Int. J. Nonlinear Anal. Appl. 13 (2022) 2, 2897–2907
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.24702.2805

Using a Laplace approximation to estimate the genetic variance components in animal models

Fatemeh Pakbaz[a], Alireza Nematollahi[a], Fatemeh Hosseini[b]

[a] *Department of Statistics, Shiraz University, Shiraz, Iran*

[b] *Department of Statistics, Semnan University, Semnan, Iran*

(Communicated by Haydar Akca)

Abstract

Animal model is a type of mixed-effects model, where covariance among data points comes from genetic and environmental effects. In this paper, the multivariate normal distribution are assumed for the genetic random effects. A new approximate maximum likelihood method is proposed to obtain the estimates of the genetic variance components and heritability. The effectiveness of the proposed method is illustrated through a simulation study.

Keywords: Maximum Likelihood, Generalized Linear Mixed Model, Heritability
2020 MSC: 92D50, 62J12

---

1 Introduction

Animal models are used to answer various scientific questions, from basic science to the development and evaluation of new vaccines or treatments. These models are kind of linear mixed models popularized by [18]. Linear Mixed Models are a kind of generalized Linear Mixed Models (GLMMs). In these models, the response variable is assumed to follow an exponential family distribution, ([3]). The likelihood function of GLMMs cannot usually be given in a closed form and the maximum likelihood estimations are very challenging. For this problem, there are numerical ways to obtain maximize likelihood estimates of the parameters in the GLMMs, see e.g. [28], [8], [10], [22], [23] and [1]. Animal model is a type of the GLMM with the genetic random effects and is used to model the genetic responses, see e.g. [15], [16], [17], [26], [2], [27], [25] and [19].

Whether changes in an animal trait are attributed to environmental or biological reasons is a fundamental question in biology. In the animal model, all known relationships from a pedigree are taken as the genetic random effects and partitioned observed variance into additive genetic variance and other (environmental) sources of variance, [16]. Obtaining an estimate of these variance components and the heritability component is one of the most important goals in studying the animal model. Heritability value describes how much of a change in a trait is due to genetic factors. The heritability is defined as $h^2 = \frac{V_G}{V_P}$, where V_G is additive genetic effects variance and $V_P = V_G + V_R$ is the phenotypic variance that is equal to the sum of genetic and environmental variance in the animal models. The restricted maximum likelihood (REML) method is useful for inferences about variance components in the animal

---

*Corresponding author
Email addresses: `f_pakbaz@yahoo.com` (Fatemeh Pakbaz), `ar.nematollahi@shirazu.ac.ir` (Alireza Nematollahi), `fatemeh.hoseini@semnan.ac.ir` (Fatemeh Hosseini)

Received: September 2021 *Accepted:* April 2022

model. In REML estimation, the maximum likelihood method is applied to the linear transformation of the data $\mathbf{K}'\mathbf{y}$, where $\mathbf{K}$ is a matrix of full rank for which $\mathbf{K}'\mathbf{Z} = \mathbf{0}$ for fixed effect matrix $\mathbf{Z}$. There are many articles on the application of this method in animal models ([11], [24] [5]). The likelihood approach in the animal model often involves calculating intractable multidimensional integrals because the integration dimension in likelihood function is equal to the number of genetic random effects. Consequently, it is intractable to obtain the maximum likelihood estimates (MLE) by directly maximizing the likelihood function and an extension of REML to the hierarchical animal models is often hampered by analytically intractable integrals. The maximization algorithms for the REML such as EM is generally not recommended when the dimensionality of the integral is high. Laplace approximations are usually used to approximate high-dimensional integrals in complex models ([20]), so these approximations seem to be helpful for an animal model. In this paper, a higher-order Laplace approximation is proposed to inference and find the MLE of the variance components and the heritabilit in the animal model.

This paper is organized as follows: In Section 2, the notations and modified Laplace approximation are defined. The model and the proposed method of approximate likelihood inference are described in Section 3. Section 4 shows the results of a simulation study.

2 Notation and approach

In this article, we use the notation proposed by [21]. Let $\mathbf{x} \in R^n$ and x_i be the ith component of $\mathbf{x}$, $f(\mathbf{x})$ be a real function and its derivative shown as $f_i = \dfrac{\partial f}{\partial x_i}$ and $f_{ij} = \dfrac{\partial^2 f}{\partial x_i \partial x_j}$. The Hessian matrix is shown by f_{xx} and f^{ij} is the (i,j) element of f_{xx}^{-1} matrix. If $p = i_1, i_2, \cdots, i_\nu$, then $f_p = \dfrac{\partial^\nu f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_\nu}}$. The calculated value of the function f in $\hat{u}$ is denoted by$\hat{f}$.

2.1 Modified Laplace approximation

According to the Laplace approximation, we have

$$\int_{R^n} e^{-Mg(x)} dx \simeq \left(\frac{2\pi}{M}\right)^n \frac{e^{-Mg(x_0)}}{|\hat{g}_{xx}|^{1/2}}, \; as \; M \to \infty, \tag{2.1}$$

in which g has a minimum at x_0 and $\hat{g}_{xx}$ is the Hessian matrix in x_0. The idea of the Laplace approximation is relevant to using Taylor expansion around the minimum point of function g instead of g in. [21] developed the formal Laplace approximation as follows:

$$I = \int_{R^p} \exp\{-g(\boldsymbol{x})\} d\boldsymbol{x}, \tag{2.2}$$

when g has derivatives of all orders and is unimodal with a unique minimum at $\hat{u} = \mathbf{0}$ and $g = O(n)$. With considering $g(\boldsymbol{x}) = g(\mathbf{0}) + \frac{1}{2}\hat{g}_{i_1 i_2} x^{i_1} x^{i_2} + \frac{1}{3!}\hat{g}_{i_1 i_2 i_3} x^{i_1} x^{i_2} x^{i_3} + \cdots$ be Taylor expansion of g around zero, we have

$$\begin{aligned} I &= \int_{R^p} \exp\{-g(\mathbf{0})\} \exp\{-\frac{1}{2}\hat{g}_{i_1 i_2} x^{i_1} x^{i_2}\} \exp\{\frac{1}{3!}\hat{g}_{i_1 i_2 i_3} x^{i_1} x^{i_2} x^{i_3} + \cdots\} d\boldsymbol{x} \\ &= e^{-\hat{g}} \det |\frac{\hat{g}_{\boldsymbol{xx}}}{2\pi}|^{-1/2} E\left[\exp\{-\frac{1}{3!}\hat{g}_{i_1 i_2 i_3} X^{i_1} X^{i_2} X^{i_3} + \cdots\}\right] \\ &= e^{-\hat{g}} \det |\frac{\hat{g}_{\boldsymbol{xx}}}{2\pi}|^{-1/2} E\left[1 - \frac{1}{3!}\hat{g}_{i_1 i_2 i_3} X^{i_1} X^{i_2} X^{i_3} + \cdots\right], \end{aligned} \tag{2.3}$$

where $\boldsymbol{X}$ is the multivariate normal random vector with mean $\mathbf{0}$ and variance and covariance matrix $\hat{g}_{\boldsymbol{xx}}^{-1}$. Then all cumulants of odd order are zero and so

$$\begin{aligned} E\left[X^{i_1} X^{i_2} \cdots X^{i_{2m}}\right] &= \sum_{q_1|\cdots|q_{2m}=Q\in \boldsymbol{P}^2_{2m}} \prod_{q=\{i,j\}\in Q} E[X^i X^j] \\ &= \sum_{q_1|\cdots|q_{2m}=Q\in \boldsymbol{P}^2_{2m}} \prod_{q=\{i,j\}\in Q} cov(X^i X^j) \\ &= \sum_{q_1|\cdots|q_{2m}=Q\in \boldsymbol{P}^2_{2m}} \prod_{q=\{i,j\}\in Q} g^{ij}. \end{aligned} \tag{2.4}$$

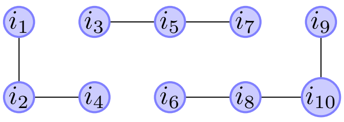
So, we can write Eq. (2.3) as follows:

$$I = e^{-\hat{g}} \det |\frac{\hat{g}_{\boldsymbol{xx}}}{2\pi}|^{-1/2} \sum_{m=0}^{\infty} \sum_{P,Q} \frac{(-1)^t}{(2m)!} \hat{g}_{p_1} \cdots \hat{g}_{p_t} \hat{g}^{q_1} \cdots \hat{g}^{q_m}, \tag{2.5}$$

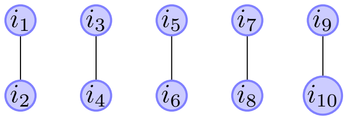
and its logarithm is equal to

$$\log I = -\hat{g} - \frac{1}{2} \log \left| \frac{\hat{g}_{\boldsymbol{xx}}}{2\pi} \right| + \sum_{m=1}^{\infty} \sum_{\substack{P,Q \\ P \vee Q = 1}} \frac{(-1)^t}{(2m)!} \hat{g}_{p_1} \cdots \hat{g}_{p_t} \hat{g}^{q_1} \cdots \hat{g}^{q_m}, \tag{2.6}$$

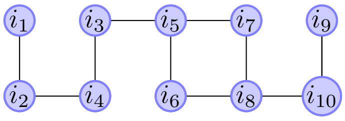
where P and Q are partition of $2m$ indices such that $P = p_1 | \cdots | p_t$ consists of t blocks, each of size 3 or more. $Q = q_1 | \cdots | q_m$ consists of m block each of size 2 and the graph of $P \vee Q$ that means union of P and Q is a connected graph. For example if P be $i_1 i_2 i_4 | i_3 i_5 i_7 | i_6 i_8 i_{10} i_9$ and Q be $i_1 i_2 | i_3 i_4 | i_5 i_6 | i_7 i_8 | i_9 i_{10}$, we can draw P, Q and $P \vee Q$ as follows



P



Q



$P \vee Q$

Often the main difficulty in applying the above formula is finding the minimum of g when the dimension of $\mathbf{x}$ is large. [21] suggested to consider function $g(\mathbf{x}) = h(\mathbf{x}) + \epsilon(\mathbf{x})$ such that $h(\mathbf{x})$ is a quadratic approximation to g in around of the minimum and $\epsilon(\mathbf{x})$ and its first and second derivative be small in the neighbourhood of that point. In this case, Eq. (2.6) can be written as

$$\log I = -\hat{g} - \frac{1}{2} \log \{ \det \left(\frac{h_{\mathbf{xx}}}{2\pi} \right) \} + \sum_{m} \frac{1}{(2m)!} \sum_{\substack{P,Q \\ P \vee Q = 1}} (-1)^t \epsilon_{p_1} \cdots \epsilon_{p_t} h^{q_1} \cdots h^{q_m}. \tag{2.7}$$

This can be a good idea for [4] to provide a generalization of Laplace approximation with a bit of change in assumptions. They suppose that $h(\mathbf{x}) = O(n)$ has a minimum at $\mathbf{0}$ and $f(\boldsymbol{x})$ and its derivatives are $O(1)$. Taylor expansion of h around $\mathbf{0}$ gives

$$h(\mathbf{x}) = \hat{h} + \frac{1}{2!} x^{i_1} x^{i_2} \hat{h}_{i_1 i_2} + \frac{1}{3!} x^{i_1} x^{i_2} x^{i_3} \hat{h}_{i_1 i_2 i_3} + \cdots, \tag{2.8}$$

such that sums are over all indices from 1 to n. $f(\mathbf{x})$ also has a Taylor series around $\mathbf{0}$,

$$f(\mathbf{x}) = \hat{f} + x^{j_1} \hat{f}_{j_1} + \frac{1}{2!} x^{j_1} x^{j_2} \hat{f}_{j_1 j_2} + \frac{1}{3!} x^{j_1} x^{j_2} x^{j_3} \hat{f}_{j_1 j_2 j_3} + \cdots. \tag{2.9}$$

Now, we have

$$\begin{aligned}
I &= \int_{R^p} \exp\{-g(\mathbf{x})\} d\mathbf{x} \\
&= \int_{R^p} \exp\{-\left(h(\mathbf{x}) + f(\mathbf{x})\right)\} d\mathbf{x} \\
&= e^{-\hat{h}} e^{-\hat{f}} \int e^{-\frac{1}{2}\mathbf{x}^T \hat{h}_{\mathbf{xx}} \mathbf{x}} \exp\{-\frac{1}{3!}\hat{h}_{[i_1 i_2 i_3]} x^{i_1} x^{i_2} x^{i_3} - \frac{1}{4!}\hat{h}_{[i_1 i_2 i_3]} x^{i_1} x^{i_2} x^{i_3} x^{i_4} - \cdots\} \\
&\times \exp\{-\hat{f}_{j_1} x^{j_1} - \frac{1}{2}\hat{f}_{j_1 j_2} x^{j_1} x^{j_2} - \cdots\} dx \\
&= e^{-\hat{h}} e^{-\hat{f}} \left|\frac{\hat{h}_{\mathbf{xx}}}{2\pi}\right|^{-1/2} E\left[\left(1 - \frac{1}{3!}\hat{h}_{[i_1 i_2 i_3]} X^{i_1} X^{i_2} X^{i_3} - \cdots\right)\left(1 - \hat{f}_{[j_1]} X^{j_1} - \frac{1}{2!}\hat{f}_{[j_1 j_2]} X^{j_1} X^{j_2} - \cdots\right)\right],
\end{aligned}$$

where $X \sim N(0, h_{\mathbf{xx}})$. From Eq. (2.5)

$$I = e^{-\hat{h}} e^{-\hat{f}} \left|\frac{\hat{h}_{\mathbf{xx}}}{2\pi}\right|^{-1/2} \sum_{r \in \{0,3,4,\cdots\}} \sum_{s=0}^{\infty} (-1)^r \frac{1}{s!\, r!} \hat{h}_{[i_1 \cdots i_r]} \hat{f}_{[j_1 \cdots j_s]} E\left[X^{i_1} \cdots X^{i_r} X^{j_1} \cdots X^{j_s}\right],$$

therefore from Eqs. (2.4) and (2.7) we can write

$$\log I = -\hat{g} - \hat{f} - \frac{1}{2} \log |\frac{\hat{g}_{\boldsymbol{xx}}}{2\pi}| + \sum_{m=1}^{\infty} \frac{1}{(2m)!} \sum_{P,Q}^{P \vee Q = 1} \chi_{p_1} \cdots \chi_{p_t} . g^{q_1} \cdots g^{q_m}, \tag{2.10}$$

such that in above formula $\chi_{p_i} = \hat{f}_{p_i}$ for $p_i < 3$ and $\chi_{p_i} = \hat{f}_{p_i} - \hat{g}_{p_i}$ for $p_i \geqslant 3$.

3 Model

Let $\boldsymbol{y} = (y_1, y_2, \cdots y_n)^T$ be the vector of n observations, then the classic form of the animal model can be written as follows

$$\boldsymbol{y} = \boldsymbol{Z}^T \boldsymbol{\beta} + \boldsymbol{u} + \boldsymbol{\epsilon}, \tag{3.1}$$

where $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n)$, $\boldsymbol{z}_i = (z_{i1}, z_{i2}, \cdots, z_{iq+1})^T$ are covariates and $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_q)^T$ is vector of fixed effects. The vector of random effects $\boldsymbol{u}$ and $\boldsymbol{\epsilon}$ are the additive genetic and the environmental origin random effects, respectively. We called $\boldsymbol{u}$ as hidden variable and $\boldsymbol{\epsilon}$ as residuals that are independent. Hidden variables are correlated with a relationship matrix $\boldsymbol{A}$, but residuals are independent, and $\boldsymbol{I}$ is $n \times n$ identity matrix. In this study we assume that $\boldsymbol{u} \sim N\left(\boldsymbol{0}, \boldsymbol{A}\sigma^2_{\boldsymbol{u}}\right)$ and $\boldsymbol{\epsilon} \sim N\left(\boldsymbol{0}, \boldsymbol{I}\sigma^2_{\boldsymbol{e}}\right)$. For the animal model with normal observations and variance-covariance matrix $\Sigma = \begin{pmatrix} \sigma^2_{\boldsymbol{u}}\boldsymbol{A} & 0 \\ 0 & \sigma^2_{\boldsymbol{e}}\boldsymbol{I} \end{pmatrix}$, the direct heritability $(h^2_{\boldsymbol{u}})$ is defined as follows

$$h^2_{\boldsymbol{u}} = \frac{\sigma^2_{\boldsymbol{u}}}{\sigma^2_{\boldsymbol{u}} + \sigma^2_{\boldsymbol{e}}},$$

which is in proportion to the phenotypic variance. The parameters of the model (3.1) are gathered in $\boldsymbol{\Theta} = (\boldsymbol{\beta}^T, \sigma^2_{\boldsymbol{u}}, \sigma^2_{\boldsymbol{e}})^T$. This study aims to estimate the unknown parameters $\boldsymbol{\Theta}$ using the approximate likelihood function with a modified Laplace method.
We know that

$$y_i | u_i \sim N(\boldsymbol{z}_i^T \boldsymbol{\beta} + u_i, {\sigma^2}_{\boldsymbol{e}}), \tag{3.2}$$

then $f_{\boldsymbol{y}|\boldsymbol{u}} = \prod_{i=1}^n f_{y_i|u_i}$ and the likelihood function can written as follows

$$\begin{aligned}
L(\boldsymbol{\Theta}) &= \int \prod_{i=1}^n f_{y_i|u_i} . f_{\boldsymbol{u}}(\boldsymbol{u}) d\boldsymbol{u} \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2_{\boldsymbol{e}}}^n} \exp\left\{\frac{-1}{2\sigma^2_{\boldsymbol{e}}} \sum_{i=1}^n \left(y_i - \boldsymbol{z}_i^T \boldsymbol{\beta} - u_i\right)^2\right\} \\
&\times \frac{1}{\sqrt{2\pi\sigma^2_{\boldsymbol{u}}}^n \det \boldsymbol{A}^{1/2}} \exp\left\{\frac{-1}{2\sigma^2_{\boldsymbol{u}}} \boldsymbol{u}^T \boldsymbol{A}^{-1} \boldsymbol{u}\right\} d\boldsymbol{u}.
\end{aligned}$$

Let $\boldsymbol{A}^{-1} = [a_{ij}]_{n\times n}$, then we have

$$\begin{aligned} L(\boldsymbol{\Theta}) &= \frac{1}{(2\pi)^n \left(\sigma^2_{\boldsymbol{u}}\sigma^2_{\boldsymbol{e}}\right)^{n/2} \det \boldsymbol{A}^{1/2}} \\ &\times \int \exp\left\{\frac{-1}{2\sigma^2_{\boldsymbol{e}}}\sum_{i=1}^{n}\left(y_i - \boldsymbol{z}_i^T\boldsymbol{\beta} - u_i\right)^2 + \frac{a_{ii}}{\sigma^2_{\boldsymbol{u}}}u_i\right\} . \exp\left\{\frac{-1}{2\sigma^2_{\boldsymbol{u}}}\sum_{i\neq j} a_{ij}u_iu_j\right\} d\boldsymbol{u}. \end{aligned} \tag{3.3}$$

Here, we reparametrize the residual variance as $\sigma^2_{\boldsymbol{e}} = k\sigma^2_{\boldsymbol{u}}$, where k is equal to any positive value. Then according to the heritability formula $h^2_{\boldsymbol{u}} = \dfrac{\sigma^2_{\boldsymbol{u}}}{\sigma^2_{\boldsymbol{u}} + \sigma^2_{\boldsymbol{e}}} = \dfrac{1}{1+k}$, and the heritability only depend on k. With this selection, the likelihood function (3.3) can be written as follows

$$\begin{aligned} L(\boldsymbol{\Theta}) &= \frac{1}{(2\pi\sigma^2{}_{\boldsymbol{u}})^n (k)^{n/2} \det \boldsymbol{A}^{1/2}} \\ &\times \int \exp\left\{\frac{-1}{2\sigma^2{}_{\boldsymbol{u}}}\sum_{i=1}^{n}\frac{\left(y_i - \boldsymbol{z}_i^T\boldsymbol{\beta} - u_i\right)^2}{k} + a_{ii}u_i\right\} \exp\left\{\frac{-1}{2\sigma^2_{\boldsymbol{u}}}\sum_{i\neq j} a_{ij}u_iu_j\right\} d\boldsymbol{u}. \end{aligned} \tag{3.4}$$

By using the modified Laplace formula for approximating the integral in Eq. (3.4), we have

$$\begin{aligned} l(\boldsymbol{\Theta}) = \log\left(L(\boldsymbol{\Theta})\right) = -n\log\left(2\pi\sigma^2{}_{\boldsymbol{u}}\right) - \frac{n}{2}\log(k) - \frac{1}{2}\log\left(\det(\boldsymbol{A})\right) \\ + \log\int \exp\left\{\frac{-1}{2\sigma^2{}_{\boldsymbol{u}}}\sum_{i=1}^{n}\frac{\left(y_i - \boldsymbol{z}_i^T\boldsymbol{\beta} - u_i\right)^2}{k} + a_{ii}u_i^2\right\} \\ . \exp\left\{\frac{-1}{2\sigma^2{}_{\boldsymbol{u}}}\sum_{i\neq j} a_{ij}u_iu_j\right\} d\boldsymbol{u}. \end{aligned} \tag{3.5}$$

and let
$\xi(\boldsymbol{u}) = \dfrac{1}{2\sigma^2_{\boldsymbol{u}}}\sum_{i=1}^{n}\dfrac{\left(y_i - \boldsymbol{z}_i^T\boldsymbol{\beta} - u_i\right)^2}{k} + a_{ii}u_i^2$ and $\tau(\boldsymbol{u}) = \dfrac{1}{2\sigma^2_{\boldsymbol{u}}}\sum_{i\neq j} a_{ij}u_iu_j$, then we can write Eq. (3.5) as follows

$$\begin{aligned} l(\boldsymbol{\Theta}) = -n\log\left(2\pi\sigma^2_{\boldsymbol{u}}\right) - \frac{n}{2}\log(k) - \frac{1}{2}\log\left(\det(\boldsymbol{A})\right) \\ + \log\int \exp\left\{-\xi(\boldsymbol{u})\right\} . \exp\left\{-\tau(\boldsymbol{u})\right\} d\boldsymbol{u}. \end{aligned} \tag{3.6}$$

Now, consider the following divisions

$$\begin{aligned} &\xi_i = \frac{dh}{du_i} = \frac{1}{k\sigma^2_{\boldsymbol{u}}}\left(-(y_i - \boldsymbol{z}_i^T\boldsymbol{\beta} - u_i) + ka_{ii}u_i\right), \\ &\xi_{ii} = \frac{dh_i}{du_i} = \frac{1}{k\sigma^2_{\boldsymbol{u}}}(1 + ka_{ii}), \\ &\xi_{ij} = 0, \qquad if\ i \neq j, \\ &\det h_{\boldsymbol{uu}} = \prod_{i=1}^{n} h_{ii} = \frac{1}{k\sigma^2_{\boldsymbol{u}}{}^n}\prod_{i=1}^{n}\frac{1}{1+ka_{ii}}, \\ &\tau_i = \frac{d\tau}{du_i} = \frac{1}{\sigma^2_{\boldsymbol{u}}}\sum_{j\neq i} a_{ij}u_j, \\ &\tau_{ii} = \frac{d\tau_i}{du_i} = 0, \\ &\tau_{ij} = \frac{d\tau_i}{du_j} = \frac{a_{ij}}{\sigma^2_{\boldsymbol{u}}}, \qquad if\ i \neq j, \end{aligned}$$

and the maximum point of the function h at $\hat{\boldsymbol{u}}$ such that $\hat{u}_i = \dfrac{y_i - \boldsymbol{z}_i^T \boldsymbol{\beta}}{1 + k a_{ii}}$, we can write Eq. (3.6) as follows

$$\begin{aligned}
l(\boldsymbol{\Theta}) &= -n \log\left(2\pi\sigma_{\boldsymbol{u}}^2\right) - \frac{n}{2}\log\left(k\right) - \frac{1}{2}\log\left(\det\left(\boldsymbol{A}\right)\right) - \hat{\xi} - \hat{\tau} \\
&\quad - \frac{1}{2}\left(-n\log\left(k\sigma_{\boldsymbol{u}}^2\right) - \sum_{i=1}^{n}\log\left(1 + k a_{ii}\right)\right) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\tau_i\tau_j - \tau_{ij}\right)\xi^{ij} + \cdots \\
&= -n\log\left(2\pi\right) - n\log\left(\sigma_{\boldsymbol{u}}\right) - \frac{1}{2}\log\left(\det\left(\boldsymbol{A}\right)\right) - \hat{\xi} - \hat{\tau} \\
&\quad + \frac{1}{2}\sum_{i=1}^{n}\log\left(1 + k a_{ii}\right) + \frac{1}{2}\sum_{i=1}^{n}{\tau_i}^2\xi^{ii} + \cdots .
\end{aligned}$$

The estimates of the variance components parameters can be obtained by differentiating the function $l(\boldsymbol{\Theta})$ with respect to the $\sigma_{\boldsymbol{u}}$ and k as shown below

$$\begin{aligned}
\frac{dl(\boldsymbol{\Theta})}{d\sigma_{\boldsymbol{u}}} &= -\frac{n}{\sigma_{\boldsymbol{u}}} - \frac{2}{\sigma_{\boldsymbol{u}}^3}\left(\sigma_{\boldsymbol{u}}^2\left(-\hat{\xi} - \hat{\tau} + \frac{1}{2}\sum_{i=1}^{n}{\tau_i}^2\xi^{ii}\right)\right) = 0, \\
\frac{dl(\boldsymbol{\Theta})}{dk} &= \frac{d\left(-\hat{\xi} - \hat{\tau} + \frac{1}{2}\sum_{i=1}^{n}{\tau_i}^2\xi^{ii}\right)}{d\boldsymbol{u}}\frac{d\boldsymbol{u}}{dk} + \sum_{i=1}^{n}\frac{a_{ii}}{1 + k a_{ii}} = 0.
\end{aligned} \tag{3.7}$$

4 Simulation

In this Section, we present one simulation study to illustrate the proposed methodology in the animal model and this method is compared with the REML approach.

For this simulation study, the pedigree file of [27] has been used. This data set consists of the records of 1309 for animals. It is used to create a relationship matrix of 1084 animals for which the effects set is available, according to [27]. The codes related to the production of the relationship matrix A are written in Mathematica software and inverse matrix calculations are also performed. R programming was used for the analyses with the modified Laplace approximation method. The results are compared with estimated values with the REML method from R package 'MM4LMM".

In the following, we generate 1084 random samples based on the pedigree. Different values for additive genetic variance (σ_u^2) and environmental (σ_e^2) are considered to examine the performance of the modified Laplace approximation method in various heritability levels (high, medium, and low heritability). First, we generate the random effects $\boldsymbol{u}$ from the normal distribution $\boldsymbol{u} \sim N_{1084}(\boldsymbol{0}, \sigma_u^2 A)$.

The environmental random effects ($\boldsymbol{\epsilon}$) are generated from $N_{1084}(0, \sigma_e^2 I)$. Also, the parameters β_0 and β_1 are assumed to be the fixed effects coefficients. Then, the observations $\boldsymbol{y}$ are created by the proposed animal model as $y_i = \boldsymbol{z}_i^T\boldsymbol{\beta} + u_i + \epsilon_i$.

Keeping the design fixed, the above data generation scheme are carried out for 100 data sets in each case. Now we can estimate the averages of mean square error (MSE) and averages of estimates of the parameters by using the proposed approximate method. Each simulated data set is analyzed by the modified Laplace approximation approach under the assumptions of the random effects are normal. By maximizing the approximate likelihood function

$$L(\boldsymbol{\theta}) = -n\log(\sigma_u) - \hat{\xi} - \hat{\tau} + \frac{1}{2}\sum_{i=1}^{n}\log(1 + a_{ii}) + \frac{1}{2}\sum_{i=1}^{n}\tau_i^2\xi^{ii},$$

The approximate values are obtained, and the simulation results are reported in table 1. The proposed approximate maximum likelihood technique is used for 1084×100 simulated datasets with different values for the variance component parameters. Furthermore, in every approach, the time spent calculating has been measured.

The results are summarized in Table 1 for a low heritability level. In this situation, although the estimation of $\sigma_{\mathbf{u}}^2$ in REML seems good, but significant bias for $\sigma_{\mathbf{e}}^2$ caused a wrong estimation for heritability (h^2). However, the results for the variance components seems not good, but the heritability (h^2) estimation has more appropriate performance in modified high Laplace approximation (Table1). The fitted values of responses are calculated, and the Mean Square

Table 1: Simulation study based on 100 data set for low heritability ($h^2 = 0.138$)

		Modified Laplace Approximation Method			REML Method		
Parameter	Real Value	Ave. Est.	Ave. MSE	Ave. Bias	Ave. Est.	Ave. MSE	Ave. Bias
σ_u^2	0.8	0.149	0.425	0.651	0.924	0.026	-0.124
σ_e^2	5	0.828	17.406	4.172	0.059	24.421	4.941
β_0	1	0.989	0.015	0.011	0.989	0.009	0.011
β_1	1	0.999	0.003	0.001	1.002	0.002	-0.002
h^2	0.138	0.153	0.001	-0.009	0.938	0.644	-0.800
$MSEP$			1.731			3.671	
$Time$			66”			394”	

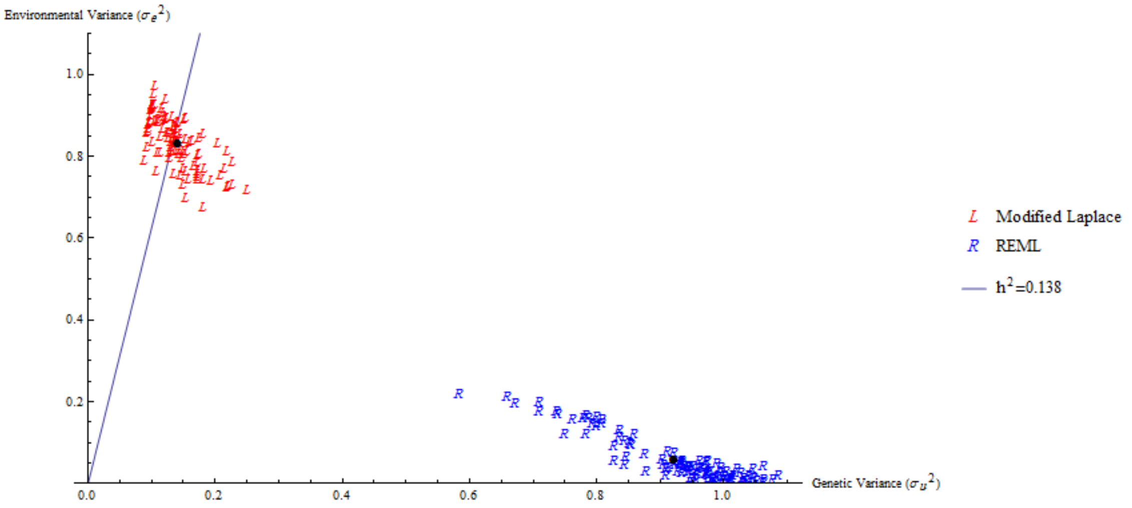


Figure 1: The high Laplace approximation and the REML results for the maximum likelihood ($\sigma^2_{\mathbf{u}} = 0.8$ and $\sigma^2_{\mathbf{u}} = 5$)

Table 2: Simulation study based on 100 data set for medium heritability ($h^2 = 0.231$)

		Modified Laplace Approximation Method			REML Method		
Parameter	Real Value	Ave. Est.	Ave. MSE	Ave. Bias	Ave. Est.	Ave. MSE	Ave. Bias
σ_u^2	0.9	0.231	0.451	0.0.679	0.976	0.032	-0.076
σ_e^2	3	1.021	3.923	1.979	0.0226	7.706	2.774
β_0	1	1.006	0.025	-0.006	1.004	0.016	-0.004
β_1	1	0.991	0.007	0.009	0.993	0.004	0.007
h^2	0.231	0.185	0.004	0.046	0.808	0.342	-0.577
$MSEP$			0.478			4.030	
$Time$			63”			394”	

Table 3: Simulation study based on 100 data set for medium heritability ($h^2 = 0.4$)

		Modified Laplace Approximation Method			REML Method		
Parameter	Real Value	Ave. Est.	Ave. MSE	Ave. Bias	Ave. Est.	Ave. MSE	Ave. Bias
σ_u^2	4	2.0722	3.866	1.928	3.970	0.284	0.134
σ_e^2	6	5.266	0.738	0.734	2.702	11.021	3.298
β_0	1	1.016	0.101	-0.016	0.995	0.080	0.005
β_1	0.8	0.794	0.027	0.006	0.802	0.023	-0.002
h^2	0.4	0.282	0.016	0.118	0.594	0.041	-0.194
$MSEP$			8.683			10.572	
$Time$			39”			396”	

Error of Predictions (MSEPs) are obtained from $MSEP = n^{-1}\sum(y-\hat{y})^2$ and reported in table 1. These results show that the averages of the MSE and the bias are small and the accuracy of results is good. The results of the modified high Laplace and REML methods can be seen in Figure 1. Genetic variance and environmental variance are the axes of the shape. Heritability is also indicated by a line with a positive slope.
In medium heritability ($0.15 < h^2 < 0.4$) again, the modified the high Laplace approximation method works better than the REML to estimate h^2(Table 2 and 3). Figures 2 and 3 show how the results are positioned for variance components. In high heritability situation ($h^2 > 0.4$), REML method performs better than high Laplace approximation (Table 4 and Figure 4).

Computing the running time of the modified Laplace approximation approach for 1084 × 100 iterations took less than 180 seconds, and the computation time is significantly reduced by this method.

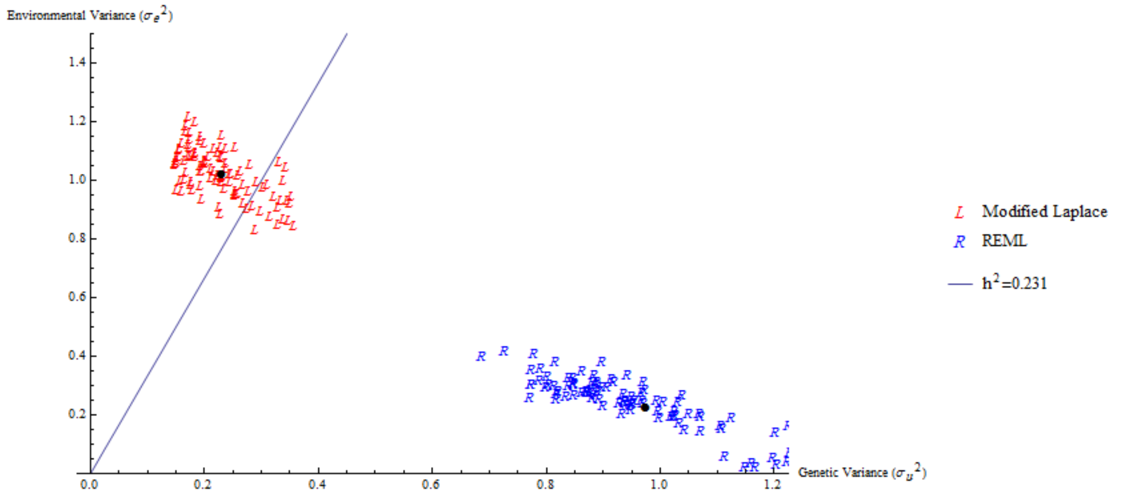


Figure 2: The high Laplace approximation and the REML results for the maximum likelihood ($\sigma^2_{\mathbf{u}} = 0.9$ and $\sigma^2_{\mathbf{e}} = 3$)

5 Discussion and Conclusions

In animal models and genetic studies, heritability is the most important consideration in determining appropriate animal evaluation methods, selection methods, and mating systems. The estimates of heritability for most traits is

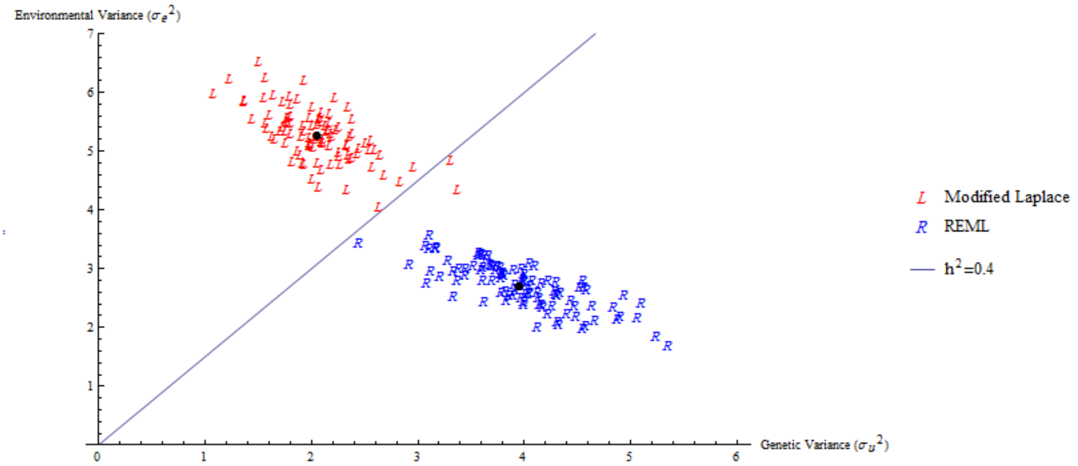


Figure 3: The high Laplace approximation and the REML results for the maximum likelihood ($\sigma^2_{\mathbf{u}} = 4$ and $\sigma^2_{\mathbf{u}} = 6$)

Table 4: Simulation study based on 100 data set for high heritability ($h^2 = 0.556$)

		Modified Laplace Approximation Method			REML Method		
Parameter	Real Value	Ave. Est.	Ave. MSE	Ave. Bias	Ave. Est.	Ave. MSE	Ave. Bias
σ_u^2	5	4.564	0.606	0.436	5.057	0.795	-0.057
σ_e^2	4	8.140	17.644	-4.140	6.178	5.280	-2.178
β_0	1	1.038	0.209	-0.038	0.983	0.161	0.017
β_1	1	0.992	0.050	0.008	1.017	0.043	-0.017
h^2	0.556	0.359	0.041	0.197	0.449	0.016	0.107
$MSEP$			17.806			15.555	
$Time$			27”			394”	

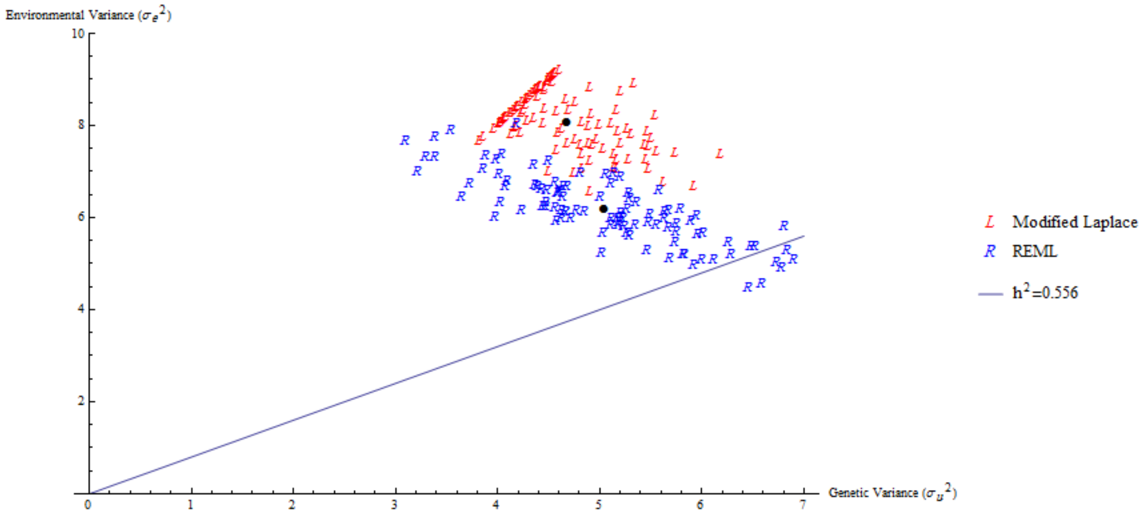


Figure 4: The high Laplace approximation and the REML results for the maximum likelihood ($\sigma^2_{\mathbf{u}} = 5$ and $\sigma^2_{\mathbf{e}} = 4$)

low ($h^2 < 0.15$) or moderate ($0.15 < h^2 < 0.4$). We show in a simulation study that the modified high Laplace approximation methods can perform better than REML in these cases.

Also, in the animal study the data size is usually too large and reducing the computation time is an important issue in these models. In this paper, we proposed a modified Laplace approximation for estimating the variance component parameters and heritability with the maximum likelihood method. The implementation and evaluation of the new proposed Laplace approximate method are examined by a simulation study. The calculated results showed that our proposed approximate likelihood method is an accurate and fast method for analyzing the animal model.

The multivariate normal distribution is usually assumed for the genetic random effects. [19] showed that misspecification of distribution of genetic random effects in an animal model affects the estimation of the parameters and used the closed skew normal distribution for modelling the genetic random effects. The family of closed skew normal includes the normal distribution and is more flexible than the normal distribution. In mixed models [6] showed wrong normal assumptions cause bias in the parameter estimates and affect the accuracy of spatial predictions, see e.g. [7], [9], [12], [13] and [14]. Therefore, as future work, we try to improve our approximate likelihood approach to the animal models with non-normal genetic random effects.

References

[1] H. Baghishani, H. Rue and M. Mohammadzadeh, *On a hybrid data cloning method and its application in generalized linear mixed models*, Statist. Comput. **22** (2012), no. 2, 613–597.

[2] B.M. Bolker, M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H.H. Stevens and J.S.S. White, *Generalized linear mixed models: a practical guide for ecology and evolution*, Trends Ecol. Evol. **24** (2009) no. 3, 127—135.

[3] N.E. Breslow and D.G. Clayton, *Approximate inference in generalized linear mixed models*, J. Amer. Statist. Assoc. **88** (1993), no. 421, 9-25.

[4] E. Evangelou, Z. Zhu and R. Smith, *Estimation and prediction for spatial generalized linear mixed models using high order Laplace approximation*, J. Statist. Plann. Inference **141** (2011), no. 11, 3564–3577.

[5] J. Guan, W. Wang, Y. Hu, M. Wang, T. Tian, J. Kong, *Estimation of genetic parameters for growth trait of turbot using Bayesian and REML approaches*, Acta Oceanologica Sinica **36** (2017), no. 6, 47–51.

[6] F. Hosseini, J. Eidsvik and M. Mohammadzadeh, *Approximate Bayesian inference in spatial GLMM with skew normal Latent variables*, Computat. Statist. Data Anal. **55** (2011), no. 4, 1791–1806.

[7] F. Hosseini and M. Mohammadzadeh, *Bayesian prediction for spatial GLMM's with closed skew normal Latent variables*, Aust. New Zealand J. Statist. **54** (2012), 43–62.

[8] F. Hosseini, *A new algorithm for estimating the parameters of the spatial generalized linear mixed models*, Envir. Ecol. Statist. **23** (2016), no. 2, 205–217.

[9] F. Hosseini and O. Karimi, *Approximate likelihood Inference in Spatial Generalized Linear Mixed Models with Closed Skew Normal latent Variables*, Commun. Statist. Simul. Comput. **49** (2020), no. 1, 121–134.

[10] F. Hosseini and O. Karimi, *Approximate composite marginal likelihood inference in spatial generalized linear mixed models*, J. Appl. Statist. **46** (2019), no. 3, 542–558.

[11] D.L. Johnson, and R. Thompson, *Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information*, J. Dairy Sci. **78** (1995), no. 2, 449–456.

[12] O. Karimi and M. Mohammadzadeh, *Bayesian spatial prediction for discrete closed skew Gaussian random field*, Math. Geosci. **43** (2011), no. 5, 565—582.

[13] O. Karimi and M. Mohammadzadeh, *Bayesian spatial regression models with closed skew normal correlated errors and missing observations*, Statist. Papers **53** (2012), no. 1, 205—218.

[14] O. Karimi, F. Hosseini and M. Mohammadzadeh, *Pairwise likelihood in spatial GLMM with skew normal latent variables*, Proc. 15th Conf. Int. Assoc. Math. Geo. Salzburg, Austria, 2011.

[15] L.E. Kruuk, T.H. Clutton-Brock, J. Slate, J.M. Pemberton, S. Brotherstone and F.E. Guinness, *Heritability of fitness in a wild mammal population*, Proc. Nat. Acad. Sci. USA. **97** (2000), no. 2, 698—703.

[16] L.E. Kruuk, *Estimating genetic parameters in natural populations using the "animal model"*, Philos. Trans. R. Soc. B Biol. Sci. **359** (2004), no. 1446, 873—890.

[17] J.M. Milner, J.M. Pemberton, S. Brotherstone and S.D. Albon, *Estimating variance components and heritabilities in the wild: a case study using the 'animal model' approach*, J. Evol. Bio. **13** (2000), no. 5, 804—813.

[18] P. McCulloch and J.A. Nelder, *Generalized linear models*, Chapman and Hall, London, 1989.

[19] F. Pakbaz, F. Hosseini and A. Nematollahi, *Modeling additive genetic effects in animal models by closed skew normal distribution*, Commun. Statist. Simul. Comput. 51 (2022), no. 3, 1186–1198.

[20] E. Ruli, N. Sartori and L. Ventura, *Improved Laplace approximation for marginal likelihoods*, Electron. J. Stat. **10** (2016), no. 2, 3986–4009.

[21] Z. Shun and P. McCullagh,*Laplace approximations of high dimensional integrals*, J. Royal Statist. Soc. Ser. B **57** (1995), no. 4, 749–760.

[22] M. Torabi, *Likelihood inference in generalized linear mixed measurement error models*, Comput. Statist. Data Anal. **57** (2013), no. 1, 549–557.

[23] M. Torabi, *Likelihood inference for spatial generalized linear mixed models*, Commun. Statist. Simul. Comput. **44** (2015), no. 7, 1692–1701.

[24] T.R. Vaez, F.W. Nicolas and H.W. Raadsma, *REML estimates of variance and covariance components for production traits in Australian Merino sheep, using an animal model. 1. Body weight from birth to 22 months*, Aust. J. Agricul. Res. **47** (1996), no. 8, 1235–1249.

[25] B.J. Wilson, F.W. Nicholas, J.W. James, C.M. Wade and P.C. Thomson*Estimated breeding values for Canine Hip Dysplasia Radiographic Traits in a cohort of Australian German Shepherd Dogs*, PLoS One **8** (2013), no. 10, e77470.

[26] A.J. Wilson, D.W. Coltman, J.M. Pemberton, A.D.J. Overall, K.A. Byrne and L.E.B. Kruuk, *Maternal genetic effects set the potential for evolution in a free-living vertebrate population*, J. Evol. Bio. **18** (2005), no. 2, 405—414.

[27] A.J. Wilson, D. Reale, M.N. Clements, M.M. Morrissey, E. Postma, C.A. Walling, L.E. Kruuk and D.H. Nussey, *An ecologist's guide to the animal model*, J. Animal Eco. **79** (2010), no. 1, 13-–26.

[28] H. Zhang, *On estimation and prediction for spatial generalized linear mixed models*, Biometrics **58** (2002), no. 1, 129-–136.