

第一天回顾

- ▶ 路线图和日程安排
- ▶ Julia 虚拟环境和 Jupyter notebook 简介
- ▶ 育种流程简介
- ▶ 统计模型和育种值
- ▶ 最小二乘原理
- ▶ 线性模型最小二乘估计
 - ▶ 优化估计（模型、目标、优化）
 - ▶ 基因组选择的最小二乘估计
 - ▶ 练习

第二天 通过系谱和表型记录估计育种值

于希江

挪威生命科学大学畜牧与水产系

二〇一九·十月
青岛



目录

数量性状基因向后代的传递

多元正态分布和一些情况下遗传力的估计

方差、协方差

多元正态分布

一些情况下遗传力的估计

根据系谱和个体表现型估计育种值

引子：一个简单的估计育种值的模型

混合模型方程组

GLS, BLUE, BLUP

数量性状座位 QTL

- ▶ 大多数性状的 QTL 个数 (N_q) 通常以千计
- ▶ 育种中, 人们通常只对加性效应感兴趣
 - ▶ 因为显性和上位效应会在形成配子的过程中分开, 在后代中重新组合
 - ▶ 而加性效应是每个等位基因的效应
- ▶ 个体 i 的育种值 $a_i = \sum_{j=1}^{N_q} x_j b_j$
- ▶ 一个后代 o 从个体 i 的 $2N_q$ 个等位基因中抽取 N_q 个
 - ▶ 于是后代育种值中来自 i 的部分的期望是 $\frac{a_i}{2}$

孟德尔抽样误差

- ▶ 由于 QTL 众多, 且大小不一, 后代 o 很难从个体 i 正好是 $\frac{a_i}{2}$
- ▶ 这个由于从众多 QTL 的等位基因抽样所造成的误差 m 叫做孟德尔抽样误差
- ▶ 这样若 $y_i = \mu + a_i + e_i$, 则个体 i, j 的后代 o 的表型模型可以写作:
 - ▶ $y_o = \mu + \frac{a_i}{2} + m_i + \frac{a_j}{2} + m_j + e_o$
- ▶ $m \sim (0, \sigma_m^2)$
 - ▶ $\sigma_{m_i}^2 \approx \sigma_{m_j}^2 \approx \frac{\sigma_a^2}{4}$

目录

数量性状基因向后代的传递

多元正态分布和一些情况下遗传力的估计

方差、协方差

多元正态分布

一些情况下遗传力的估计

根据系谱和个体表现型估计育种值

引子：一个简单的估计育种值的模型

混合模型方程组

GLS, BLUE, BLUP

目录

数量性状基因向后代的传递

多元正态分布和一些情况下遗传力的估计

方差、协方差

多元正态分布

一些情况下遗传力的估计

根据系谱和个体表现型估计育种值

引子：一个简单的估计育种值的模型

混合模型方程组

GLS, BLUE, BLUP

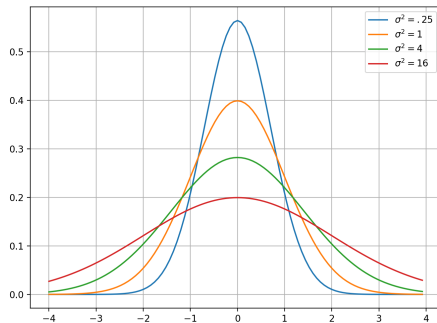
方差、协方差

方差 (σ^2 , 或者 Var , V)

► 方差是对一个随机变量离散 (spread) 程度的度量。

► $\sigma^2 = E(X - \mu)^2$

► 若由样本估计: $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$



方差、协方差

协方差

- ▶ 对两个随机变量协同变化程度的度量

- ▶ $E(x_1 - \mu_1)(x_2 - \mu_2)$

- ▶ 若由样本估计:

$$\hat{\sigma}_{12} = \frac{\sum_{i=1}^n (x_{1i} - \mu_1)(x_{2i} - \mu_2)}{n - 1}$$

方差、协方差

协方差

► 对两个随机变量协同变化程度的度量

► $E(x_1 - \mu_1)(x_2 - \mu_2)$

► 若由样本估计:

$$\hat{\sigma}_{12} = \frac{\sum_{i=1}^n (x_{1i} - \mu_1)(x_{2i} - \mu_2)}{n - 1}$$

方差协方差矩阵

由 p 个随机数的组成的向量

$\mathbf{x} = [x_1, x_2, \dots, x_p]'$, 有:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}$$

目录

数量性状基因向后代的传递

多元正态分布和一些情况下遗传力的估计

方差、协方差

多元正态分布

一些情况下遗传力的估计

根据系谱和个体表现型估计育种值

引子：一个简单的估计育种值的模型

混合模型方程组

GLS, BLUE, BLUP

多元正态分布

一元正态分布

$$n(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ 期望

σ 方差

多元正态分布

一元正态分布

$$n(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ 期望

σ 方差

多元正态分布

► 是一元正态分布的推广

$$\text{MVN}(\mathbf{x}; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)' \Sigma^{-1} (\mathbf{x}-\mu)}$$

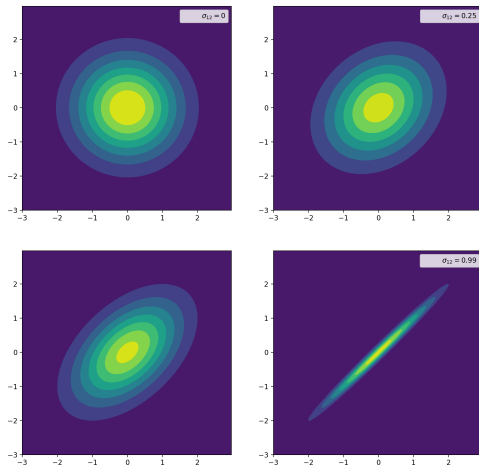
► 也记作: $\mathbf{x} \sim \text{MVN}(\mu, \Sigma)$

$$\mathbb{E}(\mathbf{x}) = \mu$$

$$\text{Cov}(\mathbf{x}) = \Sigma$$

从两个随机变量来理解方差协方差矩阵

从两个随机变量来理解方差协方差矩阵



- 右边四图中，每个变量的方差都是 1。

目录

数量性状基因向后代的传递

多元正态分布和一些情况下遗传力的估计

方差、协方差

多元正态分布

一些情况下遗传力的估计

根据系谱和个体表现型估计育种值

引子：一个简单的估计育种值的模型

混合模型方程组

GLS, BLUE, BLUP

半同胞对

- ▶ 设半同胞对 $[x_1, x_2]$ 同父异母，则：

- ▶ $x_1 = \mu + \frac{a_S}{2} + e$

- ▶ $x_2 = \mu + \frac{a_S}{2} + e$

- ▶ 其中 e 包括其它遗传效应和环境效应。

- ▶ $\text{Cov}(x_1, x_2) = \frac{\sigma_{a_S}^2}{4} = \frac{\sigma_a^2}{4}$

- ▶ 依模型： $x = \mu + a + e$ ，每个半同胞变量的方差

- ▶ $\text{Var}(x_1) = \text{Var}(x_2) = \sigma_p = \sigma_a^2 + \sigma_e^2$

半同胞对

► 设半同胞对 $[x_1, x_2]$ 同父异母，则：

► $x_1 = \mu + \frac{a_S}{2} + e$

► $x_2 = \mu + \frac{a_S}{2} + e$

► 其中 e 包括其它遗传效应和环境效应。

► $\text{Cov}(x_1, x_2) = \frac{\sigma_{a_S}^2}{4} = \frac{\sigma_a^2}{4}$

► 依模型： $x = \mu + a + e$ ，每个半同胞变量的方差

► $\text{Var}(x_1) = \text{Var}(x_2) = \sigma_p = \sigma_a^2 + \sigma_e^2$

► 于是：

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \text{MVN} \left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma_p^2 & \frac{\sigma_a^2}{4} \\ \frac{\sigma_a^2}{4} & \sigma_p^2 \end{bmatrix} \right)$$

► 注意，此处可见遗传力：

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2} = \frac{4 \cdot \sigma_{12}}{(\sigma_1^2 + \sigma_2^2)/2}$$

半同胞对模拟

真值

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{bmatrix}$$

$$h^2 = 0.4$$

见 [day-2-mvn.ipynb](#).

半同胞对模拟

真值

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{bmatrix}$$

$$h^2 = 0.4$$

见 [day-2-mvn.ipynb](#).

模拟 10 对半同胞

x_1	1.2	1.8	0.4	0.2	-1.2	0.1	-2.2	-1.6	-0.0	-0.4
x_2	-1.2	0.0	0.2	-1.7	1.1	1.9	-0.3	-2.3	0.7	1.0

$$\hat{\mu} = \begin{bmatrix} -0.2 \\ -0.1 \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} 1.5 & 0.1 \\ 0.1 & 1.8 \end{bmatrix}$$

$$\hat{h}^2 = 0.2$$

类似的

单亲与一子的对

► 模型

$$\text{► } y_i = \mu + a_i + e_i$$

$$\text{► } y_o = \mu + \frac{a_i}{2} + e_o$$

► 方差协方差矩阵

$$\text{► } \Sigma = \begin{bmatrix} \sigma_p^2 & \frac{\sigma_a^2}{2} \\ \frac{\sigma_a^2}{2} & \sigma_p^2 \end{bmatrix}$$

$$\text{► } h^2 = 2 \cdot \sigma_{12} / \sigma_p^2$$

双亲平均与一子

► 模型

$$\text{► } y_i = \mu + a_i + e_i$$

$$\text{► } y_j = \mu + a_j + e_j$$

$$\text{► } (y_i + y_j)/2 = \mu + (a_i + a_j)/2 + \bar{e}$$

$$\text{► } y_o = \mu + \frac{a_i + a_j}{2} + e_o$$

► 方差协方差矩阵

$$\text{► } \Sigma = \begin{bmatrix} \sigma_p^2/2 & \sigma_a^2/2 \\ \sigma_a^2/2 & \sigma_p^2 \end{bmatrix}$$

$$\text{► } h^2 = \sigma_{12} / \sigma_{\bar{p}}^2$$

目录

数量性状基因向后代的传递

多元正态分布和一些情况下遗传力的估计

方差、协方差

多元正态分布

一些情况下遗传力的估计

根据系谱和个体表现型估计育种值

引子：一个简单的估计育种值的模型

混合模型方程组

GLS, BLUE, BLUP

目录

数量性状基因向后代的传递

多元正态分布和一些情况下遗传力的估计

方差、协方差

多元正态分布

一些情况下遗传力的估计

根据系谱和个体表现型估计育种值

引子：一个简单的估计育种值的模型

混合模型方程组

GLS, BLUE, BLUP

简化公牛模型/Sire model

- ▶ 常见于奶牛育种
- ▶ p 头公牛
- ▶ 每头公牛各有 n_i 个女儿
- ▶ 这样，每个女儿的产奶量模型为

$$y_{ij} = \mu + u_i + e_{ij}$$

目录

数量性状基因向后代的传递

多元正态分布和一些情况下遗传力的估计

方差、协方差

多元正态分布

一些情况下遗传力的估计

根据系谱和个体表现型估计育种值

引子：一个简单的估计育种值的模型

混合模型方程组

GLS, BLUE, BLUP

效应/因子

几个概念

线性混合模型 除随机残差效应，同时包含固定和随机效应的线性模型

效应 出现在模型中（右边）的那些量

固定效应 效应的所有水平都在样本中出现

随机效应 其水平来自更大的抽样空间

效应/因子

几个概念

线性混合模型 除随机残差效应，同时包含固定和随机效应的线性模型

效应 出现在模型中（右边）的那些量

固定效应 效应的所有水平都在样本中出现

随机效应 其水平来自更大的抽样空间

$$y = Xb + Zu + e$$

其中：

y $n \times 1$ 维观测值向量

X $n \times s$ 维结构矩阵，或者指示矩阵

b $s \times 1$ 个固定效应

Z $n \times q$ 维指示矩阵

u $q \times 1$ 个随机效应

e $n \times 1$ 个随机残差效应

线性混合模型观测值举例

- 有来自 2 个奶牛场的 2 头公牛的 5 头女儿的记录如下：

个体号	场	公牛	表型值
1	1	1	11
2	1	1	15
3	2	1	10
4	1	2	19
5	2	2	25

线性混合模型观测值举例

► 公牛模型

$$\begin{array}{c} y \\ \left[\begin{array}{c} y_{111} \\ y_{112} \\ y_{213} \\ y_{121} \\ y_{222} \end{array} \right] \end{array} = \begin{array}{c} y \\ \left[\begin{array}{c} 11 \\ 15 \\ 10 \\ 19 \\ 25 \end{array} \right] \end{array} = \begin{array}{c} Xb \\ \left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{array} \right] \left[\begin{array}{c} b_1 \\ b_2 \end{array} \right] \end{array} + \begin{array}{c} Zu \\ \left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{array} \right] \left[\begin{array}{c} u_1 \\ u_2 \end{array} \right] \end{array} + \begin{array}{c} e \\ \left[\begin{array}{c} e_{111} \\ e_{112} \\ e_{213} \\ e_{121} \\ e_{222} \end{array} \right] \end{array}$$

目录

数量性状基因向后代的传递

多元正态分布和一些情况下遗传力的估计

方差、协方差

多元正态分布

一些情况下遗传力的估计

根据系谱和个体表现型估计育种值

引子：一个简单的估计育种值的模型

混合模型方程组

GLS, BLUE, BLUP

广义最小二乘 GLS

普通最小二乘

- ▶ $e \stackrel{\text{iid}}{\sim} N(0, \sigma_e^2)$
- ▶ 或者 $\epsilon \sim \text{MVN}(0, I\sigma_e^2)$
- ▶ $\Rightarrow \hat{b} = (X'X)^{-1}Xy$

广义最小二乘

- ▶ $\epsilon \sim \text{MVN}(0, \Sigma)$
- ▶ 用 Σ^{-1} 为观测值加权
 - ▶ 方差越大，权重越低

广义最小二乘 GLS

普通最小二乘

- ▶ $e \stackrel{\text{iid}}{\sim} N(0, \sigma_e^2)$
- ▶ 或者 $\epsilon \sim \text{MVN}(0, I\sigma_e^2)$
- ▶ $\Rightarrow \hat{b} = (X'X)^{-1}X'y$

广义最小二乘

- ▶ $\epsilon \sim \text{MVN}(0, \Sigma)$
- ▶ 用 Σ^{-1} 为观测值加权
 - ▶ 方差越大，权重越低

普通最小二乘的优化函数

$$(y - X\beta)'(y - X\beta)$$

广义最小二乘的优化函数

$$(y - X\beta)'\Sigma^{-1}(y - X\beta)$$

广义最小二乘 GLS

普通最小二乘

- ▶ $e \stackrel{\text{iid}}{\sim} N(0, \sigma_e^2)$
- ▶ 或者 $\epsilon \sim \text{MVN}(0, I\sigma_e^2)$
- ▶ $\Rightarrow \hat{b} = (X'X)^{-1}X'y$

广义最小二乘

- ▶ $\epsilon \sim \text{MVN}(0, \Sigma)$
- ▶ 用 Σ^{-1} 为观测值加权
 - ▶ 方差越大，权重越低

普通最小二乘的优化函数

$$(y - X\beta)'(y - X\beta)$$

广义最小二乘的优化函数

$$(y - X\beta)'\Sigma^{-1}(y - X\beta)$$

β 的广义最小二乘估计

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

OLS 和 GLS 背靠背

	普通最小二乘	广义最小二乘
模型	$y = X\beta + \epsilon$	$y = X\beta + \epsilon$
分布	$\epsilon \sim \text{MVN}(0, I\sigma_e^2)$	$\epsilon \sim \text{MVN}(0, \Sigma)$
优化解	$\hat{\beta}^{\text{OLS}} = \arg \min_{\beta} \ y - X\beta\ ^2$	$\hat{\beta}^{\text{GLS}} =$

OLS 和 GLS 背靠背

	普通最小二乘	广义最小二乘	
模型	$y = X\beta + \epsilon$	$y = X\beta + \epsilon$	▶ 设有 Σ 的平方根 $\Sigma^{1/2}$
分布	$\epsilon \sim \text{MVN}(0, I\sigma_e^2)$	$\epsilon \sim \text{MVN}(0, \Sigma)$	▶ $\Sigma = \Sigma^{1/2} \cdot (\Sigma^{1/2})^T$
优化解	$\hat{\beta}^{\text{OLS}} = \arg \min_{\beta} \ y - X\beta\ ^2$	$\hat{\beta}^{\text{GLS}} =$	▶ $\Sigma^{-1} = (\Sigma^{-1/2})^T \cdot \Sigma^{-1/2}$
			▶ 则转化模型 $\Sigma^{-1/2}y = \Sigma^{-1/2}X\beta + \Sigma^{-1/2}\epsilon$ 中
			▶ $\text{Var}(\Sigma^{-1/2}\epsilon) = I\sigma^2$

OLS 和 GLS 背靠背

	普通最小二乘	广义最小二乘	
模型	$y = X\beta + \epsilon$	$y = X\beta + \epsilon$	<ul style="list-style-type: none">▶ 设有 Σ 的平方根 $\Sigma^{1/2}$▶ $\Sigma = \Sigma^{1/2} \cdot (\Sigma^{1/2})^T$▶ $\Sigma^{-1} = (\Sigma^{-1/2})^T \cdot \Sigma^{-1/2}$▶ 则转化模型 $\Sigma^{-1/2}y = \Sigma^{-1/2}X\beta + \Sigma^{-1/2}\epsilon$ 中▶ $\text{Var}(\Sigma^{-1/2}\epsilon) = I\sigma^2$
分布	$\epsilon \sim \text{MVN}(0, I\sigma_e^2)$	$\epsilon \sim \text{MVN}(0, \Sigma)$	
优化解	$\hat{\beta}^{\text{OLS}} = \arg \min_{\beta} \ y - X\beta\ ^2$	$\hat{\beta}^{\text{GLS}} = \arg \min_{\beta} \ \Sigma^{-1/2}(y - X\beta)\ ^2$	

OLS 和 GLS 背靠背

	普通最小二乘	广义最小二乘	
模型	$y = X\beta + \epsilon$	$y = X\beta + \epsilon$	<ul style="list-style-type: none">▶ 设有 Σ 的平方根 $\Sigma^{1/2}$▶ $\Sigma = \Sigma^{1/2} \cdot (\Sigma^{1/2})^T$▶ $\Sigma^{-1} = (\Sigma^{-1/2})^T \cdot \Sigma^{-1/2}$▶ 则转化模型 $\Sigma^{-1/2}y = \Sigma^{-1/2}X\beta + \Sigma^{-1/2}\epsilon$ 中▶ $\text{Var}(\Sigma^{-1/2}\epsilon) = I\sigma^2$
分布	$\epsilon \sim \text{MVN}(0, I\sigma_e^2)$	$\epsilon \sim \text{MVN}(0, \Sigma)$	
优化解	$\hat{\beta}^{\text{OLS}} = \arg \min_{\beta} \ y - X\beta\ ^2$	$\hat{\beta}^{\text{GLS}} = \arg \min_{\beta} \ \Sigma^{-1/2}(y - X\beta)\ ^2$	
结果	$(X'X)^{-1}X'y$	$(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$	

OLS 和 GLS 背靠背

	普通最小二乘	广义最小二乘
模型	$y = X\beta + \epsilon$	$y = X\beta + \epsilon$
分布	$\epsilon \sim \text{MVN}(0, I\sigma_e^2)$	$\epsilon \sim \text{MVN}(0, \Sigma)$
优化解	$\hat{\beta}^{\text{OLS}} = \arg \min_{\beta} \ y - X\beta\ ^2$	$\hat{\beta}^{\text{GLS}} = \arg \min_{\beta} \ \Sigma^{-1/2}(y - X\beta)\ ^2$
结果	$(X'X)^{-1}X'y$	$(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$

- ▶ 设有 Σ 的平方根 $\Sigma^{1/2}$
- ▶ $\Sigma = \Sigma^{1/2} \cdot (\Sigma^{1/2})^T$
- ▶ $\Sigma^{-1} = (\Sigma^{-1/2})^T \cdot \Sigma^{-1/2}$
- ▶ 则转化模型
 $\Sigma^{-1/2}y = \Sigma^{-1/2}X\beta + \Sigma^{-1/2}\epsilon$
中
- ▶ $\text{Var}(\Sigma^{-1/2}\epsilon) = I\sigma^2$
- ▶ $\Sigma^{1/2}$ 通常是 Σ 的克莱斯基 (Cholesky) 分解。

关于 GLS 和 OLS

- ▶ GLS 模型其实也可以用 OLS 估计
- ▶ 但 GLS 估计更好，它考虑了非同质的方差，以及数据内部的相关
- ▶ 这个结果是好的，因为
 - ▶ 估计结果是向模型空间的正交投影
 - ▶ 如果 $\Sigma = I\sigma^2$ ，则估计结果是最大似然估计。不太严格地说，以最大似然估计结果做参数，观测值的概率最大
 - ▶ 根据高斯-马尔科夫 (Gauss-Markov) 定理， $\text{BLUE}(X\beta) = X\hat{\beta}$

BLUE, BLUP

- ▶ 一般称固定效应的 GLS 估计为 BLUE
- ▶ 一般称随机效应的 GLS 估计为 BLUP
- ▶ 混合线性模型: $y = Xb + Zu + e$
 - ▶ 平衡设计没有问题
 - ▶ 需要每个家系在各个群体中有分布
 - ▶ 固定效应须要由遗传差异 (随机效应) 校正
 - ▶ 随机效应须要根据固定效应校正

混合模型中的方差

$$Y = Xb + Zu + e$$

$$\text{Var}(b) = 0$$

$$\text{Var}(u) = E(uu') = G$$

$$\text{Var}(e) = E(ee') = R$$

$$\text{Var}(y) = \text{Var}(Xb + Zu + e) = ZGZ' + R$$

BLUP 的最大似然推导 (可选)

设 y, u 的联合概率密度为:

$$f(y, u) = g(y|u)h(u)$$

$$g(y|u) = g(e)$$

$$g(e) = \frac{1}{\sqrt{(2\pi)^n |\text{Var}(e)|}} e^{-\frac{1}{2} e' \text{Var}(e)^{-1} e}$$

$$h(u) = \frac{1}{\sqrt{(2\pi)^n |\text{Var}(u)|}} e^{-\frac{1}{2} u' \text{Var}(u)^{-1} u}$$

$$f(y, u) = c_1 e^{-\frac{1}{2} e' R^{-1} e} c_2 e^{-\frac{1}{2} u' G^{-1} u}$$

$$= c e^{-\frac{1}{2} e' R^{-1} e} e^{-\frac{1}{2} u' G^{-1} u}$$

$$L = f(y, u)$$

$$\log(L) = \log(c) - \frac{1}{2} e' R^{-1} e - \frac{1}{2} u' G^{-1} u$$

BLUP 的最大似然推导 (续)

$$\log(L) = \log(c) - \frac{1}{2}e'R^{-1}e - \frac{1}{2}u'G^{-1}u$$

$$e = y - Xb - Zu$$

$$\begin{aligned}\log(L) = & \log(c) \\ & - \frac{1}{2}(y - Xb - Zu)'R^{-1}(y - Xb - Zu) \\ & - \frac{1}{2}u'G^{-1}u\end{aligned}$$

BLUP 的最大似然推导 - 结果

令 $\frac{\partial(\log L)}{\partial \mathbf{b}} = \frac{\partial(\log L)}{\partial \mathbf{u}} = 0$, 得混合模型方程组

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

BLUP 的特殊情形

$$R = I\sigma_e^2$$

$$\text{由 } G^{-1} = A^{-1}\sigma_u^{-2} \text{ 和 } \lambda = \frac{\sigma_e^2}{\sigma_u^2}$$

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \qquad \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

注：逆矩阵的计算量 ($O(n^3)$) 比矩阵相乘 ($O(n^2)$) 高很多，而通常 G 的维度远少于观测值的个数。因而其逆矩阵计算比观测值方差的逆矩阵容易得多。

小结

- ▶ 微效等位基因向后代的传递
- ▶ 方差、协方差
- ▶ 多元正态分布
- ▶ 广义最小二乘的原理
- ▶ BLUE 是固定效应的广义最小二乘估计
- ▶ BLUP 和混合线性方程组

小结

- ▶ 微效等位基因向后代的传递
- ▶ 方差、协方差
- ▶ 多元正态分布
- ▶ 广义最小二乘的原理
- ▶ BLUE 是固定效应的广义最小二乘估计
- ▶ BLUP 和混合线性方程组

下一节

- ▶ A 矩阵
- ▶ A 的逆矩阵
- ▶ 几种在育种中常用的模型
- ▶ 模拟和计算