}, \quad [4]$$

$$\mathbf{D}_{(K \times p)} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}'} = \left\{ \frac{\partial \mu_k}{\partial \beta_l} \right\}, \quad [5]$$

$$\mathbf{V}_{(K \times K)} = \text{Var}(\mathbf{Y}); \quad \mathbf{Y} = \{Y_k\}. \quad [6]$$

As clearly shown by the expression in [3], the quasi-likelihood approach only requires the specification of the marginal mean vector $\boldsymbol{\mu}$ and of the variance covariance matrix \mathbf{V} of the vector \mathbf{Y} of observations.

Given the moment generating function of the multivariate normal distribution *ie* $E(\exp(\mathbf{t}'\mathbf{Y})) = \exp[\mathbf{t}'\boldsymbol{\mu} + (\mathbf{t}'\mathbf{V}\mathbf{t}/2)]$, it can be shown that:

$$\mu_k = \exp(\mathbf{x}'_k \boldsymbol{\beta}) \exp(\mathbf{z}'_k \mathbf{G} \mathbf{z}_k / 2) \quad [7]$$

(Hinde, 1982; Zeger *et al*, 1988) and

$$\frac{\partial \mu_k}{\partial \boldsymbol{\beta}} = \mu_k \mathbf{x}_k, \quad [8]$$

$$\text{Cov}(Y_k, Y_l) = \delta_{kl} \mu_k + \mu_k \mu_l [\exp(\mathbf{z}'_k \mathbf{G} \mathbf{z}_l) - 1], \quad [9]$$

(Aitchinson and Ho, 1989), δ_{kl} being the Kronecker delta, equal to 1 if $k = l$, and 0 otherwise.

Generally models used in animal breeding yield, in the absence of inbreeding, homogeneous variances so that for any k , $\sigma_k^2 = \mathbf{z}'_k \mathbf{G} \mathbf{z}_k = \sigma^2$, and $\ln(\mu_k) = \mathbf{x}'_k \boldsymbol{\beta} + (\sigma^2/2)$. Moreover, letting $\mathbf{L}_{(K \times K)} = \{\exp(\mathbf{z}'_k \mathbf{G} \mathbf{z}_l) - 1\}$ and $\mathbf{M}_{(K \times K)} = \text{Diag}\{\mu_k\}$, variances and covariances of observations defined in [9] can be expressed in matrix notations as:

$$\mathbf{V} = \mathbf{M} + \mathbf{MLM}. \quad [10]$$

Using Fisher's scoring method based on the gradient vector $\partial Q(\cdot)/\partial \boldsymbol{\beta}$, and minus the expected value of the Hessian matrix $-E[\partial^2 Q(\cdot)/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}']$, one gets an iterative algorithm which can be expressed under the form of weighted least-squares equations:

$$\mathbf{X}'\mathbf{W}^{[t]}\mathbf{X}\hat{\boldsymbol{\beta}}^{[t+1]} = \mathbf{X}'\mathbf{W}^{[t]}\mathbf{s}^{[t]}, \quad [11]$$

where

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_K)',$$

$$\mathbf{W}_{(K \times K)} = \mathbf{M}\mathbf{V}^{-1}\mathbf{M} = (\mathbf{M}^{-1} + \mathbf{L})^{-1}, \tag{12}$$

$$\mathbf{s} = \mathbf{X}\boldsymbol{\beta} + \mathbf{M}^{-1}(\mathbf{y} - \boldsymbol{\mu}), \tag{13}$$

[*t*] being the round of iteration.

As in GAR, one may consider to approximate \mathbf{V} . This can be accomplished here using a first order Taylor expansion of $\exp(\mathbf{z}'_k \mathbf{G} \mathbf{z}_k)$ around $\mathbf{G} = \mathbf{0}$, *ie* replace $\exp(\mathbf{z}'_k \mathbf{G} \mathbf{z}_k) - 1$ in \mathbf{L} , by $\mathbf{z}'_k \mathbf{G} \mathbf{z}_k$. This approximation is likely to be realistic as long as the \mathbf{u} - part of variation remains small enough in the total variation. Doing so, \mathbf{V} in [10] becomes $\tilde{\mathbf{V}} = \mathbf{M} + \mathbf{M}\mathbf{Z}\mathbf{G}\mathbf{Z}\mathbf{M}$, with $\mathbf{Z}_{(K \times q)} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k, \dots, \mathbf{z}_k)'$ being the overall incidence matrix of \mathbf{u} . Putting this formula into the inverse of \mathbf{W} in [12], one has

$$\tilde{\mathbf{W}}^{-1} = \mathbf{M}^{-1} + \mathbf{Z}\mathbf{G}\mathbf{Z}'. \tag{14}$$

This formula exhibits the classical form $(\mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$ in the usual notation) of a variance coriance matrix of data described by a linear mixed model; this allows us to solve for $\hat{\boldsymbol{\beta}}$ in [11] using the mixed model equations of Henderson (Henderson, 1984), *ie* here with:

$$\begin{bmatrix} \mathbf{X}'\mathbf{M}^{[t]}\mathbf{X} & \mathbf{X}'\mathbf{M}^{[t]}\mathbf{Z} \\ \mathbf{Z}'\mathbf{M}^{[t]}\mathbf{X} & \mathbf{Z}'\mathbf{M}^{[t]}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}^{[t+1]} \\ \hat{\mathbf{u}}^{[t+1]} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{M}^{[t]}\mathbf{s}^{[t]} \\ \mathbf{Z}'\mathbf{M}^{[t]}\mathbf{s}^{[t]} \end{bmatrix}, \tag{15}$$

or, alternatively, defining $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{u}')'$, $\mathbf{T} = (\mathbf{X}, \mathbf{Z})$ and $\boldsymbol{\Sigma}^- = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{bmatrix}$, with

$$(\mathbf{T}'\mathbf{M}^{[t]}\mathbf{T} + \boldsymbol{\Sigma}^-)\hat{\boldsymbol{\theta}}^{[t+1]} = (\mathbf{T}'\mathbf{M}^{[t]} + \boldsymbol{\Sigma}^-)\hat{\boldsymbol{\theta}}^{[t+1]} = \mathbf{T}'\mathbf{M}^{[t]}\mathbf{s}^{[t]} \tag{16}$$

The similarity between [16] and the formula given by FGI should be noted. Actually, here $\mu_k = E_u(\lambda_k)$ replaces λ_k , thus indicating the way random effects are integrated out in the GAR procedure. It should be kept in mind that the main advantage of [15] is to provide estimates of $\boldsymbol{\beta}$ which can be computed in a similar way as with mixed model equations of Henderson (1984). These equations also imply as a by-product an estimate of \mathbf{u} which, as pointed out by Knuiman and Laird (1990) about the GAR system of equations, “has no apparent justification”.

DISCUSSION

The procedure assumes \mathbf{G} known. Arguing from the mixed model structure of equations in [15], GAR have proposed an intuitive method for estimating \mathbf{G} which mimics classical EM type-formulae for linear models. FGI advocated approximate marginal likelihood procedures based on the ingredients of their iterative system in $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$. Actually, applying such procedures would mean to use a third level of

approximation; the first one was resorted to quasi-likelihood procedures and the second one to the use of [15] instead of [11]. Alternatively, pure maximum likelihood approaches based on the EM algorithm were also envisaged by Hinde (1982), viewing \mathbf{u} as missing and using Gaussian quadratures to perform the numerical integration of the random effects. More details about methods for estimating variance components in such non-linear models can be found *eg* in Ducrocq (1990), Knuiman and Laird (1990), Smith (1990), Thompson (1990), Breslow and Clayton (1992); Solomon and Cox (1992) and Tempelman and Gianola (1993).

It must be kept in mind that the mixed model structure in [2] applies to a large variety of situations. In particular, it can be used to remove extra-Poisson variation when the fit due to identified explanatory variables remains poor. In such cases, some authors (*eg* Hinde, 1982; Breslow, 1984) have suggested to improve the fit by introducing an extra variable into the random component part of [2] *ie* by modelling the Poisson mean as $\ln(\lambda_k) = \mathbf{x}'_k \boldsymbol{\beta} + \mathbf{z}'_k \mathbf{u} + \mathbf{e}_k$. This procedure can be applied *eg* to a sire model so as to fit the fraction (3/4) of the genetic variance that is not explicitly accounted for in the model.

Finally, our approach can also be used for partitioning the observed phenotypic variance (σ_y^2) into its genetic (σ_g^2) and residual (σ_r^2) components. Let us assume that the trait is determined by a purely additive genetic model on the transformed scale, *ie*, $\ln(\lambda) = \eta + a$ where, as in [2], η is the location parameter, and $a \sim N(0, \sigma_a^2)$ is the genetic value normally distributed with mean zero and variance σ_a^2 . Following Falconer (1981), the genetic value (g) on the observed scale can be defined as the mean phenotypic value of individuals having the same genotype, *ie*, $g = E(Y|a)$. Now,

$$\sigma_y^2 = E_a [\text{Var}(Y|a)] + \text{Var}_a [E(Y|a)], \tag{17}$$

with, using [7] and [9],

$$\begin{aligned} \sigma_r^2 &= E_a [\text{Var}(Y|a)] = E_a [(Y|a)] \\ &= E(Y) = \mu \quad , \end{aligned} \tag{18}$$

$$\begin{aligned} \sigma_g^2 &= \text{Var}_a [E(Y|a)] = \sigma_y^2 - \mu \\ &= \mu^2 [\exp(\sigma_a^2) - 1] \quad , \end{aligned} \tag{19}$$

Thus, the heritability in the broad sense [$H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_r^2)$] on the observed scale can be expressed as:

$$H^2 = \frac{\mu^2 [\exp(\sigma_a^2) - 1]}{\mu + \mu^2 [\exp(\sigma_a^2) - 1]} . \tag{20}$$

The additive genetic variance on the observed scale ($\sigma_{g^*}^2$) can be defined as $\sigma_{g^*}^2 = [E(\partial g / \partial a)]^2 \sigma_a^2$ (see Dempster and Lerner, 1950; p 222), or alternatively as

$\sigma_{g^*}^2 = [\text{Cov}(g, a)]^2 / \sigma_a^2$ (see Robertson, 1950; formula 1, p 234). Now, $E(\partial g / \partial a) = \mu$, and $\text{Cov}(g, a) = \mu \sigma_a^2$. Both formulae give the same result, *ie*

$$\sigma_{g^*}^2 = \mu^2 \sigma_a^2, \quad [21]$$

with a heritability coefficient in the narrow sense ($h^2 = \sigma_{g^*}^2 / \sigma_y^2$) equal to

$$h^2 = \frac{\mu^2 \sigma_a^2}{\mu + \mu^2 [\exp(\sigma_a^2) - 1]}. \quad [22]$$

Notice that, for σ_a^2 small enough, $\sigma_g^2 \rightarrow \sigma_{g^*}^2$ so that [20] and [22] tend to $\sigma_a^2 / (\sigma_a^2 + \mu^{-1})$, which can be viewed as the expression of heritability on the linear scale, as anticipated by FGI and expected from the expression of the system in [15].

CONCLUSION

Although some other more sophisticated procedures (*eg* Bayesian treatment with Gibbs sampling; Zeger and Karim, 1991) can be envisaged to make inference about GLMM parameters, it has been shown that methods based on the quasi-likelihood or related concepts are reasonably accurate for many practical situations (Breslow and Clayton, 1992).

ACKNOWLEDGMENTS

The authors are grateful to V Ducrocq, M Perez Enciso and one anonymous reviewer for their helpful comments and criticisms on previous versions of this manuscript.

REFERENCES

- Aitchinson J, Ho CH (1989) The multivariate Poisson log normal distribution. *Biometrika* 76, 643-653
- Breslow NE (1984) Extra-Poisson variation in log-linear models. *Appl Stat* 33, 38-44
- Breslow NE, Clayton DG (1992) *Approximate Inference in Generalized Linear Mixed Models*. Tech Rep No 106, Univ Washington, Seattle
- Dempster ER, Lerner IM (1950) Heritability of threshold characters. *Genetics* 35, 212-236
- Ducrocq V, Quaas RL, Pollak EJ, Casella G (1988) Length of productive life of dairy cows. 1. Justification of a Weibull model. *J Dairy Sci* 71, 2543-2553
- Ducrocq V (1990) Estimation of genetic parameters arising in nonlinear models. *In: 4th World Congr Genet Appl Livestock Prod*, (Hill WG, Thompson R, Wooliams JA, eds) Edinburgh, 23-27 July 1990, vol 13, 419-428
- Falconer DS (1981) *Introduction to Quantitative Genetics*. Longman, London, 2nd edn

- Foulley JL, Gianola D, Im S (1987) Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. *Theor Appl Genet* 73, 870-877
- Foulley JL, Manfredi E (1991) Approches statistiques de l'évaluation génétique des reproducteurs pour des caractères binaires à seuils. *Genet Sel Evol* 23, 309-338
- Gianola D, Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. *Genet Sel Evol* 15, 201-224
- Gilmour A, Anderson RD, Rae A (1985) The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72, 593-599
- Harville DA, Mee RW (1984) A mixed model procedure for analyzing ordered categorical data. *Biometrics* 40, 393-408
- Henderson CR (1984) Applications of linear models in animal breeding. Univ Guelph, Guelph, Ont
- Hinde JP (1982) Compound Poisson regression models. In: *GLIM 82* (Gilchrist R, ed) Springer Verlag, NY 109-121
- Im S (1982) Contribution à l'étude des tables de contingence à paramètres aléatoires: utilisation en biométrie. Thèse 3^e cycle, Université Paul Sabatier, Toulouse
- Knuiman M, Laird N (1990) Parameter estimation in variance component models for binary response data. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer-Verlag, Heidelberg, 177-189
- McCullagh P, Nelder J (1989) *Generalized Linear Models*. Chapman and Hall, London, 2nd edn
- Reid DD (1981) The Poisson lognormal distribution and its use as a model of plankton aggregation. In: *Statistical Distributions in Scientific Work* (Taillie C, Patil GP, Baldessari, eds) Reidel, Dordrecht, Holland, 6, 303-316
- Robertson A (1950) Proof that the additive heritability on the P scale is given by the expression $\bar{z}^2 h_x^2 / \bar{p}\bar{q}$. *Genetics* 35, 234-236
- Smith SP (1990) Estimation of genetic parameters in non-linear models. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer-Verlag, Heidelberg, 190-206
- Solomon PJ, Cox DR (1992) Non linear component of variance models. *Biometrika* 79, 1-11
- Tavernier A (1991) Genetic evaluation of horses based on ranks in competitions. *Genet Sel Evol* 23, 159-173
- Tempelman RJ, Gianola D (1993) Marginal maximum likelihood estimation of variance components in Poisson mixed models using Laplace integration. *Genet Sel Evol* (submitted)
- Thompson R (1990) Generalized linear models and applications to animal breeding. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer-Verlag, Heidelberg, 312-328
- Zeger SL, Karim MR (1991) Generalized linear models with random effects; a Gibbs sampling approach. *J Am Stat Assoc* 86, 79-86
- Zeger SL, Liang KY, Albert PS (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44, 1049-1060