华夏英才基金学术文库

动物遗传育种中的计算方法

张　勤　著

科学出版社

北　京

内 容 简 介

本书全面系统地介绍了在动物遗传育种中常用的主要计算方法，内容包括三大部分。第一部分是混合模型方程组的相关计算技术，包括加性遗传相关矩阵及其逆矩阵的计算，大型混合模型方程组的建立与求解，大型矩阵的储存与计算技术，用 REML 方法估计遗传参数的有关计算方法。第二部分是 Monte Carlo 方法，包括随机数的产生，Monte Carlo 基本方法，Monte Carlo 方法在统计学和动物遗传育种中的应用。第三部分是 MCMC 算法，包括贝叶斯推断和 Markov 链简介，Metropolis-Hasting 抽样，Gibbs 抽样，MCMC 算法在动物遗传育种中的应用。

本书可供高等院校和科研机构从事动物遗传育种的科研工作者和研究生参考。

图书在版编目(CIP)数据

动物遗传育种中的计算方法/张勤 著.—北京：科学出版社，2007
(华夏英才基金学术文库)

ISBN 978-7-03-019524-1

Ⅰ. 动… Ⅱ. 张… Ⅲ. 动物-遗传育种-计算方法 Ⅳ. Q953

中国版本图书馆 CIP 数据核字(2007)第 118320 号

责任编辑：夏 梁 莫结胜 杨 然/责任校对：陈玉凤
责任印制：钱玉芬/封面设计：陈 敬

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

双青印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2007 年 8 月第 一 版 开本：B5(720×1000)
2007 年 8 月第一次印刷 印张：11 3/4
印数：1—2 000 字数：224 000

定价：45.00 元
(如有印装质量问题，我社负责调换〈环伟〉)

前　言

动物遗传育种的理论和方法在过去的 20 余年中有了很大的发展，主要体现在：①以混合模型方程组理论为基础的动物个体遗传评估和遗传参数估计方法取代了传统的选择指数法和同胞分析或子亲回归分析，使得对个体的遗传评估更准确，对遗传参数的估计更可靠，由此大幅度提高了一些重要经济性状的遗传改良速度；②贝叶斯理论被逐渐用于动物遗传性状（尤其是非连续分布的性状）的统计分析，使我们对这些性状的统计分析更合理、更有效；③Monte Carlo 方法被广泛应用于遗传育种模拟试验，弥补了用家畜或实验动物进行试验的一些局限性，使我们能够对新的理论和方法进行更广泛的验证、比较，或探索其最佳适用条件；④分子水平的遗传信息被逐渐应用于数量性状的遗传分析及动物育种，使我们能够从分子水平了解数量性状的遗传机理，从而进一步提高育种效率，特别是对那些用现行的方法不能取得理想的改良效果或效率不高的性状。这些新的理论和方法无不涉及大量复杂的计算，这一方面需要有高性能的计算机硬件设备，另一方面则需要有先进科学的计算方法。因而伴随着遗传育种理论和方法的发展，新的计算方法也不断涌现。可以说，一方面是遗传育种理论和方法的发展带来了新的计算方法的产生，反过来，新的计算方法又促进了遗传育种理论和方法的发展。因此，计算方法已成为动物遗传育种研究和应用中不可缺少的关键技术。不了解这些方法，就很难真正理解现代遗传育种的理论和方法，不掌握这些方法，就难以开展遗传育种领域的深入研究。尽管现在已经有了一些专门的计算机软件可以帮助我们利用这些计算方法开展研究或实际应用，但如果不对这些方法本身有所了解，就难以正确使用软件，对所得到的结果也难以给出合理的解释。

本书的目的是要向我国动物遗传育种工作者全面系统地介绍在动物遗传育种中使用的一些主要计算方法。我多年从事动物遗传育种中相关计算方法的教学和研究，本书是在本人多年积累的基础上，汇集国内外本研究领域中 20 余年来的研究成果写成的。我衷心希望本书能成为对我国动物遗传育种工作者有价值的一本参考书。尽管已尽我所能，但限于水平，错误和不足之处仍在所难免，希望读者不吝赐教。

我是在我国数量遗传学奠基人吴仲贤教授的引导下逐步进入数量遗传学殿堂的，吴仲贤教授在我的成长过程中倾注了大量心血和关爱，没有他的培养和教导，就不可能有本书的问世，值此书出版之际，我谨向吴仲贤教授表示深深的谢意。在我多年的研究和教学以及在本书的写作过程中，经常得到吴常信教授、张沅教授和

其他同事的指导和帮助，在此也向他们表示衷心的感谢。最后要特别感谢华夏英才学术出版基金给予的大力资助。

张　勤

2007 年 6 月于北京

目　录

第一部分　混合模型方程组的相关计算技术

在现代家畜育种的理论和实践中，线性模型(linear model)的理论和方法占有十分重要的地位。目前，以 Henderson 为代表所发展起来的混合模型方程组方法已在家畜个体遗传评定(育种值估计)、群体遗传参数估计、杂交试验分析等方面占据了主导地位，但这种发展趋势是与计算机技术和计算方法的迅速发展密切相关的。因为线性模型方法中涉及了大量的矩阵运算，而家畜育种资料又往往十分庞大和复杂，所以在实际应用中，如何快捷、准确、有效地完成各种复杂的计算工作就成了一个十分突出的问题。本部分将介绍近年来在这方面发展起来的相关计算技术。

第 1 章　线性混合模型及混合模型方程组简介

1.1　线性混合模型

1.1.1　模型

在统计学中，模型(model)或数学模型，是指描述观察值与影响观察值变异性的各因子(factor)之间关系的数学表达式。所有的统计分析都是基于一定的模型基础上的。一个模型应恰当地反映数据资料的性质和所要解决的问题。有各种不同水平的模型：

(1)真实模型　非常准确地描述观察值的变异性，模型中不含有未知成分，对于生物学领域的数据资料来说，真实模型几乎是不可能知道的。

(2)理想模型　根据研究者所掌握的专业知识建立的尽可能接近真实模型的模型，这种模型常常由于受到数据资料的限制或过于复杂而不能用于实际分析。

(3)操作模型　用于实际统计分析的模型，它通常是理想模型的简化形式。

影响观察值的因素也称为变量(variable)，它们可分为两类，一类是离散型的，它们通常具有若干个有限的等级或水平(如季节、胎次等)，通过统计分析，我们可估计不同水平对观察值的效应的大小，并检验不同水平间有无显著差异；另一类是连续型的，它呈现连续性变异(如体重、体长等)，它们通常是作为影响观察值的协变量(如同回归模型中的自变量，也称为回归变量)来看待的，通常需要估计的是观察值对这一变量的回归系数，有时一个连续性变量也可人为地划分成若干等级而使其变为离散型变量。

离散型因子又可进一步分为固定因子和随机因子。区别一个因子是固定因子还是随机因子主要看样本的取得方法和研究的目的。如果对于一个因子我们有意识地抽取它的若干个特定的水平，而研究的目的也只是要对这些水平的效应进行估计或进行比较，则该因子就是固定因子。反之，若一个因子的若干水平可看作是来自该因子的所有可能水平所构成的一个大总体的随机样本，研究的目的是要通过该样本去推断总体，则该因子就是随机因子。

1.1.2　线性模型

在统计模型中线性模型占有很重要的地位。所谓线性模型是指在模型中观察值与各个离散型因子和连续型协变量的回归系数呈线性关系(协变量本身与观察值可以是非线性的，如多项式回归模型仍然是线性模型)。

一个线性模型应由三个部分组成：

(1)数学方程式；

(2)方程式中随机变量的期望和方差及协方差；

(3)假设及约束条件。

下面举例说明：

设有犊牛 190～210 日龄的体重资料，将日龄按每 5 天间隔分组，190～210 日龄就可分为 4 组。欲分析不同日龄组对体重的影响，可建立如下的线性模型：

$$y_{ij} = \mu + a_i + e_{ij} \tag{1.1}$$

其中，y_{ij} 是在第 i 个日龄组中的第 j 头肉牛的体重，为可观察的随机变量；μ 是总平均，是一个常量；a_i 是第 i 个日龄组的效应，它是固定效应；e_{ij} 是剩余效应，也称为随机误差。

式中随机变量的期望和方差及协方差为

$$E(e_{ij}) = 0, \quad E(y_{ij}) = \mu + a_i$$

$$\mathrm{Var}(y_{ij}) = \mathrm{Var}(e_{ij}) = \sigma^2$$

$$\mathrm{Cov}(e_{ij}, e_{ij'}) = \mathrm{Cov}(e_{ij}, e_{i'j}) = \mathrm{Cov}(e_{ij}, e_{i'j'}) = 0$$

此模型的假设和约束条件包括：

(1)所有犊牛都来自同一品种；

(2)母亲的年龄对犊牛体重无影响；

(3)犊牛的性别相同或性别对体重无影响；

(4)所有犊牛都在相同的环境下以相同的饲养方式饲养。

日龄组	犊牛体重		
1	198	204	201
2	203	206	210
3	205	212	216
4	225	220	

如果我们的资料如右表所示：

则对每一观察值都可根据上面的模型建立一个方程式：

$$\begin{aligned}
y_{11} &= 198 = \mu + a_1 + e_{11} \\
y_{12} &= 204 = \mu + a_1 + e_{12} \\
y_{13} &= 201 = \mu + a_1 + e_{13} \\
y_{21} &= 203 = \mu + a_2 + e_{21} \\
y_{22} &= 206 = \mu + a_2 + e_{22} \\
y_{23} &= 210 = \mu + a_2 + e_{23} \\
y_{31} &= 205 = \mu + a_3 + e_{31} \\
y_{32} &= 212 = \mu + a_3 + e_{32} \\
y_{33} &= 216 = \mu + a_3 + e_{33} \\
y_{41} &= 225 = \mu + a_4 + e_{41} \\
y_{42} &= 220 = \mu + a_4 + e_{42}
\end{aligned}$$

令

$$\boldsymbol{y}=\begin{bmatrix} y_{11}\\ y_{12}\\ y_{13}\\ y_{21}\\ y_{22}\\ y_{23}\\ y_{31}\\ y_{32}\\ y_{33}\\ y_{41}\\ y_{42} \end{bmatrix},\quad \boldsymbol{X}=\begin{bmatrix} 1&1&0&0&0\\ 1&1&0&0&0\\ 1&1&0&0&0\\ 1&0&1&0&0\\ 1&0&1&0&0\\ 1&0&1&0&0\\ 1&0&0&1&0\\ 1&0&0&1&0\\ 1&0&0&1&0\\ 1&0&0&0&1\\ 1&0&0&0&1 \end{bmatrix},\quad \boldsymbol{b}=\begin{bmatrix} \mu\\ a_1\\ a_2\\ a_3\\ a_4 \end{bmatrix},\quad \boldsymbol{e}=\begin{bmatrix} e_{11}\\ e_{12}\\ e_{13}\\ e_{21}\\ e_{22}\\ e_{23}\\ e_{31}\\ e_{32}\\ e_{33}\\ e_{41}\\ e_{42} \end{bmatrix}$$

则我们有

$$\begin{gathered} \boldsymbol{y}=\boldsymbol{Xb}+\boldsymbol{e}\\ E(\boldsymbol{e})=\boldsymbol{0},\quad E(\boldsymbol{y})=\boldsymbol{Xb}\\ \mathrm{Var}(\boldsymbol{y})=\mathrm{Var}(\boldsymbol{e})=\boldsymbol{I}\sigma^2 \end{gathered} \tag{1.2}$$

其中，$\boldsymbol{0}$ 是元素全为 0 的向量；$\boldsymbol{I}$ 是单位矩阵。

(1.2)式称为线性模型的矩阵表达式，矩阵 $\boldsymbol{X}$ 称为关联矩阵或结构矩阵，其中的元素均为 0 或 1，它们指示了 $\boldsymbol{y}$ 中的观察值与 $\boldsymbol{b}$ 中的各效应的关联情况。

1.1.3 线性模型的分类

线性模型可以从不同的角度进行分类，从其功能上分，可分为回归模型、方差分析模型、协方差分析模型、方差组分模型等；按模型中含有的因子个数分，可有单因子模型、双因子模型、多因子模型；按模型中因子的性质(固定的还是随机的)可分为固定效应模型、随机效应模型和混合模型。这里只介绍按因子的性质分类的情况。

1)固定效应模型

如一个模型中除了随机误差外，其余所有的效应均为固定效应，则称此模型为固定效应模型，或固定模型(fixed model)。

2)随机效应模型

若模型中除了总平均外，其余的所有效应均为随机效应，则称此模型为随机效应模型，或随机模型(random model)。

3)混合模型

若模型中除了总平均和随机误差之外，既含有固定效应，也含有随机效应，则

称之为混合模型(mixed model)。

一般地,对于任一混合模型,都可用矩阵的形式表示为

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \tag{1.3}$$

其中,$\boldsymbol{y}$ 为所有观察值构成的向量;$\boldsymbol{b}$ 为所有固定效应(包括 μ)构成的向量;$\boldsymbol{X}$ 为固定效应的关联矩阵;$\boldsymbol{u}$ 为所有随机效应构成的向量;$\boldsymbol{Z}$ 为随机效应的关联矩阵;$\boldsymbol{e}$ 为所有随机误差构成的向量。

通常我们假设

$$E(\boldsymbol{u}) = \boldsymbol{0}, \quad E(\boldsymbol{e}) = \boldsymbol{0}, \quad \mathrm{Var}(\boldsymbol{e}) = \boldsymbol{R}, \quad \mathrm{Var}(\boldsymbol{u}) = \boldsymbol{G}, \quad \mathrm{Cov}(\boldsymbol{u}, \boldsymbol{e}') = \boldsymbol{0}$$

由此可进一步得到

$$E(\boldsymbol{y}) = \boldsymbol{Xb}, \quad \mathrm{Var}(\boldsymbol{y}) = \boldsymbol{ZGZ}' + \boldsymbol{R} = \boldsymbol{V}$$

$$\mathrm{Cov}(\boldsymbol{y}, \boldsymbol{u}') = \boldsymbol{ZG}, \quad \mathrm{Cov}(\boldsymbol{y}, \boldsymbol{e}') = \boldsymbol{R}$$

在多数情况下,常假设 $\boldsymbol{R} = \boldsymbol{I}\sigma_e^2$,即随机误差具有等方差性和独立性。

若上式中的 $\boldsymbol{Zu}$ 不存在时,则它变为一固定模型

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{e}$$

若上式中 $\boldsymbol{Xb} = \boldsymbol{1}\mu$,则它变为一个随机模型

$$\boldsymbol{y} = \boldsymbol{1}\mu + \boldsymbol{Zu} + \boldsymbol{e}$$

其中,$\boldsymbol{1}$ 为元素全为 1 的向量。

因此,固定模型和随机模型均可看成是混合模型的特例。

1.1.4 统计模型与遗传模型的关系

根据数量遗传学理论,任何一个个体在任一数量性状上的表现都要受到遗传和环境两个方面的影响,这可用下面的模型来表示:

$$Y = G + E \tag{1.4}$$

其中,Y 是该个体的性状表型值,它就是统计模型中的观察值;G 是遗传效应;E 是环境效应。

遗传效应是遗传基因对所考察的性状的影响,这些基因可以是个体本身所携带的,也可以是其亲属所携带的。遗传效应又可分为加性效应(G_A)、显性效应(G_D)和上位效应(G_I)。根据数量遗传理论,数量性状要受许多基因的影响,而每个基因的作用又很小,故称为微效基因,各个基因的效应是可加的,加性效应就是所有微效基因效应累加之和。显性效应是指同一基因座内不同等位基因之间的互作,上位效应是不同基因座间的基因的互作,显性效应和上位效应统称非加性效应。由于亲代只将其基因而不是基因型传递给后代,而显性效应和上位效应都与特定的基因型有关,所以只有加性效应才能遗传给后代,因而对于种畜的选择来说,我们主要感兴趣的是 G_A,它也常被称为家畜个体的育种值。一个亲代传递给后代的加性效应的平均值称为遗传传递力(transmitting ability),由于亲代只将其

所有基因的一半传递给后代(另一半来自另一亲代),所以传递力等于 1/2 的育种值。

环境效应可进一步分为系统环境效应(E_S)和随机环境效应(E_R)。系统环境效应是指那些以固定的相同的方式影响在该环境下所有个体的效应,如畜群(包括畜群内的管理水平、饲养方式等)、性别、季节、年龄等,这些效应只要有相应的资料是可以被估计的。随机环境效应是指那些以随机的方式影响家畜个体的环境效应。当一个个体在某一性状上可重复表现时(如奶牛的多个胎次的泌乳记录、母猪的不同胎次的产仔数等),则随机环境效应又可进一步分为永久性环境效应(E_{RP})和暂时性环境效应(E_{RT})。前者是指可影响一个个体在该性状上的所有各次表现的环境效应,如某一个体在早期发育阶段营养不良,而且所造成的后果又是不可逆的,这就可能会对其产生终生的影响,即影响其所有重复观察值。暂时性环境效应只影响某一特定的观察值,如某个体在某一生产时期偶然得病,则该个体在这一时期的生产性能就会受到影响。对性状的度量误差也可看成是一种暂时性环境效应。

综上所述,我们可将(1.4)式扩展为

$$Y = E_S + G_A + G_D + G_I + E_{RP} + E$$

$$P \tag{1.5}$$

$$E^*$$

由于 G_D 和 G_I 也影响一个个体的终生,而我们常常并不需要知道它们的大小,故可将它们与 E_{RP} 合并为 P,当个体没有重复观察值时,我们不能区分永久性的和暂时性的随机环境效应,此时可将 P 和 R_{RP} 合并为 E^*。

显然,这种遗传模型也是线性的,我们只需要确定这些环境和遗传效应中哪些是固定效应,哪些是随机效应,就可将其转换为统计模型。如前所述,在环境效应中,系统环境效应通常是固定效应,随机环境效应则是随机的。遗传效应是随个体而异的,而我们所要分析的个体往往是来自某个总体的随机样本,因此遗传效应一般是随机效应。在实际的应用中,只要根据具体的问题来确定模型中应包含的效应从而得到一个操作模型。

1.1.5 遗传分析中的常用模型

在动物育种中,常根据对遗传效应的不同考虑方式将遗传分析模型分为以下几种:

(1)动物模型 将直接影响表型值的个体本身的育种值(即基因的加性效应)作为遗传效应放在模型中,这种模型被称为个体动物模型,简称动物模型(animal model),可表示为

$$y_i = \sum_{l=1}^{r} b_l + a_i + e_i \tag{1.6}$$

其中，y_i是第 i 个个体的观察值；b_l是第 l 个系统环境效应(固定效应)；a_i是该个体的加性遗传效应(育种值)；e_i是随机误差(主要由随机环境效应所致)。注意在这个模型中并没有考虑基因的非加性效应，它们实际上被归入了随机误差中。

假如我们有 n 个个体的观察值，需要对 s 个个体估计育种值($s \geqslant n$)，则对这 n 个观察值可用如下的以矩阵表示的模型来描述：

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e} \tag{1.7}$$

其中，$\boldsymbol{a}$ 是 s 个个体的育种值向量；$\boldsymbol{Z}$ 为 $\boldsymbol{a}$ 的关联矩阵，当 $\boldsymbol{a}$ 中的所有个体都有观察值时(即 $s=n$)；$\boldsymbol{Z}=\boldsymbol{I}$。通常假设 $\boldsymbol{a}$ 的方差-协方差矩阵为

$$\boldsymbol{G} = \mathrm{Var}(\boldsymbol{a}) = \boldsymbol{A}\sigma_a^2$$

其中，$\boldsymbol{A}$ 为 s 个个体间的加性遗传相关矩阵；σ_a^2 为加性遗传方差。$\boldsymbol{e}$ 是随机环境效应向量，通常假设随机环境效应间彼此独立，且具有相同的方差，故有

$$\boldsymbol{R} = \mathrm{Var}(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2$$

当个体在被考察的性状上有重复观察值时，可将个体的一个观察值表示为

$$y_{ij} = \sum_{l=1}^{r} b_l + a_i + p_i + e_{ij} \tag{1.8}$$

其中，y_{ij}为个体 i 的第 k 个观察值；b_l为影响该观察值的第 l 个系统环境效应；a_i为该个体的育种值；p_i为影响个体 i 所有观察值的永久性环境效应；e_{ij}为影响 y_{ij}的随机暂时性环境效应(随机误差)。

用矩阵形式表示，则有

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Z}_1\boldsymbol{a} + \boldsymbol{Z}_2\boldsymbol{p} + \boldsymbol{e}$$

$$\mathrm{Var}(\boldsymbol{a}) = \boldsymbol{A}\sigma_a^2, \quad \mathrm{Var}(\boldsymbol{p}) = \boldsymbol{I}\sigma_p^2, \quad \mathrm{Cov}(\boldsymbol{a}, \boldsymbol{p}') = \boldsymbol{0}, \quad \mathrm{Var}(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2 \tag{1.9}$$

若令

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{p} \end{bmatrix}, \quad \boldsymbol{Z} = [\boldsymbol{Z}_1 \quad \boldsymbol{Z}_2]$$

则(1.9)式就成为(1.3)式，且有

$$\boldsymbol{G} = \mathrm{Var}(\boldsymbol{u}) = \mathrm{Var}\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}\sigma_a^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}\sigma_p^2 \end{bmatrix}, \quad \boldsymbol{R} = \mathrm{Var}(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2$$

(2)公畜模型　在动物模型中，随机遗传效应为个体的育种值，由于个体的基因一半来自父亲，另一半来自母亲，所以任一个体的育种值 a_i都可表示为

$$a_i = 0.5a_s + 0.5a_d + m_i \tag{1.10}$$

其中，a_s为个体 i 父亲的育种值；a_d为其母亲的育种值；m_i为基因从亲代到子代传递过程中由于配子的随机抽样所造成的随机离差，称为孟德尔抽样(Mendelian sampling)离差。于是动物模型(1.6)式可写为

$$y_i = \sum_{l=1}^{r} b_l + 0.5a_s + 0.5a_d + m_i + e_i \tag{1.11}$$

所谓公畜模型，就是将(1.11)式中的最后三项($0.5a_d + m_i + e_i$)合并成随机误差项，将 $0.5a_s$ 表示为 s_i，它等于个体 i 的父亲的遗传传递力，称为父亲或公畜效应，因而有

$$y_{ij} = \sum_{l=1}^{r} b_l + s_i + \varepsilon_{ij} \tag{1.12}$$

或以矩阵形式表示

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zs} + \varepsilon \tag{1.13}$$

$$E(\boldsymbol{s}) = \boldsymbol{0}, \quad E(\varepsilon) = \boldsymbol{0}$$

$\mathrm{Var}(\boldsymbol{s}) = \boldsymbol{A}_s\sigma_s^2$，$\boldsymbol{A}_s$ 为父亲之间的加性遗传相关矩阵，σ_s^2 为公畜方差$=1/4\sigma_a^2$

$$\mathrm{Var}(\varepsilon) = \boldsymbol{I}\sigma_\varepsilon^2。$$

这个模型也可扩展成有重复观察值时的情形。

公畜模型只可用来估计公畜的育种值，而且有三个重要假设：

①公畜在群体中与母畜的交配是完全随机的；

②父亲与母亲之间、不同母亲之间没有血源关系；

③每个母亲只有一个后代，即一个公畜的所有后代都是父系半同胞。

这些假设在生产实际中一般是很难满足的。尽管如此，这个模型由于在计算上比动物模型要简单易行，因而在 19 世纪 70 年代和 80 年代前期在奶牛育种中得到了广泛应用。

(3)公畜-母畜模型　如果将(1.11)式中的最后两项($m_i + e_i$)合并为随机误差项，并将 $0.5a_d$ 表示为 d_i，它等于个体 i 的母亲的遗传传递力，称为母亲或母畜效应，于是(1.11)式可重写为

$$y_{ijk} = \sum_{l=1}^{r} b_l + s_i + d_{ij} + \varepsilon_{ijk} \tag{1.14}$$

或以矩阵形式表示

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Z}_s\boldsymbol{s} + \boldsymbol{Z}_d\boldsymbol{d} + \varepsilon \tag{1.15}$$

其中，$\boldsymbol{s}$=父亲效应向量(同公畜模型)；$\boldsymbol{d}$=母亲效应向量，$E(\boldsymbol{d}) = \boldsymbol{0}$，$\mathrm{Var}(\boldsymbol{d}) = \boldsymbol{A}_d\sigma_d^2$；$\boldsymbol{A}_d$ 为母亲之间的加性遗传相关矩阵，σ_d^2 为母畜方差$=1/4\sigma_a^2$；ε=随机误差向量，$E(\varepsilon) = \boldsymbol{0}$，$\mathrm{Var}(\varepsilon) = \boldsymbol{I}\sigma_\varepsilon^2$。

(4)外祖父模型　若将(1.10)式中的母亲育种值 a_d 按同样的方式作进一步的剖分，即表示为

$$a_d = 0.5a_{\mathrm{mgs}} + 0.5a_{\mathrm{mgd}} + m_d$$

其中，a_{mgs} 为母亲的父亲(外祖父，maternal grandsire)的育种值；a_{mgd} 为母亲的母亲(外祖母，maternal granddam)的育种值；m_d 为母亲育种值的孟德尔抽样离差。于是，个体 i 的育种值 a_i 可表示为

$$a_i = 0.5a_s + 0.5 \times (0.5a_{mgs} + 0.5a_{mgd} + m_d) + m_i$$
$$= 0.5a_s + 0.25a_{mgs} + 0.25a_{mgd} + m^*$$

其中，$m^* = 0.25m_d + m_i$。将其代入(1.6)式中，并将 $0.25a_{mgd} + m^*$ 并入随机误差中，于是得到

$$y_{ijk} = \sum_{l=1}^{r} b_l + s_i + mgs_{ij} + \varepsilon_{ijk}$$

其中，$mgs_{ij} = 0.25a_{mgs}$，称为外祖父效应。此模型即被称为外祖父模型(Maternal grandsire model)。

无论何种模型，都可表示为如(1.3)式的混合模型的一般形式。而无论从统计学还是从遗传学的观点看，动物模型都要优于其他模型，随着计算机技术和计算方法的日益完善，其他模型在育种实践中逐渐被淘汰，而动物模型的应用则越来越广泛。

1.2 混合模型方程组

Henderson(1963)证明，对于如(1.3)式表示的任意的线性混合模型，其中 $\boldsymbol{b}$ 的最佳线性无偏估计(BLUE)和 $\boldsymbol{u}$ 的最佳线性无偏预测(BLUP)可通过对以下的方程组求解而得到：

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix} \tag{1.16}$$

此方程组即称为混合模型方程组(mixed model equations，MME)。

对于动物模型(1.7)式，将 $\boldsymbol{u} = \boldsymbol{a}, \boldsymbol{G} = \boldsymbol{A}\sigma_a^2$ 和 $\boldsymbol{R} = \boldsymbol{I}\sigma_e^2$ 代入(1.16)式，可得

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{y} \end{bmatrix} \tag{1.17}$$

其中，$\lambda = \sigma_e^2 / \sigma_a^2$。

对此方程组求解，可得个体育种值的估计值。

对于动物模型(1.9)式，将

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{p} \end{bmatrix}, \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{Z}_2 \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} \boldsymbol{A}\sigma_a^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}\sigma_p^2 \end{bmatrix}, \quad \boldsymbol{R} = \boldsymbol{I}\sigma_e^2$$

代入(1.16)式，可得

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z}_1 & \boldsymbol{X}'\boldsymbol{Z}_2 \\ & \boldsymbol{Z}_1'\boldsymbol{Z}_1 + \boldsymbol{A}^{-1}\lambda_1 & \boldsymbol{Z}_1'\boldsymbol{Z}_2 \\ \text{对称} & & \boldsymbol{Z}_2'\boldsymbol{Z}_2 + \boldsymbol{I}\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{p}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{Z}_1'\boldsymbol{y} \\ \boldsymbol{Z}_2'\boldsymbol{y} \end{bmatrix} \tag{1.18}$$

其中，$\lambda_1 = \sigma_e^2 / \sigma_a^2$；$\lambda_2 = \sigma_e^2 / \sigma_p^2$。

对于如(1.12)式的公畜模型，$\boldsymbol{u}=\boldsymbol{s}$，$\boldsymbol{G}=\boldsymbol{A}_s\sigma_s^2$ 和 $\boldsymbol{R}=\boldsymbol{I}\sigma_e^2$ 代入(1.16)式，可得

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z}+\boldsymbol{A}_s^{-1}\lambda_s \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{s}} \end{bmatrix}=\begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{y} \end{bmatrix} \tag{1.19}$$

其中，$\lambda_s=\sigma_e^2/\sigma_s^2$。

第 2 章　加性遗传相关矩阵及其逆矩阵的计算

由第 1 章的介绍可看出，在利用混合模型方程组估计育种值时，首先要求得到个体间的加性遗传相关矩阵 $\boldsymbol{A}$ 的逆矩阵 $\boldsymbol{A}^{-1}$。虽然从理论上说我们总可以根据 $\boldsymbol{A}$ 的定义求出 $\boldsymbol{A}$，然后再对 $\boldsymbol{A}$ 求逆而求得 $\boldsymbol{A}^{-1}$，但这在涉及的个体数较多时是非常困难乃至是不可能实现的，因而一些学者又提出了一些求 $\boldsymbol{A}$ 和 $\boldsymbol{A}^{-1}$ 的特殊方法。

2.1　$\boldsymbol{A}$ 的计算方法

个体 i 和 j 之间的加性遗传相关是指在它们的基因组中具有同源相同(identical by decent，IBD)基因(相同的且来自同一祖先的基因)的比例，或者说从个体 i 的基因组中随机抽取的一个基因与在个体 j 的基因组中随机抽取的一个基因同源相同的概率。它也可被理解为个体 i 和 j 的育种值(加性遗传值)之间的相关(故而称之为加性遗传相关)。根据这个定义，$\boldsymbol{A}$ 阵中的元素 a_{ij}，即个体 i 和 j 之间的加性遗传相关的计算通式为

$$a_{ij} = \sum \left(\frac{1}{2} \right)^{n_i + n_j} (1 + f_A) \tag{2.1}$$

其中，n_i 和 n_j 分别为连接个体 i 和个体 j 的一个通径中由 i 和 j 到它们的共同祖先 A 的世代数；f_A 为 A 的近交系数；$\sum$ 表示当连接个体 i 和 j 的通径不止一个时，要对所有的通径求和。

$\boldsymbol{A}$ 阵中的对角线元素 a_{ii}，即个体 i 与其自身的加性遗传相关被定义为

$$a_{ii} = 1 + f_i \tag{2.2}$$

其中，f_i 为个体 i 的近交系数。

由于由(2.1)式所定义的加性遗传相关恰好是 Wright(1922)所定义的亲缘系数(coefficient of relationship)计算公式中的分子，故又称其为分子亲缘相关(numerator relationship)。

利用(2.1)式和(2.2)式计算 $\boldsymbol{A}$ 阵时，一般要求画出完整的系谱图，这在系谱大而复杂时是很困难的。下面我们介绍另一种计算 $\boldsymbol{A}$ 阵的十分简便的方法，用这种方法，无论系谱多大多复杂，都能很容易地计算出 $\boldsymbol{A}$ 阵，且便于在计算机上实现。

2.1.1　动物模型下 $\boldsymbol{A}$ 阵的计算

在动物模型下，$\boldsymbol{A}$ 阵是所有动物个体之间的加性遗传相关矩阵，$\boldsymbol{A}$ 阵的每一元

素可用以下的递推公式来计算：

$$a_{ii}=\begin{cases}1+0.5a_{s_i d_i}, & \text{当 } s_i \text{ 和 } d_i \text{ 均已知}\\ 1, & \text{当 } s_i \text{ 或 } d_i \text{ 未知}\end{cases} \tag{2.3a}$$

$$a_{ij}=a_{ji}=\begin{cases}0.5(a_{is_j}+a_{id_j}), & \text{当 } s_j \text{ 和 } d_j \text{ 均已知}\\ 0.5a_{is_j}, & \text{当 } s_j \text{ 已知，} d_j \text{ 未知}\\ 0.5a_{id_j}, & \text{当 } d_j \text{ 已知，} s_j \text{ 未知}\\ 0, & \text{当 } s_j \text{ 和 } d_j \text{ 均未知}\end{cases} \tag{2.3b}$$

其中，$s_i(s_j)$和 $d_i(d_j)$为个体 $i(j)$的父亲和母亲。

在利用以上公式计算 $\boldsymbol{A}$ 阵时，要先将系谱中的所有个体按个体号、父号和母号列成一个三列表(数据文件)，在列表时应注意：

(1)在个体一列中应包括所有在父和母列出现过的个体。

(2)在个体一列中应保证后代绝不会出现在其父母之前，一般可按出生日期排序，先出生的在前。

(3)为便于编写程序，个体应用自然数从 1 开始连续编号。

下面举例说明。

例 2.1 设有 7 个个体，列如右表。

个体	父	母
1	—	—
2	—	—
3	1	—
4	1	2
5	3	4
6	1	4
7	5	6

右表中个体 1 和个体 2 的双亲未知，个体 3 的母亲未知。对于双亲未知的个体，我们假设它们都是非近交个体，且彼此无亲缘关系，这些个体就构成了所谓的基础群(base population)。这些个体所对应的 $\boldsymbol{A}$ 阵中的子矩阵为一单位阵。对于本例来说：

$$a_{11}=1,\quad a_{22}=1,\quad a_{12}=a_{21}=0$$

从这些元素出发可计算出 $\boldsymbol{A}$ 中所有的其他元素，如

$$\begin{aligned}a_{13}&=a_{31}=0.5a_{11}=0.5\\ a_{23}&=a_{32}=0\\ a_{33}&=1\\ a_{14}&=a_{41}=0.5(a_{11}+a_{21})=0.5\\ a_{44}&=1+0.5a_{12}=1\end{aligned}$$

完整的 $\boldsymbol{A}$ 阵为

$$\boldsymbol{A}=\begin{bmatrix}1 & 0 & 0.5 & 0.5 & 0.5 & 0.75 & 0.625\\ & 1 & 0 & 0.5 & 0.25 & 0.25 & 0.25\\ & & 1 & 0.25 & 0.625 & 0.375 & 0.5\\ & & & 1 & 0.625 & 0.75 & 0.6875\\ & & & & 1.125 & 0.5625 & 0.84375\\ & & 对 & 称 & & 1.25 & 0.90625\\ & & & & & & 1.28125\end{bmatrix}$$

2.1.2 公畜模型下 $\boldsymbol{A}$ 阵的计算

在公畜模型中，$\boldsymbol{A}$ 阵为公畜个体间的加性遗传相关矩阵，而在用上述的方法构建 $\boldsymbol{A}$ 阵时，母畜也必须包括在内，因而必须从所得到的 $\boldsymbol{A}$ 阵中将母畜对应的行和列去掉，这样做过于繁琐。另一个途径是根据系谱中公畜的父亲和外祖父的信息来构造 $\boldsymbol{A}$ 阵，为此先建立一个含有公畜号、父亲号和外祖父号三列的表，列表的原则和前面相同，只是这里的公畜对应于前面的个体，外祖父对应于前面的母亲。然后用如下的递推公式计算 $\boldsymbol{A}$ 阵中的每一元素：

$$a_{ii}=\begin{cases}1+0.25a_{s_ig_i}, & \text{当 } s_i \text{ 和 } g_i \text{ 均已知}\\ 1, & \text{其他}\end{cases} \tag{2.4a}$$

$$a_{ij}=a_{ji}=\begin{cases}0.5a_{is_j}+0.25a_{ig_j}, & \text{当 } s_j \text{ 和 } g_j \text{ 均已知}\\ 0.5a_{is_j}, & \text{当 } s_j \text{ 已知，} g_j \text{ 未知}\\ 0.25a_{ig_j}, & \text{当 } g_j \text{ 已知，} s_j \text{ 未知}\\ 0, & \text{当 } s_j \text{ 和 } g_j \text{ 均未知}\end{cases} \tag{2.4b}$$

其中，s_i（s_j）和 g_i（g_j）为公畜 i（j）的父亲和外祖父。

公畜	父亲	外祖父
1	—	—
2	1	—
3	1	2
4	1	2
5	3	4
6	3	4
7	5	6

下面也举例说明：

例 2.2 设有系谱资料如左表格：

根据(2.4 a)式和(2.4b)式，有

$$a_{11}=1$$
$$a_{12}=0.5a_{11}=0.5$$
$$a_{13}=0.5a_{11}+0.25a_{12}=0.625$$
$$a_{14}=0.5a_{11}+0.25a_{12}=0.625$$
$$a_{15}=0.5a_{13}+0.25a_{14}=0.46875$$
$$\vdots$$

完整的 $\boldsymbol{A}$ 阵为

$$\boldsymbol{A}=\begin{bmatrix}1.0 & 0.5 & 0.625 & 0.625 & 0.46875 & 0.46875 & 0.3515625\\ & 1.0 & 0.5 & 0.5 & 0.375 & 0.375 & 0.28125\\ & & 1.125 & 0.4375 & 0.671875 & 0.671875 & 0.5\\ & & & 1.125 & 0.5 & 0.5 & 0.375\\ & & & & 1.109375 & 0.4609375 & 0.669921875\\ & & \text{对} & \text{称} & & 1.109375 & 0.5078125\\ & & & & & & 1.115234375\end{bmatrix}$$

2.2 $\boldsymbol{A}^{-1}$的计算方法

从理论上说$\boldsymbol{A}^{-1}$可以通过对$\boldsymbol{A}$求逆得到，但当$\boldsymbol{A}$阵很大时，对它求逆是十分困难以致不可能的。幸运的是，Henderson(1975a，1976)提出了一个可以从系谱直接构造$\boldsymbol{A}^{-1}$(不需要构造$\boldsymbol{A}$)的简捷方法。正是由于这一方法的提出，才使得BLUP法，尤其是动物模型BLUP法在家畜育种中的真正广泛使用成为可能。

2.2.1 动物模型下$\boldsymbol{A}^{-1}$的计算

由于$\boldsymbol{A}$阵为对称矩阵，它总可被分解为

$$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{L}'$$

其中，$\boldsymbol{L}$为下三角矩阵，它又可进一步分解为

$$\boldsymbol{L}=\boldsymbol{T}\boldsymbol{D}$$

其中，$\boldsymbol{D}$是对角线矩阵，其对角线元素(d_i)等于$\boldsymbol{L}$阵的对角线元素(l_{ii})，即$d_i=l_{ii}$；$\boldsymbol{T}$阵为下三角矩阵，其对角线元素全为1。

由此，$\boldsymbol{A}^{-1}$可写为

$$\boldsymbol{A}^{-1}=(\boldsymbol{L}\boldsymbol{L}')^{-1}=(\boldsymbol{T}\boldsymbol{D}\boldsymbol{D}\boldsymbol{T}')^{-1}=(\boldsymbol{T}')^{-1}\boldsymbol{D}^{-2}\boldsymbol{T}^{-1} \tag{2.5}$$

Henderson证明，$\boldsymbol{T}^{-1}$中的对角线元素全为1，在其第i行上，第i个个体的每一已知亲本所对应的元素为-0.5，其余元素为0。如对于2.1节中的例2.1来说，我们有

$$\boldsymbol{T}^{-1}=\begin{bmatrix}
1 & & & & & & \\
0 & 1 & & & & & \\
-0.5 & 0 & 1 & & & & \\
-0.5 & -0.5 & 0 & 1 & & & \\
0 & 0 & -0.5 & -0.5 & 1 & & \\
-0.5 & 0 & 0 & -0.5 & 0 & 1 & \\
0 & 0 & 0 & 0 & -0.5 & -0.5 & 1
\end{bmatrix}$$

而$\boldsymbol{D}^{-2}$仍是对角线矩阵，设d_i^*为其对角线上的第i个元素，Henderson证明

$$d_i^*=\begin{cases}
\dfrac{1}{0.5-0.25(f_s+f_d)}, & \text{当个体 } i \text{ 的双亲 } s \text{ 和 } d \text{ 已知} \\
\dfrac{1}{0.75-0.25f_p}, & \text{当个体 } i \text{ 的一个亲本 } p \text{ 已知} \\
1, & \text{当个体 } i \text{ 的双亲未知}
\end{cases} \tag{2.6}$$

其中，f_s、f_d和f_p为s、d和p的近交系数。如果这些近交系数已知，则由(2.6)式可以很容易求得$\boldsymbol{D}^{-2}$，进而求得$\boldsymbol{A}^{-1}$。

如果我们已经求出了$\boldsymbol{A}$阵，则这些近交系数可以很容易地从$\boldsymbol{A}$阵的对角线元

素求得，因为 $\boldsymbol{A}$ 阵对角线元素为 $a_{ii}=1+f_i$，因而 $f_i=a_{ii}-1$。

Quass(1976)提出了求 d_i^* 的另一种简便方法，现介绍如下。

由 $d_i=l_{ii}$，可得 $d_i^*=1/l_{ii}^2$，所以只要求出 l_{ii}，就可得到 d_i^*。将 $f_i=a_{ii}-1$ 代入(2.6)式可得

$$l_{ii}=\frac{1}{\sqrt{d_i^*}}=\begin{cases}[1-0.25(a_{ss}+a_{dd})]^{0.5}, & \text{当个体 } i \text{ 的双亲 } s \text{ 和 } d \text{ 已知}\\(1-0.25a_{pp})^{0.5}, & \text{当个体 } i \text{ 的一个亲本 } p \text{ 已知}\\1, & \text{当个体 } i \text{ 的双亲未知}\end{cases} \tag{2.7}$$

由 $\boldsymbol{A}=\boldsymbol{LL}'$，可得 $a_{ii}=\sum_{k=1}^{i}l_{ik}^2$。注意在求 a_{ii} 时要用到 $\boldsymbol{L}$ 的非对角线元素，其计算方法如下。

因为 $\boldsymbol{L}=\boldsymbol{TD}$，而 $\boldsymbol{T}$ 的对角线元素为 1，非对角线元素可由下式计算：

$$\underset{(i>j)}{t_{ij}}=\begin{cases}0.5(t_{sj}+t_{dj}), & \text{当个体 } i \text{ 的双亲 } s \text{ 和 } d \text{ 已知}\\0.5t_{pj}, & \text{当个体 } i \text{ 的一个亲本 } p \text{ 已知}\\0, & \text{当个体 } i \text{ 的双亲未知}\end{cases} \tag{2.8}$$

所以

$$\underset{(i>j)}{l_{ij}}=t_{ij}d_j=\begin{cases}0.5(l_{sj}+l_{dj}), & \text{当个体 } i \text{ 的双亲 } s \text{ 和 } d \text{ 已知}\\0.5l_{pj}, & \text{当个体 } i \text{ 的一个亲本 } p \text{ 已知}\\0, & \text{当个体 } i \text{ 的双亲未知}\end{cases} \tag{2.9}$$

综上所述，可按下列步骤计算 $\boldsymbol{A}^{-1}$：

(1)按计算 $\boldsymbol{A}$ 阵时的要求将系谱中的所有个体列表；

(2)将 $\boldsymbol{A}^{-1}$ 中的所有元素置为零；

(3)设置两个阶数为 n 的零向量 $\boldsymbol{v}$ 和 $\boldsymbol{a}$，$\boldsymbol{v}$ 用于存放 l_{ii}，并临时存放 l_{ki}($k=i+1,\cdots,n$)，$\boldsymbol{a}$ 用于存放 a_{ii}，n 为个体总数；

(4)对于 $i=1,\cdots,n$，计算

①$v_i=l_{ii}$[用(2.7)式计算]

②$a_i=a_i+v_i^2$

③对于 $k=i+1,\cdots,n$，计算

$$v_k=\begin{cases}0.5(v_{s_k}+v_{d_k}), & \text{当个体 } k \text{ 的双亲 } s_k \text{ 和 } d_k \text{ 均已知且排在 } i \text{ 之后或等于 } i\\0.5v_{p_k}, & \text{当个体 } k \text{ 的一个亲本 } p_k \text{ 已知且排在 } i \text{ 之后或等于 } i\\0, & \text{其他}\end{cases}$$

$$a_k=a_k+v_k^2$$

④$d_i^*=\frac{1}{v_i^2}$

⑤将下列数值加到 $\boldsymbol{A}^{-1}$ 中：

如 i 的双亲 s 和 d 已知：

$$d_i^* \rightarrow (i,i)$$
$$-0.5d_i^* \rightarrow (i,s),(s,i),(i,d),(d,i)$$
$$0.25d_i^* \rightarrow (s,s),(d,d),(s,d),(d,s)$$

如 i 的一个亲本 p 已知：

$$d_i^* \rightarrow (i,i)$$
$$-0.5d_i^* \rightarrow (i,p),(p,i)$$
$$0.25d_i^* \rightarrow (p,p)$$

如 i 的双亲均未知：

$$d_i^* \rightarrow (i,i)$$

对于例 2.1，按上述步骤，可得

$i=1$：

$$v_1=l_{11}=1,\ a_1=a_1+v_1^2=1$$
$$v_2=l_{21}=0,\ a_2=a_2+v_2^2=0$$
$$v_3=l_{31}=0.5v_1=0.5,\ a_3=a_3+v_3^2=0.25$$
$$v_4=l_{41}=0.5(v_1+v_2)=0.5,\ a_4=a_4+v_4^2=0.25$$
$$v_5=l_{51}=0.5(v_3+v_4)=0.5,\ a_5=a_5+v_5^2=0.25$$
$$v_6=l_{61}=0.5(v_1+v_4)=0.75,\ a_6=a_6+v_6^2=0.5625$$
$$v_7=l_{71}=0.5(v_5+v_6)=0.625,\ a_7=a_7+v_7^2=0.390625$$
$$d_1^*=\frac{1}{v_1^2}=1$$
$$\boldsymbol{A}^{-1}(1,1)=\boldsymbol{A}^{-1}(1,1)+d_1^*=0+1=1$$

$i=2$：

$$v_2=l_{22}=1,\ a_2=a_2+v_2^2=0+1=1$$
$$v_3=l_{32}=0,\ a_3=a_3+v_3^2=0.25+0=0.25$$
$$v_4=l_{42}=0.5v_2=0.5,\ a_4=a_4+v_4^2=0.25+0.5^2=0.5$$
$$v_5=l_{52}=0.5(v_3+v_4)=0.25,\ a_5=a_5+v_5^2=0.25+0.25^2=0.3125$$
$$v_6=l_{62}=0.5v_4=0.25,\ a_6=a_6+v_6^2=0.5625+0.25^2=0.625$$
$$v_7=l_{72}=0.5(v_5+v_6)=0.25,\ a_7=a_7+v_7^2=0.390625+0.25^2$$
$$=0.453125$$
$$d_2^*=\frac{1}{v_2^2}=1$$
$$\boldsymbol{A}^{-1}(2,2)=\boldsymbol{A}^{-1}(2,2)+d_2^*=0+1=1$$

$i=3$：

$$v_3=l_{33}=\sqrt{1-0.25a_3}=\sqrt{0.75},a_3=a_3+v_3^2=0.25+0.75=1$$

$$v_4=l_{43}=0,a_4=a_4+v_4^2=0.5+0=0.5$$

$$v_5=l_{53}=0.5(v_3+v_4)=0.5(\sqrt{0.75}+0)=0.5\sqrt{0.75}$$

$$a_5=a_5+v_5^2=0.3125+(0.5\sqrt{0.75})^2=0.5$$

$$v_6=l_{63}=0.5v_4=0.5\times0=0,a_6=a_6+v_6^2=0.625+0=0.625$$

$$v_7=l_{73}=0.5(v_5+v_6)=0.25\sqrt{0.75}$$

$$a_7=a_7+v_7^2=0.453125+(0.25\sqrt{0.75})^2=0.5$$

$$d_3^*=\frac{1}{v_3^2}=\frac{4}{3}$$

$$\boldsymbol{A}^{-1}(3,3)=\boldsymbol{A}^{-1}(3,3)+d_3^*=0+\frac{4}{3}=\frac{4}{3}$$

$$\boldsymbol{A}^{-1}(3,1)=\boldsymbol{A}^{-1}(1,3)=\boldsymbol{A}^{-1}(3,1)-0.5d_3^*=0-\frac{2}{3}=-\frac{2}{3}$$

$$\boldsymbol{A}^{-1}(1,1)=\boldsymbol{A}^{-1}(1,1)+0.25d_3^*=1+\frac{1}{3}=\frac{4}{3}$$

$i=4$：

$$v_4=l_{44}=\sqrt{1-0.25(a_4+a_7)}=\sqrt{0.5},a_4=a_4+v_4^2=0.5+0.5=1$$

$$v_5=l_{54}=0.5v_4=0.5\sqrt{0.5},a_5=a_5+v_5^2=0.5+(0.5\sqrt{0.5})^2=0.625$$

$$v_6=l_{64}=0.5v_4=0.5\sqrt{0.5},a_6=a_6+v_6^2=0.625+(0.5\sqrt{0.5})^2=0.75$$

$$v_7=l_{74}=0.5(v_5+v_6)=0.5\sqrt{0.5}$$

$$a_7=a_7+v_7^2=0.5+(0.5\sqrt{0.5})^2=0.625$$

$$d_4^*=\frac{1}{v_4^2}=2$$

$$\boldsymbol{A}^{-1}(4,4)=\boldsymbol{A}^{-1}(4,4)+d_4^*=0+2=2$$

$$\boldsymbol{A}^{-1}(4,1)=\boldsymbol{A}^{-1}(1,4)=\boldsymbol{A}^{-1}(4,1)-0.5d_4^*=0-1=-1$$

$$\boldsymbol{A}^{-1}(4,2)=\boldsymbol{A}^{-1}(2,4)=\boldsymbol{A}^{-1}(4,2)-0.5d_4^*=0-1=-1$$

$$\boldsymbol{A}^{-1}(1,1)=\boldsymbol{A}^{-1}(1,1)+0.25d_4^*=\frac{4}{3}+\frac{1}{2}=\frac{11}{6}$$

$$\boldsymbol{A}^{-1}(2,2)=\boldsymbol{A}^{-1}(2,2)+0.25d_4^*=1+\frac{1}{2}=\frac{3}{2}$$

$$\boldsymbol{A}^{-1}(2,1)=\boldsymbol{A}^{-1}(1,2)=\boldsymbol{A}^{-1}(2,1)+0.25d_4^*=0+\frac{1}{2}=\frac{1}{2}$$

如此计算下去直至 $i=7$，最后可得

$$\boldsymbol{v}' = (l_{11}\ l_{22}\cdots l_{77}) = \left(1 \quad 1 \quad \sqrt{0.75} \quad \sqrt{0.5} \quad \sqrt{0.5} \quad \sqrt{0.5} \quad \sqrt{0.40625}\right)$$

$$d_1^* = 1, d_2^* = 1, d_3^* = 4/3, d_4^* = 2, d_5^* = 2, d_6^* = 2, d_7^* = 32/13$$

[利用 $\boldsymbol{A}$ 阵中的对角线元素，由(2.6)式，也可得到与之完全相同的结果。]

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \frac{7}{3} & \frac{1}{2} & -\frac{2}{3} & -\frac{1}{2} & 0 & -1 & 0 \\ & \frac{3}{2} & 0 & -1 & 0 & 0 & 0 \\ & & \frac{11}{6} & \frac{1}{2} & -1 & 0 & 0 \\ & & & 3 & -1 & -1 & 0 \\ & \text{对} & \text{称} & & \frac{34}{13} & \frac{8}{13} & -\frac{16}{13} \\ & & & & & \frac{34}{13} & -\frac{16}{13} \\ & & & & & & \frac{32}{13} \end{bmatrix}$$

当群体为非近交群体(所有个体的近交系数均为零)，由(2.6)式可得

$$d_i^* = \begin{cases} 2, & \text{当个体 } i \text{ 的双亲已知} \\ 4/3, & \text{当个体 } i \text{ 的一个亲本已知} \\ 1, & \text{当个体 } i \text{ 的双亲未知} \end{cases} \tag{2.10}$$

2.2.2 公畜模型下 $\boldsymbol{A}^{-1}$ 的计算

对于公畜模型，如果根据公畜的父亲和外祖父的信息构造 $\boldsymbol{A}^{-1}$，Henderson (1975b)也提出了相应的简捷方法。此时，与(2.7)式相应的计算 l_{ii} 的公式为

$$l_{ii} = \begin{cases} (1-0.25a_{ss}-0.0625a_{gg})^{0.5}, & \text{当个体 } i \text{ 的父亲 } s \text{ 和外祖父 } g \text{ 均已知} \\ (1-0.25a_{ss})^{0.5}, & \text{当个体 } i \text{ 的父亲 } s \text{ 已知} \\ (1-0.0625a_{gg})^{0.5}, & \text{当个体 } i \text{ 的外祖父 } g \text{ 已知} \\ 1, & \text{当个体 } i \text{ 的父亲和外祖父均未知} \end{cases} \tag{2.11}$$

具体的计算步骤为：

(1)将系谱中的所有公畜按计算 $\boldsymbol{A}$ 阵的方式列表。

(2)将 $\boldsymbol{A}^{-1}$ 中的所有元素置为 0。

(3)建立阶数为 n 的零向量 $\boldsymbol{v}$ 和 $\boldsymbol{a}$，n 为公畜总数。

(4)对于 $i=1,\cdots,n$，计算

① $v_i = l_{ii}$ [用(2.11)式计算]

② $a_i = a_i + v_i^2$

③对于 $k=i+1,\cdots,n$，计算

$$v_k=\begin{cases}0.5v_{s_k}+0.25v_{g_k} & \text{当公畜 } k \text{ 的父亲 } s_k \text{ 和外祖父 } g_k \text{ 均排在 } i \text{ 之后或等于 } i\\ 0.5v_{s_k} & \text{当公畜 } k \text{ 的父亲 } s_k \text{ 排在 } i \text{ 之后或等于 } i\\ 0.25v_{g_k} & \text{当公畜 } k \text{ 的外祖父 } g_k \text{ 排在 } i \text{ 之后或等于 } i\\ 0 & \text{当公畜 } k \text{ 的父亲 } s_k \text{ 和外祖父 } g_k \text{ 均排在 } i \text{ 之前}\end{cases}$$

$$a_k = a_k + v_k^2$$

④ $d_i^* = \frac{1}{l_{ii}^2}$

⑤将下列数值加到 $\boldsymbol{A}^{-1}$ 中：

如果公畜 i 的父亲 s 和外祖父 g 已知：

$$d_i^* \rightarrow (i,i)$$
$$-0.5d_i^* \rightarrow (i,s),(s,i)$$
$$-0.25d_i^* \rightarrow (i,g),(g,i)$$
$$0.25d_i^* \rightarrow (s,s)$$
$$0.0625d_i^* \rightarrow (g,g)$$
$$0.125d_i^* \rightarrow (s,g),(g,s)$$

如果公畜 i 的父亲 s 已知，而外祖父 g 未知：

$$d_i^* \rightarrow (i,i)$$
$$-0.5d_i^* \rightarrow (i,s),(s,i)$$
$$0.25d_i^* \rightarrow (s,s)$$

如果公畜 i 的外祖父 g 已知，而父亲 s 未知：

$$d_i^* \rightarrow (i,i)$$
$$-0.25d_i^* \rightarrow (i,g),(g,i)$$
$$0.0625d_i^* \rightarrow (g,g)$$

如果公畜 i 的父亲 s 和外祖父 g 均未知：

$$d_i^* \rightarrow (i,i)$$

当群体为非近交群体时：

$$d_i^* = \begin{cases}\frac{16}{11}, & \text{当公畜 } i \text{ 的父亲和外祖父均已知}\\ \frac{4}{3}, & \text{当公畜 } i \text{ 的父亲已知，外祖父未知}\\ \frac{16}{15}, & \text{当公畜 } i \text{ 的外祖父已知，父亲未知}\\ 1, & \text{当公畜 } i \text{ 的父亲和外祖父均未知}\end{cases} \tag{2.12}$$

对于例 2.2，有

$$\boldsymbol{v}' = \begin{bmatrix} l_{11} & l_{22} & l_{33} & l_{44} & l_{55} & l_{66} & l_{77} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.866025 & 0.829156 & 0.829156 & 0.805256 & 0.805256 & 0.808282 \end{bmatrix}$$

$$\boldsymbol{A}^{-1} = \begin{bmatrix} 2.060607 & -0.303031 & -0.727273 & -0.727273 & 0.0 & 0.0 & 0.0 \\ & 1.515153 & -0.363637 & -0.363637 & 0.0 & 0.0 & 0.0 \\ & & 2.225631 & 0.385542 & -0.771085 & -0.771085 & 0.0 \\ & & & 1.647317 & -0.385542 & -0.385542 & 0.0 \\ & & & & 1.924830 & 0.191330 & -0.765322 \\ & & 对称 & & & 1.637835 & -0.382661 \\ & & & & & & 1.530644 \end{bmatrix}$$

2.3 近交系数的计算

近交系数是家畜育种中常用的重要指标，计算近交系数也就成了育种工作者的一项经常性工作。近交系数的经典计算公式是

$$f_X = \sum \left[\frac{1}{2} \right]^{n_1+n_2+1} (1 + f_A) \tag{2.13}$$

其中，f_X 是个体 X 的近交系数；n_1 和 n_2 分别为在连接个体 X 的双亲的通径中父亲和母亲到它们的共同祖先 A 的世代数；f_A 为 A 的近交系数；$\sum$ 表示当个体 X 的父亲和母亲有多个通径相连接时，要对所有通径求和。与用(2.1)式计算个体间的加性遗传相关一样，用(2.13)式计算近交系数也需画出完整的系谱，因而也是不实用的。

由 2.1 节的介绍可知，只要计算出了 $\boldsymbol{A}$ 矩阵，近交系数就可由 $\boldsymbol{A}$ 阵的对角线元素由(2.2)式计算：

$$f_i = a_{ii} - 1$$

虽然这个方法非常简便，但其缺点是在计算一个个体的近交系数时，仍需将完整的 $\boldsymbol{A}$ 阵计算出来，而我们需要的只是其对角线元素。由上节介绍的 Quaas(1976)的算法，可以通过逐列计算矩阵 $\boldsymbol{L}$ 的元素而得到 $\boldsymbol{A}$ 的对角线元素 $\left[a_{ii} = \sum_{k=1}^{i} l_{ik}^2 \right]$。这个算法用来计算近交系数的一个缺点是，当在系谱中有新的个体加入时，如果没有储存与原有个体相关的 $\boldsymbol{L}$ 矩阵，就不能根据原有个体的近交系数计算新个体的近交系数，而必须将整个 $\boldsymbol{L}$ 矩阵重新计算一遍。Meuwissen 与 Luo(1992)和 Quaas(1995)分别提出了一个更有效的算法，他们利用 $\boldsymbol{L}$ 的另一个性质，即 $\boldsymbol{L}$ 的每一行可以独立地进行计算，这样就可以为每一新增加的个体计算近交系数而无需储存 $\boldsymbol{L}$ 矩阵，现介绍如下。

2.3.1 Meuwissen 和 Luo 的算法

由 2.2 节的介绍，可知 $\boldsymbol{A}$ 阵可分解为

$$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{L}'=\boldsymbol{T}\boldsymbol{D}^{2}\boldsymbol{T}' \tag{2.14}$$

设 w_i 是 $\boldsymbol{D}^2$ 中的对角线元素，由(2.6)式可得

$$w_i=\begin{cases}0.5-0.25(f_s+f_d), & \text{当个体 } i \text{ 的双亲 } s \text{ 和 } d \text{ 已知}\\ 0.75-0.25f_p, & \text{当个体 } i \text{ 的一个亲本 } p \text{ 已知}\\ 1, & \text{当个体 } i \text{ 的双亲未知}\end{cases} \tag{2.15}$$

$\boldsymbol{T}$ 中的非对角线元素除了可用(2.8)式计算外，还可用下式计算：

$$t_{\substack{ij\\ i>j}}=\begin{cases}0.5+0.5\sum_k t_{ik}, & \text{如果 } j \text{ 是 } i \text{ 的父亲或母亲}\\ 0.5\sum_k t_{ik}, & \text{否则}\end{cases} \tag{2.16}$$

其中，k 代表既是个体 i 的祖先，又是 j 的直接后代(即 j 是 k 的父亲或母亲)的个体。利用这个公式，可以独立地计算 $\boldsymbol{T}$ 中的每一行元素。在计算第 i 行的元素时，可从 $j=i-1,i-2,\cdots,1$，即从 i 的最近的祖先到最远的祖先，依次计算 t_{ij}。

例如，对于例 2.1，其 $\boldsymbol{T}$ 矩阵中第 7 行上的元素为

$t_{76}=0.5$，$t_{75}=0.5$，$t_{74}=0.5(t_{75}+t_{76})=0.5$，$t_{73}=0.5t_{75}=0.25$，

$t_{72}=0.5t_{74}=0.25$，$t_{71}=0.5t_{73}+0.5t_{74}+0.5t_{76}=0.625$

由(2.14)式，$\boldsymbol{A}$ 中的对角线元素可由下式计算：

$$a_{ii}=\sum_{j=1}^{i}t_{ij}^{2}w_i \tag{2.17}$$

根据这一原理，可按以下步骤计算近交系数。

设要计算个体 X 的近交系数 F_X。

(1)定义数组 AN、CN 和 T，AN 用于存放个体 X 和其所有祖先在系谱中的原始编号，CN 用于存放将 X 和其祖先重新编号后的代码，T 用于存放 t_{ij} 值。

(2)将该个体 X 的原始编号加入到数组 AN 中，将 X 重新编号为 1，将该编号加入到数组 CN 中。令 $A_X=0$，T[1]=1，$k=1$。

(3)执行以下计算，直至 AN 为空数组：

①令 j=CN 中的最小值(AN 中的最年轻的个体)，$d_j=1$。

②如果 j 的父亲已知，设其原始编号为 JS，

$$d_j=d_j-0.25-0.25f_{JS}\text{(其中 } f_{JS} \text{ 是 } JS \text{ 的近交系数)}$$

如果 JS 在 AN 中不存在，将 JS 加入到 AN 中，为其重新编号为 $k=k+1$，将 k 加入到 CN 中。

$$\mathrm{T}[k]=\mathrm{T}[k]+0.5\times\mathrm{T}[j]$$

如果 JS 在 AN 中已存在，设其在 CN 中对应的新编号为 c

$$\mathrm{T}[c]=\mathrm{T}[c]+0.5\times\mathrm{T}[j]$$

③如果 j 的母亲已知，设其原始编号为 JD，

$$d_j=d_j-0.25-0.25f_{JD}\text{(其中 } f_{JD} \text{ 是 } JD \text{ 的近交系数)}$$

如果 JD 在 AN 中不存在，将 JD 加入到 AN 中，为其重新编号为 $k=k+1$，将 k 加入到 CN 中。

$$\mathrm{T}[k]=\mathrm{T}[k]+0.5\times\mathrm{T}[j]$$

如果 JD 在 AN 中已存在，设其在 CN 中对应的新编号为 c

$$\mathrm{T}[c]=\mathrm{T}[c]+0.5\times\mathrm{T}[j]$$

④ $A_x=A_x+(\mathrm{T}[j])^2\times d_j$。

⑤从 CN 中删去个体 j，从 AN 中删去与 j 对应的个体，返回①。

(4)计算 $F_x=A_x-1$。可以看出，利用这个算法，如果对群体中原有的所有个体均计算并保存了近交系数的值，则对于群体中新增加的个体，只要根据其祖先的近交系数就可算出该个体的近交系数，而无需将所有个体重算一遍，而且，也不需事先建立如计算 $\boldsymbol{A}$ 阵所需的系谱文件。

例 2.3 用 Meuwissen 和 Luo 的算法计算例 2.1 中的第 5 个个体的近交系数，假设其祖先的近交系数均已计算(在这里，其所有祖先的近交系数都为 0)。

将 5 加入 AN 中，得 AN=[5]，将它重新编号为 1，加入 CN，得 CN=[1]

$$k=1,\quad \mathrm{T}[1]=1,\quad A_5=0$$

进入以下循环：

$j=\min\{\mathrm{CN}\}=1$，$d_j=d_1=1$

$JS=3$

$d_1=d_1-0.25-0.25f_{JS}=0.75$

AN=[5 3]，将 3 重新编号为 $k=k+1=2$，CN[1 2]，

$\mathrm{T}[k]=\mathrm{T}[2]=\mathrm{T}[2]+0.5\mathrm{T}[1]=0.5$

$JD=4$

$d_1=d_1-0.25-0.25f_{JD}=0.5$

AN=[5 3 4]，将 4 重新编号为 $k=k+1=3$，CN[1 2 3]，

$\mathrm{T}[k]=\mathrm{T}[3]=\mathrm{T}[3]+0.5\mathrm{T}[1]=0.5$

$A_5=A_5+(\mathrm{T}[1])^2\times d_1=0+1^2\times0.5=0.5$

AN=[3 4]，CN=[2 3]

$j=\min\{\mathrm{CN}\}=2$，$d_j=d_2=1$

$JS=1$(个体 3 的父亲为 1)

$d_2=d_2-0.25-0.25f_{JS}=0.75$

AN=[3 4 1]，将 1 重新编号为 $k=k+1=4$，CN[2 3 4]

$\mathrm{T}[k]=\mathrm{T}[4]=\mathrm{T}[4]+0.5\mathrm{T}[2]=0.25$

$JD=0$

$A_5=A_5+(\mathrm{T}[2])^2\times d_2=0.5+0.5^2\times0.75=0.6875$

AN=[4 1]，CN=[3 4]

$j=\min\{\mathrm{CN}\}=3$，$d_j=d_3=1$