

## 利用 Gibbs 抽样方法估计丹系长白猪乳头数的遗传参数

王楚端 陈清明  
(中国农业大学动物科技学院)

周伟良 王作强 韩绍永 李振宽  
(天津市宁河原种猪场)

**摘要** 建立了单性状动物模型通过 Gibbs 抽样对丹系长白猪乳头数遗传参数进行贝叶斯(Bayesian)估计,并与约束最大似然(REML)估计进行比较。数据包括 9 898 个个总乳头数及系谱记录。模型中性别为固定效应,个体遗传效应为随机效应。结果表明,丹系长白猪乳头数的遗传方差组分、环境方差组分及遗传力的后验均值估计分别为  $1.046 \pm 0.009$ ,  $0.534 \pm 0.005$ ,  $0.661 \pm 0.004$ , 相应的 REML 估计值为 1.036, 0.539, 0.658, 两者之间差异不显著( $P > 0.05$ )。Gibbs 抽样时不同的先验值对后验均值估计有一定影响。

**关键词** 猪; 乳头数; 遗传参数; Gibbs 抽样

**分类号** S813.1; S828

## Bayesian Inference of Teat Number in Landrace Using Gibbs Sampling

Wang Chuduan Chen Qingming  
(College of Animal Sciences and Technology, CAU)

Zhou Weiliang Wang Zuoqiang Han Shaoyong Li Zhenkuan  
(Tianjin Ninghe Pig Breeding Farm)

**Abstract** Genetic parameter were estimated for teat number of Danish Landrace by using Gibbs sampling. Total number of records was 9 898. A analytical model included sex as fixed effect and additive genetic random effects. The posterior mean estimates of genetic variance, residual variance and heritability were  $1.034 \pm 0.084$ ,  $0.541 \pm 0.064$ ,  $0.656 \pm 0.037$ , respectively. The corresponding REML estimates were 1.036, 0.539, 0.658, respectively. There were no significant differences between posterior estimates and REML estimates ( $P > 0.05$ ). Several priors were used to examine their influence on inferences. The priors were influential because of the relatively small data size.

**Key words** swine; teat number; genetic parameter; bayesian inference; Gibbs sampling

猪育种中很重视后备种猪乳头的数量与分布情况,尤其是乳头数量少的,因为乳头数量决定了母猪的哺育能力具有重要作用<sup>[1,2]</sup>。这个性状的遗传参数一直受到人们的关注,特别是那些乳头数较少的品种<sup>[3]</sup>。现有关于猪乳头数遗传力的估计主要基于亲子回归或方差分析方法<sup>[1-6]</sup>。这些方法对数据资料要求较高,需要特定的实验设计,而实际育种中普遍存在世代交替及不平衡数据结构的情况,很大程度上影响了这些方法的准确性。特别是只包含单个亲本资料进行亲子回归或方差分析时,误差更为显著<sup>[7]</sup>。目前,建立动物模型进行约束最大似然

收稿日期: 1999-07-12

王楚端,北京圆明园西路 2 号中国农业大学(西校区), 100094

(REML) 估计是育种中常用的遗传参数估计方法<sup>[8,9]</sup>, 而基于 Gibbs 抽样的贝叶斯方法也越来越受到重视<sup>[10,11]</sup>。

本研究建立单性状动物模型, 通过 Gibbs 抽样对丹系长白猪乳头数遗传参数进行贝叶斯 (Bayesian) 估计, 并与 REML 估计进行比较。

## 1 材料与方法

数据来自天津市宁河原种猪场丹系长白猪核心选育群。乳头数性状为仔猪出生后 12 h 内记录的主要性状之一。数据包括 9 898 个乳头数及系谱记录, 其中公猪 4 997 头, 母猪 4 901 头, 观察值范围为 10.0~18.0, 平均值为 14.42, 标准差 1.25。Shapiro-Wilk 检验表明数据服从正态分布。

建立以下动物模型: 
$$y = x\beta + zu + e$$

式中  $y$  是乳头数观察值,  $\beta$  是固定效应 (性别),  $u$  是个体随机加性遗传效应向量,  $e$  是随机残差;  $x$  及  $z$  分别是与固定效应 ( $\beta$ ) 及随机效应 ( $u$ ) 相对应的设计矩阵。

根据贝叶斯公式, 参数的后验联合分布可写成相关的似然函数与先验分布的乘积, 即:

$$f(\sigma_g^2, \sigma_e^2, u, \beta | y) = f(y | \beta, u, \sigma, f) f(u | \sigma_g^2) \pi(\beta) \pi(\sigma_g^2) \pi(\sigma_e^2)$$

其中,  $\pi(\beta)$ ,  $\pi(\sigma_g^2)$ ,  $\pi(\sigma_e^2)$  分别是固定效应, 加性遗传方差组分 ( $\sigma_g^2$ ), 环境方差组分 ( $\sigma_e^2$ ) 的先验分布函数。在进行 Gibbs 抽样时, 首先是通过以上方程式右手项推导出各效应及参数的全条件概率密度分布函数, 进行迭代及抽样, 产生 Gibbs 链, 最终实现各效应及参数的后验边际密度分布。遗传力通过遗传方差组分及残差组分的抽样值计算得出。所有分析过程利用 MTGSAM 软件<sup>[12]</sup>进行。分析时,  $\pi(\beta)$  为均匀分布,  $\pi(\sigma_g^2)$  及  $\pi(\sigma_e^2)$  皆为期望值为 0.000 1 的反卡方分布。这样既能避免均匀先验分布可能产生的不正确结果<sup>[13]</sup>, 又使后验分布几乎完全取决于数据资料<sup>[14]</sup>。Gibbs 链总共运行 30 000 次, 其中过渡期 (burn-in period) 为 2 000 次, 抽样间隔 (thinning rate) 分别为 300 次, 同时用 Gibbsit 程序 (Raftery and Lewis, 1995) 对这些指标进行检验, 并计算相应样本间的自相关系数。研究随后设置 3 种先验值进行 Gibbs 抽样: 均匀分布, 期望值  $\sigma_g^2 = 0.5$ ,  $\sigma_e^2 = 10$  的反卡方分布, 及期望值  $\sigma_g^2 = 1.0$ ,  $\sigma_e^2 = 0.5$  的反卡方分布, 以考察不同先验值对后验均值估计的影响。

对这个性状的约束最大似然 (REML) 估计是利用 MTDFREML 程序<sup>[15]</sup>进行的。迭代收敛的标准为 2 次连续迭代所得的估计值的方差小于  $10^{-9}$ , 重复运行 2 次。

## 2 结果与讨论

本研究使用的过渡期 (2 000 次) 和抽样间隔 (300 次) 皆高于 Gibbsit 程序<sup>[16]</sup>建议的过渡期 (80 次) 及抽样间隔 (27 次), 样本间的自相关系数在 -0.003~0.042 之间, 基本实现独立抽样。

表 1 列出遗传方差组分, 环境方差组分及遗传力的后验密度分布的有关参数。3 个参数的后验密度分布的都呈现正态性, 均值和中数及众数非常接近。后验均值估计的标准差分别为其均数的 8.9%, 8.6%, 5.8%。一般后验均值估计的标准差主要与资料信息量有关, 即资料信息量越大, 参数后验均值估计的标准差越小, 估计的精确度越高<sup>[17,18]</sup>, 但后验密度分布的正态性并不与资料信息量呈正比<sup>[19,20]</sup>。

表1 丹系长白猪乳头数遗传参数的Bayesian 后验估计及REML 估计

参数	样本数	均值	标准差	标准误	中数	众数	偏度	峰度	REML
$\sigma_e^2$	94	1.046	0.089	0.009	1.042	1.049	0.284	-0.423	1.036 <sup>N</sup> *
$\sigma_c^2$	94	0.534	0.046	0.005	0.534	0.514	-0.309	-0.345	0.539 <sup>N</sup>
$h^2$	94	0.661	0.038	0.004	0.663	0.664	0.199	-0.506	0.658 <sup>N</sup>

N: 后验均值估计和REML 估计差异不显著( $P > 0.05$ )。

已有的研究对猪乳头数遗传参数的估计结果存在很大差别, 遗传力估计值范围为 0~1.0, 多数结果认为遗传力小于 0.5<sup>[1-7]</sup>。其中有人认为这个性状可能存在母体遗传效应<sup>[4,5]</sup>。这些结果的不一致主要是由于各研究中猪群的遗传背景, 统计方法, 数据规模及结构的差别造成的。有人认为不同品种间及性别间猪乳头数性状遗传方差组分存在异质性<sup>[18]</sup>。Gibbs 抽样结果表明, 丹系长白猪乳头数加性遗传方差组分, 环境方差组分及遗传力的后验均值估计分别为  $1.046 \pm 0.009$ ,  $0.534 \pm 0.005$ ,  $0.661 \pm 0.004$ , 而 3 个参数相应的 REML 估计值为 1.036, 0.539, 0.658, 后验均值估计和 REML 估计值之间差异不显著( $P > 0.05$ ) (表 1)。

参数的后验均值估计是先验分布和资料信息综合作用的结果, 其中先验分布对后验估计的影响程度随资料信息量的增加而下降<sup>[21]</sup>。而参数的似然分析是贝叶斯分析的特殊情况, 即当参数的先验分布为均匀分布时, 其后验分布就是边际似然分布<sup>[22]</sup>。本研究中, 不同先验分布产生的后验均值估计存在显著差别 (表 2), 这意味着资料信息量还不足以在后验分布中占主导地位, 不同的先验分布对后验均值估计将产生一定影响。当参数的先验分布为均匀分布 (Gibbs1) 或者期望值  $\sigma_e^2 = 0.5$ ,  $\sigma_c^2 = 1.0$  的反卡方分布 (Gibbs3) 时, 加性遗传方差组分, 环境方差组分及遗传力的后验均值估计与 REML 估计皆没有显著差异 ( $P > 0.05$ )。但另外一组先验分布 (Gibbs2) 产生的后验均值估计与 REML 估计值差异显著 ( $P < 0.05$ )。这种差异一部分来自很小的估计误差, 但更多地是由其先验期望值  $\sigma_e^2 = 0.5$ ,  $\sigma_c^2 = 1.0$  与 REML 值的偏差造成的。Van Tassell and Van Vleck (1996) 的模拟结果与本研究相似, 而且认为当参数先验分布与“真值”差别很大时, 后验均值估计的方差将随之提高。

表2 不同先验分布产生的后验均值估计 (mean  $\pm$  Std)

项目	$\sigma_e^2$	$\sigma_c^2$	$h^2$
Gibbs1*	1.046(0.081) a	0.534(0.043) a	0.661(0.035) a
Gibbs2	1.008(0.080) b	0.556(0.042) b	0.644(0.035) b
Gibbs3	1.049(0.086) a	0.531(0.047) a	0.663(0.037) a
REML	1.036 a	0.539 a	0.658 a

Gibbs1: 均匀分布; Gibbs2: 期望值  $\sigma_e^2 = 0.5$ ,  $\sigma_c^2 = 1.0$  的反卡方分布; Gibbs3: 期望值  $\sigma_e^2 = 1.0$ ,  $\sigma_c^2 = 0.5$  的反卡方分布。同一列中含不同字母者为差异显著 ( $P < 0.05$ )。

## 参 考 文 献

- 1 Enfield F D, Rempel W E. Inheritance of teat number and relationship of teat number to various maternal traits in swine. *J Anim Sci*, 1961, 20: 876~ 879
- 2 Skjervold H. Inheritance of teat number in swine and the relationship to performance. *Acta Agric Scand*,

- 1963, 13: 323~ 333
- 3 McKay R M , Rahnefeld G W . Heritability of teat number in swine. *Can J Anim Sci*, 1990, 70: 425~ 430
  - 4 Pumfrey R A , Johnson R K , Cunningham P J , Zimmernan D R . Inheritance of teat number and its relationship to maternal traits in swine. *J Anim Sci*, 1980, 50: 1057~ 1060
  - 5 Clayton G A , Powell J C , Hiley P G . Inheritance of teat number and teat inversion in pigs. *Animal Production*, 1981, 33: 299~ 304
  - 6 Jungst S B , Kuhlers D L . Effect of teat number, teat abnormalities and underline length on litter sizes and weights at 21 and 42 days in swine. *J Anim Sci*, 1983, 57: 802~ 806
  - 7 Toro M A , Dobao M T , Rodriguez J , Silió L . Heritability of a canalized trait: teat number in Iberian pigs. *Genet Sel Evol*, 1986, 18: 173~ 183
  - 8 Henderson C R . Applications of linear models in animal breeding. University of Guelph, 1984
  - 9 张源, 张勤. 畜禽育种中的线性模型. 北京: 北京农业大学出版社, 1993
  - 10 Hofer A . Variance component estimation in animal breeding: a review. *J Animal breeding and Genetics*, 1998, 115: 247~ 265
  - 11 Wang C S . Implementation issues in Bayesian analysis in animal breeding. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*, 1998, 25: 481~ 488
  - 12 Van Tassel C P , Van Vleck L D . Multiple-trait sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. *J Anim Sci*, 1996, 74: 2586~ 2597
  - 13 Hobert J P , Casella G . The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J Am Stat Assoc*, 1996, 91: 1461~ 1473
  - 14 Roehe R . Genetic determination of individual birth weight and its association with sow productivity traits using bayesian analyses. *J Anim Sci*, 1999, 77: 330~ 343
  - 15 Boldman K G , Kriese L A , Van Vleck L D , et al . A manual for use of MTDFREML. A set of programs to obtain estimates of variances and covariances (Draft). U. S. Dept. of Agriculture, Agricultural Research Service, 1995
  - 16 Raftery A E , Lewis S M . The number of iterations, convergence diagnostics and generic Metropolis algorithms. In: Gilks W R , Spiegelhalter D J , Richardson S R , eds. *Practical Markov Chain Monte Carlo*. London, U. K. : Chapman and Hall, 1995
  - 17 Van der Lugt A W , Janss L L G , Van Arendonk J A M . Estimation of variance components in large animal models using Gibbs sampling. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, 1994, 18: 329~ 332
  - 18 Lee C , Wang C D . Estimation of Variance components for teat number in swine using REML and Gibbs sampler. *J Anim Sci*, 1999, 77(Supp1 1): 76(A bstr).
  - 19 Van Tassel C P , Pollak E J . Estimates of (co)variance components for weaning weight used in antional beef cattle genetic evaluations. *J Anim Sci*, 1994, 72(Supp1 1): 148(A bstr )
  - 20 Sorensen D A , Anderson A S , Jensen J , et al . Inferences about genetic parameters using the Gibbs sampler. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, 1994, 18: 321~ 328
  - 21 Gianola D , Fernando R L . Bayesian methods in animal breeding theory. *J Anim Sci*, 1986, 63: 217~ 224
  - 22 Tanner M A . Tools for Statistical Inference: methods for the exploration of posterior distributions and likelihood functions. 3rd edition. Springer Series in Statistics. 1993