

家畜遗传评估方法研究进展

李学武^{1,2} 王瑞军¹ 王志英¹ 苏蕊¹ 张燕军¹ 刘志红¹ 李金泉^{1*}

(1. 内蒙古农业大学 动物科学学院, 呼和浩特 010018;

2. 攀枝花学院 生物与化学工程学院, 四川 攀枝花 617000)

摘要 为研究家畜育种值估计方法的原理和适应性,提高家畜育种值估计的准确性。以“家畜”、“育种值估计”、“方差组分”、“基因组选择方法”和“统计模型”为关键词,利用中国知网、百度学术、谷歌学术和 SCI-HUB 对 2010—2019 年的家畜育种值估计方法进行文献检索和归纳。研究发现不同家畜遗传评估方法所适合的研究对象不同,并且在研究过程中所选用的方法不同对育种值估计得出的结果的准确性也不相同,因此在应用家畜育种值估计时,应充分了解各种方法的原理并选择合适的估计方法,以提高家畜育种值估计的准确性。未来遗传评估模型的建立应以研究对象为目标,选择合适的遗传评估模型,以期获得更加准确的估计育种值进行早期选种,缩短世代间隔,加快遗传进展。

关键词 家畜; 育种值; 遗传评估方法; 基因组

中图分类号 S827

文章编号 1007-4333(2020)02-0054-07

文献标志码 A

Research progress of the genetic evaluation method for livestock

LI Xuewu^{1,2}, WANG Ruijun, WANG Zhiying, SU Rui, ZHANG Yanjun, LIU Zhihong, LI Jinquan^{1*}

(1. College of Animal Science, Inner Mongolia Agricultural University, Hohhot 010018, China;

2. College of Biological and Chemical Engineering, Panzhihua University, Panzhihua 617000, China)

Abstract To study the principle and adaptability of the estimation methods of breeding value for livestock and improve the accuracy of breeding value estimation, literatures related to the methods of estimating breeding value published from 2010 to 2019 are searched by using “livestock”, “estimation of breeding value”, “variance component”, “genome selection method” and “statistical model” as keywords. China National Knowledge Infrastructure, Baidu Scholar, Google Scholar and SCI-HUB are taken as data sources. It is found that different methods are suitable for different research objects, and the accuracy of estimation breeding value is different in different methods. Therefore, to estimate breeding value for livestock and improve the accuracy of estimation of breeding value of livestock, the principles of various methods and choose appropriate estimation methods should be fully considered. More accurate estimations of breeding values could be achieved by establishing genetic evaluation model for study and select suitable genetic evaluation model to assist early selection, shorten the generation interval and speed up the genetic progress in future.

Keywords livestock; breeding value; genetic evaluation method; genome

家畜遗传评估是家畜育种过程中的一项重要内容。在家畜育种过程中通过对家畜个体育种值估计确定个体种用价值的高低。并且在自然群体进化的

研究中是以数量性状的遗传为基础,通过估计数量性状的遗传基础来研究一个性状的表型是由环境还是由遗传决定,因为一个性状必须是可遗传的才具

收稿日期: 2019-07-28

基金项目: 内蒙古自治区科技计划项目(201603004);内蒙古农业大学科技成果转化专项资金动植物品种培育(选育)专项资助项目(YZGC2017002);现代农业产业技术体系建设专项资金资助(CARS-39-06)

第一作者: 李学武, 讲师, 博士, 主要从事动物遗传育种与繁殖研究, E-mail: nmgndlxw@163.com

通讯作者: 李金泉, 教授, 主要从事绒山羊遗传育种研究, E-mail: Lijinquannd@126.com

有研究价值,通过对群体内自然选择和微观演化模式剖析可以更好的理解种群的宏观演化和形态分析^[1]。

家畜育种值估计主要用于家畜数量性状的遗传分析,通过家畜数量性状的育种值估计,选择育种值高和遗传性能好的个体留作种用,提高后代的生产性能。传统育种值估计通过数量遗传学的方法和模型利用表型数据进行育种值估计,例如选择指数法、群体比较法、最佳线性无偏估计法和平均信息约束最大似然法,目前最常用的方法为 BLUP 法,但是随着数量遗传学的发展和分子标记技术的进步,育种值估计不仅仅局限于表型值估计,而是通过对基因型值计算育种值, BLUP 法也随之发展成为 GBLUP 用于估计基因型值,这种利用基因型值估计的育种值更加准确,随后又将 Bayes 法应用于基因组育种值估计中。

1 ML REML 及 AIREML

最大似然法(Maximum likelihood, ML)和约束最大似然法(Restricted maximum likelihood, REML)分别是由 Hartley 和 Patterson^[2]提出。ML 是基于观察值的似然函数最大化,而 REML 是基于对固定效应的似然函数最大化,因此 REML 的估计值消除了模型中固定效应的影响^[3]。ML 和 REML 的假设条件均是观测值服从正态分布,在正态分布的总体中进行抽样估计总体方差时,ML 估计值和 REML 估计值分别是^[3]:

$$\hat{\sigma}_{ML}^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

$$\hat{\sigma}_{REML}^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

由上式可以看出 ML 的自由度为样本量 n ,而 REML 的自由度为样本量减 1,所以 ML 估计值的有偏的,REML 估计值是无偏的。由于无偏估计是估计值接近与真实值,所以 REML 法优于 ML。目前 REML 法在动物育种中应用较多,例如 REML 模型被用来估计动物体重、机体消耗和产子数等遗传参数^[4];通过 REML 法得出母体效应对体重有影响^[5]。但是,一般情况下 ML 和 REML 的估计值均是有偏的,并且必须迭代求解。ML 和 REML 的估计值在模型假设下是最优的,但在计算上可能非常困难,特别是对于涉及具有缺失数据的多个特征的

数据进行分析。Ashida 等^[6]在 REML 法的基础上提出了平均信息约束最大似然法(Average information restricted maximum likelihood, AIREML),AIREML 法是假设动物个体模型包括动物效应(育种值)和残差效应的,AI 矩阵元素的原始表达式,是在最初的算法基础上加入了预测的残差向量。并且 AIREML 被广泛应用于家畜育种值估计中^[7-8]。李学武等^[9]利用 AIREML 算法对内蒙古绒山羊不同毛被类型进行遗传参数估计和育种值估计,研究结果表明其长毛型个体毛长的遗传进展较快;并通过利用 AIREML 算法对不同毛被类型下产绒量和体重进行遗传参数分析发现产绒量和体重的遗传力随着毛长的增加而增加^[10],这些研究均对内蒙古绒山羊的育种具有重要意义。Gebim 等^[11]利用 AIREMLF09 程序对奈洛牛的平均日增重、干物质摄入量、饲料转化率和剩余饲料摄入量的表型数据计算其遗传力和遗传相关,发现饲料转化率的遗传力是 0.19,并且与日增重的遗传相关较高,所以将饲料转化率的纳入奈洛牛的育种规划当中。

2 BLUP 法

最小二乘法是用于估计固定效应值,但是不能分别估计遗传固定效应和环境固定效应,因为在估计个体育种值时,如果在同一环境条件下环境效应值可以忽略不计,但是在不同条件下环境效应值不可以忽略。因此 Henderson 等^[12]提出了最佳线性无偏估计(Best linear unbiased prediction, BLUP)法,在动物遗传育种中 BLUP 是一个估计随机效应的方法,有助于理解固定效应和随机效应模型。

2.1 最佳线性无偏估计(BLUP)

最初动物育种值的估计方法主要是用单性状的个体本身成绩、祖先成绩、同胞成绩和后裔成绩通过选择指数(Selection index)进行育种值估计^[13]。在实际应用中,将公母畜分别用不同的体系进行评估,具有局限性,而且在计算过程中不得不舍去一些与种畜有关的信息,降低了种畜估计的准确性^[14]。BLUP 法具有估计值方差最小、估计值无偏、可消除因选择和淘汰等原因造成的偏差等特点,并且选择指数可以看作是 BLUP 的一种随机效应模型。BLUP 法的实质是线性模型^[14]:

$$y = \mathbf{X}\beta + \mathbf{Z}u + e$$

$$E(u) = 0, E(e) = 0, \text{Var} \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{bmatrix} \sigma^2$$

式中: y 是观察值向量; β 是未知的固定效应; u 是未知随机效应; \mathbf{X} 和 \mathbf{Z} 分别是已知的固定效应和随机效应的结构矩阵; \mathbf{G} 和 \mathbf{R} 分别是正定矩阵; σ^2 是常数。

在动物育种中是对 \mathbf{A}^{-1} 的计算如下:

$$\mathbf{A}^{-1} = \mathbf{L}\mathbf{L}' = (\mathbf{T}\mathbf{D}\mathbf{D}\mathbf{T}')^{-1} = \mathbf{T}'^{-1}\mathbf{D}^{-2}\mathbf{T}^{-1}$$

对于单性状的 BLUP 模型的方程组是:

$$\begin{bmatrix} \mathbf{X}\mathbf{X}' & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + k\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} \mathbf{X}\hat{\beta}\mathbf{X}'y \\ \mathbf{Z}'y \end{bmatrix}$$

式中: $k = \sigma_e^2/\sigma_u^2$; $\text{Var}(e) = \mathbf{R} = I\sigma_e^2$; $\text{Var}(u) = \mathbf{G} = I\sigma_u^2$; \mathbf{A} 是分子血缘相关矩阵(个体间加性相关矩阵), 并且 $\mathbf{A}^{-1} = \frac{1}{\sigma_u^2}I$ 。

多性状的 BLUP 模型^[14]:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}y \\ \mathbf{Z}'y \end{bmatrix}$$

式中: $\mathbf{R} = \mathbf{V} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$, 由于计算困难, 如果设 2 个性状的随机误差的协方差为 0, 并且同一性状的观测值的随机误差属于独立分布, 即 $R_{12} = R_{21} = 0, R_{11} = I_{11}\sigma_{e1}$; $R_{22} = I_{22}\sigma_{e2}$, 所以 $\mathbf{R} = \begin{bmatrix} I_{11}\sigma_{e1}^2 & 0 \\ 0 & I_{22}\sigma_{e2}^2 \end{bmatrix}$, 动物模型 BLUP 法可以加速育种的遗传进展, 尤其是对于低遗传力和和限性性状有效果明显。Muir^[15] 经过对 BLUP 估计育种值和基因组育种值进行比较发现, 只有群体足够大的时候基因组育种值才会更加准确。虽然基因组育种成本较高, 但是 BLUP 育种仍具有重要意义。

2.2 标记辅助最佳线性无偏估计 (MBLUP)

随着 BLUP 法的发展和分子标记技术的提高, 在动物的育种中可以通过标记辅助选择来达到对动物性状的早期选择和对限性性状进行选择。育种值也由原来的估计育种值变成了基因组育种值。在标记辅助选择的过程中需要利用与表型性状相关的主效基因或者与主效基因连锁的标记的信息, 从而出现了标记辅助最佳线性无偏差预测法 (Marker-assisted BLUP, MBLUP)。MBLUP 法是现代育种学的新技术, 是常规 BLUP 方法的扩展, 同时利用表型系谱和 DNA 标记信息, 对个体进行遗传评定使育种值估计的准确性提高, 从而提高选种的准确性和效率^[16]。MBLUP 方法是在动物模

型中将基因的效应分为已定位的 QTL 效应和未知基因(剩余多基因)效应, 并借助分子标记信息来推断 QTL 基因型或 QTL 等位基因的概率, 个体总育种值是 QTL 育种值和剩余多基因育种值之和^[17]。

MBLUP 的育种模型为:

$$Y = \mathbf{X}\beta + \mathbf{Z}u + \mathbf{W}v + e$$

式中: Y 是表型值向量; β 是固定效应向量(世代和性别); u 是多基因值的随机向量; v 是 QTL 等位基因向量; e 是残差效应向量; \mathbf{X} 、 \mathbf{Z} 和 \mathbf{W} 分别是 β 、 u 和 v 的关联矩阵; β 、 u 和 v 期望和方差是

$$E \begin{bmatrix} u \\ v \\ e \end{bmatrix} = 0 \quad \text{Var} \begin{bmatrix} u \\ v \\ e \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_u^2 & 0 & 0 \\ 0 & \mathbf{G}\sigma_v^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

\mathbf{A} 是加性效应的关系矩阵; \mathbf{G} 是 QTL 基因型的关联矩阵; 其混合方程组为:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Q} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}k_1\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{Q} \\ \mathbf{Q}'\mathbf{X} & \mathbf{Q}'\mathbf{Z} & \mathbf{Q}'\mathbf{Q}k_2\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \\ \hat{v} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'y \\ \mathbf{Z}'y \\ \mathbf{Q}'y \end{bmatrix}$$

式中: 其中 k_1 和 k_2 分别是 σ_e^2/σ_u^2 和 σ_e^2/σ_v^2 ; 并且求得秩相关系数为 $r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$; 其中 d 真实值与估计值之差; n 是个体数量。Lu 等^[18] 和 Goddard 等^[19] 证明观察性状的遗传力越高、QTL 方差越小或者 2 个相邻标记间图距越小, 动物模型下 MBLUP 估计的准确性越高, 相反, 当性状的遗传力较低、QTL 方差越大或者 2 个相邻标记间图距较大时, 动物模型 MBLUP 估计的准确性越低。

2.3 基因辅助最佳线性无偏估计 (GBLUP)

由于数量性状受微效多基因影响, 选择强度, 遗传模式, 基因效应的大小, 基因频率和数据结构均影响多基因的遗传力估计。Kennedy^[20] 提出了单个位点的混合线性模型的基因辅助最佳线性无偏预测 (Gene-assisted best linear unbiased prediction, GBLUP) 估计方法。GBLUP 的原理是将已知 QTL 的基因型值作为固定效应, 将其他多基因育种值作为随机遗传效应, 并借助混合线性模型同时估计出个体的 QTL 基因型值和多基因育种值^[20]。Habier 等^[21] 表示基因组 BLUP 利用连锁不平衡 (LD) 和加性遗传相关, 找到在 QTL 的共分离关系, 用来估计基因组育种值的准确性更高。

GBLUP 采用基因型模型为^[21]:

$$y_{ij} = g_i + a_{ij} + e_{ij}$$

式中： y_{ij} 是第 ij 个观察值； g_i 是单位点第 i 个的基因型效应； a_{ij} 是基因加性效应； e_{ij} 是随机环境效应。

基因型模型转换为矩阵模型为：

$$y = Xg + Za + e$$

式中： X 是单个位点基因型(g)的观测值矩阵； Z 是基因加性效应(a)矩阵； e 是残差。其混合方程组为：

$$\begin{bmatrix} \hat{g} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'ZA^{-1}\lambda \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

式中： $\lambda = \sigma_e^2/\sigma_a^2$ ； A 是加性效应关系矩阵。Misztal^[22] 设 $H = A + A_\Delta$ 时，用基因组关系矩阵(G)代替基因型动物的分子关系矩阵(A)，即^[22]：

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$H = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & G \end{bmatrix} = A + \begin{bmatrix} 0 & 0 \\ 0 & G - A_{22} \end{bmatrix}$$

即设常规方程为：

$$H \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

并且基于系谱矩阵给出了 H^{-1} ，即：

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

但是要注意 G 通常是单数，因此，不能在没有附加步骤的情况下进行倒置^[23]。随着基因组学的发展，在 GBLUP 的基础上又提出了 SS-BLUP (Single-step genomic best linear unbiased predictions, SS-BLUP)^[24]、随机回归 BLUP (Random regression best linear unbiased prediction, RR-BLUP)^[25]、TABLUP^[26]、TP-BLUP (BLUP based on trait phenotypes and pedigree)^[27] 等。并且 Gao 等^[28] 发现单步混合方法只能在基因型和非基因型动物的信息同时进行基因组预测是使用，在使用一步法时关系矩阵可能需要进行调整，因此，提出了两步法用于基因组分析。Fragomeni 等^[29] 利用 BLUP 和 SS-BLUP 对杜洛克猪种的热应激进行遗传评估，发现 SS-BLUP 的准确率高于 BLUP 估计。随着 BLUP 法的规范和完善，BLUP 法成为使用最广的估计育种值方法^[30-35]。

3 贝叶斯法 (Bayes)

贝叶斯统计原理是根据先验分布和实际观测值

组成后验信息后，对后验分布进行统计推断。一般情况下，贝叶斯统计的后验分布估计较难，为了快速和准确的估计特定性状的遗传参数，Mathew^[36] 提出了马尔可夫链蒙特卡尔理论 (Markov chain monte carlo, 简称 MCMC)，该理论是通过方差和协方差建立随机效应的最大似然函数，由于 MCMC 的计算过程较难，又将 Gibbs 抽样方法的引入 Bayes 计算^[37]，因此 Bayes 方法在统计学中开始被广泛应用，目前主要用于基因组选择和基因组育种值估计当中，包括 BayesA^[38]、BayesB^[35]、BayesC π 、BayesD π ^[40]、BayesSSVS^[41] 和 Bayesian LASSO^[42] 等。

3.1 Bayes A

Bayes A 是由 Meuwissen^[39] 提出。Bayes A 的理论基础是假设 SNPs 效应 g_i 的先验分布服从正态分布即^[39]：

$$p(g_k | \sigma_{g_i}^2) = N(0, \sigma_{g_i}^2)$$

不同的 g_i 具有不同的 $\sigma_{g_i}^2$ ，并且 $\sigma_{g_i}^2$ 服从自由度为 v 、尺度参数为 S 的逆卡方分布^[43]，即： $\sigma_{g_i}^2 \sim \chi^{-2}(v, S)$ ，那么所得到的后验分布也是一个逆卡方分布^[39]：

$$p(\sigma_{g_i}^2 | g_i) = \chi^{-2}(v + n_i, g_i'g_i)$$

残差效应方差 σ_e^2 服从自由度为 -2 ，尺度参数为 0 的逆卡方分布，即^[39]：

$$p(\sigma_e^2 | e_i) = \chi^{-2}(n - 2, e_i'e_i)$$

v 和尺度参数 S 与遗传结构直接相关 ($v = 4.012, S = 0.002$)，通过吉布斯抽样获得单倍型效应的估计值^[43]。

3.2 Bayes B

实际上，基因位点遗传变异的分布中有许多基因位点没有遗传变异(不分离)或者只有少数遗传变异。并且在整个基因组中，可能大部分染色体片段没有包含 QTL，只有少数的几个染色体片段含有 QTL。因此，Meuwissen^[39] 在 BayesA 的基础上提出了 Bayes B，Bayes B 的理论基础是假设大多数 SNP 没有效应，只有少数 SNP 有效应并效应方差不同。Bayes B 先验分布的假设为^[39]：

$$\begin{cases} \sigma_{g_i}^2 = 0 & \text{with probability } \pi \\ \sigma_{g_i}^2 \sim \chi^{-2}(v, S) & \text{with probability } 1 - \pi \end{cases}$$

Habier D 等^[40] 提出，当 $v = 4.2$ 时对 S 进行推导^[40]：

$$E(\sigma_{g_k}^2) = \frac{v_a s_a^2}{v_a - 2} = \bar{\sigma}_a^2, \text{ 因此 } S_a^2 = \frac{\bar{\sigma}_a^2 - 2}{v_a}$$

$$\tilde{\sigma}_a^2 = \frac{\tilde{\sigma}_s^2}{(1-\pi) \sum_{k=1}^K 2p_k(1-p_k)}$$

式中： $\tilde{\sigma}_a^2$ 是随机抽样位点的加性效应的方差； S_a^2 服从单变量 t 分布 $(0, v_a, S_a^2)$ ； $\tilde{\sigma}_s^2$ 是 SNPs 的加性遗传方差； p_k 是第 K 个 SNP 的等位基因频率。在 Bayes B 中 3 个参数为 $\pi=4.2339$ ， $v_a=0.0429$ ， $S_a^2=0.947^{[40]}$ 。由于 Bayes B 使用混合分布作为标记效应方差的先验，难以构建标记效应和方差各自的完全条件后验分布，所以 Bayes B 不采用 Bayes A 中加入了 Gibbs 抽样的 MCMC 算法，而是使用 MH (Metropolis-hasting) 算法进行抽样对标记效应和方差进行联合抽样。

3.3 Bayes C π 及 Bayes D π

Bayes C π 和 Bayes D π 是 Habier 等^[40]在 Bayes A 和 Bayes B 的基础上提出，Bayes C π 是设定所有的 SNP 效应具有一个共同的方差，即： $\sigma_{a_k}^2 = \sigma_a^2$ ，并且 SNP 效应以概率为 $(1-\pi)$ 的时候服从 $t(0, v_a, IS_a^2)$ 分布。Bayes D π 则是设 S_a^2 为未知参数，并且尺度函数的先验分布服从 Gamma(1, 1)，Berg 等^[44]发现 Bayes C π 的结果受遗传力，QTL 数量和标记数量的影响较大，由于低遗传力的性状中假阳性的数量偏高，所以 Bayes C π 适用于中等遗传力以上的性状分析。

3.4 Bayesian LASSO

Bayesian LASSO (Bayesian least absolute shrinkage and selection operator) 是 Tibshirani^[42]提出，是通过标准化预测因子 x_{ij} 和中心响应值 y_i 获得线性回归系数，解决了 l_1 -处罚回归问题，使 $\beta = \{\beta_j\}$ 最小化，即^[42]：

$$\sum_{i=1}^N (y_i - \sum_j x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

这相当于一个约束平方和最小化 $\sum |\beta_j| \leq s$ 。Bayesian LASSO 的经典的线性模型：

$$y = Xb + Za + e \quad e | \sigma^2 \sim \text{MVN}(0, I\sigma^2)$$

式中： b 是固定效应； a 是 SNP 效应； X, Z 分别是固定效应和 SNP 效应的结构矩阵； e 是残差向量。然后，基于单个 SNP 建模，即：

$$p(a | \sigma^2, \lambda) = \frac{\lambda}{2\sigma} \exp\left(-\frac{\lambda |a|}{\sigma}\right)$$

最后，给定 a 在 Bayesian LASSO 的方差组分为：

$$\text{Var}(a) = 2\sigma^2/\lambda^2$$

由于在模型中假定 SNP 的分布效应与未观察到的

和残差效应无关，因此 Park^[45]将数据残差 (σ_e^2) 和 SNPs 引起的变异 (σ_a^2) 写入模型，并对模型进行重建，即^[45]：

$$y = Xb + Za + e$$

$$a | \lambda, \sigma_a^2 \sim \prod_i \frac{\lambda}{2\sigma_a} \exp\left(-\frac{\lambda |a_i|}{\sigma_a}\right)$$

$$e | \sigma_e^2 \sim \text{MVN}(0, I\sigma_e^2)$$

显然，模型等于：

$$y = Xb + Za + e$$

$$a | \lambda \sim \prod_i \frac{\lambda}{2} \exp(-\lambda |a_i|)$$

$$e | \sigma_e^2 \sim \text{MVN}(0, I\sigma_e^2)$$

那么 SNPs 效应的方差组分为：

$$\text{Var}(a) = \frac{2}{\lambda^2} = \frac{\sigma_u^2}{2 \sum_i p_i(1-p_i)}$$

式中： σ_u^2 是群体遗传变异， p_i 是第 i 个标记等位基因的频率。Legarra 等^[46]经过研究比较发现在改进模型中包含 2 个方差组分(遗传方差和残差方差)的分析更适合于 SNPs 的效应分析。

4 展望

随着标记技术和数量遗传学的进步，根据不同的遗传背景和育种目标选择合适的统计方法，建立合理的统计模型进行遗传评估，可大大提高估计育种值的准确性。目前，测量基因相互作用的效应模型有 2 类：一是统计模型用于检测统计参数、估计遗传效应和遗传方差；二是功能模型通过基因型-表型图谱来计算遗传参数。Álvarez 等^[47]结合这 2 种模型建立了自然和正交的交互 (Natural and orthogonal interactions) 模型。这种模型的应用也越来越广泛，原因是即使这些数据不平衡或没有进行无偏估计，该模型也可在一定程度上反映出随机效应^[48]。近年来，研究表明通过考虑显性效应和上位效应，可以提高育种值估计的准确性^[49]。虽然基于家系的育种值估计已成功地用于动物育种中，然而，依据家系的育种值估计对遗传进展的速度有一定的限制性。随着分子标记技术的发展，利用分子标记在基因组上进行大量的位点标记，通过微卫星 (Microsatellite)、单核苷酸多态 (Single nucleotide polymorphisms) 等遗传标记，将遗传标记与表型数据进行关联分析，挖掘与性状有关的标记或数量性状基因座 (Quantitative trait loci)，建立合理的模型和采用合适的方法对家畜进行遗传评

估,以期精确地估计家畜育种值,进行早期选种,加快遗传进展。

参考文献 References

- [1] Hendry A P, Kinnison M T. Perspective: The pace of modern life: Measuring rates of contemporary microevolution [J]. *Evolution*, 1999, 53(6): 1637
- [2] Hartley H O, Rao J N K. Maximum-likelihood estimation for the mixed analysis of variance model[J]. *Biometrika*, 1967, 54(1-2): 93-108
- [3] Lele S, Taper M L. A composite likelihood approach to (Co) variance components estimation [J]. *Journal of Statistical Planning and Inference*, 2002, 103(1-2): 117-135
- [4] Beniwal B K, Hastings I M, Thompson R, Hill W G. Estimation of changes in genetic parameters in selected lines of mice using REML with an animal model. 2. Body weight, body composition and litter size[J]. *Heredity*, 1992, 69(4): 361-371
- [5] Vaez T R, Nicolas F W, Raadsma H W. 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 [J]. *Australian Journal of Agricultural Research*, 1996, 47(8): 1235-1249
- [6] Ashida I, Iwaisaki H. A numerical technique for REML estimation of variance components using average information algorithm and its computing property [J]. *Nihon Chikusan Gakkaiho*, 1998, 69(7): 631-636
- [7] Assan N, Mabuku O. Direct and maternal genetic effects for birth weight in dorper and mutton merino sheep in Zimbabwe [J]. *Animal Production Research Advances*, 2011, 7(3): 236-240
- [8] Zhang Y J, Wang Z Y, Lei H, Wang Z X, Su R, Zhang W G, Zhou J, Wei Y L, Meng R Q, Wang R J, Li J Q. Estimates of genetic parameters and genetic changes for fleece traits in Inner Mongolia Cashmere goats [J]. *Small Ruminant Research*, 2014, 117(1): 41-46
- [9] 李学武, 王瑞军, 王志英, 李宏伟, 王振宇, 苏蕊, 张燕军, 刘志红, 李金泉. 内蒙古绒山羊不同毛被类型遗传参数估计及遗传进展研究[J]. *中国农业大学学报*, 2018, 23(4): 53-59
- LI X W, Wang R J, Wang Z Y, Li H W, Wang Z Y, Su R, Zhang Y J, Liu Z H, Li J Q. Estimation of genetic parameters and research on the genetic progress of different staple length Inner Mongolian Cashmere goat [J]. *Journal of China Agricultural University*, 2018, 23(4): 53-59 (in Chinese)
- [10] 李学武, 刘燕, 王瑞军, 王志英, 娜清, 李宏伟, 王振宇, 徐冰冰, 苏蕊, 张燕军, 刘志红, 李金泉. 内蒙古绒山羊不同毛被类型产绒量和体重的遗传参数估计[J]. *中国农业科学*, 2018, 51(12): 2410-2417
- Li X W, Liu Y, Wang R J, Wang Z Y, Na Qing, Li H W, Wang Z Y, Xu B B, Su R, Zhang Y J, Liu Z H, Li J Q. Genetic parameter estimation of cashmere yield and body weight at different staple types of Inner Mongolian Cashmere goats [J]. *Scientia Agricultura Sinica*, 2018, 51(12): 2410-2417 (in Chinese)
- [11] Gebim P G, Lais G, Esperandio C M, Paulo R J, Jose B S, Miguel A S. Genetic correlations and heritability estimates for dry matter intake, weight gain and feed efficiency of Nelore cattle in feedlot [J]. *Livestock Science*, 2018: S1871141318301896
- [12] Robinson G K. That BLUP is a good thing: The estimation of random effects [J]. *Statistical Science*, 1991: 15-32
- [13] Smith A C. Medical aspects of aviation [J]. *Archives of Otolaryngology: Head & Neck Surgery*, 1936, 23(2): 139-148
- [14] Chen Y S, Sheng Z L. The principle of universal selection index [J]. *Journal of Genetics and Genomics*, 1988(3): 27-33
- [15] Muir W M. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters [J]. *Journal of Animal Breeding and Genetics*, 2015, 124(6): 342-355
- [16] 肖礼华, 史忠辉, 曾琼, 穆林, 徐猛, 罗卫星, 刘若余. 利用 MBLUP 法选育黔北麻羊 [J]. *草食家畜*, 2014(2): 24-25
- Xiao L H, Shi Z H, Zeng Q, Mu L, Xu M, Luo W X, Liu R Y. Breeding for Qianbei Brown goat by MBLUP [J]. *Grass-Feeding Livestock*, 2014(2): 24-25 (in Chinese)
- [17] 殷宗俊, 张勤. 主基因-多基因混合模型下种畜标记辅助遗传评定的效率 [J]. *安徽农业大学学报*, 2005, 32(2): 141-145
- Yin Z J, Zhang Q. Accuracy of marker assisted genetic evaluation under the major gene-polygene mixed inheritance models [J]. *Journal of Anhui Agricultural University*, 2005, 32(2): 141-145 (in Chinese)
- [18] Lu S X, Wu C X, Lian L S. Primary factors affecting the evaluation accuracy of animal model MBLUP [J]. *Acta Genetica Sinica*, 2003, 30(1): 35-39
- [19] Goddard M E. A mixed model for analyses of data on multiple genetic markers [J]. *Theoretical and Applied Genetics*, 1992, 83(6-7): 878-886
- [20] Kennedy B W, Quinton M, van Arendonk J A. Estimation of effects of single genes on quantitative traits [J]. *Journal of Animal Science*, 1992, 70(7): 2000-2012
- [21] Habier D, Fernando, Muir W M. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters [J]. *Journal of Animal Breeding and Genetics*, 2007, 124(6): 342
- [22] Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information [J]. *Journal of Dairy Science*, 2009, 92(9): 4648-4655
- [23] Aguilar I, Misztal I, Johnson D L, Legarra A, Tsuruta S, Lawlor T J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic

- evaluation of Holstein final score [J]. *Journal of Dairy Science*, 2010, 93(2): 743-752
- [24] Christensen O F, Lund M S. Genomic prediction when some animals are not genotyped [J]. *Genetics Selection Evolution*, 2010, 42(1): 2
- [25] Whittaker J C, Thompson R, Denham M C. Marker-assisted selection using ridge regression [J]. *Annals of Human Genetics*, 1999, 63(4): 366-366
- [26] Zhang Z, Liu J F, Ding X D, Bijma P, de Koning D J, Zhang Q. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix [J]. *PLoS One*, 2010, 5(9): e12648
- [27] Fernando R L, Habier D, Stricker C, Dekkers J C M, Totir L R. Genomic selection [J]. *Acta Agriculturae Scandinavica, Section a-Animal Science*, 2007, 57(4): 192-195
- [28] Gao H D, Christensen O F, Madsen P, Nielsen U S, Zhang Y, Lund M S, Su G S. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population [J]. *Genetics Selection Evolution*, 2012, 44(1): 8
- [29] Fragomeni B D, Lourenco D, Tsuruta S, Gray K A, Huang Y, Misztal I. Genetics of heat stress in purebred and crossbred pigs from different states using BLUP or ssGBLUP [J]. *Journal of Animal Science*, 2016: 94
- [30] Strand I, Mntysaari E A. Comparison of some equivalent equations to solve single-step GBLUP [C]. In: *Proceedings of 10th World Congress on Genetics Applied to Livestock Production*. Vancouver: American Society of Animal Science, 2014
- [31] Habier D, Fernando R L, Garrick D J. Genomic BLUP decoded: A look into the black box of genomic prediction [J]. *Genetics*, 2013, 194(3): 597-607
- [32] Wang X Q, Miao J, Chang T P, Xia J W, An B X, Li Y, Xu L Y, Zhang L P, Gao X, Li J Y, Gao H J. Evaluation of GBLUP, BayesB and elastic net for genomic prediction in Chinese Simmental beef cattle [J]. *PLoS One*, 2019, 14(2): e0210442
- [33] Legarra A, Ricard A, Varona L. GWAS by GBLUP: Single and multimarker EMMAX and Bayes factors, with an example in detection of a major gene for horse gait [J]. *G3: Genes, Genomes, Genetics*, 2018, 8(7): 2301-2308
- [34] Liu Y, Lu S, Liu F, Shao C W, Zhou Q, Wang N, Li Y Z, Yang Y M, Zhang Y P, Sun H J, Zheng W W, Chen S L. Genomic selection using BayesC π and GBLUP for resistance against *Edwardsiella tarda* in Japanese flounder (*Paralichthys olivaceus*) [J]. *Marine Biotechnology*, 2018, 20(5): 559-565
- [35] Zhang C Y, Kemp R A, Stothard P, Wang Z Q, Boddicker N, Krivushin K, Dekkers J, Plastow G. Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K and whole-genome sequence variants [J]. *Genetics Selection Evolution*, 2018, 50(1): 14
- [36] Mathew B, Bauer A M, Koistinen P, Reetz T C, Léon J, Sillanpää M J. Bayesian adaptive Markov chain Monte Carlo estimation of genetic parameters [J]. *Heredity*, 2012, 109(4): 235-245
- [37] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, 6(6): 721-741
- [38] Boruah A, Baruah K, Das B, Das M J, Gohain N B. A Bayesian approach for flight fare prediction based on Kalman filter [C]. In: *Advances in Intelligent Systems and Computing*. Singapore: Springer Singapore, 2018: 191-203
- [39] Meuwissen T H, Hayes B J, Goddard M E. Prediction of total genetic value using genome-wide dense marker maps [J]. *Genetics*, 2001, 157(4): 1819-29
- [40] Habier D, Fernando R L, Kizilkaya K, Garrick D J. Extension of the bayesian alphabet for genomic selection [J]. *BMC Bioinformatics*, 2011, 12(1): 186
- [41] Verbyla K L, Hayes B J, Bowman P J, Goddard M E. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle [J]. *Genetics Research*, 2009, 91(5): 307-311
- [42] Tibshirani R. Regression shrinkage and selection via the lasso: A retrospective [J]. *Journal of the Royal Statistical Society*, 1996, 58(1): 267-288
- [43] Wang C S, Rutledge J J, Gianola D. Marginal inferences about variance components in a mixed linear model using Gibbs sampling [J]. *Genetics Selection Evolution*, 1993, 25(1): 41
- [44] Berg I V D, Fritz S, Boichard D. QTL fine mapping with Bayes C (π): A simulation study [J]. *Genetics Selection Evolution*, 2013, 45(1): 1-11
- [45] Park T, Casella G. The Bayesian lasso [J]. *Journal of the American Statistical Association*, 2008, 103(482): 681-686
- [46] Legarra A, Robert-Granié C, Croiseau P, Guillaume F, Fritz S. Improved Lasso for genomic selection [J]. *Genetics Research*, 2011, 93(1): 77-87
- [47] Álvarez-Castro J M, Carlborg Ö. A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis [J]. *Genetics*, 2007, 176(2): 1151-1167
- [48] Fry J D. The mixed-model analysis of variance applied to quantitative genetics: Biological meaning of the parameters [J]. *Evolution*, 1992, 46(2): 540-550
- [49] dos Santos J P, de Castro Vasconcellos R C, Pires L P, Balestre M, von Pinho R G. Inclusion of dominance effects in the multivariate GBLUP model [J]. *PLoS One*, 2016, 11(4): e0152045